

Annual Technical Report for WIDA Alternate ACCESS English Language Proficiency Test

Series 602, 2023–2024 Administration
Annual Technical Report No. 12

Prepared by WIDA Psychometrics Team
May 2025

Contents

1.	Purpose and Design of WIDA Alternate ACCESS	7
1.1.	<i>Purpose of Alternate ACCESS for ELLs</i>	<i>7</i>
1.2.	<i>The WIDA Standards</i>	<i>8</i>
1.3.	<i>The WIDA Proficiency Levels</i>	<i>8</i>
1.4.	<i>Language Domains</i>	<i>9</i>
1.5.	<i>Grade-Level Clusters.....</i>	<i>10</i>
2.	Test Development	10
2.1.	<i>Item and Task Design.....</i>	<i>11</i>
2.1.1.	Listening Items	11
2.1.2.	Reading Items.....	13
2.1.3.	Writing Tasks	15
2.1.4.	Speaking Tasks.....	17
2.2.	<i>Test Design</i>	<i>20</i>
2.2.1.	Listening.....	20
2.2.2.	Reading	24
2.2.3.	Writing	29
2.2.4.	Speaking.....	33
2.3.	<i>Test Construction.....</i>	<i>39</i>
2.3.1.	Item Development	39
2.3.2.	Field Testing	40
2.3.3.	Item Review and Selection	41
2.4.	<i>Standard Setting.....</i>	<i>42</i>
2.4.1.	Standard Setting Event, Method, and Outcomes	42
3.	Test Administration	45
3.1.	<i>Test Delivery</i>	<i>46</i>
3.2.	<i>Operational Administration.....</i>	<i>46</i>
3.2.1.	Administering the Test.....	46
3.2.2.	Training and Resources for Districts and Schools.....	46
3.2.3.	Test Security.....	46
3.3.	<i>Fairness and Accessibility</i>	<i>47</i>
3.3.1.	Fairness and Accessibility Design.....	47
4.	Scoring	48
4.1.	<i>Listening and Reading</i>	<i>48</i>
4.2.	<i>Writing</i>	<i>49</i>
4.3.	<i>Speaking.....</i>	<i>49</i>
4.4.	<i>Scaling.....</i>	<i>50</i>
5.	Summary of Score Reports.....	51

5.1.	Individual Student Reports	51
5.2.	Other Reports	53
6.	Annual Test Results	53
6.1.	Students Excluded from Analysis	54
6.1.1.	Out-of-Grade Level Test Administration	55
6.2.	Participation by Grade-Level Cluster.....	55
6.2.1.	Participation by Grade-Level Cluster by State.....	56
6.2.2.	Participation by Grade-Level Cluster by Gender	57
6.2.3.	Participation by Grade-Level Cluster by Ethnicity	57
6.3.	Participation by Grade	58
6.3.1.	Participation by Grade by State	58
6.3.2.	Participation by Grade by Gender	60
6.3.3.	Participation by Grade by Ethnicity.....	61
6.4.	Participation by Domain	62
6.4.1.	Participation by Grade-Level Cluster by Domain.....	62
6.4.2.	Participation by Grade by Domain	63
6.5.	Participation by Disability.....	63
6.5.1.	Participation by Primary and Secondary Disability	64
6.5.2.	Participation by Primary Disability by Grade-Level Cluster	65
6.6.	Scale Scores by Domain and Composite.....	66
6.6.1.	Mean Scale Scores by Domain and Composite	67
6.7.	Scale Scores by Grade-Level Cluster.....	68
6.7.1.	Mean Scale Scores by Gender.....	69
6.7.2.	Mean Scale Scores by Ethnicity.....	70
6.8.	Scale Score By Grade.....	77
6.8.1.	Mean Scale Scores by Gender.....	77
6.8.2.	Mean Scale Scores by Ethnicity.....	82
6.9.	Scale Scores by Disability by Grade-Level Cluster	103
6.10.	Correlations among Scale Scores by Grade-Level Cluster.....	108
6.11.	Proficiency Level Results	109
6.11.1.	Proficiency Level by Grade-Level Cluster	109
6.11.2.	Proficiency Level by Grade	113
7.	Analysis of Domain	117
7.1	Tools Used for Analysis.....	117
7.1.1.	Rasch Model for Scoring.....	117
7.1.2.	Sampling.....	119
7.1.3.	Scaling	119
7.2.	Complete Item or Task Analysis Summary	119
7.2.1.	Listening Item Analysis	120
7.2.2.	Reading Item Analysis	125

7.2.3.	Speaking Item Analysis	129
7.2.4.	Writing Item Analysis	133
7.3.	DIF Analysis and Summary	137
7.3.1.	Listening DIF Analysis	139
7.3.2.	Reading DIF Analysis	147
7.3.3.	Speaking DIF Analysis	153
7.3.4.	Writing DIF Analysis	159
7.4.	Raw Score Distribution	164
7.4.1.	Listening	166
7.4.2.	Reading	169
7.4.3.	Speaking	172
7.4.4.	Writing	175
7.5.	Scale Score Distribution	177
7.5.1.	Listening	179
7.5.2.	Reading	182
7.5.3.	Speaking	185
7.5.4.	Writing	188
7.6.	Proficiency Level Distribution	191
7.6.1.	Listening	192
7.6.2.	Reading	196
7.6.3.	Speaking	199
7.6.4.	Writing	203
7.7.	Raw Score to Scale Score Proficiency Level Conversion	207
7.7.1.	Listening	208
7.7.2.	Reading	218
7.7.3.	Speaking	228
7.7.4.	Writing	236
7.8.	Equating Summary	244
7.8.1.	Spring 2024 Post Equating	245
7.8.2.	Equipercentile Linking	245
7.8.3.	Final Calibration	247
7.8.4.	Listening	248
7.8.5.	Reading	252
7.8.6.	Speaking	256
7.8.7.	Writing	260
7.9.	Test Characteristic Curve	264
7.9.1.	Listening Test Characteristic Curves	264
7.9.2.	Reading Test Characteristic Curves	266
7.9.3.	Speaking Test Characteristic Curves	268
7.9.4.	Writing Test Characteristic Curves	270
7.10.	Test Information Curve	272
7.10.1.	Listening Test Information Curves	274
7.10.2.	Reading Test Information Curves	276

7.10.3.	Speaking Test Information Curves.....	278
7.10.4.	Writing Test Information Curves	280
8.	Analysis of Composite Scores	282
8.1.	Scale Score Distribution for Composite Scores.....	282
8.1.1.	Oral Composite	282
8.1.2.	Literacy Composite	286
8.1.3.	Comprehension Composite	289
8.1.4.	Overall Composite	293
8.2.	Proficiency Level Distribution for Composite Scores	296
8.2.1.	Oral Composite	296
8.2.2.	Literacy Composite	300
8.2.3.	Comprehension Composite	303
8.2.4.	Overall Composite	307
9.	Annual Updates of Validity Evidence.....	311
9.1.	Standards.....	312
9.1.1.	Test Content	312
9.1.2.	Response Processes.....	312
9.1.3.	Internal Structure.....	312
9.1.4.	Relations to Other Structure.....	312
9.2.	Annual Validity Studies.....	312
9.2.1.	Confirmatory Factor Analysis	313
9.2.2.	Dimensionality Check.....	313
10.	Reliability	315
10.1.	Reliability of Domain Scores/Composite Scores.....	315
10.1.1.	Reliability of Domain Scores	315
10.1.2.	Reliability of Composite Scores.....	317
10.2.	Conditional Standard Errors of Measurement of the Scale Scores at the Cut Points (Composites)	322
10.2.1.	Listening.....	323
10.2.2.	Reading	325
10.2.3.	Speaking.....	327
10.2.4.	Writing	329
10.2.5.	Oral Composite	331
10.2.6.	Literacy Composite	333
10.2.7.	Comprehension Composite	335
10.2.8.	Overall Composite	337
10.3.	Interrater Agreement Rates	339
10.3.1.	Overview	339
10.3.2.	Interrater Reliability Study for Writing and Speaking	340
10.3.3.	Interrater Reliability Study for Reading and Listening	347
10.4.	Accuracy and Consistency of Domains/Composites.....	356

10.4.1.	Classification Accuracy and Consistency	356
10.4.2.	Listening Accuracy and Consistency	358
10.4.3.	Reading Accuracy and Consistency	361
10.4.4.	Speaking Accuracy and Consistency	364
10.4.5.	Writing Accuracy and Consistency	367
10.4.6.	Oral Composite Accuracy and Consistency	370
10.4.7.	Literacy Composite Accuracy and Consistency	373
10.4.8.	Comprehension Composite Accuracy and Consistency	376
10.4.9.	Overall Composite	379
11.	Quality Control	382
11.1.	Test Assembly	382
11.1.1.	Field Test Assembly	382
11.1.2.	Operational Test Assembly	384
11.2.	Test Administration and Scoring Quality Control	385
11.3.	Score Reporting Quality Control	386
	References	388

1. Purpose and Design of WIDA Alternate ACCESS

1.1. Purpose of Alternate ACCESS for ELLs

WIDA Alternate ACCESS (Alternate ACCESS) is an assessment of English language proficiency (ELP) for students in grades K–12 who are classified as English learners (ELs) and who have the most significant cognitive disabilities that prevent their meaningful participation in ACCESS for ELLs. English learners with the most significant cognitive disabilities are individuals who have one or more disabilities that significantly limit their intellectual functioning and adaptive behavior as documented in their Individualized Education Programs (IEP), and who are progressing toward English language proficiency in speaking, reading, writing, and listening.

Alternate ACCESS meets federal accountability requirements and provides educators with a measure of the English language proficiency growth of ELs with the most significant cognitive disabilities.

Accordingly, Alternate ACCESS is used to determine whether ELs with the most significant cognitive disabilities are making adequate progress in their English language proficiency development and whether those students should be reclassified, i.e., no longer be designated as ELs. The assessment also serves as a component in state, district, and school accountability models, as per the Every Student Succeeds Act of 2015 (ESSA) which requires that all students identified as ELs, including those who receive special education services, be assessed annually for English language proficiency.

WIDA Alternate ACCESS assesses students' English language proficiency in the domains of Listening, Speaking, Reading, and Writing as specified in ESSA. WIDA Alternate ACCESS provides students with additional opportunities to demonstrate their English language proficiency. Features of the test include simplified language, repetition of questions, heavy reliance on graphics rather than on text, larger size of testing materials and graphics, and availability of cues and supplemental questions. The test is based on WIDA's Alternate English Language Proficiency Level Descriptors and corresponds to the *WIDA English Language Development Standards Framework, 2020 Edition: Kindergarten–Grade 12*.

The Individuals with Disabilities Education Act of 2004 (IDEA) also mandates that students with disabilities participate in state and district assessment programs, including alternate assessments, with any accommodations documented in Individualized Education Programs (IEPs). Extensive support for adaptation to support students' accommodation needs are built into the assessment's design. During the test administration, individualized instructional supports, which are practices that are used by teachers in everyday classroom instruction to meet individual student needs, may be used. The Alternate ACCESS script contains the following unique features:

- Scripted cues and repetitions,
- Repetition and auxiliary questions that provide additional opportunities for students to demonstrate their proficiency in the Speaking Section,
- Modeling of tasks in the Writing Section

These unique features are designed for ELs with the most significant cognitive disabilities to sufficiently demonstrate their English language proficiency.

1.2. The WIDA Standards

Five foundational WIDA ELD Standards inform the design, structure, and content of ACCESS assessments:

- Standard 1: English language learners communicate in English for Social and Instructional purposes within the school setting.
- Standard 2: English language learners communicate information, ideas, and concepts necessary for academic success in the content area of Language Arts.
- Standard 3: English language learners communicate information, ideas, and concepts necessary for academic success in the content area of Mathematics.
- Standard 4: English language learners communicate information, ideas, and concepts necessary for academic success in the content area of Science.
- Standard 5: English language learners communicate information, ideas, and concepts necessary for academic success in the content area of Social Studies.

Every selected response item and every constructed-response task on Alternate ACCESS targets at least one of these five Standards.

1.3. The WIDA Proficiency Levels

The Alternate English Language Proficiency Levels (Alternate PLs) for Alternate ACCESS are designed to be derivatives of the WIDA ACCESS assessment's English language PLs and are reflective of expectations for students with the most significant cognitive disabilities. Alternate ACCESS includes items aligned to levels PL1–PL5 for all domains. Figure 1 illustrates how the Alternate ACCESS proficiency levels build upon each other.

Figure 1.3

Alternate ACCESS Proficiency Levels



Alternate English language PLs provide a global overview of the language acquisition process. The alternate English language PLs describe EL students' increasing comprehension and production in the following areas:

- Discourse Dimension/Sentence Dimension: This criterion addresses overall meaning across an entire text and contributes to the grammatical complexity of a test.
- Word/Phrase Dimension: This criterion reflects precision in communication at the word and phrase level.

Students at PL1: Entering can communicate using routine and familiar expressions, recognize single words or symbols, and produce intentional sounds or single representations.

Students at PL2: Emerging can understand and use simple expressions and single ideas, recognize short phrases, and produce chunks of language or single words.

Students at PL3: Developing can comprehend and produce simple connected statements or questions, recognize simple sentences, and write phrases or clauses about familiar ideas.

Students at PL4: Expanding can understand and communicate compound connected expressions with related ideas, recognize simple connected text, and write simple sentences with expanded ideas.

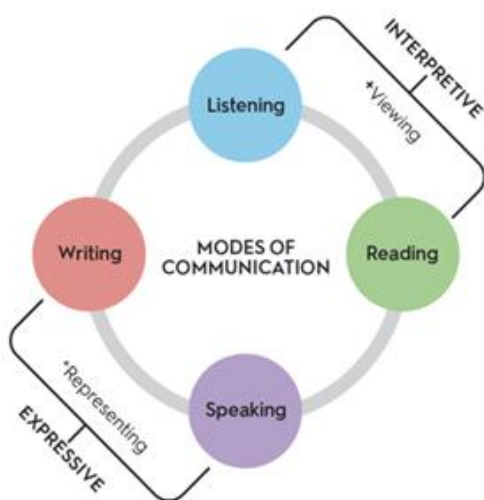
Students at PL5: Bridging can interpret and produce a variety of connected and complex statements, recognize organized text with various sentences, and write coherent sentences reflecting complex ideas.

1.4. Language Domains

Alternate ACCESS assesses students' English language proficiency in the domains of Listening, Speaking, Reading, and Writing as specified in ESSA. WIDA further operationalizes these four domains into two communication modes: interpretive and expressive. See Figure 2. The interpretive mode focuses on how we observe students processing language while the expressive mode focuses on what students can produce with language. These modes spotlight the multimodal nature of both language development and content-area learning. They position language as being more tightly integrated with other communication resources by including viewing and representing. These modes invite multiple means of engagement, representation, action, and expression, thereby extending accessibility principles to all ELs, including those with more intensive learning needs.

Figure 1.4

WIDA Modes of Communication



1.5. Grade-Level Clusters

Alternate ACCESS is administered in four grade-level clusters: kindergarten to grade 2, grades 3 to 5, grades 6 to 8, and grades 9 to 12. These levels were chosen based on the common topics identified in academic content material (i.e., English language arts, mathematics, science, and social studies) for students with the most significant cognitive disabilities in each of these clusters.

2. Test Development

The conceptual framework for the updated version of the Alternate ACCESS assessment builds upon the collective knowledge and lessons learned from 10 years of test administration. It also draws upon findings from the Alternate English Language Learning Assessment (ALTELLA) project, which identified key elements needed for the development of an alternate English language proficiency assessment that meets federal peer review.

The foundation of the Alternate ACCESS assessment is the *WIDA English Language Development Standards Framework, 2020 Edition: Kindergarten–Grade 12* (hereafter WIDA ELD Standards Framework, 2020 Edition, or 2020 Edition). The 2020 Edition views academic language to be integrated within academic content. This content–language integration means that ELs develop content and language concurrently, with academic content as a context for language learning and language as a means for learning academic content. Thus, the assessment should embed language within academic content. The WIDA ELD Standards Framework, 2020 Edition identifies the language of five academic content areas to be relevant for ELs: the language for social and instructional purposes, the English language arts, the language for mathematics, the language for science, and the language for social studies. The Alternate ACCESS assessment must correspond to the language identified in these five content areas. The connection to the 2020 Edition was operationalized through WIDA’s revised Alternate ACCESS Test Specifications (WIDA, 2021).

Another foundational element of Alternate ACCESS is the WIDA Alternate English Language Proficiency Levels Descriptors. These descriptors reflect the language expectations of ELs with the most significant cognitive disabilities. Altogether, WIDA’s decade-long experience administering Alternate ACCESS, current research on developing alternate English language proficiency assessments, the WIDA ELD Standards Framework, 2020 Edition, and the WIDA Alternate English Language Proficiency Level Descriptors guide the types of assessment items to be developed in the updated version of Alternate ACCESS.

2.1. Item and Task Design

2.1.1. Listening Items

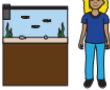
Listening items are designed to be selected response items and administered to the student one-to-one. Each item on the Listening test targets the language of one of the five WIDA ELD Standards and tests a student's ability to process language at one of the five fully delineated proficiency levels. The test administrator reads the item from the script while students respond to response options found in the Student Test Booklet. Listening items include three answer choices: one key and two distractors. Answer choices are primarily illustrations. Students may respond by verbalizing a response or by pointing to the image. The test administrator records the student response in the Student Response Booklet. Cue A of each item is aligned to the proficiency level of the item. Cues B and C offer the student additional scaffolding and support, with Cue C aligned to a lower proficiency level. A sample Listening item is provided in Figures 2.1.1.a, 2.1.1.b, and 2.1.1.c.

Figure 2.1.1.a

Listening Item: Test Administrator Script

TASK 6	
CUE A	
Paula and her friends have pet fish. The fish are all different. Paula has three zebrafish, Ali has two goldfish, and Kai has one angelfish. Paula has the most fish.	
Sweep across THREE FISH, TWO FISH, and ONE FISH. How many fish does Paula have? PAUSE.	
If correct, go to MOVING ON box.	
If incorrect or no response, repeat CUE A.	
After repeat, if incorrect or no response, go to CUE B.	

6






		
---	---	---

Figure 2.1.1.b

Listening Item: Test Administrator Script continued



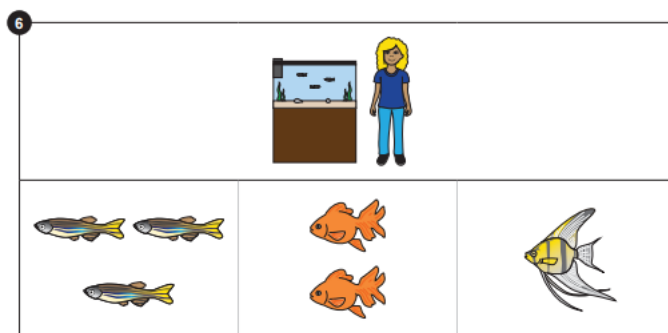
TASK 6	
CUE B	
Point to PAULA. Paula has three zebrafish, Ali has two goldfish, and Kai has one angelfish.	
Point to THREE FISH. Three fish.	
Point to TWO FISH. Two fish.	
Point to ONE FISH. One fish.	
Sweep across THREE FISH, TWO FISH, and ONE FISH. How many fish does Paula have?	
PAUSE.	
If correct, go to MOVING ON box.	
If incorrect or no response, go to CUE C .	
CUE C	
Point to PAULA. Paula has three zebrafish.	
Point to THREE FISH. Three fish.	
Point to TWO FISH. Two fish.	
Point to ONE FISH. One fish.	
Sweep across THREE FISH, TWO FISH, and ONE FISH. How many fish does Paula have?	
PAUSE.	
Go to MOVING ON box.	
MOVING ON	
If No Response or Incorrect on 3 CONSECUTIVE TASKS, stop the testing session.	
	
Record student's score in Student Response Booklet. Turn the page and go to TASK 7 . Good. Let's look at the next page.	Record student's score in Student Response Booklet. Record all remaining tasks Not Administered. Good job. We are finished.

Figure 2.1.1.c

Listening Item: Student Test Booklet



2.1.2. Reading Items

Reading items are designed to be selected response items and administered one-to-one with the student. A unique feature of the reading domain includes the use of Cloze items. They are similar in format to Listening items. The test administrator reads the item from the script while students attend to the reading passage and response options found in the Student Test Booklet. Reading items include a reading prompt and three answer choices: one key and two distractors. Answer choices are primarily text with supporting illustrations depending on the proficiency level of the item. Students may respond by verbalizing a response or by pointing to the image. The test administrator records the student response in the Student Response Booklet. Cue A of each item is aligned to the proficiency level of the item. Cues B and C offer the student additional scaffolding and support, with Cue C aligned to a lower proficiency level. A sample Reading item is provided in Figures 2.1.2.a, 2.1.2.b, and 2.1.2.c.


Figure 2.1.2.a

Reading Item: Test Administrator Script

TASK 9
CUE A
<p>Point to SENTENCES. Read these sentences. PAUSE.</p> <p>Point to QUESTION. When did Isabel get to school early? PAUSE.</p> <p>If correct, go to <i>MOVING ON</i> box.</p> <p>If incorrect or no response, repeat <i>CUE A</i>.</p> <p>After repeat, if incorrect or no response, go to <i>CUE B</i>.</p>

9

When the weather is nice, Isabel usually walks to school. Yesterday Isabel took a shortcut and got to school early. Today she walked with a friend and got to school on time.



When did Isabel get to school early?

today	yesterday	tomorrow
-------	-----------	----------

Figure 2.1.2.b

Reading Item: Test Administrator Script continued




TASK 9	
CUE B	
Point to SENTENCES. Yesterday Isabel took a shortcut and got to school early. Today she walked with a friend and got to school on time.	
Point to TODAY. Today.	
Point to YESTERDAY. Yesterday.	
Point to TOMORROW. Tomorrow.	
Sweep across TODAY, YESTERDAY, and TOMORROW. When did Isabel get to school early?	
PAUSE.	
If correct, go to MOVING ON box.	
If incorrect or no response, go to CUE C .	
CUE C	
Point to SENTENCE. Yesterday Isabel took a shortcut and got to school early.	
Point to YESTERDAY. Yesterday.	
Point to TODAY. Today.	
Point to YESTERDAY. Yesterday.	
Point to TOMORROW. Tomorrow.	
Sweep across TODAY, YESTERDAY, and TOMORROW. When did Isabel get to school early?	
PAUSE.	
Go to MOVING ON box.	
MOVING ON	
If No Response or Incorrect on 3 CONSECUTIVE TASKS, stop the testing session.	
	
Record student's score in Student Response Booklet. Turn the page and go to TASK 10. Good. Let's look at the next page.	Record student's score in Student Response Booklet. Record all remaining tasks Not Administered. Good job. We are finished.

Figure 2.1.2.c.

Item: Student Test Booklet

<p>9 When the weather is nice, Isabel usually walks to school. Yesterday Isabel took a shortcut and got to school early. Today she walked with a friend and got to school on time.</p>		
		
When did Isabel get to school early?		
today	yesterday	tomorrow

2.1.3. Writing Tasks

For the Writing test, students interact directly with the Student Response Booklet. They can use their preferred writing instrument, and they can write in the booklet, on a separate piece of paper, or on any medium they typically use during instruction, provided the tool gives the student access to all 26 letters of the alphabet. Writing items are designed to be constructed response items, eliciting language corresponding to one of the WIDA ELD Standards. The test administrator reads the item from the Test Administrator Script while the student attends to an image and supporting information in the Student Response Booklet. The Writing test contains supports and scaffolding for students. The Writing test in grades K–2 differs from the Writing test in grades 3–5, 6–8, and 9–12, as it includes additional scaffolding and modeling throughout QUESTIONS 1, 2, and 3. Often these supports include shared writing activities where the test administrator writes part of the response, and the student completes the task. This intentional design is intended to support early literacy and writing for early learners. QUESTION 1 of each task is aligned to the proficiency level of the item, and the proficiency level expectation for each Writing task is listed in the Test Administrator Script by the scaffolded question. QUESTIONS 2 and 3 offer the student additional scaffolding and support. The test administrator records the score in the Student Response Booklet. A sample Writing item is provided in Figures 2.1.3.a., 2.1.3.b., 2.1.3.c., and 2.1.3.d.

Figure 2.1.3.a.

Writing Item: Test Administrator Script

Alternate Proficiency Level – Emerging

TASK 7

QUESTION 1

Point to PICTURE. **Tomas wants to use markers, but he has colored pencils. Mona has eight markers, but she likes colored pencils. Tomas and Mona trade, so Tomas gets eight markers and Mona gets eight pencils. Why is this a good idea?**

TASK:

Point to PICTURE. **Now it's time to write some sentences. Why is this a good idea?**
Write that here. PAUSE.

Point to WORD LIST. **You can use the words in the list to help you.**
Read the words from WORD LIST to student. **You can begin writing now.** PAUSE.
(IF NECESSARY) **Try to write that.** PAUSE.

If student verbally responds:
Point to BLANK LINES. **Now write** (student response) **here.** PAUSE.
(IF NECESSARY) **You can use these words to help you.** PAUSE.

markers	pencils	trade
same	like	want

If meets, go to **MOVING ON** box.
If approaches or no response, continue on next page.

Q1 EXPECT	The student writes at least 2 simple sentences or at least 1 compound sentence.
-----------	---

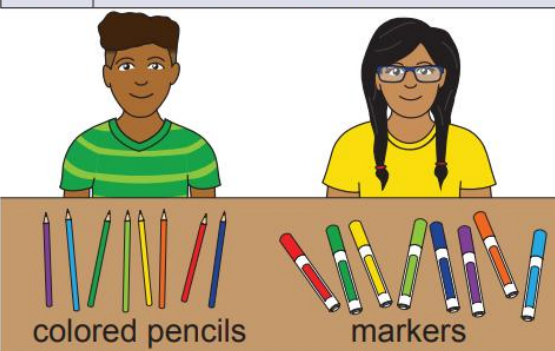


Figure 2.1.3.b.

Writing Item: Test Administrator Script continued



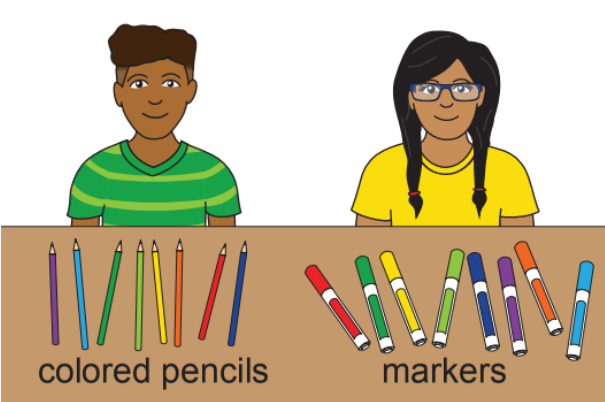
Potential Proficiency Level = 3 Emerging	
TASK 7	
QUESTION 2	
<p>Tomas wants to use markers, but he has colored pencils. Mona has markers, but she likes colored pencils. Tomas and Mona trade.</p> <p>Is this a good idea? Why or why not? PAUSE.</p> <p>Point to WORD LIST and BLANK LINES. Tell me if trading is a good idea. PAUSE.</p> <p>If the student verbally responds:</p> <p>Point to BLANK LINES. Now write that in a sentence here. PAUSE.</p> <p>(IF NECESSARY) You can use these words to help you. PAUSE.</p> <p>If meets, go to MOVING ON box.</p> <p>If approaches or no response, continue.</p>	
Q2 EXPECT	The student writes at least 2 simple sentences or at least 1 compound sentence.
QUESTION 3	
<p>Point to MARKERS. These are markers.</p> <p>Point to WORD LIST and BLANK LINES. Now try to write that here. You can use the word list. PAUSE.</p> <p>Go to MOVING ON box.</p>	
Q3 EXPECT	The student writes 1 simple sentence.
MOVING ON <p>If No Response or Incorrect on 3 CONSECUTIVE TASKS, stop the testing session.</p>	
 <p>Record student's score in Student Response Booklet. Turn the page and go to TASK 8. Good. Let's look at the next page.</p>	 <p>Record student's score in Student Response Booklet. Record all remaining tasks Not Administered. Good job. We are finished.</p>

Figure 2.1.3.c.

Writing Item Stimulus: Student Response Booklet

WRITING		Task 7
Task	7	
Q1		
Q2		
Q3		
Meets		
Approaches		
No Response		
Not Administered		



colored pencils markers

Figure 2.1.3.d.

Writing Item: Student Response Booklet

WRITING		
Task 7		
markers	pencils	trade
same	like	want

2.1.4. Speaking Tasks

The Speaking test has eight tasks that progressively increase in difficulty from level P1 to level P5. Each task in the Speaking test is made up of three questions. Speaking items are designed to be constructed response items and administered one-to-one with the student. QUESTION 1 of each task is aligned to the proficiency level of the item, and the proficiency level expectation for each Speaking task is listed in the Student Response Booklet. QUESTIONS 2 and 3 offer the student additional scaffolding and support. The test administrator reads the item from the Test Administrator Script while the student attends to an image and supporting information in the Student Test Booklet. The test administrator records the score in the Student Response Booklet. A sample Speaking item is provided in Figures 2.1.4.a., 2.1.4.b., and 2.1.4.c.

Figure 2.1.4.a.

Speaking Item: Test Administrator Script

Alternate Proficiency Level – Emerging

TASK 2

QUESTION 1

Carlos is outside. There are trees and birds.

Sweep across PICTURE. **What is outside?** PAUSE.

If meets, go to **MOVING ON** box.

If approaches or no response, repeat **QUESTION 1**.

After repeat, if approaches or no response, go to **QUESTION 2**.

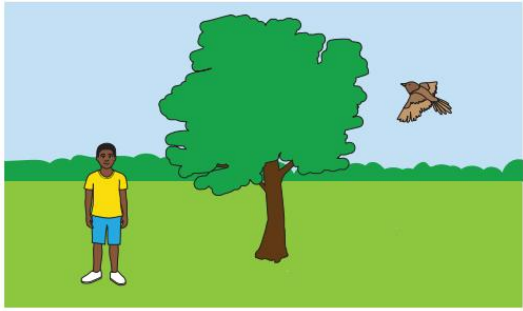


Figure 2.1.4.b.

Speaking Item: Test Administrator Script continued


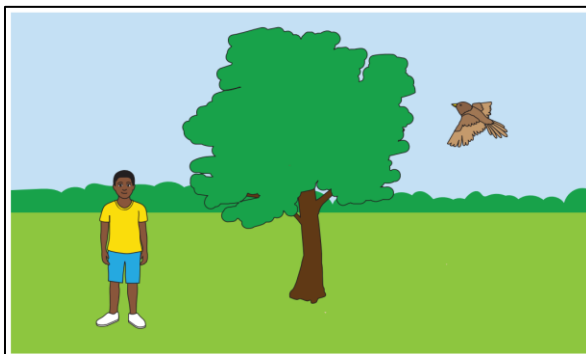
Potential Proficiency Level = 2 Emerging	
TASK 2	
QUESTION 2	
Carlos sees trees and birds outside.	
Point to PICTURE: What does Carlos see? PAUSE.	
If meets, go to MOVING ON box.	
If approaches or no response, repeat QUESTION 2 .	
After repeat, if approaches or no response, go to QUESTION 3 .	
QUESTION 3	
Carlos sees a tree.	
Point to PICTURE: Can you say tree? PAUSE.	
If meets, go to MOVING ON box.	
If approaches or no response, repeat QUESTION 3 .	
After repeat, go to MOVING ON box.	
MOVING ON	
	
Remember to record student's score in Student Response Booklet. Turn page and go to TASK 3 .	
Good. Let's turn the page and keep going.	

Figure 2.1.4.c.

Speaking Item: Student Test Booklet



2.2. Test Design

The operational WIDA Alternate ACCESS assessment (Series 602 OP) maintains many of the features of the earlier Alternate ACCESS for ELLs, but also employs many updates based on 2023 field test findings. The alternate assessment provides students with additional opportunities to demonstrate their language proficiency. Such features of the test include simplified language, repetition of questions, heavy reliance on graphics rather than on text, larger size of testing materials and graphics, and availability of cues and supplemental questions. During the test administration, individualized instructional supports, which are practices that are used by teachers in everyday classroom instruction to meet individual student needs, may be used.

Alternate ACCESS (Series 602 OP) consists of four domain subtests (Listening, Speaking, Reading, and Writing) across four grade-level clusters (K–2, 3–5, 6–8, and 9–12). For each grade-level cluster, the test form consists of 10 listening items, 10 reading items, 8 speaking items, and 8 writing items. Given the Universal Design of Alternate ACCESS, test administrators offer students multiple opportunities within each of the domains to respond to a task. With each opportunity, test administrators offer additional scaffolding and support. In Listening and Reading, the task is broken down into cues (CUE A, CUE B, CUE C) and in Speaking and Writing, the task is broken down into questions (QUESTION 1, QUESTION 2, QUESTION 3). Each test item includes graphics as an additional scaffold; however, special attention was given to make sure that the items are also accessible to students with visual impairments.

2.2.1. Listening

Tables 2.2.1.a., 2.2.1.b., 2.2.1.c., 2.2.1.d., 2.2.1.e., and 2.2.1.f. outline the distribution of tasks in the Listening domain by standard, PL, and grade-level cluster.

Table 2.2.1.a.

Range of Standards Across Grade-Level Clusters in Listening

Standard	Range
Social and Instructional Language (SIL)	2 Tasks
Language for Mathematics (LMA)	2–3 Tasks
Language for Language Arts (LLA)	1–2 Tasks
Language for Science (LSC)	2–3 Tasks
Language for Social Studies (LSS)	1–2 Tasks
Language for Language Arts/ Social Studies (LS)	1–2 Tasks
Language for Mathematics/Science (MS)	1 Task

Note: The Listening test is comprised of 10 discrete tasks

Table 2.2.1.b.**Range of Proficiency Levels across Grade-Level Clusters in Listening**

Level	Range
PL 1 Entering	1–2 Tasks
PL 2 Emerging	1–2 Tasks
PL 3 Developing	2–3 Tasks
PL 4 Expanding	2–3 Tasks
PL 5 Bridging	2–3 Tasks

Table 2.2.1.c.**Listening Form 602: Grade Level-Cluster K–2**

Task	Standard	Proficiency Level	Description
Task 1	Language for Language Arts	1	Identify characters, places, or objects from visuals and oral labels in illustrated pattern or predictable books.
Task 2	Language for Mathematics	2	Match attributes of two- or three-dimensional shapes described orally in pictures.
Task 3	Social and Instructional Language	2	Identify symbols, objects, or people associated with classrooms or school areas, personnel, or activities from pictures and oral statements.
Task 4	Language for Science	3	Interpret scientific informational texts by defining or classifying a concept or entity.
Task 5	Language for Social Studies	3	Interpret informational texts in social studies by defining and classifying attributes, characteristics, and qualities in relevant information.
Task 6	Language for Science	3	Interpret scientific informational texts by defining or classifying a concept or entity.
Task 7	Language for Science	3	Interpret scientific informational texts by defining or classifying a concept or entity.
Task 8	Language for Social Studies	4	Interpret social studies arguments by analyzing evidence gathered from source.
Task 9	Language for Mathematics	4	Interpret mathematical informational texts by identifying concept or entity.
Task 10	Language for Social Studies	5	Interpret informational texts in social studies by defining and classifying attributes, characteristics, and qualities in relevant information.

Table 2.2.1.d.**Listening Form 602: Grade-Level Cluster 3–5**

Task	Standard	Proficiency Level	Description
Task 1	Language for Language Arts	1	Interpret informational texts in language arts by identifying the main idea and key details.
Task 2	Language for Language Arts	1	Interpret informational texts in language arts by identifying the main idea and key details.
Task 3	Language for Mathematics	1	Label mathematical graphs or diagrams following oral cues.
Task 4	Language for Mathematics	2	Interpret mathematical informational texts by identifying concept or entity.
Task 5	Social and Instructional Language	2	Follow two-step oral commands supported visually or modeled.
Task 6	Language for Social Studies	3	Interpret social studies explanations by analyzing sources for event sequences and/or causes/effects.
Task 7	Language for Science	3	Distinguish among examples of states of matter from oral statements and visual support.
Task 8	Social and Instructional Language	4	Ask questions about what others have shared.
Task 9	Language for Social Studies	5	Interpret informational texts in social studies by defining and classifying attributes, characteristics, and qualities in relevant information.
Task 10	Language for Science	5	Interpret informational texts in social studies by defining and classifying attributes, characteristics, and qualities in relevant information.

Table 2.2.1.e.**Listening Form 602: Grade-Level Cluster 6–8**

Task	Standard	Proficiency Level	Description
Task 1	Language for Mathematics	1	Match quantity of objects given in oral directions.
Task 2	Social and Instructional Language	2	Identify needed resources or supplies for activities from pictures and oral statements.
Task 3	Social and Instructional Language	3	Share initial thinking with others.
Task 4	Language for Language Arts	3	Distinguish between sources of information and distractor pictures based on oral descriptions.
Task 5	Language for Language Arts	3	Identify illustrated rhyming words in recited excerpts from poems.
Task 6	Language for Science	3	Distinguish between scientific instruments and distractor pictures based on oral directions.
Task 7	Language for Mathematics	4	Interpret mathematics arguments by comparing conjectures with previously established results.
Task 8	Language for Social Studies	5	Interpret social studies explanations by analyzing sources for logical relationships among contributing factors or causes.
Task 9	Social and Instructional Language	5	Define and classify facts and interpretations; determine what is known vs. unknown
Task 10	Language for Social Studies	5	Interpret social studies explanations by analyzing sources for logical relationships among contributing factors or causes.

Table 2.2.1.f.**Listening Form 602: Grade-Level Cluster 9–12**

Task	Standard	Proficiency Level	Description
Task 1	Language for Social Studies	1	Interpret social studies arguments by identifying topic and purpose (argue in favor or against a position, present a balanced interpretation, challenge perspective).
Task 2	Social and Instructional Language	2	Sort, clarify, and summarize relationships.
Task 3	Language for Mathematics	2	Interpret mathematical informational texts by identifying concept or entity.
Task 4	Language for Science	3	Interpret scientific explanations by defining investigable questions or problems based on observations, information, and/or data about a phenomenon.
Task 5	Language for Language Arts	3	Select sources of information based on oral descriptions.
Task 6	Language for Language Arts/Social Studies	4	Interpret social studies arguments by analyzing relevant information to support and/or revise claims with reliable and valid evidence from multiple sources.
Task 7	Language for Mathematics/Science	4	Interpret mathematics arguments by evaluating relationships among evidence and mathematical principles to create generalizations.
Task 8	Language for Language Arts/Social Studies	4	Interpret social studies arguments by analyzing relevant information to support and/or revise claims with reliable and valid evidence from multiple sources.
Task 9	Language for Mathematics/Science	4	Interpret mathematics arguments by evaluating relationships among evidence and mathematical principles to create generalizations.
Task 10	Language for Science	5	Interpret scientific arguments by identifying appropriate and sufficient evidence from data, models, and/or information from investigations of a phenomenon or design solutions.

2.2.2. Reading

Tables 2.2.2.a., 2.2.2.b., 2.2.2.c., 2.2.2.d., 2.2.2.e., and 2.2.2.f. outline the distribution of tasks in the Reading domain by standard, PL, and grade-level cluster.

Table 2.2.2.a.

Range of Standards Across Grade-Level Clusters in Reading

Standard	Range
Social and Instructional Language (SIL)	2–3 Tasks
Language for Mathematics (LMA)	1–2 Tasks
Language for Language Arts (LLA)	1–3 Tasks
Language for Science (LSC)	2–3 Tasks
Language for Social Studies (LSS)	1–2 Tasks
Language for Language Arts/ Social Studies (LS)	1–2 Tasks
Language for Mathematics/Science (MS)	1 Task

Table 2.2.2.b.

Range of Levels Across Grade-Level Clusters in Reading

Level	Range
PL 1 Entering	1–2 Tasks
PL 2 Emerging	1–2 Tasks
PL 3 Developing	2–3 Tasks
PL 4 Expanding	2–3 Tasks
PL 5 Bridging	2–3 Tasks

Table 2.2.2.c.**Reading Form 602: Grade-Level Cluster K-2**

Task	Standard	Proficiency Level	Description
Task 1	Language for Language Arts	2	Interpret informational texts in language arts by identifying word choices in relation to topic or content area
Task 2	Language for Mathematics	2	Match quantity words to labeled pictures of varying quantities of objects.
Task 3	Social and Instructional Language	3	Recount and restate ideas.
Task 4	Social and Instructional Language	3	Recount and restate ideas.
Task 5	Language for Social Studies	3	Interpret informational texts in social studies by determining topic associated with a compelling or supporting question.
Task 6	Language for Science	4	Interpret scientific arguments by identifying potential evidence from data, models, and/or information from investigations of phenomena or design solutions.
Task 7	Language for Language Arts/Social Studies	4	Interpret language arts narratives by identifying characters, settings, and major events.
Task 8	Language for Language Arts/Social Studies	4	Match identical labeled pictures or photographs of living organisms.
Task 9	Language for Language Arts/Social Studies	5	Interpret language arts narratives by identifying how character attributes and actions contribute to an event.
Task 10	Language for Mathematics	5	Interpret mathematical informational texts by identifying concept or entity.

Table 2.2.2.d.**Reading Form 602: Grade-Level Cluster 3–5**

Task	Standard	Proficiency Level	Description
Task 1	Language for Science	2	Match labeled pictures representing earth materials with vocabulary.
Task 2	Language for Social Studies	2	Interpret language arts narratives by identifying how character attributes and actions contribute to an event.
Task 3	Social and Instructional Language	3	Select general themes related to leisure activities from pictures or words or phrases.
Task 4	Language for Mathematics	3	Identify large whole numbers from picture or models or phrases or short sentences.
Task 5	Social and Instructional Language	4	Support own opinions with reasons.
Task 6	Language for Social Studies	4	Interpret social studies explanations by evaluating disciplinary concepts and ideas associated with a compelling or supporting question.
Task 7	Social and Instructional Language	4	Support own opinions with reasons.
Task 8	Language for Social Studies	4	Interpret social studies explanations by evaluating disciplinary concepts and ideas associated with a compelling or supporting question.
Task 9	Language for Social Studies	4	Interpret social studies explanations by evaluating disciplinary concepts and ideas associated with a compelling or supporting question.
Task 10	Language for Language Arts	5	Interpret informational texts in language arts by describing relationship between a series of events, ideas or concepts, or procedural steps.

Table 2.2.2.e.**Reading Form 602: Grade-Level Cluster 6–8**

Task	Standard	Proficiency Level	Description
Task 1	Language for Language Arts	1	Match words from adventure book titles with words in text.
Task 2	Language for Language Arts	2	Match words from adventure book titles with words in text.
Task 3	Language for Language Arts	2	Interpret language arts arguments by identifying and summarizing central idea distinct from prior knowledge or opinions.
Task 4	Language for Mathematics	2	Match vocabulary associated with perimeter or area with graphics, symbols, or figures.
Task 5	Social and Instructional Language	3	Locate words or phrases on socially related topics (e.g. parties) from visually supported information (e.g. on invitations).
Task 6	Language for Mathematics	3	Identify visually supported examples of use of perimeter, area, volume, or circumference in real world situations
Task 7	Language for Science	4	Interpret scientific explanations by defining investigable questions or design problems based on observations, information, and/or data about a phenomenon.
Task 8	Language for Science	4	Interpret scientific explanations by defining investigable questions or design problems based on observations, information, and/or data about a phenomenon.
Task 9	Language for Social Studies	4	Interpret social studies arguments by evaluating point of view and credibility of source based on relevance and intended use.
Task 10	Language for Science	5	Interpret scientific explanations by determining central ideas in complex evidence and information to help explain how or why a phenomenon occurs.

Table 2.2.2.f.**Reading Form 602: Grade-Level Cluster 9–12**

Task	Standard	Proficiency Level	Description
Task 1	Social and Instructional Language	1	Sort, clarify, and summarize relationships.
Task 2	Social and Instructional Language	1	Sort, clarify, and summarize relationships.
Task 3	Social and Instructional Language	2	Share ideas about one's own and others' lived experiences and previous learning.
Task 4	Social and Instructional Language	3	Recount and restate ideas to sustain and move dialogue forward.
Task 5	Language for Mathematics	3	Interpret mathematical informational texts by identifying concept or entity.
Task 6	Language for Mathematics/ Science	4	Interpret scientific explanations by paraphrasing central ideas in complex evidence, concepts, processes, and information to help explain how or why a phenomenon occurs.
Task 7	Language for Social Studies	4	Interpret social studies arguments by analyzing relevant information to support and/or revise claims with reliable and valid evidence from multiple sources.
Task 8	Language for Mathematics/ Science	4	Interpret scientific explanations by paraphrasing central ideas in complex evidence, concepts, processes, and information to help explain how or why a phenomenon occurs.
Task 9	Language for Science	5	Interpret mathematical explanations by evaluating rationales, models, and/or interpretations based on evidence and mathematical principles.
Task 10	Language for Science	5	Interpret mathematical explanations by evaluating rationales, models, and/or interpretations based on evidence and mathematical principles.

2.2.3. Writing

Tables 2.2.3.a., 2.2.3.b., 2.2.3.c., 2.2.3.d., 2.2.3.e., and 2.2.3.f. outline the distribution of tasks in the Writing domain by standard, PL, and grade-level cluster.

Table 2.2.3.a.**Range of Standards Across Grade-Level Clusters in Writing**

Standards	Range
Social and Instructional Language (SIL)	1–2 Tasks
Language for Mathematics (LMA)	1 Task
Language for Language Arts (LLA)	1–2 Tasks
Language for Science (LSC)	1–3 Tasks
Language for Social Studies (LSS)	1–2 Tasks
Language for Mathematics/Science (MS)	1 Task

Table 2.2.3.b.**Range of Proficiency Levels Across Grade-Level Clusters in Writing**

Level	Range
PL 1 Entering	1–2 Tasks
PL 2 Emerging	1–2 Tasks
PL 3 Developing	2–3 Tasks
PL 4 Expanding	2–3 Tasks
PL 5 Bridging	2–3 Tasks

Table 2.2.3.c.**Writing Form 602: Grade-Level Cluster K–2**

Task	Standard	Proficiency Level	Description
Task 1	Language for Social Studies	1	Construct social studies explanations that describe components, order, causes, or cycles.
Task 2	Social and Instructional Language	2	Produce words about self, using models and pictures.
Task 3	Language for Language Arts	3	Select words related to settings or characters in illustrated stories.
Task 4	Language of Language Arts	4	Offer ideas and suggestions.
Task 5	Language for Social Studies	4	Construct informational texts in social studies that provide details about disciplinary ideas.
Task 6	Language for Social Studies	4	Construct informational texts in social studies that provide details about disciplinary ideas.
Task 7	Language for Language Arts	4	Offer ideas and suggestions.
Task 8	Language for Social Studies	5	Construct social studies arguments that show relationships between claim, evidence, and reasoning.

Table 2.2.3.d.**Writing Form 602: Grade-Level Cluster 3–5**

Task	Standard	Proficiency Level	Description
Task 1	Language for Social Studies	1	Construct social studies explanations that describe components, order, causes, or cycles.
Task 2	Language for Language Arts	2	Construct informational texts in language arts that add precision and details to define, describe, compare, and classify topic and/or entity.
Task 3	Language for Language Arts	2	Construct informational texts in language arts that add precision and details to define, describe, compare, and classify topic and/or entity.
Task 4	Language for Language Arts	2	Construct informational texts in language arts that add precision and details to define, describe, compare, and classify topic and/or entity.
Task 5	Language for Mathematics/Science	4	Construct scientific explanations that describe observations and/or data about a phenomenon.
Task 6	Language for Language Arts	4	Construct informational texts in language arts that introduce and define topic and/or entity for audience.
Task 7	Language for Mathematics/ Science	4	Construct scientific explanations that describe observations and/or data about a phenomenon.
Task 8	Language for Social Studies	5	Construct social studies explanations that introduce phenomena or events.

Table 2.2.3.e.**Writing Form 602: Grade-Level Cluster 6–8**

Task	Standard	Proficiency Level	Description
Task 1	Language for Science	1	Construct scientific arguments that introduce and contextualize topic/phenomenon in issues related to the natural and designed world(s).
Task 2	Language for Social Studies	2	Construct social studies arguments that select relevant information to support claims with evidence gathered from multiple sources.
Task 3	Language for Social Studies	2	Construct social studies arguments that select relevant information to support claims with evidence gathered from multiple sources.
Task 4	Language for Social Studies	3	Construct social studies explanations that develop reasoning, sequences with linear and nonlinear relationships, evidence, and details, acknowledging strengths and weaknesses.
Task 5	Language for Social Studies	3	Construct social studies explanations that develop reasoning, sequences with linear and nonlinear relationships, evidence, and details, acknowledging strengths and weaknesses.
Task 6	Social and Instructional Language	4	Compare changing variables, factors, and circumstances.
Task 7	Language for Science	4	Construct scientific explanations that develop reasoning to show relationships among independent and dependent variables in models and simple systems.
Task 8	Language for Language Arts	5	Construct informational texts in language arts that add precision, details, and clarity about relevant attributes, qualities, characteristics, activities, and behaviors.

Table 2.2.3.f.**Writing Form 602: Grade-Level Cluster 9–12**

Task	Standard	Proficiency Level	Description
Task 1	Language for Science	2	Construct scientific arguments that introduce and contextualize topic/phenomenon in current scientific or historical episodes in science.
Task 2	Language for Science	2	Construct scientific arguments that introduce and contextualize topic/phenomenon in current scientific or historical episodes in science.
Task 3	Language for Social Studies	3	Construct social studies arguments that select relevant information to support precise and knowledgeable claims with evidence from multiple sources.
Task 4	Language for Social Studies	3	Construct social studies arguments that select relevant information to support precise and knowledgeable claims with evidence from multiple sources.
Task 5	Social and Instructional Language	4	Compare changing variables, factors, and circumstances.
Task 6	Language for Language Arts	5	Construct informational texts in language arts that introduce and define topic and/or entity for audience.
Task 7	Language for Language Arts	5	Construct informational texts in language arts that introduce and define topic and/or entity for audience.
Task 8	Language for Science	5	Construct scientific explanations that summarize and refine solutions referencing scientific knowledge, evidence, criteria, and/or trade-offs.

2.2.4. Speaking

Tables 2.2.4.a., 2.2.4.b., 2.2.4.c., 2.2.4.d., 2.2.4.e., and 2.2.4.f. outline the distribution of tasks in the Speaking domain by standard, PL, and grade-level cluster.

Table 2.2.4.a.

Range of Standards Across Grade-Level Clusters in Speaking

Standard	Range
Social and Instructional Language (SIL)	1–2 Tasks
Language for Mathematics (LMA)	1–2 Tasks
Language for Language Arts (LLA)	1–3 Tasks
Language for Science (LSC)	1–2 Tasks
Language for Social Studies (LSS)	1–2 Tasks
Language for Language Arts/ Social Studies (LS)	1–2 Tasks
Language for Mathematics/Science (MS)	1 Task

Table 2.2.4.b.

Range of Proficiency Levels Across Grade-Level Clusters in Speaking

Level	Range
PL 1 Entering	1–2 Tasks
PL 2 Emerging	1–2 Tasks
PL 3 Developing	2–3 Tasks
PL 4 Expanding	2–3 Tasks
PL 5 Bridging	2–3 Tasks

Table 2.2.4.c.**Speaking Form 602: Grade-Level Cluster K-2**

Task	Standard	Proficiency Level	Description
Task 1	Language for Mathematics	2	Construct mathematical informational texts that compare/contrast concepts or entities.
Task 2	Language for Language Arts	3	Construct informational texts in language arts that describe attributes and characteristics with facts, definitions, and relevant details.
Task 3	Language for Language Arts	3	Construct informational texts in language arts that describe attributes and characteristics with facts, definitions, and relevant details.
Task 4	Social and Instructional Language	4	Share initial thinking with others.
Task 5	Language for Science	4	Construct scientific informational texts that summarize observations or factual information.
Task 6	Social and Instructional Language	4	Share initial thinking with others.
Task 7	Language for Language Arts/Social Studies	5	Construct language arts narratives that develop story with time and event sequences, complication, resolution, or ending.
Task 8	Social and Instructional Language	5	Define and classify objects or concepts.

Table 2.2.4.d.

Speaking Form 602: Grade-Level Cluster 3–5

Task	Standard	Proficiency Level	Description
Task 1	Language for Language Arts	3	Construct language arts arguments that support opinions with reasons and information.
Task 2	Language for Language Arts	2	Construct informational texts in language arts that introduce and define topic and/or entity for audience.
Task 3	Language for Language Arts	3	Construct language arts arguments that support opinions with reasons and information.
Task 4	Language for Language Arts	3	Construct language arts arguments that support opinions with reasons and information.
Task 5	Language for Social Studies	4	Construct social studies explanations that generalize possible reasons for a development or event.
Task 6	Language for Mathematics	4	Construct mathematical explanations and state reasoning used to generate solution.
Task 7	Language for Mathematics	5	Construct mathematical explanations that describe data and/or steps to solve problems.
Task 8	Language for Social Studies	5	Construct social studies explanations that generalize probable causes and effects of developments or events.

Table 2.2.4.e.**Speaking Form 602: Grade-Level Cluster 6–8**

Task	Standard	Proficiency Level	Description
Task 1	Language for Language Arts	2	Construct language arts narratives that develop a story, including themes with complication and resolution, time, and event sequences.
Task 2	Language for Science	3	Construct scientific explanations that describe valid and reliable evidence from sources about a phenomenon.
Task 3	Language for Science	3	Construct scientific explanations that describe valid and reliable evidence from sources about a phenomenon.
Task 4	Social and Instructional Language	4	Create closure, recap, and offer next steps.
Task 5	Language for Mathematics	4	Construct mathematical arguments that justify conclusions with evidence and mathematical facts.
Task 6	Language for Mathematics	4	Construct mathematical arguments that justify conclusions with evidence and mathematical facts.
Task 7	Language for Social Studies	5	Construct social studies explanations that generalize multiple causes and effects of developments or events.
Task 8	Language for Science	5	Construct scientific explanations that describe valid and reliable evidence from sources about a phenomenon.

Table 2.2.4.f.**Speaking Form 602: Grade-Level Cluster 9–12**

Task	Standard	Proficiency Level	Description
Task 1	Language for Language Arts	2	Construct language arts narratives that orient audience to context and one or multiple point(s) of view.
Task 2	Language for Mathematics/Science	3	Construct scientific explanations that describe reliable and valid evidence from multiple sources about a phenomenon.
Task 3	Language for Language Arts	3	Construct language arts arguments that support claims and refute counterclaims with valid reasoning and relevant and sufficient evidence.
Task 4	Language for Science	4	Construct scientific explanations that develop reasoning to illustrate and/or predict the relationships between variables in a system or between components of a system.
Task 5	Language for Science	4	Construct scientific explanations that develop reasoning to illustrate and/or predict the relationships between variables in a system or between components of a system.
Task 6	Social and Instructional Language	5	Report on explicit and inferred characteristics, patterns, or behavior.
Task 7	Language for Language Arts/Social Studies	5	Construct language arts arguments that logically organize claims, counterclaims, reasons, and evidence; offer a conclusion with recommendations.
Task 8	Language for Language Arts/Social Studies	5	Construct language arts arguments that logically organize claims, counterclaims, reasons, and evidence; offer a conclusion with recommendations.

2.3. *Test Construction*

2.3.1. Item Development

WIDA worked with Accessible Teaching, Learning, and Assessment (ATLAS) at the University of Kansas for item development. WIDA provided item specifications for 13 exemplar items that ATLAS staff developed. These items included new, innovative item types such as cloze, dictation, and matching. Each item specification included the grade band, content area, proficiency level, domain, proficiency level descriptor(s), language expectation, alternate academic content standard, item type, and accommodation considerations. The ATLAS Test Development (TD) team assigned to this project designed and wrote the first draft of the exemplars. In collaboration with WIDA, the items were revised and edited until finalized. These items were evaluated by WIDA during cognitive labs. Cognitive labs consisted of a total of 38 labs across 5 states: 28 with English learners and 10 with non-English learners. Overall, students and test administrators had positive experiences with the items. Many test administrators reported that the items could provide useful information about students' English language proficiency and indicated that the scoring rules and tables were clear. In terms of accessibility, many test administrators felt the assessment was accessible to students. However, test administrators had concerns about difficulties that students with low vision, students that used eye gaze, and students with assistive devices (e.g., AAC devices), might face with new test items. The cognitive labs examined three new item types: cloze, dictation, and matching item types. Matching items were evaluated least favorably by test administrators. Test administrators found the matching items to be difficult and confusing, with these items needing more clarity and additional scripting. The matching items took longer for some students, suggesting that students found these items to be more difficult than the other item types. Dictation items were rarely mentioned by test administrators in the interview. Students generally knew what to do with cloze items and this format allowed them to respond in multiple ways. Based on the results of the cognitive labs, the cloze item type was selected to be the new item type for the redesign of Alternate ACCESS, and the cloze item would be used exclusively with the Reading domain test.

A total of 20 representative and experienced item writers then developed 230 new items for Alternate ACCESS during the Advancing ALTELLA item-writing event in Charleston, South Carolina in May of 2022. After the initial work, the items were reviewed by ATLAS and WIDA subject matter experts and prepared for an external bias, sensitivity, and content review.

Many aspects of the WIDA item-writing event were influenced by approaches used for the Dynamic Learning Maps (DLM) assessment system (ATLAS). The event was structured to assist item writers in building their own knowledge about the EL/students with significant cognitive disabilities (SCD) population and use that knowledge to produce high-quality items. Item writers were placed together in pairs based on their expertise with grade band content. Each pair had an EL language expert and an SCD population expert. Item writing pairs worked together to write and revise items for each item specification provided to them.

The items received an EL and SCD review, including an alignment check to the item specifications and Performance Level Descriptors. ATLAS provided an editorial review and accessible graphics were developed. WIDA reviewed the items again and provided feedback to ATLAS. ATLAS reviewed and revised items and prepared them for WIDA's external review.

After an iterative review process between WIDA and ATLAS, WIDA conducted two separate, virtual Bias, Sensitivity, and Content Reviews for grades K–5 and 6–12. The purpose of the Advancing ALTELLA Alternate ACCESS for ELLs Bias, Sensitivity, and Content Review (BSC review) was to use the expertise of a trained group of educators to help ensure that Alternate ACCESS for ELLs is a fair, accurate, and unbiased assessment instrument for the diverse group of test takers.

After the external review, the WIDA team processed all the external review data and used trends and data to provide ATLAS with specific revisions to items. The revision recommendations were used to improve the quality, relevance, and accuracy of the items. External Review Panelists recommended revising graphics to be more sensitive to the students of all backgrounds and representative of all disabilities. Additionally, panelists suggested revisions to the language to make it gender neutral when possible. Writing instructions were revised to be consistent across grade bands. Based on the panelists' recommendations, the WIDA and ATLAS team revised 147 (63.91%) of the items. No items were rejected.

Additional information regarding item development can be found in *Advancing ALTELLA: Designing and Developing Items to Advance the WIDA Alternate ACCESS for ELLs Report*.

2.3.2. Field Testing

The goal of the Alternate ACCESS Field test was to collect the data needed to select items and tasks to update Alternate ACCESS, to develop an Alternate Screener, and to develop sufficient items for the creation of at least one new test form in each of the four domains of Listening, Speaking, Reading, and Writing and in the grade-level clusters of K–2, 3–5, 6–8, and 9–12.

Due to the number of items included in the field test as well as the size of the tested student population, WIDA conducted a stand-alone field test in all WIDA states, territories, and agencies using a census-based field test administration. All WIDA Consortium members were asked to administer the field test form between two to four weeks after the operational administration of ACCESS. The testing window for the Alternate ACCESS field test was February 14–April 17, 2023. A total of 21,551 students in 40 US states, territories, and agencies participated in the Alternate ACCESS Field Test.

For the field test, five test forms were spirally distributed to all WIDA members at the SEA level. The sampling plan was developed to account for student demographic characteristics and students' average Alternate ACCESS scores across groups so that each FT form had similar test-taker numbers and characteristics. Additionally, each FT form included states with both large and small populations, and their aggregated composite scores from the prior test administration were similar across the five FT forms. The field test forms consisted of 10 listening items, 10 reading items, 8 speaking items, and 8 writing items. The total estimate of participating students per FT form by grade-level cluster was projected to be at least 1,000 students. Each spiral form included both horizontal and vertical linking items. The detailed horizontal and vertical scaling design is described in the *WIDA Alternate ACCESS Field Test Technical Brief*.

2.3.3. Item Review and Selection

Months prior to the item selection meeting, WIDA and CAL participated in several meetings to identify item selection criteria for the operational forms of Alternate ACCESS and the provisional Alternate Screener. Four areas were determined to be relevant for these criteria: item fit statistics, item difficulty, item distribution, and alignment to *a priori* targeted alternate English language proficiency levels. The selection of items and forms occurred in two steps. The first step was to pre-select potential test items and forms empirically. This was done by WIDA's psychometric team. In the second step psychometricians, and content experts from WIDA and CAL met, deliberated, and finalized item selection and forms creation. The list of experts who participated in the item selection meeting can be found in the *Alternate ACCESS Post-Field Test Review and Item Selection planning document*. The following sections describe this process.

2.3.3.1. Item selection criteria

The first step in the creation of operational Alternate ACCESS was to classify field test items into three categories: "red," "yellow," and "green," based on statistical criteria of fit statistics and raw score distributions.

- RED: If an item has an infit and/or outfit greater than 2.00.
- YELLOW: If an item's infit and/or outfit is <2.0 but ≥ 1.5 . Instances were noted where the anticipated conceptual difficulty (*a priori*) did not harmonize with the empirical difficulty of certain items. This disparity resulted in a categorization of items as falling within the "yellow" classification.
- GREEN: If an item has an infit and/or outfit less than 1.50 and meets distribution criteria.

For item selection, the green and yellow items were included for selection and all red items were excluded from the item pool.

The priorities and sequence for test form selection for the Alternate ACCESS Operational tests were as follows:

- Create one Alternate ACCESS operational test form with a sufficient distribution of item difficulties to allow for 4 cut points to be established (PL2, PL3, PL4, PL5). The priority is to match the test specification document for each domain and grade-level cluster.
- If that is not possible for one or more domain/cluster forms, at a minimum: The Listening and Reading test forms must have a total of 10 items, the widest item difficulty distribution to allow for 4 cut points to be established, and representation of all WIDA standards (NOTE: items that have combined standards, e.g., mathematics and science, can represent coverage of two standards.).
- The Speaking and Writing test forms must have a total of 8 items, a sufficient item difficulty distribution to allow for 4 cut points to be established, and representation of all WIDA standards (NOTE: items that have combined standards, e.g., mathematics and science, can represent coverage of two standards.).

Based on findings of item difficulty and student ability distributions, WIDA arrived at a determination regarding a reasonable range of difficulties to manifest incremental progression across grade-level clusters within each domain. Specific attention was given to rectifying instances where certain items exhibited significantly lower or higher difficulties than the overall range. This led to the establishment of both minimum and maximum difficulty bounds that promote a sequential elevation of difficulty across grade-level clusters.

Concerted efforts were made to pre-select items with comparable incremental difficulty intervals, typically ranging around 0.2 to 0.3 logits between two adjacent items, thereby ensuring a comprehensive coverage of difficulties while maintaining equilibrium. Furthermore, a conscious effort was made to encompass a diverse array of content language specifications within the item selection process.

2.3.3.2. Item Selection (Confirmation) Meeting

The item selection meeting was a collaborative effort involving WIDA and CAL. An exhaustive review of pre-selected items ensued, encompassing both statistical attributes and actual item content. The items were scrutinized on a domain-by-domain basis, with the intention of either confirming the item pre-selection or identifying alternatives from the available items within the “green” or “yellow” item pool. The assessment encompassed considerations of item characteristics, content relevance, and alignment with *a priori* difficulty expectations.

2.4. Standard Setting

2.4.1. Standard Setting Event, Method, and Outcomes

WIDA conducted a standard setting meeting for Alternate ACCESS between July 16 and 19, 2024, at the Doubletree by Hilton Minneapolis Airport Hotel in Bloomington, MN. This meeting resulted as an outcome of the Advancing ALTELLA grant, a Competitive Grant for State Assessments awarded by the US Department of Education. The goal of this grant was to update and revise Alternate ACCESS for ELLs, as the assessment was previously known. Major changes to Alternate ACCESS required this activity. This standard setting event was the culmination of two years of planning and preparation. The goal of this meeting was to obtain recommended proficiency level cut scores on the updated version of Alternate ACCESS. A modified Yes/No Angoff for polytomous items standard setting method was used to obtain cut score recommendations. WIDA requested that its member states provide candidates to serve as panelists to provide these recommendations. Sixty-four panelists from 31 member states, territories, and federal agencies participated in this event. Staff from WIDA, the Center for Applied Linguistics (CAL), and Accessible Teaching, Learning, and Assessment Systems (ATLAS) at the University of Kansas served as facilitators, notetakers, and support staff to guide panelists in making cut score recommendations. Additionally, WIDA asked member states and the WIDA Technical Advisory Committee (TAC) to serve as observers. Staff from eight WIDA states and two TAC members served as observers. Observers’ roles were to observe and to report their findings to WIDA’s TAC and Executive Committee. Panelists, facilitators, notetakers, and observers were placed into nine grade-level cluster and domain groups. Each group provided cut score recommendations on four domain assessments, either in Listening and Speaking or Reading and Writing, across two grade-level clusters. Standard setting facilitators, notetakers, and observers received training before the event occurred. At the meeting, a general training session occurred on the first day to orient and prepare panelists for their tasks. After the general session, panelists, facilitators, notetakers, and observers convened in grade-level cluster and domain groups. Facilitators in each group used a script to train and guide panelists through the standard setting process. The script was used to provide consistency across the nine groups. Participants at the standard setting meeting were provided with materials, resources, and guides to support their activities. After training and practice, each group went through two rounds of cut score determinations. The median score of the second round was used as groups’ final recommendations.

Panelists were asked to complete several surveys during the meeting. One asked them how well the general training session prepared them for their tasks. Overall, panelists had positive responses to the general session. After each domain was complete, panelists filled out a survey asking them to comment on the usefulness of materials, support by the facilitators, and their confidence in the recommended cut scores. Most panelists felt the materials were useful, and that facilitators and notetakers supported them. Most were confident that their cut scores represented the WIDA Alternate Proficiency Level Descriptors, and most were moderately or highly confident that their cut scores could adequately be used to make reclassification and growth decisions.

After the standard setting event, WIDA reviewed the recommendations and conducted a vertical alignment analysis. The goal of this analysis was to ensure there were no inconsistencies between grade-level clusters in cut score assignments. At the completion of the vertical alignment, WIDA shared a written summary of the standard setting event with the TAC and the WIDA Executive Committee for their review and comment. This summary included WIDA's review of the standard setting meeting, WIDA's recommended cut scores, and a summary of observers' comments. At a virtual meeting on August 12, 2024, the Executive Committee was asked to endorse the process and procedures followed to conduct the standard setting. Note that the Executive Committee was not asked to approve final cut scores, just to endorse the process followed to obtain them. The responsibility to approve final cut scores lay exclusively with WIDA. Executive Committee members present at the August 12 meeting all endorsed the process and procedures followed by WIDA. By September 3, 2024, members who were not present provided their endorsements as well. WIDA sent the final scores to DRC for them to process and provide score reports to member states, districts, and schools.

A formal report of the Alternate ACCESS standard setting has been published and is housed in the WIDA SEA Secure Portal. It details the background of why the study was conducted, the methodology used, the panelist recruitment process and criteria, the roles and responsibilities of participants, the procedures employed, the panelists' recommendations, vertical alignment procedures, and final recommendations. Final cut scores from the Alternate ACCESS standard setting are shown in Table 2.4.1.a. and 2.4.1.b.

Table 2.4.1.a.**WIDA Recommended Alternate ACCESS Domain Cut Scores by Grade-Level Cluster**

Cluster	Proficiency Level	Listening	Reading	Speaking	Writing
K-2	Level 2	937	943	941	941
K-2	Level 3	943	950	948	951
K-2	Level 4	949	957	958	960
K-2	Level 5	959	963	962	968
3-5	Level 2	940	943	946	942
3-5	Level 3	948	950	953	953
3-5	Level 4	954	957	959	960
3-5	Level 5	961	965	965	968
6-8	Level 2	943	944	946	945
6-8	Level 3	950	950	954	955
6-8	Level 4	958	957	961	963
6-8	Level 5	962	967	966	972
9-12	Level 2	945	944	946	947
9-12	Level 3	951	950	954	957
9-12	Level 4	959	957	961	965
9-12	Level 5	965	968	966	975

Table 2.4.1.b.**WIDA Recommended Alternate ACCESS Composite Cut Scores by Grade-Level Cluster**

Cluster	Proficiency Level	Oral	Literacy	Comprehension	Overall
K-2	Level 2	939	942	941	941
K-2	Level 3	946	951	948	949
K-2	Level 4	954	959	955	957
K-2	Level 5	961	966	962	964
3-5	Level 2	943	943	942	943
3-5	Level 3	951	952	949	951
3-5	Level 4	957	959	956	958
3-5	Level 5	963	967	964	965
6-8	Level 2	945	945	944	945
6-8	Level 3	952	953	950	952
6-8	Level 4	960	960	957	960
6-8	Level 5	964	970	966	968
9-12	Level 2	946	946	944	946
9-12	Level 3	953	954	950	953
9-12	Level 4	960	961	958	961
9-12	Level 5	966	972	967	970

3. Test Administration

Alternate ACCESS is a paper assessment comprised of four tests, one in each of the four language domains: Listening, Reading, Speaking, and Writing. There are four separate administrations, one for each domain. A sequence of administration is not required. Domains should be administered in whatever order logistically makes sense for the student. The target time for each domain is about 20–30 minutes, but timing may vary based on the needs of the student. All of the domain tests are scored locally by the test administrators. More detailed instructions for administering Alternate ACCESS are contained in the *Test Administrator Manual* available in the WIDA Secure Portal.

Alternate ACCESS is administered in a one-to-one setting and can easily be individualized for the student based upon their needs. Alternate ACCESS is not a timed test, therefore test administrators can spread the administration time out over the course of multiple days if needed. Test administrators read from the Test Administrator Script for each domain, therefore no additional audio is necessary. Students can utilize assistive technologies, such as augmentative and alternative communication (AAC) devices, on the assessment as these are considered the student's voice.

3.1. Test Delivery

Alternate ACCESS is typically administered between December and April of the academic year, with testing windows determined at the state level. The Listening, Reading, Speaking, and Writing domains are recommended to be administered in that order, but do not have to be, as order can be determined based on the needs of the individual student. The test may be administered in several sessions within a single day or over a series of days.

3.2. Operational Administration

3.2.1. Administering the Test

Alternate ACCESS is designed to be locally administered in a one-on-one setting in four testing sessions by trained test administrators, one session for each domain. Each domain test takes about 30 minutes to administer. However, due to the adaptive nature of the test and the individual abilities and behaviors of the students who take the assessment, actual test times can vary widely. Breaks during the test administration session are appropriate for Alternate ACCESS students, and these interruptions can increase the testing time needed for test administration.

For more detailed information regarding stopping rules and scoring, please see Section 4 Scoring.

3.2.2. Training and Resources for Districts and Schools

Before, during, and after a state's testing window, educators take on various roles to ensure all tasks are carried out for successful test administration. These roles include test coordinators at the district and school level, and test administrators at the school level. The test administrator administers and scores the test and is responsible for managing student data prior to, during, and after testing. The *Test Administrator Manual* and the *District and School Test Coordinator Manual* were developed to contain all the information related to responsibilities and required training for the various roles.

A training course, which is housed in the WIDA Secure Portal, provides educators training to become certified to administer Alternate ACCESS. Additional materials and resources to assist administrators and coordinators before, during, and after a state's testing window are also found there. Training courses include test preparation and administration tutorials and online administration quizzes. Proper training and familiarity with Alternate ACCESS administration requirements is key to the validity of the test and the appropriate interpretations of Alternate ACCESS test scores.

3.2.3. Test Security

WIDA makes efforts to keep the test secure at all levels of development and administration. WIDA, CAL, and DRC (the entity responsible for printing, distributing, collecting, and scoring Alternate ACCESS) follow established policies and procedures regarding the security of the test, and every individual involved in the administration of Alternate ACCESS, from the district level to the classroom level, is trained in issues of test security.

All materials for Alternate ACCESS are considered secure test materials. All users of the WIDA Secure Portal are prompted to read and sign a Nondisclosure and User Agreement upon their first login. Use of the WIDA Assessment Management System (WIDA AMS) is also subject to the terms of use outlined there. Users are prompted to agree with the test security policy upon their first login. The security of all test materials must be maintained before, during, and after the test administration. Under no circumstances are students permitted to handle secure materials before or after test administration. Test materials should never be left unsecured. The test coordinator should track each secure booklet on the Alternate ACCESS Security Checklist. Individuals are responsible for the secure documents assigned to them. Secure documents should never be destroyed (e.g., shredded, thrown in the trash) except for soiled documents, which must be destroyed in a secure manner. District and school personnel carrying out their roles in the delivery of this assessment must follow guidelines noted in the *ACCESS for ELLs District and School Test Coordinator Manual* to maintain test security.

Test security policies are stated in the *Test Policy Handbook for State Education Agencies* and the Memorandum of Understanding (MOU)s with each state.

3.3. *Fairness and Accessibility*

WIDA is committed to providing an assessment that is accessible to every eligible learner, including those with the most significant cognitive disabilities. WIDA's approach to accessibility and inclusion incorporates Universal Design principles that increase access through design elements like embedded scaffolding and task modeling. Additional administrative considerations and accommodations allow further flexibility in the administration of Alternate ACCESS, to best suit the needs of the student being assessed.

3.3.1. Fairness and Accessibility Design

As part of the Universal Design of the Alternate ACCESS, test administrators offer students multiple opportunities within each of the domains to respond to a task. With each opportunity, test administrators offer additional scaffolding and support. In Listening and Reading, the task is broken down into Cues (Cue A, Cue B, Cue C) and in Speaking and Writing, the task is broken down into Questions (Question 1, Question 2, Question 3). See Section 4 for more information on administration and scoring of each domain.

Careful design consideration was taken when incorporating kindergarten into the grades 1–2 form. Within Reading and Writing, additional modeling and support were incorporated to support early literacy. Test administrators demonstrate to students how to respond to questions by modeling a partial response, allowing the student to respond and if necessary, support them in the completion of the task. As a student progresses towards higher proficiency levels, these supports are gradually reduced.

Alternate ACCESS uses graphic support in all domains; however, item writers paid careful attention to the scripting so that students with visual impairments are still able to access the content of the item.

Alternate ACCESS is administered in a one-to-one setting and can easily be individualized for the student based upon their needs. Alternate ACCESS is not a timed test, therefore test administrators can spread the administration time out over the course of multiple days if needed. Test administrators read from the Test Administrator Script for each domain, therefore no additional audio is necessary. Students can utilize assistive technologies, such as argumentative and alternative communication (AAC) devices; on the assessment as these are considered the student's voice.

4. Scoring

Alternate ACCESS items are scored by the test administrator in a student response booklet. Test administrators complete training created by WIDA for Alternate ACCESS-specific administration and scoring certification as per their state guidelines. This training course also includes student speaking and writing samples for speaking and writing, allowing the test administrator opportunities to practice scoring items.

Each domain test includes stopping criteria where test administrators can stop each domain test when a student offers no response, an incorrect response, or an Approaches response on three consecutive tasks. These are marked in the student response booklet. For both the Speaking and Writing tests, there is space provided in the Student Response Booklet for test administrators to transcribe student responses. Transcription is optional and provides TAs an opportunity to reflect on scoring practice, and to look holistically at scoring across domains.

After testing has been completed, all materials are sent back to DRC for final processing and score computing.

4.1. Listening and Reading

To administer an item, the test administrator reads the Cue A script (initial prompt and question of the task). If the student does not respond, the test administrator must repeat Cue A again, as indicated in the Test Administrator's Script. If the student answers incorrectly or does not respond to Cue A, the test administrator will read the Cue B script. Cue B simplifies the initial prompt and asks the question again. If the student responds incorrectly, or does not respond at all after the test administrator reads Cue B, the test administrator will administer Cue C. This cue provides the answer to the question, restates the prompt, and asks the question again. Test administrators should score the item as "Correct" under the appropriate Cue where the student responds accurately, "Incorrect" after administering all three questions but the student responds incorrectly, "No Response" when the student provides no response, or "Not Administered" if the test was stopped without administering the task.

- Correct Cue A = 4 score points
- Correct Cue B = 3 score points
- Correct Cue C = 2 score points
- Incorrect = 1 score point
- No Response and Not administered = 0 score points

4.2. *Writing*

Students have up to six opportunities (Questions 1, 2, and 3) to provide a “Meets” response to each Writing task. If the student demonstrates the “Meets” scoring criteria at any point during a task administration, the test administrator follows the script in the Moving On box to continue to the next task. Test administrators should score the item as “Approaches” after administering all three questions and when the student provides a response but there is clear evidence that the demands of the task are beyond the student’s current linguistic abilities, “No Response” when the student provides no written response in English, or “Not Administered” if the test was stopped without administering the task.

- Meets Question 1 = 4 score points
- Meets Question 2 = 3 score points
- Meets Question 3 = 2 score points
- Approaches = 1 score point
- No Response and Not administered = 0 score points

4.3. *Speaking*

A speaking task consists of three questions. Question 1 for each task is aligned to an alternate English language proficiency level, and that proficiency level’s expectation is presented in the Student Response Booklet. The administrator’s task is to compare the student’s response with the expected response. Should a student respond incorrectly or not at all, Questions 2 and 3 offer additional scaffolding and support. Students have up to six opportunities (Questions 1, 2, and 3) to provide a “Meets” response to each Speaking task. If the student demonstrates the Meets scoring criteria at any point during a task administration, the test administrator follows the script in the Moving On box to continue to the next task. Test administrators should score the item as “Approaches” after administering all three Questions and when the student provides a response but there is clear evidence that the demands of the task are beyond the student’s current linguistic abilities, “No Response” when the student provides no spoken response in English, or “Not Administered” if the test was stopped without administering the task.

- Meets Question 1 = 4 score points
- Meets Question 2 = 3 score points
- Meets Question 3 = 2 score points
- Approaches = 1 score point
- No Response and Not administered = 0 score points

4.4. Scaling

Given the substantial differences of psychometric properties between the new WIDA Alternate ACCESS assessment and the previous Alternate ACCESS for ELLs assessment, WIDA chose to establish a new scale score range for the 602 Alternate ACCESS scores. To represent changes in test length and student ability across grade-level clusters, WIDA extended the existing scale score range from 910–960 to 900–980. The extended scale allows student growth across grade-level clusters with more discrimination, which is also reflected in the new vertical scales and standard setting. The new scale scores cannot be directly compared to scales scores from prior test administrations. WIDA has provided guidance documents and correspondence tables for reference and also offered technical assistance for states to support their accountability and reclassification needs.

Scaling is the process of developing a reporting scale to make the scores on a test more usable to educators. Scale scores are calculated by transforming the student ability estimate via a scaling equation per domain.

WIDA utilized Kolen & Brennan (2014)'s mean and standard deviation linear transformation to establish a new scale range. This procedure linearly transforms raw scores to scale scores when the mean and standard deviation (SD) of the scale score are pre-specified. In this approach, the transformation was given as:

$$\frac{\sigma(sc)}{\sigma(\theta)}\theta + \left[\mu(sc) - \frac{\sigma(sc)}{\sigma(\theta)}\mu(\theta) \right]$$

where $\mu(\theta)$ and $\sigma(\theta)$ represent the mean and SD of estimated abilities for the norm group and $\mu(sc)$ and $\sigma(sc)$ denote the target mean and SD for the scaled score.

The following steps were applied:

1. Transform the ability estimates obtained into scaled scores using a linear conversion using the equation above. Here, $\mu(\theta)$ and $\sigma(\theta)$ represent the mean and SD of estimated abilities for cluster 3–5, while $\mu(sc)$ and $\sigma(sc)$ denote the target mean and SD for the scaled score.
2. Obtain slope and intercept values by plugging various combinations of $\mu(sc)$ (930–950) and $\sigma(sc)$ (10–20) so that the equation above is simplified into:
$$SC_{domain} = A_{domain} * \theta + B_{domain}$$
3. Convert ability estimates to scaled scores by using slopes and intercept above for all students at all clusters.
4. Generate plots and spreadsheets by cluster to evaluate the scale score ranges and distribution.
5. Compute each composite score using the selected scale per domain.
6. Reevaluate the scale to determine ideal scaling constants per domain.

Note that the center of 900–980 is 940. We tried to keep the center at 940 but some adjustment of mean scale scores (930–950) was attempted to fit the scale into the 900–980 range. $\sigma(sc)$ values are feasible values after $\mu(sc)$ was chosen to fit in the range of 900–980.

The following were the criteria for scaling:

- Final scale transformation constants will produce parallel-like lines among the individual domains for student ability to scale score conversion to meet the WIDA Board's requirement of scale score change per unit theta.
- Scale scores should fall within the 900–980 range per domain and composite with incremental lower and upper limits across each cluster.
- Scale scores for all correct and all incorrect theta/raw scores are allowed to go beyond this range and are truncated to fit the lower and upper limits.
- Aim for the highest possible $\sigma(\text{SC}=\text{Scale Score})$ for a given domain while staying within the 900–980 range to minimize clumped scale scores in the middle of the distribution but allow as many as unique scale scores and reflect the growth of student ability with the scale score.
- Aim for overall or Cluster 35 $\mu(\text{SC})$ for a given domain to be close to 940 and/or same $\mu(\text{SC})$ whenever possible.

The following scaling transformation constants are applied to each student ability to derive scale scores to each domain:

- Listening: (Ability Measure in Logits*7.948) + 942.606
- Reading: (Ability Measure in Logits*7.495) + 940.879
- Writing: (Ability Measure in Logits*7.297) + 943.625
- Speaking: (Ability Measure in Logits*7.678) + 941.392

5. Summary of Score Reports

5.1. Individual Student Reports

Alternate ACCESS provides two types of Individual Student Reports for each student. The Individual Student Report for Families is a detailed report of a single student's performance, including proficiency level for each language domain and an overall proficiency level. This report should be shared with parents/guardians as part of discussions around student progress and achievement. This report can be translated into 48 different languages and made available in an online portal for districts to print as necessary. An excerpt from the Individual Student Report for Families is included in Figure 5.1.a. below.

Figure 5.1.a.

Individual Student Report for Families Excerpt

ANTHONY's Individual Domain Score		Alternate English Language Proficiency Levels (PL)				
		1 Entering	2 Emerging	3 Developing	4 Expanding	5 Bridging
Listening	2					
Speaking	4					
Reading	1					
Writing	4					
ANTHONY's Overall Proficiency Level*	2					

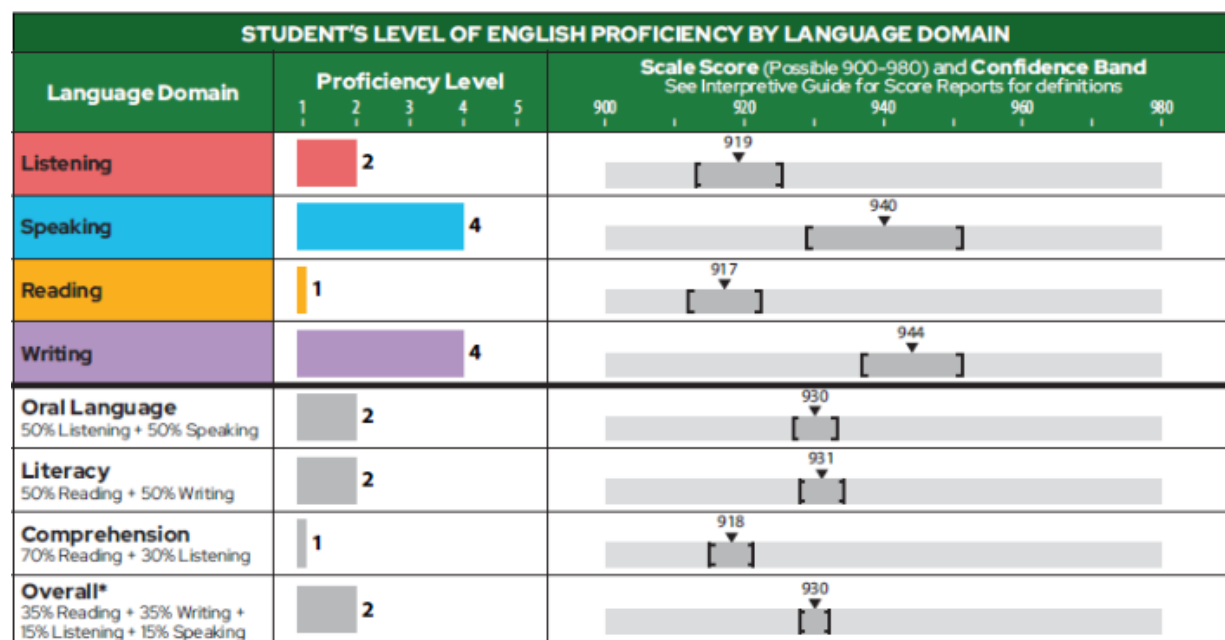
*Overall score, as shown, in the last row in the table above, is calculated only when all four domains have been assessed.

NA: Not available

The Individual Student Report for Educators is a detailed report of a single student's performance, including proficiency level and scale scores for each language domain and four composite areas. Additionally, this report shares information reported on the Individual Characteristics Questionnaire that can be used to inform conversations around reclassification. This report should be shared with the student's teachers to inform individualized classroom instruction and assessment, as well as with IEP teams when determining the student's abilities and English language needs. An excerpt from the Individual Student Report for Educators is included in Figure 5.1.b.

Figure 5.1.b.

Individual Student Report for Educators Excerpt



*Overall score, as shown, in the last row in the table above, is calculated only when all four domains have been assessed.

NA: Not available

5.2. Other Reports

The Student Roster Report provides an overview of the performances of a group of students, including proficiency level and scale scores for each language domain and composite area by school, grade, student, and grade-level cluster. It should be shared with administrators, teachers, and IEP teams to inform classroom instruction and assessment.

Additionally, Frequency Reports are made available for a single grade within a school, district, or state including the number and percentage of tested students that achieved each proficiency level for each language domain and composite area.

6. Annual Test Results

This section provides an overview of students' participation, along with the distribution of raw scores, scale scores, and proficiency levels for the Alternate ACCESS 602 administration. Results are presented through tables and figures summarizing student participation, scale scores, and proficiency levels, which are further subdivided by various demographics, including grade-level cluster, grade, state, domain, composite scores, gender, ethnicity/race, and primary disabilities. In the 2023–2024 operational administration of Alternate ACCESS, 42 WIDA Consortium states/territories participated, with a total of 32,850 students completing the 602 Alternate ACCESS tests as of December 2024.

Following the [approach of the U.S. Census Bureau](#), ethnicity is used as a binary category (Hispanic or non-Hispanic), with five categories for race (American Indian/Alaskan Native, Asian, Black/African American, Pacific Islander/Hawaiian, and White) that are not mutually exclusive. Thus, for example, Student A may be labeled as Hispanic for ethnicity and Asian for race, while Student B may be labeled as non-Hispanic for ethnicity and both American Indian/Alaskan Native and Black/African American for race. Students who are labeled Hispanic are included in the Hispanic (of any race) category, regardless of how many racial categories they are included in. Students who are identified in one racial category (e.g., Asian) who have not been identified as Hispanic are identified in only one racial category; if they are identified in more than one racial category and have not been identified as Hispanic, they are labeled non-Hispanic multiracial.

Regarding disability types, students are categorized based on both primary and secondary disabilities according to [IDEA](#), which include Autism Spectrum Disorder (AS), Deaf-blindness (DB), Developmental Delay (DD), Hearing Impairment, including Deafness (HI), Infant/Toddler with a Disability (ITD), Intellectual Disability (ID), Multiple Disability (MD), Orthopedic Impairment (OI), Other Health Impairment (OHI), Serious Emotional Disability (SED), Specific Learning Disability (SLD), Speech or Language Impairment (SLI), Traumatic Brain Injury (TBI), and Visual Impairment, including Blindness (VI). When students do not report their disability type, they are marked as having “No Primary Disability recorded” (NPD) or “No Secondary Disability recorded” (SPD).

6.1. Students Excluded from Analysis

In some circumstances there was a mismatch between a student’s reported grade and the grade-level cluster (i.e., K–2, 3–5, 6–8, or 9–12) actually administered (e.g., a student reported to be in grade 1 who was administered a test intended for students in the 3–5 grade-level cluster). Thirty-eight students were administered a test form not intended for their grade-level cluster. See Table 6.1.1. for a breakdown of the incorrect test forms assigned by grade. The data from these 38 students were eliminated from all subsequent analyses in this report.

6.1.1. Out-of-Grade Level Test Administration

Table 6.1.1.

Out-of-Grade-Level Test Administrations

Grade	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12	Total
0		0	0	0	0
1		0	0	0	0
2		9	0	0	9
3	9		0	0	9
4	1		0	0	1
5	0		2	1	3
6	0	2		0	2
7	0	1		2	3
8	0	2		6	8
9	0	0	2		2
10	0	0	0		0
11	0	0	0		0
12	0	1	0		1
Total	10	15	4	9	38

6.2. *Participation by Grade-Level Cluster*

Section 6.2 provides a breakdown of participation by grade-level cluster based on state, gender, and ethnicity. For each of the 42 WIDA states that participated in the 2023–2024 operational testing program, Table 6.2.1. details the number of test takers by grade-level cluster, along with total counts by state (final column) and grade-level cluster across all states (final row). Table 6.2.2. shows the distribution of test takers by gender (Female, Male, or Missing) for each grade-level cluster. Similarly, Table 6.2.3. presents a breakdown by ethnicity for each grade-level cluster.

Table 6.2.1. summarizes participation across the 42 WIDA states, territories, and agencies that took part in the Alternate ACCESS operational testing program in 2023–2024, organized by grade-level cluster. The 42 rows represent the number of students in each grade-level cluster by state, while the final row displays the total number of participants across all 42 states and U.S. territories. Illinois had the highest number of students (5,911), followed by Virginia (2,881). The state/territory with the fewest participants was the Virgin Islands (2 students), followed by the Northern Mariana Islands (6 students). The largest grade-level cluster was kindergarten to grade 2, with a total of 9,858 participants. The non-state abbreviations used are as follows: DC - District of Columbia; DD - Department of Defense Education Activity; MP - Northern Mariana Islands; BI - Bureau of Indian Education; and VI - Virgin Islands.

6.2.1. Participation by Grade-Level Cluster by State

Table 6.2.1.

Participation by Grade-Level Cluster by State

State	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12	Total
AK	15	14	33	50	112
AL	81	103	72	68	324
BI	7	10	6	11	34
CO	208	258	199	189	854
DC	42	41	37	29	149
DD	17	14	7	6	44
DE	8	5	4	6	23
FL	352	350	195	126	1,023
GA	295	395	315	284	1,289
HI	91	63	67	88	309
ID	19	50	46	44	159
IL	1,702	1,463	1,144	1,602	5,911
IN	334	332	283	487	1,436
KS	122	61	42	30	255
KY	211	133	78	102	524
MA	670	544	347	394	1,955
MD	209	230	171	174	784
ME	30	21	14	17	82
MI	340	311	210	232	1,093
MN	573	397	233	278	1,481
MO	98	66	46	57	267
MP	1	2	3	0	6
MT	4	9	7	1	21
NC	333	366	351	373	1,423
ND	12	5	6	8	31
NH	21	9	9	14	53
NJ	397	213	123	81	814
NM	95	131	98	97	421
NV	237	251	244	315	1,047
OK	222	259	218	190	889
PA	686	517	338	401	1,942
RI	80	64	53	65	262
SC	195	146	112	100	553
SD	14	17	18	17	66

State	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12	Total
TN	162	138	102	103	505
UT	142	196	184	178	700
VA	965	665	509	742	2,881
VI	0	2	0	0	2
VT	19	11	3	7	40
WA	759	615	441	689	2,504
WI	89	127	148	201	565
WY	1	4	5	7	17
Total	9,858	8,608	6,521	7,863	32,850

6.2.2. Participation by Grade-Level Cluster by Gender

Table 6.2.2 shows participation by grade-level cluster by gender across all 42 WIDA member states, territories, and agencies. The gender ratio was generally 40% female, 45% male, and 15% were missing gender information.

Table 6.2.2.

Participation by Grade-Level Cluster by Gender

Cluster	Female Count	Female % within Cluster	Male Count	Male % within Cluster	Missing Count	Missing % within Cluster	Total
K-2	2,340	23.74%	6,015	61.02%	1,503	15.25%	9,858
3-5	2,298	26.7%	5,055	58.72%	1,255	14.58%	8,608
6-9	1,955	29.98%	3,570	54.75%	996	15.27%	6,521
9-12	2,324	29.56%	4,098	52.12%	1,441	18.33%	7,863
Total	8917	27.14%	18738	57.04%	5195	15.81%	32,850

6.2.3. Participation by Grade-Level Cluster by Ethnicity

Table 6.2.3. shows participation by grade-level cluster by ethnicity across all 42 WIDA member states, territories, and agencies. About 64–67% of participants were Hispanic across all clusters.

Table 6.2.3.**Participation by Grade-Level Cluster by Ethnicity**

Ethnicity	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12	Total
Hispanic (of any race) Count	5,279	4,970	4,055	4,930	19,234
Hispanic (of any race) % within Cluster	53.55%	57.74%	62.18%	62.70%	58.55%
Non-Hispanic American Indian Count	55	58	37	54	204
Non-Hispanic American Indian % within Cluster	0.56%	0.67%	0.57%	0.69%	0.62%
Non-Hispanic Asian Count	1,519	1,293	828	974	4,614
Non-Hispanic Asian % within Cluster	15.41%	15.02%	12.70%	12.39%	14.05%
Non-Hispanic Black Count	1,081	707	424	565	2,777
Non-Hispanic Black % within Cluster	10.97%	8.21%	6.50%	7.19%	8.45%
Non-Hispanic Multiracial Count	61	58	38	44	201
Non-Hispanic Multiracial % within Cluster	0.62%	0.67%	0.58%	0.56%	0.61%
Non-Hispanic Pacific Islander Count	98	70	65	72	305
Non-Hispanic Pacific Islander % within Cluster	0.99%	0.81%	1.00%	0.92%	0.93%
Non-Hispanic White Count	850	784	594	625	2,853
Non-Hispanic White % within Cluster	8.62%	9.11%	9.11%	7.95%	8.68%
Missing Count	915	668	480	599	2,662
Missing % within Cluster	9.28%	7.76%	7.36%	7.62%	8.10%
Total Count	9,858	8,608	6,521	7,863	32,850

6.3. Participation by Grade**6.3.1. Participation by Grade by State**

Section 6.3 expands on the information presented in Section 6.2 by breaking down the distribution of test-takers by individual grades (kindergarten through grade 12) rather than by grade-level clusters. Table 6.3.1. details the distribution of test-takers by grade for each state, while Table 6.3.2. presents the gender distribution for each grade. Section 6.3.3. provides a breakdown of test-takers by ethnicity across grades.

The number of students per grade ranges from 1,626 to 3,774, with the highest number of students in grade 1 and the lowest in grade 11, as shown in Table 6.3.1.

Table 6.3.1.**Participation by Grade by State**

State	Grade K	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12	Total
AK	4	1	10	3	4	7	10	16	7	14	14	8	14	112
AL	17	23	41	34	37	32	30	24	18	23	13	20	12	324
BI	1	3	3	3	4	3	2	2	2	4	4	2	1	34
CO	54	79	75	84	80	94	71	64	64	58	43	45	43	854
DC	12	15	15	16	11	14	8	19	10	8	3	8	10	149
DD	6	5	6	4	5	5	3	3	1	0	4	2	0	44
DE	1	5	2	2	1	2	0	3	1	2	2	0	2	23
FL	62	146	144	113	147	90	71	71	53	34	33	31	28	1,023
GA	58	110	127	131	136	128	125	106	84	88	64	65	67	1,289
HI	22	33	36	21	21	21	22	22	23	10	19	22	37	309
ID	6	7	6	20	11	19	17	13	16	9	16	8	11	159
IL	451	660	591	519	510	434	413	364	367	329	318	302	653	5,911
IN	107	118	109	113	106	113	101	107	75	121	112	88	166	1,436
KS	54	42	26	20	18	23	18	14	10	11	4	8	7	255
KY	67	76	68	42	51	40	30	23	25	33	27	18	24	524
MA	212	241	217	220	160	164	114	107	126	100	111	110	73	1,955
MD	51	79	79	75	77	78	61	54	56	45	41	49	39	784
ME	14	9	7	8	9	4	9	1	4	0	4	6	7	82
MI	113	118	109	106	101	104	87	66	57	75	52	55	50	1,093
MN	146	228	199	123	145	129	80	76	77	66	62	56	94	1,481
MO	33	37	28	22	23	21	19	7	20	19	14	5	19	267
MP	0	0	1	1	0	1	3	0	0	0	0	0	0	6
MT	0	0	4	2	3	4	5	1	1	1	0	0	0	21
NC	45	137	151	127	110	129	103	119	129	101	80	71	121	1,423
ND	5	5	2	1	2	2	4	2	0	1	2	1	4	31
NH	4	9	8	1	6	2	2	4	3	6	3	1	4	53
NJ	105	180	112	98	69	46	52	44	27	21	23	25	12	814
NM	25	38	32	41	41	49	36	30	32	19	32	19	27	421
NV	75	84	78	76	103	72	80	81	83	76	81	73	85	1,047
OK	65	76	81	82	107	70	77	68	73	58	56	37	39	889
PA	183	257	246	202	170	145	137	103	98	119	66	76	140	1,942
RI	17	35	28	24	12	28	15	16	22	15	14	15	21	262
SC	58	74	63	55	52	39	45	36	31	14	25	24	37	553
SD	3	5	6	4	8	5	9	4	5	5	2	3	7	66
TN	32	70	60	53	52	33	39	40	23	26	33	20	24	505
UT	36	61	45	66	69	61	62	64	58	45	49	43	41	700

State	Grade K	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12	Total
VA	270	378	317	216	243	206	198	153	158	156	152	140	294	2,881
VI	0	0	0	0	2	0	0	0	0	0	0	0	0	2
VT	5	8	6	3	4	4	3	0	0	2	1	2	2	40
WA	224	283	252	230	201	184	182	139	120	153	136	131	269	2,504
WI	14	39	36	40	52	35	52	54	42	43	50	34	74	565
WY	0	0	1	2	0	2	3	2	0	1	1	3	2	17
Total	2,657	3,774	3,427	3,003	2,963	2,642	2,398	2,122	2,001	1,911	1,766	1,626	2,560	32,850

6.3.2 Participation by Grade by Gender

Table 6.3.2. reveals that the proportion of female students across all grades ranges from approximately 22% to 31%, while about 16% of students did not report their gender.

Table 6.3.2.

Participation by Grade by Gender

Grade	Female Count	Female % within Grade	Male Count	Male % within Grade	Missing Count	Missing % within Grade	Total
K	607	22.85%	1,644	61.87%	406	15.28%	2,657
1	877	23.24%	2,318	61.42%	579	15.34%	3,774
2	856	24.98%	2,053	59.91%	518	15.12%	3,427
3	765	25.47%	1,790	59.61%	448	14.92%	3,003
4	829	27.98%	1,731	58.42%	403	13.6%	2,963
5	704	26.65%	1,534	58.06%	404	15.29%	2,642
6	693	28.9%	1,343	56.01%	362	15.1%	2,398
7	659	31.06%	1,140	53.72%	323	15.22%	2,122
8	603	30.13%	1,087	54.32%	311	15.54%	2,001
9	596	31.19%	977	51.13%	338	17.69%	1,911
10	541	30.63%	939	53.17%	286	16.19%	1,766
11	477	29.34%	885	54.43%	264	16.24%	1,626
12	710	27.73%	1,297	50.66%	553	21.6%	2,560
Total	8,917	27.14%	18,738	57.04%	5,195	15.81%	32,850

6.3.3. Participation by Grade by Ethnicity

As displayed in Tables 6.3.3.a. and 6.3.3.b., the proportion of Hispanic students increases progressively from kindergarten to grade 12, peaking in grade 11 at 64.21%. The second-largest group is the Asian population, accounting for about 13% of the total student population, with a peak in grade 2 (16%). This is followed by Black and White students, each representing approximately 8% of the total population. The proportion of Black students peaks in kindergarten (13.17%), while the proportion of White students reaches its highest in grade 4 (9.21%).

Table 6.3.3.a.

Participation by Grade by Ethnicity: Kindergarten through Grade 6

Ethnicity	K	1	2	3	4	5	6
Hispanic (of any Race) Count	1,327	2,058	1,894	1,706	1,714	1,550	1,453
Hispanic (of any Race) % within Grade	49.94%	54.53%	55.27%	56.81%	57.85%	58.67%	60.59%
Non-Hispanic American Indian Count	19	26	10	15	20	23	16
Non-Hispanic American Indian % within Grade	0.72%	0.69%	0.29%	0.50%	0.67%	0.87%	0.67%
Non-Hispanic Asian Count	398	575	546	448	440	405	316
Non-Hispanic Asian % within Grade	14.98%	15.24%	15.93%	14.92%	14.85%	15.33%	13.18%
Non-Hispanic Black Count	350	387	344	248	263	196	156
Non-Hispanic Black % within Grade	13.17%	10.25%	10.04%	8.26%	8.88%	7.42%	6.51%
Non-Hispanic Multiracial Count	22	20	19	22	17	19	11
Non-Hispanic Multiracial % within Grade	0.83%	0.53%	0.55%	0.73%	0.57%	0.72%	0.46%
Non-Hispanic Pacific Islander Count	19	45	34	20	27	23	22
Non-Hispanic Pacific Islander % within Grade	0.72%	1.19%	0.99%	0.67%	0.91%	0.87%	0.92%
Non-Hispanic White Count	227	343	280	276	273	235	237
Non-Hispanic White % within Grade	8.54%	9.09%	8.17%	9.19%	9.21%	8.89%	9.88%
Missing Count	295	320	300	268	209	191	187
Missing % within Grade	11.10%	8.48%	8.75%	8.92%	7.05%	7.23%	7.80%
Total Count	2,657	3,774	3,427	3,003	2,963	2,642	2,398

Table 6.3.3.b.**Participation by Grade by Ethnicity: Grade 7 through 12**

Ethnicity	7	8	9	10	11	12	Total
Hispanic (of any Race) Count	1,348	1,254	1,177	1,130	1,044	1,579	19,234
Hispanic (of any Race) % within Grade	63.52%	62.67%	61.59%	63.99%	64.21%	61.68%	58.55%
Non-Hispanic American Indian Count	14	7	14	18	9	13	204
Non-Hispanic American Indian % within Grade	0.66%	0.35%	0.73%	1.02%	0.55%	0.51%	0.62%
Non-Hispanic Asian Count	266	246	233	195	194	352	4,614
Non-Hispanic Asian % within Grade	12.54%	12.29%	12.19%	11.04%	11.93%	13.75%	14.05%
Non-Hispanic Black Count	134	134	127	129	119	190	2,777
Non-Hispanic Black % within Grade	6.31%	6.70%	6.65%	7.30%	7.32%	7.42%	8.45%
Non-Hispanic Multiracial Count	7	20	12	13	9	10	201
Non-Hispanic Multiracial % within Grade	0.33%	1.00%	0.63%	0.74%	0.55%	0.39%	0.61%
Non-Hispanic Pacific Islander Count	20	23	15	14	13	30	305
Non-Hispanic Pacific Islander % within Grade	0.94%	1.15%	0.78%	0.79%	0.80%	1.17%	0.93%
Non-Hispanic White Count	174	183	166	151	117	191	2,853
Non-Hispanic White % within Grade	8.20%	9.15%	8.69%	8.55%	7.20%	7.46%	8.68%
Missing Count	159	134	167	116	121	195	2,662
Missing % within Grade	7.49%	6.70%	8.74%	6.57%	7.44%	7.62%	8.10%
Total Count	2,122	2,001	1,911	1,766	1,626	2,560	32,850

6.4. Participation by Domain

Section 6.4 provides a breakdown of test taker counts by domain (Listening, Reading, Speaking, and Writing).

6.4.1. Participation by Grade-Level Cluster by Domain

Table 6.4.1 summarizes the distribution by grade-level cluster and domain.

Table 6.4.1.**Participation by Grade-Level Cluster by Domain**

Cluster	Listening	Reading	Speaking	Writing
K-2	9,858	9,855	9,842	9,850
3-5	8,607	8,601	8,600	8,594
6-9	6,521	6,521	6,511	6,513
9-12	7,863	7,860	7,853	7,854
Total	32,849	32,837	32,806	32,811

6.4.2. Participation by Grade by Domain

Table 6.4.2. summarizes the participation distribution by grade.

Table 6.4.2.

Participation by Grade by Domain

Grade	Listening	Reading	Speaking	Writing
K	2,657	2,656	2,651	2,654
1	3,774	3,772	3,767	3,770
2	3,427	3,427	3,424	3,426
3	3,002	2,999	3,001	2,995
4	2,963	2,963	2,959	2,959
5	2,642	2,639	2,640	2,640
6	2,398	2,398	2,394	2,394
7	2,122	2,122	2,119	2,120
8	2,001	2,001	1,998	1,999
9	1,911	1,910	1,908	1,908
10	1,766	1,766	1,,764	1,766
11	1,626	1,626	1626	1,626
12	2,560	2,558	2,555	2,554
Total	32,849	32,837	32,806	32,811

6.5. *Participation by Disability*

Section 6.5. presents the distribution of participants by disability type, both overall and by grade-level cluster. The tables include rows representing primary disabilities and columns indicating secondary disabilities. To aid interpretation, Table 6.5.1 provides a list of acronyms for each disability category.

Table 6.5.1.a. displays the distribution of test-takers across 15 primary and secondary disability categories. Among primary disabilities, Autism Spectrum Disorder (AS) accounts for the largest proportion (12,985; 39.5%), followed by Intellectual Disability (ID) with 9,309 students (28.3%) across all clusters. Other groups comprising more than 5% of test takers include Multiple Disabilities (MD), Developmental Delay (DD), and Other Health Impairments (OHI).

6.5.1. Participation by Primary and Secondary Disability

Table 6.5.1.a.

Participation by Primary Disability (across rows) and Secondary Disability (across columns)

Disa- bility	AD	DB	DD	ED	HI	ID	MD	OHI	OI	SLD	SLI	TBI	VI	NSD	Total
AS	17	6	149	14	26	871	116	186	11	90	2,567	4	25	8,903	12,985
DB	2	0	2	0	1	6	2	2	0	0	2	0	0	9	26
DD	87	2	2	1	22	47	33	37	5	19	539	2	12	1,768	2,576
ED	1	0	1	1	0	8	0	2	1	2	8	0	0	24	48
HI	5	0	1	0	0	22	3	2	2	0	13	1	1	39	89
ID	431	7	30	34	104	14	138	461	104	68	1,858	9	84	5,967	9,309
MD	99	8	20	4	41	220	140	84	39	7	271	3	91	1,619	2,646
OHI	34	2	36	2	20	120	37	7	19	16	240	1	39	908	1,481
OI	2	0	3	0	2	10	4	7	1	3	32	0	5	51	120
SLD	5	0	1	1	2	4	2	5	1	1	95	0	1	278	396
SLI	9	0	16	0	3	8	3	9	1	7	3	0	1	201	261
TBI	2	1	0	0	1	11	3	0	2	2	12	0	6	54	94
VI	5	0	4	2	0	13	5	5	1	0	5	0	0	9	49
NPD	7	0	3	0	0	2	2	0	0	1	5	0	3	2,747	2,770
Total	706	26	268	59	222	1,356	488	807	187	216	5,650	20	268	22,577	32,850

Table 6.5.1.b.

Acronyms of Disabilities

Acronym	Category Name
NPD	No Primary Disability Recorded
NSD	No Secondary Disability Recorded
AS	Autism Spectrum Disorder
DB	Deaf-blindness
DD	Developmental Delay
HI	Hearing Impairment, including Deafness
ID	Intellectual Disability
MD	Multiple Disability
OI	Orthopedic Impairment
OHI	Other Health Impairment
SED	Serious Emotional Disability
SLD	Specific Learning Disability
SLI	Speech or Language Impairment
TBI	Traumatic Brain Injury
VI	Visual Impairment, including Blindness

6.5.2. Participation by Primary Disability by Grade-Level Cluster

Section 6.5.2. further examines the distribution of test-takers across the 15 primary disability categories by grade-level cluster. Autism represents the largest group in the K–2 cluster (5,254 students), followed by Developmental Delay (1,946 students). Similarly, Autism remains the largest group in the 3–5 cluster (3,789 students), followed by Intellectual Disability (2,421 students). In contrast, Intellectual Disability becomes the largest group in both the 6–8 and 9–12 clusters, followed by Autism. Students with Multiple Disabilities (MD) and Other Health Impairments (OHI) range between 300 and 800 students across all clusters.

Table 6.5.2.

Participation by Primary Disability (across rows) and by Grade-Level Clusters (across columns)

Primary Disability Code	Grade-Level Cluster K–2	Grade-Level Cluster 3–5	Grade-Level Cluster 6–8	Grade-Level Cluster 9–12	Total
AS	5,254	3,789	2,097	1,845	12,985
DB	7	6	5	8	26
DD	1,946	374	87	169	2,576
ED	6	14	9	19	48
HI	22	21	18	28	89
ID	902	2,421	2,631	3,355	9,309
MD	449	703	647	847	2,646
OHI	402	470	305	304	1481
OI	27	34	25	34	120
SLD	65	114	108	109	396
SLI	110	85	38	28	261
TBI	14	17	24	39	94
VI	9	13	11	16	49
NPD	645	547	516	1,062	2,770

6.6. *Scale Scores by Domain and Composite*

This section provides information on students' scale score results. Section 6.6. presents the mean scale scores by grade-level cluster for the eight evaluated components, starting with the four domains (Listening, Reading, Writing, and Speaking), followed by the four composites (Oral Language, Literacy, Comprehension, and Overall Composite). Given that Alternate ACCESS Series 602 uses the same score range across grade-level clusters, mean scale scores are generally expected to increase with grade level.

Tables 6.6.1.a. through 6.6.1.d. display the scale scores across grades. Overall, the domain and composite scores show an upward trend across grades, with the exception of grades 11 and 12, where scores are slightly lower compared to the preceding grades.

6.6.1. Mean Scale Scores by Domain and Composite

Table 6.6.1.a.

Mean Scale Scores: K–2 [Grade = G, Cluster = C]

Domain	G K Mean	G K SD	G K N	G 1 Mean	G 1 SD	G 1 N	G 2 Mean	G 2 SD	G 2 N	C K–2 Mean	C K–2 SD	C K–2 N
Listening	929.53	19.74	2,657	936.31	19.92	3,774	941.28	19.00	3,427	936.21	20.08	9,858
Reading	925.96	20.70	2,656	934.13	21.29	3,772	939.52	21.32	3,427	933.80	21.79	9,855
Speaking	920.17	19.54	2,651	926.55	21.26	3,767	931.33	21.67	3,424	926.49	21.40	9,842
Writing	917.90	18.65	2,654	926.30	22.42	3,770	932.52	23.76	3,426	926.20	22.69	9,850
Oral	925.02	18.09	2,651	931.65	19.06	3,767	936.51	18.83	3,424	931.55	19.25	9,842
Literacy	922.14	17.92	2,654	930.45	20.25	3,770	936.25	20.85	3,426	930.23	20.61	9,850
Compre- hension	927.08	19.75	2,656	934.84	20.21	3,772	940.10	19.96	3,427	934.58	20.63	9,855
Overall	922.81	17.31	2,649	930.61	19.23	3,764	936.13	19.58	3,423	930.43	19.56	9,836

Table 6.6.1.b.

Mean Scale Scores: 3–5 [Grade = G, Cluster = C]

Domain	G 3 Mean	G 3 SD	G 3 N	G 4 Mean	G 4 SD	G 4 N	G 5 Mean	G 5 SD	G 5 N	C 3–5 Mean	C 3–5 SD	C 3–5 N
Listening	942.84	17.61	3,002	945.26	17.40	2,963	947.00	17.19	2,642	944.95	17.49	8,607
Reading	939.10	18.52	2,999	941.96	19.02	2,963	944.06	18.85	2,639	941.61	18.90	8,601
Speaking	935.99	21.85	3,001	938.25	22.34	2,959	939.63	22.60	2,640	937.89	22.30	8,600
Writing	931.16	20.76	2,995	933.71	21.47	2,959	936.04	22.26	2,640	933.54	21.56	8,594
Oral	939.64	18.31	3,000	941.99	18.55	2,959	943.55	18.52	2,640	941.65	18.52	8,599
Literacy	935.35	18.17	2,994	938.09	18.87	2,959	940.29	19.19	2,639	937.81	18.84	8,592
Compre- hension	940.28	17.54	2,998	943.01	17.83	2,963	945.00	17.67	2,639	942.67	17.78	8,600
Overall	936.43	17.58	2,993	939.06	18.13	2,957	941.08	18.34	2,638	938.76	18.11	8,588

Table 6.6.1.c.**Mean Scale Scores: 6–8 [Grade = G, Cluster = C]**

Domain	G 6 Mean	G 6 SD	G 6 N	G 7 Mean	G 7 SD	G 7 N	G 8 Mean	G 8 SD	G 8 N	C 6–8 Mean	C 6–8 SD	C 6–8 N
Listening	950.42	17.86	2,398	952.17	17.76	2,122	952.61	17.77	2,001	951.66	17.82	6,521
Reading	946.09	18.10	2,398	947.67	18.21	2,122	948.76	18.95	2,001	947.42	18.43	6,521
Speaking	941.72	21.58	2,394	942.68	20.97	2,119	942.98	21.98	1,998	942.42	21.51	6,511
Writing	938.90	22.02	2,394	940.45	21.85	2,120	941.48	22.73	1,999	940.19	22.21	6,513
Oral	946.33	18.23	2,394	947.67	17.93	2,119	948.02	18.50	1,998	947.29	18.23	6,511
Literacy	942.75	18.76	2,394	944.29	18.74	2,120	945.35	19.64	1,999	944.05	19.06	6,513
Compre- hension	947.42	17.35	2,398	949.04	17.40	2,122	949.93	17.93	2,001	948.72	17.58	6,521
Overall	943.62	18.00	2,393	945.11	17.89	2,118	945.96	18.74	1,997	944.82	18.22	6,508

As shown in Table 6.6.1.d., the upward trend exhibited by other grade-level clusters is not present in grades 11 and 12, which have slightly lower scale scores than the preceding grades.

Table 6.6.1.d.**Mean Scale Scores: 9–12 [Grade = G, Cluster = C]**

Domain	G 9 Mean	G 9 SD	G 9 N	G 10 Mean	G 10 SD	G 10 N	G 11 Mean	G 11 SD	G 11 N	G 12 Mean	G 12 SD	G 12 N	C 9–12 Mean	C 9–12 SD	C 9–12 N
Listening	951.64	17.92	1,911	953.53	17.52	1,766	953.31	17.12	1,626	952.86	17.37	2,560	952.81	17.50	7,863
Reading	948.28	19.53	1,910	950.28	19.42	1,766	949.75	19.64	1,626	949.61	19.62	2,558	949.47	19.57	7,860
Speaking	942.64	23.96	1,908	945.31	23.89	1,764	944.50	23.82	1,626	944.14	23.16	2,555	944.11	23.68	7,853
Writing	941.62	22.46	1,908	943.65	22.35	1,766	943.63	23.17	1,626	944.14	22.54	2,554	943.31	22.63	7,854
Oral	947.36	19.57	1,908	949.68	19.27	1,764	949.13	19.14	1,626	948.76	18.91	2,555	948.70	19.21	7,853
Literacy	945.17	19.88	1,908	947.19	19.76	1,766	946.91	20.25	1,626	947.11	19.96	2,554	946.61	19.97	7,854
Compre- hension	949.34	18.44	1,910	951.32	18.18	1,766	950.87	18.24	1,626	950.66	18.31	2,558	950.53	18.31	7,860
Overall	945.63	19.27	1,906	947.75	19.09	1,764	947.37	19.36	1,626	947.40	19.14	2,552	947.04	19.22	7,848

6.7. *Scale Scores by Grade-Level Cluster*

Section 6.7. displays the mean scale scores (Mean), standard deviation (SD), and counts (N) by grade-level cluster across the eight scores awarded on Alternate ACCESS. These scores are first presented for each of the four domains (Listening, Reading, Speaking, and Writing) and then for each of the four composite scores (Oral, Literacy, Comprehension, and Overall). This information is further broken down by gender (Section 6.7.1.) and ethnicity (Section 6.7.2.) across grade-level clusters.

For each of the four grade-level clusters, Tables 6.7.1.a. through 6.7.1.d display the mean scale scores for each domain and composite by gender within each grade-level cluster. The tables indicate that female and male students perform similarly across grade-level clusters. Among the domains, Speaking and Writing generally have lower scale scores compared to Listening and Reading.

6.7.1. Mean Scale Scores by Gender

Table 6.7.1.a.

Mean Scale Scores by Gender: K–2 [Grade = G, Cluster = C]

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N
Listening	936.65	19.89	2,340	936.27	20.15	6,015	935.29	20.09	1,503
Reading	934.09	21.17	2,338	934.01	22.11	6,014	932.53	21.43	1,503
Speaking	926.46	21.32	2,333	926.81	21.52	6,006	925.30	21.04	1,503
Writing	925.19	21.80	2,336	926.86	23.23	6,011	925.13	21.78	1,503
Oral	931.76	19.12	2,333	931.74	19.39	6,006	930.49	18.86	1,503
Literacy	929.87	19.76	2,336	930.66	21.10	6,011	929.05	19.89	1,503
Comprehension	934.92	20.19	2,338	934.75	20.88	6,014	933.39	20.29	1,503
Overall	930.26	18.93	2,331	930.80	19.94	6,002	929.27	18.92	1,503

Table 6.7.1.b.

Mean Scale Scores by Gender: 3–5 [Grade = G, Cluster = C]

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N
Listening	946.31	16.79	2,298	944.77	17.64	5,055	943.16	17.96	1,254
Reading	941.67	18.34	2,296	942.05	19.11	5,052	939.72	18.92	1,253
Speaking	938.73	21.92	2,298	938.08	22.48	5,048	935.56	22.09	1,254
Writing	933.42	21.09	2,294	934.37	21.76	5,046	930.40	21.33	1,254
Oral	942.75	18.01	2,298	941.67	18.69	5,048	939.58	18.62	1,253
Literacy	937.79	18.33	2,294	938.44	19.04	5,045	935.32	18.74	1,253
Comprehension	943.13	17.21	2,296	942.92	17.97	5,052	940.79	17.93	1,252
Overall	939.07	17.60	2,294	939.21	18.30	5,042	936.38	18.08	1,252

Table 6.7.1.c.**Mean Scale Scores by Gender: 6–8 [Grade = G, Cluster = C]**

Domain	Female Mean	Female Std. Dev.	Female N	Male Mean	Male Std. Dev.	Male N	Missing Mean	Missing Std. Dev.	Missing N
Listening	952.45	17.95	1,955	951.64	17.80	3,570	950.20	17.57	996
Reading	947.96	18.21	1,955	947.56	18.67	3,570	945.87	17.91	996
Speaking	942.90	21.24	1,951	942.58	21.59	3,565	940.92	21.72	995
Writing	940.61	22.28	1,951	940.65	22.34	3,566	937.73	21.43	996
Oral	947.93	18.13	1,951	947.36	18.26	3,565	945.78	18.22	995
Literacy	944.53	19.02	1,951	944.35	19.24	3,566	942.03	18.36	996
Comprehension	949.34	17.49	1,955	948.80	17.73	3,570	947.19	17.12	996
Overall	945.36	18.19	1,949	945.05	18.34	3,564	942.95	17.74	995

Table 6.7.1.d.**Mean Scale Scores by Gender: 9–12 [Grade = G, Cluster = C]**

Domain	Female Mean	Female Std. Dev.	Female N	Male Mean	Male Std. Dev.	Male N	Missing Mean	Missing Std. Dev.	Missing N
Listening	951.94	17.54	2,324	953.94	17.29	4,098	950.99	17.81	1,441
Reading	948.81	19.43	2,324	950.45	19.58	4,095	947.74	19.60	1,441
Speaking	943.76	23.75	2,321	945.48	23.69	4,092	940.80	23.19	1,440
Writing	943.03	22.64	2,321	944.63	22.67	4,092	940.02	22.16	1,441
Oral	948.09	19.29	2,321	949.96	19.10	4,092	946.13	19.11	1,440
Literacy	946.15	19.91	2,321	947.76	19.97	4,092	944.10	19.84	1,441
Comprehension	949.80	18.27	2,324	951.56	18.22	4,095	948.77	18.44	1,441
Overall	946.54	19.21	2,319	948.22	19.17	4,089	944.51	19.09	1,440

6.7.2. Mean Scale Scores by Ethnicity

For each of the four grade-level clusters, Tables 6.7.2.1.a. through 6.7.2.4.h. present the mean scale scores for each domain and composite, first separately by cluster and then by ethnicity within each grade-level cluster. Comparisons are focused on the four largest ethnic groups: Hispanic, Black, Asian, and White.

In the K–2 cluster, Hispanic students have the highest Listening scores (937.14), while Black students have the lowest (933.86). Black students also have the lowest Reading scores (931.92), followed by Asian students (932.71). Speaking and Writing scores are relatively close across all racial groups, averaging around 926.

In the 3–5 cluster, similar to the K–2 cluster, Hispanic students lead in Listening scores, and Black students have the lowest Reading scores (939.53), followed by White students (940.46). Speaking and Writing scores follow a similar pattern, except for Writing, where Asian students score the highest (935.47).

In the 6–8 cluster, Hispanic students continue to have the highest Listening scores, while White students have the lowest Reading scores (943.32), followed by Asian students (944.78). In Speaking and Writing, White students have the lowest scores (938.47 for Speaking and 935.28 for Writing), while Hispanic students score the highest (943.09 for Speaking and 940.73 for Writing).

In the 9–12 cluster, Hispanic students achieve the highest scores for both Listening (953.26) and Reading (949.76). White students have the lowest scores across all domains, with scores of 950.16 for Listening, 946.78 for Reading, 942.62 for Speaking, and 939.75 for Writing.

6.7.2.1. Grade-Level Cluster K–2

Table 6.7.2.1.a.

Mean Scale Scores for Hispanic (of any race): K–2

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	937.14	934.39	926.48	925.74	932.01	930.29	935.27	930.62
SD	19.96	21.83	21.61	22.41	19.30	20.46	20.64	19.49
N	5279.00	5277.00	5267.00	5275.00	5267.00	5275.00	5277.00	5264.00

Table 6.7.2.1.b.

Mean Scale Scores Non-Hispanic American Indian: K–2

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	940.58	936.36	933.35	930.42	937.16	933.60	937.65	934.49
SD	19.66	21.94	22.32	22.80	19.45	21.01	20.79	20.03
N	55	55	55	55	55	55	55	55

Table 6.7.2.1.c.

Mean Scale Scores Non-Hispanic Asian: K–2

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	934.42	932.71	925.80	926.82	930.31	929.99	933.27	929.89
SD	19.84	21.65	20.97	23.30	18.96	20.91	20.41	19.65
N	1,519	1,519	1,519	1,519	1,519	1,519	1,519	1,519

Table 6.7.2.1.d.

Mean Scale Scores Non-Hispanic Black: K–2

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	933.86	931.92	926.03	926.84	930.16	929.61	932.56	929.62
SD	19.80	21.82	21.15	23.10	18.90	20.73	20.51	19.52
N	1,081	1,080	1,078	1,080	1,078	1,080	1,080	1,078

Table 6.7.2.1.e.**Mean Scale Scores Non-Hispanic Multiracial: K–2**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	939.59	934.25	926.84	924.46	933.43	929.56	935.89	930.51
SD	20.22	22.90	21.32	22.93	19.08	21.53	21.24	19.82
N	61	61	61	61	61	61	61	61

Table 6.7.2.1.f.**Mean Scale Scores Non-Hispanic Pacific Islander: K–2**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	928.41	926.73	921.18	920.64	924.98	923.88	927.34	923.99
SD	21.17	22.22	21.39	22.21	19.40	20.88	21.20	19.92
N	98	98	98	98	98	98	98	98

Table 6.7.2.1.g.**Mean Scale Scores Non-Hispanic White: K–2**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	935.49	933.42	926.27	925.38	931.09	929.58	934.10	929.84
SD	20.37	21.50	21.41	22.56	19.36	20.41	20.47	19.44
N	850	850	849	848	849	848	850	847

Table 6.7.2.1.h.**Mean Scale Scores Missing: K–2**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	937.62	935.38	928.60	928.33	933.30	932.09	936.12	932.25
SD	20.50	21.60	20.98	22.77	19.38	20.77	20.70	19.69
N	915	915	915	914	915	914	915	914

6.7.2.2. Grade-Level Cluster 3–5**Table 6.7.2.2.a.****Mean Scale Scores for Hispanic (of any race): 3–5**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	945.59	941.80	937.80	933.1	941.93	937.68	942.99	938.76
SD	17.38	18.65	22.65	21.3	18.59	18.57	17.59	17.94
N	4,969	4,966	4,967	4,962	4,966	4,960	4,965	4,957

Table 6.7.2.2.b.**Mean Scale Scores Non-Hispanic American Indian: 3–5**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	950.17	948.09	947.24	944.98	948.98	946.74	948.81	947.21
SD	16.21	17.20	22.47	20.95	17.91	17.47	15.81	16.66
N	58	58	58	58	58	58	58	58

Table 6.7.2.2.c.**Mean Scale Scores Non-Hispanic Asian: 3–5**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	942.76	941.17	936.19	934.47	939.69	938.07	941.70	938.35
SD	17.01	19.08	20.28	21.44	17.35	18.85	17.79	17.80
N	1,293	1,293	1,290	1,291	1,290	1,291	1,293	1,290

Table 6.7.2.2.d.**Mean Scale Scores Non-Hispanic Black: 3–5**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	942.29	939.53	936.87	932.36	939.82	936.18	940.44	937.06
SD	17.93	19.47	22.43	22.22	18.90	19.41	18.27	18.66
N	707	705	706	705	706	705	705	705

Table 6.7.2.2.e.**Mean Scale Scores Non-Hispanic Multiracial: 3–5**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	940.43	936.79	936.26	930.59	938.59	933.95	937.93	935.09
SD	18.66	21.66	22.71	21.13	19.74	20.31	20.36	19.63
N	58	58	58	58	58	58	58	58

Table 6.7.2.2.f.**Mean Scale Scores Non-Hispanic Pacific Islander: 3–5**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	942.96	938.80	936.00	932.53	939.74	935.93	940.13	936.81
SD	19.65	19.69	24.84	23.09	20.58	20.28	18.82	19.55
N	70	70	70	70	70	70	70	70

Table 6.7.2.2.g.

Mean Scale Scores Non-Hispanic White: 3–5

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	943.58	940.46	936.91	931.53	940.48	936.26	941.46	937.33
SD	17.22	18.47	22.27	21.79	18.40	18.72	17.36	17.99
N	784	783	783	782	783	782	783	782

Table 6.7.2.2.h.

Mean Scale Scores Missing: 3–5

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	948.98	944.73	943.53	937.90	946.48	941.55	946.06	942.83
SD	17.73	19.54	22.00	21.86	18.60	19.44	18.28	18.56
N	668	668	668	668	668	668	668	668

6.7.2.3. Grade-Level Cluster 6–8

Table 6.7.2.3.a.

Mean Scale Scores for Hispanic (of any race): 6–8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	952.73	948.41	943.07	940.73	948.15	944.81	949.72	945.61
SD	17.38	17.84	21.33	21.86	17.94	18.60	17.00	17.81
N	4,055	4,055	4,048	4,051	4,048	4,051	4,055	4,047

Table 6.7.2.3.b.

Mean Scale Scores Non-Hispanic American Indian: 6–8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	956.35	950.81	949.51	946.32	953.19	948.84	952.49	949.92
SD	19.28	17.98	19.62	23.75	18.62	19.28	17.74	18.69
N	37	37	37	37	37	37	37	37

Table 6.7.2.3.c.

Mean Scale Scores Non-Hispanic Asian: 6–8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	947.91	944.78	939.07	939.01	943.76	942.15	945.74	942.44
SD	18.80	20.43	21.58	22.93	18.67	20.37	19.38	19.32
N	828	828	826	827	826	827	828	826

Table 6.7.2.3.d.

Mean Scale Scores Non-Hispanic Black: 6–8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	949.79	945.91	942.39	939.18	946.31	942.78	947.09	943.61
SD	17.01	18.12	20.39	22.11	17.30	18.94	17.23	17.89
N	424	424	423	422	423	422	424	422

Table 6.7.2.3.e.

Mean Scale Scores Non-Hispanic Multiracial: 6–8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	949.08	944.16	940.87	937.00	945.16	940.79	945.66	941.89
SD	21.49	17.84	21.90	21.12	20.69	18.93	17.85	18.80
N	38	38	38	38	38	38	38	38

Table 6.7.2.3.f.

Mean Scale Scores Non-Hispanic Pacific Islander: 6–8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	949.63	946.66	944.26	939.52	947.20	943.31	947.55	944.31
SD	16.60	17.31	20.93	23.01	17.45	17.93	16.48	16.92
N	65	65	65	65	65	65	65	65

Table 6.7.2.3.g.

Mean Scale Scores Non-Hispanic White: 6–8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	948.27	943.32	938.47	935.28	943.60	939.52	944.84	940.53
SD	17.90	19.08	22.94	22.23	18.74	19.24	18.04	18.45
N	594	594	594	594	594	594	594	594

Table 6.7.2.3.h.

Mean Scale Scores Missing: 6–8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	955.09	950.19	946.96	944.57	951.25	947.59	951.68	948.53
SD	18.36	18.07	20.86	22.60	18.26	19.23	17.47	18.36
N	480	480	480	479	480	479	480	479

6.7.2.4. Grade-Level Cluster 9–12

Table 6.7.2.4.a.

Mean Scale Scores for Hispanic (of any race): 9–12

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	953.26	949.76	944.13	943.42	948.94	946.81	950.88	947.26
SD	17.56	19.74	23.94	22.56	19.35	20.01	18.43	19.28
N	4,930	4,929	4,926	4,926	4,926	4,926	4,929	4,923

Table 6.7.2.4.b.

Mean Scale Scores Non-Hispanic American Indian: 9–12

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	959.22	957.13	954.17	952.41	956.93	954.93	957.89	955.46
SD	16.63	19.61	22.04	24.97	18.03	21.47	18.06	20.01
N	54	54	54	54	54	54	54	54

Table 6.7.2.4.c.

Mean Scale Scores Non-Hispanic Asian: 9–12

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	951.95	949.09	942.95	943.98	947.71	946.77	950.01	946.85
SD	16.27	18.64	22.33	22.39	17.86	19.31	17.30	18.38
N	974	974	971	973	971	973	974	971

Table 6.7.2.4.d.

Mean Scale Scores Non-Hispanic Black: 9–12

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	951.09	948.11	944.17	942.7	947.88	945.62	949.07	946.10
SD	17.76	20.30	23.94	23.1	19.61	20.62	18.91	19.77
N	565	563	565	563	565	563	563	563

Table 6.7.2.4.e.

Mean Scale Scores Non-Hispanic Multiracial: 9–12

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	945.39	942.59	937.07	937.95	941.43	940.45	943.48	940.48
SD	18.48	19.51	23.46	22.51	19.38	19.48	18.68	19.06
N	44	44	44	44	44	44	44	44

Table 6.7.2.4.f.**Mean Scale Scores Non-Hispanic Pacific Islander: 9–12**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	950.25	947.04	941.32	943.50	946.01	945.49	948.04	945.49
SD	19.62	21.93	26.20	24.46	21.35	22.49	20.39	21.75
N	72	72	72	72	72	72	72	72

Table 6.7.2.4.g.**Mean Scale Scores Non-Hispanic White: 9–12**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	950.16	946.78	942.62	939.75	946.63	943.52	947.84	944.24
SD	17.60	18.31	23.00	21.88	19.08	19.01	17.48	18.45
N	625	625	624	624	624	624	625	624

Table 6.7.2.4.h.**Mean Scale Scores Missing: 9–12**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	955.10	951.81	947.29	945.11	951.44	948.71	952.87	949.32
SD	17.66	19.37	23.41	22.97	19.27	20.18	18.29	19.35
N	599	599	597	598	597	598	599	597

6.8. *Scale Score By Grade*

Section 6.8 provides a detailed breakdown of mean scale scores by individual grades, complementing the information presented in the prior section, which focused on grade-level clusters. Section 6.8.1. presents scale scores by gender for each grade. Section 6.8.2. presents scale scores by ethnicity.

6.8.1. Mean Scale Scores by Gender

This section presents the scale scores by gender for each grade, highlighting performance trends across grades. The patterns observed in Section 6.8. align closely with those described in Section 6.7.1. Performance trends across individual grades reveal consistent gender parity across all domains, with minor variations in Speaking and Writing scores between male and female students. Similar to the grade-level cluster data, Listening and Reading domains tend to exhibit higher mean scores compared to Speaking and Writing.

Table 6.8.1.a.**Mean Scale Scores by Gender: Kindergarten**

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N	Total Mean	Total SD	Total N
Listening	929.79	19.04	607	929.27	19.84	1,644	930.20	20.39	406	929.53	19.74	2,657
Reading	925.35	19.58	606	925.98	21.08	1,644	926.79	20.83	406	925.96	20.70	2,656
Speaking	920.02	19.12	603	920.08	19.52	1,642	920.78	20.25	406	920.17	19.54	2,651
Writing	916.66	17.55	606	918.17	19.09	1,642	918.64	18.37	406	917.90	18.65	2,654
Oral	925.03	17.52	603	924.85	18.15	1,642	925.67	18.69	406	925.02	18.09	2,651
Literacy	921.22	16.61	606	922.28	18.40	1,642	922.94	17.79	406	922.14	17.92	2,654
Compre- hension	926.74	18.76	606	927.02	20.03	1,644	927.85	20.06	406	927.08	19.75	2,656
Overall	922.17	16.26	603	922.86	17.65	1,640	923.57	17.42	406	922.81	17.31	2,649

Table 6.8.1.b.**Mean Scale Scores by Gender: Grade 1**

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N	Total Mean	Total SD	Total N
Listening	937.01	20.13	877	936.30	19.98	2,318	935.31	19.31	579	936.31	19.92	3,774
Reading	935.11	20.90	876	934.24	21.59	2,317	932.22	20.57	579	934.13	21.29	3,772
Speaking	926.33	21.05	875	927.00	21.40	2,313	925.06	21.00	579	926.55	21.26	3,767
Writing	925.96	21.77	874	926.87	22.81	2,317	924.52	21.75	579	926.30	22.42	3,770
Oral	931.92	19.07	875	931.86	19.23	2,313	930.38	18.33	579	931.65	19.06	3,767
Literacy	930.77	19.69	874	930.79	20.62	2,317	928.59	19.56	579	930.45	20.25	3,770
Compre- hension	935.75	20.06	876	934.92	20.45	2,317	933.18	19.35	579	934.84	20.21	3,772
Overall	930.93	18.86	873	930.92	19.53	2,312	928.91	18.52	579	930.61	19.23	3,764

Table 6.8.1.c.**Mean Scale Scores by Gender: Grade 2**

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N	Total Mean	Total SD	Total N
Listening	941.14	18.89	856	941.85	18.80	2,053	939.25	19.84	518	941.28	19.00	3,427
Reading	939.23	20.61	856	940.18	21.48	2,053	937.36	21.72	518	939.52	21.32	3,427
Speaking	931.13	21.88	855	931.97	21.72	2,051	929.12	20.99	518	931.33	21.67	3,424
Writing	930.43	22.73	856	933.81	24.35	2,052	930.90	22.77	518	932.52	23.76	3,426
Oral	936.35	18.88	855	937.11	18.81	2,051	934.38	18.72	518	936.51	18.83	3,424
Literacy	935.06	19.85	856	937.23	21.32	2,052	934.36	20.39	518	936.25	20.85	3,426
Compre- hension	939.86	19.52	856	940.74	20.00	2,053	937.97	20.42	518	940.10	19.96	3,427
Overall	935.27	18.88	855	937.00	19.91	2,050	934.13	19.23	518	936.13	19.58	3,423

Table 6.8.1.d.**Mean Scale Scores by Gender: Grade 3**

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N	Total Mean	Total SD	Total N
Listening	943.37	17.37	765	942.75	17.78	1,790	942.30	17.41	447	942.84	17.61	3,002
Reading	938.06	18.13	764	939.67	18.85	1,788	938.58	17.75	447	939.10	18.52	2,999
Speaking	935.94	21.42	765	936.28	22.01	1,788	934.94	21.96	448	935.99	21.85	3,001
Writing	929.75	19.92	763	932.39	21.11	1,785	928.66	20.43	447	931.16	20.76	2,995
Oral	939.87	17.97	765	939.74	18.48	1,788	938.82	18.20	447	939.64	18.31	3,000
Literacy	934.13	17.53	763	936.24	18.53	1,784	933.86	17.63	447	935.35	18.17	2,994
Compre- hension	939.71	17.22	764	940.66	17.82	1,788	939.72	16.96	446	940.28	17.54	2,998
Overall	935.65	17.01	763	937.09	17.89	1,784	935.12	17.17	446	936.43	17.58	2,993

Table 6.8.1.e.**Mean Scale Scores by Gender: Grade 4**

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N	Total Mean	Total SD	Total N
Listening	947.15	16.21	829	944.84	17.68	1,731	943.22	18.22	403	945.26	17.40	2,963
Reading	942.51	17.85	829	942.24	19.34	1,731	939.63	19.78	403	941.96	19.02	2,963
Speaking	939.05	21.89	829	938.47	22.57	1,727	935.65	22.11	403	938.25	22.34	2,959
Writing	934.00	20.68	828	934.35	21.79	1,728	930.39	21.40	403	933.71	21.47	2,959
Oral	943.32	17.82	829	941.91	18.78	1,727	939.65	18.82	403	941.99	18.55	2,959
Literacy	938.50	17.95	828	938.54	19.16	1,728	935.27	19.27	403	938.09	18.87	2,959
Compre- hension	943.98	16.67	829	943.08	18.15	1,731	940.76	18.59	403	943.01	17.83	2,963
Overall	939.76	17.30	828	939.35	18.40	1,726	936.35	18.46	403	939.06	18.13	2,957

Table 6.8.1.f.**Mean Scale Scores by Gender: Grade 5**

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N	Total Mean	Total SD	Total N
Listening	948.53	16.39	704	947.06	17.16	1,534	944.07	18.29	404	947.00	17.19	2,642
Reading	944.60	18.51	703	944.60	18.82	1,533	941.07	19.26	403	944.06	18.85	2,639
Speaking	941.40	22.18	704	939.73	22.79	1,533	936.17	22.26	403	939.63	22.60	2,640
Writing	936.72	22.18	703	936.70	22.25	1,533	932.33	22.10	404	936.04	22.26	2,640
Oral	945.19	17.88	704	943.64	18.61	1,533	940.34	18.89	403	943.55	18.52	2,640
Literacy	940.91	18.95	703	940.88	19.20	1,533	936.98	19.29	403	940.29	19.19	2,639
Compre- hension	945.86	17.24	703	945.38	17.63	1,533	942.01	18.28	403	945.00	17.67	2,639
Overall	941.97	17.98	703	941.53	18.36	1,532	937.80	18.60	403	941.08	18.34	2,638

Table 6.8.1.g.**Mean Scale Scores by Gender: Grade 6**

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N	Total Mean	Total SD	Total N
Listening	951.61	18.09	693	950.12	17.64	1,343	949.28	18.13	362	950.42	17.86	2,398
Reading	947.18	17.83	693	946.02	18.22	1,343	944.22	18.02	362	946.09	18.10	2,398
Speaking	942.46	21.14	692	941.68	21.65	1,340	940.48	22.16	362	941.72	21.58	2,394
Writing	939.51	22.20	691	939.20	22.00	1,341	936.60	21.63	362	938.90	22.02	2,394
Oral	947.30	18.12	692	946.16	18.14	1,340	945.11	18.68	362	946.33	18.23	2,394
Literacy	943.60	18.76	691	942.87	18.79	1,341	940.64	18.53	362	942.75	18.76	2,394
Compre- hension	948.55	17.31	693	947.29	17.36	1,343	945.76	17.28	362	947.42	17.35	2,398
Overall	944.51	18.01	691	943.66	17.97	1,340	941.77	17.99	362	943.62	18.00	2,393

Table 6.8.1.h.**Mean Scale Scores by Gender: Grade 7**

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N	Total Mean	Total SD	Total N
Listening	953.11	18.01	659	951.87	18.14	1,140	951.30	15.73	323	952.17	17.76	2,122
Reading	948.27	18.03	659	947.56	18.74	1,140	946.79	16.64	323	947.67	18.21	2,122
Speaking	942.52	21.29	657	943.20	20.81	1,139	941.21	20.83	323	942.68	20.97	2,119
Writing	941.20	21.80	657	941.09	22.15	1,140	936.68	20.50	323	940.45	21.85	2,120
Oral	948.06	18.23	657	947.78	18.09	1,139	946.50	16.70	323	947.67	17.93	2,119
Literacy	944.97	18.72	657	944.56	19.17	1,140	941.96	17.05	323	944.29	18.74	2,120
Compre- hension	949.74	17.35	659	948.87	17.89	1,140	948.16	15.64	323	949.04	17.40	2,122
Overall	945.70	17.97	656	945.34	18.25	1,139	943.11	16.33	323	945.11	17.89	2,118

Table 6.8.1.i.**Mean Scale Scores by Gender: Grade 8**

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N	Total Mean	Total SD	Total N
Listening	952.71	17.72	603	953.26	17.48	1,087	950.14	18.68	311	952.61	17.77	2,001
Reading	948.53	18.84	603	949.45	19.00	1,087	946.83	18.93	311	948.76	18.95	2,001
Speaking	943.81	21.29	602	943.05	22.29	1,086	941.13	22.17	310	942.98	21.98	1,998
Writing	941.25	22.87	603	941.99	22.86	1,085	940.12	22.00	311	941.48	22.73	1,999
Oral	948.51	18.06	602	948.38	18.52	1,086	945.82	19.18	310	948.02	18.50	1,998
Literacy	945.11	19.61	603	945.94	19.73	1,085	943.72	19.38	311	945.35	19.64	1,999
Compre- hension	949.81	17.83	603	950.60	17.85	1,087	947.85	18.31	311	949.93	17.93	2,001
Overall	945.98	18.62	602	946.46	18.77	1,085	944.15	18.79	310	945.96	18.74	1,997

Table 6.8.1.j.**Mean Scale Scores by Gender: Grade 9**

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N	Total Mean	Total SD	Total N
Listening	951.31	17.34	596	952.75	17.63	977	948.99	19.45	338	951.64	17.92	1,911
Reading	948.25	18.90	596	948.73	19.51	976	947.01	20.66	338	948.28	19.53	1,910
Speaking	942.05	23.76	595	944.03	23.98	975	939.67	24.02	338	942.64	23.96	1,908
Writing	942.20	22.41	595	942.33	22.42	975	938.53	22.48	338	941.62	22.46	1,908
Oral	946.90	19.22	595	948.63	19.48	975	944.54	20.17	338	947.36	19.57	1,908
Literacy	945.48	19.61	595	945.75	19.83	975	942.98	20.40	338	945.17	19.88	1,908
Compre- hension	949.23	17.82	596	949.99	18.31	976	947.66	19.77	338	949.34	18.44	1,910
Overall	945.68	18.98	594	946.41	19.20	974	943.28	19.84	338	945.63	19.27	1,906

Table 6.8.1.k.**Mean Scale Scores by Gender: Grade 10**

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N	Total Mean	Total SD	Total N
Listening	953.26	16.82	541	954.39	17.53	939	951.24	18.63	286	953.53	17.52	1,766
Reading	950.25	18.98	541	951.23	19.56	939	947.22	19.52	286	950.28	19.42	1,766
Speaking	946.11	24.37	540	946.44	23.58	938	940.12	23.38	286	945.31	23.89	1,764
Writing	944.23	22.33	541	944.75	22.48	939	938.93	21.44	286	943.65	22.35	1,766
Oral	949.97	19.13	540	950.66	19.09	938	945.93	19.72	286	949.68	19.27	1,764
Literacy	947.45	19.47	541	948.22	19.90	939	943.31	19.46	286	947.19	19.76	1,766
Compre- hension	951.20	17.71	541	952.25	18.26	939	948.49	18.58	286	951.32	18.18	1,766
Overall	948.06	18.83	540	948.75	19.14	938	943.88	18.96	286	947.75	19.09	1,764

Table 6.8.1.l.**Mean Scale Scores by Gender: Grade 11**

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N	Total Mean	Total SD	Total N
Listening	951.42	17.63	477	954.80	16.66	885	951.75	17.27	264	953.31	17.12	1,626
Reading	947.59	19.81	477	951.48	19.24	885	947.86	20.16	264	949.75	19.64	1,626
Speaking	942.40	23.25	477	946.34	23.91	885	942.17	24.14	264	944.50	23.82	1,626
Writing	941.92	22.53	477	945.63	23.15	885	940.01	23.77	264	943.63	23.17	1,626
Oral	947.12	19.14	477	950.80	18.92	885	947.20	19.39	264	949.13	19.14	1,626
Literacy	944.97	19.94	477	948.77	20.02	885	944.16	21.00	264	946.91	20.25	1,626
Compre- hension	948.78	18.58	477	952.53	17.79	885	949.09	18.63	264	950.87	18.24	1,626
Overall	945.43	19.16	477	949.17	19.13	885	944.84	19.90	264	947.37	19.36	1,626

Table 6.8.1.m.**Mean Scale Scores by Gender: Grade 12**

Domain	Female Mean	Female SD	Female N	Male Mean	Male SD	Male N	Missing Mean	Missing SD	Missing N	Total Mean	Total SD	Total N
Listening	951.82	18.16	710	953.91	17.25	1,297	951.71	16.49	553	952.86	17.37	2,560
Reading	948.99	19.90	710	950.48	19.82	1,295	948.39	18.71	553	949.61	19.62	2,558
Speaking	944.34	23.46	709	945.29	23.35	1,294	941.19	22.09	552	944.14	23.16	2,555
Writing	943.54	23.12	708	945.58	22.56	1,293	941.51	21.49	553	944.14	22.54	2,554
Oral	948.32	19.49	709	949.87	18.92	1,294	946.70	17.95	552	948.76	18.91	2,555
Literacy	946.51	20.43	708	948.26	20.00	1,293	945.18	19.10	553	947.11	19.96	2,554
Compre- hension	949.89	18.83	710	951.60	18.35	1,295	949.44	17.43	553	950.66	18.31	2,558
Overall	946.84	19.66	708	948.55	19.14	1,292	945.43	18.27	552	947.40	19.14	2,552

6.8.2. Mean Scale Scores by Ethnicity

Section 6.8.2. reports the scale scores by ethnicity, offering insights into performance differences among ethnic groups at each grade level. For ethnic groups, Hispanic students consistently demonstrate strong performance in Listening across grades, while other domains show mixed trends depending on the grade and ethnicity. Black and White students often exhibit slightly lower mean scores in certain domains like Reading and Writing. The consistency in trends across grades suggests that the broader patterns identified in grade-level clusters remain evident at the individual grade level.

6.8.2.1. Kindergarten**Table 6.8.2.1.a.****Mean Scale Scores for Hispanic (of any race): Kindergarten**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	930.44	926.45	919.50	917.27	925.13	922.08	927.70	922.79
SD	19.78	20.89	19.23	18.10	18.00	17.71	19.92	17.17
N	1,327	1,327	1,322	1,327	1,322	1,327	1,327	1,322

Table 6.8.2.1.b.**Mean Scale Scores Non-Hispanic American Indian: Kindergarten**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	932.37	925.26	922.95	917.68	927.89	921.68	927.42	923.42
SD	21.61	21.44	23.10	22.16	19.91	20.31	20.93	19.80
N	19	19	19	19	19	19	19	19

Table 6.8.2.1.c.**Mean Scale Scores Non-Hispanic Asian: Kindergarten**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	928.67	925.35	921.05	918.23	925.05	922.01	926.38	922.76
SD	19.09	19.63	19.46	18.44	17.74	17.00	18.65	16.44
N	398	398	398	398	398	398	398	398

Table 6.8.2.1.d.**Mean Scale Scores Non-Hispanic Black: Kindergarten**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	928.05	924.30	920.78	919.68	924.57	922.21	925.47	922.77
SD	19.20	20.66	19.71	20.55	17.90	18.74	19.54	17.81
N	350	349	349	349	349	349	349	349

Table 6.8.2.1.e.**Mean Scale Scores Non-Hispanic Multiracial: Kindergarten**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	933.00	932.95	927.45	925.59	930.41	929.45	933.05	929.55
SD	21.75	25.51	19.87	22.57	18.25	22.89	23.52	20.71
N	22	22	22	22	22	22	22	22

Table 6.8.2.1.f.**Mean Scale Scores Non-Hispanic Pacific Islander: Kindergarten**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	923.84	917.58	913.37	913.95	918.84	916.00	919.63	916.63
SD	20.41	19.01	17.07	17.57	16.74	17.84	18.90	17.28
N	19	19	19	19	19	19	19	19

Table 6.8.2.1.g.**Mean Scale Scores Non-Hispanic White: Kindergarten**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	927.36	924.48	918.04	915.95	922.86	920.32	925.39	920.86
SD	20.18	20.47	19.39	18.00	18.10	17.75	19.71	17.18
N	227	227	227	226	227	226	227	226

Table 6.8.2.1.h.

Mean Scale Scores Missing: Kindergarten

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	929.98	927.75	922.65	919.35	926.50	923.78	928.48	924.39
SD	20.25	21.00	20.51	18.81	18.97	18.52	20.18	17.99
N	295	295	295	294	295	294	295	294

6.8.2.2. Grade 1

Table 6.8.2.2.a.

Mean Scale Scores for Hispanic (of any race): Grade 1

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	936.92	934.36	926.53	925.81	931.94	930.32	935.19	930.61
SD	19.89	21.36	21.52	22.07	19.20	20.12	20.28	19.20
N	2,058	2,056	2,053	2,054	2,053	2,054	2,056	2,050

Table 6.8.2.2.b.

Mean Scale Scores Non-Hispanic American Indian: Grade 1

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	942.50	940.54	935.35	936.12	939.08	938.50	941.15	938.46
SD	17.09	20.48	20.34	19.19	17.88	17.94	19.10	17.32
N	26	26	26	26	26	26	26	26

Table 6.8.2.2.c.

Mean Scale Scores Non-Hispanic Asian: Grade 1

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	934.73	933.62	925.57	926.65	930.35	930.35	934.00	930.14
SD	19.72	21.22	21.02	23.21	18.87	20.68	20.09	19.49
N	575	575	575	575	575	575	575	575

Table 6.8.2.2.d.

Mean Scale Scores Non-Hispanic Black: Grade 1

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	934.52	932.77	926.09	926.95	930.59	930.09	933.35	930.13
SD	20.06	21.65	21.19	22.90	18.99	20.56	20.40	19.36
N	387	387	385	387	385	387	387	385

Table 6.8.2.2.e.

Mean Scale Scores Non-Hispanic Multiracial: Grade 1

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	943.50	937.90	929.1	924.60	936.55	931.45	939.55	932.80
SD	19.45	24.24	22.9	26.63	20.37	23.48	22.30	21.25
N	20	20	20	20	20	20	20	20

Table 6.8.2.2.f.

Mean Scale Scores Non-Hispanic Pacific Islander: Grade 1

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	927.69	928.58	924.07	919.64	926.09	924.27	928.40	924.6
SD	21.81	21.59	19.06	21.68	18.35	19.78	21.12	18.8
N	45	45	45	45	45	45	45	45

Table 6.8.2.2.g.

Mean Scale Scores Non-Hispanic White: Grade 1

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	935.94	933.91	926.53	925.76	931.45	930.06	934.57	930.28
SD	19.31	20.49	20.50	21.79	18.40	19.38	19.31	18.38
N	343	343	343	343	343	343	343	343

Table 6.8.2.2.h.

Mean Scale Scores Missing: Grade 1

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	938.06	935.52	928.47	928.85	933.46	932.41	936.35	932.51
SD	20.33	21.10	21.10	23.14	19.12	20.74	20.26	19.60
N	320	320	320	320	320	320	320	320

6.8.2.3. Grade 2

Table 6.8.2.3.a.

Mean Scale Scores for Hispanic (of any race): Grade 2

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	942.08	939.98	931.32	931.58	936.90	936.02	940.66	936.09
SD	18.74	21.27	21.94	23.61	18.81	20.68	19.86	19.46
N	1,894	1,894	1,892	1,894	1,892	1,894	1,894	1,892

Table 6.8.2.3.b.

Mean Scale Scores Non-Hispanic American Indian: Grade 2

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	951.20	946.60	947.90	939.80	949.80	943.50	948.00	945.20
SD	17.21	19.37	17.28	24.08	14.87	21.26	18.14	18.92
N	10	10	10	10	10	10	10	10

Table 6.8.2.3.c.

Mean Scale Scores Non-Hispanic Asian: Grade 2

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	938.28	937.14	929.50	933.25	934.09	935.43	937.53	934.83
SD	19.56	22.15	21.29	24.54	19.03	21.91	20.71	20.42
N	546	546	546	546	546	546	546	546

Table 6.8.2.3.d.

Mean Scale Scores Non-Hispanic Black: Grade 2

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	939.02	938.71	931.29	933.97	935.35	936.57	938.86	935.99
SD	18.57	20.75	21.26	23.60	18.25	20.37	19.40	19.14
N	344	344	344	344	344	344	344	344

Table 6.8.2.3.e.

Mean Scale Scores Non-Hispanic Multiracial: Grade 2

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	943.11	931.89	923.74	923.00	933.63	927.68	935.32	929.21
SD	18.12	18.61	22.02	20.13	19.10	18.56	17.54	17.97
N	19	19	19	19	19	19	19	19

Table 6.8.2.3.f.

Mean Scale Scores Non-Hispanic Pacific Islander: Grade 2

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	931.91	929.41	921.74	925.71	926.94	927.76	930.24	927.29
SD	20.75	23.92	25.58	24.51	21.85	23.13	22.06	22.11
N	34	34	34	34	34	34	34	34

Table 6.8.2.3.g.

Mean Scale Scores Non-Hispanic White: Grade 2

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	941.53	940.08	932.65	932.56	937.35	936.50	940.60	936.58
SD	19.64	21.07	21.89	24.10	19.13	20.83	19.97	19.64
N	280	280	279	279	279	279	280	278

Table 6.8.2.3.h.

Mean Scale Scores Missing: Grade 2

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	944.65	942.73	934.58	936.59	939.82	939.90	943.37	939.66
SD	18.26	20.16	19.65	22.73	17.79	19.83	19.01	18.47
N	300	300	300	300	300	300	300	300

6.8.2.4. Grade 3

Table 6.8.2.4.a.

Mean Scale Scores for Hispanic (of any race): Grade 3

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	943.17	938.99	935.61	930.35	939.60	934.88	940.29	936.08
SD	17.68	18.24	22.29	20.59	18.51	17.93	17.42	17.50
N	1,705	1,704	1,706	1,701	1,705	1,700	1,703	1,699

Table 6.8.2.4.b.

Mean Scale Scores Non-Hispanic American Indian: Grade 3

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	940.33	941.73	938.60	935.73	939.80	938.93	941.33	939.07
SD	19.47	20.45	24.67	22.34	20.72	20.09	18.58	19.48
N	15	15	15	15	15	15	15	15

Table 6.8.2.4.c.

Mean Scale Scores Non-Hispanic Asian: Grade 3

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	940.76	938.98	934.55	932.82	937.86	936.14	939.58	936.46
SD	17.61	19.11	19.87	20.92	17.48	18.46	17.96	17.53
N	448	448	447	447	447	447	448	447

Table 6.8.2.4.d.

Mean Scale Scores Non-Hispanic Black: Grade 3

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	939.38	936.66	934.79	929.36	937.36	933.23	937.57	934.28
SD	18.12	18.74	21.51	20.59	18.27	18.08	17.73	17.50
N	248	246	247	246	247	246	246	246

Table 6.8.2.4.e.

Mean Scale Scores Non-Hispanic Multiracial: Grade 3

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	940.23	935.23	936.32	927.82	938.55	931.77	936.73	933.50
SD	15.73	20.11	22.18	20.00	17.86	19.07	18.37	17.95
N	22	22	22	22	22	22	22	22

Table 6.8.2.4.f.

Mean Scale Scores Non-Hispanic Pacific Islander: Grade 3

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	940.15	936.30	934.75	929.95	937.65	933.40	937.50	934.45
SD	19.77	18.28	23.60	20.19	20.19	17.97	17.93	17.82
N	20	20	20	20	20	20	20	20

Table 6.8.2.4.g.

Mean Scale Scores Non-Hispanic White: Grade 3

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	942.81	939.05	936.27	930.62	939.76	935.06	940.24	936.27
SD	16.40	18.55	21.79	21.29	17.71	18.55	17.14	17.63
N	276	276	276	276	276	276	276	276

Table 6.8.2.4.h.

Mean Scale Scores Missing: Grade 3

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	948.00	942.67	941.59	935.87	945.04	939.50	944.33	940.97
SD	16.67	18.49	21.64	20.59	17.89	18.27	17.19	17.52
N	268	268	268	268	268	268	268	268

6.8.2.5. Grade 4

Table 6.8.2.5.a.

Mean Scale Scores for Hispanic (of any race): Grade 4

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	946.02	942.32	938.55	933.31	942.52	938.05	943.49	939.21
SD	17.37	18.68	22.59	21.07	18.65	18.52	17.60	17.90
N	1,714	1,714	1,713	1,712	1,713	1,712	1,714	1,711

Table 6.8.2.5.b.

Mean Scale Scores Non-Hispanic American Indian: Grade 4

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	953.65	951.2	952.35	947.95	953.30	949.75	952.10	950.60
SD	9.45	11.2	20.59	18.59	13.77	12.82	9.49	11.87
N	20	20	20	20	20	20	20	20

Table 6.8.2.5.c.

Mean Scale Scores Non-Hispanic Asian: Grade 4

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	942.97	941.43	936.75	934.98	940.09	938.48	941.95	938.74
SD	16.53	19.35	19.84	21.64	16.90	19.10	17.83	17.87
N	440	440	438	439	438	439	440	438

Table 6.8.2.5.d.

Mean Scale Scores Non-Hispanic Black: Grade 4

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	943.34	940.44	937.43	933.99	940.60	937.44	941.4	938.18
SD	18.46	19.92	23.19	22.60	19.73	19.94	18.8	19.33
N	263	263	263	263	263	263	26	263

Table 6.8.2.5.e.

Mean Scale Scores Non-Hispanic Multiracial: Grade 4

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	939.76	935.71	934.65	933.06	937.47	934.59	936.94	935.29
SD	18.28	23.74	23.56	20.78	19.71	21.62	21.64	20.73
N	17	17	17	17	17	17	17	17

Table 6.8.2.5.f.

Mean Scale Scores Non-Hispanic Pacific Islander: Grade 4

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	944.96	936.74	935.41	933.44	940.44	935.33	939.33	936.59
SD	21.05	22.33	27.16	22.69	22.70	21.85	20.98	21.30
N	27	27	27	27	27	27	27	27

Table 6.8.2.5.g.

Mean Scale Scores Non-Hispanic White: Grade 4

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	943.54	940.29	935.50	930.12	939.80	935.53	941.32	936.58
SD	16.92	18.02	22.79	20.94	18.34	18.00	16.91	17.48
N	273	273	272	272	272	272	273	272

Table 6.8.2.5.h.

Mean Scale Scores Missing: Grade 4

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	948.24	944.49	942.82	937.38	945.75	941.16	945.67	942.33
SD	17.72	20.34	22.02	22.75	18.51	20.24	18.86	19.05
N	209	209	209	209	209	209	209	209

6.8.2.6. Grade 5

Table 6.8.2.6.a.

Mean Scale Scores for Hispanic (of any race): Grade 5

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	947.78	944.31	939.39	935.90	943.83	940.35	945.40	941.21
SD	16.72	18.67	22.93	21.94	18.36	18.90	17.38	18.08
N	1,550	1,548	1,548	1,549	1,548	1,548	1,548	1,547

Table 6.8.2.6.b.

Mean Scale Scores Non-Hispanic American Indian: Grade 5

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	953.57	949.52	948.43	948.43	951.22	949.22	950.83	949.57
SD	16.54	18.80	21.88	21.03	17.74	18.32	17.26	17.18
N	23	23	23	23	23	23	23	23

Table 6.8.2.6.c.**Mean Scale Scores Non-Hispanic Asian: Grade 5**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	944.75	943.30	937.39	935.75	941.29	939.76	943.77	940.00
SD	16.63	18.51	21.10	21.72	17.54	18.85	17.34	17.88
N	405	405	405	405	405	405	405	405

Table 6.8.2.6.d.**Mean Scale Scores Non-Hispanic Black: Grade 5**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	944.55	941.92	938.76	933.95	941.87	938.17	942.76	939.05
SD	16.52	19.41	22.44	23.37	18.28	19.97	17.83	18.82
N	196	196	196	196	196	196	196	196

Table 6.8.2.6.e.**Mean Scale Scores Non-Hispanic Multiracial: Grade 5**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	941.26	939.58	937.63	931.58	939.63	935.89	940.21	936.74
SD	22.69	22.39	23.70	23.38	22.68	21.36	22.22	21.35
N	19	19	19	19	19.00	19	19	19

Table 6.8.2.6.f.**Mean Scale Scores Non-Hispanic Pacific Islander: Grade 5**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	943.04	943.39	937.78	933.7	940.74	938.83	943.35	939.13
SD	18.39	17.47	24.01	26.51	19.00	20.77	17.19	19.42
N	23	23	23	23	23	23	23	23

Table 6.8.2.6.g.**Mean Scale Scores Non-Hispanic White: Grade 5**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	944.54	942.33	939.30	934.26	942.13	938.52	943.06	939.44
SD	18.49	18.79	22.12	23.14	19.23	19.62	18.07	18.88
N	235	234	235	234	235	234	234	234

Table 6.8.2.6.h.

Mean Scale Scores Missing: Grade 5

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	951.18	947.89	947.04	941.34	949.30	944.84	948.94	946.00
SD	19.04	19.77	22.19	22.28	19.44	19.80	18.84	19.12
N	191	191	191	191	191	191	191	191

6.8.2.7. Grade 6

Table 6.8.2.7.a.

Mean Scale Scores for Hispanic (of any race): Grade 6

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	951.19	946.73	941.83	939.03	946.77	943.13	948.09	944.02
SD	17.81	17.75	21.88	21.87	18.40	18.57	17.08	17.93
N	1,453	1,453	1,450	1,450	1,450	1,450	1,453	1,449

Table 6.8.2.7.b.

Mean Scale Scores Non-Hispanic American Indian: Grade 6

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	957.88	948.94	950.12	943.12	954.25	946.31	951.56	948.50
SD	20.35	17.53	21.11	26.36	19.85	20.66	17.73	20.07
N	16	16	16	16	16	16	16	16

Table 6.8.2.7.c.

Mean Scale Scores Non-Hispanic Asian: Grade 6

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	946.86	944.33	938.99	939.28	943.23	942.10	945.14	942.21
SD	18.00	19.57	21.00	22.10	17.82	19.42	18.57	18.40
N	316	316	315	315	315	315	316	315

Table 6.8.2.7.d.

Mean Scale Scores Non-Hispanic Black: Grade 6

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	949.49	945.15	942.57	937.19	946.27	941.44	946.51	942.62
SD	16.35	17.63	19.28	21.07	16.60	17.93	16.75	16.93
N	156	156	156	156	156	156	156	156

Table 6.8.2.7.e.**Mean Scale Scores Non-Hispanic Multiracial: Grade 6**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	955.36	952.45	947.82	942.82	951.91	947.82	953.45	948.82
SD	14.23	10.79	11.47	17.58	11.67	13.98	10.64	12.64
N	11	11	11	11	11	11	11	11

Table 6.8.2.7.f.**Mean Scale Scores Non-Hispanic Pacific Islander: Grade 6**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	944.77	943.27	942.55	941.73	943.91	942.77	943.68	942.95
SD	18.42	18.49	24.34	23.95	19.91	19.15	17.69	18.61
N	22	22	22	22	22	22	22	22

Table 6.8.2.7.g.**Mean Scale Scores Non-Hispanic White: Grade 6**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	948.62	943.26	939.47	935.35	944.29	939.53	944.92	940.75
SD	17.43	18.28	22.95	22.56	18.32	19.02	17.26	18.11
N	237	237	237	237	237	237	237	237

Table 6.8.2.7.h.**Mean Scale Scores Missing: Grade 6**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	953.29	948.14	946.48	942.23	950.11	945.39	949.71	946.67
SD	18.52	18.22	19.46	22.30	17.73	19.03	17.55	18.09
N	187	187	187	187	187	187	187	187

6.8.2.8. Grade 7**Table 6.8.2.8.a.****Mean Scale Scores for Hispanic (of any race): Grade 7**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	953.56	948.78	943.92	941.13	949.00	945.19	950.23	946.15
SD	16.89	17.57	20.57	21.53	17.32	18.30	16.62	17.39
N	1,348	1,348	1,345	1,347	1,345	1,347	1,348	1,345

Table 6.8.2.8.b.

Mean Scale Scores Non-Hispanic American Indian: Grade 7

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	956.64	956.29	950.79	947.86	954.00	952.29	956.50	952.57
SD	17.27	16.91	18.18	23.27	16.79	19.01	16.55	18.10
N	14	14	14	14	14	14	14	14

Table 6.8.2.8.c.

Mean Scale Scores Non-Hispanic Asian: Grade 7

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	948.33	944.77	938.82	939.02	943.80	942.14	945.85	942.44
SD	19.82	20.50	21.83	22.76	19.34	20.35	19.76	19.52
N	266	266	266	266	266	266	266	266

Table 6.8.2.8.d.

Mean Scale Scores Non-Hispanic Black: Grade 7

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	949.46	946.28	942.49	940.29	946.2	943.51	947.19	944.08
SD	17.48	17.41	19.90	21.84	17.2	18.37	16.78	17.42
N	134	134	134	133	134	133	134	133

Table 6.8.2.8.e.

Mean Scale Scores Non-Hispanic Multiracial: Grade 7

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	944.86	943.71	936.43	937.86	940.86	941.00	944.14	940.86
SD	20.46	20.61	22.87	24.27	21.18	21.43	19.74	20.33
N	7	7	7	7	7	7	7	7

Table 6.8.2.8.f.

Mean Scale Scores Non-Hispanic Pacific Islander: Grade 7

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	953.05	949.30	945.30	941.60	949.40	945.65	950.45	946.65
SD	16.54	15.67	19.16	21.07	16.33	15.62	15.44	14.76
N	20	20	20	20	20	20	20	20

Table 6.8.2.8.g.

Mean Scale Scores Non-Hispanic White: Grade 7

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	947.49	942.22	937.11	934.41	942.52	938.55	943.83	939.51
SD	18.96	19.40	22.51	21.32	19.22	18.82	18.63	18.28
N	174	174	174	174	174	174	174	174

Table 6.8.2.8.h.

Mean Scale Scores Missing: Grade 7

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	954.03	949.43	944.21	943.09	949.36	946.48	950.83	947.13
SD	18.43	17.63	20.78	22.50	17.99	18.87	17.26	17.98
N	159	159	159	159	159	159	159	159

6.8.2.9. Grade 8

Table 6.8.2.9.a.

Mean Scale Scores for Hispanic (of any race): Grade 8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	953.62	949.96	943.59	942.26	948.83	946.34	951.07	946.88
SD	17.29	18.09	21.42	22.07	17.96	18.83	17.16	18.00
N	1,254	1,254	1,253	1,254	1,253	1,254	1,254	1,253

Table 6.8.2.9.b.

Mean Scale Scores Non-Hispanic American Indian: Grade 8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	952.29	944.14	945.57	950.57	949.14	947.71	946.57	947.86
SD	22.88	20.50	21.38	20.50	21.47	18.31	20.68	18.76
N	7	7	7	7	7	7	7	7

Table 6.8.2.9.c.

Mean Scale Scores Non-Hispanic Asian: Grade 8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	948.81	945.36	939.43	938.63	944.41	942.21	946.40	942.75
SD	18.70	21.48	22.12	24.21	19.05	21.61	20.04	20.30
N	246	246	245	246	245	246	246	245

Table 6.8.2.9.d.

Mean Scale Scores Non-Hispanic Black: Grade 8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	950.46	946.40	942.08	940.41	946.48	943.61	947.67	944.31
SD	17.38	19.45	22.20	23.53	18.29	20.65	18.30	19.45
N	134	134	133	133	133	133	134	133

Table 6.8.2.9.e.

Mean Scale Scores Non-Hispanic Multiracial: Grade 8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	947.10	939.75	938.60	933.50	942.95	936.85	941.90	938.45
SD	25.04	19.08	25.66	22.08	24.13	20.11	19.66	20.82
N	20	20	20	20	20	20	20	20

Table 6.8.2.9.f.

Mean Scale Scores Non-Hispanic Pacific Islander: Grade 8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	951.3	947.61	945.00	935.61	948.43	941.78	948.74	943.57
SD	14.3	17.72	19.68	24.19	16.13	19.15	16.17	17.52
N	23	23	23	23	23	23	23	23

Table 6.8.2.9.g.

Mean Scale Scores Non-Hispanic White: Grade 8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	948.57	944.43	938.49	936.01	943.74	940.43	945.71	941.23
SD	17.55	19.83	23.38	22.73	18.88	19.98	18.50	19.11
N	183	183	183	183	183	183	183	183

Table 6.8.2.9.h.

Mean Scale Scores Missing: Grade 8

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	958.86	953.93	950.90	949.65	955.08	952.02	955.43	952.80
SD	17.63	17.93	22.34	22.48	18.87	19.35	17.16	18.61
N	134	134	134	133	134	133	134	133

6.8.2.10. Grade 9

Table 6.8.2.10.a.

Mean Scale Scores for Hispanic (of any race): Grade 9

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	951.94	948.67	942.61	941.68	947.50	945.41	949.7	945.83
SD	17.53	19.39	24.13	22.18	19.45	19.57	18.2	19.02
N	1,177	1,177	1,176	1,176	1,176	1,176	1,177	1,175

Table 6.8.2.10.b.

Mean Scale Scores Non-Hispanic American Indian: Grade 9

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	954.86	953.0	946.71	946.93	951.07	950.21	953.64	950.36
SD	8.79	15.4	20.33	21.85	13.72	17.40	12.76	15.68
N	14	14	14	14	14	14	14	14

Table 6.8.2.10.c.

Mean Scale Scores Non-Hispanic Asian: Grade 9

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	952.35	948.05	943.62	942.05	948.20	945.30	949.38	945.94
SD	17.46	18.87	21.50	23.24	18.09	19.96	17.84	18.98
N	233	233	232	232	232	232	233	232

Table 6.8.2.10.d.

Mean Scale Scores Non-Hispanic Black: Grade 9

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	947.74	945.94	940.68	940.51	944.46	943.40	946.56	943.55
SD	20.07	21.52	25.14	23.76	21.42	21.98	20.64	21.30
N	127	126	127	126	127	126	126	126

Table 6.8.2.10.e.

Mean Scale Scores Non-Hispanic Multiracial: Grade 9

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	944.42	941.75	941.25	939.58	943.00	940.75	942.67	941.33
SD	19.19	20.37	22.34	20.91	20.21	19.24	19.81	19.11
N	12	12	12	12	12	12	12	12

Table 6.8.2.10.f.

Mean Scale Scores Non-Hispanic Pacific Islander: Grade 9

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	950.87	949.13	944.00	944.20	947.60	946.87	949.73	947.00
SD	13.07	18.24	27.67	23.46	18.72	20.59	15.36	19.46
N	15	15	15	15	15	15	15	15

Table 6.8.2.10.g.

Mean Scale Scores Non-Hispanic White: Grade 9

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	949.92	946.34	941.95	940.09	946.17	943.44	947.46	944.05
SD	18.61	18.39	23.86	20.53	19.97	18.31	17.87	18.15
N	166	166	166	166	166	166	166	166

Table 6.8.2.10.h.

Mean Scale Scores Missing: Grade 9

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	953.52	949.53	943.29	942.41	948.64	946.16	950.77	946.70
SD	19.33	21.28	25.59	24.47	20.98	22.01	20.11	21.19
N	167	167	166	167	166	167	167	166

6.8.2.11. Grade 10

Table 6.8.2.11.a.

Mean Scale Scores for Hispanic (of any race): Grade 10

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	954.26	950.45	945.23	943.78	950.00	947.34	951.66	947.95
SD	17.30	19.51	24.12	22.15	19.23	19.69	18.15	19.02
N	1,130	1,130	1,129	1,130	1,129	1,130	1,130	1,129

Table 6.8.2.11.b.

Mean Scale Scores Non-Hispanic American Indian: Grade 10

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	962.50	961.17	957.50	956.67	960.22	959.06	961.72	959.44
SD	13.97	18.27	23.16	25.81	17.25	21.42	15.77	19.58
N	18	18	18	18	18	18	18	18

Table 6.8.2.11.c.

Mean Scale Scores Non-Hispanic Asian: Grade 10

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	951.61	949.57	943.62	944.69	947.98	947.35	950.23	947.46
SD	17.49	19.89	23.66	22.71	19.05	20.00	18.62	19.14
N	195	195	194	195	194	195	195	194

Table 6.8.2.11.d.

Mean Scale Scores Non-Hispanic Black: Grade 10

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	952.47	950.03	946.80	942.04	949.91	946.29	950.80	947.13
SD	16.08	18.40	23.37	22.26	18.36	18.98	17.06	18.33
N	129	129	129	129	129	129	129	129

Table 6.8.2.11.e.

Mean Scale Scores Non-Hispanic Multiracial: Grade 10

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	949.08	948.69	938.92	941.62	944.23	945.38	948.85	944.69
SD	16.63	14.27	21.65	20.21	16.83	15.51	13.88	15.29
N	13	13	13	13	13	13	13	13

Table 6.8.2.11.f.

Mean Scale Scores Non-Hispanic Pacific Islander: Grade 10

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	953.93	951.93	950.21	949.64	952.36	951.0	952.64	951.21
SD	20.90	22.68	21.79	22.83	20.46	22.3	21.87	21.64
N	14	14	14	14	14	14	14	14

Table 6.8.2.11.g.

Mean Scale Scores Non-Hispanic White: Grade 10

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	949.74	946.95	942.91	938.79	946.53	943.11	947.85	943.91
SD	19.19	18.85	23.57	23.25	20.19	20.05	18.30	19.55
N	151	151	151	151	151	151	151	151

Table 6.8.2.11.h.

Mean Scale Scores Missing: Grade 10

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	954.85	952.75	948.65	946.19	951.97	949.67	953.42	950.13
SD	18.35	19.34	23.07	21.26	19.56	19.63	18.55	19.07
N	116	116	116	116	116	116	116	116

6.8.2.12. Grade 11

Table 6.8.2.12.a.

Mean Scale Scores for Hispanic (of any Race): Grade 11

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	953.30	949.98	944.29	943.52	949.02	946.96	951.03	947.39
SD	17.78	19.94	24.05	23.24	19.55	20.46	18.65	19.60
N	1,044	1,044	1,044	1,044	1,044	1,044	1,044	1,044

Table 6.8.2.12.b.

Mean Scale Scores Non-Hispanic American Indian: Grade 11

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	962.56	958.44	958.11	956.22	960.56	957.44	959.78	958.33
SD	14.45	19.89	23.48	24.85	17.34	21.82	18.01	20.42
N	9	9	9	9	9	9	9	9

Table 6.8.2.12.c.

Mean Scale Scores Non-Hispanic Asian: Grade 11

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	953.71	948.95	943.58	944.81	948.86	947.12	950.44	947.40
SD	15.03	18.78	22.81	22.45	17.48	19.37	16.90	18.34
N	194	194	194	194	194	194	194	194

Table 6.8.2.12.d.

Mean Scale Scores Non-Hispanic Black: Grade 11

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	950.24	947.44	944.04	942.68	947.39	945.29	948.31	945.69
SD	16.95	20.56	23.84	23.04	19.26	20.69	18.83	19.73
N	119	119	119	119	119	119	119	119

Table 6.8.2.12.e.

Mean Scale Scores Non-Hispanic Multiracial: Grade 11

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	945.33	942.33	929.78	925.33	937.78	934.11	943.33	934.89
SD	11.34	18.30	26.83	22.50	17.46	19.03	15.98	18.35
N	9	9	9	9	9	9	9	9

Table 6.8.2.12.f.

Mean Scale Scores Non-Hispanic Pacific Islander: Grade 11

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	957.31	950.54	943.92	946.38	950.85	948.69	952.54	949.15
SD	19.70	23.08	25.22	24.68	20.94	23.51	21.30	22.40
N	13	13	13	13	13	13	13	13

Table 6.8.2.12.g.

Mean Scale Scores Non-Hispanic White: Grade 11

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	952.36	947.62	942.80	941.03	947.79	944.54	949.09	945.28
SD	15.13	18.24	23.59	22.74	18.08	19.26	16.67	18.33
N	117	117	117	117	117	117	117	117

Table 6.8.2.12.h.

Mean Scale Scores Missing: Grade 11

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	956.21	953.21	950.10	946.21	953.39	949.93	954.17	950.74
SD	16.15	18.14	22.63	23.46	18.38	19.60	16.99	18.61
N	121	121	121	121	121	121	121	121

6.8.2.13. Grade 12

Table 6.8.2.13.a.

Mean Scale Scores for Hispanic (of any race): Grade 12

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	953.51	949.95	944.38	944.40	949.20	947.39	951.09	947.74
SD	17.58	20.02	23.56	22.64	19.17	20.22	18.64	19.42
N	1,579	1,578	1,577	1,576	1,577	1,576	1,578	1,575

Table 6.8.2.13.b.**Mean Scale Scores Non-Hispanic American Indian: Grade 12**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	957.08	955.08	954.85	949.77	956.15	952.54	955.85	953.46
SD	25.90	25.58	21.84	28.29	23.30	26.06	25.40	24.82
N	13	13	13	13	13	13	13	13

Table 6.8.2.13.c.**Mean Scale Scores Non-Hispanic Asian: Grade 12**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	950.89	949.60	941.8	944.39	946.59	947.24	950.06	946.80
SD	15.37	17.71	21.9	21.60	17.24	18.48	16.42	17.59
N	352	352	351	352	351	352	352	351

Table 6.8.2.13.d.**Mean Scale Scores Non-Hispanic Black: Grade 12**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	952.93	948.66	944.81	944.62	949.11	946.86	950.05	947.35
SD	17.49	20.54	23.48	23.27	19.24	20.75	18.89	19.68
N	190	189	190	189	190	189	189	189

Table 6.8.2.13.e.**Mean Scale Scores Non-Hispanic Multiracial: Grade 12**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	941.80	935.90	936.20	942.60	939.20	939.40	937.60	939.00
SD	25.75	25.25	25.95	26.19	24.76	25.32	24.96	24.86
N	10	10	10	10	10	10	10	10

Table 6.8.2.13.f.**Mean Scale Scores Non-Hispanic Pacific Islander: Grade 12**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	945.17	942.20	934.70	939.03	940.17	940.83	943.10	940.47
SD	21.16	22.84	27.28	25.91	22.62	23.23	21.36	22.56
N	30	30	30	30	30	30	30	30

Table 6.8.2.13.g.**Mean Scale Scores Non-Hispanic White: Grade 12**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	949.35	946.52	942.86	939.42	946.41	943.28	947.39	944.02
SD	16.80	17.99	21.54	21.47	18.07	18.72	17.04	17.99
N	191	191	190	190	190	190	191	190

Table 6.8.2.13.h.**Mean Scale Scores Missing: Grade 12**

Statistic	Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Mean	955.91	952.35	948.15	946.11	952.31	949.57	953.52	950.19
SD	16.65	18.36	21.84	22.28	17.95	19.16	17.24	18.23
N	195	195	194	194	194	194	195	194

6.9. *Scale Scores by Disability by Grade-Level Cluster*

Section 6.9. presents mean scale scores by disability type and by grade-level cluster. As mentioned in Section 6.5., a list of acronyms for each disability category to facilitate interpretation can be found in Table 6.5.1.b., and Table 6.5.1.a. shows the distribution of test takers across 15 primary and secondary disability categories.

Tables 6.9.a. through 6.9.h. provide mean scale scores by primary disability, across grade-level clusters, for each domain and composite score, showing how performance varies by disability and educational stage.

Listening: The lowest scale scores were observed in the following groups: students in grades K–2 with Multiple Disabilities (MD), students in grades 3–5 with Deaf-Blind (DB), students in grades 6–8 with Visual Impairment (VI), and students in grades 9–12 with Deaf-Blind (DB).

Reading: The lowest scores were seen in students in grades K–2 with Multiple Disabilities (MD), students in grades 3–5 with Deaf-Blind (DB), students in grades 6–8 with Visual Impairment (VI), and students in grades 9–12 with Deaf-Blind (DB).

Speaking: The lowest scores were observed for students in grades K–2 with Hearing Impairment (HI), students in grades 3–5 with Deaf-Blind (DB), students in grades 6–8 with Visual Impairment (VI), and students in grades 9–12 with Deaf-Blind (DB).

Writing: The lowest scores were observed for students in grades K–2 with Multiple Disabilities (MD), students in grades 3–5 with Deaf-Blind (DB), students in grades 6–8 with Deaf-Blind (DB), and students in grades 9–12 with Multiple Disabilities (MD).

Table 6.9.a.**Listening Mean Scale Scores by Disability by Grade-Level Cluster**

Primary Disability Code	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12
AS	935.08	942.61	948.87	952.35
DB	935.29	921.33	939.00	936.75
DD	938.34	948.26	953.17	953.42
ED	952.17	959.43	959.78	963.26
HI	936.82	950.29	952.22	949.50
ID	939.02	949.19	955.16	954.41
MD	929.89	937.27	942.25	941.62
OHI	935.75	943.36	954.43	955.27
OI	938.33	948.79	952.80	958.91
SLD	954.88	960.42	967.18	966.07
SLI	950.22	957.41	965.05	963.11
TBI	937.43	935.24	951.92	952.49
VI	939.44	930.23	935.27	952.50
NPD	935.20	946.29	951.12	954.85

Table 6.9.b.**Reading Mean Scale Scores by Disability by Grade-Level Cluster**

Primary Disability Code	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12
AS	933.49	940.96	946.32	950.71
DB	930.29	917.33	936.80	938.50
DD	935.26	942.64	948.10	949.46
ED	943.83	952.07	955.89	963.05
HI	933.23	951.05	947.67	944.64
ID	935.73	944.20	950.04	950.55
MD	925.86	932.99	937.54	936.46
OHI	932.07	938.88	947.64	950.78
OI	934.19	941.36	947.44	955.47
SLD	950.00	955.98	961.94	966.53
SLI	947.68	954.00	961.71	967.00
TBI	936.86	928.76	949.29	950.18
VI	934.12	924.15	932.55	948.75
NPD	931.74	942.85	946.80	951.41

Table 6.9.c.**Speaking Mean Scale Scores by Disability by Grade-Level Cluster**

Primary Disability Code	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12
AS	925.98	935.67	940.00	943.77
DB	926.29	920.33	936.80	928.00
DD	928.66	942.95	943.10	946.33
ED	942.17	958.86	953.44	955.58
HI	918.62	943.71	941.89	936.21
ID	926.04	942.29	945.90	945.73
MD	918.83	927.24	931.51	930.37
OHI	925.21	935.46	943.91	947.71
OI	925.85	940.03	946.68	950.24
SLD	946.17	958.30	962.31	964.75
SLI	943.46	953.41	957.97	959.96
TBI	939.43	932.00	944.67	945.90
VI	929.22	931.92	933.73	952.56
NPD	925.90	938.98	941.68	946.35

Table 6.9.d.**Writing Mean Scale Scores by Disability by Grade-Level Cluster**

Primary Disability Code	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12
AS	927.58	934.49	941.24	946.00
DB	923.86	908.50	926.40	934.00
DD	926.09	934.43	938.18	944.49
ED	935.83	944.43	956.00	954.42
HI	924.05	940.57	942.89	936.00
ID	923.23	935.16	941.65	943.81
MD	917.14	920.50	927.08	927.80
OHI	920.37	928.54	940.10	944.17
OI	919.56	931.82	934.42	941.06
SLD	944.77	952.63	962.58	963.77
SLI	943.28	953.08	957.38	962.21
TBI	925.36	922.65	939.88	944.36
VI	926.12	919.31	928.55	943.12
NPD	924.92	933.59	939.72	946.43

Table 6.9.e.**Oral Mean Scale Scores by Disability by Grade-Level Cluster**

Primary Disability Code	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12
AS	930.73	939.38	944.68	948.29
DB	930.86	921.00	938.20	932.50
DD	933.70	945.84	948.33	950.09
ED	947.50	959.43	956.67	959.63
HI	927.62	947.24	947.28	943.07
ID	932.74	945.97	950.77	950.32
MD	924.57	932.46	937.14	936.22
OHI	930.64	939.62	949.42	951.74
OI	932.30	944.62	950.00	954.79
SLD	950.72	959.61	964.88	965.63
SLI	947.05	955.64	961.89	961.79
TBI	938.64	933.76	948.42	949.51
VI	934.56	931.31	934.64	952.81
NPD	930.76	942.85	946.71	950.87

Table 6.9.f.**Literacy Mean Scale Scores by Disability by Grade-Level Cluster**

Primary Disability Code	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12
AS	930.76	937.98	944.02	948.59
DB	927.43	913.00	932.00	936.38
DD	930.91	938.77	943.39	947.20
ED	940.00	948.50	956.22	959.00
HI	928.86	946.05	945.56	940.50
ID	929.70	939.92	946.09	947.40
MD	921.73	926.90	932.53	932.33
OHI	926.40	933.93	944.07	947.66
OI	927.15	936.91	941.33	948.53
SLD	947.66	954.53	962.49	965.32
SLI	945.76	953.75	959.84	964.82
TBI	931.29	925.94	944.83	947.51
VI	930.38	921.85	930.64	946.12
NPD	928.54	938.45	943.53	949.18

Table 6.9.g.

Comprehension Mean Scale Scores by Disability by Grade-Level Cluster

Primary Disability Code	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12
AS	934.03	941.51	947.11	951.26
DB	931.86	918.50	937.60	938.12
DD	936.24	944.38	949.68	950.73
ED	946.33	954.36	957.11	963.21
HI	934.36	950.90	949.00	946.21
ID	936.76	945.75	951.60	951.77
MD	927.13	934.34	938.98	938.06
OHI	933.20	940.28	949.70	952.17
OI	935.48	943.61	948.96	956.62
SLD	951.54	957.39	963.48	966.48
SLI	948.49	955.08	962.74	965.89
TBI	937.00	930.71	950.12	950.95
VI	935.75	925.92	933.27	949.94
NPD	932.84	943.93	948.12	952.51

Table 6.9.h.

Overall Mean Scale Scores by Disability by Cluster

Primary Disability Code	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12
AS	930.56	938.22	944.01	948.28
DB	928.14	915.33	933.60	935.12
DD	931.54	940.66	944.69	947.87
ED	942.17	951.57	956.11	958.95
HI	928.10	946.19	945.83	941.14
ID	930.44	941.51	947.28	948.08
MD	922.42	928.37	933.74	933.31
OHI	927.46	935.45	945.50	948.70
OI	928.41	938.94	943.79	950.12
SLD	948.19	955.82	963.04	965.28
SLI	946.11	954.07	960.24	963.75
TBI	933.36	928.18	945.67	947.90
VI	931.50	924.54	931.82	947.75
NPD	929.06	939.55	944.33	949.49

6.10. Correlations among Scale Scores by Grade-Level Cluster

Section 6.10. presents the correlations among scale scores by grade-level cluster.

Tables 6.10.a. through 6.10.d. display the Pearson correlations between scale scores across the four domains (Listening, Reading, Writing, and Speaking) for each of the four grade-level clusters, along with the corresponding sample sizes.

These tables provide Pearson correlations among the four domain scale scores across all tiers, along with the number of students included in each correlation. The pattern of correlations between domains varies across clusters.

Across all clusters, the domain correlations generally range from 0.60 to 0.80. The correlation between Reading and Listening is the highest, while Writing tends to have the lowest correlations with other domains. The correlations among Reading, Speaking, and Writing are typically between 0.70 and 0.79. Notably, the grade-level cluster for grades 9–12 exhibits the highest correlations compared to other clusters.

Table 6.10.a.

Correlations Among Scale Scores: K-2

Domain	Statistic	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.85	0.712	0.645
Listening	N	9,858	9,855	9,842	9,850
Reading	Pearson Correlation	NA	1	0.736	0.711
Reading	N	NA	9,855	9,840	9,850
Speaking	Pearson Correlation	NA	NA	1	0.703
Speaking	N	NA	NA	9,842	9,836
Writing	Pearson Correlation	NA	NA	NA	1
Writing	N	NA	NA	NA	9,850

Table 6.10.b.

Correlations Among Scale Scores: 3–5

Domain	Statistic	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.816	0.721	0.661
Listening	N	8,607	8,600	8,599	8,593
Reading	Pearson Correlation	NA	1	0.728	0.729
Reading	N	NA	8,601	8,595	8,592
Speaking	Pearson Correlation	NA	NA	1	0.737
Speaking	N	NA	NA	8,600	8,590
Writing	Pearson Correlation	NA	NA	NA	1
Writing	N	NA	NA	NA	8,594

Table 6.10.c.**Correlations Among Scale Scores: 6–8**

Domain	Statistic	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.827	0.719	0.675
Listening	N	6,521	6,521	6,511	6,513
Reading	Pearson Correlation	NA	1	0.737	0.755
Reading	N	NA	6,521	6,511	6,513
Speaking	Pearson Correlation	NA	NA	1	0.739
Speaking	N	NA	NA	6,511	6,508
Writing	Pearson Correlation	NA	NA	NA	1
Writing	N	NA	NA	NA	6,513

Table 6.10.d.**Correlations Among Scale Scores: 9–12**

Domain	Statistic	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.827	0.735	0.716
Listening	N	7,863	7,860	7,853	7,854
Reading	Pearson Correlation	NA	1	0.757	0.789
Reading	N	NA	7,860	7,851	7,854
Speaking	Pearson Correlation	NA	NA	1	0.787
Speaking	N	NA	NA	7,853	7,848
Writing	Pearson Correlation	NA	NA	NA	1
Writing	N	NA	NA	NA	7,854

6.11. Proficiency Level Results

6.11.1. Proficiency Level by Grade-Level Cluster

Section 6.11, Proficiency Level Results, displays the distribution of students' language proficiency levels by grade-level cluster (Tables 6.11.1.a.–6.11.1.h.) and grade (Tables 6.11.2.a.–6.11.2.h.), with each sub-table presenting results by domain or composite.

Listening: Table 6.11.1.a. highlights the distribution of students across proficiency levels in Listening. Proficiency Level 1 (P1) accounts for the largest proportion of students at 28.67% (9,419 students), particularly dominant in the K–2 cluster (41.13%), and steadily decreasing to 21.84% in grades 9–12. Proficiency Level 3 (P3) is the second-largest group, representing 22.67% (7,448 students), with increasing proportions in higher clusters, peaking at 26.99% in grades 9–12. Higher proficiency levels (P4 and P5) grow in the older clusters, with P5 reaching 22.26% in grades 9–12, compared to just 9.12% in K–2.

Table 6.11.1.a.**Proficiency Level by Grade-Level Cluster: Listening**

Cluster	P1 Count	P1 % Within PL	P2 Count	P2 % Within PL	P3 Count	P3 % Within PL	P4 Count	P4 % Within PL	P5 Count	P5 % Within PL	Total
K-2	4,055	41.13%	983	9.97%	1,616	16.39%	2,305	23.38%	899	9.12%	9,858
3-5	2,265	26.32%	1,553	18.04%	2,049	23.81%	1,731	20.11%	1,009	11.72%	8,607
6-8	1,382	21.19%	998	15.30%	1,661	25.47%	911	13.97%	1,569	24.06%	6,521
9-12	1,717	21.84%	1,143	14.54%	2,122	26.99%	1,131	14.38%	1,750	22.26%	7,863
Total	9,419	28.67%	4,677	14.24%	7,448	22.67%	6,078	18.50%	5,227	15.91%	32,849

Reading: Table 6.11.1.b. shows a dominance of P1, representing 40.69% (13,361 students) overall. This proportion declines significantly from 56.24% in K-2 to 29.94% in grades 9-12. Proficiency Level 3 (P3) becomes more prominent in higher clusters, peaking at 21.58% in grades 6-8 and remaining high in grades 9-12 (20.64%). Higher proficiency levels (P4 and P5) increase notably, with P5 reaching 18.52% in grades 9-12, compared to 7.94% in K-2.

Table 6.11.1.b.**Proficiency Level by Grade-Level Cluster: Reading**

Cluster	P1 Count	P1 % Within PL	P2 Count	P2 % Within PL	P3 Count	P3 % Within PL	P4 Count	P4 % Within PL	P5 Count	P5 % Within PL	Total
K-2	5,542	56.24%	1,638	16.62%	1,193	12.11%	700	7.10%	782	7.94%	9,855
3-5	3,499	40.68%	2,158	25.09%	1,320	15.35%	834	9.70%	790	9.18%	8,601
6-8	1,967	30.16%	1,072	16.44%	1,407	21.58%	1,254	19.23%	821	12.59%	6,521
9-12	2,353	29.94%	1,171	14.90%	1,622	20.64%	1,258	16.01%	1,456	18.52%	7,860
Total	13,361	40.69%	6,039	18.39%	5,542	16.88%	4,046	12.32%	3,849	11.72%	32,837

Speaking: Table 6.11.1.c. reveals that P1 dominates with 57.51% (18,868 students), particularly in K-2 (71.77%) and decreasing to 47.50% in grades 9-12. Proficiency Level 2 (P2) increases in mid-grade clusters, peaking at 22.07% in grades 6-8, and then declines in higher grades (16.81% in grades 9-12). Proficiency levels P4 and P5 increase gradually in higher clusters, with P5 peaking at 14.96% in grades 9-12 compared to 3.94% in K-2.

Table 6.11.1.c.**Proficiency Level by Grade-Level Cluster: Speaking**

Cluster	P1 Count	P1 % Within PL	P2 Count	P2 % Within PL	P3 Count	P3 % Within PL	P4 Count	P4 % Within PL	P5 Count	P5 % Within PL	Total
K-2	7,064	71.77%	1,299	13.20%	863	8.77%	228	2.32%	388	3.94%	9,842
3-5	5,073	58.99%	1,394	16.21%	869	10.10%	392	4.56%	872	10.14%	8,600
6-8	3,001	46.09%	1,437	22.07%	1,083	16.63%	240	3.69%	750	11.52%	6,511
9-12	3,730	47.50%	1,320	16.81%	832	10.59%	796	10.14%	1,175	14.96%	7,853
Total	18,868	57.51%	5,450	16.61%	3,647	11.12%	1,656	5.05%	3,185	9.71%	32,806

Writing: Table 6.11.1.d. highlights that P1 remains the largest group at 60.44% (19,831 students) but decreases from 72.03% in K-2 to 52.11% in grades 9-12. Proficiency Level 2 (P2) peaks in grades 3-5 (21.14%) and remains stable in higher clusters. Proficiency Level 5 (P5) shows consistent growth in higher clusters, increasing from 5.27% in K-2 to 11.05% in grades 9-12, while P4 follows a similar trend, peaking at 7.45% in grades 9-12.

Table 6.11.1.d.**Proficiency Level by Grade-Level Cluster: Writing**

Cluster	P1 Count	P1 % Within PL	P2 Count	P2 % Within PL	P3 Count	P3 % Within PL	P4 Count	P4 % Within PL	P5 Count	P5 % Within PL	Total
K-2	7,095	72.03%	674	6.84%	1,004	10.19%	558	5.66%	519	5.27%	9,850
3-5	5,045	58.70%	1,817	21.14%	739	8.60%	403	4.69%	590	6.87%	8,594
6-8	3,598	55.24%	1,144	17.56%	606	9.30%	490	7.52%	675	10.36%	6,513
9-12	4,093	52.11%	1,393	17.74%	915	11.65%	585	7.45%	868	11.05%	7,854
Total	19,831	60.44%	5,028	15.32%	3,264	9.95%	2,036	6.21%	2,652	8.08%	32,811

Composites: For Oral, Literacy, and Comprehension composites, whose data is presented in Tables 6.11.1.e, 6.11.1.f., and 6.11.1.g., respectively, the trends are similar. Proficiency Level 1 (P1) is dominant in younger clusters (K-2 and 3-5) and decreases in higher clusters (6-8 and 9-12). Higher proficiency levels (P4 and P5) consistently grow in older clusters, especially in grades 9-12, reflecting overall improvement.

Table 6.11.1.e.**Proficiency Level by Grade-Level Cluster: Oral**

Cluster	P1 Count	P1 % Within PL	P2 Count	P2 % Within PL	P3 Count	P3 % Within PL	P4 Count	P4 % Within PL	P5 Count	P5 % Within PL	Total
K-2	5,477	55.65%	1,670	16.97%	1,604	16.30%	626	6.36%	465	4.72%	9,842
3-5	3,528	41.03%	2,125	24.71%	1,300	15.12%	734	8.54%	912	10.61%	8,599
6-8	2,196	33.73%	1,335	20.50%	1,376	21.13%	451	6.93%	1,153	17.71%	6,511
9-12	2,897	36.89%	1,377	17.53%	1,271	16.18%	755	9.61%	1,553	19.78%	7,853
Total	14,098	42.98%	6,507	19.84%	5,551	16.92%	2,566	7.82%	4,083	12.45%	32,805

Table 6.11.1.f.**Proficiency Level by Grade-Level Cluster: Literacy**

Cluster	P1 Count	P1 % Within PL	P2 Count	P2 % Within PL	P3 Count	P3 % Within PL	P4 Count	P4 % Within PL	P5 Count	P5 % Within PL	Total
K-2	6,807	69.11%	1,009	10.24%	988	10.03%	584	5.93%	462	4.69%	9,850
3-5	4,711	54.83%	1,809	21.05%	956	11.13%	596	6.94%	520	6.05%	8,592
6-8	2,934	45.05%	1,364	20.94%	793	12.18%	824	12.65%	598	9.18%	6,513
9-12	3,525	44.88%	1,325	16.87%	908	11.56%	1,144	14.57%	952	12.12%	7,854
Total	17,977	54.79%	5,507	16.79%	3,645	11.11%	3,148	9.59%	2,532	7.72%	32,809

Table 6.11.1.g.**Proficiency Level by Grade-Level Cluster: Comprehension**

Cluster	P1 Count	P1 % Within PL	P2 Count	P2 % Within PL	P3 Count	P3 % Within PL	P4 Count	P4 % Within PL	P5 Count	P5 % Within PL	Total
K-2	5,137	52.13%	1,407	14.28%	1,675	17%	990	10.05%	646	6.56%	9,855
3-5	3,143	36.55%	1,689	19.64%	1,947	22.64%	1,151	13.38%	670	7.79%	8,600
6-8	1,844	28.28%	897	13.76%	1,502	23.03%	1,399	21.45%	879	13.48%	6,521
9-12	2,186	27.81%	1,063	13.52%	1,838	23.38%	1,403	17.85%	1,370	17.43%	7,860
Total	12,310	37.49%	5,056	15.40%	6,962	21.20%	4,943	15.05%	3,565	10.86%	32,836

Overall Composite: Table 6.11.1.h. reveals that P1 dominates across clusters, accounting for 52.18% (17,104 students) overall. The proportion of P1 decreases from 65.79% in K-2 to 43.27% in grades 9-12, reflecting overall improvement in student performance as grade-level clusters increase. Proficiency Level 5 (P5) shows significant growth, rising from 3.70% in K-2 to 13.25% in grades 9-12, indicating that more students achieve higher proficiency levels in older clusters. Proficiency Level 4 (P4) also increases, from 6.57% in K-2 to 13.43% in grades 9-12, while P3 proportions remain relatively stable, peaking at 17.81% in grades 6-8. Overall, the Overall Composite reflects clear growth in proficiency as students' progress through grade-level clusters, with increasing representation of higher proficiency levels (P4 and P5) and a steady decline in lower levels (P1 and P2).

Table 6.11.1.h.

Proficiency Level by Grade-Level Cluster: Overall

Cluster	P1 Count	P1 % Within PL	P2 Count	P2 % Within PL	P3 Count	P3 % Within PL	P4 Count	P4 % Within PL	P5 Count	P5 % Within PL	Total
K-2	6,471	65.79%	1,248	12.69%	1,107	11.25%	646	6.57%	364	3.70%	9,836
3-5	4,466	52%	1,703	19.83%	1,232	14.35%	633	7.37%	554	6.45%	8,588
6-8	2,771	42.58%	1,193	18.33%	1,159	17.81%	764	11.74%	621	9.54%	6,508
9-12	3,396	43.27%	1,218	15.52%	1,140	14.53%	1,054	13.43%	1,040	13.25%	7,848
Total	17,104	52.18%	5,362	16.36%	4,638	14.15%	3,097	9.45%	2,579	7.87%	32,780

6.11.2. Proficiency Level by Grade

Table 6.11.2.a.

Proficiency Level by Grade: Listening

Grade	PL1 Count	PL1 % Within Grade	PL2 Count	PL2 % Within Grade	PL3 Count	PL3 % Within Grade	PL4 Count	PL4 % Within Grade	PL5 Count	PL5 % Within Grade	Total
0	1483	55.81%	278	10.46%	396	14.90%	416	15.66%	84	3.16%	2657
1	1510	40.01%	404	10.70%	636	16.85%	896	23.74%	328	8.69%	3774
2	1062	30.99%	301	8.78%	584	17.04%	993	28.98%	487	14.21%	3427
3	917	30.55%	583	19.42%	715	23.82%	533	17.75%	254	8.46%	3002
4	744	25.11%	551	18.60%	693	23.39%	611	20.62%	364	12.28%	2963
5	604	22.86%	419	15.86%	641	24.26%	587	22.22%	391	14.80%	2642
6	547	22.81%	409	17.06%	624	26.02%	320	13.34%	498	20.77%	2398
7	434	20.45%	302	14.23%	553	26.06%	301	14.18%	532	25.07%	2122
8	401	20.04%	287	14.34%	484	24.19%	290	14.49%	539	26.94%	2001
9	453	23.70%	298	15.59%	502	26.27%	260	13.61%	398	20.83%	1911
10	373	21.12%	253	14.33%	459	25.99%	253	14.33%	428	24.24%	1766
11	344	21.16%	218	13.41%	463	28.47%	228	14.02%	373	22.94%	1626
12	547	21.37%	374	14.61%	698	27.27%	390	15.23%	551	21.52%	2560
Total	9419	28.67%	4677	14.24%	7448	22.67%	6078	18.50%	5227	15.91%	32849

Table 6.11.2.b.**Proficiency Level by Grade: Reading**

Grade	PL1 Count	PL1 % Within Grade	PL2 Count	PL2 % Within Grade	PL3 Count	PL3 % Within Grade	PL4 Count	PL4 % Within Grade	PL5 Count	PL5 % Within Grade	Total
0	1903	71.65%	370	13.93%	208	7.83%	91	3.43%	84	3.16%	2656
1	2107	55.86%	660	17.50%	478	12.67%	259	6.87%	268	7.10%	3772
2	1532	44.70%	608	17.74%	507	14.79%	350	10.21%	430	12.55%	3427
3	1382	46.08%	782	26.08%	428	14.27%	241	8.04%	166	5.54%	2999
4	1176	39.69%	731	24.67%	488	16.47%	278	9.38%	290	9.79%	2963
5	941	35.66%	645	24.44%	404	15.31%	315	11.94%	334	12.66%	2639
6	778	32.44%	406	16.93%	542	22.60%	432	18.02%	240	10.01%	2398
7	615	28.98%	364	17.15%	461	21.72%	418	19.70%	264	12.44%	2122
8	574	28.69%	302	15.09%	404	20.19%	404	20.19%	317	15.84%	2001
9	586	30.68%	296	15.50%	429	22.46%	287	15.03%	312	16.34%	1910
10	514	29.11%	242	13.70%	373	21.12%	292	16.53%	345	19.54%	1766
11	477	29.34%	223	13.71%	326	20.05%	296	18.20%	304	18.70%	1626
12	776	30.34%	410	16.03%	494	19.31%	383	14.97%	495	19.35%	2558
Total	13361	40.69%	6039	18.39%	5542	16.88%	4046	12.32%	3849	11.72%	32837

Table 6.11.2.c.**Proficiency Level by Grade: Speaking**

Grade	PL1 Count	PL1 % Within Grade	PL2 Count	PL2 % Within Grade	PL3 Count	PL3 % Within Grade	PL4 Count	PL4 % Within Grade	PL5 Count	PL5 % Within Grade	Total
0	2206	83.21%	275	10.37%	121	4.56%	21	0.79%	28	1.06%	2651
1	2722	72.26%	487	12.93%	326	8.65%	91	2.42%	141	3.74%	3767
2	2136	62.38%	537	15.68%	416	12.15%	116	3.39%	219	6.40%	3424
3	1893	63.08%	495	16.49%	270	9.0%	107	3.57%	236	7.86%	3001
4	1706	57.65%	507	17.13%	303	10.24%	136	4.60%	307	10.38%	2959
5	1474	55.83%	392	14.85%	296	11.21%	149	5.64%	329	12.46%	2640
6	1155	48.25%	516	21.55%	369	15.41%	95	3.97%	259	10.82%	2394
7	947	44.69%	509	24.02%	360	16.99%	75	3.54%	228	10.76%	2119
8	899	44.99%	412	20.62%	354	17.72%	70	3.50%	263	13.16%	1998
9	958	50.21%	307	16.09%	198	10.38%	177	9.28%	268	14.05%	1908
10	792	44.90%	307	17.40%	183	10.37%	184	10.43%	298	16.89%	1764
11	760	46.74%	254	15.62%	182	11.19%	182	11.19%	248	15.25%	1626
12	1220	47.75%	452	17.69%	269	10.53%	253	9.90%	361	14.13%	2555
Total	18868	57.51%	5450	16.61%	3647	11.12%	1656	5.05%	3185	9.71%	32806

Table 6.11.2.d.**Proficiency Level by Grade: Writing**

Grade	PL1 Count	PL1 % Within Grade	PL2 Count	PL2 % Within Grade	PL3 Count	PL3 % Within Grade	PL4 Count	PL4 % Within Grade	PL5 Count	PL5 % Within Grade	Total
0	2296	86.51%	137	5.16%	124	4.67%	54	2.03%	43	1.62%	2654
1	2715	72.02%	256	6.79%	410	10.88%	213	5.65%	176	4.67%	3770
2	2084	60.83%	281	8.20%	470	13.72%	291	8.49%	300	8.76%	3426
3	1895	63.27%	639	21.34%	222	7.41%	98	3.27%	141	4.71%	2995
4	1733	58.57%	626	21.16%	255	8.62%	143	4.83%	202	6.83%	2959
5	1417	53.67%	552	20.91%	262	9.92%	162	6.14%	247	9.36%	2640
6	1388	57.98%	422	17.63%	206	8.60%	145	6.06%	233	9.73%	2394
7	1152	54.34%	385	18.16%	215	10.14%	168	7.92%	200	9.43%	2120
8	1058	52.93%	337	16.86%	185	9.25%	177	8.85%	242	12.11%	1999
9	1034	54.19%	355	18.61%	227	11.90%	122	6.39%	170	8.91%	1908
10	922	52.21%	310	17.55%	205	11.61%	129	7.30%	200	11.33%	1766
11	830	51.05%	291	17.90%	175	10.76%	132	8.12%	198	12.18%	1626
12	1307	51.17%	437	17.11%	308	12.06%	202	7.91%	300	11.75%	2554
Total	19831	60.44%	5028	15.32%	3264	9.95%	2036	6.21%	2652	8.08%	32811

Table 6.11.2.e.**Proficiency Level by Grade: Oral**

Grade	PL1 Count	PL1 % Within Grade	PL2 Count	PL2 % Within Grade	PL3 Count	PL3 % Within Grade	PL4 Count	PL4 % Within Grade	PL5 Count	PL5 % Within Grade	Total
0	1858	70.09%	411	15.50%	281	10.60%	63	2.38%	38	1.43%	2651
1	2080	55.22%	652	17.31%	633	16.80%	244	6.48%	158	4.19%	3767
2	1539	44.95%	607	17.73%	690	20.15%	319	9.32%	269	7.86%	3424
3	1361	45.37%	768	25.60%	429	14.30%	222	7.40%	220	7.33%	3000
4	1194	40.35%	720	24.33%	454	15.34%	253	8.55%	338	11.42%	2959
5	973	36.86%	637	24.13%	417	15.80%	259	9.81%	354	13.41%	2640
6	852	35.59%	516	21.55%	492	20.55%	158	6.60%	376	15.71%	2394
7	692	32.66%	426	20.10%	472	22.27%	140	6.61%	389	18.36%	2119
8	652	32.63%	393	19.67%	412	20.62%	153	7.66%	388	19.42%	1998
9	763	39.99%	325	17.03%	300	15.72%	167	8.75%	353	18.50%	1908
10	634	35.94%	285	16.16%	288	16.33%	168	9.52%	389	22.05%	1764
11	581	35.73%	279	17.16%	270	16.61%	166	10.21%	330	20.30%	1626
12	919	35.97%	488	19.10%	413	16.16%	254	9.94%	481	18.83%	2555
Total	14098	42.98%	6507	19.84%	5551	16.92%	2566	7.82%	4083	12.45%	32805

Table 6.11.2.f.**Proficiency Level by Grade: Literacy**

Grade	PL1 Count	PL1 % Within Grade	PL2 Count	PL2 % Within Grade	PL3 Count	PL3 % Within Grade	PL4 Count	PL4 % Within Grade	PL5 Count	PL5 % Within Grade	Total
0	2233	84.14%	204	7.69%	131	4.94%	49	1.85%	37	1.39%	2654
1	2599	68.94%	398	10.56%	401	10.64%	216	5.73%	156	4.14%	3770
2	1975	57.65%	407	11.88%	456	13.31%	319	9.31%	269	7.85%	3426
3	1794	59.92%	633	21.14%	312	10.42%	148	4.94%	107	3.57%	2994
4	1606	54.28%	631	21.32%	317	10.71%	226	7.64%	179	6.05%	2959
5	1311	49.68%	545	20.65%	327	12.39%	222	8.41%	234	8.87%	2639
6	1161	48.50%	512	21.39%	265	11.07%	272	11.36%	184	7.69%	2394
7	928	43.77%	440	20.75%	303	14.29%	272	12.83%	177	8.35%	2120
8	845	42.27%	412	20.61%	225	11.26%	280	14.01%	237	11.86%	1999
9	898	47.06%	330	17.30%	233	12.21%	258	13.52%	189	9.91%	1908
10	775	43.88%	303	17.16%	218	12.34%	240	13.59%	230	13.02%	1766
11	709	43.60%	271	16.67%	187	11.50%	255	15.68%	204	12.55%	1626
12	1143	44.75%	421	16.48%	270	10.57%	391	15.31%	329	12.88%	2554
Total	17977	54.79%	5507	16.79%	3645	11.11%	3148	9.59%	2532	7.72%	32809

Table 6.11.2.g.**Proficiency Level by Grade: Comprehension**

Grade	PL1 Count	PL1 % Within Grade	PL2 Count	PL2 % Within Grade	PL3 Count	PL3 % Within Grade	PL4 Count	PL4 % Within Grade	PL5 Count	PL5 % Within Grade	Total
0	1800	67.77%	350	13.18%	312	11.75%	133	5.01%	61	2.30%	2656
1	1938	51.38%	561	14.87%	691	18.32%	364	9.65%	218	5.78%	3772
2	1399	40.82%	496	14.47%	672	19.61%	493	14.39%	367	10.71%	3427
3	1257	41.93%	609	20.31%	658	21.95%	334	11.14%	140	4.67%	2998
4	1042	35.17%	597	20.15%	675	22.78%	407	13.74%	242	8.17%	2963
5	844	31.98%	483	18.30%	614	23.27%	410	15.54%	288	10.91%	2639
6	743	30.98%	338	14.10%	556	23.19%	493	20.56%	268	11.18%	2398
7	564	26.58%	298	14.04%	509	23.99%	474	22.34%	277	13.05%	2122
8	537	26.84%	261	13.04%	437	21.84%	432	21.59%	334	16.69%	2001
9	546	28.59%	271	14.19%	466	24.40%	342	17.91%	285	14.92%	1910
10	472	26.73%	227	12.85%	409	23.16%	322	18.23%	336	19.03%	1766
11	439	27%	202	12.42%	387	23.80%	312	19.19%	286	17.59%	1626
12	729	28.50%	363	14.19%	576	22.52%	427	16.69%	463	18.10%	2558
Total	12310	37.49%	5056	15.40%	6962	21.20%	4943	15.05%	3565	10.86%	32836

Table 6.11.2.h.**Proficiency Level by Grade: Overall**

Grade	PL1 Count	PL1 % Within Grade	PL2 Count	PL2 % Within Grade	PL3 Count	PL3 % Within Grade	PL4 Count	PL4 % Within Grade	PL5 Count	PL5 % Within Grade	Total
0	2156	81.39%	264	9.97%	152	5.74%	56	2.11%	21	0.79%	2649
1	2473	65.70%	489	12.99%	445	11.82%	230	6.11%	127	3.37%	3764
2	1842	53.81%	495	14.46%	510	14.90%	360	10.52%	216	6.31%	3423
3	1712	57.20%	603	20.15%	400	13.36%	164	5.48%	114	3.81%	2993
4	1510	51.07%	595	20.12%	425	14.37%	234	7.91%	193	6.53%	2957
5	1244	47.16%	505	19.14%	407	15.43%	235	8.91%	247	9.36%	2638
6	1083	45.26%	456	19.06%	409	17.09%	252	10.53%	193	8.07%	2393
7	873	41.22%	397	18.74%	410	19.36%	253	11.95%	185	8.73%	2118
8	815	40.81%	340	17.03%	340	17.03%	259	12.97%	243	12.17%	1997
9	861	45.17%	302	15.84%	296	15.53%	239	12.54%	208	10.91%	1906
10	741	42.01%	279	15.82%	260	14.74%	232	13.15%	252	14.29%	1764
11	684	42.07%	243	14.94%	252	15.50%	222	13.65%	225	13.84%	1626
12	1110	43.50%	394	15.44%	332	13.01%	361	14.15%	355	13.91%	2552
Total	17104	52.18%	5362	16.36%	4638	14.15%	3097	9.45%	2579	7.87%	32780

7. Analysis of Domain

Section 7 provides some background on the technical measurement and statistical tools used to analyze WIDA Alternate ACCESS. Then it explains the results for each domain and grade-level cluster.

7.1 Tools Used for Analysis

7.1.1. Rasch Model for Scoring

The measurement model that forms the basis of the analysis for the development of Alternate ACCESS is the Rasch measurement model (Wright and Stone, 1979). Additional information on its use in the development of the test is available in *WIDA Technical Report 1, Alternate ACCESS for ELLs TM, Series 100 Development and Operational Field Test: Technical Report*. The test was developed using Rasch measurement principles, and in that sense the Rasch model guided all decisions throughout the development of the assessment and was not merely a tool for the statistical analysis of the data. For example, data based on Rasch fit statistics guided the inclusion, revision, or deletion of items during the development and field testing of the test forms and will continue to guide the refinement and further development of the test. For all domains, a Rasch Rating Scale model was used. Mathematically, this can be represented as

$$\log \left(\frac{P_{nik}}{P_{nik-1}} \right) = B_n - D_i - F_k$$

where

P_{nik} = probability of person "n" on task "i" receiving a rating at level "k" on the rating scale

P_{nik-1} = probability of person "n" on task "i" receiving a rating at level "k - 1" on the rating scale (i.e., the next lowest rating)

B_n = ability of person "n"

D_i = difficulty of task "i"

F_k = calibration of step "k" on the rating scale

All Rasch analyses were conducted using the Rasch measurement software program Winsteps 3.92.1 (Linacre, 2006). When speaking of the measure of student ability, we use the term "ability measure" (rather than "theta," used commonly when discussing models based on item response theory). When speaking of the measure of how hard an item is, we use the term item "difficulty measure" (rather than "b parameter," used commonly when discussing models based on item response theory). Step measures refer to the calibration of the steps in the Rasch rating scale model previously presented. All three measures (ability, difficulty, and step) are expressed in terms of Rasch logits, which then are converted into scores on the ACCESS score scale for reporting purposes.

Fit statistics for the Rasch model are calculated by comparing the observed empirical data with the data that the Rasch model would be expected to produce if the data fit the model perfectly. Outfit mean square statistics for items and tasks are influenced by outlier responses for machine scored dichotomous items or outlier ratings for rater-scored performance tasks. For example, a difficult item that some low-ability students get correct—for reasons unknown—will have a high outfit mean square statistic. Similarly, an easy item that some high-ability students get wrong will also have a high outfit mean square statistic. Infit mean square statistics are influenced by unexpected patterns of students' responses and ratings on items and tasks that are roughly targeted for them and generally indicate a more serious measurement problem. The expectation for both statistics is 1.00, and values near 1.00 are not of great concern. Values less than 1.00 indicate that the response and rating patterns are too predictable and thus redundant, or the model is overfitting the data, but are not of great concern. High values are of greater concern.

Linacre (2002) provided more guidance on how to interpret these statistics for dichotomous items. According to Linacre (2002):

- values greater than 2.0 "distort or degrade the measurement system"
- values between 1.5 and 2.0 are "unproductive for construction of measurement, but not degrading"
- values between 0.5 and 1.5 should be considered "productive for measurement"
- values below 0.5 are considered "less productive for measurement, but not degrading"

Linacre also stated in this guidance that infit problems are more serious to the construction of measurement than are outfit problems. Because conservative guidelines were followed in the development of Alternate ACCESS, 85% of the test items have infit statistics within the range of 0.5 to 1.5, aligning with the standards for being "productive for measurement" as defined by the aforementioned guidelines.

7.1.2. Sampling

The results presented in most of the tables in Section 7 are based on the full data set of all students who were administered operational Series 602 of Alternate ACCESS in the academic year 2023–2024. The item analysis summary tables, DIF analysis summary, and the raw score-to-scale score conversion tables use item difficulties from this calibration. Equated tables are based on the spring verification data set. For spring verification post equating, WIDA drew early testing student data (N=16,213) across all four domains. Per cluster and domain, there were around N=3,000–4,000 range of students in the analysis.

7.1.3. Scaling

The table below provides the scaling equation for each domain. This equation is used to convert an examinee’s ability measure into the scale score. Each equation is used across all grade-level clusters within each domain. For detailed scaling procedures, please refer to Section 4.4 Scaling.

Table 7.1.3.

Scaling Equations for Each Domain

Domain	Scale Score
Listening	$(\text{Ability Measure in Logits} \times 7.948) + 942.606$
Reading	$(\text{Ability Measure in Logits} \times 7.495) + 940.879$
Speaking	$(\text{Ability Measure in Logits} \times 7.678) + 941.392$
Writing	$(\text{Ability Measure in Logits} \times 7.297) + 943.625$

The following paragraphs describe the subsequent tables and figures that appear in Section 7. Each description applies to each test form in each domain. Information on raw and scale score descriptive statistics, proficiency level distribution, and the equating summary are displayed for each domain.

7.2. *Complete Item or Task Analysis Summary*

Section 7.2 provides a comprehensive summary of item analyses.

In the first table of each series, denoted with “a.”, the item summary presents an overview of the items on the test form. The first column identifies the type of item, categorized as either MOSR (Multiple Opportunities for Selected Response) in Listening and Reading domains or MOCR (Multiple Opportunities for Constructed Response) in Speaking and Writing domains. The next columns include the number of items on the test form and the average difficulty value for these items, expressed in logits. Following these, the average of P-values across all items is displayed, where P-values represent the average of raw scores. This is a useful measure for understanding overall task performance. The final two columns present Rasch model item fit statistics, specifically the average infit mean square statistic and outfit mean square statistic. These statistics assess how well the data align with the Rasch model, with optimal values close to 1.00. Additionally, the step value summary in the tables denoted with “b.” provides step value estimates along with corresponding infit and outfit statistics for each raw score point per grade-level cluster and domain.

The third table in each series, denoted with “c.”, is a detailed table summarizing the analyses for all items or tasks on the test form. The first column provides descriptive names for each item, which vary by domain. These names include characters representing the domain (e.g., “R” for Reading), the target language proficiency level (e.g., “P2”), language standard (e.g., “LA”), key word for item theme (e.g., “Ball”), numeric item ID (e.g., “21603”). The second column indicates expected PL. The third column indicates language standard. The fourth column indicates item difficulty measure in logits. The fifth column indicates whether the item served as a common anchor item to align the measurement scale with field test results. The sixth column includes the average raw score points for polytomous items which reflect task difficulty, with higher values indicating easier tasks. The seventh and eighth columns present Rasch item fit statistics, including infit and outfit measures, to evaluate model-data alignment. Finally, the last column provides the point measure correlation for each item, a statistic that measures how well an item distinguishes between high- and low-performing test-takers. This correlation is a critical indicator of an item’s utility in contributing to the overall reliability and validity of the test.

The results indicate that nearly all items and tasks (96.5%) have infit mean square statistics below 2.0 across all grade-level clusters and domains, demonstrating that these items and tasks reliably measure ability within the targeted region of the ability distribution. As previously discussed, the outfit mean square statistic is particularly sensitive to outlier responses or scores that fall outside the targeted ability range. Eleven items exhibit outfit mean square statistics exceeding 2.0. Specifically, six items in the Writing clusters, and two items in the Reading and three items in the Listening clusters, fall into this category. Across the Listening, Reading, Speaking, and Writing domains, the proportion of items with infit/outfit statistics within the productive range of 0.5 to 1.5 is 77.5%, 77.5%, 93.7%, and 68.7%, respectively, across clusters.

For the Andrich threshold estimates, fit indices are generally around 1 for infit and outfit statistics, except for the Writing domain in the K–2 grade-level. This cluster showed outfit statistics higher than 2.0, leading to the exclusion of 5% of outliers during the calibration procedure to ensure the test’s validity.

Additionally, 6 items used as anchors showed displacement values higher than 0.5. These include four items each in the Writing domain for the grades 3–5 (3) and the grades 6–8 (1) clusters, the Reading domain for the grades K–2 cluster, and the Listening domain for the grades 6–8 cluster. These items were excluded from the anchor set and were freely estimated during the calibration procedure to maintain accuracy.

7.2.1. Listening Item Analysis

Table 7.2.1.a.

Item Summary: K–2

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Information withheld due to confidentiality requirements.

Table 7.2.1.b.

Threshold Summary K-2

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.1.c. Information withheld due to confidentiality requirements.

Complete Item Analysis: K-2

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated.

Information withheld due to confidentiality requirements.

Table 7.2.1.d.

Item Summary: 3-5

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Information withheld due to confidentiality requirements.

Table 7.2.1.e.

Threshold Summary: 3–5

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.1.f. Information withheld due to confidentiality requirements.

Complete Item Analysis: 3–5

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated.

Information withheld due to confidentiality requirements.

Table 7.2.1.g.

Item Summary: 6–8

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Information withheld due to confidentiality requirements.

Table 7.2.1.h.**Threshold Summary: 6–8**

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.1.i. Information withheld due to confidentiality requirements.**Complete Item Analysis: 6–8**

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated

Table 7.2.1.j. Information withheld due to confidentiality requirements.**Item Summary: 9–12**

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Information withheld due to confidentiality requirements.

Table 7.2.1.k.

Threshold Summary: 9–12

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.1.l. Information withheld due to confidentiality requirements.

Complete Item Analysis: 9–12

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated.

Information withheld due to confidentiality requirements.

7.2.2. Reading Item Analysis

Table 7.2.2.a.

Item Summary: K-2

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Table 7.2.2.b. Information withheld due to confidentiality requirements.

Threshold Summary: K-2

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.2.c. Information withheld due to confidentiality requirements.

Complete Item Analysis: K-2

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Information withheld due to confidentiality requirements.

Table 7.2.2.c

Item Summary: 3–5

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Table 7.2.2.d. **Information withheld due to confidentiality requirements.**

Threshold Summary: 3–5

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.2.e **Information withheld due to confidentiality requirements.**

Complete Item Analysis: 3–5

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated.

Information withheld due to confidentiality requirements.

Table 7.2.2.f.**Item Summary: 6–8**

Item Type	No. of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Table 7.2.2.g. Information withheld due to confidentiality requirements.**Threshold Summary: 6–8**

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.2.h. Information withheld due to confidentiality requirements.**Complete Item Analysis: 6–8**

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated.

Information withheld due to confidentiality requirements.

Table 7.2.2.i.

Item Summary: 9–12

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Table 7.2.2.j. **Information withheld due to confidentiality requirements.**

Threshold Summary: 9–12

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.2.k. **Information withheld due to confidentiality requirements.**

Complete Item Analysis: 9–12

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated.

Information withheld due to confidentiality requirements.

7.2.3. Speaking Item Analysis

Table 7.2.3.a.

Item Summary: K-2

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Table 7.2.3.b. Information withheld due to confidentiality requirements.

Threshold Summary: K-2

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.3.c. Information withheld due to confidentiality requirements.

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated.

Information withheld due to confidentiality requirements.

Table 7.2.3.d.

Item Summary: 3–5

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Table 7.2.3.e. **Information withheld due to confidentiality requirements.**

Threshold Summary: 3–5

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.3.f. **Information withheld due to confidentiality requirements.**

Complete Item Analysis: 3–5

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated.

Information withheld due to confidentiality requirements.

Table 7.2.3.g.

Item Summary: 6–8

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Table 7.2.3.h. **Information withheld due to confidentiality requirements.**

Threshold Summary: 6–8

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.3.i. **Information withheld due to confidentiality requirements.**

Complete Item Analysis: 6–8

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated.

Information withheld due to confidentiality requirements.

Table 7.2.3.j.

Item Summary: 9–12

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Table 7.2.3.k. Information withheld due to confidentiality requirements.

Threshold Summary: 9–12

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.3.l. Information withheld due to confidentiality requirements.

Complete Item Analysis: 9–12

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated.

Information withheld due to confidentiality requirements.

7.2.4. Writing Item Analysis

Table 7.2.4.a.

Item Summary: K-2

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Table 7.2.4.b. Information withheld due to confidentiality requirements.

Threshold Summary: K-2

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.4.c. Information withheld due to confidentiality requirements.

Complete Item Analysis: K-2

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated.

Information withheld due to confidentiality requirements.

Table 7.2.4.d.

Item Summary: 3–5

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Table 7.2.4.e. **Information withheld due to confidentiality requirements.**

Threshold Summary: 3–5

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.4.f. **Information withheld due to confidentiality requirements.**

Complete Item Analysis: 3–5

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated

Information withheld due to confidentiality requirements.

Table 7.2.4.g.

Item Summary: 6–8

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Table 7.2.4.h. Information withheld due to confidentiality requirements.

Threshold Summary: 6–8

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.4.i. Information withheld due to confidentiality requirements.

Complete Item Analysis: 6–8

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Information withheld due to confidentiality requirements.

Table 7.2.4.j.

Item Summary: 9–12

Item Type	Number of Items	Average Item Difficulty (in logits)	Average of P-value	Average Infit Mean Square	Average Outfit Mean Square

MOSR: multiple opportunities for selected response

Table 7.2.4.k. Information withheld due to confidentiality requirements.

Threshold Summary: 9–12

Raw Score	Frequency	Threshold	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics

Table 7.2.4.l. Information withheld due to confidentiality requirements.

Complete Item Analysis: 9–12

Name	Expected PL	Standard	Item Difficulty (in logits)	Anchored	P-value	Infit Mnsq Fit Statistics	Outfit Mnsq Fit Statistics	Point Measure

Note: A (Y) in the Anchored column indicates that anchor items with a Displacement greater than 0.5 were recalibrated.

Information withheld due to confidentiality requirements.

7.3. *DIF Analysis and Summary*

Differential Item Functioning (DIF) analysis aims to determine whether item or task performance is influenced by factors unrelated to English language proficiency, the construct being measured by the test. Essentially, DIF analysis seeks to identify items that may function differently for various groups due to irrelevant characteristics. For the Alternate ACCESS, student performance was compared across four groupings: (1) males versus females, (2) Hispanic versus non-Hispanic ethnic backgrounds, (3) race (Hispanic versus five racial groups) and (4) primary disabilities. Students with missing test scores, gender, or ethnicity were excluded from the analysis. For gender and Hispanic vs non-Hispanic and race DIF analysis, male and Hispanic groups are reference groups. For disability DIF analysis, multiple group performances were compared against the overall performance at once, instead of setting one group as a reference group and conducting multiple pairwise comparisons.

To ensure sufficient sample sizes within racial groups and disability categories, the analysis included the four largest racial groups: Hispanic, White, Black, and Asian. For disability categories, groups with fewer than 100 students were aggregated, while those with at least 100 students were analyzed separately. This approach ensured robust and reliable DIF detection across all examined subgroups.

A multiple-group analysis was used for DIF detection within the context of rating scale models, which Alternate ACCESS employs. This approach is an extension of the IRT model to multiple groups and is preferred due to its flexibility in assessing the invariance of item properties such as discrimination and difficulty (Tay et al., 2015). For DIF detection, rating scale models are estimated separately for each group with constraints. To identify DIF, one item difficulty of one group (the focal group) is compared to that of the reference group, while keeping all other difficulties consistent across groups. If the difference is statistically significant, that item exhibits DIF for the corresponding source.

Winsteps provides two types of DIF contrasts: (1) a paired DIF effect between two specific groups, with the hypothesis that an item has the same difficulty across the groups, and (2) a contrast between a specific group and the overall average difficulty across all groups, with the hypothesis that an item's difficulty is equal to its average difficulty across groups. For gender and ethnicity, the first type was used, with the male and Hispanic groups as the references. The five racial groups—White (W), Black (B), Asian (AS), American Indian/Alaskan Native (AI), and Pacific Islander/Hawaiian (PI)—are compared to the Hispanic group, which serves as the reference group. For types of disabilities, the second type was employed since there was no specific reference group; instead, each disability group's item difficulty was compared against the overall average difficulty for each item.

Following guidelines by ETS for NAEP assessment (Allen, Carlson, & Zalanak, 1999), Alternate ACCESS tasks are classified into three DIF levels:

- AA (no DIF), when the Rasch-Welch Chi-square statistic is not significant or when it is significant and $|DIF|$ is less than 0.43 logits
- BB (weak DIF), when the Rasch-Welch Chi-square statistic is significant and $|DIF|$ is greater than or equal to 0.43 but less than 0.64 logits
- CC (strong DIF), when the Rasch-Welch Chi-square statistic is significant and $|DIF|$ is greater than or equal to 0.64 logits

Note: ETS uses Delta units, where 1 Delta unit is equivalent to 0.426 logits.

The following tables are organized into four sections, divided by domains and clusters:

- a) Overall DIF Summary: This section provides a summary of the number of items identified with DIF across the three levels—AA, BB, or CC—for gender, ethnicity, and disabilities. For disabilities, each item may exhibit at least five DIF effects due to multiple comparisons among disability groups. This highlights the complexity of DIF analysis for this population, given the variety of group comparisons involved.
- b) DIF analysis for gender and ethnicity: This section details the DIF results for individual items. The second and fourth columns indicate the DIF level (AA, BB, or CC) for gender and ethnicity, respectively. The third and fifth columns identify items that favor one group over the other at each DIF level. Ideally, even when all items fall into the AA category, there should be a relatively even distribution of items favoring each group to ensure there is no systematic bias in the test.
- c) DIF analysis for race: This section presents a breakdown of the DIF results across five different ethnic groups. It provides detailed insights into item performance and potential bias related to individual ethnicity.
- d) DIF analysis for disability: This section focuses on DIF results for different disability categories. It provides a closer examination of item performance across various disability groups, ensuring that the test is equitable and free from bias across these sub-populations.

The DIF analysis results are presented across four domains (Listening, Reading, Speaking, and Writing) and focus on identifying potential bias across gender, ethnicity, race, and disability groups. The analysis highlights three DIF levels—A (negligible), B (moderate), and C (large). No B- or C-level DIF was found for gender or ethnicity. However, both B- and C-level DIF effects were observed for racial and disability groups across all domains. For racial groups, some items demonstrated B- and C-level DIF, particularly for participants identifying as American Indian or Pacific Islander. The American Indian group, in particular, showed the highest number of items flagged with B- and C-level DIF. For disability groups, a varying number of B- and C-level DIF items were identified across disability subgroups. Given the complexity of these findings, the remainder of this section focuses specifically on DIF related to disability. For reference, the term “Base” represents the baseline group, reflecting overall item difficulty levels. The “Other” group includes the following disability categories, each with fewer than 100 students: Visual Impairment (VI), Traumatic Brain Injury (TBI), Specific Learning Disability (SLD), Orthopedic Impairment (OI), Hearing Impairment (HI), Emotional Disability (ED), and Deaf-Blindness (DB). In grade-level cluster 3, the “Other” group also includes Developmental Delay (DD), in addition to the aforementioned categories. When DIF favors the baseline group, it indicates that the item was relatively easier for the general population compared to the specific disability subgroup.

In the Listening domain, one C-level DIF was observed in the 9-12 cluster for Item 4, favoring the baseline group (average difficulty) over the SLD group. Additionally, B-level DIFs were observed in the 9-12, 3-5, and 6-8 clusters. In the 3-5 cluster, Item 2 exhibited a B-level DIF favoring the baseline group over the SLD group. In the 6-8 cluster, Item 5 showed a B-level DIF favoring the SLD group. Similarly, in the 9-12 cluster, Item 10 showed a B-level DIF favoring the SLD group.

In the Reading domain, in the K-2 cluster, one B-level DIF was observed for the OHI group on Item 2, and one C-level DIF favored the SLI group on Item 2. No B- or C-level DIFs were observed in the 3-5 cluster. In the 6-8 cluster, Item 2 exhibited a C-level DIF favoring the SLD group, while in the 9-12 cluster, one B-level DIF was observed on Item 1, favoring the SLD group.

In the Speaking domain, no B-level or C-level DIFs were observed in the K-2 and 3-5 clusters. In the 6-8 cluster, two B-level DIFs were observed on Items 6 and 8, both favoring the SLD group. For Item 6, the SLD group was favored, while Item 8 favored the baseline group. In the 9-12 cluster, two B-level DIFs were observed on Item 1, favoring the SLD group, and on Item 2, favoring the "Other" group (a combination of smaller groups). Both items favored the SLD and Other groups over the baseline group.

In the Writing domain, DIF effects were observed across all clusters, particularly in the K-2 and 3-5 clusters. In the K-2 cluster, there were three B-level DIFs and one C-level DIF. The C-level DIF occurred on Item 2 for the SLI group, favoring the SLI group. B-level DIFs were observed on Item 1 for the Other group, Item 3 for the OHI and SLI groups, with the baseline group favored for the OHI group, and SLI and Other groups favored for the SLI group. In the 3-5 cluster, one C-level DIF was observed on Item 1 for the SLD group, favoring the SLD group, while a B-level DIF on Item 2 also favored the SLD group. In the 6-8 cluster, one B-level and one C-level DIF were observed. The B-level DIF was found on Item 1, favoring the SLD group, and the C-level DIF was found on Item 5, favoring the baseline group. In the 9-12 cluster, two C-level DIFs were observed on Items 1 and 2, both favoring the SLD group.

Overall, while the DIF analysis revealed no systematic bias for gender and ethnicity, it highlighted areas of potential concern for disability groups, particularly in the higher DIF levels (B and C), across all domains and grade-level clusters. These findings underscore the need for careful interpretation of DIF effects, especially for disability comparisons.

7.3.1. Listening DIF Analysis

Table 7.3.1.a.

Overall DIF Summary: List K-2

DIF Level	Gender	Ethnicity	Race	Disability

Information withheld due to confidentiality requirements.

Table 7.3.1.b.

DIF Summary for Gender and Ethnicity: K-2

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Table 7.3.1.c. Information withheld due to confidentiality requirements.

DIF Analysis for Race: K-2

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.1.d.

DIF Analysis for Disability: K-2

Item	AS DIF Level	AS Favored Group	DD DIF Level	DD Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLI DIF Level	SLI Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other"; "NPD" is excluded from the table.

Information withheld due to confidentiality requirements.

Table 7.3.1.e.

Overall DIF Summary: 3–5

DIF Level	Gender	Ethnicity	Race	Disability

Table 7.3.1.f. Information withheld due to confidentiality requirements.

DIF Summary for Gender and Ethnicity: 3–5

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Table 7.3.1.g. Information withheld due to confidentiality requirements.

DIF Analysis for Race: 3–5

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.1.h.

DIF Analysis for Disability: 3–5

Item	AS DIF Level	AS Favored Group	DD DIF Level	DD Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLD DIF Level	SLD Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other"; "NPD" is excluded from the table

Information withheld due to confidentiality requirements.

Table 7.3.1.i.

Overall DIF Summary: 6–8

DIF Level	Gender	Ethnicity	Race	Disability

Table 7.3.1.j. Information withheld due to confidentiality requirements.

DIF Summary for Gender and Ethnicity: 6–8

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.1.k.

DIF Analysis for Race: 6–8

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Table 7.3.1.l. Information withheld due to confidentiality requirements.

DIF Analysis for Disability: 6–8

Item	AS DIF Level	AS Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLD DIF Level	SLD Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into “Other”; “NPD” is excluded from the table

Information withheld due to confidentiality requirements.

Table 7.3.1.m.

Overall DIF Summary: 9–12

DIF Level	Gender	Ethnicity	Race	Disability

Information withheld due to confidentiality requirements.

Table 7.3.1.n.

DIF Summary for Gender and Ethnicity: 9–12

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Table 7.3.1.o. Information withheld due to confidentiality requirements.

DIF Analysis for Race: 9–12

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.1.p.

DIF Analysis for Disability: 9–12

Item	AS DIF Level	AS Favored Group	DD DIF Level	DD Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLD DIF Level	SLD Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other"; "NPD" is excluded from the table

Information withheld due to confidentiality requirements.

7.3.2. Reading DIF Analysis

Table 7.3.2.a.

Overall DIF Summary: K-2

DIF Level	Gender	Ethnicity	Race	Disability

Table 7.3.2.b. Information withheld due to confidentiality requirements.

DIF Summary for Gender and Ethnicity: K-2

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Table 7.3.2.c. Information withheld due to confidentiality requirements.

DIF Analysis for Race: K-2

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.2.d.

DIF Analysis for Disability: K-2

Item	AS DIF Level	AS Favored Group	DD DIF Level	DD Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLI DIF Level	SLI Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other"; "NPD" is excluded from the table

Information withheld due to confidentiality requirements.

Table 7.3.2.e.

Overall DIF Summary: 3-5

DIF Level	Gender	Ethnicity	Race	Disability

Table 7.3.2.f. Information withheld due to confidentiality requirements.

DIF Summary for Gender and Ethnicity: 3-5

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.2.g.

DIF Analysis for Race: 3–5

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Table 7.3.2.h. Information withheld due to confidentiality requirements.

DIF Analysis for Disability: 3–5

Item	AS DIF Level	AS Favored Group	DD DIF Level	DD Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLD DIF Level	SLD Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other"; "NPD" is excluded from the table

Information withheld due to confidentiality requirements.

Table 7.3.2.i.

Overall DIF Summary: 6–8

DIF Level	Gender	Ethnicity	Race	Disability

Table 7.3.2.j. Information withheld due to confidentiality requirements.

DIF Summary for Gender and Ethnicity: 6–8

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Table 7.3.2.k. Information withheld due to confidentiality requirements.

DIF Analysis for Race: 6–8

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.2.I.

DIF Analysis for Disability: 6–8

Item	AS DIF Level	AS Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLD DIF Level	SLD Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other"; "NPD" is excluded from the table

Information withheld due to confidentiality requirements.

Table 7.3.2.m.

Overall DIF Summary: 9–12

DIF Level	Gender	Ethnicity	Race	Disability

Table 7.3.2.n. Information withheld due to confidentiality requirements.

DIF Summary for Gender and Ethnicity: 9–12

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Table 7.3.2.o. Information withheld due to confidentiality requirements.

DIF Analysis for Race: 9–12

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.2.p.

DIF Analysis for Disability: 9–12

Item	AS DIF Level	AS Favored Group	DD DIF Level	DD Favored Group	ID DIF Level	ID Favore d Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLD DIF Level	SLD Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other"; "NPD" is excluded from the table

Information withheld due to confidentiality requirements.

7.3.3. Speaking DIF Analysis

Table 7.3.3.a.

Overall DIF Summary: K–2

DIF Level	Gender	Ethnicity	Race	Disability

Table 7.3.3.b. Information withheld due to confidentiality requirements.

DIF Summary for Gender and Ethnicity: K–2

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.3.c.

DIF Analysis for Race: K-2

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Table 7.3.3.d. Information withheld due to confidentiality requirements.

DIF Analysis for Disability: K-2

Item	AS DIF Level	AS Favored Group	DD DIF Level	DD Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLI DIF Level	SLI Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other"; "NPD" is excluded from the table

Information withheld due to confidentiality requirements.

Table 7.3.3.e.

Overall DIF Summary: 3–5

DIF Level	Gender	Ethnicity	Race	Disability

Table 7.3.3.f. Information withheld due to confidentiality requirements.

DIF Summary for Gender and Ethnicity: 3–5

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Table 7.3.3.g. Information withheld due to confidentiality requirements.

DIF Analysis for Race: 3–5

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.3.h.

DIF Analysis for Disability: 3–5

Item	AS DIF Level	AS Favored Group	DD DIF Level	DD Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLD DIF Level	SLD Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other"; "NPD" is excluded from the table

Information withheld due to confidentiality requirements.

Table 7.3.3.i.

Overall DIF Summary: 6–8

DIF Level	Gender	Ethnicity	Race	Disability

Table 7.3.3.j. **Information withheld due to confidentiality requirements.**

DIF Summary for Gender and Ethnicity: 6–8

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.3.k.

DIF Analysis for Race: 6–8

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Table 7.3.3.l. Information withheld due to confidentiality requirements.

DIF Analysis for Disability: 6–8

Item	AS DIF Level	AS Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLD DIF Level	SLD Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other"; "NPD" is excluded from the table

Information withheld due to confidentiality requirements.

Table 7.3.3.m.

Overall DIF Summary: 9–12

DIF Level	Gender	Ethnicity	Race	Disability

Information withheld due to confidentiality requirements.

Table 7.3.3.n.

DIF Summary for Gender and Ethnicity: 9–12

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Table 7.3.3.o. Information withheld due to confidentiality requirements.

DIF Analysis for Race: 9–12

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Table 7.3.3.p. Information withheld due to confidentiality requirements.

DIF Analysis for Disability: 9–12

Item	AS DIF Level	AS Favored Group	DD DIF Level	DD Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLD DIF Level	SLD Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other" "NPD" is excluded from the table

Information withheld due to confidentiality requirements.

7.3.4. Writing DIF Analysis

Table 7.3.4.a.

Overall DIF Summary: K-2

DIF Level	Gender	Ethnicity	Race	Disability

Table 7.3.4.b. Information withheld due to confidentiality requirements.

DIF Summary for Gender and Ethnicity: K-2

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Table 7.3.4.c. Information withheld due to confidentiality requirements.

DIF Analysis for Race: K-2

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.4.d.

DIF Analysis for Disability: K-2

Item	AS DIF Level	AS Favored Group	DD DIF Level	DD Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLI DIF Level	SLI Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other"; "NPD" is excluded from the table.

Information withheld due to confidentiality requirements.

Table 7.3.4.e.

Overall DIF Summary: 3-5

DIF Level	Gender	Ethnicity	Race	Disability

Table 7.3.4.f. **Information withheld due to confidentiality requirements.**

DIF Summary for Gender and Ethnicity: 3-5

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.4.g.

DIF Analysis for Race: 3–5

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Table 7.3.4.h. Information withheld due to confidentiality requirements.

DIF Analysis for Disability: 3–5

Item	AS DIF Level	AS Favored Group	DD DIF Level	DD Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLD DIF Level	SLD Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other"; "NPD" is excluded from the table

Table 7.3.4.i. Information withheld due to confidentiality requirements.

Overall DIF Summary: 6–8

DIF Level	Gender	Ethnicity	Race	Disability

Information withheld due to confidentiality requirements.

Table 7.3.4.j.

DIF Summary for Gender and Ethnicity: 6–8

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Table 7.3.4.k. Information withheld due to confidentiality requirements.

DIF Analysis for Race: 6–8

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Table 7.3.4.l. Information withheld due to confidentiality requirements.

DIF Analysis for Disability: 6–8

Item	AS DIF Level	AS Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLD DIF Level	SLD Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into "Other"; "NPD" is excluded from the table

Information withheld due to confidentiality requirements.

Table 7.3.4.m.

Overall DIF Summary: 9–12

DIF Level	Gender	Ethnicity	Race	Disability

Table 7.3.4.n. Information withheld due to confidentiality requirements.

DIF Summary for Gender and Ethnicity: 9–12

Name	M/F DIF Level	M/F Favored Group	H/O DIF Level	H/O Favored Group

Table 7.3.4.o. Information withheld due to confidentiality requirements.

DIF Analysis for Race: 9–12

Item	W/H DIF Level	W/H Favored Group	B/H DIF Level	B/H Favored Group	A/H DIF Level	A/H Favored Group	AI/H DIF Level	AI/H Favored Group	PS/H DIF Level	PS/H Favored Group

Information withheld due to confidentiality requirements.

Table 7.3.4.p.

DIF Analysis for Disability: 9–12

Item	AS DIF Level	AS Favored Group	DD DIF Level	DD Favored Group	ID DIF Level	ID Favored Group	MD DIF Level	MD Favored Group	OHI DIF Level	OHI Favored Group	SLD DIF Level	SLD Favored Group	Other DIF Level	Other Favored Group

Note: Groups with sample sizes less than 100 are combined into “Other”; “NPD” is excluded from the table

Information withheld due to confidentiality requirements.

7.4. Raw Score Distribution

Tables 7.4.1.a through 7.4.4.d. summarize the raw score performance for each test form per domain and grade-level cluster, including the number of students, the range of scores (minimum and maximum), the mean, and the standard deviation (SD). Figures 7.4.1.a through 7.4.4.d. illustrate the distribution of raw scores graphically.

Overall, the tables highlight several important trends:

Listening: Mean raw scores consistently increase with grade level, reflecting developmental progress. The range of scores remains consistent across grades (0–40), with a gradual decrease in standard deviation (SD), indicating reduced variability in higher grades. When examining the figures, the raw score distribution is severely right-skewed in grades K–2. As grade-level clusters progress, the distribution becomes more left-skewed, culminating in a pronounced left-skew in grades 9–12.

Reading: Similar to Listening, mean scores increase across grade-level clusters, aligning with expected growth in reading proficiency. Variability, as indicated by standard deviations, is slightly higher than in Listening but follows a similar pattern of gradual decrease across grade levels. The figures show that the raw score distribution is also severely right-skewed in grades K–2. As grade-level clusters increase, the distribution shifts toward left-skewness, although lower-end scores still occupy a notable portion of the distribution.

Speaking: Speaking scores demonstrate a clear increase in means with grade progression. The maximum score is capped at 32 for all grades, and standard deviations are generally higher compared to Listening and Reading, suggesting greater variability in spoken language performance. From the figures, the raw score distribution appears severely right-skewed across all grade-level clusters. As grade-level clusters advance, both the upper and lower ends of the distribution continue to occupy significant proportions, highlighting greater heterogeneity in speaking performance.

Writing: Writing scores steadily increase with grade level, with the maximum score fixed at 32 across all grades. Variability decreases slightly across grade levels, mirroring trends in the other domains. The figures reveal that the raw score distribution is initially severely right-skewed. This skewness becomes less pronounced across grade-level clusters but remains right-skewed overall.

The right-skewed patterns in raw scores are likely due to the increased difficulty of the new test form compared to 601.

7.4.1. Listening

Table 7.4.1.a.

Raw Score Descriptive Statistics: K–2

Grade	Number of Students	Min.	Max.	Mean	SD
K	2,657	0	40	14.25	13.31
1	3,774	0	40	19.32	14.14
2	3,427	0	40	22.88	13.96
Total	9,858	0	40	19.19	14.26

Table 7.4.1.b.

Raw Score Descriptive Statistics: 3–5

Grade	Number of Students	Min.	Max.	Mean	SD
3	3,002	0	40	23.20	13.43
4	2,963	0	40	25.15	13.20
5	2,642	0	40	26.41	12.91
Total	8,607	0	40	24.86	13.26

Table 7.4.1.c.

Raw Score Descriptive Statistics: 6–8

Grade	Number of Students	Min.	Max.	Mean	SD
6	2,398	0	40	27.50	13.11
7	2,122	0	40	28.70	12.85
8	2,001	0	40	28.96	12.74
Total	6,521	0	40	28.34	12.92

Table 7.4.1.d.

Raw Score Descriptive Statistics: 9–12

Grade	Number of Students	Min.	Max.	Mean	SD
9	1,911	0	40	27.90	12.79
10	1,766	0	40	29.05	12.32
11	1,626	0	40	29.07	12.14
12	2,560	0	40	28.78	12.29
Total	7,863	0	40	28.69	12.39

Figure 7.4.1.a.

Raw Score Distribution: K-2

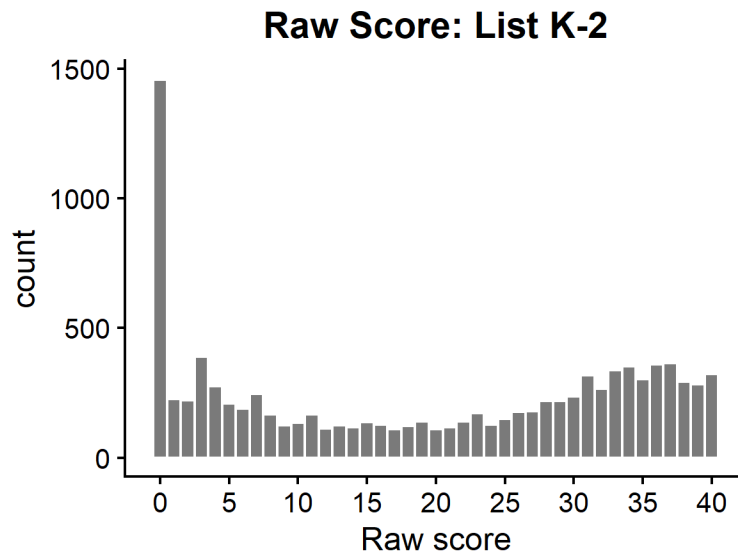


Figure 7.4.1.b.

Raw Score Distribution: 3-5

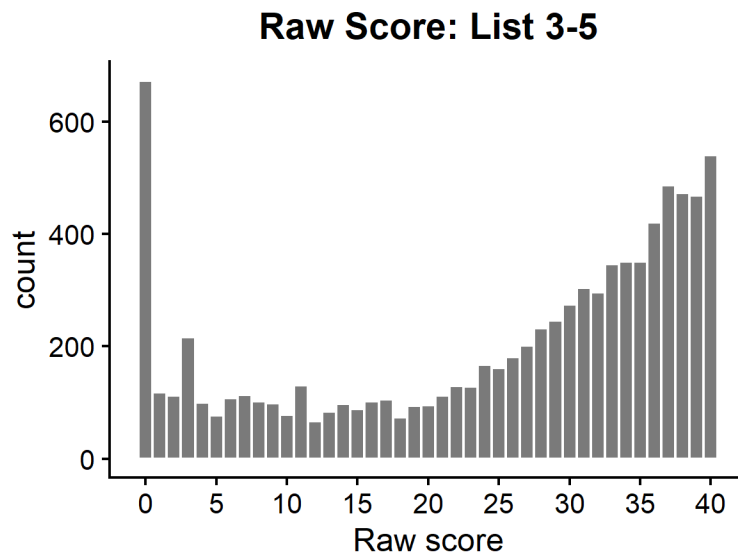


Figure 7.4.1.c.

Raw Score Distribution: 6–8

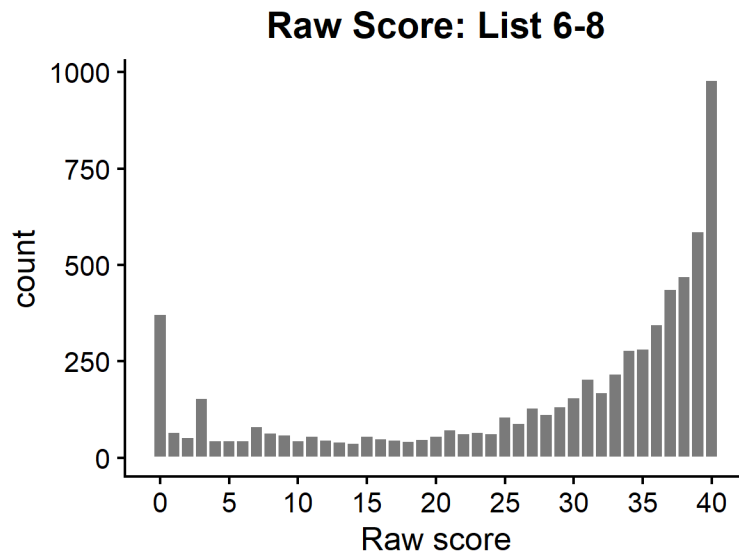
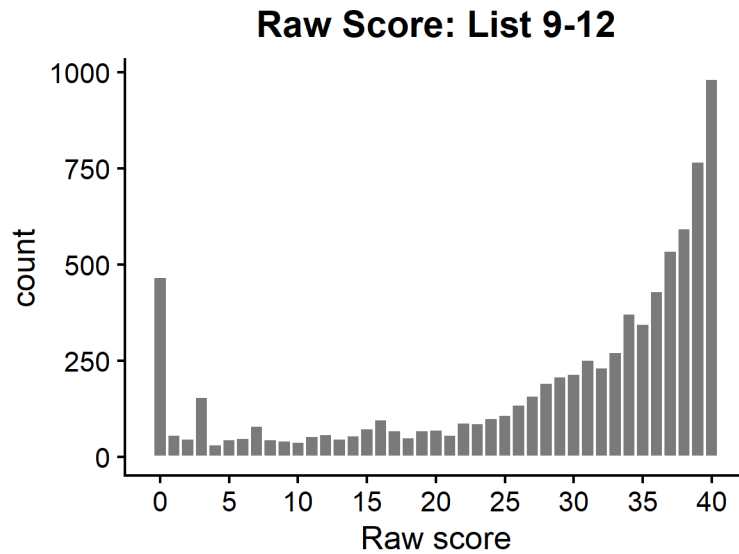


Figure 7.4.1.d.

Raw Score Distribution: 9–12



7.4.2. Reading

Table 7.4.2.a.

Raw Score Descriptive Statistics: K–2

Grade	Number of Students	Min.	Max.	Mean	SD
K	2,656	0	40	13.08	13.70
1	3,772	0	40	18.63	14.70
2	3,427	0	40	22.36	14.73
Total	9,855	0	40	18.43	14.89

Table 7.4.2.b.

Raw Score Descriptive Statistics: 3–5

Grade	Number of Students	Min.	Max.	Mean	SD
3	2,999	0	40	19.85	13.44
4	2,963	0	40	22.00	13.61
5	2,639	0	40	23.36	13.54
Total	8,601	0	40	21.67	13.60

Table 7.4.2.c.

Raw Score Descriptive Statistics: 6–8

Grade	Number of Students	Min.	Max.	Mean	SD
6	2,398	0	40	24.12	13.28
7	2,122	0	40	25.36	13.04
8	2,001	0	40	25.78	13.28
Total	6,521	0	40	25.03	13.22

Table 7.4.2.d.

Raw Score Descriptive Statistics: 9–12

Grade	Number of Students	Min.	Max.	Mean	SD
9	1,910	0	40	24.38	13.57
10	1,766	0	40	25.42	13.48
11	1,626	0	40	25.27	13.67
12	2,558	0	40	24.80	13.54
Total	7,860	0	40	24.94	13.57

Figure 7.4.2.a.

Raw Score Distribution: K-2

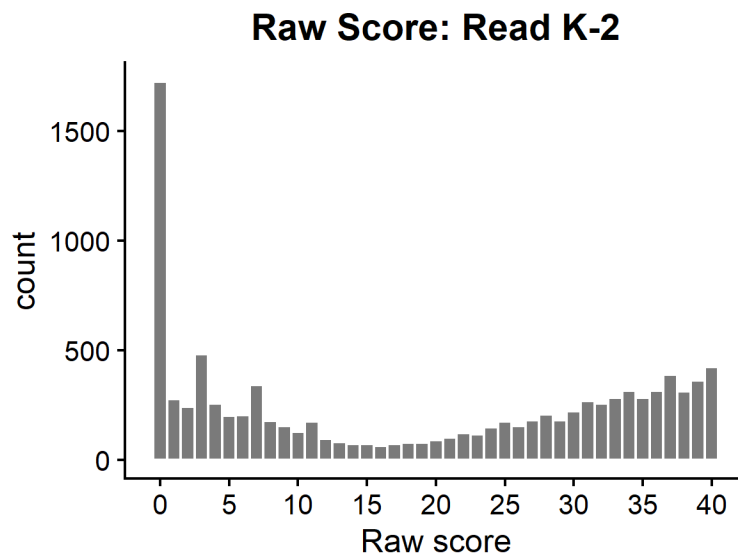


Figure 7.4.2.b.

Raw Score Distribution: 3-5

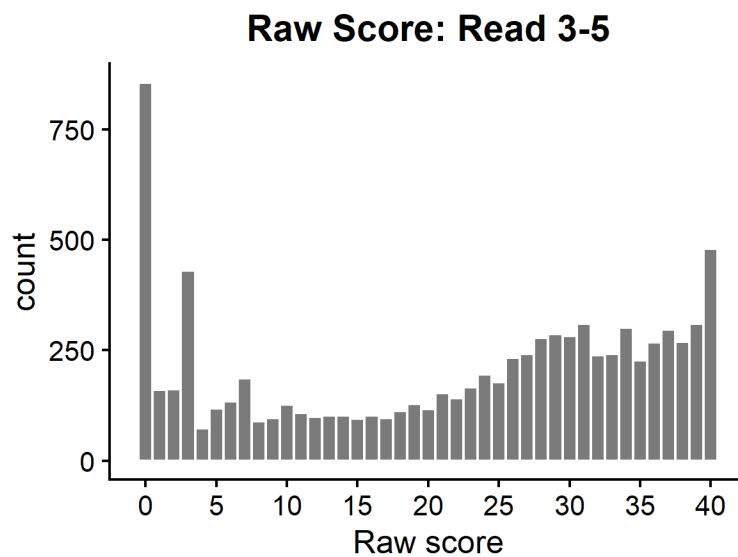


Figure 7.4.2.c.

Raw Score Distribution: 6–8

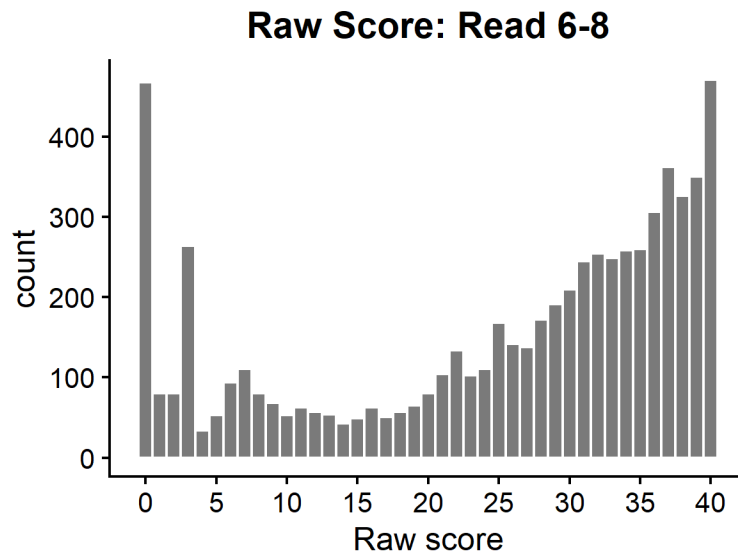
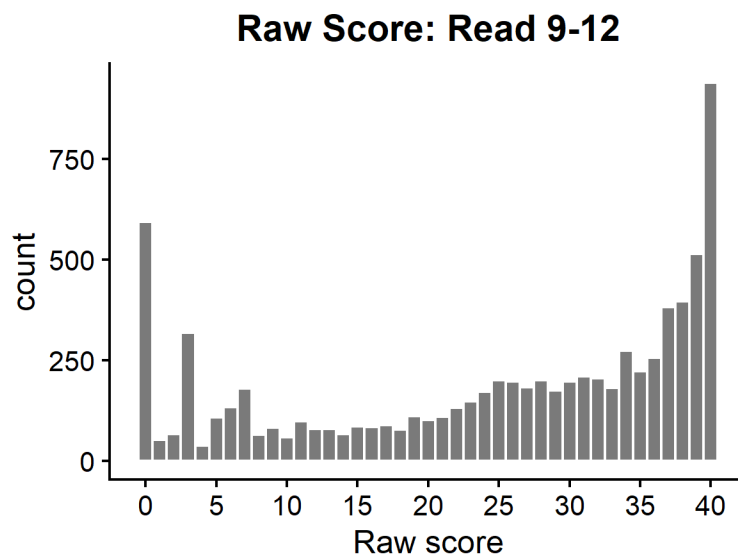


Figure 7.4.2.d.

Raw Score Distribution: 9–12



7.4.3. Speaking

Table 7.4.3.a.

Raw Score Descriptive Statistics: K–2

Grade	Number of Students	Min.	Max.	Mean	SD
K	2,651	0	32	5.59	7.55
1	3,767	0	32	8.53	9.55
2	3,424	0	32	10.90	10.48
Total	9,842	0	32	8.57	9.63

Table 7.4.3.b.

Raw Score Descriptive Statistics: 3–5

Grade	Number of Students	Min.	Max.	Mean	SD
3	3,001	0	32	12.49	10.68
4	2,959	0	32	13.77	11.12
5	2,640	0	32	14.65	11.38
Total	8,600	0	32	13.59	11.08

Table 7.4.3.c.

Raw Score Descriptive Statistics: 6–8

Grade	Number of Students	Min.	Max.	Mean	SD
6	2,394	0	32	15.19	11.21
7	2,119	0	32	15.75	10.95
8	1,998	0	32	16.02	11.40
Total	6,511	0	32	15.63	11.19

Table 7.4.3.d.

Raw Score Descriptive Statistics: 9–12

Grade	Number of Students	Min.	Max.	Mean	SD
9	1,908	0	32	14.92	12.26
10	1,764	0	32	16.30	12.28
11	1,626	0	32	15.98	12.28
12	2,555	0	32	15.58	12.02
Total	7,853	0	32	15.66	12.20

Figure 7.4.3.a.

Raw Score Distribution: K-2

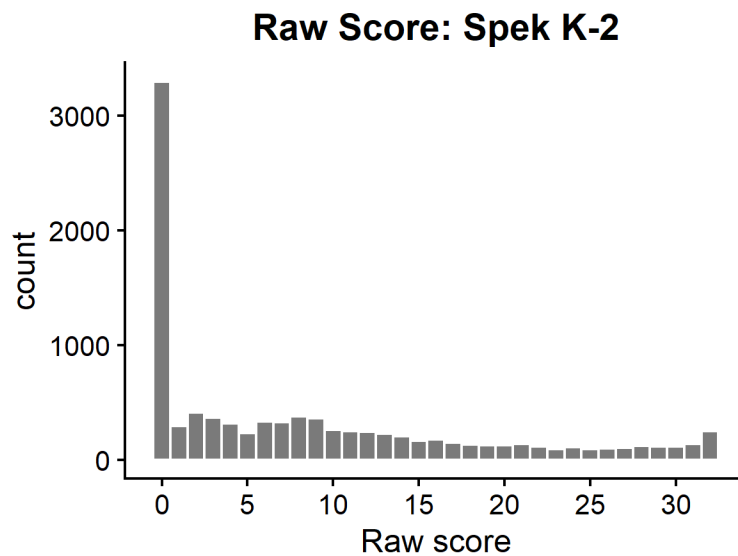


Figure 7.4.3.b.

Raw Score Distribution: 3-5

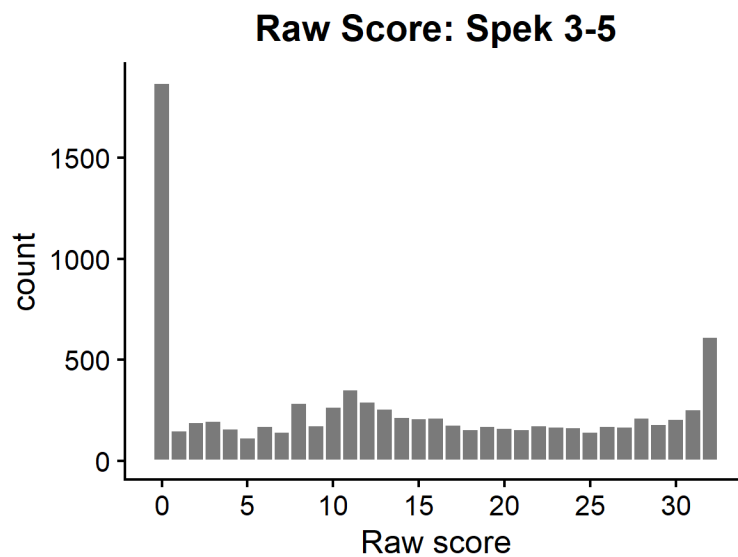


Figure 7.4.3.c.

Raw Score Distribution: 6–8

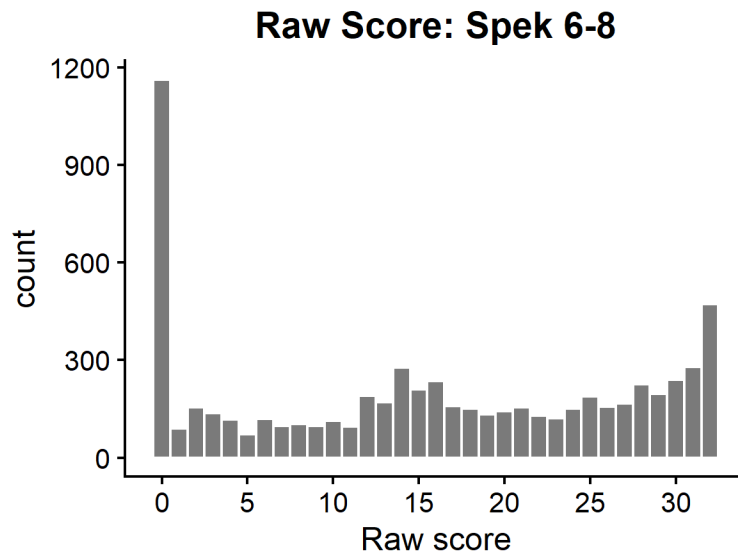
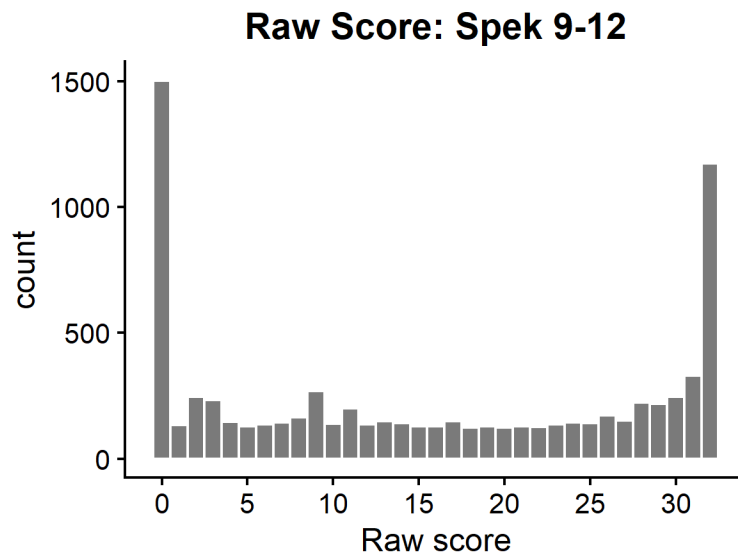


Figure 7.4.3.d.

Raw Score Distribution: 9–12



7.4.4. Writing

Table 7.4.4.a.

Raw Score Descriptive Statistics: K–2

Grade	Number of Students	Min.	Max.	Mean	SD
K	2,654	0	32	4.72	7.11
1	3,770	0	32	8.29	9.77
2	3,426	0	32	11.08	10.99
Total	9,850	0	32	8.30	9.91

Table 7.4.4.b.

Raw Score Descriptive Statistics: 3–5

Grade	Number of Students	Min.	Max.	Mean	SD
3	2,995	0	32	11.41	9.68
4	2,959	0	32	12.64	10.14
5	2,640	0	32	13.82	10.59
Total	8,594	0	32	12.57	10.17

Table 7.4.4.c.

Raw Score Descriptive Statistics: 6–8

Grade	Number of Students	Min.	Max.	Mean	SD
6	2,394	0	32	12.99	10.21
7	2,120	0	32	13.77	10.20
8	1,999	0	32	14.32	10.63
Total	6,513	0	32	13.65	10.35

Table 7.4.4.d.

Raw Score Descriptive Statistics: 9–12

Grade	Number of Students	Min.	Max.	Mean	SD
9	1908	0	32	13.12	10.62
10	1766	0	32	13.99	10.80
11	1626	0	32	14.15	11.05
12	2554	0	32	14.25	10.94
Total	7854	0	32	13.90	10.86

Figure 7.4.4.a.

Raw Score Distribution: K-2

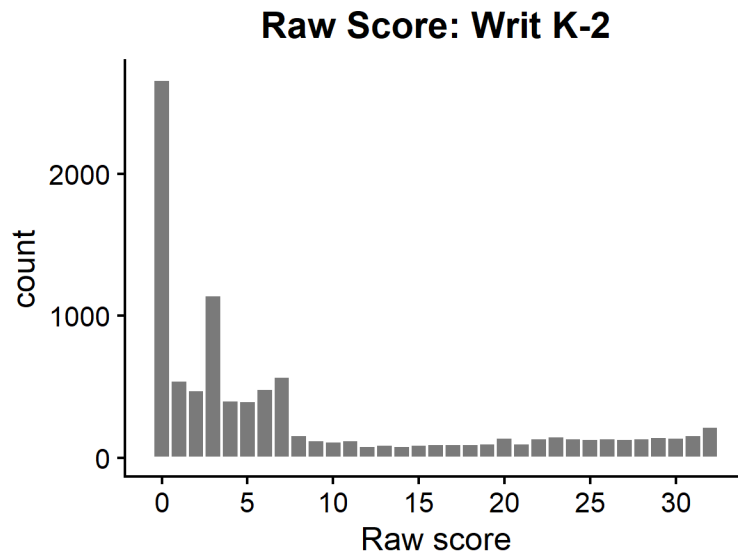


Figure 7.4.4.b.

Raw Score Distribution: 3-5

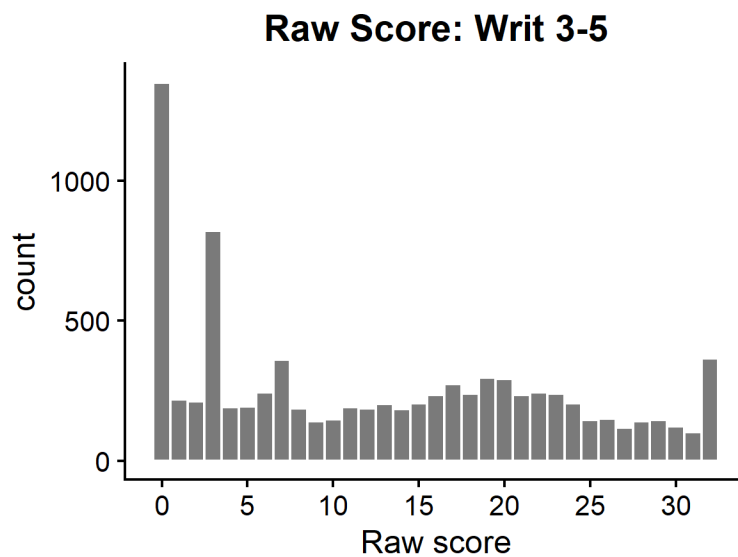


Figure 7.4.4.c.

Raw Score Distribution: 6–8

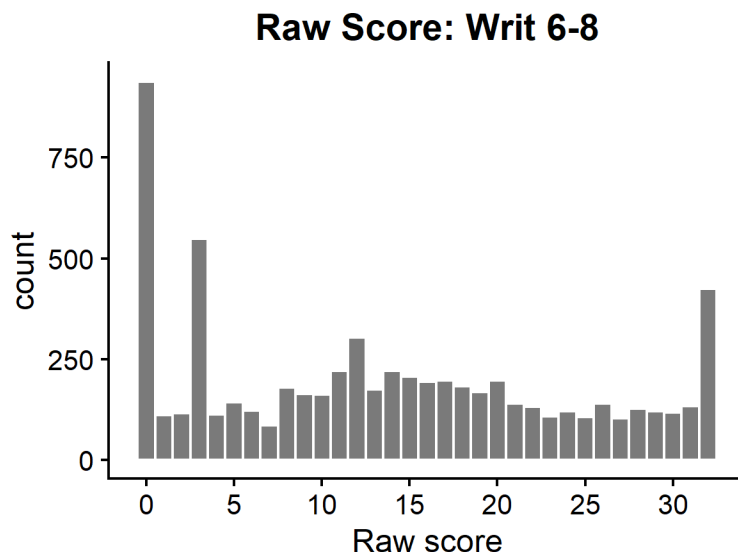
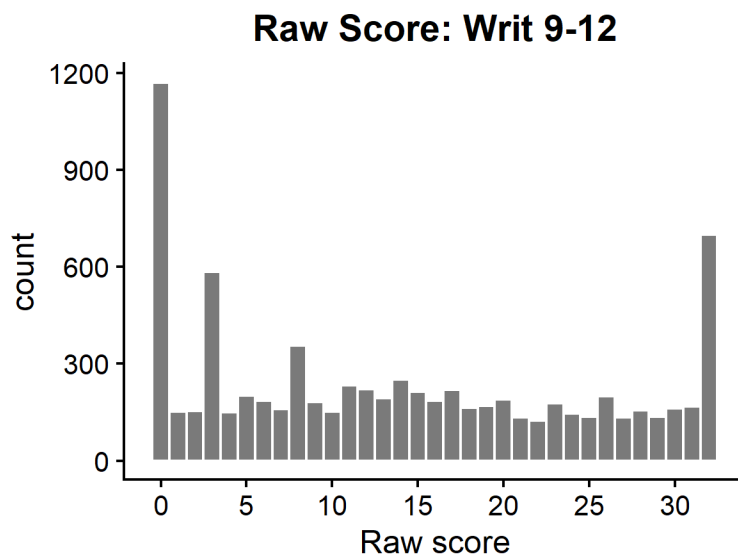


Figure 7.4.4.d

Raw Score Distribution: 9–12



7.5. Scale Score Distribution

Tables 7.5.1.a. through 7.5.4.d. summarize the scale score performance for each test form per grade-level cluster and domain, including the number of students, the range of scores (minimum and maximum), the mean, and the standard deviation (SD). Figures 7.5.1.a. through 7.5.4.d. provide graphical representation of the scale score distributions.

Listening: Mean scale scores increase consistently across grade-level clusters, from 936.21 (K–2) to 952.81 (9–12), reflecting developmental growth in listening proficiency. The distribution shows right-skewness in K–2 and left-skewness in 9–12, with scores clustering at the lower and upper end (e.g., 900 and 980). Compared to Series 601, which exhibited negatively skewed distributions in K–2 and 3–5, Series 602 appears more normally distributed, except for a high proportion of lowest scores in K–2.

Reading: Mean scores rise steadily from 933.80 (K–2) to 949.47 (9–12), with incremental growth across grades. The distribution transitions from heavily right-skewed in K–2 to a more symmetrical shape by 6–8, with slight right-skewness persisting in 9–12. Compared to Series 601, which consistently displayed negatively skewed distributions across grade-level clusters, Series 602 follows a more normal distribution, except in K–2, where the scores appear more flattened.

Speaking: Mean scores grow from 926.49 (K–2) to 944.35 (9–12), showing consistent progress across grades. Severely right-skewness dominates in K–2, but distributions gradually look normal by 6–8, while scores near the lower end cluster at the lower and upper end (e.g., 900 and 980). Compared to Series 601, which displayed a U-shaped distribution in K–2 and increasing negative skewness across grade-level clusters, Series 602 appears more normally distributed. However, 10–30% of scores are still concentrated at either the lowest or highest points.

Writing: Mean scores increase from 926.20 (K–2) to 943.31 (9–12), with steady growth across grades. Distributions remain right-skewed throughout but become more symmetrical in 6–8 and 9–12. Compared to Series 601, the primary difference is that Series 602 has a higher proportion of scores at the lowest level.

7.5.1. Listening

Table 7.5.1.a.

Scale Score Descriptive Statistics: K–2

Grade	Number of Students	Min.	Max.	Mean	SD
K	2,657	900	974	929.53	19.74
1	3,774	900	974	936.31	19.92
2	3,427	900	974	941.28	19.00
Total	9,858	900	974	936.21	20.08

Table 7.5.1.b.

Scale Score Descriptive Statistics: 3–5

Grade	Number of Students	Min.	Max.	Mean	SD
3	3,002	902	976	942.84	17.61
4	2,963	902	976	945.26	17.40
5	2,642	902	976	947.00	17.19
Total	8,607	902	976	944.95	17.49

Table 7.5.1.c.

Scale Score Descriptive Statistics: 6–8

Grade	Number of Students	Min.	Max.	Mean	SD
6	2,398	904	978	950.42	17.86
7	2,122	904	978	952.17	17.76
8	2,001	904	978	952.61	17.77
Total	6,521	904	978	951.66	17.82

Table 7.5.1.d.

Scale Score Descriptive Statistics: 9–12

Grade	Number of Students	Min.	Max.	Mean	SD
9	1,911	906	980	951.64	17.92
10	1,766	906	980	953.53	17.52
11	1,626	906	980	953.31	17.12
12	2,560	906	980	952.86	17.37
Total	7,863	906	980	952.81	17.50

Figure 7.5.1.a.

Scale Score Distribution: K-2

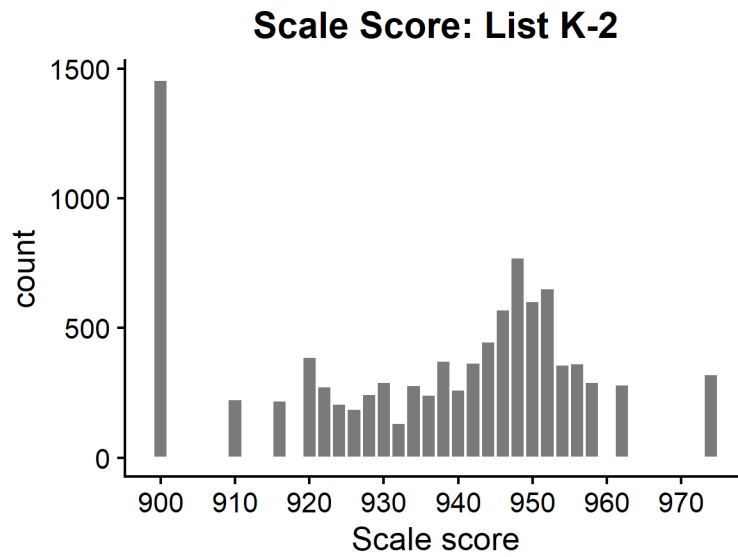


Figure 7.5.1.b.

Scale Score Distribution: 3-5

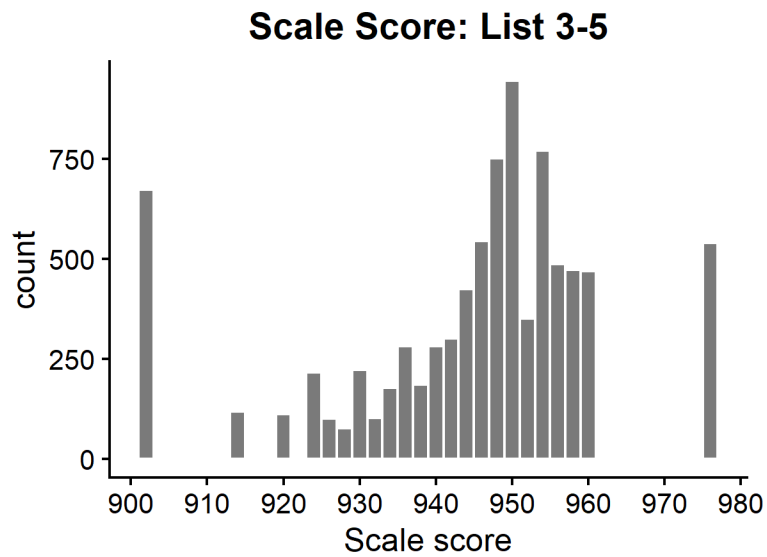


Figure 7.5.1.c.

Scale Score Distribution: 6–8

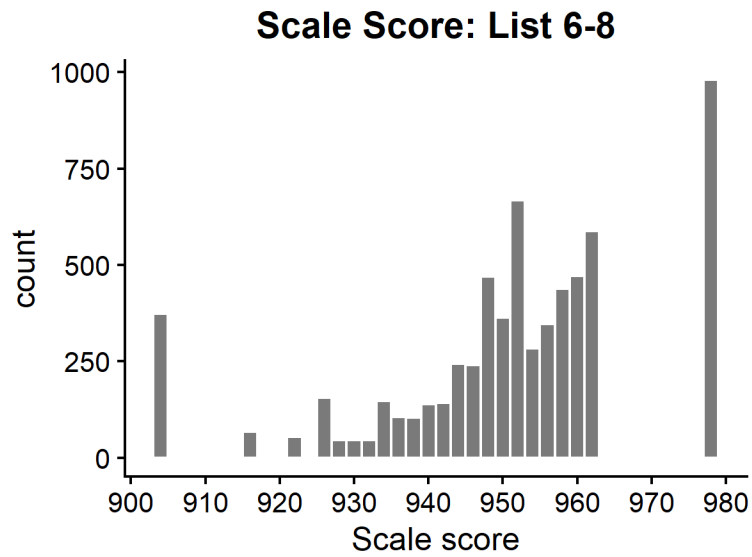
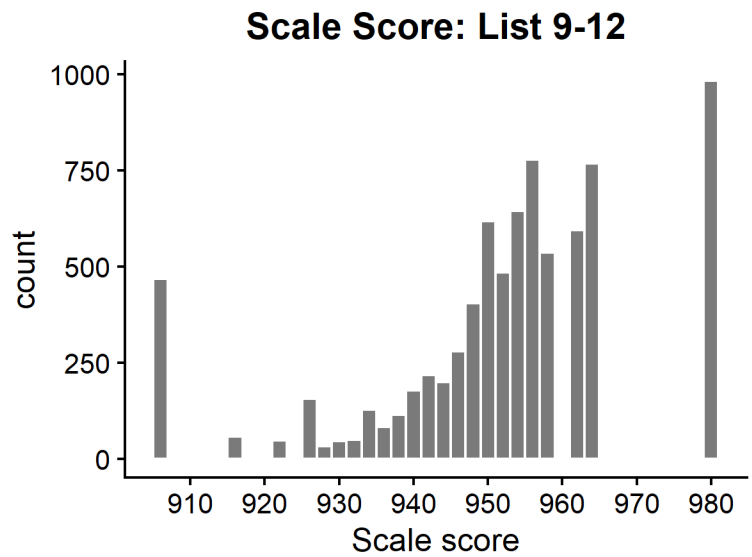


Figure 7.5.1.d.

Scale Score Distribution: 9–12



7.5.2. Reading

Table 7.5.2.a.

Scale Score Descriptive Statistics: K–2

Grade	Number of Students	Min.	Max.	Mean	SD
K	2,656	900	974	925.96	20.70
1	3,772	900	974	934.13	21.29
2	3,427	900	974	939.52	21.32
Total	9,855	900	974	933.80	21.79

Table 7.5.2.b.

Scale Score Descriptive Statistics: 3–5

Grade	Number of Students	Min.	Max.	Mean	SD
3	2,999	902	976	939.10	18.52
4	2,963	902	976	941.96	19.02
5	2,639	902	976	944.06	18.85
Total	8,601	902	976	941.61	18.90

Table 7.5.2.c.

Scale Score Descriptive Statistics: 6–8

Grade	Number of Students	Min.	Max.	Mean	SD
6	2,398	904	978	946.09	18.10
7	2,122	904	978	947.67	18.21
8	2,001	904	978	948.76	18.95
Total	6,521	904	978	947.42	18.43

Table 7.5.2.d.

Scale Score Descriptive Statistics: 9–12

Grade	Number of Students	Min.	Max.	Mean	SD
9	1,910	906	980	948.28	19.53
10	1,766	906	980	950.28	19.42
11	1,626	906	980	949.75	19.64
12	2,558	906	980	949.61	19.62
Total	7,860	906	980	949.47	19.57

Figure 7.5.2.a.

Scale Score Distribution: K-2

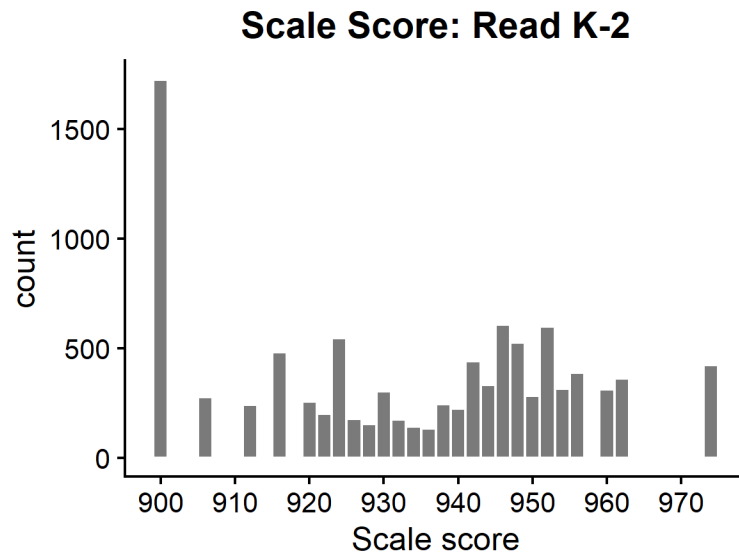


Figure 7.5.2.b.

Scale Score Distribution: 3-5

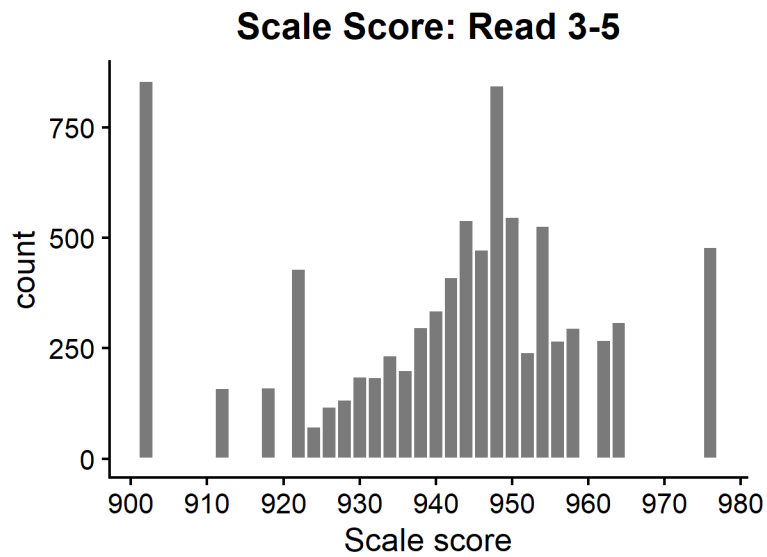


Figure 7.5.2.c.

Scale Score Distribution: 6–8

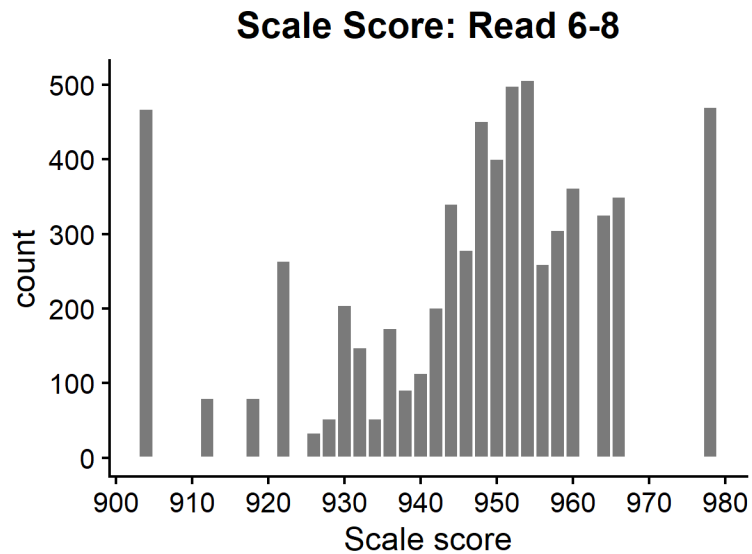
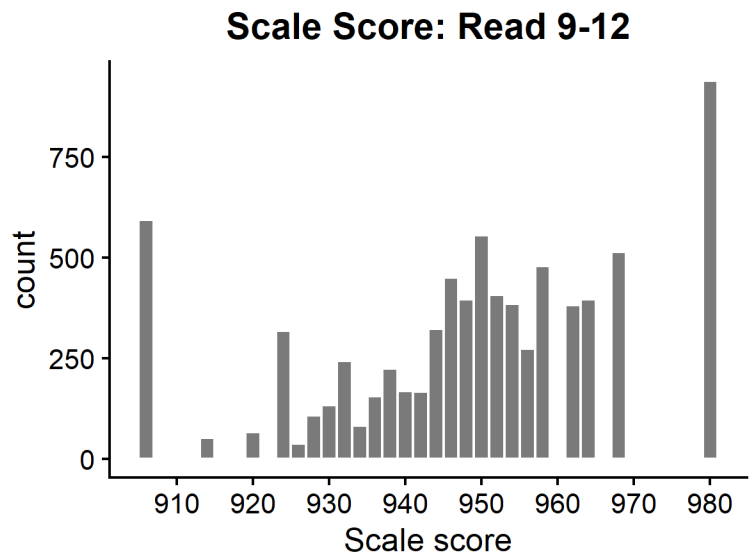


Figure 7.5.2.d.

Scale Score Distribution: 9–12



7.5.3. Speaking

Table 7.5.3.a.

Scale Score Descriptive Statistics: K–2

Grade	Number of Students	Min.	Max.	Mean	SD
K	2,651	900	974	920.17	19.54
1	3,767	900	974	926.55	21.26
2	3,424	900	974	931.33	21.67
Total	9,842	900	974	926.49	21.40

Table 7.5.3.b.

Scale Score Descriptive Statistics: 3–5

Grade	Number of Students	Min.	Max.	Mean	SD
3	3,001	902	976	935.99	21.85
4	2,959	902	976	938.25	22.34
5	2,640	902	976	939.63	22.60
Total	8,600	902	976	937.89	22.30

Table 7.5.3.c.

Scale Score Descriptive Statistics: 6–8

Grade	Number of Students	Min.	Max.	Mean	SD
6	2,394	904	978	941.72	21.58
7	2,119	904	978	942.68	20.97
8	1,998	904	978	942.98	21.98
Total	6,511	904	978	942.42	21.51

Table 7.5.3.d.

Scale Score Descriptive Statistics: 9–12

Grade	Number of Students	Min.	Max.	Mean	SD
9	1,908	906	980	942.64	23.96
10	1,764	906	980	945.31	23.89
11	1,626	906	980	944.50	23.82
12	2,555	906	980	944.14	23.16
Total	7,853	906	980	944.11	23.68

Figure 7.5.3.a.

Scale Score Distribution: K-2

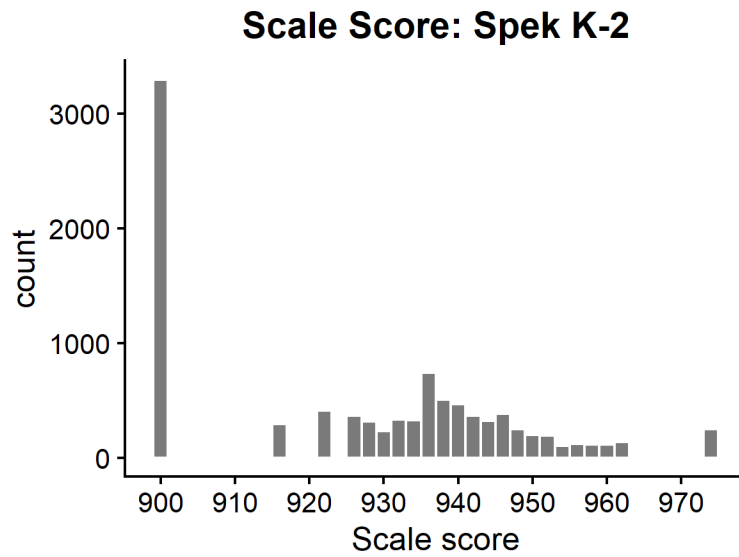


Figure 7.5.3.b.

Scale Score Distribution: 3-5

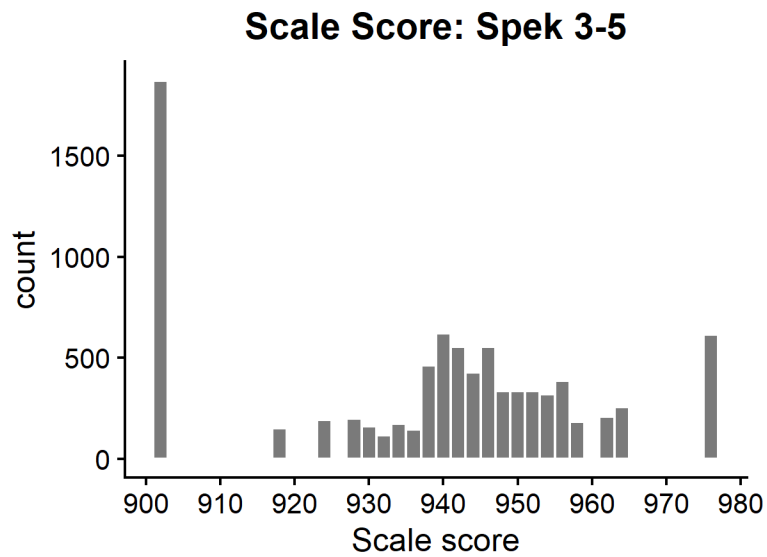


Figure 7.5.3.c.

Scale Score Distribution: 6–8

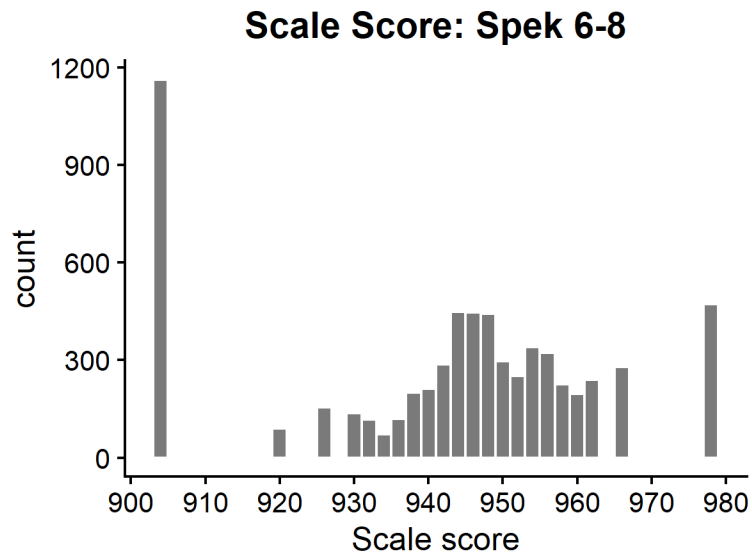
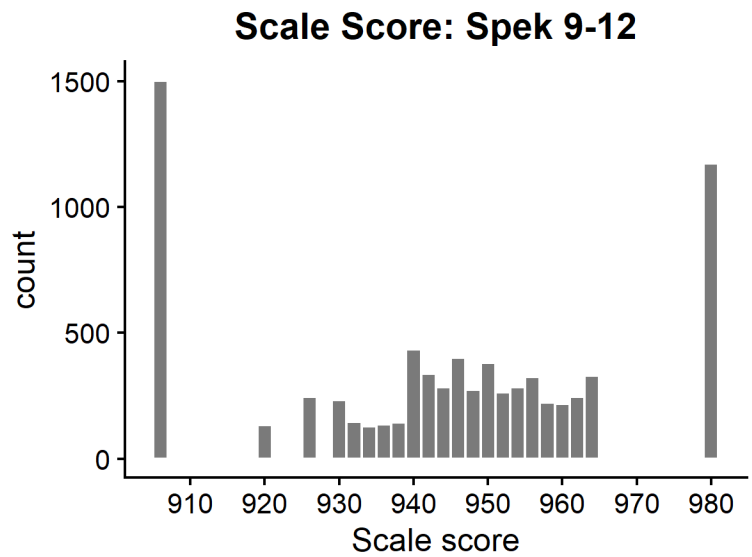


Figure 7.5.3.d.

Scale Score Distribution: 9–12



7.5.4. Writing

Table 7.5.4.a.

Scale Score Descriptive Statistics: K–2

Grade	Number of Students	Min.	Max.	Mean	SD
K	2,654	900	974	917.90	18.65
1	3,770	900	974	926.30	22.42
2	3,426	900	974	932.52	23.76
Total	9,850	900	974	926.20	22.69

Table 7.5.4.b.

Scale Score Descriptive Statistics: 3–5

Grade	Number of Students	Min.	Max.	Mean	SD
3	2,995	902	976	931.16	20.76
4	2,959	902	976	933.71	21.47
5	2,640	902	976	936.04	22.26
Total	8,594	902	976	933.54	21.56

Table 7.5.4.c.

Scale Score Descriptive Statistics: 6–8

Grade	Number of Students	Min.	Max.	Mean	SD
6	2,394	904	978	938.90	22.02
7	2,120	904	978	940.45	21.85
8	1,999	904	978	941.48	22.73
Total	6,513	904	978	940.19	22.21

Table 7.5.4.d.

Scale Score Descriptive Statistics: 9–12

Grade	Number of Students	Min.	Max.	Mean	SD
9	1,908	906	980	941.62	22.46
10	1,766	906	980	943.65	22.35
11	1,626	906	980	943.63	23.17
12	2,554	906	980	944.14	22.54
Total	7,854	906	980	943.31	22.63

Figure 7.5.4.a.

Scale Score Distribution: K-2

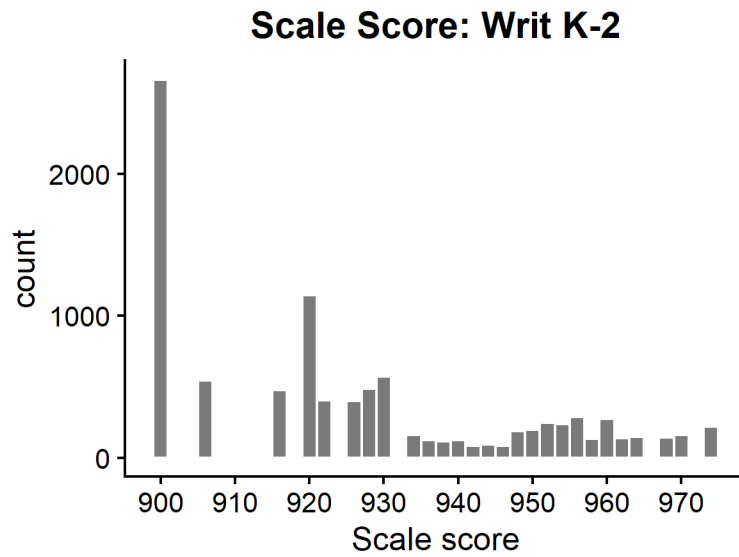


Figure 7.5.4.b.

Scale Score Distribution: 3-5

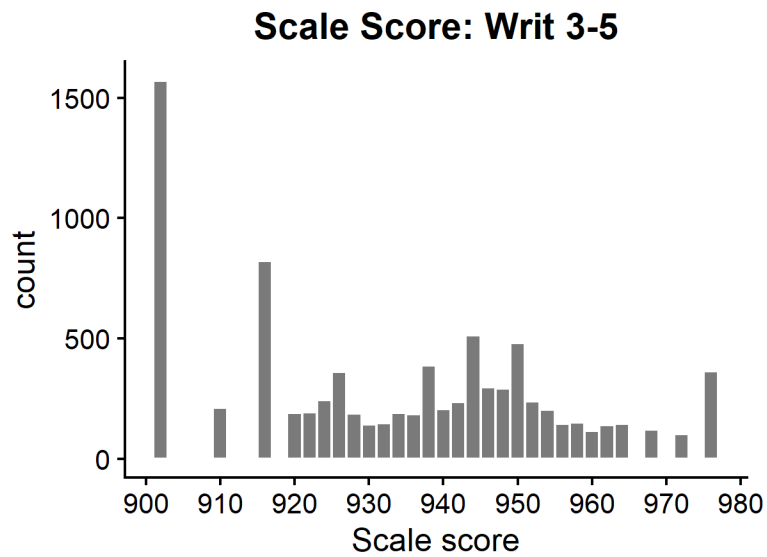


Figure 7.5.4.c.

Scale Score Distribution: 6–8

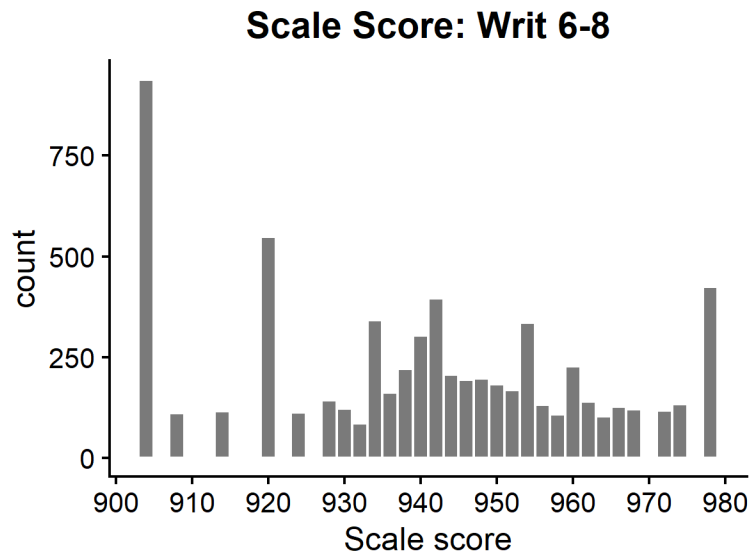
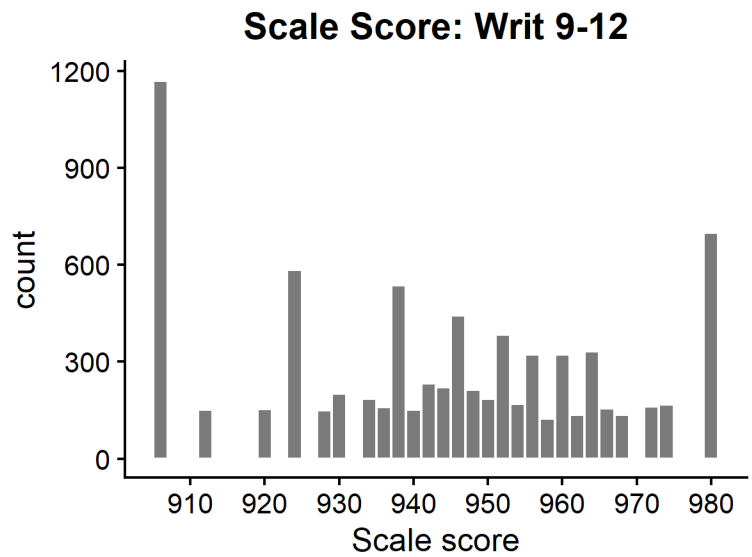


Figure 7.5.4.d.

Scale Score Distribution: 9–12



7.6. Proficiency Level Distribution

Tables 7.6.1.a. through 7.6.4.d. provide a summary of student performance by proficiency level (PL) for each test form across grades and grade-level clusters per domain. The proficiency levels are based on the new WIDA Alternate Proficiency Level Descriptors, which include five levels (P1–P5). The tables include the following information:

- The number of students (count) whose performance placed them into each proficiency level within the tested domain.
- The percentage of students, out of the total number of students taking the test form (by grade or grade-level cluster), who were classified into each proficiency level within the domain.

Figures 7.6.1.a. through 7.6.4.d. visually depict the distribution of proficiency levels across grades and grade-level clusters.

Key highlights are as follows:

Listening: The majority of students are classified into levels P1 and P2 across grade-level clusters, with proportions decreasing as grade levels increase. In K–2, 55.81% of students are at P1, compared to 23.70% in grades 9–12, reflecting overall progress in listening proficiency. The percentage of students at P5 increases from 3.16% in K–2 to 22.26% in grades 9–12, highlighting significant proficiency gains in higher grades.

Reading: A substantial proportion of students are classified at P1 across all grade-level clusters, with 71.65% in K–2 and 29.94% in grades 9–12. The percentage of students at P5 increases steadily across grades, from 3.16% in K–2 to 18.52% in grades 9–12, indicating growth in reading proficiency. Mid-level proficiency (P3 and P4) becomes more prominent in grades 6–8 and 9–12, with a noticeable balance across proficiency levels in these higher grades.

Speaking: In K–2, the vast majority of students (83.21%) are at P1, with minimal representation in higher proficiency levels. As grade levels increase, there is a marked decrease in the proportion of P1 students, dropping to 50.21% in grades 9–12. The percentage of students at P5 increases from 1.06% in K–2 to 14.96% in grades 9–12, showcasing gradual improvements in speaking proficiency.

Writing: Proficiency levels in writing follow a similar trend, with the highest percentage of students at P1 in K–2 (86.51%), decreasing to 54.19% in grades 9–12. The proportion of students achieving P5 rises consistently across grades, from 1.62% in K–2 to 11.05% in grades 9–12. Mid-level proficiency (P3 and P4) grows steadily in the upper grades, reflecting developmental progress in writing skills.

It should be noted that PL distributions are different with the new cut scores compared to Series 601. This is attributed to the fact that the new test form (602) is more difficult than the previous one (601). Such a high proportion of P1 is expected given the increased difficulty of the new test form.

7.6.1. Listening

Table 7.6.1.a.

Proficiency Level Distribution: K–2

Level	Grade K Count	Grade K Percent	Grade 1 Count	Grade 1 Percent	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	1,483	55.81%	1,510	40.01%	1,062	30.99%	4,055	41.13%
2	278	10.46%	404	10.70%	301	8.78%	983	9.97%
3	396	14.90%	636	16.85%	584	17.04%	1,616	16.39%
4	416	15.66%	896	23.74%	993	28.98%	2,305	23.38%
5	84	3.16%	328	8.69%	487	14.21%	899	9.12%
Total	2,657	100.0%	3,774	100.0%	3,427	100.0%	9,858	100.0%

Figure 7.6.1.a.

Proficiency Level Distribution: K–2

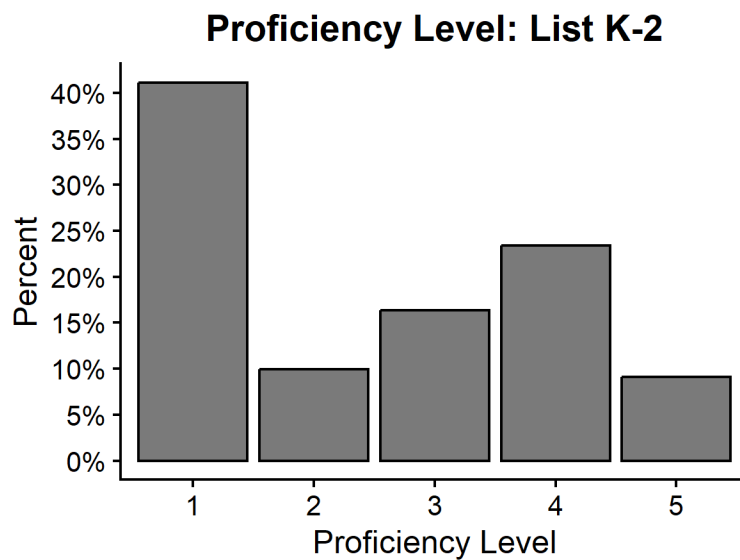


Table 7.6.1.b.

Proficiency Level Distribution: 3–5

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	917	30.55%	744	25.11%	604	22.86%	2,265	26.32%
2	583	19.42%	551	18.60%	419	15.86%	1,553	18.04%
3	715	23.82%	693	23.39%	641	24.26%	2,049	23.81%
4	533	17.75%	611	20.62%	587	22.22%	1,731	20.11%
5	254	8.46%	364	12.28%	391	14.80%	1,009	11.72%
Total	3,002	100.0%	2,963	100.0%	2,642	100.0%	8,607	100.0%

Figure 7.6.1.b.

Proficiency Level Distribution: 3–5

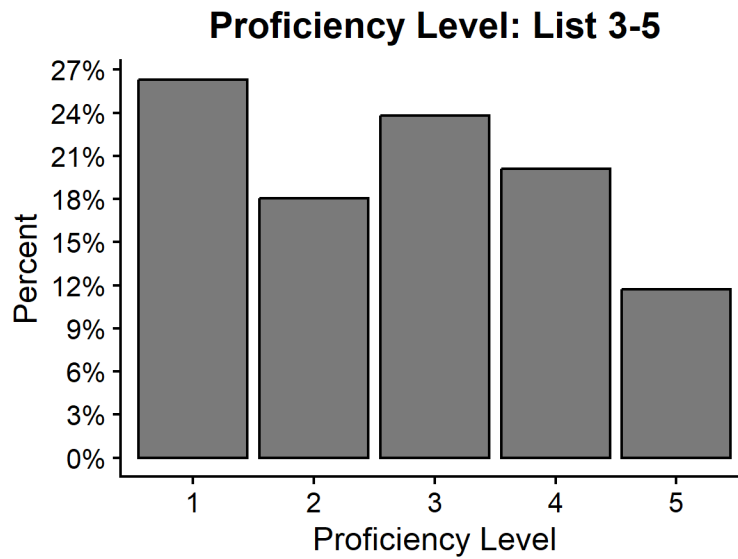


Table 7.6.1.c.

Proficiency Level Distribution: 6–8

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	547	22.81%	434	20.45%	401	20.04%	1,382	21.19%
2	409	17.06%	302	14.23%	287	14.34%	998	15.30%
3	624	26.02%	553	26.06%	484	24.19%	1,661	25.47%
4	320	13.34%	301	14.18%	290	14.49%	911	13.97%
5	498	20.77%	532	25.07%	539	26.94%	1,569	24.06%
Total	2,398	100.0%	2,122	100.0%	2,001	100.0%	6,521	100.0%

Figure 7.6.1.c.

Proficiency Level Distribution: 6–8

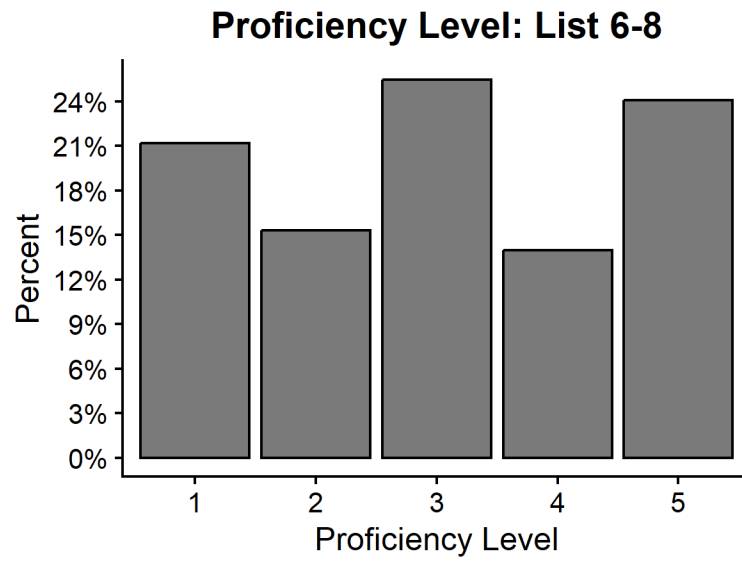


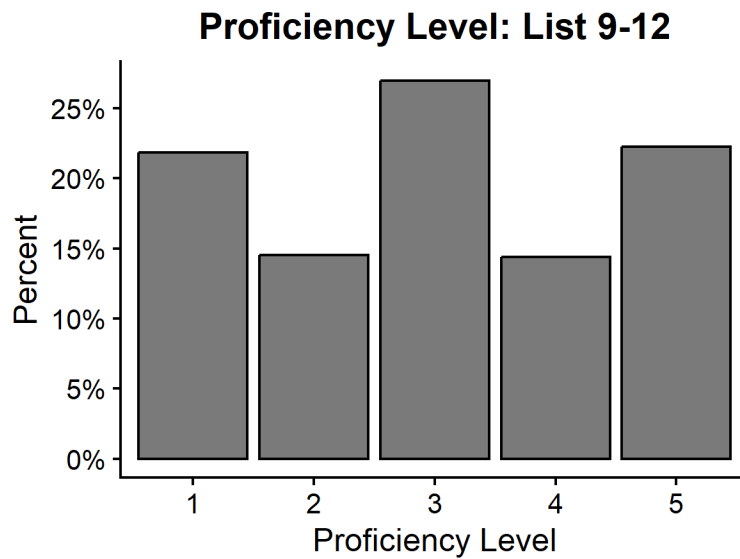
Table 7.6.1.d.

Proficiency Level Distribution: 9–12

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	453	23.70%	373	21.12%	344	21.16%	547	21.37%	1,717	21.84%
2	298	15.59%	253	14.33%	218	13.41%	374	14.61%	1,143	14.54%
3	502	26.27%	459	25.99%	463	28.47%	698	27.27%	2,122	26.99%
4	260	13.61%	253	14.33%	228	14.02%	390	15.23%	1,131	14.38%
5	398	20.83%	428	24.24%	373	22.94%	551	21.52%	1,750	22.26%
Total	1,911	100.0%	1,766	100.0%	1,626	100.0%	2,560	100.0%	7,863	100.0%

Figure 7.6.1.d.

Proficiency Level Distribution: 9–12



7.6.2. Reading

Table 7.6.2.a.

Proficiency Level Distribution: K-2

Level	Grade K Count	Grade K Percent	Grade 1 Count	Grade 1 Percent	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	1,903	71.65%	2,107	55.86%	1,532	44.70%	5,542	56.24%
2	370	13.93%	660	17.50%	608	17.74%	1,638	16.62%
3	208	7.83%	478	12.67%	507	14.79%	1,193	12.11%
4	91	3.43%	259	6.87%	350	10.21%	700	7.10%
5	84	3.16%	268	7.10%	430	12.55%	782	7.94%
Total	2,656	100.0%	3,772	100.0%	3,427	100.0%	9,855	100.0%

Figure 7.6.2.a.

Proficiency Level Distribution: K-2

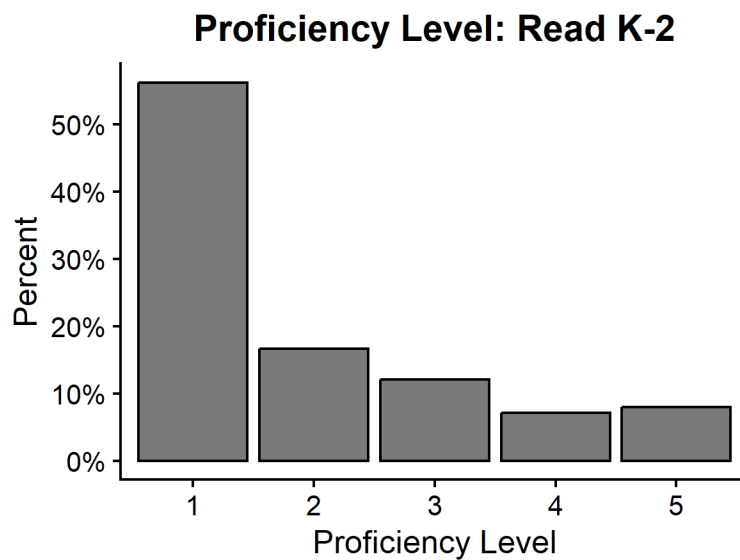


Table 7.6.2.b.

Proficiency Level Distribution: 3-5

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	1,382	46.08%	1,176	39.69%	941	35.66%	3,499	40.68%
2	782	26.08%	731	24.67%	645	24.44%	2,158	25.09%
3	428	14.27%	488	16.47%	404	15.31%	1,320	15.35%
4	241	8.04%	278	9.38%	315	11.94%	834	9.70%
5	166	5.54%	290	9.79%	334	12.66%	790	9.18%
Total	2,999	100.0%	2,963	100.0%	2,639	100.0%	8,601	100.0%

Figure 7.6.2.b.

Proficiency Level Distribution: 3–5

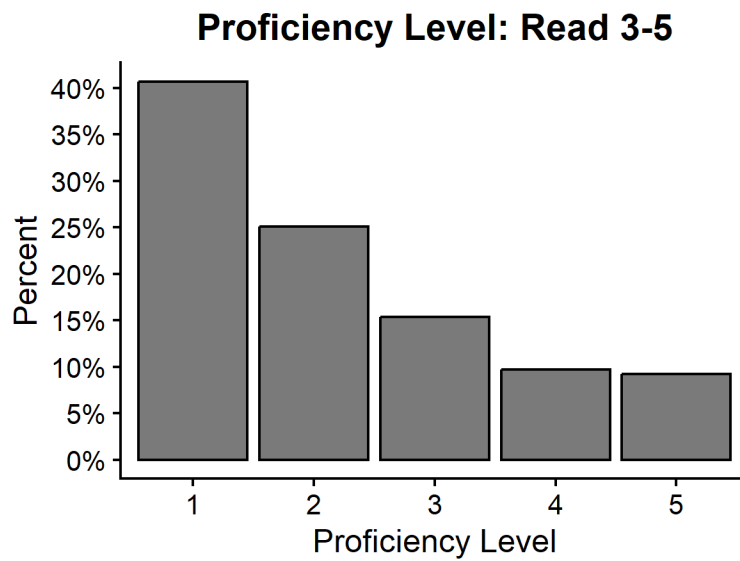


Table 7.6.2.c.

Proficiency Level Distribution: 6–8

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	778	32.44%	615	28.98%	574	28.69%	1967	30.16%
2	406	16.93%	364	17.15%	302	15.09%	1072	16.44%
3	542	22.60%	461	21.72%	404	20.19%	1407	21.58%
4	432	18.02%	418	19.70%	404	20.19%	1254	19.23%
5	240	10.01%	264	12.44%	317	15.84%	821	12.59%
Total	2398	100.0%	2122	100.0%	2001	100.0%	6521	100.0%

Figure 7.6.2.c.

Proficiency Level Distribution: 6–8

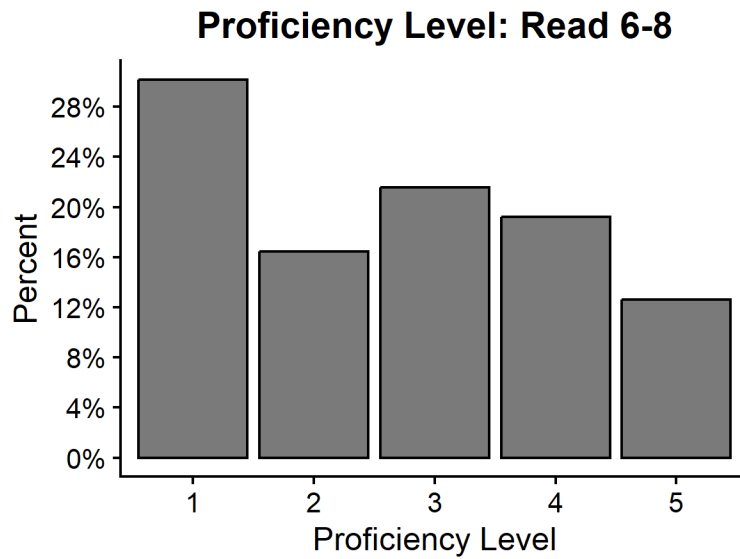


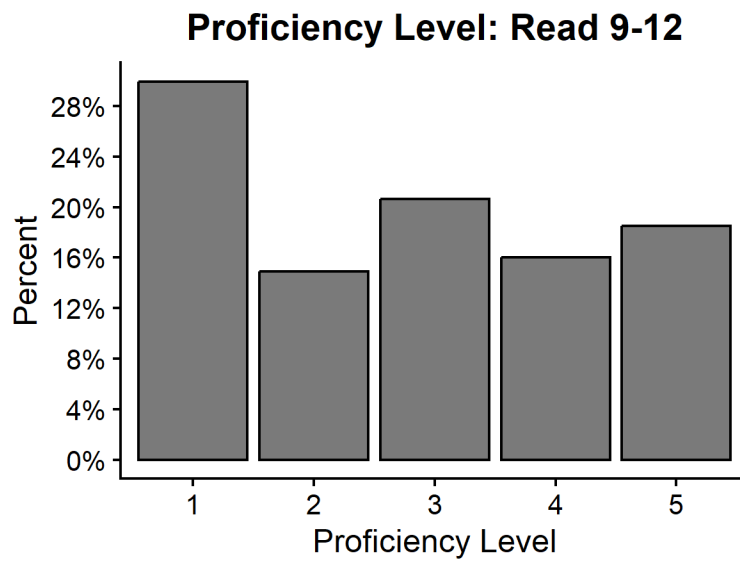
Table 7.6.2.d.

Proficiency Level Distribution: 9–12

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	586	30.68%	514	29.11%	477	29.34%	776	30.34%	2,353	29.94%
2	296	15.50%	242	13.70%	223	13.71%	410	16.03%	1,171	14.90%
3	429	22.46%	373	21.12%	326	20.05%	494	19.31%	1,622	20.64%
4	287	15.03%	292	16.53%	296	18.20%	383	14.97%	1,258	16.01%
5	312	16.34%	345	19.54%	304	18.70%	495	19.35%	1,456	18.52%
Total	1,910	100.0%	1,766	100.0%	1,626	100.0%	2,558	100.0%	7,860	100.0%

Figure 7.6.2.d.

Proficiency Level Distribution: 9–12



7.6.3. Speaking

Table 7.6.3.a.

Proficiency Level Distribution: K–2

Level	Grade K Count	Grade K Percent	Grade 1 Count	Grade 1 Percent	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	2,206	83.21%	2,722	72.26%	2,136	62.38%	7,064	71.77%
2	275	10.37%	487	12.93%	537	15.68%	1,299	13.20%
3	121	4.56%	326	8.65%	416	12.15%	863	8.77%
4	21	0.79%	91	2.42%	116	3.39%	228	2.32%
5	28	1.06%	141	3.74%	219	6.40%	388	3.94%
Total	2,651	100.0%	3,767	100.0%	3,424	100.0%	9,842	100.0%

Figure 7.6.3.a.

Proficiency Level Distribution: K-2

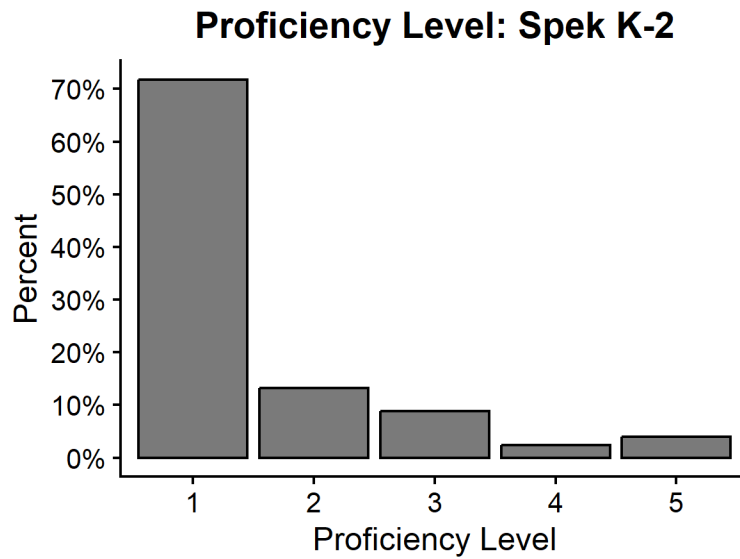


Table 7.6.3.b.

Proficiency Level Distribution: 3-5

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	1,893	63.08%	1,706	57.65%	1,474	55.83%	5,073	58.99%
2	495	16.49%	507	17.13%	392	14.85%	1,394	16.21%
3	270	9.00%	303	10.24%	296	11.21%	869	10.10%
4	107	3.57%	136	4.60%	149	5.64%	392	4.56%
5	236	7.86%	307	10.38%	329	12.46%	872	10.14%
Total	3,001	100.0%	2,959	100.0%	2,640	100.0%	8,600	100.0%

Figure 7.6.3.b.

Proficiency Level Distribution: 3–5

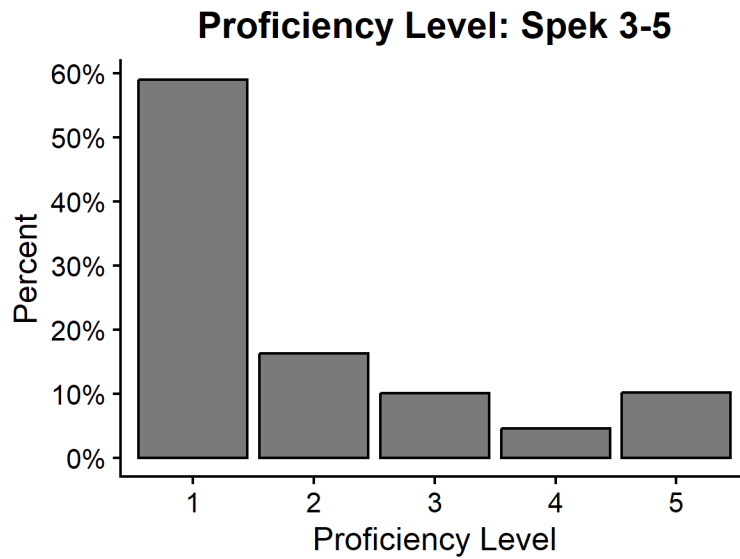


Table 7.6.3.c.

Proficiency Level Distribution: 6–8

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	1,155	48.25%	947	44.69%	899	44.99%	3,001	46.09%
2	516	21.55%	509	24.02%	412	20.62%	1,437	22.07%
3	369	15.41%	360	16.99%	354	17.72%	1,083	16.63%
4	95	3.97%	75	3.54%	70	3.50%	240	3.69%
5	259	10.82%	228	10.76%	263	13.16%	750	11.52%
Total	2,394	100.0%	2,119	100.0%	1,998	100.0%	6,511	100.0%

Figure 7.6.3.c.

Proficiency Level Distribution: 6–8

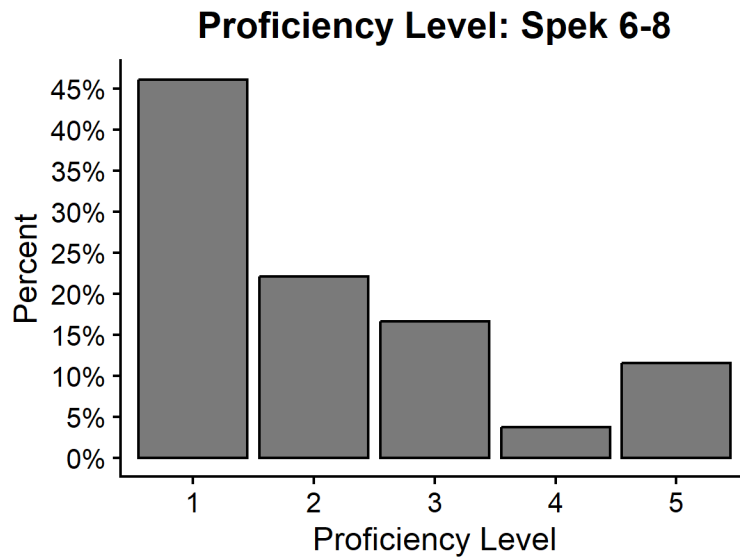


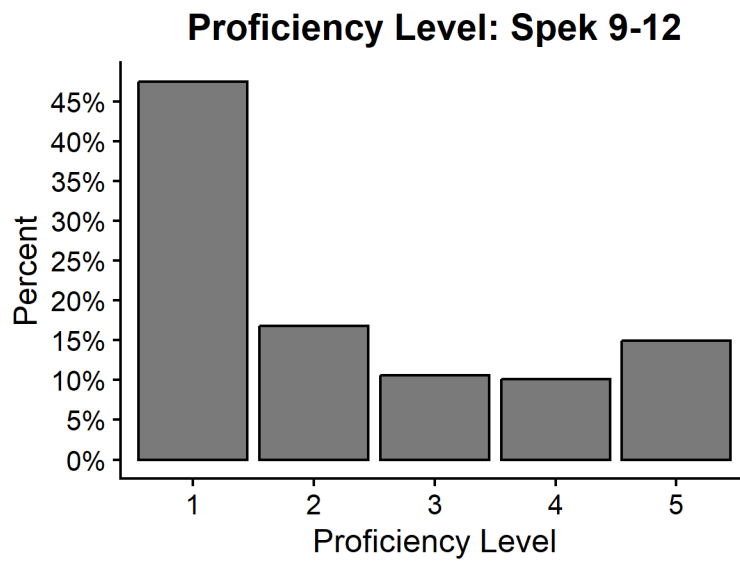
Table 7.6.3.d.

Proficiency Level Distribution: 9–12

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	958	50.21%	792	44.90%	760	46.74%	1,220	47.75%	3,730	47.50%
2	307	16.09%	307	17.40%	254	15.62%	452	17.69%	1,320	16.81%
3	198	10.38%	183	10.37%	182	11.19%	269	10.53%	832	10.59%
4	177	9.28%	184	10.43%	182	11.19%	253	9.90%	796	10.14%
5	268	14.05%	298	16.89%	248	15.25%	361	14.13%	1,175	14.96%
Total	1,908	100.0%	1,764	100.0%	1,626	100.0%	2,555	100.0%	7,853	100.0%

Figure 7.6.3.d.

Proficiency Level Distribution: 9–12



7.6.4. Writing

Table 7.6.4.a.

Proficiency Level Distribution: K–2

Level	Grade K Count	Grade K Percent	Grade 1 Count	Grade 1 Percent	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	2,296	86.51%	2,715	72.02%	2,084	60.83%	7,095	72.03%
2	137	5.16%	256	6.79%	281	8.20%	674	6.84%
3	124	4.67%	410	10.88%	470	13.72%	1,004	10.19%
4	54	2.03%	213	5.65%	291	8.49%	558	5.66%
5	43	1.62%	176	4.67%	300	8.76%	519	5.27%
Total	2,654	100.0%	3,770	100.0%	3,426	100.0%	9,850	100.0%

Figure 7.6.4.a.

Proficiency Level Distribution: K-2

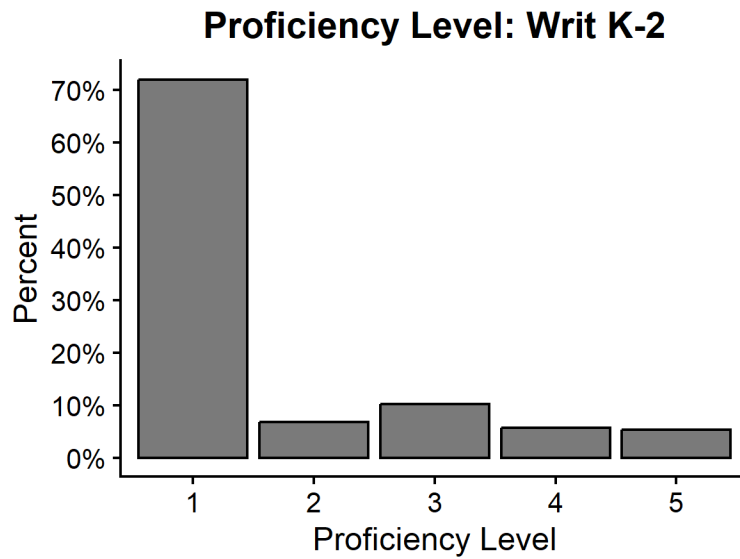


Table 7.6.4.b.

Proficiency Level Distribution: 3-5

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	1,895	63.27%	1,733	58.57%	1,417	53.67%	5,045	58.70%
2	639	21.34%	626	21.16%	552	20.91%	1,817	21.14%
3	222	7.41%	255	8.62%	262	9.92%	739	8.60%
4	98	3.27%	143	4.83%	162	6.14%	403	4.69%
5	141	4.71%	202	6.83%	247	9.36%	590	6.87%
Total	2,995	100.0%	2,959	100.0%	2,640	100.0%	8,594	100.0%

Figure 7.6.4.b.

Proficiency Level Distribution: 3–5

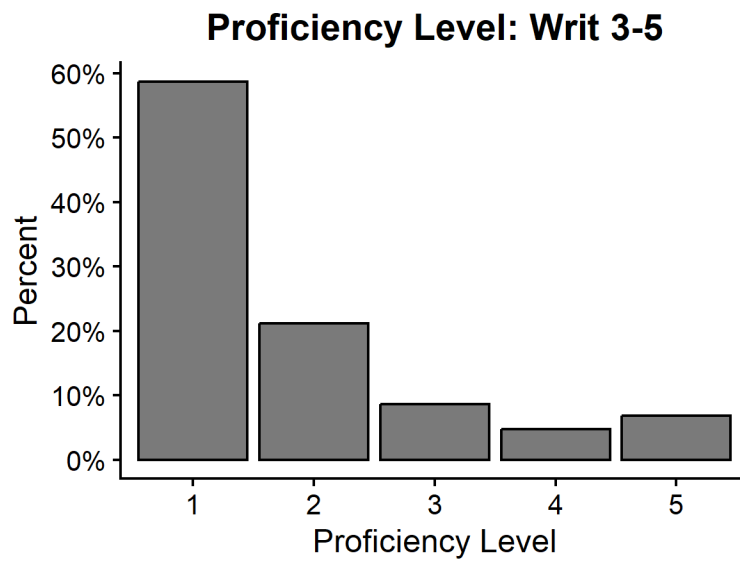


Table 7.6.4.c.

Proficiency Level Distribution: 6–8

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	1,388	57.98%	1,152	54.34%	1,058	52.93%	3,598	55.24%
2	422	17.63%	385	18.16%	337	16.86%	1,144	17.56%
3	206	8.60%	215	10.14%	185	9.25%	606	9.30%
4	145	6.06%	168	7.92%	177	8.85%	490	7.52%
5	233	9.73%	200	9.43%	242	12.11%	675	10.36%
Total	2,394	100.0%	2,120	100.0%	1,999	100.0%	6,513	100.0%

Figure 7.6.4.c.

Proficiency Level Distribution: 6–8

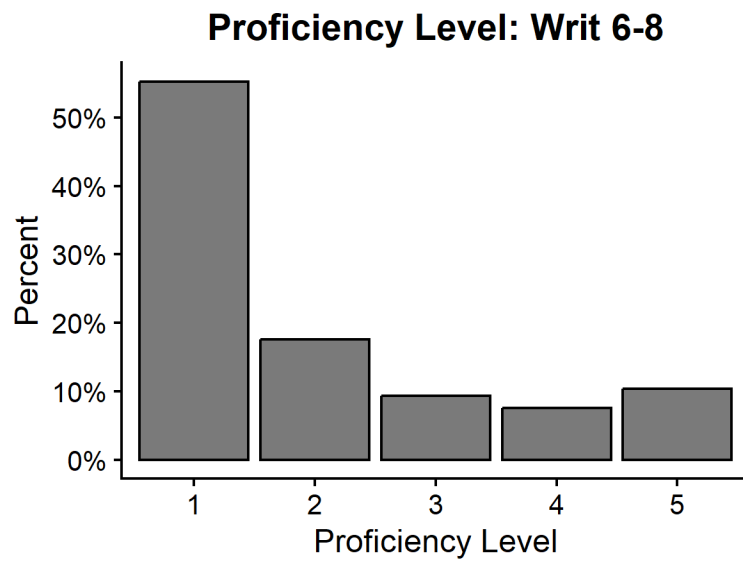


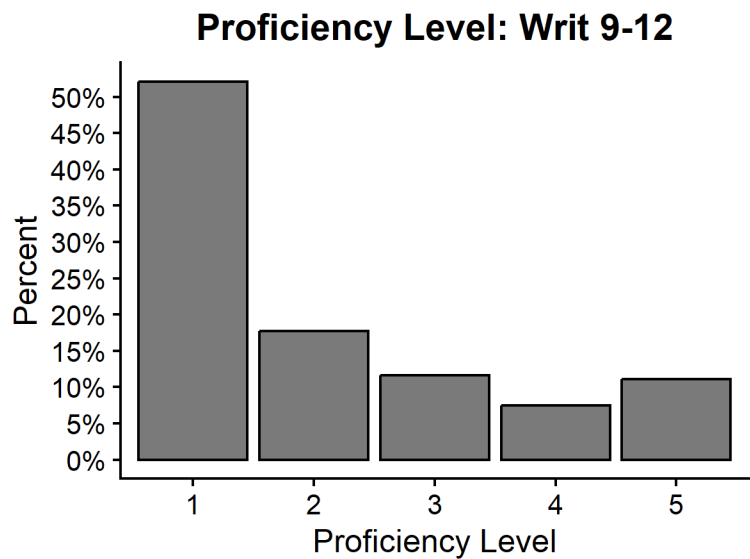
Table 7.6.4.d.

Proficiency Level Distribution: 9–12

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	1,034	54.19%	922	52.21%	830	51.05%	1,307	51.17%	4,093	52.11%
2	355	18.61%	310	17.55%	291	17.90%	437	17.11%	1,393	17.74%
3	227	11.90%	205	11.61%	175	10.76%	308	12.06%	915	11.65%
4	122	6.39%	129	7.30%	132	8.12%	202	7.91%	585	7.45%
5	170	8.91%	200	11.33%	198	12.18%	300	11.75%	868	11.05%
Total	1,908	100.0%	1,766	100.0%	1,626	100.0%	2,554	100.0%	7,854	100.0%

Figure 7.6.4.d.

Proficiency Level Distribution: 9–12



7.7. *Raw Score to Scale Score Proficiency Level Conversion*

In this section, the tables ending in a, c, e, and g present the raw score to scale score conversion for each grade-level cluster, by domain. Table 7.7.1.a., 7.7.1.c., 7.7.1.e, and 7.7.1.g. present raw score to scale score conversions for the Listening domain across grade-level clusters K–2, 3–5, 6–8, and 9–12, accordingly. Tables 7.7.2.a. to 7.7.4.g. provide similar conversions for Reading, Speaking, and Writing.

The first column (left) shows all possible raw scores. The second column shows the corresponding scale score for the grade-level cluster. The third column shows the conditional standard error (i.e., from the Rasch analysis) in the metric of the scale score. The last two columns (fourth and fifth) show a lower bound (i.e., the scale score minus one standard error) and an upper bound (i.e., the scale score plus one standard error) around the scale score. In some cases, the resulting lower bound or upper bound is below 910, which has been set as the lowest score on the scale. All domains were adjusted for an end-of-scale effect per cluster by allowing the top and bottom scale scores to increase only at the same rate as the preceding scale scores. If they were not adjusted, their effect in the composite scores might be excessive.

Thus, if the scale scores towards the high end of the raw score scale were increasing with each raw score by 9 scale points before the group of adjusted scores, then each of the adjusted scores would increase by only 9 scale points. Because the lower and upper bounds were calculated based on the original logit scores, these adjusted scores do not fall in the middle of the range; they fall toward the lower end of the range, but they always fall within the range. In other words, the adjusted scale score is a very possible observed score for that number of raw score points obtained. In addition, at the lower end of the raw score scale, scale scores are truncated when necessary, so that the lowest scale score given is the scale score corresponding to a proficiency level score of P1.

Tables 7.7.1.b., 7.7.1.d., 7.7.1.f., and 7.7.1.h display proficiency level interpretations for Listening across the same grade-level clusters, with Tables 7.7.2.b. to 7.7.4.h. covering Reading, Speaking, and Writing.

The first column in Table I shows the raw score. The remaining columns show the proficiency level score associated with each raw score/scale score for each grade in the cluster, the percentage of students in that grade who scored at that raw score/scale score/proficiency level score, and the cumulative percentage of students in that grade who scored up to that raw score/scale score/proficiency level score.

There are two things to note about this table. First, unlike scale scores, which are determined psychometrically and have a one-to-one correspondence to raw scores regardless of the grade level of the student, proficiency level scores are interpretations of the scale score. Second, for Alternate ACCESS, cut scores between proficiency levels were determined by domain and by grade-level cluster and do not change by grade level.

For students with severe cognitive disabilities, the cognitive abilities that support language proficiency development are not expected to increase dramatically from one grade level to the next, but steady growth was observed across grades and grade-level clusters in each domain. At this point in the understanding of these students' development of ELP according to the Alternate ACCESS data observed in the past, it appears appropriate to use the same cut scores per grade-level cluster levels (from grades K to 12) by domain. In this way, it becomes easier to detect growth in ELP from year to year for this population of English learners.

7.7.1. Listening

Table 7.7.1.a.

Raw Score to Scale Score Conversion: K–2

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	900	29	900	929
1	910	16	900	926
2	916	11	905	927
3	920	9	911	929
4	922	8	914	930
5	925	7	918	932
6	926	7	919	933
7	928	7	921	935
8	930	6	924	936
9	931	6	925	937
10	933	6	927	939
11	934	5	929	939
12	935	5	930	940
13	936	5	931	941
14	937	5	932	942
15	938	5	933	943

16	939	5	934	944
17	939	4	935	943
18	940	4	936	944
19	941	4	937	945
20	942	4	938	946
21	942	4	938	946
22	943	4	939	947
23	944	4	940	948
24	944	4	940	948
25	945	4	941	949
26	946	4	942	950
27	946	4	942	950
28	947	4	943	951
29	948	4	944	952
30	948	4	944	952
31	949	4	945	953
32	950	5	945	955
33	951	5	946	956
34	952	5	947	957
35	953	6	947	959
36	954	6	948	960
37	956	7	949	963
38	959	10	949	969
39	962	12	950	974
40	974	31	943	980

Table 7.7.1.b.

Raw Score to Proficiency Level Conversion: K-2

Raw Score	Grade K Proficiency Level Score	Grade K % of Students	Grade K Cumulative % of Students	Grade 1 Proficiency Level Score	Grade 1 % of Students	Grade 1 Cumulative % of Students	Grade 2 Proficiency Level Score	Grade 2 % of Students	Grade 2 Cumulative % of Students
0	P1	21.79	21.79	P1	14.76	14.76	P1	9.43	9.43
1	P1	2.75	24.54	P1	2.36	17.12	P1	1.90	11.32
2	P1	3.35	27.89	P1	2.01	19.13	P1	1.63	12.96
3	P1	5.57	33.46	P1	3.71	22.84	P1	2.98	15.93
4	P1	3.61	37.07	P1	2.76	25.60	P1	2.22	18.15
5	P1	2.60	39.67	P1	2.17	27.77	P1	1.66	19.81
6	P1	2.63	42.30	P1	1.59	29.36	P1	1.75	21.56
7	P1	3.16	45.46	P1	2.44	31.80	P1	2.07	23.64

Raw Score	Grade K Proficiency Level Score	Grade K % of Students	Grade K Cumulative % of Students	Grade 1 Proficiency Level Score	Grade 1 % of Students	Grade 1 Cumulative % of Students	Grade 2 Proficiency Level Score	Grade 2 % of Students	Grade 2 Cumulative % of Students
8	P1	2.26	47.72	P1	1.75	33.55	P1	1.23	24.86
9	P1	1.73	49.45	P1	1.03	34.58	P1	1.20	26.06
10	P1	1.54	51.00	P1	1.51	36.09	P1	1.11	27.17
11	P1	2.22	53.22	P1	1.54	37.63	P1	1.49	28.65
12	P1	1.32	54.54	P1	1.03	38.66	P1	1.14	29.79
13	P1	1.28	55.81	P1	1.35	40.01	P1	1.20	30.99
14	P2	1.32	57.13	P2	1.38	41.39	P2	0.88	31.86
15	P2	1.73	58.86	P2	1.30	42.69	P2	1.23	33.09
16	P2	1.62	60.48	P2	1.17	43.85	P2	1.20	34.29
17	P2	1.13	61.61	P2	1.30	45.15	P2	0.93	35.22
18	P2	0.90	62.51	P2	1.48	46.63	P2	1.25	36.48
19	P2	1.32	63.83	P2	1.56	48.20	P2	1.34	37.82
20	P2	1.28	65.11	P2	1.19	49.39	P2	0.88	38.69
21	P2	1.17	66.28	P2	1.32	50.72	P2	1.08	39.77
22	P3	1.66	67.93	P3	1.25	51.96	P3	1.43	41.20
23	P3	1.54	69.48	P3	1.99	53.95	P3	1.63	42.84
24	P3	1.02	70.49	P3	1.25	55.19	P3	1.55	44.38
25	P3	1.62	72.11	P3	1.32	56.52	P3	1.63	46.02
26	P3	1.43	73.54	P3	1.93	58.45	P3	1.90	47.91
27	P3	1.28	74.82	P3	2.07	60.52	P3	1.98	49.90
28	P3	2.15	76.97	P3	2.38	62.90	P3	2.07	51.97
29	P3	1.88	78.85	P3	2.33	65.24	P3	2.33	54.30
30	P3	2.33	81.18	P3	2.33	67.57	P3	2.51	56.81
31	P4	2.41	83.59	P4	3.42	70.99	P4	3.65	60.46
32	P4	2.22	85.81	P4	2.91	73.90	P4	2.86	63.32
33	P4	2.86	88.67	P4	3.52	77.42	P4	3.76	67.08
34	P4	2.30	90.97	P4	3.84	81.27	P4	4.29	71.37
35	P4	1.77	92.74	P4	2.91	84.18	P4	4.23	75.61
36	P4	2.26	94.99	P4	3.68	87.86	P4	4.73	80.33
37	P4	1.84	96.84	P4	3.44	91.31	P4	5.46	85.79
38	P5	0.94	97.78	P5	3.07	94.38	P5	4.41	90.20
39	P5	1.05	98.83	P5	2.65	97.03	P5	4.55	94.75
40	P5	1.17	100.0	P5	2.97	100.0	P5	5.25	100.0

Table 7.7.1.c.**Raw Score to Scale Score Conversion: 3–5**

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	902	31	900	933
1	914	16	900	930
2	920	11	909	931
3	924	9	915	933
4	926	8	918	934
5	928	7	921	935
6	930	6	924	936
7	931	6	925	937
8	932	5	927	937
9	934	5	929	939
10	935	5	930	940
11	936	5	931	941
12	936	5	931	941
13	937	5	932	942
14	938	4	934	942
15	939	4	935	943
16	940	4	936	944
17	940	4	936	944
18	941	4	937	945
19	942	4	938	946
20	942	4	938	946
21	943	4	939	947
22	944	4	940	948
23	944	4	940	948
24	945	4	941	949
25	946	4	942	950
26	946	4	942	950
27	947	4	943	951
28	948	4	944	952
29	948	4	944	952
30	949	4	945	953
31	950	4	946	954
32	951	5	946	956
33	951	5	946	956
34	952	5	947	957
35	954	6	948	960
36	955	6	949	961

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
37	957	8	949	965
38	959	10	949	969
39	961	11	950	972
40	976	33	943	980

Table 7.7.1.d.

Raw Score to Proficiency Level Conversion: 3–5

Raw Score	Grade 3 Proficiency Level Score	Grade 3 % of Students	Grade 3 Cumulative % of Students	Grade 4 Proficiency Level Score	Grade 4 % of Students	Grade 4 Cumulative % of Students	Grade 5 Proficiency Level Score	Grade 5 % of Students	Grade 5 Cumulative % of Students
0	P1	9.13	9.13	P1	7.59	7.59	P1	6.59	6.59
1	P1	1.73	10.86	P1	1.38	8.98	P1	0.95	7.53
2	P1	1.43	12.29	P1	1.32	10.29	P1	1.14	8.67
3	P1	2.80	15.09	P1	2.33	12.62	P1	2.38	11.05
4	P1	1.40	16.49	P1	1.08	13.70	P1	0.98	12.04
5	P1	1.07	17.55	P1	0.84	14.55	P1	0.76	12.79
6	P1	1.43	18.99	P1	1.42	15.96	P1	0.87	13.66
7	P1	1.43	20.42	P1	1.38	17.35	P1	1.14	14.80
8	P1	1.60	22.02	P1	1.08	18.43	P1	0.83	15.63
9	P1	1.23	23.25	P1	1.01	19.44	P1	1.21	16.84
10	P1	0.77	24.02	P1	0.94	20.38	P1	1.02	17.87
11	P1	2.00	26.02	P1	1.35	21.73	P1	1.17	19.04
12	P1	0.87	26.88	P1	0.78	22.51	P1	0.68	19.72
13	P1	1.13	28.01	P1	0.78	23.29	P1	1.02	20.74
14	P1	1.30	29.31	P1	0.94	24.23	P1	1.17	21.92
15	P1	1.23	30.55	P1	0.88	25.11	P1	0.95	22.86
16	P2	1.20	31.75	P2	1.42	26.53	P2	0.91	23.77
17	P2	1.20	32.94	P2	1.32	27.84	P2	1.17	24.94
18	P2	0.90	33.84	P2	1.05	28.89	P2	0.61	25.55
19	P2	1.27	35.11	P2	0.94	29.83	P2	1.06	26.61
20	P2	1.53	36.64	P2	1.01	30.85	P2	0.72	27.33
21	P2	1.40	38.04	P2	1.42	32.26	P2	1.06	28.39
22	P2	1.77	39.81	P2	1.35	33.61	P2	1.40	29.79
23	P2	1.70	41.51	P2	1.52	35.13	P2	1.21	31.00
24	P2	1.77	43.27	P2	1.92	37.06	P2	2.16	33.16
25	P2	1.80	45.07	P2	2.19	39.25	P2	1.63	34.78
26	P2	2.50	47.57	P2	2.09	41.34	P2	1.67	36.45
27	P2	2.40	49.97	P2	2.36	43.71	P2	2.27	38.72

Raw Score	Grade 3 Proficiency Level Score	Grade 3 % of Students	Grade 3 Cumulative % of Students	Grade 4 Proficiency Level Score	Grade 4 % of Students	Grade 4 Cumulative % of Students	Grade 5 Proficiency Level Score	Grade 5 % of Students	Grade 5 Cumulative % of Students
28	P3	3.20	53.16	P3	2.46	46.17	P3	2.38	41.11
29	P3	2.47	55.63	P3	2.83	49.00	P3	3.33	44.44
30	P3	3.36	58.99	P3	3.00	52.01	P3	3.18	47.62
31	P3	3.83	62.82	P3	3.31	55.32	P3	3.44	51.06
32	P3	3.46	66.29	P3	3.17	58.49	P3	3.71	54.77
33	P3	3.70	69.99	P3	4.25	62.74	P3	4.13	58.89
34	P3	3.80	73.78	P3	4.35	67.09	P3	4.09	62.98
35	P4	3.63	77.42	P4	4.49	71.58	P4	4.13	67.11
36	P4	4.20	81.61	P4	5.13	76.71	P4	5.37	72.48
37	P4	4.83	86.44	P4	5.77	82.48	P4	6.47	78.96
38	P4	5.10	91.54	P4	5.23	87.72	P4	6.25	85.20
39	P5	4.00	95.54	P5	5.94	93.66	P5	6.55	91.75
40	P5	4.46	100.0	P5	6.34	100.0	P5	8.25	100.0

Table 7.7.1.e.

Raw Score to Scale Score Conversion: 6–8

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	904	32	900	936
1	917	16	901	933
2	923	11	912	934
3	927	9	918	936
4	929	8	921	937
5	931	7	924	938
6	933	6	927	939
7	934	6	928	940
8	935	5	930	940
9	936	5	931	941
10	937	5	932	942
11	938	5	933	943
12	939	4	935	943
13	940	4	936	944
14	941	4	937	945
15	941	4	937	945
16	942	4	938	946
17	942	4	938	946
18	943	4	939	947

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
19	944	4	940	948
20	944	4	940	948
21	945	4	941	949
22	945	4	941	949
23	946	4	942	950
24	946	4	942	950
25	947	4	943	951
26	948	4	944	952
27	948	4	944	952
28	949	4	945	953
29	949	4	945	953
30	950	4	946	954
31	951	4	947	955
32	952	4	948	956
33	952	5	947	957
34	953	5	948	958
35	955	6	949	961
36	956	6	950	962
37	958	8	950	966
38	960	10	950	970
39	962	12	950	974
40	978	31	947	980

Table 7.7.1.f.

Raw Score to Proficiency Level Conversion: 6-8

Raw Score	Grade 6 Proficiency Level Score	Grade 6 % of Students	Grade 6 Cumulative % of Students	Grade 7 Proficiency Level Score	Grade 7 % of Students	Grade 7 Cumulative % of Students	Grade 8 Proficiency Level Score	Grade 8 % of Students	Grade 8 Cumulative % of Students
0	P1	6.26	6.26	P1	5.42	5.42	P1	5.40	5.40
1	P1	1.08	7.34	P1	1.13	6.55	P1	0.90	6.30
2	P1	1.13	8.47	P1	0.61	7.16	P1	0.70	7.00
3	P1	2.38	10.84	P1	2.50	9.66	P1	2.30	9.30
4	P1	0.75	11.59	P1	0.71	10.37	P1	0.60	9.90
5	P1	0.75	12.34	P1	0.71	11.07	P1	0.60	10.49
6	P1	0.83	13.18	P1	0.66	11.73	P1	0.60	11.09
7	P1	1.29	14.47	P1	1.37	13.10	P1	1.15	12.24
8	P1	1.13	15.60	P1	0.90	14.00	P1	0.95	13.19
9	P1	1.00	16.60	P1	0.75	14.75	P1	1.05	14.24

Raw Score	Grade 6 Proficiency Level Score	Grade 6 % of Students	Grade 6 Cumulative % of Students	Grade 7 Proficiency Level Score	Grade 7 % of Students	Grade 7 Cumulative % of Students	Grade 8 Proficiency Level Score	Grade 8 % of Students	Grade 8 Cumulative % of Students
10	P1	0.79	17.39	P1	0.57	15.32	P1	0.70	14.94
11	P1	0.79	18.18	P1	1.13	16.45	P1	0.70	15.64
12	P1	1.04	19.22	P1	0.71	17.15	P1	0.35	15.99
13	P1	0.71	19.93	P1	0.52	17.67	P1	0.75	16.74
14	P1	0.63	20.56	P1	0.47	18.14	P1	0.70	17.44
15	P1	0.79	21.35	P1	0.99	19.13	P1	0.85	18.29
16	P1	0.71	22.06	P1	0.71	19.84	P1	0.90	19.19
17	P1	0.75	22.81	P1	0.61	20.45	P1	0.85	20.04
18	P2	0.79	23.60	P2	0.52	20.97	P2	0.70	20.74
19	P2	0.67	24.27	P2	0.85	21.82	P2	0.75	21.49
20	P2	0.96	25.23	P2	0.71	22.53	P2	0.95	22.44
21	P2	1.04	26.27	P2	0.94	23.47	P2	1.45	23.89
22	P2	1.08	27.36	P2	1.27	24.74	P2	0.55	24.44
23	P2	1.17	28.52	P2	1.08	25.82	P2	0.85	25.29
24	P2	1.21	29.73	P2	0.99	26.81	P2	0.70	25.99
25	P2	1.96	31.69	P2	1.70	28.51	P2	1.25	27.24
26	P2	1.96	33.65	P2	0.80	29.31	P2	1.35	28.59
27	P2	2.21	35.86	P2	2.17	31.48	P2	1.60	30.18
28	P2	1.88	37.74	P2	1.13	32.61	P2	2.25	32.43
29	P2	2.13	39.87	P2	2.07	34.68	P2	1.95	34.38
30	P3	2.54	42.41	P3	2.36	37.04	P3	2.35	36.73
31	P3	3.25	45.66	P3	2.83	39.87	P3	3.35	40.08
32	P3	2.38	48.04	P3	2.97	42.84	P3	2.50	42.58
33	P3	3.25	51.29	P3	3.49	46.32	P3	3.30	45.88
34	P3	5.05	56.34	P3	4.01	50.33	P3	3.70	49.58
35	P3	3.84	60.18	P3	4.85	55.18	P3	4.40	53.97
36	P3	5.71	65.89	P3	5.56	60.74	P3	4.60	58.57
37	P4	6.71	72.60	P4	6.69	67.44	P4	6.80	65.37
38	P4	6.63	79.23	P4	7.49	74.93	P4	7.70	73.06
39	P5	7.59	86.82	P5	9.28	84.21	P5	10.44	83.51
40	P5	13.18	100.0	P5	15.79	100.0	P5	16.49	100.0

Table 7.7.1.g.

Raw Score to Scale Score Conversion: 9–12

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	906	30	900	936

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
1	917	16	901	933
2	923	11	912	934
3	927	9	918	936
4	929	8	921	937
5	931	7	924	938
6	933	6	927	939
7	934	6	928	940
8	935	5	930	940
9	936	5	931	941
10	937	5	932	942
11	938	5	933	943
12	939	4	935	943
13	940	4	936	944
14	940	4	936	944
15	941	4	937	945
16	942	4	938	946
17	942	4	938	946
18	943	4	939	947
19	944	4	940	948
20	944	4	940	948
21	945	4	941	949
22	946	4	942	950
23	946	4	942	950
24	947	4	943	951
25	948	4	944	952
26	948	4	944	952
27	949	4	945	953
28	950	4	946	954
29	950	4	946	954
30	951	4	947	955
31	952	5	947	957
32	953	5	948	958
33	954	5	949	959
34	955	5	950	960
35	956	6	950	962
36	957	6	951	963
37	959	8	951	967
38	962	10	952	972
39	965	13	952	978
40	980	38	942	980

Table 7.7.1.h.**Raw Score to Proficiency Level Conversion: 9–12**

Raw Score	Grade 9 PL Score	Grade 9 % of Students	Grade 9 Cumulative % of Students	Grade 10 PL Score	Grade 10 % of Students	Grade 10 Cumulative % of Students	Grade 11 PL Score	Grade 11 % of Students	Grade 11 Cumulative % of Students	Grade 12 PL Score	Grade 12 % of Students	Grade 12 Cumulative % of Students
0	P1	7.06	7.06	P1	5.38	5.38	P1	5.29	5.29	P1	5.94	5.94
1	P1	0.68	7.74	P1	0.62	6.00	P1	1.05	6.33	P1	0.63	6.56
2	P1	0.78	8.53	P1	0.74	6.74	P1	0.55	6.89	P1	0.39	6.95
3	P1	1.83	10.36	P1	2.49	9.23	P1	1.35	8.24	P1	2.15	9.10
4	P1	0.42	10.78	P1	0.57	9.80	P1	0.31	8.55	P1	0.39	9.49
5	P1	0.52	11.30	P1	0.74	10.53	P1	0.49	9.04	P1	0.59	10.08
6	P1	0.73	12.04	P1	0.68	11.21	P1	0.62	9.66	P1	0.55	10.62
7	P1	1.31	13.34	P1	0.62	11.83	P1	1.35	11.01	P1	0.90	11.52
8	P1	0.63	13.97	P1	0.57	12.40	P1	0.18	11.19	P1	0.82	12.34
9	P1	0.78	14.76	P1	0.17	12.57	P1	0.68	11.87	P1	0.51	12.85
10	P1	0.37	15.12	P1	0.45	13.02	P1	0.62	12.48	P1	0.59	13.44
11	P1	0.89	16.01	P1	0.40	13.42	P1	1.05	13.53	P1	0.55	13.98
12	P1	0.89	16.90	P1	0.68	14.10	P1	0.62	14.15	P1	0.82	14.80
13	P1	0.89	17.79	P1	0.34	14.44	P1	0.62	14.76	P1	0.59	15.39
14	P1	0.84	18.63	P1	0.74	15.18	P1	0.86	15.62	P1	0.51	15.90
15	P1	1.05	19.68	P1	0.85	16.02	P1	1.11	16.73	P1	0.82	16.72
16	P1	1.36	21.04	P1	1.42	17.44	P1	1.17	17.90	P1	1.05	17.77
17	P1	0.84	21.87	P1	1.13	18.57	P1	0.86	18.76	P1	0.74	18.52
18	P1	0.58	22.45	P1	0.74	19.31	P1	0.49	19.25	P1	0.74	19.26
19	P1	0.63	23.08	P1	1.08	20.39	P1	0.80	20.05	P1	1.02	20.27
20	P1	0.63	23.70	P1	0.74	21.12	P1	1.11	21.16	P1	1.09	21.37
21	P2	0.99	24.70	P2	0.68	21.80	P2	0.31	21.46	P2	0.86	22.23
22	P2	1.15	25.85	P2	0.85	22.65	P2	1.29	22.76	P2	1.25	23.48
23	P2	1.15	27.00	P2	0.96	23.61	P2	1.11	23.86	P2	1.21	24.69
24	P2	1.10	28.10	P2	1.81	25.42	P2	0.92	24.78	P2	1.29	25.98
25	P2	1.41	29.51	P2	1.47	26.90	P2	1.60	26.38	P2	1.17	27.15
26	P2	1.62	31.14	P2	1.70	28.60	P2	1.66	28.04	P2	1.87	29.02
27	P2	2.46	33.59	P2	1.42	30.01	P2	2.09	30.14	P2	2.07	31.09
28	P2	3.09	36.68	P2	2.32	32.33	P2	2.03	32.16	P2	2.34	33.44
29	P2	2.62	39.30	P2	3.11	35.45	P2	2.40	34.56	P2	2.54	35.98
30	P3	2.46	41.76	P3	2.89	38.34	P3	2.95	37.52	P3	2.73	38.71
31	P3	2.72	44.48	P3	3.17	41.51	P3	4.00	41.51	P3	3.09	41.80
32	P3	3.04	47.51	P3	2.27	43.77	P3	2.52	44.03	P3	3.63	45.43
33	P3	3.72	51.23	P3	3.11	46.89	P3	3.26	47.29	P3	3.67	49.10

Raw Score	Grade 9 PL Score	Grade 9 % of Students	Grade 9 Cumulative % of Students	Grade 10 PL Score	Grade 10 % of Students	Grade 10 Cumulative % of Students	Grade 11 PL Score	Grade 11 % of Students	Grade 11 Cumulative % of Students	Grade 12 PL Score	Grade 12 % of Students	Grade 12 Cumulative % of Students
34	P3	4.60	55.83	P3	4.70	51.59	P3	4.67	51.97	P3	4.88	53.98
35	P3	4.97	60.81	P3	4.25	55.83	P3	4.67	56.64	P3	3.91	57.89
36	P3	4.76	65.57	P3	5.61	61.44	P3	6.40	63.04	P3	5.35	63.24
37	P4	6.75	72.32	P4	6.91	68.35	P4	6.64	69.68	P4	6.91	70.16
38	P4	6.86	79.17	P4	7.42	75.76	P4	7.38	77.06	P4	8.32	78.48
39	P5	9.73	88.91	P5	10.14	85.90	P5	10.27	87.33	P5	9.18	87.66
40	P5	11.09	100.0	P5	14.10	100.0	P5	12.67	100.0	P5	12.34	100.0

7.7.2. Reading

Table 7.7.2.a.

Raw Score to Scale Score Conversion: K–2

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	900	24	900	924
1	907	15	900	922
2	913	11	902	924
3	917	9	908	926
4	920	8	912	928
5	922	7	915	929
6	924	7	917	931
7	925	6	919	931
8	927	6	921	933
9	929	6	923	935
10	930	6	924	936
11	931	5	926	936
12	932	5	927	937
13	933	5	928	938
14	934	5	929	939
15	935	5	930	940
16	936	5	931	941
17	937	4	933	941
18	938	4	934	942
19	939	4	935	943
20	939	4	935	943
21	940	4	936	944
22	941	4	937	945

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
23	942	4	938	946
24	942	4	938	946
25	943	4	939	947
26	944	4	940	948
27	945	4	941	949
28	946	4	942	950
29	946	4	942	950
30	947	5	942	952
31	948	5	943	953
32	949	5	944	954
33	950	5	945	955
34	952	6	946	958
35	953	6	947	959
36	955	7	948	962
37	957	8	949	965
38	960	10	950	970
39	963	13	950	976
40	974	26	948	980

Table 7.7.2.b.

Raw Score to Proficiency Level Conversion: K–2

Raw Score	Grade K PL Score	Grade K % of Students	Grade K Cumulative % of Students	Grade 1 PL Score	Grade 1 % of Students	Grade 1 Cumulative % of Students	Grade 2 PL Score	Grade 2 % of Students	Grade 2 Cumulative % of Students
0	P1	26.54	26.54	P1	16.22	16.22	P1	11.91	11.91
1	P1	3.61	30.16	P1	2.76	18.98	P1	2.22	14.12
2	P1	3.16	33.32	P1	2.60	21.58	P1	1.72	15.84
3	P1	6.17	39.50	P1	4.72	26.30	P1	4.06	19.90
4	P1	3.61	43.11	P1	2.78	29.08	P1	1.63	21.53
5	P1	2.48	45.59	P1	2.17	31.26	P1	1.49	23.02
6	P1	2.90	48.49	P1	1.94	33.19	P1	1.58	24.60
7	P1	4.18	52.67	P1	3.58	36.77	P1	2.77	27.37
8	P1	2.11	54.78	P1	1.80	38.57	P1	1.52	28.89
9	P1	1.96	56.74	P1	1.43	40.01	P1	1.43	30.32
10	P1	1.51	58.25	P1	1.25	41.25	P1	1.20	31.51
11	P1	1.54	59.79	P1	1.94	43.19	P1	1.75	33.27
12	P1	1.02	60.81	P1	1.09	44.27	P1	0.79	34.05
13	P1	0.79	61.60	P1	0.82	45.10	P1	0.82	34.87

Raw Score	Grade K PL Score	Grade K % of Students	Grade K Cumulative % of Students	Grade 1 PL Score	Grade 1 % of Students	Grade 1 Cumulative % of Students	Grade 2 PL Score	Grade 2 % of Students	Grade 2 Cumulative % of Students
14	P1	0.94	62.54	P1	0.66	45.76	P1	0.64	35.51
15	P1	0.83	63.37	P1	0.58	46.34	P1	0.79	36.30
16	P1	0.83	64.19	P1	0.58	46.92	P1	0.53	36.83
17	P1	0.68	64.87	P1	0.80	47.72	P1	0.70	37.53
18	P1	0.83	65.70	P1	0.82	48.54	P1	0.73	38.26
19	P1	0.75	66.45	P1	0.85	49.39	P1	0.76	39.01
20	P1	0.72	67.17	P1	1.17	50.56	P1	0.76	39.77
21	P1	1.20	68.37	P1	1.03	51.59	P1	0.90	40.68
22	P1	0.87	69.24	P1	1.46	53.05	P1	1.25	41.93
23	P1	1.17	70.41	P1	1.25	54.29	P1	1.14	43.07
24	P1	1.24	71.65	P1	1.56	55.86	P1	1.63	44.70
25	P2	1.77	73.42	P2	1.75	57.61	P2	1.81	46.51
26	P2	1.47	74.89	P2	1.75	59.36	P2	1.40	47.91
27	P2	1.51	76.39	P2	1.86	61.21	P2	2.01	49.93
28	P2	1.77	78.16	P2	2.20	63.41	P2	2.25	52.17
29	P2	1.77	79.93	P2	1.59	65.01	P2	2.13	54.30
30	P2	1.73	81.66	P2	2.41	67.42	P2	2.42	56.73
31	P2	2.26	83.92	P2	3.08	70.49	P2	2.66	59.38
32	P2	1.66	85.58	P2	2.86	73.36	P2	3.06	62.45
33	P3	2.03	87.61	P3	2.89	76.25	P3	3.44	65.89
34	P3	2.03	89.65	P3	3.37	79.61	P3	3.91	69.80
35	P3	1.66	91.30	P3	3.15	82.77	P3	3.47	73.27
36	P3	2.11	93.41	P3	3.26	86.03	P3	3.97	77.24
37	P4	1.84	95.26	P4	3.58	89.61	P4	5.92	83.16
38	P4	1.58	96.84	P4	3.29	92.90	P4	4.29	87.45
39	P5	1.69	98.53	P5	3.26	96.16	P5	5.60	93.06
40	P5	1.47	100.0	P5	3.84	100.0	P5	6.94	100.0

Table 7.7.2.c.

Raw Score to Scale Score Conversion: 3–5

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	902	28	900	930
1	913	15	900	928
2	919	11	908	930
3	922	9	913	931
4	925	8	917	933

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
5	927	7	920	934
6	929	6	923	935
7	930	6	924	936
8	932	6	926	938
9	933	5	928	938
10	934	5	929	939
11	935	5	930	940
12	936	5	931	941
13	937	4	933	941
14	938	4	934	942
15	938	4	934	942
16	939	4	935	943
17	940	4	936	944
18	941	4	937	945
19	941	4	937	945
20	942	4	938	946
21	943	4	939	947
22	943	4	939	947
23	944	4	940	948
24	945	4	941	949
25	945	4	941	949
26	946	4	942	950
27	947	4	943	951
28	948	4	944	952
29	949	4	945	953
30	949	5	944	954
31	950	5	945	955
32	951	5	946	956
33	952	5	947	957
34	954	6	948	960
35	955	6	949	961
36	957	7	950	964
37	959	8	951	967
38	962	10	952	972
39	965	13	952	978
40	976	26	950	980

Table 7.7.2.d.**Raw Score to Proficiency Level Conversion: 3–5**

Raw Score	Grade 3 PL Score	Grade 3 % of Students	Grade 3 Cumulative % of Students	Grade 4 PL Score	Grade 4 % of Students	Grade 4 Cumulative % of Students	Grade 5 PL Score	Grade 5 % of Students	Grade 5 Cumulative % of Students
0	P1	11.57	11.57	P1	9.89	9.89	P1	8.22	8.22
1	P1	1.80	13.37	P1	2.13	12.01	P1	1.63	9.85
2	P1	2.23	15.61	P1	1.89	13.90	P1	1.48	11.33
3	P1	5.90	21.51	P1	4.72	18.63	P1	4.32	15.65
4	P1	0.87	22.37	P1	0.81	19.44	P1	0.87	16.52
5	P1	1.83	24.21	P1	1.08	20.52	P1	1.21	17.73
6	P1	1.73	25.94	P1	1.32	21.84	P1	1.67	19.40
7	P1	2.43	28.38	P1	2.16	24.00	P1	1.89	21.30
8	P1	1.30	29.68	P1	0.94	24.94	P1	0.83	22.13
9	P1	1.13	30.81	P1	1.08	26.02	P1	1.14	23.27
10	P1	1.67	32.48	P1	1.38	27.40	P1	1.36	24.63
11	P1	1.23	33.71	P1	1.32	28.72	P1	1.21	25.84
12	P1	1.30	35.01	P1	1.15	29.87	P1	1.02	26.87
13	P1	1.33	36.35	P1	1.11	30.98	P1	1.10	27.97
14	P1	1.20	37.55	P1	1.25	32.23	P1	1.10	29.06
15	P1	1.40	38.95	P1	0.91	33.14	P1	0.99	30.05
16	P1	1.63	40.58	P1	0.88	34.02	P1	1.02	31.07
17	P1	1.03	41.61	P1	1.11	35.13	P1	1.21	32.28
18	P1	1.50	43.11	P1	1.32	36.45	P1	1.06	33.35
19	P1	1.87	44.98	P1	1.42	37.87	P1	1.17	34.52
20	P1	1.10	46.08	P1	1.82	39.69	P1	1.14	35.66
21	P2	1.83	47.92	P2	1.45	41.14	P2	2.08	37.74
22	P2	1.40	49.32	P2	1.86	43.00	P2	1.71	39.45
23	P2	2.17	51.48	P2	1.79	44.79	P2	1.86	41.30
24	P2	2.17	53.65	P2	2.70	47.49	P2	1.93	43.24
25	P2	2.27	55.92	P2	1.99	49.48	P2	1.93	45.17
26	P2	3.03	58.95	P2	2.23	51.70	P2	2.88	48.05
27	P2	2.97	61.92	P2	2.83	54.54	P2	2.61	50.66
28	P2	3.83	65.76	P2	3.07	57.61	P2	2.73	53.39
29	P2	3.43	69.19	P2	3.17	60.78	P2	3.41	56.80
30	P2	2.97	72.16	P2	3.58	64.36	P2	3.30	60.10
31	P3	3.17	75.33	P3	4.08	68.44	P3	3.56	63.66
32	P3	2.27	77.59	P3	2.90	71.35	P3	3.22	66.88
33	P3	2.37	79.96	P3	3.24	74.59	P3	2.84	69.72

Raw Score	Grade 3 PL Score	Grade 3 % of Students	Grade 3 Cumulative % of Students	Grade 4 PL Score	Grade 4 % of Students	Grade 4 Cumulative % of Students	Grade 5 PL Score	Grade 5 % of Students	Grade 5 Cumulative % of Students
34	P3	3.87	83.83	P3	3.61	78.20	P3	2.96	72.68
35	P3	2.60	86.43	P3	2.63	80.83	P3	2.73	75.41
36	P4	2.63	89.06	P4	2.97	83.80	P4	3.83	79.23
37	P4	2.63	91.70	P4	3.41	87.21	P4	4.43	83.67
38	P4	2.77	94.46	P4	3.00	90.21	P4	3.68	87.34
39	P5	2.23	96.70	P5	3.78	93.99	P5	4.96	92.31
40	P5	3.30	100.0	P5	6.01	100.0	P5	7.69	100.0

Table 7.7.2.e.

Raw Score to Scale Score Conversion: 6–8

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	904	27	900	931
1	913	15	900	928
2	919	11	908	930
3	923	9	914	932
4	926	8	918	934
5	928	7	921	935
6	930	6	924	936
7	931	6	925	937
8	932	5	927	937
9	933	5	928	938
10	935	5	930	940
11	936	5	931	941
12	937	5	932	942
13	937	5	932	942
14	938	4	934	942
15	939	4	935	943
16	940	4	936	944
17	941	4	937	945
18	942	4	938	946
19	942	4	938	946
20	943	4	939	947
21	944	4	940	948
22	945	4	941	949
23	945	4	941	949
24	946	4	942	950

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
25	947	4	943	951
26	948	4	944	952
27	949	4	945	953
28	949	4	945	953
29	950	4	946	954
30	951	5	946	956
31	952	5	947	957
32	953	5	948	958
33	954	5	949	959
34	955	6	949	961
35	957	6	951	963
36	959	7	952	966
37	961	8	953	969
38	964	10	954	974
39	967	13	954	980
40	978	26	952	980

Table 7.7.2.f.

Raw Score to Proficiency Level Conversion: 6–8

Raw Score	Grade 6 PL Score	Grade 6 % of Students	Grade 6 Cumulative % of Students	Grade 7 PL Score	Grade 7 % of Students	Grade 7 Cumulative % of Students	Grade 8 PL Score	Grade 8 % of Students	Grade 8 Cumulative % of Students
0	P1	7.46	7.46	P1	7.16	7.16	P1	6.85	6.85
1	P1	1.13	8.59	P1	1.18	8.34	P1	1.40	8.25
2	P1	1.46	10.05	P1	1.23	9.57	P1	0.95	9.20
3	P1	4.55	14.60	P1	3.82	13.38	P1	3.70	12.89
4	P1	0.71	15.30	P1	0.42	13.81	P1	0.40	13.29
5	P1	0.88	16.18	P1	0.75	14.56	P1	0.80	14.09
6	P1	1.63	17.81	P1	1.23	15.79	P1	1.45	15.54
7	P1	1.96	19.77	P1	1.32	17.11	P1	1.80	17.34
8	P1	1.58	21.35	P1	1.13	18.24	P1	0.90	18.24
9	P1	1.00	22.35	P1	1.13	19.37	P1	1.00	19.24
10	P1	0.96	23.31	P1	0.85	20.22	P1	0.60	19.84
11	P1	0.88	24.19	P1	0.94	21.16	P1	1.10	20.94
12	P1	0.88	25.06	P1	0.80	21.96	P1	0.95	21.89
13	P1	0.92	25.98	P1	0.61	22.57	P1	0.95	22.84
14	P1	0.83	26.81	P1	0.38	22.95	P1	0.75	23.59
15	P1	0.71	27.52	P1	0.66	23.61	P1	0.90	24.49

Raw Score	Grade 6 PL Score	Grade 6 % of Students	Grade 6 Cumulative % of Students	Grade 7 PL Score	Grade 7 % of Students	Grade 7 Cumulative % of Students	Grade 8 PL Score	Grade 8 % of Students	Grade 8 Cumulative % of Students
16	P1	1.25	28.77	P1	0.85	24.46	P1	0.75	25.24
17	P1	0.96	29.73	P1	0.85	25.31	P1	0.50	25.74
18	P1	0.79	30.53	P1	0.99	26.30	P1	0.85	26.59
19	P1	0.83	31.36	P1	1.27	27.57	P1	0.90	27.49
20	P1	1.08	32.44	P1	1.41	28.98	P1	1.20	28.69
21	P2	1.42	33.86	P2	1.79	30.77	P2	1.60	30.28
22	P2	2.38	36.24	P2	1.74	32.52	P2	2.00	32.28
23	P2	1.71	37.95	P2	1.27	33.79	P2	1.75	34.03
24	P2	1.83	39.78	P2	1.74	35.53	P2	1.50	35.53
25	P2	2.46	42.24	P2	3.06	38.60	P2	2.20	37.73
26	P2	2.54	44.79	P2	2.26	40.86	P2	1.65	39.38
27	P2	2.34	47.12	P2	2.21	43.07	P2	1.75	41.13
28	P2	2.25	49.37	P2	3.06	46.14	P2	2.65	43.78
29	P3	2.92	52.29	P3	2.73	48.87	P3	3.15	46.93
30	P3	3.63	55.92	P3	3.11	51.98	P3	2.85	49.78
31	P3	4.25	60.18	P3	3.63	55.61	P3	3.30	53.07
32	P3	4.00	64.18	P3	3.53	59.14	P3	4.15	57.22
33	P3	3.67	67.85	P3	4.67	63.81	P3	3.10	60.32
34	P3	4.13	71.98	P3	4.05	67.86	P3	3.65	63.97
35	P4	3.25	75.23	P4	4.52	72.38	P4	4.30	68.27
36	P4	4.55	79.77	P4	4.76	77.14	P4	4.80	73.06
37	P4	5.42	85.20	P4	5.51	82.66	P4	5.75	78.81
38	P4	4.80	89.99	P4	4.90	87.56	P4	5.35	84.16
39	P5	4.92	94.91	P5	5.75	93.31	P5	5.50	89.66
40	P5	5.09	100.0	P5	6.69	100.0	P5	10.34	100.0

Table 7.7.2.g.

Raw Score to Scale Score Conversion: 9–12

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	906	26	900	932
1	915	15	900	930
2	920	11	909	931
3	924	9	915	933
4	927	8	919	935
5	929	7	922	936
6	931	6	925	937

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
7	932	6	926	938
8	933	5	928	938
9	935	5	930	940
10	936	5	931	941
11	937	5	932	942
12	938	5	933	943
13	939	5	934	944
14	939	4	935	943
15	940	4	936	944
16	941	4	937	945
17	942	4	938	946
18	943	4	939	947
19	944	4	940	948
20	944	4	940	948
21	945	4	941	949
22	946	4	942	950
23	947	4	943	951
24	947	4	943	951
25	948	4	944	952
26	949	4	945	953
27	950	4	946	954
28	950	4	946	954
29	951	4	947	955
30	952	5	947	957
31	953	5	948	958
32	954	5	949	959
33	955	5	950	960
34	956	6	950	962
35	958	6	952	964
36	959	7	952	966
37	962	8	954	970
38	965	10	955	975
39	968	13	955	980
40	980	28	952	980

Table 7.7.2.h.**Raw Score to Proficiency Level Conversion: 9–12**

Raw Score	Grade 9 PL Score	Grade 9 % of Students	Grade 9 Cumulative % of Students	Grade 10 PL Score	Grade 10 % of Students	Grade 10 Cumulative % of Students	Grade 11 PL Score	Grade 11 % of Students	Grade 11 Cumulative % of Students	Grade 12 PL Score	Grade 12 % of Students	Grade 12 Cumulative % of Students
0	P1	8.85	8.85	P1	6.74	6.74	P1	7.50	7.50	P1	7.19	7.19
1	P1	0.79	9.63	P1	0.74	7.47	P1	0.92	8.43	P1	0.35	7.54
2	P1	0.73	10.37	P1	0.51	7.98	P1	0.98	9.41	P1	1.06	8.60
3	P1	3.87	14.24	P1	4.02	12.00	P1	4.06	13.47	P1	4.22	12.82
4	P1	0.31	14.55	P1	0.85	12.85	P1	0.43	13.90	P1	0.39	13.21
5	P1	1.36	15.92	P1	1.30	14.16	P1	1.41	15.31	P1	1.37	14.58
6	P1	1.99	17.91	P1	1.87	16.02	P1	1.54	16.85	P1	1.49	16.07
7	P1	1.83	19.74	P1	2.15	18.18	P1	2.21	19.07	P1	2.74	18.80
8	P1	0.73	20.47	P1	0.68	18.86	P1	0.74	19.80	P1	1.06	19.86
9	P1	0.99	21.47	P1	1.36	20.22	P1	0.98	20.79	P1	0.94	20.80
10	P1	0.73	22.20	P1	0.45	20.67	P1	0.86	21.65	P1	0.86	21.66
11	P1	1.62	23.82	P1	0.96	21.63	P1	1.17	22.82	P1	1.21	22.87
12	P1	0.73	24.55	P1	1.02	22.65	P1	0.98	23.80	P1	1.21	24.08
13	P1	0.84	25.39	P1	1.59	24.24	P1	0.55	24.35	P1	1.02	25.10
14	P1	0.73	26.13	P1	0.74	24.97	P1	1.05	25.40	P1	0.90	26.00
15	P1	1.10	27.23	P1	1.19	26.16	P1	0.74	26.14	P1	1.21	27.21
16	P1	1.05	28.27	P1	1.08	27.24	P1	1.05	27.18	P1	1.09	28.30
17	P1	1.57	29.84	P1	1.13	28.37	P1	0.86	28.04	P1	0.98	29.28
18	P1	0.84	30.68	P1	0.74	29.11	P1	1.29	29.34	P1	1.06	30.34
19	P2	1.62	32.30	P2	1.42	30.52	P2	1.23	30.57	P2	1.37	31.70
20	P2	1.31	33.61	P2	1.13	31.65	P2	0.92	31.49	P2	1.64	33.35
21	P2	1.47	35.08	P2	1.08	32.73	P2	1.60	33.09	P2	1.45	34.79
22	P2	1.52	36.60	P2	1.36	34.09	P2	1.05	34.13	P2	2.42	37.22
23	P2	2.46	39.06	P2	1.42	35.50	P2	1.48	35.61	P2	2.03	39.25
24	P2	1.94	40.99	P2	2.43	37.94	P2	2.40	38.01	P2	2.03	41.28
25	P2	2.72	43.72	P2	2.27	40.20	P2	2.52	40.53	P2	2.62	43.90
26	P2	2.46	46.18	P2	2.60	42.81	P2	2.52	43.05	P2	2.46	46.36
27	P3	2.67	48.85	P3	2.27	45.07	P3	2.09	45.14	P3	2.23	48.59
28	P3	2.36	51.20	P3	2.49	47.57	P3	2.40	47.54	P3	2.81	51.41
29	P3	2.25	53.46	P3	2.43	50.00	P3	2.21	49.75	P3	2.03	53.44
30	P3	2.72	56.18	P3	2.38	52.38	P3	2.89	52.64	P3	2.19	55.63
31	P3	2.98	59.16	P3	2.89	55.27	P3	2.64	55.29	P3	2.31	57.94
32	P3	3.04	62.20	P3	2.66	57.93	P3	2.71	58.00	P3	2.19	60.13
33	P3	3.04	65.24	P3	2.83	60.76	P3	1.91	59.90	P3	1.64	61.77
34	P3	3.40	68.64	P3	3.17	63.93	P3	3.20	63.10	P3	3.91	65.68

Raw Score	Grade 9 PL Score	Grade 9 % of Students	Grade 9 Cumulative % of Students	Grade 10 PL Score	Grade 10 % of Students	Grade 10 Cumulative % of Students	Grade 11 PL Score	Grade 11 % of Students	Grade 11 Cumulative % of Students	Grade 12 PL Score	Grade 12 % of Students	Grade 12 Cumulative % of Students
35	P4	2.72	71.36	P4	2.94	66.87	P4	3.01	66.11	P4	2.74	68.41
36	P4	2.67	74.03	P4	3.11	69.99	P4	4.18	70.30	P4	3.21	71.62
37	P4	4.71	78.74	P4	5.72	75.71	P4	5.10	75.40	P4	4.22	75.84
38	P4	4.92	83.66	P4	4.76	80.46	P4	5.90	81.30	P4	4.81	80.65
39	P5	6.75	90.42	P5	6.40	86.86	P5	7.13	88.44	P5	6.14	86.79
40	P5	9.58	100.0	P5	13.14	100.0	P5	11.56	100.0	P5	13.21	100.0

7.7.3. Speaking

Table 7.7.3.a.

Raw Score to Scale Score Conversion: K-2

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	900	36	900	936
1	917	15	902	932
2	923	11	912	934
3	926	9	917	935
4	929	8	921	937
5	931	7	924	938
6	933	6	927	939
7	934	6	928	940
8	936	6	930	942
9	937	5	932	942
10	938	5	933	943
11	939	5	934	944
12	940	5	935	945
13	941	5	936	946
14	942	5	937	947
15	943	5	938	948
16	944	5	939	949
17	945	5	940	950
18	946	5	941	951
19	947	5	942	952
20	947	5	942	952
21	948	5	943	953
22	949	5	944	954
23	950	5	945	955
24	951	5	946	956

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
25	952	5	947	957
26	953	5	948	958
27	955	6	949	961
28	956	6	950	962
29	958	7	951	965
30	960	9	951	969
31	962	11	951	973
32	974	28	946	980

Table 7.7.3.b.

Raw Score to Proficiency Level Conversion: K–2

Raw Score	Grade K PL Score	Grade K % of Students	Grade K Cumulative % of Students	Grade 1 PL Score	Grade 1 % of Students	Grade 1 Cumulative % of Students	Grade 2 PL Score	Grade 2 % of Students	Grade 2 Cumulative % of Students
0	P1	43.83	43.83	P1	33.16	33.16	P1	25.85	25.85
1	P1	3.55	47.38	P1	2.95	36.10	P1	2.57	28.42
2	P1	5.24	52.62	P1	4.11	40.22	P1	3.42	31.83
3	P1	4.22	56.85	P1	3.58	43.80	P1	3.53	35.37
4	P1	3.73	60.58	P1	3.42	47.23	P1	2.57	37.94
5	P1	2.26	62.84	P1	2.26	49.48	P1	2.60	40.54
6	P1	3.62	66.47	P1	3.56	53.04	P1	3.01	43.55
7	P1	3.28	69.75	P1	3.29	56.33	P1	3.36	46.90
8	P1	3.06	72.80	P1	4.22	60.55	P1	4.09	50.99
9	P1	3.81	76.61	P1	3.88	64.43	P1	3.30	54.29
10	P1	2.34	78.95	P1	2.60	67.03	P1	2.89	57.18
11	P1	2.34	81.29	P1	2.55	69.58	P1	2.60	59.78
12	P1	1.92	83.21	P1	2.68	72.26	P1	2.60	62.38
13	P2	2.15	85.36	P2	2.42	74.67	P2	2.25	64.63
14	P2	1.70	87.06	P2	1.99	76.67	P2	2.42	67.06
15	P2	1.09	88.16	P2	1.59	78.26	P2	2.22	69.28
16	P2	1.81	89.97	P2	1.49	79.75	P2	2.04	71.32
17	P2	0.72	90.68	P2	1.67	81.42	P2	1.93	73.25
18	P2	1.24	91.93	P2	1.19	82.61	P2	1.61	74.85
19	P2	0.91	92.83	P2	1.27	83.89	P2	1.61	76.46
20	P2	0.75	93.59	P2	1.30	85.19	P2	1.61	78.07
21	P3	0.68	94.27	P3	1.54	86.73	P3	1.72	79.79
22	P3	0.94	95.21	P3	1.17	87.89	P3	1.31	81.10
23	P3	0.41	95.62	P3	0.90	88.80	P3	1.43	82.54

Raw Score	Grade K PL Score	Grade K % of Students	Grade K Cumulative % of Students	Grade 1 PL Score	Grade 1 % of Students	Grade 1 Cumulative % of Students	Grade 2 PL Score	Grade 2 % of Students	Grade 2 Cumulative % of Students
24	P3	0.83	96.45	P3	0.96	89.75	P3	1.43	83.97
25	P3	0.41	96.87	P3	0.93	90.68	P3	1.43	85.40
26	P3	0.34	97.21	P3	0.98	91.66	P3	1.49	86.89
27	P3	0.41	97.62	P3	0.85	92.51	P3	1.69	88.58
28	P3	0.53	98.15	P3	1.33	93.84	P3	1.64	90.22
29	P4	0.38	98.53	P4	1.30	95.14	P4	1.64	91.85
30	P4	0.41	98.94	P4	1.11	96.26	P4	1.75	93.60
31	P5	0.30	99.25	P5	1.33	97.58	P5	2.37	95.97
32	P5	0.75	100.0	P5	2.42	100.0	P5	4.03	100.0

Table 7.7.3.c.

Raw Score to Scale Score Conversion: 3–5

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	902	38	900	940
1	919	15	904	934
2	925	11	914	936
3	929	9	920	938
4	931	8	923	939
5	933	7	926	940
6	935	6	929	941
7	937	6	931	943
8	938	6	932	944
9	939	5	934	944
10	940	5	935	945
11	941	5	936	946
12	942	5	937	947
13	943	5	938	948
14	944	5	939	949
15	945	4	941	949
16	946	4	942	950
17	947	4	943	951
18	947	4	943	951
19	948	4	944	952
20	949	4	945	953
21	950	5	945	955
22	951	5	946	956

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
23	952	5	947	957
24	953	5	948	958
25	954	5	949	959
26	955	5	950	960
27	956	6	950	962
28	957	6	951	963
29	959	7	952	966
30	962	9	953	971
31	965	12	953	977
32	976	29	947	980

Table 7.7.3.d.

Raw Score to Proficiency Level Conversion: 3–5

Raw Score	Grade 3 PL Score	Grade 3 % of Students	Grade 3 Cumulative % of Students	Grade 4 PL Score	Grade 4 % of Students	Grade 4 Cumulative % of Students	Grade 5 PL Score	Grade 5 % of Students	Grade 5 Cumulative % of Students
0	P1	23.43	23.43	P1	21.29	21.29	P1	20.45	20.45
1	P1	1.90	25.32	P1	1.86	23.15	P1	1.48	21.93
2	P1	2.57	27.89	P1	2.10	25.25	P1	2.01	23.94
3	P1	2.07	29.96	P1	2.94	28.19	P1	1.89	25.83
4	P1	2.47	32.42	P1	1.52	29.71	P1	1.52	27.35
5	P1	1.37	33.79	P1	1.45	31.16	P1	1.17	28.52
6	P1	2.63	36.42	P1	1.66	32.82	P1	1.70	30.23
7	P1	1.23	37.65	P1	1.99	34.81	P1	1.86	32.08
8	P1	3.67	41.32	P1	3.14	37.95	P1	3.22	35.30
9	P1	2.27	43.59	P1	1.89	39.84	P1	1.93	37.23
10	P1	3.53	47.12	P1	3.14	42.99	P1	2.58	39.81
11	P1	4.63	51.75	P1	3.55	46.54	P1	4.17	43.98
12	P1	3.30	55.05	P1	3.45	49.98	P1	3.52	47.50
13	P1	2.97	58.01	P1	3.01	52.99	P1	3.11	50.61
14	P1	2.40	60.41	P1	2.47	55.46	P1	2.73	53.33
15	P1	2.67	63.08	P1	2.20	57.65	P1	2.50	55.83
16	P2	2.73	65.81	P2	2.70	60.36	P2	2.05	57.88
17	P2	2.07	67.88	P2	2.37	62.72	P2	1.86	59.73
18	P2	1.97	69.84	P2	1.93	64.65	P2	1.55	61.29
19	P2	1.97	71.81	P2	2.03	66.68	P2	2.05	63.33
20	P2	1.90	73.71	P2	1.93	68.60	P2	1.86	65.19
21	P2	1.73	75.44	P2	1.89	70.50	P2	1.89	67.08

Raw Score	Grade 3 PL Score	Grade 3 % of Students	Grade 3 Cumulative % of Students	Grade 4 PL Score	Grade 4 % of Students	Grade 4 Cumulative % of Students	Grade 5 PL Score	Grade 5 % of Students	Grade 5 Cumulative % of Students
22	P2	2.17	77.61	P2	2.26	72.76	P2	1.67	68.75
23	P2	1.97	79.57	P2	2.03	74.79	P2	1.93	70.68
24	P3	2.00	81.57	P3	1.86	76.65	P3	1.93	72.61
25	P3	1.37	82.94	P3	1.86	78.51	P3	1.89	74.51
26	P3	1.93	84.87	P3	2.10	80.60	P3	1.97	76.48
27	P3	1.67	86.54	P3	2.03	82.63	P3	2.23	78.71
28	P3	2.03	88.57	P3	2.40	85.03	P3	3.18	81.89
29	P4	1.77	90.34	P4	2.13	87.16	P4	2.54	84.43
30	P4	1.80	92.14	P4	2.47	89.62	P4	3.11	87.54
31	P5	2.57	94.70	P5	2.57	92.19	P5	3.90	91.44
32	P5	5.30	100.0	P5	7.81	100.0	P5	8.56	100.0

Table 7.7.3.e

Raw Score to Scale Score Conversion: 6–8

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	904	37	900	941
1	920	15	905	935
2	926	11	915	937
3	930	9	921	939
4	932	8	924	940
5	934	7	927	941
6	936	6	930	942
7	938	6	932	944
8	939	6	933	945
9	940	5	935	945
10	941	5	936	946
11	942	5	937	947
12	943	5	938	948
13	944	5	939	949
14	945	5	940	950
15	946	5	941	951
16	947	4	943	951
17	948	4	944	952
18	949	4	945	953
19	949	4	945	953
20	950	4	946	954

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
21	951	4	947	955
22	952	5	947	957
23	953	5	948	958
24	954	5	949	959
25	955	5	950	960
26	956	5	951	961
27	957	6	951	963
28	959	6	953	965
29	960	7	953	967
30	963	9	954	972
31	966	12	954	978
32	978	31	947	980

Table 7.7.3.f

Raw Score to Proficiency Level Conversion: 6–8

Raw Score	Grade 6 PL Score	Grade 6 % of Students	Grade 6 Cumulative % of Students	Grade 7 PL Score	Grade 7 % of Students	Grade 7 Cumulative % of Students	Grade 8 PL Score	Grade 8 % of Students	Grade 8 Cumulative % of Students
0	P1	18.42	18.42	P1	17.08	17.08	P1	17.97	17.97
1	P1	1.63	20.05	P1	0.99	18.07	P1	1.50	19.47
2	P1	2.59	22.64	P1	2.31	20.39	P1	2.20	21.67
3	P1	2.13	24.77	P1	2.03	22.42	P1	2.15	23.82
4	P1	2.21	26.98	P1	1.46	23.88	P1	1.70	25.53
5	P1	1.04	28.03	P1	1.32	25.20	P1	0.95	26.48
6	P1	1.84	29.87	P1	1.65	26.85	P1	2.05	28.53
7	P1	1.50	31.37	P1	1.70	28.55	P1	1.25	29.78
8	P1	1.55	32.92	P1	1.65	30.20	P1	1.60	31.38
9	P1	1.80	34.71	P1	1.56	31.76	P1	1.10	32.48
10	P1	2.21	36.93	P1	1.56	33.32	P1	1.40	33.88
11	P1	1.50	38.43	P1	1.56	34.87	P1	1.30	35.19
12	P1	2.84	41.27	P1	2.83	37.71	P1	3.15	38.34
13	P1	2.46	43.73	P1	2.97	40.68	P1	2.45	40.79
14	P1	4.51	48.25	P1	4.01	44.69	P1	4.20	44.99
15	P2	3.22	51.46	P2	3.11	47.81	P2	3.35	48.35
16	P2	3.17	54.64	P2	4.81	52.62	P2	2.90	51.25
17	P2	2.42	57.06	P2	2.41	55.03	P2	2.45	53.70
18	P2	2.42	59.48	P2	2.64	57.67	P2	1.80	55.51
19	P2	1.71	61.19	P2	2.60	60.26	P2	1.90	57.41

Raw Score	Grade 6 PL Score	Grade 6 % of Students	Grade 6 Cumulative % of Students	Grade 7 PL Score	Grade 7 % of Students	Grade 7 Cumulative % of Students	Grade 8 PL Score	Grade 8 % of Students	Grade 8 Cumulative % of Students
20	P2	2.21	63.41	P2	2.36	62.62	P2	2.00	59.41
21	P2	2.34	65.75	P2	2.55	65.17	P2	2.20	61.61
22	P2	2.17	67.92	P2	1.84	67.01	P2	1.95	63.56
23	P2	1.88	69.80	P2	1.70	68.71	P2	2.05	65.62
24	P3	2.26	72.06	P3	2.27	70.98	P3	2.40	68.02
25	P3	2.80	74.85	P3	2.69	73.67	P3	3.25	71.27
26	P3	1.80	76.65	P3	3.16	76.83	P3	2.30	73.57
27	P3	2.67	79.32	P3	2.55	79.38	P3	2.45	76.03
28	P3	2.55	81.87	P3	3.82	83.20	P3	4.15	80.18
29	P3	3.34	85.21	P3	2.50	85.70	P3	3.15	83.33
30	P4	3.97	89.18	P4	3.54	89.24	P4	3.50	86.84
31	P5	3.88	93.07	P5	4.29	93.53	P5	4.75	91.59
32	P5	6.93	100.0	P5	6.47	100.0	P5	8.41	100.0

Table 7.7.3.g.

Raw Score to Scale Score Conversion: 9–12

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	906	33	900	939
1	921	15	906	936
2	927	11	916	938
3	930	9	921	939
4	933	8	925	941
5	935	7	928	942
6	937	6	931	943
7	938	6	932	944
8	940	6	934	946
9	941	5	936	946
10	942	5	937	947
11	943	5	938	948
12	944	5	939	949
13	945	5	940	950
14	946	4	942	950
15	947	4	943	951
16	947	4	943	951
17	948	4	944	952
18	949	4	945	953

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
19	950	4	946	954
20	950	4	946	954
21	951	4	947	955
22	952	5	947	957
23	953	5	948	958
24	954	5	949	959
25	955	5	950	960
26	956	5	951	961
27	957	6	951	963
28	959	6	953	965
29	961	7	954	968
30	963	9	954	972
31	965	11	954	976
32	980	35	945	980

Table 7.7.3.h.

Raw Score to Proficiency Level Conversion: 9–12

Raw Score	Grade 9 PL Score	Grade 9 % of Students	Grade 9 Cumulative % of Students	Grade 10 PL Score	Grade 10 % of Students	Grade 10 Cumulative % of Students	Grade 11 PL Score	Grade 11 % of Students	Grade 11 Cumulative % of Students	Grade 12 PL Score	Grade 12 % of Students	Grade 12 Cumulative % of Students
0	P1	21.02	21.02	P1	18.37	18.37	P1	19.19	19.19	P1	18.28	18.28
1	P1	2.36	23.38	P1	1.25	19.61	P1	1.54	20.73	P1	1.64	19.92
2	P1	3.09	26.47	P1	3.12	22.73	P1	3.44	24.17	P1	2.97	22.90
3	P1	3.20	29.66	P1	2.89	25.62	P1	2.77	26.94	P1	3.01	25.91
4	P1	1.94	31.60	P1	1.87	27.49	P1	1.78	28.72	P1	1.84	27.75
5	P1	2.04	33.65	P1	1.53	29.02	P1	1.05	29.77	P1	1.76	29.51
6	P1	1.89	35.53	P1	1.81	30.84	P1	1.54	31.30	P1	1.72	31.23
7	P1	1.73	37.26	P1	1.81	32.65	P1	1.91	33.21	P1	1.84	33.07
8	P1	1.78	39.05	P1	1.87	34.52	P1	1.78	34.99	P1	2.70	35.77
9	P1	3.83	42.87	P1	3.46	37.98	P1	2.83	37.82	P1	3.52	39.30
10	P1	1.31	44.18	P1	1.76	39.74	P1	1.66	39.48	P1	2.19	41.49
11	P1	2.36	46.54	P1	2.15	41.89	P1	3.32	42.80	P1	2.43	43.91
12	P1	1.83	48.38	P1	1.64	43.54	P1	1.72	44.53	P1	1.72	45.64
13	P1	1.83	50.21	P1	1.36	44.90	P1	2.21	46.74	P1	2.11	47.75
14	P2	1.57	51.78	P2	2.04	46.94	P2	1.97	48.71	P2	1.72	49.47
15	P2	1.26	53.04	P2	1.70	48.64	P2	1.60	50.31	P2	1.96	51.43
16	P2	1.73	54.77	P2	1.53	50.17	P2	1.72	52.03	P2	1.64	53.07
17	P2	2.15	56.92	P2	1.30	51.47	P2	1.48	53.51	P2	2.43	55.50

Raw Score	Grade 9 PL Score	Grade 9 % of Students	Grade 9 Cumulative % of Students	Grade 10 PL Score	Grade 10 % of Students	Grade 10 Cumulative % of Students	Grade 11 PL Score	Grade 11 % of Students	Grade 11 Cumulative % of Students	Grade 12 PL Score	Grade 12 % of Students	Grade 12 Cumulative % of Students
18	P2	1.26	58.18	P2	1.42	52.89	P2	1.60	55.10	P2	1.88	57.38
19	P2	1.73	59.91	P2	2.15	55.05	P2	1.41	56.52	P2	1.37	58.75
20	P2	1.47	61.37	P2	1.42	56.46	P2	1.23	57.75	P2	2.00	60.74
21	P2	1.57	62.95	P2	2.15	58.62	P2	1.35	59.10	P2	1.53	62.27
22	P2	1.68	64.62	P2	1.93	60.54	P2	1.66	60.76	P2	1.29	63.56
23	P2	1.68	66.30	P2	1.76	62.30	P2	1.60	62.36	P2	1.88	65.44
24	P3	1.83	68.13	P3	1.47	63.78	P3	2.28	64.64	P3	1.80	67.24
25	P3	1.78	69.92	P3	1.70	65.48	P3	1.91	66.54	P3	1.80	69.04
26	P3	2.25	72.17	P3	2.32	67.80	P3	2.03	68.57	P3	2.15	71.19
27	P3	1.68	73.85	P3	2.32	70.12	P3	2.15	70.73	P3	1.76	72.95
28	P3	2.83	76.68	P3	2.55	72.68	P3	2.83	73.55	P3	3.01	75.97
29	P4	2.57	79.25	P4	3.29	75.96	P4	3.08	76.63	P4	2.43	78.40
30	P4	3.04	82.29	P4	3.00	78.97	P4	3.81	80.44	P4	2.90	81.29
31	P4	3.67	85.95	P4	4.14	83.11	P4	4.31	84.75	P4	4.58	85.87
32	P5	14.05	100.0	P5	16.89	100.0	P5	15.25	100.0	P5	14.13	100.0

7.7.4. Writing

Table 7.7.4.a.

Raw Score to Scale Score Conversion: K–2

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	900	24	900	924
1	907	18	900	925
2	916	12	904	928
3	920	10	910	930
4	923	8	915	931
5	926	8	918	934
6	929	8	921	937
7	931	8	923	939
8	934	8	926	942
9	937	8	929	945
10	939	8	931	947
11	941	7	934	948
12	943	7	936	950
13	945	6	939	951
14	946	6	940	952
15	948	5	943	953

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
16	949	5	944	954
17	950	5	945	955
18	951	5	946	956
19	952	5	947	957
20	953	5	948	958
21	954	5	949	959
22	955	5	950	960
23	956	5	951	961
24	957	5	952	962
25	958	5	953	963
26	960	6	954	966
27	961	6	955	967
28	963	7	956	970
29	965	8	957	973
30	968	10	958	978
31	971	12	959	980
32	974	15	959	980

Table 7.7.4.b.

Raw Score to Proficiency Level Conversion: K–2

Raw Score	Grade K PL Score	Grade K % of Students	Grade K Cumulative % of Students	Grade 1 PL Score	Grade 1 % of Students	Grade 1 Cumulative % of Students	Grade 2 PL Score	Grade 2 % of Students	Grade 2 Cumulative % of Students
0	P1	37.30	37.30	P1	26.79	26.79	P1	19.44	19.44
1	P1	7.23	44.54	P1	5.04	31.83	P1	4.70	24.14
2	P1	6.10	50.64	P1	4.75	36.58	P1	4.03	28.17
3	P1	14.24	64.88	P1	11.56	48.14	P1	9.75	37.92
4	P1	4.33	69.22	P1	4.51	52.65	P1	3.44	41.36
5	P1	3.84	73.06	P1	4.32	56.98	P1	4.00	45.36
6	P1	4.30	77.35	P1	4.99	61.96	P1	5.31	50.67
7	P1	5.73	83.08	P1	5.46	67.43	P1	6.19	56.86
8	P1	1.39	84.48	P1	1.70	69.12	P1	1.69	58.55
9	P1	1.17	85.64	P1	1.41	70.53	P1	1.20	59.75
10	P1	0.87	86.51	P1	1.49	72.02	P1	1.08	60.83
11	P2	1.09	87.60	P2	1.17	73.18	P2	1.49	62.32
12	P2	0.79	88.39	P2	0.77	73.95	P2	1.05	63.37
13	P2	0.57	88.96	P2	1.03	74.99	P2	1.17	64.54
14	P2	0.68	89.64	P2	0.98	75.97	P2	0.90	65.44

Raw Score	Grade K PL Score	Grade K % of Students	Grade K Cumulative % of Students	Grade 1 PL Score	Grade 1 % of Students	Grade 1 Cumulative % of Students	Grade 2 PL Score	Grade 2 % of Students	Grade 2 Cumulative % of Students
15	P2	0.75	90.39	P2	1.01	76.98	P2	0.99	66.43
16	P2	0.64	91.03	P2	0.93	77.90	P2	1.28	67.72
17	P2	0.64	91.67	P2	0.90	78.81	P2	1.31	69.03
18	P3	0.49	92.16	P3	1.09	79.89	P3	1.31	70.34
19	P3	0.72	92.88	P3	1.09	80.98	P3	1.26	71.60
20	P3	0.68	93.56	P3	1.67	82.65	P3	1.84	73.44
21	P3	0.41	93.97	P3	1.30	83.95	P3	1.20	74.64
22	P3	0.60	94.57	P3	1.27	85.23	P3	2.10	76.74
23	P3	0.60	95.18	P3	1.67	86.90	P3	2.10	78.84
24	P3	0.64	95.82	P3	1.19	88.09	P3	2.16	81.00
25	P3	0.53	96.35	P3	1.59	89.68	P3	1.75	82.75
26	P4	0.68	97.02	P4	1.64	91.33	P4	1.75	84.50
27	P4	0.45	97.48	P4	1.30	92.63	P4	2.16	86.66
28	P4	0.53	98.00	P4	1.30	93.93	P4	2.13	88.79
29	P4	0.38	98.38	P4	1.41	95.33	P4	2.45	91.24
30	P5	0.53	98.91	P5	1.38	96.71	P5	2.19	93.43
31	P5	0.53	99.43	P5	1.67	98.38	P5	2.39	95.83
32	P5	0.57	100.0	P5	1.62	100.0	P5	4.17	100.0

Table 7.7.4.c.

Raw Score to Scale Score Conversion: 3–5

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	902	17	900	919
1	903	16	900	919
2	911	12	900	923
3	916	10	906	926
4	920	9	911	929
5	922	8	914	930
6	925	8	917	933
7	927	7	920	934
8	929	7	922	936
9	931	7	924	938
10	933	6	927	939
11	935	6	929	941
12	936	6	930	942
13	938	6	932	944

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
14	939	6	933	945
15	940	6	934	946
16	942	6	936	948
17	944	6	938	950
18	945	6	939	951
19	947	6	941	953
20	948	6	942	954
21	950	6	944	956
22	951	6	945	957
23	953	6	947	959
24	955	6	949	961
25	956	6	950	962
26	958	7	951	965
27	960	7	953	967
28	962	8	954	970
29	964	8	956	972
30	968	10	958	978
31	972	13	959	980
32	976	16	960	980

Table 7.7.4.d.

Raw Score to Proficiency Level Conversion: 3–5

Raw Score	Grade 3 PL Score	Grade 3 % of Students	Grade 3 Cumulative % of Students	Grade 4 PL Score	Grade 4 % of Students	Grade 4 Cumulative % of Students	Grade 5 PL Score	Grade 5 % of Students	Grade 5 Cumulative % of Students
0	P1	17.53	17.53	P1	14.97	14.97	P1	14.55	14.55
1	P1	2.77	20.30	P1	2.91	17.88	P1	1.93	16.48
2	P1	2.90	23.21	P1	2.67	20.55	P1	1.74	18.22
3	P1	9.88	33.09	P1	8.99	29.54	P1	9.89	28.11
4	P1	2.60	35.69	P1	2.26	31.80	P1	1.74	29.85
5	P1	2.24	37.93	P1	2.47	34.27	P1	2.01	31.86
6	P1	2.84	40.77	P1	2.81	37.07	P1	2.88	34.73
7	P1	4.91	45.68	P1	3.99	41.06	P1	3.64	38.37
8	P1	2.30	47.98	P1	2.64	43.70	P1	1.55	39.92
9	P1	1.64	49.62	P1	1.86	45.56	P1	1.44	41.36
10	P1	1.94	51.55	P1	1.72	47.28	P1	1.52	42.88
11	P1	2.17	53.72	P1	2.40	49.68	P1	2.08	44.96
12	P1	1.94	55.66	P1	2.03	51.71	P1	2.58	47.54

Raw Score	Grade 3 PL Score	Grade 3 % of Students	Grade 3 Cumulative % of Students	Grade 4 PL Score	Grade 4 % of Students	Grade 4 Cumulative % of Students	Grade 5 PL Score	Grade 5 % of Students	Grade 5 Cumulative % of Students
13	P1	2.70	58.36	P1	2.50	54.21	P1	1.78	49.32
14	P1	2.14	60.50	P1	2.20	56.40	P1	2.12	51.44
15	P1	2.77	63.27	P1	2.16	58.57	P1	2.23	53.67
16	P2	2.70	65.98	P2	2.57	61.14	P2	2.99	56.67
17	P2	3.17	69.15	P2	3.68	64.82	P2	2.61	59.28
18	P2	3.21	72.35	P2	2.50	67.32	P2	2.61	61.89
19	P2	3.24	75.59	P2	3.79	71.11	P2	3.33	65.23
20	P2	3.61	79.20	P2	3.31	74.42	P2	3.26	68.48
21	P2	2.77	81.97	P2	2.60	77.02	P2	2.84	71.33
22	P2	2.64	84.61	P2	2.70	79.72	P2	3.26	74.58
23	P3	2.60	87.21	P3	2.70	82.43	P3	3.07	77.65
24	P3	2.14	89.35	P3	2.03	84.45	P3	3.07	80.72
25	P3	1.44	90.78	P3	1.79	86.25	P3	1.86	82.58
26	P3	1.24	92.02	P3	2.10	88.34	P3	1.93	84.51
27	P4	1.07	93.09	P4	1.45	89.79	P4	1.59	86.10
28	P4	1.14	94.22	P4	1.59	91.38	P4	2.27	88.37
29	P4	1.07	95.29	P4	1.79	93.17	P4	2.27	90.64
30	P5	0.87	96.16	P5	1.39	94.56	P5	2.08	92.73
31	P5	0.87	97.03	P5	1.32	95.88	P5	1.44	94.17
32	P5	2.97	100.0	P5	4.12	100.0	P5	5.83	100.0

Table 7.7.4.e.

Raw Score to Scale Score Conversion: 6–8

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	904	20	900	924
1	908	16	900	924
2	915	13	902	928
3	921	11	910	932
4	925	9	916	934
5	928	8	920	936
6	930	7	923	937
7	932	7	925	939
8	934	6	928	940
9	935	6	929	941
10	937	6	931	943
11	939	6	933	945

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
12	940	6	934	946
13	942	6	936	948
14	943	6	937	949
15	945	7	938	952
16	947	7	940	954
17	949	6	943	955
18	950	6	944	956
19	952	6	946	958
20	954	6	948	960
21	955	6	949	961
22	957	6	951	963
23	958	6	952	964
24	960	6	954	966
25	961	6	955	967
26	963	6	957	969
27	965	7	958	972
28	966	7	959	973
29	969	8	961	977
30	972	10	962	980
31	975	12	963	980
32	978	15	963	980

Table 7.7.4.f.

Raw Score to Proficiency Level Conversion: 6–8

Raw Score	Grade 6 PL Score	Grade 6 % of Students	Grade 6 Cumulative % of Students	Grade 7 PL Score	Grade 7 % of Students	Grade 7 Cumulative % of Students	Grade 8 PL Score	Grade 8 % of Students	Grade 8 Cumulative % of Students
0	P1	15.16	15.16	P1	13.92	13.92	P1	14.06	14.06
1	P1	1.59	16.75	P1	1.65	15.57	P1	1.90	15.96
2	P1	2.01	18.76	P1	1.65	17.22	P1	1.60	17.56
3	P1	9.27	28.03	P1	8.40	25.61	P1	7.40	24.96
4	P1	1.71	29.74	P1	1.84	27.45	P1	1.60	26.56
5	P1	2.63	32.37	P1	1.93	29.39	P1	1.95	28.51
6	P1	2.09	34.46	P1	1.75	31.13	P1	1.75	30.27
7	P1	1.25	35.71	P1	1.46	32.59	P1	1.25	31.52
8	P1	2.84	38.55	P1	2.41	35.00	P1	3.00	34.52
9	P1	2.55	41.10	P1	2.36	37.36	P1	2.60	37.12
10	P1	3.17	44.28	P1	2.17	39.53	P1	1.95	39.07

Raw Score	Grade 6 PL Score	Grade 6 % of Students	Grade 6 Cumulative % of Students	Grade 7 PL Score	Grade 7 % of Students	Grade 7 Cumulative % of Students	Grade 8 PL Score	Grade 8 % of Students	Grade 8 Cumulative % of Students
11	P1	3.43	47.70	P1	3.58	43.11	P1	3.10	42.17
12	P1	4.55	52.26	P1	4.67	47.78	P1	4.80	46.97
13	P1	2.34	54.59	P1	3.02	50.80	P1	2.75	49.72
14	P1	3.38	57.98	P1	3.54	54.34	P1	3.20	52.93
15	P2	3.22	61.19	P2	3.25	57.59	P2	3.05	55.98
16	P2	2.84	64.04	P2	2.88	60.47	P2	3.25	59.23
17	P2	3.22	67.25	P2	2.78	63.25	P2	3.05	62.28
18	P2	2.59	69.84	P2	2.88	66.13	P2	2.95	65.23
19	P2	2.67	72.51	P2	3.11	69.25	P2	1.90	67.13
20	P2	3.09	75.61	P2	3.25	72.50	P2	2.65	69.78
21	P3	2.34	77.94	P3	2.55	75.05	P3	1.50	71.29
22	P3	1.80	79.74	P3	2.08	77.12	P3	2.20	73.49
23	P3	1.13	80.87	P3	2.08	79.20	P3	1.85	75.34
24	P3	1.71	82.58	P3	1.79	80.99	P3	2.10	77.44
25	P3	1.63	84.21	P3	1.65	82.64	P3	1.60	79.04
26	P4	1.63	85.84	P4	2.50	85.14	P4	2.40	81.44
27	P4	1.13	86.97	P4	1.51	86.65	P4	2.15	83.59
28	P4	1.84	88.81	P4	1.84	88.49	P4	2.20	85.79
29	P4	1.46	90.27	P4	2.08	90.57	P4	2.10	87.89
30	P5	1.96	92.23	P5	1.89	92.45	P5	1.50	89.39
31	P5	1.88	94.11	P5	1.93	94.39	P5	2.35	91.75
32	P5	5.89	100.0	P5	5.61	100.0	P5	8.25	100.0

Table 7.7.4.g.

Raw Score to Scale Score Conversion: 9–12

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	906	22	900	928
1	912	16	900	928
2	920	12	908	932
3	925	10	915	935
4	929	9	920	938
5	931	8	923	939
6	934	7	927	941
7	936	7	929	943
8	938	7	931	945
9	939	6	933	945

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
10	941	6	935	947
11	943	6	937	949
12	944	6	938	950
13	946	6	940	952
14	947	6	941	953
15	949	6	943	955
16	950	6	944	956
17	952	6	946	958
18	953	6	947	959
19	954	6	948	960
20	956	6	950	962
21	957	6	951	963
22	958	6	952	964
23	960	6	954	966
24	961	6	955	967
25	962	6	956	968
26	964	6	958	970
27	965	6	959	971
28	967	7	960	974
29	969	8	961	977
30	972	10	962	980
31	975	12	963	980
32	980	17	963	980

Table 7.7.4.h.

Raw Score to Proficiency Level Conversion: 9–12

Raw Score	Grade 9 PL Score	Grade 9 % of Students	Grade 9 Cumulative % of Students	Grade 10 PL Score	Grade 10 % of Students	Grade 10 Cumulative % of Students	Grade 11 PL Score	Grade 11 % of Students	Grade 11 Cumulative % of Students	Grade 12 PL Score	Grade 12 % of Students	Grade 12 Cumulative % of Students
0	P1	16.51	16.51	P1	13.87	13.87	P1	15.93	15.93	P1	13.74	13.74
1	P1	2.36	18.87	P1	1.87	15.74	P1	1.97	17.90	P1	1.64	15.39
2	P1	2.15	21.02	P1	2.04	17.78	P1	1.48	19.37	P1	2.04	17.42
3	P1	7.91	28.93	P1	7.36	25.14	P1	6.89	26.26	P1	7.52	24.94
4	P1	1.47	30.40	P1	2.32	27.46	P1	1.72	27.98	P1	2.08	27.02
5	P1	2.41	32.81	P1	2.77	30.24	P1	2.28	30.26	P1	2.70	29.72
6	P1	2.46	35.27	P1	2.49	32.73	P1	2.34	32.60	P1	2.19	31.91
7	P1	1.83	37.11	P1	2.15	34.88	P1	1.66	34.26	P1	2.35	34.26
8	P1	4.40	41.51	P1	4.13	39.01	P1	4.43	38.68	P1	4.97	39.23

Raw Score	Grade 9 PL Score	Grade 9 % of Students	Grade 9 Cumulative % of Students	Grade 10 PL Score	Grade 10 % of Students	Grade 10 Cumulative % of Students	Grade 11 PL Score	Grade 11 % of Students	Grade 11 Cumulative % of Students	Grade 12 PL Score	Grade 12 % of Students	Grade 12 Cumulative % of Students
9	P1	2.41	43.92	P1	2.60	41.62	P1	2.21	40.90	P1	2.08	41.31
10	P1	1.68	45.60	P1	2.38	44.00	P1	2.28	43.17	P1	1.61	42.91
11	P1	3.35	48.95	P1	2.89	46.89	P1	2.64	45.82	P1	2.98	45.89
12	P1	2.99	51.94	P1	2.49	49.38	P1	2.83	48.65	P1	2.90	48.79
13	P1	2.25	54.19	P1	2.83	52.21	P1	2.40	51.05	P1	2.39	51.17
14	P2	3.35	57.55	P2	2.72	54.93	P2	3.26	54.31	P2	3.37	54.54
15	P2	2.78	60.32	P2	2.49	57.42	P2	2.71	57.01	P2	2.86	57.40
16	P2	2.25	62.58	P2	1.98	59.40	P2	2.58	59.59	P2	2.55	59.95
17	P2	3.09	65.67	P2	3.00	62.40	P2	2.83	62.42	P2	2.43	62.37
18	P2	2.20	67.87	P2	2.43	64.84	P2	2.03	64.45	P2	1.80	64.17
19	P2	2.52	70.39	P2	2.38	67.21	P2	2.15	66.61	P2	1.76	65.94
20	P2	2.41	72.80	P2	2.55	69.76	P2	2.34	68.94	P2	2.35	68.29
21	P3	1.78	74.58	P3	1.81	71.57	P3	1.60	70.54	P3	1.61	69.89
22	P3	1.36	75.94	P3	1.87	73.44	P3	1.41	71.96	P3	1.64	71.53
23	P3	2.41	78.35	P3	2.60	76.05	P3	1.97	73.92	P3	2.08	73.61
24	P3	2.25	80.61	P3	1.70	77.75	P3	1.72	75.65	P3	1.76	75.37
25	P3	1.47	82.08	P3	1.53	79.28	P3	1.78	77.43	P3	2.04	77.41
26	P3	2.62	84.70	P3	2.10	81.37	P3	2.28	79.70	P3	2.94	80.34
27	P4	1.89	86.58	P4	1.70	83.07	P4	1.78	81.49	P4	1.49	81.83
28	P4	1.00	87.58	P4	2.15	85.22	P4	2.46	83.95	P4	2.27	84.10
29	P4	1.52	89.10	P4	1.64	86.86	P4	1.91	85.85	P4	1.80	85.90
30	P4	1.99	91.09	P4	1.81	88.67	P4	1.97	87.82	P4	2.35	88.25
31	P5	1.89	92.98	P5	2.32	91.00	P5	2.34	90.16	P5	2.08	90.33
32	P5	7.02	100.0	P5	9.00	100.0	P5	9.84	100.0	P5	9.67	100.0

7.8. Equating Summary

In the 2023–2024 testing year, a revised version of the Alternate ACCESS test, aligned with the WIDA English Language Development Standards Framework, 2020 Edition, and the new Alternate Proficiency Level Descriptors (Alt PLDs), was released and administered. This updated version reflects a substantial increase in rigor, with 85% of the test items newly developed to align with the updated standards, resulting in a more challenging assessment. Most of the new items and a small portion of existing items were field tested in 2022–2023. Based on the field test item measures, a new operational test was created for the Series 602 Alternate ACCESS tests.

7.8.1. Spring 2024 Post Equating

In the spring of 2024, WIDA conducted a post equating to verify the field test item measures using the first year's operational data. The calibration was conducted by grade-level cluster and by domain. The task difficulty parameters (individual item measures and step measures) were anchored to the values derived from the 602 Field Test study. Displacement statistics were evaluated to determine whether these parameters need to be re-estimated based on the Series 602 verification sample data. The criterion was displacement $|\geq 0.5|$. The following is the summary of post-equating calibration per domain:

Speaking: no items with displacement value greater than $|\geq 0.5|$.

Listening: one item in G68 (displacement value = 0.5284) was released due to the displacement value greater than $|\geq 0.5|$.

Reading: one item in GK2 (displacement value = -0.5657) was released due to the displacement value greater than $|\geq 0.5|$.

Writing: three items in G35 (displacement value = 1.0378, -0.5748, -0.5557) were released due to the displacement value greater than $|\geq 0.5|$.

Please note that three K–2 grade-level cluster Writing items retained from the old version of the Series 601 Alternate ACCESS for ELLs test, which had a converted score of 0–2 during the field test administration, now have a converted score of 0–4.

Therefore, these three K–2 grade-level cluster Writing items were free estimated (unanchored) during the Series 602 Alternate ACCESS verification study.

7.8.2. Equipercentile Linking

To accommodate these changes of new items and of a new vertical scale, WIDA has rescaled the Series 602 Alternate ACCESS test. The previous scale (used in 2022–2023) ranged from 910 to 960, while the new scale for 2023–2024 spans a broader range of 900 to 980, reflecting the increased difficulty and growth measures across grade-level clusters. Due to these modifications, scores from the two cycles are not directly comparable since the Series 602 Alternate ACCESS started on a new calibration and scale. Consequently, scores between these versions cannot be directly compared without adjustments.

To ensure comparability between the Series 602 WIDA Alternate ACCESS and its predecessor, Series 601 Alternate ACCESS for ELLs, WIDA has implemented the Equipercentile Equating Method for score alignment. This method allows scores from the Series 602 WIDA Alternate ACCESS to be equated to those from the Series 601 Alternate ACCESS for ELLs, enabling the use of the old proficiency levels (PLs) for consistent interpretation. By applying this approach, WIDA aims to provide practitioners with a framework for understanding and interpreting the updated test scores within the context of the previous assessment.

Equipercentile equating is particularly effective because it directly matches scores from two test forms based on their percentile ranks among test takers. If a score on one test corresponds to the same percentile rank as a score on another test, the two scores are considered equivalent (Livingston, 2014). Unlike methods that assume normal score distributions, equipercentile equating uses observed ranks, making it adaptable to changes in content or difficulty, such as those seen in the updated Alternate ACCESS.

Despite its strengths, challenges have arisen in applying equipercentile equating to the Series 602 Alternate ACCESS. One significant issue is the absence of some scale score points, particularly at the high and low extremes, which complicates the interpretation of performances across the full proficiency range.

In the writing domain of the Series 602 Alternate ACCESS, specific problems include the absence of P3 proficiency level scores when using existing Series 601 Alternate ACCESS for ELLs' cut scores and a higher frequency of zero scores compared to Series 601 Alternate ACCESS. These issues make it difficult to accurately interpret writing proficiency and progression. The equating results also show that the P3 proficiency level does not appear in the equipercentile results. This absence suggests that final scores fall below the P3 cut established in the Series 601 Alternate ACCESS, creating challenges for practitioners attempting to interpret the data. This issue is consistent across all grade-level clusters for writing, indicating a systemic gap in the equating process that requires further attention.

To address these challenges, the Circle-Arc Method was applied as a follow-up to the initial Equipercentile equating. This method, introduced by Livingston and Kim (2008, 2009), was designed for equating in small sample sizes and scenarios where substantial differences in test difficulty occur. The Circle-Arc approach smooths the equating curve by using three reference points—Lower, Middle, and Upper—to account for score distributions at the extremes (LaFlair et al., 2017).

The primary advantage of the Circle-Arc method in this context is its ability to preserve the full range of scale scores, ensuring that both minimum and maximum values are included. This precision is crucial for identifying the P3 proficiency level at the upper end of the Series 601 Alternate ACCESS scale, enabling a comprehensive assessment of test takers' abilities.

Using the Circle-Arc method alongside the initial equipercentile equating effectively addressed the issue of missing scale scores at the extremes. The results, represented in the Circle-Arc column, provide a side-by-side comparison of traditional equipercentile equating and the adjusted scores from the Circle-Arc approach. This method successfully preserved the full range of scale scores, ensuring that all proficiency levels (PLs) are now included in the equated results.

Importantly, the Circle-Arc method maintained the overall proportions of proficiency levels, ensuring consistency with the original distribution observed in Series 601 Alternate ACCESS.

In conclusion, while the combined equating process has resolved many of the challenges in the writing domain, careful evaluation of the results remains essential. Differences in score distributions between Series 601 and 602 Alternate ACCESS must be scrutinized to understand their impact on score interpretation and the validity of crosswalk comparisons. The results of equipercentile linking are presented in Tables 7.8.4.a. through 7.8.7.d. (Equating Summary) by domain and grade-level cluster.

7.8.3. Final Calibration

With the population data of Series 602 Alternate ACCESS, the item measures of spring 2024 post-equating were evaluated to monitor item parameters drift. For data cleaning, students whose person outfit values were greater than 2.0 were removed up to 5% per grade-level cluster and domain when there are items with fit statistics (> 2.0) or negative item-total correlation. All item measures of the spring 2024 post-equating calibration were anchored on the Series 602 population data and its item displacement values were checked against our criterion of $|0.5|$. Six items (of the 144 total test items) showed displacement greater than .5. Item fit and item-total correlation were also reported. 8% of the test items have infit/outfit statistics greater than 2.0, and 11% of the test items have infit/outfit statistics between 1.5 and 2.0. The other 81% of test items have infit/outfit statistics below 1.5 aligning with the standards for being “productive for measurement” as defined by the aforementioned guideline. All items showed item-total correlation greater than 0.7. The Andrich threshold step measures derived from the field test calibration showed infit/outfit statistics values under 2.0 criterion in all domains and grade-level clusters, except the Writing domain K–2 grade-level cluster.

The results of item analysis of the final calibration are presented in Sections 7.8.4, 7.8.5, 7.8.6, and 7.8.7. Equating summary tables per grade-level cluster and domain present scale scores of Series 602 Alternate ACCESS, scale scores of Series 601 Alternate ACCESS for ELLs linked by equipercentile linking, PLs of Series 601 Alternate ACCESS linked to Series 601 scale scores, and PL of Series 602 Alternate ACCESS that were derived from the July 2024 standard setting. The old and new scale scores and PLs are linked to each other.

7.8.4. Listening

Table 7.8.4.a.

Equating Summary: K–2

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
900	910	A1	P1
907	921	A1	P1
916	924	A2	P1
920	929	A2	P1
923	930	A2	P1
926	931	A3	P1
929	932	A3	P1
931	934	A3	P1
934	934	A3	P1
937	934	A3	P2
939	934	A3	P2
941	935	A3	P2
943	935	A3	P3
945	935	A3	P3
946	935	A3	P3
948	935	A3	P3
949	935	A3	P4
950	935	A3	P4
951	936	A3	P4
952	937	A3	P4
953	938	P1	P4
954	939	P1	P4
955	940	P1	P4
956	941	P1	P4
957	942	P1	P4
958	943	P1	P4
960	945	P1	P5
961	946	P1	P5
963	947	P2	P5
965	948	P2	P5
968	948	P2	P5
971	950	P2	P5
974	953	P3	P5

Table 7.8.4.b.**Equating Summary: 3–5**

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
902	910	A1	P1
903	918	A1	P1
911	919	A1	P1
916	925	A2	P1
920	928	A2	P1
922	928	A2	P1
925	928	A2	P1
927	929	A2	P1
929	929	A2	P1
931	929	A2	P1
933	930	A2	P1
935	931	A3	P1
936	931	A3	P1
938	932	A3	P1
939	932	A3	P1
940	932	A3	P2
942	932	A3	P2
944	933	A3	P2
945	933	A3	P2
947	935	A3	P2
948	937	A3	P3
950	938	P1	P3
951	940	P1	P3
953	942	P1	P3
955	944	P1	P4
956	945	P1	P4
958	946	P1	P4
960	947	P2	P4
962	948	P2	P5
964	949	P2	P5
968	950	P2	P5
972	951	P2	P5
976	953	P3	P5

Table 7.8.4.c.**Equating Summary: 6–8**

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
904	910	A1	P1
908	917	A1	P1
915	918	A1	P1
921	923	A2	P1
925	926	A2	P1
928	927	A2	P1
930	927	A2	P1
932	927	A2	P1
934	928	A2	P1
935	928	A2	P1
937	928	A2	P1
939	929	A2	P1
940	929	A2	P1
942	930	A2	P1
943	930	A2	P2
945	930	A2	P2
947	932	A3	P2
949	933	A3	P2
950	934	A3	P3
952	935	A3	P3
954	936	A3	P3
955	938	P1	P3
957	939	P1	P3
958	940	P1	P4
960	941	P1	P4
961	942	P1	P4
963	943	P1	P5
965	943	P1	P5
966	944	P1	P5
969	945	P1	P5
972	946	P1	P5
975	948	P2	P5
978	953	P3	P5

Table 7.8.4.d.**Equating Summary: 9–12**

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
906	910	A1	P1
912	919	A1	P1
920	920	A1	P1
925	924	A2	P1
929	926	A2	P1
931	926	A2	P1
934	927	A2	P1
936	928	A2	P1
938	928	A2	P1
939	929	A2	P1
941	929	A2	P1
943	929	A2	P1
944	930	A2	P1
946	930	A2	P2
947	931	A3	P2
949	933	A3	P2
950	933	A3	P2
952	934	A3	P3
953	936	A3	P3
954	937	A3	P3
956	938	P1	P3
957	939	P1	P3
958	940	P1	P3
960	941	P1	P4
961	942	P1	P4
962	943	P1	P4
964	944	P1	P4
965	945	P1	P5
967	945	P1	P5
969	946	P1	P5
972	948	P2	P5
975	949	P2	P5
980	953	P3	P5

7.8.5. Reading

Table 7.8.5.a.

Equating Summary: K–2

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
900	910	A1	P1
907	921	A1	P1
916	924	A2	P1
920	929	A2	P1
923	930	A2	P1
926	931	A3	P1
929	932	A3	P1
931	934	A3	P1
934	934	A3	P1
937	934	A3	P1
939	934	A3	P1
941	935	A3	P1
943	935	A3	P2
945	935	A3	P2
946	935	A3	P2
948	935	A3	P2
949	935	A3	P2
950	935	A3	P3
951	936	A3	P3
952	937	A3	P3
953	938	P1	P3
954	939	P1	P3
955	940	P1	P3
956	941	P1	P3
957	942	P1	P4
958	943	P1	P4
960	945	P1	P4
961	946	P1	P4
963	947	P2	P5
965	948	P2	P5
968	948	P2	P5
971	950	P2	P5
974	953	P3	P5

Table 7.8.5.b.**Equating Summary: 3–5**

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
902	910	A1	P1
903	918	A1	P1
911	919	A1	P1
916	925	A2	P1
920	928	A2	P1
922	928	A2	P1
925	928	A2	P1
927	929	A2	P1
929	929	A2	P1
931	929	A2	P1
933	930	A2	P1
935	931	A3	P1
936	931	A3	P1
938	932	A3	P1
939	932	A3	P1
940	932	A3	P1
942	932	A3	P1
944	933	A3	P2
945	933	A3	P2
947	935	A3	P2
948	937	A3	P2
950	938	P1	P3
951	940	P1	P3
953	942	P1	P3
955	944	P1	P3
956	945	P1	P3
958	946	P1	P4
960	947	P2	P4
962	948	P2	P4
964	949	P2	P4
968	950	P2	P5
972	951	P2	P5
976	953	P3	P5

Table 7.8.5.c.**Equating Summary: 6–8**

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
904	910	A1	P1
908	917	A1	P1
915	918	A1	P1
921	923	A2	P1
925	926	A2	P1
928	927	A2	P1
930	927	A2	P1
932	927	A2	P1
934	928	A2	P1
935	928	A2	P1
937	928	A2	P1
939	929	A2	P1
940	929	A2	P1
942	930	A2	P1
943	930	A2	P1
945	930	A2	P2
947	932	A3	P2
949	933	A3	P2
950	934	A3	P3
952	935	A3	P3
954	936	A3	P3
955	938	P1	P3
957	939	P1	P4
958	940	P1	P4
960	941	P1	P4
961	942	P1	P4
963	943	P1	P4
965	943	P1	P4
966	944	P1	P4
969	945	P1	P5
972	946	P1	P5
975	948	P2	P5
978	953	P3	P5

Table 7.8.5.d.**Equating Summary: 9–12**

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
906	910	A1	P1
912	919	A1	P1
920	920	A1	P1
925	924	A2	P1
929	926	A2	P1
931	926	A2	P1
934	927	A2	P1
936	928	A2	P1
938	928	A2	P1
939	929	A2	P1
941	929	A2	P1
943	929	A2	P1
944	930	A2	P2
946	930	A2	P2
947	931	A3	P2
949	933	A3	P2
950	933	A3	P3
952	934	A3	P3
953	936	A3	P3
954	937	A3	P3
956	938	P1	P3
957	939	P1	P4
958	940	P1	P4
960	941	P1	P4
961	942	P1	P4
962	943	P1	P4
964	944	P1	P4
965	945	P1	P4
967	945	P1	P4
969	946	P1	P5
972	948	P2	P5
975	949	P2	P5
980	953	P3	P5

7.8.6. Speaking

Table 7.8.6.a.

Equating Summary: K–2

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
900	910	A1	P1
907	921	A1	P1
916	924	A2	P1
920	929	A2	P1
923	930	A2	P1
926	931	A3	P1
929	932	A3	P1
931	934	A3	P1
934	934	A3	P1
937	934	A3	P1
939	934	A3	P1
941	935	A3	P2
943	935	A3	P2
945	935	A3	P2
946	935	A3	P2
948	935	A3	P3
949	935	A3	P3
950	935	A3	P3
951	936	A3	P3
952	937	A3	P3
953	938	P1	P3
954	939	P1	P3
955	940	P1	P3
956	941	P1	P3
957	942	P1	P3
958	943	P1	P4
960	945	P1	P4
961	946	P1	P4
963	947	P2	P5
965	948	P2	P5
968	948	P2	P5
971	950	P2	P5
974	953	P3	P5

Table 7.8.6.b.**Equating Summary: 3–5**

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
902	910	A1	P1
903	918	A1	P1
911	919	A1	P1
916	925	A2	P1
920	928	A2	P1
922	928	A2	P1
925	928	A2	P1
927	929	A2	P1
929	929	A2	P1
931	929	A2	P1
933	930	A2	P1
935	931	A3	P1
936	931	A3	P1
938	932	A3	P1
939	932	A3	P1
940	932	A3	P1
942	932	A3	P1
944	933	A3	P1
945	933	A3	P1
947	935	A3	P2
948	937	A3	P2
950	938	P1	P2
951	940	P1	P2
953	942	P1	P3
955	944	P1	P3
956	945	P1	P3
958	946	P1	P3
960	947	P2	P4
962	948	P2	P4
964	949	P2	P4
968	950	P2	P5
972	951	P2	P5
976	953	P3	P5

Table 7.8.6.c**Equating Summary: 6–8**

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
904	910	A1	P1
908	917	A1	P1
915	918	A1	P1
921	923	A2	P1
925	926	A2	P1
928	927	A2	P1
930	927	A2	P1
932	927	A2	P1
934	928	A2	P1
935	928	A2	P1
937	928	A2	P1
939	929	A2	P1
940	929	A2	P1
942	930	A2	P1
943	930	A2	P1
945	930	A2	P1
947	932	A3	P2
949	933	A3	P2
950	934	A3	P2
952	935	A3	P2
954	936	A3	P3
955	938	P1	P3
957	939	P1	P3
958	940	P1	P3
960	941	P1	P3
961	942	P1	P4
963	943	P1	P4
965	943	P1	P4
966	944	P1	P5
969	945	P1	P5
972	946	P1	P5
975	948	P2	P5
978	953	P3	P5

Table 7.8.6.d.**Equating Summary: 9–12**

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
906	910	A1	P1
912	919	A1	P1
920	920	A1	P1
925	924	A2	P1
929	926	A2	P1
931	926	A2	P1
934	927	A2	P1
936	928	A2	P1
938	928	A2	P1
939	929	A2	P1
941	929	A2	P1
943	929	A2	P1
944	930	A2	P1
946	930	A2	P2
947	931	A3	P2
949	933	A3	P2
950	933	A3	P2
952	934	A3	P2
953	936	A3	P2
954	937	A3	P3
956	938	P1	P3
957	939	P1	P3
958	940	P1	P3
960	941	P1	P3
961	942	P1	P4
962	943	P1	P4
964	944	P1	P4
965	945	P1	P4
967	945	P1	P5
969	946	P1	P5
972	948	P2	P5
975	949	P2	P5
980	953	P3	P5

7.8.7. Writing

Table 7.8.7.a.

Equating Summary: K–2

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
900	910	A1	P1
907	921	A1	P1
916	924	A2	P1
920	929	A2	P1
923	930	A2	P1
926	931	A3	P1
929	932	A3	P1
931	934	A3	P1
934	934	A3	P1
937	934	A3	P1
939	934	A3	P1
941	935	A3	P2
943	935	A3	P2
945	935	A3	P2
946	935	A3	P2
948	935	A3	P2
949	935	A3	P2
950	935	A3	P2
951	936	A3	P3
952	937	A3	P3
953	938	P1	P3
954	939	P1	P3
955	940	P1	P3
956	941	P1	P3
957	942	P1	P3
958	943	P1	P3
960	945	P1	P4
961	946	P1	P4
963	947	P2	P4
965	948	P2	P4
968	948	P2	P5
971	950	P2	P5
974	953	P3	P5

Table 7.8.7.b.**Equating Summary: 3–5**

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
902	910	A1	P1
903	918	A1	P1
911	919	A1	P1
916	925	A2	P1
920	928	A2	P1
922	928	A2	P1
925	928	A2	P1
927	929	A2	P1
929	929	A2	P1
931	929	A2	P1
933	930	A2	P1
935	931	A3	P1
936	931	A3	P1
938	932	A3	P1
939	932	A3	P1
940	932	A3	P1
942	932	A3	P2
944	933	A3	P2
945	933	A3	P2
947	935	A3	P2
948	937	A3	P2
950	938	P1	P2
951	940	P1	P2
953	942	P1	P3
955	944	P1	P3
956	945	P1	P3
958	946	P1	P3
960	947	P2	P4
962	948	P2	P4
964	949	P2	P4
968	950	P2	P5
972	951	P2	P5
976	953	P3	P5

Table 7.8.7.c.**Equating Summary: 6–8**

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
904	910	A1	P1
908	917	A1	P1
915	918	A1	P1
921	923	A2	P1
925	926	A2	P1
928	927	A2	P1
930	927	A2	P1
932	927	A2	P1
934	928	A2	P1
935	928	A2	P1
937	928	A2	P1
939	929	A2	P1
940	929	A2	P1
942	930	A2	P1
943	930	A2	P1
945	930	A2	P2
947	932	A3	P2
949	933	A3	P2
950	934	A3	P2
952	935	A3	P2
954	936	A3	P2
955	938	P1	P3
957	939	P1	P3
958	940	P1	P3
960	941	P1	P3
961	942	P1	P3
963	943	P1	P4
965	943	P1	P4
966	944	P1	P4
969	945	P1	P4
972	946	P1	P5
975	948	P2	P5
978	953	P3	P5

Table 7.8.7.d.**Equating Summary: 9–12**

Scale Score Series 602	Scale Score Series 601	PL 601	PL 602
906	910	A1	P1
912	919	A1	P1
920	920	A1	P1
925	924	A2	P1
929	926	A2	P1
931	926	A2	P1
934	927	A2	P1
936	928	A2	P1
938	928	A2	P1
939	929	A2	P1
941	929	A2	P1
943	929	A2	P1
944	930	A2	P1
946	930	A2	P1
947	931	A3	P2
949	933	A3	P2
950	933	A3	P2
952	934	A3	P2
953	936	A3	P2
954	937	A3	P2
956	938	P1	P2
957	939	P1	P3
958	940	P1	P3
960	941	P1	P3
961	942	P1	P3
962	943	P1	P3
964	944	P1	P3
965	945	P1	P4
967	945	P1	P4
969	946	P1	P4
972	948	P2	P4
975	949	P2	P5
980	953	P3	P5

7.9. Test Characteristic Curve

For each test form, the test characteristic curve graphically shows the relationship between the ability measure (in logits) on the horizontal axis and the expected raw score on the vertical axis. Four vertical lines indicate the four cut scores, dividing the figure into five sections for each of the WIDA Alternate Proficiency Levels (P1–P5) for the domain being tested. As would be expected, higher raw scores are expected to be placed into higher language proficiency levels. The relative width of each section between the cut score lines, however, gives an indication of how many points must be earned to be placed into a WIDA Alternate Proficiency Level.

In item response theory, the definition of an expected score according to Andrich (1978) is used. The formula for a true score is given in this equation:

$$\text{Expected Score}(\theta_n) = \sum_{i=1}^I \left[\sum_{k=0}^K [k \times P_{nik}] \right]$$

where n is an examinee, i denotes an item, and k is k item category; P_{nik} is the probability of person n scoring k on item i based on the Rating Scale model. ES_n is the expected score for an examinee with ability level θ_n .

7.9.1. Listening Test Characteristic Curves

Figure 7.9.1.a.

Test Characteristic Curve: K–2

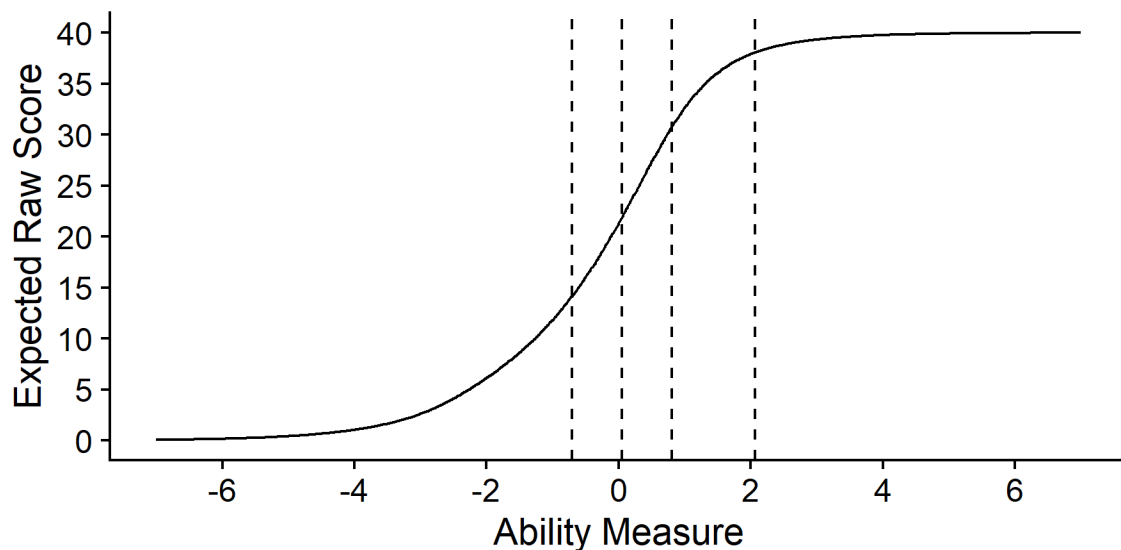


Figure 7.9.1.b.

Test Characteristic Curve: 3–5

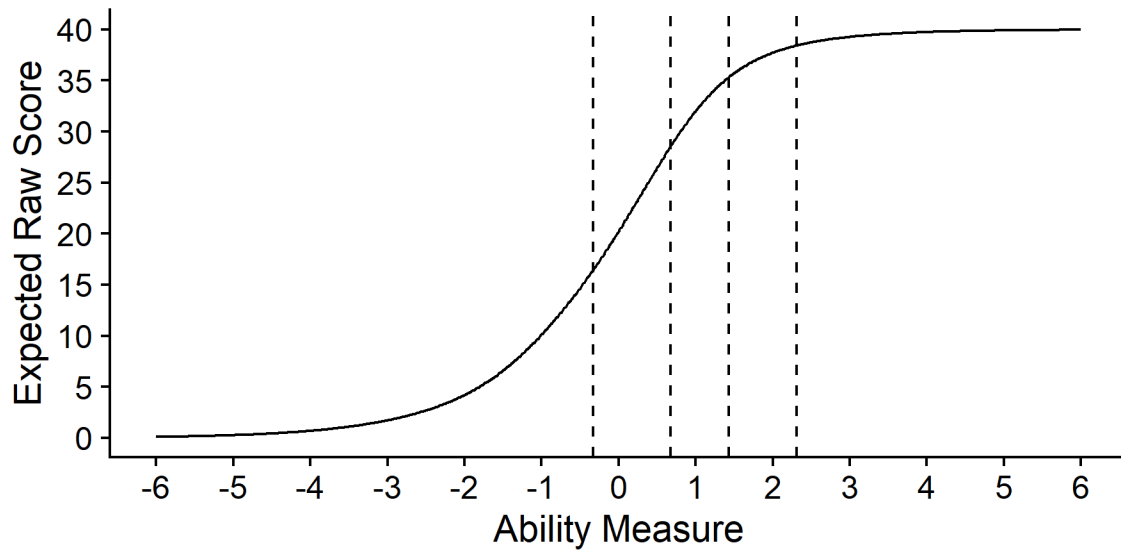


Figure 7.9.1.c.

Test Characteristic Curve: 6–8

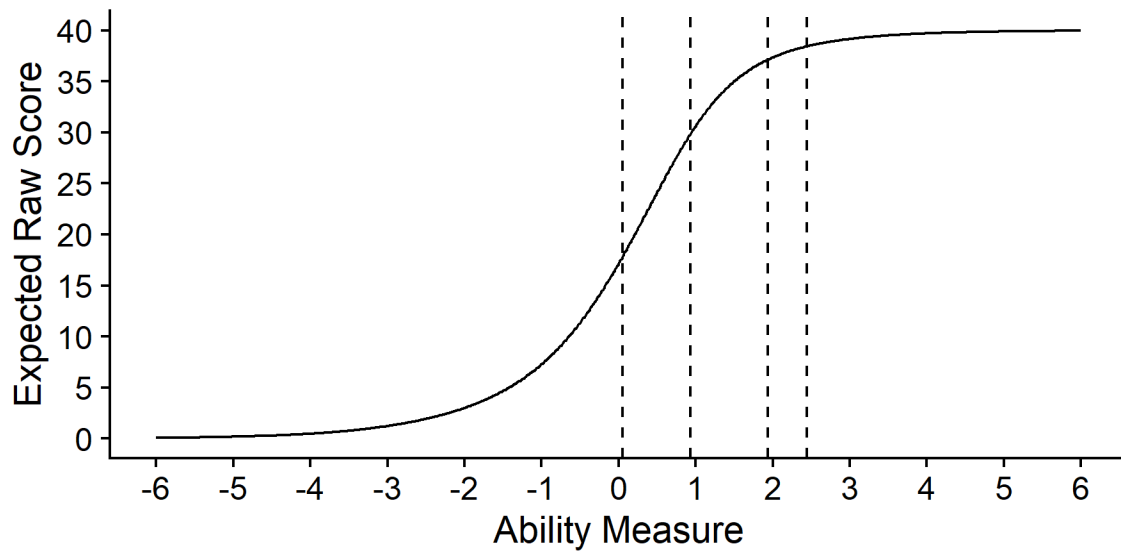
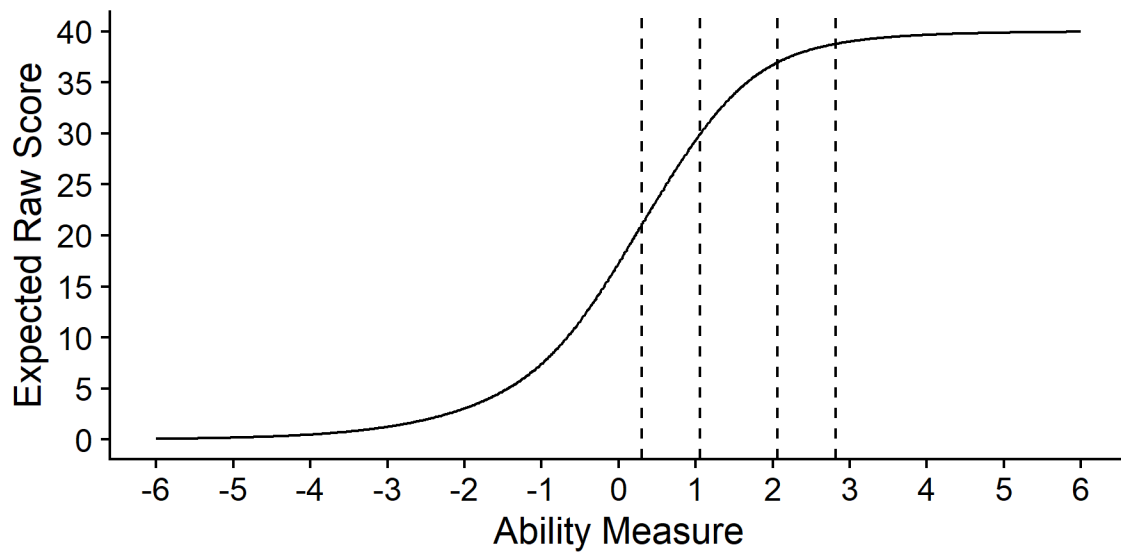


Figure 7.9.1.d.

Test Characteristic Curve: 9–12



7.9.2. Reading Test Characteristic Curves

Figure 7.9.2.a.

Test Characteristic Curve: K–2

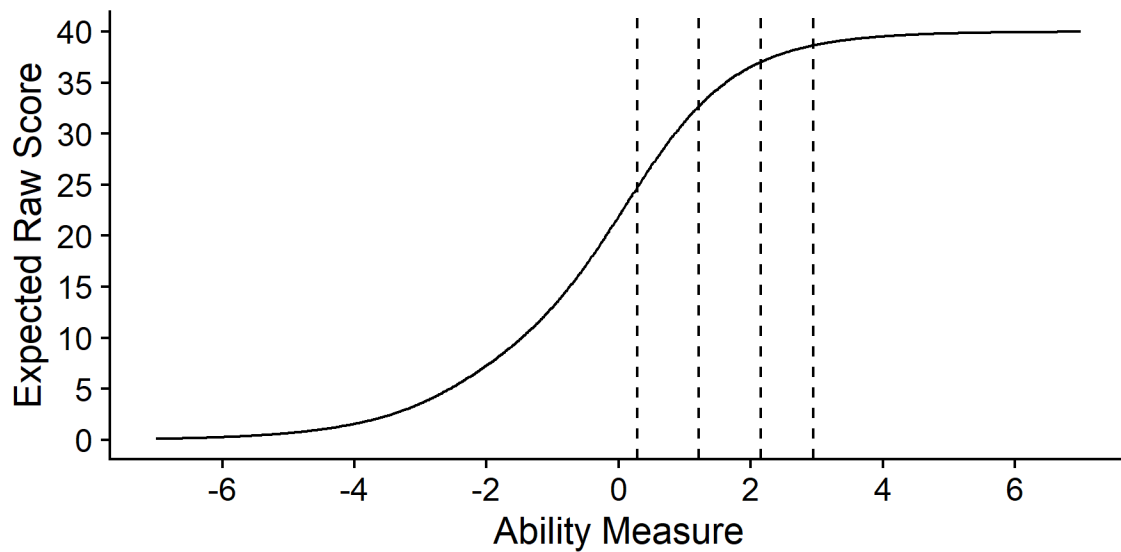


Figure 7.9.2.b.

Test Characteristic Curve: 3–5

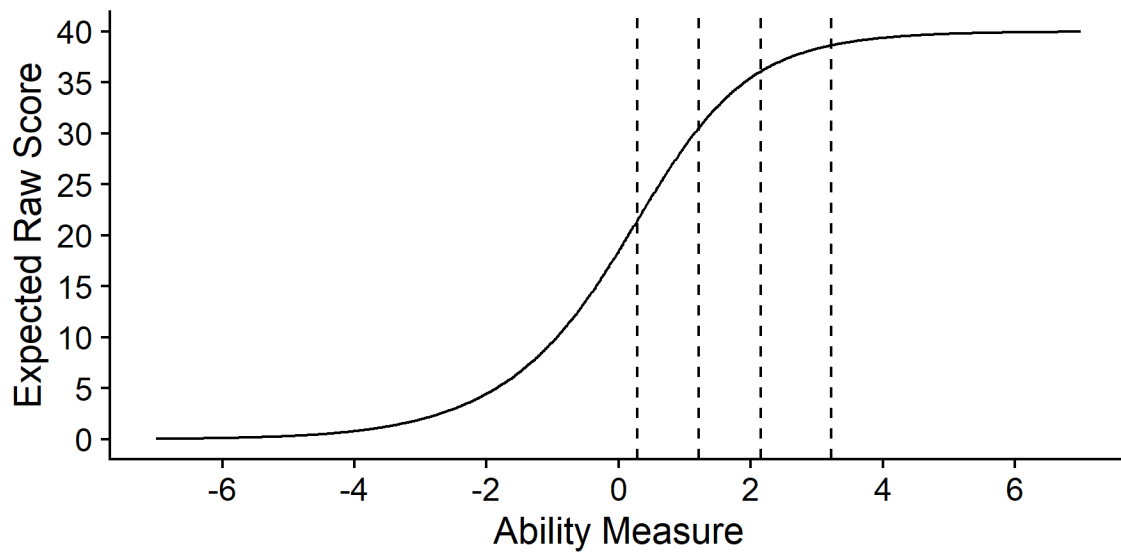


Figure 7.9.2.c.

Test Characteristic Curve: 6–8

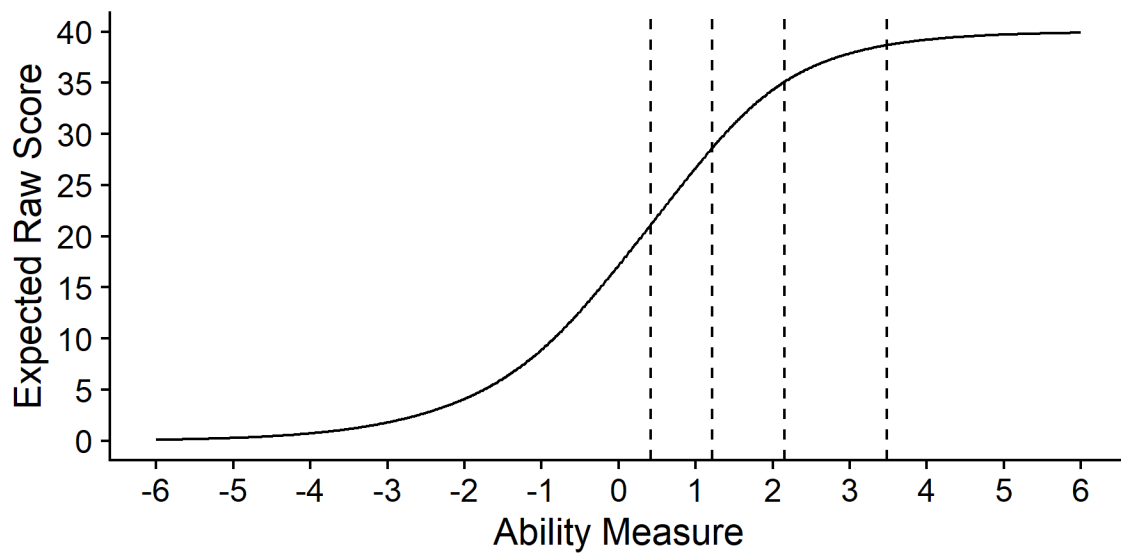
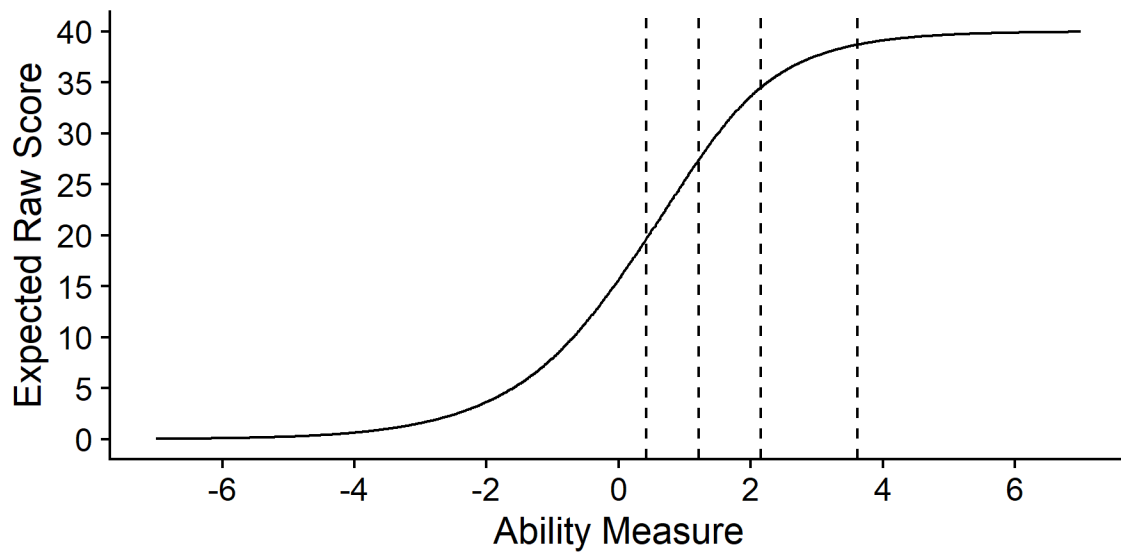


Figure 7.9.2.d.

Test Characteristic Curve: 9–12



7.9.3. Speaking Test Characteristic Curves

Figure 7.9.3.a.

Test Characteristic Curve: K–2

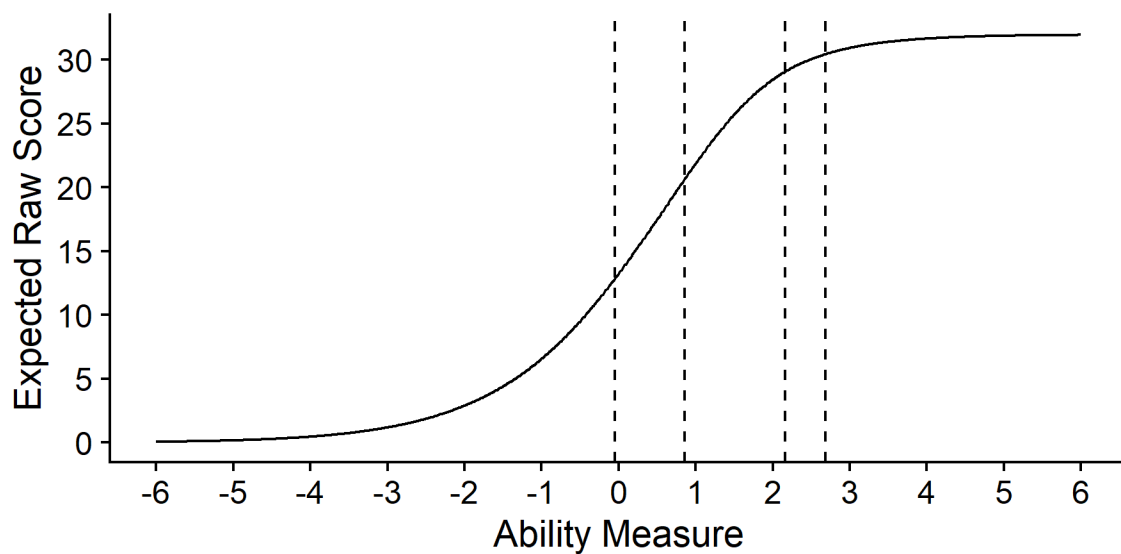


Figure 7.9.3.b.

Test Characteristic Curve: 3–5

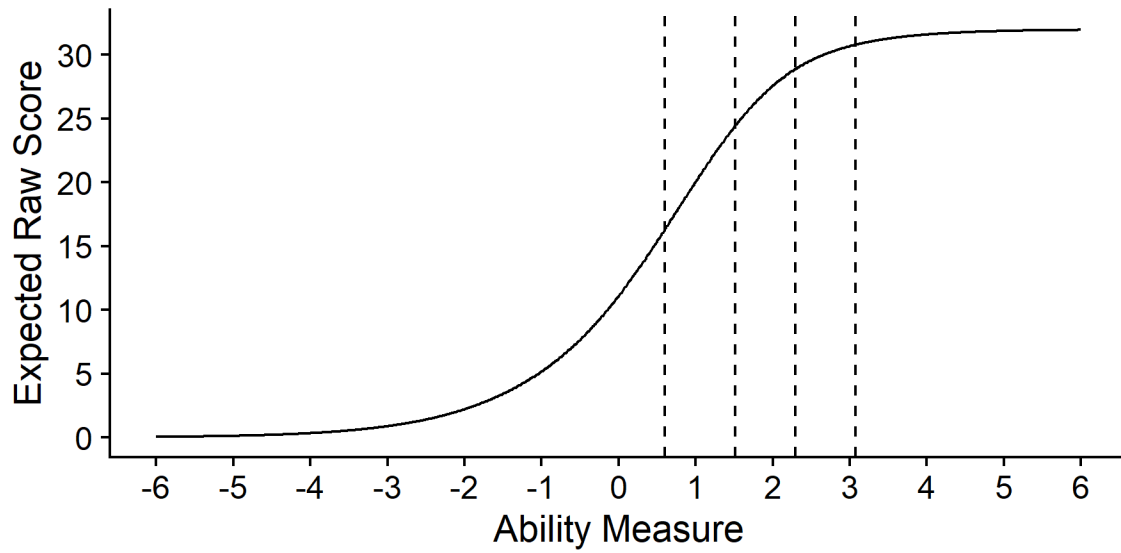


Figure 7.9.3.c.

Test Characteristic Curve: 6–8

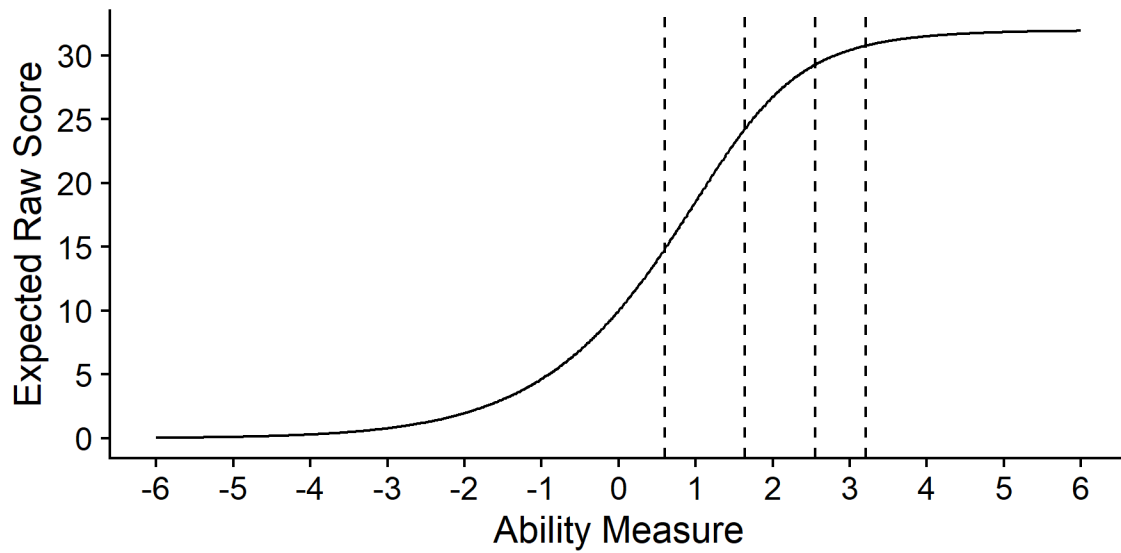
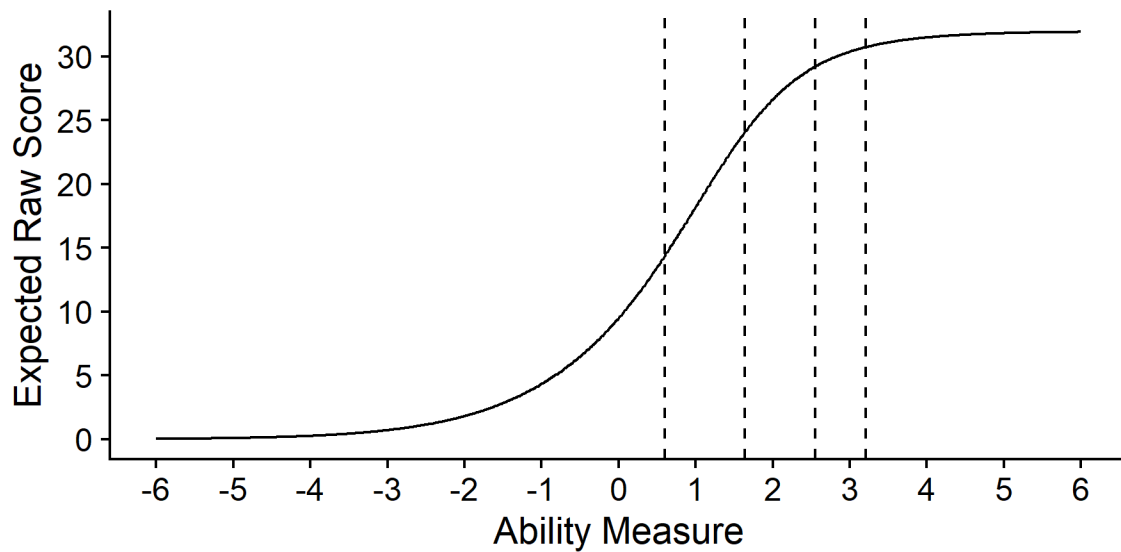


Figure 7.9.3.d.

Test Characteristic Curve: 9–12



7.9.4. Writing Test Characteristic Curves

Figure 7.9.4.a.

Test Characteristic Curve: K–2

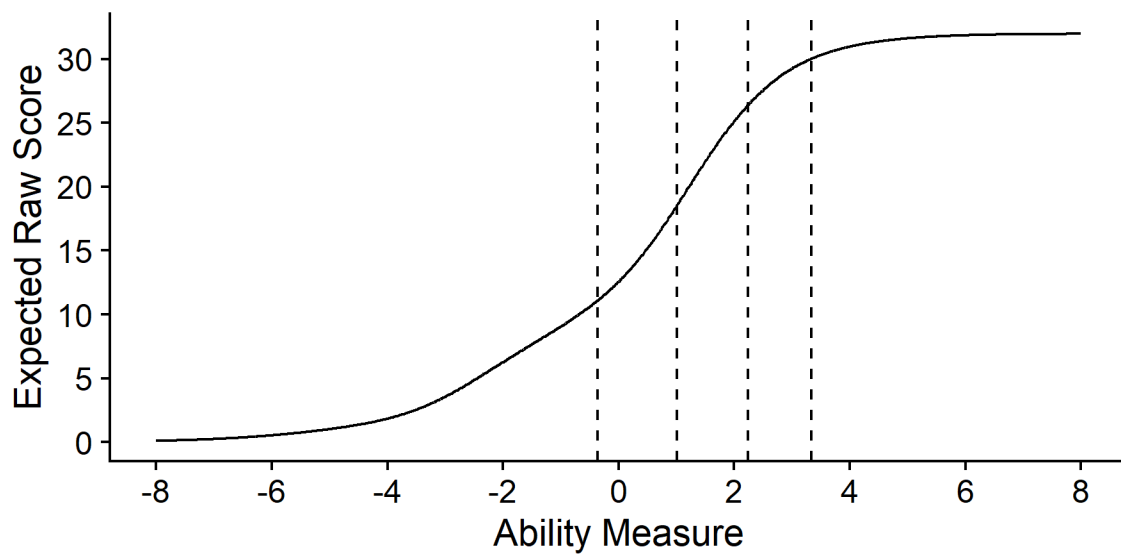


Figure 7.9.4.b.

Test Characteristic Curve: 3-5

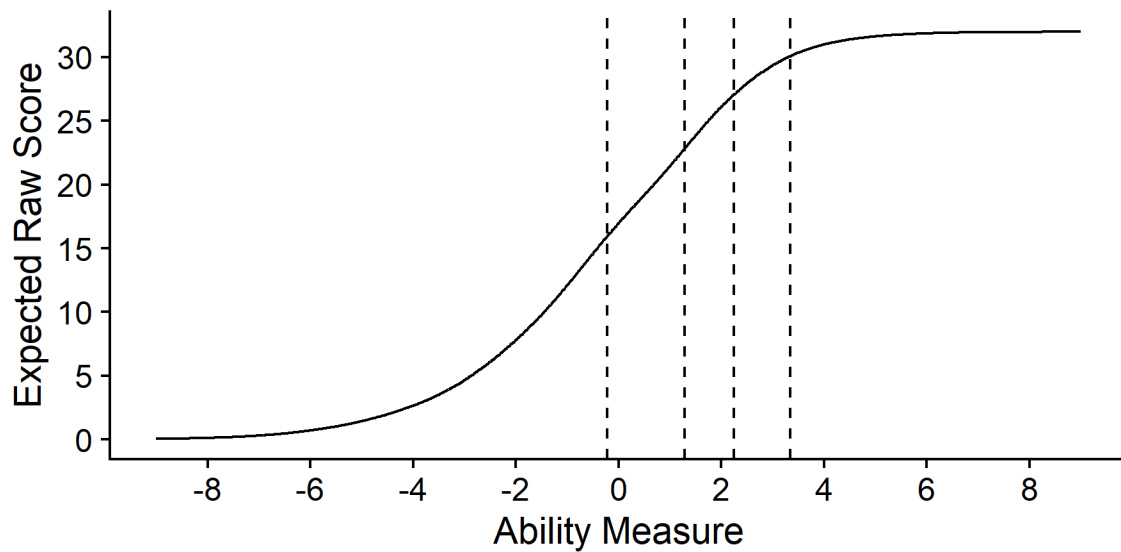


Figure 7.9.4.c.

Test Characteristic Curve: 6-8

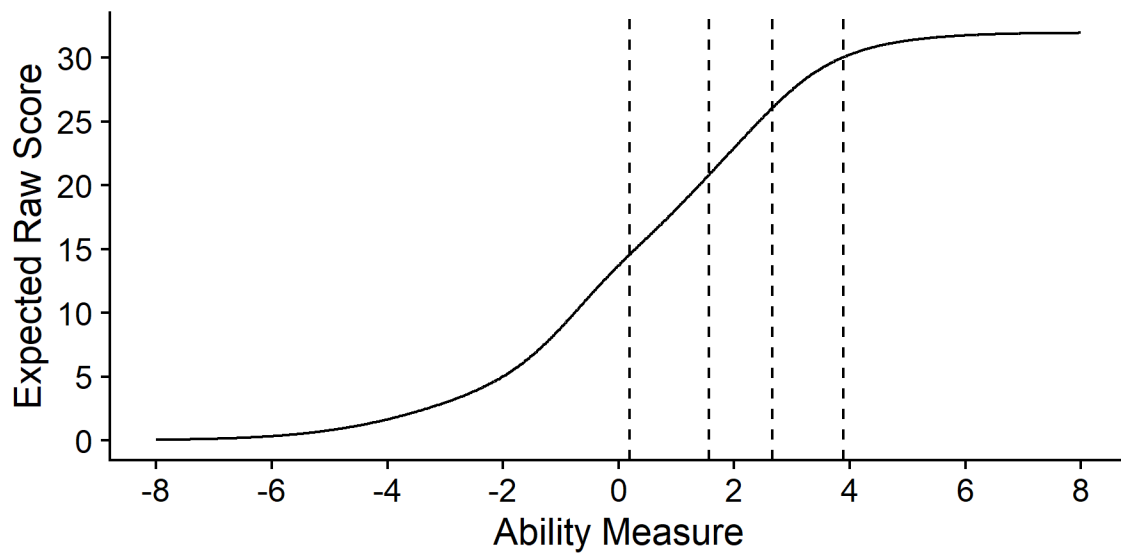
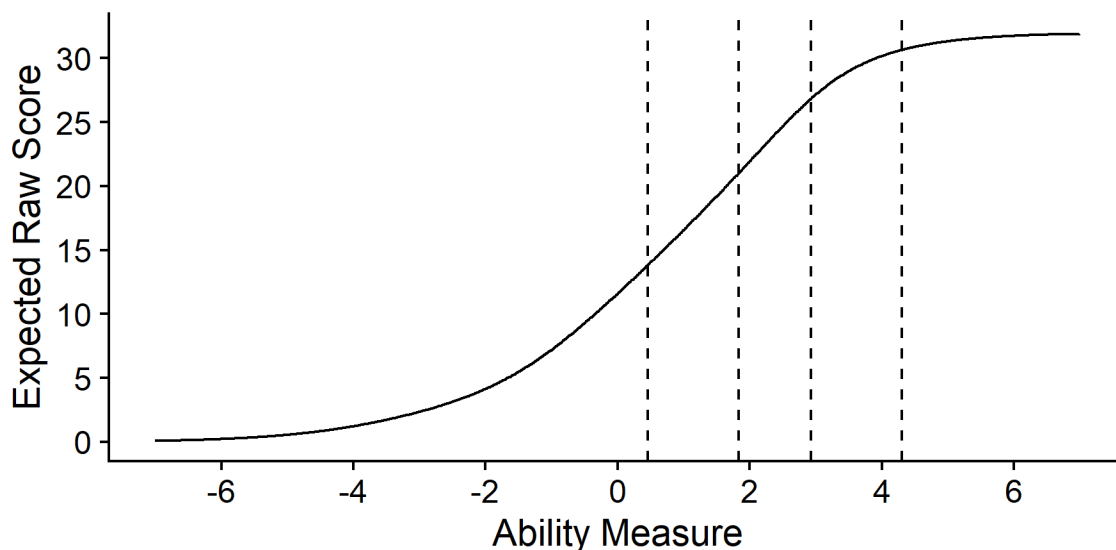


Figure 7.9.4.d.

Test Characteristic Curve: 9–12



7.10. *Test Information Curve*

With the Rasch measurement model, as with any measurement model following Item Response Theory (IRT), the relationship between the ability measure (in logits) and the accuracy of test scores can be modeled. It is recognized that tests measure most accurately when the abilities of the examinees and the difficulty of the items are most appropriate for each other. If a test is too difficult for an examinee (i.e., the examinee scores close to zero), or if the test is too easy for an examinee (i.e., the examinee “tops out”), accurate measurement of the examinee’s ability cannot be made. The test information function shows graphically how well the test is measuring across the ability measure spectrum in terms of measurement error. High values indicate more accuracy in measurement. Thus, for each test form, Figure E shows the relationship between the ability measure (in logits) on the horizontal axis and measurement accuracy, represented as the Fisher information value (which is the inverse squared of the standard error), on the vertical axis. The test information function, then, reflects the conditional standard error of measurement.

The test information function is an advanced IRT concept. It is important mainly because it provides indices analogous to reliability and SEM in classical test theory. Without using statistical formulations, we can conceptualize the idea this way: in a well-designed test, every item responded to correctly provides a bit of information about what a student knows and can do, and every item responded to incorrectly indicates what a student does not know and can’t do. When there are a sufficient number of items, information accumulates to provide an accurate estimate of student ability. In this sense, information is directly related to the reliability of test scores: the more information, the higher the reliability and the smaller the SEM.

Test information varies as a function of student ability. The same test can provide a significant amount of information for some students, but little information for other students. Usually, an achievement assessment is designed for students ranging from relatively low ability to relatively high ability. A student in this range is expected to answer some items correctly and some items incorrectly. However, if a student has extremely high ability which is far beyond the ability level required by the test, he or she might answer all items correctly. This is good from an educational point of view, but it is tricky from an ability-estimation point of view, since this test provides little information about the student's true level of ability. We certainly know the student has high ability, but there is no way to determine how high it is. To determine the true ability would require the administration of several additional items at the top of the difficulty range. From this example, it is clear that IRT test information is conditioned on ability. Usually, the test information curve has a bell shape—intermediate abilities provide for the greatest test information and high reliability, whereas extreme abilities correspond to less information and low reliability.

Statistically, at every ability point, the test information function is inversely proportional to the square of the CSEM. This relationship is used to calculate the CSEM for each obtainable scale score point.

The TIF for the RSM is defined as follows:

$$I(\theta) = \sum_{i=1}^L I_i(\theta)$$

where $I_i(\theta)$ is $\sum_{k=0}^m k^2 P_{ik} - (\sum_{k=0}^m k P_{ik})^2$; i denotes an item, k is k item category; P_{ik} is the probability of scoring k on item i given θ based on the Rating Scale model. $I(\theta)$ is the amount of test information at an ability level of θ .

Again, as in the Figures in Section 7.9., four vertical lines in the Figures in Section 7.10. indicate the four cut scores, dividing the figure into five sections for each of the WIDA Alternate Proficiency Levels (P1–P5) for the domain being tested. It is important that each test form measure most accurately in the areas for which it is primarily used to make classification decisions. In other words, optimally the test information function should be high for the cuts between P1/P2, P2/P3, P3/P4, and P4/P5.

7.10.1. Listening Test Information Curves

Figure 7.10.1.a.

Test Information Curve: K-2

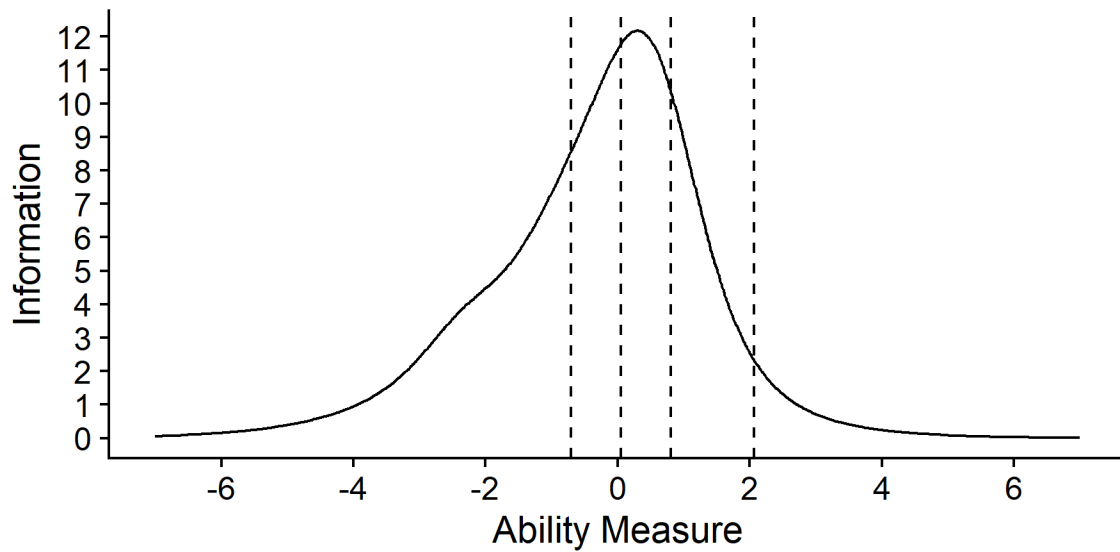


Figure 7.10.1.b.

Test Information Curve: 3-5

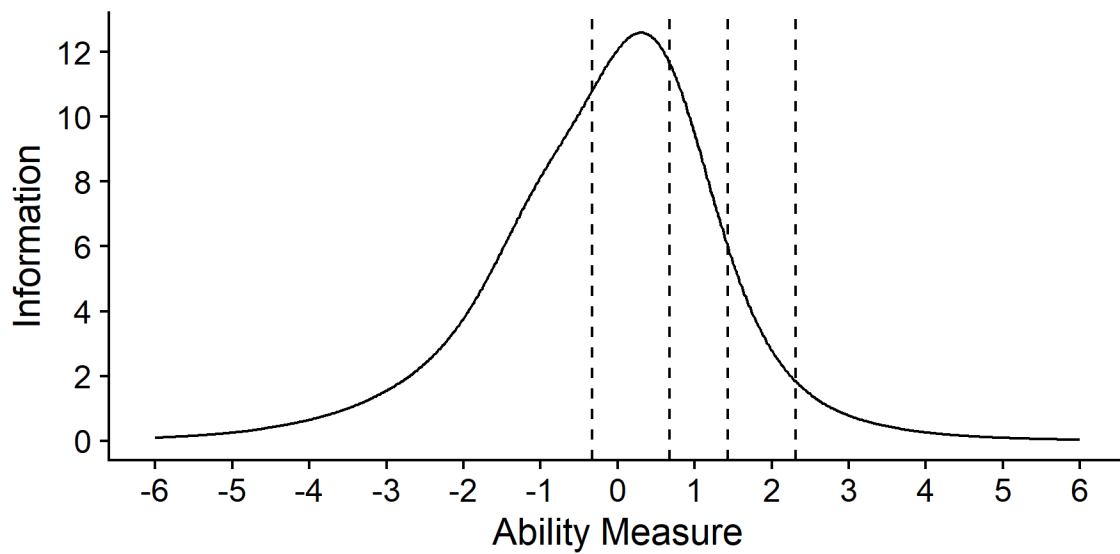


Figure 7.10.1.c.

Test Information Curve: 6–8

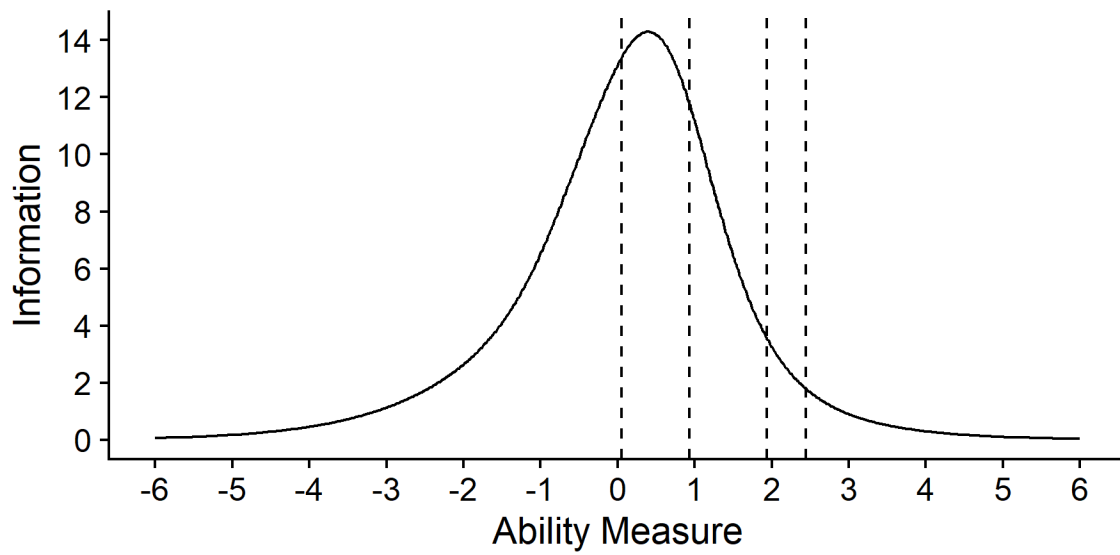
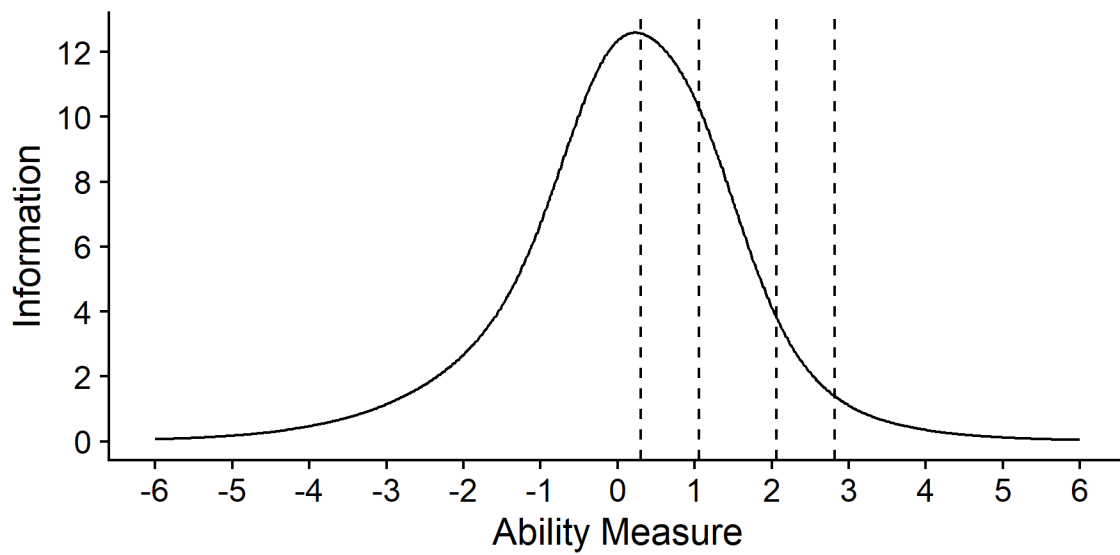


Figure 7.10.1.d.

Test Information Curve: 9–12



7.10.2. Reading Test Information Curves

Figure 7.10.2.a.

Test Information Curve: K-2

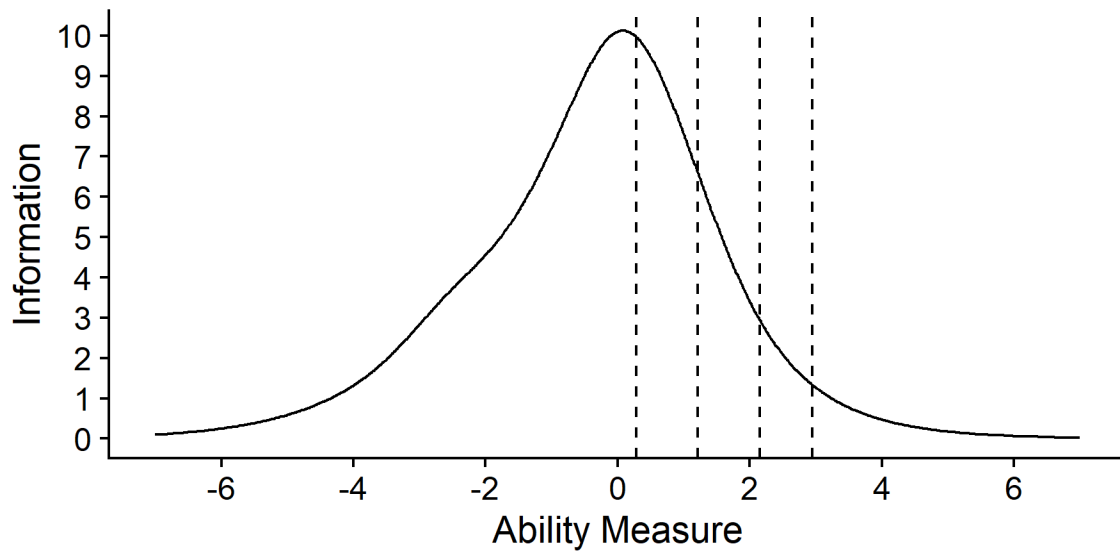


Figure 7.10.2.b.

Test Information Curve: 3-5

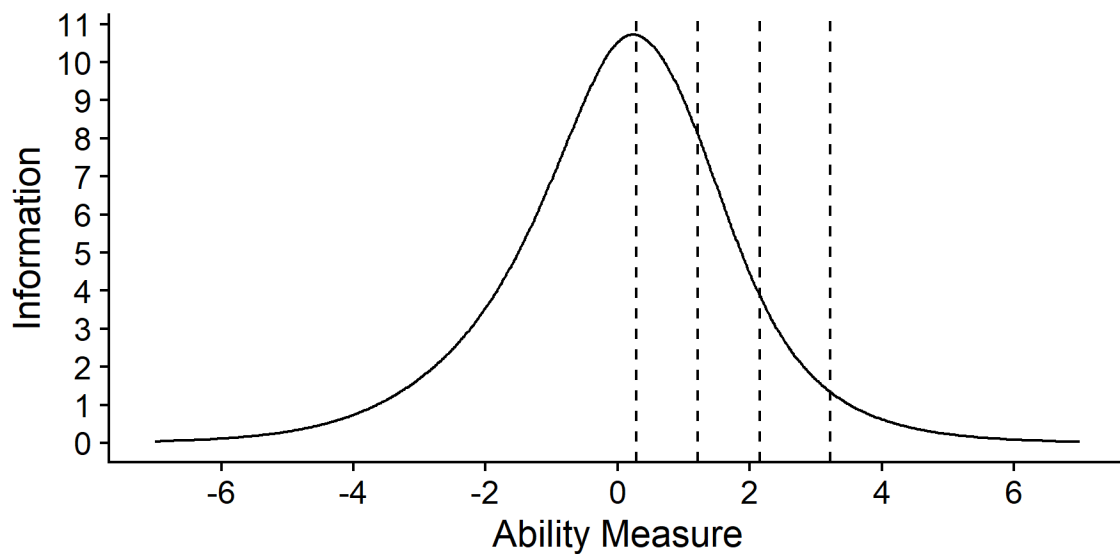


Figure 7.10.2.c.

Test Information Curve: 6–8

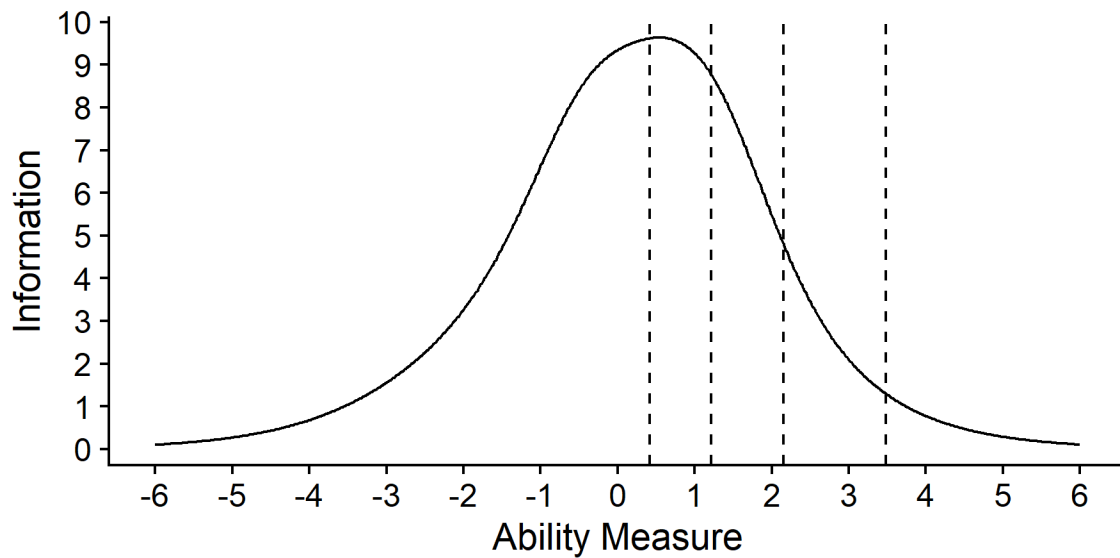
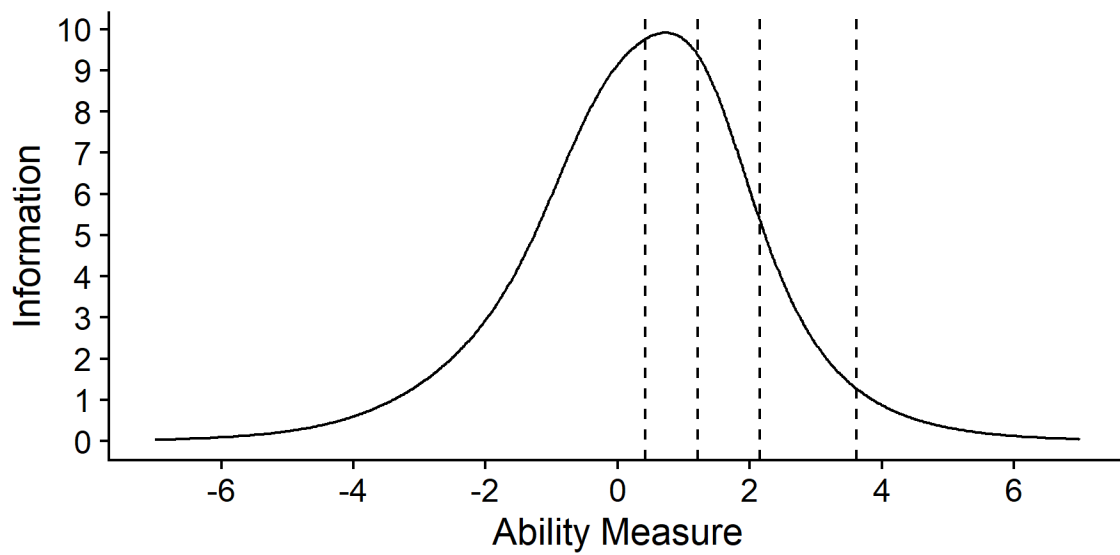


Figure 7.10.2.d.

Test Information Curve: 9–12



7.10.3. Speaking Test Information Curves

Figure 7.10.3.a.

Test Information Curve: K-2

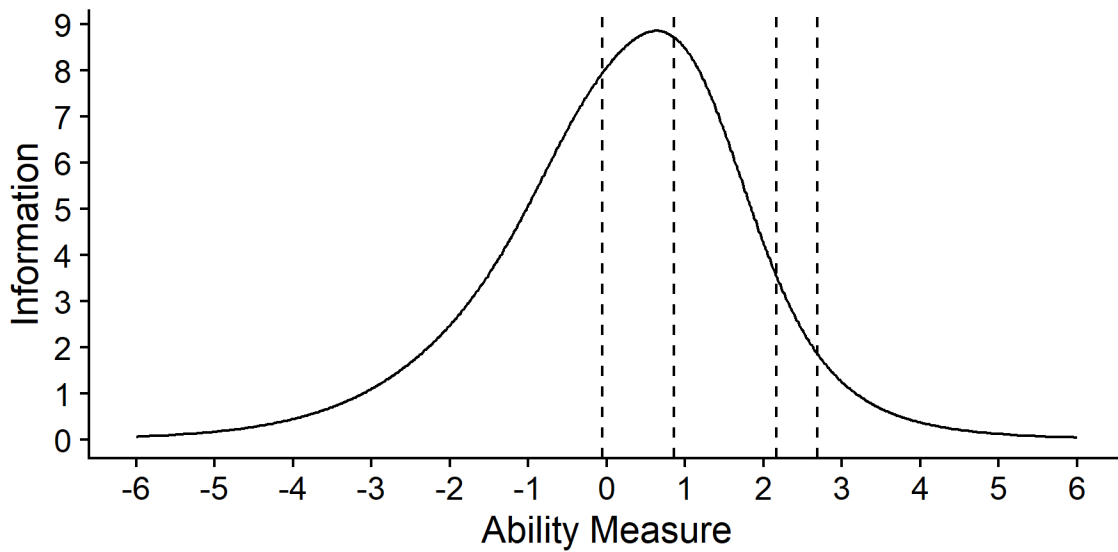


Figure 7.10.3.b.

Test Information Curve: 3-5

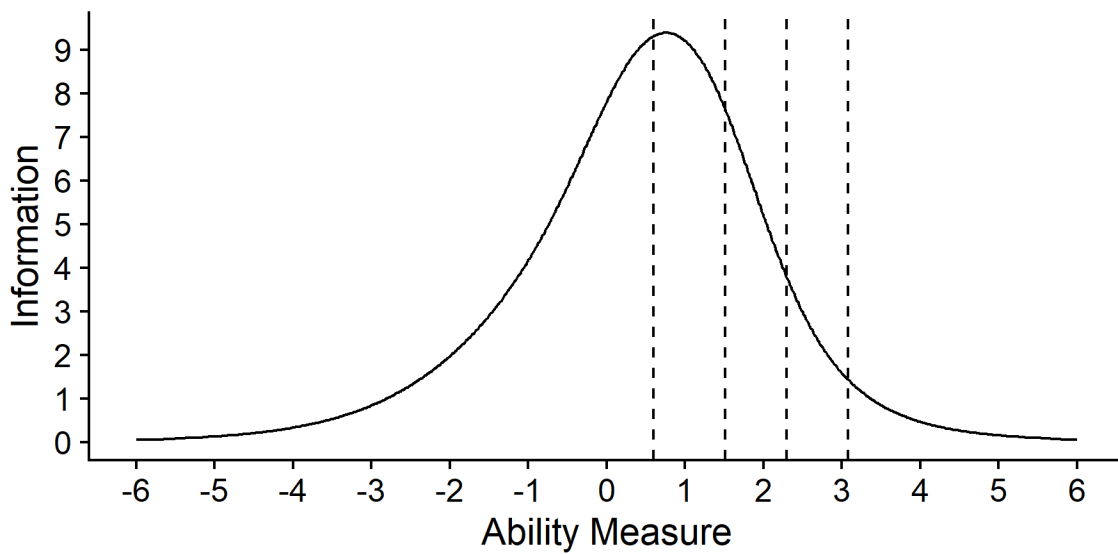


Figure 7.10.3.c.

Test Information Curve: 6–8

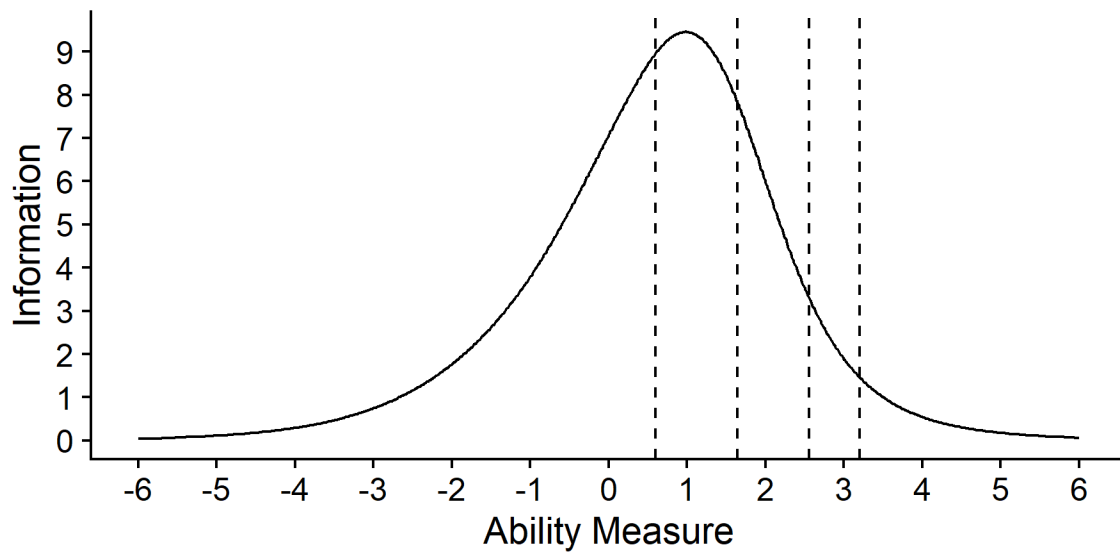
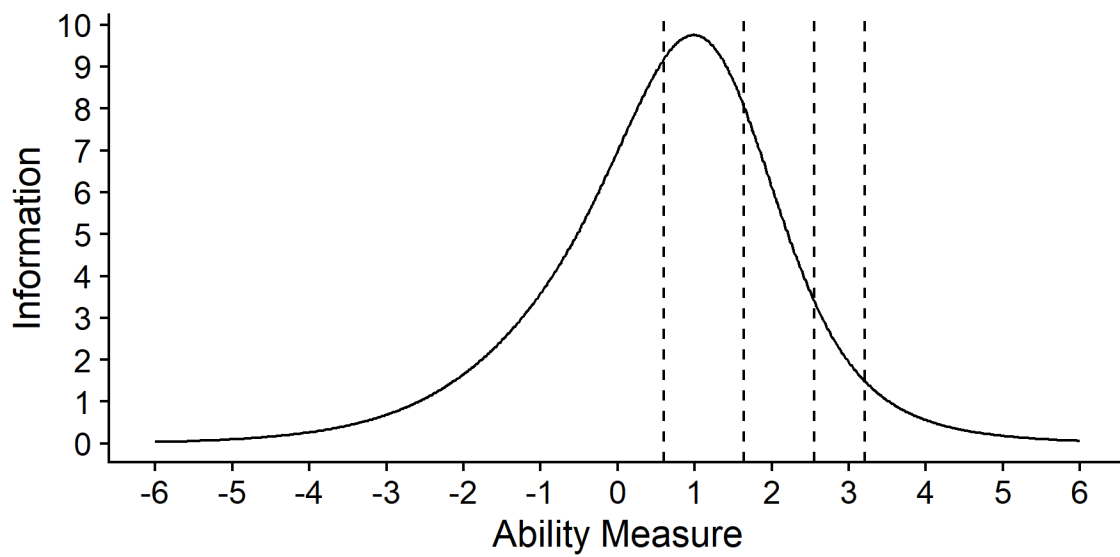


Figure 7.10.3.d.

Test Information Curve: 9–12



7.10.4. Writing Test Information Curves

Figure 7.10.4.a.

Test Information Curve: K-2

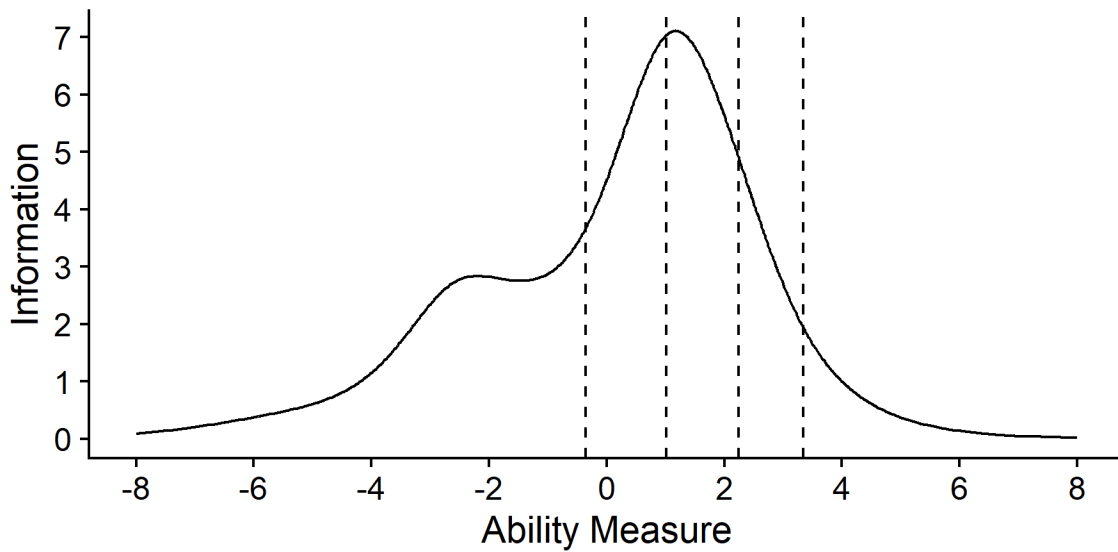


Figure 7.10.4.b.

Test Information Curve: 3-5

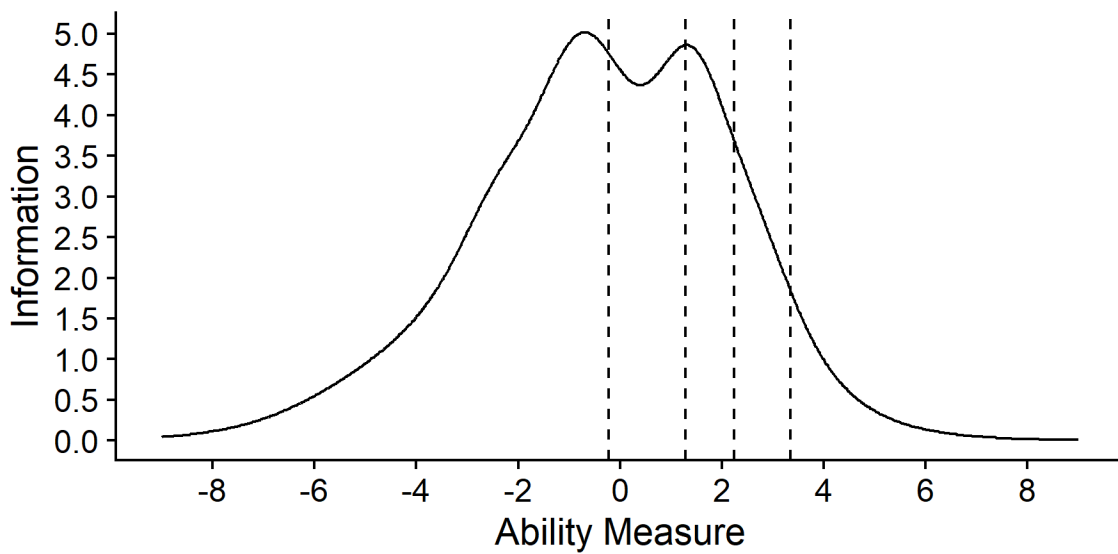


Figure 7.10.4.c.

Test Information Curve: 6–8

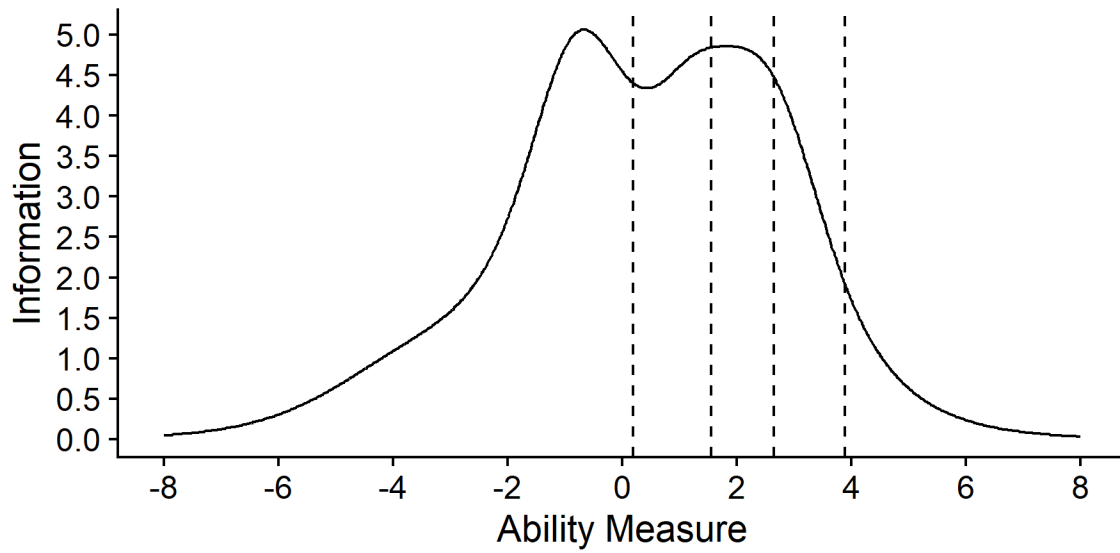
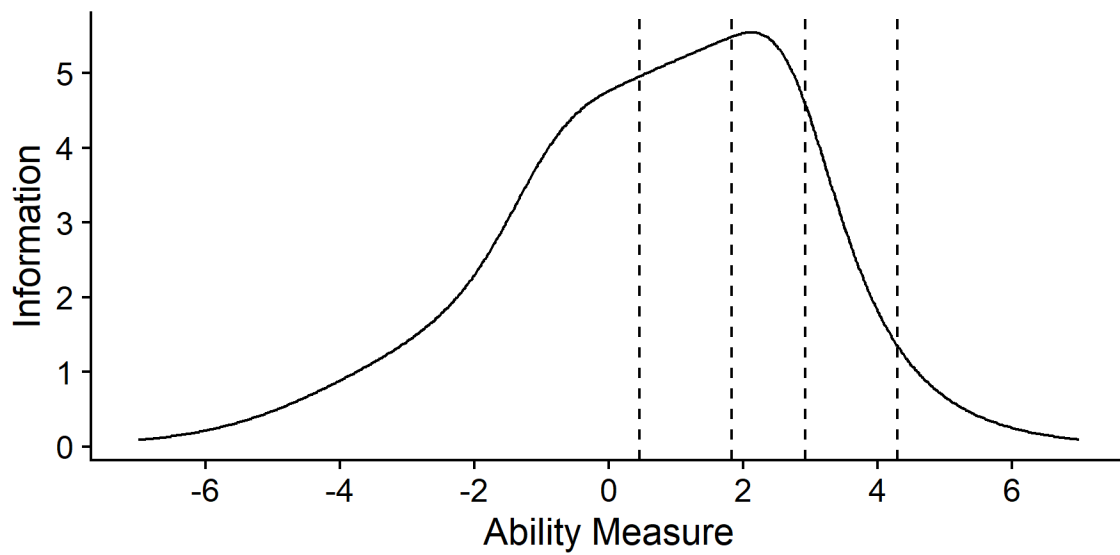


Figure 7.10.4.d.

Test Information Curve: 9–12



8. Analysis of Composite Scores

Alternate ACCESS scores are reported as both scale scores and proficiency level scores for all four language domains. Additionally, four composite scores are reported as Oral, Literacy, Comprehension, and Overall. Raw scores are converted to scale scores through a process known as scaling (see Section 4.4 for details). Scaling ensures that scores are reported on a consistent scale, familiar to test users, and stable across test forms and grade-level clusters. The scale scores range from 900 to 980.

The composite scores are calculated using weighted contributions from each domain, as follows:

- Oral = 50% Speaking + 50% Listening
- Literacy = 50% Reading + 50% Writing
- Comprehension = 70% Reading + 30% Listening
- Overall = 35% Reading + 35% Writing + 15% Listening + 15% Speaking

A policy decision by the WIDA Board, made before the first operational administration of ACCESS, resulted in the weighting, and is based on the view that literacy skills are paramount in developing academic language proficiency.

Scale score distributions and proficiency levels for composite scores are presented in Tables 8.1.1.a. through 8.1.4.d. for scale score distributions and Tables 8.2.1.a. through 8.2.4.d. for proficiency levels. These tables are organized by grade, grade-level cluster, domain, and composite scores. It is important to note that composite scores do not have raw scores associated with them. Therefore, any table or figure that relies on raw scores is not included for composite scores.

8.1. Scale Score Distribution for Composite Scores

Tables 8.1.1.a through 8.1.4.d. provide scale score distributions for each composite across grade-level clusters. The tables include information on grades, the number of students analyzed (count), minimum and maximum observed scale scores, the mean (average) scale score, and the standard deviation of the scale scores. This detailed breakdown helps illustrate the spread and central tendencies of composite scores for each grade-level cluster.

Figures 8.1.1.a. through 8.1.4.d depict the distribution of composite scale scores for each grade-level cluster. The horizontal axis shows the 8 to 10 scale score points, and each bar represents the number of students within each scale score interval. The vertical axis indicates the number of students for each scale score level, providing a visual summary of how students are distributed across the score range.

8.1.1. Oral Composite

Oral composite: Mean scores show consistent growth, increasing from 931.55 in grades K–2 to 948.70 in grades 9–12. This reflects expected developmental progress in oral proficiency. Standard deviations remain relatively stable, ranging from 18.23 to 19.25, indicating similar variability across clusters. The score range is uniform, with minimum scores at 900 and maximums near 980. Score distributions shift with grade levels. K–2 is right-skewed, with more students scoring at the lower end. By 3–5, the distribution evens out, and in 6–8 and 9–12, scores cluster more toward the higher end.

Table 8.1.1.a.

Scale Score Descriptive Statistics: Oral K–2

Grade	Number of Students	Min.	Max.	Mean	SD
K	2,651	900	974	925.02	18.09
1	3,767	900	974	931.65	19.06
2	3,424	900	974	936.51	18.83
Total	9,842	900	974	931.55	19.25

Table 8.1.1.b.

Scale Score Descriptive Statistics: Oral 3–5

Grade	Number of Students	Min.	Max.	Mean	SD
3	3,000	902	976	939.64	18.31
4	2,959	902	976	941.99	18.55
5	2,640	902	976	943.55	18.52
Total	8,599	902	976	941.65	18.52

Table 8.1.1.c.

Scale Score Descriptive Statistics: Oral 6–8

Grade	Number of Students	Min.	Max.	Mean	SD
6	2394	904	978	946.33	18.23
7	2119	904	978	947.67	17.93
8	1998	904	978	948.02	18.50
Total	6511	904	978	947.29	18.23

Table 8.1.1.d.

Scale Score Descriptive Statistics: Oral 9–12

Grade	Number of Students	Min.	Max.	Mean	SD
9	1,908	906	980	947.36	19.57
10	1,764	906	980	949.68	19.27
11	1,626	906	980	949.13	19.14
12	2,555	906	980	948.76	18.91
Total	7,853	906	980	948.70	19.21

Figure 8.1.1.a.

Scale Score Distribution: Oral K-2

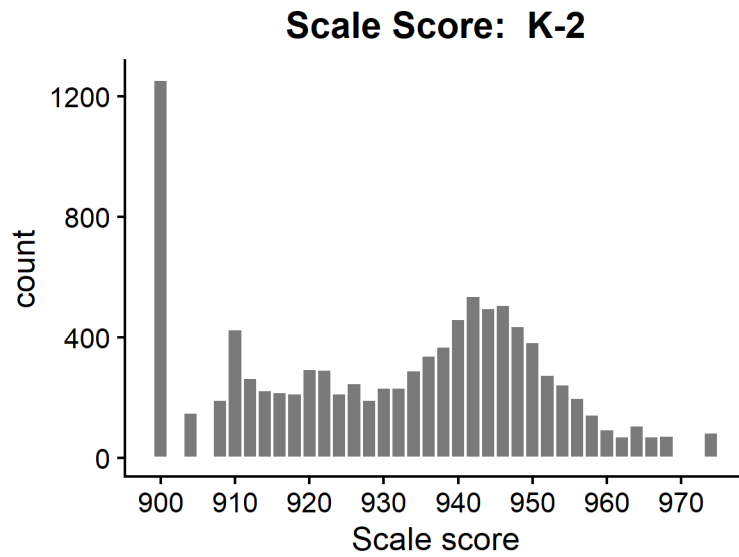


Figure 8.1.1.b.

Scale Score Distribution: Oral 3-5

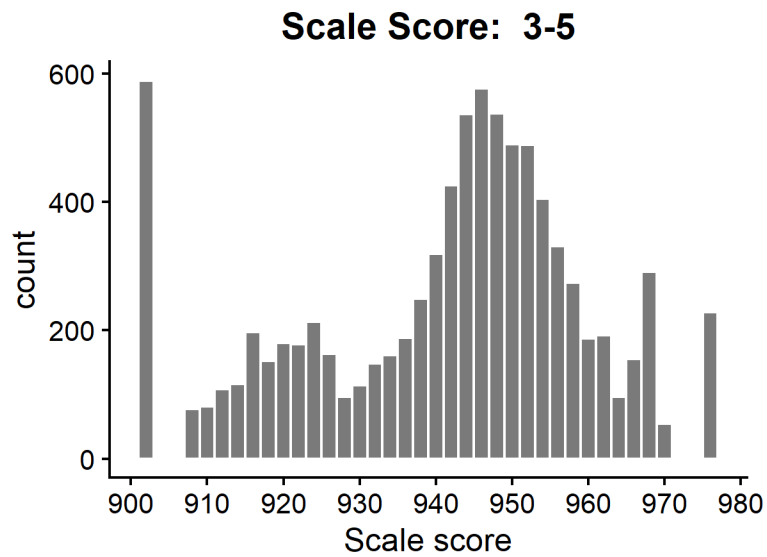


Figure 8.1.1.c.

Scale Score Distribution: Oral 6–8

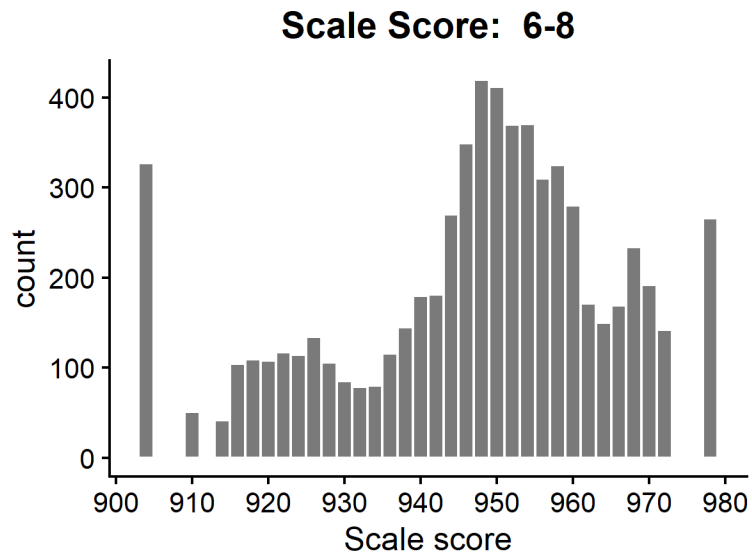
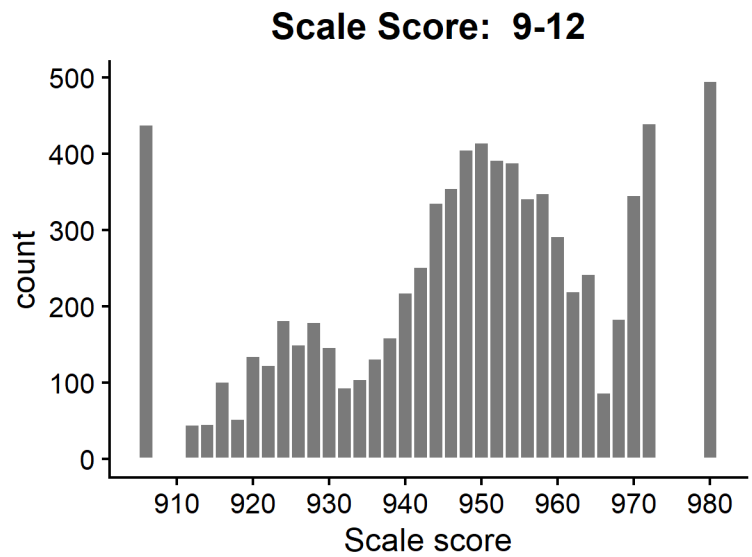


Figure 8.1.1.d.

Scale Score Distribution: Oral 9–12



8.1.2. Literacy Composite

Literacy composite: The mean scale scores show consistent growth across grade-level clusters, beginning at 930.23 for grades K–2 and increasing to 946.61 for grades 9–12. Within each cluster, steady progression is observed; for instance, in grades 3–5, the mean increases from 935.35 in grade 3 to 940.29 in grade 5, while in grades 6–8, it rises from 942.75 in grade 6 to 945.35 in grade 8. The scale score distributions, depicted in Figures 8.1.2.a. through 8.1.2.d., reveal how scores shift across grade-level clusters. In grades K–2, the distribution is heavily right-skewed, with many students scoring near the minimum value (900). As grades progress, the distributions become more symmetrical, peaking closer to the middle and upper ranges. By grades 9–12, the scores are concentrated near the upper end of the scale (around 950–960).

Table 8.1.2.a.

Scale Score Descriptive Statistics: Lit K–2

Grade	Number of Students	Min.	Max.	Mean	SD
K	2,654	900	974	922.14	17.92
1	3,770	900	974	930.45	20.25
2	3,426	900	974	936.25	20.85
Total	9,850	900	974	930.23	20.61

Table 8.1.2.b.

Scale Score Descriptive Statistics: Lit 3–5

Grade	Number of Students	Min.	Max.	Mean	SD
3	2,994	902	976	935.35	18.17
4	2,959	902	976	938.09	18.87
5	2,639	902	976	940.29	19.19
Total	8,592	902	976	937.81	18.84

Table 8.1.2.c.

Scale Score Descriptive Statistics: Lit 6–8

Grade	Number of Students	Min.	Max.	Mean	SD
6	2,394	904	978	942.75	18.76
7	2,120	904	978	944.29	18.74
8	1,999	904	978	945.35	19.64
Total	6,513	904	978	944.05	19.06

Table 8.1.2.d.

Scale Score Descriptive Statistics: Lit 9–12

Grade	Number of Students	Min.	Max.	Mean	SD
9	1,908	906	980	945.17	19.88
10	1,766	906	980	947.19	19.76
11	1,626	906	980	946.91	20.25
12	2,554	906	980	947.11	19.96
Total	7,854	906	980	946.61	19.97

Figure 8.1.2.a.

Scale Score Distribution: Lit K–2

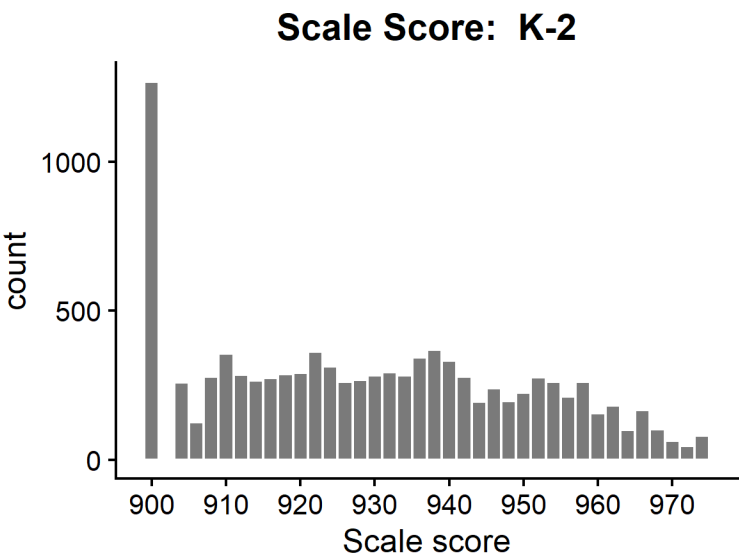


Figure 8.1.2.b.

Scale Score Distribution: Lit 3–5

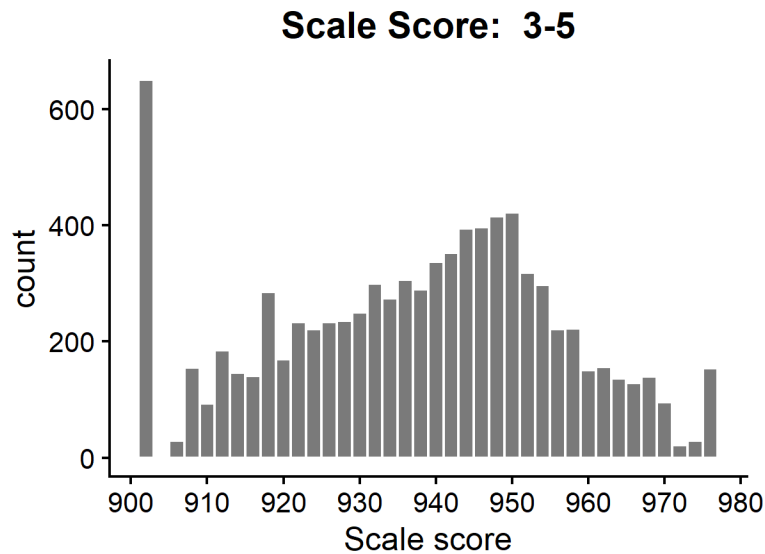


Figure 8.1.2.c.

Scale Score Distribution: Lit 6–8

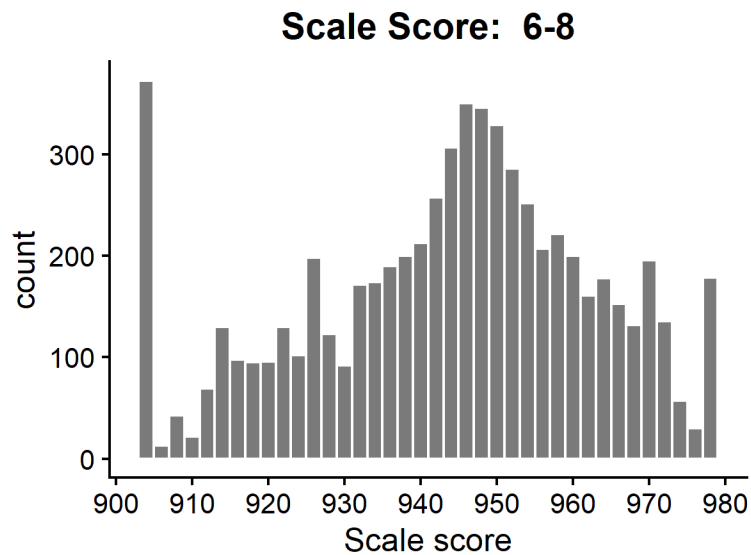
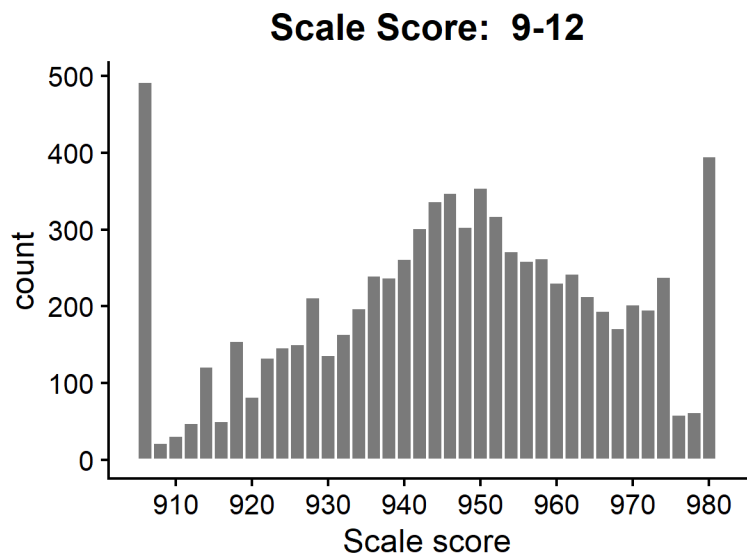


Figure 8.1.2.d.

Scale Score Distribution: Lit 9–12



8.1.3. Comprehension Composite

Comprehension composite: The mean scale scores show consistent growth across grade-level clusters, beginning at 934.58 for grades K–2 and increasing to 950.53 for grades 9–12. Within each cluster, steady progression is observed; for example, in grades 3–5, the mean rises from 940.28 in grade 3 to 945.07 in grade 5, while in grades 6–8, it increases from 947.42 in grade 6 to 949.93 in grade 8. The scale score distributions, depicted in Figures 8.1.3.a. through 8.1.3.d., illustrate how scores shift across grade-level clusters. In grades K–2, the distribution is heavily right-skewed, with many students scoring at the minimum value (900). As grades progress, the distributions become more symmetrical, peaking closer to the middle and upper ranges. By grades 9–12, scores are concentrated at the upper end of the scale (around 950–960).

Table 8.1.3.a.

Scale Score Descriptive Statistics: Comp K–2

Grade	Number of Students	Min.	Max.	Mean	SD
K	2,656	900	974	927.08	19.75
1	3,772	900	974	934.84	20.21
2	3,427	900	974	940.10	19.96
Total	9,855	900	974	934.58	20.63

Table 8.1.3.b.

Scale Score Descriptive Statistics: Comp 3–5

Grade	Number of Students	Min.	Max.	Mean	SD
3	2,998	902	976	940.28	17.54
4	2,963	902	976	943.01	17.83
5	2,639	902	976	945.00	17.67
Total	8,600	902	976	942.67	17.78

Table 8.1.3.c.

Scale Score Descriptive Statistics: Comp 6–8

Grade	Number of Students	Min.	Max.	Mean	SD
6	2,398	904	978	947.42	17.35
7	2,122	904	978	949.04	17.40
8	2,001	904	978	949.93	17.93
Total	6,521	904	978	948.72	17.58

Table 8.1.3.d.

Scale Score Descriptive Statistics: Comp 9–12

Grade	Number of Students	Min.	Max.	Mean	SD
9	1,910	906	980	949.34	18.44
10	1,766	906	980	951.32	18.18
11	1,626	906	980	950.87	18.24
12	2,558	906	980	950.66	18.31
Total	7,860	906	980	950.53	18.31

Figure 8.1.3.a.

Scale Score Distribution: Comp K-2

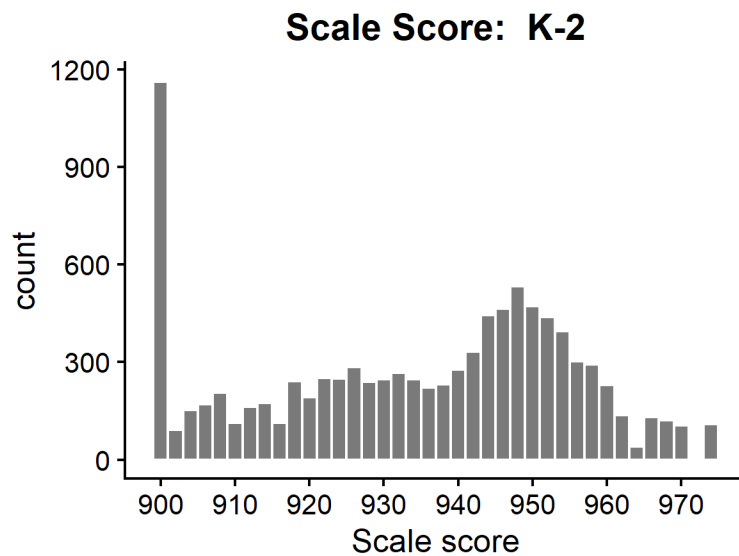


Figure 8.1.3.b.

Scale Score Distribution: Comp 3-5

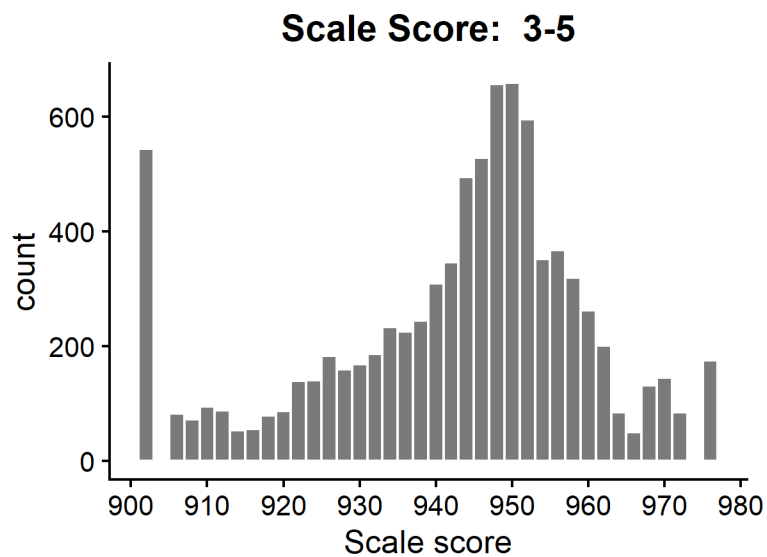


Figure 8.1.3.c.

Scale Score Distribution: Comp 6-8

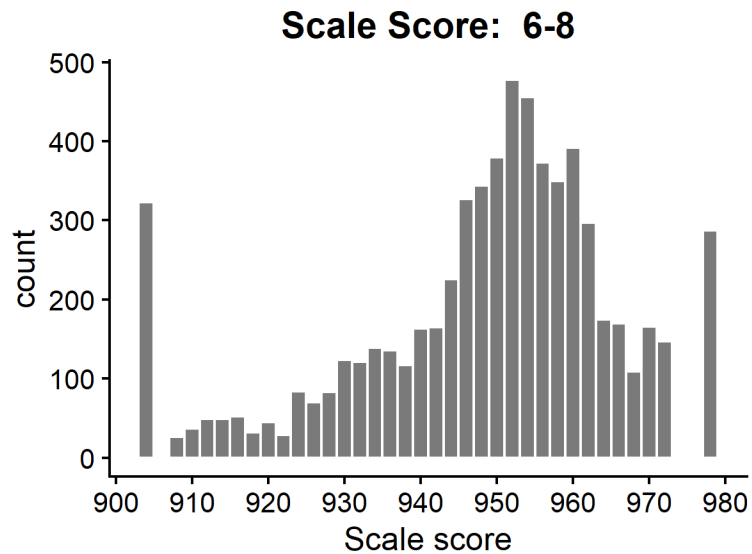
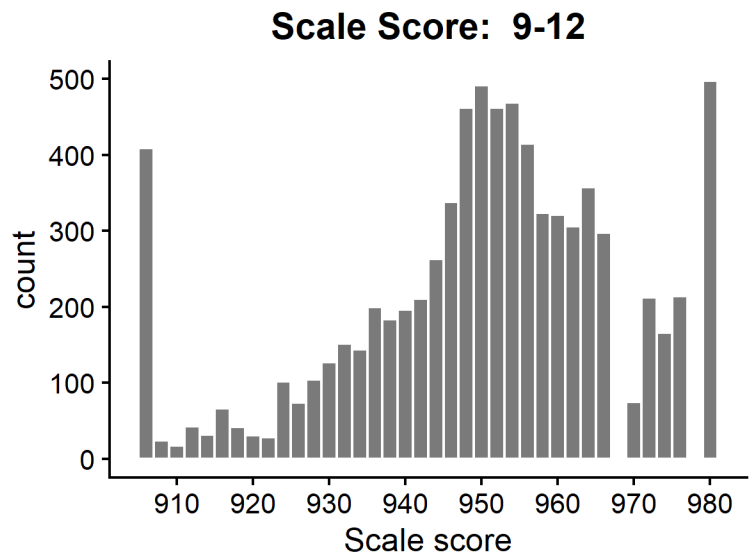


Figure 8.1.3.d.

Scale Score Distribution: Comp 9-12



8.1.4. Overall Composite

Overall composite: The mean scale scores for the overall composite indicate consistent growth across grade clusters, starting at 930.43 for grades K–2 and increasing to 947.04 for grades 9–12. Within each cluster, a steady progression is observed. For instance, in grades 3–5, the mean score increases from 936.43 in grade 3 to 941.08 in grade 5. Similarly, in grades 6–8, the mean score rises from 943.62 in grade 6 to 945.96 in grade 8. The scale score distributions, depicted in Figures 8.1.4.a. through 8.1.4.d., highlight how scores shift across grade clusters. In grades K–2, the distribution is heavily right-skewed, with a concentration of students scoring near the minimum value (900). As grades progress, the distributions become more symmetrical and centered, peaking closer to the middle and upper ranges. By grades 9–12, the distributions are more concentrated toward the upper end of the scale (around 940–960), reflecting higher overall proficiency among older students.

Table 8.1.4.a.

Scale Score Descriptive Statistics: Overall K–2

Grade	Number of Students	Min.	Max.	Mean	SD
K	2,649	900	974	922.81	17.31
1	3,764	900	974	930.61	19.23
2	3,423	900	974	936.13	19.58
Total	9,836	900	974	930.43	19.56

Table 8.1.4.b.

Scale Score Descriptive Statistics: Overall 3–5

Grade	Number of Students	Min.	Max.	Mean	SD
3	2,993	902	976	936.43	17.58
4	2,957	902	976	939.06	18.13
5	2,638	902	976	941.08	18.34
Total	8,588	902	976	938.76	18.11

Table 8.1.4.c.

Scale Score Descriptive Statistics: Overall 6–8

Grade	Number of Students	Min.	Max.	Mean	SD
6	2,393	904	978	943.62	18.00
7	2,118	904	978	945.11	17.89
8	1,997	904	978	945.96	18.74
Total	6,508	904	978	944.82	18.22

Table 8.1.4.d.

Scale Score Descriptive Statistics: Overall 9–12

Grade	Number of Students	Min.	Max.	Mean	SD
9	1,906	906	980	945.63	19.27
10	1,764	906	980	947.75	19.09
11	1,626	906	980	947.37	19.36
12	2,552	906	980	947.40	19.14
Total	7,848	906	980	947.04	19.22

Figure 8.1.4.a.

Scale Score Distribution: Overall K–2

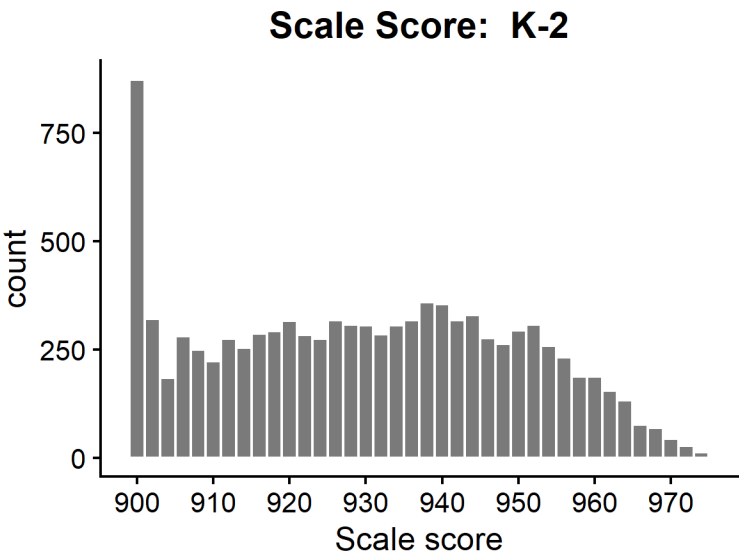


Figure 8.1.4.b.

Scale Score Distribution: Overall 3–5

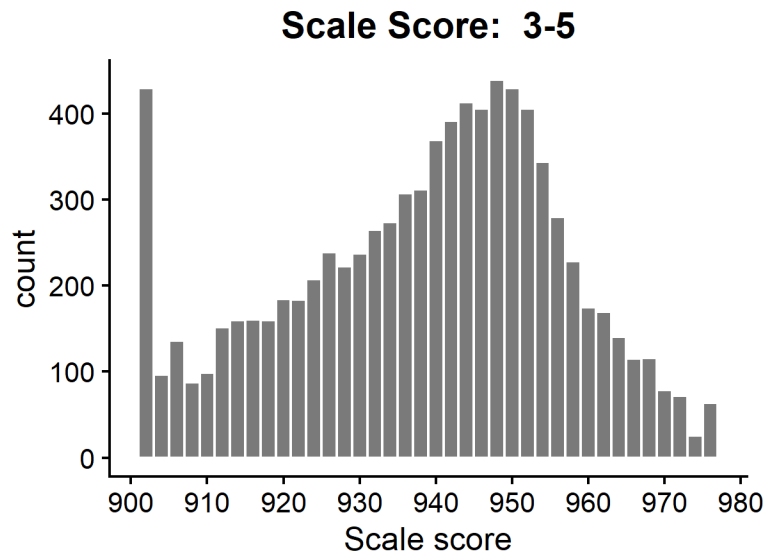


Figure 8.1.4.c.

Scale Score Distribution: Overall 6–8

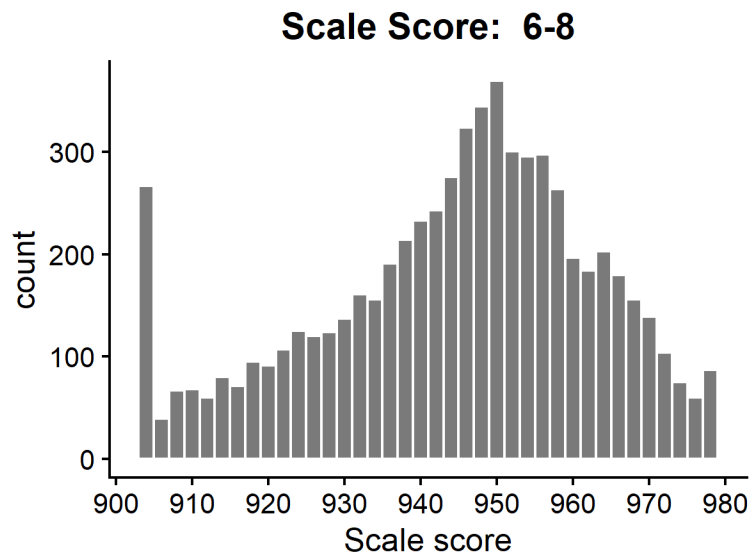
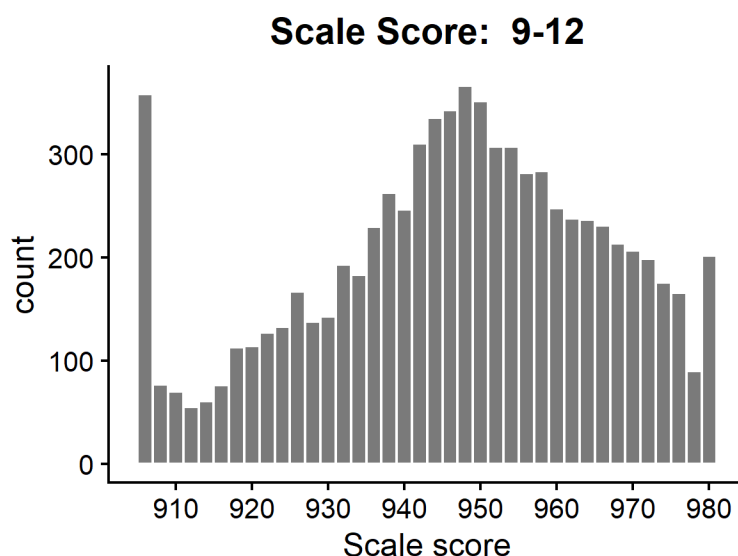


Figure 8.1.4.d.

Scale Score Distribution: Overall 9–12



8.2. *Proficiency Level Distribution for Composite Scores*

Tables 8.2.1.a. through 8.2.4.d. provide proficiency level information for each composite score across all grade-level clusters. These tables present data by individual grade and as a total for the grade-level cluster. Specifically, they include the WIDA proficiency level designations (P1–P5), the number of students whose performance placed them into each proficiency level for the domain being tested, and the percentage of students, out of the total number taking the test, who were placed into each proficiency level for the domain being tested.

Figures 8.2.1.a. through 8.2.4.d. illustrate the proficiency level distribution for each composite score across the grade-level clusters. In each figure, the horizontal axis represents the five WIDA proficiency levels, while the vertical axis shows the percentage of students. Each bar indicates the percentage of students assigned to each proficiency level within the tested domain for the specific test form.

8.2.1. Oral Composite

Oral Composite: Across all grade-level clusters, the majority of students are placed in P1 and P2 levels. In K–2, nearly 70% of students are in P1, with a gradual decline in this proportion as grades increase. By grades 9–12, a more balanced distribution across P2 to P4 emerges, though P1 remains the most frequent.

Table 8.2.1.a.**Proficiency Level Distribution: Oral K–2**

Level	Grade K Count	Grade K Percent	Grade 1 Count	Grade 1 Percent	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	1,858	70.09%	2,080	55.22%	1,539	44.95%	5,477	55.65%
2	411	15.50%	652	17.31%	607	17.73%	1,670	16.97%
3	281	10.60%	633	16.80%	690	20.15%	1,604	16.30%
4	63	2.38%	244	6.48%	319	9.32%	626	6.36%
5	38	1.43%	158	4.19%	269	7.86%	465	4.72%
Total	2,651	100.0%	3,767	100.0%	3,424	100.0%	9,842	100.0%

Table 8.2.1.b.**Proficiency Level Distribution: Oral 3–5**

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	1,361	45.37%	1,194	40.35%	973	36.86%	3,528	41.03%
2	768	25.60%	720	24.33%	637	24.13%	2,125	24.71%
3	429	14.30%	454	15.34%	417	15.80%	1,300	15.12%
4	222	7.40%	253	8.55%	259	9.81%	734	8.54%
5	220	7.33%	338	11.42%	354	13.41%	912	10.61%
Total	3,000	100.0%	2,959	100.0%	2,640	100.0%	8,599	100.0%

Table 8.2.1.c.**Proficiency Level Distribution: Oral 6–8**

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	852	35.59%	692	32.66%	652	32.63%	2,196	33.73%
2	516	21.55%	426	20.10%	393	19.67%	1,335	20.50%
3	492	20.55%	472	22.27%	412	20.62%	1,376	21.13%
4	158	6.60%	140	6.61%	153	7.66%	451	6.93%
5	376	15.71%	389	18.36%	388	19.42%	1,153	17.71%
Total	2,394	100.0%	2,119	100.0%	1,998	100.0%	6,511	100.0%

Table 8.2.1.d.

Proficiency Level Distribution: Oral 9–12

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	763	39.99%	634	35.94%	581	35.73%	919	35.97%	2,897	36.89%
2	325	17.03%	285	16.16%	279	17.16%	488	19.10%	1,377	17.53%
3	300	15.72%	288	16.33%	270	16.61%	413	16.16%	1,271	16.18%
4	167	8.75%	168	9.52%	166	10.21%	254	9.94%	755	9.61%
5	353	18.50%	389	22.05%	330	20.30%	481	18.83%	1,553	19.78%
Total	1,908	100.0%	1,764	100.0%	1,626	100.0%	2,555	100.0%	7,853	100.0%

Figure 8.2.1.a.

Proficiency Level Distribution: Oral K–2

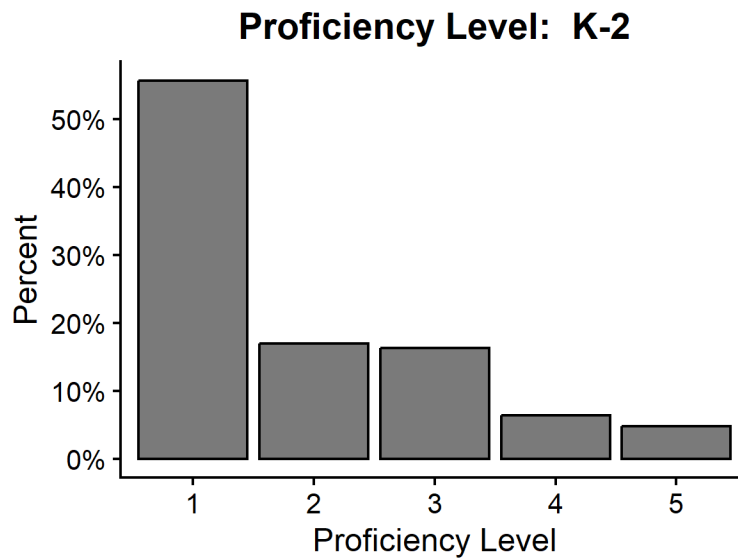


Figure 8.2.1.b.

Proficiency Level Distribution: Oral 3-5

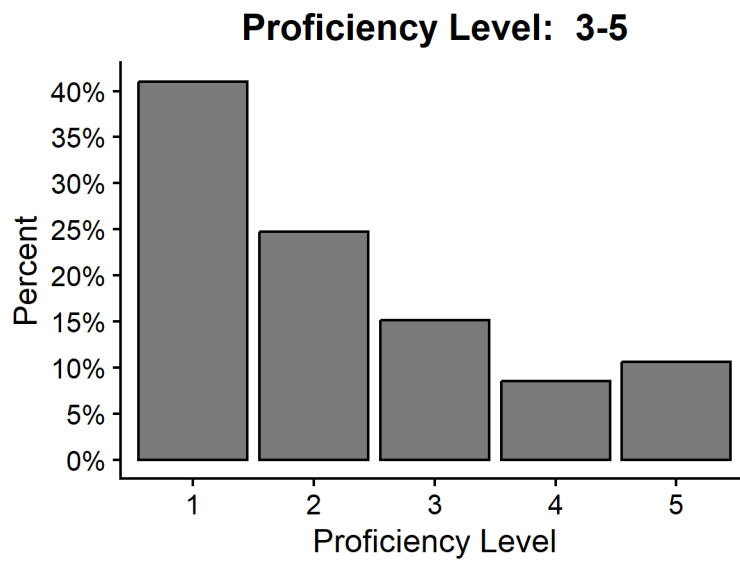


Figure 8.2.1.c.

Proficiency Level Distribution: Oral 6-8

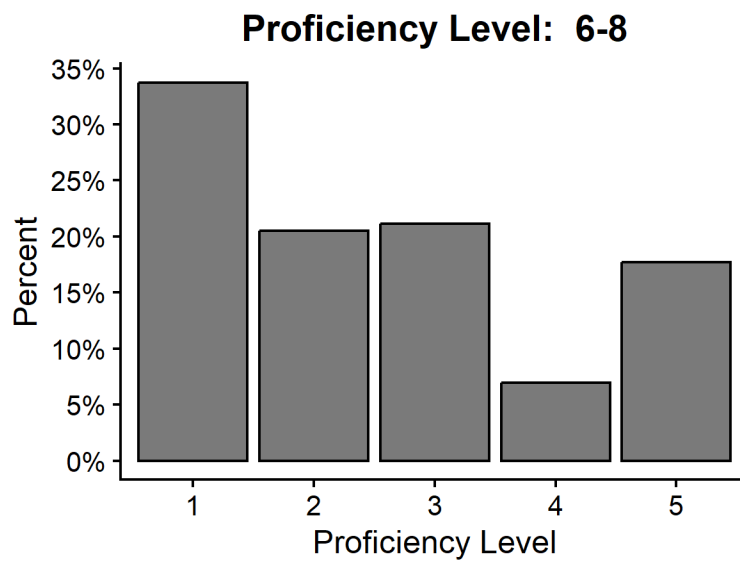
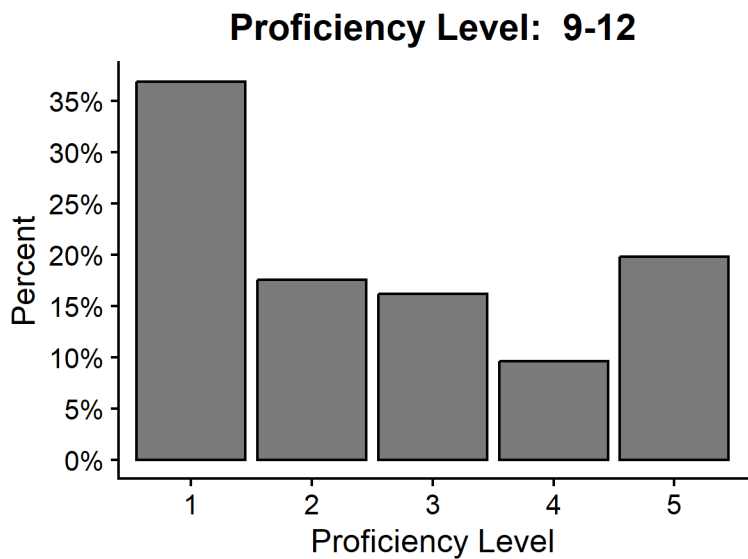


Figure 8.2.1.d.

Proficiency Level Distribution: Oral 9–12



8.2.2. Literacy Composite

Literacy Composite: In grades K–2, over 80% of students are in P1, indicating that most are at the beginning levels of literacy proficiency. As grade levels increase, there is a notable shift toward P3 and P4. By grades 9–12, P3 and P4 represent a substantial portion of the distribution, reflecting steady improvement in literacy proficiency with grade progression.

Table 8.2.2.a.

Proficiency Level Distribution: Lit K–2

Level	Grade K Count	Grade K Percent	Grade 1 Count	Grade 1 Percent	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	2,233	84.14%	2,599	68.94%	1,975	57.65%	6,807	69.11%
2	204	7.69%	398	10.56%	407	11.88%	1,009	10.24%
3	131	4.94%	401	10.64%	456	13.31%	988	10.03%
4	49	1.85%	216	5.73%	319	9.31%	584	5.93%
5	37	1.39%	156	4.14%	269	7.85%	462	4.69%
Total	2,654	100.0%	3,770	100.0%	3,426	100.0%	9,850	100.0%

Table 8.2.2.b.**Proficiency Level Distribution: Lit 3–5**

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	1,794	59.92%	1,606	54.28%	1,311	49.68%	4,711	54.83%
2	633	21.14%	631	21.32%	545	20.65%	1,809	21.05%
3	312	10.42%	317	10.71%	327	12.39%	956	11.13%
4	148	4.94%	226	7.64%	222	8.41%	596	6.94%
5	107	3.57%	179	6.05%	234	8.87%	520	6.05%
Total	2,994	100.0%	2,959	100.0%	2,639	100.0%	8,592	100.0%

Table 8.2.2.c.**Proficiency Level Distribution: Lit 6–8**

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	1,161	48.50%	928	43.77%	845	42.27%	2,934	45.05%
2	512	21.39%	440	20.75%	412	20.61%	1,364	20.94%
3	265	11.07%	303	14.29%	225	11.26%	793	12.18%
4	272	11.36%	272	12.83%	280	14.01%	824	12.65%
5	184	7.69%	177	8.35%	237	11.86%	598	9.18%
Total	2,394	100.0%	2,120	100.0%	1,999	100.0%	6,513	100.0%

Table 8.2.2.d.**Proficiency Level Distribution: Lit 9–12**

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	898	47.06%	775	43.88%	709	43.60%	1143	44.75%	3525	44.88%
2	330	17.30%	303	17.16%	271	16.67%	421	16.48%	1325	16.87%
3	233	12.21%	218	12.34%	187	11.50%	270	10.57%	908	11.56%
4	258	13.52%	240	13.59%	255	15.68%	391	15.31%	1144	14.57%
5	189	9.91%	230	13.02%	204	12.55%	329	12.88%	952	12.12%
Total	1908	100.0%	1766	100.0%	1626	100.0%	2554	100.0%	7854	100.0%

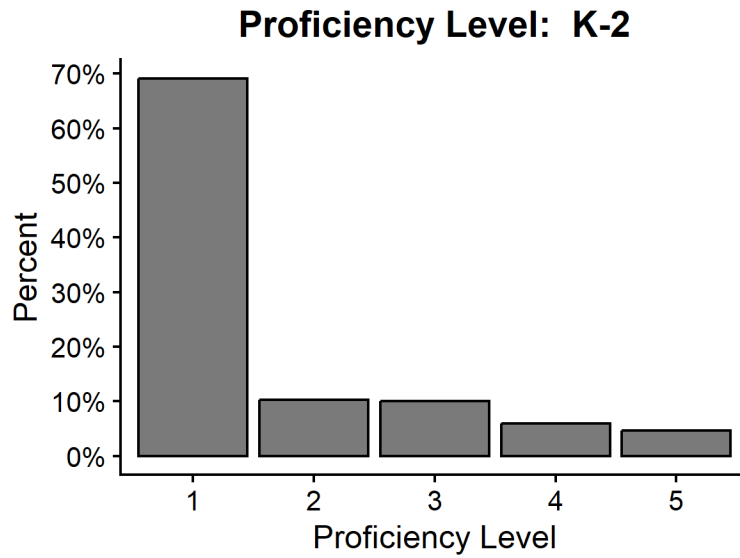


Figure 8.2.2.b.

Proficiency Level Distribution: Lit 3-5

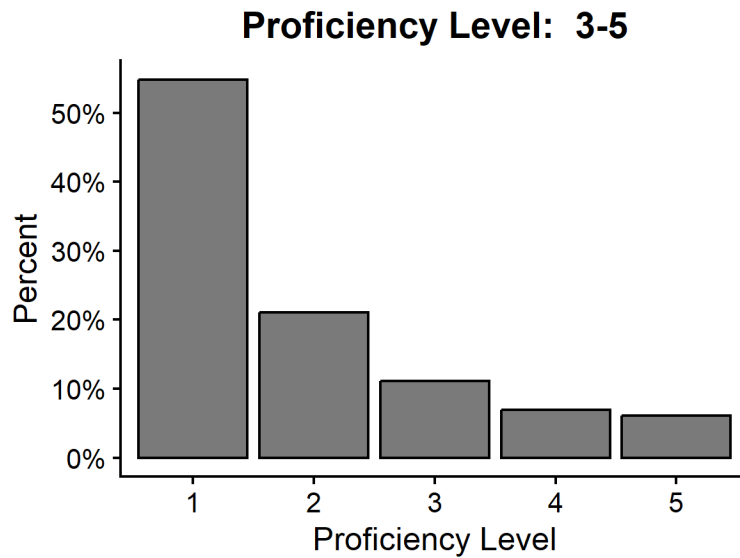


Figure 8.2.2.c.

Proficiency Level Distribution: Lit 6–8

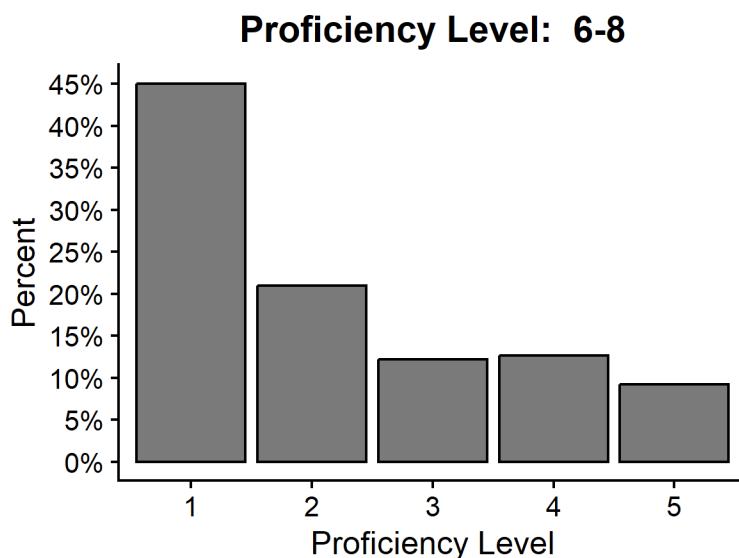
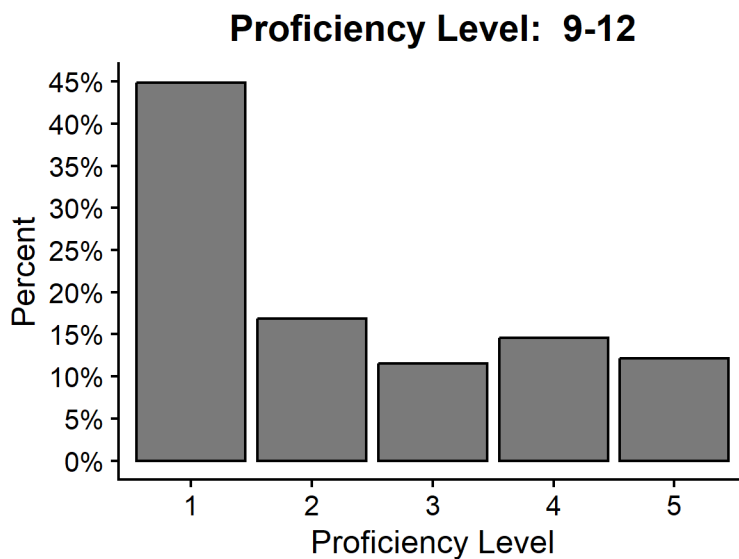


Figure 8.2.2.d.

Proficiency Level Distribution: Lit 9–12



8.2.3. Comprehension Composite

Comprehension Composite: For grades K–2, more than two-thirds of students are at P1, gradually decreasing as grade clusters advance. The proportion of students in P3 increases significantly by grades 6–8 and 9–12, with a relatively smaller percentage reaching P4 or P5.

Table 8.2.3.a.**Proficiency Level Distribution: Comp K-2**

Level	Grade K Count	Grade K Percent	Grade 1 Count	Grade 1 Percent	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	1,800	67.77%	1,938	51.38%	1,399	40.82%	5,137	52.13%
2	350	13.18%	561	14.87%	496	14.47%	1,407	14.28%
3	312	11.75%	691	18.32%	672	19.61%	1,675	17.00%
4	133	5.01%	364	9.65%	493	14.39%	990	10.05%
5	61	2.30%	218	5.78%	367	10.71%	646	6.56%
Total	2,656	100.0%	3,772	100.0%	3,427	100.0%	9,855	100.0%

Table 8.2.3.b.**Proficiency Level Distribution: Comp 3-5**

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	1,257	41.93%	1,042	35.17%	844	31.98%	3,143	36.55%
2	609	20.31%	597	20.15%	483	18.30%	1,689	19.64%
3	658	21.95%	675	22.78%	614	23.27%	1,947	22.64%
4	334	11.14%	407	13.74%	410	15.54%	1,151	13.38%
5	140	4.67%	242	8.17%	288	10.91%	670	7.79%
Total	2,998	100.0%	2,963	100.0%	2,639	100.0%	8,600	100.0%

Table 8.2.3.c.**Proficiency Level Distribution: Comp 6-8**

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	743	30.98%	564	26.58%	537	26.84%	1,844	28.28%
2	338	14.10%	298	14.04%	261	13.04%	897	13.76%
3	556	23.19%	509	23.99%	437	21.84%	1,502	23.03%
4	493	20.56%	474	22.34%	432	21.59%	1,399	21.45%
5	268	11.18%	277	13.05%	334	16.69%	879	13.48%
Total	2,398	100.0%	2,122	100.0%	2,001	100.0%	6,521	100.0%

Table 8.2.3.d.

Proficiency Level Distribution: Comp 9–12

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	546	28.59%	472	26.73%	439	27.00%	729	28.50%	2,186	27.81%
2	271	14.19%	227	12.85%	202	12.42%	363	14.19%	1,063	13.52%
3	466	24.40%	409	23.16%	387	23.80%	576	22.52%	1,838	23.38%
4	342	17.91%	322	18.23%	312	19.19%	427	16.69%	1,403	17.85%
5	285	14.92%	336	19.03%	286	17.59%	463	18.10%	1,370	17.43%
Total	1,910	100.0%	1,766	100.0%	1,626	100.0%	2,558	100.0%	7,860	100.0%

Figure 8.2.3.a.

Proficiency Level Distribution: Comp K–2

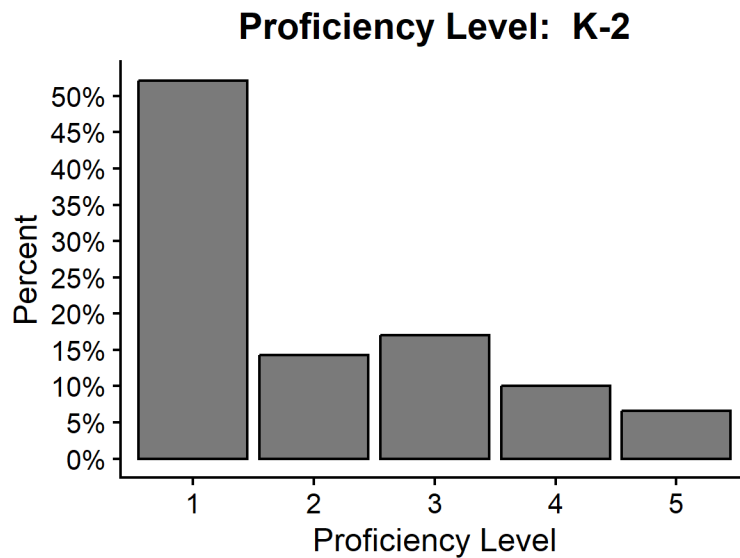


Figure 8.2.3.b.

Proficiency Level Distribution: Comp 3-5

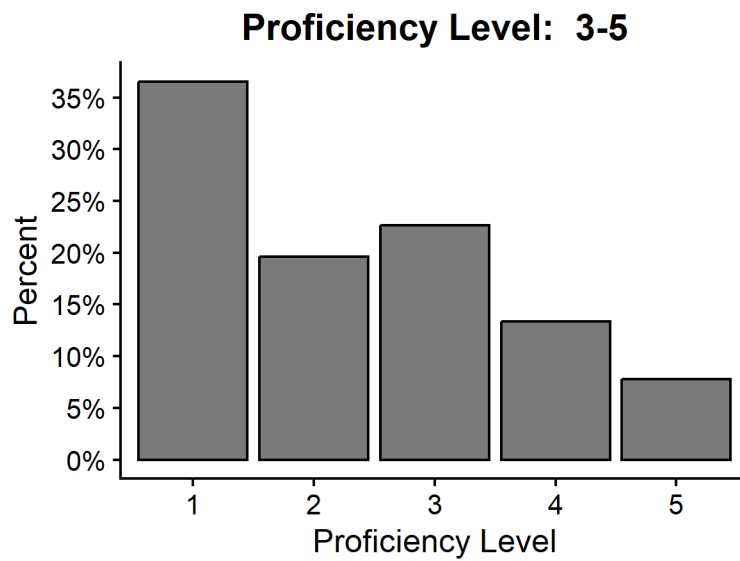


Figure 8.2.3.c.

Proficiency Level Distribution: Comp 6-8

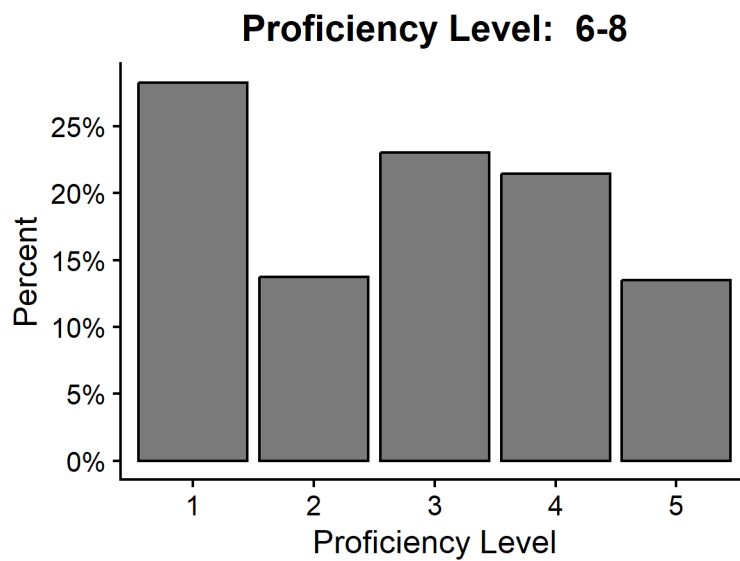


Figure 8.2.3.d.

Proficiency Level Distribution: Comp 9–12



8.2.4. Overall Composite

Overall Composite: A similar trend is observed in the Overall composite, with most students in P1 during grades K–2. By grades 9–12, the proportion of students in P2, P3, and P4 increases, reflecting overall growth across all domains. The percentage of students at P5 remains low across all grade clusters, indicating room for further development.

Table 8.2.4.a.

Proficiency Level Distribution: Overall K–2

Level	Grade K Count	Grade K Percent	Grade 1 Count	Grade 1 Percent	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	2,156	81.39%	2,473	65.70%	1,842	53.81%	6,471	65.79%
2	264	9.97%	489	12.99%	495	14.46%	1,248	12.69%
3	152	5.74%	445	11.82%	510	14.90%	1,107	11.25%
4	56	2.11%	230	6.11%	360	10.52%	646	6.57%
5	21	0.79%	127	3.37%	216	6.31%	364	3.70%
Total	2,649	100.0%	3,764	100.0%	3,423	100.0%	9,836	100.0%

Table 8.2.4.b.**Proficiency Level Distribution: Overall 3–5**

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	1,712	57.20%	1,510	51.07%	1,244	47.16%	4,466	52.00%
2	603	20.15%	595	20.12%	505	19.14%	1,703	19.83%
3	400	13.36%	425	14.37%	407	15.43%	1,232	14.35%
4	164	5.48%	234	7.91%	235	8.91%	633	7.37%
5	114	3.81%	193	6.53%	247	9.36%	554	6.45%
Total	2,993	100.0%	2,957	100.0%	2,638	100.0%	8,588	100.0%

Table 8.2.4.c.**Proficiency Level Distribution: Overall 6–8**

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	1,083	45.26%	873	41.22%	815	40.81%	2,771	42.58%
2	456	19.06%	397	18.74%	340	17.03%	1,193	18.33%
3	409	17.09%	410	19.36%	340	17.03%	1,159	17.81%
4	252	10.53%	253	11.95%	259	12.97%	764	11.74%
5	193	8.07%	185	8.73%	243	12.17%	621	9.54%
Total	2,393	100.0%	2,118	100.0%	1,997	100.0%	6,508	100.0%

Table 8.2.4.d.**Proficiency Level Distribution: Overall 9–12**

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	861	45.17%	741	42.01%	684	42.07%	1110	43.50%	3,396	43.27%
2	302	15.84%	279	15.82%	243	14.94%	394	15.44%	1,218	15.52%
3	296	15.53%	260	14.74%	252	15.50%	332	13.01%	1,140	14.53%
4	239	12.54%	232	13.15%	222	13.65%	361	14.15%	1,054	13.43%
5	208	10.91%	252	14.29%	225	13.84%	355	13.91%	1,040	13.25%
Total	1,906	100.0%	1,764	100.0%	1,626	100.0%	2,552	100.0%	7,848	100.0%

Figure 8.2.4.a.

Proficiency Level Distribution: Overall K-2

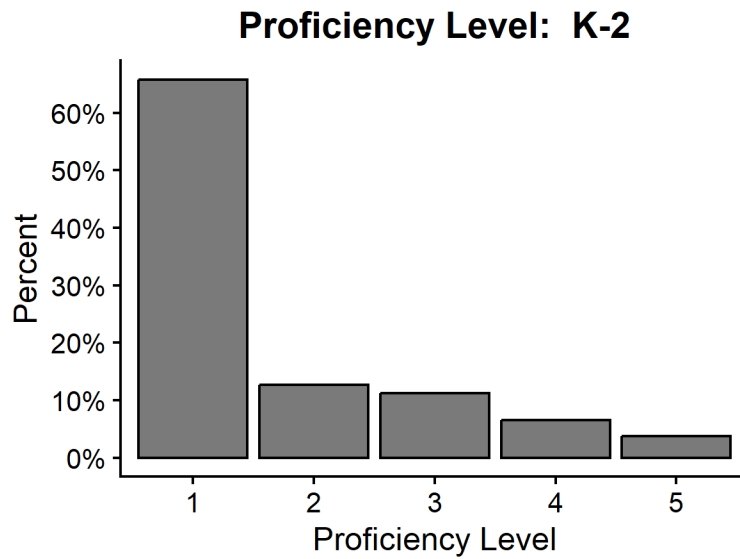


Figure 8.2.4.b.

Proficiency Level Distribution: Overall 3-5

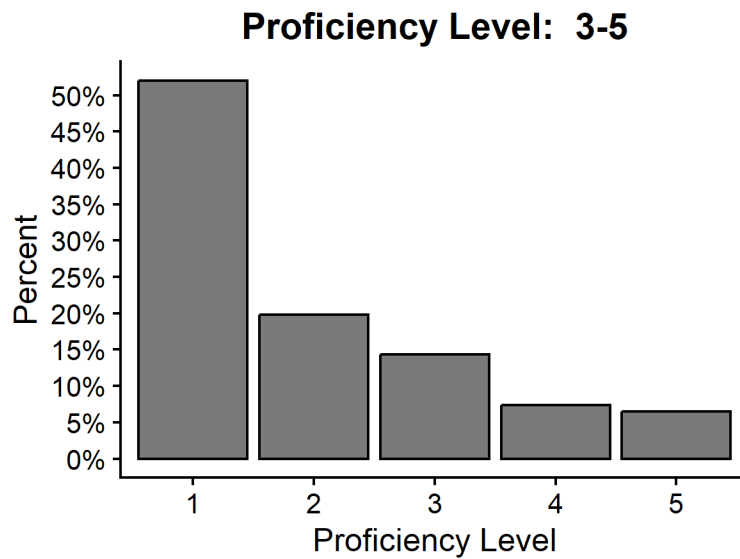


Figure 8.2.4.c.

Proficiency Level Distribution: Overall 6–8

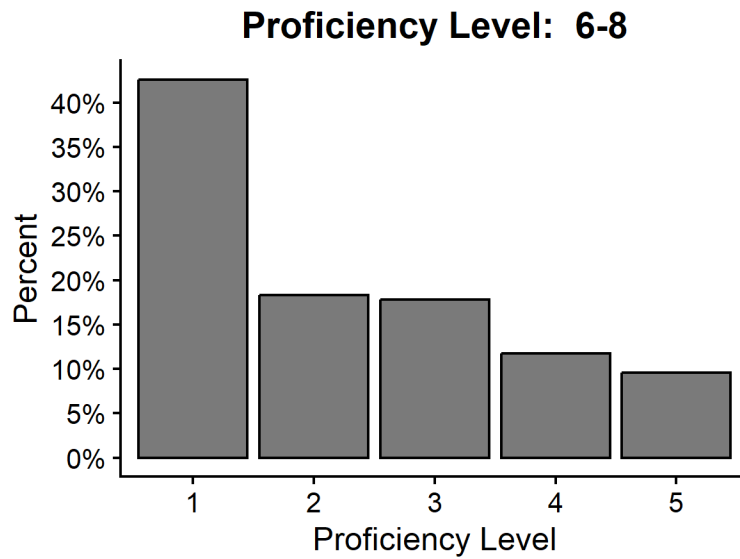
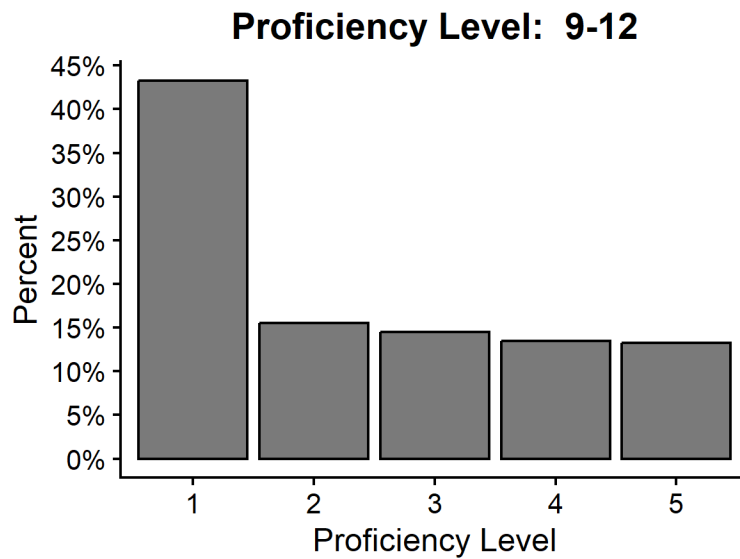


Figure 8.2.4.d.

Proficiency Level Distribution: Overall 9–12



9. Annual Updates of Validity Evidence

This section presents studies conducted as validity evidence for the WIDA Alternate ACCESS assessment. According to the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), validity is the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed use. Particular interpretations for specified uses begin by specifying the construct the test is intended to measure. Rather than referring to distinct types of validity, the aforementioned *Standards* refer to types of validity evidence. According to the *Standards*, the evidence can be based on (1) test content, (2) response processes, (3) internal structure, and (4) relation to other variables.

The validity evidence of the Standards is also observed in “A State’s Guide to the U.S. Department of Education’s Assessment Peer Review Process” document (Department of Education, 2018) to support states’ use of ELP assessments for reviewing validity evidence. It is also linked to the Assessment Use Argument (AUA) to support the validity claims of WIDA Alternate ACCESS. WIDA structures its validity arguments using the AUA model in lieu of the model highlighted in the *Standards for Educational and Psychological Testing*. AUA has similar elements; however, they are organized differently. Below is a short summary of each AUA claim. For the full AUA validity claims, please refer to the *WIDA Assessment Use Argument* document.

Claim 1 (Consequences): With the use of Alternate ACCESS, the intended decisions will have beneficial consequences for stakeholders, in terms of using Alternate ACCESS and the decisions made based on Alternate ACCESS.

Claim 2 (Decisions): Decisions based on Alternate ACCESS test results are made by individuals, in a timely manner, and affect a variety of stakeholders. Two types of decisions that are made based on ACCESS results are classification and programming decisions. The decisions take into consideration educational and societal values, and relevant laws, rules, and regulations, and they are equitable for the intended stakeholders.

Claim 3 (Interpretations): The interpretations of students’ academic English language proficiency in four domains are *relevant* to the classification, placement and programming decisions; *sufficient*, in conjunction with additional information as outlined in state and local policies, to make such decisions; *meaningful* with respect to the WIDA English Language Development Standards; *generalizable* to the academic English language used in K–12 instructional settings, and *impartial* to all students.

Claim 4 (Assessment records: Scores): Alternate ACCESS scores are consistent across different aspects of test administration, different test tasks, and different groups of students. Test forms and metrics accurately represent the construct being measured and result in expected test taker performances.

9.1. Standards

9.1.1. Test Content

The relationship between the content of a test and the construct to measure is called content validity. Test content includes the themes, wording, and format of the items, tasks, or questions on a test. Administration and scoring may also be part of the content. Empirical or logical evidence can show how appropriately the content reflects the domain as we interpret test scores.

9.1.2. Response Processes

Empirical analysis of how test takers process tests provide evidence of the nature between performance and the construct. Examples of this validity include analyzing individual item responses, different response processes in answering questions by subgroups, or evaluating test takers' performance.

9.1.3. Internal Structure

Validity related to internal structure indicates how test items/components agree with the construct score interpretation is based on. The internal structure of the construct can be unidimensional or contain multidimensional components.

9.1.4. Relations to Other Structure

The interpretation of the test scores with an external indicator provides valuable validity evidence. We often ask how accurately the test score predicts the criterion variable. The test criterion validity has two different validities: concurrent and predictive validity. Predictive validity is how accurately test scores predict the future performance of criterion scores. Concurrent validity indicates how test scores relate to criterion scores at the same time.

9.2. *Annual Validity Studies*

Annual validity studies are conducted to ensure that the test measures the intended constructs accurately. These studies focus on construct validity by examining the relationship between test scores and the theoretical constructs they are designed to assess.

To evaluate construct validity, confirmatory factor analyses (CFA) are conducted for the four assessed domains: listening, reading, speaking, and writing. These analyses test the internal structure of the test to ensure that the hypothesized factor structure aligns with the observed data. Fit indices are reviewed to confirm the appropriateness and robustness of the measurement model.

Additionally, dimensionality checks are performed to verify whether each domain within individual clusters exhibits unidimensionality. This step ensures that test items within each domain reflect a single underlying construct, strengthening the validity of the test scores and their interpretations.

9.2.1. Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA) was conducted using all items across domains within each grade-level cluster to determine whether the items align with their respective domains. Given that the items were polytomous, a graded response model was employed, utilizing a Weighted Least Squares (WLS) estimator. The model included four factors corresponding to the Listening, Reading, Speaking, and Writing domains, with 10, 10, 8, and 8 items per domain, respectively.

Evaluation of model fit was conducted using indices such as the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA). The criteria for an acceptable model fit were a CFI and TLI of at least 0.9 and an RMSEA below 0.08.

Table 9.2.1 presents the model fit indices for the internal structures across clusters. For the four-factor structure, CFI and TLI values were consistently 0.99, and RMSEA values were below 0.08 for all clusters except the K-2 grade-level cluster. The K-2 cluster showed an RMSEA of 0.082; however, given the high CFI and TLI values, the internal structure of the test was deemed satisfactory for all clusters.

Table 9.2.1.

Fit indices for internal structures across clusters

Grade-Level Cluster	Number of Parameters	Chi-square	df	CFI	TLI	RMSEA
K-2	152	41,448.45	622	0.997	0.997	0.082
3-5	152	20,960.92	622	0.998	0.998	0.062
6-8	152	21,119.06	622	0.997	0.997	0.071
9-12	152	25,240.90	622	0.998	0.998	0.071

9.2.2 Dimensionality Check

Dimensionality checks evaluated whether each domain within individual grade-level clusters adhered to unidimensionality, a critical requirement for Rasch model applications. This assessment ensures that items within each domain collectively measure a single construct. To assess this, Winsteps software was used, employing Principal Components Analysis (PCA) of residuals and item fit statistics.

The primary criterion for unidimensionality was the size of the eigenvalue associated with residuals. An eigenvalue below 2 typically suggests unidimensionality, whereas a significantly higher eigenvalue may indicate the presence of an additional dimension. When an eigenvalue exceeds 2, the percentage of variance explained by the secondary dimension is also considered. If this percentage is relatively low, it may indicate that the secondary dimension contributes minimally to the variance, even with a higher eigenvalue. Both the eigenvalues and the percentage of variance explained by potential secondary dimensions were reviewed to determine whether unidimensionality could be reasonably supported or if the items were assessing the same conceptual domain.

Tables 9.2.2.a. through 9.2.2.d. present the eigenvalues for each domain across grade-level clusters. In the K–2 cluster, the Reading, Speaking, and Writing domains had eigenvalues exceeding 2, while the Listening domain had an eigenvalue close to 2. For the 3–5 grade-level cluster, the Reading and Writing domains also had eigenvalues exceeding 2. In the 6–8 cluster, only the Writing domain showed an eigenvalue greater than 2. For the 9–12 cluster, the Reading and Writing domains had eigenvalues greater than 2, while the Speaking domain had an eigenvalue close to 2.

Although several domains exhibited eigenvalues greater than 2, the proportion of variance explained by these secondary dimensions ranged from 5% to 6% for most grade-level clusters, except for the Writing domain in the K–2 cluster, where it reached up to 9%. Given that all percentages were below 10%, the domains can still be considered unidimensional for practical purposes.

Table 9.2.2.a.

Dimensionality checks for K–2

Domain	Eigenvalue	Percentage among total variance
Listening	1.9915	6.1
Reading	2.2401	6.0
Speaking	2.0817	7.5
Writing	3.0364	9.0

Table 9.2.2.b.

Dimensionality checks for 3–5

Domain	Eigenvalue	Percentage among total variance
Listening	1.9118	6.9
Reading	2.0425	5.9
Speaking	1.8912	6.4
Writing	2.4863	6.6

Table 9.2.2.c.

Dimensionality checks for 6–8

Domain	Eigenvalue	Percentage among total variance
Listening	1.8051	6.6
Reading	1.8428	5.8
Speaking	1.7122	5.9
Writing	2.3981	6.5

Table 9.2.2.d

Dimensionality checks for 9–12

Domain	Eigenvalue	Percentage among total variance
Listening	1.7901	6.5
Reading	2.0693	5.7
Speaking	1.9566	6.0
Writing	2.3389	6.4

10. Reliability

Reliability, along with classification accuracy and consistency, is presented in Tables 10.1.1.a. through 10.1.2.4.d. These tables are organized by grade-level cluster and domain or composite scores.

The tables include Cronbach’s alpha and the standard error of measurement (SEM) for each grade-level cluster by domain and composite. Cronbach’s alpha measures the internal consistency of the test, with values above 0.90 typically considered high, indicating strong reliability. SEM provides an estimate of the precision of Cronbach’s alpha, reflecting the amount of expected error in an estimated alpha.

Reliability coefficients (Cronbach’s alpha) are consistently high across all domains and composite scores, ranging from 0.95 to 0.97 across grade-level clusters for domains and from 0.95 to 0.98 for composite scores. These values indicate strong internal consistency, suggesting that the items within each domain and composite are effectively measuring the intended constructs.

10.1. Reliability of Domain Scores/Composite Scores

10.1.1. Reliability of Domain Scores

The reliability information, based on Classical Test Theory, includes the following metrics:

- The number of students
- The number of items
- Cronbach’s coefficient alpha (as a measure of internal consistency)
- The classical standard error of measurement (SEM) in terms of raw scores

Cronbach’s coefficient alpha is widely used as an estimate of reliability, particularly of the internal consistency of test items. It expresses how well the items on a test appear to measure the same construct. Conceptually, it may be thought of as the correlation obtained between performances on two halves of the test, if every possibility of dividing the test items in two were attempted. Thus, Cronbach’s alpha may be low if some items are measuring something other than what the majority of the items are measuring. As with any reliability index, it is affected by the number of test items (or test score points that may be awarded). That is, all things being equal, the greater the number of items, the higher the reliability.

Cronbach's alpha is also affected by the distribution of ability within the group of students tested. All things being equal, the greater the heterogeneity of abilities within the group of students tested (i.e., the more widely the scores are distributed), the higher the reliability. In this sense, Cronbach's alpha is sample dependent. It is widely recognized that reliability can be as much a function of the test as of the sample of students tested. That is, the exact same test can produce widely disparate reliability indices based on ability distribution of the group of students tested.

The formula for Cronbach's alpha is

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_t^2} \right]$$

where

n = number of items

σ_i^2 = variance of score on item

σ_t^2 = variance of total score

The last column in the following tables presents the standard error of measurement (SEM) based on classical test theory. Unlike IRT, in this approach, SEM is seen as a constant across the spread of test scores (ability continuum). Thus, it is not conditional on ability being measured. It is, however, a function of two statistics: the reliability of the test and the (observed) standard deviation of the test scores. It is calculated as

$$SEM = SD\sqrt{1 - \text{reliability}}$$

Traditionally, SEM has been used to create a band around an examinee's observed score. The assertion in the view of classical test theory is that the examinee's true score (i.e., what the examinee's score would be if it could be measured without error) would lie with a certain degree of probability within this band. Therefore, the statistical expectation is that an examinee's true score has a 68% probability of lying within the band, extending from the observed score minus 1 SEM to the observed score plus 1 SEM.

Table 10.1.1.a.

Reliability: Listening

Grade-Level Cluster	Number of Students	Number of Items	Cronbach's Alpha	SEM
K-2	9,858	10	0.955	3.0095
3-5	8,607	10	0.950	2.9511
6-8	6,521	10	0.957	2.6833
9-12	7,863	10	0.956	2.6105

Table 10.1.1.b.**Reliability: Reading**

Grade-Level Cluster	Number of Students	Number of Items	Cronbach's Alpha	SEM
K-2	9,855	10	0.965	2.7949
3-5	8,601	10	0.965	2.5485
6-8	6,521	10	0.961	2.6254
9-12	7,860	10	0.967	2.4503

Table 10.1.1.c.**Reliability: Speaking**

Grade-Level Cluster	Number of Students	Number of Items	Cronbach's Alpha	SEM
K-2	9,842	8	0.958	1.9629
3-5	8,600	8	0.968	1.9922
6-8	6,511	8	0.970	1.9457
9-12	7,853	8	0.975	1.9345

Table 10.1.1.d.**Reliability: Writing**

Grade-Level Cluster	Number of Students	Number of Items	Cronbach's Alpha	SEM
K-2	9,850	8	0.953	2.1447
3-5	8,594	8	0.952	2.2315
6-8	6,513	8	0.954	2.2275
9-12	7,854	8	0.961	2.1378

10.1.2. Reliability of Composite Scores

Four composite scores are reported for Alternate ACCESS: Oral Language Composite (Oral), Literacy Composite (Lit), Comprehension Composite (Comp), and Overall Composite (Overall). To estimate the reliability of these composite scores, a stratified Cronbach's alpha coefficient (e.g., Kamata, Turhan, & Darandari, 2003; Kane, & Case, 2004; Rudner, 2001) is computed, weighted by the contribution of each domain score into the composite. Specifically, the formula is

$$\alpha_c = 1 - \frac{\sum_{j=1}^k w_j^2 \sigma_j^2 (1 - p_j)}{\sigma_c^2}$$

where

k = number of components j

w_j = domain weight of component j

σ_j^2 = variance of component j

σ_c^2 = variance of composite

ρ_j = reliability coefficient of component j

10.1.2.1 Oral Composite

Table 10.1.2.1.a.

Reliability: Oral K–2

Component	Weight	Variance	Reliability
Listening	0.5	403.29	0.9555
Speaking	0.5	458.04	0.9584
Oral	NA	370.52	0.9750

Table 10.1.2.1.b.

Reliability: Oral 3–5

Component	Weight	Variance	Reliability
Listening	0.5	305.97	0.9504
Speaking	0.5	497.14	0.9677
Oral	NA	343.06	0.9772

Table 10.1.2.1.c.

Reliability: Oral 6–8

Component	Weight	Variance	Reliability
Listening	0.5	317.61	0.9569
Speaking	0.5	462.71	0.9698
Oral	NA	332.17	0.9792

Table 10.1.2.1.d.

Reliability: Oral 9–12

Component	Weight	Variance	Reliability
Listening	0.5	306.25	0.9556
Speaking	0.5	560.52	0.9749
Oral	NA	369.17	0.9813

10.1.2.2. Literacy Composite

Table 10.1.2.2.a.

Reliability: Lit K–2

Component	Weight	Variance	Reliability
Reading	0.5	474.93	0.9648
Writing	0.5	514.98	0.9532
Literacy	NA	424.87	0.9760

Table 10.1.2.2.b.

Reliability: Lit 3–5

Component	Weight	Variance	Reliability
Reading	0.5	357.09	0.9649
Writing	0.5	464.86	0.9519
Literacy	NA	354.81	0.9754

Table 10.1.2.2.c.

Reliability: Lit 6–8

Component	Weight	Variance	Reliability
Reading	0.5	339.72	0.9606
Writing	0.5	493.12	0.9537
Literacy	NA	363.11	0.9750

Table 10.1.2.2.d.

Reliability: Lit 9–12

Component	Weight	Variance	Reliability
Reading	0.5	382.92	0.9674
Writing	0.5	512.08	0.9613
Literacy	NA	398.89	0.9797

10.1.2.3. Comprehension Composite

Table 10.1.2.3.a.

Reliability: Comp K–2

Component	Weight	Variance	Reliability
Listening	0.3	403.29	0.9555
Reading	0.7	474.93	0.9648
Comprehension	NA	425.64	0.9769

Table 10.1.2.3.b.

Reliability: Comp 3–5

Component	Weight	Variance	Reliability
Listening	0.3	305.97	0.9504
Reading	0.7	357.09	0.9649
Comprehension	NA	316.23	0.9763

Table 10.1.2.3.c.

Reliability: Comp 6–8

Component	Weight	Variance	Reliability
Listening	0.3	317.61	0.9569
Reading	0.7	339.72	0.9606
Comprehension	NA	308.91	0.9748

Table 10.1.2.3.d.

Reliability: Comp 9–12

Component	Weight	Variance	Reliability
Listening	0.3	306.25	0.9556
Reading	0.7	382.92	0.9674
Comprehension	NA	335.21	0.9781

10.1.2.4. Overall Composite

Table 10.1.2.4.a.

Reliability: Overall K–2

Component	Weight	Variance	Reliability
Listening	0.15	403.29	0.9555
Speaking	0.15	458.04	0.9584
Reading	0.35	474.93	0.9648
Writing	0.35	514.98	0.9532
Overall	NA	382.56	0.9847

Table 10.1.2.4.b.

Reliability: Overall 3–5

Component	Weight	Variance	Reliability
Listening	0.15	305.97	0.9504
Speaking	0.15	497.14	0.9677
Reading	0.35	357.09	0.9649
Writing	0.35	464.86	0.9519
Overall	NA	327.80	0.9848

Table 10.1.2.4.c.

Reliability: Overall 6–8

Component	Weight	Variance	Reliability
Listening	0.15	317.61	0.9569
Speaking	0.15	462.71	0.9698
Reading	0.35	339.72	0.9606
Writing	0.35	493.12	0.9537
Overall	NA	331.88	0.9847

Table 10.1.2.4.d.

Reliability: Overall 9–12

Component	Weight	Variance	Reliability
Listening	0.15	306.25	0.9556
Speaking	0.15	560.52	0.9749
Reading	0.35	382.92	0.9674
Writing	0.35	512.08	0.9613
Overall	NA	369.33	0.9876

10.2. Conditional Standard Errors of Measurement of the Scale Scores at the Cut Points (Composites)

In addition to evaluating test score reliability in terms of estimates of internal consistency, we can calculate the amount of measurement error in students' test scores in two different ways. One way is to hypothesize that there is an error-free measure of each student's true ability, referred to as the true score in classical test theory. The true score is a theoretical value, so it is not a known quantity. Rather, we view it as the hypothetical average score over repeated replications of the same testing condition (Livingston et al, 2018). Under the assumptions of classical test theory, the error of measurement over a replication of a testing condition provides an estimate of the amount of variability from students' true scores that we would expect. In practical testing contexts, it is generally not possible to replicate a testing condition (i.e., have students take the same test form multiple times), so it is not possible to estimate the standard error of each student's score using a repeated measures design. Instead, we calculate the average error of measurement over the population of students who take the test, and then we use that as an indication of the amount of variation in any individual student's score that we would expect. Classical test theory refers to this average as the standard error of measurement (SEM), which provides an indication of how much students' scores differ from their true scores, on average, on the raw score metric. Because it is a standard deviation of the distribution of errors of measurement, we can construct a confidence interval to indicate how the errors of measurement are affecting the scores. Test scores with large SEMs pose a challenge to the interpretation of the reliability of any single test score.

A second way to address the impact of measurement errors on students' test scores is to estimate the SEM for specific scores using IRT. IRT addresses reliability using the test information function, which indicates the precision with which we can use student performances on items and tasks to estimate the latent (i.e., true) ability of each student (i.e., latent scores). The square root of the inverse of the information function at any point on the latent ability distribution is the conditional standard error of measurement (CSEM). The CSEM provides information about the amount of error we would expect in any student's score at that point on the underlying latent ability scale, which IRT refers to in terms of the latent score metric (i.e., the IRT metric for expressing student ability, as opposed to the raw score metric). In addition, by using IRT, we can estimate indices analogous to traditional reliability coefficients such as Cronbach's coefficient alpha from the test information function and the distribution of the latent scores in the same student population.

The tables in this section present information about the conditional standard errors of measurement (CSEM) values of scale scores at the most important points at which policy makers make decisions such as reclassification about students based on performance on Alternate ACCESS—the cut points between language proficiency levels. The CSEM provides information about the amount of measurement error we would expect in any student's scale score at that point on the underlying latent ability scale. We first computed CSEM values on the theta metric, which is the square root of the inverse of the Test Information Function. Next, we used the multiplicative constant of the linear equation for the domain to linearly transform those logit-based CSEM values so that we could report them on the Alternate ACCESS score scale.

We use the CSEM to construct an error band, quantifying the amount of uncertainty in a student's scale score. One CSEM below a student's scale score and one CSEM above that scale score indicates an approximate 68% confidence interval. To interpret this confidence interval, consider a student who takes the test 100 times. Assuming measurement error is normally distributed, the student's true proficiency would fall within the confidence interval 68% of the time (or 68 times out of 100).

Figure 10.2.5.a. through Figure 10.2.8.d. present conditional standard error of measurement (CSEM) for composite scores. CSEM is measurement error computed by applying weights of individual domain scale scores in each composite score. The CSEM curves are presented by each proficiency levels in composite scores. This figure informs the amount of error variability on scale score level. Higher CSEM informs more measurement error and lower CSEM indicates more reliability.

10.2.1. Listening

Figure 10.2.1.a.

CSEM for Listening K-2

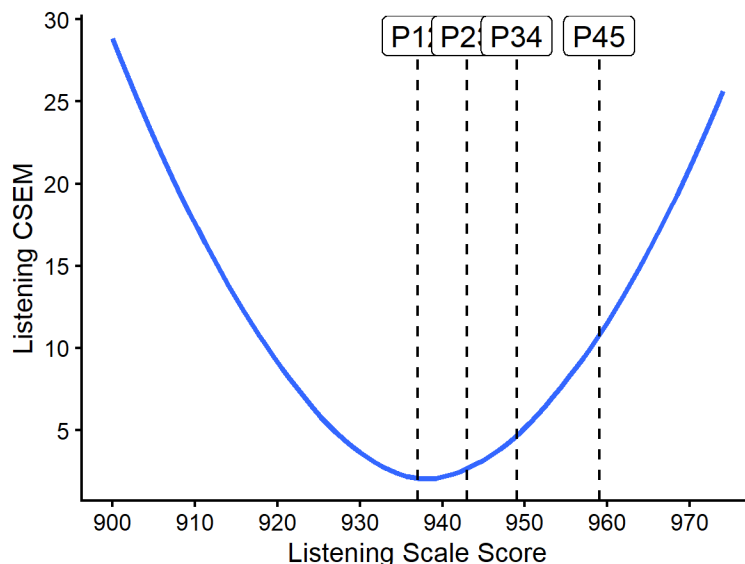


Figure 10.2.1.b.

CSEM for Listening 3-5

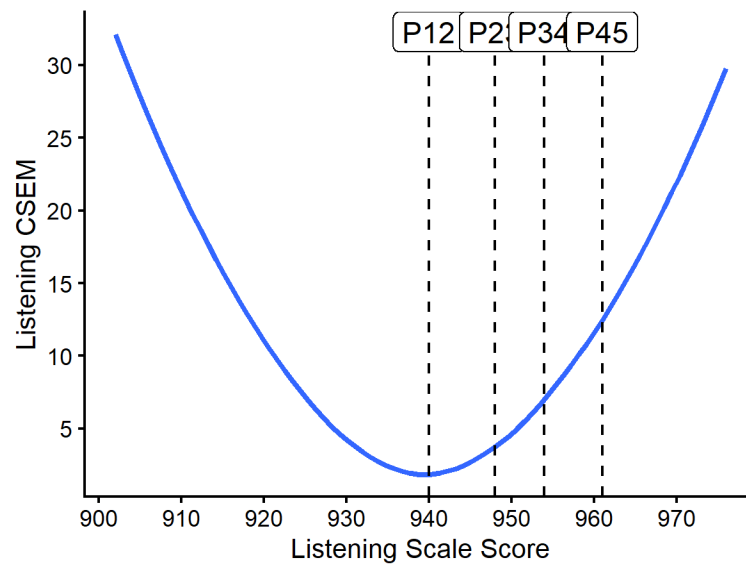


Figure 10.2.1c.

CSEM for Listening 6-8

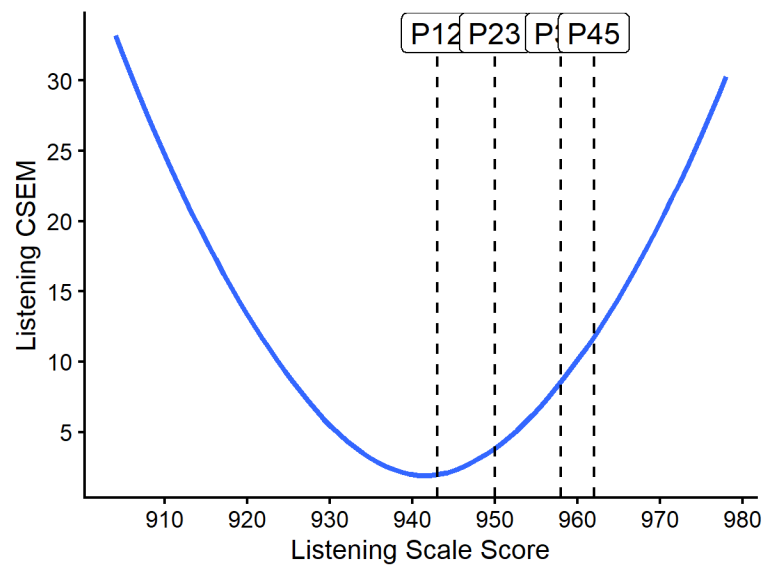
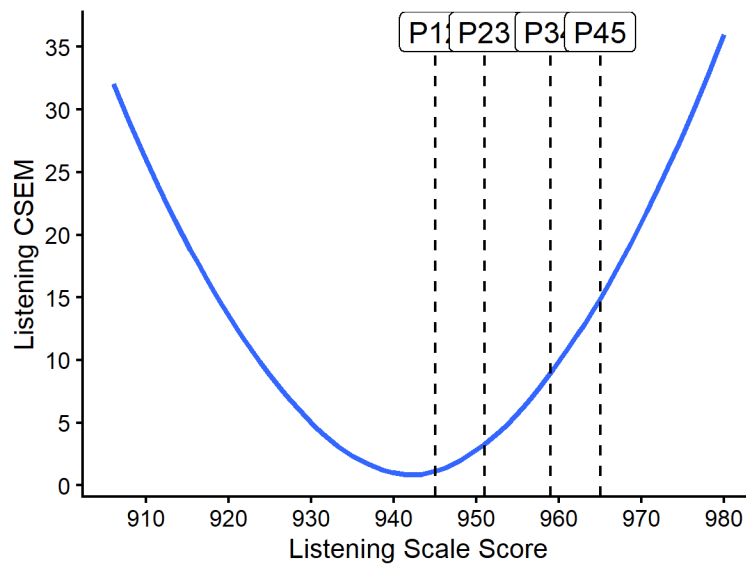


Figure 10.2.1.d.

CSEM for Listening 9–12



10.2.2. Reading

Figure 10.2.2.a.

CSEM for Reading K–2

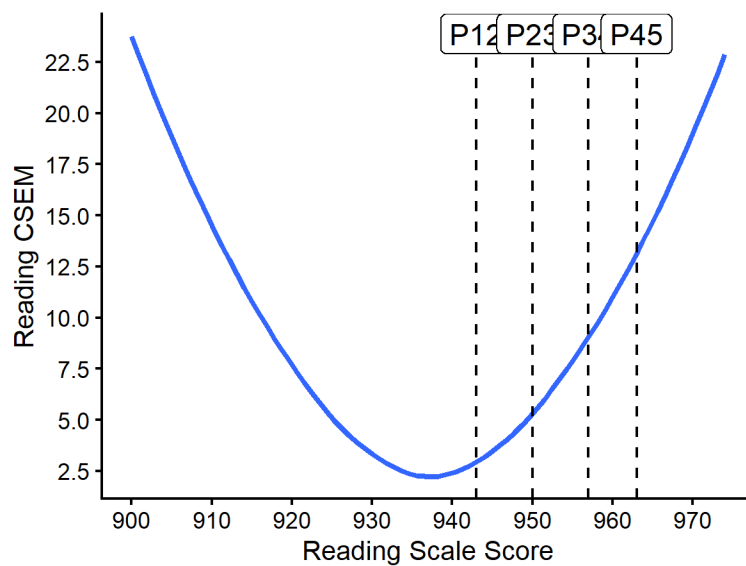


Figure 10.2.2.a.

CSEM for Reading 3-5

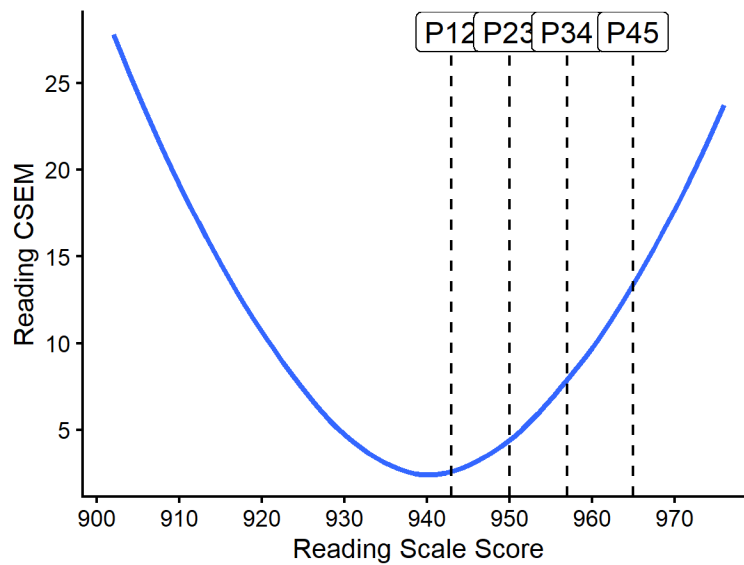


Figure 10.2.2.c.

CSEM for Reading 6-8

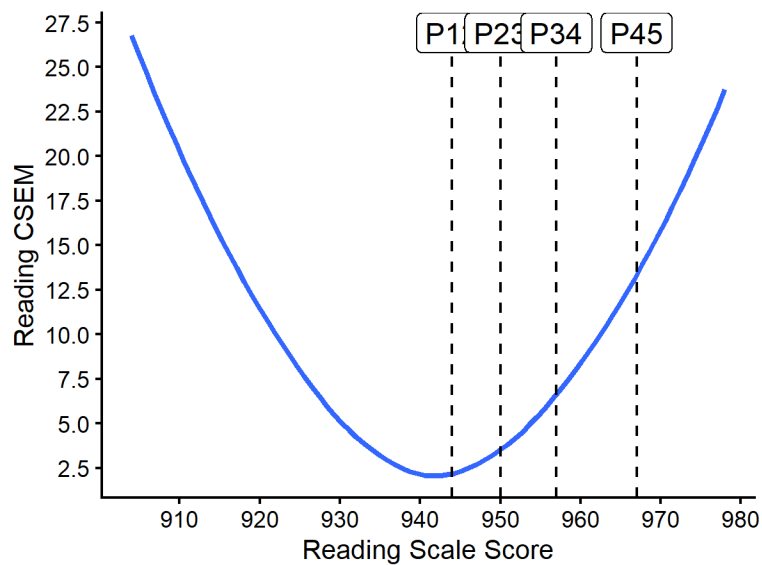
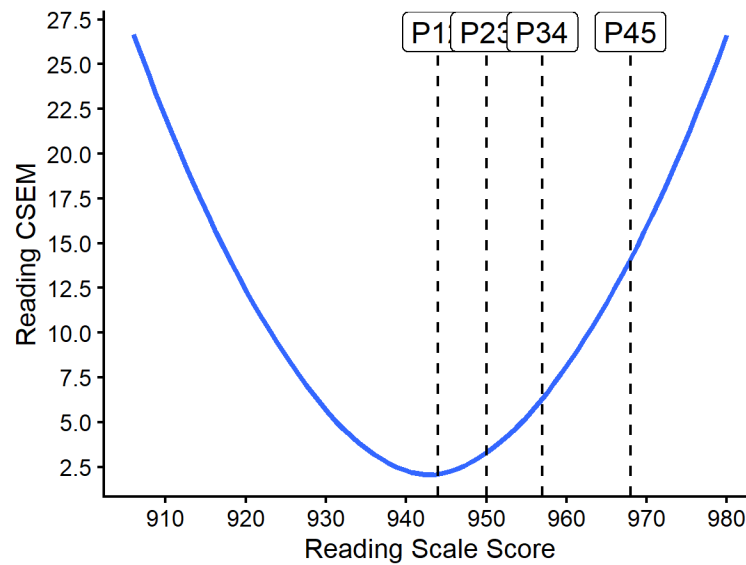


Figure 10.2.2.d.

CSEM for Reading 9–12



10.2.3. Speaking

Figure 10.2.3.a.

CSEM for Speaking K–2

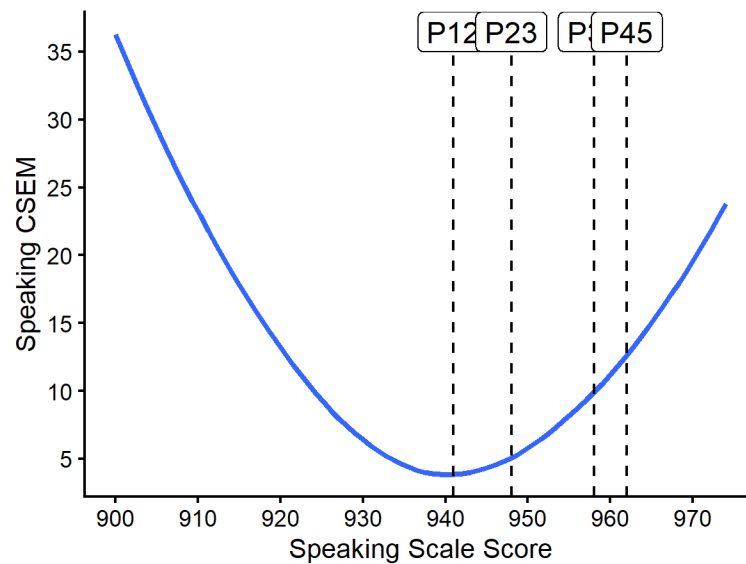


Figure 10.2.3.b.

CSEM for Speaking 3-5

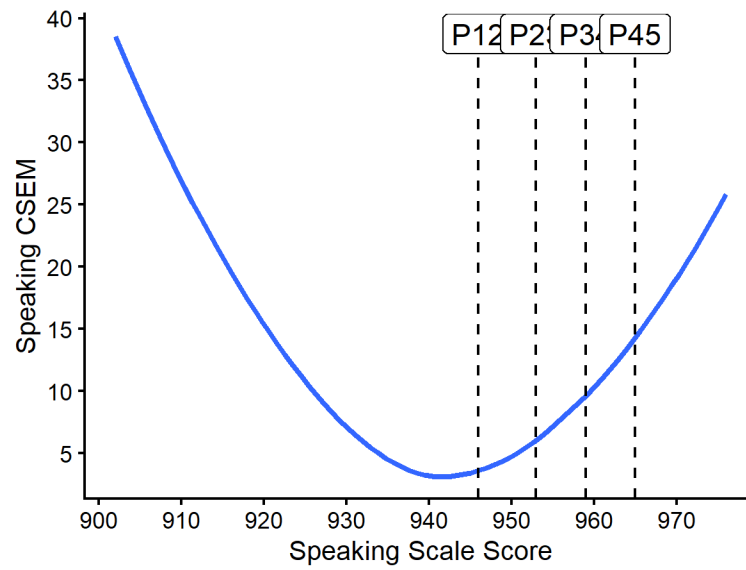


Figure 10.2.3.c.

CSEM for Speaking 6-8

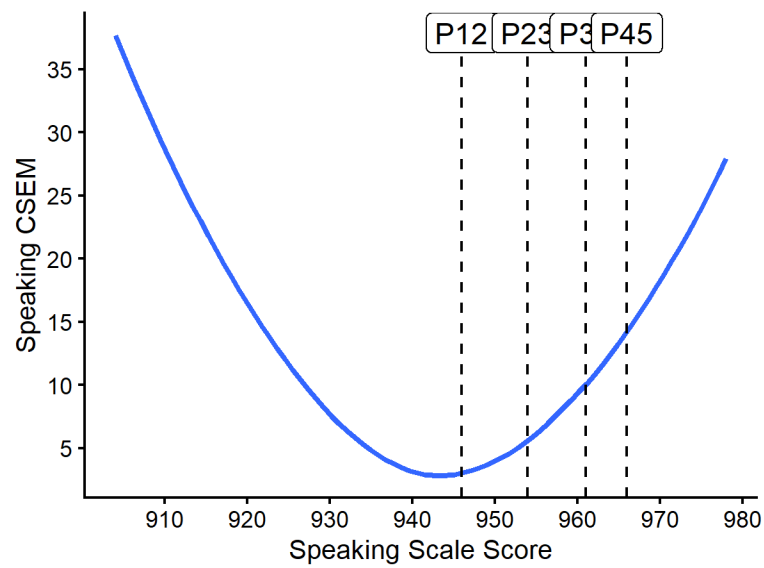
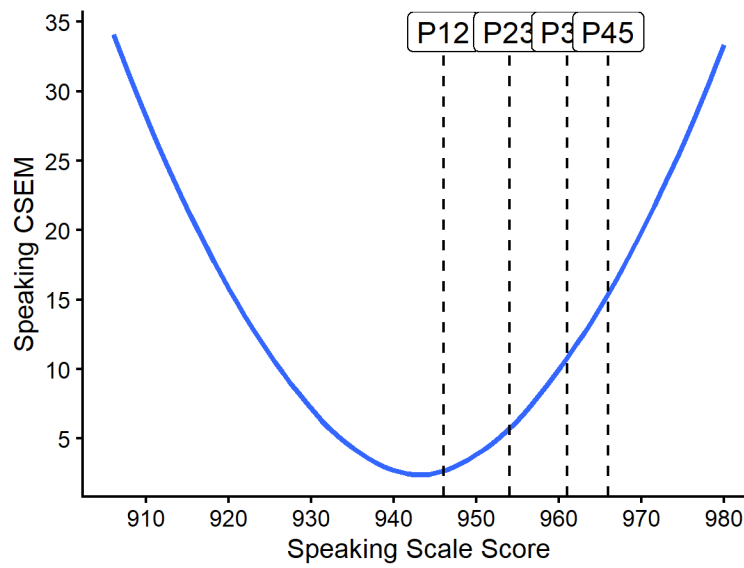


Figure 10.2.3.d.

CSEM for Speaking 9–12



10.2.4. Writing

Figure 10.2.4.a.

CSEM for Writing K–2

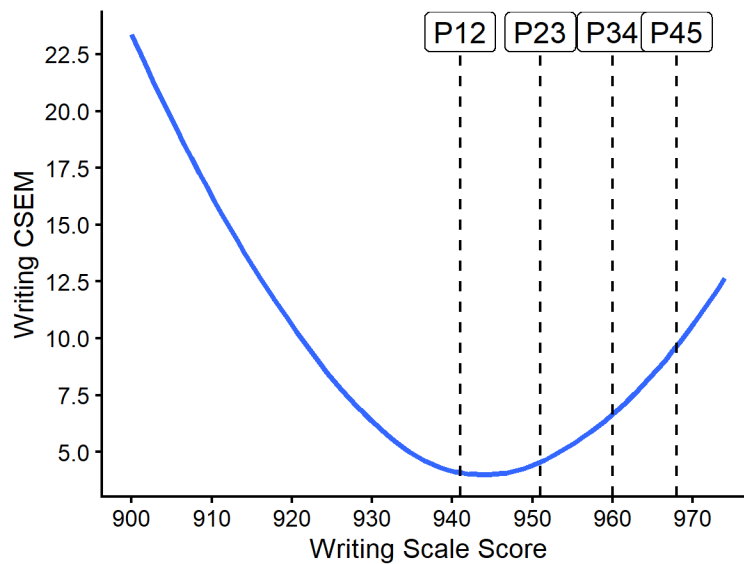


Figure 10.2.4.b.

CSEM for Writing 3–5

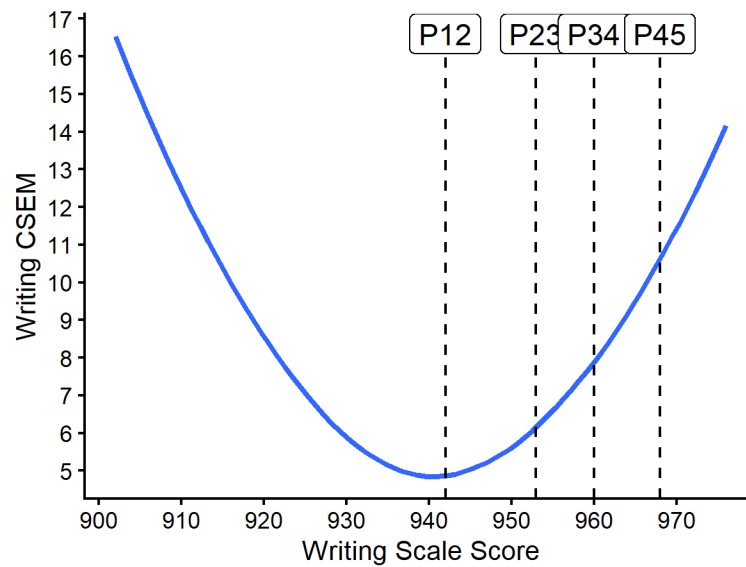


Figure 10.2.4.c.

CSEM for Writing 6–8

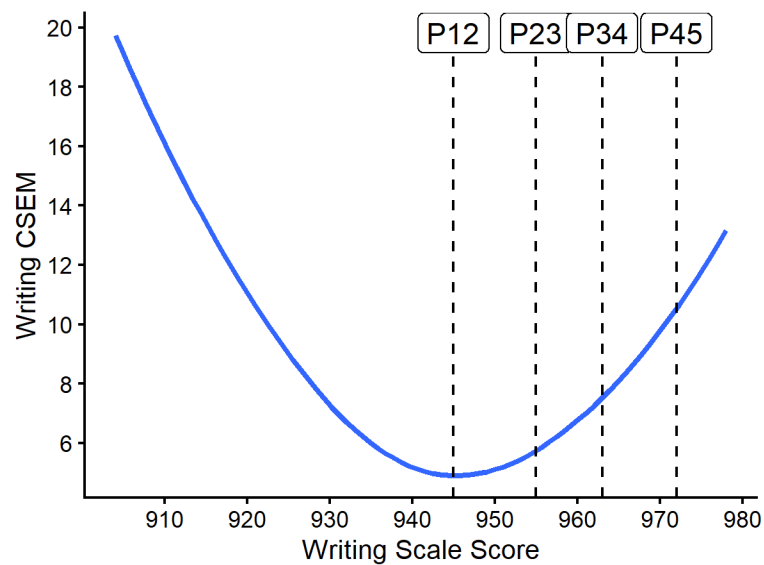
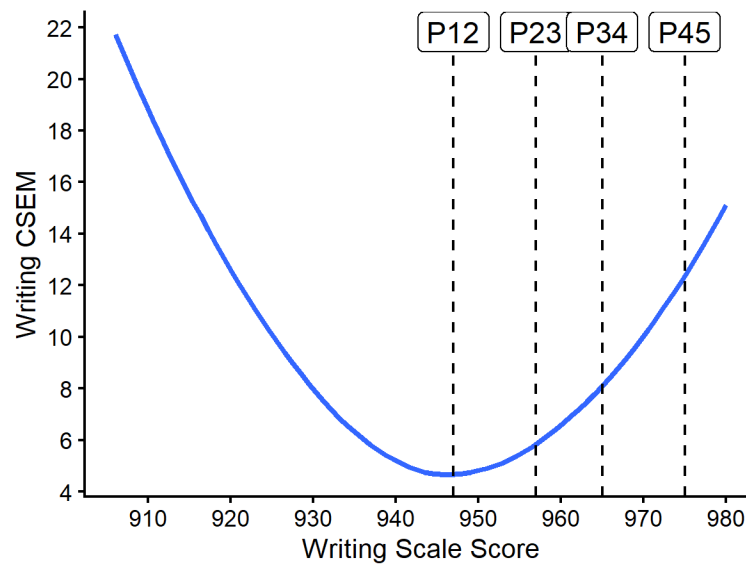


Figure 10.2.4.d.

CSEM for Writing 9–12



10.2.5. Oral Composite

Figure 10.2.5.a.

CSEM for Oral K–2

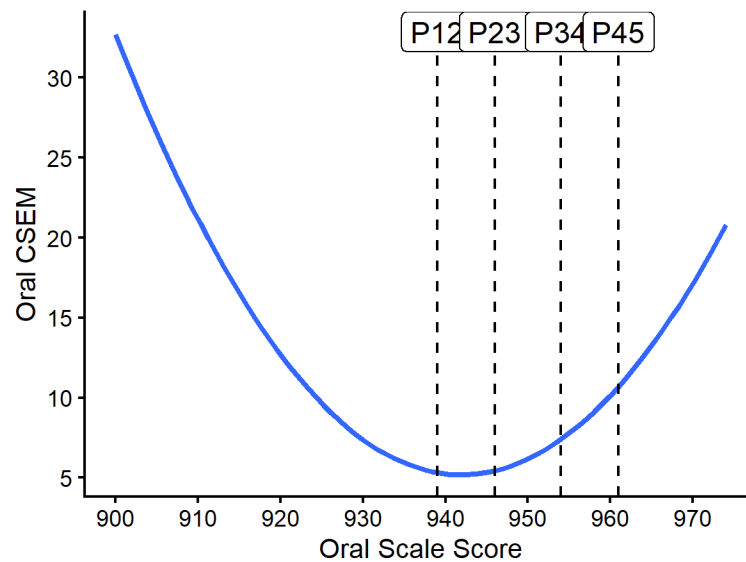


Figure 10.2.5.b.

CSEM for Oral 3–5

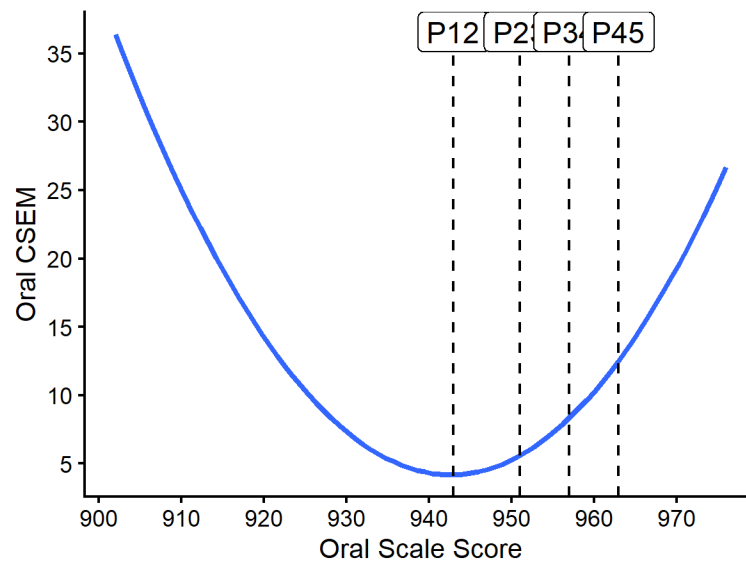


Figure 10.2.5.c.

CSEM for Oral 6–8

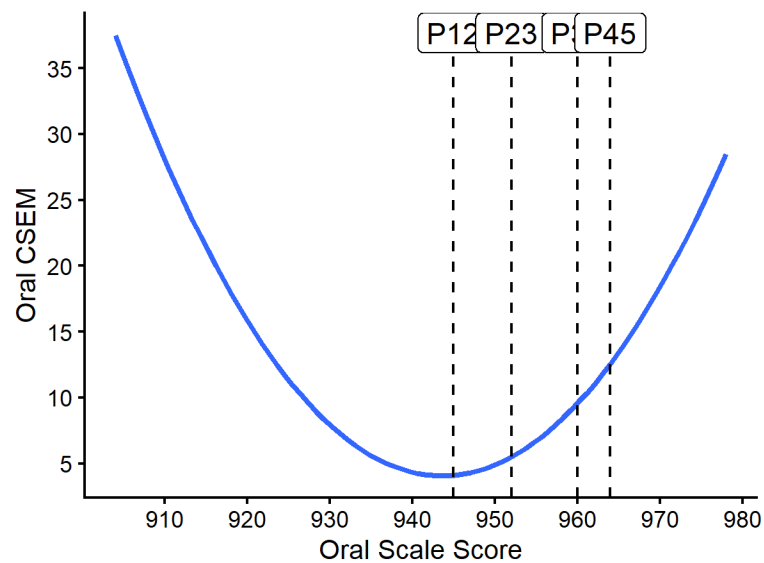
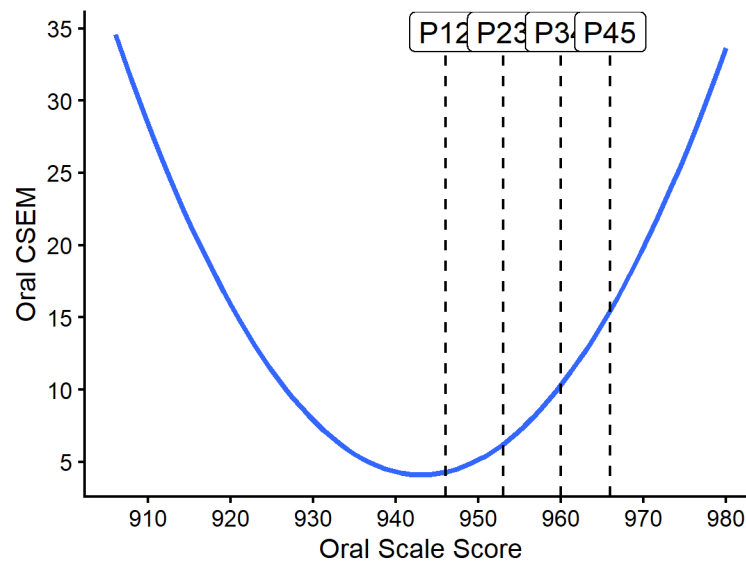


Figure 10.2.5.d.

CSEM for Oral 9–12



10.2.6. Literacy Composite

Figure 10.2.6.a.

CSEM for Lit K–2

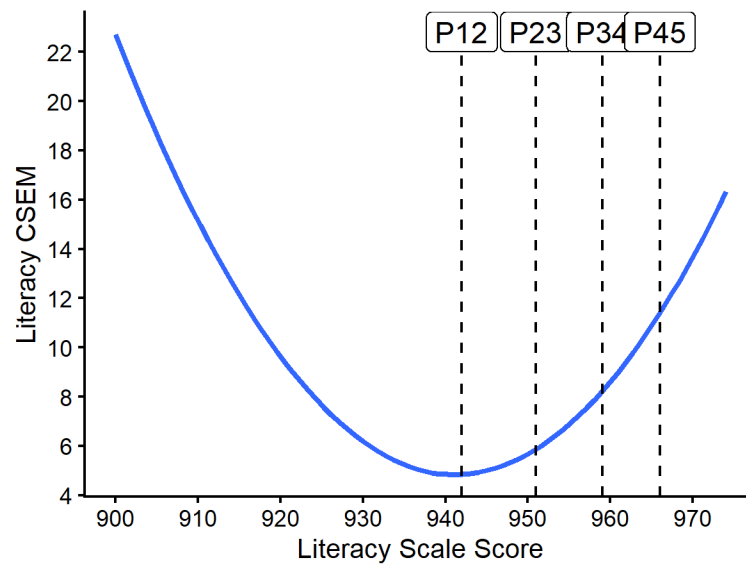


Figure 10.2.6.b.

CSEM for Lit 3–5

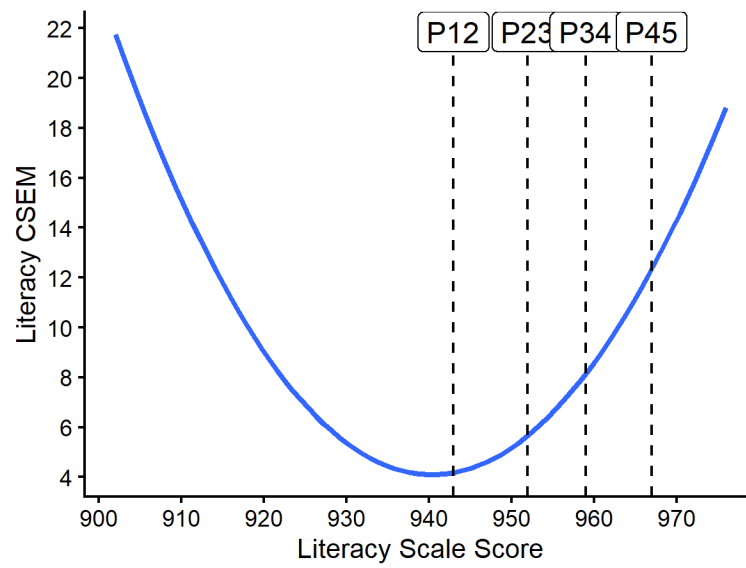


Figure 10.2.6.c.

CSEM for Lit 6–8

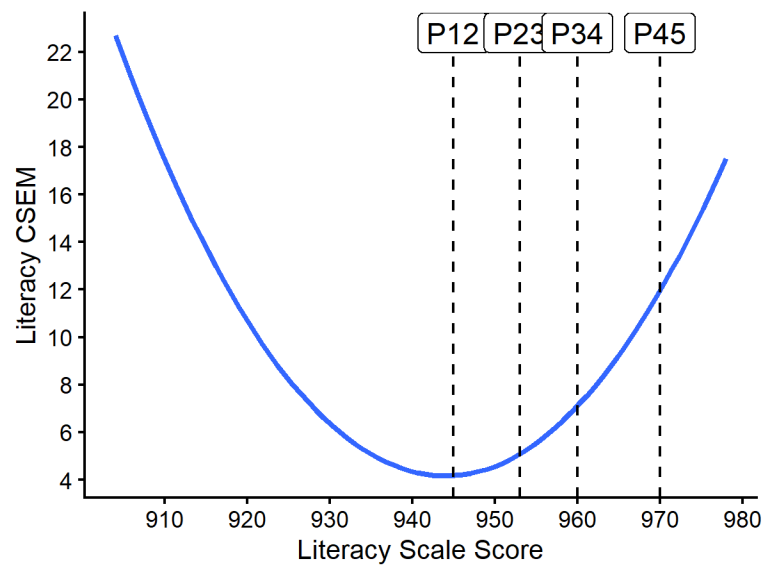
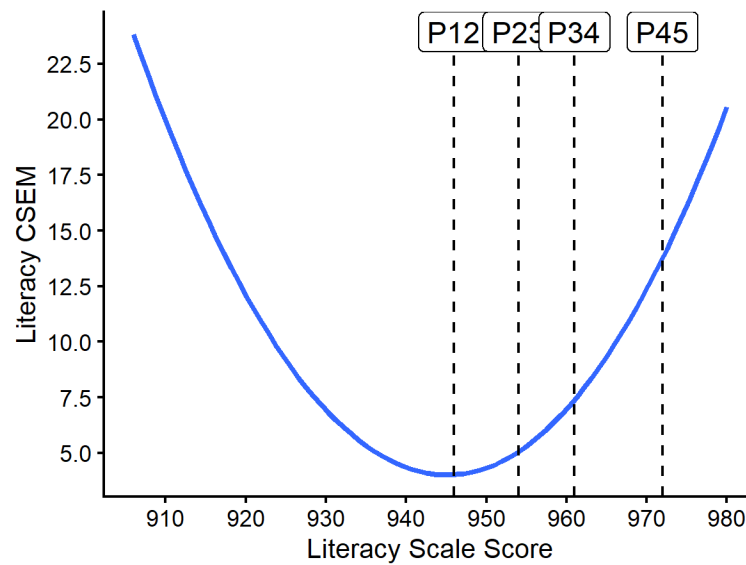


Figure 10.2.6.d.

CSEM for Lit 9–12



10.2.7. Comprehension Composite

Figure 10.2.7.a.

CSEM for Comp K–2

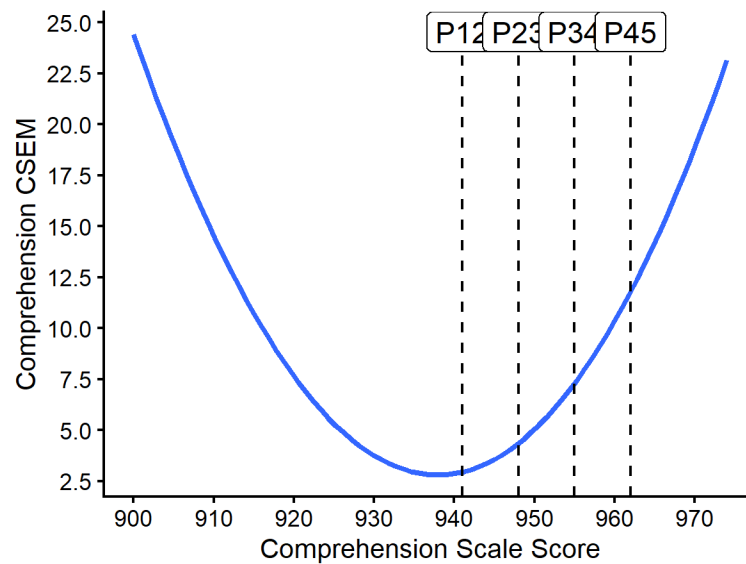


Figure 10.2.7.b.

CSEM for Comp 3-5

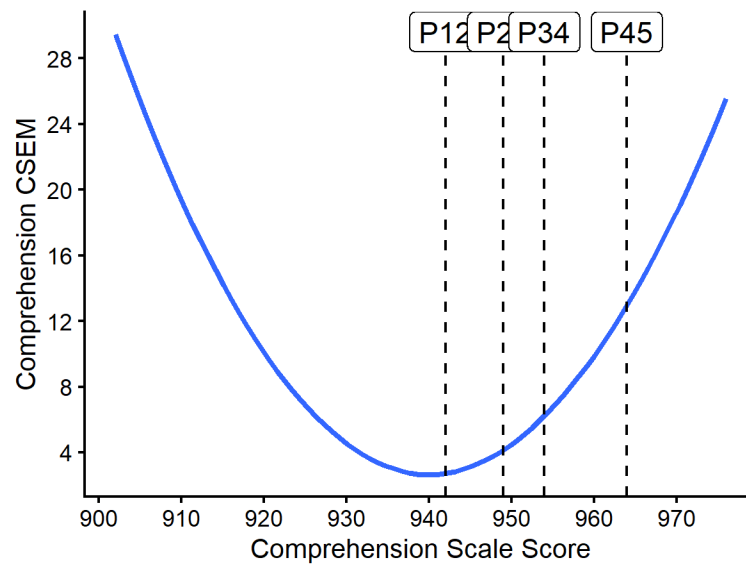


Figure 10.2.7.c.

CSEM for Comp 6-8

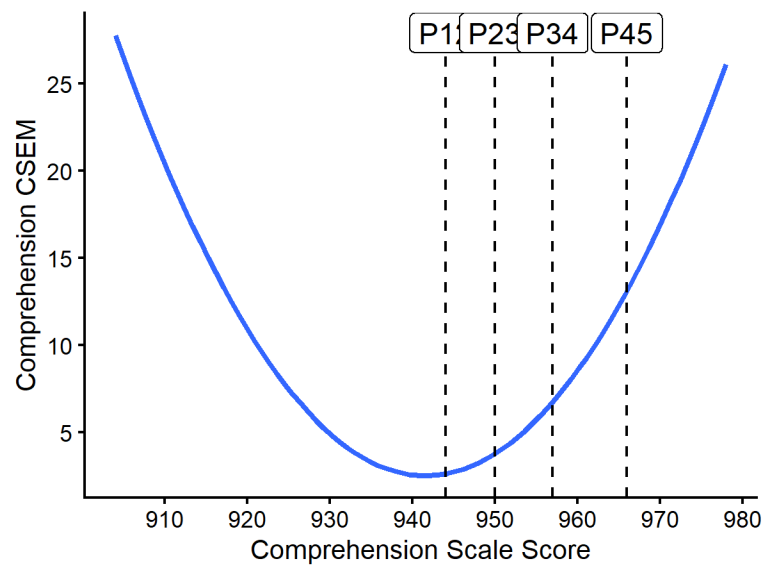
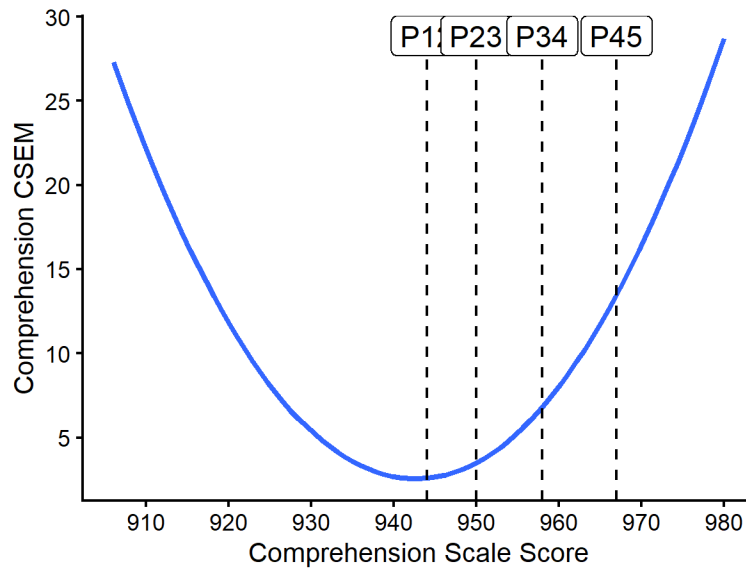


Figure 10.2.3.d.

CSEM for Comp 9–12



10.2.8. Overall Composite

Figure 10.2.8.a.

CSEM for Overall K–2

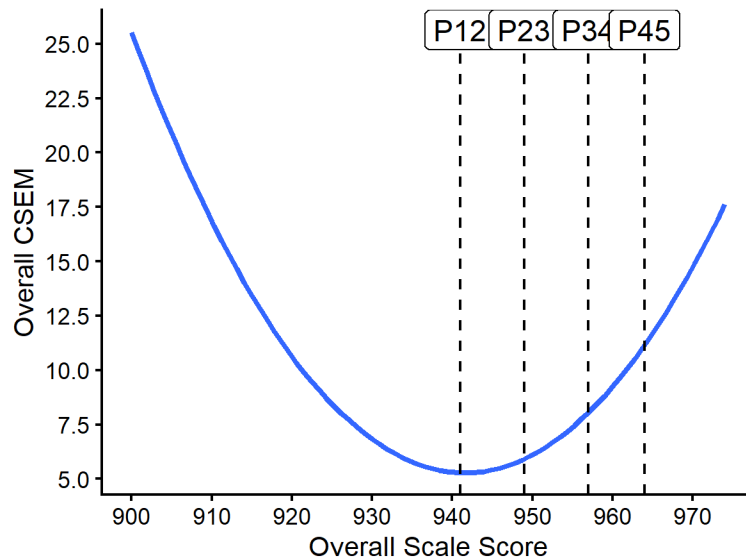


Figure 10.2.8.b.

CSEM for Overall 3-5

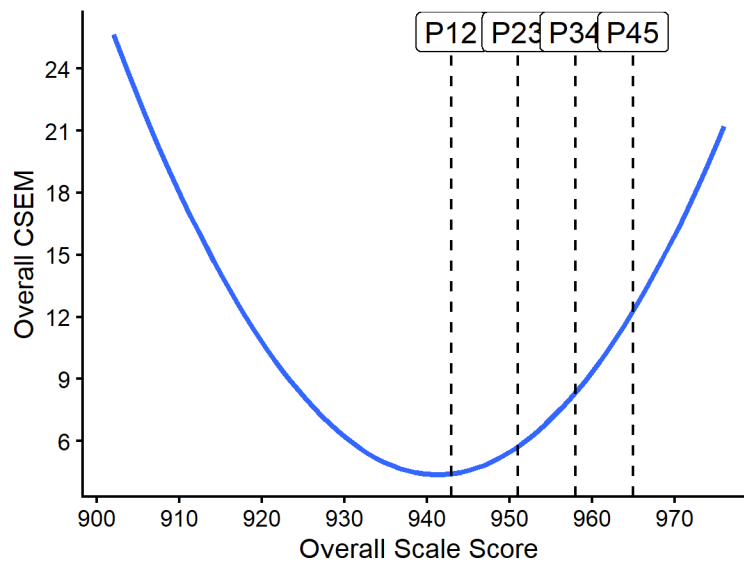


Figure 10.2.8.c.

CSEM for Overall 6-8

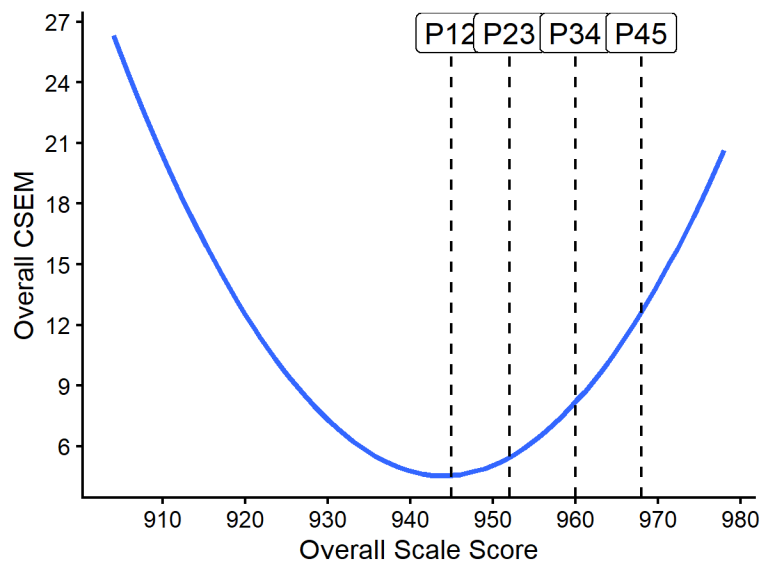
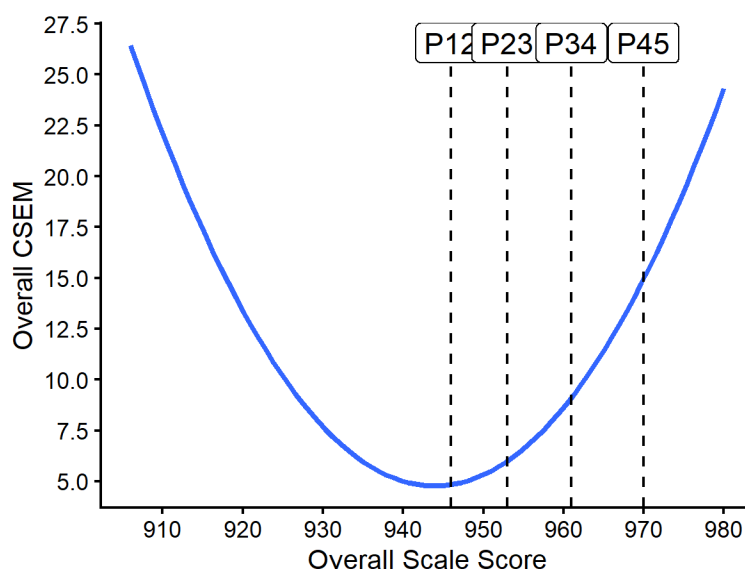


Figure 10.2.8.d.

CSEM for Overall 9–12



10.3. Interrater Agreement Rates

10.3.1. Overview

Two studies about Alternate ACCESS interrater reliability are described in this section. The first study was on the Writing and Speaking tests. It was conducted in the 2022–2023 school year during the Alternate ACCESS Field Test. It found that there were generally high levels of rater agreement on Writing and Speaking tests, although there were some areas where agreement levels were below expected criteria. The purpose of this study was to examine interrater agreement, but it was also to identify areas that could be improved in rater training and scoring. The study showed that there was good interrater agreement. A key takeaway from the study was the need for more clarity in the Expect Boxes and more detailed explanations of grammatical terms. The Writing and Speaking operational rater training and scoring materials were updated accordingly.

The second study was on Reading and Listening tests. It occurred during the 2023–2024 operational test administration of Alternate ACCESS. To obtain interrater reliability information, raters needed to be present during test administration. WIDA trained raters, which included staff members from both WIDA and member states who traveled to schools and scored students' test performances alongside local test administrators. High rater agreement was observed between WIDA raters and local test administrators. However, only 50 students were included in this study. Because of the small sample, strong generalizations about interrater agreement across the consortium should be made with caution. Results do indicate that the Alternate ACCESS Reading and Listening training materials and scoring resources can be used effectively to obtain acceptable agreement. Descriptions of both studies along with study findings are presented in the following subsections.

10.3.2. Interrater Reliability Study for Writing and Speaking

The purpose of this study was to examine the scoring protocols of the Alternate ACCESS Field Tests' items in the domains of Speaking and Writing through examining the interrater reliability of test administrators, formally trained WIDA raters, and WIDA expert raters. Three significant outcomes were anticipated from this study. First, developers of the updated version of Alternate ACCESS will gain information on the quality of scoring protocols in Speaking and Writing. Second, using the information gained, scoring protocols and associated training materials can be improved upon. Third, validity evidence for the interrater reliability of the updated version of Alternate ACCESS scores will be obtained. Together, these outcomes provide evidence of the quality of this test.

10.3.2.1 Study design, hypotheses, and analyses

This study was conducted in three phases. The first phase collected samples of student responses and test administrator scores from the 2022–2023 Alternate ACCESS Field Test administration. From that administration, a subset of student scores was used to conduct a benchmarking activity, and another subset was used for the interrater reliability study.

In phase two, expert raters rated a small set of Speaking and Writing responses and discussed their ratings. The goal of this activity was to identify the consistency among expert raters' judgments, come to consensus about how to rate student performances, and prepare more detailed training materials for the interrater reliability study.

In phase three, expert raters and WIDA staff trained as raters (WIDA trained raters) scored the interrater reliability sample. There were four expert raters and four WIDA trained raters. One expert rater was paired with one WIDA trained rater, and each pair rated the Speaking and Writing responses for a single grade-level cluster.

The hypothesis was that the agreement rates between expert raters and WIDA trained raters will be highest, the agreement rate between WIDA trained raters and the local test administrator ratings would be second highest, and the lowest agreement rates would be between expert raters and local test administrators. This assumption was made because expert raters were part of the development process of the Alternate ACCESS Field Test, and as such, they would have the best sense of how to score students responses based on the test's scoring protocols. Also, the WIDA trained raters were trained by the expert raters and would have insights on scoring that the local test administrators did not receive.

Two agreement statistics were used for the interrater reliability study. The first was the percent agreement between raters. The raters' task was to agree whether student samples were a Q1-Meets (the highest score). Did they agree with local test administrators' determination of Q1-Meets or not? Percentages were calculated from those determinations. The criteria for rater agreement varies in the literature; however, for this study agreement rates greater than 80% are considered high levels of agreement.

The second statistic was Cohen's kappa coefficient (k). The kappa coefficient is a well-known metric for rater agreement. It is expressed as follows:

$$k = \frac{P_o - P_e}{1 - P_e}$$

where P_o is the observed agreement among raters and P_e is the likelihood of chance agreement. Given the hypothesis of rater agreements, the kappa coefficient was only calculated between expert raters and WIDA trained raters. Kappa coefficient interpretation criteria used for this study are shown in Table 10.3.2.

Table 10.3.2.

Kappa Coefficient Interpretation

Cohen's Kappa	Interpretation
No agreement	0.00
Little agreement	0.01 to 0.20
Fair agreement	0.21 to 0.40
Moderate agreement	0.41 to 0.60
High agreement	0.61 to 0.80
Near perfect agreement	0.61 to 1.00

10.3.2.2. Raters

Expert Rater Qualifications: There were four expert raters. Two are principal investigators on the Advancing ALTELLA project, work at WIDA, and have decades of experience designing and developing assessments and working with English learners and students with significant cognitive disabilities. Another rater serves as the content specialist on the Advancing ALTELLA project and has many years' experience working with English learners and students with disabilities. The final expert rater serves as an inclusion researcher at WIDA, taught university in-service and pre-service courses in special education, and has decades of experience in the classroom teaching special education students.

WIDA Trained Rater Qualifications: Four WIDA trained raters participated in this study. All raters are experts in working with English learners and work at WIDA in a variety of areas such as professional learning, assessment content development, and assessment research. All WIDA trained raters were required to take the online Alternate ACCESS Field Test training course. Additionally, all raters were required to participate in another training activity facilitated by the expert raters. This additional training was created as a direct result of what was learned from the expert rater benchmarking activity.

Test Administrator Rater Qualifications: The specific qualifications of local test administrators are unknown. Given the process of randomly selecting the sample, test administrators are likely to represent the typical makeup of those who administer Alternate ACCESS. Most individuals who administer Alternate ACCESS are ESL/Bilingual teachers or Title III Coordinators, with smaller numbers of educators being school or district administrators and special education teachers. It is reasonable to assume that similar distributions of test administrators were represented in the sample. Local test administrators were required to take an online training course to administer the field test. An excerpt from the WIDA Secure Portal describes this course:

The purpose of this course is to help test administrators successfully prepare for and administer the Alternate ACCESS for ELLs assessment. The course includes an overview of the assessment, its structure, and each domain (Listening, Reading, Speaking, Writing); materials needed to prepare for and administer the assessment; how to administer and score each domain of the test during the session; and what needs to be done following the session. Upon completing the course, a quiz will assess your knowledge and readiness to administer and score the assessment. A score of at least 80% is required in order to deliver the assessment. (<https://portal.wida.us/course/detail/Alternate-ACCESS>)

10.3.2.3 Study Sample

In total, 480 student responses were sought for the benchmarking activity and interrater reliability study. To achieve this, 1,200 student responses with a score of Q1-Meets (i.e., score=4) were selected. Specifically, 150 student responses were sought for each domain (Speaking and Writing) and each grade-level cluster (i.e., 150 x 2 [domains] x 4 [grade-level clusters]). From this sample, 80 student responses were randomly selected for the benchmarking activity, and 400 responses were randomly selected for the interrater reliability study. All 480 student responses were reviewed by WIDA researchers who were not part of the expert rater or WIDA trained rater groups. These researchers determined whether student responses were legible and whether there was any personally identifiable information. If either of these were true, that student's response was replaced with a previously non-selected sample.

Table 10.3.2.3.a. shows the number of selected samples for the benchmarking activity. The goal in selecting samples for the benchmarking activity was to broadly cover proficiency levels across grade-level clusters and domains.

Table 10.3.2.3.a.

Benchmarking Sampled Task Counts by Grade-Level Cluster and Proficiency Level (PL)

Grade-Level Cluster	PL1	PL2	PL3	PL4	PL5	Total
K-2 Speaking	0	0	10	0	0	10
K-2 Writing	0	0	0	0	10	10
3-5 Speaking	10	0	0	0	0	10
3-5 Writing	0	10	0	0	0	10
6-8 Speaking	0	0	0	0	10	10
6-8 Writing	0	0	0	10	0	10
9-12 Speaking	0	0	0	0	10	10
9-12 Writing	10	0	0	0	0	10
Total	20	10	10	10	30	80

Note: The proficiency levels listed in the table are those established by test developers.

Table 10.3.2.3.b. displays the number of student samples used for the interrater reliability study. The original design was to select items at proficiency levels (PLs) 1, 3, and 5 for the study. This was possible for the domain of Speaking. However, it was not possible for Writing due to limited response samples. If an item had insufficient numbers (or didn't exist) at PL3, then a PL2 sample replaced it. Likewise, if an item had an insufficient number of responses at PL5, a PL4 sample replaced it. At each domain and grade-level cluster, 50 student responses were evaluated. Because of the likelihood that reclassification criteria for this assessment would be at or greater than the PL3 level, PL3 and PL5 items were oversampled.

Table 10.3.2.3.b.

Interrater Reliability Study Counts by Grade-Level Cluster and Proficiency Level (PL)

Grade-Level Cluster	PL1	PL2	PL3	PL4	PL5	Total
K-2 Speaking	10	0	20	0	20	50
K-2 Writing	10	0	20	0	20	50
3-5 Speaking	10	0	20	0	20	50
3-5 Writing	10	20	0	20	0	50
6-8 Speaking	10	0	20	0	20	50
6-8 Writing	10	0	20	20	0	50
9-12 Speaking	10	0	20	0	20	50
9-12 Writing	10	0	20	0	20	50
Total	80	20	140	40	120	400

Note: The proficiency levels listed in the table are those established by test developers.

10.3.2.4 Expert Rater Benchmarking

The benchmarking activity rated a smaller number of student samples, one in each domain in each grade-level cluster. Expert raters rated all items in all grade-level clusters. The activity was meant to identify any potential issues with the collected student samples, get a sense of the consistency of expert raters, and come to consensus about how to interpret Expect Boxes. Expert raters had an initial meeting at which the benchmarking task was discussed and an opportunity to ask questions was provided. Once complete, raters were given five days to score student responses. Expert raters then met and discussed their scores.

Table 10.3.2.4.a. and Table 10.3.2.4.b. display the agreement between expert raters before discussions. The percentages are calculated by counting the number of student samples in which all, three-quarters, or half of raters agreed, divided by the number of samples. For example, in K-2 Speaking, in 9 of the 10 rated samples, raters agreed (or disagreed) with local test administrators' ratings, i.e., 9/10=90%.

Full agreement was higher in lower grade-level clusters for the domain of Speaking. Across all grade-level clusters in the Speaking domain, 95% of raters either fully or three-quarters agreed in their ratings. That is, there was generally high agreement in expert raters' judgements of Speaking. Writing has similar findings except for the K-2 grade-level cluster. In this grade-level cluster, 8 of the 10 rated samples had half of the expert raters agree with the local test administrator, and half did not.

Table 10.3.2.4.a.**Agreement Rates Among Expert Raters at Benchmarking for Speaking Domain (i=10)**

Grade-Level Cluster	All Agree	3/4 Agree	1/2 Agree
K-2	90%	10%	0%
3-5	80%	10%	10%
6-8	60%	40%	0%
9-12	70%	20%	10%
Total	75%	20%	5%

Table 10.3.2.4.b.**Agreement Rates Among Expert Raters at Benchmarking for Writing Domain (i=10)**

Grade-Level Cluster	All Agree	3/4 Agree	1/2 Agree
K-2	0%	20%	80%
3-5	90%	0%	10%
6-8	60%	40%	0%
9-12	80%	20%	0%
Total	58%	20%	23%

After discussions, experts felt that clarification of the Expect Boxes was needed for Writing. For example, an Expect box for a K-2 Writing item at proficiency level 5 states, "The student writes a simple and compound sentence(s) with detail related to task." This expectation is confusing. Does it mean the student should write a simple sentence and a compound sentence, i.e., two sentences are necessary, or would a compound sentence suffice for a full mark? This was unclear. Experts felt this should read as, "The student writes two or more simple sentences or a compound sentence with detail related to task." Several other clarifications for Expect Boxes were identified and those clarifications were added to the WIDA rater training materials. Also, further clarity of grammatical terms and expectations was needed. These too were updated in the WIDA rater training materials.

10.3.2.5. WIDA Rater Training

All raters participating in WIDA rater training would have already completed the online training course. Three main topics were covered in the training. First, an overview of the study was presented. Second, raters needed to be trained in how to access and then record their individual ratings (and if desired, comments on their ratings) into online spreadsheets. Third, three exemplar student responses were shared that typified the clarification issues found during benchmarking. All raters were then given the opportunity to ask questions or seek clarification on how to rate student responses. Raters were then given three weeks to complete their rating tasks.

10.3.2.6 Ratings

After raters completed rating, score information was aggregated and summarized. Raters were also asked to comment on specific students' responses that were difficult to rate or were noteworthy in some fashion.

Table 10.3.2.6.a. identifies the average agreement rates and kappa coefficients across grade-level clusters. There was a high level of agreement (greater than 80%) for the domain of Speaking across all comparison groups. The level of agreement for Writing was slightly less, with a low of 75% agreement between the expert raters and local test administrators (local raters). Both Speaking and Writing Kappa coefficients are in the moderate range.

Table 10.3.2.6.a.

Agreement Rates across All Grade-Level Clusters

Domain	Average Agreement Rates: Expert Rater & WIDA Trained Rater	Average Agreement Rates: Expert Rater & Local Rater	Average Agreement Rates: WIDA Trained Rater & Local Rater	Kappa Coefficient: Expert Rater & WIDA Trained Rater
Speaking	88%	80%	84%	0.542
Writing	90%	75%	76%	0.566

Except for grades 9–12 Speaking, as seen in Table 10.3.2.6.b., all agreement rates are high between expert raters and WIDA trained raters. The agreement rates between expert raters and local raters tended to be the lowest of the comparison groups, which was hypothesized. Nonetheless, agreement rates in grade-level clusters 3–5 and 6–8 were high, and except for K–2 Writing, agreement rates were moderate to moderately high. The lower agreement rates in K–2 Writing can be attributed, in large part, to the lack of clarity in the Expect Boxes for items in that grade-level cluster. That issue has been addressed.

The Kappa coefficients for the K–2 grade-level cluster are extremely high. In the 9–12 grade-level cluster for Speaking, the Kappa coefficient was 0.000, indicating no agreement. Specifically, that rater did not differ at all from the local rater. In reviewing this raters’ ratings, errors in coding were discovered, so the observed Kappa coefficient should be viewed with skepticism.

Table 10.3.2.6.b.**Agreement Rates by Grade-Level Cluster, Domain, and Rater Group**

Grade-Level Cluster	Domain	Agreement Rates: Expert Rater & WIDA Trained Rater	Agreement Rates: Expert Rater & Local Rater	Agreement Rates: WIDA Trained Rater & Local Rater	Kappa Coefficient: Expert Rater & WIDA Trained Rater
K-2	Speaking	94%	70%	64%	0.865
K-2	Writing	96%	42%	42%	0.918
3-5	Speaking	90%	84%	86%	0.608
3-5	Writing	92%	92%	96%	0.296
6-8	Speaking	94%	92%	86%	0.696
6-8	Writing	86%	88%	82%	0.455
9-12	Speaking	74%	74%	100%	0.000
9-12	Writing	86%	78%	86%	0.596

There were 111 recorded rater comments. Forty-six comments were about the Speaking test responses and 65 were about the Writing test responses. There are four categories of comments: general, not complete, not interpretable, and not on task. The overwhelming majority of comments were about tasks being incomplete (not complete=96). An example of an incomplete comment was, "This one-word response doesn't meet expectations for this P3 task." Eight comments were about not being on task. There were four general comments, and three comments about tasks being not interpretable. Raters' comments were reviewed by the development team and used in helping to update training materials for the first operational administration (2023–2024).

10.3.2.7 Interrater Reliability Findings

Overall, the findings were quite positive. There were generally high levels of agreement on the assignment of Q1-Meets in the domains of Speaking and Writing among comparison groups. With one exception, Kappa coefficients between the expert and WIDA trained raters were moderate to high. Through discussions with and comments by expert and WIDA trained raters, new insights were gained on the scoring protocols, especially the Writing Expect Boxes. There was an identified need to develop definitions of grammatical terms that are found in the Expect Boxes. The information from this study was used to make slight revisions to Expect Boxes, and these changes became part of the operational version of Alternate ACCESS for the 2023–2024 administration. The Advancing ALTELLA test development team also updated the Test Administrator Manual so that key grammatical definitions are included to help local raters better interpret students' responses in Speaking and Writing.

10.3.3. Interrater Reliability Study for Reading and Listening

The Alternate ACCESS Reading and Listening tests measure receptive language skills and require raters to be present during test administrations to score. During the 2023–2024 operational administration of Alternate ACCESS, WIDA trained raters were sent to schools in member states to observe and concurrently score students' Reading and Listening test performances alongside local test administrators. WIDA trained raters were either State Educational Agency (SEA) staff who oversee the administration of Alternate ACCESS in their states, or WIDA staff. The purpose in sending WIDA raters was to collect data for interrater reliability analyses. This study shares results from that activity. The following sections describe the study's design, recruitment, rater qualifications, study sample, and interrater reliability results.

10.3.3.1 Study Design and Analysis

This study focuses specifically on the interrater reliability between local test administrators and WIDA trained raters. There were several steps to this study. The first step was to obtain University of Wisconsin's Institutional Review Board approval. Following approval or exemption, schools and test administrators willing to participate in the study were identified. Study and rater scoring materials were developed, and WIDA raters were trained to use them. WIDA raters then traveled to participating schools and observed and concurrently scored students' Alternate ACCESS test administrations. Local test administrators returned their scoring materials through the normal test administration process. WIDA raters returned their test materials to DRC (WIDA's production and scoring vendor), and their study and scoring materials to WIDA for processing. In WIDA rater scoring materials, students' DRC lithocode numbers were recorded. Lithocodes are anonymized identifiers of test booklets. They are unique to each student. After the Alternate ACCESS test administration, lithocode numbers of participating students and their associated Reading and Listening scores were obtained from DRC. Ratings by local test administrators and WIDA raters were linked through lithocode numbers, and a dataset was created for analyses. Three types of interrater reliability analyses were conducted between local test administrators and WIDA raters: descriptive statistic comparisons, total score correlations, and exact score agreement rates on rated test items.

10.3.3.2 Recruitment

WIDA member states were contacted in a variety of ways to identify school districts that were willing to participate. In districts and schools willing to participate, parents or guardians received an information letter describing the study. If parents declined permission, WIDA would not observe their children. Test administrators were asked if they would be willing to have WIDA raters scoring their students' Reading and Listening assessments. If they opted out, WIDA would not observe their students.

10.3.3.3 WIDA/SEA Rater Qualifications

Eight WIDA raters participated in this study. Four were English learner or special education specialist staff from WIDA member states. The remaining four raters were WIDA staff. All member state raters have master's degrees. One WIDA staff member has a Ph.D., two have master's degrees, and one has a bachelor's degree. All WIDA raters are highly qualified educators with degrees and experience in working with English learners, students with disabilities, and English learners with disabilities. They are leaders in their states or at WIDA. Several WIDA raters were heavily involved in the development of the updated version of Alternate ACCESS. All WIDA raters went through the Alternate ACCESS scorer training course. All raters participated in an additional training activity that described the interrater reliability study and their role in it.

10.3.3.4 Test Administrator Qualifications

Thirty-two local test administrators participated in this study. They came from seven WIDA member states: Florida (5), Idaho (1), Indiana (3), Massachusetts (1), Minnesota (10), North Dakota (7), and Utah (5). When WIDA raters visited their schools, they asked them for their qualifications. Information from 23 of the 32 (72%) test administrators was collected. The tables below highlight those qualifications. Nine educators (39%) had students in the K–2 cluster, eight (35%) in the 3–5 cluster, five (22%) in the 6–8 cluster, and one (4%) in the 9–12 cluster.

Table 10.3.3.4.a.

Participating Test Administrators' Current Positions

Position	Cluster K–2	Cluster 3–5	Cluster 6–8	Cluster 9–12	Total
Assessment Department	0	0	2	0	2
Assessment Department Assistant	0	0	1	0	1
District EL Coach	1	2	0	0	3
District EL support (coordinator)	1	2	0	0	3
EL/bilingual/Title III teacher	7	1	0	1	9
Testing Assistant	0	3	2	0	5
Grand Total	9	8	5	1	23

Three local test administrators report that they are in assessment departments. Six serve as support staff at the district level, and fourteen worked at schools.

Table 10.3.3.4.b.

Participating Test Administrators' Years of Experience

Experience	Cluster K–2	Cluster 3–5	Cluster 6–8	Cluster 9–12	Total
0–2 years	2	1	0	1	4
3–5 years	2	0	0	0	2
6–10 years	0	3	3	0	6
Over 10 years	5	4	2	0	11
Grand Total	9	8	5	1	23

Seventeen test (74%) test administrators had six or more years of experience as educators and slightly less than half had ten or more years of experience.

Table 10.3.3.4.c.**Participating Test Administrators' Highest Degree**

Highest Degree	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12	Total
Associate degree	0	3	2	0	5
Bachelor's degree	2	2	0	0	4
Master's degree	7	3	2	1	13
Grand Total	9	8	4	1	22

The majority (57%) of test administrators had master's degrees. Five (22%) had associate's degrees.

Table 10.3.3.4.d.**Participating Test Administrators' State Certifications**

State Certification	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12	Total
No	0	3	3	0	6
Yes - Both	4	0	0	1	5
Yes - ESL/Bilingual	5	5	0	0	10
Yes - Special Education	0	0	2	0	2
Grand Total	9	8	5	1	23

Almost three-quarters (74%) of test administrators reported that they had state certifications. Six did not record any state certifications.

Table 10.3.3.4.e.**Participating Test Administrators' Last Time Taking the Alternate ACCESS Training Course**

Last Alternate ACCESS Training Course Taken	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12	Total
Within the last 1-2 months	8	3	3	1	15
Within the last 3-6 months	1	4	2	0	7
Grand Total	9	7	5	1	22

Prior to administering Alternate ACCESS, WIDA strongly recommends that test administrators take the Alternate ACCESS scoring training course. Almost two-thirds (65%) report that they had taken this training within the last 1-2 months. The remaining third report that they took it within the last 3-6 months.

WIDA does not routinely collect test administrator qualifications during Alternate ACCESS test administrations. WIDA relies on state, district, and school staff to assure test administrators are qualified. Drawing upon WIDA's prior research on Alternate ACCESS raters' qualifications, the information reported above is consistent with what has been observed in the past.

10.3.3.5. Study sample

Fifty-nine students were jointly scored by local test administrators and WIDA raters. Nine students' records were removed because they had missing data or both sets of raters scored all items at 0, i.e., students' total Listening or Reading score was zero. Fifty students' scores were used in these analyses. Thirty-two students (64%) had both Listening and Reading scores. Thirteen (26%) had only Listening test scores and 5 (10%) had only Reading scores.

Table 10.3.12 shows the number of students rated by WIDA raters by state. WIDA raters from SEAs rated 23 students, and WIDA staff rated 27 students.

Table 10.3.3.5.a.

WIDA Raters by State and Students Sampled

Rater	State	Students
R1	FL	5
R2	ID	1
R3	IN	3
R4	MN	3
R5	MA	3
R6	MN	10
R7	ND	16
R8	UT	9

Table 10.3.3.5.b. shows the number of test administrators by state. What may become apparent from Table 10.3.3.5.a. and Table 10.3.3.5.b. is that many WIDA raters rated only one student with one test administrator. Eight (25%) test administrators co-rated students with four (50%) WIDA raters. Twenty-six students in this sample were rated by both a WIDA rater and a local test administrator.

Table 10.3.3.5.b.

Number of Test Administrators by State

State	TAs
FL	5
ID	1
IN	3
MA	1
MN	10
ND	7
UT	5
Total	32

Table 10.3.3.5.c. shows the number of students by state. Slightly over three-quarters of rated students came from MN, ND, and UT.

Table 10.3.3.5.c.**Number of Sampled Students by State**

State	Students
FL	5
ID	1
IN	3
MA	3
MN	13
ND	16
UT	9
Total	50

Table 10.3.3.5.d. shows the number of students rated by state and grade-level cluster across domains. Table 10.3.3.5.e. and Table 10.3.3.5.f. show the number of students rated by WIDA raters and local test administrators in Listening (45 students) and Reading (37 students) respectively.

Table 10.3.3.5.d.**Number of Students Rated by Raters Across Domains by Grade-Level Cluster**

State	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12	Total
FL	0	0	5	0	5
ID	0	0	1	0	1
IN	0	0	0	3	3
MA	1	0	0	2	3
MN	4	3	2	4	13
ND	5	3	5	3	16
UT	1	4	4	0	9
Total	11	10	17	12	50

Table 10.3.3.5.e.**Students Sampled by State and Grade-Level Cluster: Listening**

State	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12	Total
FL	0	0	5	0	5
ID	0	0	1	0	1
IN	0	0	0	3	3
MA	1	0	0	2	3
MN	2	3	1	3	9
ND	5	3	4	3	15
UT	1	4	4	0	9
Total	9	10	15	11	45

Table 10.3.3.5.f.**Students Sampled by State and Grade-Level Cluster: Reading**

State	Cluster K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12	Total
FL	0	0	5	0	5
ID	0	0	1	0	1
IN	0	0	0	2	2
MA	0	0	0	2	2
MN	2	0	1	3	6
ND	3	3	4	3	13
UT	1	4	3	0	8
Total	6	7	14	10	37

In both the domains, the student sample has slightly more secondary students than primary students. Interrater reliability analyses were conducted with the above student sample. The next section provides findings from those analyses.

10.3.3.6. Interrater reliability findings

Prior to discussing findings, it is important to mention what WIDA raters observed during testing. Part of the study's protocol was for WIDA raters to write down any notable test administration observations. Several comments were made about test administration irregularities. The most frequent comment had to do with test administrators not repeating Cue A for items, which is not an egregious error. Not repeating Cue A (a part of the test administrator's script) does limit a student's ability to show what they know. Seven students were mentioned to be challenging, i.e., students were easily distracted or difficult to keep on task. Two test administrators stated that they were not familiar with test materials. Two comments were made about room distractions, and one student had a health issue. We view these irregularities as minor and not likely to affect rater judgments or rater agreement rates.

Three analyses are presented below: descriptive statistic comparisons, correlations, and exact agreement rates. The two tables below display Listening and Reading test descriptive statistic comparisons and correlations between WIDA raters and local test administrators. The first column shows grade-level clusters. Next is the number of students jointly scored by raters (N). The third and fourth columns show the average raw scores for each group of raters, and the last column shows the correlations between raters. Note that samples sizes within each grade-level cluster for both domains is small; accordingly, strong inferences about the similarities or differences between groups should be made with caution.

Table 10.3.3.6.a.

Listening Average Raw Scores by Grade-Level Cluster and Rater Group

Cluster	N	Average WIDA Raters	Average Local Test Administrator	Correlation
K-2	9	26.7	26.7	0.991
3-5	10	26.8	26.7	0.990
6-8	15	25.7	25.6	0.981
9-12	11	24.0	23.9	0.997
All	45	28.5	29.0	0.988

The average Listening scores between rater groups are similar. The largest difference is the aggregate scores between raters, and even there it is only 0.5 raw score points. Likewise, the correlations are very high.

Table 10.3.3.6.b.

Reading Average Raw Scores by Grade-Level Cluster and Rater Group

Cluster	N	Average WIDA Raters	Average Local Test Administrator	Correlation
K-2	6	22.8	19.4	0.989
3-5	7	21.9	18.7	0.998
6-8	14	21.7	18.2	0.947
9-12	10	22.3	19.1	0.887
All	37	28.0	25.2	0.947

WIDA raters' average Reading raw scores are around three points higher than those of local test administrators. The correlations are high with the lowest being 0.887 for the 9-12 grade-level cluster.

While there are differences between raters in Reading, all correlations in both domains are high, ranging from a low of $r_{xy}=0.887$ in 9-12 Reading to a high of $r_{xy}=0.997$ in 9-12 Listening. This suggests that both groups of raters rate similarly (especially in Listening).

Another way to determine interrater reliability is to calculate the exact agreement rate on each item. Agreement rates are typically calculated between common sets of raters. For example, the number of students jointly rated on the 3–5 Listening test is ten. Ideally, the same WIDA rater and local test administrator would rate the same ten students. This is not the case here. Those ten students came from three different states (MN, ND, and UT) and represent seven different WIDA rater and local test administrator pairings. Only two of the seven pairings were on more than one student. Because of the small sample and the different pairings, the agreement rates in the tables below represent comparisons between classes of raters, i.e., WIDA trained raters and locally trained test administrators. We are assuming that WIDA trained raters and local test administrators each share common characteristics. This is a reasonable assumption for WIDA raters because they all went through the same training. That is likely not so with local test administrators. Given this assumption, we might expect to see large variability in agreement rates. That is not what was observed.

To compare agreement rates, we used the following criteria. Exact agreement rates at or above 80% are defined as acceptable. Agreement rates between 60% to 79% are marginally acceptable, and agreement rates less than 60% are considered not acceptable.

Table 10.3.3.6.c.

Exact Agreement Rates: Listening

Item	Cluster K–2	Cluster 3–5	Cluster 6–8	Cluster 9–12	Total
1	89%	70%	93%	91%	87%
2	78%	80%	93%	100%	89%
3	78%	90%	93%	100%	89%
4	89%	80%	93%	91%	89%
5	100%	80%	93%	91%	91%
6	78%	70%	100%	100%	89%
7	100%	90%	87%	100%	93%
8	89%	100%	87%	91%	91%
9	100%	90%	80%	82%	87%
10	100%	90%	87%	82%	89%

Five items are below the 80% acceptable criterion, but all those items are in the marginally acceptable category. Three marginally acceptable agreement rates are in the K–2 grade-level cluster, and two are in the 3–5 cluster. All other items across grade-level clusters are at acceptable rates.

Table 10.3.3.6.d.**Exact Agreement Rates: Reading**

Item	K-2	Cluster 3-5	Cluster 6-8	Cluster 9-12	Total
1	100%	86%	93%	70%	87%
2	100%	100%	93%	80%	92%
3	83%	71%	79%	90%	82%
4	100%	100%	93%	90%	95%
5	100%	71%	93%	80%	87%
6	83%	100%	93%	80%	89%
7	83%	100%	93%	90%	92%
8	67%	86%	100%	90%	89%
9	100%	100%	86%	90%	92%
10	100%	100%	93%	90%	95%

Five items have marginally acceptable agreement rates. There is one marginally acceptable item in each grade-level cluster, with the 3–5 cluster having two. The remainder of agreement rates are acceptable.

No item’s agreement rate was unacceptable. This suggests that there is good agreement between WIDA raters and local test administrators. However, making strong generalizations about high agreement across all raters of Alternate ACCESS Listening and Reading tests is not warranted. The number of raters represented, and students sampled, was small. It is encouraging to see how few items were not at an acceptable agreement rate, and it indicates that the training materials and scoring resources can be used effectively to obtain acceptable agreement.

We recommended that WIDA’s Alternate ACCESS test development team review the items identified below the acceptable level. There may be obvious improvements to scorer training on these items. Encouragingly, the score comparisons, correlations, and exact score agreement rates indicate high interrater reliability for raters who participated in this study.

10.4. Accuracy and Consistency of Domains/Composites

10.4.1. Classification Accuracy and Consistency

For each domain across grade-level clusters, as well as for the four composite scores, tables were produced that indicate estimates of the accuracy and consistency of classification of examinees into the Alternate ACCESS Proficiency Levels based on their performances on the test. It is important to know the reliability of any student's test score and the degree of precision with which it has been measured (i.e., the estimate of the invariant standard error of measure [SEM] of classical test theory and the estimate of the variable conditional standard error of the Rasch measurement model). However, because decisions about students are ultimately made based on their classification into language proficiency levels using their performance on Alternate ACCESS, it is important to know how well these classifications are made. The analyses that we employed make use of the methods outlined and implemented in Livingston and Lewis (1995) and Young and Yoon (1998), as implemented in the software program BB-CLASS (Brennan, 2004) (cf. also Lee, Hanson, & Brennan, 2002).

In the approach of Livingston and Lewis (1995), the accuracy of a decision is the extent to which decisions made on the basis of the administered test (i.e., the observed scores) would agree with the decisions that would be made if each student could somehow be tested with all possible parallel forms of the assessments; that is, decisions based on the examinees' "true score." On the other hand, the consistency of a decision is the extent to which decisions made on the administered test would agree with the decisions that would be made if the students had taken a different but parallel form of the test. Thus, in every analysis of classification, two parallel analyses are made: accuracy (that is, vis-à-vis "true scores") and consistency (that is, vis-à-vis a second form).

In terms of classifications around a single cut point, students can be misclassified in one of two ways. Students who were below the proficiency cut score (based on their "true score") but were classified on the basis of the assessment as being above the cut score, are considered to be false positives. Students who were above the proficiency cut score (based on their "true score"), but were classified as being below a cut score, are considered to be false negatives. All other students are considered to be accurately placed either above or below the cut score.

Since a "true score" is a theoretical construct, it is unknown for any given student. The approach taken by Livingston and Lewis (1995) and implemented here to model true scores uses information about the reliability of the test, the cut scores, and the observed distribution of scores. Then, using a four-parameter beta distribution, we modeled the distribution of the true scores and of scores on a parallel form. Overall accuracy and consistency indices are produced by comparing the percentage of students classified across all categories the same way by both the observed distribution and modeled distribution. These indices indicate the percent of all students who would be classified into the same language proficiency level by both the administered test and either the true score distribution (accuracy) or a parallel test (consistency). Our tables also provide an estimate of Cohen's kappa statistic, which is a very conservative estimate of the overall classification since it corrects for chance.

We also look at accuracy and consistency conditioned on the language proficiency level. These indices examine the percent of students classified into a level divided by all students classified into that level according either to the true score distribution (accuracy) or based on a parallel test (consistency).

Finally, we look at what may be the most important set of indices, which are the indices at the cut points. That is, at every cut point, using the true score distribution (e.g., accuracy), we provide the percent of students who are consistently placed above and below the cut score, as well as those who are false positives and false negatives. For consistency, only the percent of students classified consistently above and below the cut score is calculated. Thus, for example, to evaluate the degree of confidence that one can have in a decision made based on the Overall Composite score as to whether students are being accurately classified into Alternate WIDA language proficiency level P2 ("Beginning") or not, one can look at the accuracy index provided in the table for the cut score P1/P2.

The tables in Sections 10.4.2 through 10.4.9 present information related to the accuracy and consistency of placement into proficiency categories based on Alternate ACCESS (see above). The first table of each grade-level cluster series (designated as "a") provides overall indices related to the accuracy and consistency of classification, as well as Cohen's kappa. The second table ("b") shows accuracy and consistency information conditional on level. The third table ("c") provides indices of classification accuracy and consistency at the cut points. These indices are perhaps the most important of all when using any of these as an absolute cut- point for placement decisions. Note that the consistency is generally higher at the cut points than over the levels. For practical purposes, the primary scores used for such decisions are the Overall Composite scores. In general, the accuracy and consistency of classification for the Overall Composite reach 0.7 and 0.75, respectively, indicating that 70–75% of classifications are accurate and consistent.

10.4.1.1. Overall Classification Accuracy and Consistency

Overall classification accuracy indicates the percentage of all students whom we would classify into the same language proficiency level by both their domain scale scores and their true scale scores (i.e., the percentage of students whom we accurately classified). Overall classification consistency indicates the percentage of all students whom we would classify into the same language proficiency levels by their performances on both the administered test and on a parallel test.

10.4.1.2. Marginal Classification Accuracy and Consistency at the Cut Points

Overall classification accuracy and consistency indices indicate the degree to which we accurately and consistently classify students into the same WIDA proficiency levels, but not the degree to which we accurately or consistently classify students into the proficiency levels below or above the specific cut point (e.g., at the P4/P5 cut point). The indices that can address this question are marginal classification accuracy and consistency indices based on domain scale scores at the cut points. From an accountability perspective, the most important indices for test users and policy makers to examine are the marginal classification accuracy and consistency indices, because they show how reliably and consistently a test assigns students to correct PL categories at specific cut points. Specially, it is important for decision-making at the exiting PL to ensure a student receives an appropriate level of support. To help decision makers interpret results, we report the range of the marginal classification accuracy and consistency indices for each domain across grades and then highlight the grades (and the cut points within those grades) that had the lowest marginal classification accuracy and the lowest classification consistency. Highlighting the grades and cut points with the lowest marginal classification accuracy and consistency ensures that any vulnerabilities in the test's classification decisions are transparent and can be addressed appropriately, supporting fair and effective accountability measures.

Assessment experts have issued little guidance to aid in making judgments about the ideal or expected levels of decision consistency and accuracy needed for educational assessments since many different factors affect the calculation of these indices, as discussed earlier. To help test users and policy makers interpret the results from our classification analyses, for each of the test domains, we report the range of the overall classification accuracy and consistency indices across grades. Additionally, we highlight the grade with the lowest classification accuracy and consistency indices. Since the overall accuracy and consistency indices are summaries of the degree of classification accuracy and consistency across all proficiency level cut points, we also report the marginal classification accuracy and consistency indices for these grades to identify the specific source(s) of low classification accuracy and consistency.

The marginal classification accuracy indices based on domain scale scores at the cut points report the percentage of students whom we accurately placed into proficiency levels above and below each cut point based on their domain scale scores. The marginal classification consistency indices based on domain scale scores at the cut points report the percentage of students whom we would classify consistently above and below each cut point based on their domain scale scores.

Note that the marginal accuracy and consistency indices are generally higher for students' domain scale scores at the cut points than are the overall classification accuracy and consistency indices (Livingston et al., 2018). This is because the marginal accuracy and consistency indices report the classification decisions at one cut point at a time while the overall accuracy and consistency indices report the classification decisions at all five cut points at the same time.

10.4.2. Listening Accuracy and Consistency

As shown in Tables 10.4.2.1.a through 10.4.2.4.c, overall classification accuracy for Listening ranged from 0.568 to 0.693, and overall consistency ranged from 0.504 to 0.612, with kappa values between 0.36 and 0.476. Grade-level cluster 3–5 had the lowest overall classification accuracy and consistency. The marginal classification accuracy for Listening at cut points ranged from 0.84 to 0.942, and consistency ranged from 0.772 to 0.917. Grade-level cluster 3–5 at the P3/4 cut point showed the lowest marginal accuracy and consistency indices.

10.4.2.1 Grade-Level Cluster K–2

Table 10.4.2.1.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.657	0.568	0.399

Table 10.4.2.1.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.933	0.884
P2	0.307	0.193
P3	0.314	0.227
P4	0.531	0.494
P5	NA	0.203

Table 10.4.2.1.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.933	0.027	0.04	0.903
P2/P3	0.903	0.033	0.065	0.854
P3/P4	0.84	0.045	0.115	0.776
P4/P5	0.911	0.089	0	0.858

10.4.2.2. Grade-Level Cluster 3–5**Table 10.4.2.2.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.586	0.504	0.359

Table 10.4.2.2.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.921	0.865
P2	0.448	0.305
P3	0.37	0.257
P4	0.462	0.431
P5	NA	0.309

Table 10.4.2.2.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.942	0.024	0.035	0.917
P2/P3	0.909	0.033	0.058	0.865
P3/P4	0.84	0.046	0.113	0.772
P4/P5	0.852	0.148	0	0.795

10.4.2.3. Grade-Level Cluster 6–8**Table 10.4.2.3.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.693	0.612	0.476

Table 10.4.2.3.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.893	0.84
P2	0.386	0.282
P3	0.446	0.326
P4	0.267	0.188
P5	0.805	0.753

Table 10.4.2.3.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.936	0.03	0.034	0.91
P2/P3	0.922	0.035	0.043	0.889
P3/P4	0.896	0.046	0.058	0.851
P4/P5	0.871	0.051	0.078	0.815

10.4.2.4 Grade-Level Cluster 9–12

Table 10.4.2.4.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.648	0.561	0.425

Table 10.4.2.4.b.

Accuracy and Consistency of Conditional on Level Indices

Level	Accuracy	Consistency
P1	0.897	0.846
P2	0.346	0.25
P3	0.452	0.326
P4	0.39	0.296
P5	0.725	0.645

Table 10.4.2.4.c.

Accuracy and Consistency of Indices at Cut Points

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.938	0.029	0.033	0.912
P2/P3	0.924	0.035	0.041	0.892
P3/P4	0.893	0.048	0.059	0.845
P4/P5	0.842	0.068	0.089	0.784

10.4.3. Reading Accuracy and Consistency

As shown in Tables 10.4.3.1.a through 10.4.3.4.c, the overall classification accuracy for Reading ranged from 0.646 to 0.669, with consistency between 0.504 and 0.612, and kappa between 0.39 and 0.408. Grade-level cluster 3–5 had the lowest indices. The marginal classification accuracy at cut points for Reading ranged from 0.839 to 0.93, while consistency ranged from 0.788 to 0.901. Grade-level cluster 3–5 at the P3/4 cut point had the lowest indices.

10.4.3.1. Grade-Level Cluster K–2

Table 10.4.3.1.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.669	0.62	0.39

Table 10.4.3.1.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.934	0.889
P2	0.324	0.229
P3	0.348	0.303
P4	0.326	0.274
P5	NA	0.197

Table 10.4.3.1.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.912	0.037	0.051	0.873
P2/P3	0.882	0.048	0.07	0.831
P3/P4	0.852	0.079	0.069	0.819
P4/P5	0.937	0.063	0	0.899

10.4.3.2. Grade-Level Cluster 3–5**Table 10.4.3.2.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.646	0.567	0.394

Table 10.4.3.2.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.922	0.869
P2	0.373	0.259
P3	0.398	0.315
P4	0.495	0.435
P5	NA	0.157

Table 10.4.3.2.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.919	0.033	0.048	0.885
P2/P3	0.889	0.045	0.066	0.839
P3/P4	0.839	0.062	0.099	0.788
P4/P5	0.946	0.054	0	0.908

10.4.3.3. Grade-Level Cluster 6–8**Table 10.4.3.3.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.656	0.562	0.408

Table 10.4.3.3.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.912	0.86
P2	0.332	0.233
P3	0.394	0.276
P4	0.599	0.553
P5	NA	0.254

Table 10.4.3.3.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.93	0.031	0.039	0.901
P2/P3	0.914	0.036	0.05	0.876
P3/P4	0.871	0.053	0.077	0.814
P4/P5	0.901	0.099	0	0.852

10.4.3.4. Grade-Level Cluster 9–12

Table 10.4.3.4.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.667	0.587	0.45

Table 10.4.3.4.b.

Accuracy and Consistency of Conditional on Level Indices

Level	Accuracy	Consistency
P1	0.904	0.85
P2	0.311	0.227
P3	0.364	0.269
P4	0.557	0.449
P5	0.72	0.603

Table 10.4.3.4.c.

Accuracy and Consistency of Indices at Cut Points

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.922	0.034	0.044	0.891
P2/P3	0.915	0.038	0.047	0.88
P3/P4	0.903	0.048	0.049	0.863
P4/P5	0.881	0.065	0.054	0.838

10.4.4. Speaking Accuracy and Consistency

For Speaking, Tables 10.4.4.1.a. through 10.4.4.4.c. show overall accuracy ranging from 0.592 to 0.712, with consistency ranging from 0.528 to 0.664, and kappa between 0.347 and 0.457. Grade-level cluster 6–8 showed the lowest overall classification indices. The marginal accuracy for Speaking ranged from 0.807 to 0.932, and consistency ranged from 0.751 to 0.964. Grade-level cluster K–2 at the P2/3 cut point had the lowest marginal accuracy, while cluster 6–8 at P3/4 showed the lowest marginal consistency.

10.4.4.1. Grade-Level Cluster K–2

Table 10.4.4.1.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.712	0.664	0.361

Table 10.4.4.1.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.948	0.862
P2	0.299	0.269
P3	0.411	0.37
P4	0	0.075
P5	NA	0.05

Table 10.4.4.1.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.876	0.03	0.093	0.82
P2/P3	0.816	0.07	0.113	0.787
P3/P4	0.952	0.048	0	0.916
P4/P5	0.982	0.018	0	0.965

10.4.4.2. Grade-Level Cluster 3–5**Table 10.4.4.2.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.605	0.562	0.347

Table 10.4.4.2.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.939	0.863
P2	0.315	0.233
P3	0.291	0.273
P4	0	0.257
P5	NA	0.156

Table 10.4.4.2.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.906	0.029	0.065	0.858
P2/P3	0.836	0.045	0.119	0.773
P3/P4	0.807	0.193	0	0.773
P4/P5	0.932	0.068	0	0.885

10.4.4.3. Grade-Level Cluster 6–8**Table 10.4.4.3.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.592	0.528	0.348

Table 10.4.4.3.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.932	0.868
P2	0.366	0.245
P3	0.344	0.326
P4	0	0.24
P5	NA	0.19

Table 10.4.4.3.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.923	0.029	0.049	0.884
P2/P3	0.849	0.042	0.108	0.781
P3/P4	0.783	0.217	0	0.751
P4/P5	0.91	0.09	0	0.854

10.4.4.4. Grade-Level Cluster 9–12

Table 10.4.4.4.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.691	0.627	0.457

Table 10.4.4.4.b.

Accuracy and Consistency of Conditional on Level Indices

Level	Accuracy	Consistency
P1	0.92	0.881
P2	0.31	0.22
P3	0.294	0.207
P4	0.232	0.181
P5	0.708	0.619

Table 10.4.4.4.c.

Accuracy and Consistency of Indices at Cut Points

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.919	0.038	0.043	0.886
P2/P3	0.908	0.043	0.048	0.869
P3/P4	0.887	0.048	0.065	0.839
P4/P5	0.867	0.059	0.074	0.818

10.4.5. Writing Accuracy and Consistency

As seen in Tables 10.4.5.1.a. through 10.4.5.4.c., Writing showed an overall classification accuracy range of 0.7 to 0.784 and consistency ranging from 0.627 to 0.741, with kappa between 0.441 and 0.456. Grade-level cluster 9–12 had the lowest overall indices. The marginal classification accuracy ranged from 0.892 to 0.95, and consistency from 0.849 to 0.948. Grade-level cluster 9–12 at the P3/4 cut point showed the lowest indices for both accuracy and consistency.

10.4.5.1. Grade-Level Cluster K–2

Table 10.4.5.1.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.784	0.741	0.456

Table 10.4.5.1.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.945	0.924
P2	0.384	0.287
P3	0.385	0.293
P4	0.415	0.33
P5	NA	0.291

Table 10.4.5.1.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.923	0.039	0.038	0.892
P2/P3	0.929	0.04	0.031	0.9
P3/P4	0.935	0.039	0.026	0.91
P4/P5	0.963	0.037	0	0.948

10.4.5.2. Grade-Level Cluster 3–5**Table 10.4.5.2.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.731	0.671	0.442

Table 10.4.5.2.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.932	0.901
P2	0.464	0.35
P3	0.348	0.275
P4	0.451	0.367
P5	NA	0.191

Table 10.4.5.2.c.

Accuracy and Consistency of Indices at Cut Points

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.916	0.041	0.043	0.881
P2/P3	0.907	0.047	0.047	0.868
P3/P4	0.904	0.05	0.046	0.871
P4/P5	0.964	0.036	0	0.941

10.4.5.3. Grade-Level Cluster 6–8

Table 10.4.5.3.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.703	0.635	0.441

Table 10.4.5.3.b.

Accuracy and Consistency of Conditional on Level Indices

Level	Accuracy	Consistency
P1	0.927	0.89
P2	0.425	0.317
P3	0.38	0.289
P4	0.519	0.439
P5	NA	0.232

Table 10.4.5.3.c.

Accuracy and Consistency of Indices at Cut Points

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.917	0.039	0.045	0.882
P2/P3	0.906	0.047	0.047	0.867
P3/P4	0.895	0.056	0.049	0.854
P4/P5	0.95	0.05	0	0.924

10.4.5.4. Grade-Level Cluster 9–12

Table 10.4.5.4.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.7	0.627	0.441

Table 10.4.5.4.b.

Accuracy and Consistency of Conditional on Level Indices

Level	Accuracy	Consistency
P1	0.923	0.884
P2	0.414	0.308
P3	0.377	0.283
P4	0.556	0.478
P5	NA	0.219

Table 10.4.5.4.c.

Accuracy and Consistency of Indices at Cut Points

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.916	0.039	0.045	0.881
P2/P3	0.907	0.046	0.047	0.869
P3/P4	0.892	0.056	0.053	0.849
P4/P5	0.948	0.052	0	0.919

10.4.6. Oral Composite Accuracy and Consistency

Tables 10.4.6.1.a. through 10.4.6.4.c. show Oral Composite accuracy ranging from 0.623 to 0.698, and consistency from 0.551 to 0.64. Grade-level cluster 6–8 had the lowest accuracy, while cluster 3–5 had the lowest consistency. The marginal accuracy for Oral Composite ranged from 0.85 to 0.923, with consistency between 0.801 and 0.884. Grade-level cluster 6–8 at P4/5 had the lowest accuracy, while cluster 3–5 at P3/4 had the lowest consistency.

10.4.6.1. Grade-Level Cluster K–2

Table 10.4.6.1.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.698	0.64	0.414

Table 10.4.6.1.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.928	0.889
P2	0.362	0.259
P3	0.418	0.37
P4	0.386	0.303
P5	NA	0.11

Table 10.4.6.1.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.911	0.041	0.048	0.872
P2/P3	0.885	0.046	0.068	0.837
P3/P4	0.875	0.091	0.034	0.846
P4/P5	0.971	0.029	0	0.949

10.4.6.2. Grade-Level Cluster 3–5**Table 10.4.6.2.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.625	0.559	0.388

Table 10.4.6.2.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.921	0.87
P2	0.433	0.311
P3	0.361	0.283
P4	0.4	0.351
P5	NA	0.271

Table 10.4.6.2.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.918	0.034	0.048	0.884
P2/P3	0.89	0.045	0.065	0.842
P3/P4	0.852	0.06	0.088	0.801
P4/P5	0.91	0.09	0	0.869

10.4.6.3. Grade-Level Cluster 6–8**Table 10.4.6.3.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.623	0.551	0.403

Table 10.4.6.3.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.912	0.858
P2	0.384	0.28
P3	0.453	0.338
P4	0.277	0.232
P5	0.615	0.505

Table 10.4.6.3.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.923	0.033	0.044	0.891
P2/P3	0.905	0.044	0.052	0.864
P3/P4	0.866	0.061	0.073	0.812
P4/P5	0.85	0.076	0.074	0.807

10.4.6.4. Grade-Level Cluster 9–12

Table 10.4.6.4.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.666	0.593	0.45

Table 10.4.6.4.b.

Accuracy and Consistency of Conditional on Level Indices

Level	Accuracy	Consistency
P1	0.904	0.856
P2	0.362	0.269
P3	0.38	0.28
P4	0.362	0.274
P5	0.745	0.644

Table 10.4.6.4.c.

Accuracy and Consistency of Indices at Cut Points

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.918	0.038	0.044	0.884
P2/P3	0.907	0.048	0.045	0.869
P3/P4	0.896	0.052	0.052	0.853
P4/P5	0.884	0.059	0.057	0.839

10.4.7. Literacy Composite Accuracy and Consistency

For Literacy Composite, which is shown in Tables 10.4.7.1.a. through 10.4.7.4.c., overall accuracy ranged from 0.684 to 0.769, and consistency from 0.609 to 0.717. Grade-Level cluster 9–12 had the lowest indices. The marginal accuracy ranged from 0.89 to 0.97, with consistency from 0.846 to 0.954. Grade-level cluster 6–8 at P3/4 had the lowest marginal accuracy and consistency.

10.4.7.1. Grade-Level Cluster K–2

Table 10.4.7.1.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.769	0.717	0.451

Table 10.4.7.1.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.934	0.913
P2	0.419	0.315
P3	0.409	0.318
P4	0.437	0.339
P5	NA	0.24

Table 10.4.7.1.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.917	0.045	0.039	0.883
P2/P3	0.92	0.044	0.036	0.887
P3/P4	0.932	0.043	0.025	0.906
P4/P5	0.97	0.03	0	0.954

10.4.7.2. Grade-Level Cluster 3–5**Table 10.4.7.2.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.718	0.649	0.445

Table 10.4.7.2.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.925	0.887
P2	0.466	0.358
P3	0.411	0.322
P4	0.524	0.423
P5	NA	0.167

Table 10.4.7.2.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.911	0.041	0.048	0.874
P2/P3	0.903	0.052	0.045	0.863
P3/P4	0.903	0.054	0.043	0.868
P4/P5	0.972	0.028	0	0.953

10.4.7.3. Grade-Level Cluster 6–8**Table 10.4.7.3.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.688	0.609	0.44

Table 10.4.7.3.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.915	0.869
P2	0.418	0.313
P3	0.394	0.297
P4	0.601	0.521
P5	NA	0.229

Table 10.4.7.3.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.913	0.039	0.048	0.878
P2/P3	0.903	0.046	0.051	0.862
P3/P4	0.89	0.058	0.052	0.846
P4/P5	0.949	0.051	0	0.922

10.4.7.4. Grade-Level Cluster 9–12

Table 10.4.7.4.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.684	0.609	0.45

Table 10.4.7.4.b.

Accuracy and Consistency of Conditional on Level Indices

Level	Accuracy	Consistency
P1	0.908	0.865
P2	0.395	0.297
P3	0.373	0.277
P4	0.589	0.501
P5	0.612	0.433

Table 10.4.7.4.c.

Accuracy and Consistency of Indices at Cut Points

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.913	0.042	0.045	0.877
P2/P3	0.906	0.048	0.045	0.868
P3/P4	0.905	0.048	0.046	0.867
P4/P5	0.922	0.054	0.024	0.896

10.4.8. Comprehension Composite Accuracy and Consistency

For the Comprehension Composite, Tables 10.4.8.1.a. through 10.4.8.4.c. show overall accuracy ranging from 0.632 to 0.671, and consistency between 0.538 and 0.601. Grade-level cluster 6–8 had the lowest overall indices. The marginal accuracy ranged from 0.842 to 0.928, while consistency ranged from 0.781 to 0.913. Grade-level cluster 3–5 at P3/4 showed the lowest marginal accuracy and consistency.

10.4.8.1. Grade-Level Cluster K–2

Table 10.4.8.1.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.659	0.601	0.391

Table 10.4.8.1.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.932	0.883
P2	0.337	0.236
P3	0.357	0.302
P4	0.388	0.331
P5	NA	0.176

Table 10.4.8.1.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.914	0.035	0.051	0.875
P2/P3	0.881	0.047	0.072	0.829
P3/P4	0.846	0.068	0.086	0.807
P4/P5	0.943	0.057	0	0.906

10.4.8.2. Grade-Level Cluster 3–5**Table 10.4.8.2.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.671	0.576	0.414

Table 10.4.8.2.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.918	0.866
P2	0.4	0.275
P3	0.314	0.23
P4	0.612	0.56
P5	NA	0.134

Table 10.4.8.2.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.927	0.031	0.042	0.895
P2/P3	0.893	0.045	0.063	0.843
P3/P4	0.843	0.053	0.104	0.781
P4/P5	0.95	0.05	0	0.913

10.4.8.3. Grade-Level Cluster 6–8**Table 10.4.8.3.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.632	0.538	0.387

Table 10.4.8.3.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.91	0.855
P2	0.348	0.244
P3	0.41	0.284
P4	0.563	0.522
P5	NA	0.288

Table 10.4.8.3.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.935	0.028	0.037	0.907
P2/P3	0.917	0.034	0.049	0.881
P3/P4	0.866	0.052	0.082	0.806
P4/P5	0.877	0.123	0	0.825

10.4.8.4. Grade-Level Cluster 9–12

Table 10.4.8.4.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.643	0.558	0.423

Table 10.4.8.4.b.

Accuracy and Consistency of Conditional on Level Indices

Level	Accuracy	Consistency
P1	0.903	0.845
P2	0.33	0.241
P3	0.433	0.32
P4	0.516	0.416
P5	0.662	0.56

Table 10.4.8.4.c.

Accuracy and Consistency of Indices at Cut Points

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.928	0.031	0.041	0.898
P2/P3	0.916	0.036	0.048	0.882
P3/P4	0.896	0.046	0.058	0.852
P4/P5	0.862	0.064	0.074	0.816

10.4.9. Overall Composite

Tables 10.4.9.1.a. through 10.4.9.4.c. present Overall Composite accuracy ranging from 0.665 to 0.755, and consistency ranging from 0.585 to 0.702. Grade-level cluster 6–8 had the lowest overall indices. The marginal accuracy for the Overall Composite ranged from 0.877 to 0.92, with consistency from 0.83 to 0.962. Grade-level cluster 6–8 at P3/4 showed the lowest indices for both accuracy and consistency.

10.4.9.1. Grade-Level Cluster K–2

Table 10.4.9.1.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.755	0.702	0.444

Table 10.4.9.1.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.937	0.907
P2	0.403	0.302
P3	0.426	0.347
P4	0.435	0.33
P5	NA	0.152

Table 10.4.9.1.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.914	0.041	0.046	0.878
P2/P3	0.91	0.044	0.045	0.874
P3/P4	0.92	0.051	0.03	0.893
P4/P5	0.978	0.022	0	0.962

10.4.9.2. Grade-Level Cluster 3–5**Table 10.4.9.2.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.69	0.622	0.424

Table 10.4.9.2.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.923	0.88
P2	0.442	0.33
P3	0.418	0.34
P4	0.473	0.38
P5	NA	0.162

Table 10.4.9.2.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.91	0.039	0.05	0.873
P2/P3	0.894	0.048	0.058	0.85
P3/P4	0.885	0.066	0.049	0.848
P4/P5	0.965	0.035	0	0.941

10.4.9.3. Grade-Level Cluster 6–8**Table 10.4.9.3.a.****Accuracy and Consistency of Overall Indices**

Accuracy	Consistency	Kappa (k)
0.665	0.585	0.42

Table 10.4.9.3.b.**Accuracy and Consistency of Conditional on Level Indices**

Level	Accuracy	Consistency
P1	0.915	0.865
P2	0.391	0.288
P3	0.459	0.352
P4	0.533	0.461
P5	NA	0.239

Table 10.4.9.3.c.**Accuracy and Consistency of Indices at Cut Points**

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.916	0.037	0.048	0.881
P2/P3	0.901	0.047	0.051	0.86
P3/P4	0.877	0.062	0.061	0.83
P4/P5	0.938	0.062	0	0.906

10.4.9.4. Grade-Level Cluster 9–12

Table 10.4.9.4.a.

Accuracy and Consistency of Overall Indices

Accuracy	Consistency	Kappa (k)
0.668	0.592	0.436

Table 10.4.9.4.b.

Accuracy and Consistency of Conditional on Level Indices

Level	Accuracy	Consistency
P1	0.906	0.862
P2	0.366	0.273
P3	0.435	0.326
P4	0.528	0.438
P5	0.612	0.458

Table 10.4.9.4.c.

Accuracy and Consistency of Indices at Cut Points

Cut Point	Accuracy	Accuracy - False Positives	Accuracy - False Negatives	Consistency
P1/P2	0.914	0.041	0.046	0.878
P2/P3	0.907	0.05	0.044	0.869
P3/P4	0.9	0.049	0.051	0.859
P4/P5	0.91	0.054	0.036	0.88

11. Quality Control

11.1. Test Assembly

11.1.1. Field Test Assembly

The updated Alternate ACCESS assessment was field tested in the 2022–23 school year. For each grade-level cluster, the Center for Applied Linguistics (CAL) assembled five field test forms, with common items across forms and across grade-level clusters, per a spiraling plan embedded into a test blueprint prepared by CAL and approved by WIDA.

To assemble the forms, CAL implemented a rigorous quality control process. First, it must be noted that CAL did not develop the test content; rather, Accessible Teaching, Learning, and Assessment Systems (ATLAS) developed all linguistic and graphical content for the assessment, which was reviewed, revised, and ultimately signed off by WIDA. In order to coordinate the test assembly process, CAL, WIDA, and ATLAS collaborated on a number of steps. First, CAL and WIDA agreed on the layout of the test materials, including:

- The Test Booklets, which contain the graphical and text item stimulus and response options for the Listening and Reading domains, and the graphical item stimulus for the Speaking domain;
- The Test Administrator (TA) Scripts, which contain the scripting that the test administrator reads aloud to the students, along with instructions for moving through the test, and;
- The Student Response Booklets (SRB), which contain the Individual Characteristics Questionnaire (ICQ), the spaces for the TA to record student responses to Listening, Reading, and Speaking domain items, and the graphical item stimulus and the response spaces and score recording spaces for the Writing domain.

Once the format of the materials was finalized, CAL produced Adobe InDesign templates for all of the test materials. These templates reflected the intended layout of the test, and contained paragraph styles for the way the various types of text appear in the documents (in terms of font selection, size, and style). These templates, along with instructions for use, were then provided to ATLAS, who populated the templates with the item content. Each item was populated into InDesign individually; in other words, for each item, there was a separate InDesign file for the test booklet, the TA script, and the SRB, as appropriate for the item domain. CAL also provided ATLAS with naming conventions for each item, which were included in the test blueprint.

After ATLAS delivered the populated InDesign files to CAL, CAL Production Team staff and consultants first reviewed each InDesign file for compliance with the necessary style and formatting conventions. Then, using detailed production blueprints that CAL developed from the test blueprint, CAL assembled all of the forms from the individual InDesign files into form-level InDesign files. These files were then exported to PDF for quality control reviews.

CAL's Test Assembly Manager coordinated a detailed review of each test form, which other CAL Test Development team members (including Test Development Managers, Language Testing Specialists, and Test Development Assistants) participated in. Reviewers used a detailed checklist to review that the test content was correct and all layout and formatting was accurate. The reviews were conducted in terms of mock administrations; two reviewers were paired up for each form, with one reviewer serving as the "mock TA" and one serving as the "mock student". As the reviewers moved through the mock administration, both used the review checklist to check the materials, with the mock TA responsible for checking the TA script and the SRB, and the mock student responsible for checking the Test Booklet. The reviewers added their review observations into a tracking spreadsheet, with detailed information regarding the test materials, the page number, and the specific edit needed. A separate tracking spreadsheet was used for each form.

Once all mock administrations were complete, CAL Production Team members implemented the necessary revisions to the InDesign files and re-exported the files to PDF. The CAL Test Assembly Manager then sent all forms to an external copyeditor for professional proofing. CAL provided the copyeditor with a style guide and notes on conventions specific to Alternate ACCESS, along with spreadsheets to track all copyediting edits. Upon completion of copyediting, CAL Production Team staff again implemented all revisions to the InDesign files. The CAL Test Assembly Manager then reviewed and signed off on all edits, with iterative reviews and revisions as needed.

Once the forms were complete, CAL then delivered them to WIDA, along with review tracking spreadsheets, for final review and signoff. WIDA reviewed all forms, requested a small number of edits, which CAL implemented and checked. At the end of November 2022, CAL delivered final print-ready PDFs to DRC for printing and distribution to the states.

In addition to assembling the test forms themselves, CAL also entered all item metadata and form metadata into their internal item database. The database was then used to produce the Alternate ACCESS field test item inventory file, a spreadsheet containing all test metadata that DRC uses to program their scanning and scoring systems, and that WIDA and CAL psychometrics teams use for data cleaning and analysis. CAL also updated DRC's Master Materials Specification List with the relevant print specifications for the Alternate ACCESS field test forms.

11.1.2. Operational Test Assembly

Operational Series 602 of Alternate ACCESS was administered in the 2023–24 school year. After the completion of Alternate ACCESS field testing, DRC scanned the test materials and delivered the data to WIDA and CAL for analysis. CAL then analyzed the data, delivered the analysis to WIDA, and CAL and WIDA collaborated to plan a Post Field Test Review and Item Selection meeting, which was held in Washington, DC on June 25–29, 2023. In the meeting, the statistics of each item were reviewed, and the final items for each grade-level cluster and domain were selected and ordered.

One of the final products of the meeting was a test map to be used for operational test assembly. CAL used this test map to enter the form metadata for operational Alternate ACCESS Series 602 into their internal item database and prepared the operational item inventory file. CAL also used the test map to prepare production blueprints for the operational test forms.

CAL then assembled the operational forms based on the operational test map and production blueprints. During assembly, CAL also made several edits to the test materials per WIDA request. This included revisions to administration instructions and Expect Boxes in the TA Scripts based on field test observations conducted by WIDA, changes to the ICQ, and other editorial revisions. CAL Production Team members made these revisions to item level InDesign files, which were then assembled into form-level InDesign files and exported to PDF for quality control reviews.

The CAL Test Assembly Manager then reviewed the forms, checking against the list of WIDA's requested revisions, and using the same checklist as was used in preparation of the field test forms, focusing on the accuracy of content and formatting. Once the forms were reviewed and revised accordingly, CAL then submitted the forms to an external copyeditor for proofing. CAL then revised the forms and submitted them to WIDA for final signoff. WIDA then requested final revisions, which CAL implemented and checked internally prior to final signoff by WIDA.

Final operational test materials, along with the operational item inventory file and updates to the Master Materials Specification List, were delivered to WIDA and DRC in August 2023.

11.2. Test Administration and Scoring Quality Control

With the updated Alternate ACCESS assessment, we have updated our quality control (QC) processes to match what we use for ACCESS. There are many steps involved in ensuring the correctness of the assessment results including forms review, scanner testing and user-acceptance testing. However, this write-up is concerned with the scoring QC.

The scoring QC takes place after the assessment is given, and before scores are reported to the states. DRC sends State Student Response (SSR) files to WIDA for approval. Here at WIDA, we process these files and after finding that no issues are in the file, we send DRC our approval to print reports and send the files onto the states.

Our scoring QC process is concerned with verifying that the scores contained in the SSR files are correct. This process begins with examining the raw response strings reported for each student for each domain in the SSR file. For Alternate ACCESS, these strings represent the bubbles that the test administrator bubbled in for their assessment of the quality of a student's response. These raw response strings are assumed to be correct for this process. This correctness should have been established while DRC tests their scanners and during the user-acceptance testing step where we verify the values in the SSR file match the intended, bubbled-in values.

Each item in the Alternate ACCESS assessment is assessed and scored by the test administrator. The student does not directly bubble in any responses, and there is no answer key. When scoring the Alternate ACCESS assessment, each of the raw responses is associated with a score. These scores are concatenated into a scored response string for each student for each domain. The first step of the scoring QC process is to ensure that scored response strings are correct given the raw response string. Essentially, WIDA verifies that a student receives the proper number of points for each item given the provided raw response.

Before continuing to calculate a student's score, we need to determine if a student should receive a score for a domain. Most students should receive a score, but there are two conditions that would cause a student to not receive a score. The first is that a domain didn't meet criteria to be considered an attempt. This would occur if a test administrator didn't fill in any bubbles for a domain. The other condition would be if a Do Not Score code was entered for a domain. In these cases, we ensure that a student does not receive a scale score, Proficiency Level, CSEM, or Confidence range for the domain or any dependent composites.

Once the scored response strings are verified, we can calculate a raw score for each student for each domain by summing the digits of the scored response string. This raw score isn't reported in the SSR files but is instead used to look up an entry in a scoring table. DRC uses a scoring table called the omnibus scoring table to use a student's grade level, domain, grade-level cluster, and their raw score to look up a student's scale score, CSEM, and Proficiency Level. We check that these values match the values in the scoring tables provided by CAL.

Once the scale score, Proficiency Level, and CSEM values are validated, we check the Confidence Range for each student. This is calculated by adding or subtracting the CSEM value from the student's scale score. However, this value is bounded and needs to remain in the scoring scale range of 900–980. Any values outside of those bounds are set to the closest value in the scoring scale range.

Once domains are validated, composite scores can be calculated. Each composite scale score is a weighted average of its respective domain scores. Once those scale scores are calculated, just like the domains, the Proficiency Level and CSEM values are found in the Omnibus Scoring Table for each composite by the scale score and student's grade level. The Confidence Range is determined the same way.

Once all scale scores, Proficiency Levels, CSEM, and Confidence Ranges have been verified for all students, for all domains and composites, the file will be approved. This approval is required for reports to be printed, and for the SSR files to be released to the states.

11.3. *Score Reporting Quality Control*

Score reporting quality control takes place in two separate phases. In the first phase, we collect and ingest State Student Response (or SSR) files. These files contain data on each test administrator's rating for each response from each student along with their domain and composite scale scores, proficiency levels, and confidence bands. In this step, we verify that all domain scores correctly reflect the test administrator ratings. We check that all composite scale scores are properly calculated, and that all proficiency level and CSEM values are associated with each score. When checking the scores in the SSR files, no errors were found.

The second phase is checking the reports to ensure they properly reflect the data in the SSR files. No issues were found with the roster, district summary, and school summary reports. However, several issues were found in the Family Individual Student Report (Family ISR) and Educator Individual Student Report (Educator ISR). These issues included missing identifier information in the header blocks when the display length was longer than anticipated. The table showing student performance didn't handle missing domains properly. There is also list of student abilities that corresponds with a student's overall proficiency level that was blank when a student was missing a domain and therefore didn't have an overall proficiency level. In the Educator ISR, there are also tables that display the information from the ICQ about a student's abilities both in English and a language other than English. In this table, both the English and language other than English were reported as "Yes" regardless of which was selected.

These issues were reported to DRC, who worked to create fixes for these issues. For the headers, they increased the space allocated to the identifying information to accommodate the longest values in the states in the QC set.

DRC updated the student performance table on each ISR to properly display a domain without a score by displaying the performance level as "N/A", and not filling in any of the blocks used to indicate performance.

The list of abilities properly shows the "N/A" overall proficiency level when missing but left the text and check marks even though no text was displayed. In future years, this behavior will be changed to hide this section when no overall proficiency level can be reported.

The ICQ section of the Educator ISR was updated to only show “Yes” for the proper language context. However, the text “No Response” was still displayed for each line in the table for each language context when neither option was selected. A blank was now displayed opposite a “Yes” when only one language context was selected. In future years, the text “No Response” will be removed from the table, instead leaving unselected options as blank. The “Yes” option will be replaced with a checkmark to indicate it was the option that was selected. Text to explain a blank IQC section will be added to the paragraph above the table to explain how this section may consist of all blank values.

Of the errors found in the ISR reports, the issues that caused missing or incorrect data to be reported were corrected. Also formatting issues with missing domains or composites in the student performance table were also corrected. Other formatting issues such as the blank student ability descriptors with a missing overall composite score, or the information display of the ICQ tables, were pushed to the next year in the interest of both providing reports on time, as well as allowing time to consider and agree on solutions instead of trying to rush a solution at the last minute.

References

- Allen, N. L., Carlson, J. E., & Zelenak, C. A. (1999). *The 1996 NAEP Technical Report*. ED Pubs, PO Box 1398, Jessup, MD 20794-1398.
- American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*.
https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561-573.
- Brennan, R. L. (2004). Manual for BB-Class: A computer program that uses the betabinomial model for classification consistency and accuracy. (CASMA Rep. No. 9). [Computer software manual] Retrieved from <https://education.uiowa.edu/casma/computer-programs>.
- Every Student Succeeds Act (ESSA) of 2015, Pub. L. No. 114-95, 129 Stat. 1802 (2015). Retrieved from <https://www.congress.gov/bill/114th-congress/senate-bill/1177>.
- Individuals with Disabilities Education Act (IDEA) of 2004, 20 U.S.C. § 1400 et seq. Retrieved from <https://sites.ed.gov/idea>.
- Kamata, A., Turhan, A., & Darandari, E. (2003, April). *Estimating reliability for multidimensional composite scale scores*. In annual meeting of American Educational Research Association, Chicago, IL.
- Kane, M., & Case, S. M. (2004). The reliability and validity of weighted composite scores. *Applied Measurement in Education*, 17(3), 221-240.
- Kolen, M. J. & Brennan, R. L. (2014). *Linking* (pp. 487-536). New York: Springer.
- LaFlair, G., Hanson, B. A., & Brennan, R. L. (2017). Equipercentile equating issues in small sample sizes. *Educational Measurement: Issues and Practice*, 36(4), 23-33.
- Lee, W., Hanson, B. A., & Brennan, R. L. (2002). Classification consistency and accuracy indices for composite scores. *Applied Psychological Measurement*, 26(4), 412-432.
- Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch measurement transactions*, 16(2), 878.
- Linacre, J. M. (2006). A user's guide to WINSTEPS: Rasch-model computer programs. *Winsteps.com*.
- Livingston, S. A., Carlson, J., & Bridgeman, B. (2018). *Test reliability-basic concepts*. Research Memorandum No. RM-18-01). Princeton, NJ: Educational Testing Service.
- Livingston, S. A., & Kim, S. (2008). Small-sample equating by the circle-arc method. *ETS Research Report Series*, 2008(2), i-16.
- Livingston, S. A., & Kim, S. (2009). The circle-arc method for equating in small samples. *Journal of Educational Measurement*, 46(3), 330-343.
- Livingston, S. A., & Kim, S. (2008). The circle-arc method for equating: Theory and applications. *Journal of Educational Measurement*, 45(3), 225-245.

- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32(2), 179-197.
- Rudner, L. (2001). Informed test component weighting. *Educational Measurement: Issues and Practice*, 20(1), 16-19.
- Tay, L., Meade, A. W., & Cao, M. (2015). An overview and practical guide to IRT measurement equivalence analysis. *Organizational Research Methods*, 18(1), 3-46.
- U.S. Census Bureau. (n.d.) About the Topic of Race.
<https://www.census.gov/topics/population/race/about.html>
- U.S. Department of Education. (2018). *A State's Guide to the U.S. Department of Education's Assessment Peer Review Process*.
<https://www.ed.gov/sites/ed/files/2020/07/assessmentpeerreview.pdf>
- WIDA. (n.d.). *ACCESS for ELLs District and School Test Coordinator Manual*. Board of Regents of the University of Wisconsin System.
- WIDA. (n.d.). *ACCESS for ELLs Test Administrator Manual*. Board of Regents of the University of Wisconsin System.
- WIDA. (n.d.) Alternate ACCESS: Administration and Scoring training course.
<https://portal.wida.us/course/detail/Alternate-ACCESS>
- WIDA. (n.d.). *Test Policy Handbook*. Retrieved from
<https://sea.wida.us/system/files/documents/SEA-support/test-policy-handbook.pdf>.
- WIDA. (May 2024). *WIDA Alternate ACCESS Field Test Technical Brief*. Board of Regents of the University of Wisconsin System. Retrieved from <https://sea.wida.us/documents/alternate-access-field-test-technical-brief>.
- WIDA. (December 2024). *WIDA Alternate ACCESS Assessment July 16–July 19, 2024, Standard Setting Meeting Report*. Board of Regents of the University of Wisconsin System. Retrieved from <https://sea.wida.us/documents/wida-alternate-access-standard-setting-meeting-report>.
- WIDA. (2021). *WIDA Alternate ACCESS Test Specifications*.
- WIDA. (2020). *WIDA English Language Development Standards Framework, 2020 Edition: Kindergarten–Grade 12*. Board of Regents of the University of Wisconsin System.
- Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch Measurement*. Chicago, IL: Mesa Press.
- Young, M. J., & Yoon, H. J. (1998). Estimating the accuracy of decision classifications using a latent class model. *Journal of Educational and Behavioral Statistics*, 23(4), 367-383.