



Annual Technical Report for  
ACCESS for ELLs  
Online English Language Proficiency Test  
Series 503, 2021–2022 Administration

Annual Technical Report No. 18A

Prepared by:

Center for Applied Linguistics

Language Assessment Division  
Psychometrics and Quantitative Research Team

Nov 2022

# **1. The WIDA ACCESS for ELLs Technical Advisory Committee**

This report has been reviewed by the WIDA ACCESS for ELLs Technical Advisory Committee (TAC), which comprises the following members:

- Gregory J. Cizek, Ph.D., Guy B. Phillips Distinguished Professor, Educational Measurement and Evaluation, University of North Carolina at Chapel Hill
- Claudia Flowers, Ph.D., Professor, Educational Research, Measurement, and Evaluation, University of North Carolina at Charlotte
- Akihito Kamata, Ph.D., Professor, Department of Education Policy and Leadership, Department of Psychology, Southern Methodist University
- Timothy Kurtz, Teacher (retired), Hanover High School, Hanover, New Hampshire
- Carol Myford, Ph.D., Professor Emerita, Educational Psychology, University of Illinois at Chicago
- Micheline Chalhoub-Deville, Ph.D., Professor, Educational Research Methodology, University of North Carolina at Greensboro

## Executive Summary

This is the 18th annual technical report on the ACCESS for ELLs English Language Proficiency Test and the seventh report on the ACCESS for ELLs assessment given in Online format.

This technical report is produced as a service to members and potential members of the WIDA Consortium and to support states' submissions for U.S. Department of Education English language proficiency assessment peer review. The technical information herein is intended for use by those who have technical knowledge of test construction and measurement procedures, as stated in *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). WIDA also produces an annual *Year in Review Report*, intended for a general audience, for readers who are interested in a nontechnical overview of the 2021–2022 ACCESS assessment.

ACCESS for ELLs is intended to assess reliably and validly the English language development of English language learners (ELLs) in Grades K–12 according to the WIDA 2012 Amplification of the English Language Development Standards Kindergarten–Grade 12 (WIDA Consortium, 2012). Results on ACCESS for ELLs are used by WIDA Consortium states for monitoring the progress of students, for making decisions about exiting students from language support services, and for accountability. WIDA additionally provides screening instruments for initial identification purposes; however, decision processes on how these are incorporated into identification decisions are at individual states' discretion.

ACCESS for ELLs assesses students in the four domains of Listening, Reading, Writing, and Speaking, as required by federal law (Elementary and Secondary Education Act of 1965, amended 2015; §1111(b)(1)(F); §1111(b)(2)(G)) and provides composite scores as required by the same statute (§3121).

ACCESS for ELLs Online Series 503 was administered in school year 2021–2022 in 35 states, the Bureau of Indian Education, the Department of Defense Education Activity, the Northern Mariana Islands, the Virgin Islands, and the District of Columbia for a total of 40 state entities (henceforth “states”).

The Series 503 Online data set used in this report included the results of 1,841,814 students as of September 2022. The final number of students who participated in the Series 503 Online ACCESS tests is 1,882,183. The largest grade in this report was Grade 2 with 212,617 students, while the smallest was Grade 12 with 70,706 students. Of the participating WIDA states, the largest was IL with 205,824 students, while the smallest was the U.S. Virgin Islands with 990 students.

During the 2021–2022 testing year, states returned to in-person schooling. Additionally, WA joined the WIDA consortium. Based on a comparison with prior years' numbers of participating students, there are 67% more student population participated in ACCESS Series 503 testing than the ACCESS Series 502 testing.

ACCESS for ELLs Series 503 was offered in two administrative formats, an online format (Grades 1–12) and a paper format (Kindergarten–Grade 12). The current report (WIDA ACCESS

Technical Report 18A) provides technical information pertaining to ACCESS for ELLs Series 503 Online. A second report (WIDA ACCESS Technical Report 18B) provides technical information for the ACCESS for ELLs Series 503 Paper assessment, including the Kindergarten assessment.



## **Part 1**

### **Purpose, Design, Implementation**

# Contents

1 Purpose and Design of ACCESS .....	1-1
1.1 Purpose Statement.....	1-1
1.2 The WIDA Standards.....	1-2
1.3 The WIDA Proficiency Levels .....	1-3
1.4 Language Domains .....	1-5
1.5 Grade-Level Clusters .....	1-6
1.6 Tiers .....	1-6
2 Test Development .....	2-8
2.1 Item and Task Design .....	2-8
2.1.1 Listening Items .....	2-8
2.1.2 Reading Items .....	2-12
2.1.3 Writing Tasks .....	2-15
2.1.4 Speaking Tasks.....	2-21
2.2 Test Design .....	2-24
2.2.1 Listening .....	2-24
2.2.2 Reading.....	2-27
2.2.3 Writing.....	2-29
2.2.4 Speaking .....	2-33
2.3 Test Construction .....	2-36
2.3.1 Item and Task Development.....	2-36
2.3.2 Field Testing.....	2-43
2.3.3 Item/Task Review and Selection .....	2-53
3 Test Administration stopped here .....	3-60
3.1 Test Delivery.....	3-60
3.1.1 Listening and Reading.....	3-60

3.1.2 Writing.....	3-60
3.1.3 Speaking .....	3-60
3.2 Operational Administration .....	3-61
3.2.1 Administering the Test Practice .....	3-61
3.2.2 Listening Test Administration.....	3-62
3.2.3 Reading Test Administration.....	3-62
3.2.4 Writing Test Administration.....	3-62
3.2.5 Speaking Test Administration .....	3-63
3.2.6 Test Security .....	3-64
3.3 Fairness and Accessibility .....	3-65
3.3.1 Support Provided to All ELLs .....	3-66
3.3.2. Support Provided to ELLs with IEPs or 504 Plans .....	3-66
4 Scoring .....	4-69
4.1 Multiple Choice Scoring: Listening and Reading.....	4-69
4.2 Scoring Performance-Based Tasks: Writing and Speaking .....	4-69
4.3 Writing Scoring Scale .....	4-78
4.4 Speaking Scoring Scale.....	4-81
5 Summary of Score Reports .....	5-85
5.1 Individual Student Report.....	5-85
5.2 Other Reports .....	5-89

# 1. Purpose and Design of ACCESS

## 1.1 Purpose Statement

The purpose of ACCESS for ELLs is to assess the developing English language proficiency of English language learners (ELLs) in Grades K–12 in the 41 U.S. states, territories, and federal agencies in the WIDA Consortium, first in the English Language Proficiency Standards (Gottlieb, 2004; WIDA Consortium, 2007) and then in the amplified 2012 English Language Development (ELD) Standards (WIDA Consortium, 2012). The WIDA ELD Standards, which correspond to the academic language used in state academic content standards, describe six levels of developing English language proficiency and form the core of the WIDA Consortium’s approach to instructing and testing ELLs. ACCESS may thus be described as a standards-based English language proficiency test designed to measure the social and academic language proficiency of ELLs in English. It assesses social and instructional English as well as the academic language associated with language arts, mathematics, science, and social studies, within the school context, across the four language domains (Listening, Reading, Writing, and Speaking).

Other purposes of ACCESS include

- Identifying the English language proficiency level of students with respect to the WIDA ELD Standards used in all member states of the WIDA Consortium
- Identifying students who have attained English language proficiency
- Assessing annual English language proficiency gains using a standards-based assessment instrument
- Providing districts with information that will help them to evaluate the effectiveness of their language instructional educational programs and determine staffing requirements
- Providing data for meeting federal and state statutory requirements with respect to student assessment
- Providing information that enhances instruction and learning in programs for ELLs.

ACCESS for ELLs is offered in two formats: ACCESS for ELLs Online (also referred to as *ACCESS Online*), described in this report, and ACCESS for ELLs Paper, described in a companion report.

## **1.2 The WIDA Standards**

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

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

For practical purposes, the five Standards are abbreviated as follows in this report:

- Social and Instructional Language: SIL
- Language of Language Arts: LoLA
- Language of Math: LoMa
- Language of Science: LoSc
- Language of Social Studies: LoSS

Every selected response item and every performance-based task on ACCESS for ELLs targets at least one of these five Standards. In Speaking and Writing tasks, the Standards are combined as follows:

- Integrated Social and Instructional Language (SIL), Language of Language Arts (LoLA), and Language of Social Studies (LoSS): IT (Writing only)

- Language of Math (LoMa) and Language of Science (LoSc): MS (Speaking and Writing)
- Language of Language Arts (LoLA) and Language of Social Studies (LoSS): LS (Speaking and Writing)

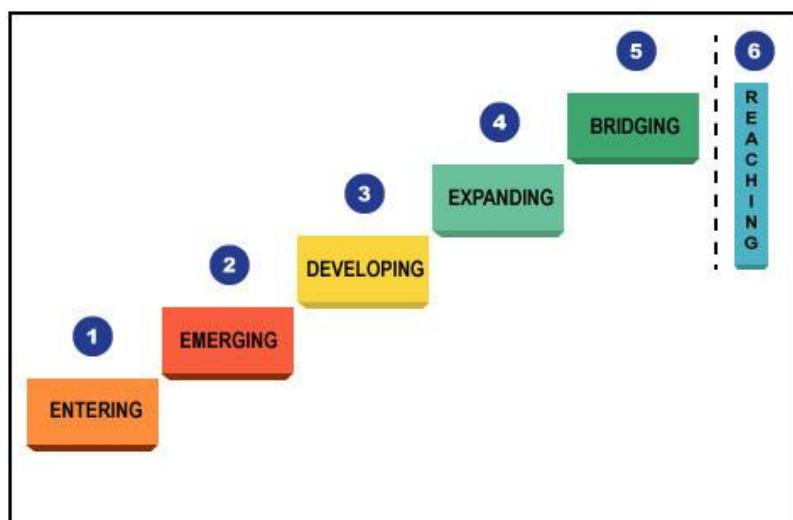
The overarching goal of ACCESS for ELLs Online is to measure the academic English language proficiency of students. Proficiency is measured according to a scale, as defined by the WIDA ELD Standards Framework as comprising five levels of proficiency, which are in turn defined in the performance definitions (WIDA Consortium, 2012).

The five WIDA ELD Standards should not be thought of in the same sense as content standards (Allen, Carlson, & Zelenak, 1999); rather, they provide the context for assessing a student's language proficiency in a given domain, so the skills that contribute to academic English language proficiency in a domain are the same across the five ELD Standards. In other words, the construct being measured across the five ELD Standards is the same within a domain.

Because of this conceptualization of the WIDA ELD Standards, scores are not reported for each of the Standards, and it is not necessary to assess all five Standards in one domain if each of the Standards is measured on the assessment in some capacity (although ACCESS for ELLs Online does strive to represent all five WIDA Standards in each domain test).

### **1.3 The WIDA Proficiency Levels**

The WIDA ELD Standards describe the continuum of language development via five language proficiency levels (PLs) that are fully delineated in the WIDA ELD Standards document (WIDA Consortium, 2012), with scores indicating progression through each level. These levels are *Entering*, *Emerging*, *Developing*, *Expanding*, and *Bridging*. There is also a final stage known as *Reaching*, which is used to describe students who have progressed across the entire WIDA English language proficiency continuum; as this is the end of the continuum, scores do not indicate progression through this level. The proficiency levels are shown graphically in Figure 1.



*Figure 1. The language proficiency levels of the WIDA ELD Standards.*

These language proficiency levels are embedded in the WIDA ELD Standards in two ways.

First, they appear in the **performance definitions**. The performance definitions describe the stages of language acquisition, providing details about the language that students can comprehend and produce at each proficiency level. The performance definitions are based on three criteria: (1) vocabulary usage at the word/phrase level; (2) language forms and conventions at the sentence level; and (3) linguistic complexity at the discourse level. Vocabulary usage refers to students' increasing comprehension and production of the technical language required for success in the academic content areas. Language forms and conventions refers to the increasing development of phonological, syntactic, and semantic understanding in receptive skills or control of usage in productive language skills. Linguistic complexity refers to students' demonstration of oral interaction or writing of increasing quantity and variety.

Second, language proficiency levels are represented through connections to the accompanying **Model Performance Indicators** (MPIs). The MPIs provide a model of the expectations for ELL students in each of the five Standards, by grade-level cluster, across the four language domains, for each of the language proficiency levels up to level 5. The grouping of MPIs at proficiency levels 1 through 5 for a given WIDA Standard, grade-level cluster, domain, and topic is called a strand. These MPIs together describe a logical progression and accumulation of skills on the path from the lowest level of English language proficiency to full English language proficiency for

academic success. The final level, PL 6: *Reaching*, represents the end of the continuum rather than another level of language proficiency.

Each MPI has a tripartite structure, consisting of a language function, a content stem, and support. The MPIs used on ACCESS can be taken directly from the WIDA English Language Proficiency Standards (WIDA Consortium, 2007) or the amplified 2012 ELD Standards (WIDA Consortium, 2012). In addition, given that the MPIs in the WIDA Standards are truly “models” and do not cover all possible topics within each Standard for each grade-level cluster and language domain, MPIs can be “transformed” to accommodate the needs of classroom instruction, as described in the amplified 2012 ELD Standards (WIDA Consortium, 2012, p. 11). MPIs are also transformed for the purposes of the assessment. When MPIs are transformed, one or more of the three aspects of the base MPI are changed. For example, if an MPI from the amplified 2012 ELD Standards (WIDA Consortium, 2012) has “categorize” as its language function, which could be transformed to “compare/contrast” or “infer.” Likewise, if the content stem for a Grades 9–10 Language of Social Studies strand of MPIs is “supply and demand,” it could be transformed to “freedom and democracy.” Each item specification document for a given WIDA Standard, grade-level cluster, and language domain contains an MPI for each item or task, such that the MPI is the core construct that the given item/task intends to measure. Each selected-response item or performance-based task on ACCESS for ELLs is carefully developed, reviewed, piloted, and field tested to ensure that it allows students to demonstrate accomplishment of the targeted MPI.

In reporting proficiency, WIDA reports scores for each of the domains, in addition to composite scores and an overall score (WIDA Consortium, 2021d). So, for each of the domain scores, WIDA reports measures of academic English language proficiency in that domain. More specifically, the score for Speaking is a measure of academic English language proficiency in the domain of Speaking, and likewise for Writing.

## **1.4 Language Domains**

The WIDA ELD Standards describe developing English language proficiency for each of the four language domains: Listening, Reading, Writing, and Speaking. Thus, ACCESS for ELLs contains four sections, each assessing an individual language domain.



## 1.5 Grade-Level Clusters

The grade-level cluster structure for ACCESS for ELLs Online is as follows: 1, 2–3, 4–5, 6–8, and 9–12. Note that the Kindergarten (K) form is not administered online and thus is not covered in this report.

## 1.6 Tiers

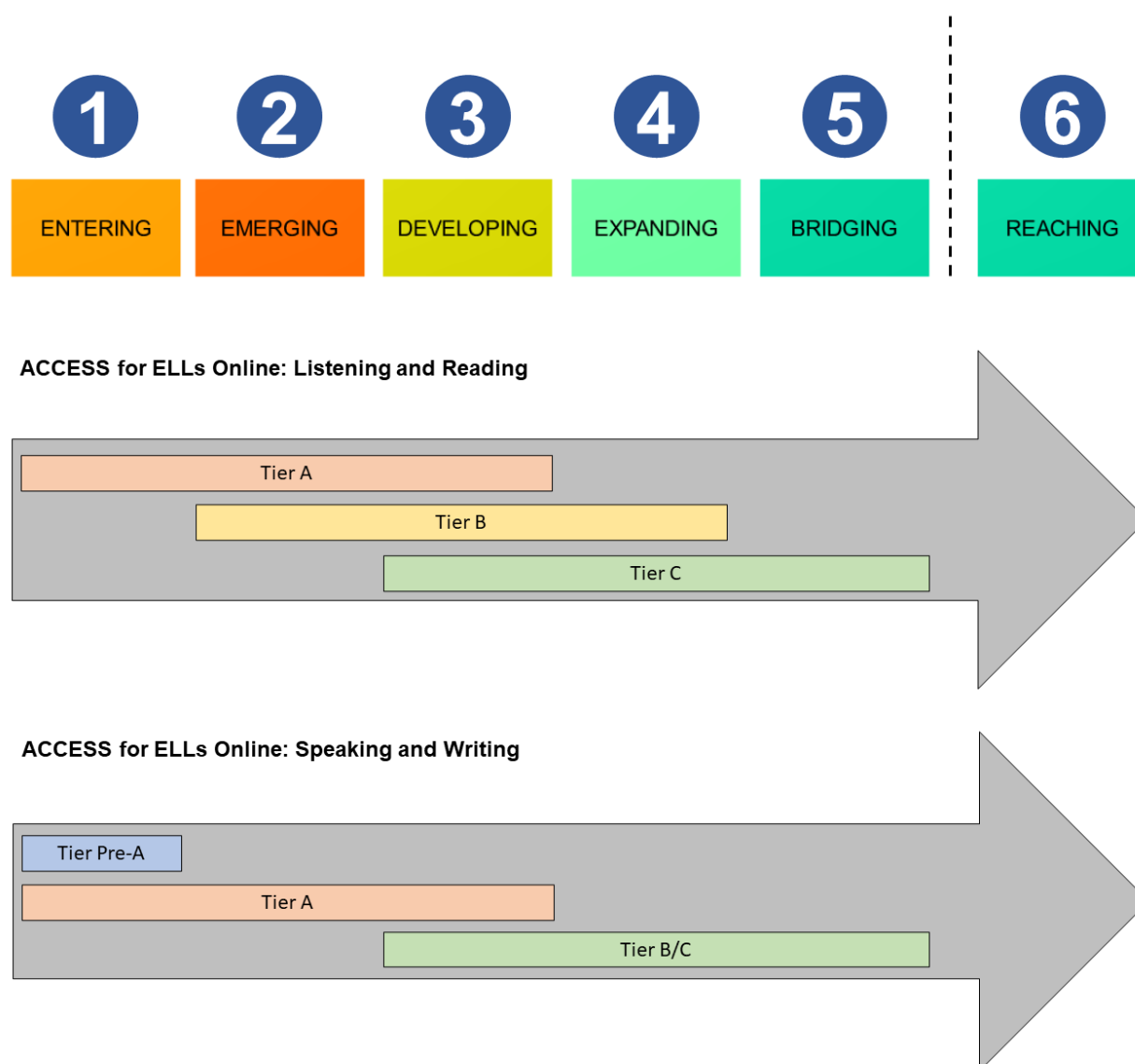
ACCESS is designed so that test paths or forms are appropriate to the proficiency level of individual students across the wide range of proficiencies described in the WIDA ELD Standards (*Figure 2*). Tests must be at the appropriate difficulty level for each individual student to facilitate valid and reliable interpretations of scores. While the grade-level cluster structure is a design feature intended to ensure that the language expectations are developmentally appropriate for students in different age ranges, within each grade-level cluster, students display a range of abilities. Test items and tasks designed for Entering (PL 1) or Emerging (PL 2) students to demonstrate accomplishment of the MPIs at their proficiency level will not allow Expanding (PL 4) or Bridging (PL 5) students to demonstrate the full extent of their language proficiency. Likewise, items and tasks that allow Expanding (PL 4) and Bridging (PL 5) students to demonstrate accomplishment of the MPIs at their level would be far too challenging for Entering (PL 1) or Emerging (PL 2) students. Items that are far too easy for students may be boring and lead to inattentiveness; items that are far too difficult for students may be frustrating and discourage them from performing their best. But more importantly, items that are too easy or too hard for a student add very little to the accuracy or quality of the measurement of that student's language proficiency.

In the Listening and Reading multistage adaptive tests, students are routed to folders that vary in difficulty, designated as A, B, or C level folders.<sup>1</sup> Tier A folders are intended for students at beginning levels of English language proficiency (PLs 1–3), Tier B folders for students at intermediate levels (PLs 2–4), and Tier C folders for students at more advanced proficiency levels (PLs 3–5). In the domain of Writing, the test forms are designated as either Tier A, which

---

<sup>1</sup> In Listening and Reading, a *Thematic folder*, or folder for short, is a collection of three items constructed around a common theme. For Writing, a thematic folder consists of one or two tasks written to a common theme. For Speaking, a thematic folder consists of two tasks written to a common theme.

includes tasks written to elicit language up to PL 3, or Tier B/C, which includes tasks written to elicit language up to PL 4 or PL 5. In the domain of Speaking, test forms are designed so that students at very beginning levels of proficiency take a pre-A form, which is designed to elicit language at PL 1; students at early levels of proficiency take the Tier A form, with tasks designed to elicit language at PL 1 and PL 3; and more proficient students take the Tier B/C form, with tasks designed to elicit language at PL 3 and PL 5.



**Figure 2. Tiers and proficiency levels.**

## **2. Test Development**

### **2.1 Item and Task Design**

This section describes how the Center for Applied Linguistics (CAL) Test Development (TD) team designs items and tasks to collect the necessary evidence required for the purposes of the assessment. Items and tasks are discussed by language domain. Readers who are interested in seeing illustrative examples of items and tasks can find these on the ACCESS Test Practice and Sample Items page on WIDA's website.

When the task models for ACCESS for ELLs Online were first developed, CAL and WIDA addressed issues of fairness by ensuring that principles of Universal Design of Assessments (UDA) were adhered to in this design phase (National Center on Educational Outcomes, 2021). Therefore, CAL, WIDA, and Data Recognition Corporation (DRC) collaborated to design the item and task layout on the screen to be maximally readable/legible with sufficient whitespace, to be accessed intuitively by students, to be accompanied by instructions and practice items to allow students to become accustomed to the test interface, and to contain universal accessibility tools (magnifier, line guide) as well as tools for accommodation (such as control of test audio and extended response time for the Speaking test). The ways in which the CAL TD team ensures fairness by adhering to principles of UDA in item development, in addition to the process by which bias and sensitivity review panels evaluate items and tasks to ensure accessibility and fairness for all students, are described in Section 2.3.1 below.

#### **2.1.1 Listening Items**

All Listening items include a prerecorded stimulus passage and question stem. Listening items are selected-response items, with one key and two distractors as answer choices. Answer choices are primarily graphics (illustrations, photographs, charts/diagrams); for Grades 2–12, items that test Listening proficiency at PLs 3–5 may consist of short written text response options that are written to be about two PLs lower than the targeted PL of the Listening item.

Most items on the operational Listening test are traditional multiple choice, though some operational items and some items embedded for field testing purposes may involve enhanced item presentations, including hot spot items (i.e., the student clicks on an area of the screen to

respond) and drag-and-drop items (i.e., the student drags an image/text to a specified screen area to respond).

For traditional multiple-choice items, students choose an answer from a set of ordered response options. The response options may be images or text. Students select their answer by clicking anywhere within the box that denotes the response options, including inside the circle that appears to the left of the text or image. Students can change their answer by clicking on a different response option. A screenshot of a sample Listening multiple choice item is provided in Figure 3.

The screenshot displays the 'Listening Practice' section of the ACCESS for ELLs Online Listening test. On the left, there is a header 'Listening Practice' in a blue bar. Below it is a small portrait of a female student. The main area shows a large illustration of a teacher and several children walking along a dirt path in a park-like setting with trees, rocks, and a squirrel. To the right of the illustration is a vertical list of three response options, each with a radio button to its left. The options are: 1. A squirrel (labeled with a '1' in a box above it). 2. An orange flower. 3. A large dark rock. At the bottom of the interface is a blue navigation bar containing icons for settings, a 'Pause Test' button, a help icon, a 'Line Guide' button, and a 'Next' button.

Figure 3: Multiple choice item layout for the ACCESS for ELLs Online Listening test

For hot spot items, students see a large response area. The response area may be an image, a paragraph of text, or some combination of images and text, such as a timeline or a webpage. The answer choices may be pictures or text and are embedded in the response area inside blue boxes. Students answer the question by clicking on one of the boxes in the response area. Each answer choice changes color when selected. Students can change their answers by clicking on a different blue box or by clicking on the reset eraser button, which clears the original response, and clicking on a different blue box. A screenshot of a sample hot spot item is provided in Figure 4.



**Figure 4: Layout of a Hot Spot item for the ACCESS for ELLs Online Listening test**

Drag and drop items have two possible formats. In one format, students see one object, either a small image or a line of text, above the response area, which may be an image, a paragraph of text, or some combination of images and text, such as a timeline, a webpage, etc. The response area has three or four blue boxes in it. To show their answer, students click and drag/move the

small object into a blue box within the response area. Students do not have to place the object exactly in the blue box; the object snaps into place when students release the mouse button. In this type of drag-and-drop item, students can change their answer by dragging their object into a different blue box in the response area or by clicking on the reset eraser button, which clears the original response, and then dragging the object into a different blue box in the response area. Alternatively, students may see three small objects above the response area. In this case, students select one object to drag into the single blue box within the response area. A screenshot of a sample drag and drop item is provided in Figure 5.



**Figure 5: Layout of a drag and drop item for the ACCESS for ELLs Online Listening test**

The number of enhanced items on the Listening test is not specified in the test or item specifications, so the appearance of enhanced items on the test is emergent from the content. In

other words, if the content of a given item lends itself well to an enhanced item type, then it is operationalized as such.

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.<sup>2</sup> Folders group together three test items that are written around a common theme, with each item targeting a progressively higher proficiency level.

- Tier A folders are constructed to target PLs 1 through 3.
- Tier B folders are constructed to target PLs 2 through 4.
- Tier C folders are constructed to target PLs 3 through 5.

In the ACCESS Online Listening test, students take a multistage adaptive test form, which routes students to Tier A, B, or C folders as appropriate to their ability level.

Each Listening item appears on its own screen with associated graphic support. Scripts containing the item orientation, stimulus, and question stem are audio recorded with professional voice actors, and a professional recording studio produces the items. Audio playback of test item content is automatic when students advance to the next screen. Listening test content is played one time for students unless the student has a predetermined accommodation allowing for a single repetition of the item stimulus and question stem. Further detail on accommodations can be found in Section 3.3.2.

### **2.1.2 Reading Items**

Reading items are similar in format to Listening items. The stimulus and question stems for Reading items are written text, and answer choices are also primarily written text, though response options for items targeting PLs 1 and 2 may be graphics (illustrations, photographs, charts/diagrams) or text. As with Listening items, Reading items are grouped into thematic folders of three test items each.

- Tier A folders target PLs 1 through 3.

---

<sup>2</sup> Level 6 is defined as “language that meets all criteria through Level 5, Bridging” and does not have descriptors at the word, sentence, and discourse levels like the other levels.

- Tier B folders target PLs 2 through 4.
- Tier C folders target PLs 3 through 5.

In the ACCESS Online Reading tests, students take a multistage adaptive test form, which routes them to Tier A, B, or C folders as appropriate to their ability level.

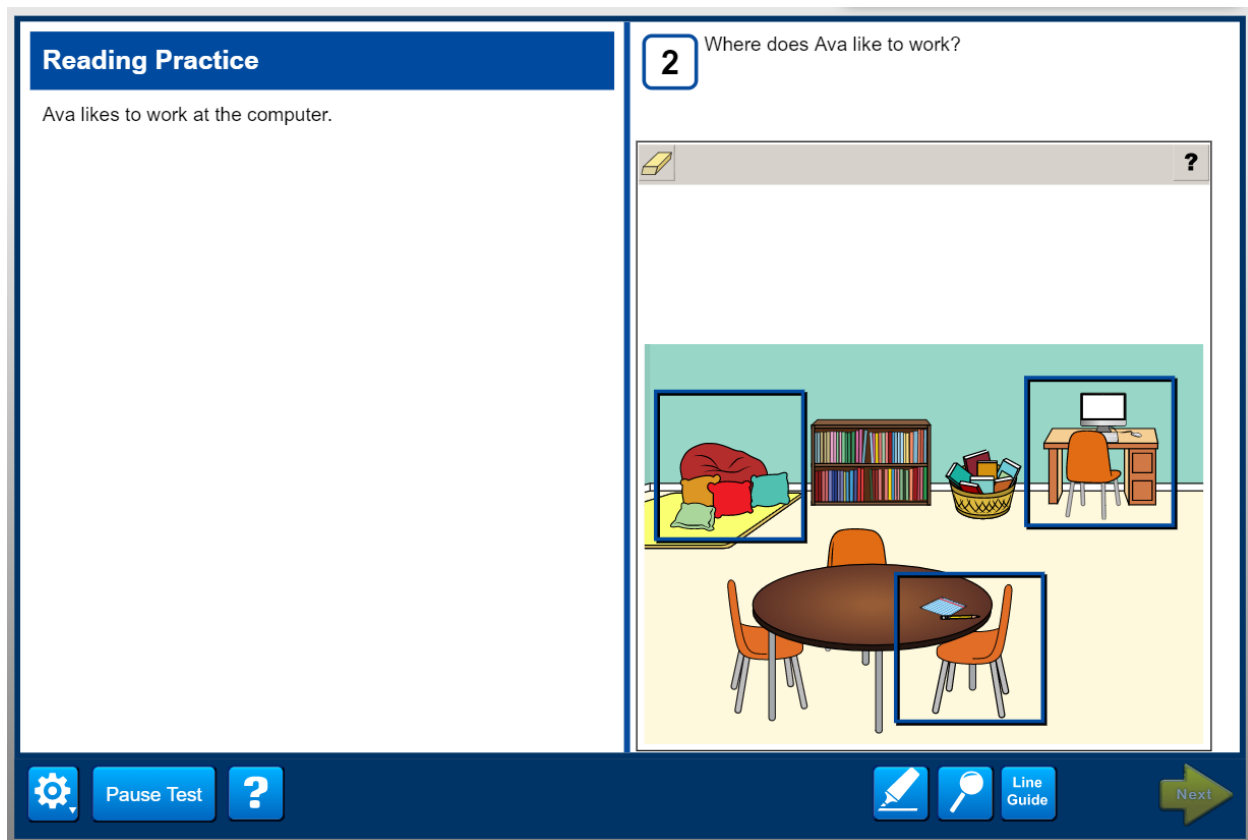
Most items on the operational Reading test are traditional multiple choice. A screenshot of a sample Reading multiple choice item is provided in Figure 6.

The screenshot displays the ACCESS Online Reading test interface. On the left, a blue header bar reads "Reading Practice". Below it, the text states: "Robert, Ava, and Mr. Green are reading about fish." Underneath this text is an illustration of three people (a man, a woman, and a boy) sitting around a round table in a room with bookshelves, a desk, and a basket of books. On the right side of the interface, a question box is labeled "1" and contains the text "What are they reading about?". Below the question box are three radio button options: "Trees", "Fish", and "Birds". At the bottom of the interface, there is a navigation bar with icons for settings, a "Pause Test" button, a help icon, a pencil icon, a magnifying glass icon, a "Line Guide" button, and a "Next" button.

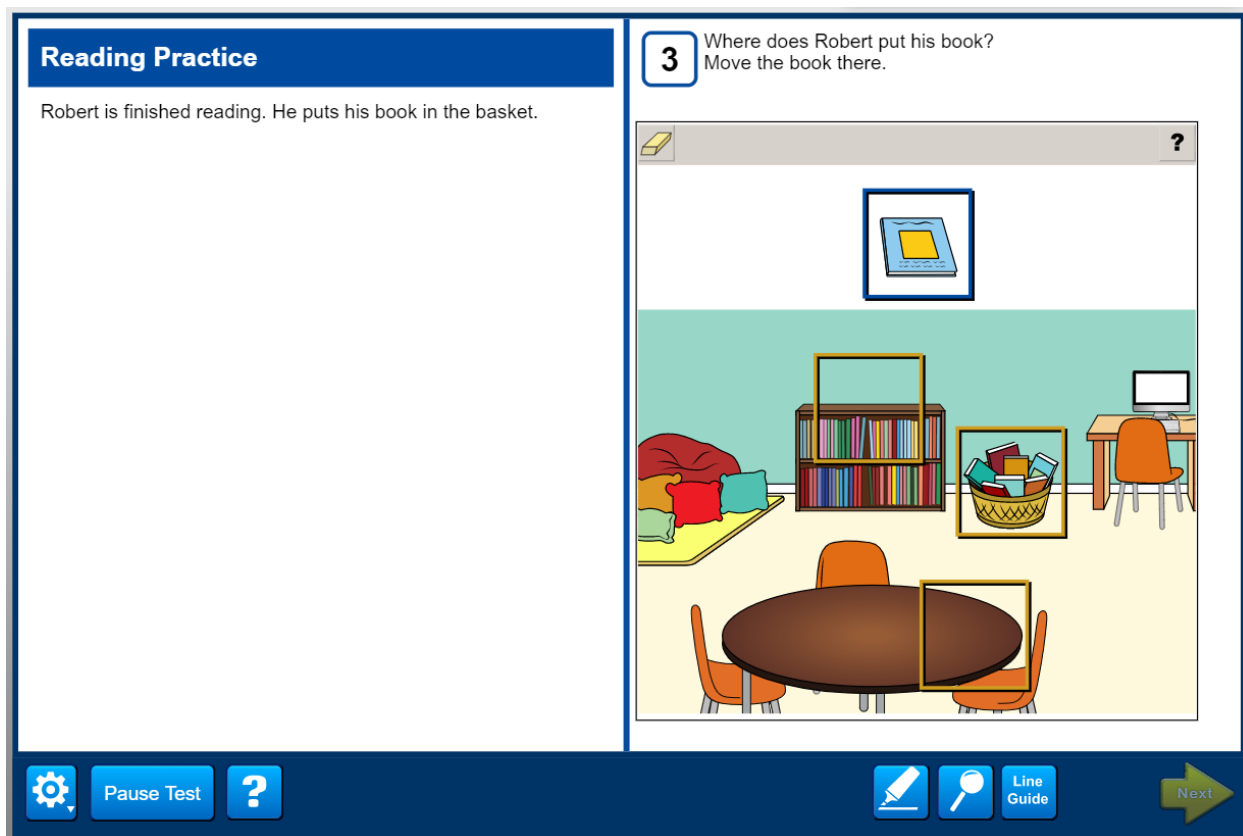
**Figure 6: Multiple choice item layout for the ACCESS for ELLs Online Reading test**

As with the Listening test, some operational items and some items embedded for field testing purposes involve enhanced item presentations, including hot spot and drag and drop items. The layouts of the Reading hot spot and drag and drop items are presented in Figure 7 and Figure 8 respectively.





**Figure 7: Layout of a hot spot item for the ACCESS for ELLs Online Reading test**



**Figure 8: Layout of a drag and drop item for the ACCESS for ELLs Online Reading test**

The number of enhanced items on the Reading test is not specified in the test or item specifications, so the appearance of enhanced items on the test is emergent from the content. In other words, if the content of a given item lends itself well to an enhanced item type, then it is operationalized as such.




Items have one key and either two or three distractors, depending upon the grade-level cluster and the targeted proficiency level. For Grades 1 and 2–3, all items have a key and two distractors. For Grades 4–5, 6–8, and 9–12, items targeting PLs 1 and 2 have a key and two distractors, and items targeting PLs 3, 4, and 5 have a key and three distractors.

### 2.1.3 Writing Tasks

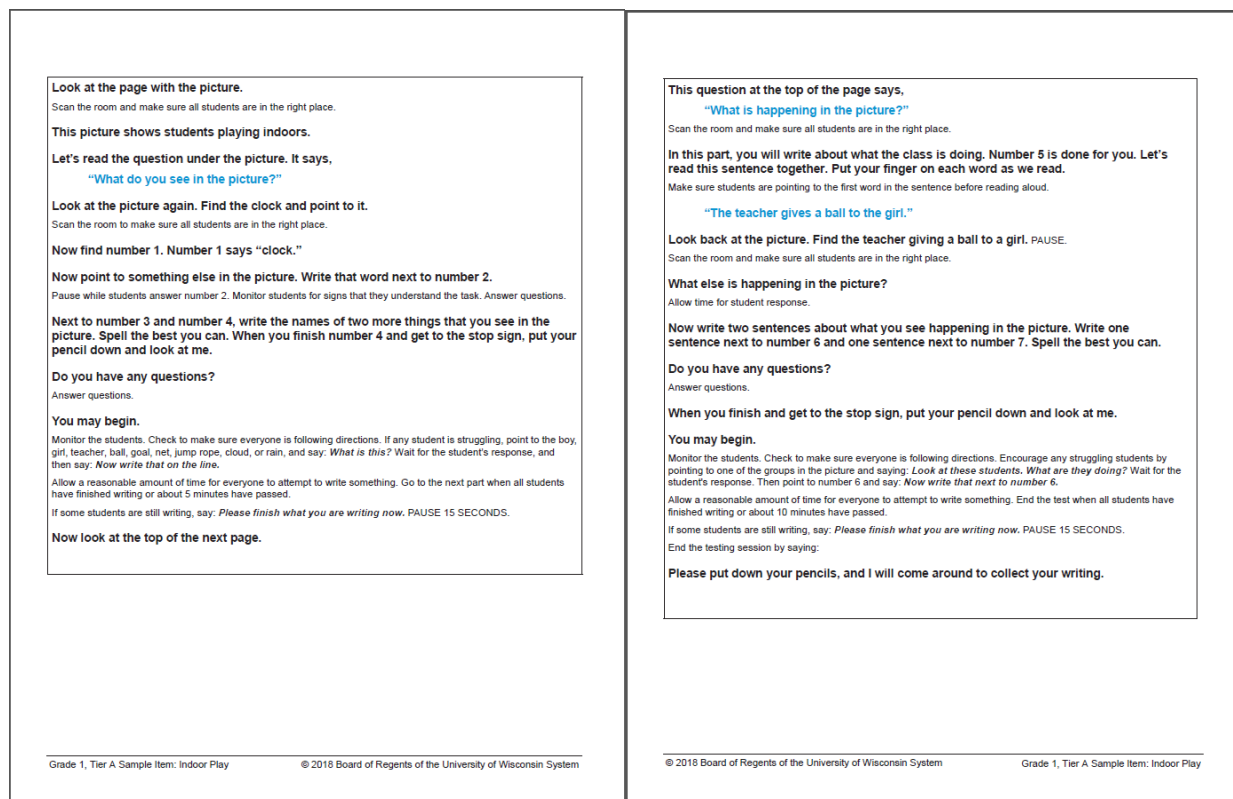
Writing tasks are designed to elicit language corresponding to one or more of the WIDA ELD Standards. Tasks appearing on the Tier A test form are designed to give students the opportunity to produce writing samples that fulfill linguistic expectations up to PL 3. DRC raters score

students' written responses to these tasks using the entire breadth of the scoring scale. (For more information about scoring the Writing test, see Section 2.2.3 below.) Therefore, students may achieve proficiency levels higher than PL 3, although the tasks are not designed to elicit extended responses, so the scores are limited by task design. Tasks appearing on the Tier B/C form are designed to give students the opportunity to produce writing samples that fulfill linguistic expectations up to PL 5. Again, although these tasks are designed to elicit extended responses, DRC raters score the responses using all nine categories of the scoring scale, so students' actual performance may extend above or below the PL 5 range.

For students in Grades 1–3, the test is not administered via computer. For students in these grades, the Test Administrator reads from a script and the students respond in a printed test booklet. CAL and WIDA made this design decision when ACCESS Online was first developed, based on the challenge that students at this age have with keyboarding their responses, as CAL and WIDA observed in cognitive labs. Figure 9 provides an example of the paper test booklet, and Figure 10 provides an example of the accompanying script.

<p style="text-align: right;">Name: _____</p> <div style="text-align: center;">  </div> <p>What do you see in the picture?</p> <div style="display: flex; flex-wrap: wrap;"> <div style="width: 50%;"> <p>1 <u>clock</u></p> <p>2 _____</p> </div> <div style="width: 50%;"> <p>3 _____</p> <p>4 _____</p> </div> </div> <div style="text-align: right;">  </div> <p style="font-size: small;">Grade 1, Tier A Sample Item: Indoor Play      © 2018 Board of Regents of the University of Wisconsin System</p>	<p style="text-align: right;">Name: _____</p> <p>What is happening in the picture?</p> <p>5 <u>The teacher gives a ball</u> <u>to the girl.</u></p> <p>6 _____</p> <p>7 _____</p> <div style="text-align: right;">  </div> <p style="font-size: small;">© 2018 Board of Regents of the University of Wisconsin System      Grade 1, Tier A Sample Item: Indoor Play</p>
--	--

**Figure 9: Example test booklet for the ACCESS for ELLs Online Writing test, Grades 1–3**



**Figure 10: Example script for the ACCESS for ELLs Online Writing test, Grades 1–3**

For students in Grades 4–12, writing prompts appear on the computer screen. In the spirit of providing maximal support and making every provision to ensure that students are given the opportunity to demonstrate the full extent of their English language proficiency, modeling is sometimes used to make task expectations as clear as possible to students. For example, the first of a series of questions may already be partially completed, or a sentence starter may be provided. In addition to the task screens, all tasks on the ACCESS for ELLs Online Writing test contain one or more orientation screens, which introduce the students to the context of the task and provide stimuli to serve as input to the tasks. Figure 11, Figure 12, and Figure 13 show the screen layouts for the tasks on the computer-delivered Writing test for Tier A and Tier B/C respectively.

## Carlos's Rainy Day

Here is a story about Carlos. He is getting ready for school.



1



2



3



4


Pause Test





Line Guide



**Figure 11:** Example orientation screen for the ACCESS for ELLs Online Writing test, Grades 4–12, both tiers

**Figure 12: Example layout for the ACCESS for ELLs Online Writing test, Grades 4–12, Tier A**

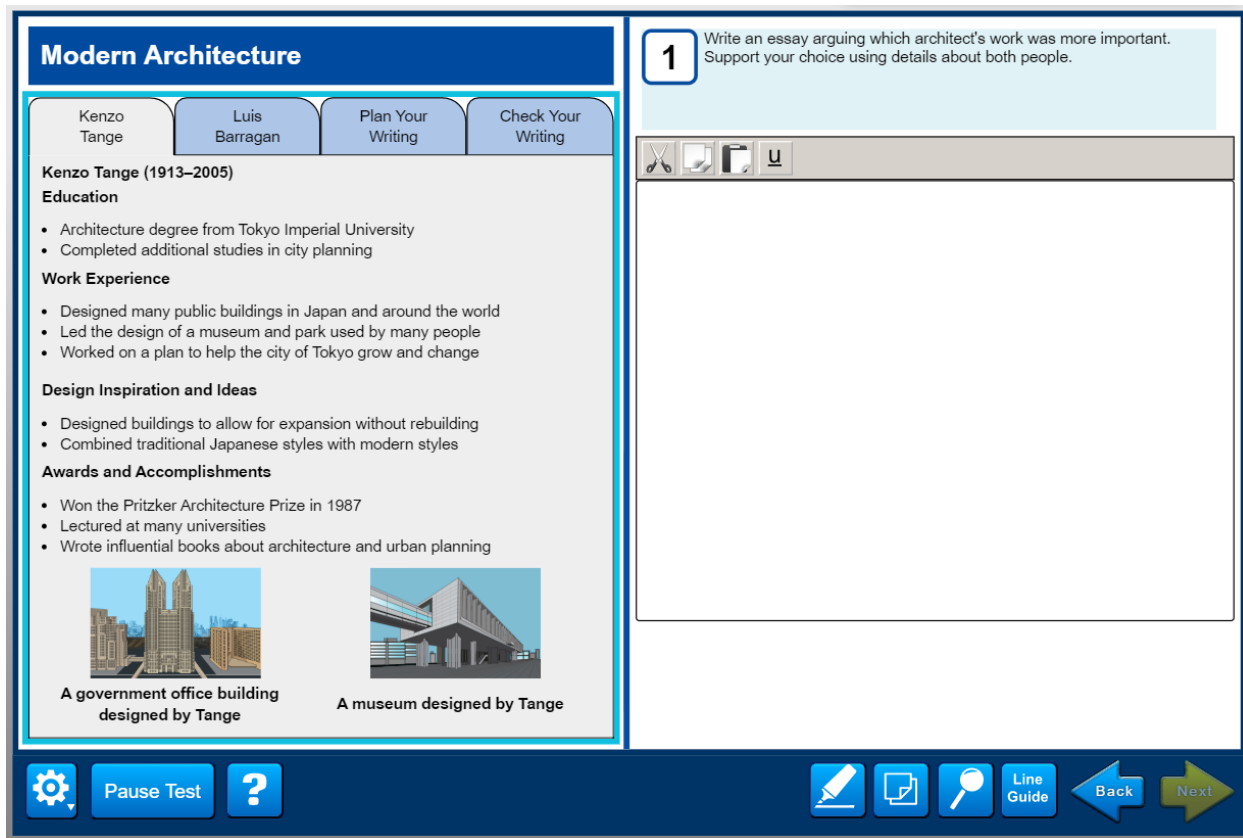


Figure 13: Example layout for the ACCESS for ELLs Online Writing test, Grades 4–12, Tier B/C

Students in Grades 4–5 provide either handwritten or keyboarded responses, with the default response mode determined in advance at the state or district level. For students in Grades 6–12, keyboarding is the default response mode, with a handwriting option offered as an accommodation.

For students who respond by handwriting in a writing response booklet, the test tasks have a slightly different appearance on the screen when compared to the tasks experienced by students who keyboard their responses. As shown in Figure 14, instead of a writing response space on the right side of the screen, an image of the test booklet appears on the screen to indicate to students where in their writing response booklet they should write.

Writing Practice

This picture shows a classroom.

A girl is...

1 What is happening in the picture? Write 1 sentence.

Writing Practice

1

2

Pause Test

?

Line Guide

Back

Next

**Figure 14: Example layout for an ACCESS for ELLs Online Tier A writing task with the handwritten (HW) response mode**

## 2.1.4 Speaking Tasks

Stimuli on the Speaking test include graphics, audio, and text. All stimuli are presented by a virtual Test Administrator (VTA). The VTA serves as a narrator who guides students through the test and acts as a virtual interlocutor. The VTA is introduced to students during the test directions to establish the testing context.

Task modeling is an essential component of the Speaking test design. In addition to the VTA, students are introduced to a virtual model student during the test directions. Prior to responding to each task, students first listen as the model student responds to a parallel task. The purpose of the model is to demonstrate task expectations to both students and to DRC raters, who score all Speaking task responses.

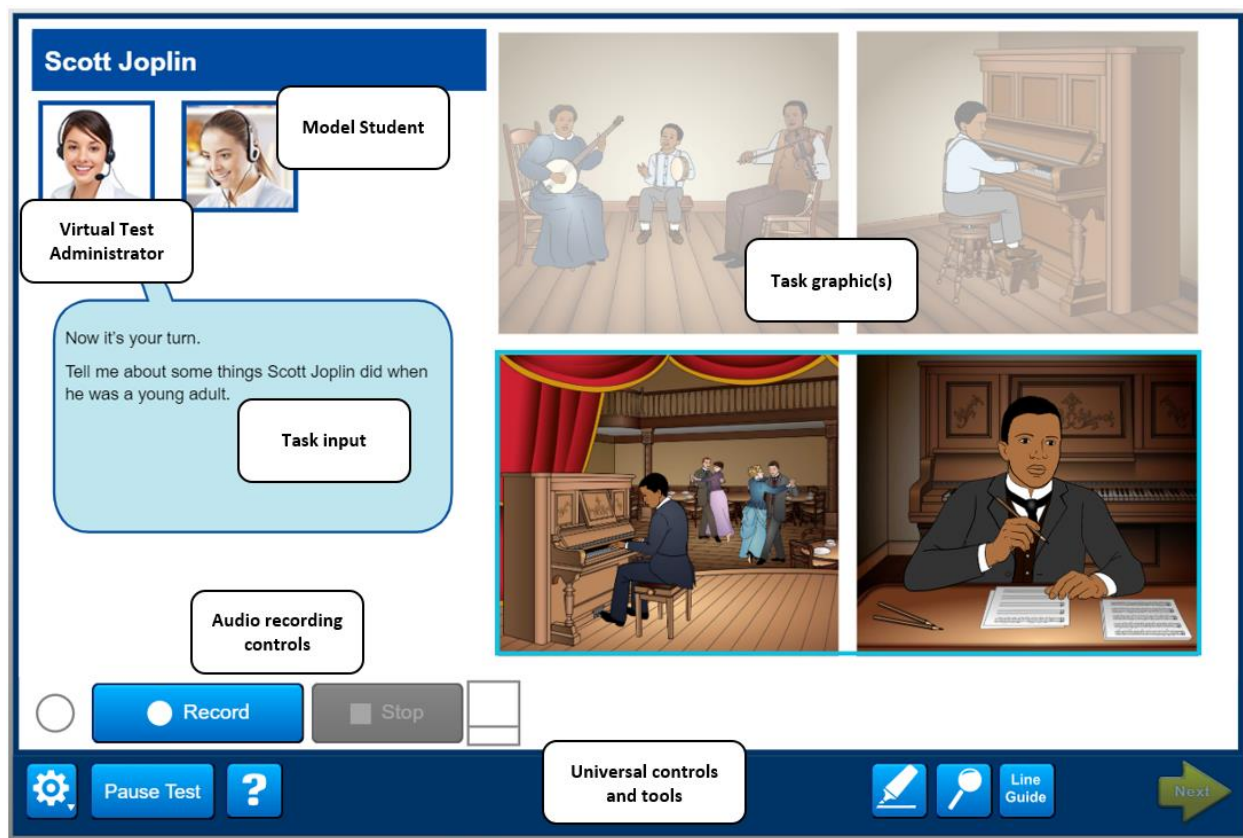


Students navigate through the Speaking test independently and at their own pace. They must listen to all audio on a screen before the test allows them to advance to the next screen. Most students can only listen to the audio stimuli once, although students with a specific accommodation related to audio stimuli may listen to the audio as many times as they wish. The amount of time that students are allowed for recording their responses varies by grade-level cluster and the target proficiency level of the task; tasks targeting a higher proficiency level are permitted more recording time.<sup>3</sup> The amount and complexity of task input varies by grade-level cluster and task level. The purpose of the input is to provide academic content for students to draw on in their responses.

Figure 15 shows the general screen layout of the Speaking test.

---

<sup>3</sup> During the piloting of the Speaking test design before ACCESS Online was operational, the response recording time was one of the variables investigated. CAL and WIDA jointly determined the recording times. These times were a compromise between the minimum and maximum times considered. This allows for more time than minimally necessary, while not allowing so much time that students who have already provided a sufficient response feel the need to fill all of the available time.



**Figure 15: Visualization of the Speaking test screen layout**

Both the VTA and the model student are represented within the testing interface by static images. They are portrayed wearing computer headsets with microphones to reflect the actual testing scenario. Test input and stimuli are presented both aurally and in speech bubbles on the screen. Students respond orally to the tasks, with their responses recorded and transmitted to DRC for later scoring.

All Speaking tasks for a given grade cluster and WIDA Standard are designed in terms of *panels*; a panel is a thematically related set of three tasks, targeting the elicitation of PL 1, PL 3, and PL 5 language. When the tasks are field tested, the panels are split out into folders, with each folder containing one or two tasks. Tier Pre-A folders contain a single task targeting PL 1; Tier A folders contain two tasks targeting PL 1 and PL 3; and Tier C folders contain two tasks targeting PLs 3 and 5. For a given pair of Tier A and Tier C folders based on a single panel, the PL 3 task is identical in both folders (see Figure 19 in Section 2.2.4 for an illustration).

## **2.2 Test Design**

This section describes how ACCESS for ELLs Online is assembled to ensure that the evidence collected is (1) sufficient to make the required decisions based on the test results, and (2) appropriate for the student’s level of proficiency. To tailor the test closely to student ability levels while still including items and tasks that assess all the Standards, adaptivity has been built into the test. The Listening and Reading tests both use a multistage adaptive test design. The Writing and Speaking tests are tiered, and placement into the tiers depends on performance on the Listening and Reading tests.

For all four domains, the test design is broken into different tiers (as described in Section 1.6 above) and stages (as described in this section). For each tier and stage within a given grade cluster, a single folder is earmarked for that “slot” on the test. Items selected for each slot must meet strict criteria (in terms of difficulty) to be placed in that slot. This ensures that the item pool is adequate to support the multistage administrations, including the adaptive component in Listening and Reading.

### **2.2.1 Listening**

For the ACCESS Listening test, Table 1 shows, for each grade-level cluster and tier pool, the number of items, the targeted range of WIDA proficiency levels, the proportion of items by item type (MC – multiple choice; DD – drag-and-drop; HS – hot spot), the response format, and the scoring procedure.

Table 1. Number and Types of Items on the ACCESS 503 Listening Test

Grade-Level Cluster	Tier Pool	Number of Items	Targeted PL range	Item Types and Percentages*			Response Formats	Scoring Procedures
				MC	DD	HS		
1	Entry	6	PL1–PL4	83%	0%	17%	Dichotomous selected response	Machine scored
1	A	12	PL1–PL3	100%	0%	0%		
1	B	18	PL2–PL4	89%	0%	11%		
1	C	18	PL3–PL5	100%	0%	0%		
2–3	Entry	6	PL1–PL4	100%	0%	0%	Dichotomous selected response	Machine scored
2–3	A	12	PL1–PL3	100%	0%	0%		
2–3	B	18	PL2–PL4	78%	0%	22%		
2–3	C	18	PL3–PL5	94%	0%	6%		
4–5	Entry	6	PL1–PL4	100%	0%	0%	Dichotomous selected response	Machine scored
4–5	A	12	PL1–PL3	100%	0%	0%		
4–5	B	18	PL2–PL4	83%	0%	17%		
4–5	C	18	PL3–PL5	100%	0%	0%		
6–8	Entry	6	PL1–PL4	83%	0%	17%	Dichotomous selected response	Machine scored
6–8	A	12	PL1–PL3	92%	0%	8%		
6–8	B	18	PL2–PL4	55%	28%	17%		
6–8	C	18	PL3–PL5	94%	0%	6%		
9–12	Entry	6	PL1–PL4	100%	0%	0%	Dichotomous selected response	Machine scored
9–12	A	12	PL1–PL3	92%	0%	8%		
9–12	B	18	PL2–PL4	84%	5%	11%		

9–12	C	18	PL3–PL5	100%	0%	0%		
------	---	----	---------	------	----	----	--	--

\*Item types are MC – multiple choice; DD – drag-and-drop; HS – hot spot.

The Listening test uses a multistage adaptive design, as illustrated in Figure 16. All students begin the Listening test with two entry folders (with three items each) at Stage 1 and Stage 2, both targeting Social and Instructional Language (see Section 1.2 for the WIDA ELD Standards). At that point, the test engine estimates the student’s ability based on the student’s performance on those six items, and the engine then uses that ability estimate to determine which of the three leveled Language of Language Arts folders in Stage 3 is administered next. Students whose ability estimate predicts a PL score of 5.0 or higher are routed into the folder at the highest level (C in Figure 16); students whose ability estimate predicts a PL score of 2.5 or lower are routed into the folder at the lowest level (A in Figure 16); all others are routed into the B folder. Throughout the test, the test engine re-estimates a student’s underlying measure of ability with the completion of each folder, and the engine then uses that information to choose the level of the next folder to be administered, following the decision rules above. Thus, each student will trace a tailor-made path through the test according to ability level, but the order of the stages is invariant across students. In total, there are eight possible stages, but a student whose ability estimate falls below PL 2.5 after the sixth stage ends the test at that point. This shortening of the test for students at the lower proficiency levels allows them to demonstrate what they know without subjecting them to additional content, when their ability is not near the cut point where the EL reclassification decision is made. The intent of this design is to ensure coverage of the Standards while delivering a test that closely matches the student’s PL, thus minimizing measurement error. Although timing guidance is included in the Test Administrator Manual (WIDA Consortium, 2021a), the Listening test is untimed.



Table 2 shows, for each grade-level cluster and tier pool, the number of items, the targeted range of WIDA proficiency levels, the proportion of items by item type (MC – multiple choice; DD – drag-and-drop; HS – hot spot), the response format, and the scoring procedure.

Table 2. Number and Types of Items on the ACCESS 503 Reading Test

Grade-Level Cluster	Tier Pool	Number of Items	Targeted PL range	Item Types and Percentages*			Response Formats	Scoring Procedures
				MC	DD	HS		
1	Entry	6	PL1–PL4	100%	0%	0%	Dichotomous selected response	Machine scored
1	A	18	PL1–PL3	100%	0%	0%		
1	B	24	PL2–PL4	96%	0%	4%		
1	C	24	PL3–PL5	100%	0%	0%		
2–3	Entry	6	PL1–PL4	100%	0%	0%	Dichotomous selected response	Machine scored
2–3	A	18	PL1–PL3	100%	0%	0%		
2–3	B	24	PL2–PL4	96%	0%	4%		
2–3	C	24	PL3–PL5	100%	0%	0%		
4–5	Entry	6	PL1–PL4	100%	0%	0%	Dichotomous selected response	Machine scored
4–5	A	18	PL1–PL3	94%	0%	6%		
4–5	B	24	PL2–PL4	96%	4%	0%		
4–5	C	24	PL3–PL5	96%	4%	0%		
6–8	Entry	6	PL1–PL4	100%	0%	0%	Dichotomous selected response	Machine scored
6–8	A	18	PL1–PL3	100%	0%	0%		
6–8	B	24	PL2–PL4	100%	0%	0%		
6–8	C	24	PL3–PL5	100%	0%	0%		
9–12	Entry	6	PL1–PL4	100%	0%	0%	Dichotomous selected response	Machine scored
9–12	A	18	PL1–PL3	100%	0%	0%		
9–12	B	24	PL2–PL4	100%	0%	0%		



9–12	C	24	PL3–PL5	100%	0%	0%		
------	---	----	---------	------	----	----	--	--

\*Item types are MC – multiple choice; DD – drag-and-drop; HS – hot spot.

Figure 17 shows the format of the Reading test. The format and adaptivity are like those of the Listening test, but the Reading test consists of 10 stages rather than eight. This reflects the greater weight given to Reading in calculating the composite scores (see Part 2, Chapter 3, “Analyses of Composite Scores”), as well as the view that literacy skills are paramount in developing academic language proficiency. The greater weight afforded to Reading and Writing resulted from a policy decision by the WIDA Board before the first operational administration of ACCESS. A student whose ability estimate falls below PL 2.5 after the eighth stage ends the test at that point. Although timing guidance is included in the Test Administrator Manual, the Reading test is untimed.

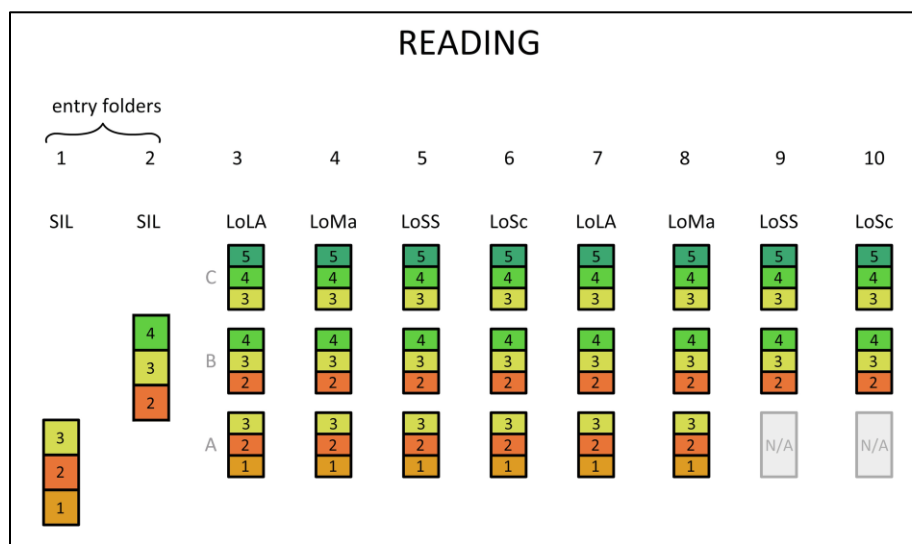


Figure 17. Format of the Reading test

## 2.2.3 Writing

For the ACCESS Writing test,

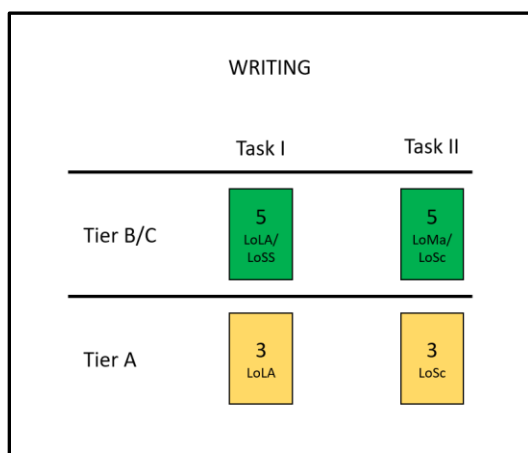
Table 3 shows, for each grade-level cluster and tier, the number of tasks, the targeted range of WIDA proficiency levels, the task type, the response format, and the scoring procedure.

**Table 3. Number and Types of Tasks on the Writing Test**

Grade-Level Cluster	Tier	Number of Tasks	Targeted PL Range	Task Type	Response Formats	Scoring Procedures
1	A	2	PL1–PL3	Writing constructed response	Polytomous constructed response; handwritten in test booklet	Human scored: centrally scored by DRC
1	B/C	2	PL2–PL5			
2–3	A	2	PL1–PL3	Writing constructed response	Polytomous constructed response; handwritten in test booklet	Human scored: centrally scored by DRC
2–3	B/C	2	PL2–PL5			
4–5	A	2	PL1–PL3	Writing constructed response	Polytomous constructed response; handwritten in response booklet or keyboarded in test platform	Human scored: centrally scored by DRC
4–5	B/C	2	PL2–PL5			
6–8	A	2	PL1–PL3	Writing constructed response	Polytomous constructed response; handwritten in response booklet or keyboarded in test platform	Human scored: centrally scored by DRC
6–8	B/C	2	PL2–PL5			
9–12	A	2	PL1–PL3	Writing constructed response	Polytomous constructed response; handwritten in response booklet or	Human scored: centrally
9–12	B/C	2	PL2–PL5			

					keyboarded in test platform	scored by DRC
--	--	--	--	--	-----------------------------	---------------

As shown in Figure 18, the format of the Writing test is tiered. As Writing tasks are polytomous and elicit a range of student performances, each task is targeted to elicit language across a range of proficiency levels, rather than targeted to a single proficiency level. Tier A consists of tasks written to elicit language up to PL 3, while Tier B/C tasks are designed to elicit language up to PL 5. This is indicated by the large number in the colored rectangle in the figure. However, for both tiers of the test, DRC raters score students' responses to all tasks using the entire breadth of the scoring scale. Students can theoretically score anywhere from 0 to 9 on any task (in terms of the raw scores in the scoring scale), although the design of some tasks limits the possible scores. For example, Tier A tasks are not designed to elicit extended responses, so although the tasks are scored using the entire scale, these tasks do not elicit language above PL 4. Likewise, although Tier B/C tasks are designed to elicit extended discourse so that students can display proficiency at PL 5 or even PL 6, students' performances on these tasks may range from PL 1 to PL 6.



**Figure 18. Format of the Writing test**

Beginning with Series 501, both tiers consist of two tasks. Prior to Series 501, all test forms had three tasks, except for Grade 1 Tier A, which consisted of four tasks. This change was made starting with Series 501 to accommodate an embedded field test design for field testing Series

502 Writing tasks. Tier A tasks target a single WIDA Standard; for all grade level clusters except Grade 1, Task I targets Language of Language Arts and Task II targets Language of Science, while for Grade 1, Task I targets Language of Science and Task II targets Language of Language Arts. Tier B/C tasks integrate more than one WIDA Standard; Task I integrates Language of Language Arts and Language of Social Studies, and Task II integrates Language of Math and Language of Science. The ways in which the Standards are targeted by these tasks vary across grade levels and are spelled out in the generative item specifications.

The design of the embedded Writing field test for Series 503 is described in greater detail in Section 2.3.2.3 below.

Placement into tiers on the Writing test depends on the scores that students receive based on their performances on the Listening and Reading tests (which the test engine scores automatically). To determine how to best place each student into an appropriate tier, the CAL psychometrics team carried out logistic regression analyses to examine the relationship between student performance on the Listening and Reading tests administered in 2015–2016 and their performance on the Writing test. They then used this information to program an algorithm into the ACCESS Online test that the test engine uses to determine which tier of the Writing test to administer to each student. The purpose of the algorithm is to place students who are predicted to score above PL 3.0 into Tier B/C for the Writing test. All other students are placed into Tier A.

Although timing guidance is included in the Test Administrator Manual, the Writing test is untimed.

## 2.2.4 Speaking

For the ACCESS Speaking test, Table 4 shows, for each grade-level cluster and tier, the number of tasks, the targeted range of WIDA proficiency levels, the task type, the response format, and the scoring procedure.

**Table 4. Number and Types of Tasks on the Speaking Test**

Grade-Level Cluster	Tier	Number of Tasks	Targeted PL range	Task Type	Response Formats	Scoring Procedures
---------------------	------	-----------------	-------------------	-----------	------------------	--------------------

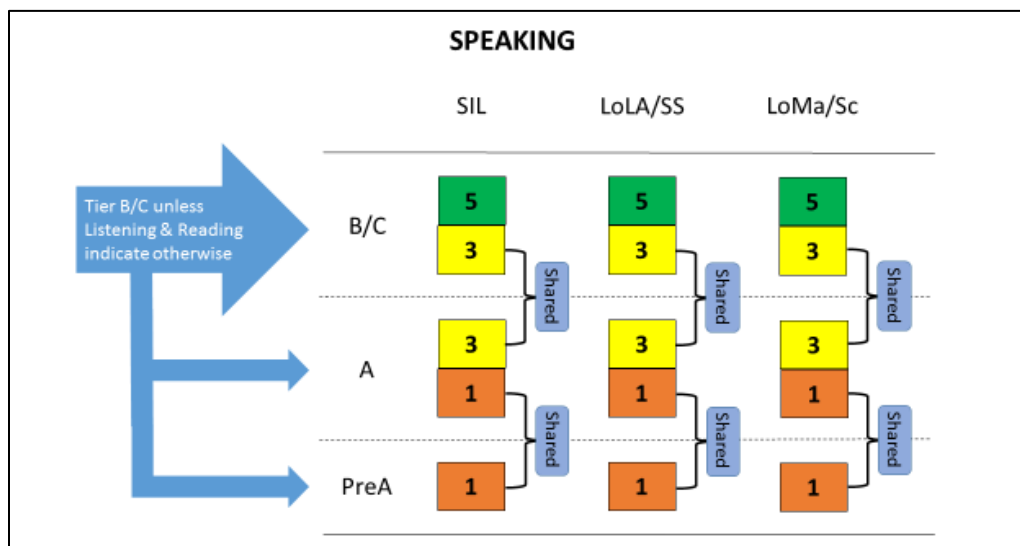
1	Pre-A	3	PL1	Speaking constructed response	Polytomous constructed response	Human scored; centrally scored by DRC
1	A	6	PL1–PL3			
1	B/C	6	PL3–PL5			
2–3	Pre-A	3	PL1	Speaking constructed response	Polytomous constructed response	Human scored; centrally scored by DRC
2–3	A	6	PL1–PL3			
2–3	B/C	6	PL3–PL5			
4–5	Pre-A	3	PL1	Speaking constructed response	Polytomous constructed response	Human scored; centrally scored by DRC
4–5	A	6	PL1–PL3			
4–5	B/C	6	PL3–PL5			
6–8	Pre-A	3	PL1	Speaking constructed response	Polytomous constructed response	Human scored; centrally scored by DRC
6–8	A	6	PL1–PL3			
6–8	B/C	6	PL3–PL5			
9–12	Pre-A	3	PL1	Speaking constructed response	Polytomous constructed response	Human scored; centrally scored by DRC
9–12	A	6	PL1–PL3			
9–12	B/C	6	PL3–PL5			

Figure 19 shows the format of the Speaking test. The Speaking test includes tasks that target language elicitation at three PLs: 1, 3, and 5. The tasks are grouped into thematic folders, each of which is aligned to one or two of the WIDA Standards. These folders are generally presented in the same order as the folders on the Listening and Reading tests; folders aligned to Social and Instructional Language are presented first, then folders aligned to Language of Language Arts, then folders aligned to Language of Math.

As shown in Figure 19, the Speaking test includes three tiers: Tier Pre-A, Tier A, and Tier B/C. Tier Pre-A includes tasks that target elicitation of language at PL 1. Tier A includes tasks that

target elicitation of language at PLs 1 and 3. Tier B/C includes tasks that target elicitation of language at PLs 3 and 5.

A thematic panel refers to the folders across all tiers within a grade-level cluster that relate to a particular WIDA ELD Standard. In other words, the Tier B/C, Tier A, and Tier Pre-A folders that address Social and Instructional Language in each grade cluster make up a single thematic panel, with the PL 1 and PL 3 tasks shared across tiered folders in a panel. For example, within a Social and Instructional Language panel, the same PL 3 task appears on both the Tier A and the Tier B/C forms of the test, and the same PL 1 task appears on both the Tier Pre-A and Tier A forms of the test.



**Figure 19. Format of the Speaking test**

As with the Writing test, placement of students into the three tiers on the Speaking test depends on their performance on the Listening and Reading tests. Unlike Writing, the Speaking test has one additional tier, Tier Pre-A. Students are placed into Tier Pre-A when their scores on both the Listening and Reading tests are below PL 2.0. The Speaking Pre-A tier is designed to meet the needs of students in the very early stages of English language development. As noted previously, these tasks are targeted to the P1 level. DRC raters score students' responses to these tasks using a modified version of the full Speaking rating scale (see Section 3.2.4).

The process for placing students into Tiers A and B/C for the Writing test is analogous to the process used for tier placement for the Speaking test. The CAL psychometrics team carried out logistic regression analyses using test data for all students who were administered the assessment in 2015–2016 (i.e., the first year of the ACCESS Online assessment) to examine the relationship between students’ performances on the Listening and Reading tests and their performance on the Speaking test. They used this information to program an algorithm into the ACCESS 2.0 Online test that the test engine used to determine which tier of the Speaking test to administer to each student. The purpose of the algorithm is to place students who are predicted to score above PL 3.0 into Tier B/C for the Speaking test, based on their performances in the Listening and Reading tests, and to place all other students into Tier A (except for those students who, as noted previously, are routed into Tier Pre-A).

Although timing guidance is included in the Test Administrator Manual, the Speaking test is untimed.

## **2.3 Test Construction**

### **2.3.1 Item and Task Development**

The ACCESS item/task development process spans approximately three years and follows a standardized test development cycle. Each cycle begins with the development of a Refreshment Plan. The CAL TD team develops the Refreshment Plan, taking several factors into consideration, including empirical item/task performance, length of time that folders have been on the test, item/task-specification level information, and the success (or lack thereof) in refreshing the test for each targeted slot in the previous cycle. The CAL TD team presents the Refreshment Plan to the WIDA Assessment team for approval, with ultimate signoff by WIDA’s Director of Test Development.

Upon receiving sign-off on the Refreshment Plan, the CAL TD team then determines which item/task specifications need to be updated or replaced and which can move forward as is. Generally, the CAL TD team updates or replaces item/task specifications for two reasons:

- The CAL TD team analyzes prior items and tasks that could not be used operationally due to fit issues, or in cases where the item or task fit was acceptable, an item or task difficulty measure that was outside of the range for the intended



slot on the test. The purpose of this analysis is to determine if the poor performance is due to item/task mechanics (e.g., an issue with the wording of the passage or stem, a distractor that is too attractive) or if there is a deeper item/task-specification issue that cannot be resolved (e.g., the specification is difficult to operationalize successfully). In the latter case, the CAL TD team can update the specification (usually focused on updating the MPIs) or completely replace it, depending on the specific situation.

- The CAL TD team also updates or replaces item/task specifications as content standards change. As noted previously, the ACCESS item/task specifications include explicit connections to the content standards. If an update to the relevant content standard makes an ACCESS item/task specification obsolete, the CAL TD team revises or replaces the specification.

Once updates to item/task specifications are complete, item and task development begins. The generation of initial item/task content occurs in two interconnected steps. First, the CAL TD team initiates a process of theme generation. In the ACCESS item/task specifications, the CAL TD team writes each specification to a broad topic related to the given WIDA ELD Standard, and a theme is a more focused instantiation of the topic. For example, if the topic for a Language of Social Studies item/task specification for Grades 4–5 is U.S. history, an example of an appropriate theme might be “the Industrial Revolution.”

The CAL TD team and WIDA TD staff are responsible for recruiting classroom English as a second language (ESL) and content teachers with experience teaching the academic content associated with one or more of the WIDA ELD Standards (including educators with experience working with English learners with disabilities), and the CAL TD team provides these educators with key parts of the item/task specification document (i.e., the topic, the MPIs, and guidelines for selecting a good theme). Then, the CAL TD team asks educators to propose themes related to the topic, along with possible directions for each item or task, which are grade-level appropriate. After the theme generation process is complete, the CAL TD team reviews the list of themes to identify those that will become the focus of item/task writing. This determination is based on several factors, including operationalizability on a large-scale assessment (since many ideas from educators are well suited for the classroom but do not clearly translate to the assessment context), themes currently in use on the assessment, and bias and sensitivity considerations.

The team then assigns themes to professional item/task writers to develop the initial item/task content. The team recruits individuals with prior experience developing ESL or English language arts items/tasks, preferably in the context of large-scale, standardized assessments, but individuals with other experience (such as experience writing items/tasks for language tests in languages other than English, and experience with English placement tests for the college/university setting) are also considered. All item/task writers, both new item/task writers and those returning from the previous test development cycle, participate in an introductory training, and the team provides them with extensive documentation regarding writing items/tasks for ACCESS, including an Item Writing Handbook and ancillary documents (i.e., checklists, item/task specifications, templates) to complete their assignments. One or more CAL Language Testing Specialists work with each item/task writer, providing feedback on the item content.

After item/task writing is complete, CAL Language Testing Specialists and Test Development Managers review the folders, using a standard checklist, to determine which folders will undergo further development and which will be retired. Folders then go to their first external review, the Standards Expert review.

During the Standards Expert review, educators provide feedback about the overall grade-level appropriateness of the language and content of the items/tasks to ensure that no drift, in terms of grade-level appropriateness of the content or the language, has occurred between initial theme generation and item/task writing. The CAL TD team and WIDA TD staff are jointly responsible for recruiting educators with ESL and content-area expertise to serve as Standards Experts. CAL Language Testing Specialists prepare a short questionnaire with both yes/no and open-ended questions about each folder and send the questionnaires and folders to the Standards Experts.

Subsequent to the Standards Expert review, all content proceeds through a rigorous folder refinement stage internal to CAL. Folder refinement includes numerous steps, including additional research and sourcing/fact-checking, meticulous review against a comprehensive, industry-standard item/task development checklist with peer review that other Language Testing Specialists carry out, as well as review by Test Development Managers and the Director of Test Development and successive rounds of revision before sign-off. During this stage, all aspects of the items and tasks are scrutinized: the WIDA proficiency level of the stimulus, the graphic support, the question stems, and response options (for the Listening and Reading tests) and task prompts (for the Speaking and Writing tests). The CAL TD team also conducts mock

administrations. During this phase, CAL Language Testing Specialists produce other ancillary materials, such as Test Administrator scripts. Upon sign-off, the CAL TD team works with the CAL Production and Tech teams to generate the graphics used on the test and to begin the development of the question and test interoperability (QTI) packages for the online assessment. A QTI package is a collection of files that contain all the item/task content, including assets such as graphics and audio files, coded so that the test engine can read them. There is one QTI package for each folder on ACCESS. Once the graphics are generated, CAL Language Testing Specialists inserted them into the folders and conducted layout review and fact-checking (with Test Development Manager sign-off) to ensure that the items and tasks are ready for external Content Review and Bias and Sensitivity Review.

Content Review and Bias and Sensitivity Review are external reviews that educators and WIDA TD staff carry out on ACCESS items and tasks. WIDA TD staff are responsible for assembling these panels by recruiting educators of multilingual learners from around the consortium, including culturally, racially, and linguistically diverse educators who reflect the population of students that take WIDA assessments. WIDA employs several criteria when recruiting educators to perform these tasks. The criteria used to recruit educators to conduct Content Reviews differ somewhat from the criteria used to recruit educators to conduct Bias and Sensitivity Reviews.

Educators conduct Content Reviews by grade-level cluster (G1, G2-3, G4-5, G6-8). The educators who are recruited to review a particular grade cluster's content (four reviewers per grade cluster) have experience teaching English language learners and are either currently teaching students who are in that grade cluster or have extensive prior experience teaching students who are in that grade cluster. Additionally, educators serving on each panel represent different content areas. WIDA TD staff seek to ensure that each panel includes at least one educator who has teaching experience in each of the following content areas: ELA, Science, Math, Social Studies, and Special Education. Additionally, during the recruitment process, WIDA TD staff seek to ensure diversity and balance across (1) consortium states, (2) school locale (rural/suburban/urban), and (3) years of teaching experience. The CAL TD team and WIDA TD staff first train the Content Review Panel on the procedures and scope of the review. The panelists are introduced to the test layout, instructed on the logistics of the review, and trained to use the review checklist. The panel members then individually review each item and task, followed by a collective discussion of each item and task to determine (1) whether the

content is accessible and relevant to students in the targeted grade-level cluster, (2) is at the targeted WIDA proficiency level, and (3) matches the Model Performance Indicator from the WIDA English Language Development Standards that it is intended to assess.

The Bias and Sensitivity Review Panel ensures that test items and tasks are free of material that (1) might favor any subgroup of students over another on the basis on gender, race/ethnicity, home language, religion, culture, region, or socioeconomic status, and (2) might be upsetting to students. Educators conduct Bias and Sensitivity Reviews by grade groupings (e.g., G1-3, G4-5, G6-8, and G9-12). The educators who are recruited to review a particular grade cluster's content (5 or 6 reviewers per grade grouping) are educators or school administrators who have experience teaching English language learners and are either currently teaching students who are in that grade cluster or have extensive prior experience teaching students who are in that grade cluster. WIDA TD staff employ additional criteria to ensure that a variety of perspectives are represented on each panel. These criteria include recruiting at least one educator with experience in Special Education to serve on each panel. Additionally, during the recruitment process, WIDA TD staff seek to ensure diversity and balance across (1) consortium states, (2) school locale (rural/suburban/urban), and (3) years of teaching experience. The CAL TD team and WIDA TD staff conduct training for all new and returning reviewers before any items or tasks are reviewed. The panel members then individually review each item and task, followed by a collective discussion of each item and task to determine if any bias or sensitive topics are detected in the items/tasks, and if so, what the CAL TD team can do to remediate the issues. The CAL TD team and WIDA TD staff facilitate the reviews and take extensive notes to capture all feedback during the reviews. WIDA TD staff also conduct a separate, asynchronous review around the time of the Content Review and Bias and Sensitivity Review, using the same materials that the educators review, and provide written feedback on the materials.

The CAL Language Testing Specialists compile all the Content Review and Bias and Sensitivity Review feedback from educators and from WIDA TD staff, and then work to implement the feedback, with the CAL Test Development Manager sign-off as a final step. The CAL Test Production and Tech teams then revise the graphics and the QTI packages. The input and feedback from educators at various stages in the item/task development process serves as evidence that each item and task is appropriate for the age and grade-level cluster for which it is intended.

Tasks in the Writing domain undergo one additional step: a small-scale tryout with educators and students. Given the changes to the Writing test over the past few years, including a change from three to two operational tasks, along with changes to task specifications to better align the Writing tasks with classroom practice, these tryouts allow the CAL TD team to evaluate whether each Writing task will effectively elicit language at its targeted WIDA proficiency level. For the Writing tryouts, the CAL TD team and WIDA TD staff jointly recruit educators with appropriate numbers of students at the targeted proficiency levels (approximately 15 students per task) to participate. The CAL Test Development Manager for Writing prepares a recruitment flyer, which the CAL TD team and WIDA TD staff circulate to educators. The CAL TD team circulates the flyer to educators who have previously participated in the tryouts, and WIDA TD staff circulate the flyer through WIDA's regular SEA/LEA communications emails. Due to the small-scale nature of the tryouts, the recruitment is ultimately a convenience sample, although CAL and WIDA strive to obtain a sample with geographic diversity (i.e., educators distributed throughout the consortium, with a mix of urban, suburban, and rural representation). In addition, the tasks target different proficiency levels/tiers at the different grade level clusters, so recruiting educators with students at these targeted proficiency levels and grade level clusters is another requirement. Finally, we do not recruit educators who have already reviewed the tasks in development during Bias and Sensitivity and Content review to participate in tryouts. If the CAL TD team determines that the first round of recruitment has failed to find educators with students at the appropriate proficiency levels for all grade clusters and tiers, the CAL Test Development Manager for Writing identifies the grade-level clusters and proficiency levels/tiers with gaps and provides this information to the WIDA TD staff, who can then do targeted recruitment based on existing databases of educators who have indicated willingness to participate in test development activities.

The educators administer the tasks to their students and send the students' written responses back to CAL for analysis. The students and the educators also fill out short surveys about the tasks. The students each fill out a six-question/prompt survey answering questions like "I understood what to do." and "This is an interesting topic to write about." The educators complete an eight-question survey focusing on the effectiveness and appropriateness of the task input and graphics, the comparability of the task to first-draft writing in class, and student familiarity and engagement with the task content.

CAL Language Testing Specialists conduct qualitative analyses of the student responses and the survey data and use the results to inform any final revisions to the tasks prior to field testing. For some tiers, the tryouts also inform which task moves on to field testing and which is postponed, in cases where only a single task is field tested. (See Section 2.3.2 for more information regarding the field test design.)

After the CAL Language Testing Specialists complete edits from the Content Review and Bias and Sensitivity Review (and tryout edits for Writing), they then prepare the folders for final production. Additionally, they produce audio recording scripts for professional audio recording, arrange for recording the audio files, complete extensive quality control checks for both content and technical specifications of the audio (e.g., file types, recording quality, and compression levels), conduct final layout reviews, and perform key checks for the Listening and Reading tests. Both the CAL TD team and WIDA TD staff conduct quality control checks of the QTI. The WIDA TD staff sign off on all materials before DRC builds the final test forms in the test engine. Items and tasks that reach this point then go through field testing processes, described in the next subsection by domain.

Throughout the item/task development process, the CAL TD team focuses on issues of fairness. The team applies the seven Universal Design of Assessment (UDA) principles when creating items and tasks:

- Inclusive assessment population
- Precisely defined constructs
- Accessible, nonbiased items/tasks
- Amenability to accommodations
- Simple, clear, and intuitive instructions and procedures
- Maximum readability and comprehensibility
- Maximum legibility

Additionally, when CAL's TD managers, WIDA TD staff, and external reviewers conduct Standards Expert reviews, Content Reviews, and Bias and Sensitivity Reviews, they use checklists that ask them to consider the seven principles of universal design as they are reviewing each item and task.

In recent years, WIDA has placed additional focus on ensuring that the items and tasks, and especially the graphics, are amenable to accommodations by involving WIDA's Accessibility and Accommodations Team directly in the item/task review process. WIDA's Accessibility and Accommodations Team helped CAL's TD team develop principles for graphics development and for eliminating language that is biased towards students with sight, and WIDA's Accessibility and Accommodations Team also reviews the items and tasks during development to help CAL identify areas that still need to be addressed.

Through maintaining a focus on fairness throughout the test development cycle by integrating the principles of UDA in various steps, the CAL TD team strives to ensure that ACCESS Online items and tasks are best positioned to be maximally fair for all populations.

## **2.3.2 Field Testing**

### *2.3.2.1 Listening*

DRC field tested the Listening items developed for Series 503 as embedded folders during the operational administration of Series 502. The embedded field test folders contained items that featured innovative formats, including hot spot items (i.e., the student clicks on an area of the screen to respond) and drag-and-drop items (i.e., the student drags an image/text to a specified screen area to respond).

For Series 503, DRC field tested a total of 120 Listening items (40 folders), across all five grade-level clusters, as indicated in Table 5.

Table 5. Number of Field Test Folders and Items for the Series 503 Listening Test

Grade-Level Cluster	Tier Pool	Number of Folders to Refresh	Number of Overage Folders	Total Number of Field Test Folders	Total Number of Field Test Items	Standards Addressed in FT
1	Entry	1	0	1	3	SIL
1	A	0	0	0	0	
1	B	1	0	1	3	LoMa
1	C	2	2	4	12	LoMa, LoSc
2–3	Entry	1	1	2	6	SIL
2–3	A	0	0	0	0	
2–3	B	3	3	6	18	LoMa, LoSc
2–3	C	3	1	4	12	LoLA, LoMa, LoSc
4–5	Entry	1	0	1	3	SIL
4–5	A	1	0	1	3	LoSS
4–5	B	0	0	0	0	
4–5	C	3	2	5	15	LoMa, LoSS, LoSc
6–8	Entry	1	0	1	3	SIL
6–8	A	0	0	0	0	
6–8	B	2	2	4	12	LoLA, LoSS
6–8	C	1	0	1	3	LoLA
9–12	Entry	1	0	1	3	SIL



9–12	A	2	2	4	12	LoLA, LoMa
9–12	B	0	0	0	0	
9–12	C	2	2	4	12	LoLA, LoMa
Total		25	15	40	120	

Each student received one Listening field test folder embedded in the operational test. Field test folders are targeted to refresh a specific operational folder on the test, and field test folder specifications include the stage, WIDA ELD Standard, and tier pool target (i.e., Entry, A, B, or C) of the folder. Students received the embedded field test folder at the stage targeted for refreshment, with administration randomized so that half of the students saw the field test folder before the corresponding operational folder, and half saw the operational folder before the field test folder. Field test folders were administered to those students who were routed to take the operational folder that was either at the same tier or adjacent to the tier that the field test folder targeted. When DRC drew the field test samples, 50% of the sample were students who were routed to the tier that the field test folder targeted, and the other 50% were students who were routed to adjacent tiers. (If there were adjacent tiers both above and below the field test target, then 25% of the sample were students routed to each of those tiers.) In cases where the field test folder was to be placed in one of the entry stages, students receiving that field test folder took it directly after the pair of operational entry folders. CAL set the field test sample targets for the Listening test at a minimum of 3,000 responses per folder.

Because CAL's psychometrics team used the Listening field test data in the pre-equating analysis, their sample size requirement of 3,000 was much higher than the minimum of 250 per form for high-stakes tests that Linacre (1994) proposed, to ensure that the pre-equated parameter estimates would be stable. Linacre (1994), citing Wright and Douglas's (1977) formulation, explained how to determine the minimum sample required for calibrating dichotomous items to achieve various levels of estimation precision and confidence intervals. With a sample size of

3,000, one can be 95% confident that no item parameter will be more than  $\pm 0.1$  logit away from its true value. The sample sizes for all field test folders exceeded the minimum requirement of 3,000, except for one Listening grade Cluster 4–5 Tier A folder, which had a sample size of 2,800, due to the fact that the population of grade 4-5 tier A students comprise a smaller portion of the population than students in other grade level clusters and tiers.

After CAL's psychometrics team accessed the field test data, they analyzed students' responses to the items in the field test folders to determine each item's psychometric properties, and folders for which all three items met established psychometric standards (as described below) were eligible for inclusion in the next year's operational test.

The team then classified each item using the classification system shown in

Table 6. If all three items in a folder were green, the entire folder was eligible for operational use. If one or more items were red, the folder was no longer considered appropriate for operational use. If one or more items were yellow, the Post–Field Test Review Panel reviewed the content of each item, along with relevant statistics contained in the distractor analyses (e.g., the mean ability of students selecting the key vs. the mean ability of students selecting the distractors; infit and outfit statistics for each response option, the point measure correlation of each response option, the percentage of students selecting each response option), to determine if each item would be reclassified as green or red. If all yellow items in a folder were reclassified as green (and there were no red items in the folder), the folder was deemed appropriate for operational testing.

**Table 6. CAL's Post-Field Test Review Classification System for Series 503**

Color	Interpretation	Definition
Green	Appropriate for operational testing	A- or B-level DIF AND a $p$ value $\geq .85$ OR infit and outfit $\leq 1.20$
Yellow	Content review is required to confirm item is appropriate for operational testing	C-level DIF OR infit/outfit $> 1.20$ and $\leq 1.50$  Three-response choice item with $p$ value $\leq .40$ and outfit $< 1.75$  Four-response choice item with $p$ value $\leq .35$ and outfit $< 1.75$
Red	Not appropriate for operational testing	Infit/outfit $> 1.50$

#### *2.3.2.2 Reading*

DRC field tested the Reading items developed for Series 503 as embedded items during the operational administration of Series 502. The embedded field test folders contained items that featured innovative formats, including hot spot items (i.e., the student clicks on an area of the screen to respond) but no drag-and-drop items.

For Series 503, DRC field tested a total of 204 Reading items (68 folders), across all five grade-level clusters, as indicated in Table 7.

**Table 7. Number of Field Test Folders and Items for the Series 503 Reading Field Test**

Grade-Level Cluster	Tier Pool	Number of Folders to Refresh	Number of Overage Folders	Total Number of Field Test Folders	Total Number of Field Test Items	Standards Addressed in FT
1	Entry	1	0	1	3	SIL
1	A	0	0	0	0	
1	B	2	2	4	12	LoSc, LoMa
1	C	3	3	6	18	LoLA, LoMa, LoSc
2–3	Entry	1	0	1	3	SIL
2–3	A	4	4	8	24	LoLA, LoMa, LoSS
2–3	B	0	0	0	0	
2–3	C	1	1	2	6	LoSc
4–5	Entry	1	1	2	6	SIL
4–5	A	1	1	2	6	LoSS
4–5	B	4	4	8	24	LoLA, LoMa, LoSS
4–5	C	2	2	4	12	LoMa, LoSS
6–8	Entry	1	0	1	3	SIL
6–8	A	3	2	5	15	LoLA, LoMa, LoSc
6–8	B	2	2	4	12	LoMa
6–8	C	3	3	6	18	LoLA, LoMa
9–12	Entry	1	0	1	3	SIL

9–12	A	1	1	2	6	LoSc
9–12	B	2	2	4	12	LoLA, LoSc
9–12	C	4	3	7	21	LoLA, LoSS, LoSc
Total		37	31	68	204	

DRC administered the embedded Reading field test in the same way as the embedded Listening field test. As with the Listening test, CAL set the field test sample targets for the Reading test at a minimum of 3,000 responses per folder. The sample sizes for all field test folders exceeded the minimum requirement of 3,000.

After CAL’s psychometrics team accessed the field test data, they analyzed students’ responses to the items in the field test folders to determine each item’s psychometric properties, and folders for which all three items met established psychometric standards (as described in Section 2.3.2.1 above) were eligible for inclusion in the next year’s operational test.

### 2.3.2.3 Writing

DRC administered the Series 503 Writing tasks in an embedded field-test model. For Series 503, a total of 17 Writing tasks were field tested, as indicated in Table 8.

**Table 8. Number of Field Test Tasks for Series 503 Writing**

Grade-Level Cluster	Tier	Number of Folders to Refresh	Number of Folders Field Tested	Standards Addressed in FT
1	A	1	2	LoLA
1	BC	1	2	LoLA/LoSS
2–3	A	1	2	LoLA
2–3	BC	1	2	LoLA/LoSS

4–5	A	1	1	LoLA
4–5	BC	1	2	LoLA/LoSS
6–8	A	1	1	LoLA
6–8	BC	1	2	LoLA/LoSS
9–12	A	1	1	LoLA
9–12	BC	1	2	LoLA/LoSS
Total		10	17	

All students received a field test folder that was appended to their operational assessment. Students received a field test folder in the tier that corresponded to their operational tier. CAL targeted a sample of 500 students per task. This was much higher than the minimum of 250 per form for high-stakes tests that Linacre (1994) proposed, making it likely that, for each of the nine scale categories, there would be at least 10 students whose responses to the task would warrant receiving scores in that category, as Linacre (2002a) recommended for polytomously scored tasks. If raters assign fewer than 10 scores in each scale category, then the category statistics for that category tend to be unstable. Historically, the distribution of scores that raters assign to students' responses to the Writing tasks tends to be highly concentrated in the middle of the score distribution (i.e., exhibit a central tendency effect), with raters assigning relatively fewer scores in the categories at the high end of the score scale. Therefore, CAL targeted a sample size of 500 to ensure that there would be students for analysis whose responses to the task would warrant receiving scores at the high end of the 9-category score scale. Use of this larger sample size also provided examples of students' responses that received scores in the higher scale categories that trainers could use as anchor papers for rater training.

DRC administered the field test under standard testing conditions. The field test used the online interface with keyboarded responses for Grades 4–12 and paper booklets with handwritten responses for Grades 1–3. For the Writing field test, DRC raters scored the students' responses to

the field test tasks. DRC performed a 20% read-behind as a quality control measure, with the first score assigned as the score of record.<sup>4</sup>

#### *2.3.2.4 Speaking*

All Tier A and B/C students received a Speaking field test folder that was appended to their operational Speaking assessment. Tier Pre-A was not included in the field test. DRC field tested a total of 30 folders (15 panels) for Series 503, with a target sample size of 500 students per folder. This is much higher than the minimum of 250 per form for high-stakes tests that Linacre (1994) proposed, making it likely that, for each of the four scale categories, there would be at least 10 students whose responses to the task would warrant receiving scores in that category, as Linacre (2002a) recommended for polytomously scored tasks. Historically, the distribution of scores that raters assign to students' responses to the Speaking tasks tends to be highly concentrated in the middle of the score distribution (i.e., exhibit a central tendency effect), with raters assigning relatively fewer scores in the categories at the high end of the score scale. Therefore, CAL targeted a sample size of 500 to ensure that there would be students for analysis whose responses to the task would warrant receiving scores at the high end of the 4-category score scale.<sup>5</sup> Use of this larger sample size also provided examples of students' responses that received scores in the higher scale categories that trainers could then use as anchor papers for rater training.

DRC-trained raters scored students' responses to the field test Speaking tasks, with a 20% read-behind as a quality control measure and the first score as the score of record.

---

<sup>4</sup> The purpose of the 20% read-behind is to monitor rater performance on a daily basis. (See Section 3.2.2 below.) If the read-behinds detect that one rater is consistently scoring inaccurately, DRC can rescore all of the students' responses to tasks scored by that rater, and the rater can be retrained or terminated. Ratets go through significant training and qualification prior to live scoring, and they are monitored daily through validity and recalibration tasks, so a scenario where a rater is consistently anomalous in his or her ratings would be uncommon, and it would be detected and corrected immediately.

<sup>5</sup> Technically, the score scale includes 5 categories, including "No Response (in English)."



Students received a Speaking field test folder in the tier that corresponded to their operational tier. For Series 503, CAL field tested a total of 30 Speaking folders, as indicated in Table 9.

**Table 9. Number of Field Test Tasks for Series 503 Speaking**

Grade-Level Cluster	Tier	Number of Folders to Refresh	Number of Folders Field Tested	Standards Addressed in FT
1	A	2	3	SIL, LoMa/LoSc
1	BC	2	3	SIL, LoMa/LoSc
2–3	A	2	3	SIL, LoMa/LoSc
2–3	BC	2	3	SIL, LoMa/LoSc
4–5	A	2	3	SIL, LoMa/LoSc
4–5	BC	2	3	SIL, LoMa/LoSc
6–8	A	2	3	SIL, LoMa/LoSc
6–8	BC	2	3	SIL, LoMa/LoSc
9–12	A	2	3	SIL, LoMa/LoSc
9–12	BC	2	3	SIL, LoMa/LoSc
Total		20	30	

### 2.3.3 Item/Task Review and Selection

After the analysis of field test data, a panel consisting of members of the WIDA TD and psychometrics staff, the CAL TD Team, and the CAL psychometrics team conducted an item/task selection meeting to determine which of the field-tested folders would be placed on the Series 503 operational assessment. Results from qualitative and quantitative analyses guided the selection of operational items and tasks.

In the domains of Listening and Reading, item selection was a two-step process. First, the Item Selection Panel reviewed the field test results. CAL’s psychometrics team used a three-tier color-coding system for field test review. Items are coded as “green,” “yellow,” or “red,” and CAL’s psychometrics team then assigned each folder a color based on the least favorable item in the folder. In other words, a folder with a red item was always coded as red, a folder with a yellow

item (but no red items) was coded yellow, and folders were coded green only when all items were green.

Items were coded by color according to the following criteria:

- If an item showed C-level or CC-level differential item functioning (DIF), it was automatically coded yellow. Any items that showed this level of DIF were subject to an extra round of review (to determine if anything in the item could be detected that clearly indicates bias) prior to item selection (see Part 2, Section 2.2 for further detail). The CAL psychometrics team provided the Item Selection Panel with the report of the DIF review.
- Items were coded as green if they had infit and outfit values  $\leq 1.20$ . As outfit and infit values are sensitive to students' unexpected responses to items that are very easy for them, any item with a  $p$  value  $> 0.85$  was automatically coded as green, even if it had fit values outside of these thresholds.
- Items with infit and outfit values  $> 1.20$  and  $< 1.50$  were coded as yellow. As outfit values are also sensitive to students' unexpected responses to items that are very hard for them, items with  $p$  values close to chance (0.40 for a three-response item and 0.35 for a four-response item) were coded as yellow if outfit was  $> 1.20$  and  $< 1.75$ .
- Items that did not meet these criteria were coded as red.

The task of the Item Selection Panel in this first stage was to review all yellow folders and recode them as “green,” meaning “appropriate for operational use,” or “red,” meaning “not appropriate for operational use.” The panel reviewed the content of each yellow item, along with relevant statistics like those derived from distractor analyses (e.g., the mean ability of students selecting the key vs. the mean ability of students selecting the distractors; infit and outfit statistics for each response option, the point measure correlation of each response option, the percentage of students selecting each response option), to determine if the item would be reclassified as green or red. If all yellow items in a folder were reclassified as green (and there were no red items in the folder), the folder was deemed appropriate for operational testing.

In the next stage, the set of green folders, which the panel had deemed appropriate for operational use, became the pool of folders for item selection. The panelists selected folders (through a process of discussion and consensus building) with attention to the difficulty of each

item within a folder, the mean item difficulty of the items within a folder, and the content of a folder.

Table 10 and

Table 11 provide the numbers of continuing and new items per grade-level cluster for the Listening and Reading tests. For further detail on item statistics, including a summary of the number of items used as anchors across years, see Part 2 of this report, Sections 2.1 and 2.7.

Table 10. Number of New and Continuing Items on ACCESS Online, Series 503 Listening Test, by Grade-Level Cluster

Grade-Level Cluster	Number of New Items	Number of Continuing Items	Total Number of Items
1	6	48	54
2–3	12	42	54
4–5	9	45	54
6–8	9	45	54
9–12	12	42	54

Table 11. Number of New and Continuing Items on ACCESS Online, Series 503 Reading Test, by Grade-Level Cluster

Grade-Level Cluster	Number of New Items	Number of Continuing Items	Total Number of Items
1	18	54	72
2–3	21	51	72
4–5	9	63	72
6–8	15	57	72
9–12	15	57	72

In the domains of Writing and Speaking, the Task Selection Panel considered results from both qualitative and quantitative analyses of the students’ responses to the tasks. The CAL TD team reviewed student responses and DRC raters’ comments on each of the field-tested tasks. They then integrated those observations with task statistics, including fit statistics, raw score distributions, and rater agreement indices. Based on the information they compiled, the team made a recommendation to present to the panel for each task. If the panel needed to choose between two tasks, the team identified the task that was most appropriate to place on the operational test, based on the evidence they had compiled. Alternatively, the team could recommend that the slot remain unrefreshed. In cases where there was only a single task, the team determined whether the associated evidence was sufficient to support placing the task on the operational test or whether that slot should remain unrefreshed.

Although the CAL TD team considered rater agreement indices and fit statistics when making their recommendations, they based those recommendations primarily on the results from their analyses of the following: (1) the qualitative data (i.e., whether students could successfully score in the intended range, and/or whether DRC raters observed major anomalies that could indicate that a given task was not performing as intended), (2) the raw score distributions of students’ task performance, and (3) the task difficulty measures. The field-test tasks and the operational tasks that were tagged for refreshment should have comparable raw score distributions and task difficulty measures, they reasoned. The CAL TD team took this approach to ensure that the

vertical scale was maintained from year to year. The panel then reviewed each recommendation and associated evidence and either accepted or rejected the recommendation; recommendations were generally rejected if the task difficulty measures and the raw score distributions of the field-test tasks varied too much from those of the operational tasks.

Table 12 and Table 13 provide the numbers of continuing and new tasks, per grade-level cluster, for the Writing and Speaking tests. For further detail on task statistics, including a summary of the number of tasks used as anchors across years, see Part 2 of this report, Sections 2.1 and 2.7.

Table 12

Number of New and Continuing Tasks on ACCESS Online Series 503 Writing Test, by Grade-Level Cluster

Grade-Level Cluster	Tier	Number of New Tasks	Number of Continuing Tasks	Total Number of Tasks
1	A	1	1	2
1	B/C	1	1	2
2–3	A	1	1	2
2–3	B/C	1	1	2
4–5	A	1	1	2
4–5	B/C	1	1	2
6–8	A	1	1	2
6–8	B/C	1	1	2
9–12	A	1	1	2
9–12	B/C	0	2	2

Table 13

Number of New and Continuing Tasks on ACCESS Online Series 503 Speaking Test, by Grade-Level Cluster

Grade-Level Cluster	Tier	Number of New Tasks	Number of Continuing Tasks	Total Number of Tasks
1	Pre-A	1	2	3
1	A	2	4	6



1	B/C	2	4	6
2-3	Pre-A	2	1	3
2-3	A	4	2	6
2-3	B/C	4	2	6
4-5	Pre-A	2	1	3
4-5	A	4	2	6
4-5	B/C	4	2	6
6-8	Pre-A	2	1	3
6-8	A	4	2	6
6-8	B/C	4	2	6
9-12	Pre-A	2	1	3
9-12	A	4	2	6
9-12	B/C	4	2	6

## **3. Test Administration**

### **3.1 Test Delivery**

ACCESS Online is typically administered between December and April of the academic year, with testing windows determined at the state level. The Reading and Listening tests are administered first (in either order), followed by Writing and Speaking (in either order). The test may be administered in several sessions within a single day or over a series of days.

#### **3.1.1 Listening and Reading**

Listening and Reading are the first domains assessed. Students may take these in either order. Students sit at individual computer monitors and take the Listening and Reading tests online. They use headsets to listen to directions for the Listening and Reading tests, as well as listen to the Listening items. Students use the computer interface to select their answers. Once a student selects an answer and clicks the Next button, the answer is final, and the student is not permitted to go back and change an answer. The Listening and Reading tests are untimed, but approximate administration times are provided in the following sections.

#### **3.1.2 Writing**

Students in Grades 1–3 perform the Writing tasks on paper and handwrite their response.

Students in Grades 4–12 perform the Writing tasks online. A student may provide handwritten or keyboarded responses, with the choice dependent on a combination of local, state, and consortium-wide policies, as follows:

- Grades 4–5: A decision is made at the local or state level as to whether handwriting or keyboarding is the default response mode. In districts where keyboarding is the default, the option exists to use handwriting as an accommodation.
- Grades 6–12: Keyboarding is the default, with the option to use handwriting as an accommodation.

#### **3.1.3 Speaking**

Speaking tasks are delivered online. Students listen to prompts via headsets that are equipped with microphones to capture their responses. The student receives extensive support via illustrations and multimodal (text and audio) input designed to provide sufficient context for the

response, as well as a model student response that provides guidance on the level of linguistic complexity required to respond adequately (see Section 2.2.4).

## **3.2 Operational Administration**

Before, during, and after a state’s testing window, there are various roles that educators hold to ensure all tasks are carried out for successful test administration. These roles include Test Coordinators at the district and school level, Test Administrators, and, for online administration, Technology Coordinators. The Test Administrator administers and monitors 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* contain more information related to responsibilities and required training for the various roles. These manuals can be found on the WIDA Secure Portal (<https://portal.wida.us/>).

The training course within the WIDA Secure Portal (<https://portal.wida.us/>) where educators can access both training to become certified to administer ACCESS for ELLs as well as additional materials and resources to assist administrators and coordinators before, during, and after a state’s testing window. Training courses include test preparation and administration tutorials and online administration quiz.

It cannot be understated that the roles of test administrator and technology coordinator are critical for the proper administration of the assessments. Proper training and familiarity with ACCESS for ELLs administration requirements is key to the validity of the test and the appropriate interpretations of ACCESS for ELLs test scores.

### **3.2.1 Administering the Test Practice**

A test practice experience is provided to each student immediately prior to the administration of the an individual test domain. The test practice acclimates the student to the test interface and the types of items the student may experience in the test. The test practice takes approximately 5 to 10 minutes, depending on how many questions students have about the directions or practice items. Additional time should be scheduled for students to go through the test practice again if needed. The narration within the test practice is included both as spoken audio and as text captioning displayed directly on the screen, allowing the student to be able to read along as the script is read aloud.

The test practice for each domain and grade cluster are available as stand-alone materials on the WIDA website (<https://wida.wisc.edu/assess/access/preparing-students/practice>) to help educators prepare students to take ACCESS for ELLs. Before each domain test of ACCESS for ELLs, each student is required to take the test practice for that domain. No data are collected regarding the test practice; these items/tasks are presented to the students specifically to help ensure that they understand how to navigate the test interface.

### **3.2.2 Listening Test Administration**

The Listening test (including test practice items) is designed to take approximately 30 to 40 minutes. Students in all grades view the Listening prompts on the desktop, laptop, or tablet. Note that the approximate test administration time does not include convening students, taking attendance, or explaining test directions.

All Listening items are forced choices; in other words, students must respond to an item before they can proceed to the next item. In addition, once the students proceed to the next screen, they cannot return to any previous screens.

### **3.2.3 Reading Test Administration**

The Reading test (including directions and practice items) is designed to take approximately 35 minutes. Students in all grades view the Reading prompts on the desktop, laptop, or tablet. Note that the approximate test administration time does not include convening students, taking attendance, or explaining test directions.

All Reading items are forced choices; in other words, students must respond to an item before they can proceed to the next item. In addition, once the students proceed to the next screen, they cannot return to any previous screens.

### **3.2.4 Writing Test Administration**

All students in Grades 1–3 complete the ACCESS for ELLs Writing test on paper. The test is group administered. For Grades 6–12, all students view the Writing prompts on the desktop, laptop, or tablet. The default response mode is keyboarding. For Grades 4–5, all students also view the Writing prompts on the device. However, each state determines whether the default response mode for students in Grades 4–5 will be keyboarding or handwriting. If keyboarding is the default response mode, and upon logging in and starting the test a student expresses

discomfort, concern, or anxiety about keyboarding, administrators may switch the student to responding to the Writing test on paper. For Grades 6–12, all students view the Writing prompts on the desktop, laptop, or tablet. The default response mode is keyboarding.

The Writing test is designed to take approximately 45 to 60 minutes. For all grade-level clusters, the Tier B/C Writing tests have recommended timing guidelines for Parts A, B, and C of 10, 20, and 30 minutes, respectively. Note that the approximate test administration time does not include convening students, taking attendance, distributing, and collecting test materials, or explaining test directions, including the directions and practice that precede the test.

#### 3.2.4.1 Writing Test Tiers

Student performance on the Listening and Reading tests determines the appropriate tier that the student will take in the Writing and Speaking tests. For Grades 4–12, the test engine automatically routes students to the appropriate tier for Writing. For Grades 1–3 Writing, once the students have completed the Listening and Reading tests, Test Coordinators run a Tier Placement Report that identifies the Writing tier each student is assigned to take. This report is necessary for test administrators to know which tier Writing form to administer to which student, since the Writing test for Grades 1–3 is entirely paper based (see section 2.1.3 Writing Tasks for more information about the design of the Writing test). The Writing test has two tiers: A and B/C. In Grades 1–3, students must be tested in groups organized by grade-level cluster and tier.

### 3.2.5 Speaking Test Administration

The Speaking test (including directions and practice) is designed to take approximately 30 minutes. All students in grades complete the ACCESS for ELLs Speaking test on desktop, laptop, or tablet.

Recording response time on every task on the Speaking test has a preset time limit, which varies depending on the grade-level cluster, tier, and task level. Students learn about the time limits in the test directions and practice. Students see a circle change color and then disappear as the time to respond elapses. While there is a limit to how long students can take to record their response, students can navigate the directions, practice, and test items at their own pace. Students click the Next button when they are ready to move on from a screen, without time limits. The test does not advance automatically.

### *3.2.5.1 Speaking Test Tiers*

For each grade-level cluster, the Speaking test has three different tiered forms, Pre-A, A, and B/C. For all grade-level clusters, the tier the student takes is determined by the student's Listening and Reading test results; the test engine automatically routes students to the appropriate Speaking tier. The Pre-A tier is designed to address the needs of newcomer students and to allow those students at the beginning stages of English language development an opportunity to respond to tasks appropriate to what they can do. Tier Pre-A also includes a simplified version of the Speaking test practice to ease the burden of learning how to respond to Speaking tasks on the screen for newcomer students. Most students are placed in either Tier A or Tier B/C.

### *3.2.5.2 Group vs. Individual Delivery*

The Speaking test is administered to small groups of students. For students in all grade-level clusters taking the Tier A and Tier B/C forms, it is recommended that the Speaking test be administered to groups of three to five students.

It is recommended that students taking the Pre-A form be administered the test individually so Test Administrators can provide additional support during the test. For students in all tiers, the Speaking test may be administered individually or in smaller groups of students as mentioned above, if needed. Test administrators use their professional judgment to consider whether students with high test anxiety or students requiring extra support should be given the test individually or in a very small group.

## **3.2.6 Test Security**

Every effort is made 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 the printed tests) follow established policies and procedures regarding the security of the test, and every individual involved in the administration of ACCESS, from the district level to the classroom level, is trained in issues of test security.

All materials for ACCESS for ELLs are considered secure test materials. All users of the WIDA website are prompted to read and sign a Nondisclosure and User Agreement upon their first login. Use of the WIDA Assessment Management System and INSIGHT test engine are also subject to the terms of use outlined in the WIDA Assessment Management System. 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 (i.e., the Grades 1-3 Writing test booklets and any Grades 4-12 handwritten student response booklets) on the ACCESS for ELLs 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 ACCESS for ELLs District and School Test Coordinator Manual guidelines to maintain test security. Test security policies are stated in the Test Policy Handbook (<https://sea.wida.us/system/files/documents/SEA-support/test-policy-handbook.pdf>) and the Memorandum of Understanding (MOU)s with states.

### **3.3 Fairness and Accessibility**

The WIDA Accessibility and Accommodations Framework provides support for all ELLs, as well as targeted accommodations for students with individualized education plans (IEPs) or 504 plans. These supports are intended to increase the accessibility for the assessments for all ELLs. (Please see Accessibility and Accommodations Supplement for detailed information: <https://wida.wisc.edu/resources/accessibility-and-accommodations-supplement>). Fairness and accessibility are considered throughout the assessment process (i.e., test design, test development, item selection, forms creation, and test administration). For details, please refer to the universal design principles throughout test and item design to the *WIDA consortium English Language Proficiency Assessment for grades 1-12 Test and Item Design Plan ACCESS for ELLs Online Annual Summative Assessment and WIDA Screener Online*.

### 3.3.1 Support Provided to All ELLs

**Universal design.** ACCESS for ELLs incorporates universal design principles to provide greater accessibility for all ELLs. The test items are presented using multiple modalities, including supporting prompts with appropriate animations and graphics, embedded scaffolding, tasks broken into chunks, and modeling that uses task prototypes and guides. These aspects of universal design are built into CAL's item specifications and item review checklists, and CAL test development managers train the CAL language testing specialists on these principles of universal design through training on the use of the specifications and checklists.

**Administrative considerations** include adaptive and specialized equipment or furniture, alternative microphone, familiar Test Administrator, frequent or additional supervised breaks, individual or small group setting, monitoring of the placement of responses in the test booklet or on screen, participation in different testing formats (Paper vs Online), reading aloud to self, specific seating, short segments, verbal praise or tangible reinforcement for on-task or appropriate behavior, and verbal redirection of students' attention to the test (in English or native language).

**Universal tools** are available to all students taking ACCESS for ELLs to address their accessibility needs. These may either be embedded in the online test or provided by Test Administrators during testing. The universal tools provided on ACCESS for ELLs Online are described in Section 3.3.2 below. The Test Demo videos available on the WIDA website (<https://wida.wisc.edu/assess/access/preparing-students/practice>) instruct students how to use the universal tools. During testing, students choose whether to use the tools or not, but they are available to all students throughout testing.

### 3.3.2. Support Provided to ELLs with IEPs or 504 Plans

**Accommodations** include allowable changes to the test presentation, response method, timing, and setting in which assessments are administered. Accommodations are intended to provide testing conditions that do not result in changes in what the test measures; that provide test results comparable to those of students who do not receive accommodations; and that do not affect the validity and reliability of the interpretation of the scores for their intended purposes.

Accommodations are available only to ELLs with disabilities when listed in an approved IEP or 504 plan, and only when the student requires the accommodation(s) to participate in ACCESS



for ELLs meaningfully and appropriately. Accommodations are delivered locally by a Test Administrator. More information regarding accommodations is provided in the [ACCESS for ELLs Accessibility and Accommodations manual](#).

WIDA also offers Braille Test for ELLs and Large Print Test. The Braille test is paper based, and the translation and graphics are provided in either contracted or uncontracted Braille for Tier B (Grades 1–12). This test is used to provide access to the test for ELLs who are blind. The Large Print Test is used for students with visual impairments. The font size on the large print paper test is increased to 18 point. For the online test, the magnification/zoom tool increases the on-screen font size up to 1.5× or 2×, depending on the size of the computer monitor.

**Universal tools** are also available to all ELLs taking ACCESS for ELLs. Examples of universal tools include highlighter, line guide, magnification, and color overlay. All universal tools are available to all ELLs during testing; specific designation is not required prior to testing to make them available to the student during testing. The Test Demo videos available on the WIDA website (<https://wida.wisc.edu/assess/access/preparing-students/practice>) instruct students how to use the universal tools. During testing, students choose whether to use the tools or not, but they are available to all students throughout testing. Features available during online-based test administration include the following:

- Audio amplification device (provided by student)
- Highlight tool
- Line guide
- Zoom tool (magnifier)
- Sticky notes—which allow students to take notes to prepare responses to Writing items. This tool is only available in the Writing domain.
- Color overlay—which allows students to change the background color that appears behind text, graphics, and response areas. Five colors are available: pink, yellow, blue, green, and orange.
- Color contrast—which allows students to select from a variety of background/text color combinations
- Keyboard shortcuts/equivalents—which are alternatives to using a mouse (for navigating through the test and using online test tools)

- Scratch/blank paper (to be submitted with the test or disposed of according to state policy)

**Allowable test administration procedures** are variations in standard test administration procedures that provide flexibility to schools and districts in determining the conditions under which ACCESS for ELLs can be administered most effectively. These procedures are available to any student, as needed, at the discretion of the Test Coordinator (or principal or designee), provided that all security conditions and staffing requirements are met. Examples of allowable test administration procedures include tests administered by familiar school personnel, in an individual or small group setting, in a separate room, with frequent supervised breaks, or in short segments. For detailed information on the allowable test administration procedures, consult the *ACCESS for ELLs Test Administration Manual*.

Schools and districts should consider how accessibility features and allowable test administration procedures can support accessibility to the test for *all* ELLs. The accommodations, accessibility features, and allowable test administration procedures are based on (1) accepted practices in English language proficiency assessment; (2) existing accommodation policies of WIDA Consortium member states; (3) consultation with representatives of WIDA member states who are experts in the education and assessment of ELLs and students with disabilities; and (4) the expertise of the CAL test developers.

WIDA offers *Alternate ACCESS for ELLs*. This test is intended only for those ELLs who have cognitive disabilities that are so significant as to prevent meaningful participation in ACCESS testing, even with accommodations. The results of the Alternate ACCESS for ELLs operational administration appear in a separate technical report.

## **4. Scoring**

### ***4.1 Multiple Choice Scoring: Listening and Reading***

Listening and Reading items are scored dichotomously, as correct or incorrect. Scale scores for each domain are calculated based on the items administered to the student and the set of those items that the student answers correctly. For details on how scale scores for Listening and Reading are calculated, see Part 2, Chapter 2, “Analysis of Domains.”

### ***4.2 Scoring Performance-Based Tasks: Writing and Speaking***

Trained raters scored student responses to the performance-based tasks in the domains of Writing and Speaking. DRC retains many raters from year to year; the return rater rate was approximately 60% in 2021 and, overall, most raters scoring for ACCESS for ELLs were experienced DRC raters. DRC drew together this pool of experienced raters to staff the scoring pool for ACCESS for ELLs. To complete the rater staffing, DRC held recruiting events, after which DRC’s recruiting staff screened applications for rater positions, and DRC staff then personally interviewed likely candidates. As part of the hiring process, DRC required each candidate to provide an on-demand writing sample, an on-demand math sample, references, and proof of a 4-year college degree. In this screening process, DRC gave preference to candidates who had previous experience scoring students’ responses to tasks included in large-scale assessments and candidates with degrees in English language arts. The rater pool consisted of educators, writers, editors, and other professionals with content-specific backgrounds.

Prior to scoring live student responses, the raters underwent thorough training and qualifying. Training was task-specific to ensure that raters understood the nuances of each unique Writing or Speaking task. DRC selected team leaders based on their prior performance as raters and for their leadership skills and assigned them to small groups of raters; typically, there were 7 to 10 raters on each team. The team leaders were responsible for monitoring the performance of their team members and providing ongoing feedback to support accurate scoring. DRC promoted scoring directors, who earned their positions by demonstrating quality work as raters and as team leaders on previous projects, from within. Scoring directors were responsible for a specific set of tasks within a single domain. The scoring directors trained and oversaw the teams of raters assigned to these tasks. What follows are general scoring procedures that DRC utilized.

## Rater Training and Qualifying

- DRC assigned each rater a unique ID number and password.
- The scoring director conducted a team leader training session before training the raters. This session followed the same procedures as rater training but was more rigorous and in-depth due to the extra responsibilities required of team leaders. During team leader training, all WIDA materials were reviewed and discussed. To facilitate scoring consistency, it was imperative that all team leaders imparted the same rationale for each response. Once the team leaders were qualified, leadership responsibilities were reviewed, and team assignments were given.
- Rater training began with the scoring director going through the ACCESS for ELLs PowerPoint presentation provided by CAL. The PowerPoint gave scorers a good overview of ACCESS for ELLs and the WIDA scoring process.
- Rater training continued with the scoring director providing an intensive review of the ACCESS for ELLs Scoring Scale, the model student response for Speaking items, and task-specific anchor sets created by CAL. The anchor set contained a collection of student responses that were used to exemplify each possible score point. Each response included a scoring annotation that explained the scoring rationale. Scorers used the ACCESS for ELLs Scoring Scale, the model student response for Speaking, and the anchor sets as primary references during scoring.
- Next, raters practiced by independently scoring responses in training sets. Training sets were created by DRC scoring directors from responses approved by WIDA and CAL. The responses were selected to show raters the range of each score point (e.g., high, mid, and low 2s). This process helped raters recognize the various ways that a student could respond in order to earn each score point outlined and defined in the scoring guidelines. After each training set was taken, the scoring director led a thorough discussion of the responses.
- Once the scoring scale, anchor sets, and training sets were thoroughly discussed, each rater was required to demonstrate understanding of the scoring criteria by qualifying (i.e., scoring with acceptable agreement to the true scores) on at least one of the qualifying sets. Raters who failed to achieve at least 70 percent exact agreement on the first qualifying set were given additional training, either individually or in a small group

setting. Raters who did not perform at the required level of agreement by the end of the qualifying process were not allowed to score any student responses. These individuals were removed from the pool of potential raters in DRC's imaging system and released from the project. Qualifying sets were created by DRC scoring directors from responses approved by WIDA and CAL.

- Throughout training, the scoring director provided detailed directions for use of DRC's computerized scoring system and remote communication tools for raters.
- Once raters were trained, qualified, and began live scoring, DRC used recalibration sets and validity responses to keep the raters calibrated on the tasks they were scoring. Recalibration sets were pre-scored sets of responses that were approved by WIDA and CAL and were used to help refocus raters on WIDA scoring guidelines. Validity responses were also approved by WIDA and CAL and were responses that were pre-scored and used to ensure raters were adhering to WIDA scoring criteria. Recalibration and validity are explained in greater detail below.

#### Calculating Score Agreement for Score Monitoring

- DRC's handscoring system generated handscoring reports, detailing agreement rates for each rater and task. The reports were automatically generated overnight throughout the course of handscoring and could also be run on demand. DRC provided weekly interrater reliability reports to WIDA throughout the handscoring process to ensure that DRC maintained sufficient quality control throughout the course of scoring.
- For Writing, DRC defines **agreement** as two adjacent scores, reported as %AG. (See Section 4.3 for a description of the Writing Scoring Scale.) For example, using the Writing Scoring Scale, DRC considers scores of 2 and 2+ as agreement, as well as scores of 2 and 2 or scores of 2+ and 3. However, DRC considers scores of 2 and 3 on the Writing Scoring Scale as **adjacent**, while considering scores of 2 and 3+ as **nonadjacent**.
- For Speaking, DRC defines **agreement** as two scores that are exactly the same, reported as %EX. (See Section 4.4 for a description of the Speaking Scoring Scale.) Unlike in Writing, where DRC considers two adjacent scores as "Agreement," raters scoring

responses to Speaking tasks must demonstrate Exact Agreement (EX) in order to be considered in “agreement.”

- WIDA stipulates a minimum interrater agreement rate of 70% for both Writing and Speaking.

#### Routing Responses to Ensure “Blind” Second Scores

- The DRC scoring system routed and rerouted responses to raters until raters were assigned the prescribed number of scores for all responses. All responses were scored once, and at least twenty percent of the responses were scored a second time. The responses that were used for the twenty percent read- and listen-behinds were randomly chosen by the imaging system at the item level. Additional read- and listen-behinds by the team leaders and scoring directors were done to further ensure reliability. Raters did not see the scores the other raters assigned, and they did not know if they were the first or second rater.
- The purpose of the first and second scores was to monitor interrater reliability by comparing the scores that two separate raters assigned to the same response. When calculating final scores, the first score assigned was the score of record.

#### Monitoring Scoring (Quality Control)

- Rater accuracy was monitored throughout the scoring session by means of daily and on-demand reports. These reports ensured that an acceptable level of scoring accuracy was maintained throughout the project. Interrater reliability was tracked and monitored with multiple quality control reports. These reports and other quality control documents were generated at the scoring centers, where they were reviewed by the scoring directors, team leaders, and project managers. DRC provided WIDA with access to these reports on a regular basis throughout the scoring process to provide assurance that the quality control metrics met or exceeded expectations. If a scorer did not meet scoring expectations a portion of, or all, their scores could be dropped if the scores had not been reported.
- During the handscoring process, the scoring directors communicated regularly with their team leaders to review the statistics generated from the previous day’s

work, including interrater reliability, score point distributions, and validity reports.

- Throughout handscoring, team leaders conducted routine read- and listen-behinds to observe, in real time, raters' performance. Team leaders utilized live, scored responses to provide ongoing feedback and, if necessary, retraining for raters.
- The DRC system generated interrater reliability reports daily to monitor how often each rater's scores matched other raters' scores, and scoring leaders continually monitored individual rater statistics, comparing them to the group average. If the agreement rate for a rater fell below 70%, supervisors increased monitoring and retraining activities with the rater. If the rater failed to demonstrate improved reliability, DRC released the rater from scoring responses to that task.
- Since the interrater agreement rates were all at or above 70%, the target that WIDA stipulated, the focus turned to raters with lower-than-average agreement rates—even if their agreement rate was at or above 70%. Even when all agreement rates were at or above 70%, scoring supervisors continued to seek opportunities to increase reliability by providing ongoing feedback and retraining raters based on the specific performance of each rater, as evidenced by the quality control reports and observations made when reviewing scores that a rater assigned.
- DRC can retrieve students' responses on demand (e.g., specific grade-level clusters, specific students) should the need arise during or after the scoring process.
- If needed, DRC can re-score a student's response to a task based on task- or response-level information, such as task number, date, score assigned, or rater ID.
- For both Speaking and Writing, DRC used both recalibration sets and validity responses to monitor handscoring quality control. DRC, CAL, and WIDA collaborated to develop these recalibration sets and validity responses. CAL developed an initial pool of responses for use as recalibration and validity checks by selecting responses from a previous administration of the tasks (e.g., a field test). WIDA staff reviewed and approved this pool of responses and their scores.

DRC supervisors supplemented this pool of responses as needed by selecting additional responses, which CAL and WIDA approved before use. For each of the first 5 days that raters scored student responses to a task, they scored one recalibration set of five responses. The recalibration sets did not differ from rater to rater. For example, DRC identified a recalibration set to use for the first day that a rater scored students' responses to a specific task; every rater who was working on that task took this same recalibration set on the first day that they worked on that task. After the raters assigned scores to the recalibration set, the scoring director or team leader reviewed the set using descriptors from the scoring scale and the anchor responses to confirm the rationale behind each response's score. Starting on the sixth day that a rater was working on a task, DRC used validity responses to continue monitoring rater performance. DRC seeded the validity responses into the operational scoring so that the raters did not know which responses were operational and which were validity responses. Reports generated daily compared the scores that each rater assigned to the "true" score for each validity response. When a rater was working on a task, DRC seeded the validity responses in random order into the rater's queue for scoring. Given enough time, every rater working on a task would score every validity response for that task, but the order in which the raters would see the validity responses would differ.

### Handling Unusual Responses

The following processes were in place at DRC to manage specific types of "unusual" responses:

- **Scoring questions.** If a rater had questions about the application of the scoring guidelines to a response (e.g., if they were uncertain as to the proper score that they should assign), the rater forwarded the response to their team leader for assistance. The team leader then reviewed the response with the rater and assigned the proper score. If the rater needed further clarifications, the team leader worked with the rater to review scoring guidelines.



- **Nonscore codes.** Unusual or aberrant responses for which raters could not assign a score based on the scoring guidelines received a nonscorable code (e.g., Writing responses that are entirely blank or consist entirely of scribbles or pictures). DRC’s handscoring team collaborated with WIDA and CAL to define what specifically constituted a nonscorable response to ensure consistency when applying nonscorable codes, and CAL provided this information to DRC along with other task-specific training materials that DRC then used to train its raters. During scoring, when raters assigned a nonscorable code (except for Blank), DRC’s imaging system automatically forwarded the response to a handscoring supervisor for review and approval. If the handscoring supervisor had any questions about the application of non-score codes to specific responses, the supervisor contacted WIDA and CAL representatives for further review and discussion.
- **Alerts.** To handle possible alert responses (i.e., student responses indicating potential issues related to the student’s safety and/or well-being that may require attention at the local level, as well as potential plagiarism and potential teacher interference), DRC’s imaging system gave raters the ability to alert questionable student responses. When a rater flagged a response with the alert status, the imaging system automatically routed the response to handscoring supervisors for review. The states are notified within 24 hours. If the response was related to the student’s safety and/or well-being, and the handscoring supervisors concurred with the alert, it was then forwarded to WIDA’s project management team who provided the response to the appropriate local education agency.
- **Request for originals.** When a rater came across a scanned student response that was difficult to read (for example, having some partially erased text), the rater flagged the response with a “request original” status. If a rater flagged a response as “request original,” DRC’s imaging system automatically forwarded the response to a handscoring supervisor. If the handscoring supervisor agreed that the original student response needed to be reviewed to properly apply the scoring guidelines, the supervisor forwarded the request to staff in DRC’s Operations

Services, who located the original student response so the handscoring supervisor could review the response and score it.

#### Remote Scoring Procedures due to the COVID-19 Pandemic

Prior to 2020, DRC's handscoring centers managed all WIDA handscoring. In 2020, due to the COVID-19 pandemic, DRC shifted from site-based handscoring to remote handscoring to continue meeting all the handscoring deadlines. All WIDA handscoring continued to be remote in 2021. DRC designed the remote scoring to very closely emulate the work carried out in the physical scoring locations. The platform, content, and expectations for quality remained the same. Using a variety of modes of interactive technology (i.e., web screen sharing, webcast, video chat, and chat), DRC conducted rater training and discussions live (virtually). DRC equipped scoring leaders with a variety of tools to ensure that every rater was successful in understanding and applying scoring criteria to student responses.

Remote scoring began with a training session to guide supervisors and raters using the tools that DRC utilized for remote scoring. Once supervisors and raters were trained on the remote scoring process, handscoring commenced for the ACCESS assessments. A description of DRC's remote scoring process follows.

- **System tools—scoring, training, chat.** ScoreBoard is DRC's secure, web-based scoring application that is designed to be used in a distributed environment. The platform is used within DRC's scoring centers and in remote locations (e.g., in a rater's home). Integrated training resources provide the capability to securely maintain digital training materials within the scoring platform itself.
- DRC conducted live, interactive rater training using the Moodle Learning Management System, which mirrored aspects of the scoring room and provided a versatile platform for training. It also served as a place to share files of important documents, including daily scoring statistics and platform user guides. Through embedded communication tools, scoring directors, assistant scoring directors, and team leaders facilitated group and one-on-one training sessions and discussions using audio and video.

- To facilitate instant communication between supervisors and raters, DRC utilized a chat tool called Zulip in conjunction with ScoreBoard and Moodle. Zulip provided a tool for raters to directly ask supervisors questions about responses and allowed supervisors to direct individuals or groups of raters to join Moodle training rooms for important discussions and retraining.
- **Security.** Security is essential to the handscoring process. When users logged into ScoreBoard, they were required to read and accept the security policy before they were allowed to access the project. DRC also required raters to read and sign nondisclosure agreements. During training and large-group discussions, trainers continuously emphasized what security means, the importance of maintaining security, and how all staff accomplish this. In the remote environment, DRC could give these security reminders daily. DRC requires raters working remotely to work in a private environment away from other people (including family members). Raters working in ScoreBoard were not allowed to print from their computers in order to protect the security of the student responses, test questions, and training materials. Restrictions built into ScoreBoard defined the hours during the day that raters were able to log into the system, ensuring that raters were only scoring responses while supervisors were in place to monitor handscoring and answer any questions.
- **Rater training with Moodle.** DRC conducted rater training remotely as an interactive, comprehensive, hands-on experience. For Writing training, scoring directors trained groups of raters by screensharing PDFs of training materials. Raters individually viewed each training example, with supervisors directing raters to relevant text.
- For Speaking training, scoring directors trained groups of raters by playing the responses aloud over Moodle during live, remote training sessions.
- As with site-based training sessions, supervisors guided the discussion, and raters posed questions to supervisors. The scoring director directed the team leaders and raters to take training and qualifying sets, following the same training flow as they would in the scoring facility.

- **Quality control.** DRC utilized its robust quality control processes and handscoring metrics for all scoring sessions. Scored responses were monitored with second reads, and team leaders conducted read- and listen-behinds. DRC's handscoring system allowed scoring supervisors to determine specific read- and listen-behind rates (frequency of monitoring) for each rater. Any retraining and/or conversations needed because of the monitoring were held in one-on-one video chat sessions. Handscoring quality reports were available daily and on demand for handscoring supervisors and DRC's project leadership, and DRC also provided WIDA staffing with handscoring reports. If a rater fell below 70% exact agreement and failed to improve after retraining and feedback, DRC removed the rater from the project and assigned the responses to other raters to score.

### **4.3 Writing Scoring Scale**

The Writing Scoring Scale has six whole score points that range from 1 to 6. The scale descriptors include three different yet interrelated dimensions: discourse, sentence, and word/phrase. These scale descriptors guide raters as they consider all three dimensions to make holistic judgments about which score point best suits a response. The dimensions are distinguished as follows:

- The descriptors for the discourse dimension focus on the degree of organization and the extent to which the response is tailored to the context (e.g., purpose, situation, and audience).
- The descriptors for the sentence dimension evaluate the complexity and grammatical accuracy of sentence structures used in the response.
- The descriptors for the word/phrase dimension specify the range and appropriateness of the original vocabulary used (i.e., text other than that copied and adapted from the stimulus and prompt).

Figure 20 shows the Writing Scoring Scale.

<b>ACCESS for ELLS 2.0 Writing Scoring Scale, Grades 1–12</b>		
<b>5+</b>	<b>Score Point 6</b>	<p>D: Sophisticated organization of text that clearly demonstrates an overall sense of unity throughout, tailored to context (e.g., purpose, situation, and audience)</p> <p>S: Purposeful use of a variety of sentence structures that are essentially error-free</p> <p>W: Precise use of vocabulary with just the right word in just the right place</p>
	<b>Score Point 5</b>	<p>D: Strong organization of text that supports an overall sense of unity, appropriate to context (e.g., purpose, situation, and audience)</p> <p>S: A variety of sentence structures with very few grammatical errors</p> <p>W: A wide range of vocabulary, used appropriately and with ease</p>
	<b>Score Point 4</b>	<p>D: Organized text that presents a clear progression of ideas, demonstrating an awareness of context (e.g., purpose, situation, and audience)</p> <p>S: Complex and some simple sentence structures, containing occasional grammatical errors that don't generally interfere with comprehensibility</p> <p>W: A variety of vocabulary beyond the stimulus and prompt, generally conveying the intended meaning</p>
<b>3+</b>	<b>Score Point 3</b>	<p>D: Text that shows developing organization including the use of elaboration and detail, though the progression of ideas may not always be clear</p> <p>S: Simple and some complex sentence structures, whose meaning may be obscured by noticeable grammatical errors</p> <p>W: Some vocabulary beyond the stimulus and prompt, although usage is noticeably awkward at times</p>
	<b>Score Point 2</b>	<p>D: Text that shows emerging organization of ideas but with heavy dependence on the stimulus and prompt and/or resembles a list of simple sentences (which may be linked by simple connectors)</p> <p>S: Simple sentence structures; meaning is frequently obscured by noticeable grammatical errors when attempting beyond simple sentences</p> <p>W: Vocabulary primarily drawn from the stimulus and prompt</p>
	<b>Score Point 1</b>	<p>D: Minimal text that represents an idea or ideas</p> <p>S: Primarily words, chunks of language, and short phrases rather than complete sentences</p> <p>W: Distinguishable English words that are often limited to high frequency words or reformulated expressions from the stimulus and prompt</p>
<p><i>D: Discourse Level      S: Sentence Level      W: Word/Phrase Level</i></p>		

**Figure 20. Writing Scoring Scale**

When assigning a score, a rater makes an initial judgment about which whole score point (1–6) best describes a response and then determines whether the three descriptors for that whole score point suit that response. If all three descriptors suit the response, the rater assigns the score associated with that score point (e.g., if all three descriptors for score point 3 are appropriate, the rater would assign a score of 3). However, if there is clear evidence that one or two descriptors from an adjacent score point are a better fit, the rater would assign a plus score between the two applicable whole score points (e.g., if two descriptors for score point 3 seem to fit, but one descriptor for score point 4 is a better fit than the associated descriptor for score point 3, the rater would assign a score of 3+).

In addition to scale descriptors, scoring rules address special cases where responses are nonscorable, completely or partially off task, and completely or partially off topic. These are defined as follows:

**Nonscorable:** The response is blank; consists only of verbatim copied text; consists only of text that is completely off task; or is entirely in a language other than English; or appears to have been plagiarized from an outside source during testing. More information on how plagiarized responses are handled by DRC is provided in Section 4.2 Scoring Performance-Based Tasks: Writing and Speaking., *Handling Unusual Responses, Alerts* above.

**Completely off-task response:** The entire response shows no understanding of or interaction with the prompt. It may be a memorized, previously practiced response or appear to answer another, unrelated prompt. A response that is entirely off task is nonscorable.

**Completely off-topic response:** The entire response shows a misinterpretation or misunderstanding of the prompt. An off-topic response is related to the prompt but does not seem to address it as intended. However, the response is clearly not a memorized, previously practiced response. Raters score these responses in their entirety using the scoring scale; however, the maximum score for a completely off-topic response is 2+.

**Partially off-task response:** The response contains both off-task and on-task writing. Raters score these responses by ignoring the off-task portion (which may be memorized and previously practiced) and scoring only the on-task portion using the scoring scale.

**Partially off-topic response:** The response contains both off-topic and on-topic writing (i.e., a portion of the response shows a misinterpretation or misunderstanding of the prompt). Raters score these responses in their entirety using the scoring scale.

Each student responds to two Writing tasks. One rater assigns a score to each student's response for each task. To calculate a student's total raw score by task, the scores that the raters assigned are converted to whole numbers ranging from 0 to 9, as shown in Table 14. The Writing scoring scale was designed to go up to score point 6. However, we did not have enough responses to estimate the rating scale parameters at each of the 5, 5+, and 6 score points in the empirical data. Therefore, these score points were collapsed into one category for psychometric purposes. Students' scores are then added across tasks, resulting in a total raw score that ranges from 0 to 18.

Table 14

**Rating to Raw Score Conversion (Writing)**

Rating	Raw score
Nonscorable	0
1	1
1+	2
2	3
2+	4
3	5
3+	6
4	7
4+	8
5	9
5+	9
6	9

The ACCESS Writing Scoring Scale is distinct from the WIDA Writing Rubric, which is a tool for evaluating student writing in classrooms and for interpreting student scores from ACCESS Online. CAL and WIDA designed the ACCESS Writing Scoring Scale for trained raters to use to evaluate students' responses to ACCESS writing tasks; thus, it is not appropriate for any other purposes.

#### **4.4 Speaking Scoring Scale**

The Speaking Scoring Scale defines five score points: *Exemplary*, *Strong*, *Adequate*, *Attempted*, and *No Response*. The *No Response* score point applies only if the rater uses one of three

nonscorable codes: R = dead air or white noise; F = foreign language response; I = nonscorable utterance; K = suspected plagiarism. A nonscorable utterance is defined as one of the following:

- The quality of the audio recording is too poor for any words to be understood. It may be too garbled or too quiet.
- The response contains sounds but no words in English (e.g., *hmmm, la la la, blah blah blah*).
- The response consists only of a teacher giving instruction or some other overlaying sound (from another student, PA system, etc.).
- The rater believes that the response may have been plagiarized. More information on how plagiarized responses are handled by DRC is provided in Section 4.2 Scoring Performance-Based Tasks: Writing and Speaking, *Handling Unusual Responses, Alerts* above.

Raters assign scores based on the proficiency level expectations of each task, that is, the level of language proficiency that each task is designed to elicit. The model student response exemplifies these expectations (see Section 2.2.4). In this way, the model response serves as a scoring benchmark. Raters listen to the model response and then score student responses relative to the model. A score of 4 (*Exemplary*) means that the student response demonstrates English language use that is equal to or beyond the English language use that the model student response illustrates.

Figure 21 shows the Speaking Scoring Scale.



<b>ACCESS for ELLs 2.0 Speaking Scoring Scale</b>	
<b>Score point</b>	<b>Response characteristics</b>
<b>Exemplary</b> use of oral language to provide an elaborated response	<ul style="list-style-type: none"> <li>• Language use comparable to or going beyond the model in sophistication</li> <li>• Clear, automatic, and fluent delivery</li> <li>• Precise and appropriate word choice</li> </ul>
<b>Strong</b> use of oral language to provide a detailed response	<ul style="list-style-type: none"> <li>• Language use approaching that of model in sophistication, though not as rich</li> <li>• Clear delivery</li> <li>• Appropriate word choice</li> </ul>
<b>Adequate</b> use of oral language to provide a satisfactory response	<ul style="list-style-type: none"> <li>• Language use not as sophisticated as that of model</li> <li>• Generally comprehensible use of oral language</li> <li>• Adequate word choice</li> </ul>
<b>Attempted</b> use of oral language to provide a response in English	<ul style="list-style-type: none"> <li>• Language use does not support an adequate response</li> <li>• Comprehensibility may be compromised</li> <li>• Word choice may not be fully adequate</li> </ul>
<b>No response (in English)</b>	<ul style="list-style-type: none"> <li>• Does not respond (in English)</li> </ul>

**Figure 21. Speaking Scoring Scale**

The Speaking Scoring Scale includes descriptors for overall language use, response sophistication, language delivery, and word choice.

Each student responds to three (or six) Speaking tasks, depending upon how the test engine routes the student. A single rater assigns a score to each of those responses, as shown in Table 15. To calculate a total raw score, the scores are then summed, based on the following guidelines:

- For tasks targeting language elicitation at PL 1, there are only three possible score points: *No Response*, *Attempted*, and *Adequate and Above*. This is the case because appropriate responses to PL 1 tasks are single words and short chunks of language, so it is not possible to reliably distinguish between *Adequate*, *Strong*, and *Exemplary* performances.
- For tasks targeting language elicitation at PL 3 and PL 5, each task can be scored on the entire breadth of the scale.
- Each student routed to Tier Pre-A responds to three PL 1 Speaking tasks. Thus, for students in this tier, the total raw score can range from 0 to 6.

- Students routed to Tier A respond to six Speaking tasks, three at PL 1 and three at PL 3. For students in this tier, the total raw score can range from 0 to 18.
- When scoring students' responses to Speaking tasks included in Tier B/C, six points are added to the total raw score, representing a score of *Adequate and Above* for three tasks targeting language at PL 1. Though a Tier B/C student would not be administered any tasks targeting the PL 1 level, it is assumed that a student who had been routed to Tier B/C would easily achieve a score of *Adequate and Above* on these tasks. Thus, for a student routed to Tier B/C, the total raw score can range from 6 to 30.

Table 15

**Score to Raw Score Conversion (Speaking)**

Score	Raw score
No Response (R, F, I, or K)*	0
Attempted	1
Adequate/Adequate and Above	2
Strong	3
Exemplary	4

\*R = Dead air or white noise; F = Foreign language response; I = Nonscorable utterance; K = suspected plagiarism.

DRC trained raters evaluate students' responses to the Speaking tasks using the ACCESS Speaking Scoring Scale. The Speaking Scoring Scale is distinct from the WIDA Speaking Rubric, which is a tool for classroom use and score interpretation. CAL and WIDA designed the ACCESS Speaking Scoring Scale for raters to use to evaluate students' responses to ACCESS speaking tasks; thus, it is not intended to be used for classroom assessment purposes.

## 5. Summary of Score Reports

### 5.1 Individual Student Report

Score reports (district, school, and student level reports) are made available in WIDA Assessment Management System (AMS) as soon as they are available for each state, and WIDA ships printed reports to school districts and schools at the same time or shortly thereafter. Score reports are available for states to use to identify students' language performance and properly determine language support for ELLs. Each state and school district determines when and how students' parents or guardians will receive individual score reports. WIDA provides resources that schools, districts and states may use to aid in score interpretation. (See links below.) How these stakeholders use the material to communicate assessments results is determined locally.

Individual student reports are available in various languages in WIDA AMS, and alternate formats (i.e., Braille or large print) of score reports are available upon request.

WIDA offers several online resources to help communicate test score information to educators, families, and students. (See ACCESS for ELLs Score and Reports

<https://wida.wisc.edu/assess/access/scores-reports>; Family Engagement

<https://wida.wisc.edu/teach/learners/engagement>.) WIDA also provides a post-testing Q & A webinar about score interpretation (<https://portal.wida.us/webinar/detail/702b69ef-0265-eb11-a2dd-0050568beee8>).

According to Kim et al. (2016, 2020), educators find interpreting technical information supplied in score reports to be challenging, which suggests a need for more clarity when describing student performance. WIDA plans to convene focus groups to gain an understanding of how various test users (i.e., educators, parents/guardians, students) interpret the information conveyed in current score reports in order to guide efforts to revise those reports for greater clarity.

The Individual Student Report (Figure 22) contains detailed information about the performance of a single student in Grades K–12. Its primary users are students, parents/guardians, teachers, and school teams. It provides information about one indicator of a student's English language proficiency: the language needed to access content and succeed in school.



**Sample Student**

Birth Date: mm/dd/yyyy | Grade: sample grade

Tier: sample tier

District ID: XXXXXXXXXXXXXXXX | State ID: XXXXXXXXXXXXXXXX

School: sample school

District: sample district

State: sample state

## Individual Student Report 20XX

This report provides information about the student's scores on the ACCESS for ELLs 2.0 English language proficiency test. This test is based on the WIDA English Language Development Standards and is used to measure students' progress in learning English. Scores are reported as Language Proficiency Levels and as Scale Scores.

Language Domain	Proficiency Level (Possible 1.0-6.0)						Scale Score (Possible 100-600) and Confidence Band See Interpretive Guide for Score Reports for definitions					
	1	2	3	4	5	6	100	200	300	400	500	600
<b>Listening</b>				4.0						368		
<b>Speaking</b>			2.2							320		
<b>Reading</b>				3.4						356		
<b>Writing</b>				3.5						355		
<b>Oral Language</b> 50% Listening + 50% Speaking				3.2						344		
<b>Literacy</b> 50% Reading + 50% Writing				3.5						356		
<b>Comprehension</b> 70% Reading + 30% Listening				3.7						360		
<b>Overall*</b> 35% Reading + 35% Writing + 15% Listening + 15% Speaking				3.4						352		

\*Overall score is calculated only when all four domains have been assessed. NA: Not available

Domain	Proficiency Level	Students at this level generally can...	
<b>Listening</b>	<b>4</b>	understand oral language in English related to specific topics in school and can participate in class discussions, for example: <ul style="list-style-type: none"><li>• Exchange information and ideas with others</li><li>• Connect people and events based on oral information</li></ul>	<ul style="list-style-type: none"><li>• Apply key information about processes or concepts presented orally</li><li>• Identify positions or points of view on issues in oral discussions</li></ul>
<b>Speaking</b>	<b>2</b>	communicate ideas and information orally in English using language that contains short sentences and everyday words and phrases, for example: <ul style="list-style-type: none"><li>• Share about what, when, or where something happened</li><li>• Compare objects, people, pictures, events</li></ul>	<ul style="list-style-type: none"><li>• Describe steps in cycles or processes</li><li>• Express opinions</li></ul>
<b>Reading</b>	<b>3</b>	understand written language related to common topics in school and can participate in class discussions, for example: <ul style="list-style-type: none"><li>• Classify main ideas and examples in written information</li><li>• Identify main information that tells who, what, when or where something happened</li></ul>	<ul style="list-style-type: none"><li>• Identify steps in written processes and procedures</li><li>• Recognize language related to claims and supporting evidence</li></ul>
<b>Writing</b>	<b>3</b>	communicate in writing in English using language related to common topics in school, for example: <ul style="list-style-type: none"><li>• Describe familiar issues and events</li><li>• Create stories or short narratives</li></ul>	<ul style="list-style-type: none"><li>• Describe processes and procedures with some details</li><li>• Give opinions with reasons in a few short sentences</li></ul>

[OBJ]

**Figure 22. Individual Student Report**

The Individual Student Report includes four language domain scores (Listening, Speaking, Reading, and Writing) and four language domain composite scores (Oral Language, Literacy, Comprehension, and Overall), as shown in the first table of the score report. In the first column of the last four rows of that table, test users can see how WIDA uses a student's domain scores to

calculate each composite score (e.g., for Oral Language, WIDA calculates the composite score based on a student's performance on the Listening and Speaking tests, with scores on each of those tests contributing equally to the composite score). For students who are unable to complete all four domains due to their disabilities, WIDA provides states methods to compute alternative composite scores based on their available domain scores upon request (Sahakyan,2020).

The proficiency level that a student attained in each language domain is presented both graphically and as a whole number followed by a decimal. These are interpretive scores that are based on, but separate from, scale scores. The shaded bar of the graph describes a student's performance in terms of the 6-level English Language Proficiency Scale. The whole number indicates a student's English language proficiency level (1–Entering, 2–Emerging, 3–Developing, 4–Expanding, 5–Bridging, and 6–Reaching) in accordance with the WIDA ELD Standards. ELLs who attain Level 6, Reaching, have moved through the entire second language continuum, as defined by the test and the WIDA ELD Standards.

The decimal indicates the proportion within the proficiency level range that the student's scale score represents, rounded to the nearest tenth. For example, a proficiency level score of 3.5 is halfway between English language proficiency levels 3.0 and 4.0.

To the right of the proficiency level is the reported scale score and the associated confidence band for each domain and composite. A scale score represents a student's performance that has been put on a standardized scale. Students' performance relies on the number of items and item difficulties they respond to correctly. Scale scores allow comparison across different forms and grades. In ACCESS, the scale score ranges between 100-600 in all grades. The confidence band reflects the standard error of measurement for the scale score, a statistical calculation of a student's likelihood of scoring within a particular range of scores if he or she were to take the same test repeatedly without any change in ability. For ACCESS scale scores, the confidence band reflects a 95% probability level.

The second table in the Individual Student Report provides information about the student's proficiency levels expressed as whole numbers. The third column of the table describes what that student should generally be able to do in each of the four language domains, given his or her level of proficiency. For example, as shown in Figure 22, this student received a proficiency level score of 2 for Speaking, which suggests that the student should generally be able to

“communicate ideas and information orally in English using language that contains short sentences and everyday words and phrases.”

If a student was not tested in one (or more) of the language domains, a code of NA (Not Available) will appear in the score report for the impacted language domain(s) and for all composite scores that are calculated using those domain scores. For these students, WIDA provides states with information about statistical methods that can be used to compute alternative composite scores based on a student’s available domain scores (Sahakyan, 2020).

When interpreting scores, test users are cautioned to keep in mind these points:

- The report provides information on English proficiency. It does not provide information on a student’s academic achievement or knowledge of content areas.
- Students do not typically acquire proficiency in Listening, Speaking, Reading, and Writing at the same pace. Generally,
  - Oral language (L+S) is acquired faster than literacy (R+W).
  - Receptive language (L+R) is acquired faster than productive language (S+W).
  - Writing is usually the last domain to be mastered.
- The students’ foundation in their home or primary language is a predictor of their English language development. Those who have strong literacy backgrounds in their native language will most likely acquire literacy in English at a quicker pace than students who do not.
- The Overall score is helpful as a summary of other scores and is used because a single number may be needed for reference. However, it is important to remember that it is compensatory, averaged using weights; a particularly high score in one domain may effectively offset a low score in another domain and vice versa. Similar Overall scores can mask very different performances on individual tests.
- No single scale score or language proficiency level, including the Overall score (composite), should be used as the sole determiner for making decisions regarding a student’s English language proficiency. School work and local assessment throughout the school year also provide evidence of a student’s English language development.
- Scale scores can be used to make comparisons across grade levels, but not across domains. Each domain has its own score scale, so scale scores should not be used for

comparing performance across domains. For example, a scale score of 350 in Listening at grade 3 is not equivalent to a scale score of 350 in Speaking at grade 3. For performance comparisons across domains, proficiency levels should be used.

- Either scale scores or proficiency levels can be used to compare test performance from different years, although it is easier to see changes when examining scale scores.

For detailed information about score reports, please refer to the Interpretive Guide.

## **5.2 Other Reports**

**Student Roster Report.** The Student Roster Report contains information on a group of students within a single school and grade. It provides scale scores for individual students in each language domain and composite scores, identical to those appearing in the Individual Student Report. Its intended users are teachers, program coordinators/directors, and administrators.

**Frequency Reports.** The primary audiences for frequency reports are typically program coordinators/directors, administrators, and boards of education. There are three types of frequency reports:

- School Frequency Report
- District Frequency Report
- State Frequency Report

Each shows the number and percentage of tested students who attained each proficiency level within a given population.

**Part 2:**  
**Technical Results**



## Contents

1. Annual Test Results.....	1
1.1 Participation .....	1
1.1.1 Grade-Level Cluster.....	1
1.1.2 Grade .....	4
1.1.3 Tier.....	10
1.2 Scale Score Results .....	16
1.2.1 Mean Scale Score Across Domain and Composite Score by Cluster.....	16
1.2.2 Mean Scale Score Across Domain and Composite Score by Grade .....	21
1.2.3 Correlations.....	31
1.3 Proficiency Level Results .....	36
1.3.1 Domains.....	36
1.3.2 Composites .....	59
2. Analysis of Domains.....	80
2.1 Complete Item or Task Analysis and Summary.....	84
2.2 DIF Analysis and Summary .....	88
2.3 Raw Score Distribution .....	92
2.3.1 Listening.....	93
2.3.2 Reading.....	106
2.3.3 Writing.....	119
2.3.4 Speaking .....	132
2.4 Scale Score Distribution .....	145
2.4.1 Listening.....	146
2.4.2 Reading.....	165
2.4.3 Writing.....	184
2.4.4 Speaking .....	203
2.5 Proficiency Level Distribution .....	222
2.5.1 Listening.....	223
2.5.2 Reading.....	242
2.5.3 Writing .....	261

2.5.4	Speaking .....	280
2.6	Raw Score to Scale Score to Proficiency Level Conversion .....	299
2.6.1	Listening .....	300
2.6.2	Reading .....	312
2.6.3	Writing .....	325
2.6.4	Speaking .....	340
2.7	Equating and Recalibration Summary .....	353
2.8	Test Characteristic Curve .....	357
2.8.1	Listening .....	358
2.8.2	Reading .....	365
2.8.3	Writing .....	372
2.8.4	Speaking .....	379
2.9	Test Information Function .....	385
2.9.1	Listening .....	388
2.9.2	Reading .....	397
2.9.3	Writing .....	407
2.9.4	Speaking .....	417
3.	Analyses of Composite Scores .....	427
3.1	Scale Score Distribution for Composites .....	427
3.1.1	Oral .....	428
3.1.2	Literacy .....	435
3.1.3	Comprehension .....	442
3.1.4	Overall .....	449
3.2	Proficiency Level Distribution for Composites .....	456
3.2.1	Oral .....	457
3.2.2	Literacy .....	464
3.2.3	Comprehension .....	471
3.2.4	Overall .....	478
4.	Annual Updates of Validity Evidence .....	485
4.1.	Standards .....	486

4.1.1. Test Content .....	486
4.1.2. Response Processes.....	486
4.1.3. Internal Structure .....	486
4.1.4. Relation to Other Variables .....	486
5 Reliability.....	487
5.1 Reliability of the Domain Scores.....	492
5.1.1 Listening.....	495
5.1.2 Reading.....	497
5.1.3 Writing.....	499
5.1.4 Speaking .....	501
5.2. Interrater Agreement Rates .....	502
5.2.1 Listening .....	503
5.2.2 Reading .....	503
5.2.3 Writing .....	503
5.3 Conditional Standard Errors of Measurement at Cut Score .....	507
5.3.1 Listening .....	509
5.3.2 Reading .....	513
5.3.3 Writing .....	517
5.3.4 Speaking .....	521
5.4 Accuracy and Consistency .....	524
5.4.1 Listening.....	530
5.4.2 Reading.....	531
5.4.3 Writing.....	533
5.4.4 Speaking .....	534
5.5 Reliabilities of Students' Composite Scale Scores.....	536
5.5.1 Oral .....	539
5.5.2 Literacy .....	543
5.5.3 Comprehension.....	547
5.5.4 Overall.....	551
5.6 Conditional Standard Error of Measurement for Composites.....	555

5.6.1	Oral .....	557
5.6.2	Literacy .....	562
5.6.3	Comprehension .....	568
5.6.4	Overall.....	574
5.7	Accuracy and Consistency of Composites.....	578
5.7.1	Oral .....	582
5.7.2	Literacy .....	583
5.7.3	Comprehension .....	585
5.7.4	Overall.....	586
6	Quality Control .....	588
6.1	Content Development Quality Control .....	588
6.2	Test Administration Quality Control.....	590
6.3	Rater Quality Control .....	591
6.4	Score Reporting Quality Control.....	593
6.5	Data Forensic Quality Control.....	594

# 1. Annual Test Results

In this section of the report, detail is provided on students' participation in the assessment and on scale score and proficiency level (PL) results. These data are disaggregated in several ways, including by grade-level cluster, grade, and tier, and also by gender, ethnicity, and race.

Analyses use the Census Bureau approach to reporting race and ethnicity (<https://www.census.gov/topics/population/race/about.html>). Ethnicity is conceptualized as a binary category (Hispanic or non-Hispanic). There are five categories for race: American Indian/Alaskan Native, Asian, Black/African American, Pacific Islander/Hawaiian, and White. The race and ethnicity categories 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. Starting with Series 202, students who identify as Hispanic are included in the Hispanic (of any race) category, regardless of how many racial categories they are included in. Students who are identified as one of the racial categories (e.g., Asian) and 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, then they are labeled non-Hispanic multiracial.

A total of 8 students were excluded from the analyses due to mismatches in students' tiers across domains. In addition, 10,695 students taking Paper ACCESS tests in Colorado used equated scores to the Online ACCESS tests; therefore, their score analyses were not included in this 503 Paper Annual Technical Report. For the equated scoring procedure, please refer to the WIDA mode-adjustment procedure report.

## 1.1 Participation

Participation in ACCESS Paper is shown in three ways: by grade-level cluster, by grade, and by tier. Participation data are reported by state, by gender, and ethnicity.

### 1.1.1 Grade-Level Cluster

Table 1.1.1.1 shows participation across the 41 WIDA states and U.S. territories that participated in the operational testing program of ACCESS Paper in 2021–2022 by grade level. The rows provide data for the number of students in that grade-level cluster who took the test by state, with the final row showing the total number of participants across all 41 states and territories. Some states' sample sizes are small except for Kindergarten, which is only in Paper form, since most students take the Online form of the tests. The biggest state was Florida, which constitutes about 53% of the students who take Paper ACCESS. Illinois, Washington, and Virginia were the next largest states. The full names of acronyms of U.S. territories are the following: BI, Bureau of Indian Education; DC, District of Columbia; DD, Department of Defense Education Activity; MP, Northern Mariana Islands; and VI, Virgin Islands.

**Table 1.1.1.1**

Participation by Grade-Level Cluster by State, S503 Paper

State	Cluster							Total
	K	1	2	3	4–5	6–8	9–12	
AK	834	3	2	4	15	23	83	964
AL	4017	0	0	1	1	2	1	4022
BI	539	168	143	146	243	319	104	1662
CO	9525	185	185	166	238	247	149	10695
DC	1041	1	0	2	2	0	1	1047
DD	693	0	1	1	7	0	0	702
DE	1626	3	0	0	6	3	0	1638
FL	34523	31513	31191	30394	45506	45422	42737	261286
GA	15002	148	115	20	40	29	20	15374
HI	1412	2	0	2	1	1	3	1421
ID	1588	5	8	6	10	7	14	1638
IL	22923	174	149	134	262	186	139	23967
IN	7950	14	20	18	26	16	8	8052
KY	4468	7	5	11	14	3	5	4513
MA	11546	40	15	26	65	32	27	11751
MD	10982	2	4	11	23	17	14	11053
ME	475	2	1	5	4	2	3	492
MI	8797	94	108	113	175	215	224	9726
MN	7972	29	42	38	96	59	33	8269
MO	4459	10	11	13	18	9	2	4522
MP	37	0	0	0	0	0	0	37
MT	216	0	0	0	0	0	0	216
NC	11827	5	12	8	20	16	8	11896
ND	339	3	2	4	5	10	7	370
NH	517	2	2	2	5	1	3	532
NJ	13187	84	77	40	43	29	33	13493
NM	3950	0	0	2	4	3	34	3993
NV	5520	0	0	0	0	1	0	5521
OK	6236	8	4	11	21	59	13	6352
PA	7384	339	267	209	369	361	339	9268
RI	1437	2	0	2	2	5	4	1452
SC	4038	5	8	4	15	8	5	4083
SD	907	24	22	19	36	30	0	1038
TN	7644	5	8	8	9	4	9	7687
UT	3875	1	0	0	0	20	29	3925
VA	14061	941	219	136	171	117	137	15782
VI	73	0	0	0	0	1	0	74
VT	164	3	1	3	4	3	0	178

WA	15336	7	7	8	7	7	11	15383
WI	4492	25	32	31	55	44	4	4683
WY	253	2	1	0	4	7	4	271
Total	251865	33856	32662	31598	47522	47318	44207	489028

Table 1.1.1.2 shows participation by grade-level cluster and by gender across all states and territories for the population of students who participated in ACCESS Paper, while Table 1.1.1.3 shows participation by grade-level cluster and by ethnicity. The gender ratio was 45% female and 51% male in general. The Hispanic ethnicity percentage was about 78% in all clusters except Kindergarten, which was 66%.

**Table 1.1.1.2**

Participation by Grade-Level Cluster by Gender, S503 Paper

Cluster		Gender			Total
		F	M	Missing	
K	Count	109747	123013	19105	251865
	% within Cluster	43.57%	48.84%	7.59%	100%
1	Count	15795	17952	109	33856
	% within Cluster	46.66%	53.02%	0.32%	100%
2	Count	15302	17273	87	32662
	% within Cluster	46.85%	52.89%	0.26%	100%
3	Count	14431	17067	100	31598
	% within Cluster	45.67%	54.01%	0.32%	100%
4–5	Count	21622	25743	157	47522
	% within Cluster	45.5%	54.17%	0.33%	100%
6–8	Count	21717	25461	140	47318
	% within Cluster	45.9%	53.8%	0.3%	100%
9–12	Count	20390	23626	191	44207
	% within Cluster	46.13%	53.44%	0.43%	100%
Total	Count	219004	250135	19889	489028
	% within Cluster	44.78%	51.15%	4.07%	100%

**Table 1.1.1.3**

Participation by Grade-Level Cluster by Ethnicity, S503 Paper

Cluster		Ethnicity			Total
		Hispanic	Non-Hispanic	Unknown	
K	Count	166565	69543	15757	251865
	% within Cluster	66.13%	27.61%	6.26%	100%
1	Count	26635	6736	485	33856
	% within Cluster	78.67%	19.9%	1.43%	100%
2	Count	25777	6454	431	32662
	% within Cluster	78.92%	19.76%	1.32%	100%
3	Count	25098	6125	375	31598
	% within Cluster	79.43%	19.38%	1.19%	100%
4–5	Count	37571	9285	666	47522
	% within Cluster	79.06%	19.54%	1.4%	100%
6–8	Count	37412	9153	753	47318
	% within Cluster	79.07%	19.34%	1.59%	100%
9–12	Count	34204	9203	800	44207
	% within Cluster	77.37%	20.82%	1.81%	100%
Total	Count	353262	116499	19267	489028
	% within Cluster	72.24%	23.82%	3.94%	100%

**1.1.2 Grade**

This section provides data similar to that in the previous section but broken out by grade rather than by grade-level cluster. As shown in Table 1.1.2.1, the largest grade was Kindergarten, which comprised almost 51% of the Paper ACCESS population.



**Table 1.1.2.1**

Participation by Grade by State, S503 Paper

State	Grade													Total
	K	1	2	3	4	5	6	7	8	9	10	11	12	
AK	834	3	2	4	5	10	6	10	7	28	19	16	20	964
AL	4017	0	0	1	1	0	0	2	0	1	0	0	0	4022
BI	539	168	143	146	133	110	128	100	91	33	20	25	26	1662
CO	9525	185	185	166	125	113	77	82	88	19	32	43	55	10695
DC	1041	1	0	2	0	2	0	0	0	1	0	0	0	1047
DD	693	0	1	1	2	5	0	0	0	0	0	0	0	702
DE	1626	3	0	0	3	3	2	1	0	0	0	0	0	1638
FL	34523	31513	31191	30394	22757	22749	16119	14954	14349	13317	11897	10563	6960	261286
GA	15002	148	115	20	23	17	8	10	11	8	6	4	2	15374
HI	1412	2	0	2	0	1	0	0	1	0	0	1	2	1421
ID	1588	5	8	6	6	4	3	3	1	7	3	2	2	1638
IL	22923	174	149	134	137	125	66	53	67	33	35	35	36	23967
IN	7950	14	20	18	13	13	9	2	5	2	3	1	2	8052
KY	4468	7	5	11	10	4	1	1	1	2	2	1	0	4513
MA	11546	40	15	26	45	20	14	10	8	12	2	5	8	11751
MD	10982	2	4	11	15	8	4	9	4	6	2	5	1	11053
ME	475	2	1	5	1	3	0	0	2	1	1	1	0	492
MI	8797	94	108	113	83	92	72	76	67	64	57	53	50	9726
MN	7972	29	42	38	59	37	20	21	18	15	10	5	3	8269
MO	4459	10	11	13	9	9	2	4	3	1	1	0	0	4522
MP	37	0	0	0	0	0	0	0	0	0	0	0	0	37
MT	216	0	0	0	0	0	0	0	0	0	0	0	0	216
NC	11827	5	12	8	11	9	7	5	4	2	4	2	0	11896
ND	339	3	2	4	3	2	2	4	4	3	2	1	1	370
NH	517	2	2	2	4	1	0	1	0	2	1	0	0	532

NJ	13187	84	77	40	25	18	13	11	5	12	7	6	8	13493
NM	3950	0	0	2	2	2	0	0	3	8	7	14	5	3993
NV	5520	0	0	0	0	0	1	0	0	0	0	0	0	5521
OK	6236	8	4	11	10	11	11	16	32	2	7	2	2	6352
PA	7384	339	267	209	203	166	140	108	113	134	105	50	50	9268
RI	1437	2	0	2	1	1	0	4	1	1	1	1	1	1452
SC	4038	5	8	4	9	6	3	2	3	1	2	1	1	4083
SD	907	24	22	19	17	19	15	3	12	0	0	0	0	1038
TN	7644	5	8	8	7	2	1	3	0	3	5	0	1	7687
UT	3875	1	0	0	0	0	0	11	9	11	6	4	8	3925
VA	14061	941	219	136	95	76	31	45	41	65	27	26	19	15782
VI	73	0	0	0	0	0	0	0	1	0	0	0	0	74
VT	164	3	1	3	1	3	1	0	2	0	0	0	0	178
WA	15336	7	7	8	5	2	3	2	2	0	5	4	2	15383
WI	4492	25	32	31	31	24	19	14	11	2	0	1	1	4683
WY	253	2	1	0	1	3	4	1	2	1	0	2	1	271
Total	251865	33856	32662	31598	23852	23670	16782	15568	14968	13797	12269	10874	7267	489028

**Table 1.1.2.2**

Participation by Grade by Gender, S503 Paper

Grade		Gender			Total
		F	M	Missing	
K	Count	109747	123013	19105	251865
	% within Grade	43.57%	48.84%	7.59%	100%
1	Count	15795	17952	109	33856
	% within Grade	46.66%	53.02%	0.32%	100%
2	Count	15302	17273	87	32662
	% within Grade	46.85%	52.88%	0.27%	100%
	Count	14431	17067	100	31598

3	% within Grade	45.67%	54.01%	0.32%	100%
4	Count	10908	12851	93	23852
	% within Grade	45.73%	53.88%	0.39%	100%
5	Count	10714	12892	64	23670
	% within Grade	45.26%	54.47%	0.27%	100%
6	Count	7812	8915	55	16782
	% within Grade	46.55%	53.12%	0.33%	100%
7	Count	7131	8389	48	15568
	% within Grade	45.81%	53.88%	0.31%	100%
8	Count	6774	8157	37	14968
	% within Grade	45.26%	54.49%	0.25%	100%
9	Count	6201	7516	80	13797
	% within Grade	44.94%	54.48%	0.58%	100%
10	Count	5710	6506	53	12269
	% within Grade	46.54%	53.03%	0.43%	100%
11	Count	5029	5806	39	10874
	% within Grade	46.25%	53.39%	0.36%	100%
12	Count	3450	3798	19	7267
	% within Grade	47.47%	52.27%	0.26%	100%
Total	Count	219004	250135	19889	489028
	% within Grade	44.78%	51.15%	4.07%	100%

**Table 1.1.2.3**

Participation by Grade by Ethnicity, S503 Paper

Grade		Ethnicity			Total
		Hispanic	Non-Hispanic	Unknown	
K	Count	166565	69543	15757	251865
	% within Grade	66.13%	27.61%	6.26%	100%
1	Count	26635	6736	485	33856
	% within Grade	78.67%	19.9%	1.43%	100%
2	Count	25777	6454	431	32662
	% within Grade	78.92%	19.76%	1.32%	100%
3	Count	25098	6125	375	31598
	% within Grade	79.43%	19.38%	1.19%	100%
4	Count	18820	4672	360	23852
	% within Grade	78.9%	19.59%	1.51%	100%
5	Count	18751	4613	306	23670
	% within Grade	79.22%	19.49%	1.29%	100%
6	Count	13225	3237	320	16782
	% within Grade	78.8%	19.29%	1.91%	100%
7	Count	12285	3074	209	15568
	% within Grade	78.91%	19.75%	1.34%	100%
8	Count	11902	2842	224	14968
	% within Grade	79.52%	18.99%	1.49%	100%
9	Count	10915	2518	364	13797
	% within Grade	79.11%	18.25%	2.64%	100%
10	Count	9571	2492	206	12269
	% within Grade	78.01%	20.31%	1.68%	100%
11	Count	8374	2370	130	10874

	% within Grade	77.01%	21.79%	1.2%	100%
12	Count	5344	1823	100	7267
	% within Grade	73.53%	25.1%	1.37%	100%
Total	Count	353262	116499	19267	489028
	% within Grade	72.24%	23.82%	3.94%	100%

### 1.1.3 Tier

This section provides information on participation by tier. The tables show this information in several ways:

- By grade-level cluster, tier, and domain
- By grade, tier, and domain
- By grade-level cluster and tier for gender
- By grade-level cluster and tier for ethnicity

Table 1.1.3.1 shows the number of students in each tier per cluster. In Grade 1, 55% of students were in Tier A and 45% in Tier B/C. In Grade 2, 29% of students were in Tier A and 71% in Tier B/C. In Grade 3 and Grades 4–5, 21-22% were in Tier A and 79-78% in Tier B/C. In Grades 6–8 and 9–12, there were about 28-29% of students in Tier A and 71-72% in Tier B/C. In all domains these percentages remained the same since students were placed in one tier throughout the test.

**Table 1.1.3.1**

Participation by Grade-Level Cluster by Tier by Domain, S503 Paper

Cluster			Domain			
			Listening	Reading	Speaking	Writing
K	Tier	-	251859	251860	251860	251860
1	Tier	A	18615	18623	18624	18617
		B	15226	15230	15227	15228
	Total		33841	33853	33851	33845
2	Tier	A	9436	9436	9434	9437
		B	23218	23222	23221	23223
	Total		32654	32658	32655	32660
3	Tier	A	7022	7024	7024	7023
		B	24571	24573	24572	24573
	Total		31593	31597	31596	31596
4–5	Tier	A	9842	9844	9844	9844
		B	37677	37677	37675	37678
	Total		47519	47521	47519	47522

6–8	Tier	A	13466	13467	13466	13467
		B	33847	33847	33849	33847
	Total		47313	47314	47315	47314
9–12	Tier	A	13210	13210	13211	13209
		B	30984	30986	30981	30982
	Total		44194	44196	44192	44191

**Table 1.1.3.2**

Participation by Grade by Tier by Domain, S503 Paper

Cluster			Domain			
			Listening	Reading	Speaking	Writing
K	Tier	-	251859	251860	251860	251860
1	Tier	A	18615	18623	18624	18617
		B	15226	15230	15227	15228
	Total		33841	33853	33851	33845
2	Tier	A	9436	9436	9434	9437
		B	23218	23222	23221	23223
	Total		32654	32658	32655	32660
3	Tier	A	7022	7024	7024	7023
		B	24571	24573	24572	24573
	Total		31593	31597	31596	31596
4	Tier	A	5033	5034	5034	5034
		B	18817	18817	18816	18818
	Total		23850	23851	23850	23852
5	Tier	A	4809	4810	4810	4810
		B	18860	18860	18859	18860
	Total		23669	23670	23669	23670
6	Tier	A	4579	4579	4578	4579
		B	12202	12202	12202	12202
	Total		16781	16781	16780	16781
7	Tier	A	4449	4450	4450	4450
		B	11116	11116	11118	11115
	Total		15565	15566	15568	15565
8	Tier	A	4438	4438	4438	4438
		B	10529	10529	10529	10530



	Total		14967	14967	14967	14968
9	Tier	A	5129	5129	5131	5128
		B	8661	8662	8661	8661
	Total		13790	13791	13792	13789
10	Tier	A	3605	3605	3604	3605
		B	8660	8662	8659	8660
	Total		12265	12267	12263	12265
11	Tier	A	3019	3019	3019	3019
		B	7854	7853	7852	7854
	Total		10873	10872	10871	10873
12	Tier	A	1457	1457	1457	1457
		B	5809	5809	5809	5807
	Total		7266	7266	7266	7264

**Table 1.1.3.3**

Participation by Grade-Level Cluster by Tier by Gender, S503 Paper

Cluster	Tier		Gender			Total
			F	M	Missing	
K	-	Count	109747	123013	19105	251865
		% within Tier	43.6%	48.8%	7.6%	100%
1	A	Count	8501	10036	89	18626
		% within Tier	45.6%	53.9%	0.5%	100%
	BC	Count	7294	7916	20	15230
		% within Tier	47.9%	52%	0.1%	100%
2	A	Count	4166	5219	53	9438
		% within Tier	44.1%	55.3%	0.6%	100%
	BC	Count	11136	12054	34	23224
		% within Tier	48%	51.9%	0.1%	100%
3	A	Count	3088	3893	43	7024
		% within Tier	44%	55.4%	0.6%	100%
	BC	Count	11343	13174	57	24574

		% within Tier	46.2%	53.6%	0.2%	100%
4–5	A	Count	4379	5386	79	9844
		% within Tier	44.5%	54.7%	0.8%	100%
	BC	Count	17243	20357	78	37678
		% within Tier	45.8%	54%	0.2%	100%
6–8	A	Count	6195	7188	85	13468
		% within Tier	46%	53.4%	0.6%	100%
	BC	Count	15522	18273	55	33850
		% within Tier	45.9%	54%	0.2%	100%
9–12	A	Count	5978	7113	123	13214
		% within Tier	45.2%	53.8%	0.9%	100%
	BC	Count	14412	16513	68	30993
		% within Tier	46.5%	53.3%	0.2%	100%

Table 1.1.3.4 presents percentages of Hispanic and other ethnic groups in tiers. Overall, the percentages of Hispanic students in Tier A were 3% to 7% higher than in Tier B/C except in Grades 2 and 3.

**Table 1.1.3.4**

Participation by Grade-Level Cluster by Tier by Ethnicity, S503 Paper

Cluster	Tier		Ethnicity			Total
			Hispanic	Other	Unknown	
K	-	Count	166565	69543	15757	251865
		% within Tier	66.1%	27.6%	6.3%	100%
1	A	Count	14976	3334	316	18626
		% within Tier	80.4%	17.9%	1.7%	100%
	BC	Count	11659	3402	169	15230
		% within Tier	76.6%	22.3%	1.1%	100%
2	A	Count	7474	1739	225	9438
		% within Tier	79.2%	18.4%	2.4%	100%
	BC	Count	18303	4715	206	23224
		% within Tier	78.8%	20.3%	0.9%	100%
3	A	Count	5615	1265	144	7024
		% within Tier	79.9%	18%	2.1%	100%
	BC	Count	19483	4860	231	24574
		% within Tier	79.3%	19.8%	0.9%	100%
4–5	A	Count	7994	1621	229	9844
		% within Tier	81.2%	16.5%	2.3%	100%
	BC	Count	29577	7664	437	37678
		% within Tier	78.5%	20.3%	1.2%	100%
6–8	A	Count	11266	2000	202	13468
		% within Tier	83.7%	14.9%	1.5%	100%
	BC	Count	26146	7153	551	33850
		% within Tier	77.2%	21.1%	1.6%	100%
9–12	A	Count	10841	2055	318	13214
		% within Tier	82%	15.6%	2.4%	100%
	BC	Count	23363	7148	482	30993
		% within Tier	75.4%	23.1%	1.6%	100%

## 1.2 Scale Score Results

### 1.2.1 Mean Scale Score Across Domain and Composite Score by Cluster

This section shows mean (average) scale scores by grade-level cluster across the eight scores awarded on ACCESS, first for the four domains (Listening, Speaking, Reading, and Writing) and then for the four composites (Oral Language, Literacy, Comprehension, and Overall). The mean scale scores are expected to increase as grade increases, as ACCESS is vertically scaled; however, there is also an intersection between this principle and the population of test-takers. In this section, under each average, the number of students in each group is also given. Tables are provided for the total student population, for the student population by gender, and for the student population by race and ethnicity. In Table 1.2.1.1, the order of average scale scores among single domains in descending order were Listening, Reading, Speaking, and then Writing in clusters of 1, 2–3, 4–5, and 6–8. Kindergarten had the average scale scores in the order of Speaking, Listening, Writing, and then Reading. Cluster 9–12 had the order of Reading, Listening, Writing, and then Speaking.

**Table 1.2.1.1**

Mean Scale Scores by Grade-Level Cluster, S503 Paper

Cluster		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
K	Mean	252.9	170.69	183.2	255.26	254.34	177.16	195.34	200.11
	N	241979	241977	241972	241969	241967	241972	241974	241960
1	Mean	300.77	284.37	242.45	263.19	283.85	264.57	289.92	271.13
	N	28412	25868	33645	33450	28240	25859	22913	22785
2	Mean	326.85	309.88	271.59	285.35	307.27	291.94	315.55	297.08
	N	30027	26864	32461	32244	29834	26860	25383	25221
3	Mean	355.59	335.18	289.43	299.65	328.4	313.12	341.62	318
	N	29384	26242	31423	31233	29200	26239	25022	24867
4–5	Mean	374.24	349.04	325.91	326.71	351.43	338.37	357.07	342.71
	N	45263	41512	47273	46910	44921	41508	40204	39905
6–8	Mean	379.59	356.54	321.92	342.99	362.36	339.71	363.76	346.63
	N	44657	40976	47047	46721	44335	40963	39389	39108
9–12	Mean	380.94	383.6	351	344.21	363.93	367.95	383.31	367.1
	N	41414	38650	44024	43564	40962	38642	37004	36604

Table 1.2.1.2 demonstrated that female groups performed higher than male groups in general.

**Table 1.2.1.2**

Mean Scale Scores by Grade-Level Cluster by Gender, S503 Paper

Cluster	Gender		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
K	F	Mean	257.77	170.71	187.08	261.87	260.08	179.12	196.81	203.19
		N	104958	104958	104955	104957	104957	104955	104957	104954
	M	Mean	247.17	170.27	180.16	246.77	247.23	175.43	193.33	196.76
		N	117931	117928	117927	117922	117922	117927	117928	117920
	Missing	Mean	261.49	173.23	180.69	271.38	266.71	177.17	199.7	203.83
		N	19090	19091	19090	19090	19088	19090	19089	19086
1	F	Mean	302.32	284.89	247.62	265.68	285.65	267.2	290.82	273.44
		N	13499	12197	15701	15607	13414	12193	10913	10847
	M	Mean	299.52	283.97	238.09	261.23	282.38	262.35	289.2	269.16
		N	14821	13575	17835	17734	14734	13570	11913	11851
	Missing	Mean	274.76	274.08	212.22	224.35	254.51	246.01	275.39	250.07
		N	92	96	109	109	92	96	87	87
2	F	Mean	328.38	310.81	277.46	287.97	309.21	295.17	316.62	299.86
		N	14264	12769	15217	15123	14178	12768	12186	12113
	M	Mean	325.55	309.15	266.49	283.12	305.59	289.1	314.67	294.61
		N	15681	14020	17157	17034	15574	14017	13124	13035
	Missing	Mean	310.04	288.59	251.15	265.66	289.35	272.31	295.12	276.89
		N	82	75	87	87	82	75	73	73
3	F	Mean	354.88	335.44	295.59	300.7	328.56	316.26	341.56	320.19
		N	13581	12165	14362	14270	13492	12163	11699	11623
	M	Mean	356.32	335.05	284.33	298.86	328.36	310.49	341.77	316.17
		N	15709	13995	16961	16864	15615	13994	13246	13168
	Missing	Mean	335.26	319.93	270.55	283.46	311.43	295.46	325.01	301.14
		N	94	82	100	99	93	82	77	76
4–5	F	Mean	373.97	349.75	333.02	327.41	351.61	342.15	357.41	345.33
		N	20727	19120	21499	21330	20570	19119	18606	18469
	M	Mean	374.62	348.54	320.15	326.33	351.46	335.27	356.9	340.6
		N	24392	22262	25617	25423	24207	22259	21473	21311
	Missing	Mean	348.04	331.18	292.25	291.96	322.15	314.03	336.03	315.88
		N	144	130	157	157	144	130	125	125
6–8	F	Mean	379.54	357.7	326.89	341.33	361.6	342.75	364.59	348.58
		N	20537	18982	21602	21428	20366	18976	18257	18111
	M	Mean	379.81	355.62	317.85	344.72	363.24	337.2	363.14	345.09
		N	23992	21869	25305	25155	23843	21862	21017	20884
	Missing	Mean	346.52	342.44	289.31	282.97	316.52	316.42	343.99	316.9
		N	128	125	140	138	126	125	115	113
9–12	F	Mean	381.65	386.11	356.14	343.55	363.86	371.81	385.18	369.7
		N	19228	18003	20313	20113	19027	18001	17328	17151
	M	Mean	380.49	381.51	346.73	345.14	364.24	364.7	381.78	364.96
		N	22024	20482	23520	23262	21774	20476	19527	19305
	Missing	Mean	356.94	368.42	329.6	299.99	331.49	350.25	365.71	345.51
		N	162	165	191	189	161	165	149	148

Table 1.2.1.3 presents scale score performance by ethnic groups. The top three performing ethnic groups were Asian students, multiracial students, and White students.

**Table 1.2.1.3**

Mean Scale Scores by Grade-Level Cluster by Ethnicity, S503 Paper

Cluster	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
K	Non-Hispanic Asian	Mean	282.26	212.62	223.34	286.38	284.59	218.23	233.5	237.91
		N	28467	28467	28467	28466	28466	28467	28467	28466
	Non-Hispanic Pacific Islander	Mean	253.8	159.72	173.21	261.03	257.67	166.66	187.95	193.78
		N	1494	1494	1494	1494	1494	1494	1494	1494
	Non-Hispanic Black	Mean	265.81	189.05	194.28	281.02	273.68	191.9	212.07	216.22
		N	13227	13227	13225	13227	13227	13225	13227	13225
	Hispanic (of Any Race)	Mean	244.36	160.15	173.14	244.95	244.92	166.86	185.41	190.07
		N	158997	158995	158993	158990	158988	158993	158992	158984
	Non-Hispanic American Indian	Mean	269.95	166.58	172.37	273.71	272.1	169.68	197.58	200.21
		N	1829	1829	1829	1829	1829	1829	1829	1829
	Non-Hispanic Multiracial	Mean	279.42	198.65	207.06	288.57	284.26	203.11	222.87	227.2
		N	1266	1266	1266	1266	1266	1266	1266	1266
	Non-Hispanic White	Mean	272.62	187.54	203.06	277.92	275.54	195.53	213.05	219.32
		N	20994	20994	20993	20992	20992	20993	20994	20991
	Unknown	Mean	244.61	162.65	176.73	245.85	245.48	169.89	187.23	192.36
		N	15705	15705	15705	15705	15705	15705	15705	15705
1	Non-Hispanic Asian	Mean	307.89	299.32	260.22	275.05	293.09	280.87	302.7	285.51
		N	1158	1037	1316	1309	1150	1037	941	937
	Non-Hispanic Pacific Islander	Mean	296.39	279.97	235.08	250.51	277.21	256.31	285.5	263.31
		N	33	29	39	39	33	29	26	26
	Non-Hispanic Black	Mean	299.9	285.98	236.64	268.17	285.94	263.31	290.65	270.99
		N	2021	1849	2504	2478	2001	1848	1594	1580
	Hispanic (of Any Race)	Mean	300.22	283.02	241.77	260.94	282.41	263.45	288.81	269.85
		N	22414	20416	26465	26325	22290	20409	18091	17997
	Non-Hispanic American Indian	Mean	293.78	278.1	226.32	252.5	275.08	252.82	284.33	261.58
		N	144	126	164	163	142	125	116	113
	Non-Hispanic Multiracial	Mean	310.28	294.81	254.55	286.6	301.18	275.46	300.16	285.33
		N	175	160	206	206	174	160	138	137
	Non-Hispanic White	Mean	305.1	290.28	250.9	276.44	292.77	272.12	295.4	279.19
		N	2052	1860	2466	2448	2038	1860	1650	1640
	Unknown	Mean	291.78	277.63	219.16	255.59	277.63	250.32	282.12	259.24
		N	415	391	485	482	412	391	357	355

Cluster	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
2	Non-Hispanic Asian	Mean	333.28	325.35	288.36	296.33	316	308.1	328.5	311.13
		N	1082	988	1152	1144	1073	987	940	932
	Non-Hispanic Pacific Islander	Mean	308.23	303.23	264.5	278.09	292.13	281.85	302.96	283.88
		N	30	26	32	32	30	26	25	25
	Non-Hispanic Black	Mean	323.79	306.33	264.12	287.78	307.33	287.14	312.16	293.75
		N	2380	2058	2635	2605	2353	2056	1932	1905
	Hispanic (of Any Race)	Mean	326.57	309.13	271.23	283.66	306.27	291.26	314.93	296.31
		N	23741	21289	25619	25452	23596	21288	20121	20001
	Non-Hispanic American Indian	Mean	312.49	299.45	256.19	271.48	292.56	280.25	302.89	284.03
		N	107	91	116	115	106	91	87	87
	Non-Hispanic Multiracial	Mean	336.21	319.97	278.47	295.75	316.74	300.05	324.93	304.68
		N	168	159	184	184	168	159	150	150
	Non-Hispanic White	Mean	332.66	316.23	279.9	298.01	316.14	299.29	321.79	304.76
		N	2121	1903	2292	2282	2111	1903	1799	1793
	Unknown	Mean	314.59	296.57	251.6	274.2	296.37	275.82	302.35	282.67
		N	398	350	431	430	397	350	329	328
3	Non-Hispanic Asian	Mean	366.35	344.8	300.75	307.16	337.64	323.82	352.43	329.35
		N	938	824	1003	995	930	824	777	771
	Non-Hispanic Pacific Islander	Mean	344.89	337.03	287.1	305.87	326.17	312.68	339.86	316.47
		N	38	38	40	38	36	38	36	34
	Non-Hispanic Black	Mean	354.56	332.54	284.06	302.53	329.58	308.81	339.45	315.46
		N	2583	2303	2826	2807	2566	2303	2154	2141
	Hispanic (of Any Race)	Mean	355.06	334.77	289.4	298.36	327.44	312.87	341.11	317.44
		N	23400	20929	24944	24800	23259	20926	20015	19893
	Non-Hispanic American Indian	Mean	347.59	333.56	289.44	290.52	320.36	312.45	338.32	315.82
		N	115	102	118	117	114	102	100	99
	Non-Hispanic Multiracial	Mean	366.2	344.76	298.14	311.87	340.13	322.27	352.84	329.22
		N	143	131	155	151	139	131	122	118
	Non-Hispanic White	Mean	360.18	339.95	293.97	311	336.57	317.94	346.59	324.19
		N	1819	1627	1962	1953	1810	1627	1547	1541
	Unknown	Mean	345.32	327.64	274.92	281.87	314.79	304.03	333.54	308.01
		N	348	288	375	372	346	288	271	270

Cluster	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
4-5	Non-Hispanic Asian	Mean	387.09	361.28	338.18	341.91	364.91	350.23	369.36	354.84
		N	1280	1184	1328	1315	1267	1184	1152	1141
	Non-Hispanic Pacific Islander	Mean	372.55	349.31	335.07	335	356.17	344.64	358.89	350.49
		N	53	45	56	55	52	45	44	43
	Non-Hispanic Black	Mean	374.67	348.47	323.67	333.78	355.21	336.86	356.71	342.65
		N	4399	3919	4645	4606	4362	3919	3773	3742
	Hispanic (of Any Race)	Mean	373.44	348.41	325.58	324.6	349.98	337.87	356.38	341.92
		N	35836	32951	37369	37093	35575	32947	31944	31711
	Non-Hispanic American Indian	Mean	370.03	346.91	325.67	322.02	347.12	336.75	353.55	340.1
		N	234	215	246	245	233	215	205	205
	Non-Hispanic Multiracial	Mean	383.4	356.29	335.21	350.1	367.69	346.36	365.8	353.86
		N	208	191	218	214	204	191	183	180
	Non-Hispanic White	Mean	380.67	354.53	330.47	338.66	360.7	343.79	362.94	349.43
		N	2621	2433	2745	2720	2599	2433	2352	2334
6-8	Unknown	Mean	362.83	339.2	312.84	309.81	337.86	327.66	347.21	331.72
		N	632	574	666	662	629	574	551	549
	Non-Hispanic Asian	Mean	397.32	370.85	337.99	364.13	381.64	355	379.2	363.19
		N	1192	1124	1263	1256	1186	1123	1075	1069
	Non-Hispanic Pacific Islander	Mean	382.84	353.27	323.54	362.45	373	338.39	362.29	349.54
		N	56	51	59	58	55	51	49	48
	Non-Hispanic Black	Mean	383.37	358.12	323.35	358.83	371.78	341.52	365.95	350.64
		N	4161	3718	4449	4401	4117	3717	3533	3495
	Hispanic (of Any Race)	Mean	377.74	355.3	320.67	339.2	359.6	338.39	362.3	344.85
		N	35380	32462	37177	36931	35133	32452	31267	31053
	Non-Hispanic American Indian	Mean	386.88	358.63	327.68	357.33	372.5	344.42	367.59	353.77
		N	247	225	265	264	246	225	213	212
	Non-Hispanic Multiracial	Mean	391.65	364.01	330.5	357.74	375.31	346.51	371.58	354.52
		N	193	176	202	200	191	176	172	170
	Non-Hispanic White	Mean	389.41	363.53	328.31	357.38	374.53	346.82	371.98	355.55
		N	2727	2543	2879	2860	2706	2542	2443	2424
	Unknown	Mean	376.19	355.15	319.34	335.81	357.91	337.49	362.14	344.47
		N	701	677	753	751	701	677	637	637



Cluster	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
9–12	Non-Hispanic Asian	Mean	401.33	397.78	368.72	366.29	384.75	383.8	399.44	384.17
		N	1320	1252	1390	1370	1301	1252	1206	1189
	Non-Hispanic Pacific Islander	Mean	383.5	383.97	363.44	353.8	371.26	373.65	383.17	372.53
		N	66	60	70	70	66	60	58	58
	Non-Hispanic Black	Mean	381.49	385.02	349.71	358.79	371.23	368.13	384.32	369.22
		N	4451	3983	4824	4770	4399	3981	3755	3710
	Hispanic (of Any Race)	Mean	378.78	381.96	349.77	340.19	360.86	366.44	381.47	365.12
		N	32113	30065	34044	33705	31777	30060	28837	28540
	Non-Hispanic American Indian	Mean	396.04	389.81	358.76	346.58	373.65	374.19	392.3	374.81
		N	166	167	181	180	165	167	155	155
	Non-Hispanic Multiracial	Mean	395.31	391.36	362.49	365.46	381.68	377.78	393.66	380.21
		N	175	164	183	182	174	164	158	157
	Non-Hispanic White	Mean	397.62	395.38	361.96	364.44	382.57	379.68	396.88	381.07
		N	2392	2264	2532	2499	2359	2263	2180	2150
	Unknown	Mean	373.83	379.37	340.28	319.01	348.96	361.06	378.81	358.87
		N	731	695	800	788	721	695	655	645

### 1.2.2 Mean Scale Score Across Domain and Composite Score by Grade

This section shows the mean scale scores broken down by grade rather than by grade-level cluster. Tables are provided for the total student population, for the student population by gender, and for the student population by race and ethnicity. Table 1.2.2.1 shows the increment of scale scores by grade. Across domains, mean scale scores increased with grades. Mean scale scores increased as grades increased, with slight decreases in grades 6 and 9 in Listening. In Speaking domain there was a slight decrease in grade 9. Table 1.2.2.2 exhibits student performance by gender. Female student groups mostly scored higher than male groups throughout grades, except for a few grades across domains.

**Table 1.2.2.1**

Mean Scale Scores by Grade, S503 Paper

Grade		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
K	Mean	252.9	170.69	183.2	255.26	254.34	177.16	195.34	200.11
	N	241979	241977	241972	241969	241967	241972	241974	241960
1	Mean	300.77	284.37	242.45	263.19	283.85	264.57	289.92	271.13
	N	28412	25868	33645	33450	28240	25859	22913	22785
2	Mean	326.85	309.88	271.59	285.35	307.27	291.94	315.55	297.08
	N	30027	26864	32461	32244	29834	26860	25383	25221
3	Mean	355.59	335.18	289.43	299.65	328.4	313.12	341.62	318
	N	29384	26242	31423	31233	29200	26239	25022	24867
4	Mean	369.69	345.13	319.78	323.05	347.37	333.31	352.92	337.88
	N	22591	20621	23725	23529	22409	20620	19905	19741
5	Mean	378.77	352.9	332.09	330.39	355.48	343.37	361.14	347.43
	N	22672	20891	23548	23381	22512	20888	20299	20164
6	Mean	375.71	351.15	317.64	341.16	359.32	334.76	358.64	342.1
	N	15872	14348	16698	16574	15747	14345	13801	13695
7	Mean	379.09	356.25	321.82	341.18	361.26	339.49	363.38	346.14
	N	14688	13501	15474	15371	14586	13496	12989	12900
8	Mean	384.49	362.74	326.82	346.91	366.91	345.33	369.75	352.08
	N	14097	13127	14875	14776	14002	13122	12599	12513
9	Mean	373.16	375.28	341.9	330.26	353.4	359.22	375.18	357.89
	N	12869	11950	13765	13630	12742	11948	11388	11277
10	Mean	381.82	383.61	351.54	344.18	364.43	368.1	383.51	367.21
	N	11511	10669	12228	12096	11381	10667	10224	10111
11	Mean	384.44	388.76	356.06	349.71	368.31	373.05	387.88	371.92
	N	10196	9556	10825	10717	10088	9555	9166	9072
12	Mean	388.87	391.33	359.87	362.69	376.45	376.29	391.1	376.7
	N	6838	6475	7206	7121	6751	6472	6226	6144

**Table 1.2.2.2**

Mean Scale Scores by Grade by Gender, S503 Paper

Grade	Gender		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
K	F	Mean	257.77	170.71	187.08	261.87	260.08	179.12	196.81	203.19
		N	104958	104958	104955	104957	104957	104955	104957	104954
	M	Mean	247.17	170.27	180.16	246.77	247.23	175.43	193.33	196.76
		N	117931	117928	117927	117922	117922	117927	117928	117920
	Missing	Mean	261.49	173.23	180.69	271.38	266.71	177.17	199.7	203.83
		N	19090	19091	19090	19090	19088	19090	19089	19086
1	F	Mean	302.32	284.89	247.62	265.68	285.65	267.2	290.82	273.44
		N	13499	12197	15701	15607	13414	12193	10913	10847
	M	Mean	299.52	283.97	238.09	261.23	282.38	262.35	289.2	269.16
		N	14821	13575	17835	17734	14734	13570	11913	11851
	Missing	Mean	274.76	274.08	212.22	224.35	254.51	246.01	275.39	250.07
		N	92	96	109	109	92	96	87	87
2	F	Mean	328.38	310.81	277.46	287.97	309.21	295.17	316.62	299.86
		N	14264	12769	15217	15123	14178	12768	12186	12113
	M	Mean	325.55	309.15	266.49	283.12	305.59	289.1	314.67	294.61
		N	15681	14020	17157	17034	15574	14017	13124	13035
	Missing	Mean	310.04	288.59	251.15	265.66	289.35	272.31	295.12	276.89
		N	82	75	87	87	82	75	73	73
3	F	Mean	354.88	335.44	295.59	300.7	328.56	316.26	341.56	320.19
		N	13581	12165	14362	14270	13492	12163	11699	11623
	M	Mean	356.32	335.05	284.33	298.86	328.36	310.49	341.77	316.17
		N	15709	13995	16961	16864	15615	13994	13246	13168
	Missing	Mean	335.26	319.93	270.55	283.46	311.43	295.46	325.01	301.14
		N	94	82	100	99	93	82	77	76
4	F	Mean	369.35	345.66	326.31	323.92	347.56	336.7	353.11	340.18
		N	10412	9517	10844	10754	10331	9517	9239	9168
	M	Mean	370.14	344.75	314.42	322.47	347.35	330.5	352.87	336
		N	12092	11026	12788	12682	11991	11025	10589	10496
	Missing	Mean	348.95	333.6	294.44	301.06	327.6	316.87	338.77	320.7
		N	87	78	93	93	87	78	77	77
5	F	Mean	378.64	353.8	339.85	330.97	355.69	347.56	361.66	350.39
		N	10315	9603	10655	10576	10239	9602	9367	9301
	M	Mean	379.03	352.25	325.86	330.17	355.49	339.95	360.82	345.06
		N	12300	11236	12829	12741	12216	11234	10884	10815
	Missing	Mean	346.65	327.56	289.06	278.73	313.82	309.77	331.65	308.15
		N	57	52	64	64	57	52	48	48

Grade	Gender		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
6	F	Mean	375.83	352.29	322.94	339.2	358.37	338.04	359.57	344.2
		N	7425	6754	7778	7710	7358	6752	6521	6463
	M	Mean	375.77	350.2	313.17	343.23	360.4	331.96	357.9	340.37
		N	8397	7545	8865	8810	8340	7544	7235	7188
	Missing	Mean	347.8	340.24	288.96	284.69	318.65	314.9	342.91	317.39
		N	50	49	55	54	49	49	45	44
7	F	Mean	378.28	357.11	326.73	339.38	360.1	342.36	363.72	347.79
		N	6734	6233	7092	7040	6682	6232	5998	5955
	M	Mean	379.96	355.61	317.83	343.03	362.49	337.17	363.21	344.89
		N	7909	7225	8334	8284	7860	7221	6951	6906
	Missing	Mean	346	339.14	287.75	286.28	317.41	313.51	341.25	314.82
		N	45	43	48	47	44	43	40	39
8	F	Mean	385.2	364.41	331.62	345.85	366.95	348.46	371.21	354.39
		N	6378	5995	6732	6678	6326	5992	5738	5693
	M	Mean	384.07	361.39	323	348.11	367.12	342.79	368.61	350.3
		N	7686	7099	8106	8061	7643	7097	6831	6790
	Missing	Mean	345.27	350	291.84	276.27	312.15	322.45	349.27	318.9
		N	33	33	37	37	33	33	30	30
9	F	Mean	374.74	378.37	347.91	330.84	354.27	363.88	377.72	361.35
		N	5804	5429	6185	6134	5753	5428	5191	5146
	M	Mean	372.14	372.83	337.13	330.27	353.05	355.47	373.23	355.21
		N	6996	6452	7500	7416	6920	6451	6132	6066
	Missing	Mean	343.45	360.93	323.32	285.04	314.84	343.01	355.91	334.42
		N	69	69	80	80	69	69	65	65
10	F	Mean	382.77	386.22	356.41	344.18	364.97	371.91	385.63	370.07
		N	5414	5020	5698	5630	5347	5020	4850	4791
	M	Mean	381.06	381.37	347.45	344.43	364.05	364.82	381.66	364.71
		N	6055	5605	6477	6414	5992	5603	5338	5284
	Missing	Mean	367.83	372.18	328.28	313.29	349.88	351.14	373.53	354.19
		N	42	44	53	52	42	44	36	36
11	F	Mean	384.01	390.75	360.15	347.49	366.72	376.14	388.98	373.42
		N	4746	4454	5007	4964	4701	4454	4297	4261
	M	Mean	384.98	387.14	352.63	352.01	369.95	370.45	387.04	370.73
		N	5417	5066	5779	5714	5354	5065	4837	4779
	Missing	Mean	359.36	371.22	339.62	294.87	327.58	356.61	368.06	349.75
		N	33	36	39	39	33	36	32	32
12	F	Mean	388.63	392.84	364.7	359.74	374.94	379.31	391.95	378.25
		N	3264	3100	3423	3385	3226	3099	2990	2953
	M	Mean	389.14	389.96	355.59	365.5	377.9	373.56	390.36	375.32
		N	3556	3359	3764	3718	3508	3357	3220	3176
	Missing	Mean	378.83	384.06	339.11	339.11	361.24	364.69	383.25	363.73
		N	18	16	19	18	17	16	16	15

**Table 1.2.2.3**

Mean Scale Scores by Grade by Ethnicity, S503 Paper

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
K	Non-Hispanic Asian	Mean	282.26	212.62	223.34	286.38	284.59	218.23	233.5	237.91
		N	28467	28467	28467	28466	28466	28467	28467	28466
	Non-Hispanic Pacific Islander	Mean	253.8	159.72	173.21	261.03	257.67	166.66	187.95	193.78
		N	1494	1494	1494	1494	1494	1494	1494	1494
	Non-Hispanic Black	Mean	265.81	189.05	194.28	281.02	273.68	191.9	212.07	216.22
		N	13227	13227	13225	13227	13227	13225	13227	13225
	Hispanic (of Any Race)	Mean	244.36	160.15	173.14	244.95	244.92	166.86	185.41	190.07
		N	158997	158995	158993	158990	158988	158993	158992	158984
	Non-Hispanic American Indian	Mean	269.95	166.58	172.37	273.71	272.1	169.68	197.58	200.21
		N	1829	1829	1829	1829	1829	1829	1829	1829
	Non-Hispanic Multiracial	Mean	279.42	198.65	207.06	288.57	284.26	203.11	222.87	227.2
		N	1266	1266	1266	1266	1266	1266	1266	1266
	Non-Hispanic White	Mean	272.62	187.54	203.06	277.92	275.54	195.53	213.05	219.32
		N	20994	20994	20993	20992	20992	20993	20994	20991
	Unknown	Mean	244.61	162.65	176.73	245.85	245.48	169.89	187.23	192.36
		N	15705	15705	15705	15705	15705	15705	15705	15705
1	Non-Hispanic Asian	Mean	307.89	299.32	260.22	275.05	293.09	280.87	302.7	285.51
		N	1158	1037	1316	1309	1150	1037	941	937
	Non-Hispanic Pacific Islander	Mean	296.39	279.97	235.08	250.51	277.21	256.31	285.5	263.31
		N	33	29	39	39	33	29	26	26
	Non-Hispanic Black	Mean	299.9	285.98	236.64	268.17	285.94	263.31	290.65	270.99
		N	2021	1849	2504	2478	2001	1848	1594	1580
	Hispanic (of Any Race)	Mean	300.22	283.02	241.77	260.94	282.41	263.45	288.81	269.85
		N	22414	20416	26465	26325	22290	20409	18091	17997
	Non-Hispanic American Indian	Mean	293.78	278.1	226.32	252.5	275.08	252.82	284.33	261.58
		N	144	126	164	163	142	125	116	113
	Non-Hispanic Multiracial	Mean	310.28	294.81	254.55	286.6	301.18	275.46	300.16	285.33
		N	175	160	206	206	174	160	138	137
	Non-Hispanic White	Mean	305.1	290.28	250.9	276.44	292.77	272.12	295.4	279.19
		N	2052	1860	2466	2448	2038	1860	1650	1640
	Unknown	Mean	291.78	277.63	219.16	255.59	277.63	250.32	282.12	259.24
		N	415	391	485	482	412	391	357	355

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
2	Non-Hispanic Asian	Mean	333.28	325.35	288.36	296.33	316	308.1	328.5	311.13
		N	1082	988	1152	1144	1073	987	940	932
	Non-Hispanic Pacific Islander	Mean	308.23	303.23	264.5	278.09	292.13	281.85	302.96	283.88
		N	30	26	32	32	30	26	25	25
	Non-Hispanic Black	Mean	323.79	306.33	264.12	287.78	307.33	287.14	312.16	293.75
		N	2380	2058	2635	2605	2353	2056	1932	1905
	Hispanic (of Any Race)	Mean	326.57	309.13	271.23	283.66	306.27	291.26	314.93	296.31
		N	23741	21289	25619	25452	23596	21288	20121	20001
	Non-Hispanic American Indian	Mean	312.49	299.45	256.19	271.48	292.56	280.25	302.89	284.03
		N	107	91	116	115	106	91	87	87
	Non-Hispanic Multiracial	Mean	336.21	319.97	278.47	295.75	316.74	300.05	324.93	304.68
		N	168	159	184	184	168	159	150	150
	Non-Hispanic White	Mean	332.66	316.23	279.9	298.01	316.14	299.29	321.79	304.76
		N	2121	1903	2292	2282	2111	1903	1799	1793
	Unknown	Mean	314.59	296.57	251.6	274.2	296.37	275.82	302.35	282.67
		N	398	350	431	430	397	350	329	328
3	Non-Hispanic Asian	Mean	366.35	344.8	300.75	307.16	337.64	323.82	352.43	329.35
		N	938	824	1003	995	930	824	777	771
	Non-Hispanic Pacific Islander	Mean	344.89	337.03	287.1	305.87	326.17	312.68	339.86	316.47
		N	38	38	40	38	36	38	36	34
	Non-Hispanic Black	Mean	354.56	332.54	284.06	302.53	329.58	308.81	339.45	315.46
		N	2583	2303	2826	2807	2566	2303	2154	2141
	Hispanic (of Any Race)	Mean	355.06	334.77	289.4	298.36	327.44	312.87	341.11	317.44
		N	23400	20929	24944	24800	23259	20926	20015	19893
	Non-Hispanic American Indian	Mean	347.59	333.56	289.44	290.52	320.36	312.45	338.32	315.82
		N	115	102	118	117	114	102	100	99
	Non-Hispanic Multiracial	Mean	366.2	344.76	298.14	311.87	340.13	322.27	352.84	329.22
		N	143	131	155	151	139	131	122	118
	Non-Hispanic White	Mean	360.18	339.95	293.97	311	336.57	317.94	346.59	324.19
		N	1819	1627	1962	1953	1810	1627	1547	1541
	Unknown	Mean	345.32	327.64	274.92	281.87	314.79	304.03	333.54	308.01
		N	348	288	375	372	346	288	271	270

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
4	Non-Hispanic Asian	Mean	383.91	356.95	333.17	341.47	363.28	345.32	365.32	350.93
		N	711	664	743	734	702	664	642	634
	Non-Hispanic Pacific Islander	Mean	356.39	335.67	325	329.52	343.22	330	343.33	335.17
		N	23	18	23	23	23	18	18	18
	Non-Hispanic Black	Mean	369.41	344.22	316.86	327.8	349.56	331.43	352.19	337.18
		N	2046	1842	2181	2157	2023	1842	1766	1746
	Hispanic (of Any Race)	Mean	368.83	344.49	319.38	320.94	345.89	332.75	352.19	337.02
		N	17862	16303	18718	18568	17724	16302	15758	15633
	Non-Hispanic American Indian	Mean	368.23	347.53	318.46	317.9	343.69	333.92	353.37	336.96
		N	118	109	126	126	118	109	102	102
	Non-Hispanic Multiracial	Mean	382.53	357.11	328.73	354.63	368.68	343.3	366.02	351.75
		N	106	91	110	109	105	91	89	88
	Non-Hispanic White	Mean	376.11	350	325.1	335.9	357.17	338.94	358.45	345
		N	1388	1281	1464	1452	1377	1281	1234	1224
	Unknown	Mean	358.13	335.18	306.19	305.81	333.84	322.07	343.09	326.76
		N	337	313	360	360	337	313	296	296
5	Non-Hispanic Asian	Mean	391.05	366.82	344.56	342.46	366.93	356.5	374.43	359.73
		N	569	520	585	581	565	520	510	507
	Non-Hispanic Pacific Islander	Mean	384.93	358.41	342.09	338.94	366.45	354.41	369.65	361.52
		N	30	27	33	32	29	27	26	25
	Non-Hispanic Black	Mean	379.25	352.24	329.69	339.05	360.11	341.67	360.69	347.43
		N	2353	2077	2464	2449	2339	2077	2007	1996
	Hispanic (of Any Race)	Mean	378.01	352.24	331.81	328.27	354.04	342.89	360.46	346.68
		N	17974	16648	18651	18525	17851	16645	16186	16078
	Non-Hispanic American Indian	Mean	371.87	346.26	333.25	326.38	350.63	339.67	353.73	343.2
		N	116	106	120	119	115	106	103	103
	Non-Hispanic Multiracial	Mean	384.31	355.54	341.81	345.4	366.64	349.15	365.6	355.88
		N	102	100	108	105	99	100	94	92
	Non-Hispanic White	Mean	385.8	359.57	336.6	341.83	364.67	349.18	367.89	354.32
		N	1233	1152	1281	1268	1222	1152	1118	1110
	Unknown	Mean	368.19	344.01	320.66	314.58	342.51	334.37	352	337.53
		N	295	261	306	302	292	261	255	253

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
6	Non-Hispanic Asian	Mean	394.08	366.29	333.49	364.35	380.25	350.34	375.07	359.5
		N	394	372	419	415	391	372	356	353
	Non-Hispanic Pacific Islander	Mean	383.21	348.12	317.16	358.53	371.05	334.31	361.25	348.38
		N	19	16	19	19	19	16	16	16
	Non-Hispanic Black	Mean	378.63	352.24	318.39	354.43	366.89	335.66	360.34	344.88
		N	1503	1330	1594	1569	1478	1330	1270	1250
	Hispanic (of Any Race)	Mean	374.06	350.06	316.63	337.98	356.97	333.68	357.33	340.61
		N	12542	11318	13154	13064	12451	11315	10917	10840
	Non-Hispanic American Indian	Mean	391.33	360.35	330.89	369.05	380.63	348.14	370.55	359.55
		N	95	81	100	100	95	81	78	78
	Non-Hispanic Multiracial	Mean	388.26	361.23	328.39	348.97	368.87	344.89	368.44	351.1
		N	69	61	74	73	68	61	59	58
	Non-Hispanic White	Mean	383.55	356.18	322.09	352.06	368.57	340.06	364.77	348.7
		N	957	881	1018	1014	952	881	839	834
	Unknown	Mean	372.71	349.35	313.7	329.92	353.22	331.18	357.02	338.69
		N	293	289	320	320	293	289	266	266
7	Non-Hispanic Asian	Mean	393.77	368.62	335.98	360.42	377.34	353.09	376.19	360.15
		N	414	393	434	432	413	393	378	377
	Non-Hispanic Pacific Islander	Mean	383.24	348.18	318.79	362.32	370.06	331.82	358.81	343.31
		N	17	17	19	19	17	17	16	16
	Non-Hispanic Black	Mean	383.52	358.05	324.17	358.3	371.92	341.81	365.66	350.67
		N	1389	1238	1503	1492	1379	1238	1166	1157
	Hispanic (of Any Race)	Mean	377.24	354.9	320.4	337.21	358.4	338.04	361.88	344.27
		N	11614	10689	12202	12122	11532	10685	10303	10232
	Non-Hispanic American Indian	Mean	383.38	355.76	327.14	344.95	364.27	342.56	364.53	349.61
		N	84	78	93	92	83	78	72	71
	Non-Hispanic Multiracial	Mean	395.45	363.88	334.54	359.13	377.9	347.18	373.12	355.82
		N	62	56	63	63	62	56	56	56
	Non-Hispanic White	Mean	388.91	364.5	329.09	356.66	374.05	347.72	372.54	356.16
		N	911	842	951	944	903	841	817	810
	Unknown	Mean	373.06	356.88	319.31	330.87	354.58	338.78	362.89	344.75
		N	197	188	209	207	197	188	181	181



Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
8	Non-Hispanic Asian	Mean	404.47	378.03	344.72	367.82	387.7	361.93	386.87	370.42
		N	384	359	410	409	382	358	341	339
	Non-Hispanic Pacific Islander	Mean	382.15	362.67	333.62	366.3	377.58	348.22	366.53	356.94
		N	20	18	21	20	19	18	17	16
	Non-Hispanic Black	Mean	388.8	365.01	328.28	364.56	377.35	347.98	372.77	357.23
		N	1269	1150	1352	1340	1260	1149	1097	1088
	Hispanic (of Any Race)	Mean	382.37	361.37	325.43	342.61	363.77	343.86	368.13	350.07
		N	11224	10455	11821	11745	11150	10452	10047	9981
	Non-Hispanic American Indian	Mean	384.99	359.91	323.93	356.89	371.18	342.06	367.43	351.3
		N	68	66	72	72	68	66	63	63
	Non-Hispanic Multiracial	Mean	391.63	367	329	366.39	379.87	347.56	373.3	356.75
		N	62	59	65	64	61	59	57	56
	Non-Hispanic White	Mean	396.47	370.45	334.44	364.1	381.7	353.16	379.09	362.23
		N	859	820	910	902	851	820	787	780
	Unknown	Mean	383.95	361.9	327.42	348.78	367.52	345.4	368.59	352.29
		N	211	200	224	224	211	200	190	190
9	Non-Hispanic Asian	Mean	393.72	389.64	359.8	359.31	377.04	375.13	390.52	375.15
		N	372	346	396	394	370	346	333	331
	Non-Hispanic Pacific Islander	Mean	383.67	392.93	370.95	348.15	372.61	382.67	389.36	380
		N	18	15	20	20	18	15	14	14
	Non-Hispanic Black	Mean	376.03	378.24	341.29	348.09	362.94	360.64	378.18	361.55
		N	1119	982	1229	1210	1104	981	919	907
	Hispanic (of Any Race)	Mean	370.83	373.54	340.68	326.59	350.5	357.64	373.2	355.87
		N	10205	9513	10886	10783	10105	9512	9076	8988
	Non-Hispanic American Indian	Mean	401.21	388.1	362.23	343.55	375.46	375.63	393.38	378.09
		N	57	60	62	62	57	60	55	55
	Non-Hispanic Multiracial	Mean	395.66	387.6	355.37	349.94	373.5	372.33	390.53	372.89
		N	64	58	65	65	64	58	57	57
	Non-Hispanic White	Mean	392.03	388.57	353.42	351.64	373.78	372.25	390.82	373.76
		N	698	662	743	736	691	662	635	629
	Unknown	Mean	362.77	369.33	329.82	297.92	332.4	350.91	368.79	347.18
		N	336	314	364	360	333	314	299	296

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
10	Non-Hispanic Asian	Mean	400.67	397.24	367.87	362.63	382.35	382.55	398.74	382.26
		N	364	351	383	376	357	351	336	330
	Non-Hispanic Pacific Islander	Mean	374	370.54	353.27	339.6	355.29	362.69	371.38	360.77
		N	14	13	15	15	14	13	13	13
	Non-Hispanic Black	Mean	383.25	385.22	352.01	360.72	373.45	369.19	384.94	370.52
		N	1199	1053	1284	1275	1188	1052	995	985
	Hispanic (of Any Race)	Mean	379.43	381.8	349.92	339.72	361	366.37	381.51	365.03
		N	8999	8364	9537	9438	8903	8363	8036	7954
	Non-Hispanic American Indian	Mean	396.32	389.42	348.83	350.4	371.85	367.56	389.94	366.51
		N	40	36	42	42	40	36	35	35
	Non-Hispanic Multiracial	Mean	396.76	391.86	364.49	375	390.81	379.78	397.12	387.03
		N	37	36	41	41	37	36	33	33
	Non-Hispanic White	Mean	400.21	396.47	365.04	366.43	384.83	381.44	398.43	382.77
		N	677	634	720	707	664	634	609	597
	Unknown	Mean	378.63	384.93	343.92	328.43	357.4	364.74	383.16	362.94
		N	181	182	206	202	178	182	167	164
11	Non-Hispanic Asian	Mean	404.81	402.49	375.28	372.12	390.41	389.78	404.73	391
		N	323	303	340	333	317	303	292	287
	Non-Hispanic Pacific Islander	Mean	385.1	381.42	359.48	371.14	381.4	369.63	382	371.78
		N	20	19	21	21	20	19	18	18
	Non-Hispanic Black	Mean	383.94	388.87	353.87	360.67	373.25	371.99	387.67	372.51
		N	1147	1033	1250	1234	1131	1033	975	962
	Hispanic (of Any Race)	Mean	382.61	387.55	355.06	345.93	365.43	371.85	386.43	370.2
		N	7878	7415	8330	8255	7801	7414	7127	7058
	Non-Hispanic American Indian	Mean	389.87	387.75	354.41	341.48	370.96	371.81	390	373.14
		N	46	48	51	50	45	48	44	44
	Non-Hispanic Multiracial	Mean	383.4	385.4	359.59	362.09	372.55	372.22	384.24	372.38
		N	43	40	46	45	42	40	38	37
	Non-Hispanic White	Mean	398.86	398.19	364.36	369.05	385.77	382.72	398.77	383.66
		N	624	589	657	650	618	589	572	567
	Unknown	Mean	378.05	384.15	347.99	326.77	355	367.17	383.64	365.7
		N	115	109	130	129	114	109	100	99

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
12	Non-Hispanic Asian	Mean	408.82	404.06	374.71	374.49	392.19	390.23	406.21	391.05
		N	261	252	271	267	257	252	245	241
	Non-Hispanic Pacific Islander	Mean	390.5	390.77	369.57	351.07	371	380.08	389.92	377.31
		N	14	13	14	14	14	13	13	13
	Non-Hispanic Black	Mean	382.7	387.7	351.78	366.57	375.54	370.59	386.37	372.14
		N	986	915	1061	1051	976	915	866	856
	Hispanic (of Any Race)	Mean	387.75	390.33	359.86	360.02	374.54	375.69	390.05	375.69
		N	5031	4773	5291	5229	4968	4771	4598	4540
	Non-Hispanic American	Mean	395.09	399.22	375.08	357.46	377.57	385.74	398.24	383.57
		N	23	23	26	26	23	23	21	21
	Non-Hispanic Multiracial	Mean	409.42	405.97	379.1	390.29	400.06	393.33	407.73	396.27
		N	31	30	31	31	31	30	30	30
	Non-Hispanic White	Mean	401.13	401.08	368.17	376.81	389.3	385.04	401.88	386.99
		N	393	379	412	406	386	378	364	357
	Unknown	Mean	397.7	397.34	360.84	367.32	383.53	381.62	398.88	383.52
		N	99	90	100	97	96	90	89	86

### 1.2.3 Correlations

The tables in this section show Pearson correlations among the four domain scale scores by grade-level clusters across all tiers, as well as the number of students included in each correlation. Results are provided by grade-level cluster. In Kindergarten, the correlation between Listening and Speaking, and the correlation between Reading and Writing were pronounced. In Grades 1 to 12, the highest correlations were between Listening and Reading and between Reading and Writing in general.

**Table 1.2.3.1**

Correlations Among Scale Scores: K, S503 Paper

<b>Domains</b>	<b>Correlations and N counts</b>	<b>Listening</b>	<b>Reading</b>	<b>Writing</b>	<b>Speaking</b>
<b>Listening</b>	Pearson Correlation	1	0.523	0.571	0.803
	N	241979	241974	241969	241967
<b>Reading</b>	Pearson Correlation		1	0.718	0.481
	N		241977	241972	241966
<b>Writing</b>	Pearson Correlation			1	0.540
	N			241972	241962
<b>Speaking</b>	Pearson Correlation				1
	N				241969

**Table 1.2.3.2**

Correlations Among Scale Scores: Grade 1, S503 Paper

<b>Domains</b>	<b>Correlations and N counts</b>	<b>Listening</b>	<b>Reading</b>	<b>Writing</b>	<b>Speaking</b>
<b>Listening</b>	Pearson Correlation	1	0.497	0.484	0.551
	N	28412	22913	28399	28240
<b>Reading</b>	Pearson Correlation		1	0.516	0.463
	N		25868	25859	25729
<b>Writing</b>	Pearson Correlation			1	0.481
	N			33645	33431
<b>Speaking</b>	Pearson Correlation				1
	N				33450

**Table 1.2.3.3**

Correlations Among Scale Scores: Grade 2, S503 Paper

<b>Domains</b>	<b>Correlations and N counts</b>	<b>Listening</b>	<b>Reading</b>	<b>Writing</b>	<b>Speaking</b>
<b>Listening</b>	Pearson Correlation	1	0.599	0.537	0.571
	N	30027	25383	30022	29834
<b>Reading</b>	Pearson Correlation		1	0.650	0.519
	N		26864	26860	26687
<b>Writing</b>	Pearson Correlation			1	0.525
	N			32461	32239
<b>Speaking</b>	Pearson Correlation				1
	N				32244

**Table 1.2.3.4**

Correlations Among Scale Scores: Grade 3, S503 Paper

<b>Domains</b>	<b>Correlations and N counts</b>	<b>Listening</b>	<b>Reading</b>	<b>Writing</b>	<b>Speaking</b>
<b>Listening</b>	Pearson Correlation	1	0.629	0.523	0.515
	N	29384	25022	29378	29200
<b>Reading</b>	Pearson Correlation		1	0.590	0.525
	N		26242	26239	26088
<b>Writing</b>	Pearson Correlation			1	0.538
	N			31423	31227
<b>Speaking</b>	Pearson Correlation				1
	N				31233

**Table 1.2.3.5**

Correlations Among Scale Scores: Grades 4–5, S503 Paper

<b>Domains</b>	<b>Correlations and N counts</b>	<b>Listening</b>	<b>Reading</b>	<b>Writing</b>	<b>Speaking</b>
<b>Listening</b>	Pearson Correlation	1	0.714	0.611	0.606
	N	45263	40204	45259	44921
<b>Reading</b>	Pearson Correlation		1	0.678	0.612
	N		41512	41508	41200
<b>Writing</b>	Pearson Correlation			1	0.626
	N			47273	46907
<b>Speaking</b>	Pearson Correlation				1
	N				46910

**Table 1.2.3.6**

Correlations Among Scale Scores: Grades 6–8, S503 Paper

<b>Domains</b>	<b>Correlations and N counts</b>	<b>Listening</b>	<b>Reading</b>	<b>Writing</b>	<b>Speaking</b>
<b>Listening</b>	Pearson Correlation	1	0.735	0.714	0.681
	N	44657	39389	44641	44335
<b>Reading</b>	Pearson Correlation		1	0.6934	0.616
	N		40976	40963	40692
<b>Writing</b>	Pearson Correlation			1	0.711
	N			47047	46701
<b>Speaking</b>	Pearson Correlation				1
	N				46721

**Table 1.2.3.7**

Correlations Among Scale Scores: Grades 9–12, S503 Paper

<b>Domains</b>	<b>Correlations and N counts</b>	<b>Listening</b>	<b>Reading</b>	<b>Writing</b>	<b>Speaking</b>
<b>Listening</b>	Pearson Correlation	1	0.758	0.669	0.658
	N	41414	37004	41394	40962
<b>Reading</b>	Pearson Correlation		1	0.696	0.638
	N		38650	38642	38243
<b>Writing</b>	Pearson Correlation			1	0.673
	N			44024	43538
<b>Speaking</b>	Pearson Correlation				1
	N				43564

### 1.3 Proficiency Level Results

Proficiency level results show the distribution of students falling into the six language proficiency levels outlined by the WIDA ELD Standards. The results are presented in eight subsections—four domains and four composites--by count and percentage.

Each table in this section shows either the number or percentage of students classified into each language proficiency level. Results are first presented by grade-level cluster and tier, then by grade and tier, and then by grade alone.

Performance of PL 5 and PL 6 was observed in the descending order of Listening, Reading, Speaking, and Writing. Percentages of PL 5 and 6 across grades ranged 25%-61% (Listening), 8%-28% (Reading), 6%-35% (Speaking), and 0%-2% (Writing). Grades with the highest percentages in PL5 and PL6 in each domain are grades 4-5(Listening), grades 3-5(Reading), Kindergarten(Speaking), and grade 5(Writing).

#### 1.3.1 Domains

##### 1.3.1.1 Listening

##### 1.3.1.1.1 By Cluster by Tier

**Table 1.3.1.1.1**

Proficiency Level by Cluster (Count): Listening, S503 Paper

Cluster	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	82558	22607	19140	12497	32231	72946	241979
1	A	1721	2380	2535	1674	3997	2931	15238
	BC	169	526	3007	1280	3017	5175	13174
2	A	1304	1928	1698	1083	2376	0	8389
	BC	72	1239	3381	4754	3149	9043	21638
3	A	215	1715	1798	1157	855	678	6418
	BC	17	425	2894	4267	5658	9705	22966
4-5	A	803	2944	2357	1345	869	784	9102
	BC	81	672	3287	5347	13405	13369	36161
6-8	A	3833	4771	1902	1029	839	125	12499
	BC	45	801	3925	8238	8502	10647	32158



9–12	A	5769	3509	1863	716	192	0	12049
	BC	615	2870	7588	8349	5328	4615	29365

**Table 1.3.1.1.1.2**

Proficiency Level by Cluster (Percent): Listening, S503 Paper

Cluster	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	34.1%	9.3%	7.9%	5.2%	13.3%	30.1%	100%
1	A	11.3%	15.6%	16.6%	11%	26.2%	19.2%	100%
	BC	1.3%	4%	22.8%	9.7%	22.9%	39.3%	100%
2	A	15.5%	23%	20.2%	12.9%	28.3%	0%	100%
	BC	0.3%	5.7%	15.6%	22%	14.6%	41.8%	100%
3	A	3.3%	26.7%	28%	18%	13.3%	10.6%	100%
	BC	0.1%	1.9%	12.6%	18.6%	24.6%	42.3%	100%
4–5	A	8.8%	32.3%	25.9%	14.8%	9.5%	8.6%	100%
	BC	0.2%	1.9%	9.1%	14.8%	37.1%	37%	100%
6–8	A	30.7%	38.2%	15.2%	8.2%	6.7%	1%	100%
	BC	0.1%	2.5%	12.2%	25.6%	26.4%	33.1%	100%
9–12	A	47.9%	29.1%	15.5%	5.9%	1.6%	0%	100%
	BC	2.1%	9.8%	25.8%	28.4%	18.1%	15.7%	100%



1.3.1.1.2 By Grade by Tier

**Table 1.3.1.1.2.1**

Proficiency Level by Grade (Count): Listening, S503 Paper

Grade	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	82558	22607	19140	12497	32231	72946	241979
1	A	1721	2380	2535	1674	3997	2931	15238
	BC	169	526	3007	1280	3017	5175	13174
2	A	1304	1928	1698	1083	2376	0	8389
	BC	72	1239	3381	4754	3149	9043	21638
3	A	215	1715	1798	1157	855	678	6418
	BC	17	425	2894	4267	5658	9705	22966
4	A	304	1478	1248	697	558	340	4625
	BC	30	295	1638	2650	6527	6826	17966
5	A	499	1466	1109	648	311	444	4477
	BC	51	377	1649	2697	6878	6543	18195
6	A	1088	1812	602	442	267	56	4267
	BC	11	217	1098	3249	3058	3972	11605
7	A	1424	1320	804	191	323	69	4131
	BC	21	355	1316	3129	2753	2983	10557
8	A	1321	1639	496	396	249	0	4101
	BC	13	229	1511	1860	2691	3692	9996
9	A	2018	1570	632	331	74	0	4625
	BC	34	672	1581	2600	1664	1693	8244
10	A	1669	745	665	120	85	0	3284
	BC	91	617	2244	2540	1443	1292	8227
11	A	1338	901	355	162	21	0	2777
	BC	203	732	2323	1559	1407	1195	7419

12	A	744	293	211	103	12	0	1363
	BC	287	849	1440	1650	814	435	5475

**Table 1.3.1.1.2.2**

Proficiency Level by Grade (Percent): Listening, S503 Paper

Grade	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	34.1%	9.3%	7.9%	5.2%	13.3%	30.1%	100%
1	A	11.3%	15.6%	16.6%	11%	26.2%	19.2%	100%
	BC	1.3%	4%	22.8%	9.7%	22.9%	39.3%	100%
2	A	15.5%	23%	20.2%	12.9%	28.3%	0%	100%
	BC	0.3%	5.7%	15.6%	22%	14.6%	41.8%	100%
3	A	3.3%	26.7%	28%	18%	13.3%	10.6%	100%
	BC	0.1%	1.9%	12.6%	18.6%	24.6%	42.3%	100%
4	A	6.6%	32%	27%	15.1%	12.1%	7.4%	100%
	BC	0.2%	1.6%	9.1%	14.8%	36.3%	38%	100%
5	A	11.1%	32.7%	24.8%	14.5%	6.9%	9.9%	100%
	BC	0.3%	2.1%	9.1%	14.8%	37.8%	36%	100%
6	A	25.5%	42.5%	14.1%	10.4%	6.3%	1.3%	100%
	BC	0.1%	1.9%	9.5%	28%	26.4%	34.2%	100%
7	A	34.5%	32%	19.5%	4.6%	7.8%	1.7%	100%
	BC	0.2%	3.4%	12.5%	29.6%	26.1%	28.3%	100%
8	A	32.2%	40%	12.1%	9.7%	6.1%	0%	100%
	BC	0.1%	2.3%	15.1%	18.6%	26.9%	36.9%	100%
9	A	43.6%	33.9%	13.7%	7.2%	1.6%	0%	100%
	BC	0.4%	8.2%	19.2%	31.5%	20.2%	20.5%	100%
10	A	50.8%	22.7%	20.2%	3.7%	2.6%	0%	100%
	BC	1.1%	7.5%	27.3%	30.9%	17.5%	15.7%	100%
11	A	48.2%	32.4%	12.8%	5.8%	0.8%	0%	100%
	BC	2.7%	9.9%	31.3%	21%	19%	16.1%	100%
12	A	54.6%	21.5%	15.5%	7.6%	0.9%	0%	100%
	BC	5.2%	15.5%	26.3%	30.1%	14.9%	7.9%	100%



1.3.1.1.3 By Grade

**Table 1.3.1.1.3.1**

Proficiency Level by Grade (Count): Listening, S503 Paper

Grade	Listening Proficiency Range						Total
	1	2	3	4	5	6	
K	82558	22607	19140	12497	32231	72946	241979
1	1890	2906	5542	2954	7014	8106	28412
2	1376	3167	5079	5837	5525	9043	30027
3	232	2140	4692	5424	6513	10383	29384
4	334	1773	2886	3347	7085	7166	22591
5	550	1843	2758	3345	7189	6987	22672
6	1099	2029	1700	3691	3325	4028	15872
7	1445	1675	2120	3320	3076	3052	14688
8	1334	1868	2007	2256	2940	3692	14097
9	2052	2242	2213	2931	1738	1693	12869
10	1760	1362	2909	2660	1528	1292	11511
11	1541	1633	2678	1721	1428	1195	10196
12	1031	1142	1651	1753	826	435	6838

**Table 1.3.1.1.3.2**

Proficiency Level by Grade (Percent): Listening, S503 Paper

Grade	Listening Proficiency Range						Total
	1	2	3	4	5	6	
K	34.1%	9.3%	7.9%	5.2%	13.3%	30.1%	100%
1	6.7%	10.2%	19.5%	10.4%	24.7%	28.5%	100%
2	4.6%	10.5%	16.9%	19.4%	18.4%	30.1%	100%
3	0.8%	7.3%	16%	18.5%	22.2%	35.3%	100%
4	1.5%	7.8%	12.8%	14.8%	31.4%	31.7%	100%
5	2.4%	8.1%	12.2%	14.8%	31.7%	30.8%	100%
6	6.9%	12.8%	10.7%	23.3%	20.9%	25.4%	100%
7	9.8%	11.4%	14.4%	22.6%	20.9%	20.8%	100%
8	9.5%	13.3%	14.2%	16%	20.9%	26.2%	100%
9	15.9%	17.4%	17.2%	22.8%	13.5%	13.2%	100%
10	15.3%	11.8%	25.3%	23.1%	13.3%	11.2%	100%

11	15.1%	16%	26.3%	16.9%	14%	11.7%	100%
12	15.1%	16.7%	24.1%	25.6%	12.1%	6.4%	100%

### 1.3.1.2 Reading

#### 1.3.1.2.1 By Cluster by Tier

**Table 1.3.1.2.1.1**

Proficiency Level by Cluster (Count): Reading, S503 Paper

Cluster	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	190408	5718	15534	10739	19578	0	241977
1	A	6836	4615	1523	423	659	466	14522
	BC	87	2096	4601	1884	1391	1287	11346
2	A	4595	1546	714	668	390	0	7913
	BC	946	6149	3692	1881	2957	3326	18951
3	A	3258	1399	497	156	284	197	5791
	BC	111	2013	6271	4562	5699	1795	20451
4–5	A	5145	1544	1012	310	452	32	8495
	BC	366	6013	10605	4770	7042	4221	33017
6–8	A	6465	3707	1132	307	344	153	12108
	BC	950	9754	8997	3640	3909	1618	28868
9–12	A	4604	4704	1404	445	561	234	11952
	BC	585	6492	7648	2857	4795	4321	26698

**Table 1.3.1.2.1.2**

Proficiency Level by Cluster (Percent): Reading, S503 Paper

Cluster	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	78.7%	2.4%	6.4%	4.4%	8.1%	0%	100%
1	A	47.1%	31.8%	10.5%	2.9%	4.5%	3.2%	100%
	BC	0.8%	18.5%	40.6%	16.6%	12.3%	11.3%	100%
2	A	58.1%	19.5%	9%	8.4%	4.9%	0%	100%
	BC	5%	32.4%	19.5%	9.9%	15.6%	17.6%	100%
3	A	56.3%	24.2%	8.6%	2.7%	4.9%	3.4%	100%
	BC	0.5%	9.8%	30.7%	22.3%	27.9%	8.8%	100%
4–5	A	60.6%	18.2%	11.9%	3.6%	5.3%	0.4%	100%
	BC	1.1%	18.2%	32.1%	14.4%	21.3%	12.8%	100%
	A	53.4%	30.6%	9.3%	2.5%	2.8%	1.3%	100%



6-8	BC	3.3%	33.8%	31.2%	12.6%	13.5%	5.6%	100%
9-12	A	38.5%	39.4%	11.7%	3.7%	4.7%	2%	100%
	BC	2.2%	24.3%	28.6%	10.7%	18%	16.2%	100%

1.3.1.2.2. By Grade by Tier

**Table 1.3.1.2.2.1**

Proficiency Level by Grade (Count): Reading, S503 Paper

Grade	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	190408	5718	15534	10739	19578	NA	241977
1	A	6836	4615	1523	423	659	466	14522
	BC	87	2096	4601	1884	1391	1287	11346
2	A	4595	1546	714	668	390	NA	7913
	BC	946	6149	3692	1881	2957	3326	18951
3	A	3258	1399	497	156	284	197	5791
	BC	111	2013	6271	4562	5699	1795	20451
4	A	2581	852	450	154	223	32	4292
	BC	135	2456	5749	2490	3477	2022	16329
5	A	2564	692	562	156	229	NA	4203
	BC	231	3557	4856	2280	3565	2199	16688
6	A	1931	1501	421	87	104	43	4087
	BC	212	3627	3298	1253	1446	425	10261
7	A	2263	1101	406	98	79	65	4012
	BC	333	3095	3300	1132	1104	525	9489
8	A	2271	1105	305	122	161	45	4009
	BC	405	3032	2399	1255	1359	668	9118
9	A	1755	1849	561	133	221	79	4598
	BC	69	1662	2246	900	1154	1321	7352
10	A	1224	1298	431	119	104	99	3275
	BC	125	1723	2122	881	1268	1275	7394
11	A	1066	1045	282	129	191	32	2745
	BC	213	1547	1871	776	1211	1193	6811
12	A	559	512	130	64	45	24	1334
	BC	178	1560	1409	300	1162	532	5141



**Table 1.3.1.2.2.2**

Proficiency Level by Grade (Percent): Reading, S503 Paper

Grade	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	78.7%	2.4%	6.4%	4.4%	8.1%	0%	100%
1	A	47.1%	31.8%	10.5%	2.9%	4.5%	3.2%	100%
	BC	0.8%	18.5%	40.6%	16.6%	12.3%	11.3%	100%
2	A	58.1%	19.5%	9%	8.4%	4.9%	0%	100%
	BC	5%	32.4%	19.5%	9.9%	15.6%	17.6%	100%
3	A	56.3%	24.2%	8.6%	2.7%	4.9%	3.4%	100%
	BC	0.5%	9.8%	30.7%	22.3%	27.9%	8.8%	100%
4	A	60.1%	19.9%	10.5%	3.6%	5.2%	0.7%	100%
	BC	0.8%	15%	35.2%	15.2%	21.3%	12.4%	100%
5	A	61%	16.5%	13.4%	3.7%	5.4%	0%	100%
	BC	1.4%	21.3%	29.1%	13.7%	21.4%	13.2%	100%
6	A	47.2%	36.7%	10.3%	2.1%	2.5%	1.1%	100%
	BC	2.1%	35.3%	32.1%	12.2%	14.1%	4.1%	100%
7	A	56.4%	27.4%	10.1%	2.4%	2%	1.6%	100%
	BC	3.5%	32.6%	34.8%	11.9%	11.6%	5.5%	100%
8	A	56.6%	27.6%	7.6%	3%	4%	1.1%	100%
	BC	4.4%	33.3%	26.3%	13.8%	14.9%	7.3%	100%
9	A	38.2%	40.2%	12.2%	2.9%	4.8%	1.7%	100%
	BC	0.9%	22.6%	30.5%	12.2%	15.7%	18%	100%
10	A	37.4%	39.6%	13.2%	3.6%	3.2%	3%	100%
	BC	1.7%	23.3%	28.7%	11.9%	17.1%	17.2%	100%
11	A	38.8%	38.1%	10.3%	4.7%	7%	1.2%	100%
	BC	3.1%	22.7%	27.5%	11.4%	17.8%	17.5%	100%
12	A	41.9%	38.4%	9.7%	4.8%	3.4%	1.8%	100%
	BC	3.5%	30.3%	27.4%	5.8%	22.6%	10.3%	100%

1.3.1.2.3. By Grade

**Table 1.3.1.2.3.1**

Proficiency Level by Grade (Count): Reading, S503 Paper

Grade	Reading Proficiency Range						Total
	1	2	3	4	5	6	
K	190408	5718	15534	10739	19578	0	241977
1	6923	6711	6124	2307	2050	1753	25868
2	5541	7695	4406	2549	3347	3326	26864
3	3369	3412	6768	4718	5983	1992	26242
4	2716	3308	6199	2644	3700	2054	20621
5	2795	4249	5418	2436	3794	2199	20891
6	2143	5128	3719	1340	1550	468	14348
7	2596	4196	3706	1230	1183	590	13501
8	2676	4137	2704	1377	1520	713	13127
9	1824	3511	2807	1033	1375	1400	11950
10	1349	3021	2553	1000	1372	1374	10669
11	1279	2592	2153	905	1402	1225	9556
12	737	2072	1539	364	1207	556	6475

**Table 1.3.1.2.3.2**

Proficiency Level by Grade (Percent): Reading, S503 Paper

Grade	Reading Proficiency Range						Total
	1	2	3	4	5	6	
K	78.7%	2.4%	6.4%	4.4%	8.1%	0%	100%
1	26.8%	25.9%	23.7%	8.9%	7.9%	6.8%	100%
2	20.6%	28.6%	16.4%	9.5%	12.5%	12.4%	100%
3	12.8%	13%	25.8%	18%	22.8%	7.6%	100%
4	13.2%	16%	30.1%	12.8%	17.9%	10%	100%
5	13.4%	20.3%	25.9%	11.7%	18.2%	10.5%	100%
6	14.9%	35.7%	25.9%	9.3%	10.8%	3.3%	100%
7	19.2%	31.1%	27.4%	9.1%	8.8%	4.4%	100%
8	20.4%	31.5%	20.6%	10.5%	11.6%	5.4%	100%
9	15.3%	29.4%	23.5%	8.6%	11.5%	11.7%	100%
10	12.6%	28.3%	23.9%	9.4%	12.9%	12.9%	100%
11	13.4%	27.1%	22.5%	9.5%	14.7%	12.8%	100%

12	11.4%	32%	23.8%	5.6%	18.6%	8.6%	100%
----	-------	-----	-------	------	-------	------	------

### 1.3.1.3 Writing

#### 1.3.1.3.1 By Cluster by Tier

**Table 1.3.1.3.1.1**

Proficiency Level by Cluster (Count): Writing, S503 Paper

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	177157	35127	23131	6557	0	0	241972
1	A	9577	8289	611	0	0	0	18477
	BC	3211	4914	6696	343	4	0	15168
2	A	4338	2882	2117	4	0	0	9341
	BC	2226	6494	13743	657	0	0	23120
3	A	2877	2600	1477	15	0	0	6969
	BC	1280	3832	18502	838	1	1	24454
4–5	A	3740	2337	3678	38	0	0	9793
	BC	872	1534	21549	12780	718	27	37480
6–8	A	6647	4549	2220	12	0	0	13428
	BC	1038	2138	24401	6039	3	0	33619
9–12	A	5404	4131	3372	253	1	0	13161
	BC	2092	2461	15371	10675	262	2	30863

**Table 1.3.1.3.1.2**

Proficiency Level by Cluster (Percent): Writing, S503 Paper

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	73.2%	14.5%	9.6%	2.7%	0%	0%	100%
1	A	51.8%	44.9%	3.3%	0%	0%	0%	100%
	BC	21.2%	32.4%	44.1%	2.3%	0%	0%	100%
2	A	46.4%	30.9%	22.7%	0%	0%	0%	100%
	BC	9.6%	28.1%	59.4%	2.8%	0%	0%	100%
3	A	41.3%	37.3%	21.2%	0.2%	0%	0%	100%
	BC	5.2%	15.7%	75.7%	3.4%	0%	0%	100%
	A	38.2%	23.9%	37.6%	0.4%	0%	0%	100%

4-5	BC	2.3%	4.1%	57.5%	34.1%	1.9%	0.1%	100%
6-8	A	49.5%	33.9%	16.5%	0.1%	0%	0%	100%
	BC	3.1%	6.4%	72.6%	18%	0%	0%	100%
9-12	A	41.1%	31.4%	25.6%	1.9%	0%	0%	100%
	BC	6.8%	8%	49.8%	34.6%	0.8%	0%	100%

1.3.1.3.2 By Grade by Tier

**Table 1.3.1.3.2.1**

Proficiency Level by Grade (Count): Writing, S503 Paper

Grade	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	177157	35127	23131	6557	0	0	241972
1	A	9577	8289	611	0	0	0	18477
	BC	3211	4914	6696	343	4	0	15168
2	A	4338	2882	2117	4	0	0	9341
	BC	2226	6494	13743	657	0	0	23120
3	A	2877	2600	1477	15	0	0	6969
	BC	1280	3832	18502	838	1	1	24454
4	A	2076	1192	1718	22	0	0	5008
	BC	522	836	11698	5446	205	10	18717
5	A	1664	1145	1960	16	0	0	4785
	BC	350	698	9851	7334	513	17	18763
6	A	2121	1555	886	2	0	0	4564
	BC	300	813	8617	2403	1	0	12134
7	A	2185	1672	575	7	0	0	4439
	BC	334	805	8078	1817	1	0	11035
8	A	2341	1322	759	3	0	0	4425
	BC	404	520	7706	1819	1	0	10450
9	A	2034	1583	1343	156	1	0	5117
	BC	315	645	3776	3812	99	1	8648
10	A	1306	1285	947	49	0	0	3587
	BC	539	512	4540	2974	75	1	8641
11	A	1313	976	691	27	0	0	3007
	BC	596	596	4063	2484	79	0	7818
12	A	751	287	391	21	0	0	1450
	BC	642	708	2992	1405	9	0	5756

**Table 1.3.1.3.2.2**

Proficiency Level by Grade (Percent): Writing, S503 Paper

Grade	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	73.2%	14.5%	9.6%	2.7%	0%	0%	100%
1	A	51.8%	44.9%	3.3%	0%	0%	0%	100%
	BC	21.2%	32.4%	44.1%	2.3%	0%	0%	100%
2	A	46.4%	30.9%	22.7%	0%	0%	0%	100%
	BC	9.6%	28.1%	59.4%	2.8%	0%	0%	100%
3	A	41.3%	37.3%	21.2%	0.2%	0%	0%	100%
	BC	5.2%	15.7%	75.7%	3.4%	0%	0%	100%
4	A	41.5%	23.8%	34.3%	0.4%	0%	0%	100%
	BC	2.8%	4.5%	62.5%	29.1%	1.1%	0.1%	100%
5	A	34.8%	23.9%	41%	0.3%	0%	0%	100%
	BC	1.9%	3.7%	52.5%	39.1%	2.7%	0.1%	100%
6	A	46.5%	34.1%	19.4%	0%	0%	0%	100%
	BC	2.5%	6.7%	71%	19.8%	0%	0%	100%
7	A	49.2%	37.7%	13%	0.2%	0%	0%	100%
	BC	3%	7.3%	73.2%	16.5%	0%	0%	100%
8	A	52.9%	29.9%	17.2%	0.1%	0%	0%	100%
	BC	3.9%	5%	73.7%	17.4%	0%	0%	100%
9	A	39.7%	30.9%	26.2%	3%	0%	0%	100%
	BC	3.6%	7.5%	43.7%	44.1%	1.1%	0%	100%
10	A	36.4%	35.8%	26.4%	1.4%	0%	0%	100%
	BC	6.2%	5.9%	52.5%	34.4%	0.9%	0%	100%
11	A	43.7%	32.5%	23%	0.9%	0%	0%	100%
	BC	7.6%	7.6%	52%	31.8%	1%	0%	100%
12	A	51.8%	19.8%	27%	1.4%	0%	0%	100%
	BC	11.2%	12.3%	52%	24.4%	0.2%	0%	100%



1.3.1.3.3 By Grade

**Table 1.3.1.3.3.1**

Proficiency Level by Grade (Count): Writing, S503 Paper

Grade	Writing Proficiency Range						Total
	1	2	3	4	5	6	
K	177157	35127	23131	6557	0	0	241972
1	12788	13203	7307	343	4	0	33645
2	6564	9376	15860	661	0	0	32461
3	4157	6432	19979	853	1	1	31423
4	2598	2028	13416	5468	205	10	23725
5	2014	1843	11811	7350	513	17	23548
6	2421	2368	9503	2405	1	0	16698
7	2519	2477	8653	1824	1	0	15474
8	2745	1842	8465	1822	1	0	14875
9	2349	2228	5119	3968	100	1	13765
10	1845	1797	5487	3023	75	1	12228
11	1909	1572	4754	2511	79	0	10825
12	1393	995	3383	1426	9	0	7206

**Table 1.3.1.3.3.2**

Proficiency Level by Grade (Percent): Writing, S503 Paper

Grade	Writing Proficiency Range						Total
	1	2	3	4	5	6	
K	73.2%	14.5%	9.6%	2.7%	0%	0%	100%
1	38%	39.2%	21.7%	1%	0%	0%	100%
2	20.2%	28.9%	48.9%	2%	0%	0%	100%
3	13.2%	20.5%	63.6%	2.7%	0%	0%	100%
4	11%	8.5%	56.5%	23%	0.9%	0%	100%
5	8.6%	7.8%	50.2%	31.2%	2.2%	0.1%	100%
6	14.5%	14.2%	56.9%	14.4%	0%	0%	100%
7	16.3%	16%	55.9%	11.8%	0%	0%	100%
8	18.5%	12.4%	56.9%	12.2%	0%	0%	100%

9	17.1%	16.2%	37.2%	28.8%	0.7%	0%	100%
10	15.1%	14.7%	44.9%	24.7%	0.6%	0%	100%
11	17.6%	14.5%	43.9%	23.2%	0.7%	0%	100%
12	19.3%	13.8%	46.9%	19.8%	0.1%	0%	100%

#### 1.3.1.4 Speaking

##### 1.3.1.4.1 By Cluster by Tier

**Table 1.3.1.4.1.1**

Proficiency Level by Cluster (Count): Speaking, S503 Paper

Cluster	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	73135	49674	16594	18543	24352	59671	241969
1	A	5299	5969	4069	2584	449	0	18370
	BC	401	4393	4955	3504	1521	306	15080
2	A	3653	2169	2564	662	227	0	9275
	BC	1320	4541	9245	5280	1619	964	22969
3	A	3309	1663	1311	642	0	0	6925
	BC	1083	4722	10401	5687	1031	1384	24308
4–5	A	5911	1757	1450	434	151	0	9703
	BC	1001	4293	10967	14209	4669	2068	37207
6–8	A	8214	2504	1349	1022	239	0	13328
	BC	1475	3130	9265	12513	4135	2875	33393
9–12	A	9019	1450	1842	609	83	0	13003
	BC	3449	4146	10945	7300	1572	3149	30561

**Table 1.3.1.4.1.2**

Proficiency Level by Cluster (Percent): Speaking, S503 Paper

Cluster	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	

K	-	30.2%	20.5%	6.9%	7.7%	10.1%	24.7%	100%
1	A	28.8%	32.5%	22.2%	14.1%	2.4%	0%	100%
	BC	2.7%	29.1%	32.9%	23.2%	10.1%	2%	100%
2	A	39.4%	23.4%	27.6%	7.1%	2.4%	0%	100%
	BC	5.7%	19.8%	40.2%	23%	7%	4.2%	100%
3	A	47.8%	24%	18.9%	9.3%	0%	0%	100%
	BC	4.5%	19.4%	42.8%	23.4%	4.2%	5.7%	100%
4–5	A	60.9%	18.1%	14.9%	4.5%	1.6%	0%	100%
	BC	2.7%	11.5%	29.5%	38.2%	12.5%	5.6%	100%
6–8	A	61.6%	18.8%	10.1%	7.7%	1.8%	0%	100%
	BC	4.4%	9.4%	27.7%	37.5%	12.4%	8.6%	100%
9–12	A	69.4%	11.2%	14.2%	4.7%	0.6%	0%	100%
	BC	11.3%	13.6%	35.8%	23.9%	5.1%	10.3%	100%

1.3.1.4.2 By Grade by Tier

**Table 1.3.1.4.2.1**

Proficiency Level by Grade (Count): Speaking, S503 Paper

Grade	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	73135	49674	16594	18543	24352	59671	241969
1	A	5299	5969	4069	2584	449	0	18370
	BC	401	4393	4955	3504	1521	306	15080
2	A	3653	2169	2564	662	227	0	9275
	BC	1320	4541	9245	5280	1619	964	22969
3	A	3309	1663	1311	642	0	0	6925
	BC	1083	4722	10401	5687	1031	1384	24308
4	A	2829	941	898	225	72	0	4965
	BC	423	1870	5605	7140	2667	859	18564
5	A	3082	816	552	209	79	0	4738
	BC	578	2423	5362	7069	2002	1209	18643
6	A	2594	1075	455	283	123	0	4530
	BC	378	1090	3029	5105	1186	1256	12044
7	A	2876	702	441	337	48	0	4404
	BC	542	1213	2915	3830	1758	709	10967
8	A	2744	727	453	402	68	0	4394
	BC	555	827	3321	3578	1191	910	10382
9	A	3820	441	591	134	83	0	5069
	BC	890	1116	2728	2584	406	837	8561
10	A	2486	325	510	209	0	0	3530
	BC	923	1097	3299	1970	442	835	8566
	A	1909	365	507	190	0	0	2971

11	BC	979	1141	2647	1730	422	827	7746
12	A	804	319	234	76	0	0	1433
	BC	657	792	2271	1016	302	650	5688

**Table 1.3.1.4.2.2**

Proficiency Level by Grade (Percent): Speaking, S503 Paper

Grade	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	30.2%	20.5%	6.9%	7.7%	10.1%	24.7%	100%
1	A	28.8%	32.5%	22.2%	14.1%	2.4%	0%	100%
	BC	2.7%	29.1%	32.9%	23.2%	10.1%	2%	100%
2	A	39.4%	23.4%	27.6%	7.1%	2.4%	0%	100%
	BC	5.7%	19.8%	40.2%	23%	7%	4.2%	100%
3	A	47.8%	24%	18.9%	9.3%	0%	0%	100%
	BC	4.5%	19.4%	42.8%	23.4%	4.2%	5.7%	100%
4	A	57%	19%	18.1%	4.5%	1.5%	0%	100%
	BC	2.3%	10.1%	30.2%	38.5%	14.4%	4.6%	100%
5	A	65%	17.2%	11.7%	4.4%	1.7%	0%	100%
	BC	3.1%	13%	28.8%	37.9%	10.7%	6.5%	100%
6	A	57.3%	23.7%	10%	6.2%	2.7%	0%	100%
	BC	3.1%	9.1%	25.1%	42.4%	9.8%	10.4%	100%
7	A	65.3%	15.9%	10%	7.7%	1.1%	0%	100%
	BC	4.9%	11.1%	26.6%	34.9%	16%	6.5%	100%
8	A	62.4%	16.5%	10.3%	9.1%	1.5%	0%	100%
	BC	5.3%	8%	32%	34.5%	11.5%	8.8%	100%
9	A	75.4%	8.7%	11.7%	2.6%	1.6%	0%	100%
	BC	10.4%	13%	31.9%	30.2%	4.7%	9.8%	100%
10	A	70.4%	9.2%	14.4%	5.9%	0%	0%	100%
	BC	10.8%	12.8%	38.5%	23%	5.2%	9.7%	100%
11	A	64.3%	12.3%	17.1%	6.4%	0%	0%	100%
	BC	12.6%	14.7%	34.2%	22.3%	5.4%	10.7%	100%
12	A	56.1%	22.3%	16.3%	5.3%	0%	0%	100%
	BC	11.6%	13.9%	39.9%	17.9%	5.3%	11.4%	100%

1.3.1.4.3 By Grade

**Table 1.3.1.4.3.1**

Proficiency Level by Grade (Count): Speaking, S503 Paper

Grade	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
K	73135	49674	16594	18543	24352	59671	241969
1	5700	10362	9024	6088	1970	306	33450
2	4973	6710	11809	5942	1846	964	32244
3	4392	6385	11712	6329	1031	1384	31233
4	3252	2811	6503	7365	2739	859	23529
5	3660	3239	5914	7278	2081	1209	23381
6	2972	2165	3484	5388	1309	1256	16574
7	3418	1915	3356	4167	1806	709	15371
8	3299	1554	3774	3980	1259	910	14776
9	4710	1557	3319	2718	489	837	13630
10	3409	1422	3809	2179	442	835	12096
11	2888	1506	3154	1920	422	827	10717
12	1461	1111	2505	1092	302	650	7121

**Table 1.3.1.4.3.2**

Proficiency Level by Grade (Percent): Speaking, S503 Paper

Grade	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
K	30.2%	20.5%	6.9%	7.7%	10.1%	24.7%	100%
1	17%	31%	27%	18.2%	5.9%	0.9%	100%
2	15.4%	20.8%	36.6%	18.4%	5.7%	3%	100%
3	14.1%	20.4%	37.5%	20.3%	3.3%	4.4%	100%
4	13.8%	11.9%	27.6%	31.3%	11.6%	3.7%	100%
5	15.7%	13.9%	25.3%	31.1%	8.9%	5.2%	100%
6	17.9%	13.1%	21%	32.5%	7.9%	7.6%	100%
7	22.2%	12.5%	21.8%	27.1%	11.7%	4.6%	100%
8	22.3%	10.5%	25.5%	26.9%	8.5%	6.2%	100%
9	34.6%	11.4%	24.4%	19.9%	3.6%	6.1%	100%
10	28.2%	11.8%	31.5%	18%	3.7%	6.9%	100%
11	26.9%	14.1%	29.4%	17.9%	3.9%	7.7%	100%

12	20.5%	15.6%	35.2%	15.3%	4.2%	9.1%	100%
----	-------	-------	-------	-------	------	------	------

### 1.3.2 Composites

Performance of composites is observed in their percentage in PL5 and 6 in grades: Comprehension (10-38%), Oral (11-27%), Overall (1-8%), and Literacy (0-5%). In Literacy and Overall, there are fewer students in PL 5 and 6 than Comprehension and Oral.

#### 1.3.2.1 Oral

##### 1.3.2.1.1 By Cluster by Tier

**Table 1.3.2.1.1.1**

Proficiency Level by Cluster (Count): Oral, S503 Paper

Cluster	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	80866	34059	26746	20675	36484	43137	241967
1	A	2716	3507	5168	2610	1026	116	15143
	BC	118	1458	4378	3971	2375	797	13097
2	A	2292	2072	2482	1266	222	0	8334
	BC	183	1955	7161	7715	3634	852	21500
3	A	1881	1804	1636	834	220	0	6375
	BC	69	1377	7174	9161	3994	1050	22825
4-5	A	3655	2521	1710	927	195	12	9020
	BC	115	1393	8217	14176	9147	2853	35901
6-8	A	6570	2853	1807	984	183	8	12405
	BC	190	1496	6908	13251	7406	2679	31930
9-12	A	7324	2375	1601	569	26	0	11895
	BC	974	3657	10005	9809	3554	1068	29067

**Table 1.3.2.1.1.2**

Proficiency Level by Cluster (Percent): Oral, S503 Paper

Cluster	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	33.4%	14.1%	11.1%	8.5%	15.1%	17.8%	100%
1	A	17.9%	23.2%	34.1%	17.2%	6.8%	0.8%	100%
	BC	0.9%	11.1%	33.4%	30.3%	18.1%	6.1%	100%
2	A	27.5%	24.9%	29.8%	15.2%	2.7%	0%	100%
	BC	0.9%	9.1%	33.3%	35.9%	16.9%	4%	100%
3	A	29.5%	28.3%	25.7%	13.1%	3.5%	0%	100%
	BC	0.3%	6%	31.4%	40.1%	17.5%	4.6%	100%
4-5	A	40.5%	27.9%	19%	10.3%	2.2%	0.1%	100%
	BC	0.3%	3.9%	22.9%	39.5%	25.5%	7.9%	100%
6-8	A	53%	23%	14.6%	7.9%	1.5%	0.1%	100%
	BC	0.6%	4.7%	21.6%	41.5%	23.2%	8.4%	100%
9-12	A	61.6%	20%	13.5%	4.8%	0.2%	0%	100%
	BC	3.4%	12.6%	34.4%	33.7%	12.2%	3.7%	100%



1.3.2.1.2 By Grade by Tier

**Table 1.3.2.1.2.1**

Proficiency Level by Grade (Count): Oral, S503 Paper

Grade	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	80866	34059	26746	20675	36484	43137	241967
1	A	2716	3507	5168	2610	1026	116	15143
	BC	118	1458	4378	3971	2375	797	13097
2	A	2292	2072	2482	1266	222	0	8334
	BC	183	1955	7161	7715	3634	852	21500
3	A	1881	1804	1636	834	220	0	6375
	BC	69	1377	7174	9161	3994	1050	22825
4	A	1696	1377	882	505	117	12	4589
	BC	49	613	4069	6917	4572	1600	17820
5	A	1959	1144	828	422	78	0	4431
	BC	66	780	4148	7259	4575	1253	18081
6	A	2029	1134	616	362	84	8	4233
	BC	39	411	2353	4790	2861	1060	11514
7	A	2265	867	604	310	55	0	4101
	BC	67	572	2385	4263	2381	817	10485
8	A	2276	852	587	312	44	0	4071
	BC	84	513	2170	4198	2164	802	9931
9	A	2881	925	536	221	17	0	4580
	BC	156	926	2482	3109	1114	375	8162
10	A	2004	609	445	167	4	0	3229
	BC	275	1034	2668	2854	1013	308	8152
11	A	1671	553	399	111	5	0	2739
	BC	300	950	2718	2201	901	279	7349
12	A	768	288	221	70	0	0	1347
	BC	243	747	2137	1645	526	106	5404

**Table 1.3.2.1.2.2**

Proficiency Level by Grade (Percent): Oral, S503 Paper

Grade	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	33.4%	14.1%	11.1%	8.5%	15.1%	17.8%	100%
1	A	17.9%	23.2%	34.1%	17.2%	6.8%	0.8%	100%
	BC	0.9%	11.1%	33.4%	30.3%	18.1%	6.1%	100%
2	A	27.5%	24.9%	29.8%	15.2%	2.7%	0%	100%
	BC	0.9%	9.1%	33.3%	35.9%	16.9%	4%	100%
3	A	29.5%	28.3%	25.7%	13.1%	3.5%	0%	100%
	BC	0.3%	6%	31.4%	40.1%	17.5%	4.6%	100%
4	A	37%	30%	19.2%	11%	2.5%	0.3%	100%
	BC	0.3%	3.4%	22.8%	38.8%	25.7%	9%	100%
5	A	44.2%	25.8%	18.7%	9.5%	1.8%	0%	100%
	BC	0.4%	4.3%	22.9%	40.1%	25.3%	6.9%	100%
6	A	47.9%	26.8%	14.6%	8.6%	2%	0.2%	100%
	BC	0.3%	3.6%	20.4%	41.6%	24.8%	9.2%	100%
7	A	55.2%	21.1%	14.7%	7.6%	1.3%	0%	100%
	BC	0.6%	5.5%	22.7%	40.7%	22.7%	7.8%	100%
8	A	55.9%	20.9%	14.4%	7.7%	1.1%	0%	100%
	BC	0.8%	5.2%	21.9%	42.3%	21.8%	8.1%	100%
9	A	62.9%	20.2%	11.7%	4.8%	0.4%	0%	100%
	BC	1.9%	11.3%	30.4%	38.1%	13.6%	4.6%	100%
10	A	62.1%	18.9%	13.8%	5.2%	0.1%	0%	100%
	BC	3.4%	12.7%	32.7%	35%	12.4%	3.8%	100%
11	A	61%	20.2%	14.6%	4.1%	0.2%	0%	100%
	BC	4.1%	12.9%	37%	29.9%	12.3%	3.8%	100%
12	A	57%	21.4%	16.4%	5.2%	0%	0%	100%
	BC	4.5%	13.8%	39.5%	30.4%	9.7%	2%	100%

1.3.2.1.3 By Grade

**Table 1.3.2.1.3.1**

Proficiency Level by Grade (Count): Oral, S503 Paper

Grade	Oral Language Proficiency Range						Total
	1	2	3	4	5	6	
K	80866	34059	26746	20675	36484	43137	241967
1	2834	4965	9546	6581	3401	913	28240
2	2475	4027	9643	8981	3856	852	29834
3	1950	3181	8810	9995	4214	1050	29200
4	1745	1990	4951	7422	4689	1612	22409
5	2025	1924	4976	7681	4653	1253	22512
6	2068	1545	2969	5152	2945	1068	15747
7	2332	1439	2989	4573	2436	817	14586
8	2360	1365	2757	4510	2208	802	14002
9	3037	1851	3018	3330	1131	375	12742
10	2279	1643	3113	3021	1017	308	11381
11	1971	1503	3117	2312	906	279	10088
12	1011	1035	2358	1715	526	106	6751

**Table 1.3.2.1.3.2**

Proficiency Level by Grade (Percent): Oral, S503 Paper

Grade	Oral Language Proficiency Range						Total
	1	2	3	4	5	6	
K	33.4%	14.1%	11.1%	8.5%	15.1%	17.8%	100%
1	10%	17.6%	33.8%	23.3%	12%	3.2%	100%
2	8.3%	13.5%	32.3%	30.1%	12.9%	2.9%	100%
3	6.7%	10.9%	30.2%	34.2%	14.4%	3.6%	100%
4	7.8%	8.9%	22.1%	33.1%	20.9%	7.2%	100%
5	9%	8.5%	22.1%	34.1%	20.7%	5.6%	100%
6	13.1%	9.8%	18.9%	32.7%	18.7%	6.8%	100%
7	16%	9.9%	20.5%	31.4%	16.7%	5.6%	100%
8	16.9%	9.7%	19.7%	32.2%	15.8%	5.7%	100%
9	23.8%	14.5%	23.7%	26.1%	8.9%	2.9%	100%

10	20%	14.4%	27.4%	26.5%	8.9%	2.7%	100%
11	19.5%	14.9%	30.9%	22.9%	9%	2.8%	100%
12	15%	15.3%	34.9%	25.4%	7.8%	1.6%	100%

### 1.3.2.2 Literacy

#### 1.3.2.2.1 By Cluster by Tier

**Table 1.3.2.2.1.1**

Proficiency Level by Cluster (Count): Literacy, S503 Paper

Cluster	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	191019	23709	19082	8162	0	0	241972
1	A	7174	5493	1842	6	0	0	14515
	BC	1138	3606	5506	968	120	6	11344
2	A	4035	2231	1616	29	0	0	7911
	BC	1300	4817	9920	2709	201	2	18949
3	A	2581	1962	1169	78	0	0	5790
	BC	324	2461	14506	3055	97	6	20449
4-5	A	4081	2254	2006	153	0	0	8494
	BC	340	1592	17471	11715	1699	197	33014
6-8	A	6188	4135	1706	74	1	0	12104
	BC	448	3353	19272	5586	197	3	28859
9-12	A	4747	4252	2572	375	4	0	11950
	BC	737	3254	11804	8993	1846	58	26692

**Table 1.3.2.2.1.2**

Proficiency Level by Cluster (Percent): Literacy, S503 Paper

Cluster	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	78.9%	9.8%	7.9%	3.4%	0%	0%	100%
1	A	49.4%	37.8%	12.7%	0%	0%	0%	100%
	BC	10%	31.8%	48.5%	8.5%	1.1%	0.1%	100%
2	A	51%	28.2%	20.4%	0.4%	0%	0%	100%
	BC	6.9%	25.4%	52.4%	14.3%	1.1%	0%	100%
3	A	44.6%	33.9%	20.2%	1.3%	0%	0%	100%
	BC	1.6%	12%	70.9%	14.9%	0.5%	0%	100%
4-5	A	48%	26.5%	23.6%	1.8%	0%	0%	100%
	BC	1%	4.8%	52.9%	35.5%	5.1%	0.6%	100%

6–8	A	51.1%	34.2%	14.1%	0.6%	0%	0%	100%
	BC	1.6%	11.6%	66.8%	19.4%	0.7%	0%	100%
9–12	A	39.7%	35.6%	21.5%	3.1%	0%	0%	100%
	BC	2.8%	12.2%	44.2%	33.7%	6.9%	0.2%	100%

1.3.2.2.2 By Grade by Tier

**Table 1.3.2.2.1**

Proficiency Level by Grade (Count): Literacy, S503 Paper

Grade	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	191019	23709	19082	8162	0	0	241972
1	A	7174	5493	1842	6	0	0	14515
	BC	1138	3606	5506	968	120	6	11344
2	A	4035	2231	1616	29	0	0	7911
	BC	1300	4817	9920	2709	201	2	18949
3	A	2581	1962	1169	78	0	0	5790
	BC	324	2461	14506	3055	97	6	20449
4	A	2105	1100	1005	82	0	0	4292
	BC	190	776	9294	5376	619	73	16328
5	A	1976	1154	1001	71	0	0	4202
	BC	150	816	8177	6339	1080	124	16686
6	A	2019	1416	625	27	0	0	4087
	BC	123	1134	7100	1840	58	3	10258
7	A	2047	1401	542	21	0	0	4011
	BC	155	1130	6441	1700	59	0	9485
8	A	2122	1318	539	26	1	0	4006
	BC	170	1089	5731	2046	80	0	9116
9	A	1836	1643	944	171	3	0	4597
	BC	100	670	3164	2811	573	33	7351
10	A	1248	1169	750	108	0	0	3275
	BC	133	854	3230	2611	549	15	7392
11	A	1091	975	610	67	1	0	2744
	BC	206	852	2967	2263	513	10	6811
12	A	572	465	268	29	0	0	1334
	BC	298	878	2443	1308	211	0	5138

**Table 1.3.2.2.2.2**

Proficiency Level by Grade (Percent): Literacy, S503 Paper

Grade	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	78.9%	9.8%	7.9%	3.4%	0%	0%	100%
1	A	49.4%	37.8%	12.7%	0%	0%	0%	100%
	BC	10%	31.8%	48.5%	8.5%	1.1%	0.1%	100%
2	A	51%	28.2%	20.4%	0.4%	0%	0%	100%
	BC	6.9%	25.4%	52.4%	14.3%	1.1%	0%	100%
3	A	44.6%	33.9%	20.2%	1.3%	0%	0%	100%
	BC	1.6%	12%	70.9%	14.9%	0.5%	0%	100%
4	A	49%	25.6%	23.4%	1.9%	0%	0%	100%
	BC	1.2%	4.8%	56.9%	32.9%	3.8%	0.4%	100%
5	A	47%	27.5%	23.8%	1.7%	0%	0%	100%
	BC	0.9%	4.9%	49%	38%	6.5%	0.7%	100%
6	A	49.4%	34.6%	15.3%	0.7%	0%	0%	100%
	BC	1.2%	11.1%	69.2%	17.9%	0.6%	0%	100%
7	A	51%	34.9%	13.5%	0.5%	0%	0%	100%
	BC	1.6%	11.9%	67.9%	17.9%	0.6%	0%	100%
8	A	53%	32.9%	13.5%	0.6%	0%	0%	100%
	BC	1.9%	11.9%	62.9%	22.4%	0.9%	0%	100%
9	A	39.9%	35.7%	20.5%	3.7%	0.1%	0%	100%
	BC	1.4%	9.1%	43%	38.2%	7.8%	0.4%	100%
10	A	38.1%	35.7%	22.9%	3.3%	0%	0%	100%
	BC	1.8%	11.6%	43.7%	35.3%	7.4%	0.2%	100%
11	A	39.8%	35.5%	22.2%	2.4%	0%	0%	100%
	BC	3%	12.5%	43.6%	33.2%	7.5%	0.1%	100%
12	A	42.9%	34.9%	20.1%	2.2%	0%	0%	100%
	BC	5.8%	17.1%	47.5%	25.5%	4.1%	0%	100%

1.3.2.2.3 By Grade

**Table 1.3.2.2.3.1**

Proficiency Level by Grade (Count): Literacy, S503 Paper

Grade	Literacy Proficiency Range						Total
	1	2	3	4	5	6	
K	191019	23709	19082	8162	0	0	241972
1	8312	9099	7348	974	120	6	25859
2	5335	7048	11536	2738	201	2	26860
3	2905	4423	15675	3133	97	6	26239
4	2295	1876	10299	5458	619	73	20620
5	2126	1970	9178	6410	1080	124	20888
6	2142	2550	7725	1867	58	3	14345
7	2202	2531	6983	1721	59	0	13496
8	2292	2407	6270	2072	81	0	13122
9	1936	2313	4108	2982	576	33	11948
10	1381	2023	3980	2719	549	15	10667
11	1297	1827	3577	2330	514	10	9555
12	870	1343	2711	1337	211	0	6472

**Table 1.3.2.2.3.2**

Proficiency Level by Grade (Percent): Literacy, S503 Paper

Grade	Literacy Proficiency Range						Total
	1	2	3	4	5	6	
K	78.94%	9.8%	7.89%	3.37%	0%	0%	100%
1	32.14%	35.19%	28.42%	3.77%	0.46%	0.02%	100%
2	19.86%	26.24%	42.95%	10.19%	0.75%	0.01%	100%
3	11.07%	16.86%	59.74%	11.94%	0.37%	0.02%	100%
4	11.13%	9.1%	49.95%	26.47%	3%	0.35%	100%
5	10.18%	9.43%	43.94%	30.69%	5.17%	0.59%	100%
6	14.93%	17.78%	53.85%	13.01%	0.4%	0.02%	100%
7	16.32%	18.75%	51.74%	12.75%	0.44%	0%	100%
8	17.47%	18.34%	47.78%	15.79%	0.62%	0%	100%
9	16.2%	19.36%	34.38%	24.96%	4.82%	0.28%	100%



10	12.95%	18.97%	37.31%	25.49%	5.15%	0.14%	100%
11	13.57%	19.12%	37.44%	24.39%	5.38%	0.1%	100%
12	13.44%	20.75%	41.89%	20.66%	3.26%	0%	100%

### 1.3.2.3 Comprehension

#### 1.3.2.3.1 By Cluster by Tier

**Table 1.3.2.3.1.1**

Proficiency Level by Cluster (Count): Comprehension, S503 Paper

Cluster	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	173035	15932	17777	8868	21110	5252	241974
1	A	2990	4047	3342	951	985	369	12684
	BC	16	578	3584	2265	2315	1471	10229
2	A	2680	2327	1174	577	584	0	7342
	BC	101	2518	5472	2991	3625	3334	18041
3	A	1231	2509	792	326	351	235	5444
	BC	8	315	4939	5126	6239	2951	19578
4–5	A	3142	2619	1053	489	628	125	8056
	BC	24	1547	8443	7155	9406	5573	32148
6–8	A	5180	4027	1375	443	382	80	11487
	BC	91	3721	9546	5991	6019	2534	27902
9–12	A	4619	4270	1449	506	341	46	11231
	BC	216	4370	7468	4848	5287	3584	25773

**Table 1.3.2.3.1.2**

Proficiency Level by Cluster (Percent): Comprehension, S503 Paper

Cluster	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	71.5%	6.6%	7.3%	3.7%	8.7%	2.2%	100%
1	A	23.6%	31.9%	26.3%	7.5%	7.8%	2.9%	100%
	BC	0.2%	5.7%	35%	22.1%	22.6%	14.4%	100%
2	A	36.5%	31.7%	16%	7.9%	8%	0%	100%
	BC	0.6%	14%	30.3%	16.6%	20.1%	18.5%	100%
3	A	22.6%	46.1%	14.5%	6%	6.4%	4.3%	100%
	BC	0%	1.6%	25.2%	26.2%	31.9%	15.1%	100%
4–5	A	39%	32.5%	13.1%	6.1%	7.8%	1.6%	100%
	BC	0.1%	4.8%	26.3%	22.3%	29.3%	17.3%	100%

6–8	A	45.1%	35.1%	12%	3.9%	3.3%	0.7%	100%
	BC	0.3%	13.3%	34.2%	21.5%	21.6%	9.1%	100%
9–12	A	41.1%	38%	12.9%	4.5%	3%	0.4%	100%
	BC	0.8%	17%	29%	18.8%	20.5%	13.9%	100%

### 1.3.2.3.2 By Grade by Tier

**Table 1.3.2.3.2.1**

Proficiency Level by Grade (Count): Comprehension, S503 Paper

Grade	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	173035	15932	17777	8868	21110	5252	241974
1	A	2990	4047	3342	951	985	369	12684
	BC	16	578	3584	2265	2315	1471	10229
2	A	2680	2327	1174	577	584	0	7342
	BC	101	2518	5472	2991	3625	3334	18041
3	A	1231	2509	792	326	351	235	5444
	BC	8	315	4939	5126	6239	2951	19578
4	A	1410	1504	521	242	287	99	4063
	BC	4	552	4004	3841	4783	2658	15842
5	A	1732	1115	532	247	341	26	3993
	BC	20	995	4439	3314	4623	2915	16306
6	A	1577	1536	475	151	122	38	3899
	BC	9	1112	3515	2382	2127	757	9902
7	A	1738	1284	492	148	123	24	3809
	BC	31	1365	3238	1846	1840	860	9180
8	A	1865	1207	408	144	137	18	3779
	BC	51	1244	2793	1763	2052	917	8820
9	A	1717	1711	517	171	135	25	4276
	BC	18	871	2039	1511	1622	1051	7112
10	A	1187	1225	409	141	101	14	3077
	BC	29	1109	2129	1370	1357	1153	7147
	A	1142	898	350	130	71	7	2598

11	BC	62	1249	1783	1075	1468	931	6568
12	A	573	436	173	64	34	0	1280
	BC	107	1141	1517	892	840	449	4946

**Table 1.3.2.3.2.2**

Proficiency Level by Grade (Percent): Comprehension, S503 Paper

Grade	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	71.5%	6.6%	7.3%	3.7%	8.7%	2.2%	100%
1	A	23.6%	31.9%	26.3%	7.5%	7.8%	2.9%	100%
	BC	0.2%	5.7%	35%	22.1%	22.6%	14.4%	100%
2	A	36.5%	31.7%	16%	7.9%	8%	0%	100%
	BC	0.6%	14%	30.3%	16.6%	20.1%	18.5%	100%
3	A	22.6%	46.1%	14.5%	6%	6.4%	4.3%	100%
	BC	0%	1.6%	25.2%	26.2%	31.9%	15.1%	100%
4	A	34.7%	37%	12.8%	6%	7.1%	2.4%	100%
	BC	0%	3.5%	25.3%	24.2%	30.2%	16.8%	100%
5	A	43.4%	27.9%	13.3%	6.2%	8.5%	0.7%	100%
	BC	0.1%	6.1%	27.2%	20.3%	28.4%	17.9%	100%
6	A	40.4%	39.4%	12.2%	3.9%	3.1%	1%	100%
	BC	0.1%	11.2%	35.5%	24.1%	21.5%	7.6%	100%
7	A	45.6%	33.7%	12.9%	3.9%	3.2%	0.6%	100%
	BC	0.3%	14.9%	35.3%	20.1%	20%	9.4%	100%
8	A	49.4%	31.9%	10.8%	3.8%	3.6%	0.5%	100%
	BC	0.6%	14.1%	31.7%	20%	23.3%	10.4%	100%
9	A	40.2%	40%	12.1%	4%	3.2%	0.6%	100%
	BC	0.3%	12.2%	28.7%	21.2%	22.8%	14.8%	100%
10	A	38.6%	39.8%	13.3%	4.6%	3.3%	0.5%	100%
	BC	0.4%	15.5%	29.8%	19.2%	19%	16.1%	100%
11	A	44%	34.6%	13.5%	5%	2.7%	0.3%	100%
	BC	0.9%	19%	27.1%	16.4%	22.4%	14.2%	100%
12	A	44.8%	34.1%	13.5%	5%	2.7%	0%	100%
	BC	2.2%	23.1%	30.7%	18%	17%	9.1%	100%

1.3.2.3.3 By Grade

**Table 1.3.2.3.3.1**

Proficiency Level by Grade (Count): Comprehension, S503 Paper

Grade	Comprehension Proficiency Range						Total
	1	2	3	4	5	6	
K	173035	15932	17777	8868	21110	5252	241974
1	3006	4625	6926	3216	3300	1840	22913
2	2781	4845	6646	3568	4209	3334	25383
3	1239	2824	5731	5452	6590	3186	25022
4	1414	2056	4525	4083	5070	2757	19905
5	1752	2110	4971	3561	4964	2941	20299
6	1586	2648	3990	2533	2249	795	13801
7	1769	2649	3730	1994	1963	884	12989
8	1916	2451	3201	1907	2189	935	12599
9	1735	2582	2556	1682	1757	1076	11388
10	1216	2334	2538	1511	1458	1167	10224
11	1204	2147	2133	1205	1539	938	9166
12	680	1577	1690	956	874	449	6226

**Table 1.3.2.3.3.2**

Proficiency Level by Grade (Percent): Comprehension, S503 Paper

Grade	Comprehension Proficiency Range						Total
	1	2	3	4	5	6	
K	71.51%	6.58%	7.35%	3.66%	8.72%	2.17%	100%
1	13.12%	20.19%	30.23%	14.04%	14.4%	8.03%	100%
2	10.96%	19.09%	26.18%	14.06%	16.58%	13.13%	100%
3	4.95%	11.29%	22.9%	21.79%	26.34%	12.73%	100%
4	7.1%	10.33%	22.73%	20.51%	25.47%	13.85%	100%
5	8.63%	10.39%	24.49%	17.54%	24.45%	14.49%	100%
6	11.49%	19.19%	28.91%	18.35%	16.3%	5.76%	100%
7	13.62%	20.39%	28.72%	15.35%	15.11%	6.81%	100%
8	15.21%	19.45%	25.41%	15.14%	17.37%	7.42%	100%
9	15.24%	22.67%	22.44%	14.77%	15.43%	9.45%	100%

10	11.89%	22.83%	24.82%	14.78%	14.26%	11.41%	100%
11	13.14%	23.42%	23.27%	13.15%	16.79%	10.23%	100%
12	10.92%	25.33%	27.14%	15.35%	14.04%	7.21%	100%

#### 1.3.2.4 Overall

##### 1.3.2.4.1 By Cluster by Tier

**Table 1.3.2.4.1.1**

Proficiency Level by Grade-Level Cluster (Count): Overall, S503 Paper

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	156407	36030	28363	18171	2989	0	241960
1	A	3735	5304	3457	107	0	0	12603
	BC	582	1770	5897	1671	244	18	10182
2	A	2711	2637	1813	133	0	0	7294
	BC	457	2968	9792	4256	449	5	17927
3	A	1929	2039	1253	181	2	0	5404
	BC	143	1297	11960	5713	340	10	19463
4-5	A	3497	2295	1911	279	0	0	7982
	BC	221	902	12793	15142	2677	188	31923
6-8	A	5857	3385	1943	216	1	0	11402
	BC	189	1607	13770	11465	660	15	27706
9-12	A	5312	3350	2109	323	1	0	11095
	BC	458	2708	11032	9488	1759	64	25509

**Table 1.3.2.4.1.2**

Proficiency Level by Grade-Level Cluster (Percent): Overall, S503 Paper

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	64.6%	14.9%	11.7%	7.5%	1.2%	0%	100%
1	A	29.6%	42.1%	27.4%	0.8%	0%	0%	100%
	BC	5.7%	17.4%	57.9%	16.4%	2.4%	0.2%	100%
2	A	37.2%	36.2%	24.9%	1.8%	0%	0%	100%
	BC	2.5%	16.6%	54.6%	23.7%	2.5%	0%	100%
3	A	35.7%	37.7%	23.2%	3.3%	0%	0%	100%
	BC	0.7%	6.7%	61.4%	29.4%	1.7%	0.1%	100%

4–5	A	43.8%	28.8%	23.9%	3.5%	0%	0%	100%
	BC	0.7%	2.8%	40.1%	47.4%	8.4%	0.6%	100%
6–8	A	51.4%	29.7%	17%	1.9%	0%	0%	100%
	BC	0.7%	5.8%	49.7%	41.4%	2.4%	0.1%	100%
9–12	A	47.9%	30.2%	19%	2.9%	0%	0%	100%
	BC	1.8%	10.6%	43.2%	37.2%	6.9%	0.3%	100%

1.3.2.4.2 By Grade by Tier

**Table 1.3.2.4.2.1**

Proficiency Level by Grade (Count): Overall, S503 Paper

Grade	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	156407	36030	28363	18171	2989	0	241960
1	A	3735	5304	3457	107	0	0	12603
	BC	582	1770	5897	1671	244	18	10182
2	A	2711	2637	1813	133	0	0	7294
	BC	457	2968	9792	4256	449	5	17927
3	A	1929	2039	1253	181	2	0	5404
	BC	143	1297	11960	5713	340	10	19463
4	A	1731	1189	941	170	0	0	4031
	BC	118	396	6631	7351	1134	80	15710
5	A	1766	1106	970	109	0	0	3951
	BC	103	506	6162	7791	1543	108	16213
6	A	1837	1252	699	80	0	0	3868
	BC	48	469	4962	4112	227	9	9827
7	A	1977	1089	656	62	0	0	3784
	BC	67	567	4651	3629	198	4	9116
8	A	2043	1044	588	74	1	0	3750
	BC	74	571	4157	3724	235	2	8763
9	A	2006	1313	772	140	1	0	4232
	BC	58	526	2929	2908	590	34	7045
10	A	1412	939	574	105	0	0	3030
	BC	111	685	2981	2767	517	20	7081
11	A	1272	719	516	59	0	0	2566
	BC	139	739	2818	2345	455	10	6506
12	A	622	379	247	19	0	0	1267
	BC	150	758	2304	1468	197	0	4877



**Table 1.3.2.4.2.2**

Proficiency Level by Grade (Percent): Overall, S503 Paper

Grade	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	64.6%	14.9%	11.7%	7.5%	1.2%	0%	100%
1	A	29.6%	42.1%	27.4%	0.8%	0%	0%	100%
	BC	5.7%	17.4%	57.9%	16.4%	2.4%	0.2%	100%
2	A	37.2%	36.2%	24.9%	1.8%	0%	0%	100%
	BC	2.5%	16.6%	54.6%	23.7%	2.5%	0%	100%
3	A	35.7%	37.7%	23.2%	3.3%	0%	0%	100%
	BC	0.7%	6.7%	61.4%	29.4%	1.7%	0.1%	100%
4	A	42.9%	29.5%	23.3%	4.2%	0%	0%	100%
	BC	0.8%	2.5%	42.2%	46.8%	7.2%	0.5%	100%
5	A	44.7%	28%	24.6%	2.8%	0%	0%	100%
	BC	0.6%	3.1%	38%	48.1%	9.5%	0.7%	100%
6	A	47.5%	32.4%	18.1%	2.1%	0%	0%	100%
	BC	0.5%	4.8%	50.5%	41.8%	2.3%	0.1%	100%
7	A	52.2%	28.8%	17.3%	1.6%	0%	0%	100%
	BC	0.7%	6.2%	51%	39.8%	2.2%	0%	100%
8	A	54.5%	27.8%	15.7%	2%	0%	0%	100%
	BC	0.8%	6.5%	47.4%	42.5%	2.7%	0%	100%
9	A	47.4%	31%	18.2%	3.3%	0%	0%	100%
	BC	0.8%	7.5%	41.6%	41.3%	8.4%	0.5%	100%
10	A	46.6%	31%	18.9%	3.5%	0%	0%	100%
	BC	1.6%	9.7%	42.1%	39.1%	7.3%	0.3%	100%
11	A	49.6%	28%	20.1%	2.3%	0%	0%	100%
	BC	2.1%	11.4%	43.3%	36%	7%	0.2%	100%
12	A	49.1%	29.9%	19.5%	1.5%	0%	0%	100%
	BC	3.1%	15.5%	47.2%	30.1%	4%	0%	100%

1.3.2.4.3 By Grade

**Table 1.3.2.4.3.1**

Proficiency Level by Grade (Count): Overall, S503 Paper

Grade	Overall Proficiency Range						Total
	1	2	3	4	5	6	
K	156407	36030	28363	18171	2989	0	241960
1	4317	7074	9354	1778	244	18	22785
2	3168	5605	11605	4389	449	5	25221
3	2072	3336	13213	5894	342	10	24867
4	1849	1585	7572	7521	1134	80	19741
5	1869	1612	7132	7900	1543	108	20164
6	1885	1721	5661	4192	227	9	13695
7	2044	1656	5307	3691	198	4	12900
8	2117	1615	4745	3798	236	2	12513
9	2064	1839	3701	3048	591	34	11277
10	1523	1624	3555	2872	517	20	10111
11	1411	1458	3334	2404	455	10	9072
12	772	1137	2551	1487	197	0	6144

**Table 1.3.2.4.3.2**

Proficiency Level by Grade (Percent): Overall, S503 Paper

Grade	Overall Proficiency Range						Total
	1	2	3	4	5	6	
K	64.64%	14.89%	11.72%	7.51%	1.24%	0%	100%
1	18.95%	31.05%	41.05%	7.8%	1.07%	0.08%	100%
2	12.56%	22.22%	46.01%	17.4%	1.78%	0.02%	100%
3	8.33%	13.42%	53.13%	23.7%	1.38%	0.04%	100%
4	9.37%	8.03%	38.36%	38.1%	5.74%	0.41%	100%
5	9.27%	7.99%	35.37%	39.18%	7.65%	0.54%	100%
6	13.76%	12.57%	41.34%	30.61%	1.66%	0.07%	100%
7	15.84%	12.84%	41.14%	28.61%	1.53%	0.03%	100%
8	16.92%	12.91%	37.92%	30.35%	1.89%	0.02%	100%
9	18.3%	16.31%	32.82%	27.03%	5.24%	0.3%	100%
10	15.06%	16.06%	35.16%	28.4%	5.11%	0.2%	100%
11	15.55%	16.07%	36.75%	26.5%	5.02%	0.11%	100%

12	12.57%	18.51%	41.52%	24.2%	3.21%	0%	100%
----	--------	--------	--------	-------	-------	----	------

## 2. Analysis of Domains

The measurement model that forms the basis of the analysis for the development of ACCESS for ELLs is the Rasch measurement model (Wright & Stone, 1979). Additional information on its use in the development of the ACCESS for ELLs assessment program is available in WIDA Consortium Technical Report No. 1, *Development and Field Test of ACCESS for ELLs* (Kenyon, 2006). The original ACCESS test developers used Rasch measurement principles, and in that sense, the Rasch model guided all decisions throughout the development of the assessment and was not just a tool for the statistical analysis of the data. Thus, 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. All Rasch analyses are conducted using the Rasch measurement software program *Winsteps* (Linacre, 2006).

### Rasch Model for Dichotomous Scoring

For Listening and Reading, the dichotomous Rasch model was used as the measurement model. Mathematically, the measurement model may be presented as

$$\log\left(\frac{P_{ni1}}{P_{ni0}}\right) = B_n - D_i$$

where

$P_{ni1}$  = probability of providing a correct response “1” by student “n” to item “i”

$P_{ni0}$  = probability of providing an incorrect response “0” by student “n” to item “i”

$B_n$  = ability of student “n”

$D_i$  = difficulty of item “i”

When the probability of a student providing a correct answer to an item equals the probability of a student providing an incorrect answer (i.e., 50% probability of getting it right and 50% probability of getting it wrong),  $P_{ni1}/P_{ni0}$  is equal to 1. The log of 1 is 0. This is the point at which a student’s ability equals the difficulty of an item. For example, a student whose ability estimate is 1.56 on the Rasch logit scale encountering an item whose difficulty is 1.56 on the Rasch logit scale would have a 50% probability of providing a correct answer to that item.

### Rasch Model for Polytomous Scoring

For the Writing and Speaking tasks, a Rasch-grouped rating scale model, which is an extension of Andrich’s rating scale model (Andrich, 1978), is used. Mathematically, this can be represented as

$$\log\left(\frac{P_{ngik}}{P_{ngi(k-1)}}\right) = \beta_n - D_{gi} - F_{gk}$$

where

$P_{ngik}$  = probability of student “n” on task “i” receiving a rating at level “k” on rating scale “g”

$P_{ngi(k-1)}$  = probability of student “n” on task “i” receiving a rating at level “k – 1” on rating scale “g” (i.e., the next lowest rating)

$\beta_n$  = ability of student “n”

$D_{gi}$  = difficulty of task “i” specific to rating scale “g”

$F_{gk}$  = step calibration value of category “k” relative to category “k-1” on rating scale “g”

The subscript “g” is a group index specifying the group of tasks to which task “i” belongs. It also identifies the rating scale that was used for the group of tasks. There is only one rating scale ( $g = 1$ ) in the Writing domain and two grouped rating scales ( $g = 2$ ) in the Speaking domain. As with the dichotomous Rasch model, there is an item difficulty parameter ( $D_{gi}$ ) for each item for rating scale “g” modeled by the Rasch rating scale model (Andrich, 1978). In addition, there is a step calibration value or *step measure* ( $F_{gk}$ ) that corresponds to the location on the latent variable where the probability of being observed in the “k” and “k – 1” category for rating scale “g” is equal relative to the difficulty measure of the task. The step measures are also the points where adjacent category probability “k – 1” and “k” curves for rating scale “g” intercept. All tasks that belong to the same rating scale group have the same step measures.

As described in Part 1, Section 3.2.2, ratings on the ACCESS Writing Scoring Scale range from 0, 1, 1+, ..., 6, and the possible raw scores range from 0 to 9. All Writing tasks are scored using this scoring scale except for Grade 1 Tier A Tasks 1 and 2. The profiles of the responses to these two tasks do not fit the generic scoring scale well, so additional task-specific instructions are provided to raters. These instructions guide raters in applying a limited number of score points on the scoring scale to responses elicited by these two tasks. The possible ratings for Grade 1 Tier A Task 1 are 0 or 1, and the possible ratings for Grade 1 Tier A Task 2 are 0, 1, 1+, or 2. To simplify the year-to-year linking process, the Grade 1 Writing Tier A Task 1 is treated as a dichotomously scored task. The Grade 1 Writing Tier A Task 2 is modeled using a rating scale with a possible raw score of 0 to 3. All other Writing tasks are modeled using a rating scale with possible raw scores of 0 to 9. Thus, a total of two rating scales are modeled for ACCESS Writing. One rating scale is associated with the Grade 1 Writing Tier A Task 2, and the other rating scale is associated with all Writing tasks that are scored using the rating scale with raw score values of 0–9. We conducted a study in the summer of 2016 to reconstruct the logit scales. Detailed information about the derivation of the Writing rating scales as well as the psychometric properties of Writing rating scales are available in the scaling report (see Center for Applied Linguistics, 2017).

For Speaking, we model Proficiency Level 1 tasks as a group on a 0–2 scale, and PL 3 and PL 5 tasks as a group on a 0–4 scale (see Part 1, Section 3.2.4). We conducted a study in the summer of 2016 to reconstruct the logit scales and detailed information about the derivation as well as the psychometric properties of Speaking rating scales are available in the scaling report (Center for Applied Linguistics, 2017).

### Scale Scores and Proficiency Level Scores

Scale scores are calculated by transforming the student ability estimate via a scaling equation.

For Paper ACCESS Grades 1–12, the following scaling equations are used to convert ability measures in logits to scale scores:

- L: (Ability Measure in Logits \* 37.571) + 316.637
- R: (Ability Measure in Logits \* 26.000) + 323.272
- W: (Ability Measure in Logits \* 26.851) + 303.332
- S: (Ability Measure in Logits \* 29.248) + 265.076

In the domains of Listening and Reading, we established the current ACCESS scale for the original paper-only version of the test and maintained this scale through the transition to an online and paper delivered test in the 2015–2016 school year (Series 400). Evidence for scale maintenance in the transitional year is described elsewhere (Center for Applied Linguistics, 2016). In the domains of Writing and Speaking, we conducted a study in the summer of 2016 to reconstruct the logit scale (see Center for Applied Linguistics, 2017).

Note that these new scales were not applied to the Kindergarten test, which is a static form. The following scaling equations are used for the Kindergarten test:

- L: (Ability Measure in Logits \* 37.571) + 316.637
- R: (Ability Measure in Logits \* 26.000) + 323.272
- W: (Ability Measure in Logits \* 31.097) + 317.068
- S: (Ability Measure in Logits \* 20.084) + 322.686

Proficiency level scores are interpretations of these scale scores in terms of the proficiency levels described in the WIDA ELD Standards. These interpretations derive from a series of standard setting studies, in which educators reviewed evidence from the test, either in the form of items for the selected response sections (Listening and Reading) or student portfolios for the constructed response sections (Writing and Speaking), to establish cut scores between the proficiency levels. The first standard setting study for ACCESS took place in 2005; it established cut scores for all four domains by grade-level cluster (Kenyon, 2006). The second cut score study took place in 2007; it established cut scores for all four domains by grade level (Kenyon, Ryu, & MacGregor, 2013). These cut scores were used to derive proficiency level scores through the 2015–2016 administration (Series 400) of ACCESS for ELLs. WIDA and CAL conducted a third cut score study in summer 2016 (Cook & MacGregor, 2017). The purpose of this study was to re-examine cut scores for each of the proficiency levels in light of the migration from the paper-and-pencil-only assessment to both online and paper delivery, the revision of the Speaking test, and the influence of college- and career-ready standards. These new cut scores were first used for ACCESS Series 401 (2016–2017 school year).

A proficiency level score consists of a two-digit decimal number (e.g., 4.5). The first digit represents the student's overall proficiency level range based on the student's scale score. The number to the right of the decimal is an indication of the proportion of the range between cut scores that the student's scale score represents. A score of 4.5, for example, tells us that the student is in PL 4 and that the student's scale score is halfway between the cut scores for PLs 4 and 5.

Unlike the scale scores, which form an interval scale and are continuous across grades from Kindergarten to Grade 12, PL scores are dependent upon the grade a student was in when the student took the assessment. For example, a score of 350 in Listening would be interpreted as a PL score of 5.8 for a Grade 2 student, a 3.8 for a Grade 5 student, a 3.1 for a Grade 8 student, and a 2.3 for a Grade 12 student.

Because the bands between cut scores on the score scale vary in width, PL scores do not form an interval scale. Only scale scores should be used as interval measures. PL scores are at even intervals within a grade and proficiency level (e.g., in Grade 3, the distance between 3.1 and 3.2 is the same as the distance between 3.7 and 3.8), but they do not form an interval scale across proficiency levels.



## 2.1 Complete Item or Task Analysis and Summary

The tables in this section provide information on the psychometric qualities of the items and tasks. We provide values for item or task difficulties in logits, the number of items or tasks on the form, the average  $p$  value (for forms with selected response items), and the Rasch model fit statistics. For Writing and Speaking, we also provide raw score distributions by task.

Tables in this section have either two parts (in the case of Listening and Reading) or three parts (in the case of Writing and Speaking). The first part of the table gives a summary of the total set of items or tasks on the form. The second part provides statistics pertaining to the individual items or tasks, and the third part (for Writing and Speaking only) expresses raw score distributions by task.

All Rasch analyses were conducted using the Rasch measurement software program *Winsteps 5.2.4.0* (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 are then 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 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. He wrote:

- Values greater than 2.0 “distort or degrade<sup>1</sup> 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 “less productive for measurement, but not degrading.”

Linacre also stated in his guidance that infit problems are more serious to the construction of measurement than are outfit problems.

---

<sup>1</sup> We interpret “degrade” here in the sense of lowering the quality of the measurement system.



Because we follow conservative guidelines in the development of ACCESS for ELLs, it is desired that the dichotomous items on the test forms have mean square fit statistics in the range of 0.5 to 1.5; and thus, they fit the range that is “productive for measurement” according to the guidelines above. See below for the percentages of dichotomous items which have mean square statistics within this range, by domain.

Since performance tasks are constructed and scored very differently from dichotomous items, it is not as straightforward to apply this same guidance to interpret these fit statistics for performance tasks that raters scored polytomously on a rubric scale. We design some performance tasks to elicit a restricted range of performances (for example, very easy tasks where we expect that most students will get the highest rating), and these tasks can cause the model to predict the data too well (overfitting). Conversely, when raters score performance tasks using a very wide rubric scale such as the ACCESS for ELLs Writing rubric, sometimes unmodeled noise or other sources of variance in the ratings of the students’ responses to the task will cause the model to underpredict those ratings (underfitting). Overall, for ACCESS for ELLs performance tasks, overfitting is more common than underfitting. Underfitting indicates that the task is less productive for measurement, but, according to Linacre (2002), including the rating of the student’s performance on the task when calculating that student’s score does not degrade the measurement of the student’s performance.

Tables in this section are presented by test form (i.e., by grade cluster and tier) for Listening, Reading, and Writing. For the Speaking test, due to the design of the test, a number of items are shared between tiers. To best present the results of the Speaking task analysis, all Speaking items in a grade-level cluster are presented in one single table.

The first section of the Complete Item/Task Analysis and Summary table provides information about the total set of items or tasks and includes the item type (selected response or constructed response), the average item difficulty measure (in logits), the number of items, the average  $p$  value (for Listening and Reading only), the average infit mean square statistic, and the average outfit mean square statistic.

The second section of these tables presents results from the analyses of all the items or tasks on the test form. The first column provides the unique item name. The second column in this section presents the item or task difficulty measure in logits. For Listening and Reading, the third column shows the  $p$  value (percentage of correct answers on that item), while the next column for Writing and Speaking shows the task difficulty in logits. The final two columns show the Rasch fit statistics for the item or task. Folders with items that have fit statistics greater than 2.0 are evaluated by the test development team to determine whether and when the folders can be refreshed in the next test refreshment cycle.

In addition, Writing and Speaking tables have a section at the bottom of the table that provides raw score distributions by task.

For the Grades 1–12 tests, all items and tasks across domains have infit mean square statistics less than 2, indicating that the items and tasks provide good measurement for students around the ability range that the items and tasks are targeting. One task in Writing Grade 1 Tier A form has an outfit mean square statistic greater than 2. This is the easiest task for this test form, and there might be some high-ability students receiving a low rating, causing the outfit mean square statistics to be inflated.

The results show that for the Kindergarten test, all items and tasks across domains have infit mean square statistics less than 2, except for the fifth task in the Writing domain, indicating that the items and tasks provide trustworthy measures of ability for those students whose ability measures are in the region of the ability distribution that the items and tasks are targeting. As discussed earlier, the outfit mean square statistic is sensitive to outlier responses and ratings that are not close to the ability range that the items

and tasks are targeting. Five items in the Listening domain, 11 items in the Reading domain, two tasks in the Writing domain, and two tasks in the Speaking domain have outfit mean square statistics greater than 2. For the most part, these are very easy items or tasks (with  $p$  values  $> 0.85$ ) early in the test. These outfit values are likely due to high-ability students getting these early test items incorrect. The test design includes multiple easy items at the onset of the test to ensure that Kindergarten students, who are often unfamiliar with standardized testing, are not presented with discouraging difficult items at the beginning of their test administration.

Outfit values are exceedingly high (9.90) for the first three Reading items. The Kindergarten ACCESS technical brief notes that the items in this folder are prereading items and that children with high reading ability who are not familiar with these items may not answer correctly, leading to high outfit values.

All Listening Grades 1-12 items have infit mean square statistics and outfit mean square statistics between 0.5 and 1.5. All Reading Grades 1-12 items have infit mean square statistics between 0.5 and 1.5. All Reading items have outfit mean square statistics between 0.5 and 1.5, except the forms in Reading G1A, G1BC, G2A, G2BC, G3A, G45A and G68A, which have 96% of items that have outfit mean square statistics that falls between 0.5 and 1.5. The percentage of Listening Kindergarten items that have infit mean square statistics between 0.5 and 1.5 is 97% and the percentage for outfit mean square statistics between 0.5 and 1.5 is 73%. The percentages of Kindergarten Reading items with infit mean spear statistics between 0.5 and 1.5 is 100, while the percentage of those items with outfit mean square statistics between 0.5 and 1.5 is 37%.

Note: The redacted version of the annual technical report does not have item related information (tables are removed from section 2.1 Complete Item or Task Analysis and Summary and section 2.2 DIF Analysis and Summary).

## 2.2 *DIF Analysis and Summary*

Prior to field testing, the Bias and Sensitivity Review Panel ensures that test items and tasks are free of material that (1) might favor any subgroup of students over another on the basis on gender, race/ethnicity, home language, religion, culture, region, or socioeconomic status, and (2) might be upsetting to students. This process is qualitatively driven, while the DIF analysis, described below, is data driven. Please see Part 1, section 2.3.1 for more information on Bias and Sensitivity panels.

CAL uses differential item functioning (DIF) analysis to investigate whether factors extraneous to English language proficiency (i.e., the construct being measured on the test) may have influenced some students' performances on items. DIF attempts to find items that may be functioning differently for different groups based on criteria irrelevant to the construct that is purportedly being measured. We compare the performance of students on ACCESS for ELLs Paper items and tasks by dividing students into two different groupings: first, males versus females; second, students of Hispanic ethnic background versus students of all other backgrounds. For the former analysis, females is the reference group, while males is the focal group. For the latter analysis, Hispanics is the reference group, while Non-Hispanics is the focal group. We exclude students for whom gender or ethnicity<sup>2</sup> was unknown from both analyses. We used two commonly used procedures for detecting DIF: one for dichotomously scored items (Listening and Reading) and one for polytomously scored items (Writing and Speaking).

It should be noted that for ACCESS Paper Listening, Reading, Writing and Speaking, static forms are used. As such, the DIF analysis was conducted the first year these forms were used operationally, using ATR data. However, for Grades 6, 8, 9, 10, 11, and 12, the DIF analysis was conducted using only the field test data, as directed by WIDA, except for the Listening domain, Grades 9-12, Tier BC, which did not have a sufficient sample size to conduct a DIF analysis. CAL computed these DIF statistics using 503 ATR data. Please see section 2.7, below, for further explanation and operational history of forms.

### **Dichotomous Items**

We used the Mantel-Haenszel (M-H) chi-square statistic (Mantel & Haenszel, 1959) procedure for dichotomous items, originally proposed by the Educational Testing Service (ETS). This procedure compares item-level performances of students in the two groups (e.g., males versus females) who are divided into subgroups based on their performance on the total test. We assume that if there is no DIF, a similar percentage of students in each group should get the item correct at any ability level (based on performance on the total test). We use the M-H chi-square statistic to check the probability that the two groups performed comparably on each item across the ability groupings. The statistic is transformed into the "M-H delta" scale. This scale is symmetrical around zero, with a delta zero interpreted as indicating that neither group is favored. A positive result indicates that the focal group is favored; a negative result indicates that the reference group is favored.

Because DIF is measured on a continuous scale, and because most items are likely to show some degree of DIF, it is useful to have guidelines to determine when the level of DIF requires further review of the item. We follow the guidance provided by ETS (Zieky, 1993) to classify items into DIF levels as follows:

---

<sup>2</sup> In the dataset, Hispanic ethnicity, as well as each of the race categories, is coded as a binary variable (Y/blank). Ethnicity information is counted as "Unknown" in cases where the student is recorded as blank for Hispanic ethnicity and blank for every race category.

- A (no DIF) when the absolute value of delta is  $<1.0$
- B (weak DIF) when the absolute value of delta is 1.0 to 1.5
- C (strong DIF) when the absolute value of the delta is  $>1.5$

We used the software program *EZDIF* (Waller, 1998) to run the DIF analyses for all forms containing dichotomous items. For each test form, the greatest number of ability-level groupings is used; however, for many test forms, students scoring some of the lowest and highest raw scores need to be grouped together to have enough cases in each cell for the statistic to be appropriately calculated. (Note that this software program uses a two-step purification process; that is, items with C-level DIF in the first pass are removed from the matching variable in the second stage, and the DIF is then recalculated for the remaining items.)

## Polytomous Items

For polytomous items (i.e., Writing and Speaking tasks), we take a similar approach. Our approach is based on the M-H chi-square statistic and the standardized mean difference following procedures that ETS developed (Allen, Carlson, & Zalanak, 1999; Zwick, Donoghue, & Grima, 1993). These DIF procedures for polytomous items were used to identify tasks that exhibit DIF. We used JMetrik (Meyer, 2018), an open-source computer program for psychometric analysis, to conduct the analyses. The procedures implemented in JMetrik first calculate the Cochran-Mantel-Haenszel chi-square statistic for testing statistical significance. This statistic gives an indication of the probability that observed differences are the result of chance but does not indicate how significant that difference is. To indicate how significant the difference is, we calculate the standardized mean difference between the performances of the two comparison groups. The standardized mean difference compares the means of the two groups, adjusting for differences in the distribution of the groups across the values of the total raw scores. To standardize the outcome, this difference is divided by the item score range and serves as an effect size measure for the Cochran-Mantel-Haenszel chi-square statistic. This effect size measure (reported as standardized P-DIF in JMetrik) ranges from -1 to 1, which may present some interpretation challenges. To mitigate the negative value, the absolute value of the Cochran-Mantel-Haenszel chi-square statistic is used in JMetrik (Meyer, 2018) and the range of the rescaled effect size (standardized P-DIF\*) is restricted to fall between 0 and 1. The effect size flagging criterion for polytomous items that ETS proposed (Allen et al., 1999) is also rescaled to the standardized P-DIF\* metric (Meyer, 2018).

Following guidance that ETS proposed for the National Assessment of Educational Progress (Allen et al., 1999), we classify ACCESS for ELLs Writing and Speaking tasks into three DIF levels as follows:

- AA (no DIF), when the Cochran-Mantel-Haenszel chi-square statistic is not significant or when it is significant and standardized P-DIF\* is  $<0.05$
- BB (weak DIF), when the Cochran-Mantel-Haenszel chi-square statistic is significant and standardized P-DIF\* is  $\geq 0.05$  but  $<0.10$
- CC (strong DIF) when the Cochran-Mantel-Haenszel chi-square statistic is significant and standardized P-DIF\* is  $\geq 0.10$

The tables in this section provide a summary of the findings of the DIF analyses at the top, followed by information for any item or task that showed B, BB, C, or CC-level DIF. The first column gives the DIF

level: A, B, or C for dichotomous items or AA, BB, or CC for polytomous tasks (i.e., Writing and Speaking tasks). The next columns show the contrasting groups in the DIF analyses: either male (focal group) versus female (reference group) or Hispanic (reference group) versus non-Hispanic ethnicities (focal group). The top part of the table summarizes the number of items that exhibit DIF falling into each of the three categories (A, B, or C for Listening and Reading, and AA, BB, or CC for Writing and Speaking). Any items that show B (or BB) or C (or CC)–level DIF are reported in the bottom part of the table.

Paper ACCESS is administered as two rotating static forms. Bias and sensitivity panels reviewed these items prior to any field testing, as described in Section 2.3.1. We conducted DIF analysis prior to the final selection of the two static forms. For any items or tasks that showed C-level (or CC-level) DIF, an additional DIF review panel was convened to re-examine the item for bias concerns.

The DIF panel manager, from CAL, draws panelists from CAL staff members. Members are chosen so that a diverse background is represented. Therefore, the panel manager considers gender, first/second language backgrounds, and ethnicity when empaneling judges. The manager also ensures that some members have expertise in English as a Second Language instruction and/or professional development for teachers of ESL students. The panel is asked to discuss the item and come to a consensus on whether they believe or do not believe that the item demonstrates bias against a particular group and is or is not appropriate to place on the operational test.

Two items, one from Grade 3 Listening Tier A and other item from Grade 4-5 Listening Tier A showed C-level DIF, both items favoring ‘Females’ for gender and ‘Other’ for ethnicity in Series 503. The items illustrating C- level DIF were retained from series 501 and underwent a DIF panel at that time. No new items added to 503 showed C- level DIF.

Note: The redacted version of the annual technical report does not have item related information (tables are removed from section 2.1 Complete Item or Task Analysis and Summary and section 2.2 DIF Analysis and Summary).

## 2.3 Raw Score Distribution

Figures and tables in this section provide detail on the distribution of raw scores. For each grade-level cluster and tier combination, the figure shows the distribution of the raw scores. The horizontal axis shows the raw scores. The vertical axis shows the number of students (count). Each bar shows how many students received each raw score.

Each table in this section summarizes results for a grade-level cluster and tier combination (e.g., Speaking 4–5 Tier A). For each table, results are broken down by grade and presented for the grade-level cluster as a whole for that tier. The following information is included in each table:

- The number of students in the analyses (the number of students who were not absent, invalid, refused, exempt, or in the wrong grade-level cluster)
- The minimum observed raw score
- The maximum observed raw score
- The mean (average) raw score
- The standard deviation (std. dev.) of the raw scores

Test design and student population impact the distribution of raw scores. In general, raw score distributions tend to be smoothly distributed with a single peak; however, there are a number of exceptions. Understanding these distributions supports the understanding of other statistical properties of the test forms.

In the domain of Writing, in Tier B/C, the three tasks are weighted once, twice, and three times, respectively. The impact of this weighting is that the raw scores are not smoothly distributed.

In the domain of Speaking, on Tier A forms, three of the six tasks are scored on a restricted portion of the rubric (with possible raw scores of 0 to 2). Most students score all six of these points; however, less proficient students may score only one or two points consistently on the remaining tasks. On Tier B/C, students are automatically awarded these six points (as it is assumed they would have the ability to achieve the maximum possible points on the easiest tasks). These aspects of the test design impact raw score distribution.

As mentioned, students routed to the A form take three P1 tasks, scored 0 to 2. They also take three P3 tasks, scored 0 to 4, for a total raw score range of 0 to 18. Students routed to take the B/C form do not take the P1 tasks, as it is assumed that they would be able to get the full two points on these very easy P1 tasks. These students take three P3 and three P5 tasks, each scored 0 to 4, and they are awarded two points on each of three P1 tasks. The total raw score range for Tier B/C form is 6 to 30.

The Kindergarten test design includes skipping and stopping rules intended to reduce testing time for young children; these rules also have an impact on the distribution of raw scores, leading to less smooth distributions.



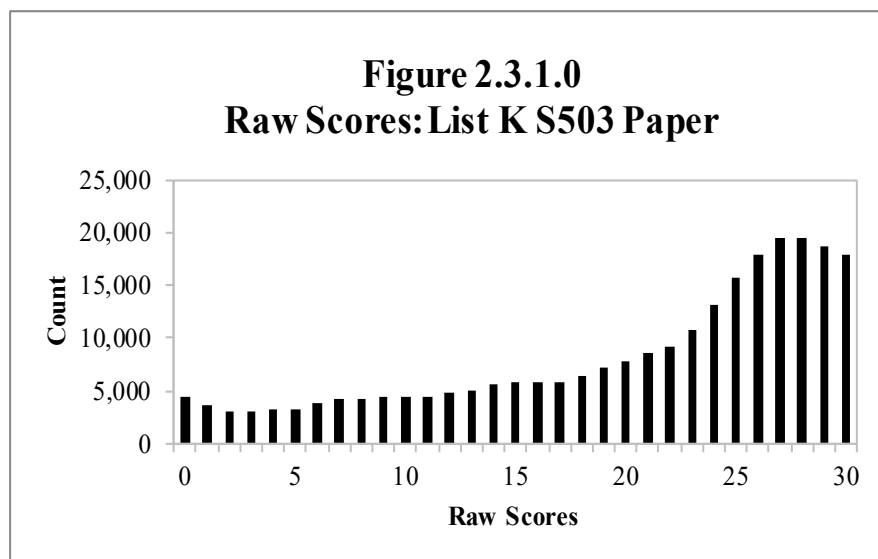
## 2.3.1 Listening

### 2.3.1.0 Kindergarten

**Table 2.3.1.0**

Raw Score Descriptive Statistics: List K S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	251,493	0	30	20.32	8.43
Total	251,493	0	30	20.32	8.43

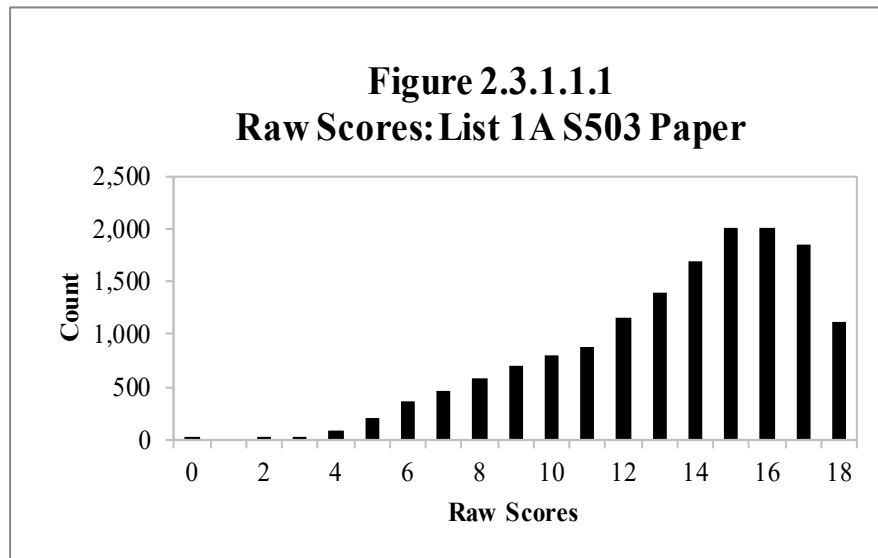


### 2.3.1.1 Grade 1

**Table 2.3.1.1.1**

Raw Score Descriptive Statistics: List 1 A S503 Paper

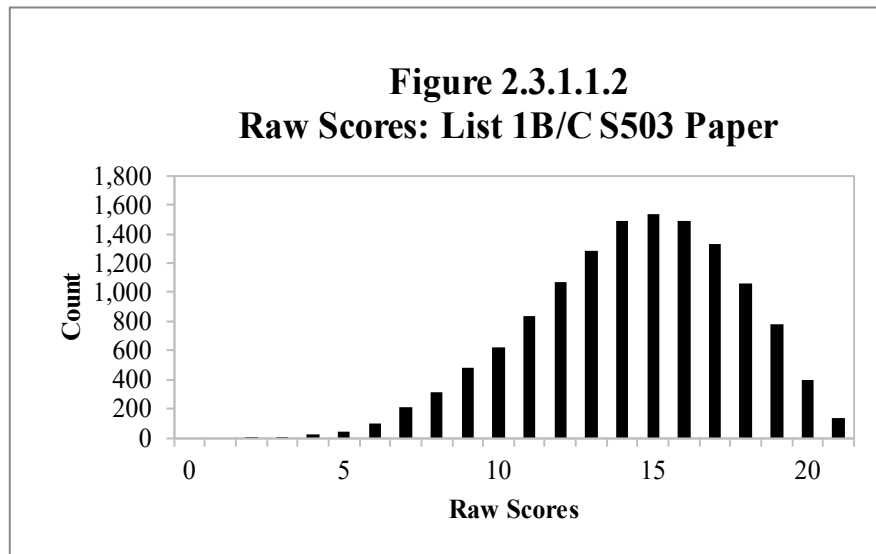
Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,349	0	18	13.36	3.42
Total	15,349	0	18	13.36	3.42



**Table 2.3.1.1.2**

Raw Score Descriptive Statistics: List 1 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	13,219	2	21	14.33	3.32
Total	13,219	2	21	14.33	3.32

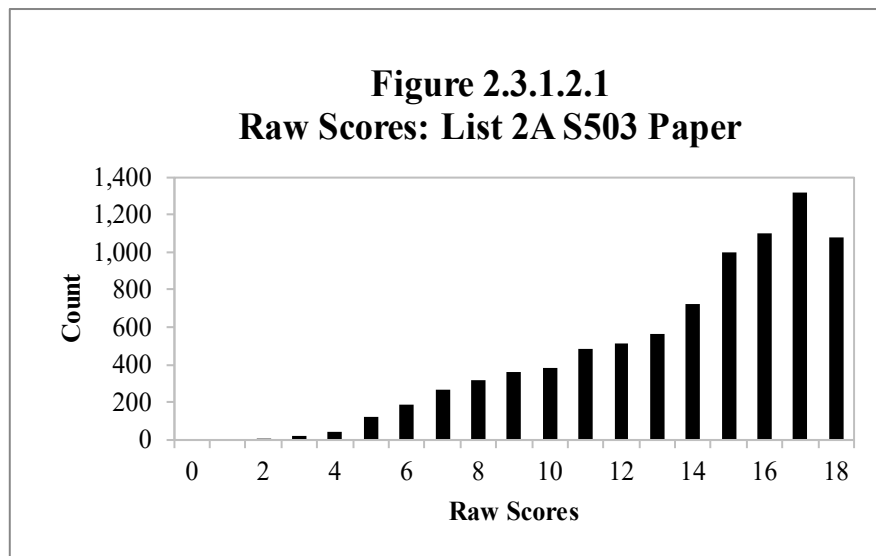


### 2.3.1.2 Grade 2

**Table 2.3.1.2.1**

Raw Score Descriptive Statistics: List 2 A S503 Paper

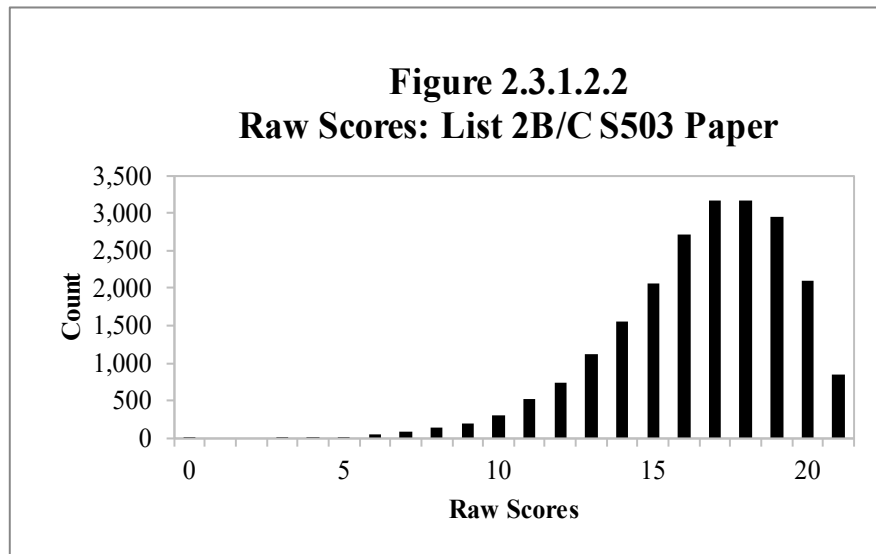
Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	8,467	2	18	13.78	3.58
Total	8,467	2	18	13.78	3.58



**Table 2.3.1.2.2**

Raw Score Descriptive Statistics: List 2 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	21,727	0	21	16.51	2.88
Total	21,727	0	21	16.51	2.88

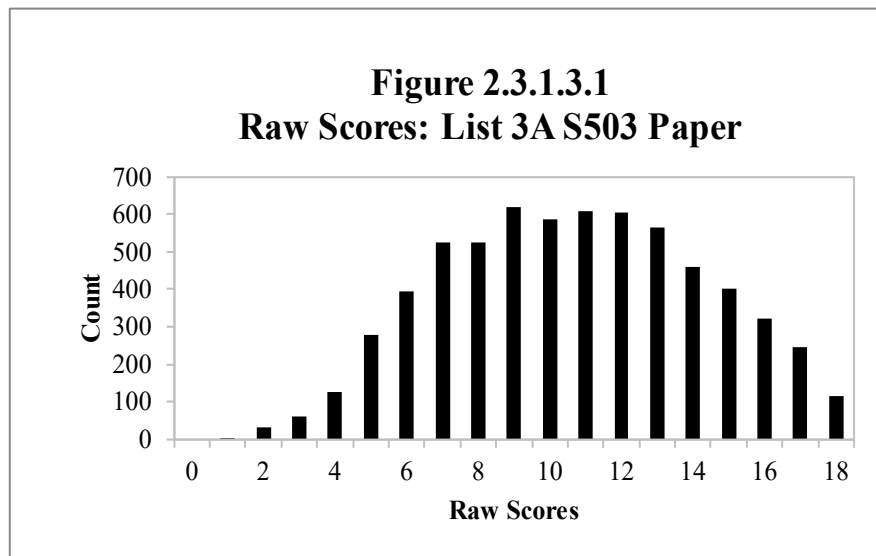


### 2.3.1.3 Grade 3

**Table 2.3.1.3.1**

Raw Score Descriptive Statistics: List 3 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	6,466	1	18	10.66	3.60
Total	6,466	1	18	10.66	3.60

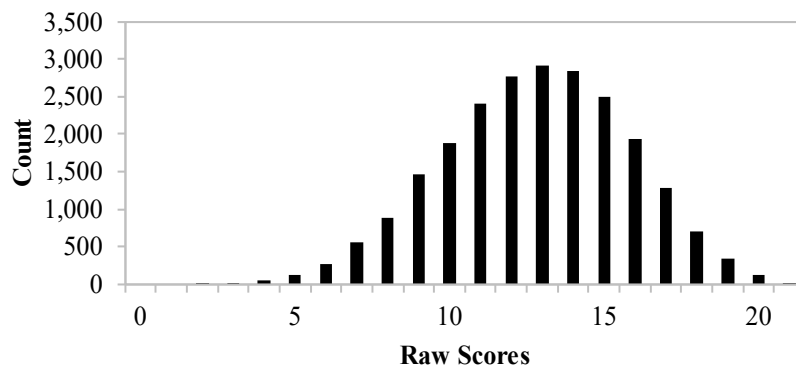


**Table 2.3.1.3.2**

Raw Score Descriptive Statistics: List 3 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	23,072	2	21	12.79	2.99
Total	23,072	2	21	12.79	2.99

**Figure 2.3.1.3.2**  
**Raw Scores: List 3B/C S503 Paper**

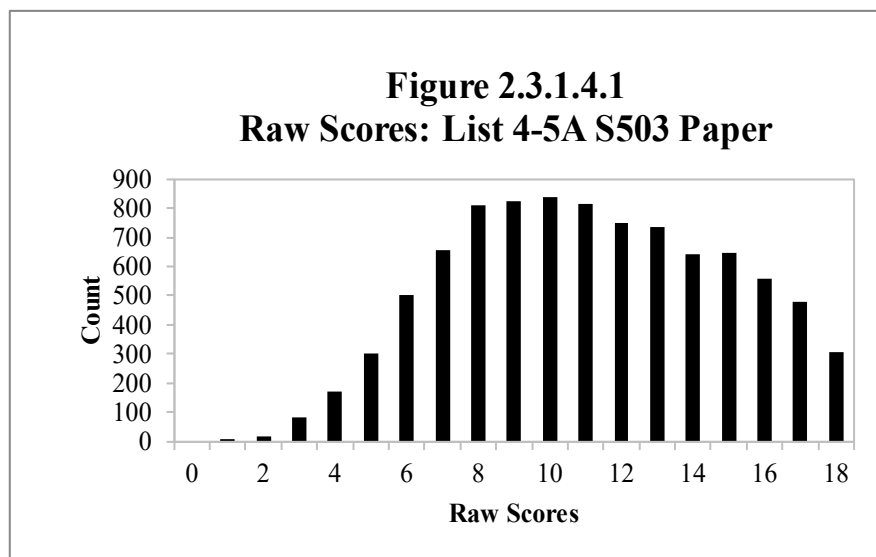


#### 2.3.1.4 Grades 4-5

**Table 2.3.1.4.1**

Raw Score Descriptive Statistics: List 4-5 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	4,649	1	18	10.86	3.66
5	4,498	1	18	11.27	3.78
Total	9,147	1	18	11.07	3.73

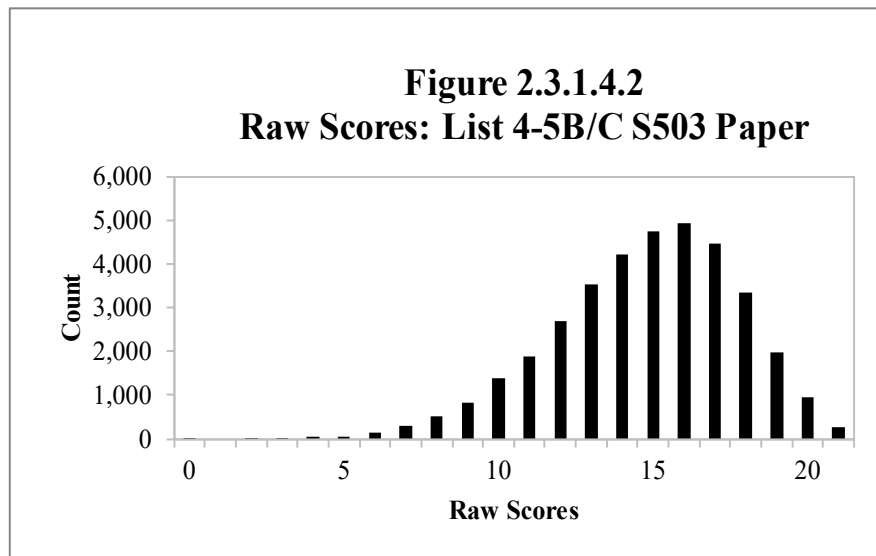




**Table 2.3.1.4.2**

Raw Score Descriptive Statistics: List 4-5 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	18,063	2	21	14.32	2.96
5	18,281	0	21	15.16	2.94
Total	36,344	0	21	14.75	2.98

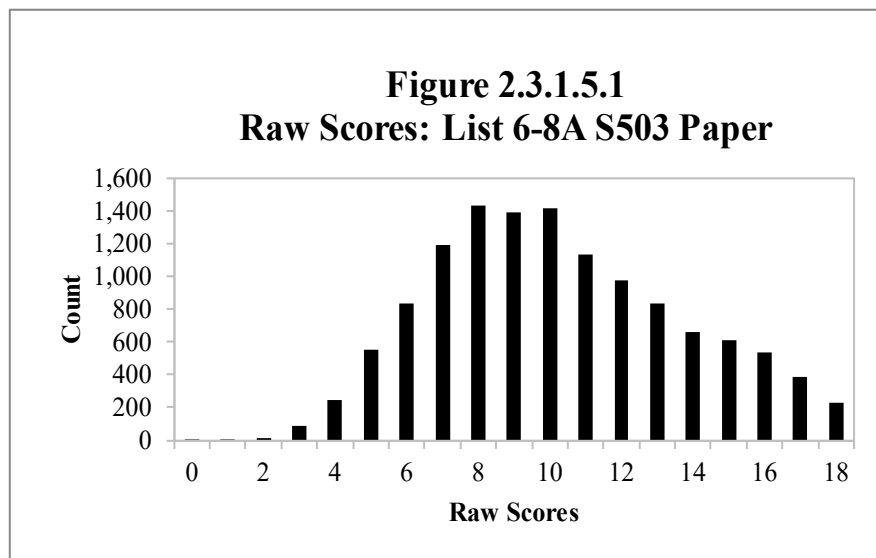


### 2.3.1.5 Grades 6-8

**Table 2.3.1.5.1**

Raw Score Descriptive Statistics: List 6-8 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	4,277	0	18	10.02	3.41
7	4,139	1	18	10.23	3.46
8	4,111	2	18	10.46	3.49
<b>Total</b>	<b>12,527</b>	<b>0</b>	<b>18</b>	<b>10.24</b>	<b>3.46</b>

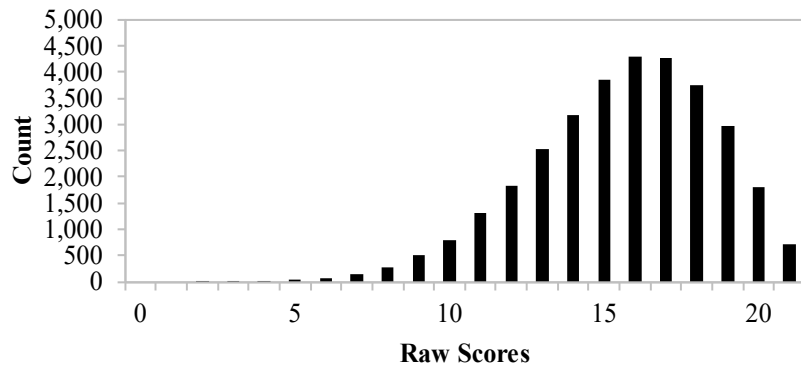


**Table 2.3.1.5.2**

Raw Score Descriptive Statistics: List 6-8 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	11,661	3	21	15.11	2.91
7	10,618	2	21	15.53	2.99
8	10,069	4	21	16.18	2.93
Total	32,348	2	21	15.58	2.98

**Figure 2.3.1.5.2**  
**Raw Scores: List 6-8B/C S503 Paper**



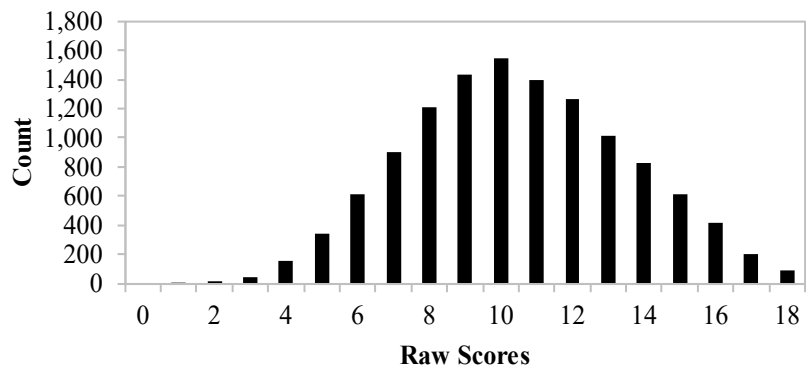
### 2.3.1.6 Grades 9-12

**Table 2.3.1.6.1**

Raw Score Descriptive Statistics: List 9-12 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	4,629	1	18	10.08	3.06
10	3,295	2	18	10.53	3.07
11	2,784	2	18	10.76	2.99
12	1,369	4	18	11.20	2.97
Total	12,077	1	18	10.49	3.06

**Figure 2.3.1.6.1**  
**Raw Scores: List 9-12A S503 Paper**

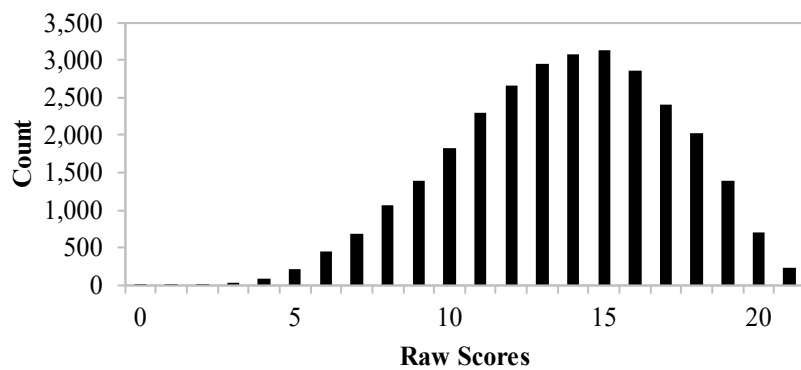


**Table 2.3.1.6.2**

Raw Score Descriptive Statistics: List 9-12 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	8,255	2	21	13.51	3.42
10	8,244	0	21	13.69	3.54
11	7,452	0	21	13.82	3.57
12	5,522	1	21	13.58	3.56
Total	29,473	0	21	13.65	3.52

**Figure 2.3.1.6.2**  
**Raw Scores: List 9-12B/C S503 Paper**



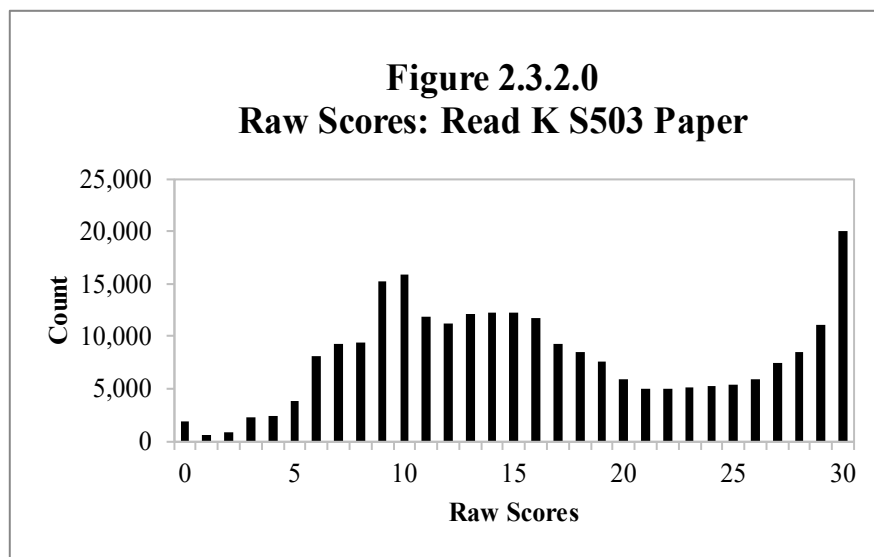
## 2.3.2 Reading

### 2.3.2.0 Kindergarten

**Table 2.3.2.0**

Raw Score Descriptive Statistics: Read K S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	251,491	0	30	16.60	7.99
Total	251,491	0	30	16.60	7.99

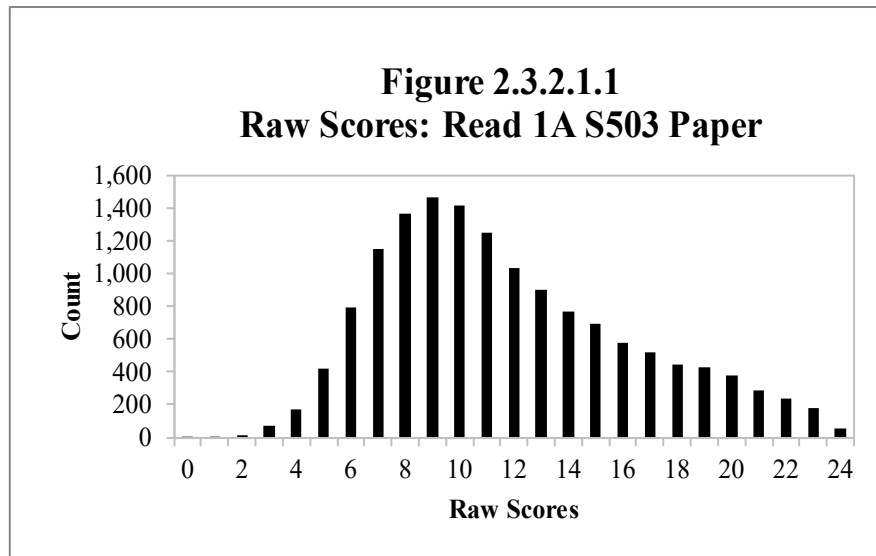


### 2.3.2.1 Grade 1

**Table 2.3.2.1.1**

Raw Score Descriptive Statistics: Read 1 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	14,624	0	24	11.73	4.58
Total	14,624	0	24	11.73	4.58

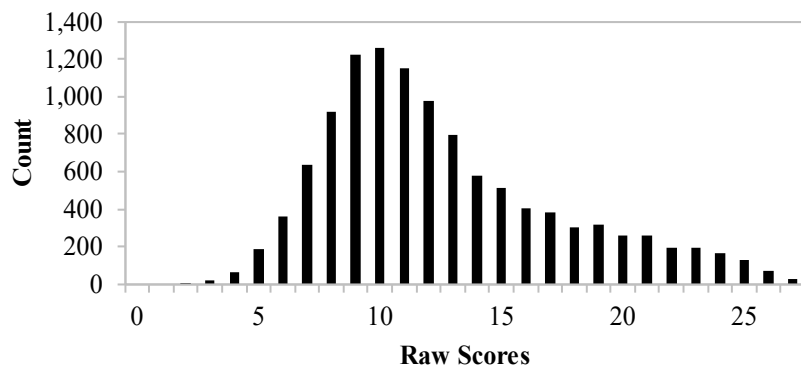


**Table 2.3.2.1.2**

Raw Score Descriptive Statistics: Read 1 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	11,385	2	27	12.54	4.85
Total	11,385	2	27	12.54	4.85

**Figure 2.3.2.1.2**  
**Raw Scores: Read 1B/C S503 Paper**



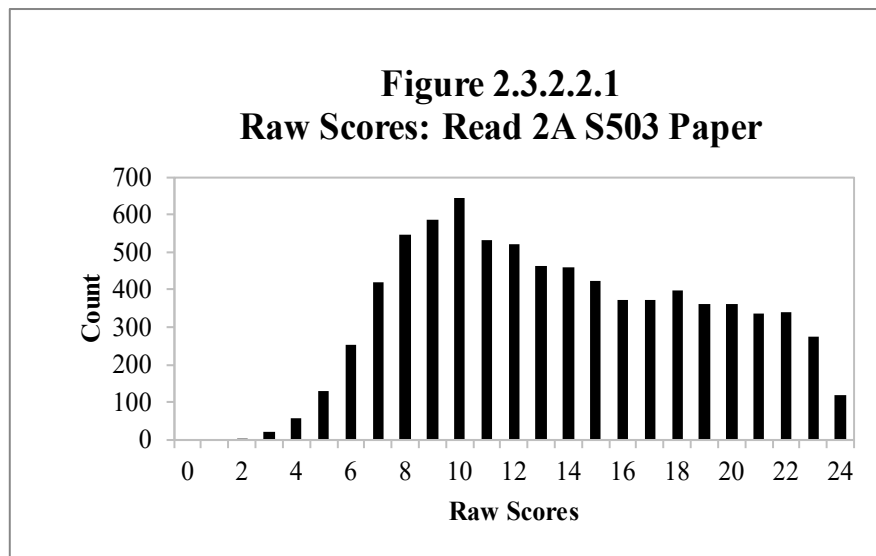


### 2.3.2.2 Grade 2

**Table 2.3.2.2.1**

Raw Score Descriptive Statistics: Read 2 A S503 Paper

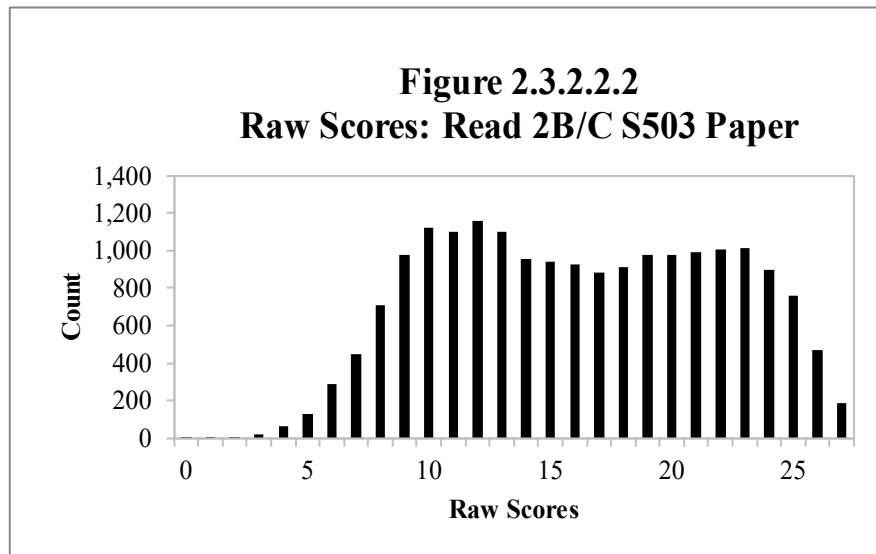
Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	7,993	2	24	13.72	5.17
Total	7,993	2	24	13.72	5.17



**Table 2.3.2.2.2**

Raw Score Descriptive Statistics: Read 2 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	19,027	0	27	16.19	5.69
Total	19,027	0	27	16.19	5.69

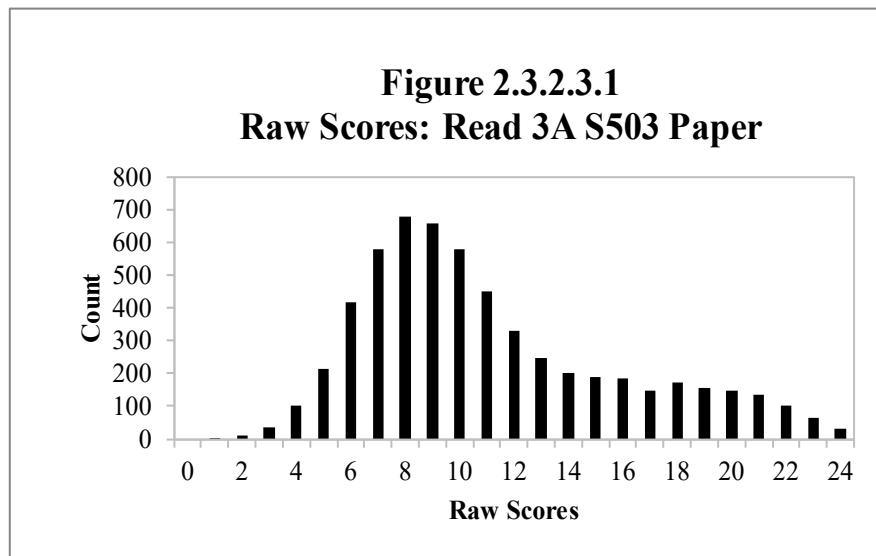


### 2.3.2.3 Grade 3

**Table 2.3.2.3.1**

Raw Score Descriptive Statistics: Read 3 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	5,840	1	24	11.12	4.75
Total	5,840	1	24	11.12	4.75

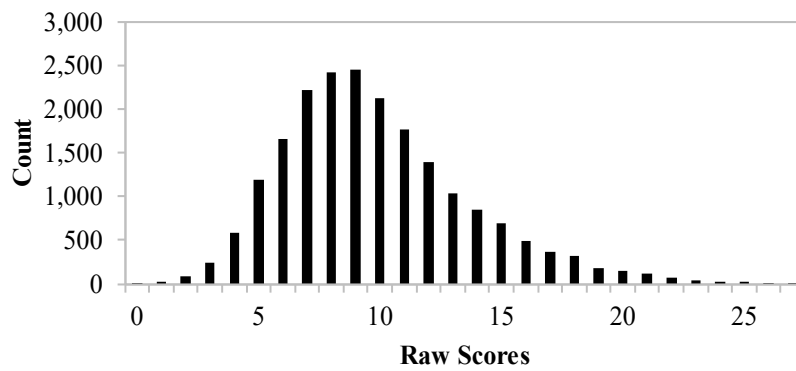


**Table 2.3.2.3.2**

Raw Score Descriptive Statistics: Read 3 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	20,551	0	27	9.83	3.86
Total	20,551	0	27	9.83	3.86

**Figure 2.3.2.3.2**  
**Raw Scores: Read 3B/C S503 Paper**

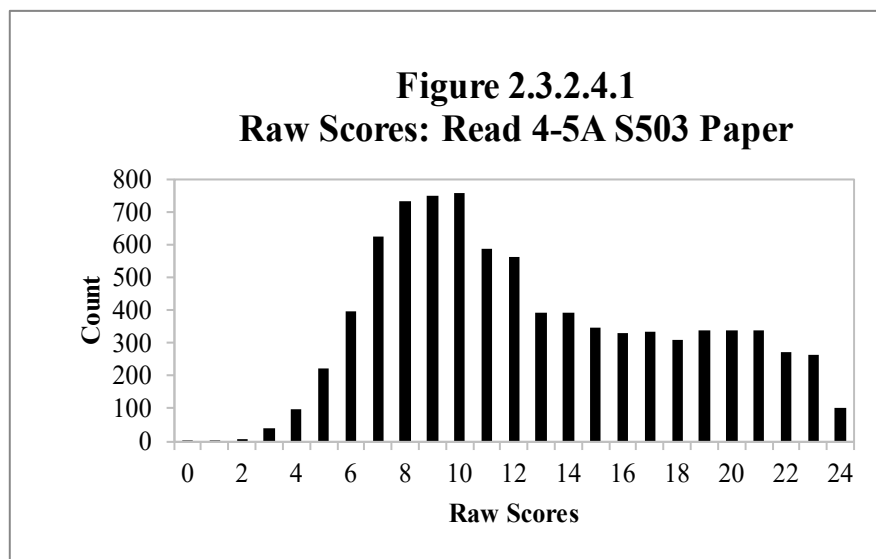


#### 2.3.2.4 Grades 4-5

**Table 2.3.2.4.1**

Raw Score Descriptive Statistics: Read 4-5 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	4,314	0	24	12.28	5.10
5	4,222	1	24	13.26	5.41
Total	8,536	0	24	12.76	5.28

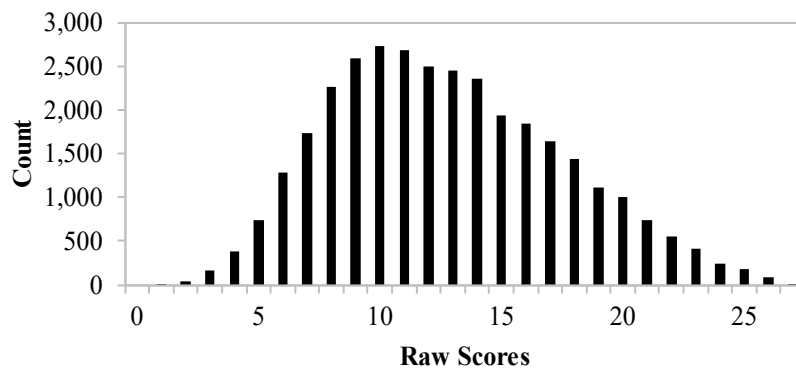


**Table 2.3.2.4.2**

Raw Score Descriptive Statistics: Read 4-5 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	16,407	1	27	11.93	4.45
5	16,770	1	27	13.56	4.87
Total	33,177	1	27	12.75	4.74

**Figure 2.3.2.4.2**  
**Raw Scores: Read 4-5B/C S503 Paper**



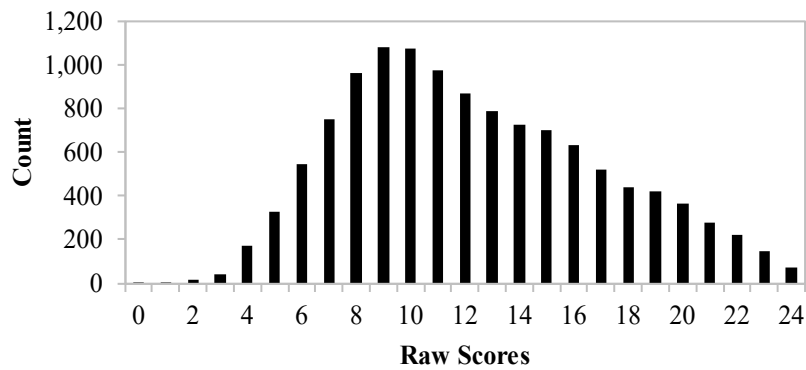
### 2.3.2.5 Grades 6-8

**Table 2.3.2.5.1**

Raw Score Descriptive Statistics: Read 6-8 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	4,098	0	24	11.54	4.48
7	4,019	1	24	12.31	4.69
8	4,018	2	24	13.05	4.75
<b>Total</b>	12,135	0	24	12.30	4.68

**Figure 2.3.2.5.1**  
**Raw Scores: Read 6-8A S503 Paper**

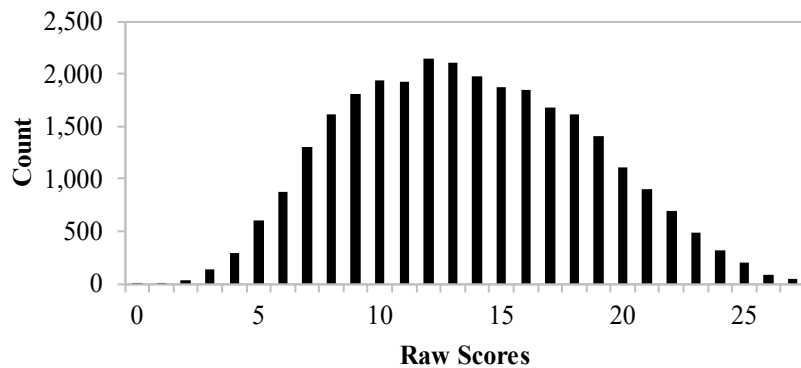


**Table 2.3.2.5.2**

Raw Score Descriptive Statistics: Read 6-8 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	10,316	1	27	12.30	4.46
7	9,553	1	27	13.54	4.84
8	9,192	0	27	15.17	5.11
Total	29,061	0	27	13.61	4.94

**Figure 2.3.2.5.2**  
**Raw Scores: Read 6-8B/C S503 Paper**





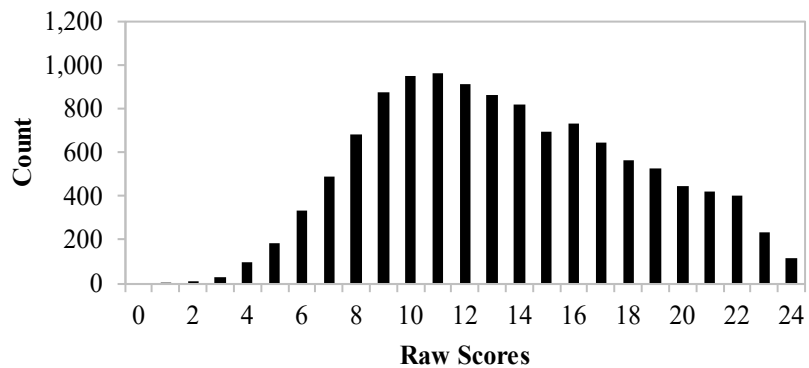
### 2.3.2.6 Grades 9-12

**Table 2.3.2.6.1**

Raw Score Descriptive Statistics: Read 9-12 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	4,605	1	24	12.54	4.63
10	3,289	2	24	13.60	4.71
11	2,754	2	24	14.33	4.74
12	1,339	2	24	14.83	4.66
Total	11,987	1	24	13.50	4.76

**Figure 2.3.2.6.1**  
**Raw Scores: Read 9-12A S503 Paper**

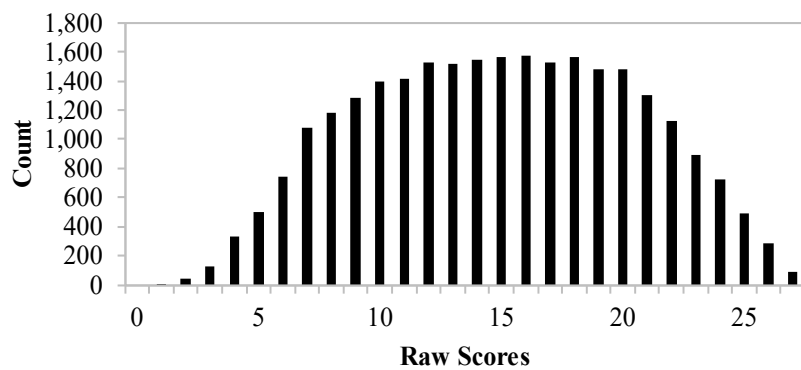


**Table 2.3.2.6.2**

Raw Score Descriptive Statistics: Read 9-12 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	7,362	1	27	13.99	5.40
10	7,411	1	27	14.83	5.49
11	6,837	1	27	15.65	5.62
12	5,181	2	27	15.30	5.58
Total	26,791	1	27	14.90	5.55

**Figure 2.3.2.6.2**  
**Raw Scores: Read 9-12B/C S503 Paper**



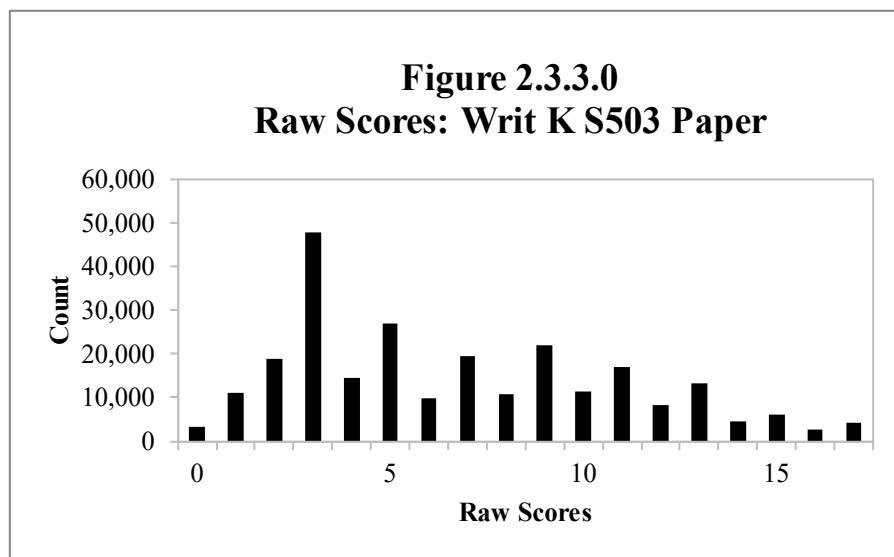
## 2.3.3 Writing

### 2.3.3.0 Kindergarten

**Table 2.3.3.0**

Raw Score Descriptive Statistics: Writ K S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	251,486	0	17	6.76	4.19
Total	251,486	0	17	6.76	4.19

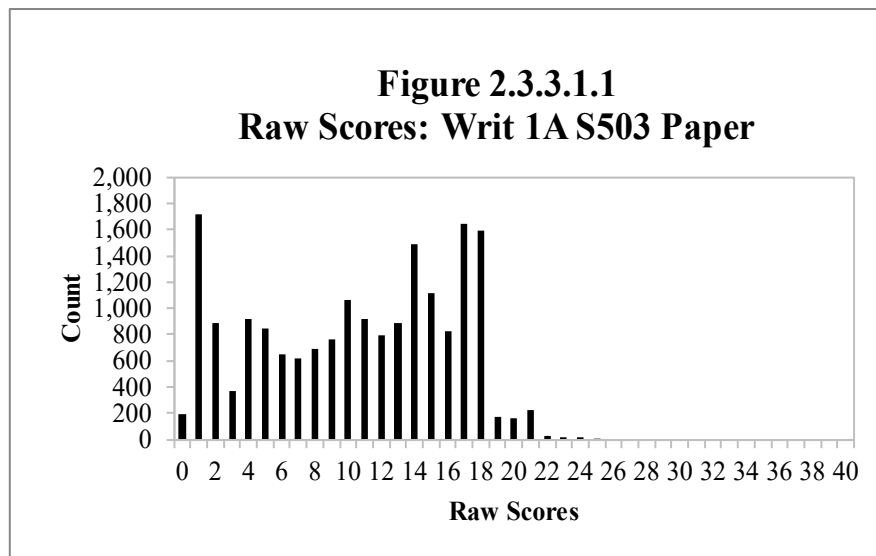


### 2.3.3.1 Grade 1

**Table 2.3.3.1.1**

Raw Score Descriptive Statistics: Writ 1 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	18,607	0	25	10.48	5.89
Total	18,607	0	25	10.48	5.89

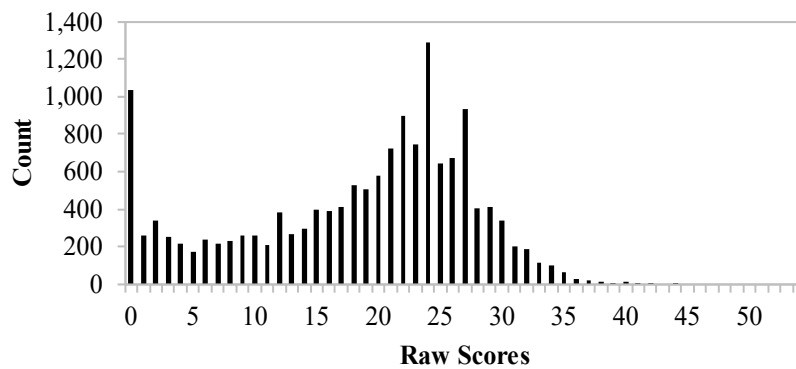


**Table 2.3.3.1.2**

Raw Score Descriptive Statistics: Writ 1 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,223	0	44	18.18	9.36
Total	15,223	0	44	18.18	9.36

**Figure 2.3.3.1.2**  
**Raw Scores: Writ 1B/C S503 Paper**

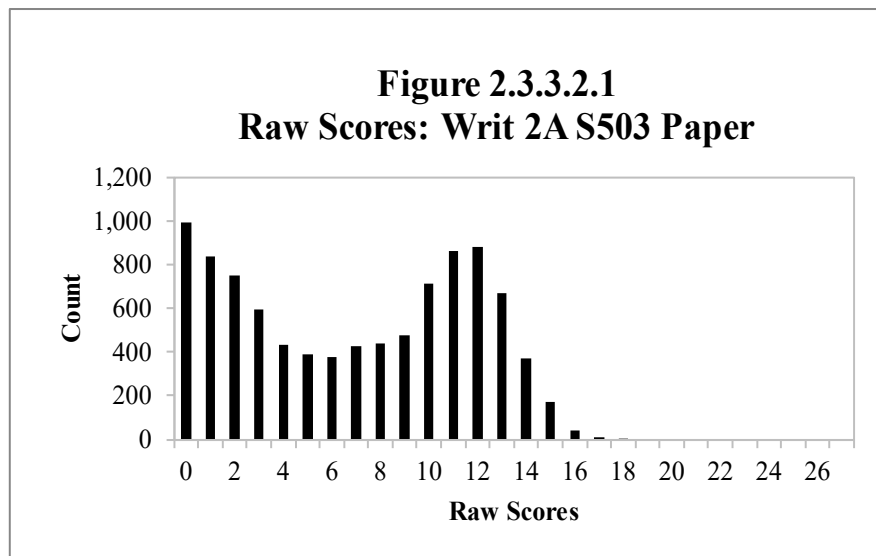


### 2.3.3.2 Grade 2

**Table 2.3.3.2.1**

Raw Score Descriptive Statistics: Writ 2 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	9,430	0	18	6.93	4.78
Total	9,430	0	18	6.93	4.78

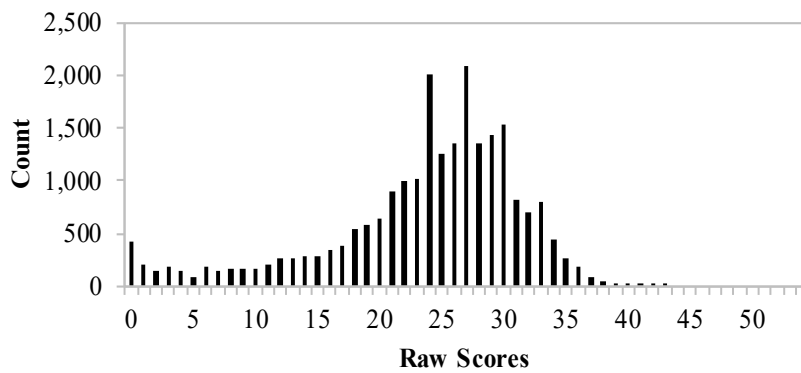


**Table 2.3.3.2.2**

Raw Score Descriptive Statistics: Writ 2 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	23,216	0	43	23.54	8.03
Total	23,216	0	43	23.54	8.03

**Figure 2.3.3.2.2**  
**Raw Scores: Writ 2B/C S503 Paper**

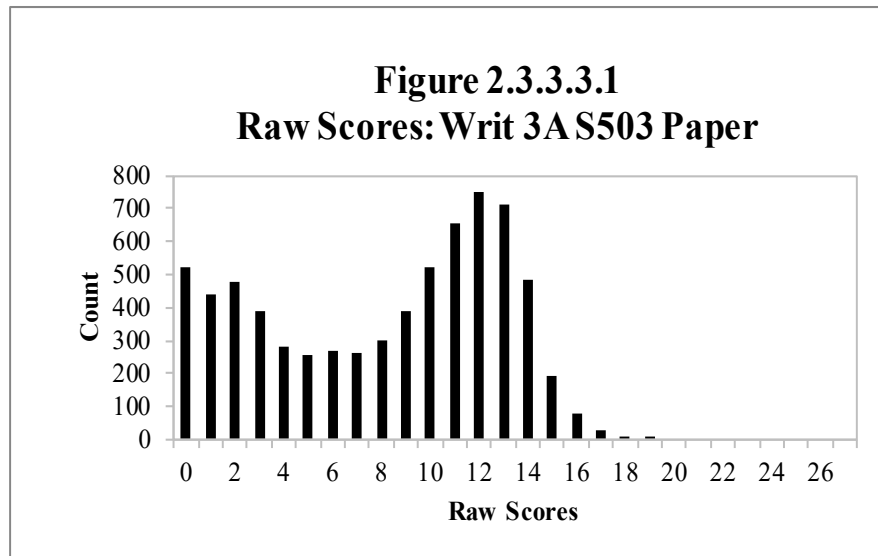


### 2.3.3.3 Grade 3

**Table 2.3.3.3.1**

Raw Score Descriptive Statistics: Writ 3 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	7,021	0	19	8.07	4.83
Total	7,021	0	19	8.07	4.83



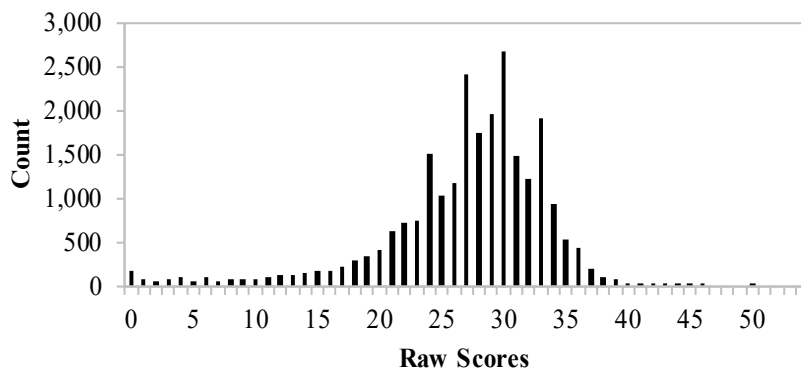


**Table 2.3.3.3.2**

Raw Score Descriptive Statistics: Writ 3 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	24,568	0	50	26.85	6.74
Total	24,568	0	50	26.85	6.74

**Figure 2.3.3.3.2**  
**Raw Scores: Writ 3B/C S503 Paper**



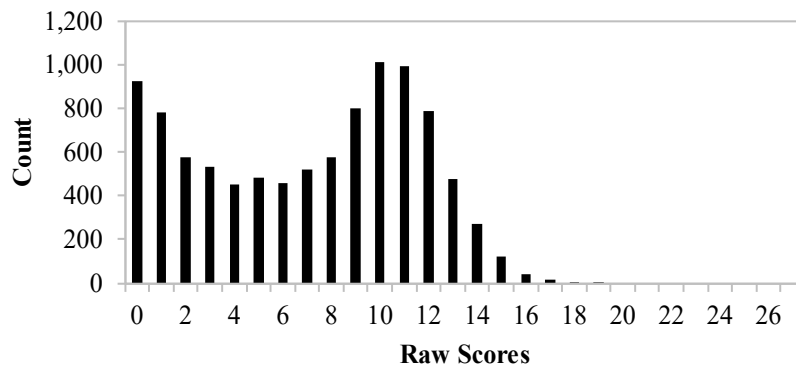
#### 2.3.3.4 Grades 4-5

**Table 2.3.3.4.1**

Raw Score Descriptive Statistics: Writ 4-5 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	5,034	0	19	6.69	4.44
5	4,809	0	19	7.43	4.46
Total	9,843	0	19	7.05	4.46

**Figure 2.3.3.4.1**  
**Raw Scores: Writ 4-5A S503 Paper**

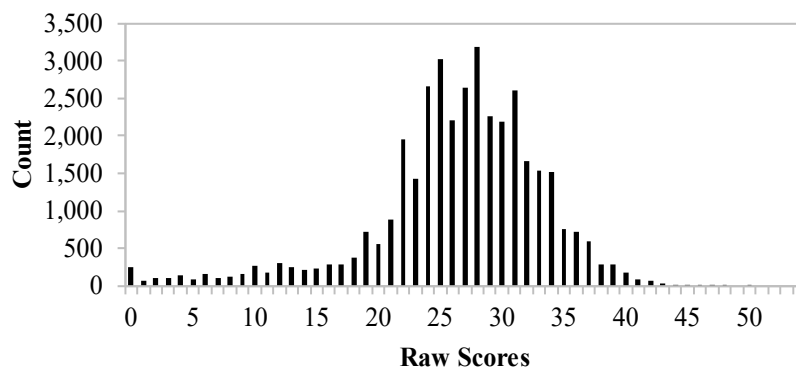


**Table 2.3.3.4.2**

Raw Score Descriptive Statistics: Writ 4-5 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	18,816	0	48	25.16	6.84
5	18,852	0	50	27.70	6.81
Total	37,668	0	50	26.43	6.94

**Figure 2.3.3.4.2**  
**Raw Scores: Writ 4-5B/C S503 Paper**

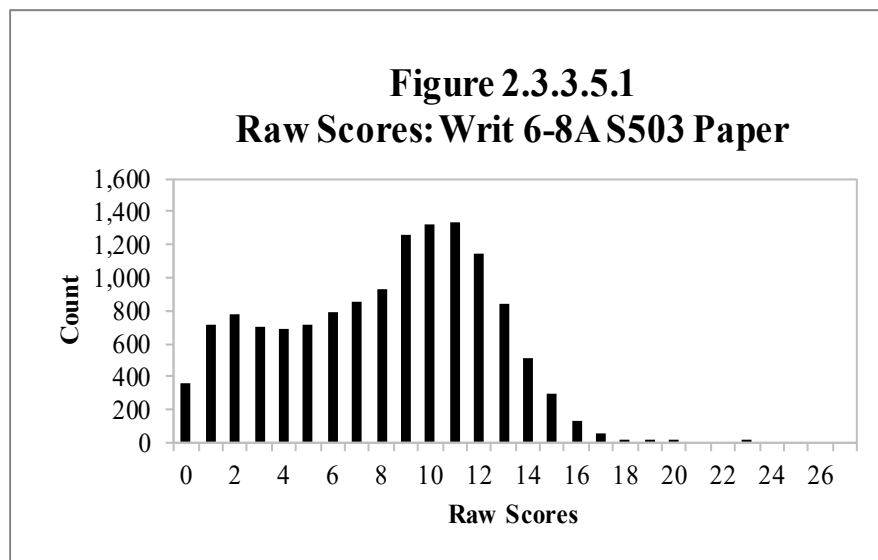


### 2.3.3.5 Grades 6-8

**Table 2.3.3.5.1**

Raw Score Descriptive Statistics: Writ 6-8 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	4,578	0	19	7.52	4.18
7	4,448	0	19	7.98	4.10
8	4,435	0	23	8.58	4.10
<b>Total</b>	<b>13,461</b>	<b>0</b>	<b>23</b>	<b>8.02</b>	<b>4.15</b>

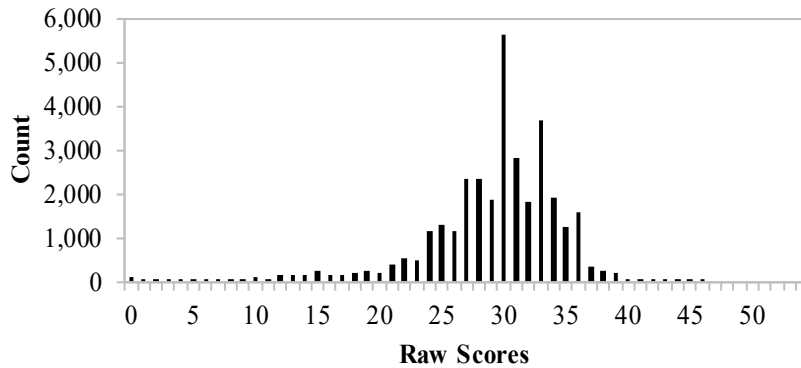


**Table 2.3.3.5.2**

Raw Score Descriptive Statistics: Writ 6-8 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	12,197	0	43	27.85	6.06
7	11,108	0	45	28.89	6.11
8	10,528	0	46	30.05	5.89
Total	33,833	0	46	28.88	6.09

**Figure 2.3.3.5.2**  
**Raw Scores: Writ 6-8B/C S503 Paper**



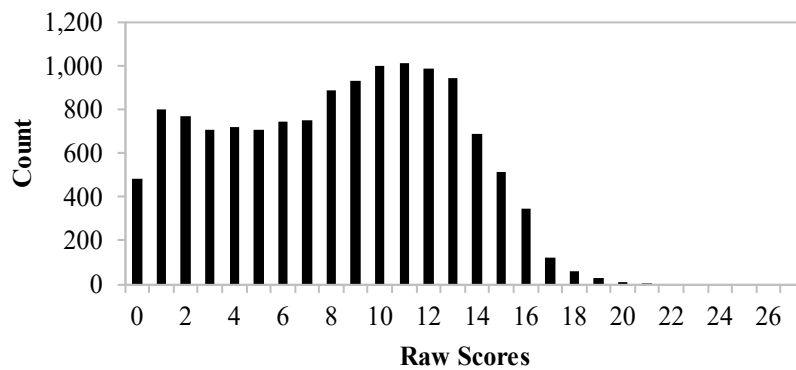
### 2.3.3.6 Grades 9-12

**Table 2.3.3.6.1**

Raw Score Descriptive Statistics: Writ 9-12 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	5,123	0	21	7.26	4.71
10	3,601	0	21	8.26	4.51
11	3,016	0	21	8.86	4.50
12	1,456	0	19	9.82	4.12
<b>Total</b>	<b>13,196</b>	<b>0</b>	<b>21</b>	<b>8.18</b>	<b>4.63</b>

**Figure 2.3.3.6.1**  
**Raw Scores: Writ 9-12A S503 Paper**

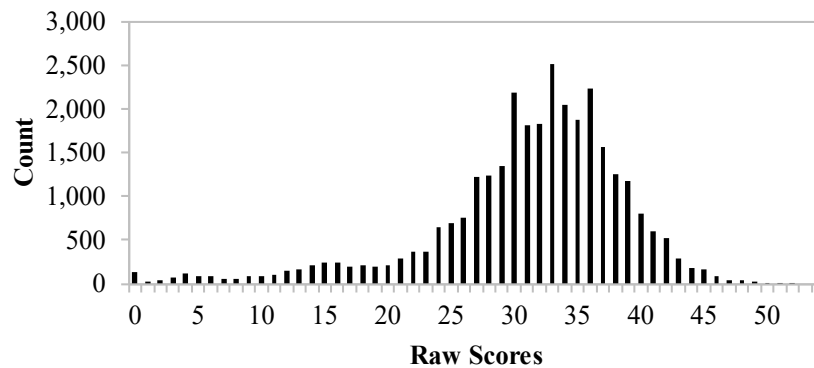


**Table 2.3.3.6.2**

Raw Score Descriptive Statistics: Writ 9-12 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	8,659	0	51	30.54	7.52
10	8,657	0	52	31.10	7.74
11	7,852	0	52	31.70	7.87
12	5,804	0	51	31.15	8.20
Total	30,972	0	52	31.11	7.81

**Figure 2.3.3.6.2**  
**Raw Scores: Writ 9-12B/C S503 Paper**



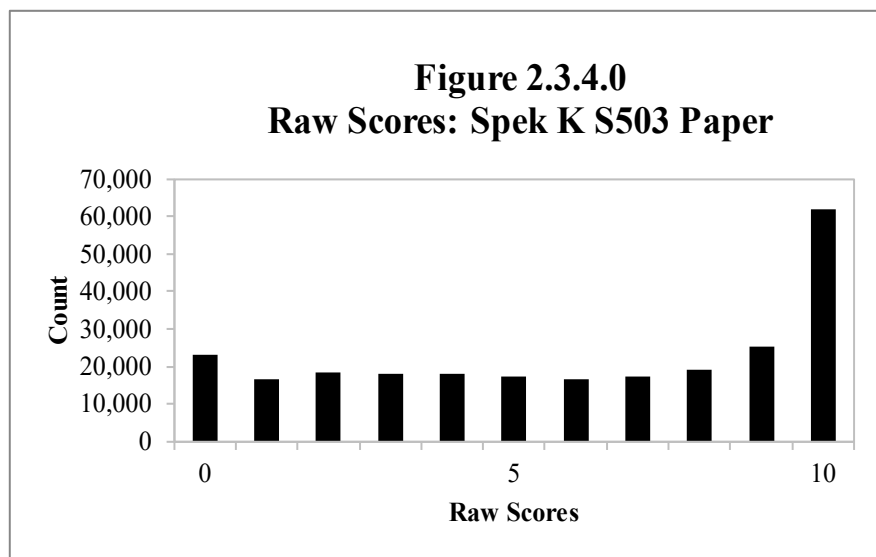
## 2.3.4 Speaking

### 2.3.4.0 Kindergarten

**Table 2.3.4.0**

Raw Score Descriptive Statistics: Spek K S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	251,483	0	10	5.91	3.51
Total	251,483	0	10	5.91	3.51



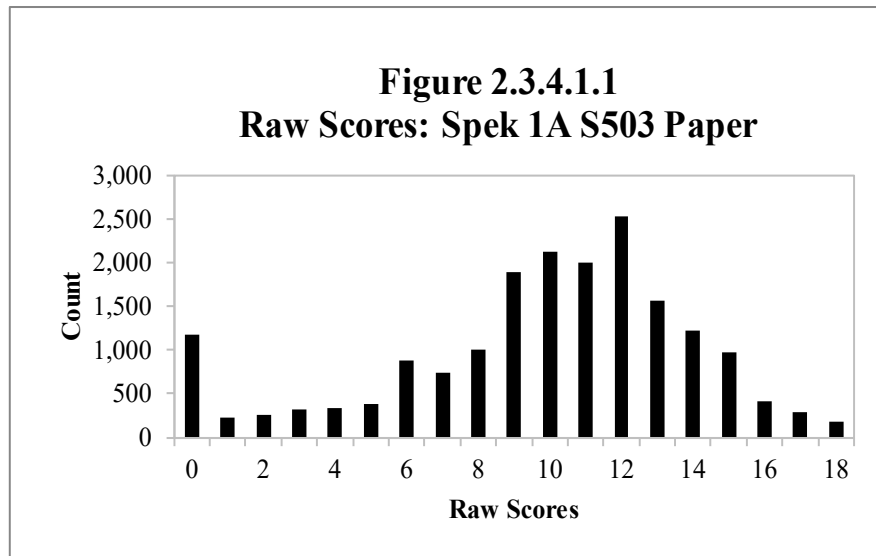


### 2.3.4.1 Grade 1

**Table 2.3.4.1.1**

Raw Score Descriptive Statistics: Spek 1 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	18,500	0	18	9.78	4.20
Total	18,500	0	18	9.78	4.20

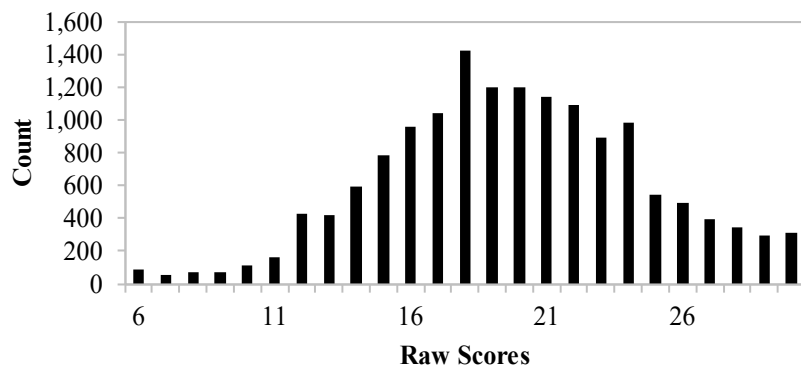


**Table 2.3.4.1.2**

Raw Score Descriptive Statistics: Spek 1 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,135	6	30	19.73	4.79
Total	15,135	6	30	19.73	4.79

**Figure 2.3.4.1.2**  
**Raw Scores: Spek 1B/C S503 Paper**

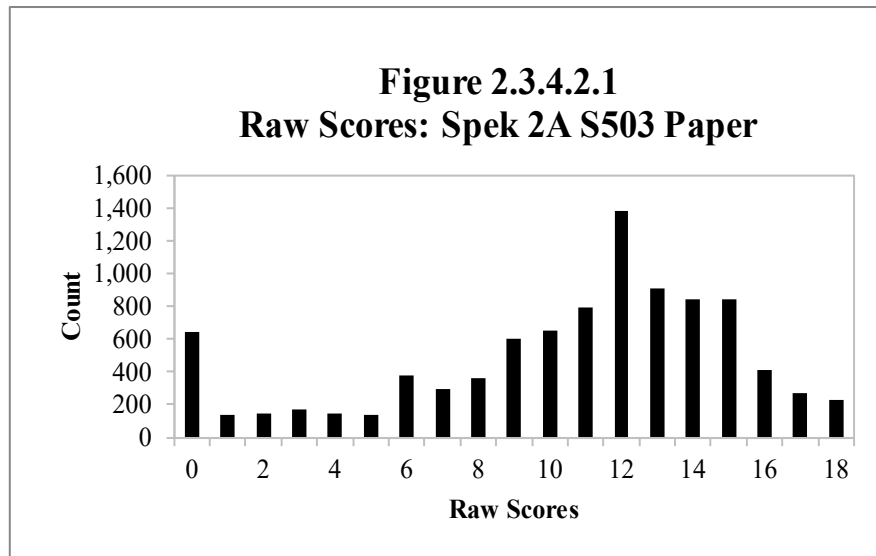


### 2.3.4.2 Grade 2

**Table 2.3.4.2.1**

Raw Score Descriptive Statistics: Spek 2 A S503 Paper

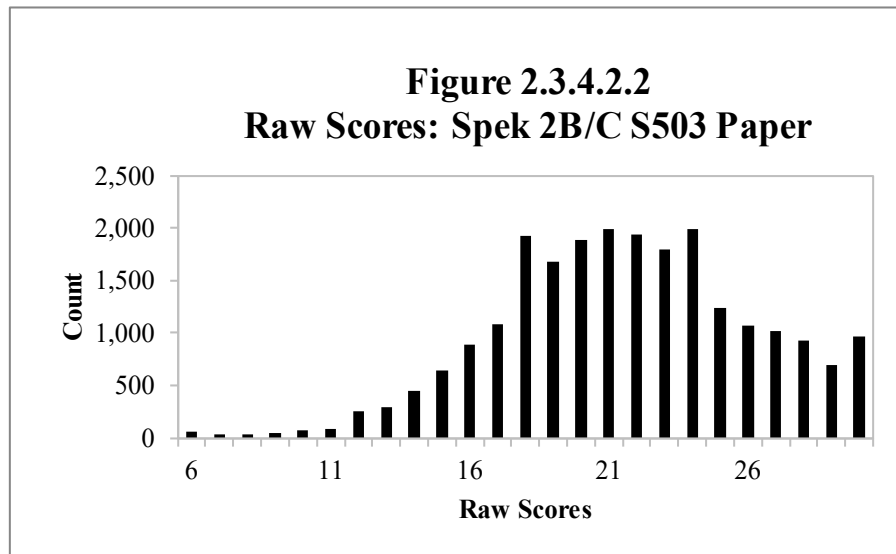
Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	9,364	0	18	10.51	4.65
Total	9,364	0	18	10.51	4.65



**Table 2.3.4.2.2**

Raw Score Descriptive Statistics: Spek 2 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	23,064	6	30	21.58	4.51
Total	23,064	6	30	21.58	4.51

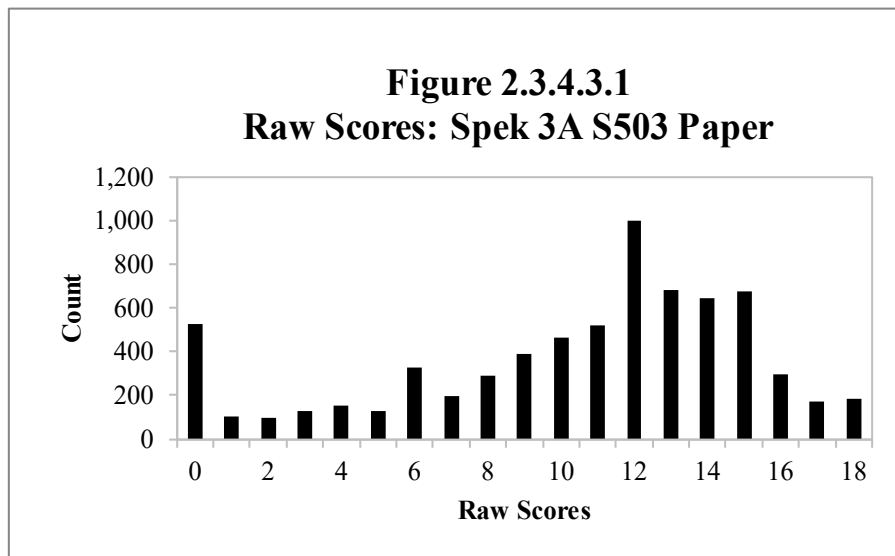


### 2.3.4.3 Grade 3

**Table 2.3.4.3.1**

Raw Score Descriptive Statistics: Spek 3 A S503 Paper

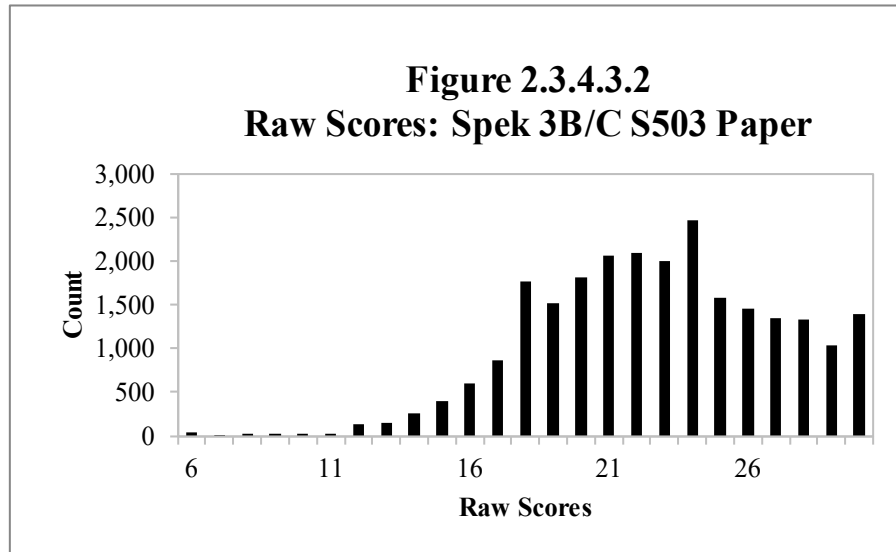
Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	6,977	0	18	10.39	4.77
Total	6,977	0	18	10.39	4.77



**Table 2.3.4.3.2**

Raw Score Descriptive Statistics: Spek 3 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	24,422	6	30	22.66	4.26
Total	24,422	6	30	22.66	4.26

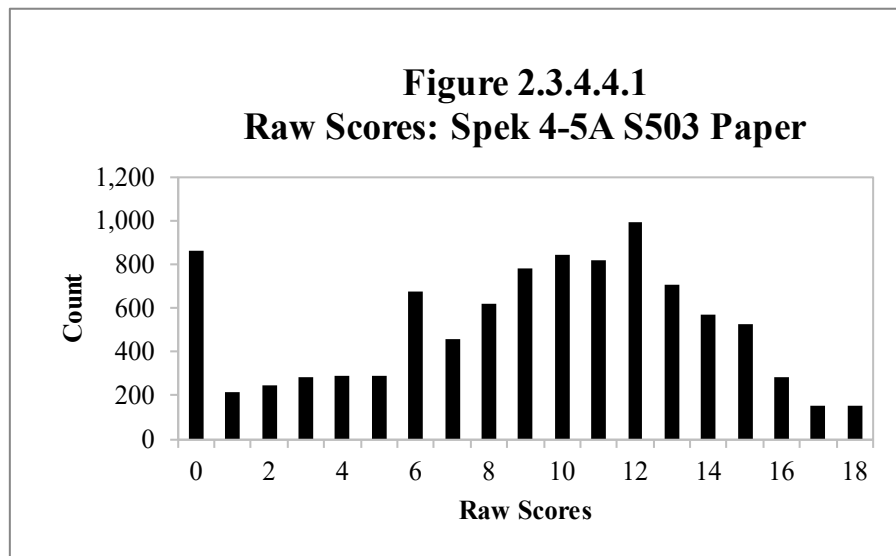


#### 2.3.4.4 Grades 4-5

**Table 2.3.4.4.1**

Raw Score Descriptive Statistics: Spek 4-5 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	4,991	0	18	8.96	4.75
5	4,762	0	18	9.01	4.74
Total	9,753	0	18	8.98	4.74

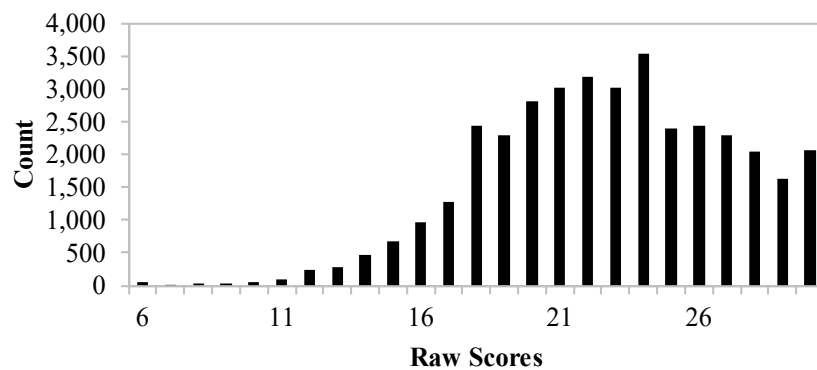


**Table 2.3.4.4.2**

Raw Score Descriptive Statistics: Spek 4-5 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	18,662	6	30	22.33	4.32
5	18,732	6	30	22.97	4.33
Total	37,394	6	30	22.65	4.34

**Figure 2.3.4.4.2**  
**Raw Scores: Spek 4-5B/C S503 Paper**



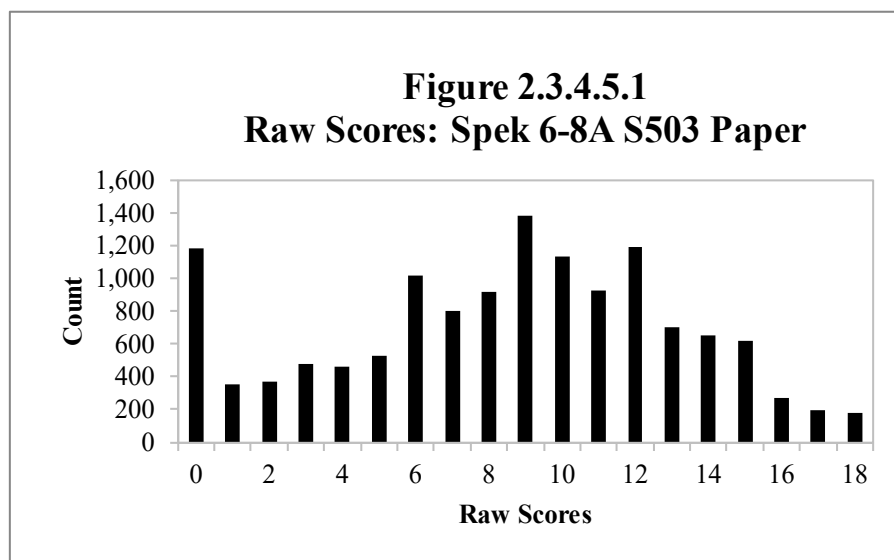


### 2.3.4.5 Grades 6-8

**Table 2.3.4.5.1**

Raw Score Descriptive Statistics: Spek 6-8 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	4,543	0	18	8.31	4.66
7	4,413	0	18	8.27	4.66
8	4,404	0	18	8.72	4.61
Total	13,360	0	18	8.43	4.65

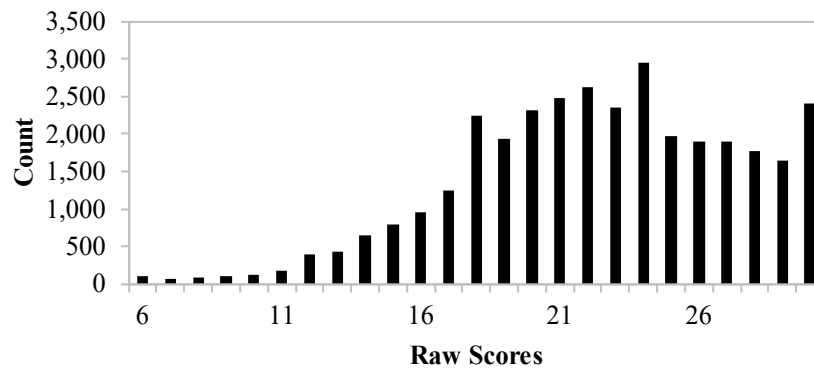


**Table 2.3.4.5.2**

Raw Score Descriptive Statistics: Spek 6-8 B/C S503 Paper

<b>Grade</b>	<b>No. of Students</b>	<b>Min.</b>	<b>Max.</b>	<b>Mean</b>	<b>Std. Dev.</b>
<b>6</b>	12,107	6	30	22.03	4.82
<b>7</b>	11,040	6	30	22.24	4.93
<b>8</b>	10,460	6	30	22.84	4.87
<b>Total</b>	33,607	6	30	22.35	4.89

**Figure 2.3.4.5.2**  
**Raw Scores: Spek 6-8B/C S503 Paper**

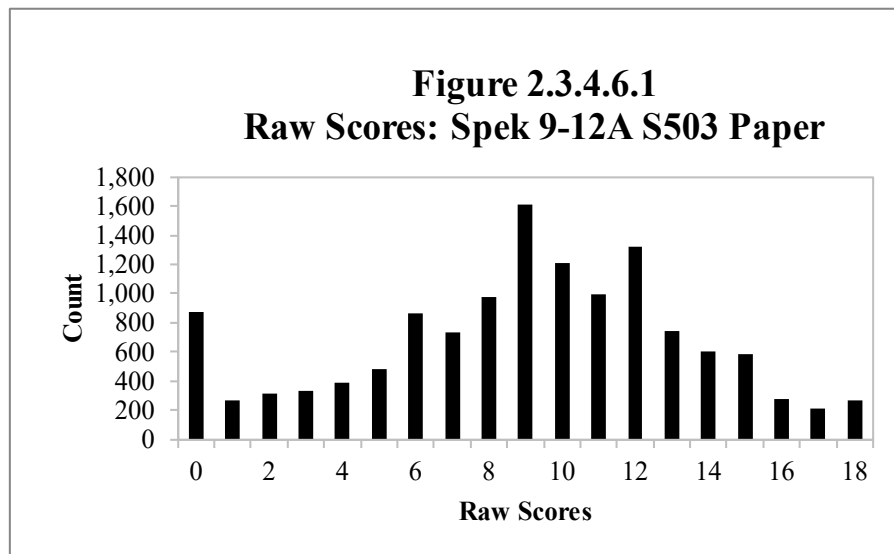


### 2.3.4.6 Grades 9-12

**Table 2.3.4.6.1**

Raw Score Descriptive Statistics: Spek 9-12 A S503 Paper

<b>Grade</b>	<b>No. of Students</b>	<b>Min.</b>	<b>Max.</b>	<b>Mean</b>	<b>Std. Dev.</b>
<b>9</b>	5,074	0	18	8.13	4.59
<b>10</b>	3,543	0	18	9.08	4.27
<b>11</b>	2,980	0	18	9.54	4.28
<b>12</b>	1,438	0	18	10.44	4.05
<b>Total</b>	13,035	0	18	8.97	4.44

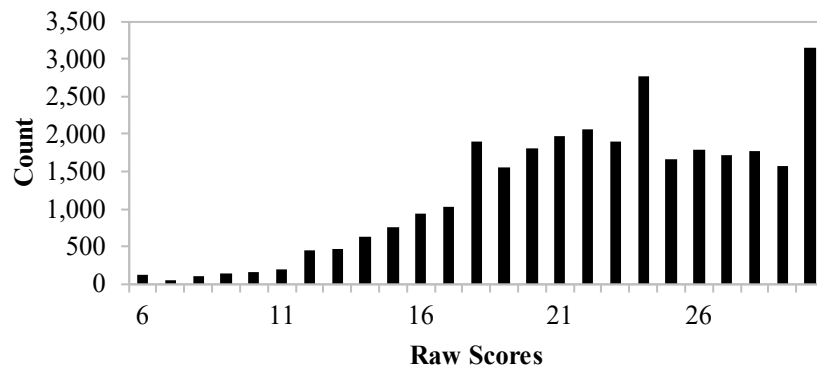


**Table 2.3.4.6.2**

Raw Score Descriptive Statistics: Spek 9-12 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	8,571	6	30	22.35	5.24
10	8,583	6	30	22.45	5.28
11	7,780	6	30	22.67	5.21
12	5,737	6	30	22.93	5.12
Total	30,671	6	30	22.57	5.23

**Figure 2.3.4.6.2**  
**Raw Scores: Spek 9-12B/C S503 Paper**



## ***2.4 Scale Score Distribution***

Figures and tables in this section relate to the ACCESS for ELLs scale scores on each test form. For each test form, we converted raw scores to vertically equated scale scores. Scale score distribution is presented by grade-level cluster and tier, and by grade-level cluster, combining tiers.

For each test form, the figure shows the distribution of the scale scores. Scale scores are plotted on the horizontal axis, grouped into units of five scale score points (e.g., 100–104, 105–109, 110–114, etc.). The number of students with scale scores falling into each range is plotted on the vertical axis. ACCESS Paper is tiered; therefore, depending on the tiers the students were placed in, their range of possible scale scores will vary.

The tables in this section show, by grade and by total for the grade-level cluster:

- The number of students in the analyses (count)
- The minimum observed scale score
- The maximum observed scale score
- The mean (average) scale score
- The standard deviation (std. dev.) of the scale score

As is the case for raw scores, scale score distributions are impacted by the test design and student population. Scale score distribution figures for the grade-level cluster incorporate distributions from Tier A and Tier B/C test forms and so will not appear smooth.

In the domain of Writing, task weighting results in raw scores that are not smoothly distributed. This distribution is also apparent in the distribution of scale scores.

The Kindergarten test design includes skipping and stopping rules intended to reduce testing time for young children; these rules also have an impact on the distribution of raw scores and subsequently on the distribution of scale scores, leading to less smooth distributions.

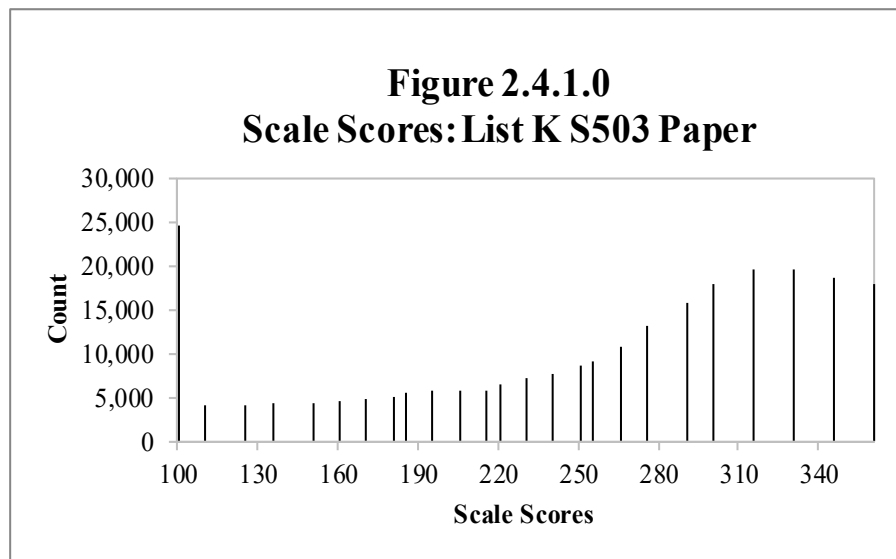
## 2.4.1 Listening

### 2.4.1.0 Kindergarten

**Table 2.4.1.0**

Scale Score Descriptive Statistics: List K S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	251,493	100	363	252.88	81.42
Total	251,493	100	363	252.88	81.42

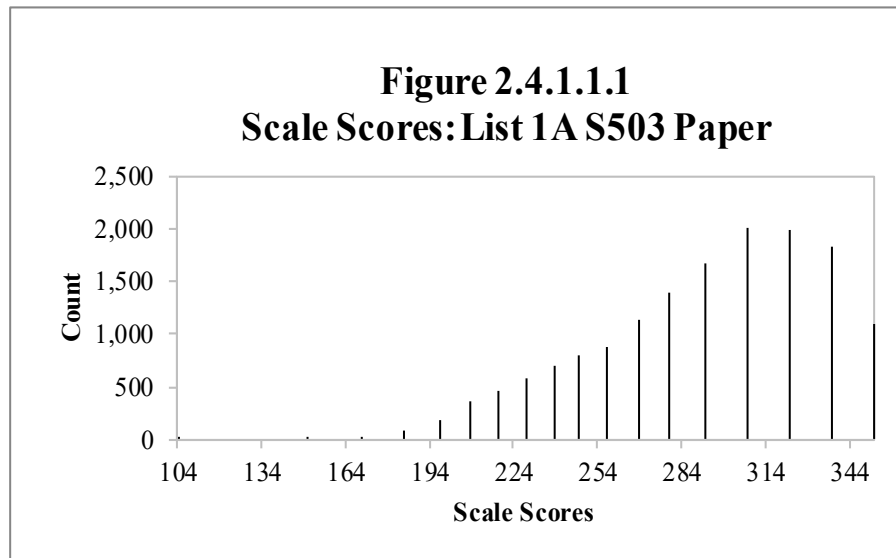


#### 2.4.1.1 Grade 1

**Table 2.4.1.1.1**

Scale Score Descriptive Statistics: List 1 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,238	104	352	288.97	41.53
Total	15,238	104	352	288.97	41.53

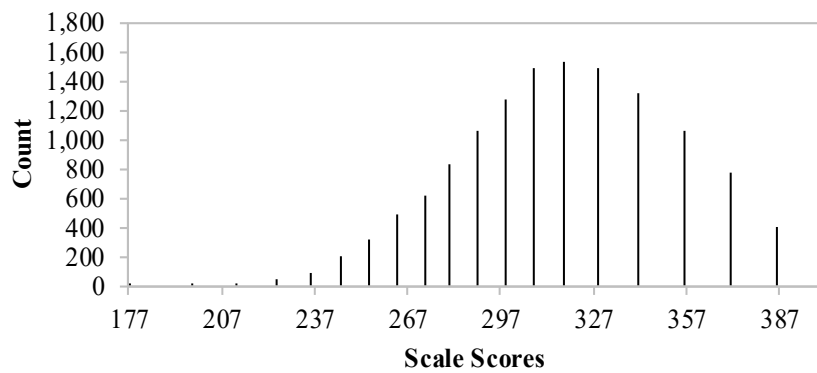


**Table 2.4.1.1.2**

Scale Score Descriptive Statistics: List 1 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	13,174	177	401	314.42	35.46
Total	13,174	177	401	314.42	35.46

**Figure 2.4.1.1.2**  
**Scale Scores: List 1B/C S503 Paper**



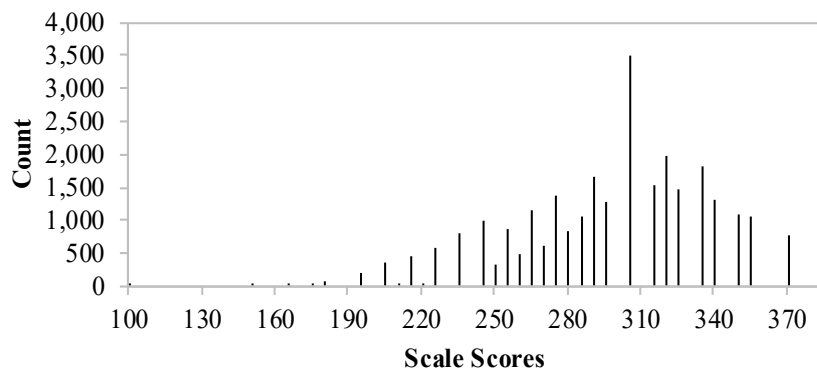


**Table 2.4.1.1.3**

Scale Score Descriptive Statistics: List 1 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	28,412	104	401	300.77	40.85
Total	28,412	104	401	300.77	40.85

**Figure 2.4.1.1.3**  
**Scale Scores: List 1 S503 Paper**

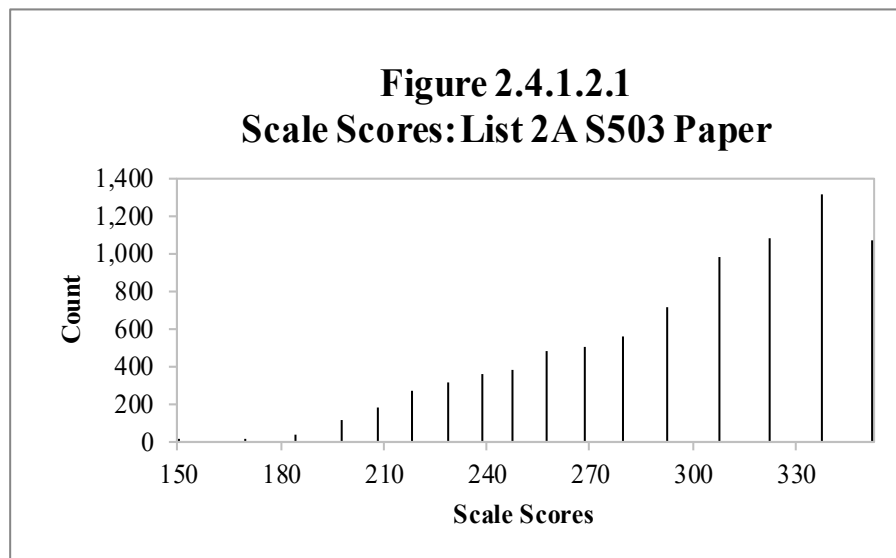


#### 2.4.1.2 Grade 2

**Table 2.4.1.2.1**

Scale Score Descriptive Statistics: List 2 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	8,389	150	352	294.86	44.09
Total	8,389	150	352	294.86	44.09

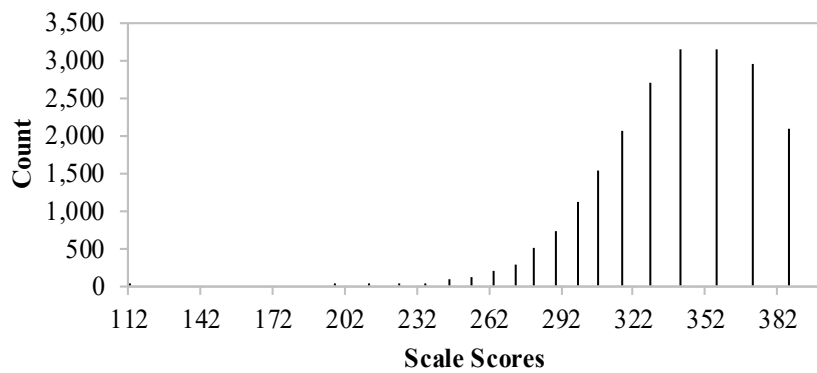


**Table 2.4.1.2.2**

Scale Score Descriptive Statistics: List 2 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	21,638	112	401	339.25	34.31
Total	21,638	112	401	339.25	34.31

**Figure 2.4.1.2.2**  
**Scale Scores: List 2B/C S503 Paper**

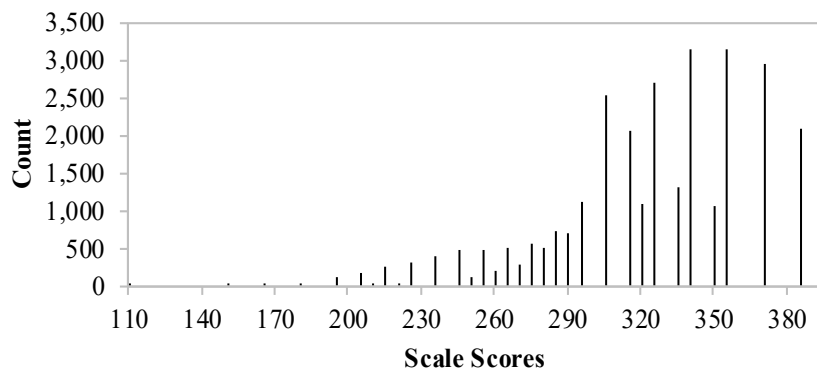


**Table 2.4.1.2.3**

Scale Score Descriptive Statistics: List 2 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	30,027	112	401	326.85	42.28
Total	30,027	112	401	326.85	42.28

**Figure 2.4.1.2.3**  
**Scale Scores: List 2 S503 Paper**

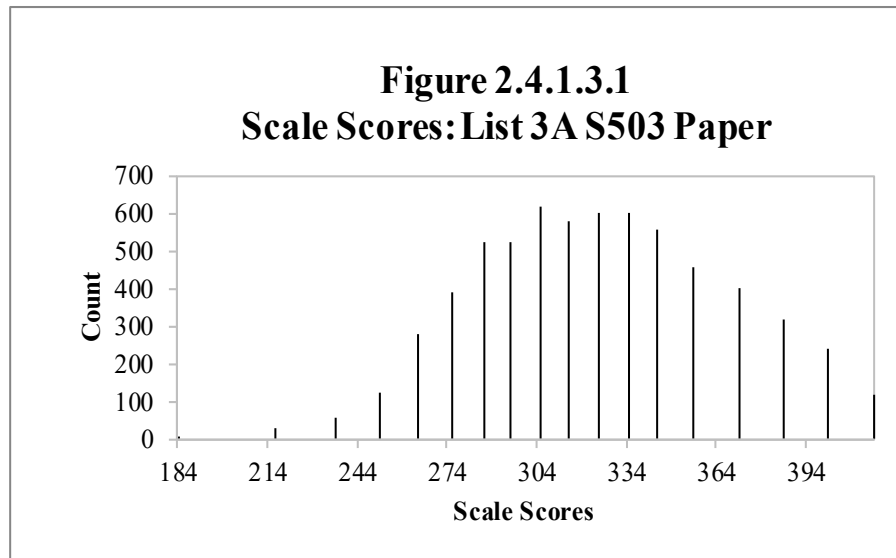


### 2.4.1.3 Grade 3

**Table 2.4.1.3.1**

Scale Score Descriptive Statistics: List 3 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	6,418	184	416	323.00	39.83
Total	6,418	184	416	323.00	39.83

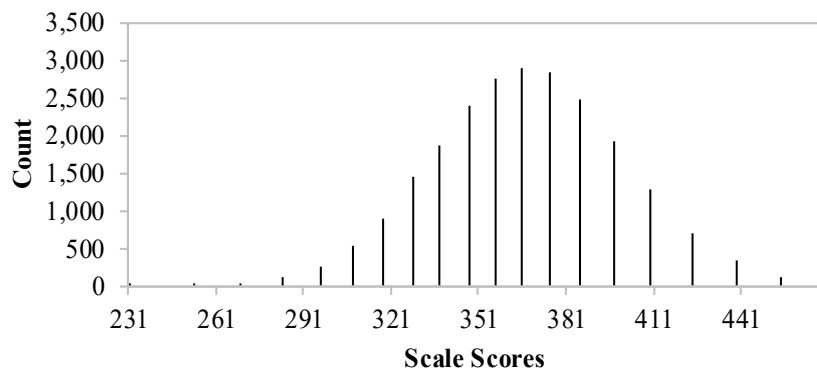


**Table 2.4.1.3.2**

Scale Score Descriptive Statistics: List 3 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	22,966	231	469	364.70	30.99
Total	22,966	231	469	364.70	30.99

**Figure 2.4.1.3.2**  
**Scale Scores: List 3B/C S503 Paper**

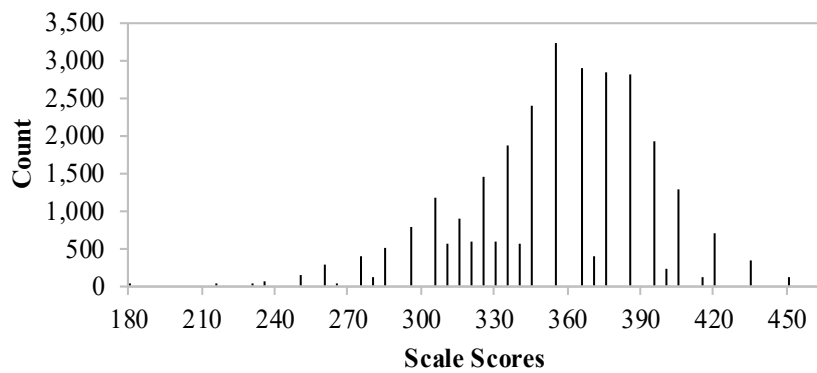


**Table 2.4.1.3.3**

Scale Score Descriptive Statistics: List 3 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	29,384	184	469	355.59	37.34
Total	29,384	184	469	355.59	37.34

**Figure 2.4.1.3.3**  
**Scale Scores: List 3 S503 Paper**

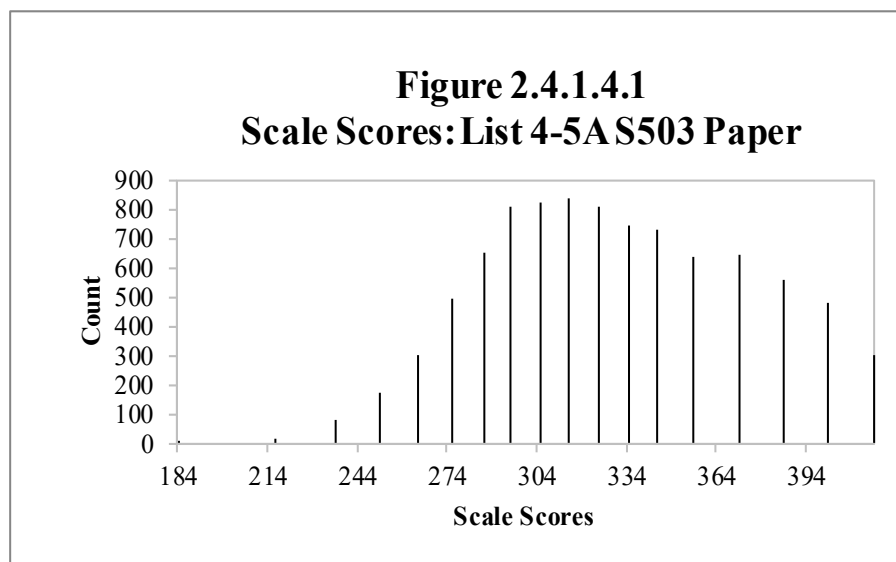


#### 2.4.1.4 Grades 4-5

**Table 2.4.1.4.1**

Scale Score Descriptive Statistics: List 4-5 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	4,625	184	416	325.48	40.94
5	4,477	184	416	330.37	42.80
Total	9,102	184	416	327.89	41.93



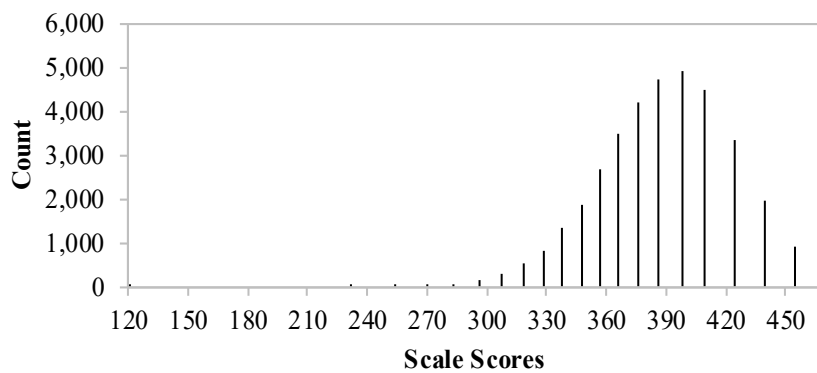


**Table 2.4.1.4.2**

Scale Score Descriptive Statistics: List 4-5 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	17,966	231	469	381.08	32.32
5	18,195	120	469	390.68	33.38
Total	36,161	120	469	385.91	33.21

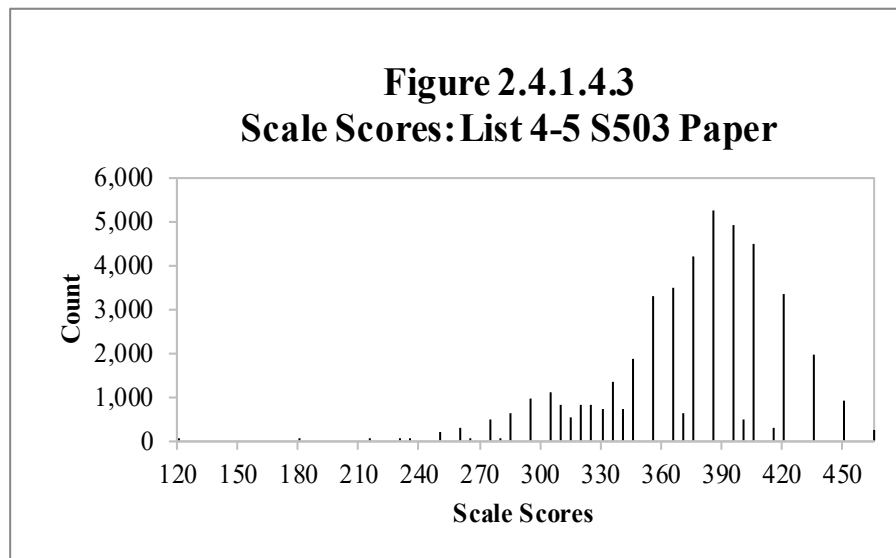
**Figure 2.4.1.4.2**  
**Scale Scores: List 4-5B/C S503 Paper**



**Table 2.4.1.4.3**

Scale Score Descriptive Statistics: List 4-5 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	22,591	184	469	369.69	40.95
5	22,672	120	469	378.77	42.81
Total	45,263	120	469	374.24	42.13

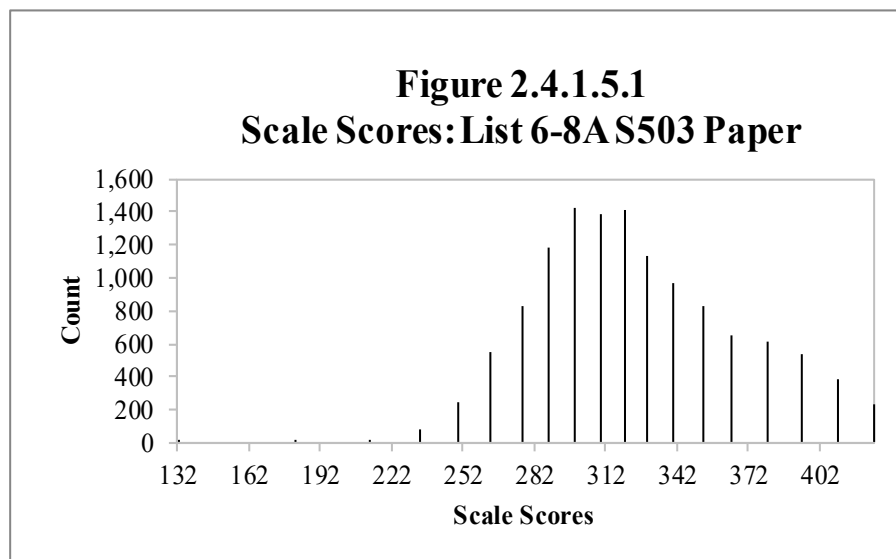


#### 2.4.1.5 Grades 6-8

**Table 2.4.1.5.1**

Scale Score Descriptive Statistics: List 6-8 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	4,267	132	424	320.37	40.02
7	4,131	181	424	323.00	40.83
8	4,101	212	424	325.76	41.34
<b>Total</b>	<b>12,499</b>	<b>132</b>	<b>424</b>	<b>323.01</b>	<b>40.78</b>

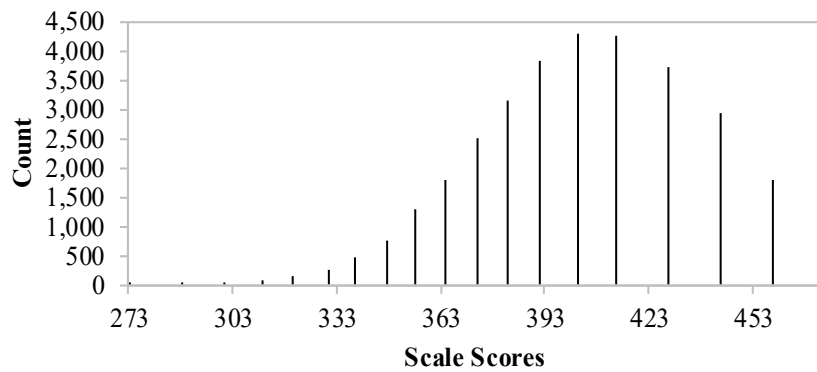


**Table 2.4.1.5.2**

Scale Score Descriptive Statistics: List 6-8 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	11,605	273	473	396.06	31.24
7	10,557	288	473	401.04	32.80
8	9,996	288	473	408.58	33.23
Total	32,158	273	473	401.59	32.78

**Figure 2.4.1.5.2**  
**Scale Scores: List 6-8B/C S503 Paper**

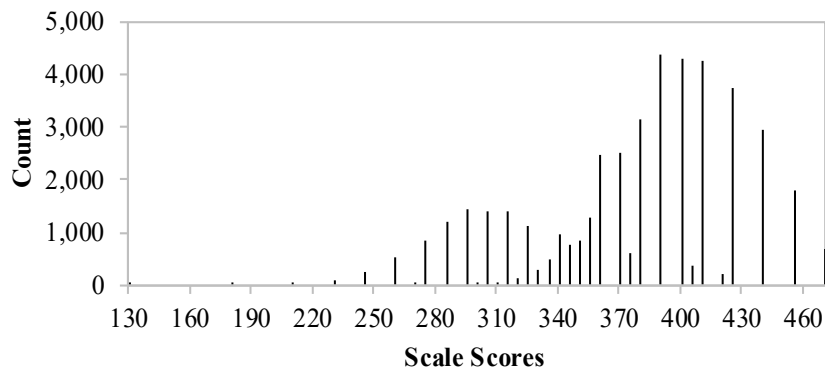


**Table 2.4.1.5.3**

Scale Score Descriptive Statistics: List 6-8 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	15,872	132	473	375.71	47.65
7	14,688	181	473	379.09	49.73
8	14,097	212	473	384.49	51.91
Total	44,657	132	473	379.59	49.84

**Figure 2.4.1.5.3**  
**Scale Scores: List 6-8 S503 Paper**

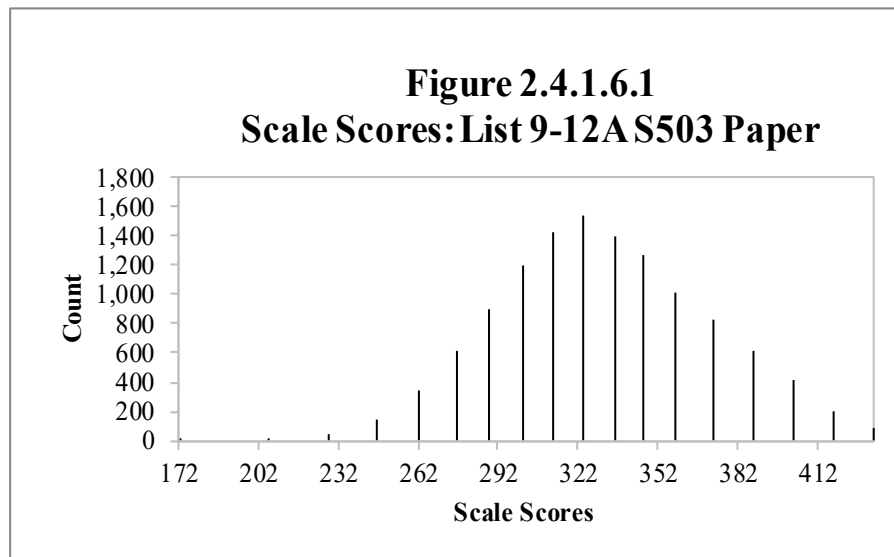


#### 2.4.1.6 Grades 9-12

**Table 2.4.1.6.1**

Scale Score Descriptive Statistics: List 9-12 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	4,625	172	432	324.73	37.92
10	3,284	205	432	330.42	38.04
11	2,777	205	432	333.01	37.26
12	1,363	246	432	338.87	36.97
<b>Total</b>	<b>12,049</b>	<b>172</b>	<b>432</b>	<b>329.79</b>	<b>37.98</b>

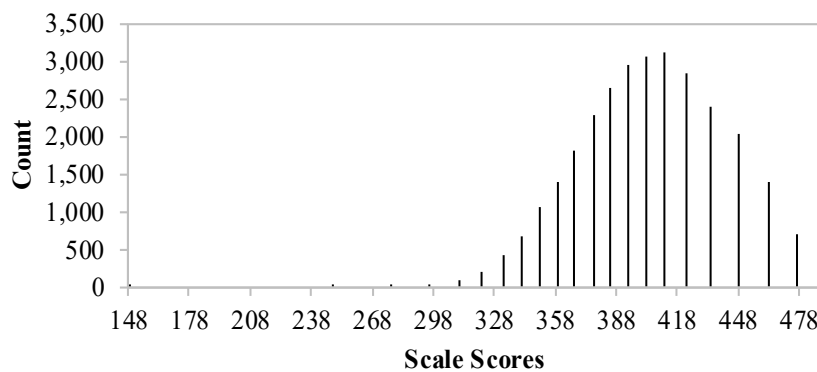


**Table 2.4.1.6.2**

Scale Score Descriptive Statistics: List 9-12 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	8,244	277	490	400.33	33.88
10	8,227	148	490	402.33	35.50
11	7,419	148	490	403.70	35.91
12	5,475	248	490	401.32	35.46
Total	29,365	148	490	401.92	35.17

**Figure 2.4.1.6.2**  
**Scale Scores: List 9-12B/C S503 Paper**

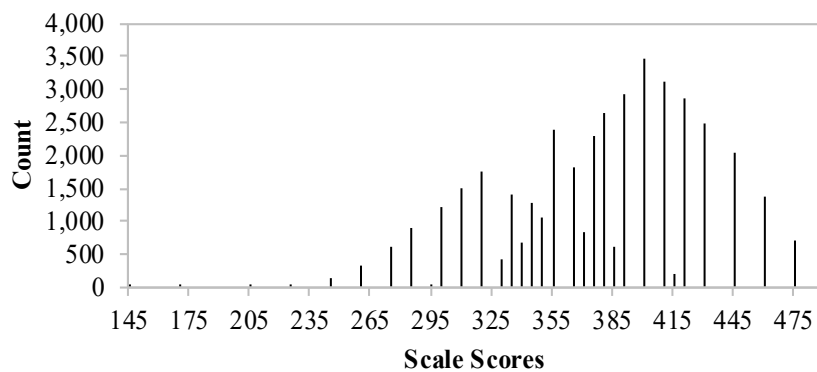


**Table 2.4.1.6.3**

Scale Score Descriptive Statistics: List 9-12 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	12,869	172	490	373.16	50.67
10	11,511	148	490	381.82	48.66
11	10,196	148	490	384.44	48.02
12	6,838	246	490	388.87	43.60
Total	41,414	148	490	380.94	48.68

**Figure 2.4.1.6.3**  
**Scale Scores: List 9-12 S503 Paper**





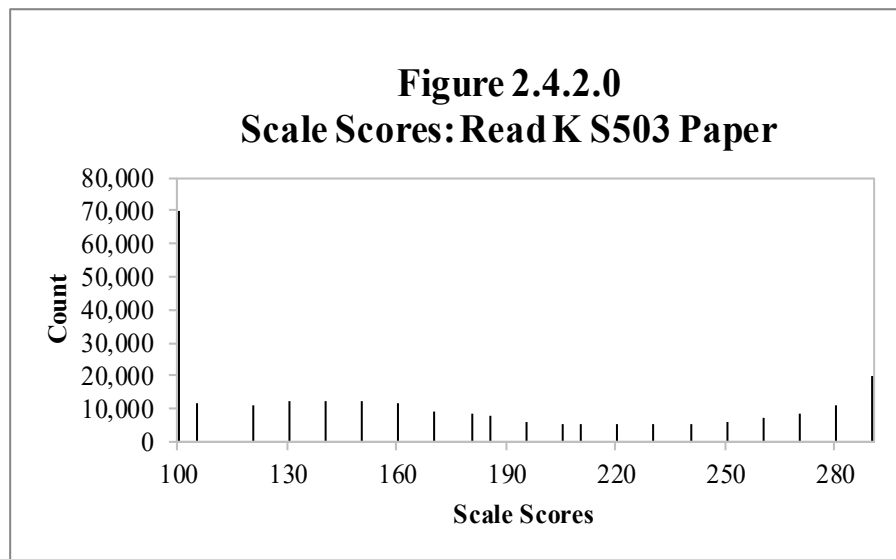
## 2.4.2 Reading

### 2.4.2.0 Kindergarten

**Table 2.4.2.0**

Scale Score Descriptive Statistics: Read K S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	251,491	100	290	170.27	67.13
Total	251,491	100	290	170.27	67.13

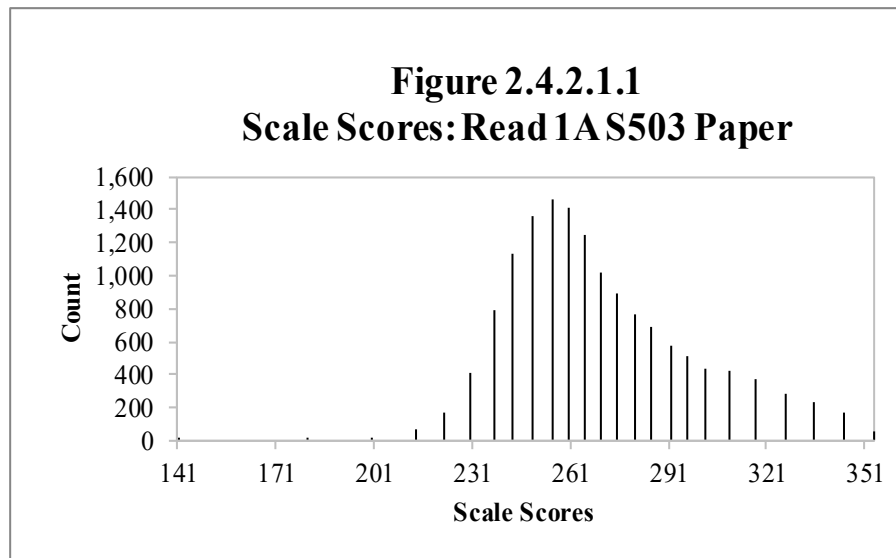


### 2.4.2.1 Grade 1

**Table 2.4.2.1.1**

Scale Score Descriptive Statistics: Read 1 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	14,522	141	353	269.08	26.53
Total	14,522	141	353	269.08	26.53

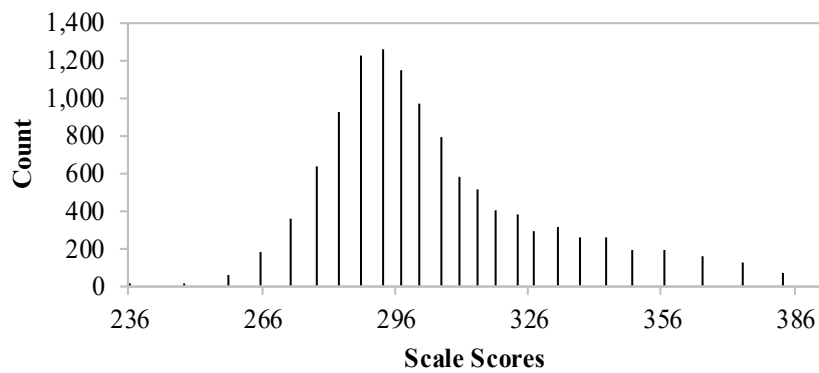


**Table 2.4.2.1.2**

Scale Score Descriptive Statistics: Read 1 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	11,346	236	392	303.94	23.66
Total	11,346	236	392	303.94	23.66

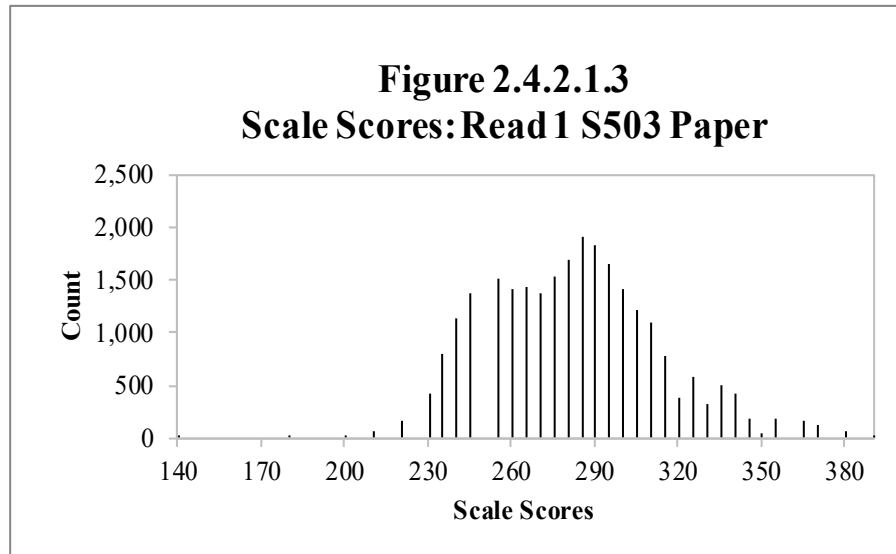
**Figure 2.4.2.1.2**  
**Scale Scores: Read 1B/C S503 Paper**



**Table 2.4.2.1.3**

Scale Score Descriptive Statistics: Read 1 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	25,868	141	392	284.37	30.66
Total	25,868	141	392	284.37	30.66

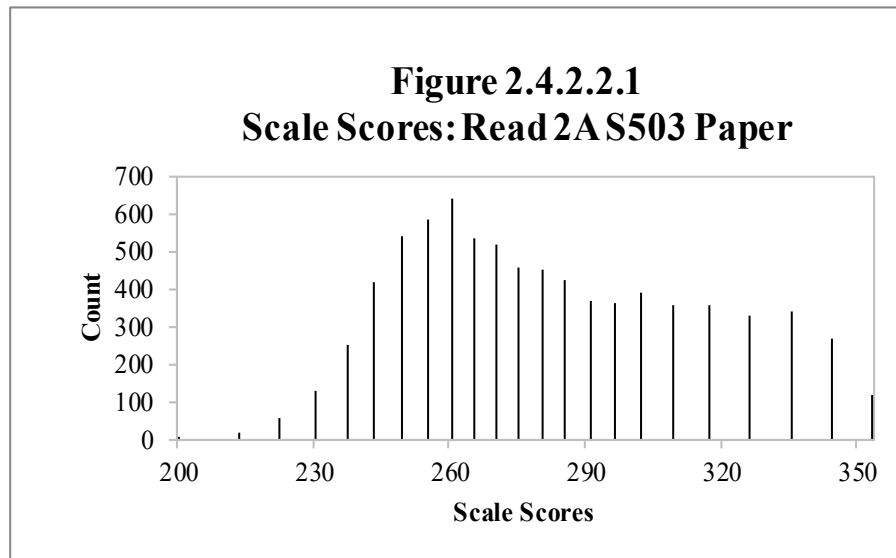


#### 2.4.2.2 Grade 2

**Table 2.4.2.2.1**

Scale Score Descriptive Statistics: Read 2 A S503 Paper

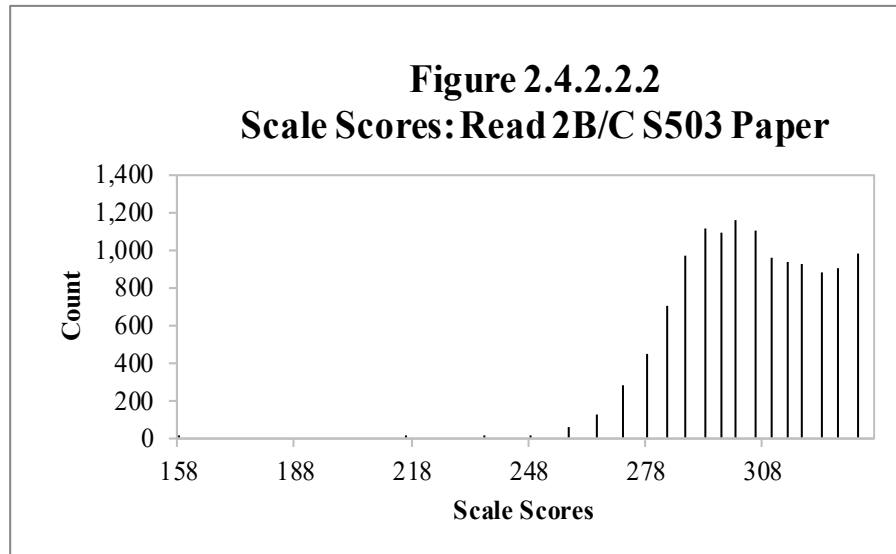
Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	7,913	200	353	280.89	30.91
Total	7,913	200	353	280.89	30.91



**Table 2.4.2.2.2**

Scale Score Descriptive Statistics: Read 2 B/C S503 Paper

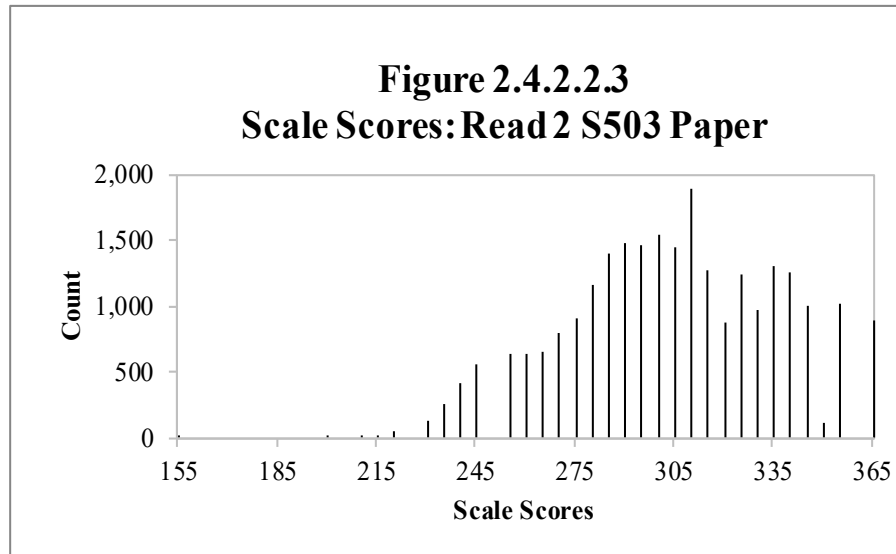
Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	18,951	158	392	321.98	29.32
Total	18,951	158	392	321.98	29.32



**Table 2.4.2.2.3**

Scale Score Descriptive Statistics: Read 2 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	26,864	158	392	309.88	35.20
Total	26,864	158	392	309.88	35.20

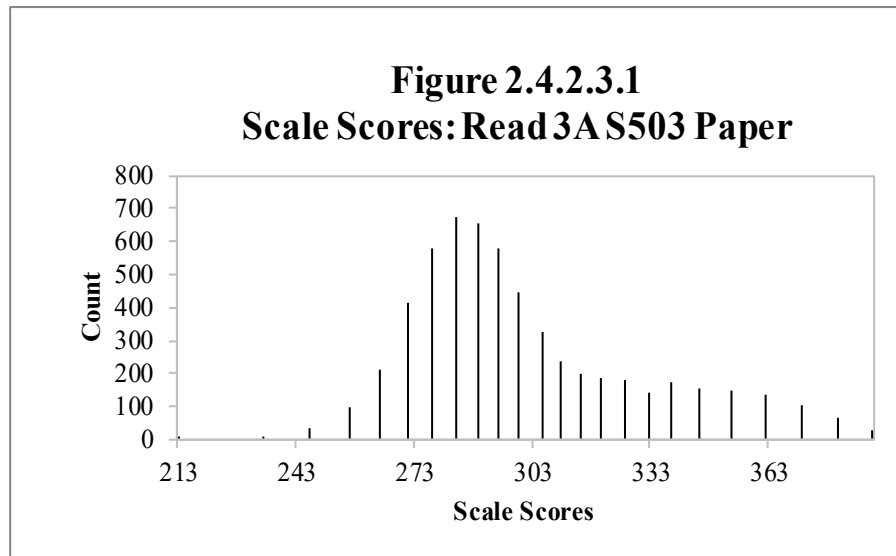


### 2.4.2.3 Grade 3

**Table 2.4.2.3.1**

Scale Score Descriptive Statistics: Read 3 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	5,791	213	389	300.37	28.42
Total	5,791	213	389	300.37	28.42



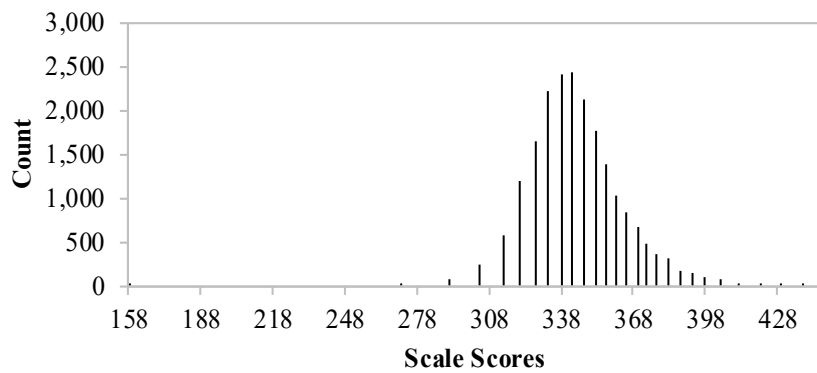


**Table 2.4.2.3.2**

Scale Score Descriptive Statistics: Read 3 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	20,451	158	447	345.04	19.15
Total	20,451	158	447	345.04	19.15

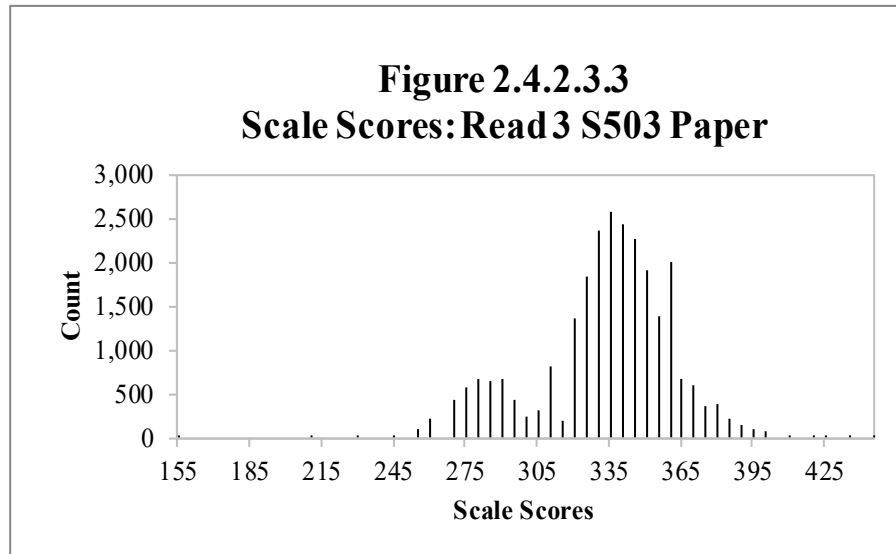
**Figure 2.4.2.3.2**  
**Scale Scores: Read 3B/C S503 Paper**



**Table 2.4.2.3.3**

Scale Score Descriptive Statistics: Read 3 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	26,242	158	447	335.18	28.41
Total	26,242	158	447	335.18	28.41

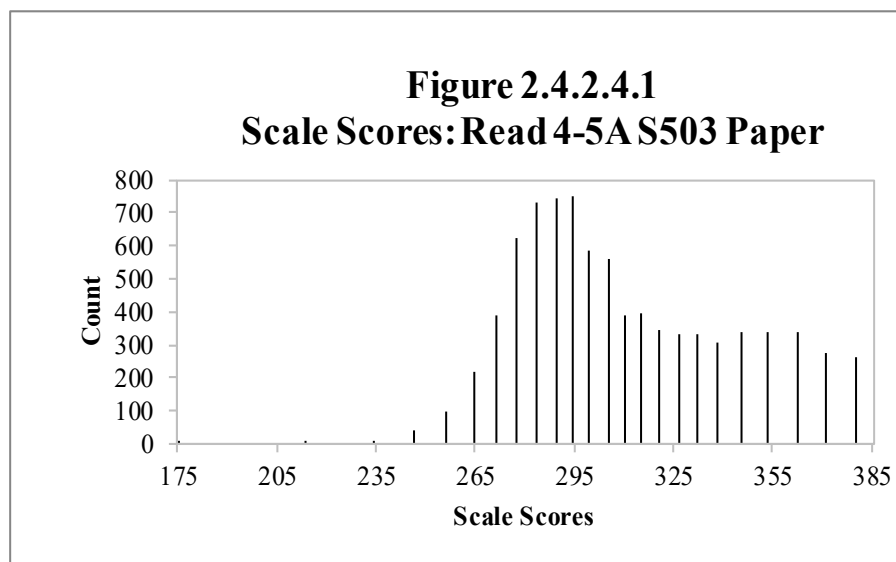


#### 2.4.2.4 Grades 4-5

**Table 2.4.2.4.1**

Scale Score Descriptive Statistics: Read 4-5 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	4,292	175	389	307.37	30.92
5	4,203	213	389	313.39	33.20
Total	8,495	175	389	310.35	32.21

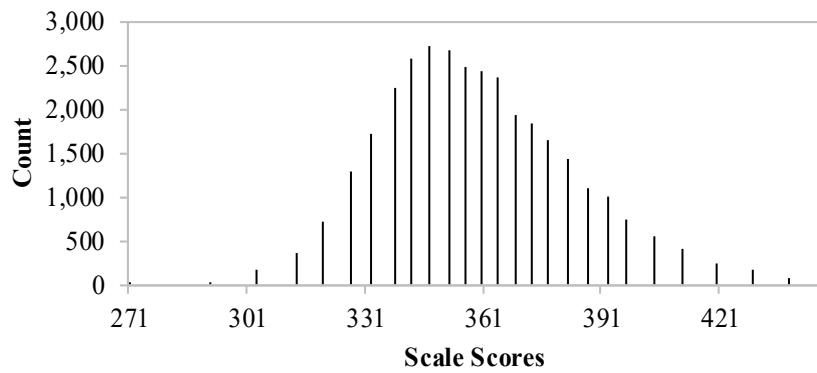


**Table 2.4.2.4.2**

Scale Score Descriptive Statistics: Read 4-5 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	16,329	271	447	355.05	21.29
5	16,688	271	447	362.85	23.52
Total	33,017	271	447	358.99	22.78

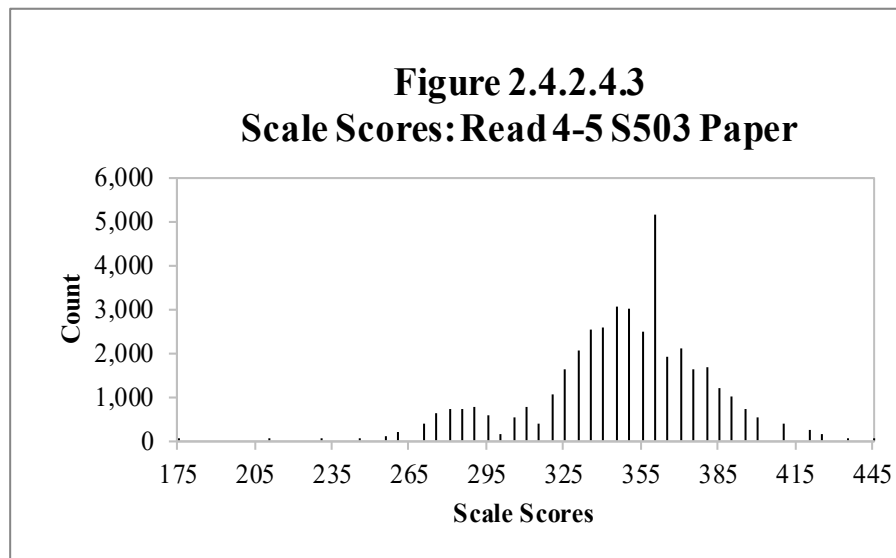
**Figure 2.4.2.4.2**  
**Scale Scores: Read 4-5B/C S503 Paper**



**Table 2.4.2.4.3**

Scale Score Descriptive Statistics: Read 4-5 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	20,621	175	447	345.13	30.54
5	20,891	213	447	352.90	32.51
Total	41,512	175	447	349.04	31.78

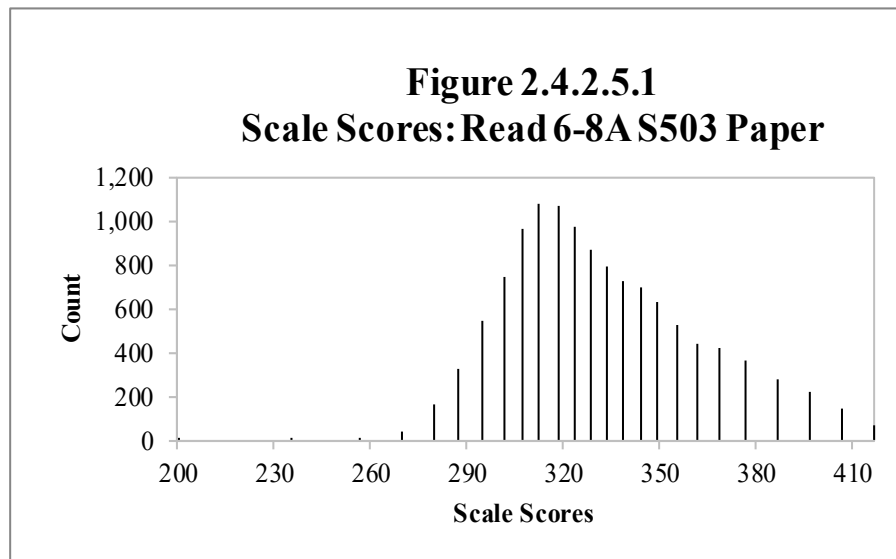


#### 2.4.2.5 Grades 6-8

**Table 2.4.2.5.1**

Scale Score Descriptive Statistics: Read 6-8 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	4,087	200	416	325.93	26.39
7	4,012	235	416	330.53	28.02
8	4,009	256	416	334.99	28.85
<b>Total</b>	<b>12,108</b>	<b>200</b>	<b>416</b>	<b>330.46</b>	<b>28.01</b>

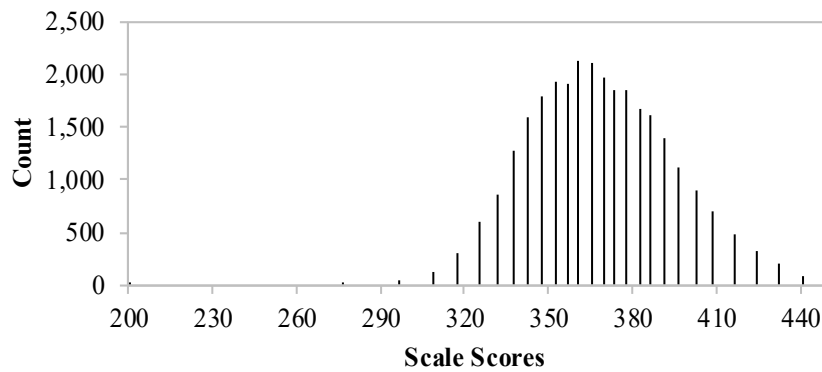


**Table 2.4.2.5.2**

Scale Score Descriptive Statistics: Read 6-8 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	10,261	276	448	361.19	21.13
7	9,489	276	448	367.12	23.12
8	9,118	200	448	374.94	25.05
Total	28,868	200	448	367.48	23.75

**Figure 2.4.2.5.2**  
**Scale Scores: Read 6-8B/C S503 Paper**

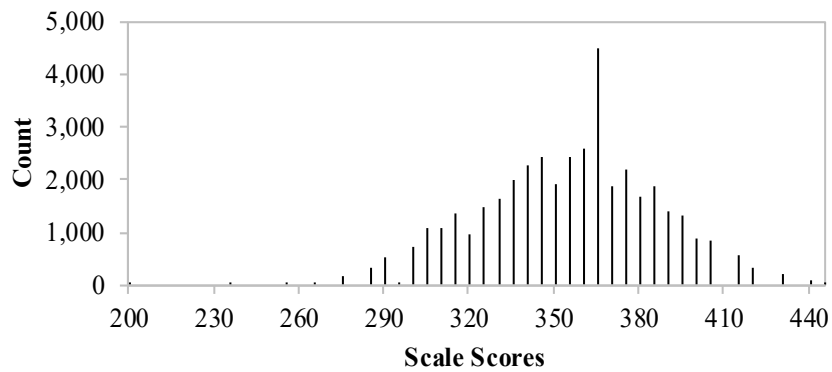


**Table 2.4.2.5.3**

Scale Score Descriptive Statistics: Read 6-8 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	14,348	200	448	351.15	27.77
7	13,501	235	448	356.25	29.81
8	13,127	200	448	362.74	32.07
Total	40,976	200	448	356.54	30.24

**Figure 2.4.2.5.3**  
**Scale Scores: Read 6-8 S503 Paper**



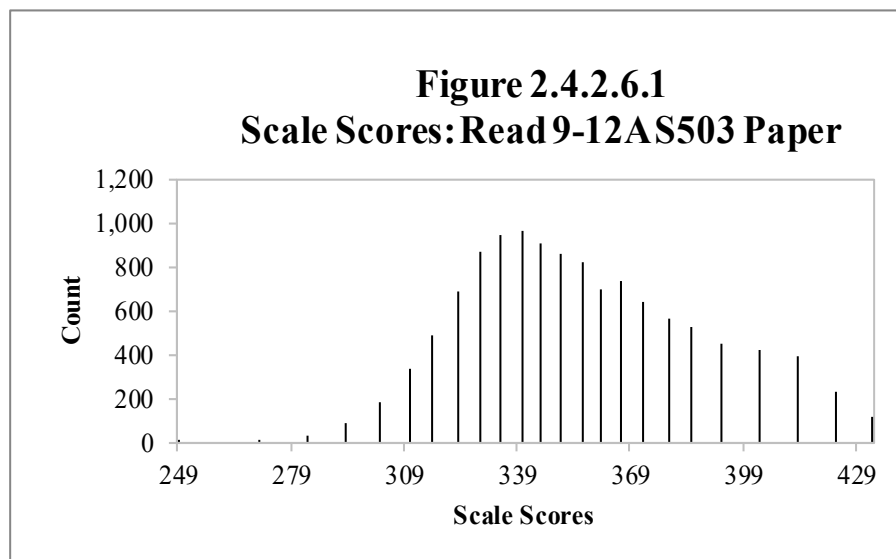


#### 2.4.2.6 Grades 9-12

**Table 2.4.2.6.1**

Scale Score Descriptive Statistics: Read 9-12 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	4,598	249	433	348.57	28.28
10	3,275	270	433	355.12	29.13
11	2,745	270	433	359.50	29.79
12	1,334	270	433	362.79	29.63
Total	11,952	249	433	354.46	29.48

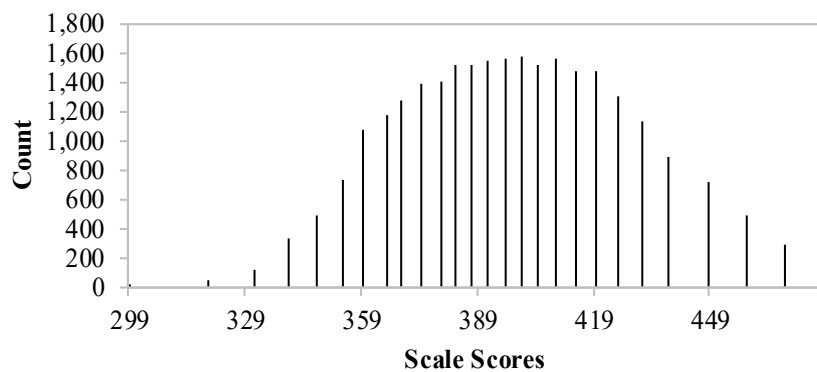


**Table 2.4.2.6.2**

Scale Score Descriptive Statistics: Read 9-12 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	7,352	299	478	391.98	26.89
10	7,394	299	478	396.23	27.60
11	6,811	299	478	400.55	28.75
12	5,141	319	478	398.73	28.36
Total	26,698	299	478	396.65	28.04

**Figure 2.4.2.6.2**  
**Scale Scores: Read 9-12B/C S503 Paper**

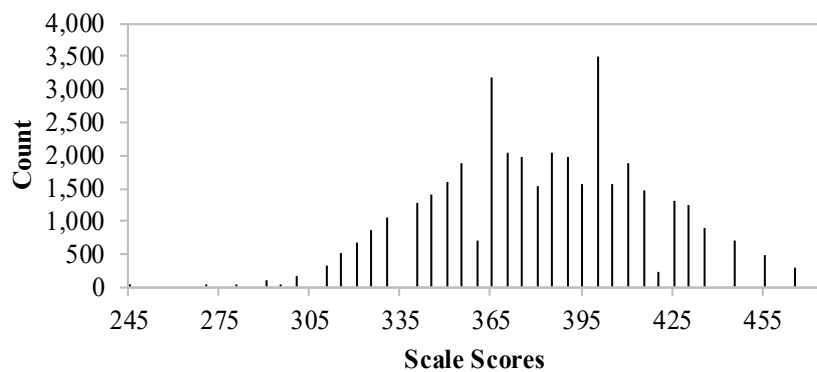


**Table 2.4.2.6.3**

Scale Score Descriptive Statistics: Read 9-12 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	11,950	249	478	375.28	34.62
10	10,669	270	478	383.61	33.88
11	9,556	270	478	388.76	34.48
12	6,475	270	478	391.33	32.10
Total	38,650	249	478	383.60	34.52

**Figure 2.4.2.6.3**  
**Scale Scores: Read 9-12 S503 Paper**



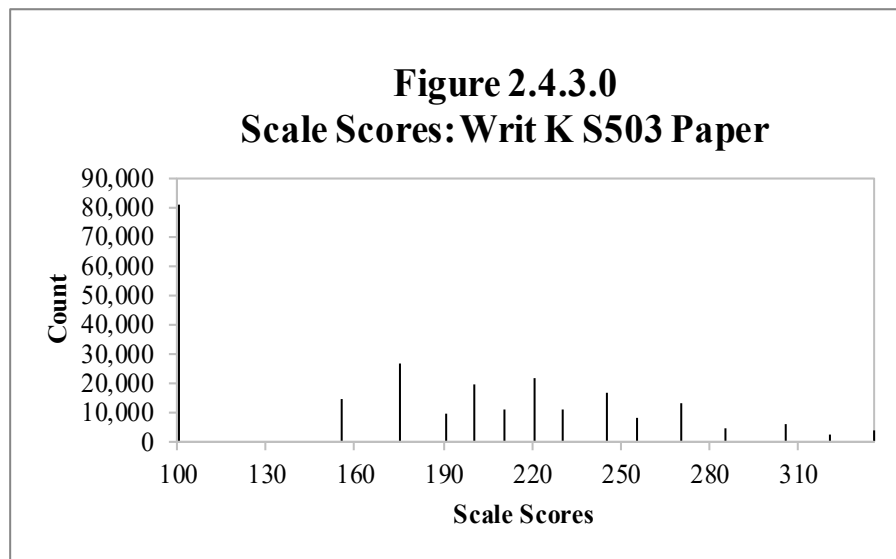
## 2.4.3 Writing

### 2.4.3.0 Kindergarten

**Table 2.4.3.0**

Scale Score Descriptive Statistics: Writ K S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	251,486	100	339	183.05	67.91
Total	251,486	100	339	183.05	67.91

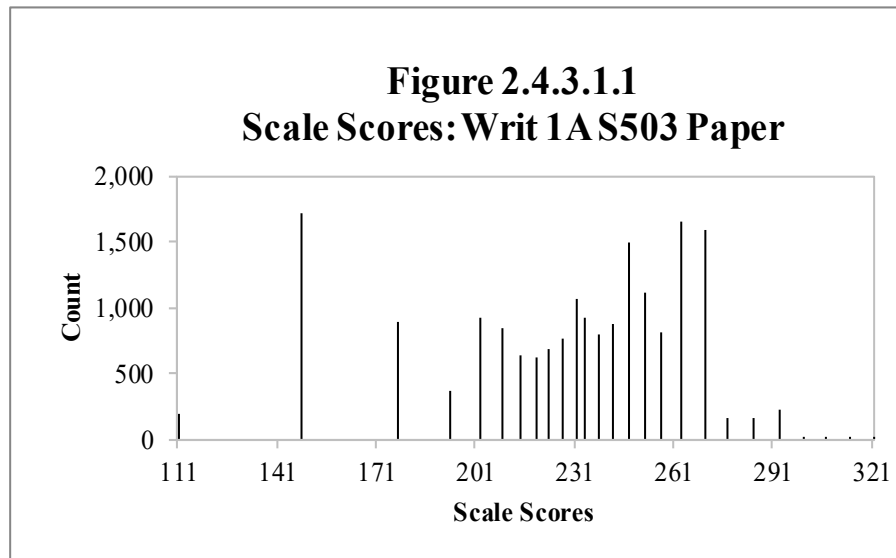


### 2.4.3.1 Grade 1

**Table 2.4.3.1.1**

Scale Score Descriptive Statistics: Writ 1 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	18,607	111	321	227.58	37.94
Total	18,607	111	321	227.58	37.94

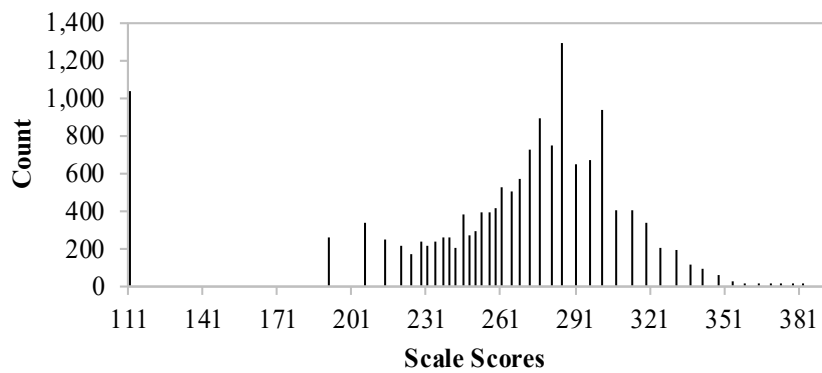


**Table 2.4.3.1.2**

Scale Score Descriptive Statistics: Writ 1 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,223	111	391	260.60	51.04
Total	15,223	111	391	260.60	51.04

**Figure 2.4.3.1.2**  
**Scale Scores: Writ 1B/C S503 Paper**

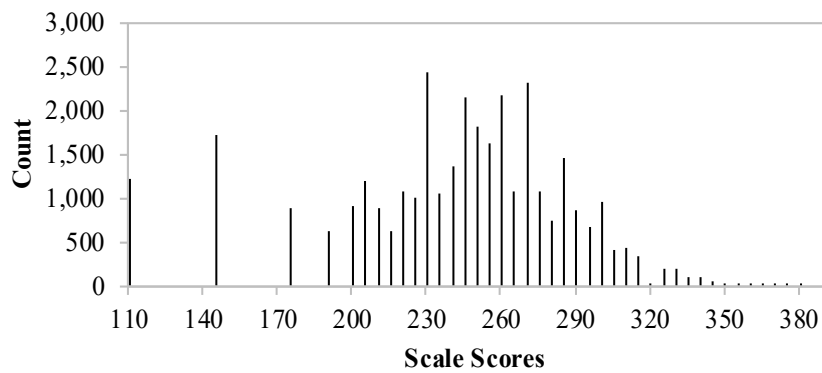


**Table 2.4.3.1.3**

Scale Score Descriptive Statistics: Writ 1 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	33,830	111	391	242.44	47.27
Total	33,830	111	391	242.44	47.27

**Figure 2.4.3.1.3**  
**Scale Scores: Writ 1 S503 Paper**

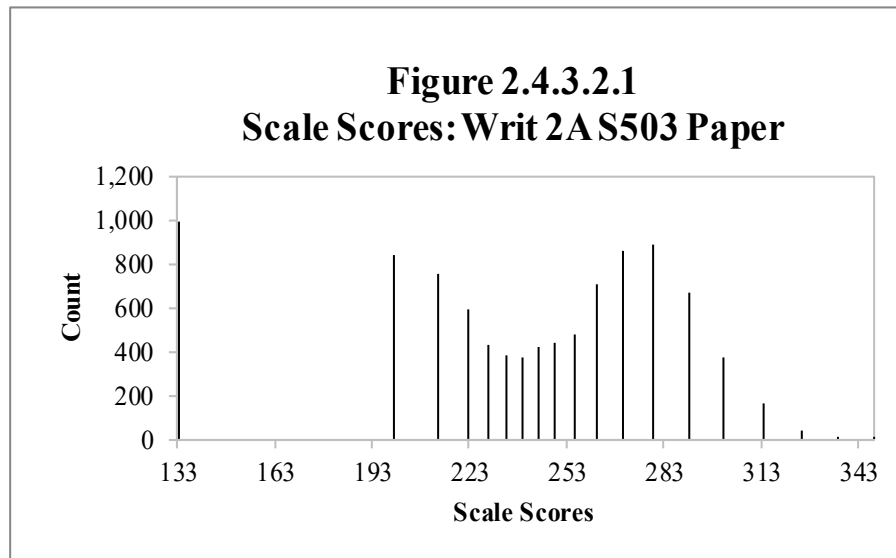


#### 2.4.3.2 Grade 2

**Table 2.4.3.2.1**

Scale Score Descriptive Statistics: Writ 2 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	9,430	133	347	238.16	46.66
Total	9,430	133	347	238.16	46.66



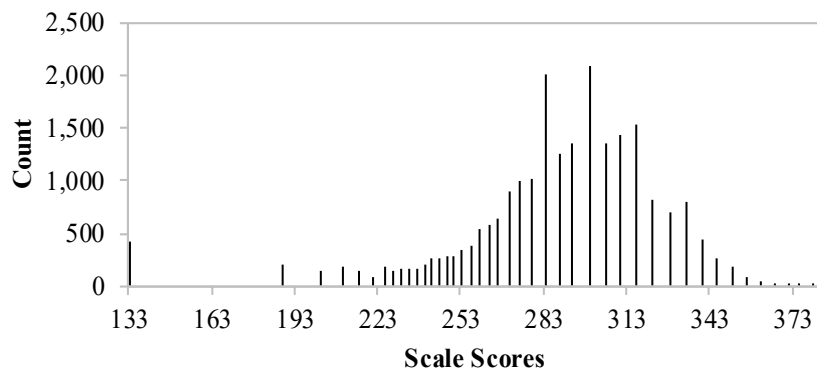


**Table 2.4.3.2.2**

Scale Score Descriptive Statistics: Writ 2 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	23,216	133	384	285.15	38.00
Total	23,216	133	384	285.15	38.00

**Figure 2.4.3.2.2**  
**Scale Scores: Writ B/C S503 Paper**

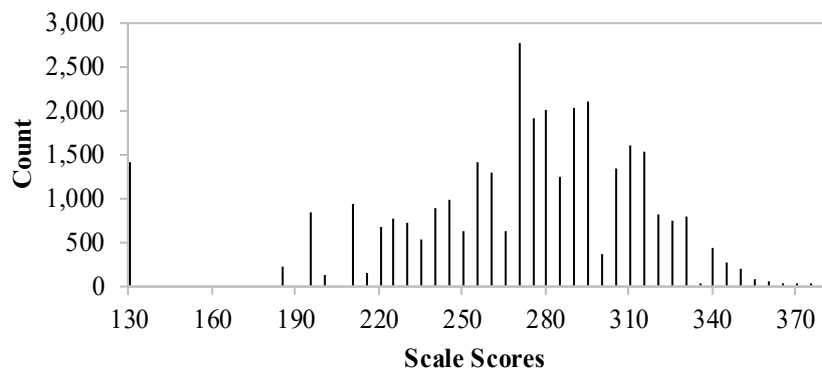


**Table 2.4.3.2.3**

Scale Score Descriptive Statistics: Writ 2 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	32,646	133	384	271.57	45.93
Total	32,646	133	384	271.57	45.93

**Figure 2.4.3.2.3**  
**Scale Scores: Writ 2 S503 Paper**

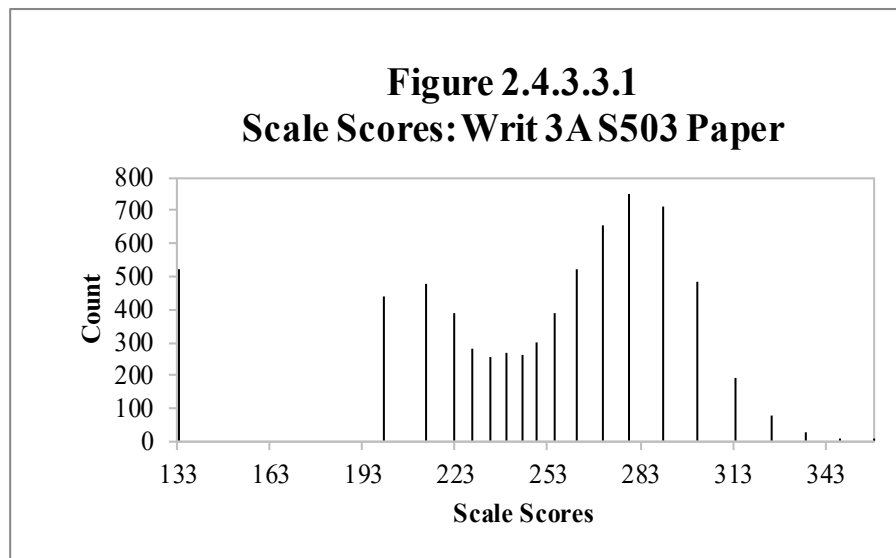


### 2.4.3.3 Grade 3

**Table 2.4.3.3.1**

Scale Score Descriptive Statistics: Writ 3 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	7,021	133	358	248.70	45.38
Total	7,021	133	358	248.70	45.38

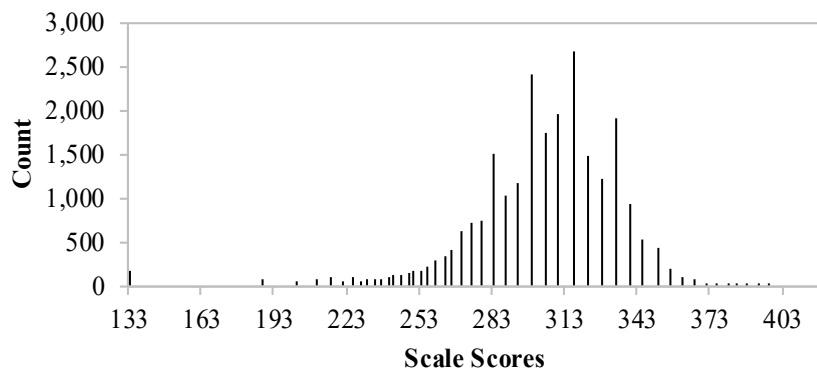


**Table 2.4.3.3.2**

Scale Score Descriptive Statistics: Writ 3 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	24,568	133	420	301.17	32.69
Total	24,568	133	420	301.17	32.69

**Figure 2.4.3.3.2**  
**Scale Scores: Writ 3B/C S503 Paper**

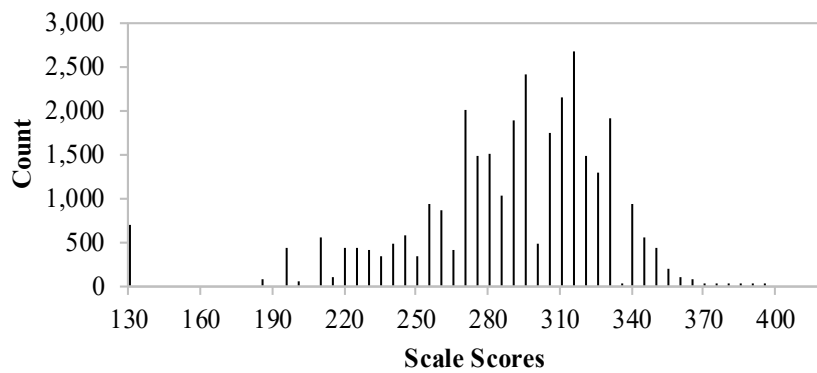


**Table 2.4.3.3.3**

Scale Score Descriptive Statistics: Writ 3 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	31,589	133	420	289.51	42.01
Total	31,589	133	420	289.51	42.01

**Figure 2.4.3.3.3**  
**Scale Scores: Writ 3 S503 Paper**

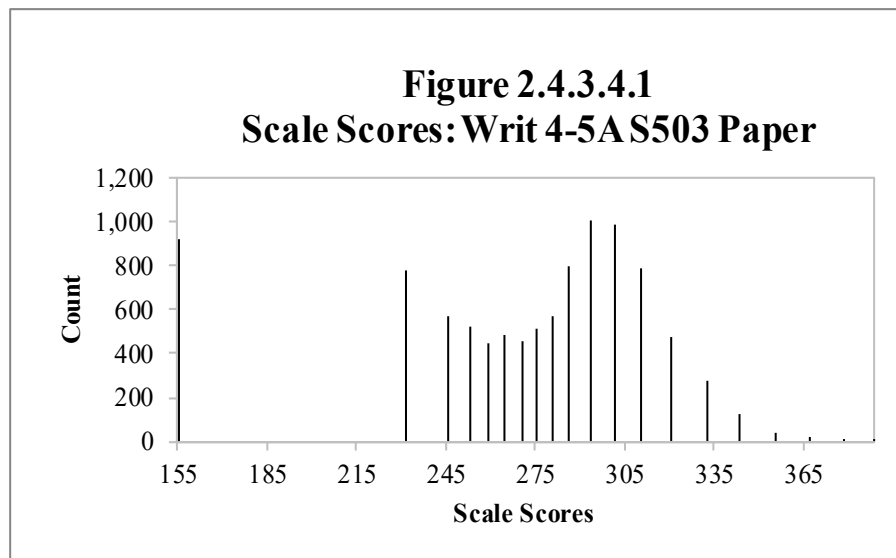


#### 2.4.3.4 Grades 4-5

**Table 2.4.3.4.1**

Scale Score Descriptive Statistics: Writ 4-5 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	5,008	155	388	266.04	46.12
5	4,785	155	388	272.83	45.00
Total	9,793	155	388	269.36	45.70

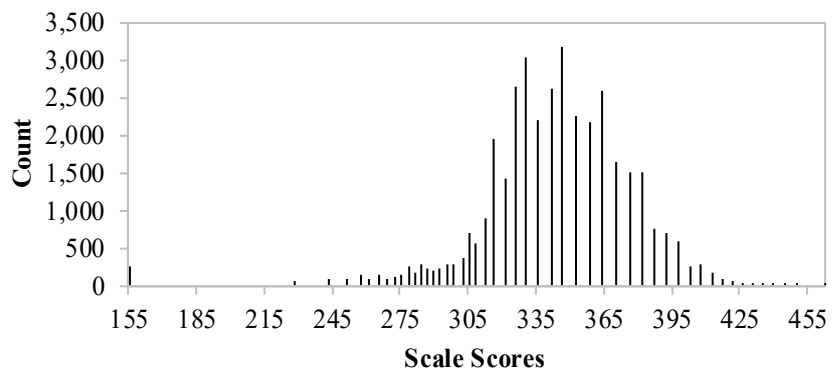


**Table 2.4.3.4.2**

Scale Score Descriptive Statistics: Writ 4-5 B/C S503 Paper

<b>Grade</b>	<b>No. of Students</b>	<b>Min.</b>	<b>Max.</b>	<b>Mean</b>	<b>Std. Dev.</b>
<b>4</b>	18,717	155	450	334.16	33.41
<b>5</b>	18,763	155	462	347.20	34.23
<b>Total</b>	37,480	155	462	340.69	34.45

**Figure 2.4.3.4.2**  
**Scale Scores: Writ 4-5B/C S503 Paper**

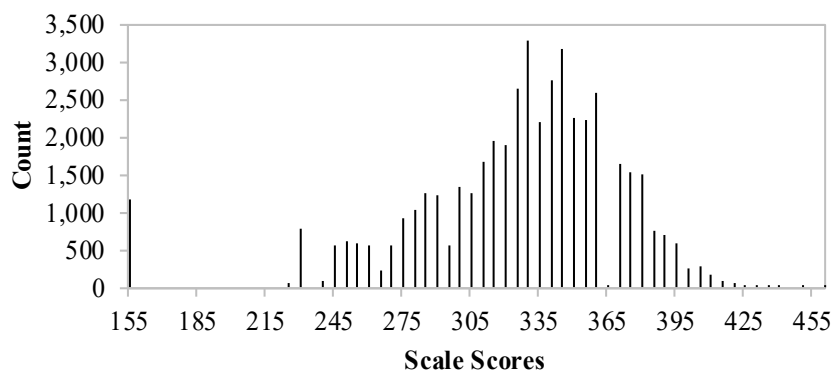


**Table 2.4.3.4.3**

Scale Score Descriptive Statistics: Writ 4-5 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	23,725	155	450	319.78	45.85
5	23,548	155	462	332.09	47.34
Total	47,273	155	462	325.91	47.00

**Figure 2.4.3.4.3**  
**Scale Scores: Writ 4-5 S503 Paper**



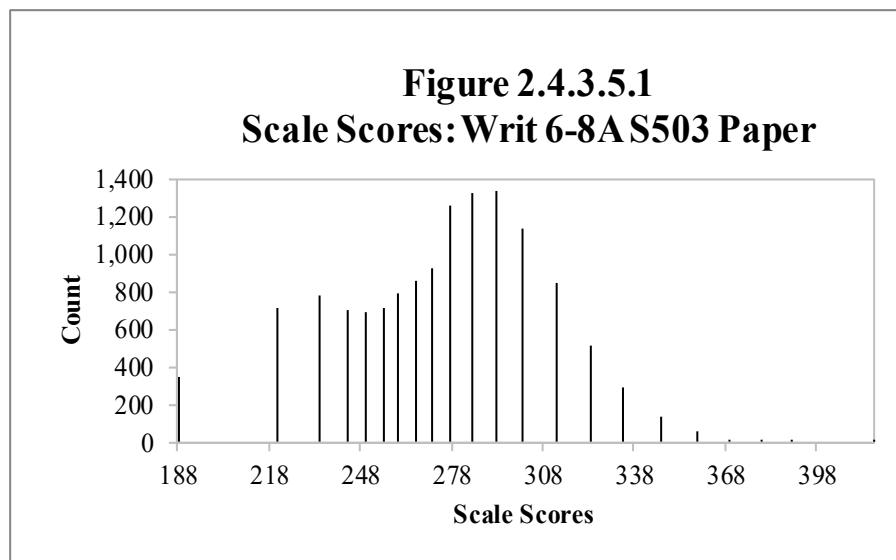


#### 2.4.3.5 Grades 6-8

**Table 2.4.3.5.1**

Scale Score Descriptive Statistics: Writ 6-8 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	4,564	188	379	269.35	32.29
7	4,439	188	379	272.89	31.50
8	4,425	188	416	277.70	31.72
Total	13,428	188	416	273.27	32.02

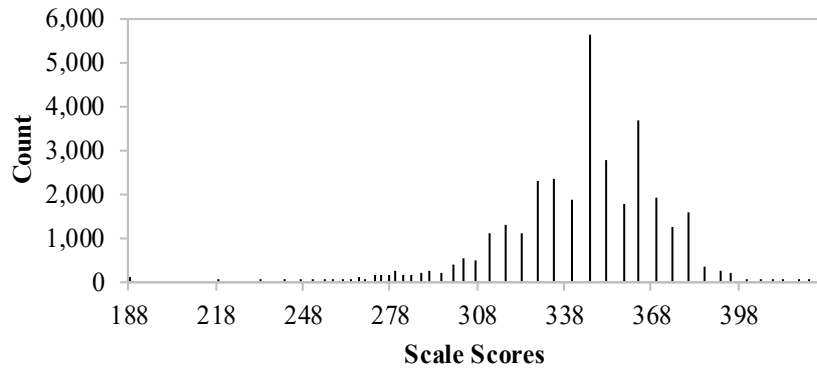


**Table 2.4.3.5.2**

Scale Score Descriptive Statistics: Writ 6-8 B/C S503 Paper

<b>Grade</b>	<b>No. of Students</b>	<b>Min.</b>	<b>Max.</b>	<b>Mean</b>	<b>Std. Dev.</b>
<b>6</b>	12,134	188	413	335.80	29.00
<b>7</b>	11,035	188	422	341.50	29.56
<b>8</b>	10,450	188	427	347.62	28.98
<b>Total</b>	33,619	188	427	341.35	29.57

**Figure 2.4.3.5.2**  
**Scale Scores: Writ 6-8B/C S503 Paper**

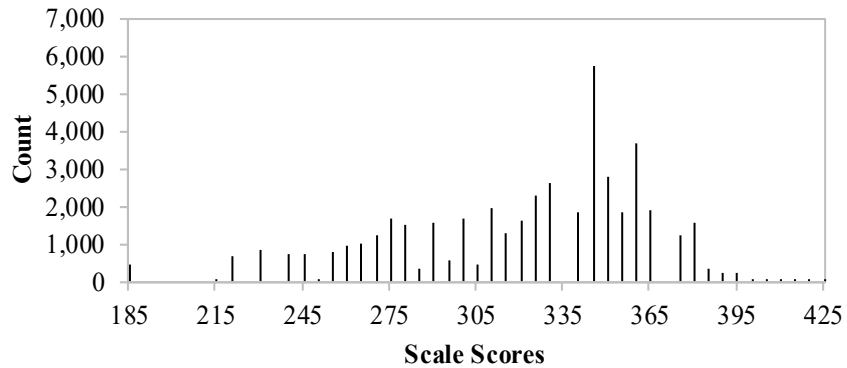


**Table 2.4.3.5.3**

Scale Score Descriptive Statistics: Writ 6-8 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	16,698	188	413	317.64	42.11
7	15,474	188	422	321.82	43.25
8	14,875	188	427	326.82	43.72
Total	47,047	188	427	321.92	43.16

**Figure 2.4.3.5.3**  
**Scale Scores: Writ 6-8 S503 Paper**

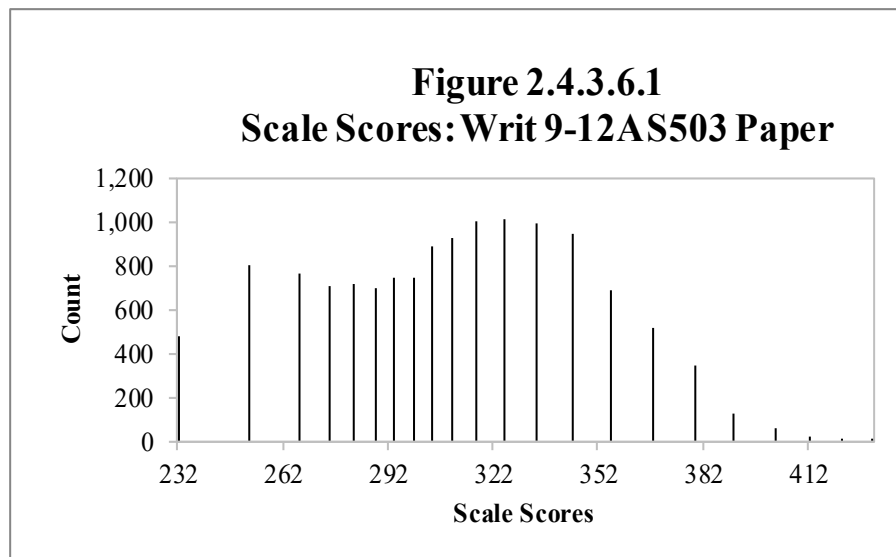


#### 2.4.3.6 Grades 9-12

**Table 2.4.3.6.1**

Scale Score Descriptive Statistics: Writ 9-12 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	5,117	232	430	301.02	37.07
10	3,587	232	430	309.02	35.24
11	3,007	232	430	313.55	35.62
12	1,450	232	412	320.91	33.03
<b>Total</b>	<b>13,161</b>	<b>232</b>	<b>430</b>	<b>308.25</b>	<b>36.43</b>

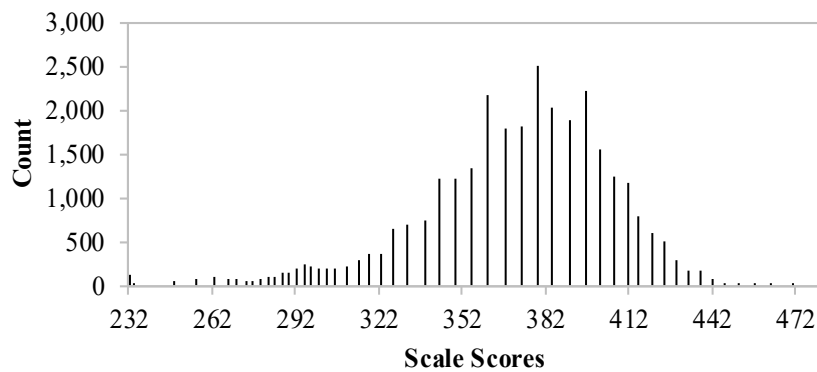


**Table 2.4.3.6.2**

Scale Score Descriptive Statistics: Writ 9-12 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	8,648	232	471	366.09	36.44
10	8,641	232	482	369.19	37.57
11	7,818	232	482	372.41	38.20
12	5,756	232	471	369.69	39.54
Total	30,863	232	482	369.23	37.87

**Figure 2.4.3.6.2**  
**Scale Scores: Writ 9-12B/C S503 Paper**

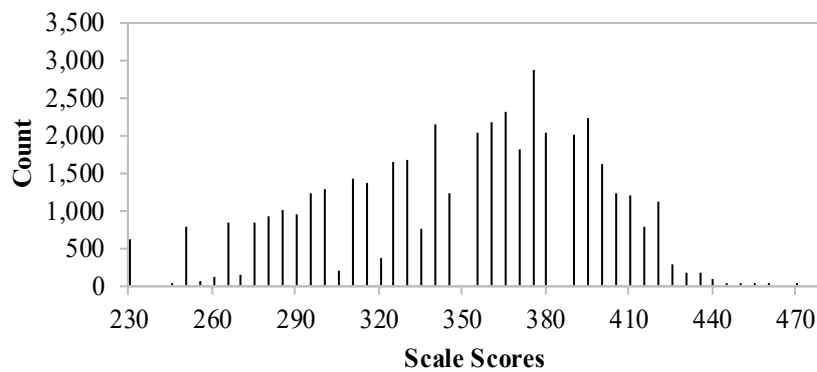


**Table 2.4.3.6.3**

Scale Score Descriptive Statistics: Writ 9-12 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	13,765	232	471	341.90	48.31
10	12,228	232	482	351.54	45.96
11	10,825	232	482	356.06	45.84
12	7,206	232	471	359.87	43.02
Total	44,024	232	482	351.00	46.70

**Figure 2.4.3.6.3**  
**Scale Scores: Writ 9-12 S503 Paper**



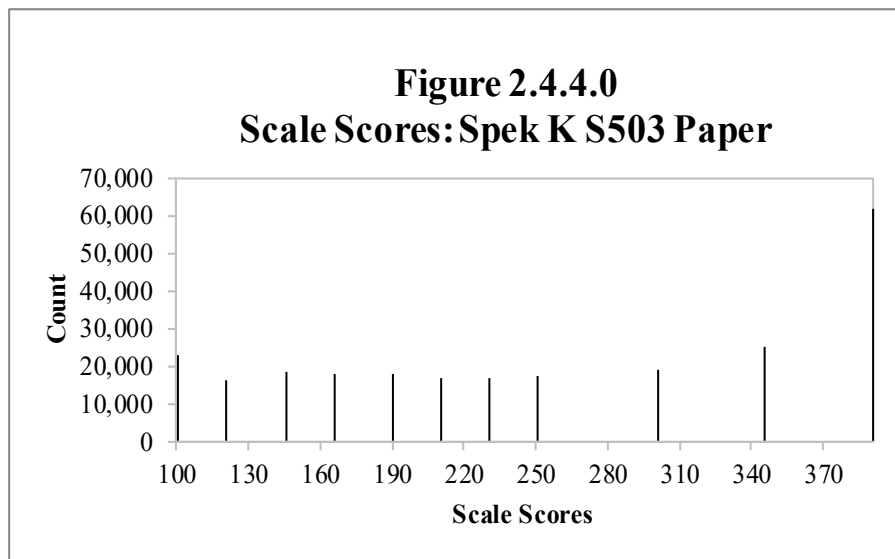
## 2.4.4 Speaking

### 2.4.4.0 Kindergarten

**Table 2.4.4.0**

Scale Score Descriptive Statistics: Spek K S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	251,483	100	392	255.29	104.28
Total	251,483	100	392	255.29	104.28

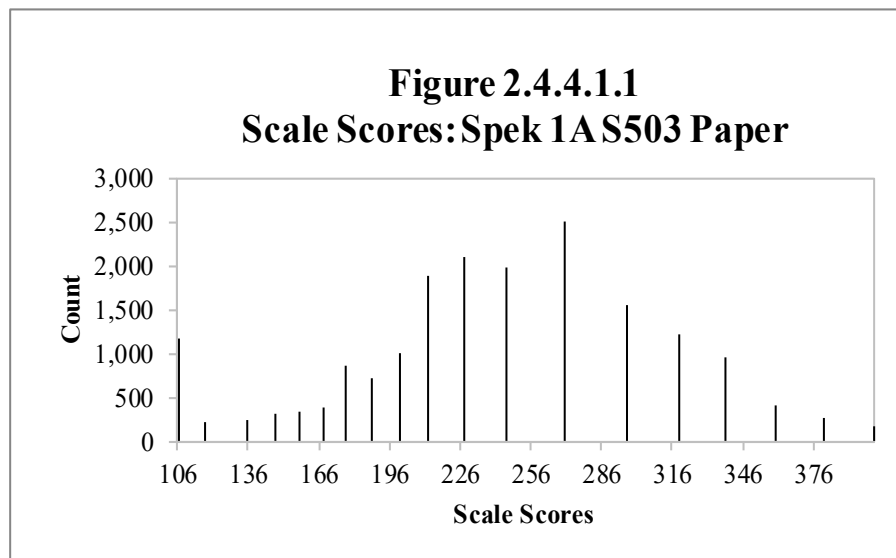


#### 2.4.4.1 Grade 1

**Table 2.4.4.1.1**

Scale Score Descriptive Statistics: Spek 1 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	18,370	106	401	238.14	67.13
Total	18,370	106	401	238.14	67.13



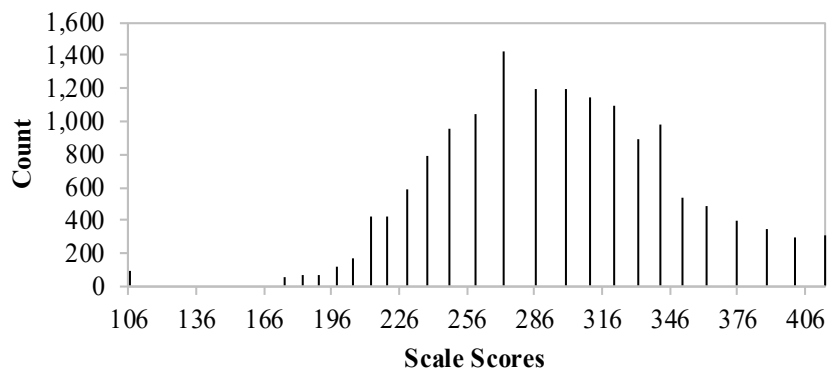


**Table 2.4.4.1.2**

Scale Score Descriptive Statistics: Spek 1 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,080	106	414	293.70	54.01
Total	15,080	106	414	293.70	54.01

**Figure 2.4.4.1.2**  
**Scale Scores: Spek 1B/C S503 Paper**

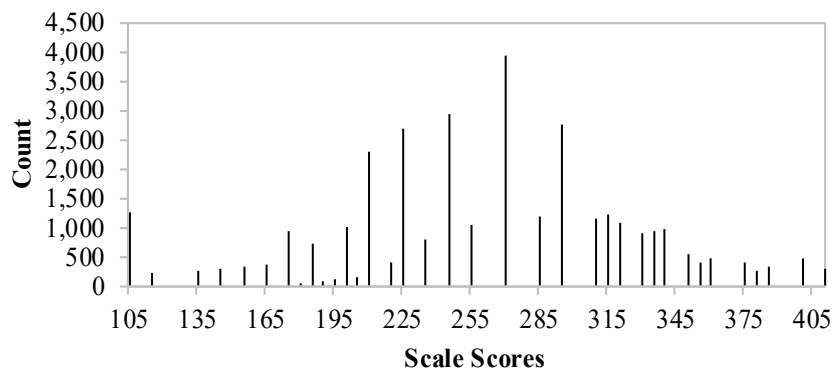


**Table 2.4.4.1.3**

Scale Score Descriptive Statistics: Spek 1 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	33,450	106	414	263.19	67.49
Total	33,450	106	414	263.19	67.49

**Figure 2.4.4.1.3**  
**Scale Scores: Spek 1 S503 Paper**

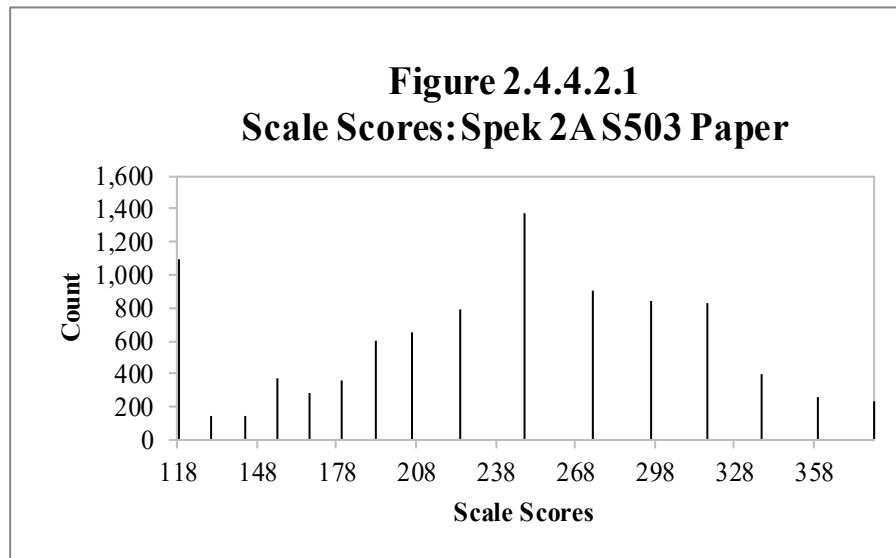


#### 2.4.4.2 Grade 2

**Table 2.4.4.2.1**

Scale Score Descriptive Statistics: Spek 2 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	9,275	118	380	234.87	71.79
Total	9,275	118	380	234.87	71.79

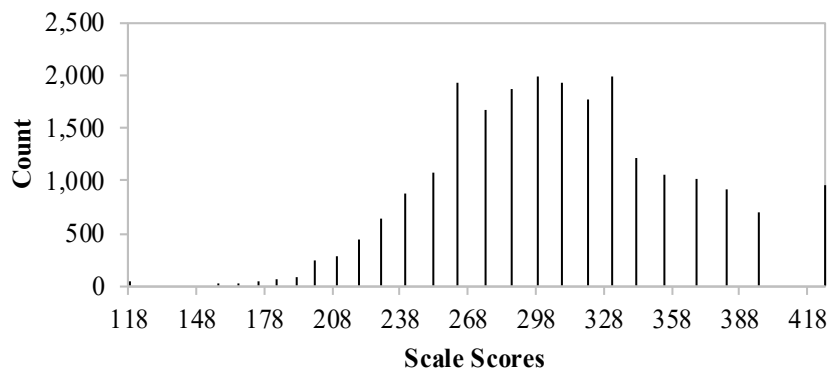


**Table 2.4.4.2.2**

Scale Score Descriptive Statistics: Spek 2 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	22,969	118	425	305.73	53.74
Total	22,969	118	425	305.73	53.74

**Figure 2.4.4.2.2**  
**Scale Scores: Spek 2B/C S503 Paper**

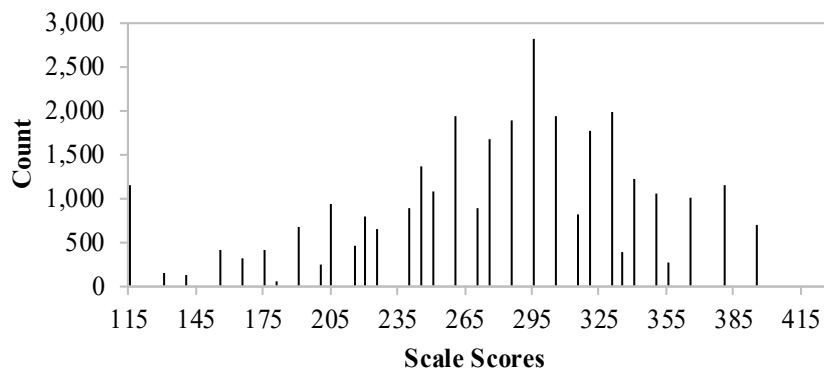


**Table 2.4.4.2.3**

Scale Score Descriptive Statistics: Spek 2 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	32,244	118	425	285.35	67.59
Total	32,244	118	425	285.35	67.59

**Figure 2.4.4.2.3**  
**Scale Scores: Spek 2 S503 Paper**

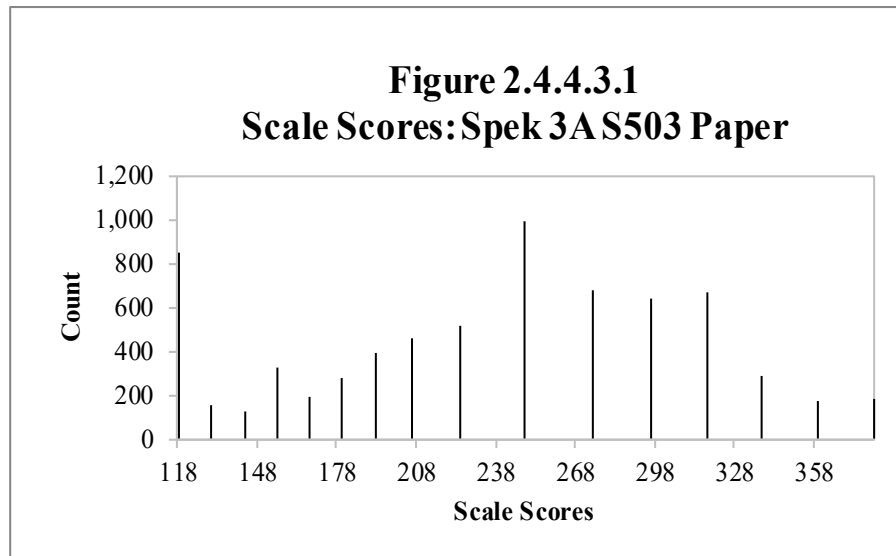


#### 2.4.4.3 Grade 3

**Table 2.4.4.3.1**

Scale Score Descriptive Statistics: Spek 3 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	6,925	118	380	233.81	73.20
Total	6,925	118	380	233.81	73.20

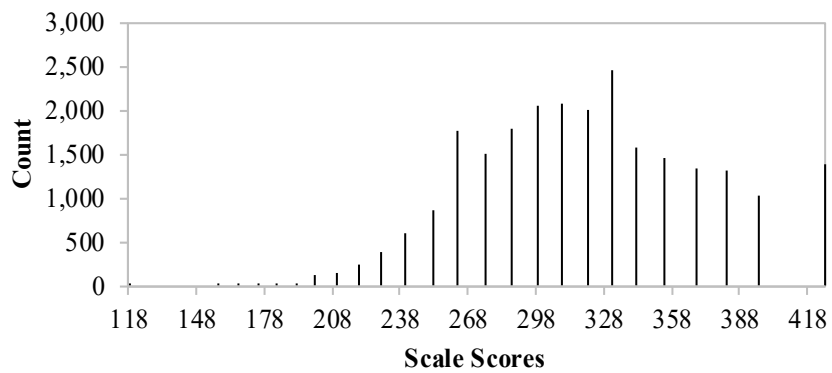


**Table 2.4.4.3.2**

Scale Score Descriptive Statistics: Spek 3 B/C S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	24,308	118	425	318.41	51.77
Total	24,308	118	425	318.41	51.77

**Figure 2.4.4.3.2**  
**Scale Scores: Spek 3B/C S503 Paper**

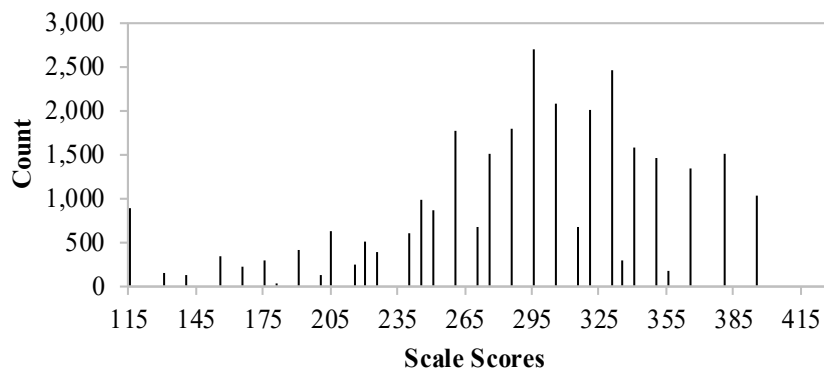


**Table 2.4.4.3.3**

Scale Score Descriptive Statistics: Spek 3 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	31,233	118	425	299.65	67.15
Total	31,233	118	425	299.65	67.15

**Figure 2.4.4.3.3**  
**Scale Scores: Spek 3 S503 Paper**



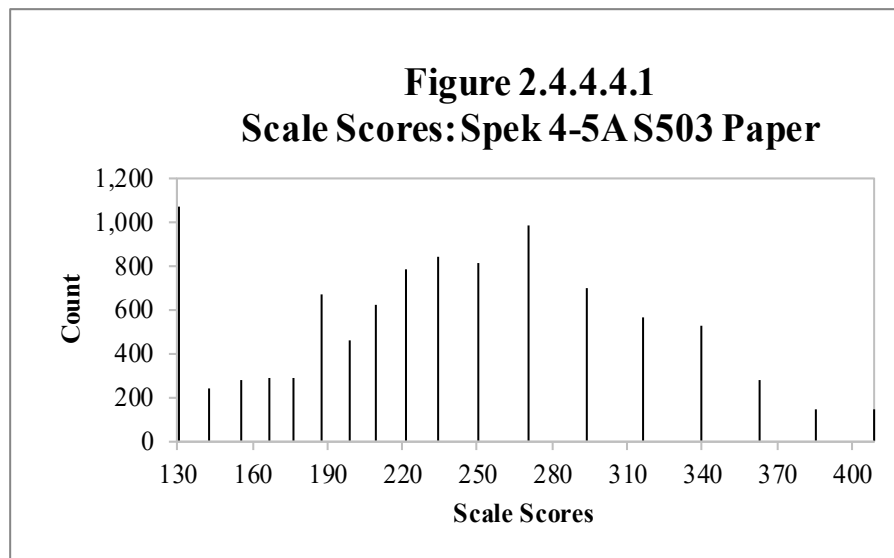


#### 2.4.4.4 Grades 4-5

**Table 2.4.4.4.1**

Scale Score Descriptive Statistics: Spek 4-5 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	4,965	130	408	234.63	69.01
5	4,738	130	408	235.55	69.34
Total	9,703	130	408	235.08	69.17

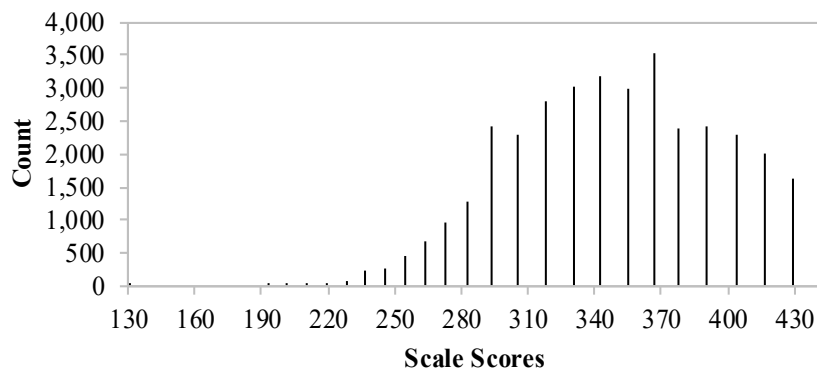


**Table 2.4.4.4.2**

Scale Score Descriptive Statistics: Spek 4-5 B/C S503 Paper

<b>Grade</b>	<b>No. of Students</b>	<b>Min.</b>	<b>Max.</b>	<b>Mean</b>	<b>Std. Dev.</b>
<b>4</b>	18,564	130	443	346.69	51.69
<b>5</b>	18,643	130	443	354.50	52.04
<b>Total</b>	37,207	130	443	350.60	52.01

**Figure 2.4.4.4.2**  
**Scale Scores: Spek 4-5B/C S503 Paper**

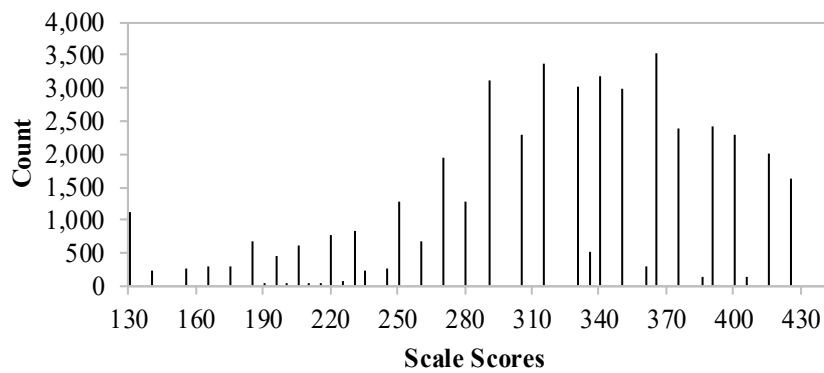


**Table 2.4.4.4.3**

Scale Score Descriptive Statistics: Spek 4-5 S503 Paper

<b>Grade</b>	<b>No. of Students</b>	<b>Min.</b>	<b>Max.</b>	<b>Mean</b>	<b>Std. Dev.</b>
<b>4</b>	23,529	130	443	323.05	72.14
<b>5</b>	23,381	130	443	330.39	73.62
<b>Total</b>	46,910	130	443	326.71	72.97

**Figure 2.4.4.4.3**  
**Scale Scores: Spek 4-5 S503 Paper**

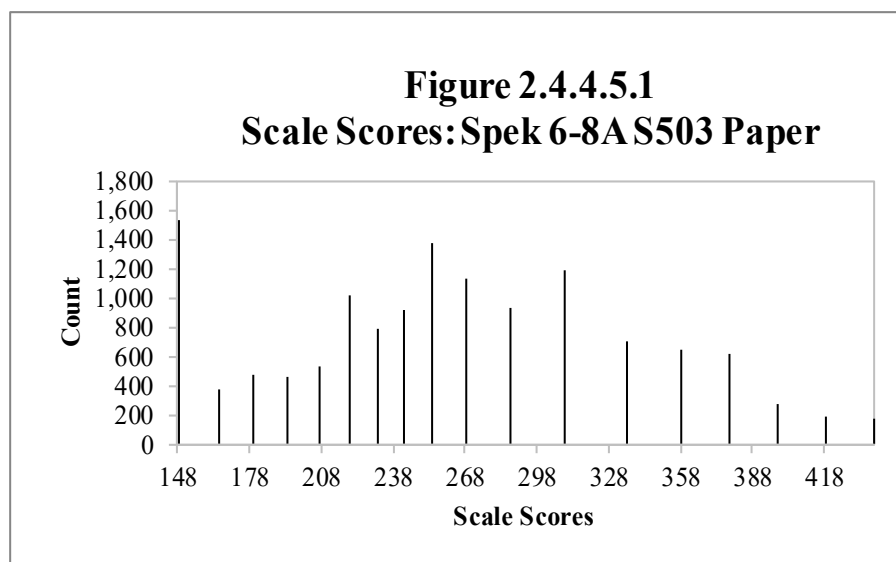


#### 2.4.4.5 Grades 6-8

**Table 2.4.4.5.1**

Scale Score Descriptive Statistics: Spek 6-8 A S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	4,530	148	438	257.10	72.38
7	4,404	148	438	256.48	72.01
8	4,394	148	438	263.36	72.78
Total	13,328	148	438	258.96	72.45

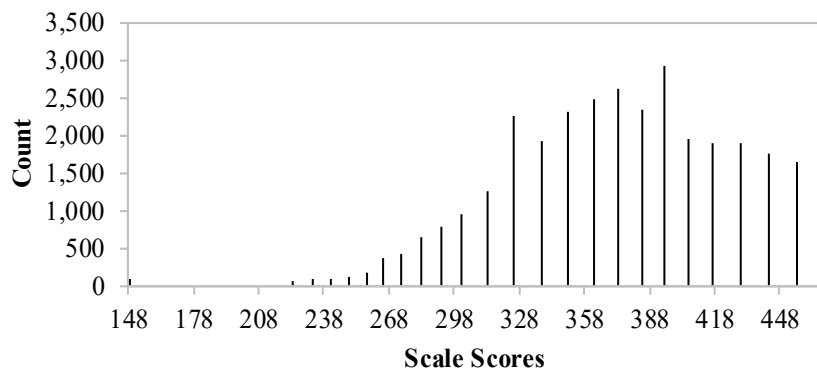


**Table 2.4.4.5.2**

Scale Score Descriptive Statistics: Spek 6-8 B/C S503 Paper

<b>Grade</b>	<b>No. of Students</b>	<b>Min.</b>	<b>Max.</b>	<b>Mean</b>	<b>Std. Dev.</b>
<b>6</b>	12,044	148	468	372.78	55.50
<b>7</b>	10,967	148	468	375.20	56.88
<b>8</b>	10,382	148	468	382.27	56.03
<b>Total</b>	33,393	148	468	376.52	56.26

**Figure 2.4.4.5.2**  
**Scale Scores: Spek 6-8B/C S503 Paper**

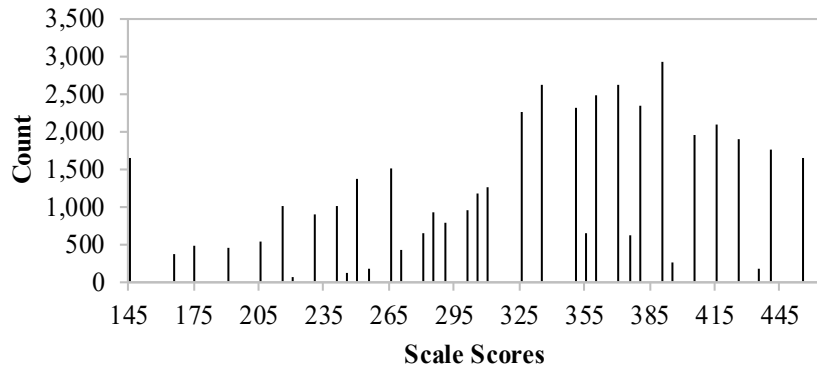


**Table 2.4.4.5.3**

Scale Score Descriptive Statistics: Spek 6-8 S503 Paper

<b>Grade</b>	<b>No. of Students</b>	<b>Min.</b>	<b>Max.</b>	<b>Mean</b>	<b>Std. Dev.</b>
<b>6</b>	16,574	148	468	341.16	79.55
<b>7</b>	15,371	148	468	341.18	81.71
<b>8</b>	14,776	148	468	346.91	82.07
<b>Total</b>	46,721	148	468	342.99	81.10

**Figure 2.4.4.5.3**  
**Scale Scores: Spek 6-8 S503 Paper**

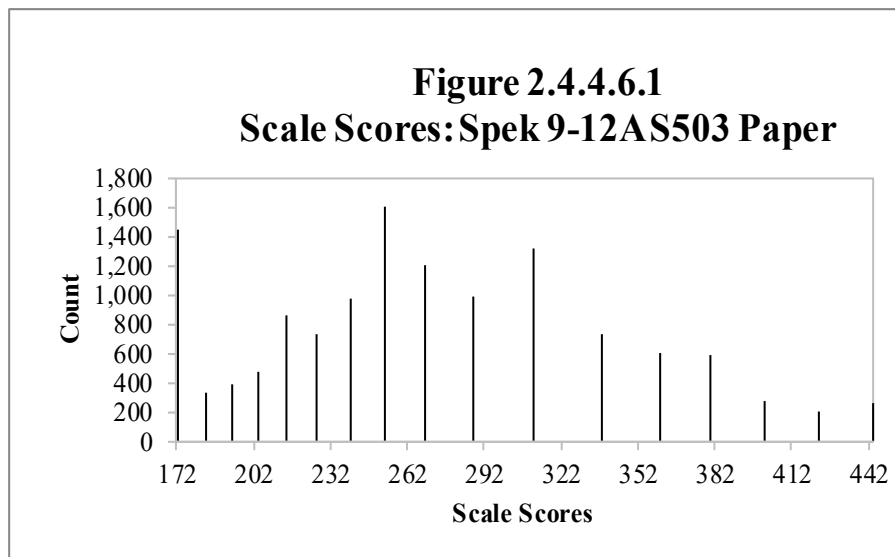


#### 2.4.4.6 Grades 9-12

**Table 2.4.4.6.1**

Scale Score Descriptive Statistics: Spek 9-12 A S503 Paper

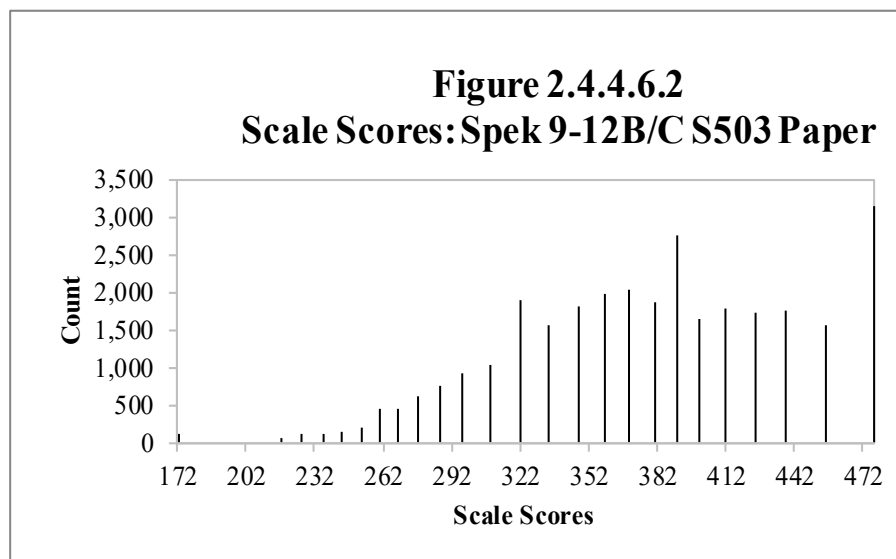
Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	5,069	172	443	256.46	67.38
10	3,530	172	443	269.09	67.42
11	2,971	172	443	276.71	68.58
12	1,433	172	443	291.09	68.85
<b>Total</b>	<b>13,003</b>	<b>172</b>	<b>443</b>	<b>268.33</b>	<b>68.76</b>



**Table 2.4.4.6.2**

Scale Score Descriptive Statistics: Spek 9-12 B/C S503 Paper

<b>Grade</b>	<b>No. of Students</b>	<b>Min.</b>	<b>Max.</b>	<b>Mean</b>	<b>Std. Dev.</b>
<b>9</b>	8,561	172	476	373.96	62.19
<b>10</b>	8,566	172	476	375.12	62.49
<b>11</b>	7,746	172	476	377.71	62.10
<b>12</b>	5,688	172	476	380.73	61.25
<b>Total</b>	30,561	172	476	376.50	62.12



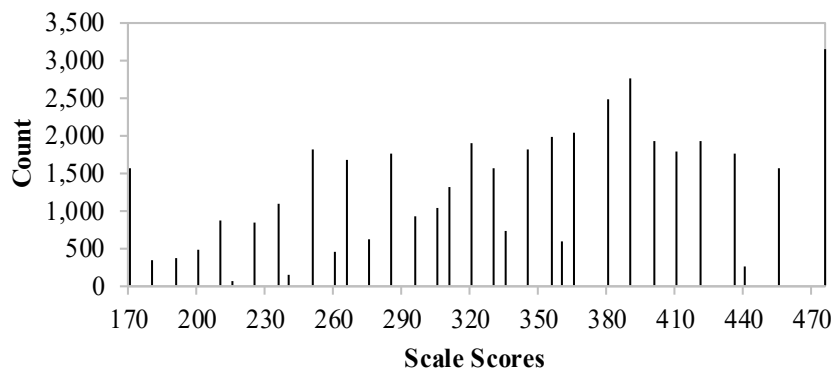


**Table 2.4.4.6.3**

Scale Score Descriptive Statistics: Spek 9-12 S503 Paper

<b>Grade</b>	<b>No. of Students</b>	<b>Min.</b>	<b>Max.</b>	<b>Mean</b>	<b>Std. Dev.</b>
<b>9</b>	13,630	172	476	330.26	85.68
<b>10</b>	12,096	172	476	344.18	80.10
<b>11</b>	10,717	172	476	349.71	78.33
<b>12</b>	7,121	172	476	362.69	72.40
<b>Total</b>	43,564	172	476	344.21	81.05

**Figure 2.4.4.6.3**  
**Scale Scores: Spek 9-12 S503 Paper**



## ***2.5 Proficiency Level Distribution***

Figures and tables in this section provide information on the proficiency level distribution for each of the domains for each grade-level cluster. In each figure, the horizontal axis shows the six WIDA proficiency levels. The vertical axis shows the percentage of students. Each bar shows the percentage of students who were placed into each proficiency level in the domain being tested on this test form.

The tables in this section present, by grade and by total for the grade-level cluster:

- The WIDA proficiency level designation (1–6)
- The number of students (count) whose performance on the test form placed them into that proficiency level in the domain being tested
- The percentage of students, out of the total number of students taking the form, who were placed into that proficiency level in the domain being tested

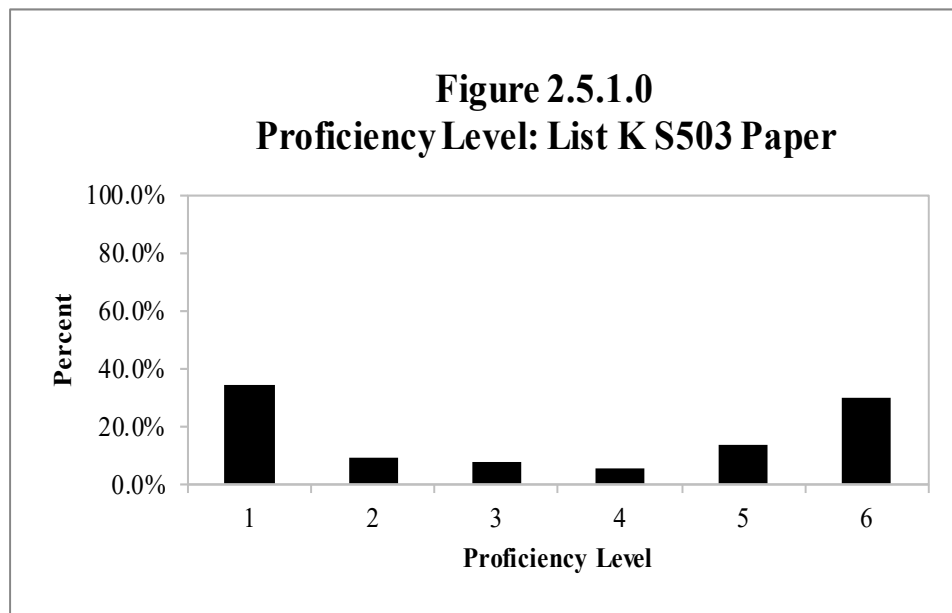
## 2.5.1 Listening

### 2.5.1.0 Kindergarten

**Table 2.5.1.0**

Proficiency Level Distribution: List K S503 Paper

Level	Grade K		Total	
	Count	Percent	Count	Percent
1	85,719	34.08%	85,719	34.08%
2	23,569	9.37%	23,569	9.37%
3	19,950	7.93%	19,950	7.93%
4	13,073	5.20%	13,073	5.20%
5	33,592	13.36%	33,592	13.36%
6	75,590	30.06%	75,590	30.06%
Total	251,493	100.00%	251,493	100.00%

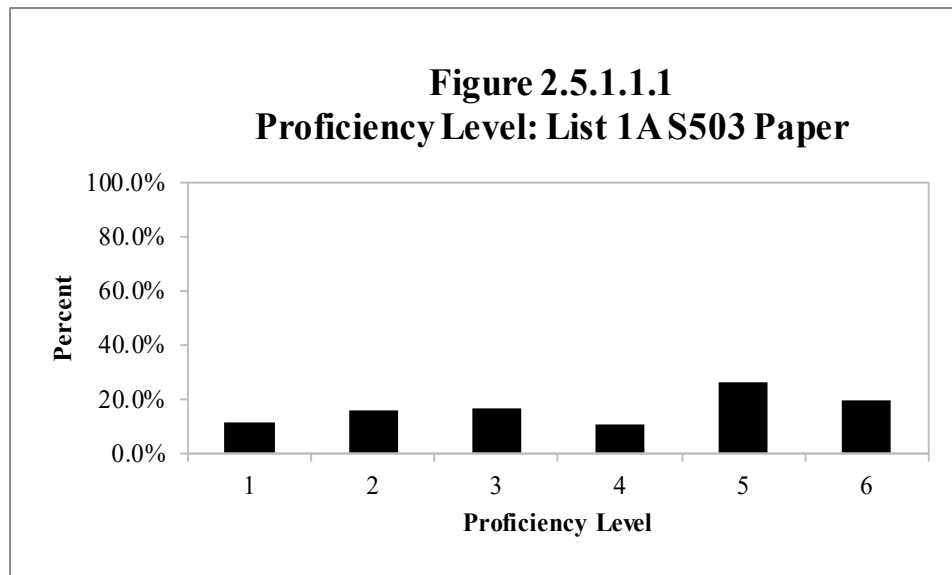


### 2.5.1.1 Grade 1

**Table 2.5.1.1.1**

Proficiency Level Distribution: List 1 A S503 Paper

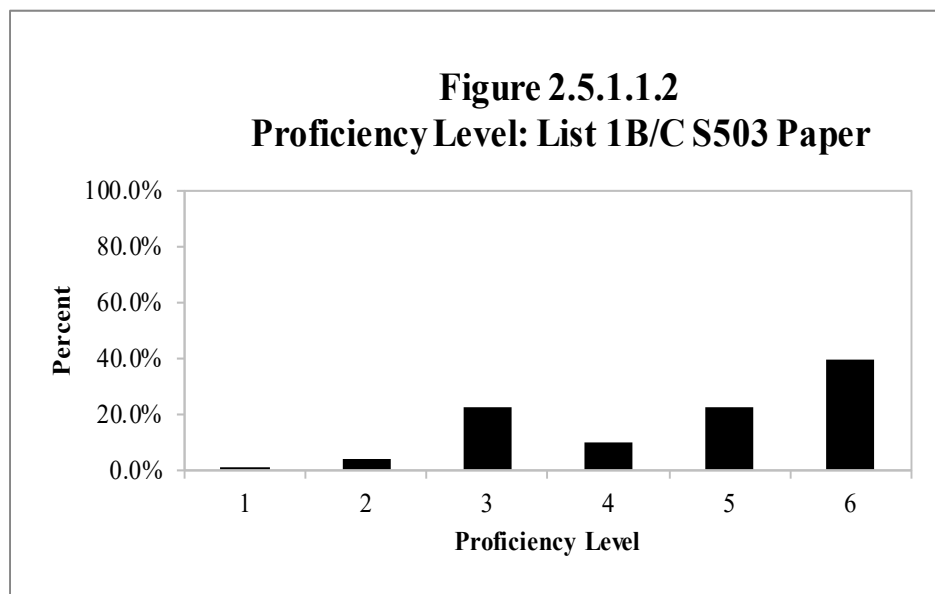
Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	1,721	11.29%	1,721	11.29%
2	2,380	15.62%	2,380	15.62%
3	2,535	16.64%	2,535	16.64%
4	1,674	10.99%	1,674	10.99%
5	3,997	26.23%	3,997	26.23%
6	2,931	19.23%	2,931	19.23%
Total	15,238	100.00%	15,238	100.00%



**Table 2.5.1.1.2**

Proficiency Level Distribution: List 1 B/C S503 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	169	1.28%	169	1.28%
2	526	3.99%	526	3.99%
3	3,007	22.83%	3,007	22.83%
4	1,280	9.72%	1,280	9.72%
5	3,017	22.90%	3,017	22.90%
6	5,175	39.28%	5,175	39.28%
Total	13,174	100.00%	13,174	100.00%

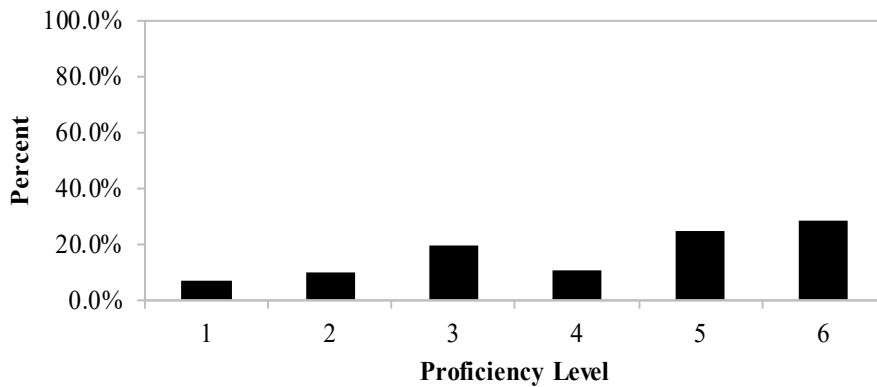


**Table 2.5.1.1.3**

Proficiency Level Distribution: List 1 S503 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	1,890	6.65%	1,890	6.65%
2	2,906	10.23%	2,906	10.23%
3	5,542	19.51%	5,542	19.51%
4	2,954	10.40%	2,954	10.40%
5	7,014	24.69%	7,014	24.69%
6	8,106	28.53%	8,106	28.53%
Total	28,412	100.00%	28,412	100.00%

**Figure 2.5.1.1.3**  
**Proficiency Level: List 1 S503 Paper**

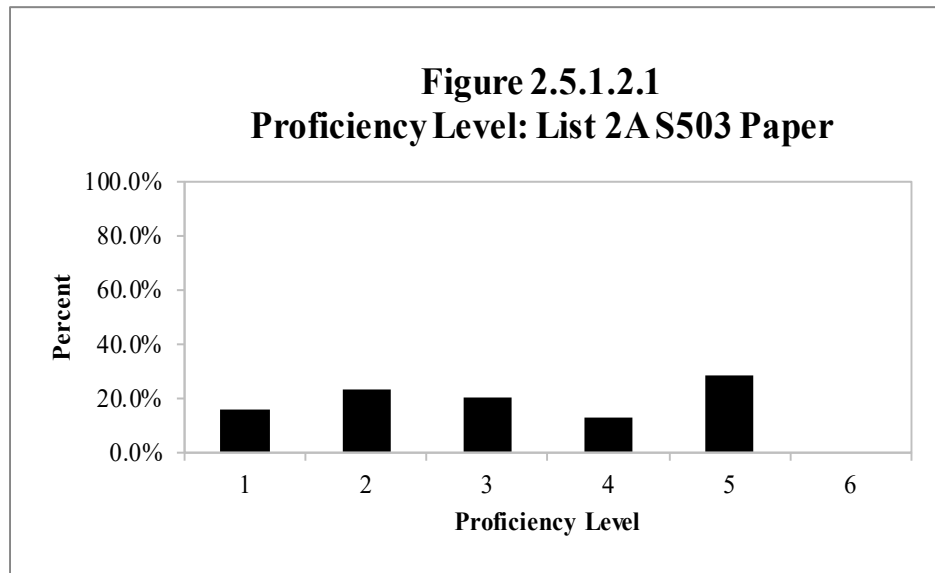


### 2.5.1.2 Grade 2

**Table 2.5.1.2.1**

Proficiency Level Distribution: List 2 A S503 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	1,304	15.54%	1,304	15.54%
2	1,928	22.98%	1,928	22.98%
3	1,698	20.24%	1,698	20.24%
4	1,083	12.91%	1,083	12.91%
5	2,376	28.32%	2,376	28.32%
6	0	0.00%	0	0.00%
Total	8,389	100.00%	8,389	100.00%

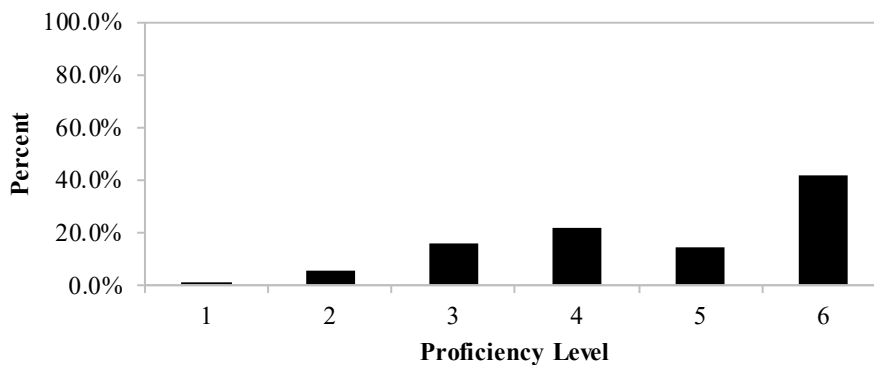


**Table 2.5.1.2.2**

Proficiency Level Distribution: List 2 B/C S503 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	72	0.33%	72	0.33%
2	1,239	5.73%	1,239	5.73%
3	3,381	15.63%	3,381	15.63%
4	4,754	21.97%	4,754	21.97%
5	3,149	14.55%	3,149	14.55%
6	9,043	41.79%	9,043	41.79%
<b>Total</b>	<b>21,638</b>	<b>100.00%</b>	<b>21,638</b>	<b>100.00%</b>

**Figure 2.5.1.2.2**  
**Proficiency Level: List 2B/C S503 Paper**

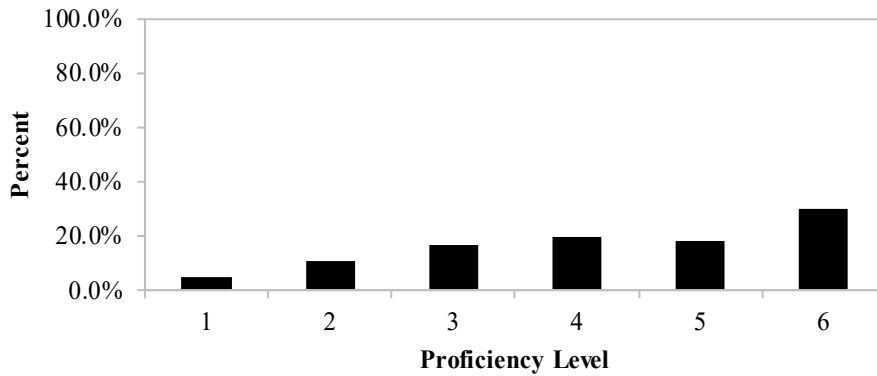
**Table 2.5.1.2.3**

Proficiency Level Distribution: List 2 S503 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	1,376	4.58%	1,376	4.58%
2	3,167	10.55%	3,167	10.55%
3	5,079	16.91%	5,079	16.91%
4	5,837	19.44%	5,837	19.44%
5	5,525	18.40%	5,525	18.40%
6	9,043	30.12%	9,043	30.12%
<b>Total</b>	<b>30,027</b>	<b>100.00%</b>	<b>30,027</b>	<b>100.00%</b>



**Figure 2.5.1.2.3**  
**Proficiency Level: List 2 S503 Paper**



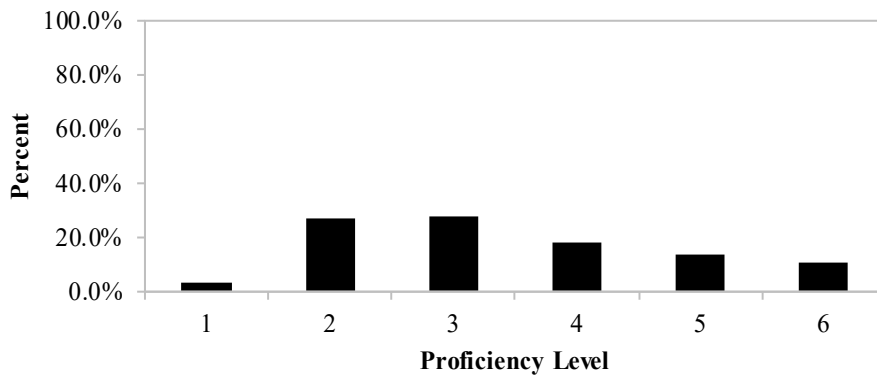
### 2.5.1.3 Grade 3

**Table 2.5.1.3.1**

Proficiency Level Distribution: List 3 A S503 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	215	3.35%	215	3.35%
2	1,715	26.72%	1,715	26.72%
3	1,798	28.01%	1,798	28.01%
4	1,157	18.03%	1,157	18.03%
5	855	13.32%	855	13.32%
6	678	10.56%	678	10.56%
<b>Total</b>	6,418	100.00%	6,418	100.00%

**Figure 2.5.1.3.1**  
**Proficiency Level: List 3A S503 Paper**

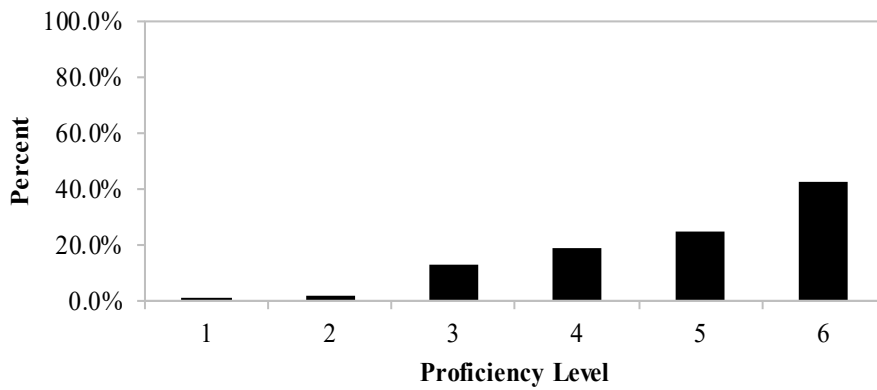


**Table 2.5.1.3.2**

Proficiency Level Distribution: List 3 B/C S503 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	17	0.07%	17	0.07%
2	425	1.85%	425	1.85%
3	2,894	12.60%	2,894	12.60%
4	4,267	18.58%	4,267	18.58%
5	5,658	24.64%	5,658	24.64%
6	9,705	42.26%	9,705	42.26%
<b>Total</b>	22,966	100.00%	22,966	100.00%

**Figure 2.5.1.3.2**  
**Proficiency Level: List 3B/C S503 Paper**

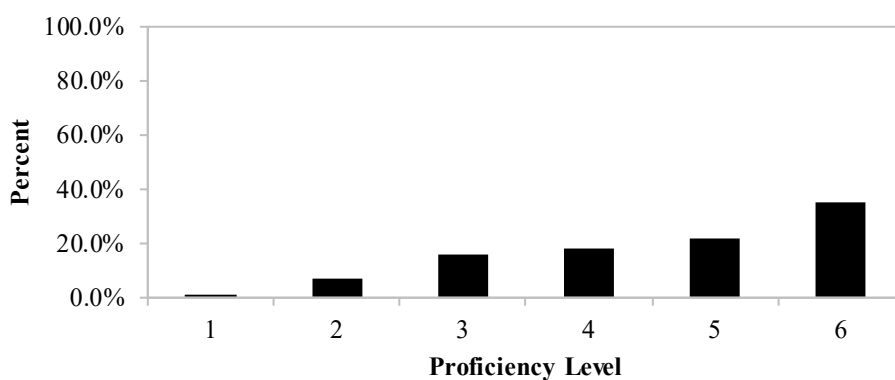


**Table 2.5.1.3.3**

Proficiency Level Distribution: List 3 S503 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	232	0.79%	232	0.79%
2	2,140	7.28%	2,140	7.28%
3	4,692	15.97%	4,692	15.97%
4	5,424	18.46%	5,424	18.46%
5	6,513	22.17%	6,513	22.17%
6	10,383	35.34%	10,383	35.34%
<b>Total</b>	29,384	100.00%	29,384	100.00%

**Figure 2.5.1.3.3**  
**Proficiency Level: List 3 S503 Paper**



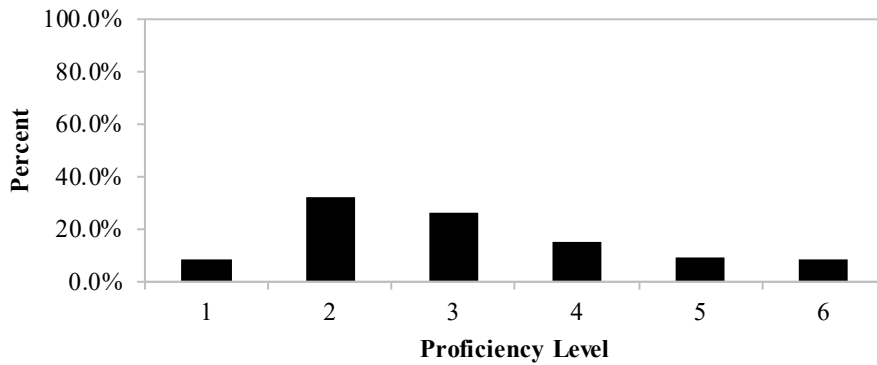
#### 2.5.1.4 Grades 4-5

**Table 2.5.1.4.1**

Proficiency Level Distribution: List 4-5 A S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	304	6.57%	499	11.15%	803	8.82%
2	1,478	31.96%	1,466	32.75%	2,944	32.34%
3	1,248	26.98%	1,109	24.77%	2,357	25.90%
4	697	15.07%	648	14.47%	1,345	14.78%
5	558	12.06%	311	6.95%	869	9.55%
6	340	7.35%	444	9.92%	784	8.61%
<b>Total</b>	4,625	100.00%	4,477	100.00%	9,102	100.00%

**Figure 2.5.1.4.1**  
**Proficiency Level: List 4-5A S503 Paper**

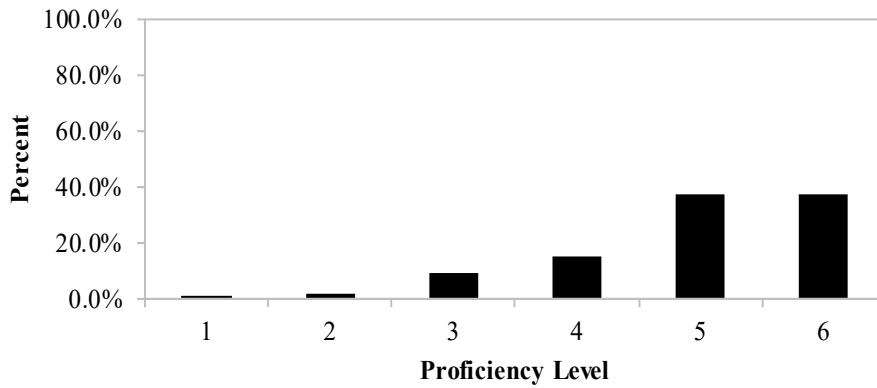


**Table 2.5.1.4.2**

Proficiency Level Distribution: List 4-5 B/C S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	30	0.17%	51	0.28%	81	0.22%
2	295	1.64%	377	2.07%	672	1.86%
3	1,638	9.12%	1,649	9.06%	3,287	9.09%
4	2,650	14.75%	2,697	14.82%	5,347	14.79%
5	6,527	36.33%	6,878	37.80%	13,405	37.07%
6	6,826	37.99%	6,543	35.96%	13,369	36.97%
Total	17,966	100.00%	18,195	100.00%	36,161	100.00%

**Figure 2.5.1.4.2**  
**Proficiency Level: List 4-5B/C S503 Paper**

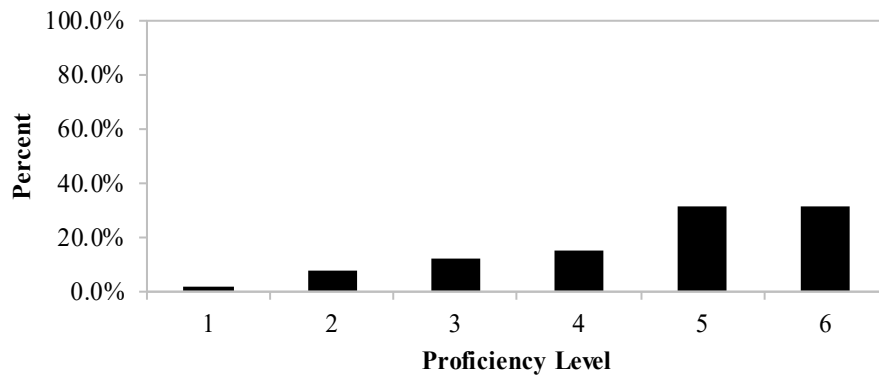


**Table 2.5.1.4.3**

Proficiency Level Distribution: List 4-5 S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	334	1.48%	550	2.43%	884	1.95%
2	1,773	7.85%	1,843	8.13%	3,616	7.99%
3	2,886	12.77%	2,758	12.16%	5,644	12.47%
4	3,347	14.82%	3,345	14.75%	6,692	14.78%
5	7,085	31.36%	7,189	31.71%	14,274	31.54%
6	7,166	31.72%	6,987	30.82%	14,153	31.27%
<b>Total</b>	22,591	100.00%	22,672	100.00%	45,263	100.00%

**Figure 2.5.1.4.3**  
**Proficiency Level: List 4-5 S503 Paper**

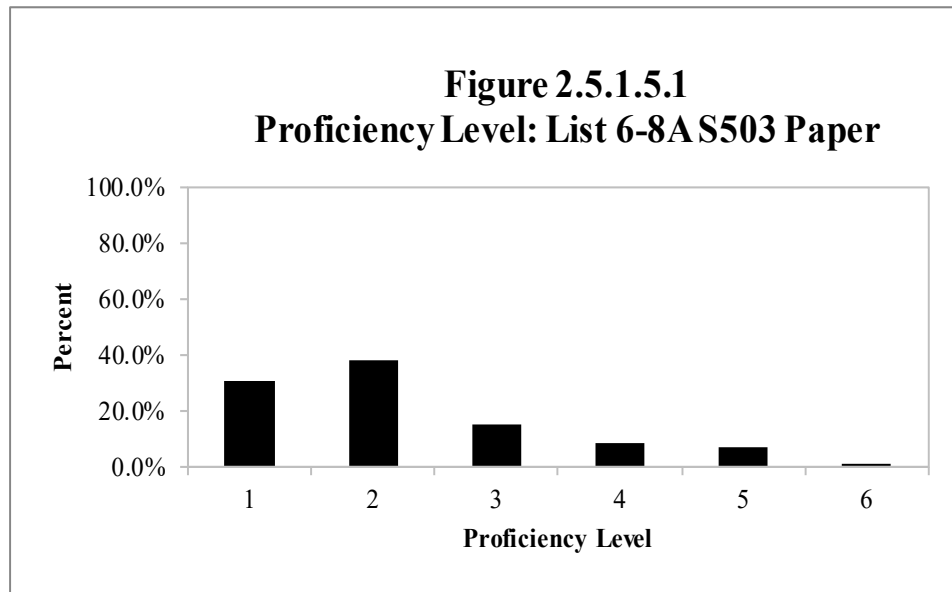


### 2.5.1.5 Grades 6-8

**Table 2.5.1.5.1**

Proficiency Level Distribution: List 6-8 A S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
<b>1</b>	1,088	25.50%	1,424	34.47%	1,321	32.21%	3,833	30.67%
<b>2</b>	1,812	42.47%	1,320	31.95%	1,639	39.97%	4,771	38.17%
<b>3</b>	602	14.11%	804	19.46%	496	12.09%	1,902	15.22%
<b>4</b>	442	10.36%	191	4.62%	396	9.66%	1,029	8.23%
<b>5</b>	267	6.26%	323	7.82%	249	6.07%	839	6.71%
<b>6</b>	56	1.31%	69	1.67%	0	0.00%	125	1.00%
<b>Total</b>	4,267	100.00%	4,131	100.00%	4,101	100.00%	12,499	100.00%



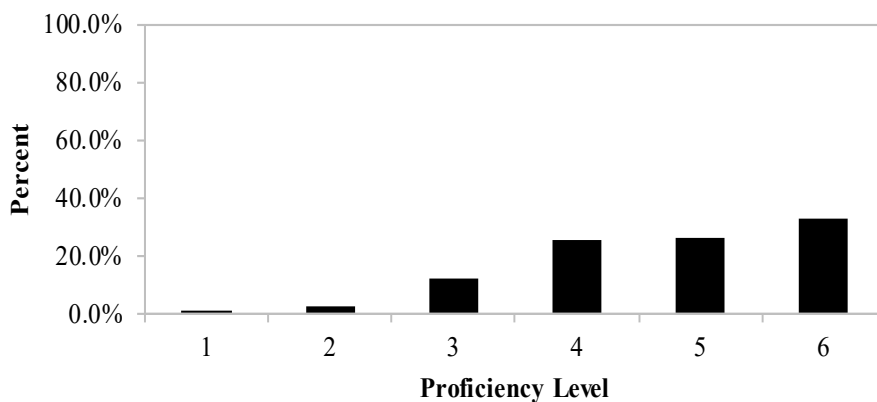


**Table 2.5.1.5.2**

Proficiency Level Distribution: List 6-8 B/C S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
<b>1</b>	11	0.09%	21	0.20%	13	0.13%	45	0.14%
<b>2</b>	217	1.87%	355	3.36%	229	2.29%	801	2.49%
<b>3</b>	1,098	9.46%	1,316	12.47%	1,511	15.12%	3,925	12.21%
<b>4</b>	3,249	28.00%	3,129	29.64%	1,860	18.61%	8,238	25.62%
<b>5</b>	3,058	26.35%	2,753	26.08%	2,691	26.92%	8,502	26.44%
<b>6</b>	3,972	34.23%	2,983	28.26%	3,692	36.93%	10,647	33.11%
<b>Total</b>	11,605	100.00%	10,557	100.00%	9,996	100.00%	32,158	100.00%

**Figure 2.5.1.5.2**  
**Proficiency Level: List 6-8B/C S503 Paper**

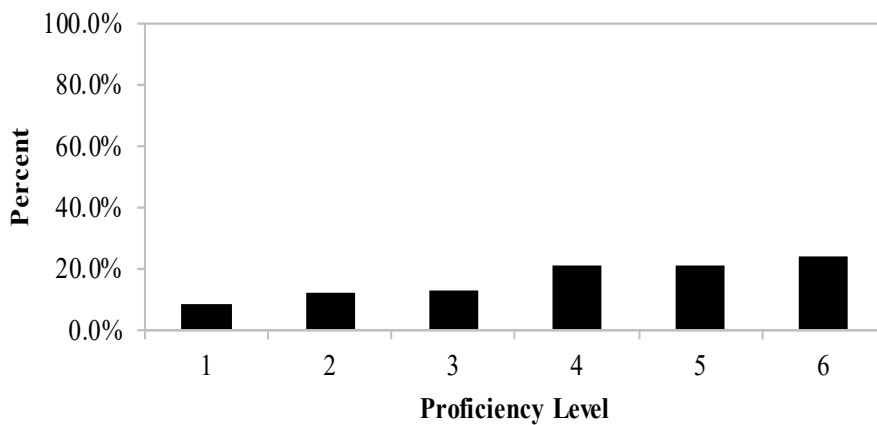


**Table 2.5.1.5.3**

Proficiency Level Distribution: List 6-8 S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
<b>1</b>	1,099	6.92%	1,445	9.84%	1,334	9.46%	3,878	8.68%
<b>2</b>	2,029	12.78%	1,675	11.40%	1,868	13.25%	5,572	12.48%
<b>3</b>	1,700	10.71%	2,120	14.43%	2,007	14.24%	5,827	13.05%
<b>4</b>	3,691	23.25%	3,320	22.60%	2,256	16.00%	9,267	20.75%
<b>5</b>	3,325	20.95%	3,076	20.94%	2,940	20.86%	9,341	20.92%
<b>6</b>	4,028	25.38%	3,052	20.78%	3,692	26.19%	10,772	24.12%
<b>Total</b>	15,872	100.00%	14,688	100.00%	14,097	100.00%	44,657	100.00%

**Figure 2.5.1.5.3**  
**Proficiency Level: List 6-8 S503 Paper**

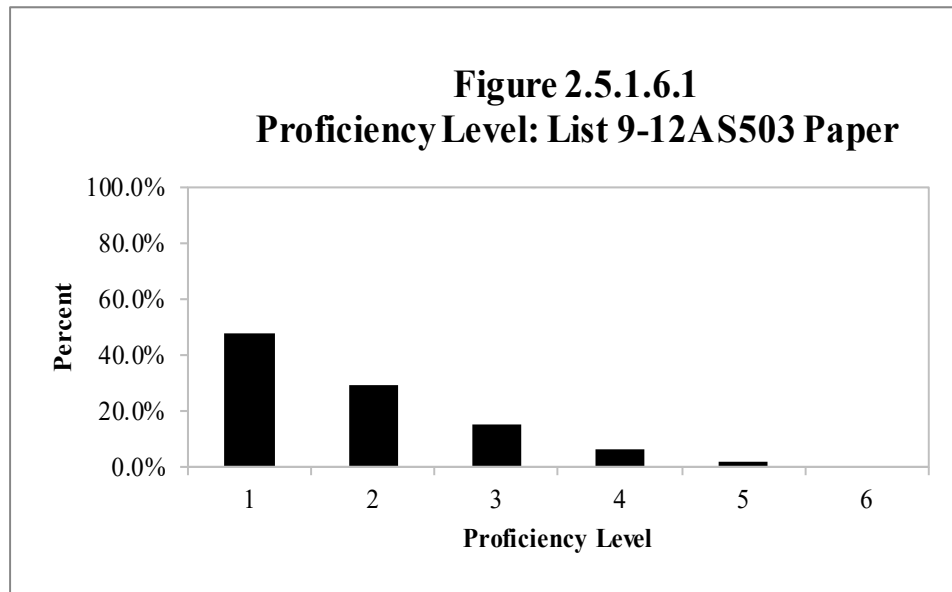


### 2.5.1.6 Grades 9-12

**Table 2.5.1.6.1**

Proficiency Level Distribution: List 9-12 A S503 Paper

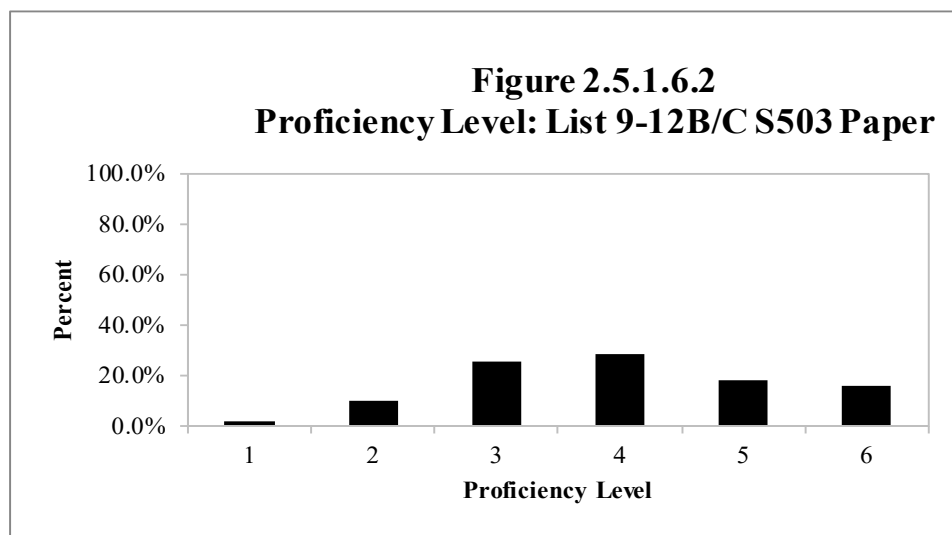
Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,018	43.63%	1,669	50.82%	1,338	48.18%	744	54.59%	5,769	47.88%
2	1,570	33.95%	745	22.69%	901	32.45%	293	21.50%	3,509	29.12%
3	632	13.66%	665	20.25%	355	12.78%	211	15.48%	1,863	15.46%
4	331	7.16%	120	3.65%	162	5.83%	103	7.56%	716	5.94%
5	74	1.60%	85	2.59%	21	0.76%	12	0.88%	192	1.59%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
<b>Total</b>	4,625	100.00%	3,284	100.00%	2,777	100.00%	1,363	100.00%	12,049	100.00%



**Table 2.5.1.6.2**

Proficiency Level Distribution: List 9-12 B/C S503 Paper

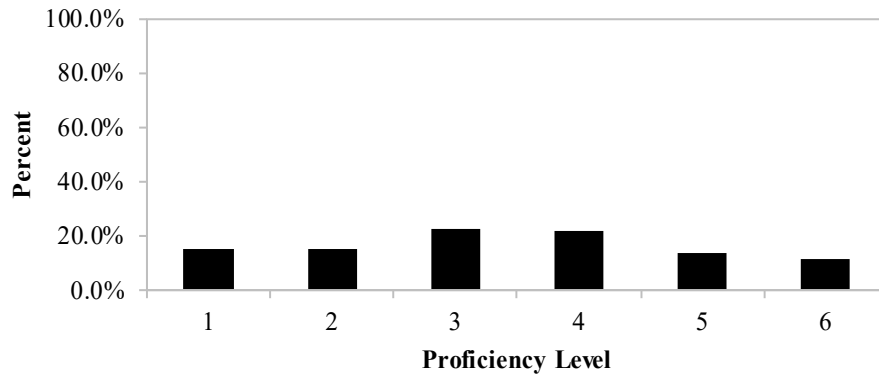
Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	34	0.41%	91	1.11%	203	2.74%	287	5.24%	615	2.09%
2	672	8.15%	617	7.50%	732	9.87%	849	15.51%	2,870	9.77%
3	1,581	19.18%	2,244	27.28%	2,323	31.31%	1,440	26.30%	7,588	25.84%
4	2,600	31.54%	2,540	30.87%	1,559	21.01%	1,650	30.14%	8,349	28.43%
5	1,664	20.18%	1,443	17.54%	1,407	18.96%	814	14.87%	5,328	18.14%
6	1,693	20.54%	1,292	15.70%	1,195	16.11%	435	7.95%	4,615	15.72%
Total	8,244	100.00%	8,227	100.00%	7,419	100.00%	5,475	100.00%	29,365	100.00%

**Table 2.5.1.6.3**

Proficiency Level Distribution: List 9-12 S503 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,052	15.95%	1,760	15.29%	1,541	15.11%	1,031	15.08%	6,384	15.42%
2	2,242	17.42%	1,362	11.83%	1,633	16.02%	1,142	16.70%	6,379	15.40%
3	2,213	17.20%	2,909	25.27%	2,678	26.27%	1,651	24.14%	9,451	22.82%
4	2,931	22.78%	2,660	23.11%	1,721	16.88%	1,753	25.64%	9,065	21.89%
5	1,738	13.51%	1,528	13.27%	1,428	14.01%	826	12.08%	5,520	13.33%
6	1,693	13.16%	1,292	11.22%	1,195	11.72%	435	6.36%	4,615	11.14%
Total	12,869	100.00%	11,511	100.00%	10,196	100.00%	6,838	100.00%	41,414	100.00%

**Figure 2.5.1.6.3**  
**Proficiency Level: List 9-12 S503 Paper**



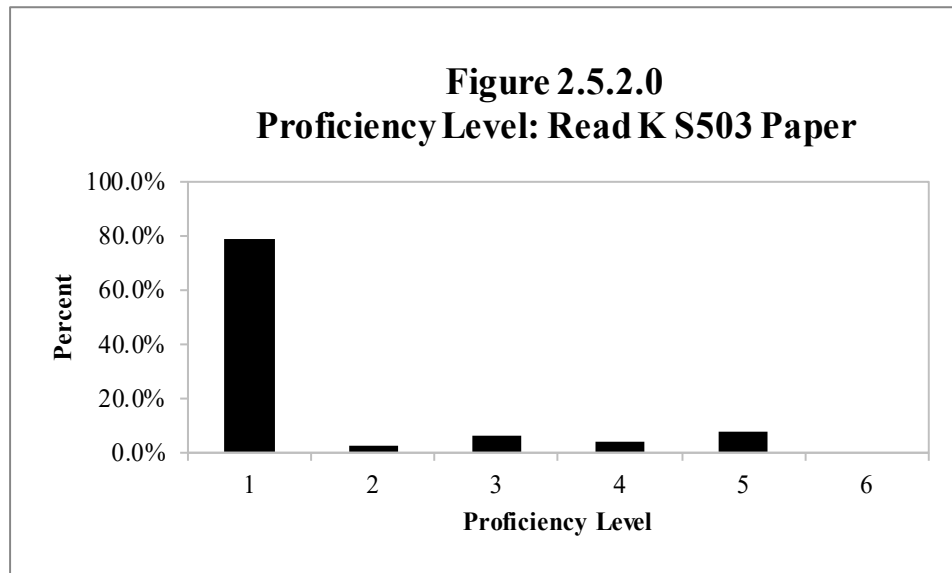
## 2.5.2 Reading

### 2.5.2.0 Kindergarten

**Table 2.5.2.0**

Proficiency Level Distribution: Read K S503 Paper

Level	Grade K		Total	
	Count	Percent	Count	Percent
1	198,565	78.96%	198,565	78.96%
2	5,909	2.35%	5,909	2.35%
3	15,958	6.35%	15,958	6.35%
4	11,057	4.40%	11,057	4.40%
5	20,002	7.95%	20,002	7.95%
6	0	0.00%	0	0.00%
Total	251,491	100.00%	251,491	100.00%

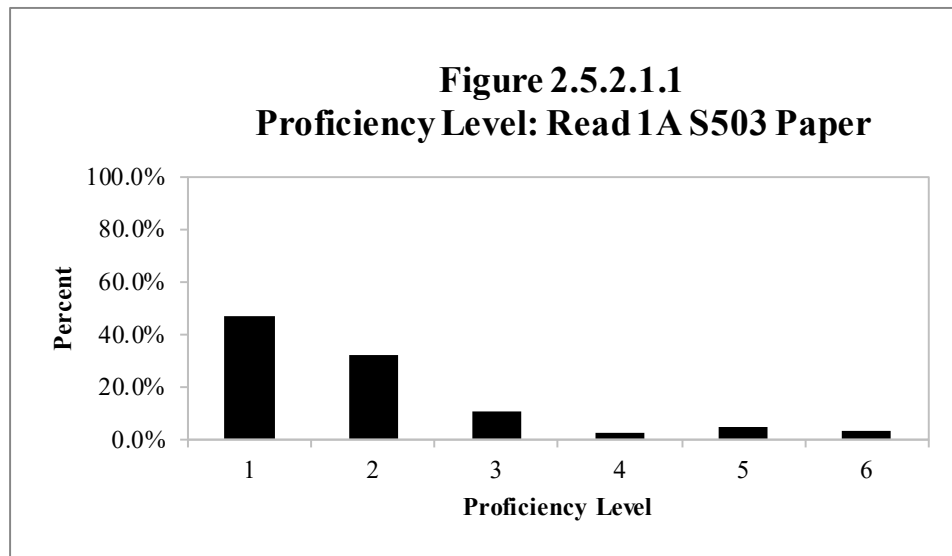


### 2.5.2.1 Grade 1

**Table 2.5.2.1.1**

Proficiency Level Distribution: Read 1 A S503 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	6,836	47.07%	6,836	47.07%
2	4,615	31.78%	4,615	31.78%
3	1,523	10.49%	1,523	10.49%
4	423	2.91%	423	2.91%
5	659	4.54%	659	4.54%
6	466	3.21%	466	3.21%
Total	14,522	100.00%	14,522	100.00%

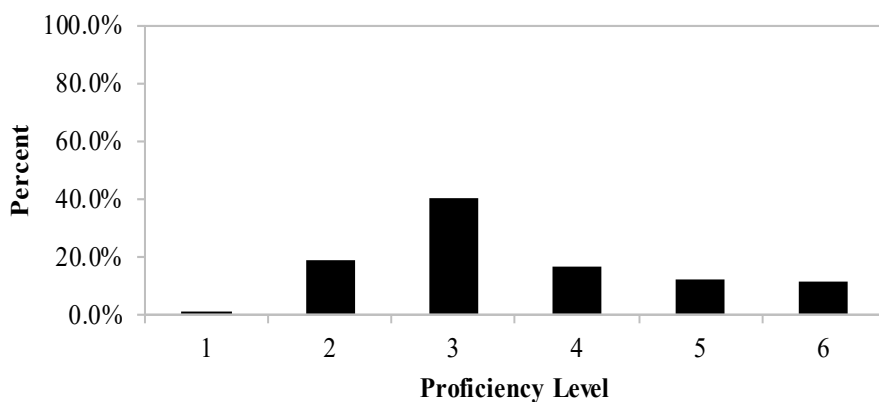


**Table 2.5.2.1.2**

Proficiency Level Distribution: Read 1 B/C S503 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	87	0.77%	87	0.77%
2	2,096	18.47%	2,096	18.47%
3	4,601	40.55%	4,601	40.55%
4	1,884	16.60%	1,884	16.60%
5	1,391	12.26%	1,391	12.26%
6	1,287	11.34%	1,287	11.34%
Total	11,346	100.00%	11,346	100.00%

**Figure 2.5.2.1.2**  
**Proficiency Level: Read 1B/C S503 Paper**



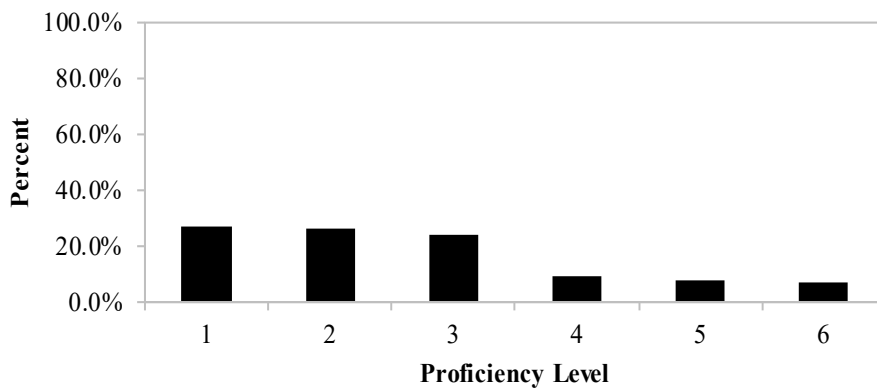


**Table 2.5.2.1.3**

Proficiency Level Distribution: Read 1 S503 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	6,923	26.76%	6,923	26.76%
2	6,711	25.94%	6,711	25.94%
3	6,124	23.67%	6,124	23.67%
4	2,307	8.92%	2,307	8.92%
5	2,050	7.92%	2,050	7.92%
6	1,753	6.78%	1,753	6.78%
Total	25,868	100.00%	25,868	100.00%

**Figure 2.5.2.1.3**  
**Proficiency Level: Read 1 S503 Paper**

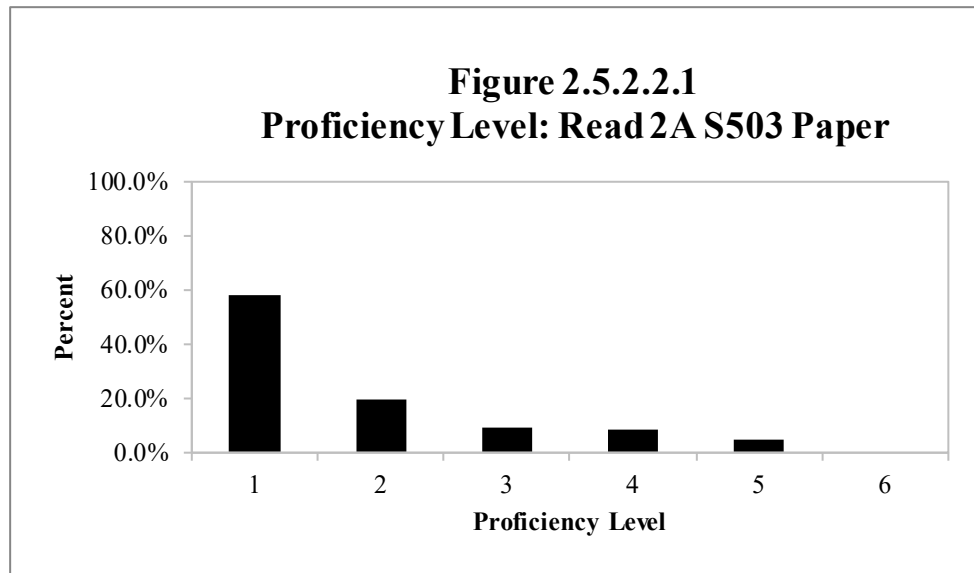


### 2.5.2.2 Grade 2

**Table 2.5.2.2.1**

Proficiency Level Distribution: Read 2 A S503 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	4,595	58.07%	4,595	58.07%
2	1,546	19.54%	1,546	19.54%
3	714	9.02%	714	9.02%
4	668	8.44%	668	8.44%
5	390	4.93%	390	4.93%
6	0	0.00%	0	0.00%
Total	7,913	100.00%	7,913	100.00%

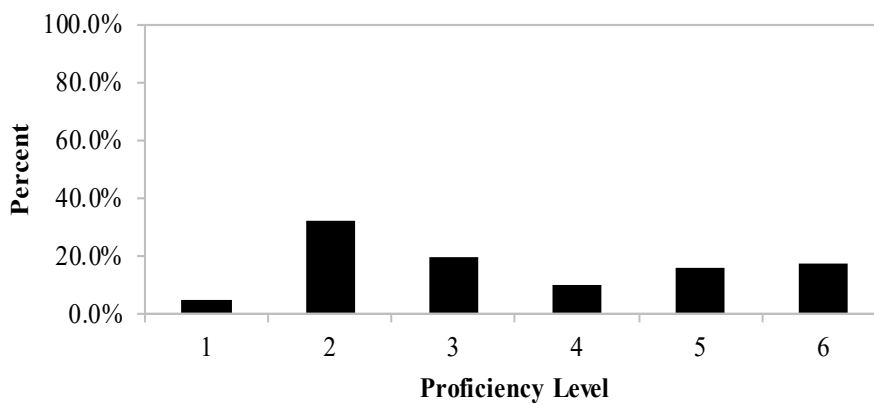


**Table 2.5.2.2.2**

Proficiency Level Distribution: Read 2 B/C S503 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	946	4.99%	946	4.99%
2	6,149	32.45%	6,149	32.45%
3	3,692	19.48%	3,692	19.48%
4	1,881	9.93%	1,881	9.93%
5	2,957	15.60%	2,957	15.60%
6	3,326	17.55%	3,326	17.55%
Total	18,951	100.00%	18,951	100.00%

**Figure 2.5.2.2.2**  
**Proficiency Level: Read 2B/C S503 Paper**

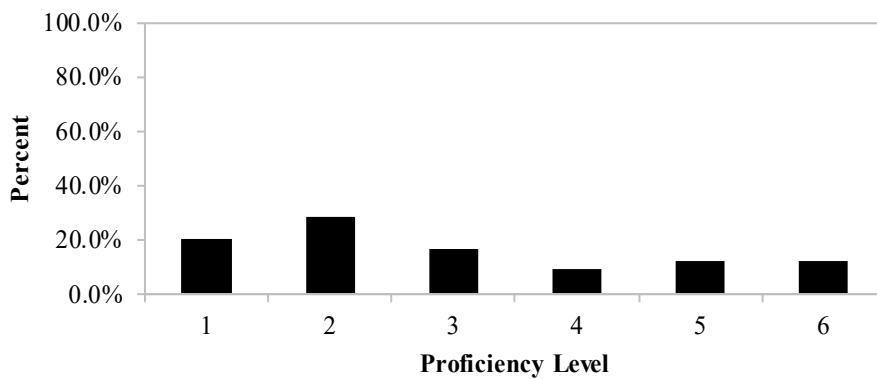


**Table 2.5.2.2.3**

Proficiency Level Distribution: Read 2 S503 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	5,541	20.63%	5,541	20.63%
2	7,695	28.64%	7,695	28.64%
3	4,406	16.40%	4,406	16.40%
4	2,549	9.49%	2,549	9.49%
5	3,347	12.46%	3,347	12.46%
6	3,326	12.38%	3,326	12.38%
Total	26,864	100.00%	26,864	100.00%

**Figure 2.5.2.2.3**  
**Proficiency Level: Read 2 S503 Paper**

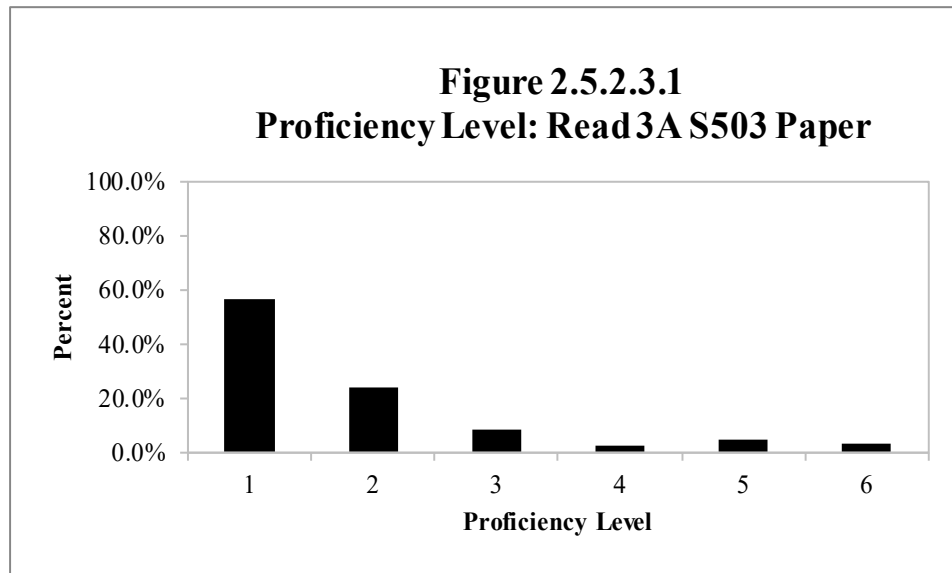


### 2.5.2.3 Grade 3

**Table 2.5.2.3.1**

Proficiency Level Distribution: Read 3 A S503 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	3,258	56.26%	3,258	56.26%
2	1,399	24.16%	1,399	24.16%
3	497	8.58%	497	8.58%
4	156	2.69%	156	2.69%
5	284	4.90%	284	4.90%
6	197	3.40%	197	3.40%
Total	5,791	100.00%	5,791	100.00%

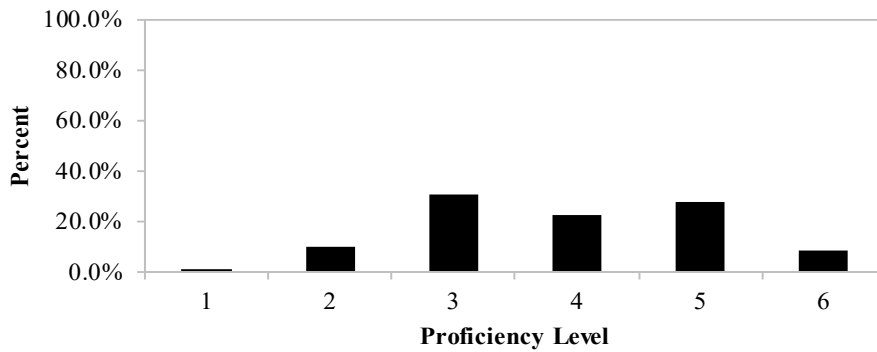


**Table 2.5.2.3.2**

Proficiency Level Distribution: Read 3 B/C S503 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	111	0.54%	111	0.54%
2	2,013	9.84%	2,013	9.84%
3	6,271	30.66%	6,271	30.66%
4	4,562	22.31%	4,562	22.31%
5	5,699	27.87%	5,699	27.87%
6	1,795	8.78%	1,795	8.78%
Total	20,451	100.00%	20,451	100.00%

**Figure 2.5.2.3.2**  
**Proficiency Level: Read 3B/C S503 Paper**

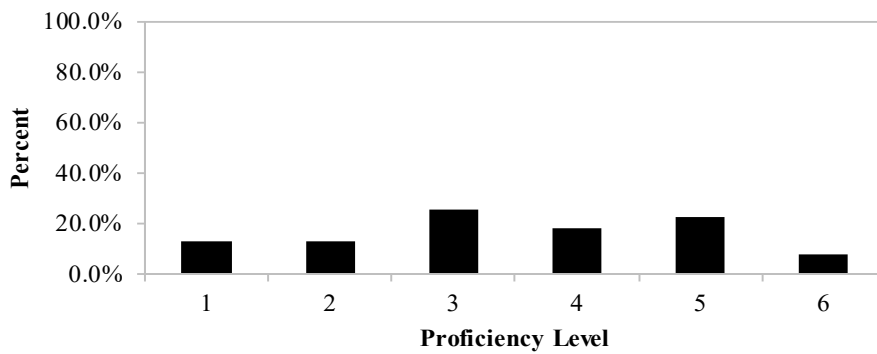


**Table 2.5.2.3.3**

Proficiency Level Distribution: Read 3 S503 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	3,369	12.84%	3,369	12.84%
2	3,412	13.00%	3,412	13.00%
3	6,768	25.79%	6,768	25.79%
4	4,718	17.98%	4,718	17.98%
5	5,983	22.80%	5,983	22.80%
6	1,992	7.59%	1,992	7.59%
Total	26,242	100.00%	26,242	100.00%

**Figure 2.5.2.3.3**  
**Proficiency Level: Read 3 S503 Paper**

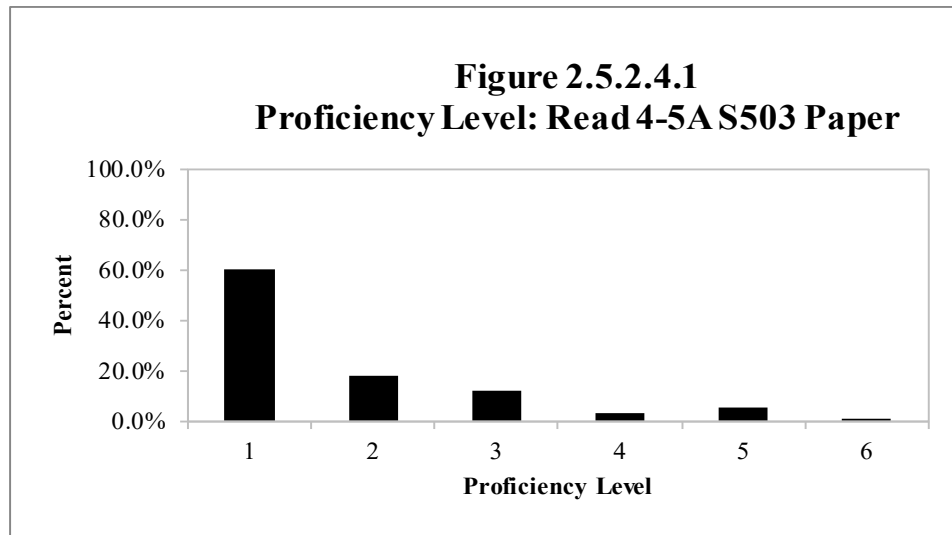


#### 2.5.2.4 Grades 4-5

**Table 2.5.2.4.1**

Proficiency Level Distribution: Read 4-5 A S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	2,581	60.14%	2,564	61.00%	5,145	60.57%
2	852	19.85%	692	16.46%	1,544	18.18%
3	450	10.48%	562	13.37%	1,012	11.91%
4	154	3.59%	156	3.71%	310	3.65%
5	223	5.20%	229	5.45%	452	5.32%
6	32	0.75%	0	0.00%	32	0.38%
<b>Total</b>	4,292	100.00%	4,203	100.00%	8,495	100.00%



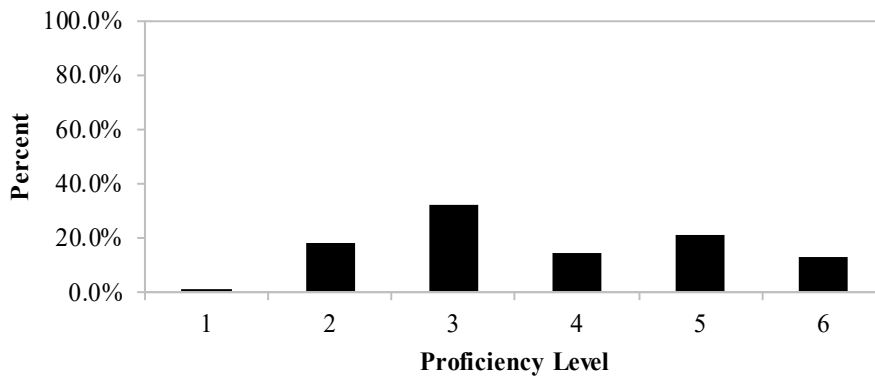


**Table 2.5.2.4.2**

Proficiency Level Distribution: Read 4-5 B/C S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	135	0.83%	231	1.38%	366	1.11%
2	2,456	15.04%	3,557	21.31%	6,013	18.21%
3	5,749	35.21%	4,856	29.10%	10,605	32.12%
4	2,490	15.25%	2,280	13.66%	4,770	14.45%
5	3,477	21.29%	3,565	21.36%	7,042	21.33%
6	2,022	12.38%	2,199	13.18%	4,221	12.78%
Total	16,329	100.00%	16,688	100.00%	33,017	100.00%

**Figure 2.5.2.4.2**  
**Proficiency Level: Read 4-5B/C S503 Paper**

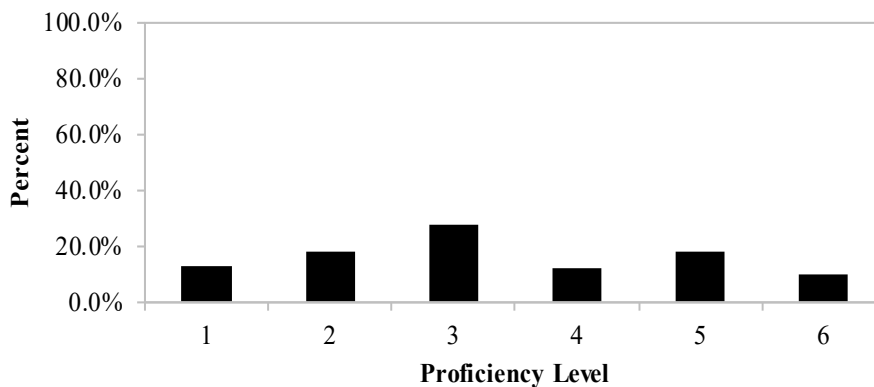


**Table 2.5.2.4.3**

Proficiency Level Distribution: Read 4-5 S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	2,716	13.17%	2,795	13.38%	5,511	13.28%
2	3,308	16.04%	4,249	20.34%	7,557	18.20%
3	6,199	30.06%	5,418	25.93%	11,617	27.98%
4	2,644	12.82%	2,436	11.66%	5,080	12.24%
5	3,700	17.94%	3,794	18.16%	7,494	18.05%
6	2,054	9.96%	2,199	10.53%	4,253	10.25%
Total	20,621	100.00%	20,891	100.00%	41,512	100.00%

**Figure 2.5.2.4.3**  
**Proficiency Level: Read 4-5 S503 Paper**

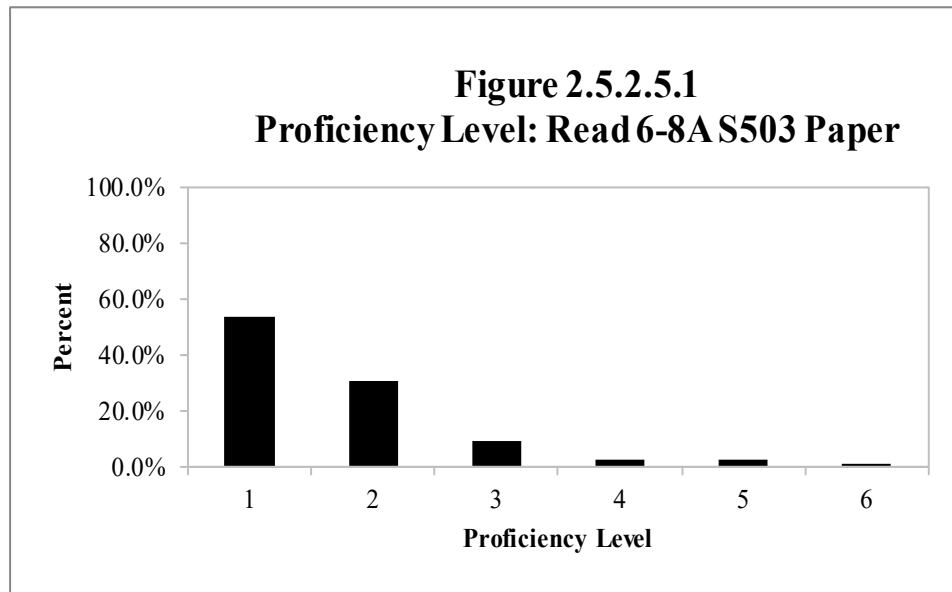


### 2.5.2.5 Grades 6-8

**Table 2.5.2.5.1**

Proficiency Level Distribution: Read 6-8 A S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
<b>1</b>	1,931	47.25%	2,263	56.41%	2,271	56.65%	6,465	53.39%
<b>2</b>	1,501	36.73%	1,101	27.44%	1,105	27.56%	3,707	30.62%
<b>3</b>	421	10.30%	406	10.12%	305	7.61%	1,132	9.35%
<b>4</b>	87	2.13%	98	2.44%	122	3.04%	307	2.54%
<b>5</b>	104	2.54%	79	1.97%	161	4.02%	344	2.84%
<b>6</b>	43	1.05%	65	1.62%	45	1.12%	153	1.26%
<b>Total</b>	4,087	100.00%	4,012	100.00%	4,009	100.00%	12,108	100.00%

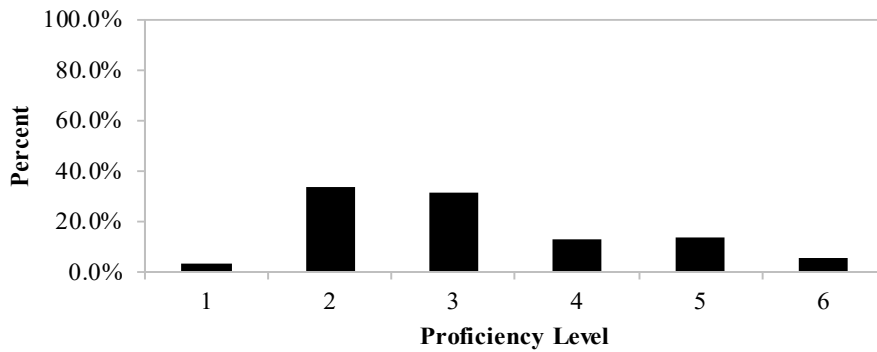


**Table 2.5.2.5.2**

Proficiency Level Distribution: Read 6-8 B/C S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	212	2.07%	333	3.51%	405	4.44%	950	3.29%
2	3,627	35.35%	3,095	32.62%	3,032	33.25%	9,754	33.79%
3	3,298	32.14%	3,300	34.78%	2,399	26.31%	8,997	31.17%
4	1,253	12.21%	1,132	11.93%	1,255	13.76%	3,640	12.61%
5	1,446	14.09%	1,104	11.63%	1,359	14.90%	3,909	13.54%
6	425	4.14%	525	5.53%	668	7.33%	1,618	5.60%
Total	10,261	100.00%	9,489	100.00%	9,118	100.00%	28,868	100.00%

**Figure 2.5.2.5.2**  
**Proficiency Level: Read 6-8B/C S503 Paper**

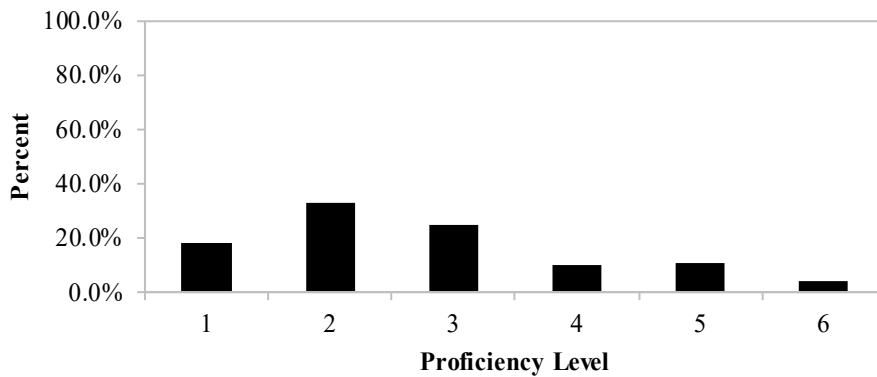


**Table 2.5.2.5.3**

Proficiency Level Distribution: Read 6-8 S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,143	14.94%	2,596	19.23%	2,676	20.39%	7,415	18.10%
2	5,128	35.74%	4,196	31.08%	4,137	31.52%	13,461	32.85%
3	3,719	25.92%	3,706	27.45%	2,704	20.60%	10,129	24.72%
4	1,340	9.34%	1,230	9.11%	1,377	10.49%	3,947	9.63%
5	1,550	10.80%	1,183	8.76%	1,520	11.58%	4,253	10.38%
6	468	3.26%	590	4.37%	713	5.43%	1,771	4.32%
Total	14,348	100.00%	13,501	100.00%	13,127	100.00%	40,976	100.00%

**Figure 2.5.2.5.3**  
**Proficiency Level: Read 6-8 S503 Paper**

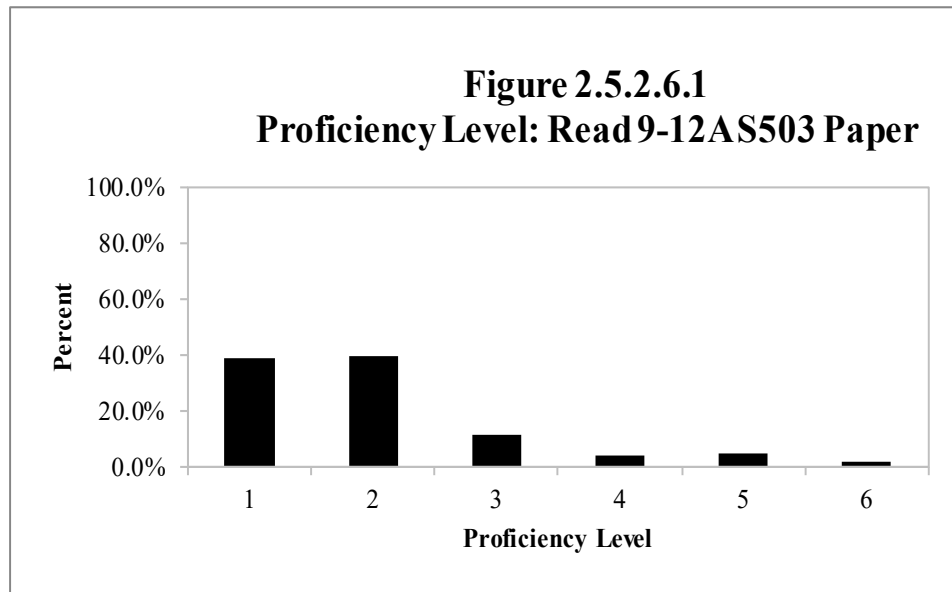


### 2.5.2.6 Grades 9-12

**Table 2.5.2.6.1**

Proficiency Level Distribution: Read 9-12 A S503 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,755	38.17%	1,224	37.37%	1,066	38.83%	559	41.90%	4,604	38.52%
2	1,849	40.21%	1,298	39.63%	1,045	38.07%	512	38.38%	4,704	39.36%
3	561	12.20%	431	13.16%	282	10.27%	130	9.75%	1,404	11.75%
4	133	2.89%	119	3.63%	129	4.70%	64	4.80%	445	3.72%
5	221	4.81%	104	3.18%	191	6.96%	45	3.37%	561	4.69%
6	79	1.72%	99	3.02%	32	1.17%	24	1.80%	234	1.96%
<b>Total</b>	<b>4,598</b>	<b>100.00%</b>	<b>3,275</b>	<b>100.00%</b>	<b>2,745</b>	<b>100.00%</b>	<b>1,334</b>	<b>100.00%</b>	<b>11,952</b>	<b>100.00%</b>

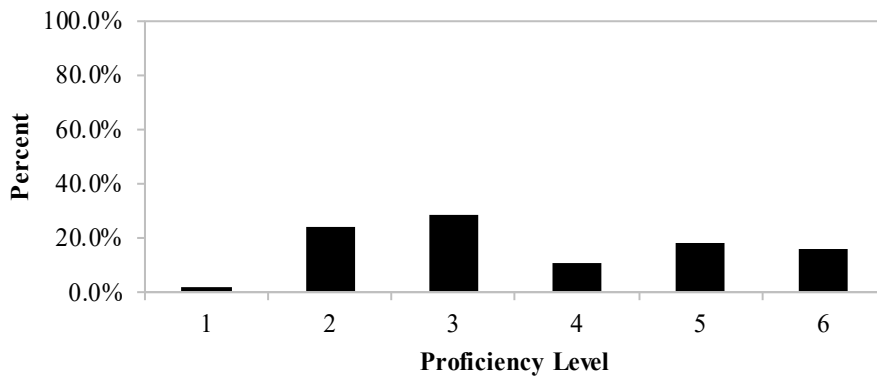


**Table 2.5.2.6.2**

Proficiency Level Distribution: Read 9-12 B/C S503 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	69	0.94%	125	1.69%	213	3.13%	178	3.46%	585	2.19%
2	1,662	22.61%	1,723	23.30%	1,547	22.71%	1,560	30.34%	6,492	24.32%
3	2,246	30.55%	2,122	28.70%	1,871	27.47%	1,409	27.41%	7,648	28.65%
4	900	12.24%	881	11.92%	776	11.39%	300	5.84%	2,857	10.70%
5	1,154	15.70%	1,268	17.15%	1,211	17.78%	1,162	22.60%	4,795	17.96%
6	1,321	17.97%	1,275	17.24%	1,193	17.52%	532	10.35%	4,321	16.18%
Total	7,352	100.00%	7,394	100.00%	6,811	100.00%	5,141	100.00%	26,698	100.00%

**Figure 2.5.2.6.2**  
**Proficiency Level: Read 9-12B/C S503 Paper**

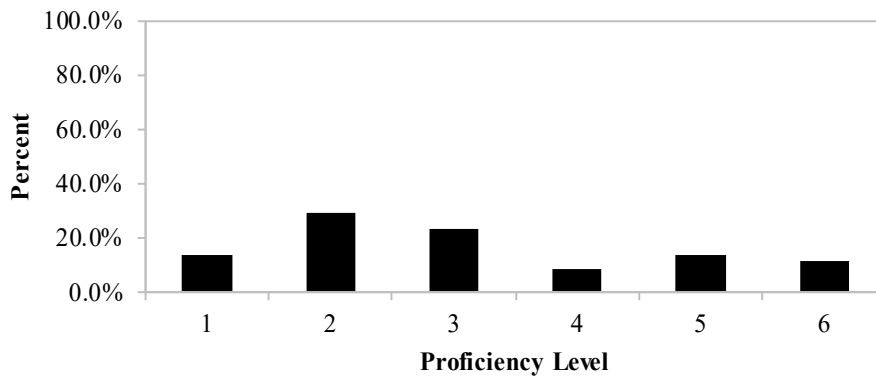


**Table 2.5.2.6.3**

Proficiency Level Distribution: Read 9-12 S503 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,824	15.26%	1,349	12.64%	1,279	13.38%	737	11.38%	5,189	13.43%
2	3,511	29.38%	3,021	28.32%	2,592	27.12%	2,072	32.00%	11,196	28.97%
3	2,807	23.49%	2,553	23.93%	2,153	22.53%	1,539	23.77%	9,052	23.42%
4	1,033	8.64%	1,000	9.37%	905	9.47%	364	5.62%	3,302	8.54%
5	1,375	11.51%	1,372	12.86%	1,402	14.67%	1,207	18.64%	5,356	13.86%
6	1,400	11.72%	1,374	12.88%	1,225	12.82%	556	8.59%	4,555	11.79%
<b>Total</b>	11,950	100.00%	10,669	100.00%	9,556	100.00%	6,475	100.00%	38,650	100.00%

**Figure 2.5.2.6.3**  
**Proficiency Level: Read 9-12 S503 Paper**





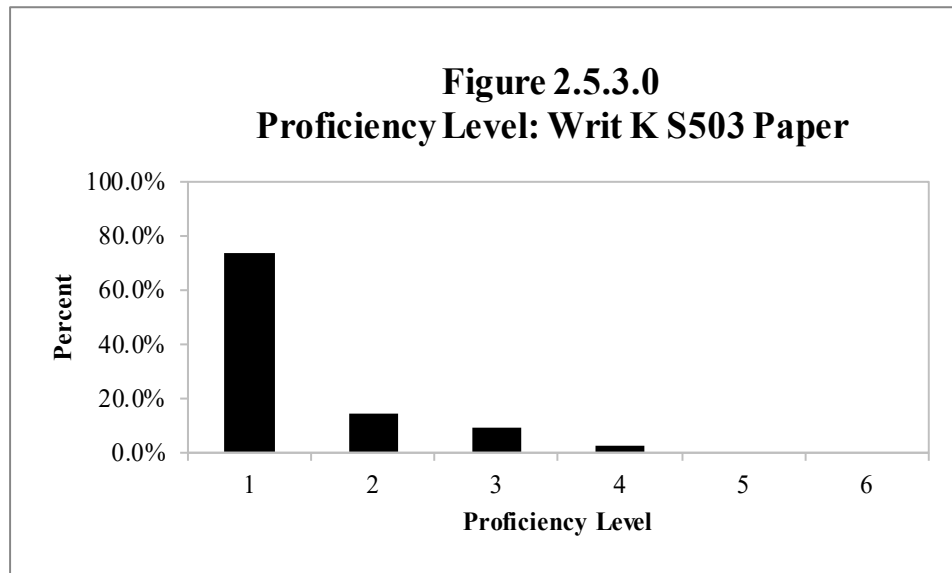
## 2.5.3 Writing

### 2.5.3.0 Kindergarten

**Table 2.5.3.0**

Proficiency Level Distribution: Writ K S503 Paper

Level	Grade K		Total	
	Count	Percent	Count	Percent
1	184,440	73.34%	184,440	73.34%
2	36,463	14.50%	36,463	14.50%
3	23,886	9.50%	23,886	9.50%
4	6,697	2.66%	6,697	2.66%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	251,486	100.00%	251,486	100.00%

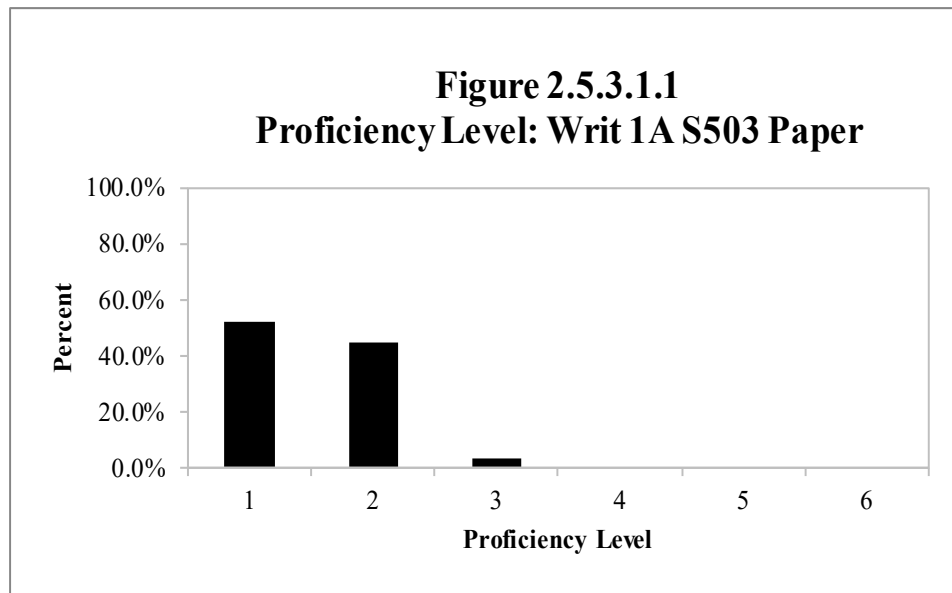


### 2.5.3.1 Grade 1

**Table 2.5.3.1.1**

Proficiency Level Distribution: Writ 1 A S503 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	9,646	51.84%	9,646	51.84%
2	8,344	44.84%	8,344	44.84%
3	617	3.32%	617	3.32%
4	0	0.00%	0	0.00%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	18,607	100.00%	18,607	100.00%

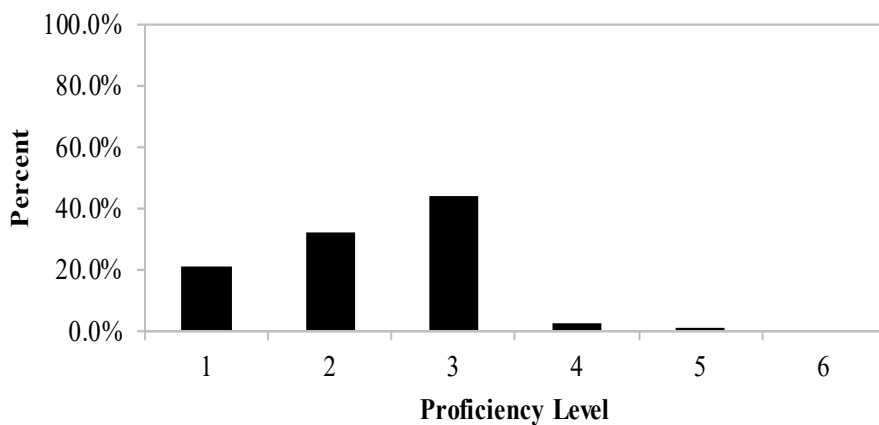


**Table 2.5.3.1.2**

Proficiency Level Distribution: Writ 1 B/C S503 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	3,218	21.14%	3,218	21.14%
2	4,933	32.40%	4,933	32.40%
3	6,722	44.16%	6,722	44.16%
4	346	2.27%	346	2.27%
5	4	0.03%	4	0.03%
6	0	0.00%	0	0.00%
Total	15,223	100.00%	15,223	100.00%

**Figure 2.5.3.1.2**  
**Proficiency Level: Writ 1B/C S503 Paper**

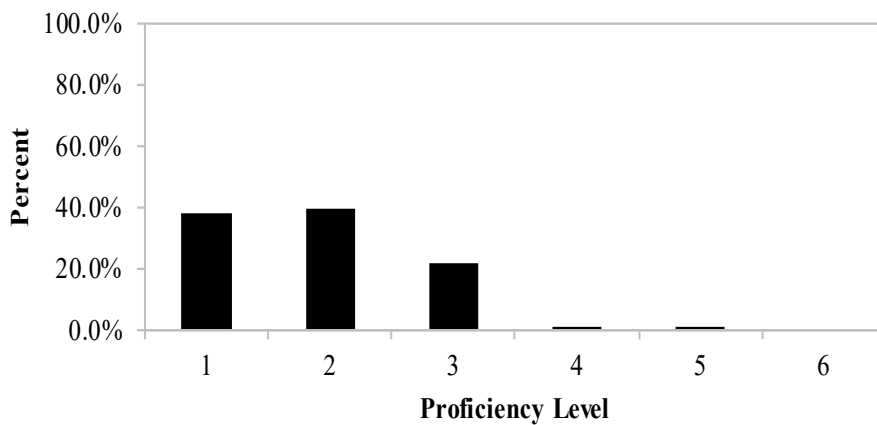


**Table 2.5.3.1.3**

Proficiency Level Distribution: Writ 1 S503 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	12,864	38.03%	12,864	38.03%
2	13,277	39.25%	13,277	39.25%
3	7,339	21.69%	7,339	21.69%
4	346	1.02%	346	1.02%
5	4	0.01%	4	0.01%
6	0	0.00%	0	0.00%
Total	33,830	100.00%	33,830	100.00%

**Figure 2.5.3.1.3**  
**Proficiency Level: Writ 1 S503 Paper**

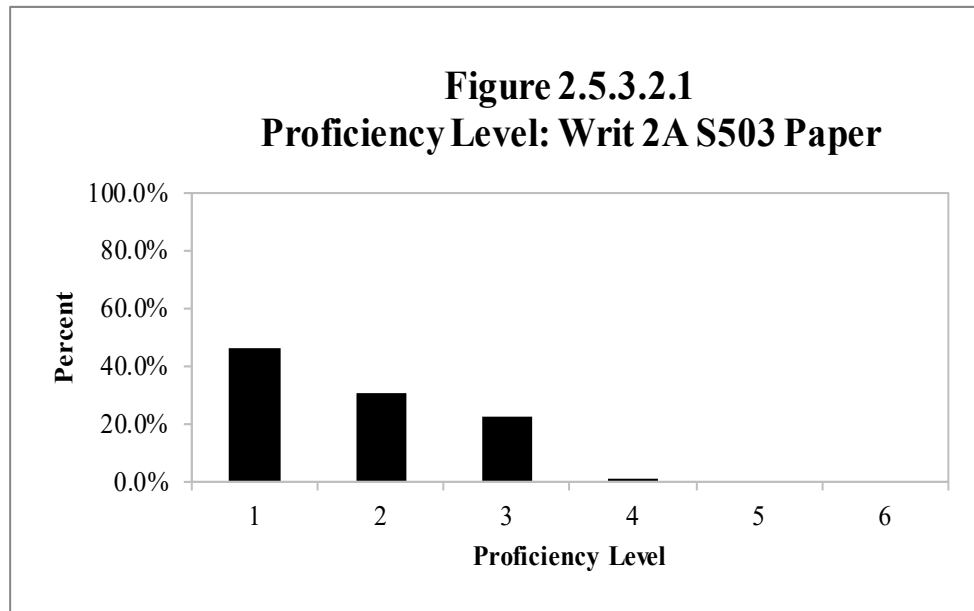


### 2.5.3.2 Grade 2

**Table 2.5.3.2.1**

Proficiency Level Distribution: Writ 2 A S503 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	4,369	46.33%	4,369	46.33%
2	2,915	30.91%	2,915	30.91%
3	2,142	22.71%	2,142	22.71%
4	4	0.04%	4	0.04%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	9,430	100.00%	9,430	100.00%

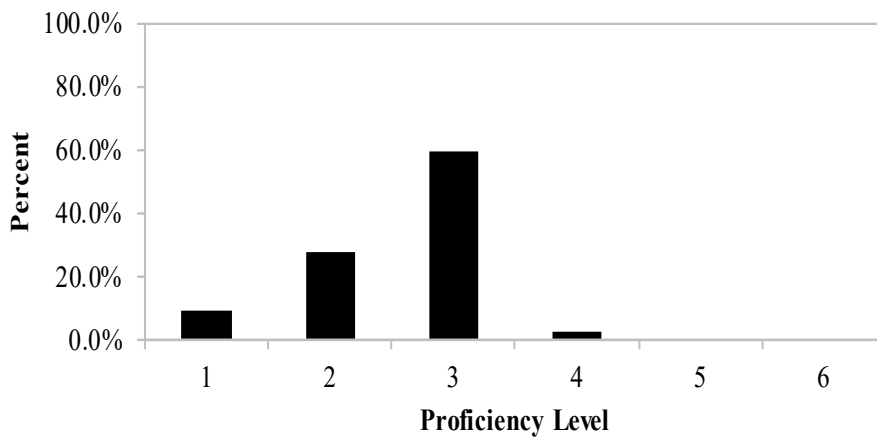


**Table 2.5.3.2.2**

Proficiency Level Distribution: Writ 2 B/C S503 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	2,228	9.60%	2,228	9.60%
2	6,514	28.06%	6,514	28.06%
3	13,812	59.49%	13,812	59.49%
4	662	2.85%	662	2.85%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	23,216	100.00%	23,216	100.00%

**Figure 2.5.3.2.2**  
**Proficiency Level: Writ 2B/C S503 Paper**

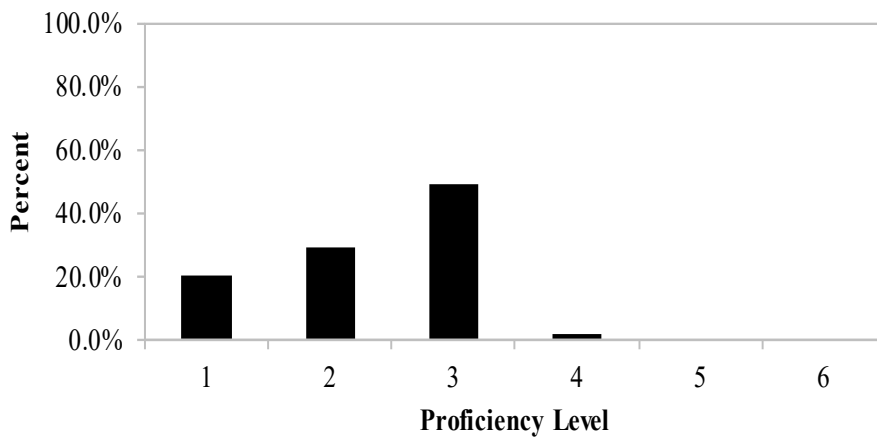


**Table 2.5.3.2.3**

Proficiency Level Distribution: Writ 2 S503 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	6,597	20.21%	6,597	20.21%
2	9,429	28.88%	9,429	28.88%
3	15,954	48.87%	15,954	48.87%
4	666	2.04%	666	2.04%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	32,646	100.00%	32,646	100.00%

**Figure 2.5.3.2.3**  
**Proficiency Level: Writ 2 S503 Paper**



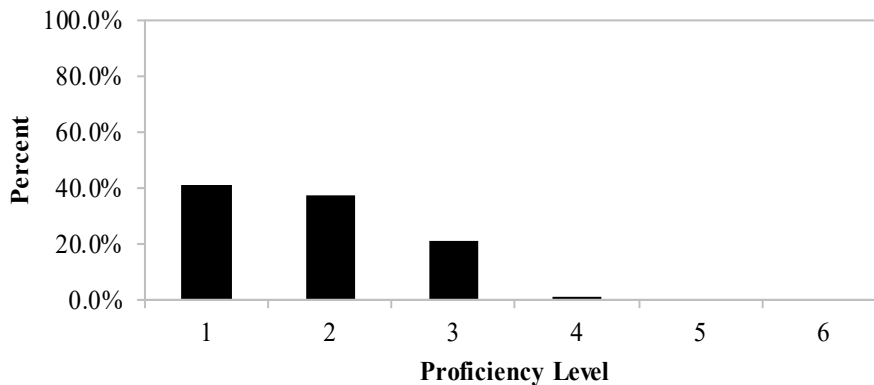
### 2.5.3.3 Grade 3

**Table 2.5.3.3.1**

Proficiency Level Distribution: Writ 3 A S503 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	2,895	41.23%	2,895	41.23%
2	2,614	37.23%	2,614	37.23%
3	1,497	21.32%	1,497	21.32%
4	15	0.21%	15	0.21%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	7,021	100.00%	7,021	100.00%

**Figure 2.5.3.3.1**  
**Proficiency Level: Writ 3A S503 Paper**



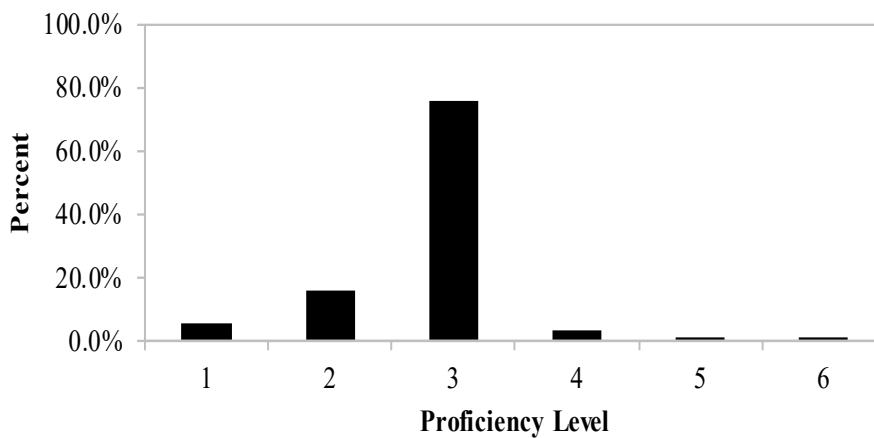


**Table 2.5.3.3.2**

Proficiency Level Distribution: Writ 3 B/C S503 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	1,280	5.21%	1,280	5.21%
2	3,845	15.65%	3,845	15.65%
3	18,572	75.59%	18,572	75.59%
4	869	3.54%	869	3.54%
5	1	0.00%	1	0.00%
6	1	0.00%	1	0.00%
Total	24,568	100.00%	24,568	100.00%

**Figure 2.5.3.3.2**  
**Proficiency Level: Writ 3B/C S503 Paper**

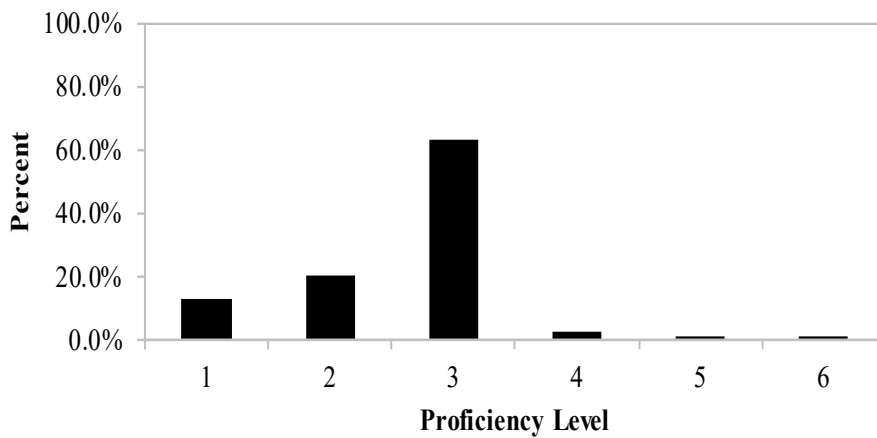


**Table 2.5.3.3.3**

Proficiency Level Distribution: Writ 3 S503 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	4,175	13.22%	4,175	13.22%
2	6,459	20.45%	6,459	20.45%
3	20,069	63.53%	20,069	63.53%
4	884	2.80%	884	2.80%
5	1	0.00%	1	0.00%
6	1	0.00%	1	0.00%
Total	31,589	100.00%	31,589	100.00%

**Figure 2.5.3.3.3**  
**Proficiency Level: Writ 3 S503 Paper**

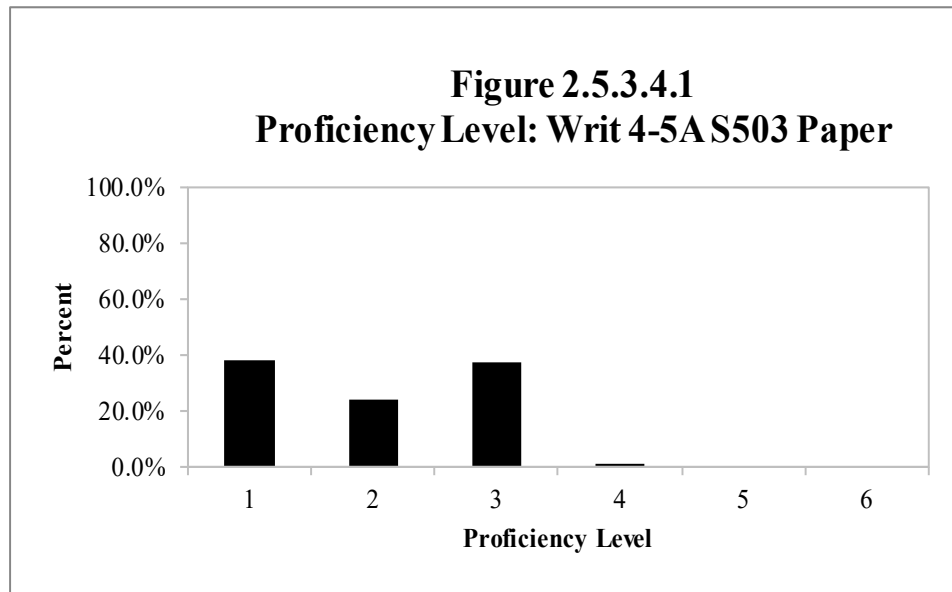


#### 2.5.3.4 Grades 4-5

**Table 2.5.3.4.1**

Proficiency Level Distribution: Writ 4-5 A S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	2,076	41.45%	1,664	34.78%	3,740	38.19%
2	1,192	23.80%	1,145	23.93%	2,337	23.86%
3	1,718	34.31%	1,960	40.96%	3,678	37.56%
4	22	0.44%	16	0.33%	38	0.39%
5	0	0.00%	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%	0	0.00%
<b>Total</b>	5,008	100.00%	4,785	100.00%	9,793	100.00%

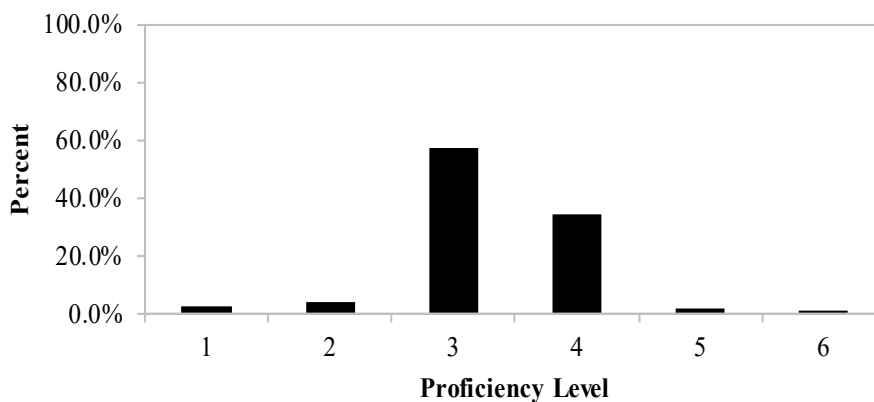


**Table 2.5.3.4.2**

Proficiency Level Distribution: Writ 4-5 B/C S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	522	2.79%	350	1.87%	872	2.33%
2	836	4.47%	698	3.72%	1,534	4.09%
3	11,698	62.50%	9,851	52.50%	21,549	57.49%
4	5,446	29.10%	7,334	39.09%	12,780	34.10%
5	205	1.10%	513	2.73%	718	1.92%
6	10	0.05%	17	0.09%	27	0.07%
Total	18,717	100.00%	18,763	100.00%	37,480	100.00%

**Figure 2.5.3.4.2**  
**Proficiency Level: Writ 4-5B/C S503 Paper**

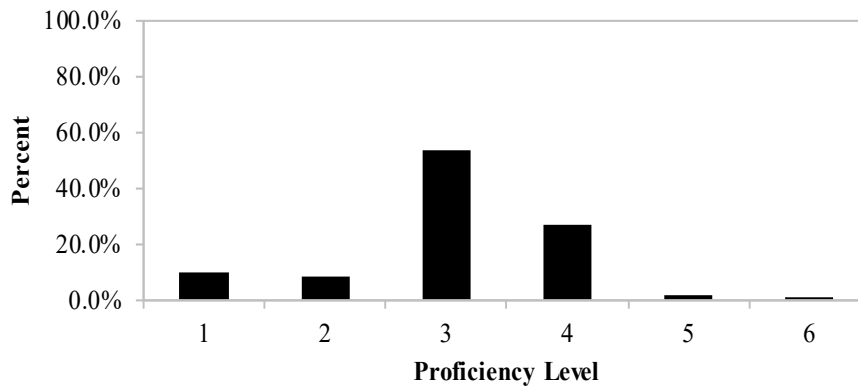


**Table 2.5.3.4.3**

Proficiency Level Distribution: Writ 4-5 S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	2,598	10.95%	2,014	8.55%	4,612	9.76%
2	2,028	8.55%	1,843	7.83%	3,871	8.19%
3	13,416	56.55%	11,811	50.16%	25,227	53.36%
4	5,468	23.05%	7,350	31.21%	12,818	27.11%
5	205	0.86%	513	2.18%	718	1.52%
6	10	0.04%	17	0.07%	27	0.06%
Total	23,725	100.00%	23,548	100.00%	47,273	100.00%

**Figure 2.5.3.4.3**  
**Proficiency Level: Writ 4-5 S503 Paper**

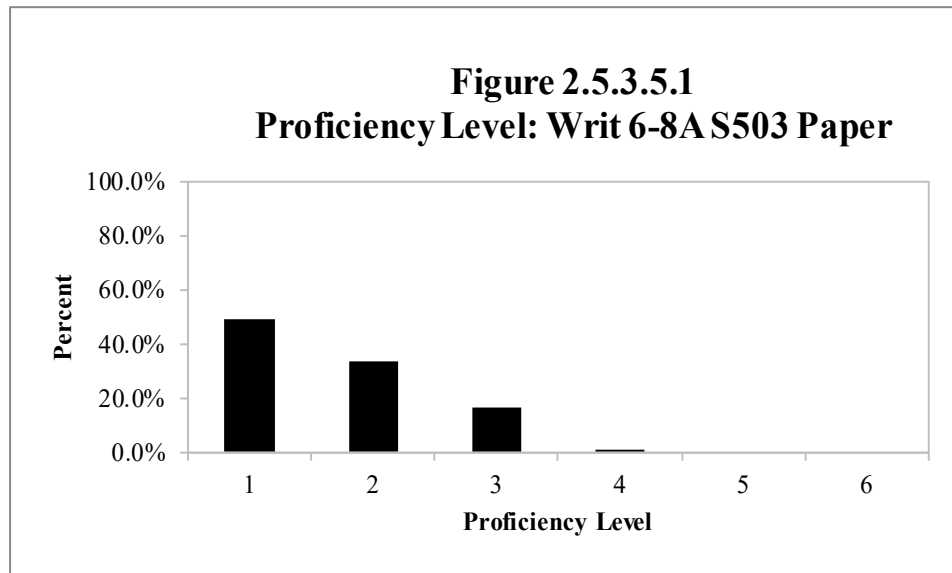


### 2.5.3.5 Grades 6-8

**Table 2.5.3.5.1**

Proficiency Level Distribution: Writ 6-8 A S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
<b>1</b>	2,121	46.47%	2,185	49.22%	2,341	52.90%	6,647	49.50%
<b>2</b>	1,555	34.07%	1,672	37.67%	1,322	29.88%	4,549	33.88%
<b>3</b>	886	19.41%	575	12.95%	759	17.15%	2,220	16.53%
<b>4</b>	2	0.04%	7	0.16%	3	0.07%	12	0.09%
<b>5</b>	0	0.00%	0	0.00%	0	0.00%	0	0.00%
<b>6</b>	0	0.00%	0	0.00%	0	0.00%	0	0.00%
<b>Total</b>	4,564	100.00%	4,439	100.00%	4,425	100.00%	13,428	100.00%

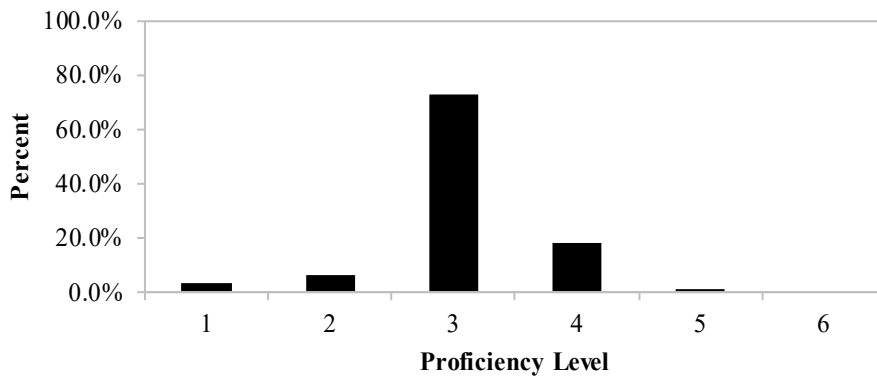


**Table 2.5.3.5.2**

Proficiency Level Distribution: Writ 6-8 B/C S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	300	2.47%	334	3.03%	404	3.87%	1,038	3.09%
2	813	6.70%	805	7.29%	520	4.98%	2,138	6.36%
3	8,617	71.02%	8,078	73.20%	7,706	73.74%	24,401	72.58%
4	2,403	19.80%	1,817	16.47%	1,819	17.41%	6,039	17.96%
5	1	0.01%	1	0.01%	1	0.01%	3	0.01%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	12,134	100.00%	11,035	100.00%	10,450	100.00%	33,619	100.00%

**Figure 2.5.3.5.2**  
**Proficiency Level: Writ 6-8B/C S503 Paper**

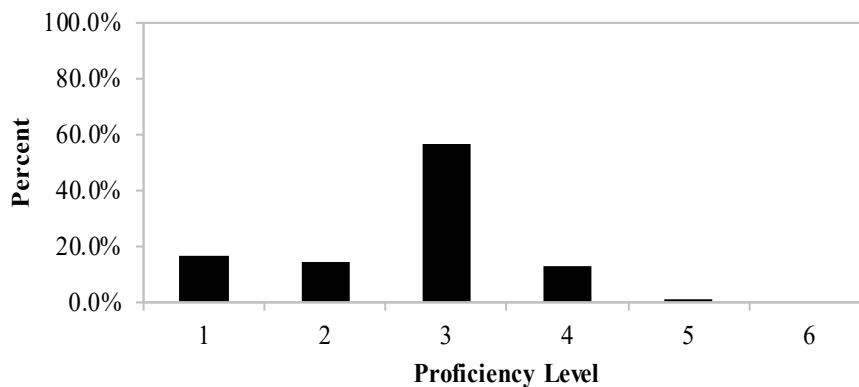


**Table 2.5.3.5.3**

Proficiency Level Distribution: Writ 6-8 S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,421	14.50%	2,519	16.28%	2,745	18.45%	7,685	16.33%
2	2,368	14.18%	2,477	16.01%	1,842	12.38%	6,687	14.21%
3	9,503	56.91%	8,653	55.92%	8,465	56.91%	26,621	56.58%
4	2,405	14.40%	1,824	11.79%	1,822	12.25%	6,051	12.86%
5	1	0.01%	1	0.01%	1	0.01%	3	0.01%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	16,698	100.00%	15,474	100.00%	14,875	100.00%	47,047	100.00%

**Figure 2.5.3.5.3**  
**Proficiency Level: Writ 6-8 S503 Paper**



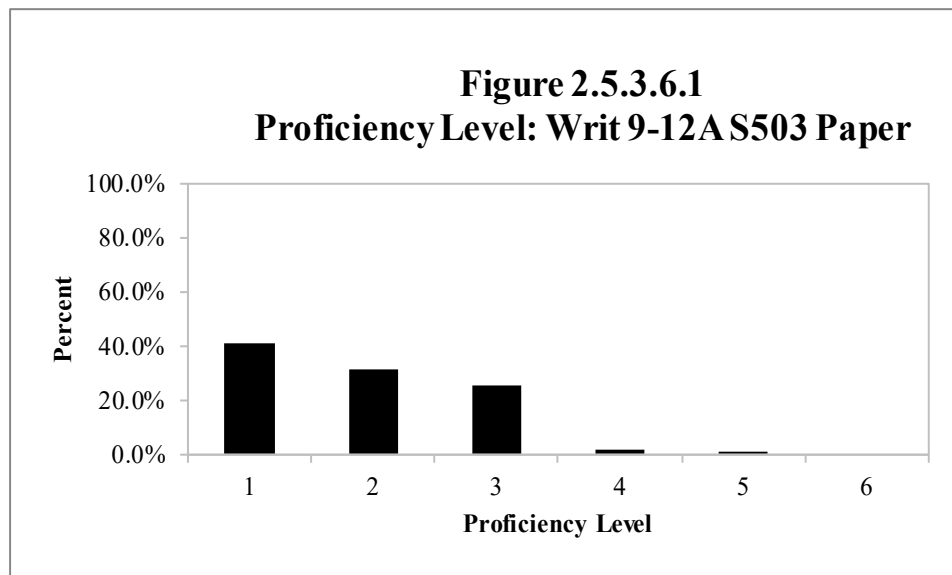


### 2.5.3.6 Grades 9-12

**Table 2.5.3.6.1**

Proficiency Level Distribution: Writ 9-12 A S503 Paper

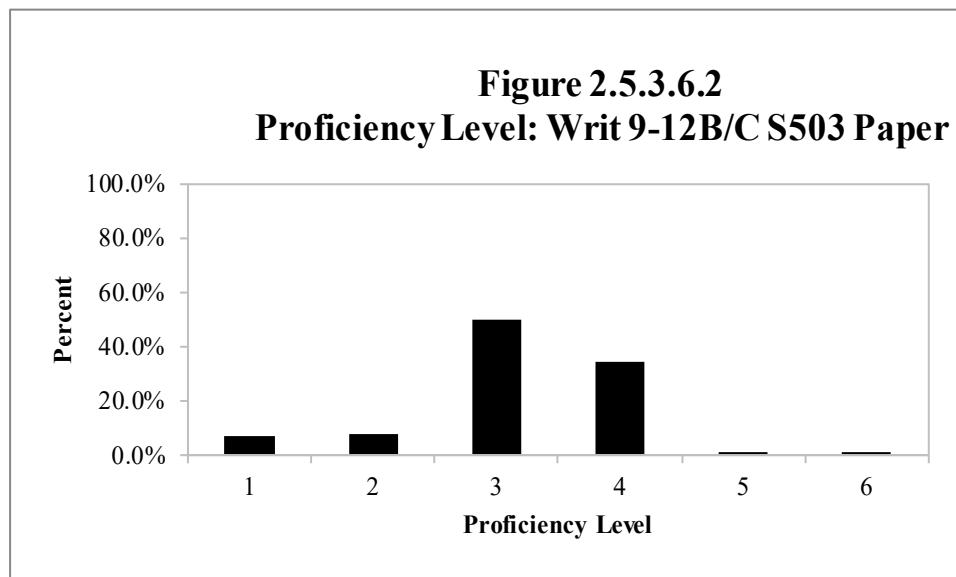
Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,034	39.75%	1,306	36.41%	1,313	43.66%	751	51.79%	5,404	41.06%
2	1,583	30.94%	1,285	35.82%	976	32.46%	287	19.79%	4,131	31.39%
3	1,343	26.25%	947	26.40%	691	22.98%	391	26.97%	3,372	25.62%
4	156	3.05%	49	1.37%	27	0.90%	21	1.45%	253	1.92%
5	1	0.02%	0	0.00%	0	0.00%	0	0.00%	1	0.01%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
<b>Total</b>	5,117	100.00%	3,587	100.00%	3,007	100.00%	1,450	100.00%	13,161	100.00%



**Table 2.5.3.6.2**

Proficiency Level Distribution: Writ 9-12 B/C S503 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	315	3.64%	539	6.24%	596	7.62%	642	11.15%	2,092	6.78%
2	645	7.46%	512	5.93%	596	7.62%	708	12.30%	2,461	7.97%
3	3,776	43.66%	4,540	52.54%	4,063	51.97%	2,992	51.98%	15,371	49.80%
4	3,812	44.08%	2,974	34.42%	2,484	31.77%	1,405	24.41%	10,675	34.59%
5	99	1.14%	75	0.87%	79	1.01%	9	0.16%	262	0.85%
6	1	0.01%	1	0.01%	0	0.00%	0	0.00%	2	0.01%
<b>Total</b>	8,648	100.00%	8,641	100.00%	7,818	100.00%	5,756	100.00%	30,863	100.00%

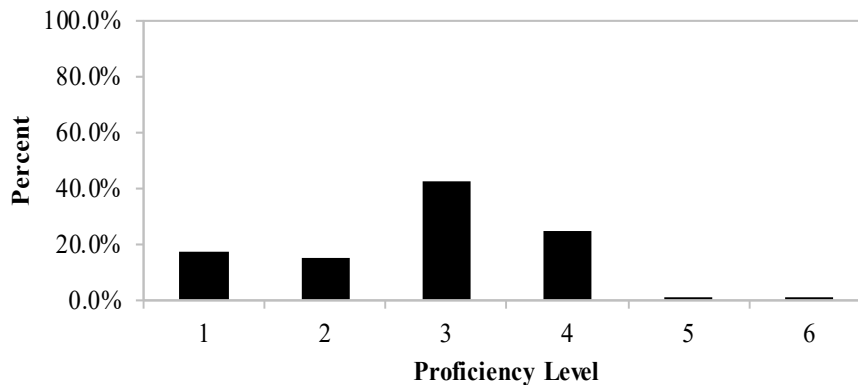


**Table 2.5.3.6.3**

Proficiency Level Distribution: Writ 9-12 S503 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,349	17.07%	1,845	15.09%	1,909	17.64%	1,393	19.33%	7,496	17.03%
2	2,228	16.19%	1,797	14.70%	1,572	14.52%	995	13.81%	6,592	14.97%
3	5,119	37.19%	5,487	44.87%	4,754	43.92%	3,383	46.95%	18,743	42.57%
4	3,968	28.83%	3,023	24.72%	2,511	23.20%	1,426	19.79%	10,928	24.82%
5	100	0.73%	75	0.61%	79	0.73%	9	0.12%	263	0.60%
6	1	0.01%	1	0.01%	0	0.00%	0	0.00%	2	0.00%
Total	13,765	100.00%	12,228	100.00%	10,825	100.00%	7,206	100.00%	44,024	100.00%

**Figure 2.5.3.6.3**  
**Proficiency Level: Writ 9-12 S503 Paper**



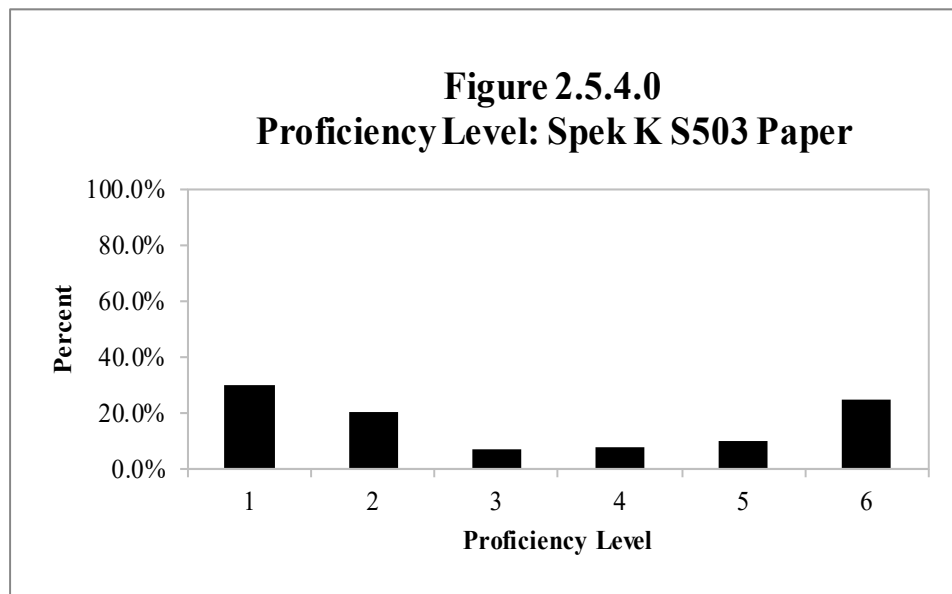
## 2.5.4 Speaking

### 2.5.4.0 Kindergarten

**Table 2.5.4.0**

Proficiency Level Distribution: Spek K S503 Paper

Level	Grade K		Total	
	Count	Percent	Count	Percent
1	75,897	30.18%	75,897	30.18%
2	51,731	20.57%	51,731	20.57%
3	17,261	6.86%	17,261	6.86%
4	19,247	7.65%	19,247	7.65%
5	25,379	10.09%	25,379	10.09%
6	61,968	24.64%	61,968	24.64%
Total	251,483	100.00%	251,483	100.00%

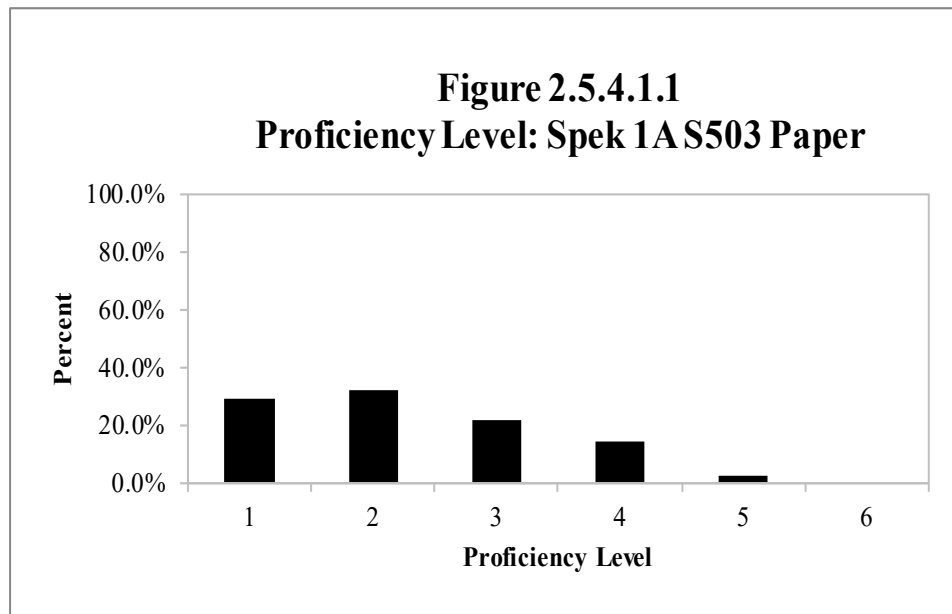


#### 2.5.4.1 Grade 1

**Table 2.5.4.1.1**

Proficiency Level Distribution: Spek 1 A S503 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	5,299	28.85%	5,299	28.85%
2	5,969	32.49%	5,969	32.49%
3	4,069	22.15%	4,069	22.15%
4	2,584	14.07%	2,584	14.07%
5	449	2.44%	449	2.44%
6	0	0.00%	0	0.00%
Total	18,370	100.00%	18,370	100.00%

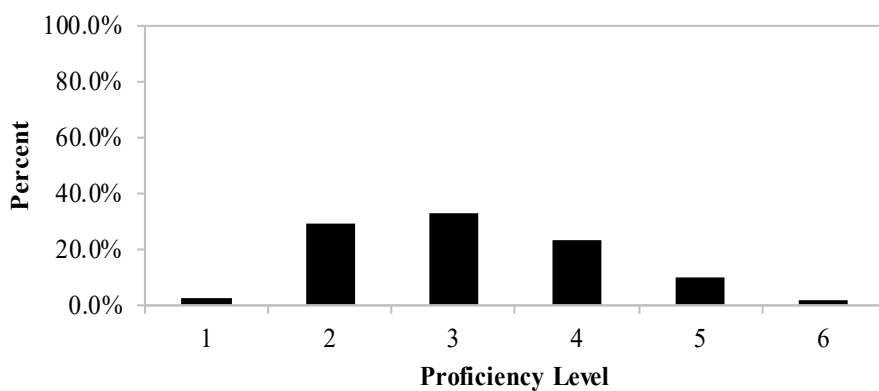


**Table 2.5.4.1.2**

Proficiency Level Distribution: Spek 1 B/C S503 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	401	2.66%	401	2.66%
2	4,393	29.13%	4,393	29.13%
3	4,955	32.86%	4,955	32.86%
4	3,504	23.24%	3,504	23.24%
5	1,521	10.09%	1,521	10.09%
6	306	2.03%	306	2.03%
Total	15,080	100.00%	15,080	100.00%

**Figure 2.5.4.1.2**  
**Proficiency Level: Spek 1B/C S503 Paper**

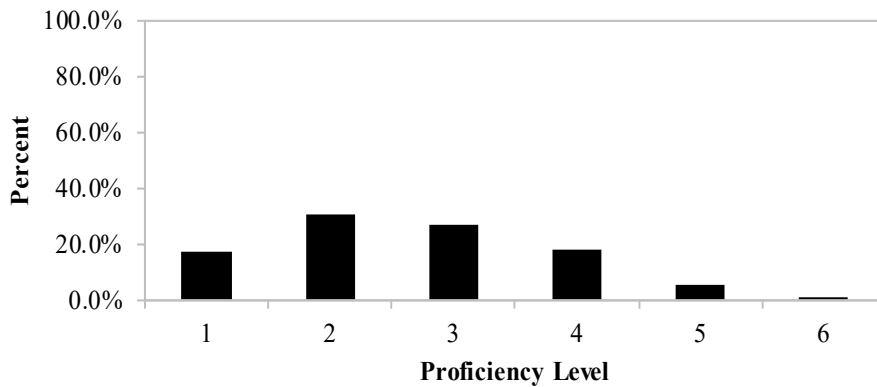


**Table 2.5.4.1.3**

Proficiency Level Distribution: Spek 1 S503 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	5,700	17.04%	5,700	17.04%
2	10,362	30.98%	10,362	30.98%
3	9,024	26.98%	9,024	26.98%
4	6,088	18.20%	6,088	18.20%
5	1,970	5.89%	1,970	5.89%
6	306	0.91%	306	0.91%
Total	33,450	100.00%	33,450	100.00%

**Figure 2.5.4.1.3**  
**Proficiency Level: Spek 1 S503 Paper**

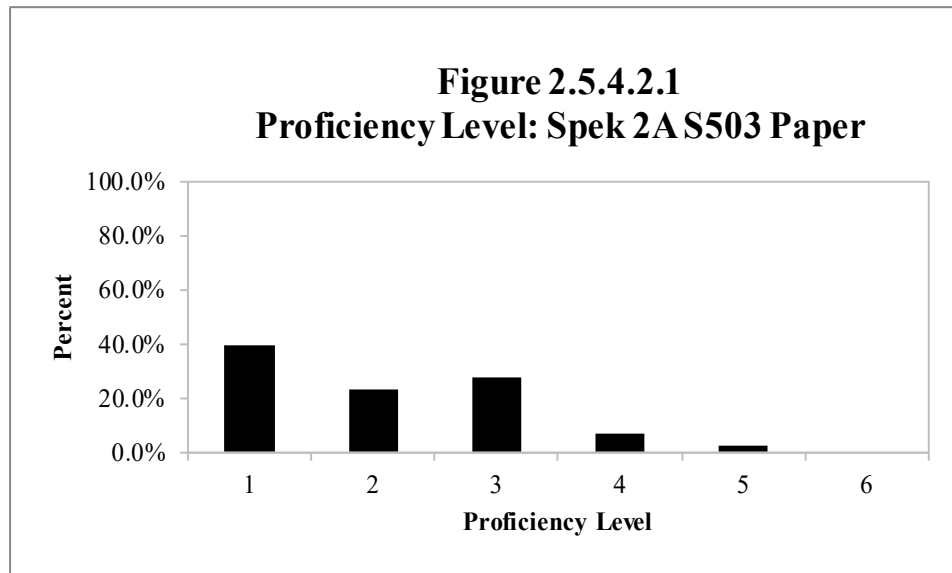


#### 2.5.4.2 Grade 2

**Table 2.5.4.2.1**

Proficiency Level Distribution: Spek 2 A S503 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	3,653	39.39%	3,653	39.39%
2	2,169	23.39%	2,169	23.39%
3	2,564	27.64%	2,564	27.64%
4	662	7.14%	662	7.14%
5	227	2.45%	227	2.45%
6	0	0.00%	0	0.00%
Total	9,275	100.00%	9,275	100.00%



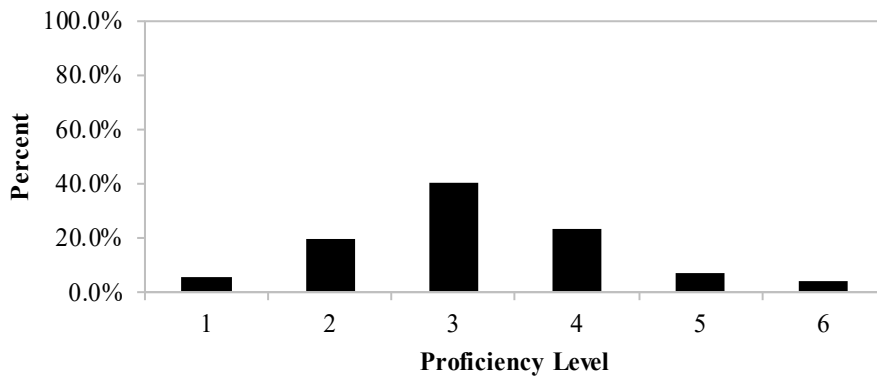


**Table 2.5.4.2.2**

Proficiency Level Distribution: Spek 2 B/C S503 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	1,320	5.75%	1,320	5.75%
2	4,541	19.77%	4,541	19.77%
3	9,245	40.25%	9,245	40.25%
4	5,280	22.99%	5,280	22.99%
5	1,619	7.05%	1,619	7.05%
6	964	4.20%	964	4.20%
Total	22,969	100.00%	22,969	100.00%

**Figure 2.5.4.2.2**  
**Proficiency Level: Spek 2B/C S503 Paper**

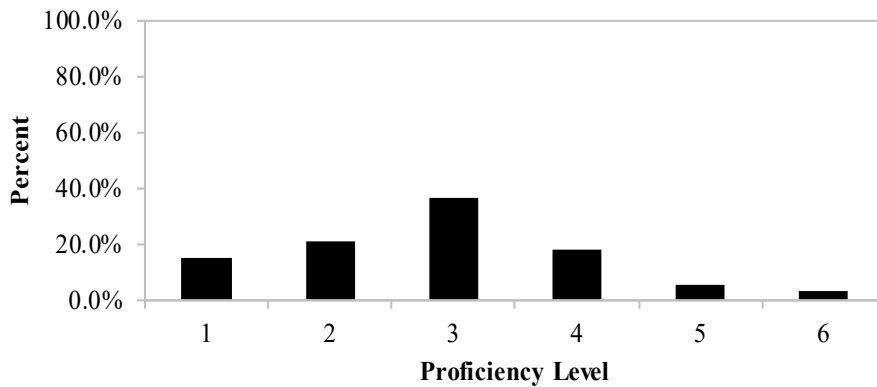


**Table 2.5.4.2.3**

Proficiency Level Distribution: Spek 2 S503 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	4,973	15.42%	4,973	15.42%
2	6,710	20.81%	6,710	20.81%
3	11,809	36.62%	11,809	36.62%
4	5,942	18.43%	5,942	18.43%
5	1,846	5.73%	1,846	5.73%
6	964	2.99%	964	2.99%
Total	32,244	100.00%	32,244	100.00%

**Figure 2.5.4.2.3**  
**Proficiency Level: Spek 2 S503 Paper**

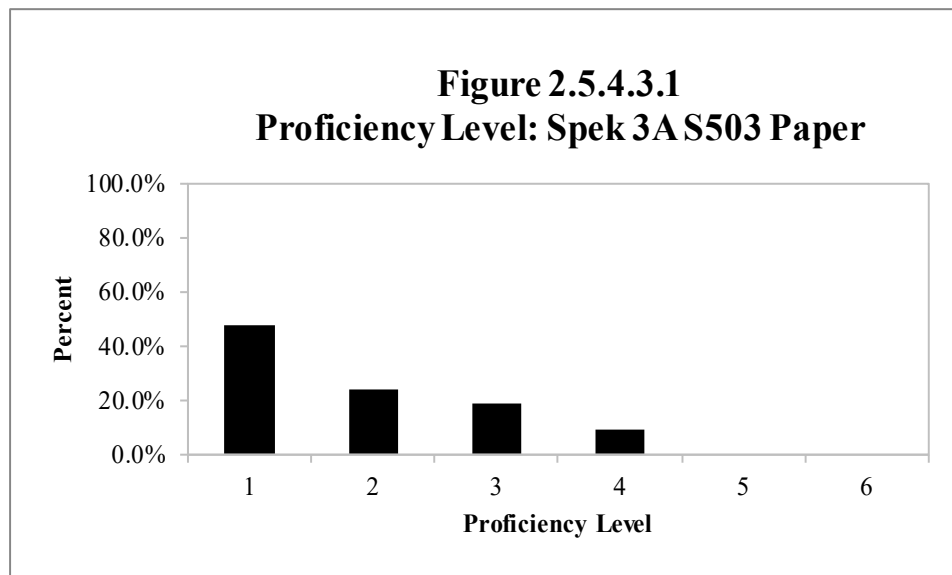


### 2.5.4.3 Grade 3

**Table 2.5.4.3.1**

Proficiency Level Distribution: Spek 3 A S503 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	3,309	47.78%	3,309	47.78%
2	1,663	24.01%	1,663	24.01%
3	1,311	18.93%	1,311	18.93%
4	642	9.27%	642	9.27%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	6,925	100.00%	6,925	100.00%

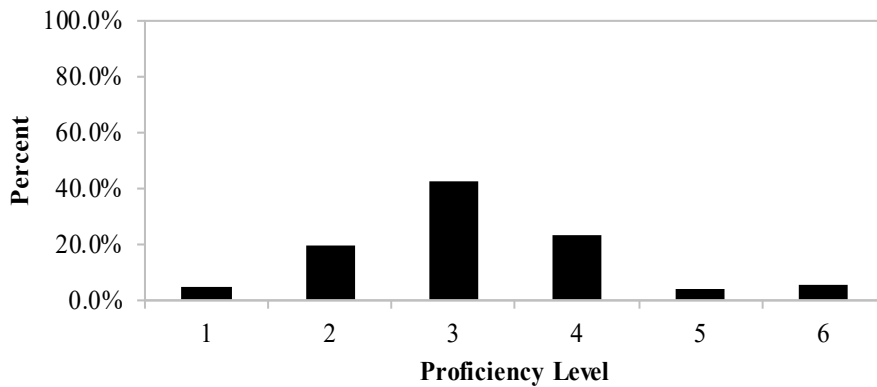


**Table 2.5.4.3.2**

Proficiency Level Distribution: Spek 3 B/C S503 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	1,083	4.46%	1,083	4.46%
2	4,722	19.43%	4,722	19.43%
3	10,401	42.79%	10,401	42.79%
4	5,687	23.40%	5,687	23.40%
5	1,031	4.24%	1,031	4.24%
6	1,384	5.69%	1,384	5.69%
Total	24,308	100.00%	24,308	100.00%

**Figure 2.5.4.3.2**  
**Proficiency Level: Spek 3B/C S503 Paper**

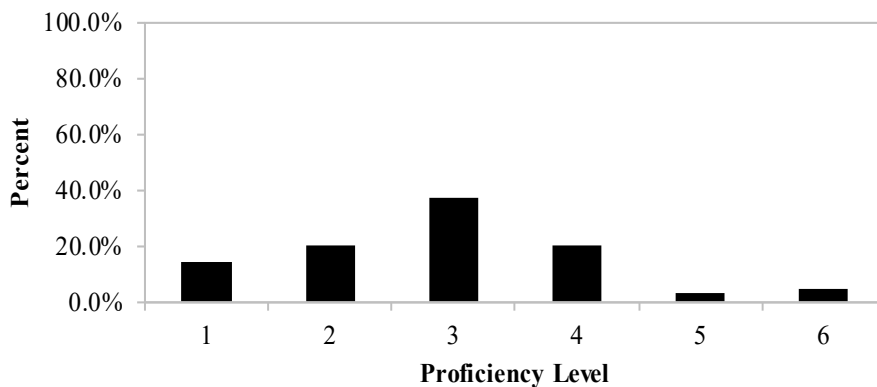


**Table 2.5.4.3.3**

Proficiency Level Distribution: Spek 3 S503 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	4,392	14.06%	4,392	14.06%
2	6,385	20.44%	6,385	20.44%
3	11,712	37.50%	11,712	37.50%
4	6,329	20.26%	6,329	20.26%
5	1,031	3.30%	1,031	3.30%
6	1,384	4.43%	1,384	4.43%
Total	31,233	100.00%	31,233	100.00%

**Figure 2.5.4.3.3**  
**Proficiency Level: Spek 3 S503 Paper**

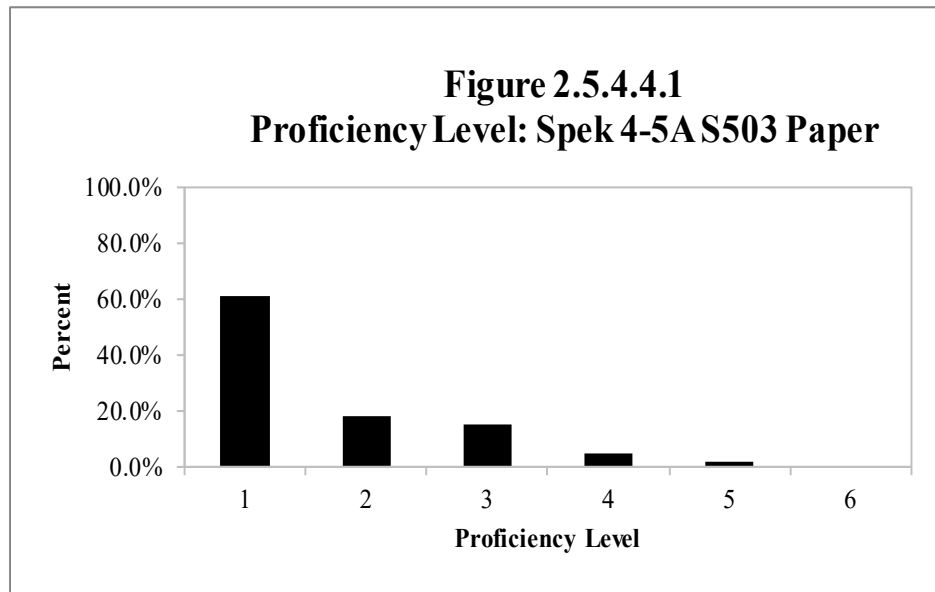


#### 2.5.4.4 Grades 4-5

**Table 2.5.4.4.1**

Proficiency Level Distribution: Spek 4-5 A S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
<b>1</b>	2,829	56.98%	3,082	65.05%	5,911	60.92%
<b>2</b>	941	18.95%	816	17.22%	1,757	18.11%
<b>3</b>	898	18.09%	552	11.65%	1,450	14.94%
<b>4</b>	225	4.53%	209	4.41%	434	4.47%
<b>5</b>	72	1.45%	79	1.67%	151	1.56%
<b>6</b>	0	0.00%	0	0.00%	0	0.00%
<b>Total</b>	4,965	100.00%	4,738	100.00%	9,703	100.00%

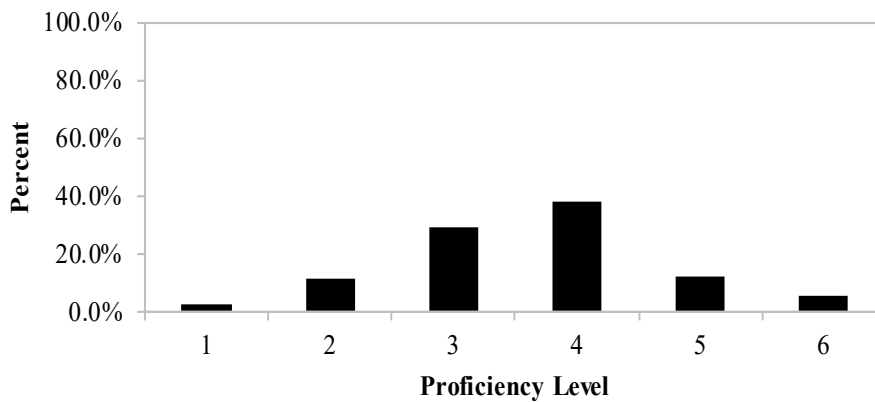


**Table 2.5.4.4.2**

Proficiency Level Distribution: Spek 4-5 B/C S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	423	2.28%	578	3.10%	1,001	2.69%
2	1,870	10.07%	2,423	13.00%	4,293	11.54%
3	5,605	30.19%	5,362	28.76%	10,967	29.48%
4	7,140	38.46%	7,069	37.92%	14,209	38.19%
5	2,667	14.37%	2,002	10.74%	4,669	12.55%
6	859	4.63%	1,209	6.49%	2,068	5.56%
Total	18,564	100.00%	18,643	100.00%	37,207	100.00%

**Figure 2.5.4.4.2**  
**Proficiency Level: Spek 4-5B/C S503 Paper**

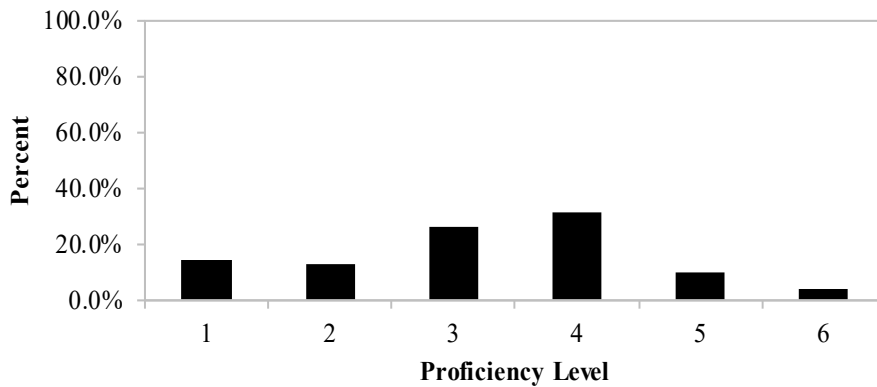


**Table 2.5.4.4.3**

Proficiency Level Distribution: Spek 4-5 S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	3,252	13.82%	3,660	15.65%	6,912	14.73%
2	2,811	11.95%	3,239	13.85%	6,050	12.90%
3	6,503	27.64%	5,914	25.29%	12,417	26.47%
4	7,365	31.30%	7,278	31.13%	14,643	31.22%
5	2,739	11.64%	2,081	8.90%	4,820	10.27%
6	859	3.65%	1,209	5.17%	2,068	4.41%
Total	23,529	100.00%	23,381	100.00%	46,910	100.00%

**Figure 2.5.4.4.3**  
**Proficiency Level: Spek 4-5 S503 Paper**



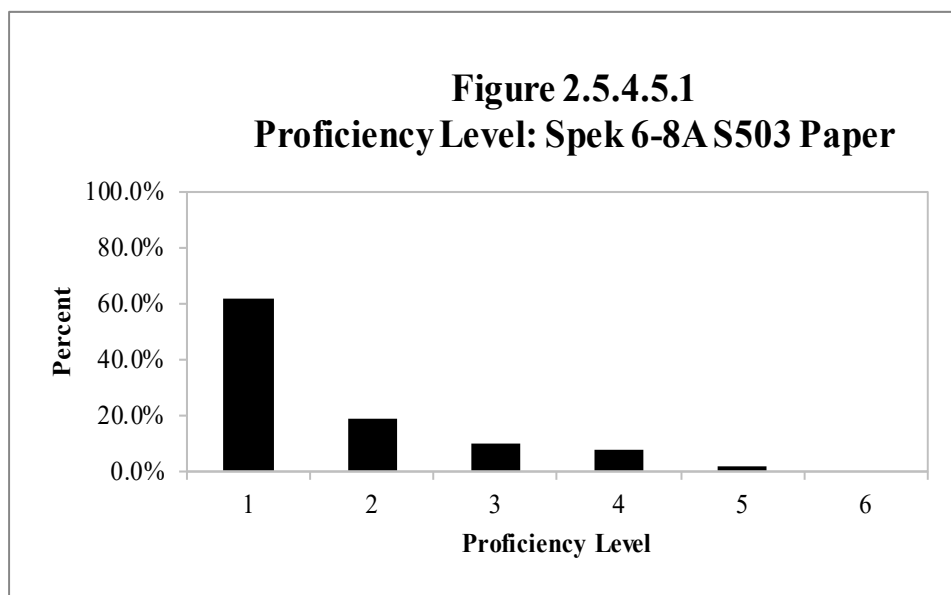


#### 2.5.4.5 Grades 6-8

**Table 2.5.4.5.1**

Proficiency Level Distribution: Spek 6-8 A S503 Paper

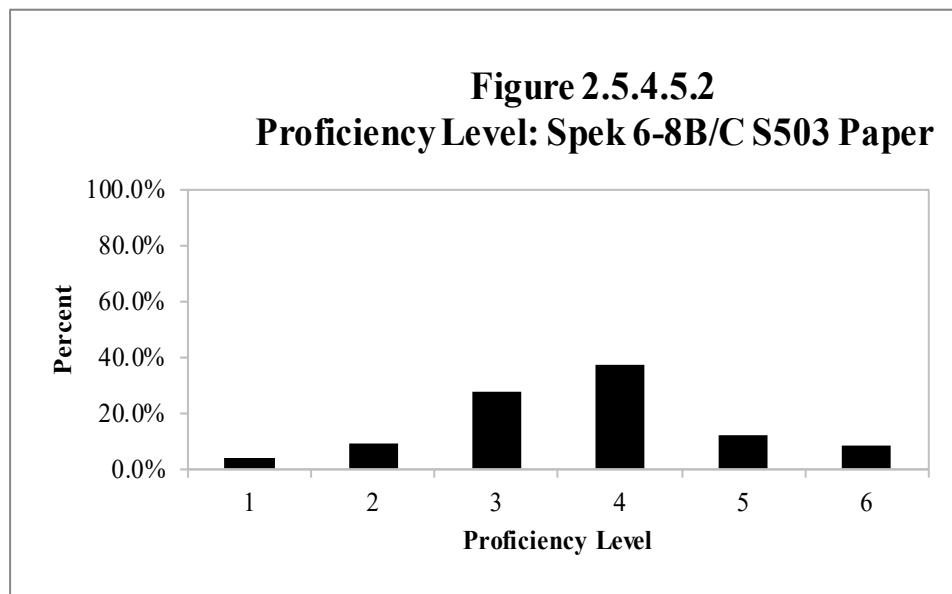
Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,594	57.26%	2,876	65.30%	2,744	62.45%	8,214	61.63%
2	1,075	23.73%	702	15.94%	727	16.55%	2,504	18.79%
3	455	10.04%	441	10.01%	453	10.31%	1,349	10.12%
4	283	6.25%	337	7.65%	402	9.15%	1,022	7.67%
5	123	2.72%	48	1.09%	68	1.55%	239	1.79%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
<b>Total</b>	4,530	100.00%	4,404	100.00%	4,394	100.00%	13,328	100.00%



**Table 2.5.4.5.2**

Proficiency Level Distribution: Spek 6-8 B/C S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	378	3.14%	542	4.94%	555	5.35%	1,475	4.42%
2	1,090	9.05%	1,213	11.06%	827	7.97%	3,130	9.37%
3	3,029	25.15%	2,915	26.58%	3,321	31.99%	9,265	27.75%
4	5,105	42.39%	3,830	34.92%	3,578	34.46%	12,513	37.47%
5	1,186	9.85%	1,758	16.03%	1,191	11.47%	4,135	12.38%
6	1,256	10.43%	709	6.46%	910	8.77%	2,875	8.61%
Total	12,044	100.00%	10,967	100.00%	10,382	100.00%	33,393	100.00%

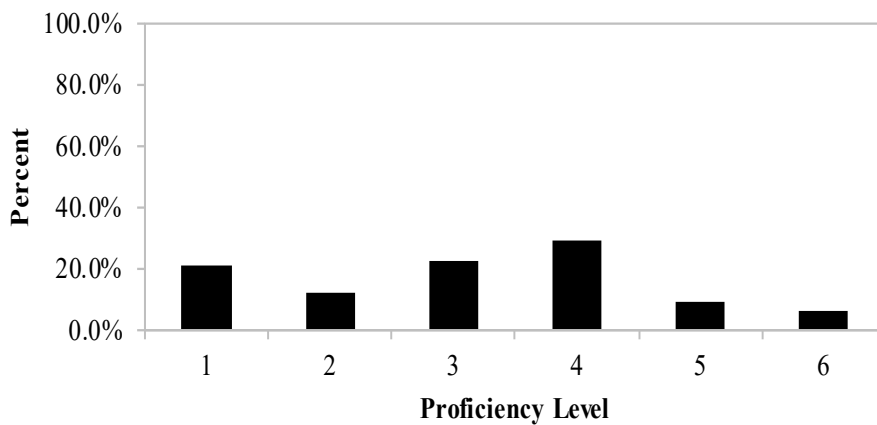


**Table 2.5.4.5.3**

Proficiency Level Distribution: Spek 6-8 S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,972	17.93%	3,418	22.24%	3,299	22.33%	9,689	20.74%
2	2,165	13.06%	1,915	12.46%	1,554	10.52%	5,634	12.06%
3	3,484	21.02%	3,356	21.83%	3,774	25.54%	10,614	22.72%
4	5,388	32.51%	4,167	27.11%	3,980	26.94%	13,535	28.97%
5	1,309	7.90%	1,806	11.75%	1,259	8.52%	4,374	9.36%
6	1,256	7.58%	709	4.61%	910	6.16%	2,875	6.15%
Total	16,574	100.00%	15,371	100.00%	14,776	100.00%	46,721	100.00%

**Figure 2.5.4.5.3**  
**Proficiency Level: Spek 6-8 S503 Paper**

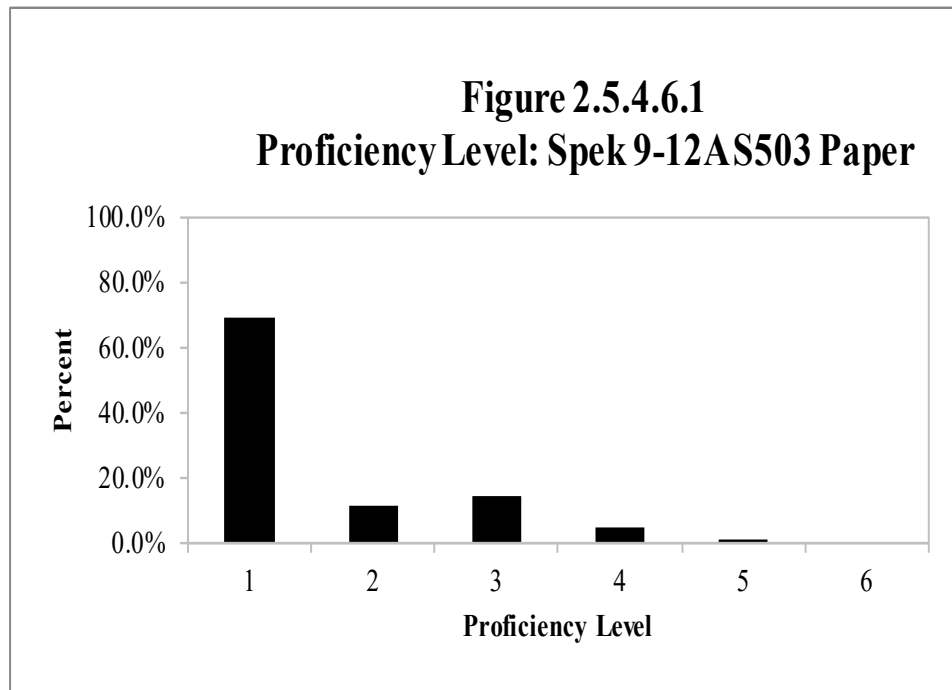


### 2.5.4.6 Grades 9-12

**Table 2.5.4.6.1**

Proficiency Level Distribution: Spek 9-12 A S503 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	3,820	75.36%	2,486	70.42%	1,909	64.25%	804	56.11%	9,019	69.36%
2	441	8.70%	325	9.21%	365	12.29%	319	22.26%	1,450	11.15%
3	591	11.66%	510	14.45%	507	17.06%	234	16.33%	1,842	14.17%
4	134	2.64%	209	5.92%	190	6.40%	76	5.30%	609	4.68%
5	83	1.64%	0	0.00%	0	0.00%	0	0.00%	83	0.64%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
<b>Total</b>	5,069	100.00%	3,530	100.00%	2,971	100.00%	1,433	100.00%	13,003	100.00%

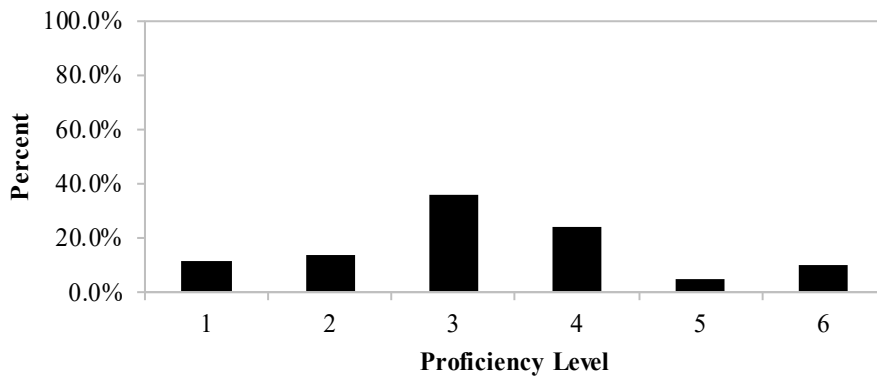


**Table 2.5.4.6.2**

Proficiency Level Distribution: Spek 9-12 B/C S503 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	890	10.40%	923	10.78%	979	12.64%	657	11.55%	3,449	11.29%
2	1,116	13.04%	1,097	12.81%	1,141	14.73%	792	13.92%	4,146	13.57%
3	2,728	31.87%	3,299	38.51%	2,647	34.17%	2,271	39.93%	10,945	35.81%
4	2,584	30.18%	1,970	23.00%	1,730	22.33%	1,016	17.86%	7,300	23.89%
5	406	4.74%	442	5.16%	422	5.45%	302	5.31%	1,572	5.14%
6	837	9.78%	835	9.75%	827	10.68%	650	11.43%	3,149	10.30%
<b>Total</b>	8,561	100.00%	8,566	100.00%	7,746	100.00%	5,688	100.00%	30,561	100.00%

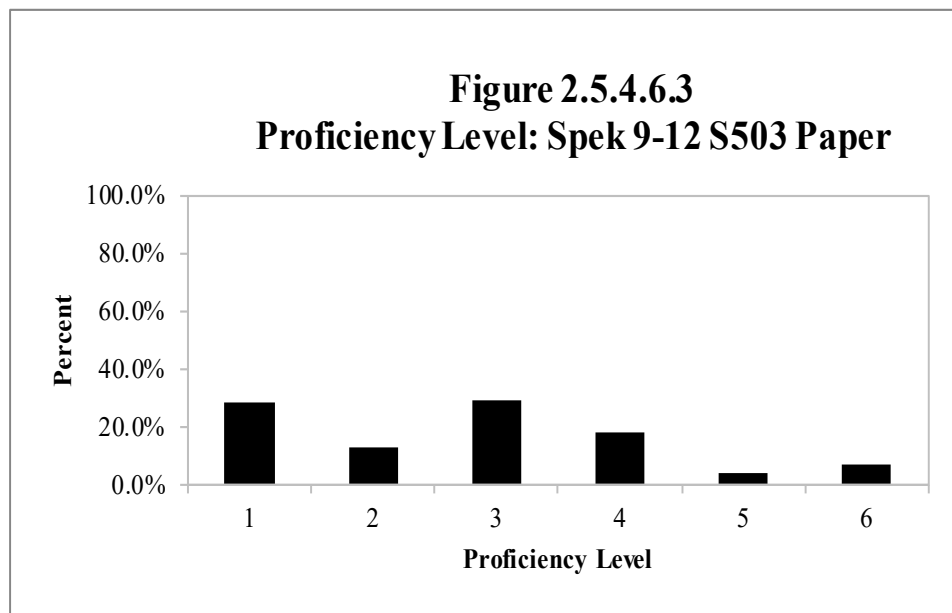
**Figure 2.5.4.6.2**  
**Proficiency Level: Spek 9-12B/C S503 Paper**



**Table 2.5.4.6.3**

Proficiency Level Distribution: Spek 9-12 S503 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
<b>1</b>	4,710	34.56%	3,409	28.18%	2,888	26.95%	1,461	20.52%	12,468	28.62%
<b>2</b>	1,557	11.42%	1,422	11.76%	1,506	14.05%	1,111	15.60%	5,596	12.85%
<b>3</b>	3,319	24.35%	3,809	31.49%	3,154	29.43%	2,505	35.18%	12,787	29.35%
<b>4</b>	2,718	19.94%	2,179	18.01%	1,920	17.92%	1,092	15.33%	7,909	18.15%
<b>5</b>	489	3.59%	442	3.65%	422	3.94%	302	4.24%	1,655	3.80%
<b>6</b>	837	6.14%	835	6.90%	827	7.72%	650	9.13%	3,149	7.23%
<b>Total</b>	13,630	100.00%	12,096	100.00%	10,717	100.00%	7,121	100.00%	43,564	100.00%



## ***2.6 Raw Score to Scale Score to Proficiency Level Conversion***

This section presents raw score to scale score conversions and associated proficiency levels for each test form.

The first column in the tables shows all possible raw scores. The following column shows the corresponding scale score. The next column shows the conditional standard error of measurement (CSEM) in the metric of the scale score, multiplied by 1.96. The resulting number (CSEM x 1.96) is used to construct the confidence band as reported on students' score reports. For example, if a student receives a scale score of 199 and if the CSEM multiplied by 1.96 is 45, then there is a 95% chance that the student's true scale score will be found somewhere between 154-244. For additional detail on conditional standard error of measurement, see Section 5, Reliability. Following the CSEM, columns provide the proficiency level interpretation for each grade in the grade-level cluster.

Performances that gain very few score points, and performances from students who gain all or almost all the score points, will have high CSEM values. The model does not precisely estimate these students' abilities; they may be well below or well above the range that is measured by the test, and therefore the error of measurement is large. We provide further detail on the CSEM because of its importance in interpreting student performance. Information on the CSEM can be found in Section 5.3, which provides CSEM values for proficiency level cuts.

Note that we truncate raw scores of zero where necessary so that the lowest scale score given is the scale score corresponding to a proficiency level score of 1.0.

## 2.6.1 Listening

### 2.6.1.0 Kindergarten

**Table 2.6.1.0**

Raw Score to Scale Score to Proficiency Level Conversion: List K S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for K
0	100	45	1.0
1	100	45	1.0
2	100	45	1.0
3	100	45	1.0
4	100	45	1.0
5	100	45	1.0
6	100	45	1.0
7	114	44	1.1
8	127	41	1.2
9	139	40	1.3
10	150	39	1.3
11	160	38	1.4
12	170	37	1.5
13	180	36	1.6
14	189	36	1.6
15	198	35	1.7
16	207	35	1.8
17	215	35	1.8
18	224	35	1.9
19	232	35	2.1
20	241	35	2.5
21	250	36	2.9
22	259	36	3.2
23	269	37	3.6
24	279	39	4.1
25	290	41	5.1
26	303	44	5.7
27	318	49	6.0
28	333	55	6.0
29	348	64	6.0
30	363	74	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.



### 2.6.1.1 Grade 1

**Table 2.6.1.1.1**

Raw Score to Scale Score to Proficiency Level Conversion: List 1 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	104	93	1.0
1	121	77	1.1
2	150	56	1.3
3	169	48	1.4
4	184	43	1.6
5	197	41	1.7
6	208	39	1.7
7	218	38	1.8
8	228	37	1.9
9	238	37	2.0
10	247	37	2.4
11	257	38	2.9
12	268	39	3.2
13	279	41	3.6
14	292	44	4.0
15	307	49	5.1
16	322	55	5.7
17	337	63	6.0
18	352	73	6.0

Note: The test form is shared between 1A and 2A.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.1.1.2**

Raw Score to Scale Score to Proficiency Level Conversion: List 1 B/C S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	104	129	1.0
1	147	77	1.3
2	177	57	1.5
3	197	48	1.7
4	211	43	1.8
5	224	40	1.9
6	235	38	1.9
7	245	37	2.3
8	254	36	2.7
9	263	35	3.1
10	272	35	3.4
11	280	35	3.6
12	289	35	3.9
13	298	36	4.5
14	307	37	5.1
15	317	38	5.5
16	328	41	6.0
17	341	44	6.0
18	356	49	6.0
19	371	55	6.0
20	386	64	6.0
21	401	74	6.0

Note: The test form is shared between 1B/C and 2B/C.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.1.2 Grade 2

**Table 2.6.1.2.1**

Raw Score to Scale Score to Proficiency Level Conversion: List 2 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G2
0	112	84	1.0
1	121	77	1.0
2	150	56	1.3
3	169	48	1.4
4	184	43	1.5
5	197	41	1.6
6	208	39	1.7
7	218	38	1.8
8	228	37	1.8
9	238	37	1.9
10	247	37	2.0
11	257	38	2.3
12	268	39	2.6
13	279	41	2.8
14	292	44	3.2
15	307	49	3.7
16	322	55	4.5
17	337	63	5.2
18	352	73	5.9

Note: The test form is shared between 1A and 2A.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.1.2.2**

Raw Score to Scale Score to Proficiency Level Conversion: List 2 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G2</b>
0	112	117	1.0
1	147	77	1.2
2	177	57	1.5
3	197	48	1.6
4	211	43	1.7
5	224	40	1.8
6	235	38	1.9
7	245	37	2.0
8	254	36	2.2
9	263	35	2.4
10	272	35	2.7
11	280	35	2.9
12	289	35	3.1
13	298	36	3.4
14	307	37	3.7
15	317	38	4.1
16	328	41	4.8
17	341	44	5.4
18	356	49	6.0
19	371	55	6.0
20	386	64	6.0
21	401	74	6.0

Note: The test form is shared between 1B/C and 2B/C.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.1.3 Grade 3

**Table 2.6.1.3.1**

Raw Score to Scale Score to Proficiency Level Conversion: List 3 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G3
0	112	181	1.0
1	184	79	1.4
2	216	58	1.6
3	236	49	1.8
4	251	44	1.9
5	264	41	2.0
6	275	39	2.3
7	286	37	2.6
8	295	37	2.8
9	305	36	3.1
10	314	36	3.4
11	324	37	3.7
12	334	38	4.1
13	344	40	4.7
14	356	43	5.2
15	371	47	5.8
16	386	54	6.0
17	401	63	6.0
18	416	74	6.0

Note: The test form is shared between 3A and 4-5A.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.1.3.2**

Raw Score to Scale Score to Proficiency Level Conversion: List 3 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G3</b>
0	112	215	1.0
1	198	80	1.5
2	231	60	1.7
3	253	51	1.9
4	269	46	2.1
5	283	43	2.5
6	296	41	2.8
7	307	39	3.2
8	318	38	3.5
9	328	37	3.9
10	337	36	4.3
11	347	36	4.8
12	356	36	5.2
13	365	37	5.6
14	375	37	6.0
15	385	39	6.0
16	397	40	6.0
17	409	43	6.0
18	424	48	6.0
19	439	55	6.0
20	454	63	6.0
21	469	74	6.0

Note: The test form is shared between 3B/C and 4-5B/C.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

#### 2.6.1.4 Grades 4–5

**Table 2.6.1.4.1**

Raw Score to Scale Score to Proficiency Level Conversion: List 4-5 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5
0	120	164	1.0	1.0
1	184	79	1.4	1.3
2	216	58	1.6	1.5
3	236	49	1.7	1.7
4	251	44	1.8	1.7
5	264	41	1.9	1.8
6	275	39	2.0	1.9
7	286	37	2.2	2.0
8	295	37	2.5	2.2
9	305	36	2.7	2.5
10	314	36	3.0	2.7
11	324	37	3.3	3.0
12	334	38	3.7	3.3
13	344	40	4.0	3.6
14	356	43	4.6	4.0
15	371	47	5.3	4.8
16	386	54	5.9	5.4
17	401	63	6.0	6.0
18	416	74	6.0	6.0

Note: The test form is shared between 3A and 4-5A.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.1.4.2**

Raw Score to Scale Score to Proficiency Level Conversion: List 4-5 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G4</b>	<b>PL for G5</b>
0	120	195	1.0	1.0
1	198	80	1.5	1.4
2	231	60	1.7	1.6
3	253	51	1.8	1.8
4	269	46	1.9	1.9
5	283	43	2.2	1.9
6	296	41	2.5	2.2
7	307	39	2.8	2.5
8	318	38	3.1	2.8
9	328	37	3.5	3.1
10	337	36	3.8	3.4
11	347	36	4.2	3.7
12	356	36	4.6	4.0
13	365	37	5.0	4.5
14	375	37	5.4	5.0
15	385	39	5.8	5.3
16	397	40	6.0	5.8
17	409	43	6.0	6.0
18	424	48	6.0	6.0
19	439	55	6.0	6.0
20	454	63	6.0	6.0
21	469	74	6.0	6.0

Note: The test form is shared between 3B/C and 4-5B/C.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.1.5 Grades 6–8

**Table 2.6.1.5.1**

Raw Score to Scale Score to Proficiency Level Conversion: List 6-8 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
0	132	137	1.0	1.0	1.0
1	181	78	1.3	1.3	1.2
2	212	59	1.5	1.4	1.4
3	233	51	1.6	1.6	1.5
4	249	46	1.7	1.6	1.6
5	263	43	1.8	1.7	1.7
6	276	41	1.8	1.8	1.8
7	287	40	1.9	1.9	1.8
8	298	39	2.1	1.9	1.9
9	309	38	2.3	2.1	2.0
10	319	38	2.6	2.4	2.2
11	329	39	2.9	2.7	2.5
12	340	39	3.2	3.0	2.8
13	352	41	3.6	3.4	3.1
14	364	44	4.0	3.8	3.5
15	379	48	4.7	4.3	4.0
16	394	55	5.3	5.0	4.6
17	409	63	5.9	5.5	5.2
18	424	73	6.0	6.0	5.8

Note: Score reports provided to students include the CSEM value multiplied by 1.96.



**Table 2.6.1.5.2**

Raw Score to Scale Score to Proficiency Level Conversion: List 6-8 B/C S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
0	132	245	1.0	1.0	1.0
1	225	76	1.5	1.5	1.5
2	255	56	1.7	1.7	1.6
3	273	48	1.8	1.8	1.8
4	288	43	1.9	1.9	1.8
5	300	40	2.1	1.9	1.9
6	311	38	2.4	2.2	2.0
7	320	36	2.6	2.4	2.3
8	330	35	2.9	2.7	2.5
9	338	35	3.1	2.9	2.7
10	347	34	3.4	3.2	3.0
11	355	34	3.7	3.5	3.2
12	364	35	4.0	3.8	3.5
13	373	35	4.4	4.1	3.8
14	382	36	4.8	4.5	4.2
15	391	37	5.2	4.8	4.5
16	402	39	5.6	5.3	5.0
17	413	42	6.0	5.7	5.4
18	428	47	6.0	6.0	6.0
19	443	54	6.0	6.0	6.0
20	458	63	6.0	6.0	6.0
21	473	75	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.1.6 Grades 9–12

**Table 2.6.1.6.1**

Raw Score to Scale Score to Proficiency Level Conversion: List 9-12 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	148	103	1.0	1.0	1.0	1.0
1	172	80	1.2	1.1	1.1	1.1
2	205	61	1.3	1.3	1.3	1.2
3	228	53	1.5	1.4	1.4	1.4
4	246	48	1.6	1.5	1.5	1.5
5	262	45	1.7	1.6	1.6	1.5
6	276	43	1.7	1.7	1.6	1.6
7	288	42	1.8	1.8	1.7	1.7
8	301	41	1.9	1.8	1.8	1.7
9	312	40	1.9	1.9	1.8	1.8
10	323	40	2.2	1.9	1.9	1.9
11	335	40	2.5	2.3	2.0	1.9
12	346	41	2.8	2.6	2.3	2.1
13	358	42	3.1	3.0	2.7	2.6
14	372	45	3.6	3.4	3.2	3.1
15	387	49	4.1	3.9	3.7	3.6
16	402	55	4.7	4.5	4.3	4.1
17	417	63	5.3	5.0	4.8	4.6
18	432	73	5.9	5.6	5.4	5.2

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.1.6.2**

Raw Score to Scale Score to Proficiency Level Conversion: List 9-12 B/C S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	148	269	1.0	1.0	1.0	1.0
1	248	76	1.6	1.5	1.5	1.5
2	277	56	1.7	1.7	1.6	1.6
3	295	47	1.8	1.8	1.7	1.7
4	310	42	1.9	1.9	1.8	1.8
5	321	39	2.1	1.9	1.9	1.8
6	332	37	2.4	2.2	1.9	1.9
7	341	36	2.6	2.4	2.2	1.9
8	350	35	2.9	2.7	2.5	2.3
9	359	34	3.2	3.0	2.8	2.6
10	367	34	3.4	3.2	3.1	2.9
11	376	34	3.7	3.5	3.4	3.2
12	384	35	4.0	3.8	3.6	3.5
13	393	35	4.3	4.1	3.9	3.8
14	402	36	4.7	4.5	4.3	4.1
15	411	38	5.0	4.8	4.6	4.4
16	422	40	5.5	5.2	5.0	4.8
17	434	43	6.0	5.7	5.5	5.3
18	448	47	6.0	6.0	6.0	5.8
19	462	54	6.0	6.0	6.0	6.0
20	476	61	6.0	6.0	6.0	6.0
21	490	72	6.0	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

## 2.6.2 Reading

### 2.6.2.0 Kindergarten

**Table 2.6.2.0**

Raw Score to Scale Score to Proficiency Level Conversion: Read K S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for K
0	100	33	1.0
1	100	33	1.0
2	100	33	1.0
3	100	33	1.0
4	100	33	1.0
5	100	33	1.0
6	100	33	1.0
7	100	33	1.0
8	100	33	1.0
9	100	33	1.0
10	100	33	1.0
11	109	34	1.0
12	120	33	1.1
13	132	33	1.2
14	142	32	1.2
15	152	31	1.3
16	162	30	1.4
17	171	29	1.5
18	180	29	1.5
19	188	29	1.6
20	196	29	1.6
21	205	29	1.7
22	213	29	1.8
23	222	29	1.8
24	230	30	1.9
25	240	31	1.9
26	250	32	2.5
27	260	35	3.0
28	270	38	3.5
29	280	43	4.1
30	290	49	5.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.2.1 Grade 1

**Table 2.6.2.1.1**

Raw Score to Scale Score to Proficiency Level Conversion: Read 1 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	141	105	1.0
1	180	52	1.3
2	200	38	1.4
3	213	32	1.5
4	222	29	1.6
5	230	27	1.7
6	237	25	1.7
7	243	24	1.8
8	249	23	1.8
9	255	23	1.9
10	260	22	1.9
11	265	22	2.0
12	270	22	2.2
13	275	22	2.5
14	280	22	2.7
15	285	22	2.9
16	291	23	3.2
17	296	24	3.5
18	302	25	3.8
19	309	26	4.4
20	317	28	5.1
21	326	32	5.5
22	335	36	6.0
23	344	41	6.0
24	353	48	6.0

Note: The test form is shared between 1A and 2A.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.2.1.2**

Raw Score to Scale Score to Proficiency Level Conversion: Read 1 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G1</b>
0	141	203	1.0
1	216	53	1.6
2	236	38	1.7
3	248	32	1.8
4	258	28	1.9
5	265	26	2.0
6	272	24	2.3
7	278	23	2.6
8	283	22	2.8
9	288	21	3.1
10	293	21	3.3
11	297	20	3.6
12	301	20	3.8
13	306	20	4.1
14	310	20	4.5
15	314	20	4.9
16	318	20	5.1
17	323	21	5.4
18	327	21	5.6
19	332	22	5.8
20	337	23	6.0
21	343	24	6.0
22	349	25	6.0
23	356	28	6.0
24	365	31	6.0
25	374	36	6.0
26	383	41	6.0
27	392	48	6.0

Note: The test form is shared between 1B/C and 2B/C.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.2.2 Grade 2

**Table 2.6.2.2.1**

Raw Score to Scale Score to Proficiency Level Conversion: Read 2 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G2
0	158	77	1.0
1	180	52	1.2
2	200	38	1.3
3	213	32	1.4
4	222	29	1.5
5	230	27	1.6
6	237	25	1.6
7	243	24	1.6
8	249	23	1.7
9	255	23	1.7
10	260	22	1.8
11	265	22	1.8
12	270	22	1.9
13	275	22	1.9
14	280	22	1.9
15	285	22	2.0
16	291	23	2.3
17	296	24	2.5
18	302	25	2.7
19	309	26	3.1
20	317	28	3.5
21	326	32	4.0
22	335	36	4.8
23	344	41	5.3
24	353	48	5.8

Note: The test form is shared between 1A and 2A.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.2.2.2**

Raw Score to Scale Score to Proficiency Level Conversion: Read 2 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G2</b>
0	158	149	1.0
1	216	53	1.4
2	236	38	1.6
3	248	32	1.7
4	258	28	1.8
5	265	26	1.8
6	272	24	1.9
7	278	23	1.9
8	283	22	2.0
9	288	21	2.2
10	293	21	2.4
11	297	20	2.5
12	301	20	2.7
13	306	20	2.9
14	310	20	3.1
15	314	20	3.3
16	318	20	3.5
17	323	21	3.8
18	327	21	4.0
19	332	22	4.5
20	337	23	5.0
21	343	24	5.3
22	349	25	5.6
23	356	28	6.0
24	365	31	6.0
25	374	36	6.0
26	383	41	6.0
27	392	48	6.0

Note: The test form is shared between 1B/C and 2B/C.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.



### 2.6.2.3 Grade 3

**Table 2.6.2.3.1**

Raw Score to Scale Score to Proficiency Level Conversion: Read 3 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G3
0	158	144	1.0
1	213	52	1.3
2	234	38	1.5
3	246	32	1.6
4	256	29	1.7
5	264	27	1.7
6	271	25	1.8
7	277	24	1.8
8	283	23	1.8
9	289	23	1.9
10	294	22	1.9
11	299	22	2.0
12	305	22	2.3
13	310	22	2.5
14	315	22	2.6
15	320	23	2.8
16	326	23	3.1
17	332	24	3.4
18	338	25	3.7
19	345	27	4.3
20	353	29	5.0
21	362	32	5.5
22	371	36	6.0
23	380	42	6.0
24	389	48	6.0

Note: The test form is shared between 3A and 4-5A.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.2.3.2**

Raw Score to Scale Score to Proficiency Level Conversion: Read 3 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G3</b>
0	158	438	1.0
1	271	52	1.8
2	291	38	1.9
3	303	32	2.2
4	313	28	2.6
5	320	26	2.8
6	327	24	3.2
7	332	23	3.4
8	338	22	3.7
9	342	21	4.0
10	347	21	4.5
11	352	20	5.0
12	356	20	5.2
13	360	20	5.4
14	364	20	5.6
15	369	20	5.9
16	373	20	6.0
17	377	21	6.0
18	382	21	6.0
19	387	22	6.0
20	392	23	6.0
21	397	24	6.0
22	404	25	6.0
23	411	28	6.0
24	420	31	6.0
25	429	36	6.0
26	438	42	6.0
27	447	48	6.0

Note: The test form is shared between 3B/C and 4-5B/C.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

#### 2.6.2.4 Grades 4–5

**Table 2.6.2.4.1**

Raw Score to Scale Score to Proficiency Level Conversion: Read 4-5 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5
0	175	104	1.0	1.0
1	213	52	1.3	1.2
2	234	38	1.4	1.4
3	246	32	1.5	1.5
4	256	29	1.6	1.5
5	264	27	1.6	1.6
6	271	25	1.7	1.6
7	277	24	1.7	1.7
8	283	23	1.8	1.7
9	289	23	1.8	1.8
10	294	22	1.9	1.8
11	299	22	1.9	1.8
12	305	22	1.9	1.9
13	310	22	2.1	1.9
14	315	22	2.2	1.9
15	320	23	2.4	2.1
16	326	23	2.6	2.3
17	332	24	2.8	2.5
18	338	25	3.1	2.7
19	345	27	3.5	3.0
20	353	29	3.9	3.4
21	362	32	4.8	3.8
22	371	36	5.3	4.7
23	380	42	5.8	5.3
24	389	48	6.0	5.8

Note: The test form is shared between 3A and 4-5A.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.2.4.2**

Raw Score to Scale Score to Proficiency Level Conversion: Read 4-5 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G4</b>	<b>PL for G5</b>
0	175	315	1.0	1.0
1	271	52	1.7	1.6
2	291	38	1.8	1.8
3	303	32	1.9	1.9
4	313	28	2.2	1.9
5	320	26	2.4	2.1
6	327	24	2.7	2.3
7	332	23	2.8	2.5
8	338	22	3.1	2.7
9	342	21	3.3	2.8
10	347	21	3.6	3.1
11	352	20	3.8	3.3
12	356	20	4.2	3.5
13	360	20	4.6	3.7
14	364	20	5.0	4.0
15	369	20	5.2	4.5
16	373	20	5.5	5.0
17	377	21	5.7	5.2
18	382	21	6.0	5.5
19	387	22	6.0	5.7
20	392	23	6.0	6.0
21	397	24	6.0	6.0
22	404	25	6.0	6.0
23	411	28	6.0	6.0
24	420	31	6.0	6.0
25	429	36	6.0	6.0
26	438	42	6.0	6.0
27	447	48	6.0	6.0

Note: The test form is shared between 3B/C and 4-5B/C.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.2.5 Grades 6–8

**Table 2.6.2.5.1**

Raw Score to Scale Score to Proficiency Level Conversion: Read 6-8 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
0	200	99	1.1	1.0	1.0
1	235	53	1.3	1.3	1.2
2	256	39	1.5	1.4	1.4
3	269	33	1.6	1.5	1.5
4	279	29	1.6	1.6	1.5
5	287	27	1.7	1.6	1.6
6	294	25	1.7	1.7	1.6
7	301	24	1.8	1.7	1.7
8	307	23	1.8	1.8	1.7
9	312	23	1.9	1.8	1.8
10	318	22	1.9	1.9	1.8
11	323	22	2.0	1.9	1.9
12	328	22	2.1	1.9	1.9
13	333	22	2.3	2.1	1.9
14	338	22	2.5	2.2	2.0
15	344	23	2.7	2.4	2.2
16	349	23	2.8	2.6	2.4
17	355	24	3.1	2.8	2.6
18	361	25	3.4	3.0	2.8
19	368	27	3.7	3.4	3.1
20	376	29	4.3	3.8	3.5
21	386	32	5.2	4.6	4.0
22	396	37	5.8	5.4	5.0
23	406	43	6.0	6.0	5.6
24	416	51	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.2.5.2**

Raw Score to Scale Score to Proficiency Level Conversion: Read 6-8 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G6</b>	<b>PL for G7</b>	<b>PL for G8</b>
0	200	216	1.1	1.0	1.0
1	276	52	1.6	1.6	1.5
2	296	38	1.8	1.7	1.7
3	308	32	1.8	1.8	1.8
4	317	28	1.9	1.9	1.8
5	325	26	2.0	1.9	1.9
6	331	24	2.2	2.0	1.9
7	337	23	2.4	2.2	2.0
8	342	22	2.6	2.4	2.2
9	347	21	2.8	2.5	2.3
10	352	21	2.9	2.7	2.5
11	356	20	3.1	2.8	2.6
12	360	20	3.3	3.0	2.8
13	365	20	3.6	3.2	2.9
14	369	20	3.8	3.4	3.1
15	373	20	4.0	3.6	3.3
16	377	20	4.4	3.8	3.5
17	382	21	5.0	4.2	3.8
18	386	21	5.2	4.6	4.0
19	391	22	5.5	5.1	4.5
20	396	23	5.8	5.4	5.0
21	402	24	6.0	5.7	5.4
22	408	25	6.0	6.0	5.7
23	416	28	6.0	6.0	6.0
24	424	31	6.0	6.0	6.0
25	432	35	6.0	6.0	6.0
26	440	40	6.0	6.0	6.0
27	448	45	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.2.6 Grades 9–12

**Table 2.6.2.6.1**

Raw Score to Scale Score to Proficiency Level Conversion: Read 9-12 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	233	69	1.1	1.1	1.0	1.0
1	249	53	1.3	1.2	1.2	1.1
2	270	39	1.4	1.4	1.3	1.3
3	283	33	1.5	1.5	1.4	1.4
4	293	30	1.6	1.6	1.5	1.5
5	302	28	1.7	1.6	1.6	1.5
6	310	26	1.7	1.7	1.6	1.6
7	316	25	1.8	1.7	1.7	1.6
8	323	24	1.8	1.8	1.7	1.7
9	329	24	1.9	1.8	1.8	1.8
10	334	23	1.9	1.9	1.8	1.8
11	340	23	2.0	1.9	1.9	1.8
12	345	22	2.1	2.0	1.9	1.9
13	350	22	2.3	2.1	2.0	1.9
14	356	23	2.5	2.3	2.2	2.1
15	361	23	2.6	2.5	2.3	2.2
16	366	23	2.8	2.6	2.5	2.4
17	372	24	3.0	2.8	2.7	2.5
18	379	25	3.3	3.1	2.9	2.7
19	385	26	3.6	3.4	3.1	2.9
20	393	29	4.1	3.8	3.5	3.3
21	403	32	5.1	4.6	4.1	3.8
22	413	37	5.7	5.4	5.1	4.8
23	423	43	6.0	6.0	5.7	5.5
24	433	51	6.0	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.2.6.2**

Raw Score to Scale Score to Proficiency Level Conversion: Read 9-12 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G9</b>	<b>PL for G10</b>	<b>PL for G11</b>	<b>PL for G12</b>
0	233	178	1.1	1.1	1.0	1.0
1	299	52	1.6	1.6	1.6	1.5
2	319	38	1.8	1.8	1.7	1.7
3	331	31	1.9	1.8	1.8	1.8
4	340	28	2.0	1.9	1.9	1.8
5	347	26	2.2	2.0	1.9	1.9
6	354	24	2.4	2.3	2.1	2.0
7	359	23	2.5	2.4	2.3	2.2
8	365	22	2.7	2.6	2.5	2.3
9	369	21	2.9	2.7	2.6	2.5
10	374	21	3.1	2.9	2.7	2.6
11	379	20	3.3	3.1	2.9	2.7
12	383	20	3.5	3.3	3.0	2.9
13	387	20	3.7	3.5	3.2	3.0
14	391	20	3.9	3.7	3.4	3.2
15	396	20	4.4	3.9	3.7	3.4
16	400	20	4.8	4.3	3.9	3.6
17	404	21	5.1	4.7	4.2	3.8
18	409	21	5.4	5.1	4.8	4.2
19	414	22	5.7	5.4	5.2	5.0
20	419	23	6.0	5.7	5.5	5.2
21	425	24	6.0	6.0	5.8	5.6
22	431	26	6.0	6.0	6.0	5.9
23	438	28	6.0	6.0	6.0	6.0
24	448	31	6.0	6.0	6.0	6.0
25	458	37	6.0	6.0	6.0	6.0
26	468	43	6.0	6.0	6.0	6.0
27	478	51	6.0	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.



## 2.6.3 Writing

### 2.6.3.0 Kindergarten

**Table 2.6.3.0**

Raw Score to Scale Score to Proficiency Level Conversion: Writ K S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for K</b>
0	100	107	1.0
1	100	107	1.0
2	100	107	1.0
3	100	107	1.0
4	155	60	1.4
5	177	44	1.5
6	191	37	1.6
7	202	35	1.7
8	213	34	1.8
9	223	35	1.9
10	234	37	2.0
11	246	37	2.3
12	258	39	2.6
13	271	41	3.0
14	288	48	3.4
15	305	57	3.8
16	322	65	4.1
17	339	71	4.5

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.3.1 Grade 1

**Table 2.6.3.1.1**

Raw Score to Scale Score to Proficiency Level Conversion: Writ 1 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G1	Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	111	97	1.0	34	381	25	4.9
1	148	63	1.2	35	387	26	5.2
2	177	46	1.5	36	395	28	5.5
3	193	34	1.6	37	403	31	5.9
4	202	28	1.7	38	414	37	6.0
5	209	24	1.7	39	433	51	6.0
6	214	22	1.8	40	464	95	6.0
7	219	20	1.8				
8	223	20	1.8				
9	227	19	1.9				
10	231	19	1.9				
11	234	19	1.9				
12	238	20	2.0				
13	242	20	2.1				
14	247	21	2.2				
15	252	23	2.3				
16	257	24	2.5				
17	263	25	2.6				
18	270	27	2.8				
19	277	27	3.0				
20	285	28	3.1				
21	293	27	3.2				
22	300	27	3.4				
23	307	26	3.5				
24	314	26	3.6				
25	321	26	3.7				
26	328	26	3.8				
27	334	26	3.9				
28	341	26	4.0				
29	348	26	4.2				
30	355	26	4.4				
31	362	25	4.5				
32	368	25	4.6				
33	374	25	4.8				

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.3.1.2**

Raw Score to Scale Score to Proficiency Level Conversion: Writ 1 B/C S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G1	Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	111	259	1.0	34	342	24	4.1
1	191	47	1.6	35	348	23	4.2
2	206	32	1.7	36	354	23	4.3
3	214	26	1.8	37	359	23	4.4
4	220	22	1.8	38	364	22	4.6
5	224	20	1.8	39	369	22	4.7
6	228	19	1.9	40	373	21	4.8
7	231	18	1.9	41	378	21	4.9
8	234	17	1.9	42	382	21	5.0
9	237	16	1.9	43	387	21	5.2
10	240	16	2.0	44	391	21	5.3
11	242	16	2.1	45	395	21	5.5
12	245	16	2.1	46	400	21	5.7
13	248	16	2.2	47	405	22	6.0
14	250	16	2.3	48	410	23	6.0
15	253	16	2.4	49	415	24	6.0
16	256	16	2.4	50	422	26	6.0
17	258	17	2.5	51	429	30	6.0
18	261	18	2.6	52	440	36	6.0
19	265	18	2.7	53	458	51	6.0
20	268	19	2.8	54	490	95	6.0
21	272	20	2.9				
22	276	21	3.0				
23	281	21	3.0				
24	285	22	3.1				
25	291	23	3.2				
26	296	23	3.3				
27	301	24	3.4				
28	307	24	3.5				
29	313	24	3.6				
30	319	24	3.7				
31	325	24	3.8				
32	331	24	3.9				
33	337	24	4.0				

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.3.2 Grade 2

**Table 2.6.3.2.1**

Raw Score to Scale Score to Proficiency Level Conversion: Writ 2 A S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G2</b>
0	133	213	1.0
1	199	45	1.6
2	213	32	1.7
3	222	26	1.8
4	228	24	1.8
5	234	23	1.9
6	239	22	1.9
7	244	23	2.0
8	249	24	2.1
9	255	25	2.3
10	262	27	2.5
11	270	29	2.7
12	279	31	3.0
13	290	33	3.1
14	301	34	3.3
15	313	35	3.5
16	325	34	3.7
17	336	34	3.9
18	347	33	4.1
19	358	31	4.3
20	367	30	4.5
21	376	30	4.7
22	385	29	4.9
23	394	30	5.2
24	403	32	5.6
25	415	38	6.0
26	434	51	6.0
27	465	94	6.0

Note: The test form is shared between 2A and 3A.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.3.2.2**

Raw Score to Scale Score to Proficiency Level Conversion: Writ 2 B/C S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G2	Raw Score	Scale Score	CSEM x 1.96	PL for G2
0	133	159	1.0	34	340	24	3.9
1	188	47	1.5	35	345	23	4.0
2	202	32	1.6	36	351	23	4.2
3	210	26	1.7	37	356	23	4.3
4	216	22	1.7	38	361	22	4.4
5	221	20	1.8	39	366	22	4.5
6	225	19	1.8	40	371	21	4.6
7	228	18	1.8	41	375	21	4.7
8	231	17	1.9	42	380	21	4.8
9	234	17	1.9	43	384	21	4.9
10	237	16	1.9	44	388	21	5.0
11	240	16	1.9	45	393	21	5.2
12	242	16	2.0	46	397	21	5.3
13	245	16	2.0	47	402	22	5.6
14	248	16	2.1	48	407	23	5.8
15	250	16	2.2	49	413	24	6.0
16	253	17	2.2	50	420	27	6.0
17	256	17	2.3	51	428	30	6.0
18	259	18	2.4	52	438	36	6.0
19	263	18	2.5	53	456	51	6.0
20	266	19	2.6	54	488	95	6.0
21	270	20	2.7				
22	274	20	2.8				
23	278	21	2.9				
24	283	22	3.0				
25	288	23	3.1				
26	293	23	3.2				
27	299	24	3.3				
28	305	24	3.4				
29	310	24	3.5				
30	316	24	3.5				
31	322	24	3.6				
32	328	24	3.7				
33	334	24	3.8				

Note: The test form is shared between 2B/C and 3B/C.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.3.3 Grade 3

**Table 2.6.3.3.1**

Raw Score to Scale Score to Proficiency Level Conversion: Writ 3 A S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G3</b>
0	133	213	1.0
1	199	45	1.5
2	213	32	1.7
3	222	26	1.7
4	228	24	1.8
5	234	23	1.8
6	239	22	1.9
7	244	23	1.9
8	249	24	2.0
9	255	25	2.2
10	262	27	2.4
11	270	29	2.6
12	279	31	2.8
13	290	33	3.1
14	301	34	3.2
15	313	35	3.4
16	325	34	3.6
17	336	34	3.8
18	347	33	4.0
19	358	31	4.2
20	367	30	4.4
21	376	30	4.6
22	385	29	4.8
23	394	30	5.0
24	403	32	5.3
25	415	38	5.8
26	434	51	6.0
27	465	94	6.0

Note: The test form is shared between 2A and 3A.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.3.3.2**

Raw Score to Scale Score to Proficiency Level Conversion: Writ 3 B/C S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G3	Raw Score	Scale Score	CSEM x 1.96	PL for G3
0	133	159	1.0	34	340	24	3.9
1	188	47	1.4	35	345	23	3.9
2	202	32	1.6	36	351	23	4.1
3	210	26	1.6	37	356	23	4.2
4	216	22	1.7	38	361	22	4.3
5	221	20	1.7	39	366	22	4.4
6	225	19	1.8	40	371	21	4.5
7	228	18	1.8	41	375	21	4.6
8	231	17	1.8	42	380	21	4.7
9	234	17	1.8	43	384	21	4.7
10	237	16	1.9	44	388	21	4.8
11	240	16	1.9	45	393	21	4.9
12	242	16	1.9	46	397	21	5.1
13	245	16	1.9	47	402	22	5.3
14	248	16	2.0	48	407	23	5.5
15	250	16	2.0	49	413	24	5.7
16	253	17	2.1	50	420	27	6.0
17	256	17	2.2	51	428	30	6.0
18	259	18	2.3	52	438	36	6.0
19	263	18	2.4	53	456	51	6.0
20	266	19	2.5	54	488	95	6.0
21	270	20	2.6				
22	274	20	2.7				
23	278	21	2.8				
24	283	22	3.0				
25	288	23	3.0				
26	293	23	3.1				
27	299	24	3.2				
28	305	24	3.3				
29	310	24	3.4				
30	316	24	3.5				
31	322	24	3.6				
32	328	24	3.7				
33	334	24	3.8				

Note: The test form is shared between 2B/C and 3B/C.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

#### 2.6.3.4 Grades 4–5

**Table 2.6.3.4.1**

Raw Score to Scale Score to Proficiency Level Conversion: Writ 4-5 A S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G4</b>	<b>PL for G5</b>
0	155	253	1.0	1.0
1	231	45	1.7	1.6
2	245	32	1.8	1.8
3	253	26	1.8	1.8
4	259	24	1.9	1.9
5	264	23	1.9	1.9
6	270	22	2.1	2.1
7	275	23	2.4	2.3
8	280	23	2.6	2.5
9	286	25	2.9	2.7
10	293	27	3.0	3.0
11	301	29	3.2	3.1
12	310	31	3.3	3.2
13	320	33	3.5	3.4
14	332	34	3.6	3.6
15	343	35	3.8	3.7
16	355	34	4.0	3.9
17	367	34	4.3	4.2
18	378	33	4.5	4.4
19	388	31	4.7	4.6
20	398	30	4.9	4.8
21	407	29	5.2	5.0
22	416	29	5.6	5.3
23	424	30	5.9	5.6
24	434	32	6.0	6.0
25	446	37	6.0	6.0
26	464	51	6.0	6.0
27	496	94	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.



**Table 2.6.3.4.2**

Raw Score to Scale Score to Proficiency Level Conversion: Writ 4-5 B/C S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5	Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5
0	155	229	1.0	1.0	34	381	24	4.6	4.4
1	228	47	1.6	1.6	35	387	23	4.7	4.6
2	243	32	1.8	1.7	36	392	23	4.8	4.7
3	251	26	1.8	1.8	37	397	22	4.9	4.8
4	257	22	1.9	1.9	38	403	22	5.0	4.9
5	261	20	1.9	1.9	39	407	22	5.2	5.0
6	265	19	1.9	1.9	40	412	21	5.4	5.1
7	269	18	2.1	2.0	41	417	21	5.6	5.3
8	272	17	2.2	2.1	42	421	21	5.8	5.5
9	275	17	2.4	2.3	43	426	21	6.0	5.7
10	278	17	2.5	2.4	44	430	21	6.0	5.8
11	281	17	2.6	2.5	45	435	21	6.0	6.0
12	284	17	2.8	2.6	46	439	22	6.0	6.0
13	286	17	2.9	2.7	47	444	22	6.0	6.0
14	289	17	3.0	2.8	48	450	23	6.0	6.0
15	292	17	3.0	2.9	49	455	25	6.0	6.0
16	295	17	3.1	3.0	50	462	27	6.0	6.0
17	298	17	3.1	3.0	51	470	30	6.0	6.0
18	302	18	3.2	3.1	52	481	36	6.0	6.0
19	305	18	3.2	3.1	53	499	51	6.0	6.0
20	308	19	3.3	3.2	54	531	95	6.0	6.0
21	312	19	3.3	3.3					
22	316	20	3.4	3.3					
23	321	21	3.5	3.4					
24	325	22	3.5	3.5					
25	330	22	3.6	3.5					
26	335	23	3.7	3.6					
27	341	23	3.8	3.7					
28	346	24	3.9	3.8					
29	352	24	4.0	3.9					
30	358	24	4.1	4.0					
31	364	24	4.2	4.1					
32	370	24	4.3	4.2					
33	376	24	4.5	4.3					

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.3.5 Grades 6–8

**Table 2.6.3.5.1**

Raw Score to Scale Score to Proficiency Level Conversion: Writ 6-8 A S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G6</b>	<b>PL for G7</b>	<b>PL for G8</b>
0	188	103	1.2	1.1	1.0
1	220	45	1.5	1.4	1.3
2	234	32	1.6	1.5	1.4
3	243	27	1.7	1.6	1.5
4	249	24	1.8	1.7	1.6
5	255	23	1.8	1.8	1.7
6	260	23	1.9	1.8	1.7
7	266	23	1.9	1.9	1.8
8	271	24	2.1	1.9	1.8
9	277	25	2.3	2.1	1.9
10	284	27	2.5	2.3	2.1
11	292	29	2.8	2.5	2.3
12	301	31	3.0	2.8	2.6
13	312	33	3.2	3.1	3.0
14	323	34	3.3	3.2	3.1
15	334	34	3.5	3.4	3.3
16	346	34	3.7	3.6	3.5
17	358	34	3.9	3.8	3.7
18	369	33	4.1	4.0	3.9
19	379	31	4.3	4.2	4.1
20	389	30	4.5	4.4	4.3
21	398	30	4.7	4.5	4.5
22	407	30	4.8	4.7	4.6
23	416	30	5.1	4.9	4.8
24	425	33	5.4	5.1	5.0
25	438	38	5.8	5.6	5.4
26	457	52	6.0	6.0	5.9
27	488	94	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.3.5.2**

Raw Score to Scale Score to Proficiency Level Conversion: Writ 6-8 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G6</b>	<b>PL for G7</b>	<b>PL for G8</b>
0	188	96	1.2	1.1	1.0
1	218	47	1.5	1.4	1.3
2	233	32	1.6	1.5	1.4
3	241	26	1.7	1.6	1.5
4	247	22	1.7	1.7	1.6
5	251	20	1.8	1.7	1.6
6	255	18	1.8	1.8	1.7
7	258	17	1.9	1.8	1.7
8	261	17	1.9	1.8	1.7
9	264	16	1.9	1.9	1.8
10	267	16	1.9	1.9	1.8
11	269	16	2.0	1.9	1.8
12	272	16	2.1	1.9	1.9
13	274	16	2.2	2.0	1.9
14	277	16	2.3	2.1	1.9
15	279	16	2.3	2.1	1.9
16	282	16	2.4	2.2	2.0
17	285	17	2.5	2.3	2.1
18	288	17	2.6	2.4	2.2
19	291	18	2.7	2.5	2.3
20	295	19	2.9	2.6	2.4
21	299	20	3.0	2.8	2.6
22	303	21	3.0	2.9	2.7
23	307	21	3.1	3.0	2.8
24	312	22	3.2	3.1	3.0
25	317	23	3.3	3.1	3.0
26	323	23	3.3	3.2	3.1
27	328	24	3.4	3.3	3.2
28	334	24	3.5	3.4	3.3
29	340	24	3.6	3.5	3.4
30	346	24	3.7	3.6	3.5
31	352	24	3.8	3.7	3.6
32	358	24	3.9	3.8	3.7
33	363	24	4.0	3.9	3.8

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G6</b>	<b>PL for G7</b>	<b>PL for G8</b>
34	369	24	4.1	4.0	3.9
35	375	23	4.2	4.1	4.0
36	380	23	4.3	4.2	4.1
37	386	23	4.4	4.3	4.2
38	391	22	4.5	4.4	4.3
39	395	22	4.6	4.5	4.4
40	400	21	4.7	4.6	4.5
41	405	21	4.8	4.7	4.6
42	409	21	4.9	4.8	4.7
43	413	21	5.0	4.8	4.7
44	418	21	5.1	4.9	4.8
45	422	21	5.3	5.0	4.9
46	427	21	5.5	5.2	5.0
47	431	22	5.6	5.3	5.2
48	436	23	5.8	5.5	5.3
49	442	24	6.0	5.7	5.5
50	448	26	6.0	5.9	5.6
51	456	30	6.0	6.0	5.9
52	467	36	6.0	6.0	6.0
53	485	51	6.0	6.0	6.0
54	516	95	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.3.6 Grades 9–12

**Table 2.6.3.6.1**

Raw Score to Scale Score to Proficiency Level Conversion: Writ 9-12 A S503 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	232	77	1.3	1.2	1.1	1.0
1	252	45	1.5	1.4	1.3	1.2
2	266	32	1.7	1.6	1.5	1.3
3	275	27	1.8	1.7	1.6	1.5
4	282	25	1.9	1.8	1.7	1.5
5	288	24	1.9	1.8	1.7	1.6
6	293	23	2.1	1.9	1.8	1.7
7	299	23	2.3	2.0	1.8	1.7
8	304	24	2.5	2.2	1.9	1.8
9	310	25	2.7	2.4	2.0	1.9
10	317	27	2.9	2.6	2.3	1.9
11	325	29	3.1	2.9	2.6	2.2
12	334	31	3.2	3.1	2.9	2.6
13	344	33	3.4	3.3	3.1	3.0
14	355	34	3.6	3.4	3.3	3.2
15	367	34	3.8	3.6	3.5	3.4
16	379	34	4.0	3.8	3.7	3.6
17	390	34	4.2	4.0	3.9	3.8
18	402	33	4.4	4.3	4.2	4.0
19	412	31	4.6	4.5	4.4	4.2
20	421	30	4.8	4.7	4.6	4.4
21	430	30	5.0	4.8	4.7	4.6
22	439	30	5.2	5.0	4.9	4.8
23	448	31	5.4	5.2	5.1	5.0
24	458	33	5.7	5.5	5.3	5.2
25	471	38	6.0	5.8	5.6	5.4
26	489	52	6.0	6.0	5.9	5.7
27	521	94	6.0	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.3.6.2**

Raw Score to Scale Score to Proficiency Level Conversion: Writ 9-12 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G9</b>	<b>PL for G10</b>	<b>PL for G11</b>	<b>PL for G12</b>
0	232	49	1.3	1.2	1.1	1.0
1	233	47	1.3	1.2	1.1	1.0
2	248	32	1.5	1.4	1.3	1.1
3	256	26	1.6	1.5	1.4	1.2
4	262	22	1.7	1.5	1.4	1.3
5	267	20	1.7	1.6	1.5	1.4
6	270	18	1.7	1.6	1.5	1.4
7	274	17	1.8	1.7	1.6	1.4
8	276	17	1.8	1.7	1.6	1.5
9	279	16	1.8	1.7	1.6	1.5
10	282	16	1.9	1.8	1.7	1.5
11	284	16	1.9	1.8	1.7	1.6
12	287	16	1.9	1.8	1.7	1.6
13	289	16	2.0	1.8	1.7	1.6
14	292	16	2.1	1.9	1.8	1.6
15	295	16	2.2	1.9	1.8	1.7
16	297	16	2.2	1.9	1.8	1.7
17	300	17	2.3	2.0	1.9	1.7
18	303	17	2.4	2.1	1.9	1.8
19	306	18	2.5	2.2	1.9	1.8
20	310	19	2.7	2.4	2.0	1.9
21	314	20	2.8	2.5	2.2	1.9
22	318	21	2.9	2.7	2.3	2.0
23	322	21	3.0	2.8	2.5	2.1
24	327	22	3.1	3.0	2.7	2.3
25	332	23	3.2	3.1	2.8	2.5
26	338	23	3.3	3.2	3.0	2.7
27	343	24	3.4	3.2	3.1	2.9
28	349	24	3.5	3.3	3.2	3.0
29	355	24	3.6	3.4	3.3	3.2
30	361	24	3.7	3.5	3.4	3.3
31	367	24	3.8	3.6	3.5	3.4
32	373	24	3.9	3.7	3.6	3.5
33	379	24	4.0	3.8	3.7	3.6

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G9</b>	<b>PL for G10</b>	<b>PL for G11</b>	<b>PL for G12</b>
34	384	24	4.1	3.9	3.8	3.7
35	390	23	4.2	4.0	3.9	3.8
36	396	23	4.3	4.2	4.1	3.9
37	401	23	4.4	4.3	4.2	4.0
38	406	22	4.5	4.4	4.3	4.1
39	411	22	4.6	4.5	4.4	4.2
40	415	21	4.7	4.5	4.4	4.3
41	420	21	4.8	4.6	4.5	4.4
42	424	21	4.8	4.7	4.6	4.5
43	429	21	4.9	4.8	4.7	4.6
44	433	21	5.0	4.9	4.8	4.7
45	437	21	5.1	5.0	4.9	4.7
46	442	21	5.3	5.1	5.0	4.8
47	446	22	5.4	5.2	5.1	4.9
48	451	23	5.5	5.3	5.2	5.0
49	457	24	5.6	5.4	5.3	5.1
50	463	26	5.8	5.6	5.4	5.2
51	471	30	6.0	5.8	5.6	5.4
52	482	36	6.0	6.0	5.8	5.6
53	500	51	6.0	6.0	6.0	5.9
54	531	95	6.0	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

## 2.6.4 Speaking

### 2.6.4.0 Kindergarten

**Table 2.6.4.0**

Raw Score to Scale Score to Proficiency Level Conversion: Spek K S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for K</b>
0	100	183	1.0
1	123	133	1.2
2	147	83	1.5
3	169	63	1.7
4	191	55	2.0
5	211	52	2.3
6	230	48	2.6
7	250	41	3.0
8	301	32	4.0
9	349	44	5.0
10	392	105	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.



#### 2.6.4.1 Grade 1

**Table 2.6.4.1.1**

Raw Score to Scale Score to Proficiency Level Conversion: Spek 1 A S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G1</b>
0	106	65	1.0
1	117	52	1.1
2	135	39	1.2
3	147	35	1.4
4	157	33	1.5
5	167	33	1.6
6	177	34	1.7
7	188	35	1.8
8	200	36	1.9
9	212	39	2.1
10	227	42	2.3
11	245	48	2.7
12	270	54	3.1
13	296	52	3.7
14	318	48	4.1
15	338	47	4.5
16	359	50	4.9
17	380	59	5.4
18	401	75	5.9

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.4.1.2**

Raw Score to Scale Score to Proficiency Level Conversion: Spek 1 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G1</b>
6	106	60	1.0
7	175	29	1.6
8	183	28	1.7
9	190	28	1.8
10	198	28	1.9
11	205	28	2.0
12	213	29	2.1
13	220	29	2.2
14	229	31	2.4
15	238	32	2.5
16	248	34	2.7
17	259	37	2.9
18	272	38	3.2
19	286	38	3.5
20	299	37	3.7
21	310	35	3.9
22	321	34	4.2
23	331	33	4.4
24	341	33	4.6
25	351	34	4.8
26	362	35	5.0
27	375	39	5.3
28	388	44	5.6
29	401	51	5.9
30	414	60	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

#### 2.6.4.2 Grade 2

**Table 2.6.4.2.1**

Raw Score to Scale Score to Proficiency Level Conversion: Spek 2 A S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G2</b>
0	118	38	1.0
1	118	38	1.0
2	118	38	1.0
3	118	38	1.0
4	130	37	1.1
5	143	37	1.2
6	155	36	1.3
7	167	36	1.5
8	179	37	1.6
9	192	38	1.7
10	206	42	1.8
11	224	48	2.0
12	248	54	2.5
13	274	52	3.0
14	296	48	3.4
15	317	47	3.8
16	338	50	4.3
17	359	60	4.7
18	380	76	5.1

Note: The test form is shared between 2A and 3A.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.4.2.2**

Raw Score to Scale Score to Proficiency Level Conversion: Spek 2 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G2</b>
6	118	36	1.0
7	157	32	1.4
8	166	31	1.5
9	175	30	1.5
10	183	30	1.6
11	192	31	1.7
12	200	31	1.8
13	209	32	1.8
14	219	33	1.9
15	229	34	2.1
16	240	35	2.3
17	252	36	2.6
18	263	36	2.8
19	275	36	3.0
20	287	35	3.2
21	298	35	3.5
22	309	34	3.7
23	320	34	3.9
24	331	35	4.1
25	342	36	4.3
26	354	38	4.6
27	368	41	4.8
28	382	46	5.1
29	396	52	5.5
30	425	74	6.0

Note: The test form is shared between 2B/C and 3B/C.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

### 2.6.4.3 Grade 3

**Table 2.6.4.3.1**

Raw Score to Scale Score to Proficiency Level Conversion: Spek 3 A S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G3</b>
0	118	38	1.0
1	118	38	1.0
2	118	38	1.0
3	118	38	1.0
4	130	37	1.1
5	143	37	1.2
6	155	36	1.3
7	167	36	1.4
8	179	37	1.5
9	192	38	1.6
10	206	42	1.7
11	224	48	1.9
12	248	54	2.2
13	274	52	2.8
14	296	48	3.2
15	317	47	3.6
16	338	50	4.1
17	359	60	4.5
18	380	76	4.8

Note: The test form is shared between 2A and 3A.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.4.3.2**

Raw Score to Scale Score to Proficiency Level Conversion: Spek 3 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G3</b>
6	118	36	1.0
7	157	32	1.3
8	166	31	1.4
9	175	30	1.4
10	183	30	1.5
11	192	31	1.6
12	200	31	1.7
13	209	32	1.7
14	219	33	1.8
15	229	34	1.9
16	240	35	2.1
17	252	36	2.3
18	263	36	2.5
19	275	36	2.8
20	287	35	3.0
21	298	35	3.3
22	309	34	3.5
23	320	34	3.7
24	331	35	3.9
25	342	36	4.1
26	354	38	4.4
27	368	41	4.6
28	382	46	4.9
29	396	52	5.2
30	425	74	6.0

Note: The test form is shared between 2B/C and 3B/C.

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

#### 2.6.4.4 Grades 4–5

**Table 2.6.4.4.1**

Raw Score to Scale Score to Proficiency Level Conversion: Spek 4-5 A S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G4</b>	<b>PL for G5</b>
0	130	48	1.0	1.0
1	130	48	1.0	1.0
2	142	40	1.1	1.0
3	155	36	1.2	1.1
4	166	34	1.3	1.2
5	176	34	1.4	1.3
6	187	34	1.5	1.4
7	198	34	1.6	1.5
8	209	35	1.6	1.6
9	221	37	1.7	1.7
10	234	40	1.9	1.8
11	250	44	2.0	1.9
12	270	49	2.5	2.2
13	293	51	3.0	2.7
14	316	50	3.4	3.2
15	339	50	3.9	3.7
16	362	53	4.3	4.2
17	385	61	4.7	4.6
18	408	77	5.2	5.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.4.4.2**

Raw Score to Scale Score to Proficiency Level Conversion: Spek 4-5 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G4</b>	<b>PL for G5</b>
6	130	47	1.0	1.0
7	192	32	1.5	1.4
8	201	31	1.6	1.5
9	210	31	1.7	1.6
10	219	31	1.7	1.6
11	228	31	1.8	1.7
12	236	31	1.9	1.8
13	245	31	1.9	1.8
14	254	31	2.1	1.9
15	263	32	2.3	2.1
16	272	33	2.5	2.3
17	282	34	2.7	2.5
18	293	35	3.0	2.7
19	305	36	3.2	3.0
20	317	37	3.4	3.3
21	330	37	3.7	3.5
22	342	36	4.0	3.8
23	354	36	4.2	4.0
24	366	36	4.4	4.2
25	377	36	4.6	4.4
26	390	37	4.8	4.7
27	403	40	5.1	4.9
28	416	45	5.5	5.2
29	429	51	5.8	5.6
30	443	60	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.



#### 2.6.4.5 Grades 6–8

**Table 2.6.4.5.1**

Raw Score to Scale Score to Proficiency Level Conversion: Spek 6-8 A S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G6</b>	<b>PL for G7</b>	<b>PL for G8</b>
0	148	51	1.0	1.0	1.0
1	148	51	1.0	1.0	1.0
2	165	42	1.2	1.1	1.1
3	179	39	1.3	1.2	1.2
4	193	39	1.4	1.3	1.3
5	207	38	1.5	1.4	1.4
6	219	36	1.6	1.5	1.5
7	231	35	1.7	1.6	1.6
8	242	36	1.8	1.7	1.6
9	254	37	1.8	1.8	1.7
10	268	41	2.0	1.9	1.8
11	286	47	2.4	2.2	2.0
12	309	54	2.9	2.8	2.6
13	335	52	3.5	3.3	3.2
14	358	48	3.9	3.7	3.6
15	378	47	4.3	4.1	4.0
16	398	50	4.6	4.5	4.3
17	418	59	5.0	4.8	4.7
18	438	73	5.6	5.4	5.1

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.4.5.2**

Raw Score to Scale Score to Proficiency Level Conversion: Spek 6-8 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G6</b>	<b>PL for G7</b>	<b>PL for G8</b>
6	148	50	1.0	1.0	1.0
7	223	32	1.6	1.6	1.5
8	232	31	1.7	1.6	1.6
9	241	30	1.7	1.7	1.6
10	249	30	1.8	1.7	1.7
11	257	30	1.9	1.8	1.8
12	265	30	1.9	1.9	1.8
13	273	30	2.1	1.9	1.9
14	282	31	2.3	2.1	1.9
15	291	32	2.5	2.3	2.1
16	301	34	2.7	2.6	2.4
17	313	36	3.0	2.9	2.7
18	325	37	3.3	3.1	3.0
19	338	37	3.5	3.4	3.2
20	350	36	3.8	3.6	3.5
21	362	35	4.0	3.8	3.7
22	373	34	4.2	4.0	3.9
23	384	34	4.4	4.2	4.1
24	394	34	4.5	4.4	4.3
25	405	35	4.7	4.6	4.5
26	416	36	4.9	4.8	4.6
27	429	39	5.3	5.1	4.9
28	442	44	5.7	5.5	5.3
29	455	51	6.0	5.9	5.7
30	468	60	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

#### 2.6.4.6 Grades 9–12

**Table 2.6.4.6.1**

Raw Score to Scale Score to Proficiency Level Conversion: Spek 9-12 A S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G9</b>	<b>PL for G10</b>	<b>PL for G11</b>	<b>PL for G12</b>
0	172	38	1.1	1.0	1.0	1.0
1	172	38	1.1	1.0	1.0	1.0
2	172	38	1.1	1.0	1.0	1.0
3	183	34	1.2	1.1	1.1	1.0
4	193	33	1.2	1.2	1.2	1.1
5	203	33	1.3	1.3	1.2	1.2
6	214	35	1.4	1.4	1.3	1.3
7	226	37	1.5	1.4	1.4	1.4
8	239	38	1.6	1.5	1.5	1.5
9	253	40	1.7	1.6	1.6	1.6
10	268	43	1.8	1.8	1.7	1.7
11	287	48	1.9	1.9	1.9	1.8
12	311	54	2.5	2.4	2.3	2.2
13	337	52	3.1	3.0	3.0	2.9
14	360	48	3.5	3.4	3.3	3.3
15	380	47	3.9	3.7	3.6	3.6
16	401	50	4.2	4.1	4.0	3.9
17	422	59	4.6	4.5	4.4	4.3
18	443	75	5.1	4.9	4.8	4.7

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

**Table 2.6.4.6.2**

Raw Score to Scale Score to Proficiency Level Conversion: Spek 9-12 B/C S503 Paper

<b>Raw Score</b>	<b>Scale Score</b>	<b>CSEM x 1.96</b>	<b>PL for G9</b>	<b>PL for G10</b>	<b>PL for G11</b>	<b>PL for G12</b>
6	172	37	1.1	1.0	1.0	1.0
7	217	31	1.4	1.4	1.3	1.3
8	226	31	1.5	1.4	1.4	1.4
9	235	31	1.5	1.5	1.5	1.4
10	243	30	1.6	1.6	1.5	1.5
11	252	30	1.7	1.6	1.6	1.6
12	260	30	1.7	1.7	1.7	1.6
13	268	30	1.8	1.8	1.7	1.7
14	277	31	1.9	1.8	1.8	1.8
15	286	33	1.9	1.9	1.9	1.8
16	296	34	2.1	2.0	1.9	1.9
17	308	36	2.4	2.3	2.2	2.1
18	321	38	2.8	2.6	2.5	2.5
19	334	38	3.1	3.0	2.9	2.8
20	347	37	3.3	3.2	3.1	3.1
21	358	35	3.5	3.4	3.3	3.2
22	369	34	3.7	3.6	3.5	3.4
23	380	33	3.9	3.7	3.6	3.6
24	390	33	4.0	3.9	3.8	3.7
25	400	34	4.2	4.1	4.0	3.9
26	411	36	4.4	4.3	4.2	4.1
27	424	39	4.7	4.5	4.4	4.3
28	437	44	4.9	4.8	4.7	4.6
29	455	54	5.5	5.3	5.1	5.0
30	476	72	6.0	6.0	6.0	6.0

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

## **2.7    *Equating and Recalibration Summary***

All ACCESS Series 503 Paper test forms are pre-equated static forms (see Part 1, Section 2.3). For technical details on the Kindergarten test, see MacGregor, Kenyon, Gibson, and Evans (2009). For the ACCESS Series 503 Grades 1–12, we provide detail below on prior years that test forms have been used, where relevant, and on pre-equating and calibration processes that were in place at the time the forms were constructed.

### **Listening and Reading**

ACCESS Paper Listening and Reading Grades 1–12 Tier A and Grades 1-5 Tier B/C were used in Series 501, which was newly constructed in 2019-20. Aside for one Grade 9-12 Tier A Listening item which was replaced with a newly created item starting in Series 501 because it showed a C-Level DIF, all Tier A items were constructed using S402, S400, and S302 Paper item pools, while Grades 1-5 Tier B/C was created using Series 302 and Series 303 Tier B and C item pools. For ACCESS Paper Listening and Reading Grades 6-12 B/C, we used a newly created form which is made of continuing items from the operational test and newly items (See Part 1 Section 2.3.2.) Table 2.7.1 summarizes the sources of Listening and Reading forms for Paper Series 503.

**Table 2.7.1**

Sources of Series 503 Paper Listening and Reading Forms

	<b>Listening</b>		<b>Reading</b>	
Tier A <sup>3</sup>	Years previously used: Series 501 Paper, created with items from:	2019-20	Years previously used: Series 501 Paper, created with items from:	2019-20
	Series 403 Paper Series 402 Paper Series 401 Paper Series 400 Paper Series 303 Paper	2018-19 2017-18 2016-17 2015-16 2014-15	Series 402 Paper Series 400 Paper Series 302 Paper	2017-18 2015-16 2013-14
Grades 1-5, Tier B/C	Series 501, created with items from:	2019-20	Series 501, created with items from:	2019-20
	S303 Tier BC S302 Tier BC	2014-15 2013-14	S303 Tier BC S302 Tier BC	2014-15 2013-14
Grades 6-12, Tier BC	Newly created form	2021-22	Newly created form	2021-22

Since Series 401 Paper, Series 400 Paper, and Series 303 Listening Grades 1–12 test forms are identical, and since the Series 401 Paper population is more current than the Series 400 Paper population, we refined the item parameters for the Series 303 Listening Grades 1–12 forms using Series 401 Paper population data. In the 2018 recalibration analyses, we initially anchored the difficulty measures of the Series 303 test items to their previously calibrated values from the Series 303 annual equating study. After the first calibration run, some items that were initially anchored proved to have changed in their difficulty measure, which is measured by the “Displacement” statistic. This statistic shows the difference between the difficulty value of the anchored item and what the difficulty value would have been had it not been anchored. If this value was large (i.e., above 0.30 or below -0.30), we unanchored that item in the final calibration run (i.e., its parameter was re-estimated). For Series 503 Paper Reading Grades 1–12 Tier A forms and Grades 1-5 B/C forms, a similar process was used to refine Series 302 and Series 303 item parameters using Series 400 and 401 Paper student population data, respectively.

---

<sup>3</sup> Effective OP School Year 2021-2022 and annually thereafter, one item from Part F no longer appears on the operational Grades 9-12 Listening Tier A test (due to C-level DIF). The folder is replaced with another item from Part F.

For Listening Tier A, we applied these refined parameters to the intact Tier A forms from Series 303. For Reading Tier A, we applied these refined parameters to the intact Tier A forms from Series 303.

For Grades 1-5 Listening and Reading Tier B/C, we used the refined parameters derived from the 2018 recalibration studies to conduct a form selection meeting in 2018. We constructed the Series 402 and 403 Paper Listening and Reading Grades 1–12 Tier B/C forms at this meeting. These two forms have been used on a rotating basis, as static forms, since then.

For Grades 6-12 Listening and Reading Tier B/C, we used the parameters derived from the 2021 Paper Grades 6-12 Tier B/C field test calibration study (see Part 1 Section 2.3.2) to conduct a form selection meeting in 2021. In this Paper Grades 6-12 Tier B/C field test, new test items were appended to the end of the Series 502 test books and administered to the entire state of Florida. Test data collected from this field test administration was used to conduct the calibration study and item selection. In the calibration analyses, we initially anchored the difficulty measures of the Series 502 test items to their previously calibrated values while the difficulty measures of the new test items were estimated. After the first calibration run, some items that were initially anchored proved to have changed in their difficulty measure, which is measured by the “Displacement” statistic. This statistic shows the difference between the difficulty value of the anchored item and what the difficulty value would have been had it not been anchored. If this value was large (i.e., above 0.30 or below -0.30), we unanchored that item in the final calibration run (i.e., its parameter was re-estimated). We constructed the Series 503 Paper Listening and Reading Grades 6-12 B/C forms at this meeting.

## **Writing and Speaking**

Writing and Speaking are also static forms. Table 2.7.2 summarizes prior uses of these forms. Please see the Annual Technical Report for ACCESS for ELLs Paper Series 401 (CAL, 2018) for equating summaries for Writing and Speaking.

**Table 2.7.2**

Sources of Series 502 Paper Writing and Speaking Forms

	<b>Writing</b>		<b>Speaking</b>	
Tier A	Years previously used: Series 501 Series 402 Series 400	2019-20 2017-18 2015-16	Years previously used: Series 501 Series 402	2019-20 2017-18
Tier B/C	Years previously used: Series 501 Series 402 Series 400	2019-20 2017-18 2015-16	Years previously used: Series 501 Series 402	2019-20 2017-18



## ***2.8 Test Characteristic Curve***

Test characteristic curves (TCCs) graphically show the relationship between the ability measure (in logits) on the horizontal axis and the expected raw score or the estimated true score on the vertical axis. For a given ability measure, the corresponding expected raw score can be found via the TCC. For reporting purposes, ability measures are used to determine students' proficiency levels. Since TCC transforms ability measures to expected raw scores, this representation allows test users to relate student performance to the number of items on the test.

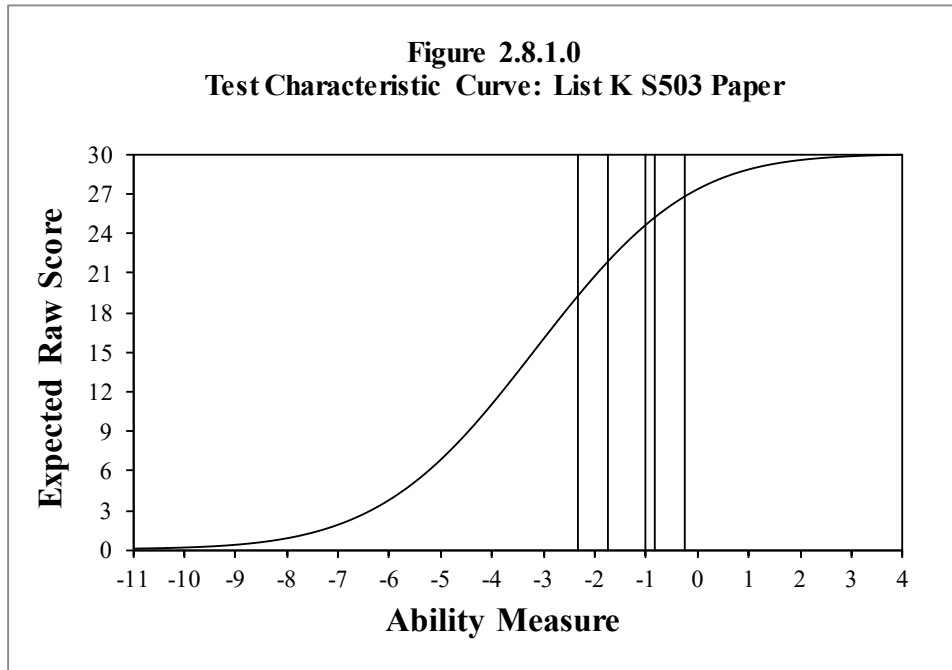
Mathematically, TCC is the sum of all item characteristic functions on the test form (Lord, 1980). Thus, the TCC depends on the item characteristic functions (Lord, 1980) of the items on the test form. The shape of TCC depends on several factors, including the number and the characteristics of items, the item response theory model used, and the values of the item parameters. Because of this, there is no explicit formula for TCC, and there are no parameters for the curve. The general form of the TCC is monotonically increasing. In most cases when the test form consists of multiple-choice items, such as in the Listening and Reading domains, the TCC curve is a smooth S shape. It is flat in the lower ability range, rises steeply in the middle, and becomes flat again on the right, at the level of proficiency above which students are expected to respond correctly to all items. In other cases, however, it will increase smoothly and then have a small plateau before increasing again. In all cases, it will be asymptotic to the value of the total number of items or total expected raw score points in the upper tail. The area where the TCC is the steepest is the area where the test provides higher discrimination and better measurement as compared to the area where the TCC is flat.

For tests consisting of polytomous tasks, the shape of the TCC is also affected by the values of the item category parameters. Since polytomous tasks have more score categories than multiple-choice items, each task has a wide range of values on the proficiency scale. The adjacent category boundaries are sometimes far apart as a result. In this situation, the TCC will have a less smooth curve or a small plateau in the area between the adjacent category boundaries. This pattern can be observed in Writing and Speaking, where the TCC may not form a perfect "S" shape. Such a pattern is also observed in other tests with polytomous items such as the National Assessment of Educational Progress Writing assessment (Muraki, 1993). Conversely, the closer the adjacent category boundaries are, the smoother the rise of the TCC will be along the ability levels.

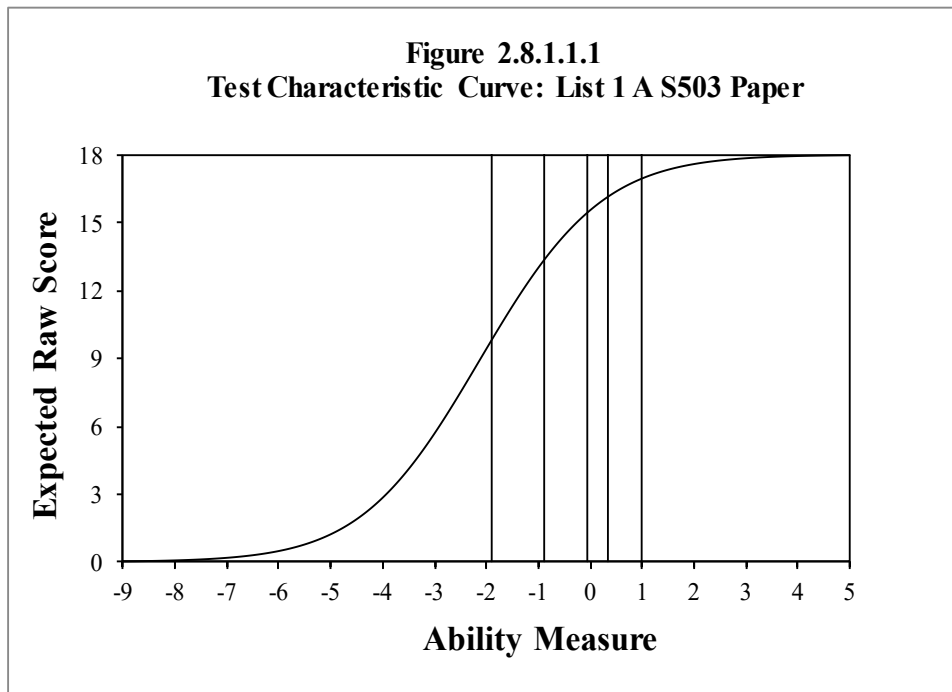
There are five vertical lines in each of the TCC plots indicating the five cut scores for the highest grade in the grade-level cluster for the test form, dividing the figure into six sections for each of the WIDA proficiency levels (PLs 1–6) for the domain being tested. (Note that for Kindergarten and Tier A tests in some domains, it was not possible to place into all six proficiency levels.) As would be expected, higher raw scores are required for placement in higher proficiency levels. The relative width of each section between the cut score lines, however, gives an indication of how many items on that form must be answered correctly (for Listening or Reading) or how many points must be earned (for Writing or Speaking) to be placed into a WIDA proficiency level.

## 2.8.1 Listening

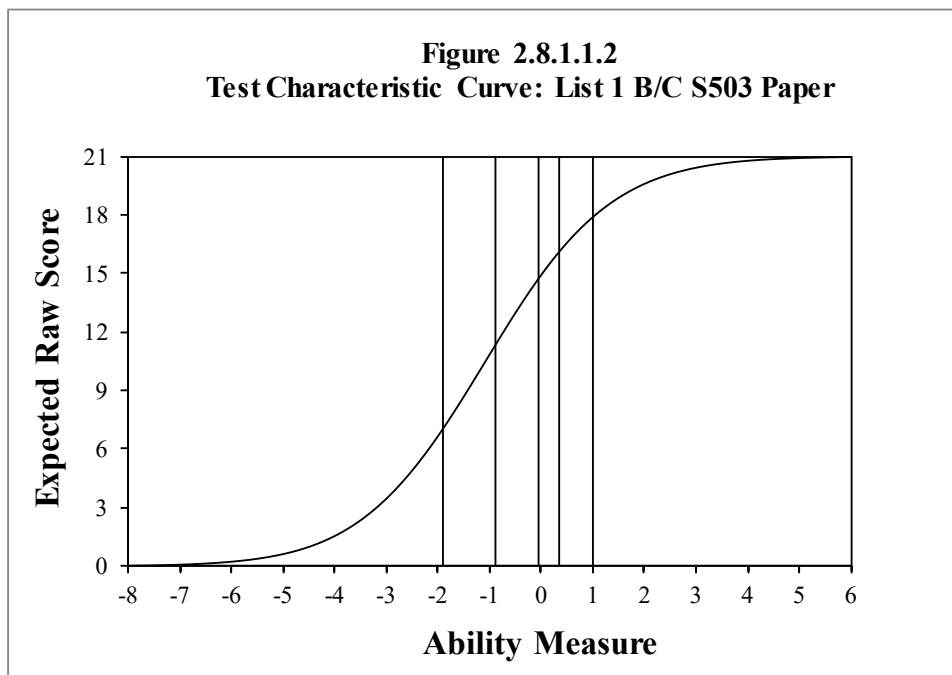
### 2.8.1.0 Kindergarten



2.8.1.1 Grade 1



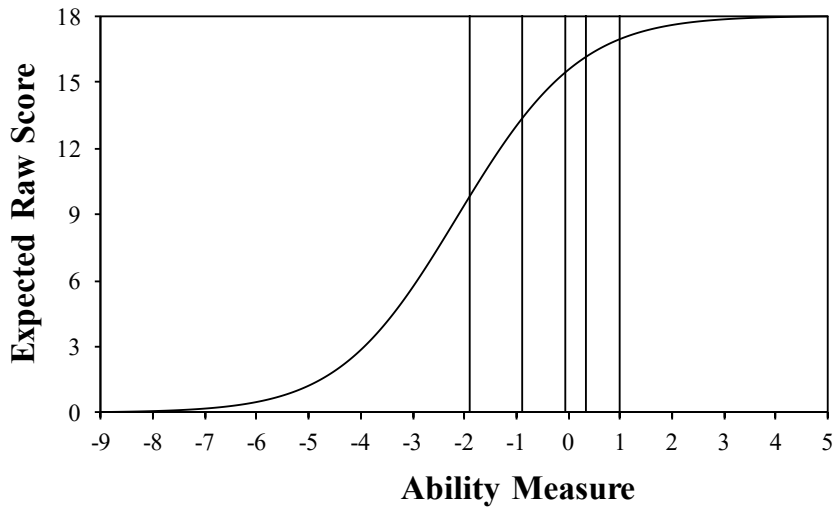
Note: The test form is shared between 1A and 2A.



Note: The test form is shared between 1B/C and 2B/C.

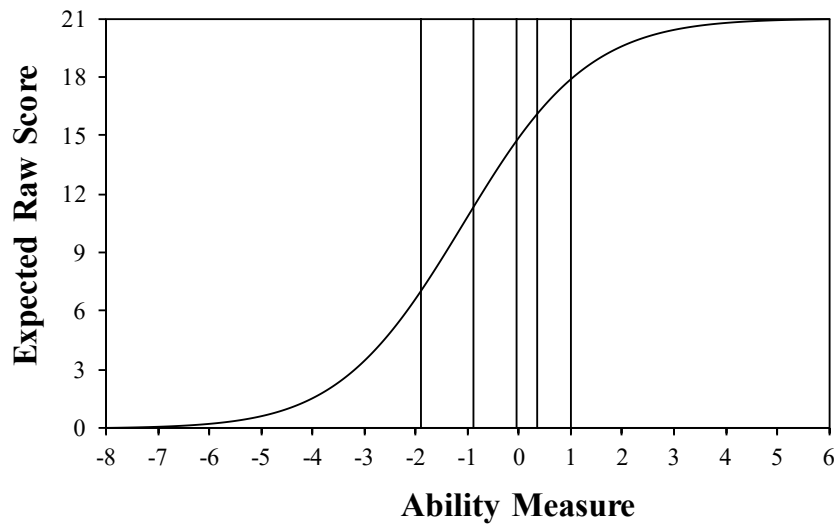
2.8.1.2 Grade 2

**Figure 2.8.1.2.1**  
**Test Characteristic Curve: List 2 A S503 Paper**



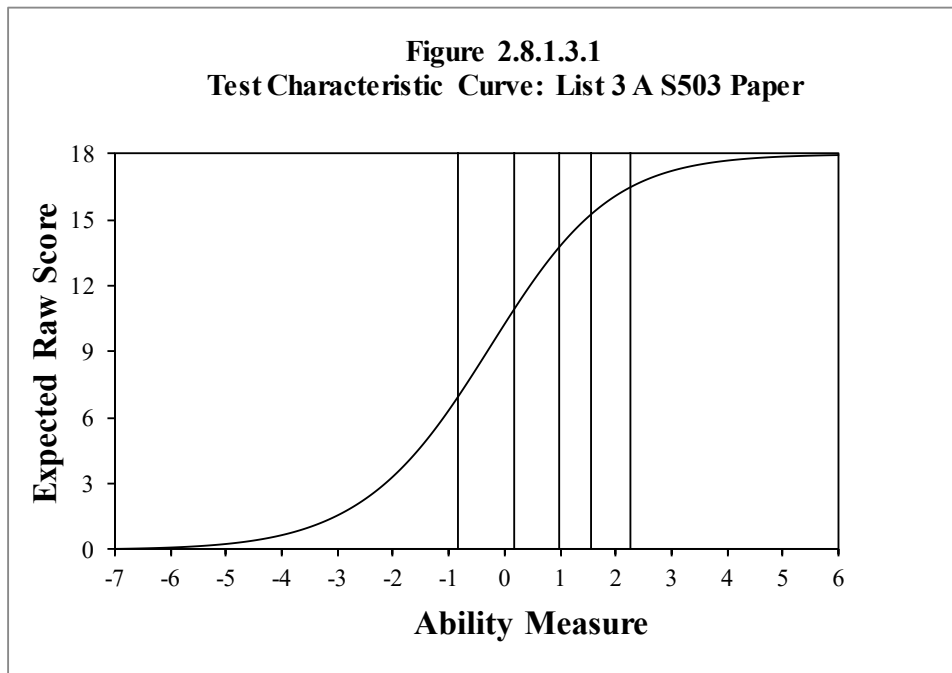
Note: The test form is shared between 1A and 2A.

**Figure 2.8.1.2.2**  
**Test Characteristic Curve: List 2 B/C S503 Paper**

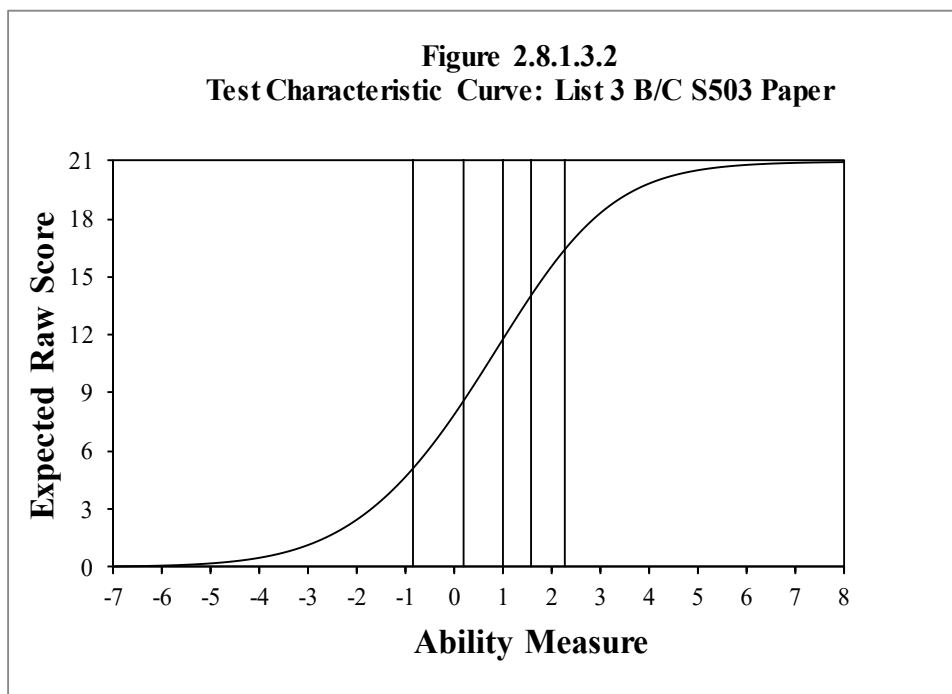


Note: The test form is shared between 1B/C and 2B/C.

2.8.1.3 Grade 3

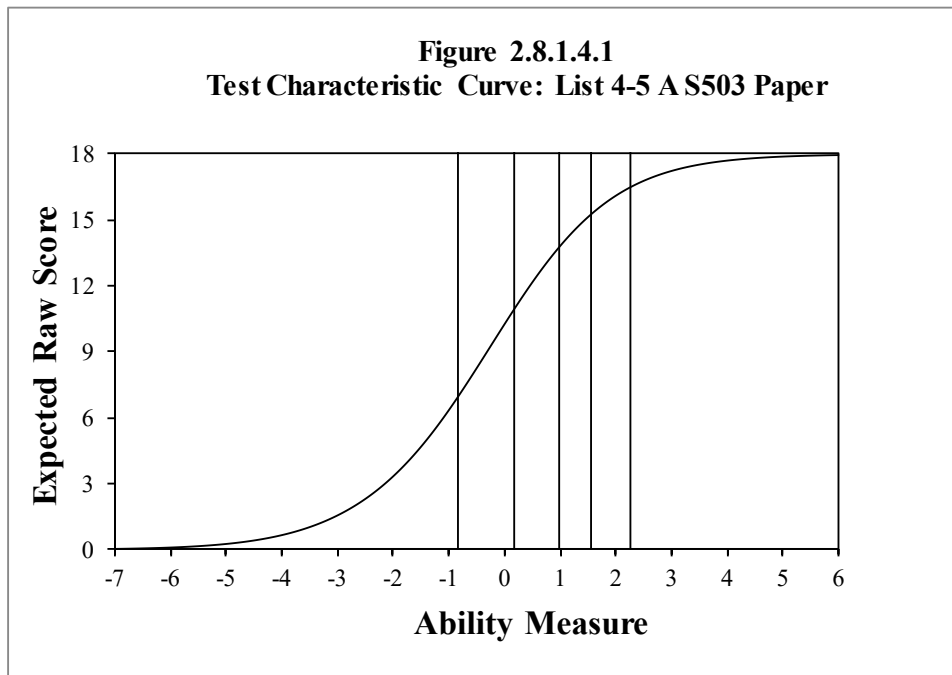


Note: The test form is shared between 3A and 4-5A.

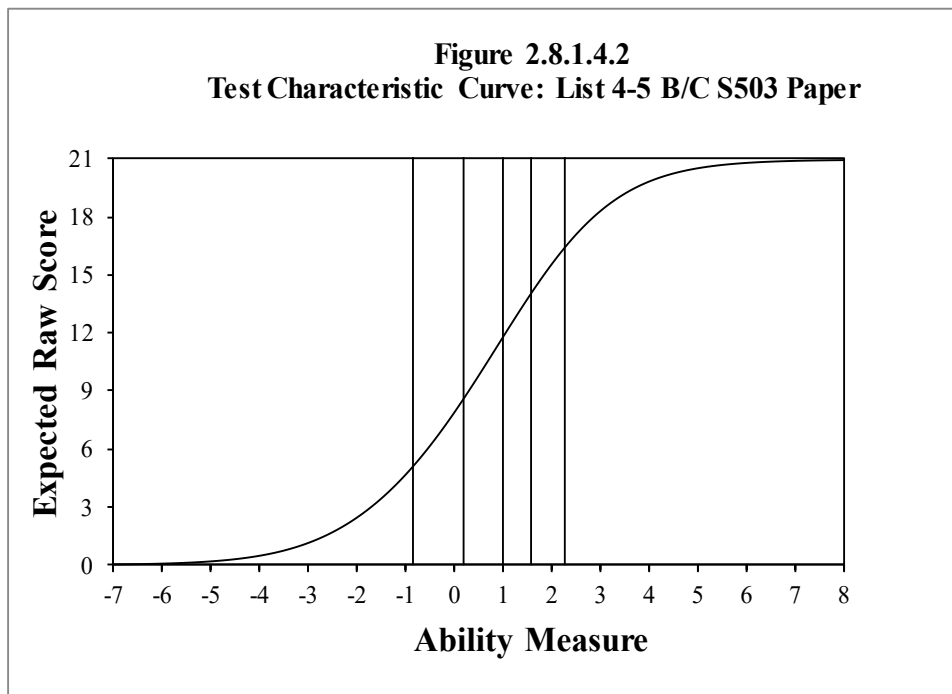


Note: The test form is shared between 3B/C and 4-5B/C.

2.8.1.4 Grades 4-5



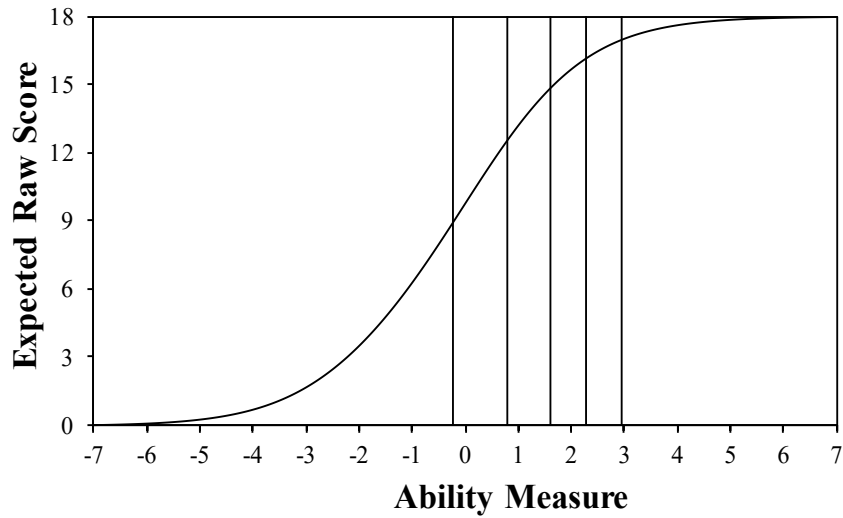
Note: The test form is shared between 3A and 4-5A.



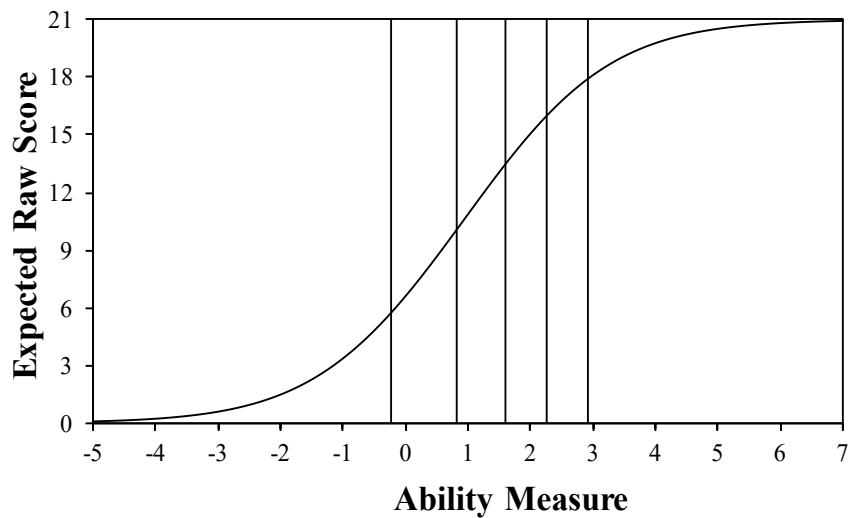
Note: The test form is shared between 3B/C and 4-5B/C.

2.8.1.5 Grades 6-8

**Figure 2.8.1.5.1**  
**Test Characteristic Curve: List 6-8 A S503 Paper**

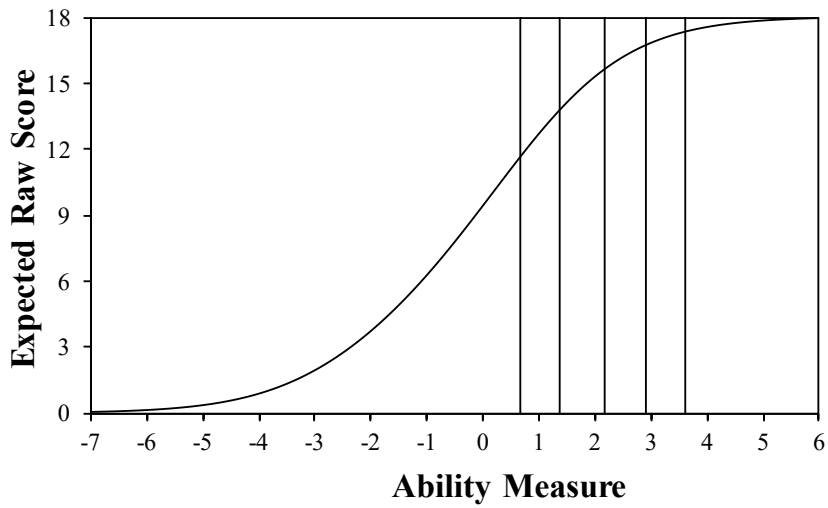


**Figure 2.8.1.5.2**  
**Test Characteristic Curve: List 6-8 B/C S503 Paper**

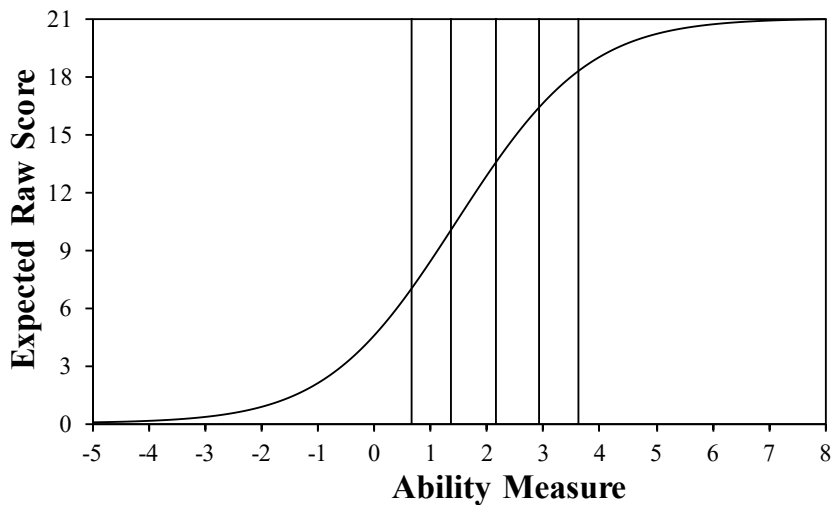


2.8.1.6 Grades 9-12

**Figure 2.8.1.6.1**  
**Test Characteristic Curve: List 9-12 A S503 Paper**



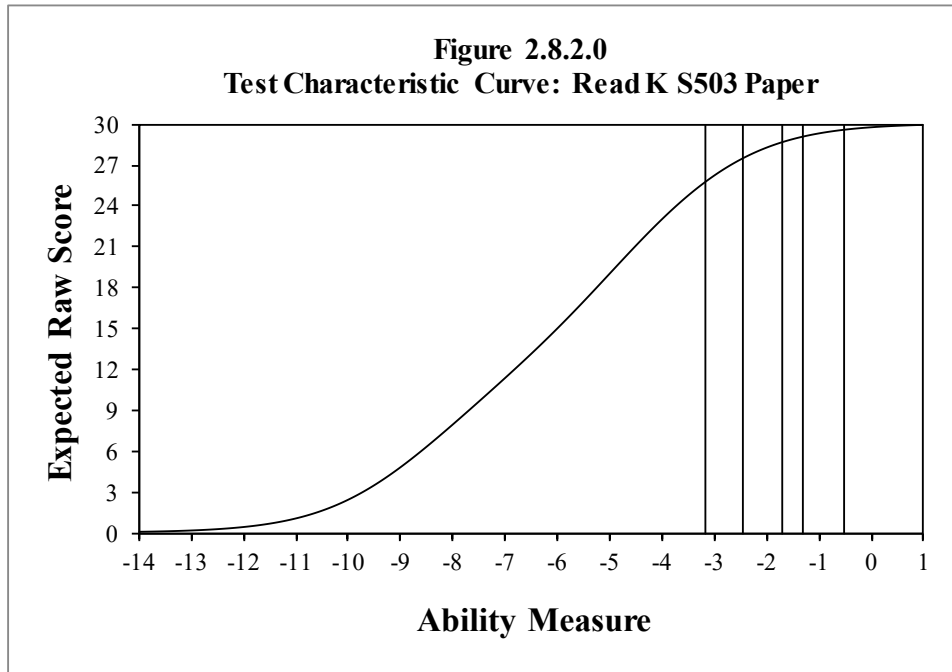
**Figure 2.8.1.6.2**  
**Test Characteristic Curve: List 9-12 B/C S503 Paper**





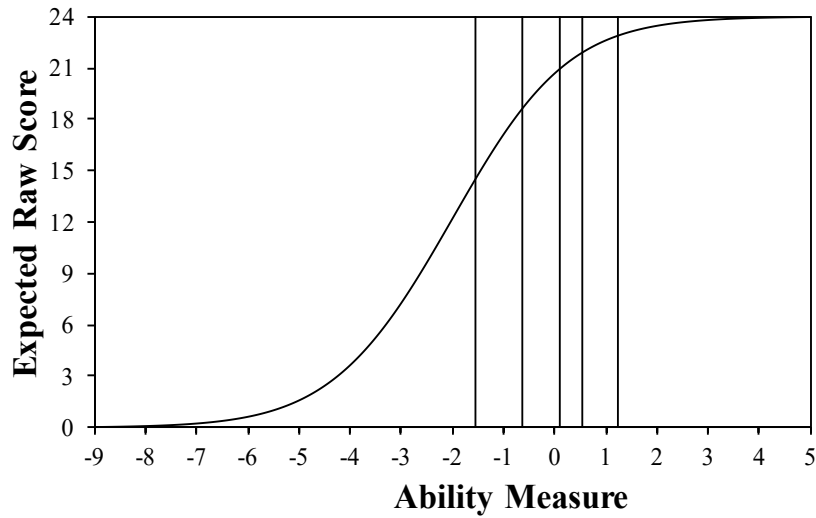
## 2.8.2 Reading

### 2.8.2.0 Kindergarten



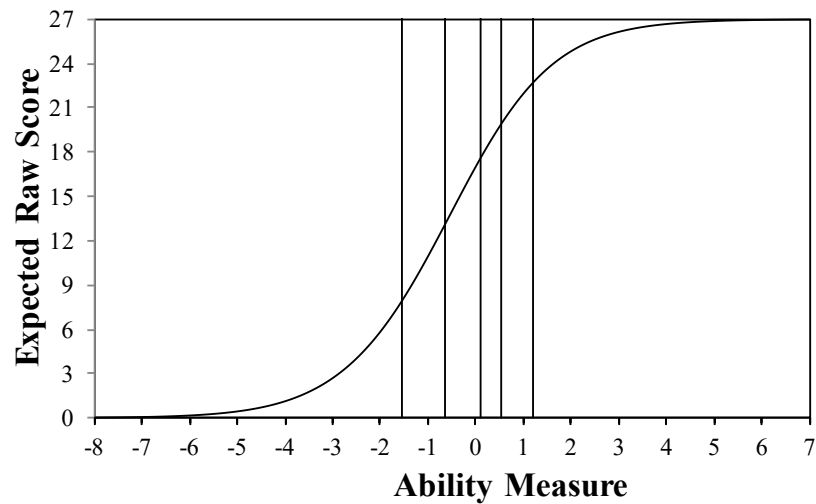
2.8.2.1 Grade 1

**Figure 2.8.2.1.1**  
**Test Characteristic Curve: Read 1 A S503 Paper**



Note: The test form is shared between 1A and 2A.

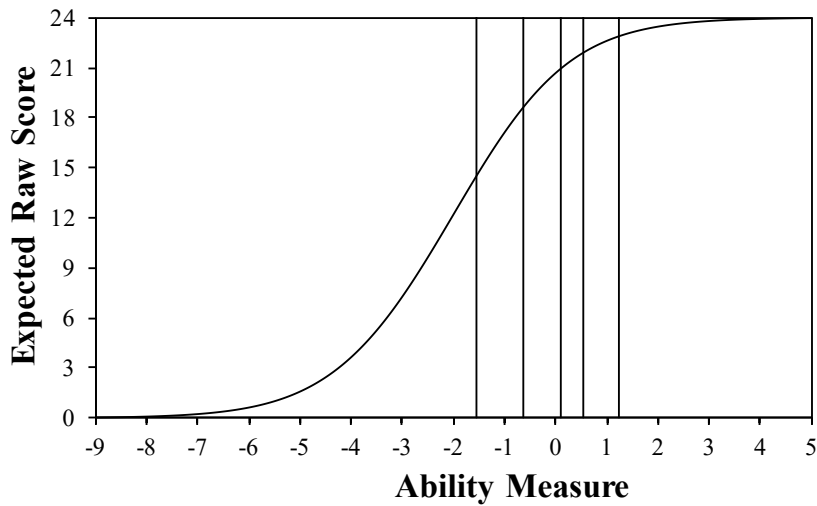
**Figure 2.8.2.1.2**  
**Test Characteristic Curve: Read 1 B/C S503 Paper**



Note: The test form is shared between 1B/C and 2B/C.

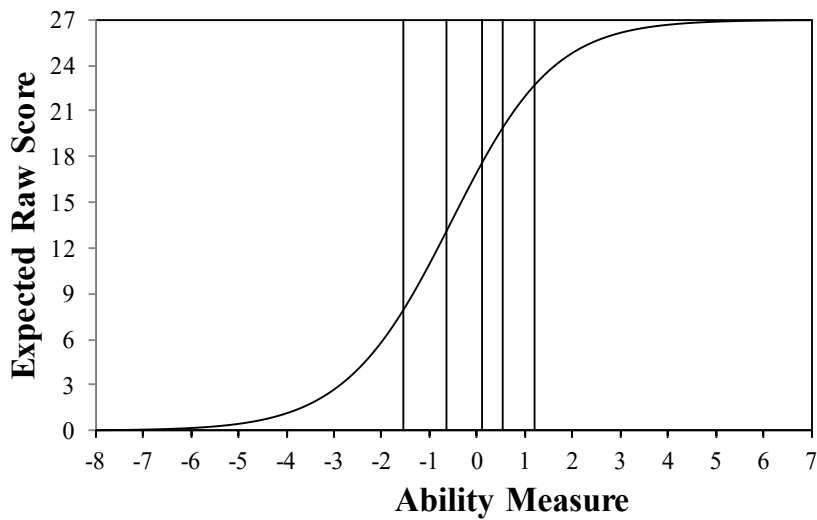
2.8.2.2 Grade 2

**Figure 2.8.2.2.1**  
**Test Characteristic Curve: Read 2 A S503 Paper**



Note: The test form is shared between 1A and 2A.

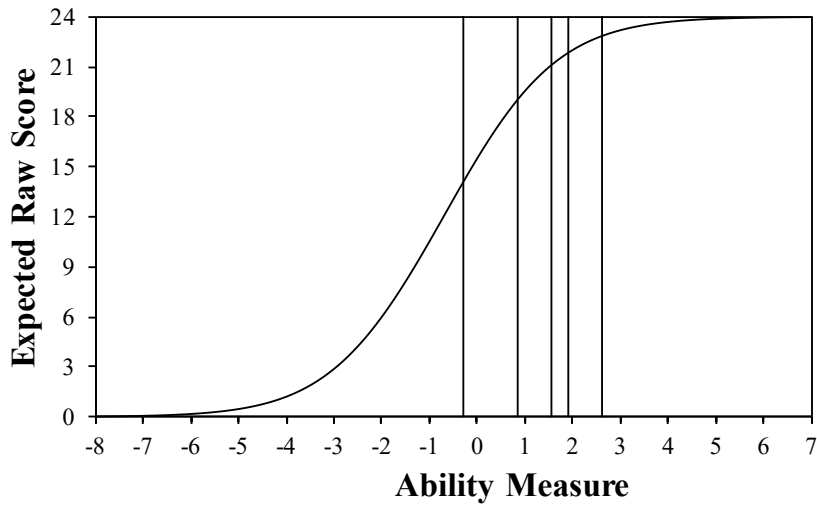
**Figure 2.8.2.2.2**  
**Test Characteristic Curve: Read 2 B/C S503 Paper**



Note: The test form is shared between 1B/C and 2B/C.

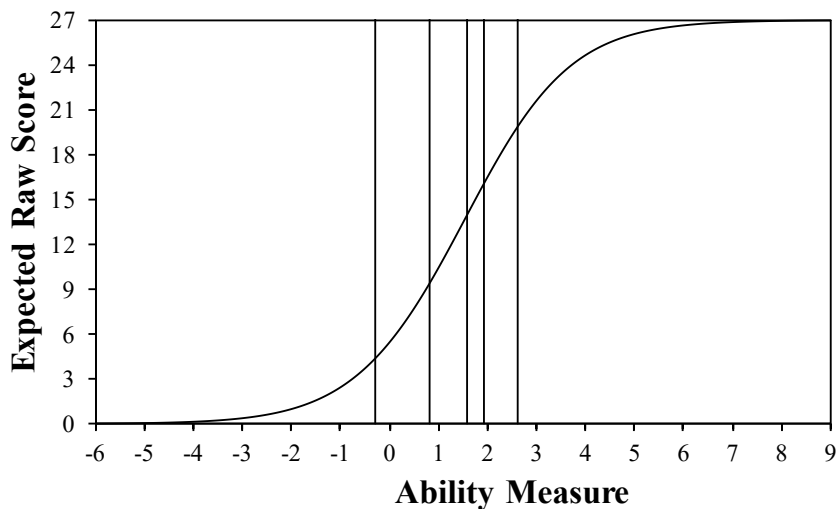
### 2.8.2.3 Grade 3

**Figure 2.8.2.3.1**  
**Test Characteristic Curve: Read 3 A S503 Paper**



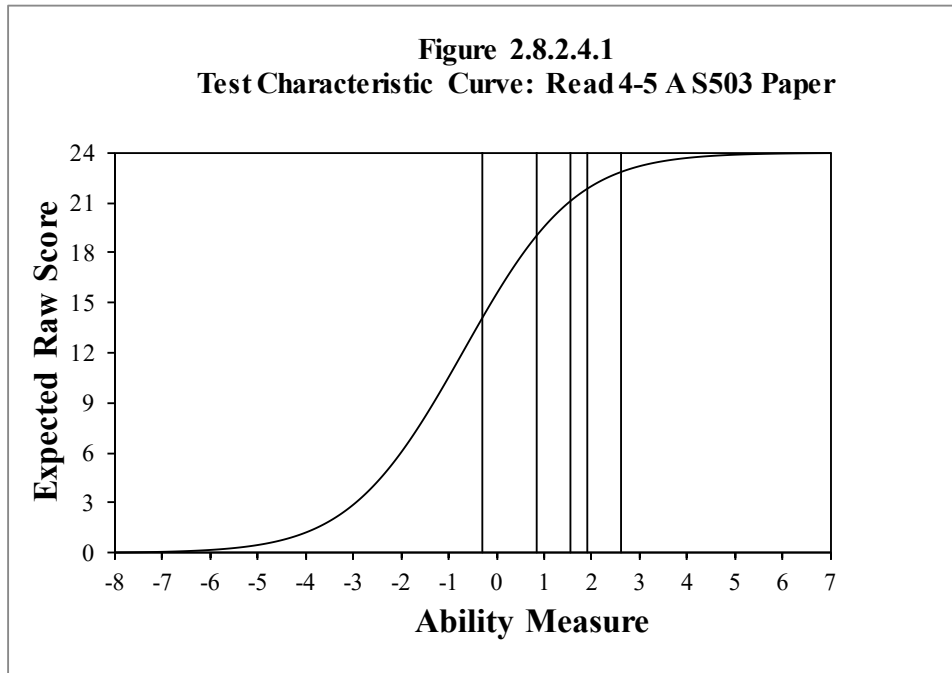
Note: The test form is shared between 3A and 4-5A.

**Figure 2.8.2.3.2**  
**Test Characteristic Curve: Read 3 B/C S503 Paper**



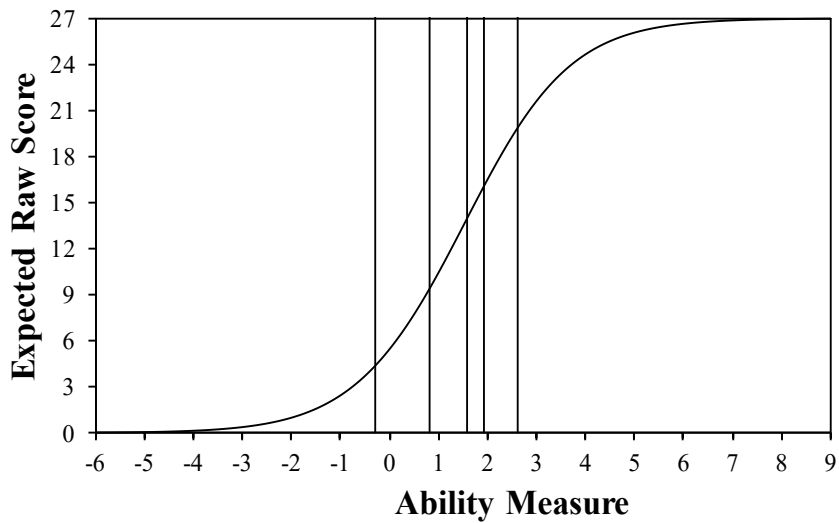
Note: The test form is shared between 3B/C and 4-5B/C.

2.8.2.4 Grades 4-5



Note: The test form is shared between 3A and 4-5A.

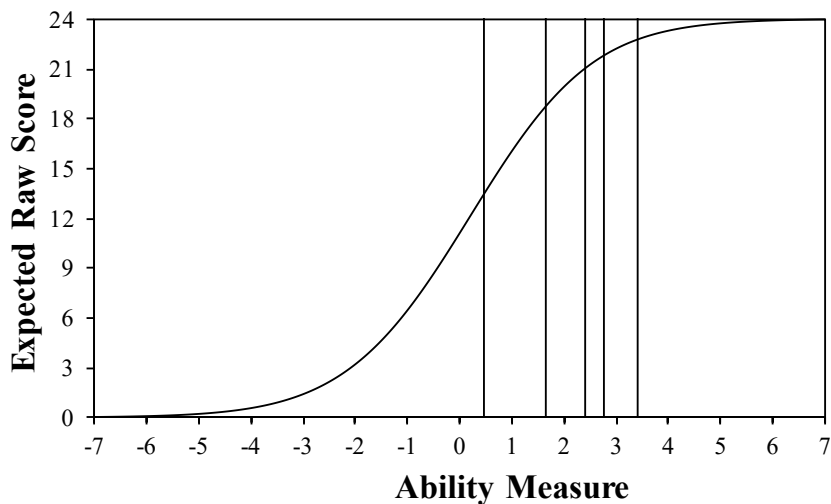
**Figure 2.8.2.4.2**  
**Test Characteristic Curve: Read 4-5 B/C S503 Paper**



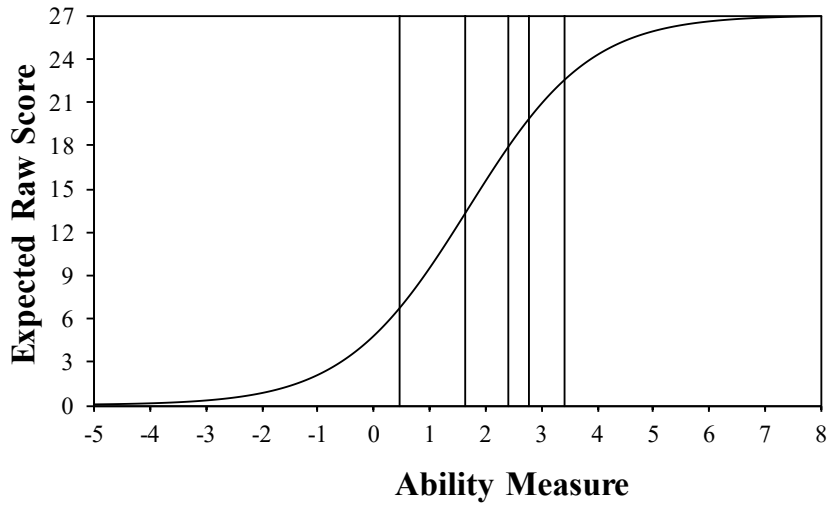
Note: The test form is shared between 3B/C and 4-5B/C.

#### 2.8.2.5 Grades 6-8

**Figure 2.8.2.5.1**  
**Test Characteristic Curve: Read 6-8 A S503 Paper**

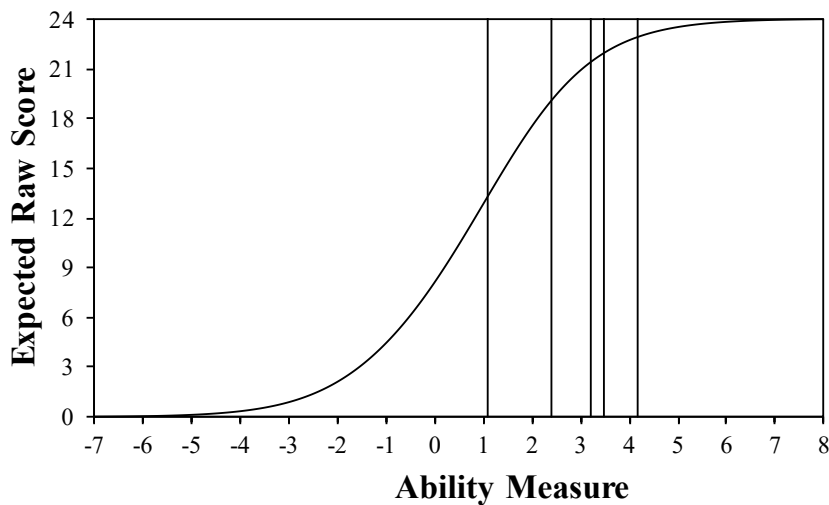


**Figure 2.8.2.5.2**  
**Test Characteristic Curve: Read 6-8 B/C S503 Paper**

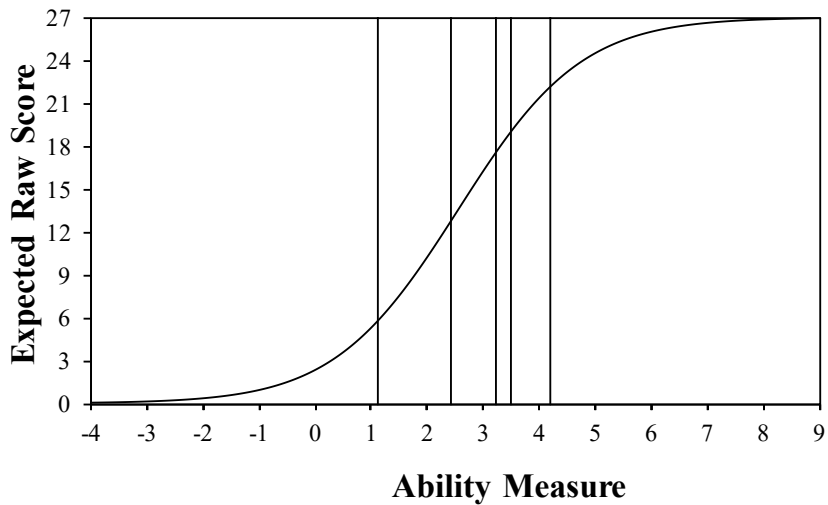


#### 2.8.2.6 Grades 9-12

**Figure 2.8.2.6.1**  
**Test Characteristic Curve: Read 9-12 A S503 Paper**



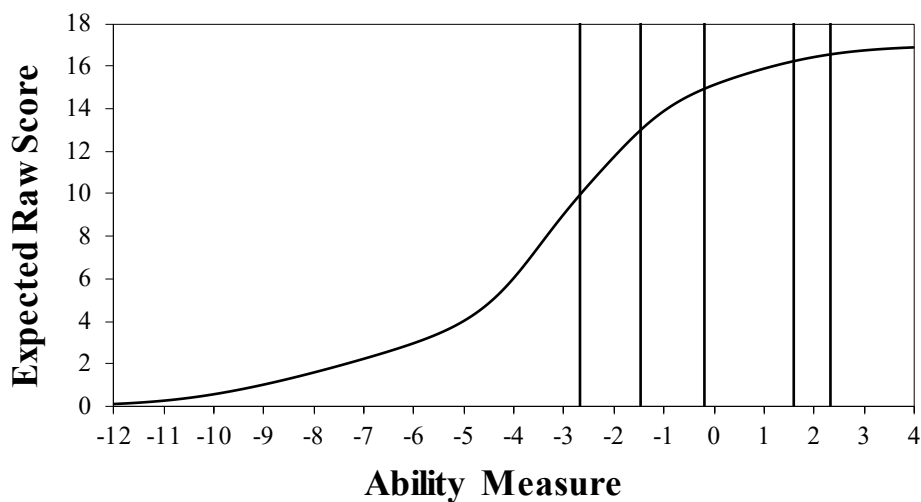
**Figure 2.8.2.6.2**  
**Test Characteristic Curve: Read9-12 B/C S503 Paper**



## 2.8.3 Writing

### 2.8.3.0 Kindergarten

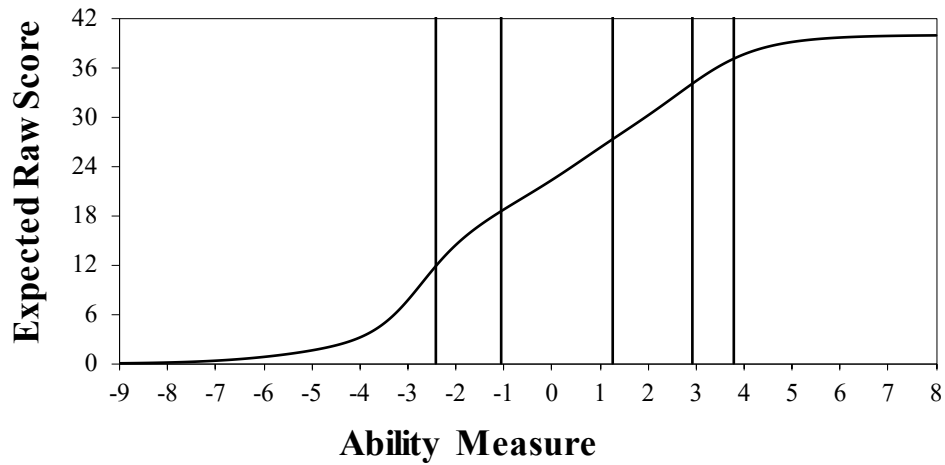
**Figure 2.8.3.0**  
**Test Characteristic Curve: Writ K S503 Paper**



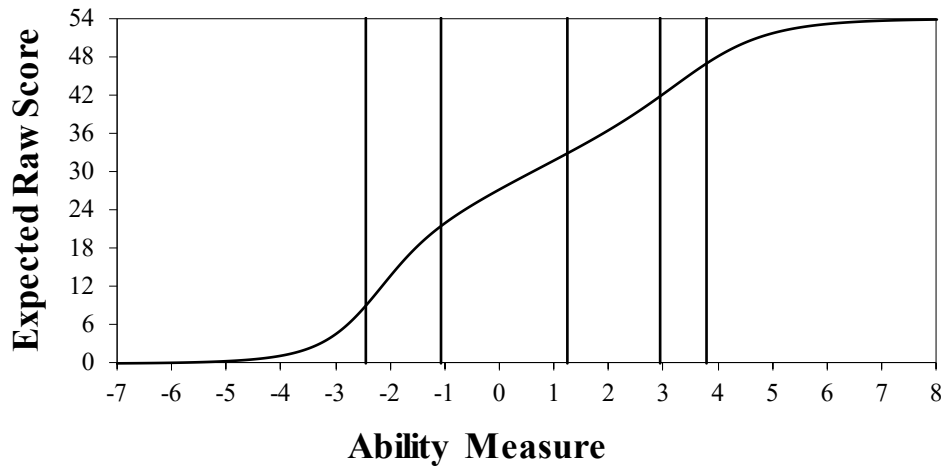


2.8.3.1 Grade 1

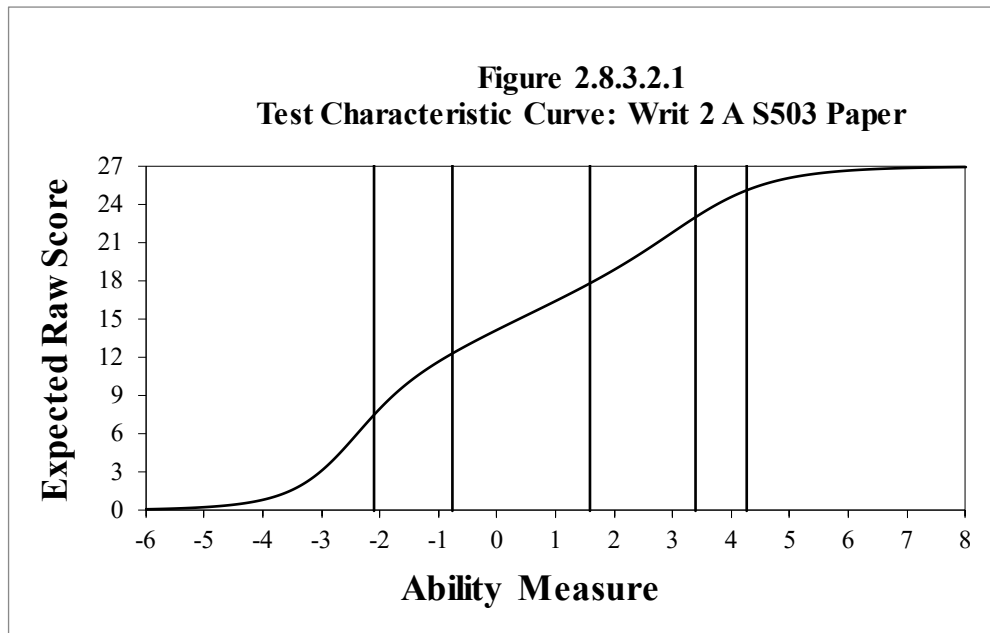
**Figure 2.8.3.1.1**  
**Test Characteristic Curve: Writ 1 A S503 Paper**



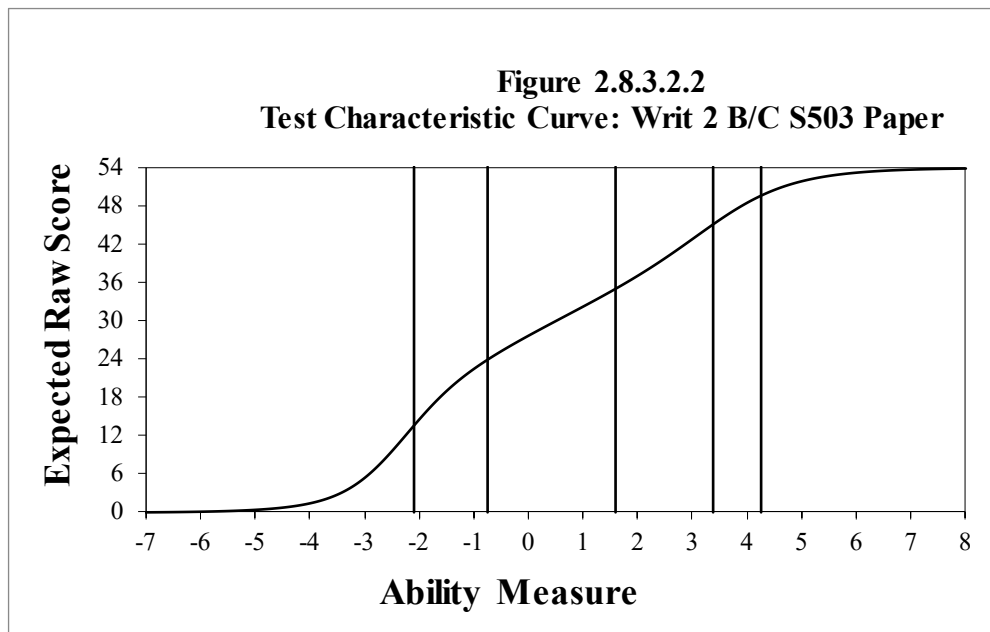
**Figure 2.8.3.1.2**  
**Test Characteristic Curve: Writ 1 B/C S503 Paper**



2.8.3.2 Grade 2

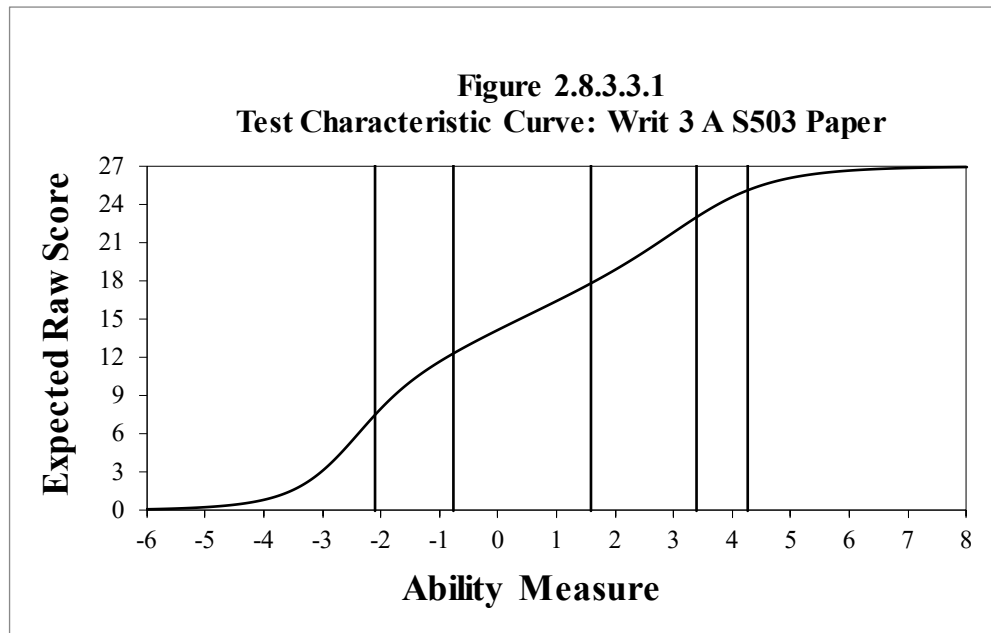


Note: The test form is shared between 2A and 3A.

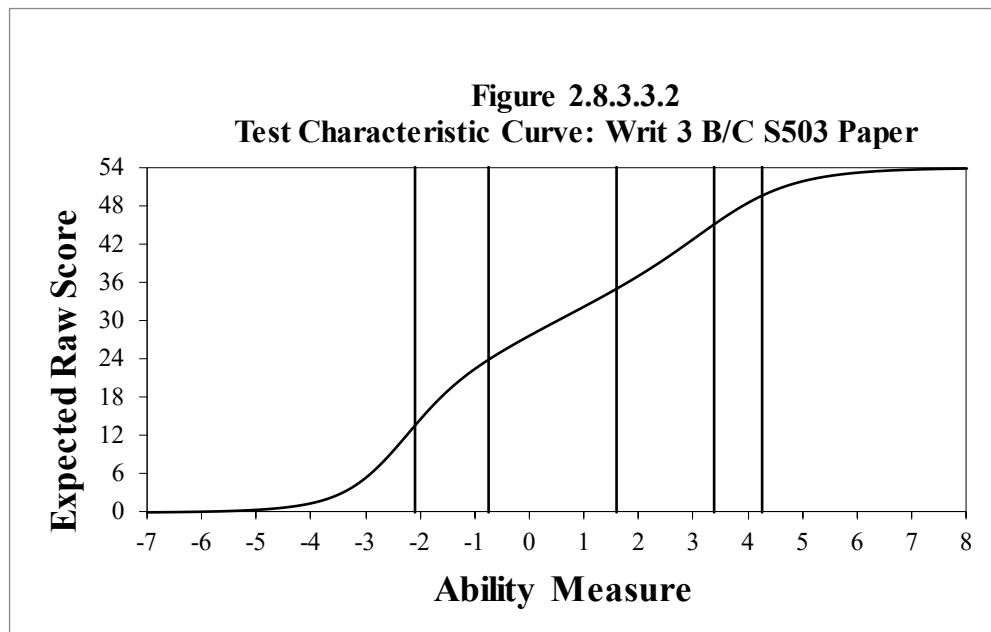


Note: The test form is shared between 2B/C and 3B/C.

### 2.8.3.3 Grade 3

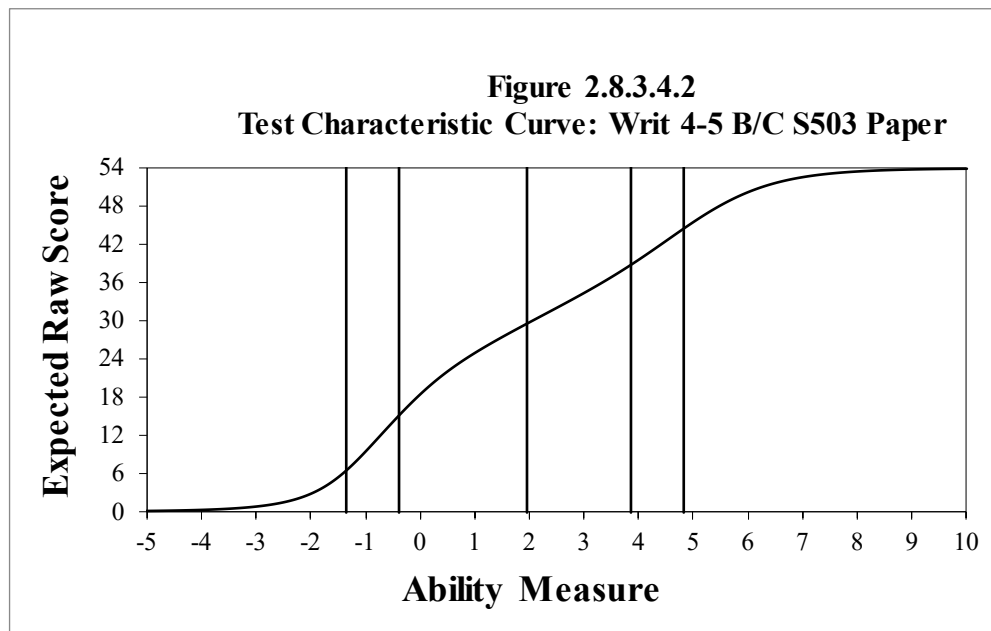
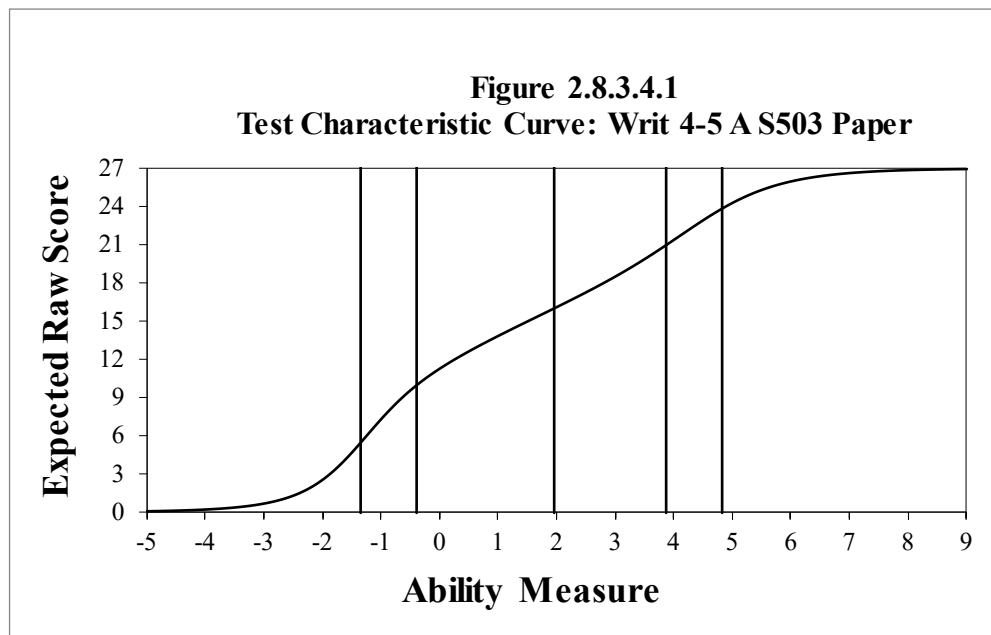


Note: The test form is shared between 2A and 3A.

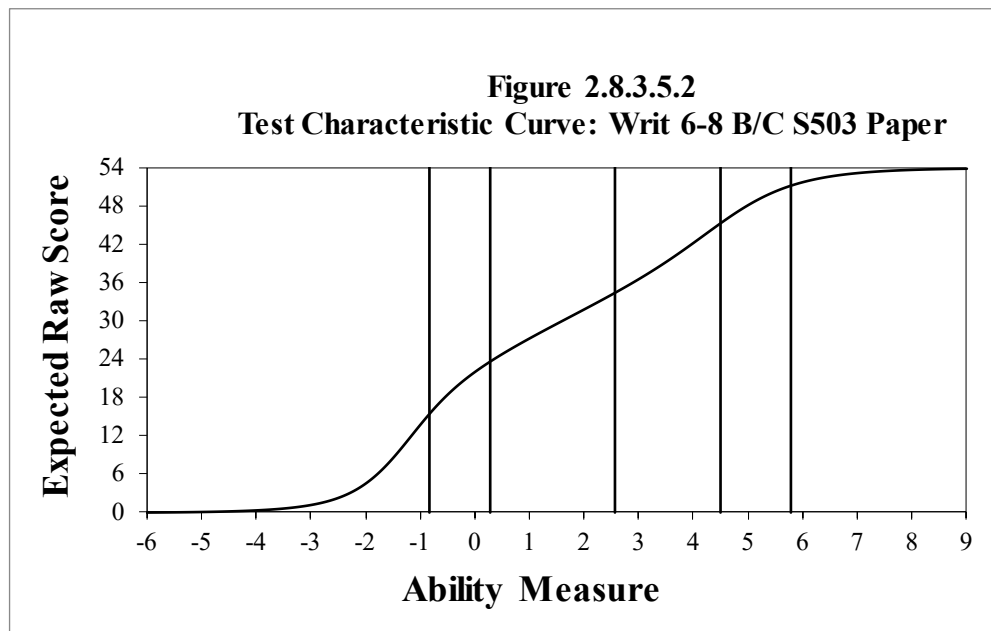
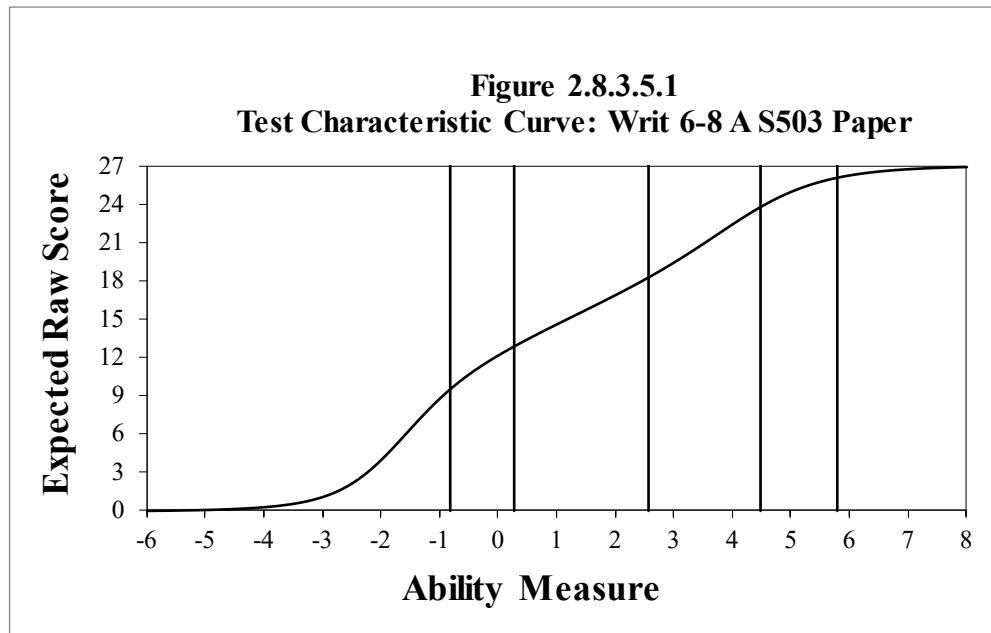


Note: The test form is shared between 2B/C and 3B/C.

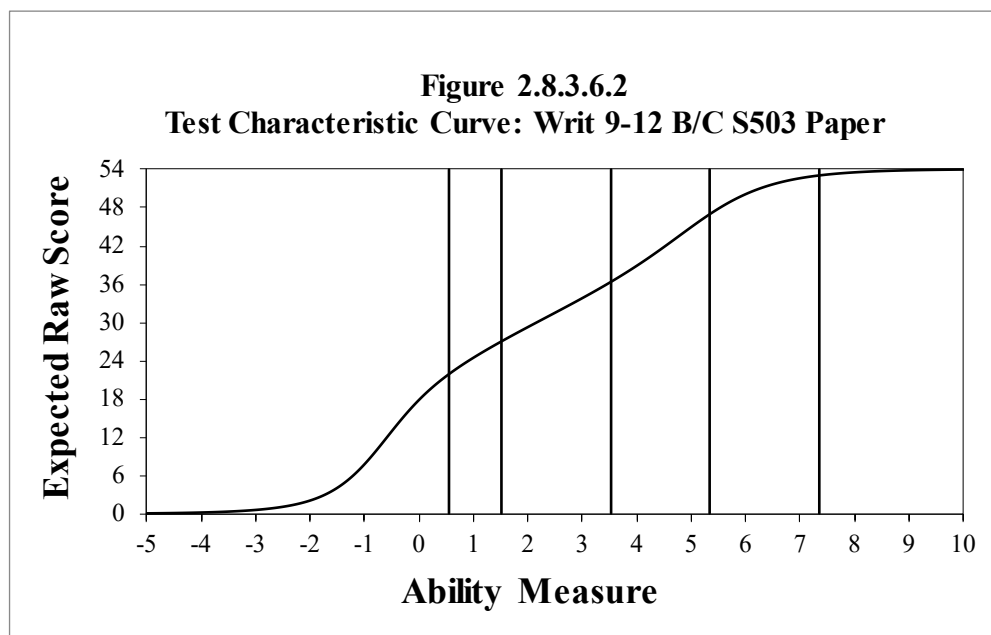
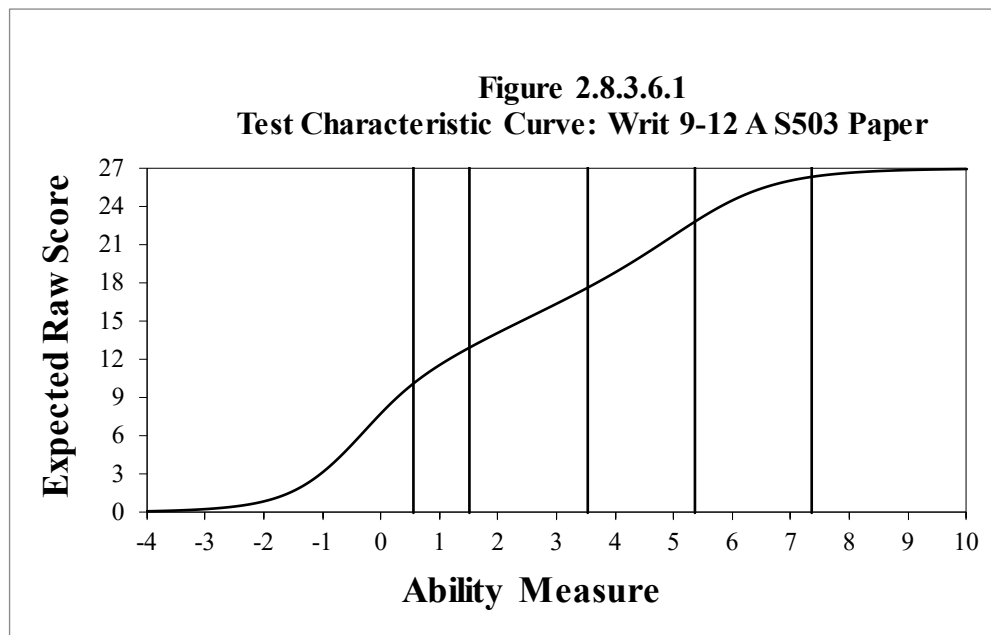
2.8.3.4 Grades 4-5



2.8.3.5 Grades 6-8

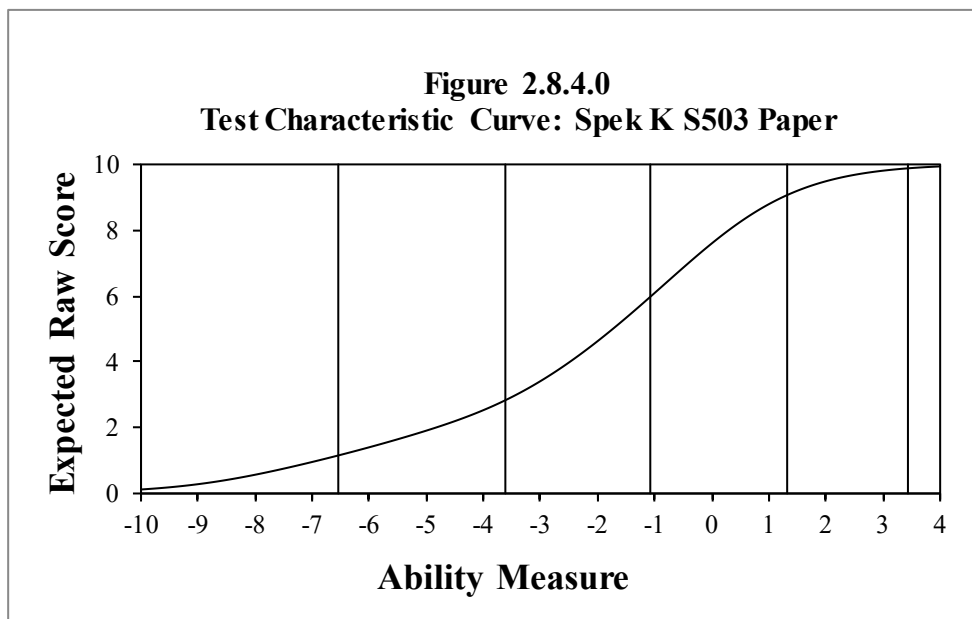


2.8.3.6 Grades 9-12

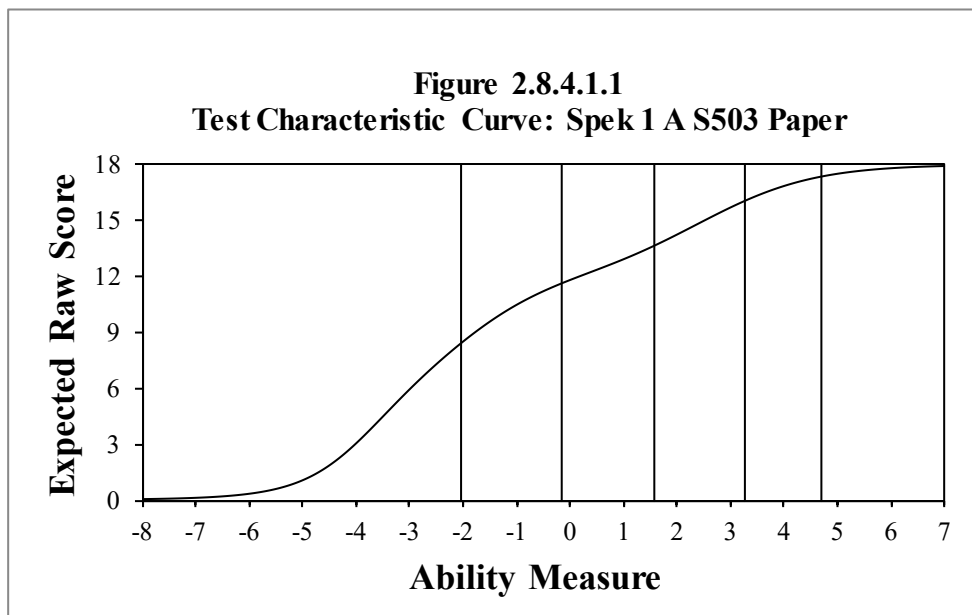


## 2.8.4 Speaking

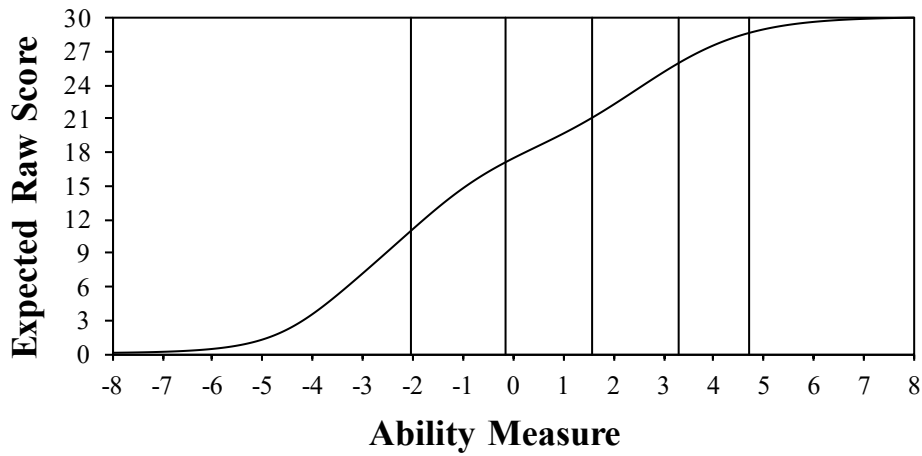
### 2.8.4.0 Kindergarten



### 2.8.4.1 Grade 1



**Figure 2.8.4.1.2**  
**Test Characteristic Curve: Spek 1 B/C S503 Paper**



*2.8.4.2 Grade 2*

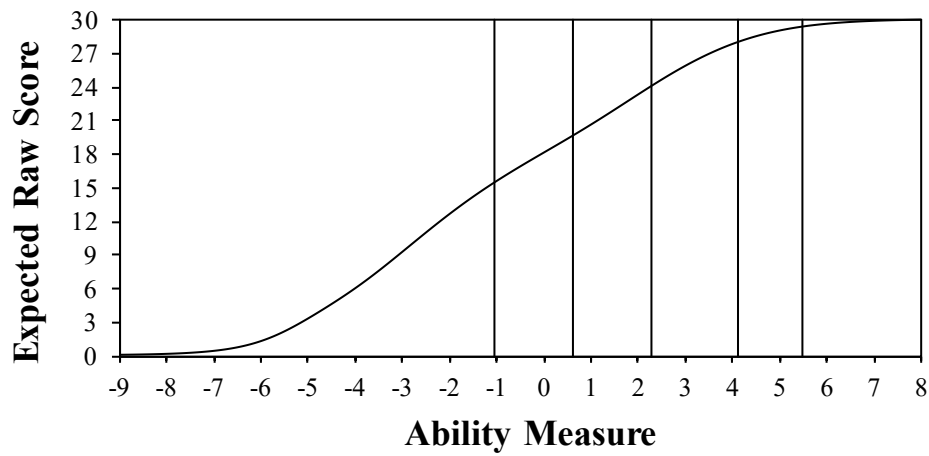
**Figure 2.8.4.2.1**  
**Test Characteristic Curve: Spek 2 A S503 Paper**



Note: The test form is shared between 2A and 3A.



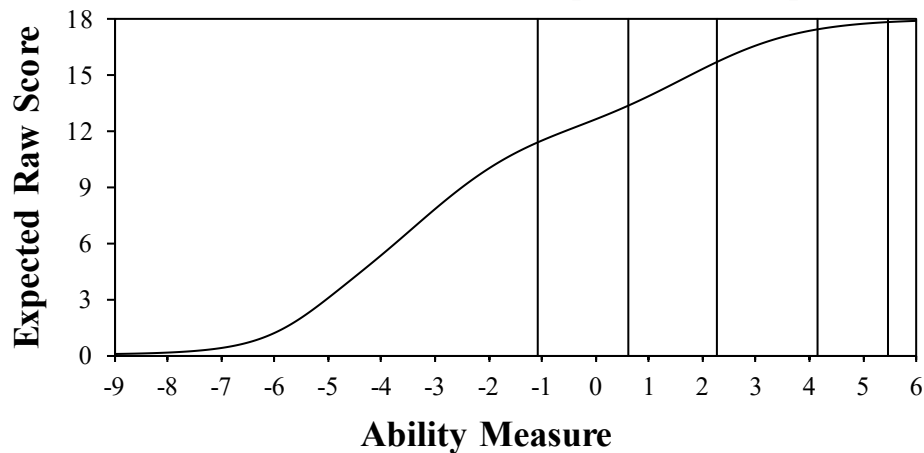
**Figure 2.8.4.2.2**  
**Test Characteristic Curve: Spek 2 B/C S503 Paper**



Note: The test form is shared between 2B/C and 3B/C.

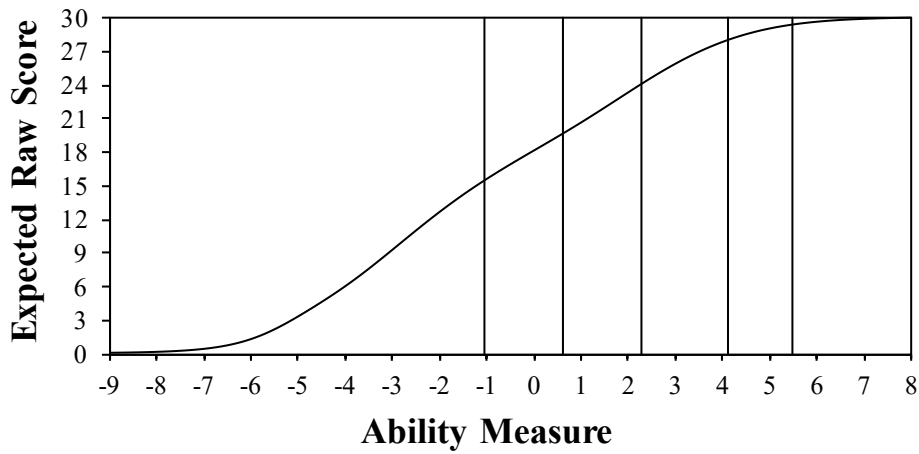
### 2.8.4.3 Grade 3

**Figure 2.8.4.3.1**  
**Test Characteristic Curve: Spek 3 A S503 Paper**



Note: The test form is shared between 2A and 3A.

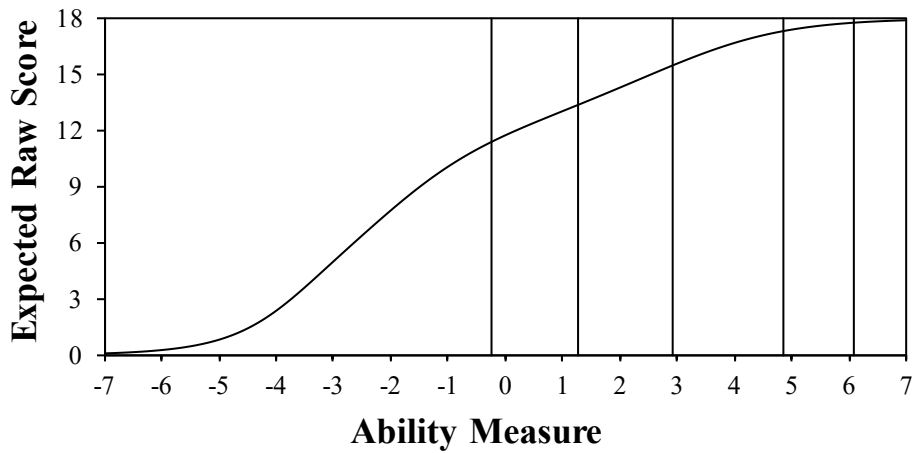
**Figure 2.8.4.3.2**  
**Test Characteristic Curve: Spek 3 B/C S503 Paper**



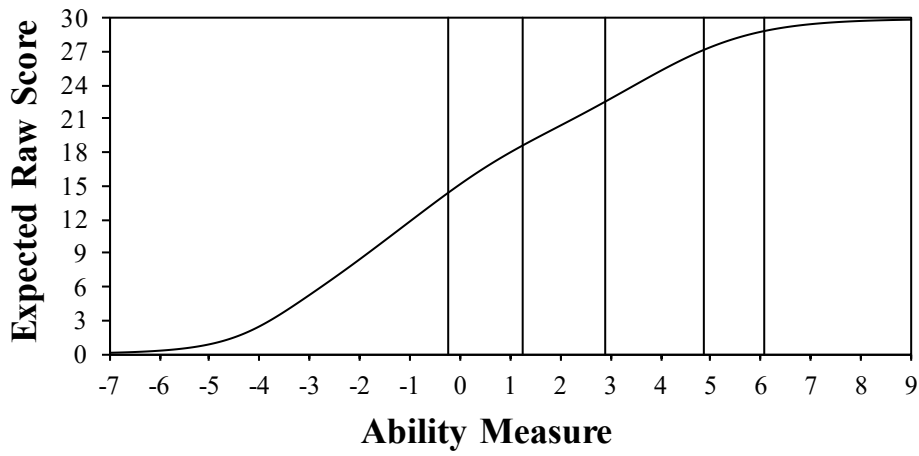
Note: The test form is shared between 2B/C and 3B/C.

#### 2.8.4.4 Grades 4-5

**Figure 2.8.4.4.1**  
**Test Characteristic Curve: Spek 4-5 A S503 Paper**

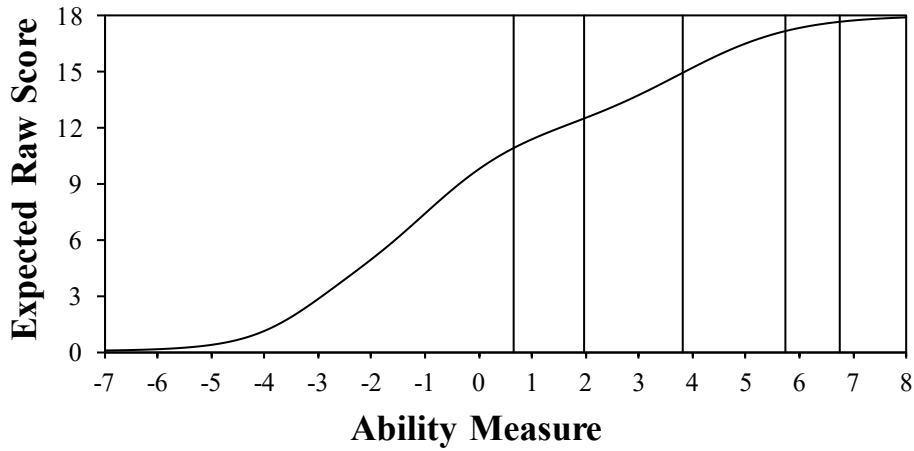


**Figure 2.8.4.4.2**  
**Test Characteristic Curve: Spek 4-5 B/C S503 Paper**

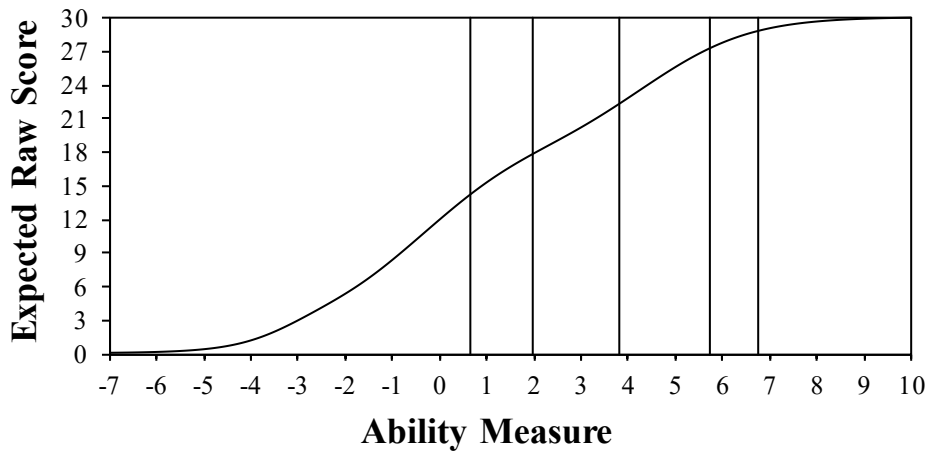


2.8.4.5 Grades 6-8

**Figure 2.8.4.5.1**  
**Test Characteristic Curve: Spek 6-8 A S503 Paper**

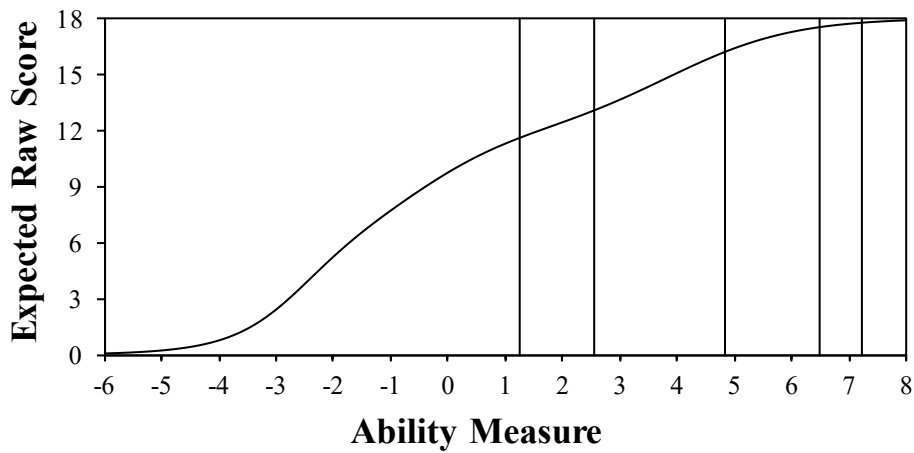


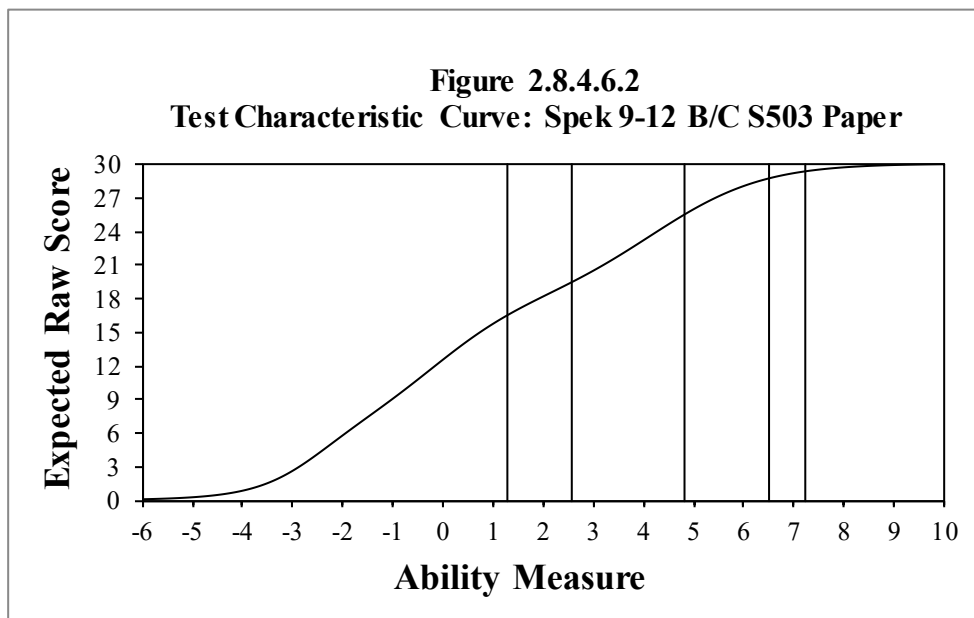
**Figure 2.8.4.5.2**  
**Test Characteristic Curve: Spek 6-8 B/C S503 Paper**



*2.8.4.6 Grades 9-12*

**Figure 2.8.4.6.1**  
**Test Characteristic Curve: Spek 9-12 A S503 Paper**





## 2.9 Test Information Function

With the Rasch measurement model, as with any measurement model following item response theory, one can use the item information function (Lord, 1980) to model the relationship between the ability measure (in logits) and the accuracy of the ability measure by item. The item information function indicates the amount of information we have about the ability estimate provided by the item, as a function of the ability level. The more information we have about the ability estimate, the more certain or confident we are about the ability estimate. If the amount of information is large, we can estimate the student whose true ability is at that level with a higher degree of certainty, and all the estimates will be close to the true values. Conversely, if the amount of information is small, we can estimate the student whose true ability is at that level with a lower degree of certainty, and estimates will be further away from the true values.

The **item/task information function** indicates the amount of information student responses to that item (or task) provides to help reduce our uncertainty regarding a student's true ability measure. The more information we have about the ability measure, the more certain or confident we can be in that estimate of the student's ability. If the amount of information is large, that means that we have estimated with a higher degree of certainty a student whose true ability is at that level. Therefore, the ability measures for students whose scores lie within that region of the ability continuum will be close to their true values. Conversely, if the amount of information is small, that means that we have estimated with a lower degree of certainty the student whose true ability is at that level. Consequently, the ability measures for students whose scores lie within that region of the ability continuum will be further away from their true values.

Mathematically, the amount of item information at a given ability level is the reciprocal of the variance of the ability estimate at the level for the item. In other words, item information value is the inverse squared of the standard errors of measurement of a given ability measure for the item. Therefore, for that item (or

task), the information value also provides information about the precision of the ability measure along the ability continuum.

The **test information function** (TIF) aggregates the item information functions across all the items on the test form or item pool. Since the item information value is the inverse squared of the standard errors of measurement of a given ability measure for the item, the test information value reflects the standard errors of measurement of a given ability level for the test. When the TIF is presented graphically as the test information curve, it shows how well the test is measuring across the continuum of student ability in terms of the amount of information, certainty, or the amount of measurement precision the test provides at each ability level. The higher the curve, the more information the test provides at the ability level.

Since the TIF is the sum of all item/task information functions on the test form (Lord, 1980), the TIF depends on the information functions (Lord, 1980) of the individual items/tasks included on the test form or in the item pool. The shape of the test information curve depends on several factors, including the number and characteristics of items/tasks, the item response theory used, and the values of the item/task parameters. With some exceptions, there is a general pattern to the shape of test information curves. Test information curves peak in the region of the student ability continuum where the test provides higher discrimination and more precise measurement as compared to other regions where the curve is less peaked, normally at the lower and upper ends of the ability continuum. When the test form consists of multiple-choice items such as on the Listening and Reading domains, the test information curve is usually unimodal.

The parameter values for the individual categories on the scoring tools that raters use to evaluate students' responses to the tasks, in addition to the factors mentioned earlier, affect the shape of the test information curves for the Writing and Speaking tests. Accordingly, some refer to these test information curves as "category information functions" (Engelhard & Wind, 2018). The scoring scales that the raters use have more score categories than the scoring schemes used for evaluating students' responses to multiple-choice items, which typically have just two categories—"right" or "wrong." Additionally, we designed the scoring scales to measure a wide range of student performance on a task. Consequently, the resulting adjacent score category boundaries may not be equidistant, and, indeed, in some cases, they may even be far apart if raters assign few scores in certain categories. In this situation, a test information curve will have one (or more) dips in the region(s) between the adjacent score category boundaries, indicating the loss of information in the corresponding ability range(s) and a decrease in the amount of information that certain score categories provide (Engelhard & Wind, 2018). Therefore, the shape of a test information curve for an ACCESS Writing or Speaking test may not be unimodal and instead may have two (or more) peaks. For example, suppose that a test information curve reveals a dip in the region of the student writing ability continuum where raters would have assigned a score of 3. That suggests that students who received a score of 3 may have displayed potentially substantively meaningful differences in writing ability that the raters were not able to adequately distinguish when they used the 9-point Writing Scoring Scale to assign scores or, alternatively, that the score categories did not describe salient characteristics of students' writing that would make it possible for the raters to distinguish reliability among the students' responses in that region of the student ability continuum (Engelhard & Wind, 2018, pp. 316-319). The ACCESS Writing and Speaking tests are not the only assessments that have test information curves with these unusual shapes. The test information curves for other tests composed of open-ended tasks, such as the National Assessment of Educational Progress Writing assessment, also show a similar "dipping" pattern (Muraki, 1993).

The figures in this section plot the TIFs and show graphically the amount of information that the test provided across the continuum of student ability. For each test form, the five vertical lines in the figure

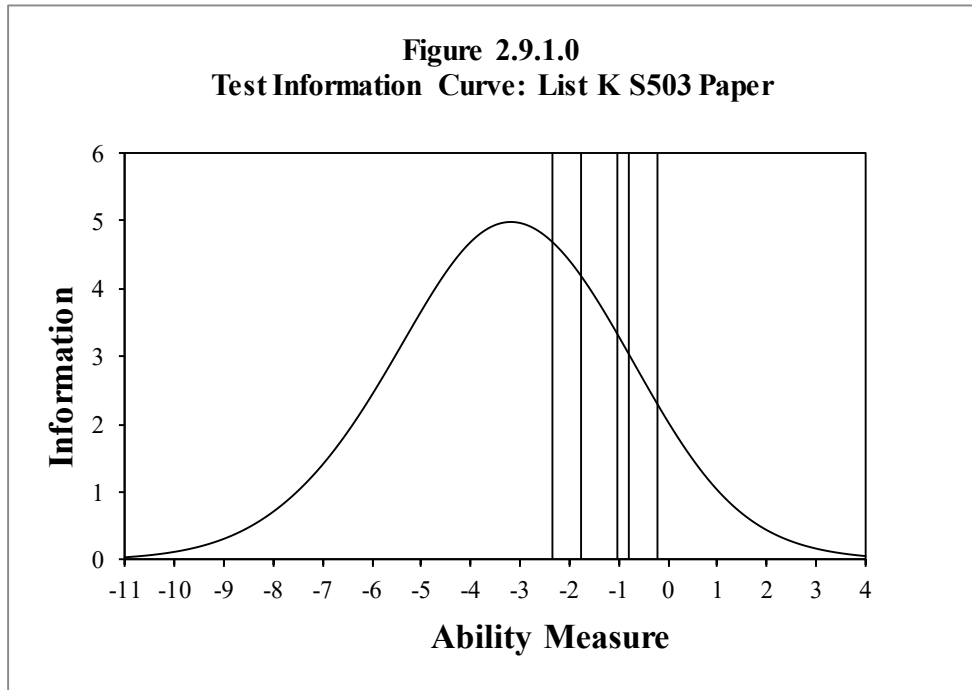
indicate the ACCESS cut scores for the highest grade in each grade-level cluster, dividing the figure into six sections denoting the WIDA proficiency levels (PL 1– PL 6) for the domain. The test information curve and the corresponding ACCESS cut-score lines are both expressed on the ACCESS logit scale. Note that for the Speaking test, in Tier Pre-A, all scores are within the PL 1.0 range, so for some graphs there are no vertical lines showing the cut scores between proficiency levels.

Inclusion of the ACCESS cut-score lines in these figures are meant only to facilitate the visual interpretation of the test information curves relative to the ACCESS cut scores by domains. These lines provide a benchmark for WIDA and CAL assessment experts to examine the ability range for which each test seems to be more (or less) accurate in estimating student ability. Readers should note that most states that use ACCESS for ELLs do not make reclassification decisions based solely on students' domain scale scores. Rather, the majority of these states set their reclassification (or exit) criterion based on a student's Overall composite scale score, which is a weighted sum of a student's four domain scale scores. Only a few states use those four domain scale scores in addition to the student's Overall composite scale score when making a reclassification decision. Therefore, from the WIDA policy perspective, it is more important to ensure that we minimize the measurement error near the cut score that most states use to set their reclassification criterion on the Overall composite scale score. We report the conditional standard errors of measurement (CSEMs) for the Overall composite scale scores in Section 5.6.

In addition to the TIF graphs by tier, we provide plots of the TIFs across tiers, by grade cluster, in the same graph. Test users may find it useful to compare the ability ranges across tiers, noting for each tier where the curve displays its highest peaks (i.e., where the most measurement information is provided). For example, as shown in Figure 2.9.3.1.3, the test information curve across tiers for Writing Grade 1 reveals that the Writing Grade 1 Tier A form provided more information about student ability measures that were either just below the PL 2 cut score or just below the PL 5 cut score. By contrast, the Writing Grade 1 Tier B/C form provided more information about the student ability measures that were either just above the PL 2 cut score or just above the PL 5 cut score. The plot also shows that the Writing Grade 1 Tier A form provided more information for those student ability measures in the lowest range (i.e., ability measures of -0.5 logits or lower), while the Writing Grade 1 Tier B/C form provided more information than the Grade 1 Tier A form for the rest of the student ability measures, especially those in the higher ability range. Lastly, consistent with the purposes of the test design, there is also considerable overlap between the ranges of writing ability that the two forms cover.

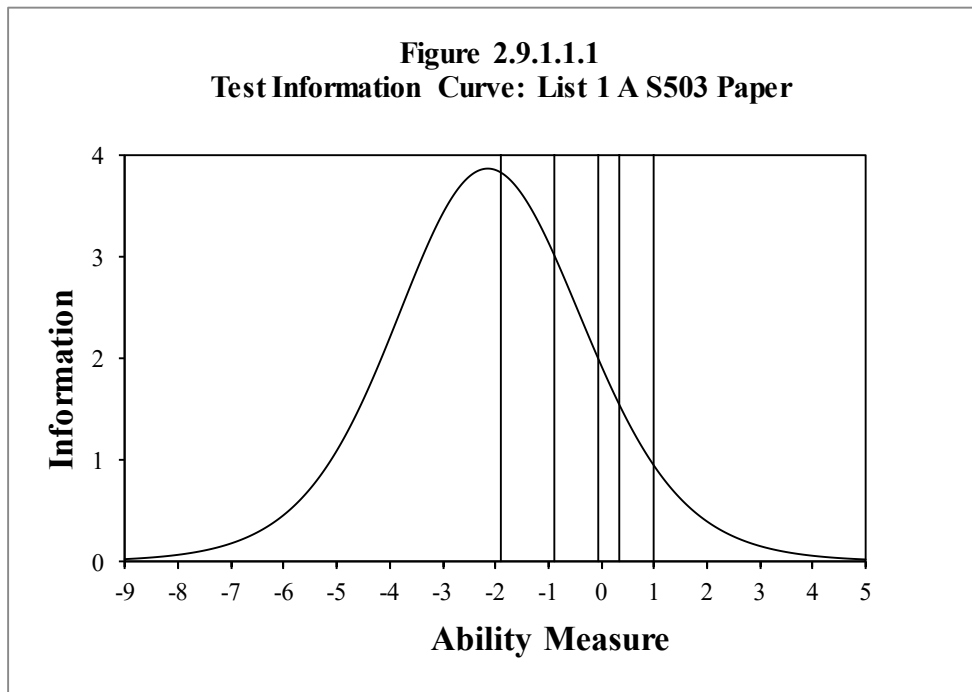
## 2.9.1 Listening

### 2.9.1.0 Kindergarten

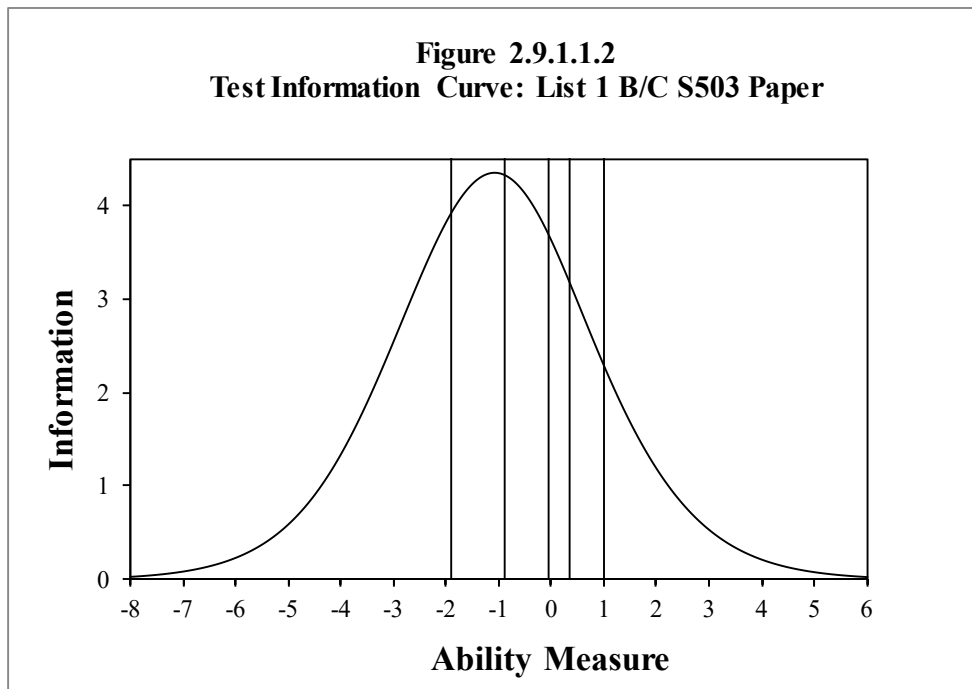




2.9.1.1 Grade 1

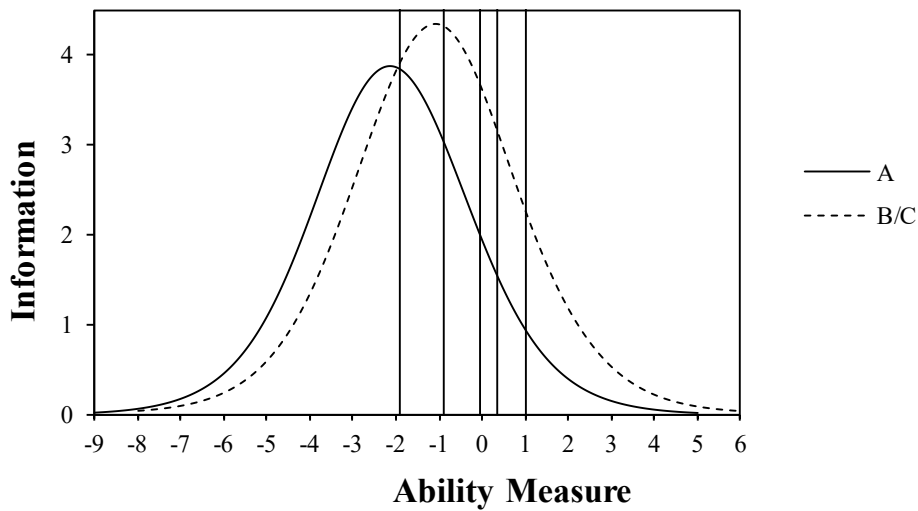


Note: The test form is shared between 1A and 2A.



Note: The test form is shared between 1B/C and 2B/C.

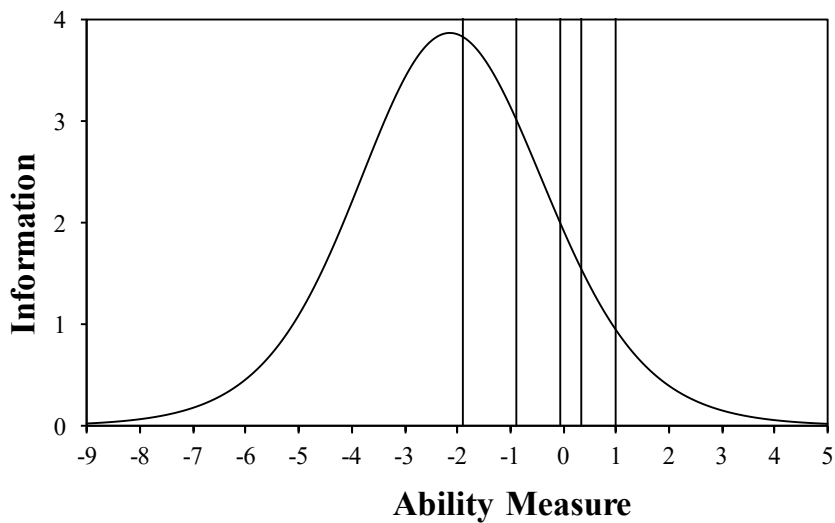
**Figure 2.9.1.1.3**  
**Test Information Curve: List 1 S503 Paper**



Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

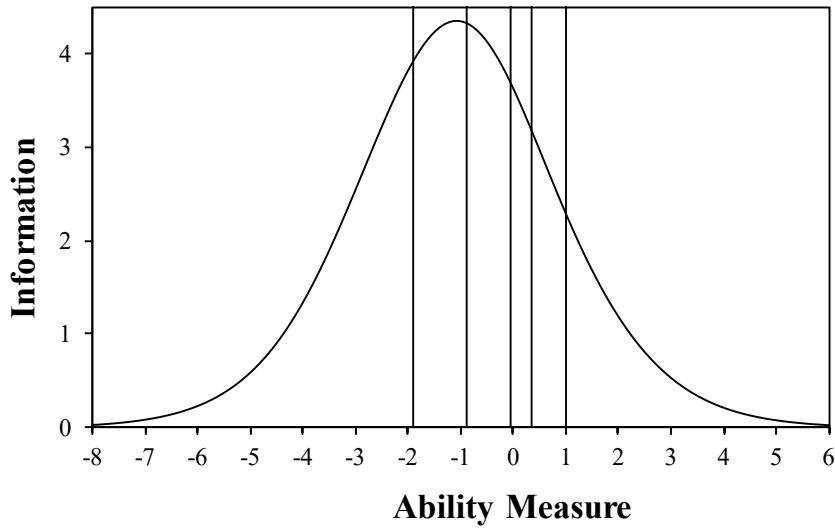
#### 2.9.1.2 Grade 2

**Figure 2.9.1.2.1**  
**Test Information Curve: List 2 A S503 Paper**



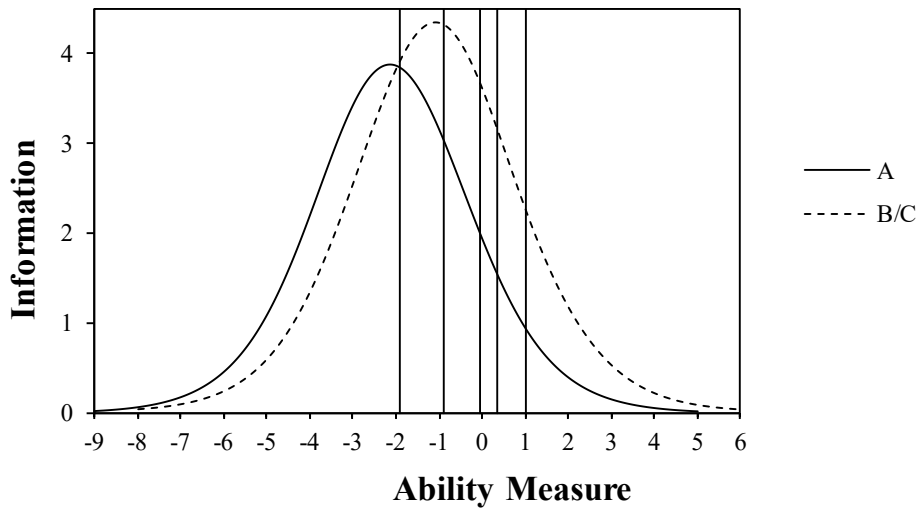
Note: The test form is shared between 1A and 2A.

**Figure 2.9.1.2.2**  
**Test Information Curve: List 2 B/C S503 Paper**



Note: The test form is shared between 1B/C and 2B/C.

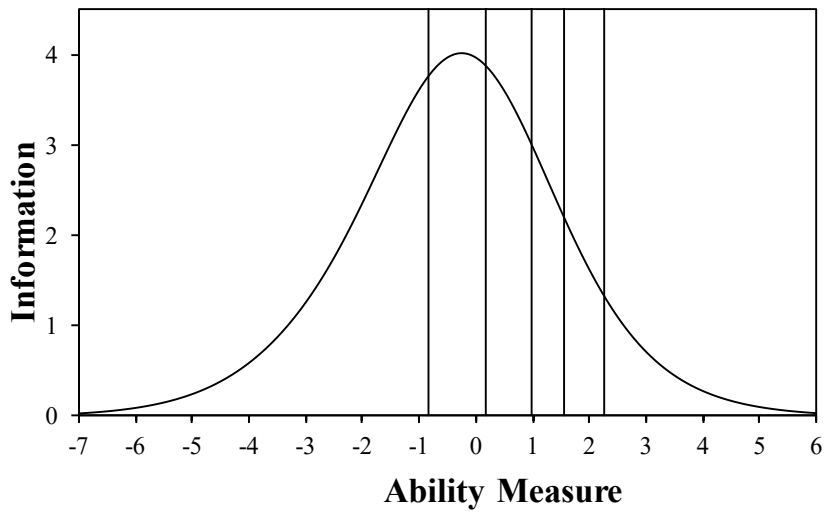
**Figure 2.9.1.2.3**  
**Test Information Curve: List 2 S503 Paper**



Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

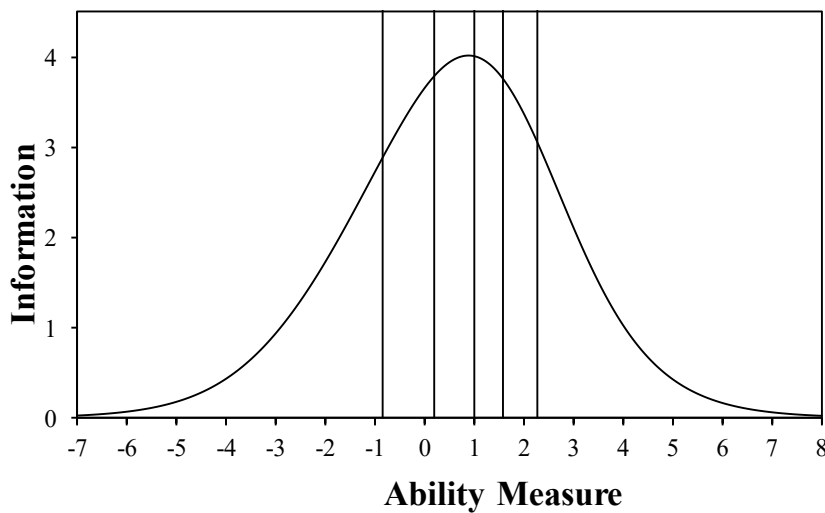
2.9.1.3 Grade 3

**Figure 2.9.1.3.1**  
**Test Information Curve: List 3 A S503 Paper**



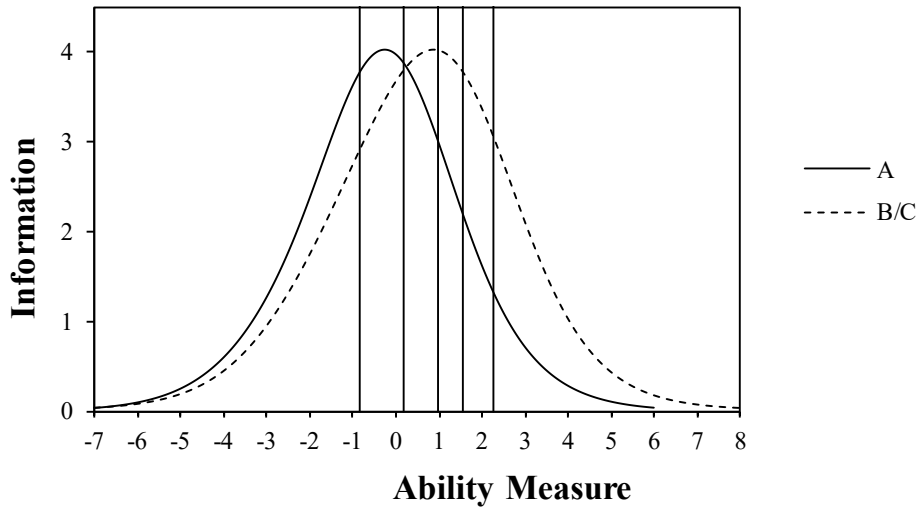
Note: The test form is shared between 3A and 4-5A.

**Figure 2.9.1.3.2**  
**Test Information Curve: List 3 B/C S503 Paper**



Note: The test form is shared between 3B/C and 4-5B/C.

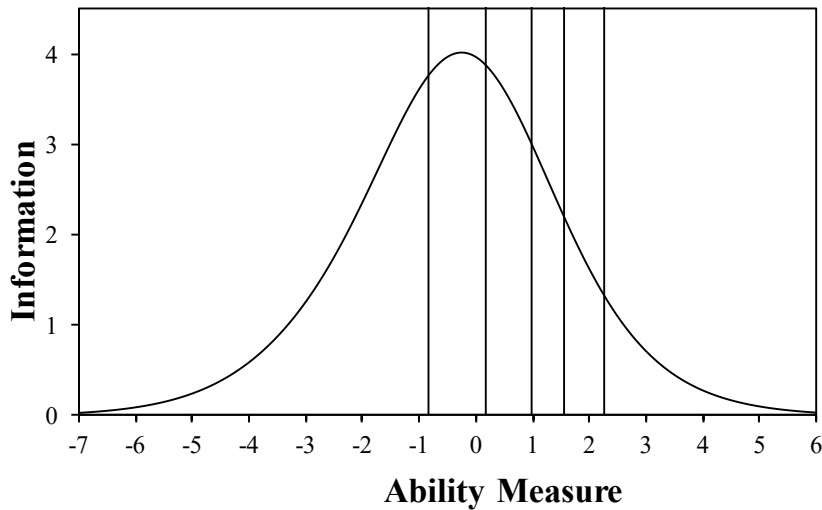
**Figure 2.9.1.3.3**  
**Test Information Curve: List 3 S503 Paper**



Note: The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

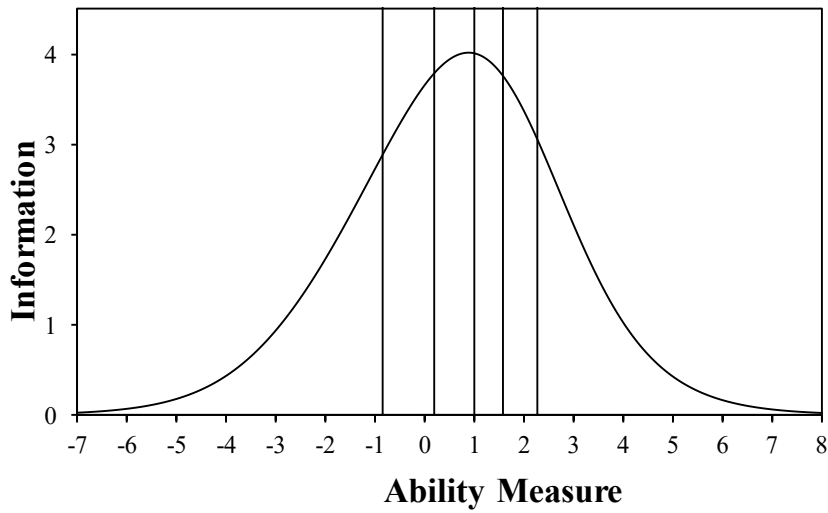
#### 2.9.1.4 Grades 4-5

**Figure 2.9.1.4.1**  
**Test Information Curve: List 4-5 A S503 Paper**



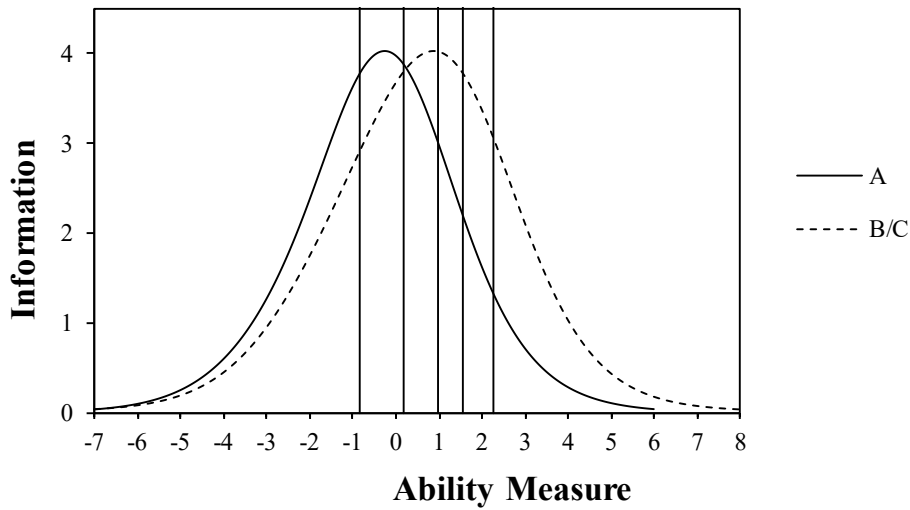
Note: The test form is shared between 3A and 4-5A.

**Figure 2.9.1.4.2**  
**Test Information Curve: List 4-5 B/C S503 Paper**



Note: The test form is shared between 3B/C and 4-5B/C.

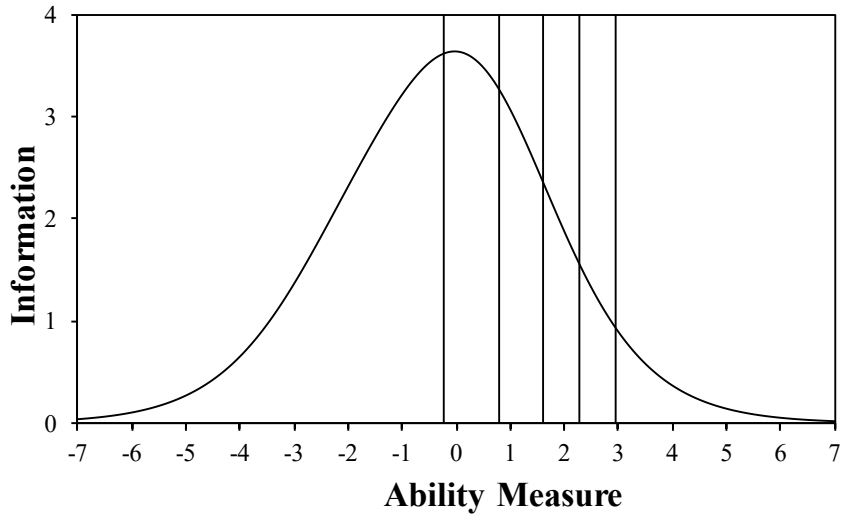
**Figure 2.9.1.4.3**  
**Test Information Curve: List 4-5 S503 Paper**



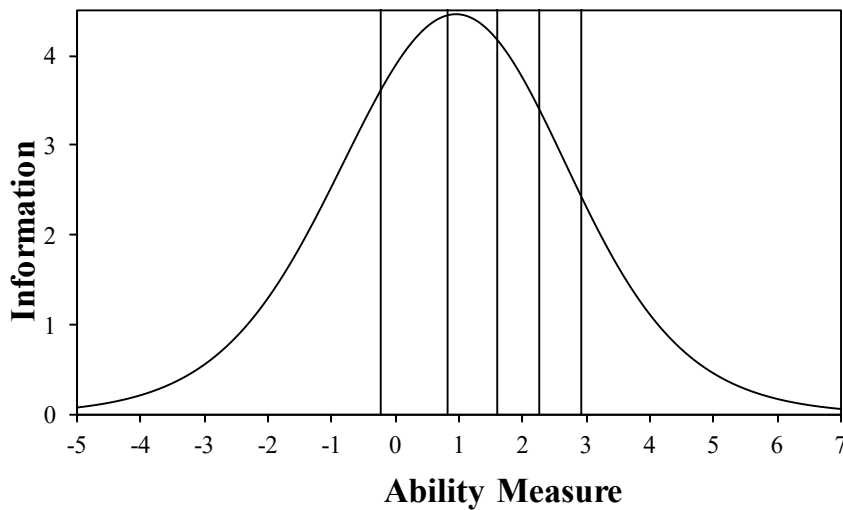
Note: The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

2.9.1.5 Grades 6-8

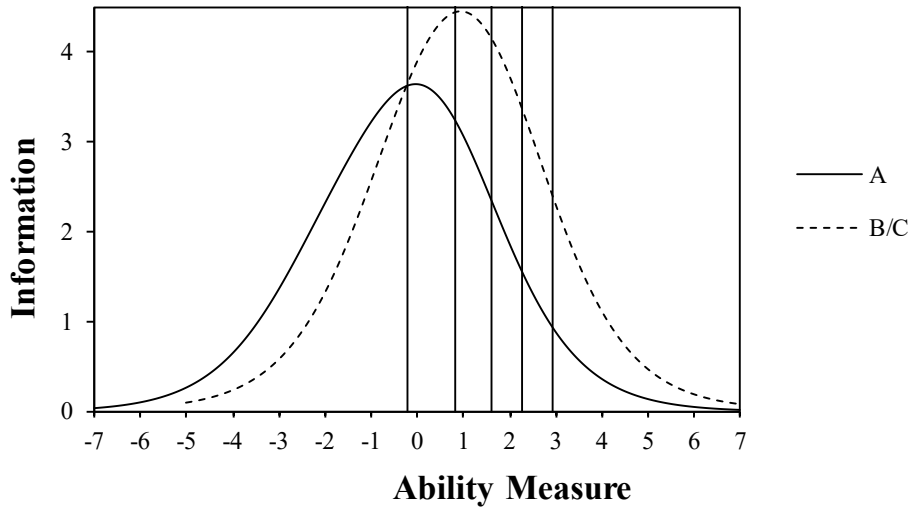
**Figure 2.9.1.5.1**  
**Test Information Curve: List 6-8 A S503 Paper**



**Figure 2.9.1.5.2**  
**Test Information Curve: List 6-8 B/C S503 Paper**

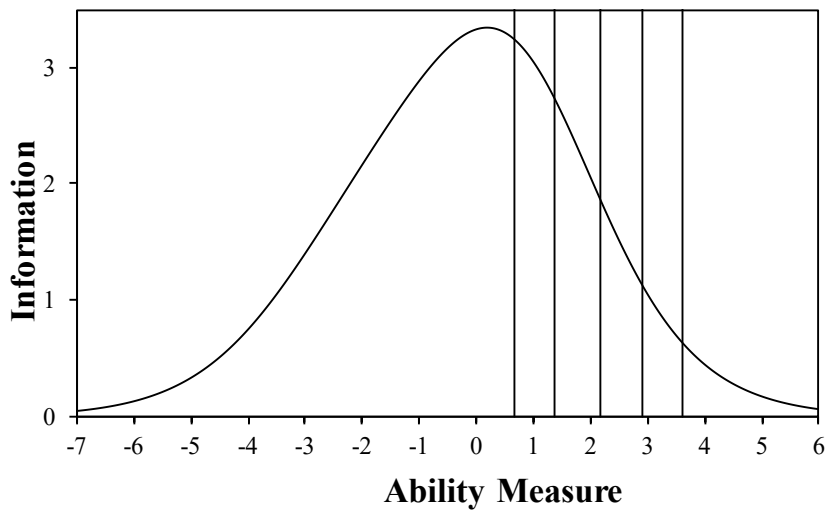


**Figure 2.9.1.5.3**  
**Test Information Curve: List 6-8 S503 Paper**



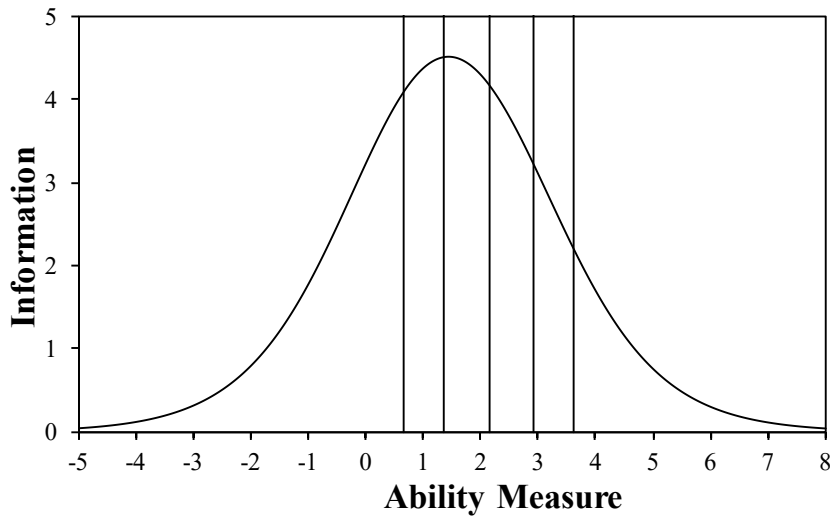
2.9.1.6 Grades 9-12

**Figure 2.9.1.6.1**  
**Test Information Curve: List 9-12 A S503 Paper**

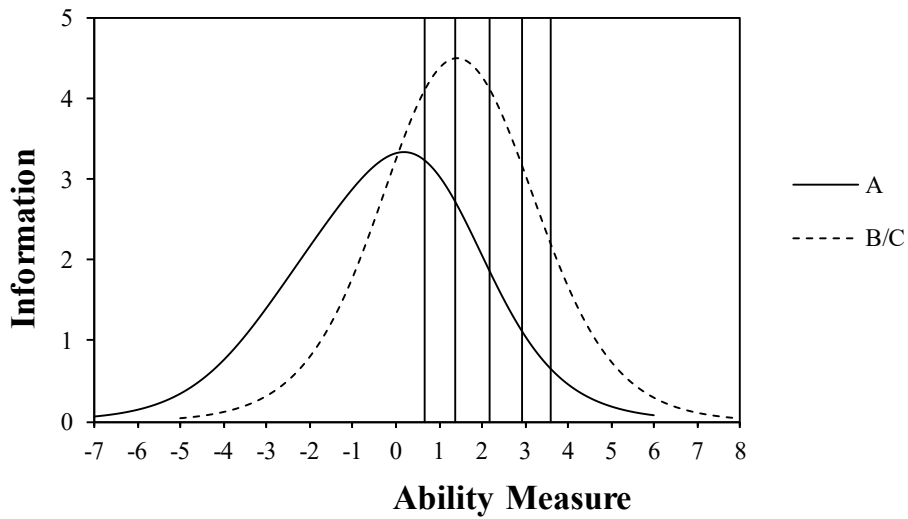




**Figure 2.9.1.6.2**  
**Test Information Curve: List 9-12 B/C S503 Paper**

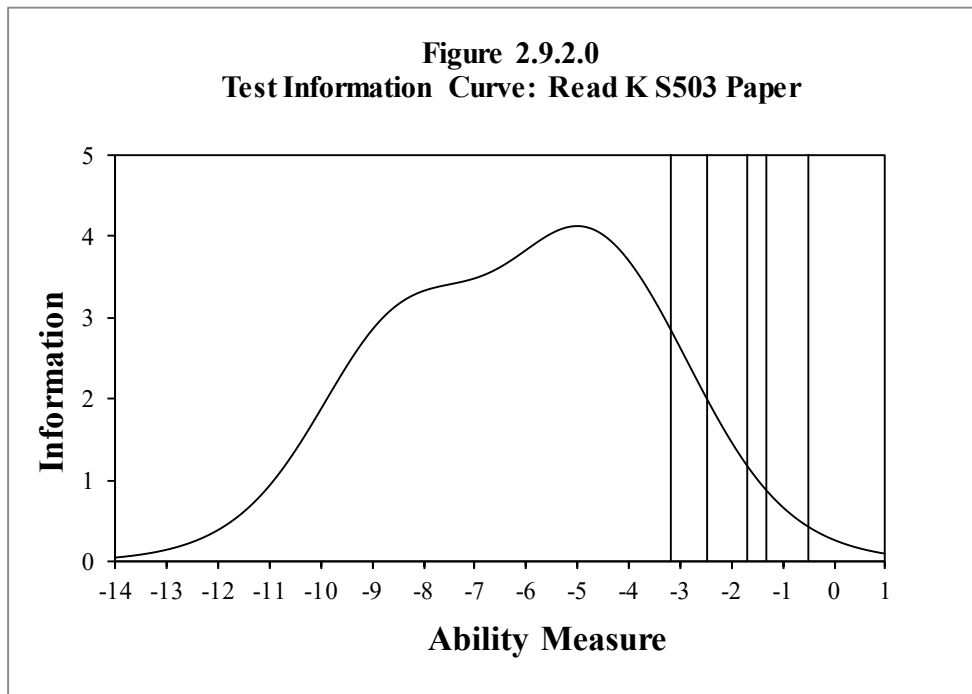


**Figure 2.9.1.6.3**  
**Test Information Curve: List 9-12 S503 Paper**

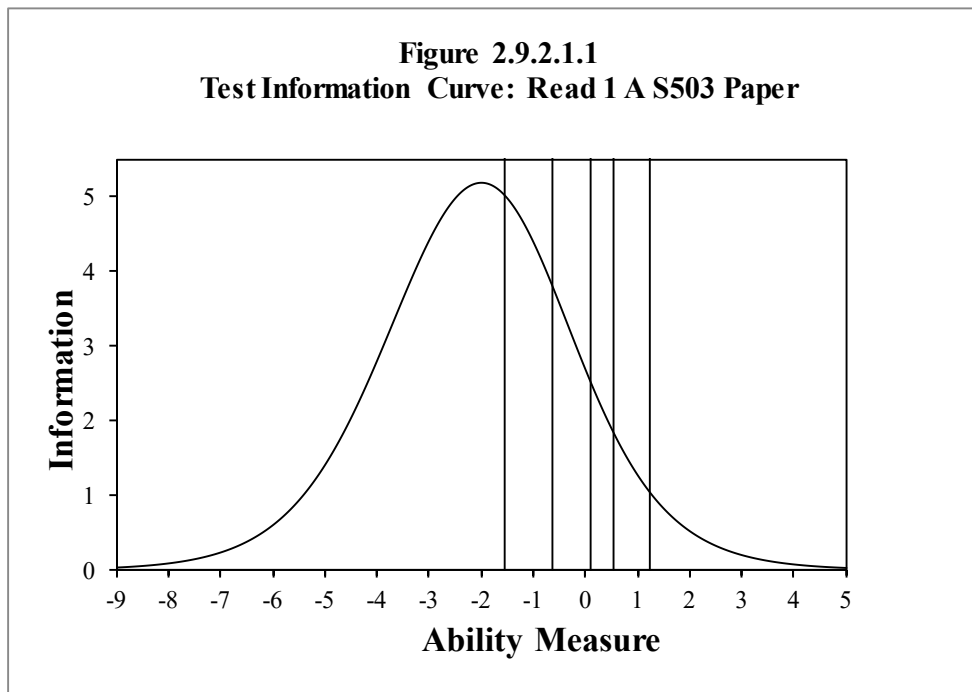


## 2.9.2 Reading

### 2.9.2.0 Kindergarten

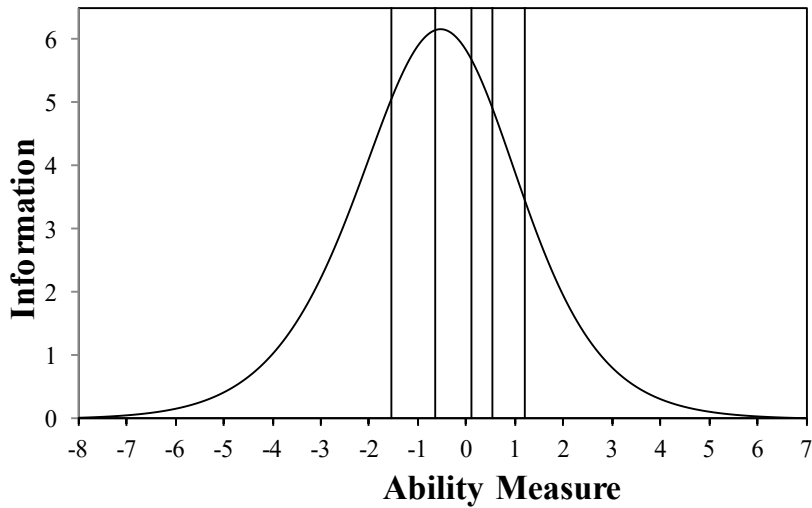


### 2.9.2.1 Grade 1



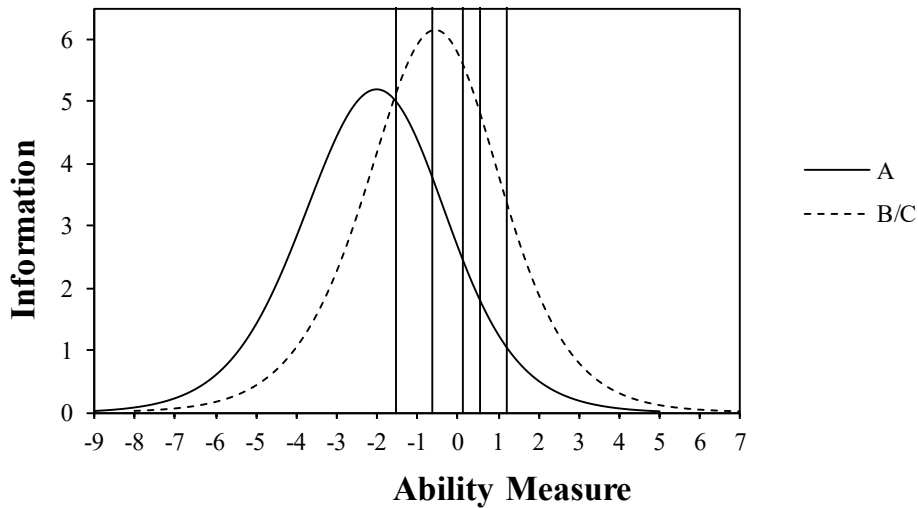
Note: The test form is shared between 1A and 2A.

**Figure 2.9.2.1.2**  
**Test Information Curve: Read 1 B/C S503 Paper**



Note: The test form is shared between 1B/C and 2B/C.

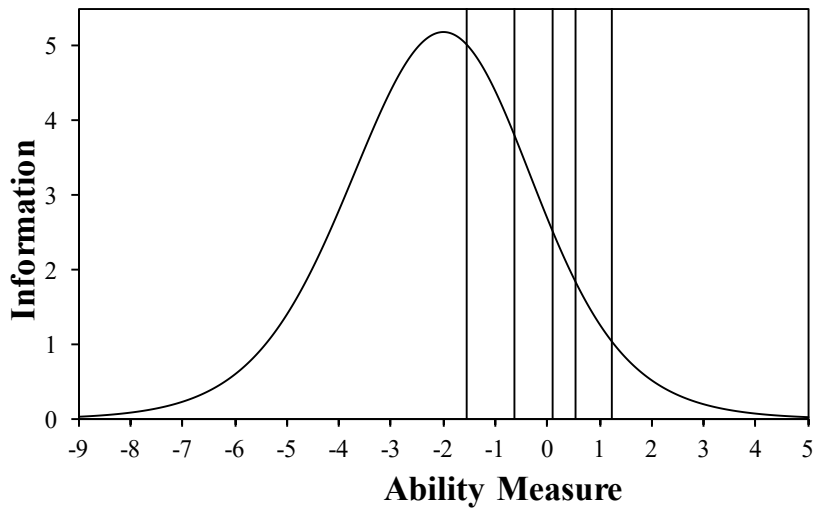
**Figure 2.9.2.1.3**  
**Test Information Curve: Read 1 S503 Paper**



Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

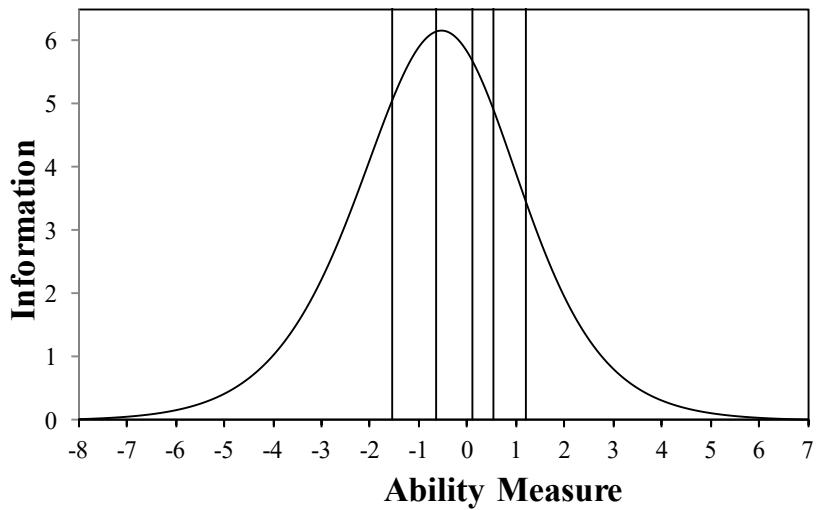
2.9.2.2 Grade 2

**Figure 2.9.2.2.1**  
**Test Information Curve: Read 2 A S503 Paper**



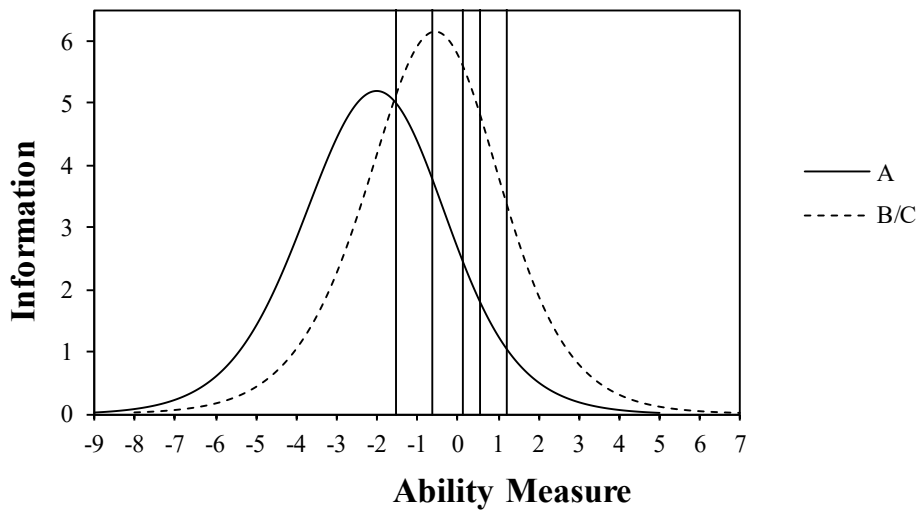
Note: The test form is shared between 1A and 2A.

**Figure 2.9.2.2.2**  
**Test Information Curve: Read 2 B/C S503 Paper**



Note: The test form is shared between 1B/C and 2B/C.

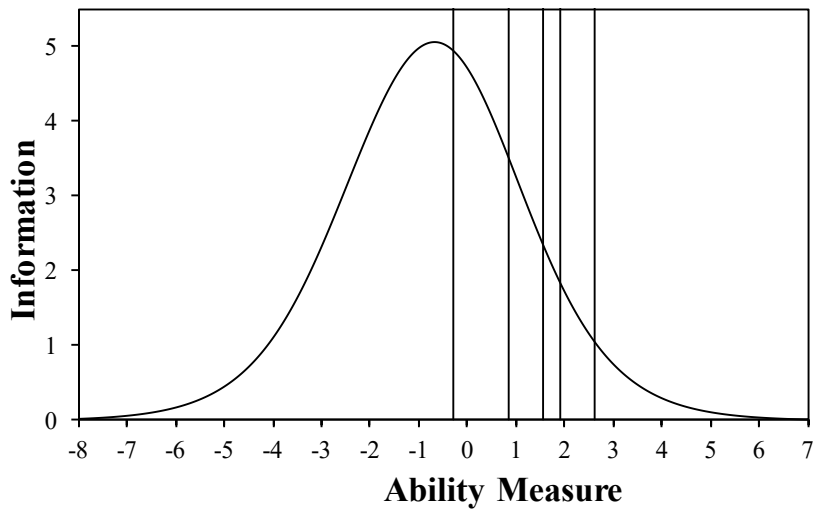
**Figure 2.9.2.2.3**  
**Test Information Curve: Read 2 S503 Paper**



Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

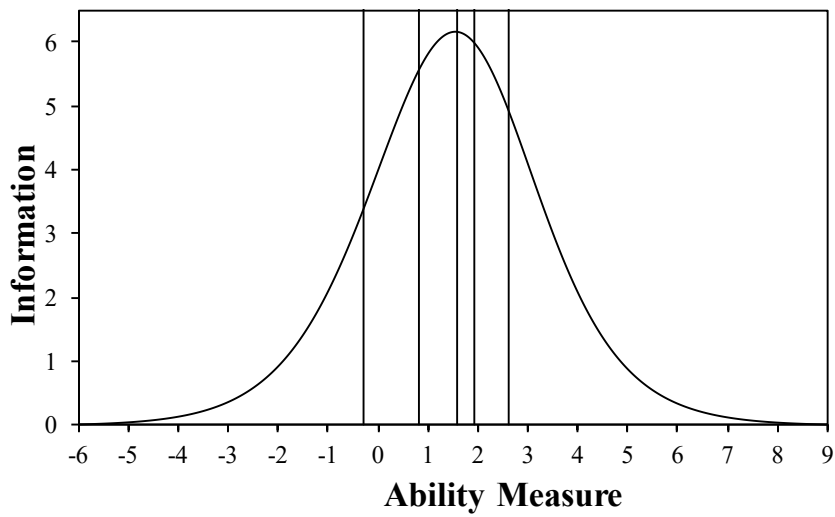
### 2.9.2.3 Grade 3

**Figure 2.9.2.3.1**  
**Test Information Curve: Read 3 A S503 Paper**



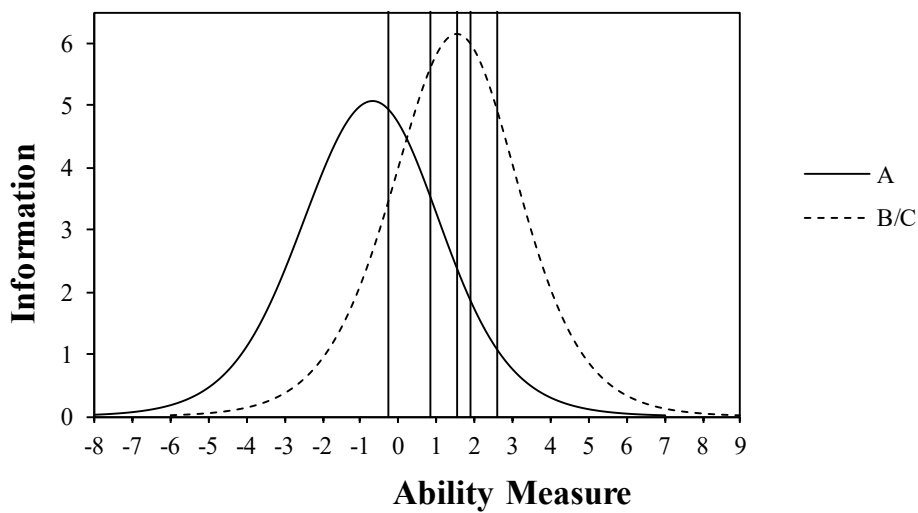
Note: The test form is shared between 3A and 4-5A.

**Figure 2.9.2.3.2**  
**Test Information Curve: Read 3 B/C S503 Paper**



Note: The test form is shared between 3B/C and 4-5B/C.

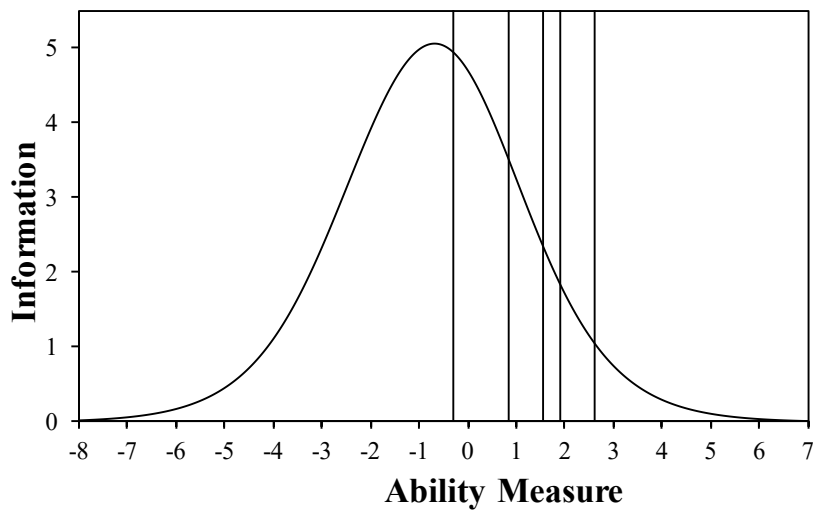
**Figure 2.9.2.3.3**  
**Test Information Curve: Read 3 S503 Paper**



Note: The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

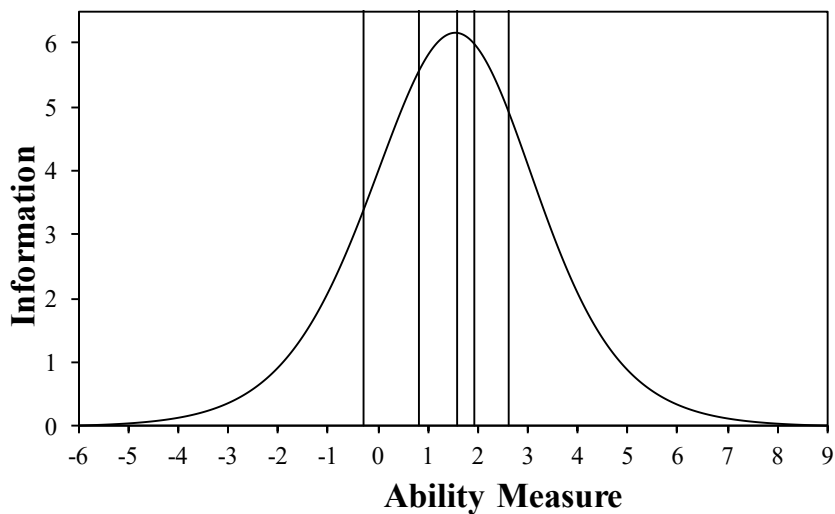
2.9.2.4 Grades 4-5

**Figure 2.9.2.4.1**  
**Test Information Curve: Read 4-5 A S503 Paper**



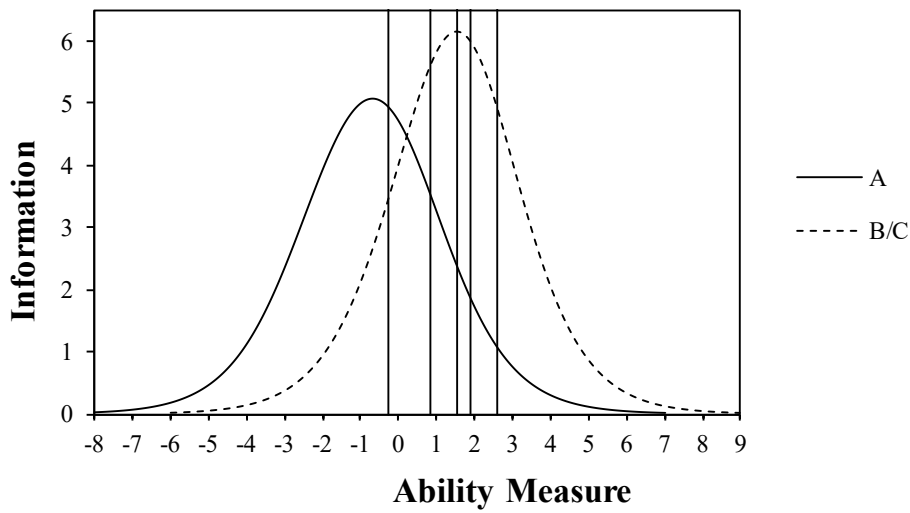
Note: The test form is shared between 3A and 4-5A.

**Figure 2.9.2.4.2**  
**Test Information Curve: Read 4-5 B/C S503 Paper**



Note: The test form is shared between 3B/C and 4-5B/C.

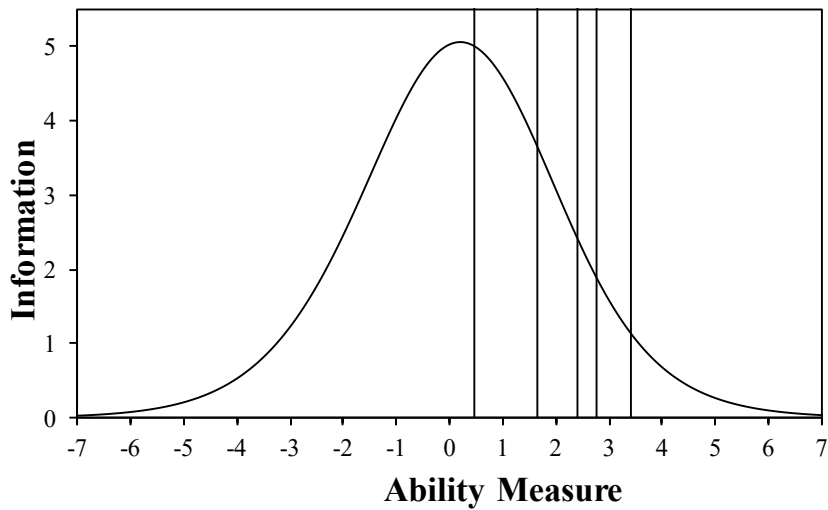
**Figure 2.9.2.4.3**  
**Test Information Curve: Read 4-5 S503 Paper**



Note: The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

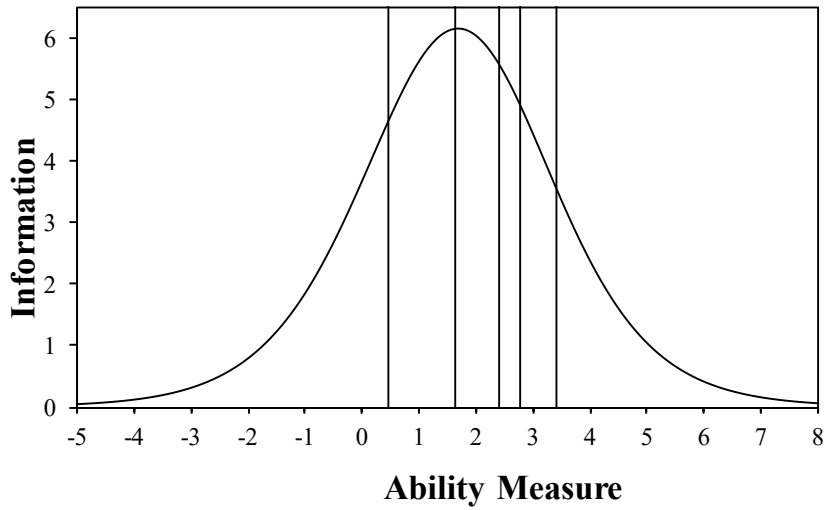
#### 2.9.2.5 Grades 6-8

**Figure 2.9.2.5.1**  
**Test Information Curve: Read 6-8 A S503 Paper**

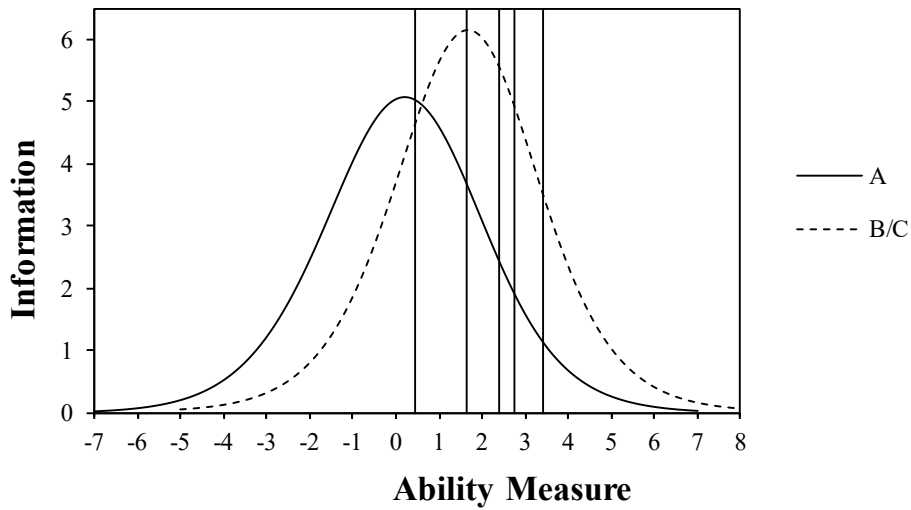




**Figure 2.9.2.5.2**  
**Test Information Curve: Read 6-8 B/C S503 Paper**

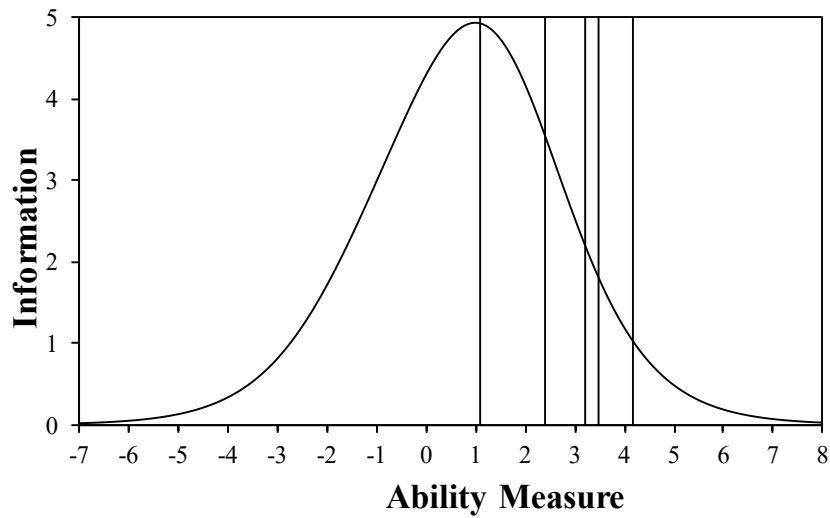


**Figure 2.9.2.5.3**  
**Test Information Curve: Read 6-8 S503 Paper**

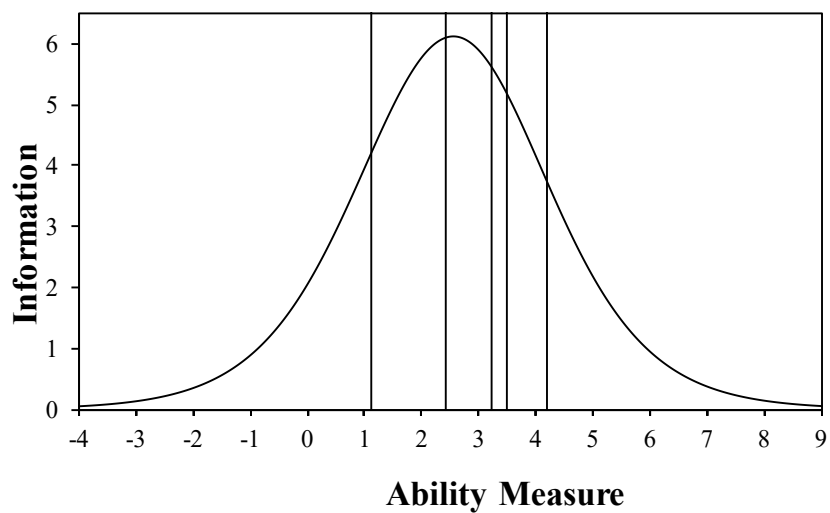


2.9.2.6 Grades 9-12

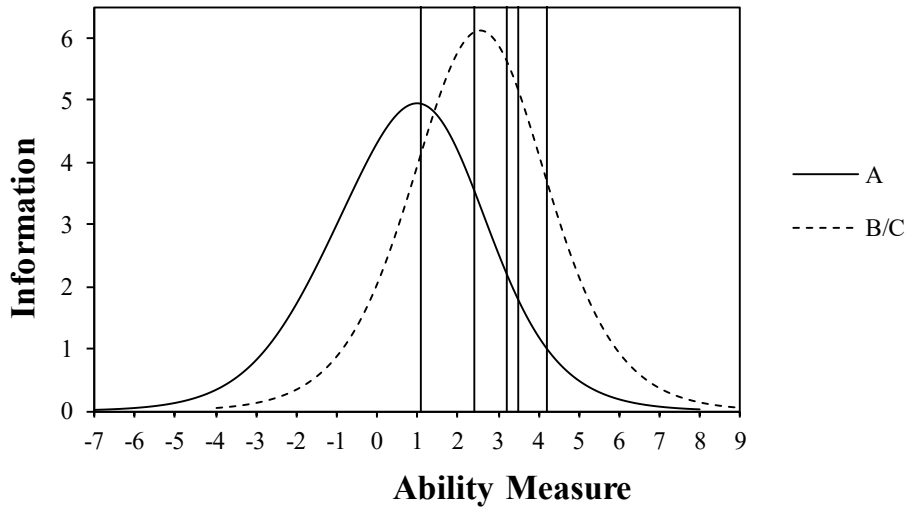
**Figure 2.9.2.6.1**  
**Test Information Curve: Read 9-12 A S503 Paper**



**Figure 2.9.2.6.2**  
**Test Information Curve: Read 9-12 B/C S503 Paper**



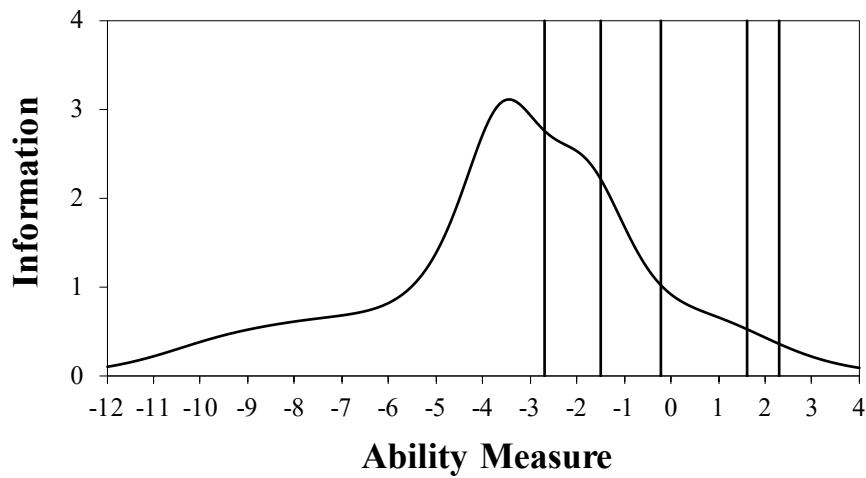
**Figure 2.9.2.6.3**  
**Test Information Curve: Read 9-12 S503 Paper**



## 2.9.3 Writing

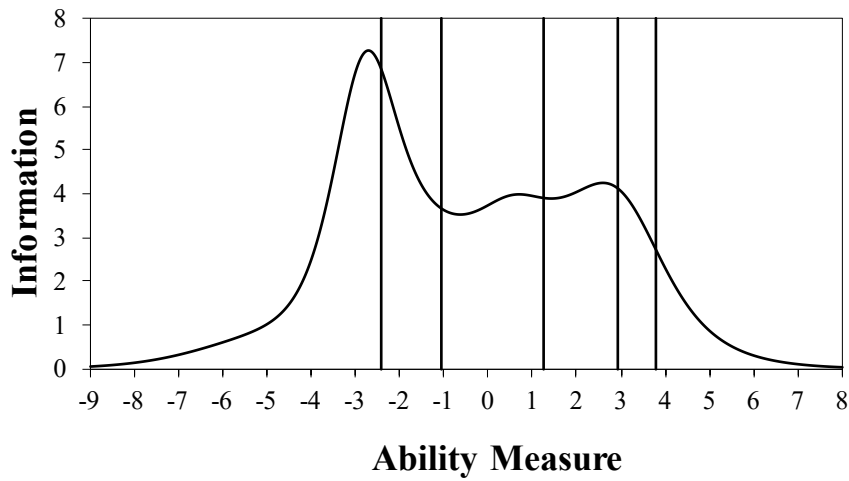
### 2.9.3.0 Kindergarten

**Figure 2.9.3.0**  
**Test Information Curve: Writ K S503 Paper**

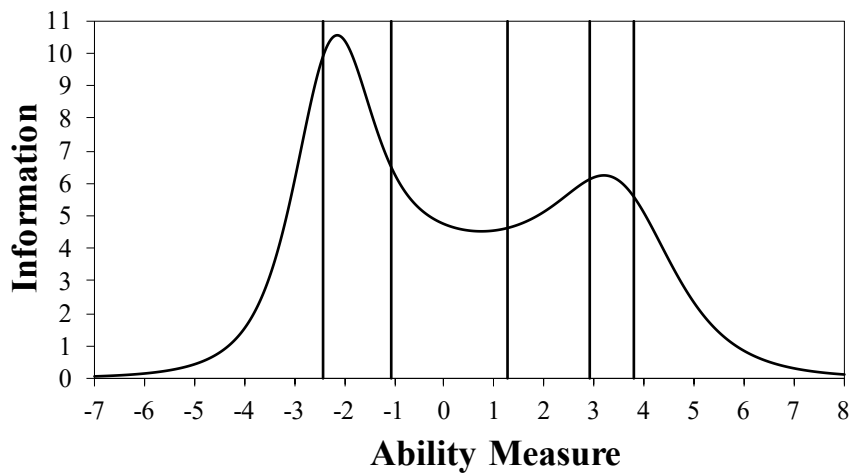


2.9.3.1 Grade 1

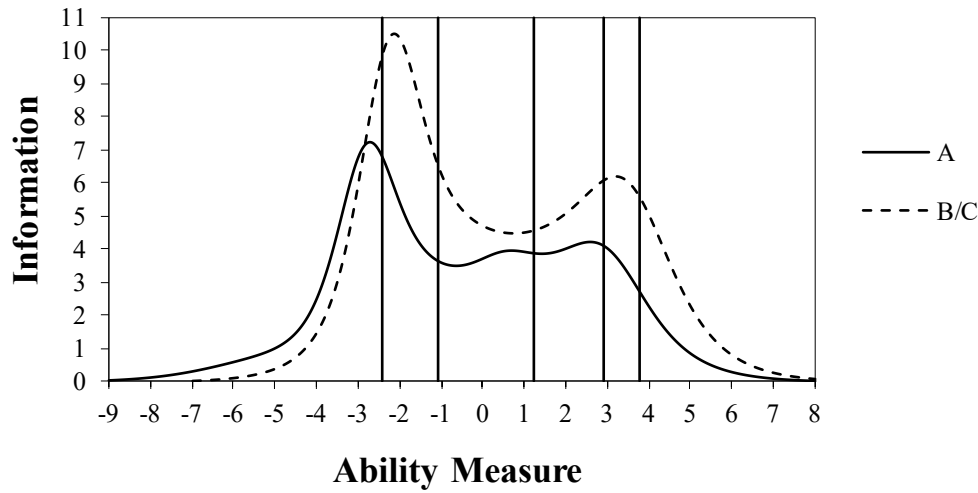
**Figure 2.9.3.1.1**  
**Test Information Curve: Writ 1 A S503 Paper**



**Figure 2.9.3.1.2**  
**Test Information Curve: Writ 1 B/C S503 Paper**

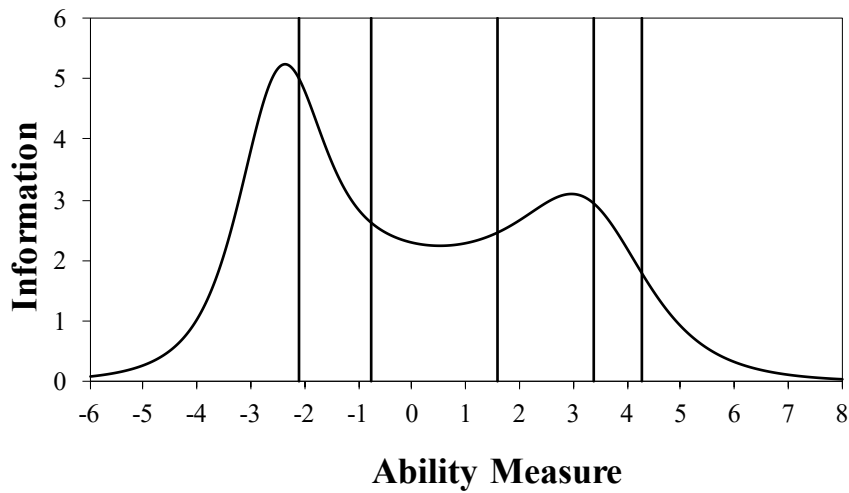


**Figure 2.9.3.1.3**  
**Test Information Curve: Writ 1 S503 Paper**



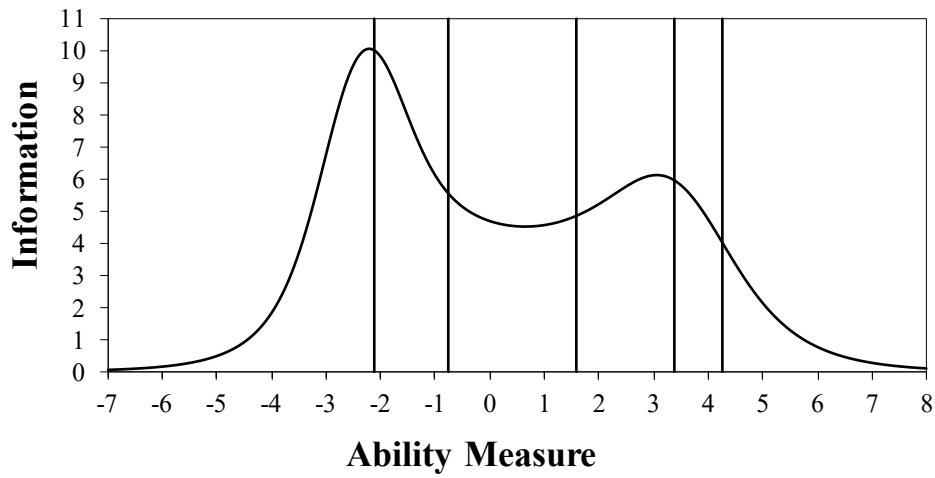
#### 2.9.3.2 Grade 2

**Figure 2.9.3.2.1**  
**Test Information Curve: Writ 2 A S503 Paper**



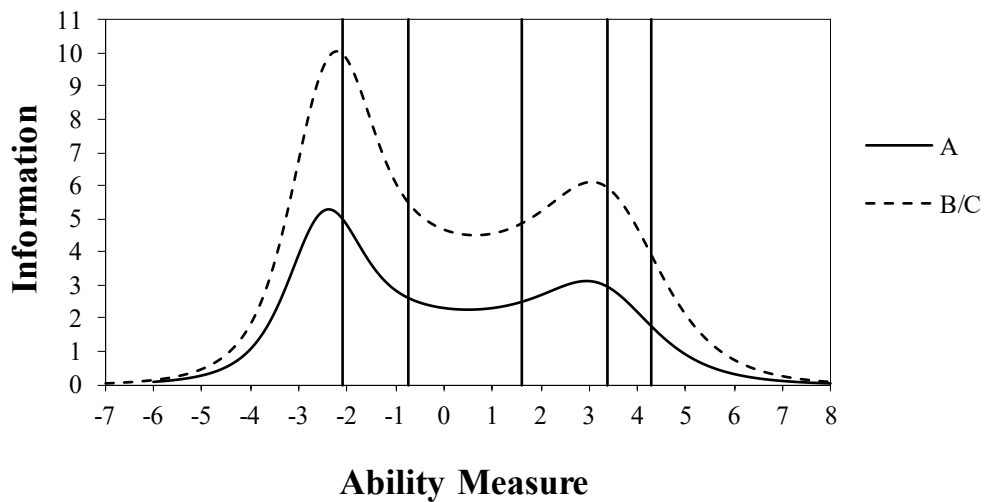
Note: The test form is shared between 2A and 3A.

**Figure 2.9.3.2.2**  
**Test Information Curve: Writ 2 B/C S503 Paper**



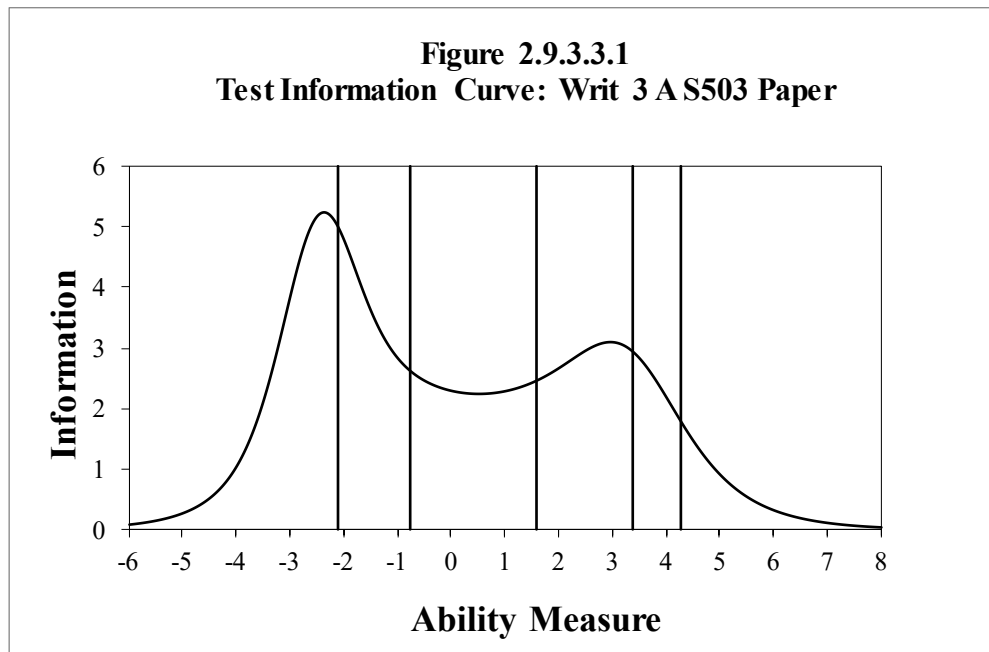
Note: The test form is shared between 2B/C and 3B/C.

**Figure 2.9.3.2.3**  
**Test Information Curve: Writ 2 S503 Paper**

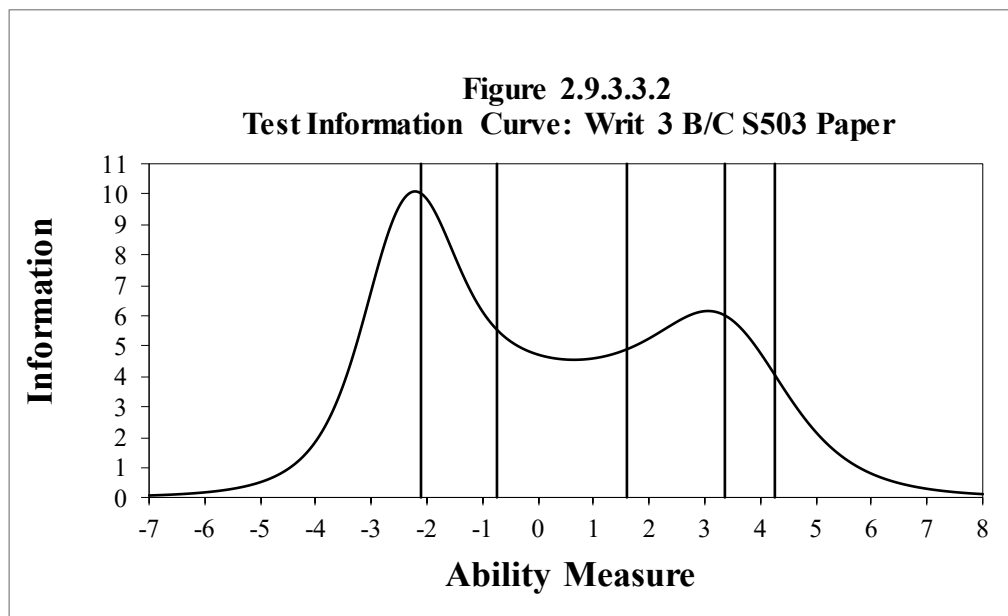


Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

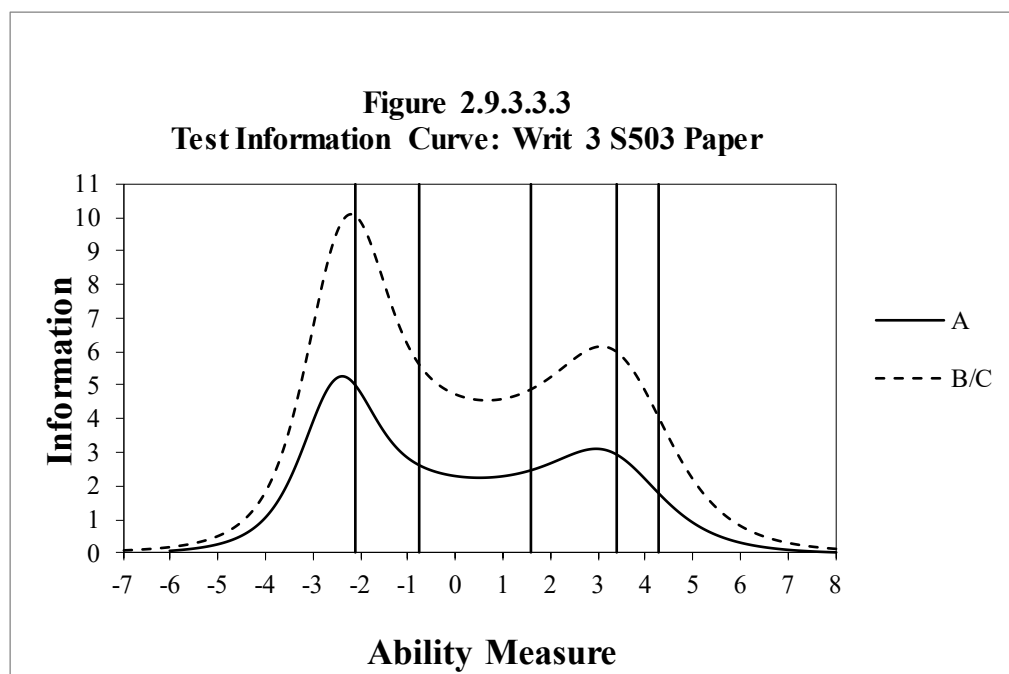
2.9.3.3 Grade 3



Note: The test form is shared between 2A and 3A.

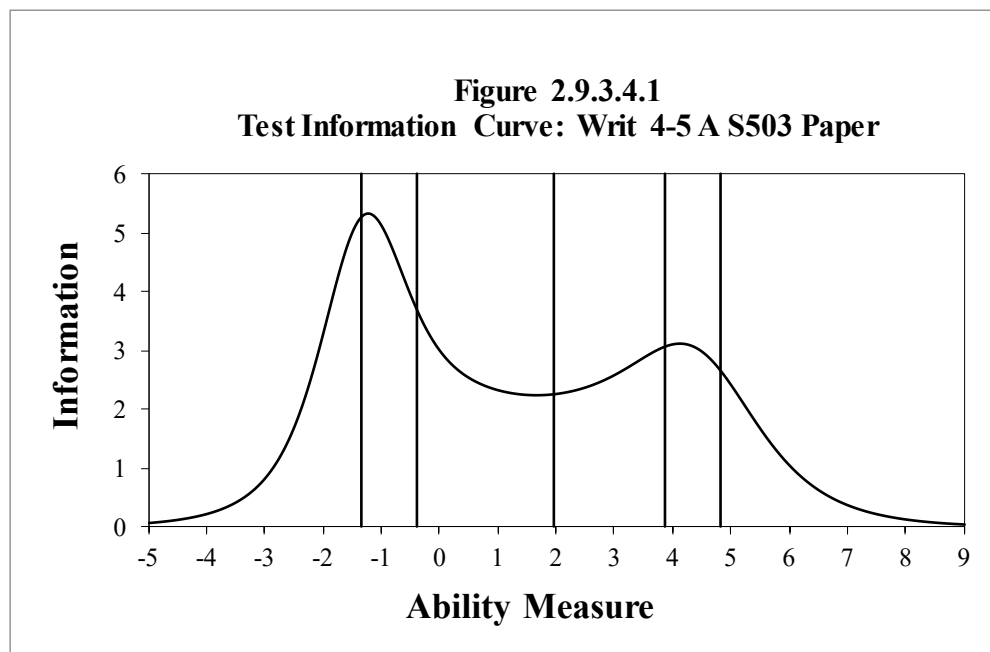


Note: The test form is shared between 2B/C and 3B/C.



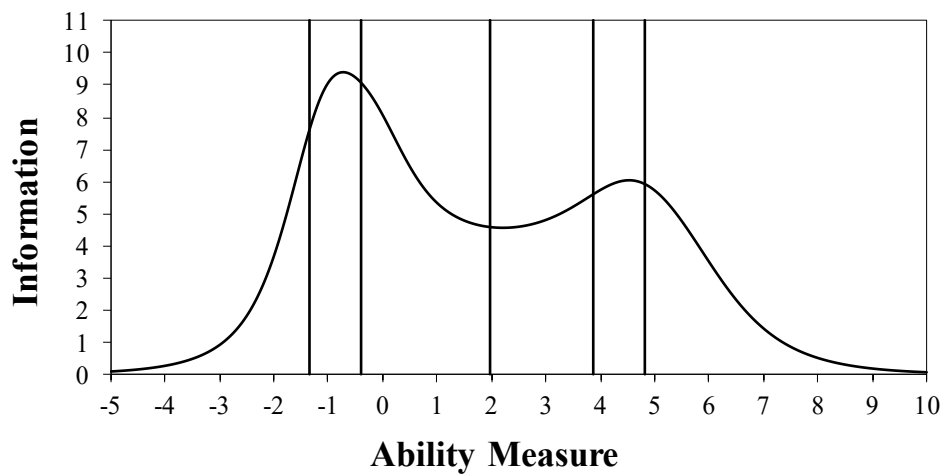
Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

#### 2.9.3.4 Grades 4-5

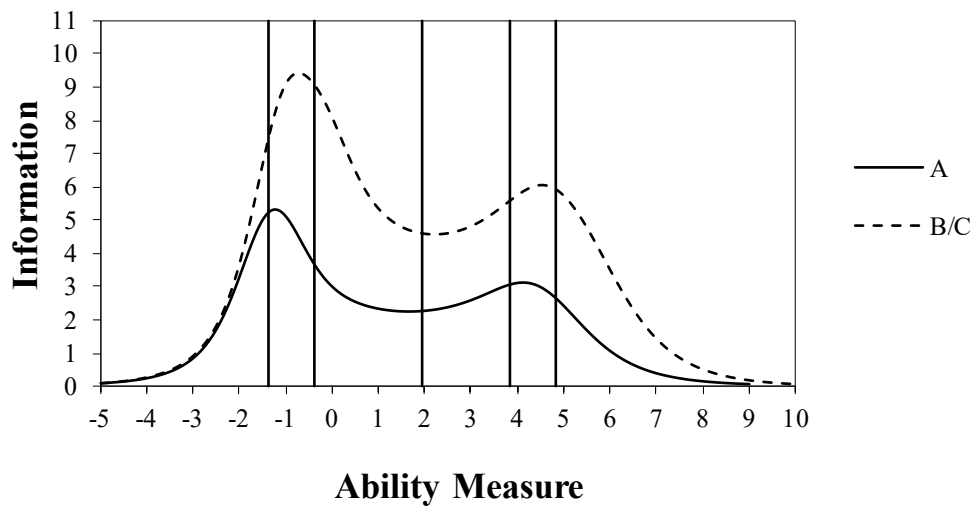




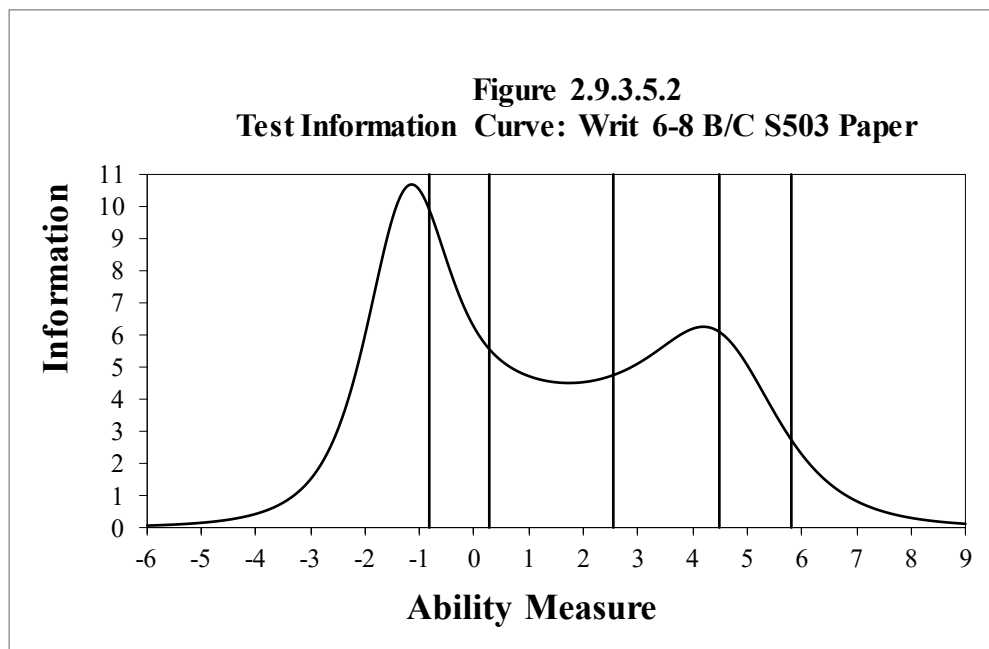
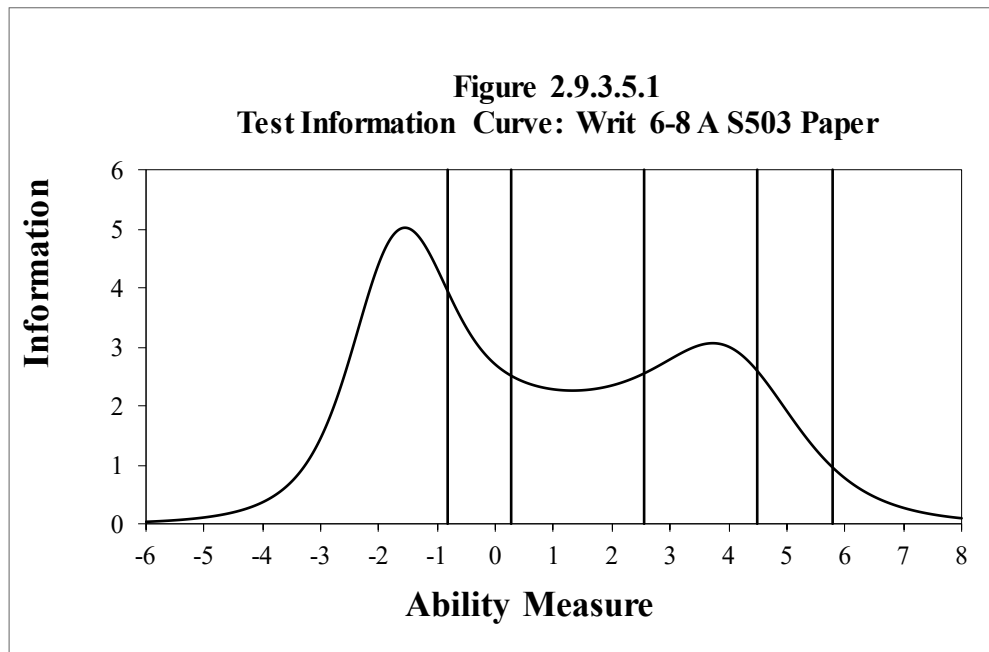
**Figure 2.9.3.4.2**  
**Test Information Curve: Writ 4-5 B/C S503 Paper**

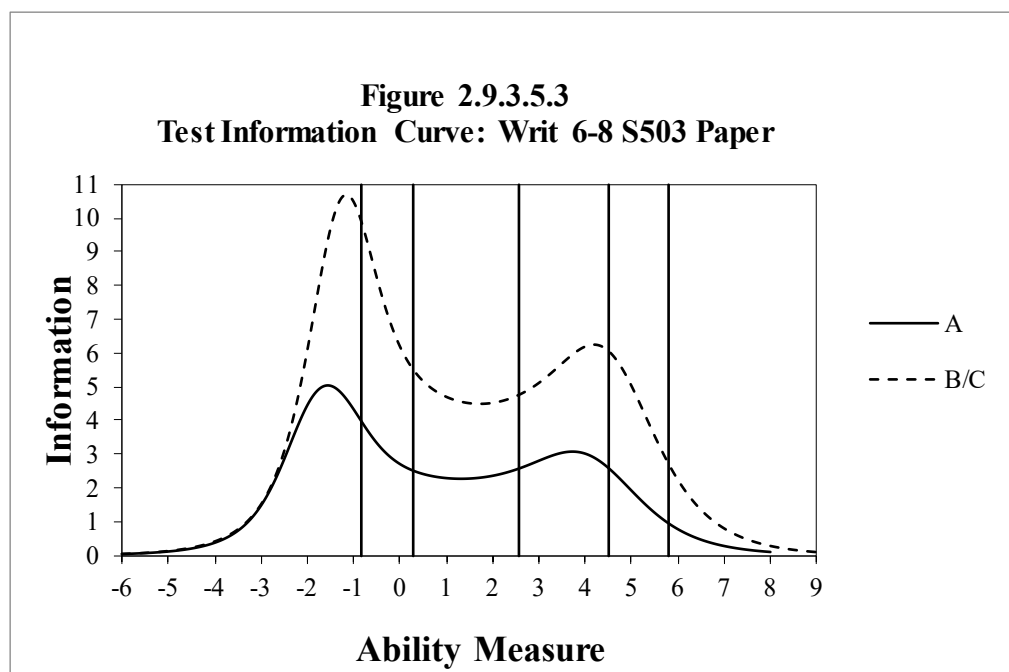


**Figure 2.9.3.4.3**  
**Test Information Curve: Writ 4-5 S503 Paper**

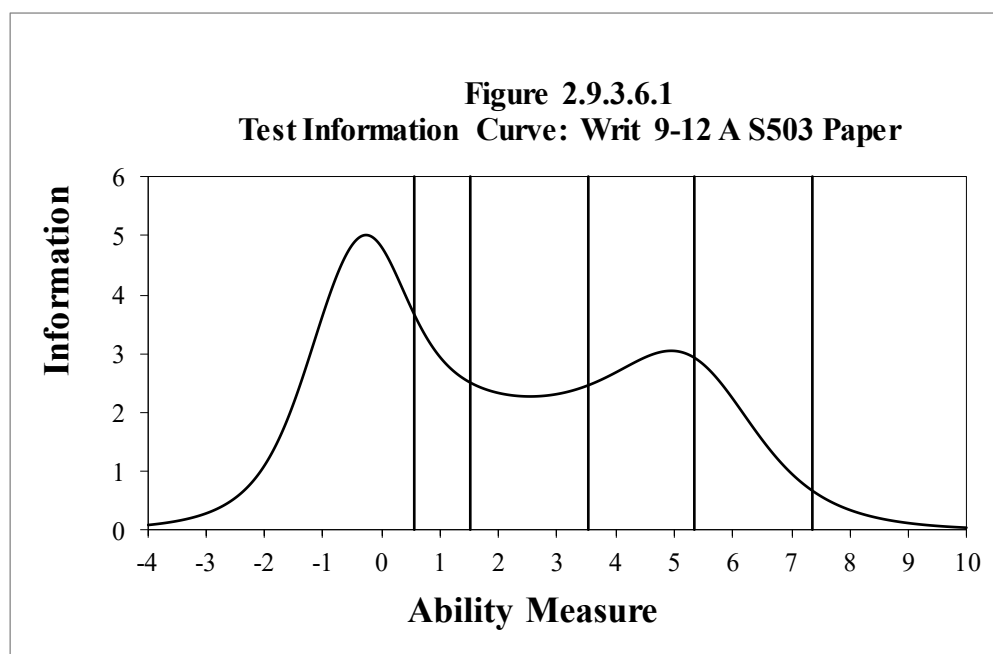


2.9.3.5 Grades 6-8

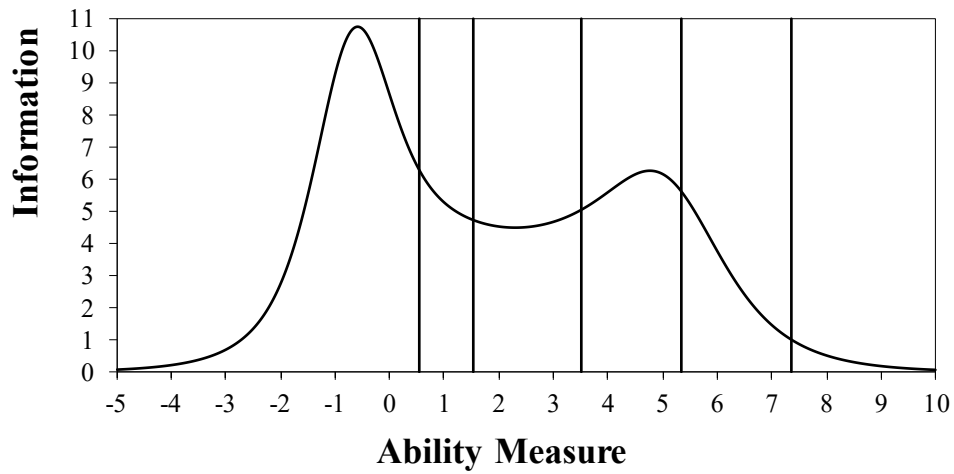




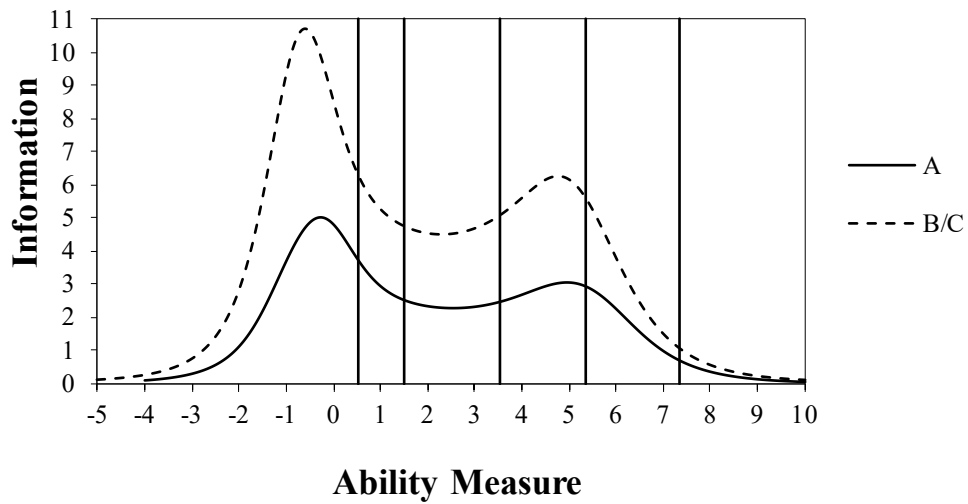
#### 2.9.3.6 Grades 9-12



**Figure 2.9.3.6.2**  
**Test Information Curve: Writ 9-12 B/C S503 Paper**

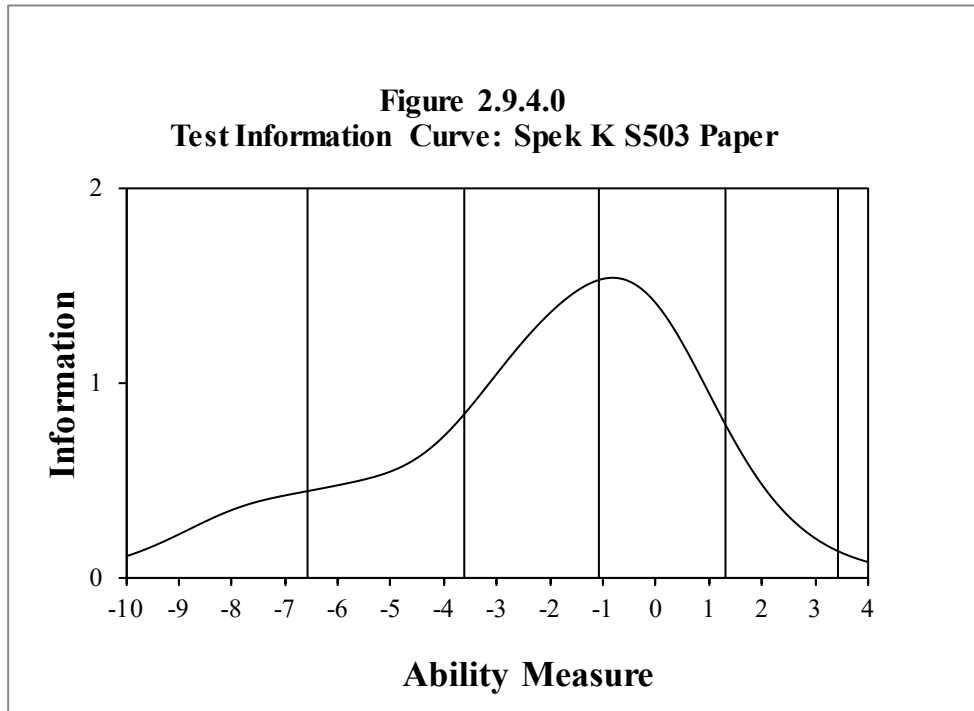


**Figure 2.9.3.6.3**  
**Test Information Curve: Writ 9-12 S503 Paper**



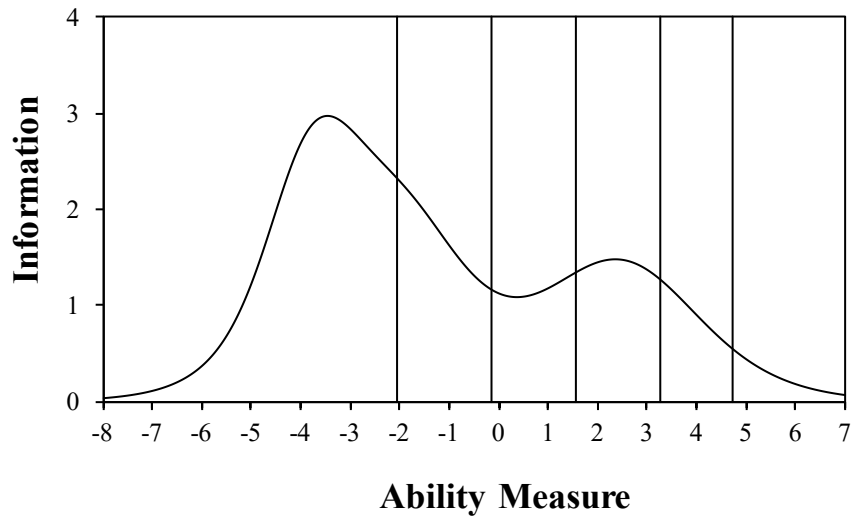
## 2.9.4 Speaking

### 2.9.4.0 Kindergarten

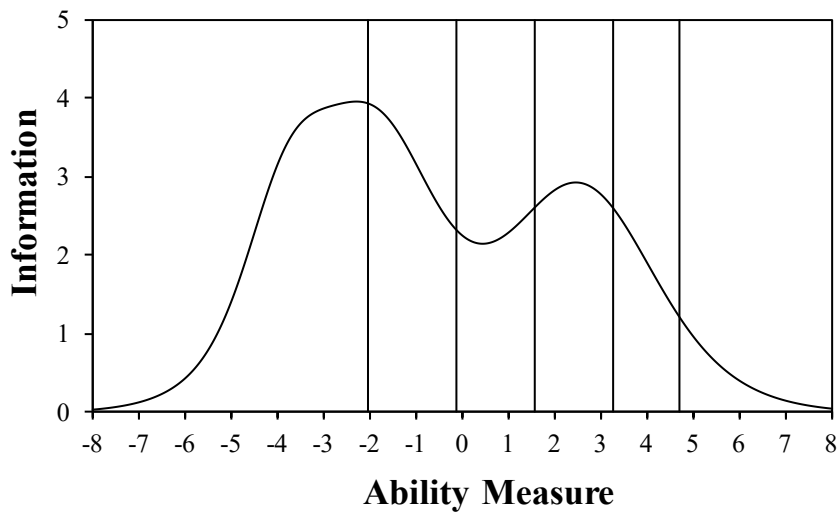


2.9.4.1 Grade 1

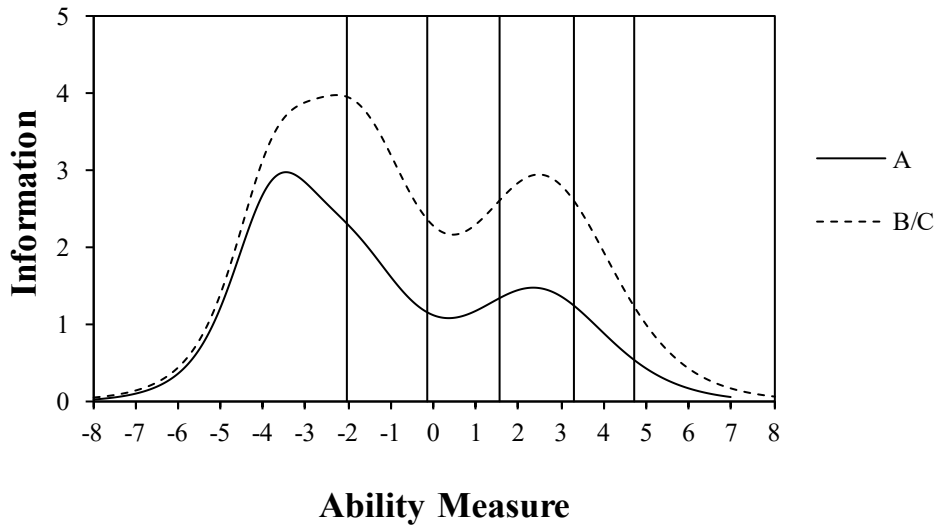
**Figure 2.9.4.1.1**  
**Test Information Curve: Spek 1 A S503 Paper**



**Figure 2.9.4.1.2**  
**Test Information Curve: Spek 1 B/C S503 Paper**

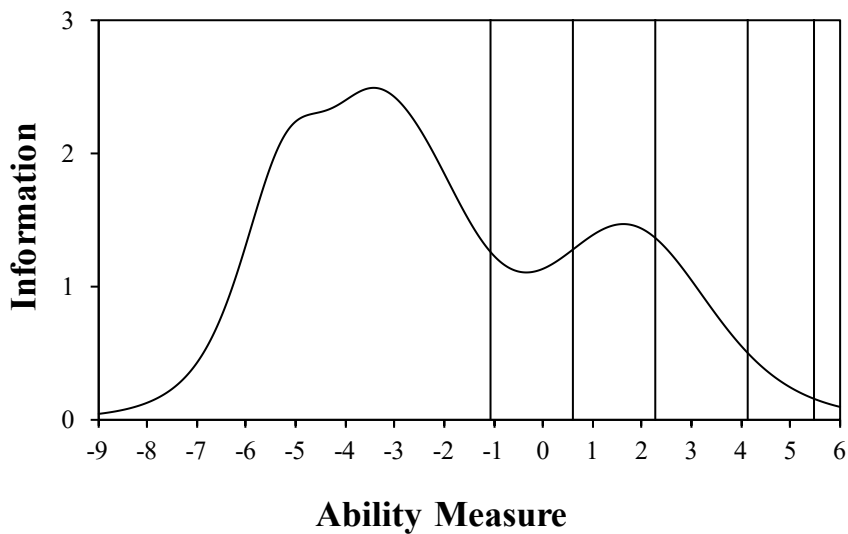


**Figure 2.9.4.1.3**  
**Test Information Curve: Spek 1 S503 Paper**

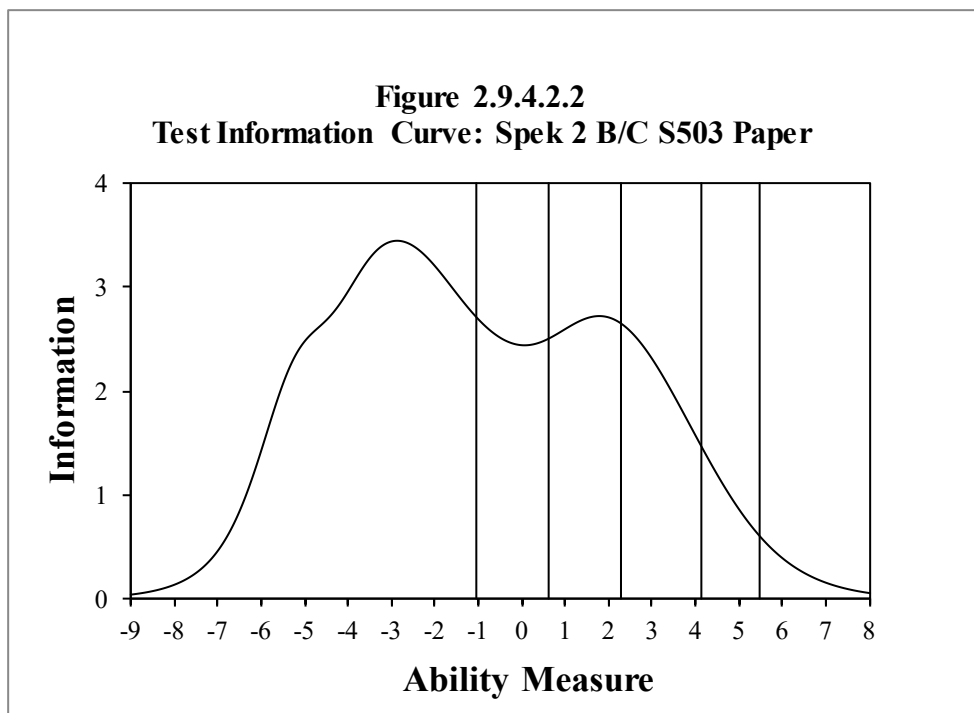


#### 2.9.4.2 Grade 2

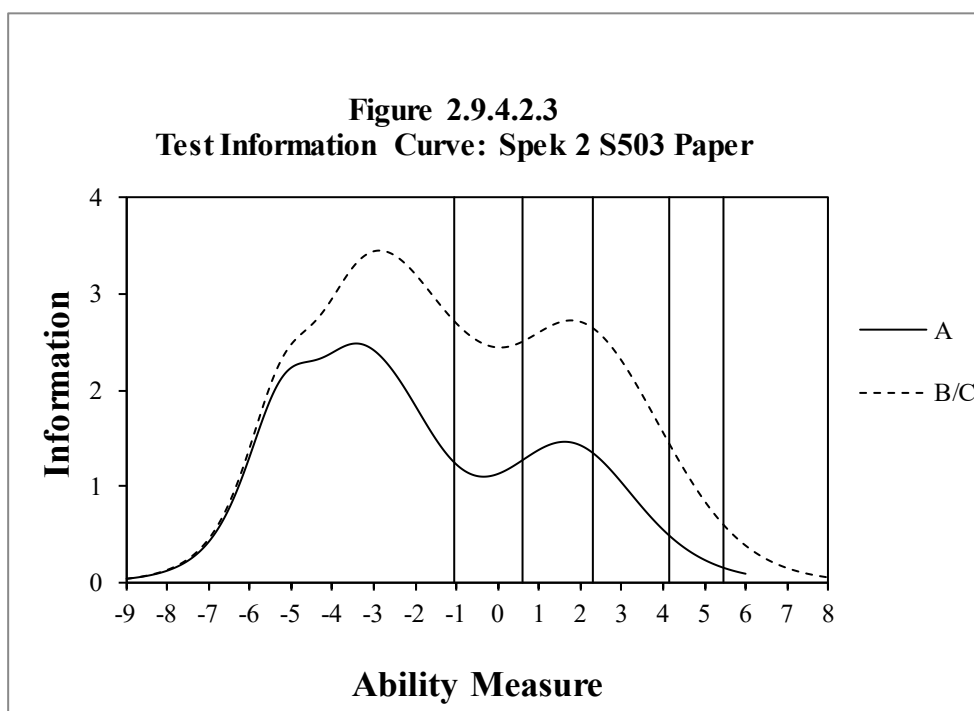
**Figure 2.9.4.2.1**  
**Test Information Curve: Spek 2 A S503 Paper**



Note: The test form is shared between 2A and 3A.



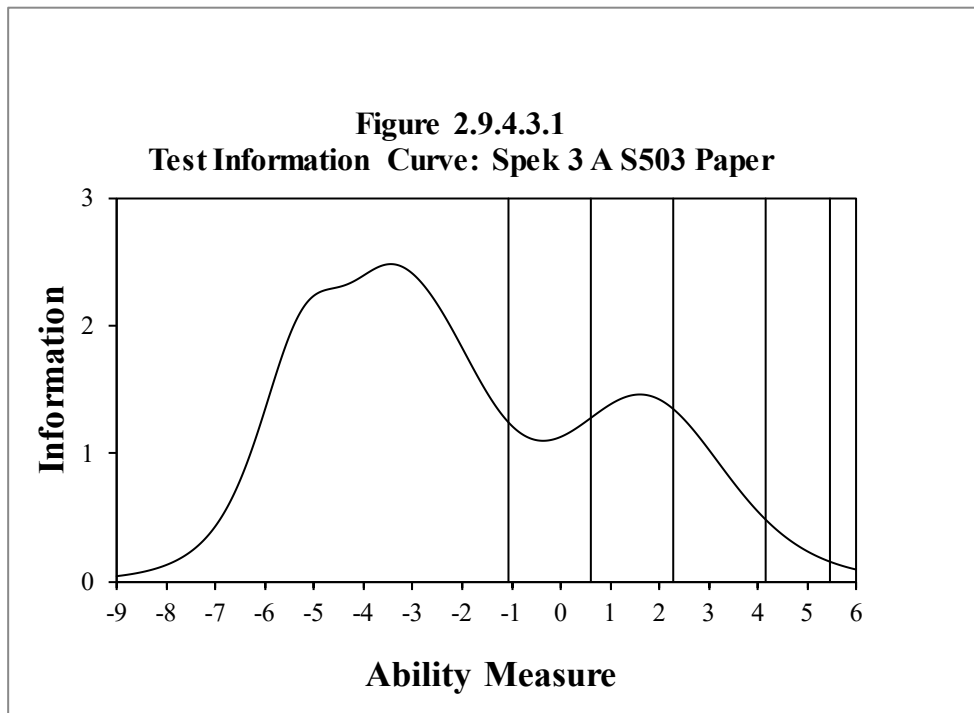
Note: The test form is shared between 2B/C and 3B/C.



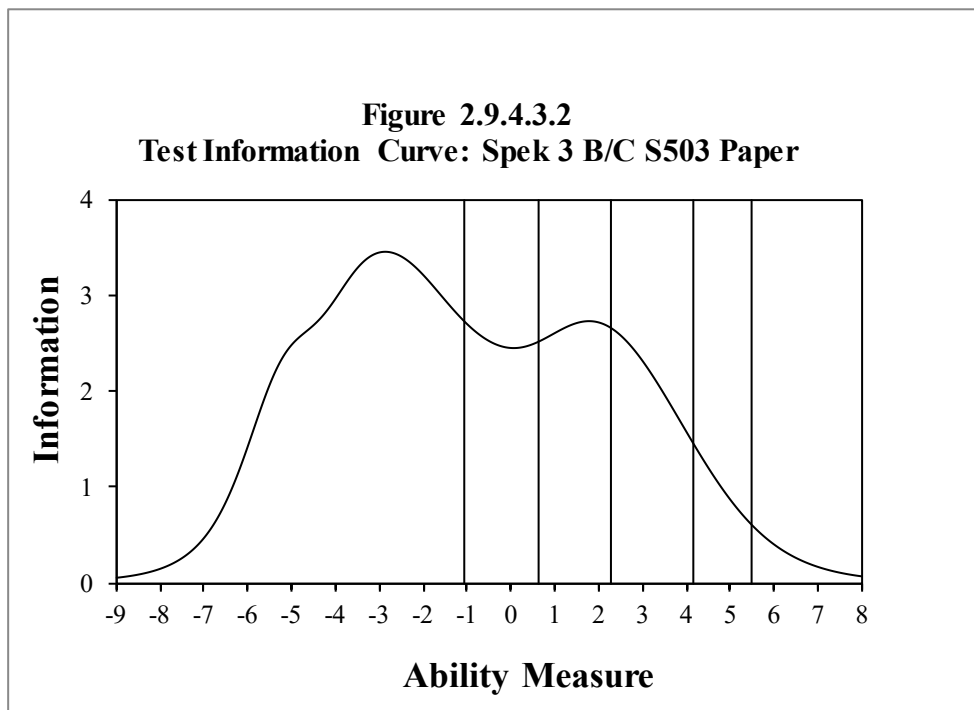
Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.



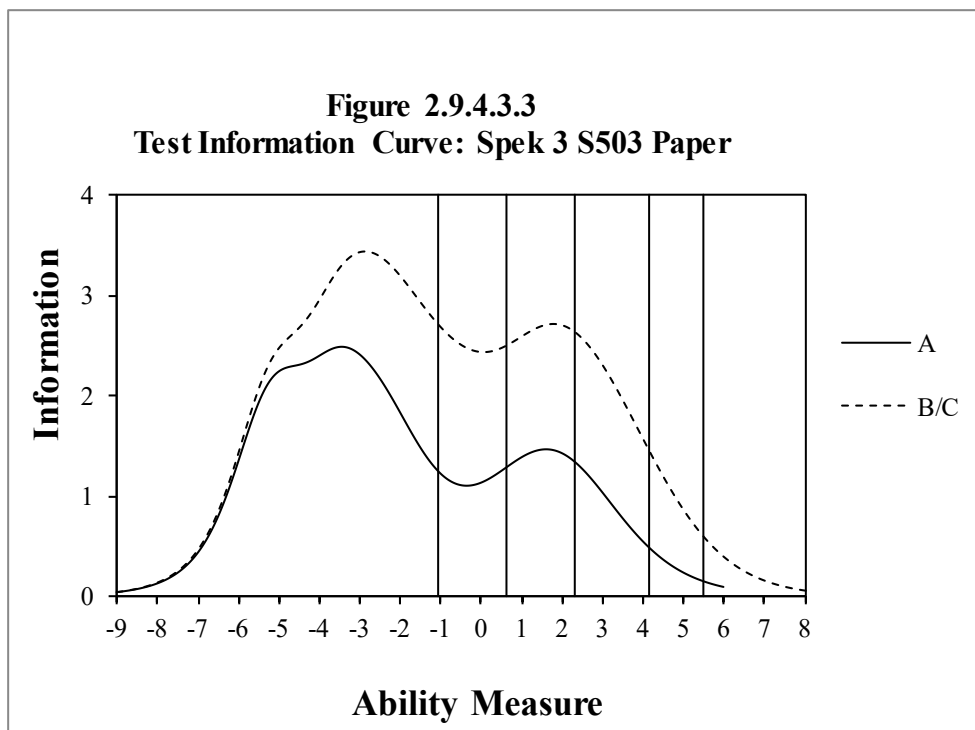
2.9.4.3 Grade 3



Note: The test form is shared between 2A and 3A.

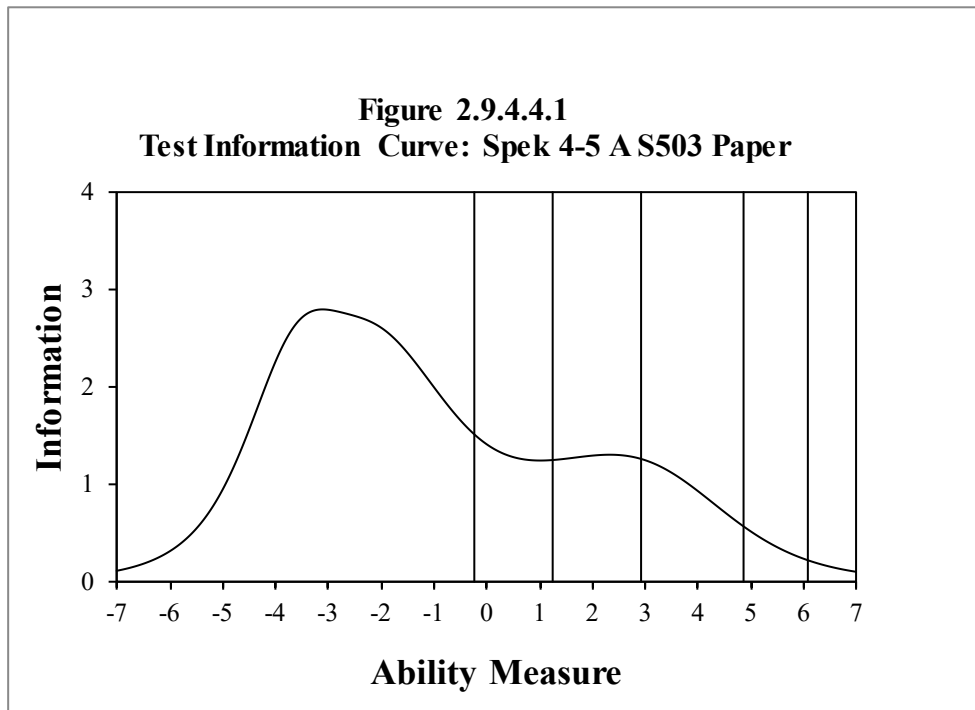


Note: The test form is shared between 2B/C and 3B/C.

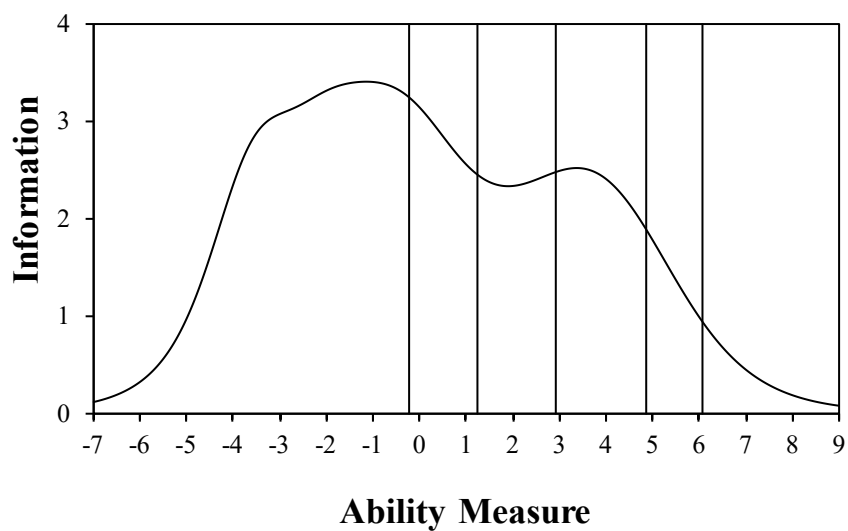


Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

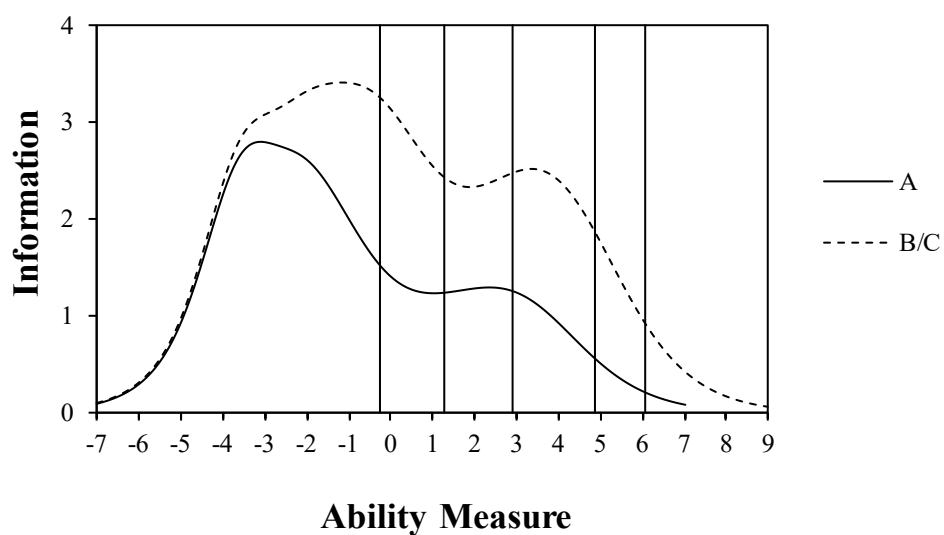
#### 2.9.4.4 Grades 4-5



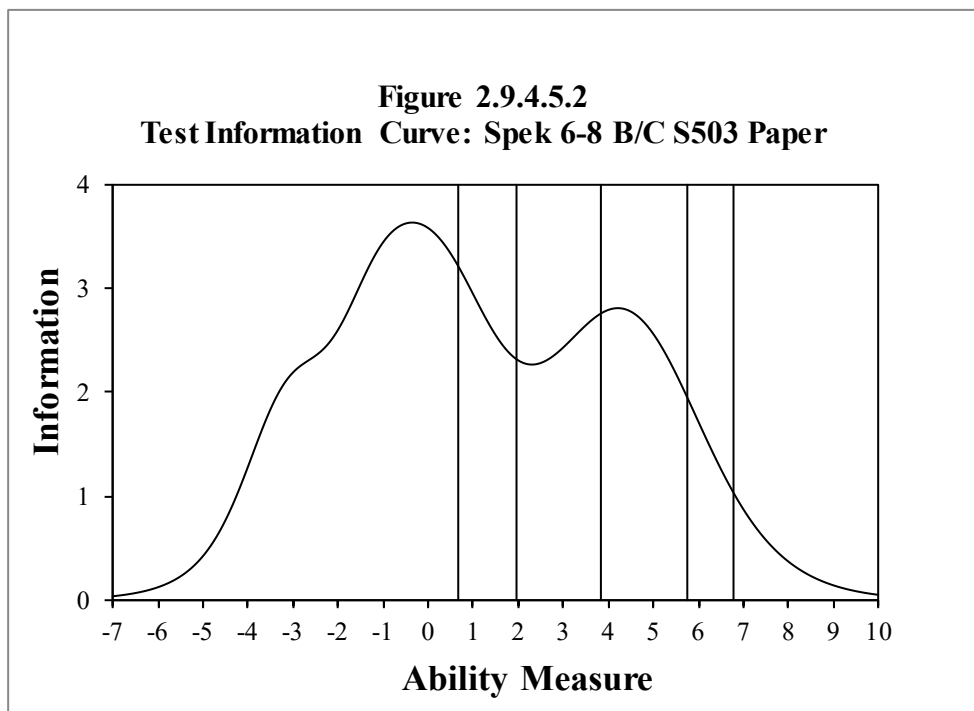
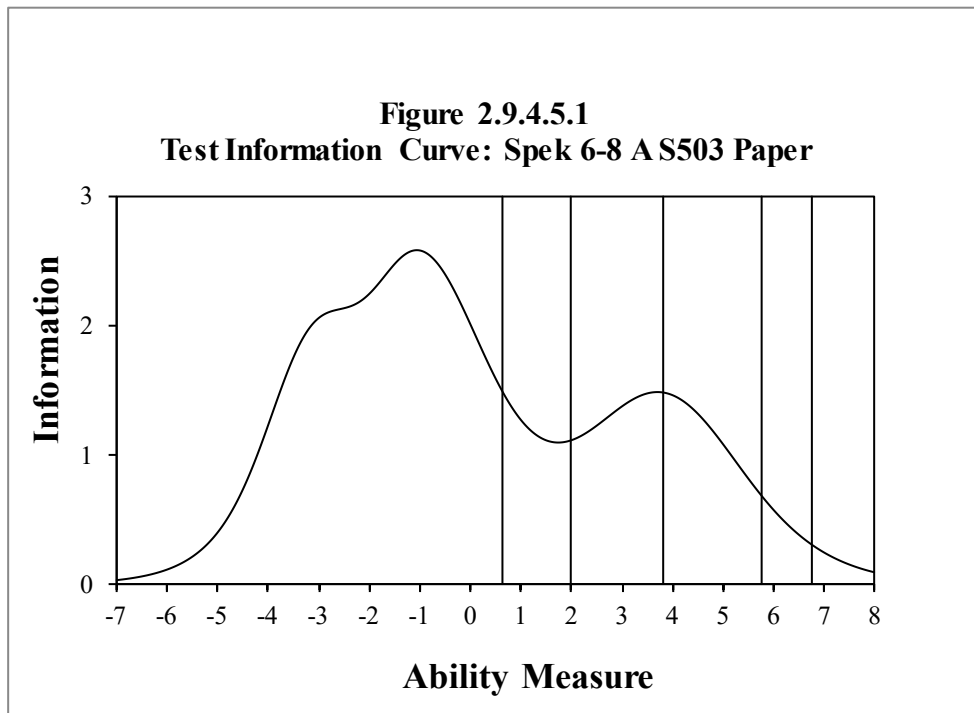
**Figure 2.9.4.4.2**  
**Test Information Curve: Spek 4-5 B/C S503 Paper**



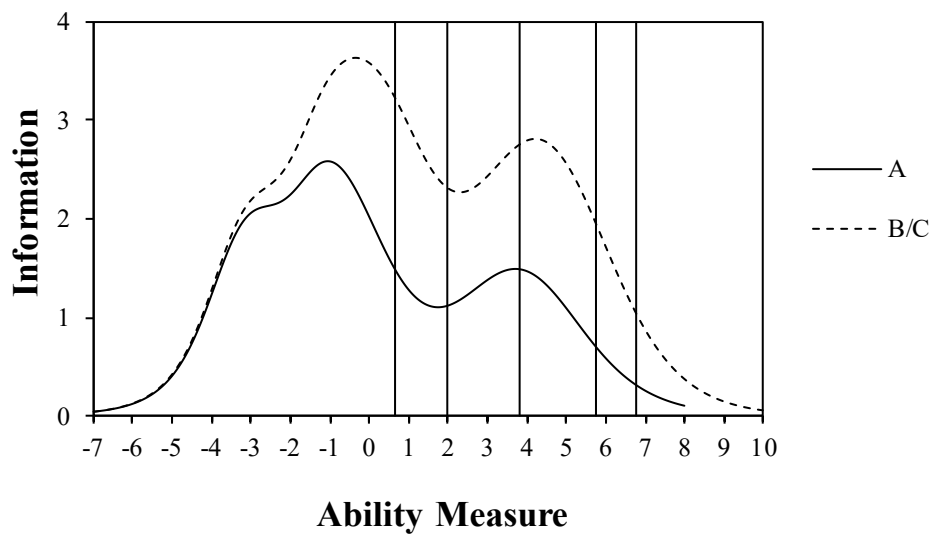
**Figure 2.9.4.4.3**  
**Test Information Curve: Spek 4-5 S503 Paper**



2.9.4.5 Grades 6-8

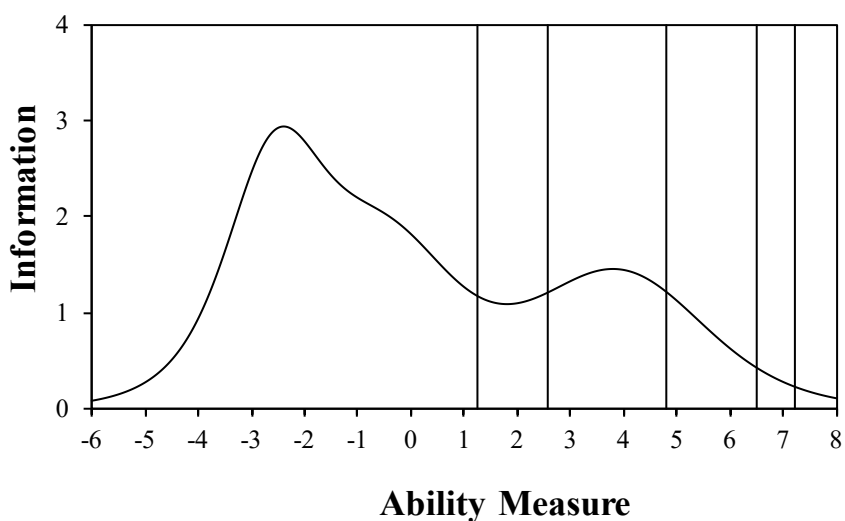


**Figure 2.9.4.5.3**  
**Test Information Curve: Spek 6-8 S503 Paper**

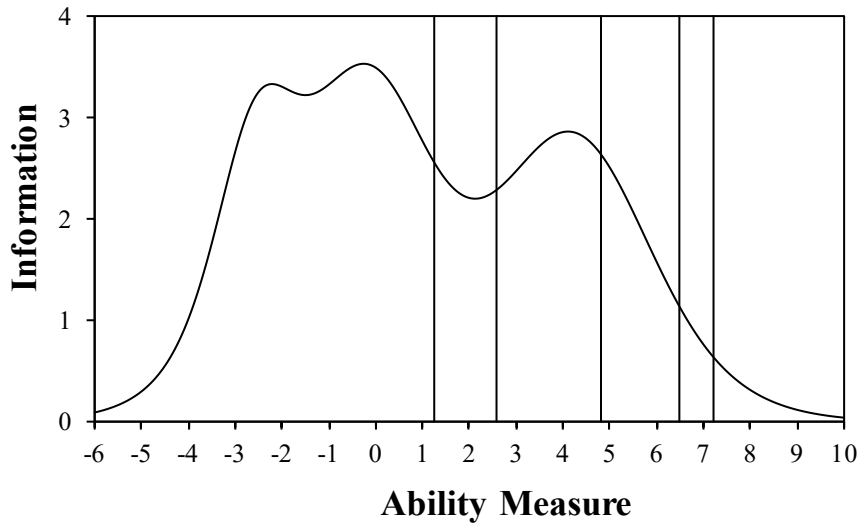


*2.9.4.6 Grades 9-12*

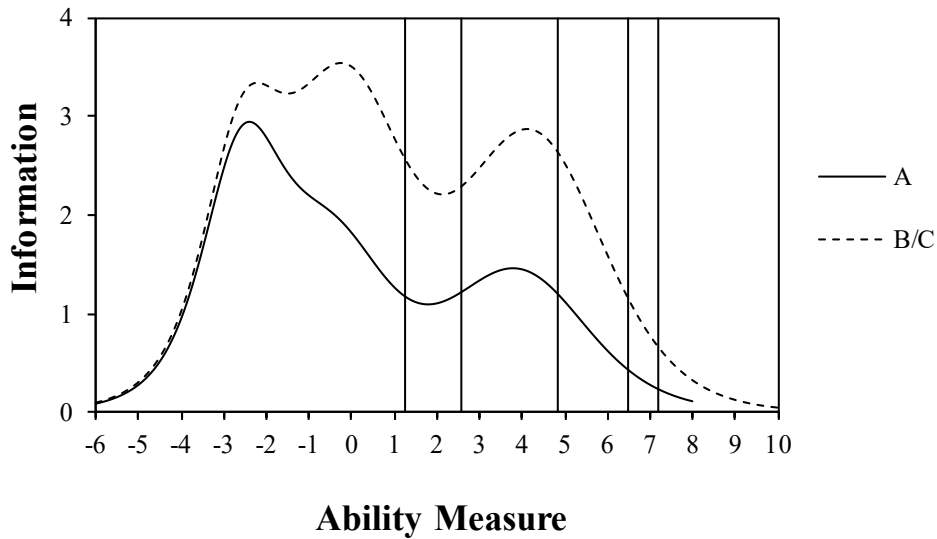
**Figure 2.9.4.6.1**  
**Test Information Curve: Spek 9-12 A S503 Paper**



**Figure 2.9.4.6.2**  
**Test Information Curve: Spek 9-12 B/C S503 Paper**



**Figure 2.9.4.6.3**  
**Test Information Curve: Spek 9-12 S503 Paper**



### 3. Analyses of Composite Scores

We calculate four composite scores for ACCESS Online: Oral Language, Literacy, Comprehension, and Overall. We calculate these composite scores as weighted averages of domain scale scores, as follows:

- Oral Language: 50% Listening + 50% Speaking
- Literacy: 50% Reading + 50% Writing
- Comprehension: 30% Listening + 70% Reading
- Overall Composite: 15% Listening + 15% Speaking + 35% Reading + 35% Writing

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.

#### 3.1 *Scale Score Distribution for Composites*

Figures and tables in this section provide scale score distributions for each of the composites, for each grade-level cluster.

For each cluster, the figure shows the distribution of the scale scores for the composite. We plotted the scale scores, grouped into units of five scale score points (e.g., 100–104, 105–109, 110–114, etc.), on the horizontal axis and the number of students with scale scores falling into each range on the vertical axis.

Each table shows, by grade and by total for the grade-level cluster:

- The number of students in the analyses (count)
- The minimum observed scale score
- The maximum observed scale score
- The mean (average) scale score
- The standard deviation (std. dev.) of the scale score

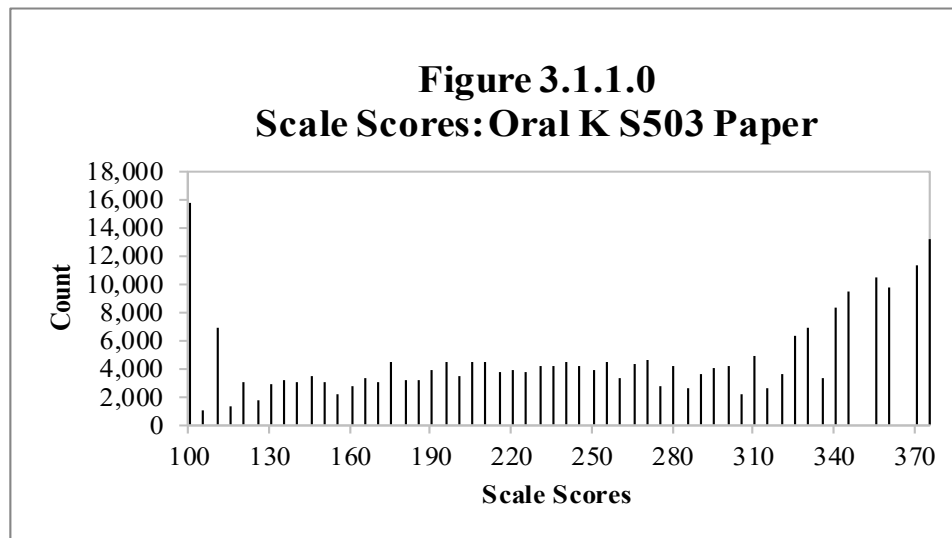
### 3.1.1 Oral

#### 3.1.1.0 Kindergarten

**Table 3.1.1.0**

Scale Score Descriptive Statistics: Oral K S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	251,481	100	378	254.35	88.23



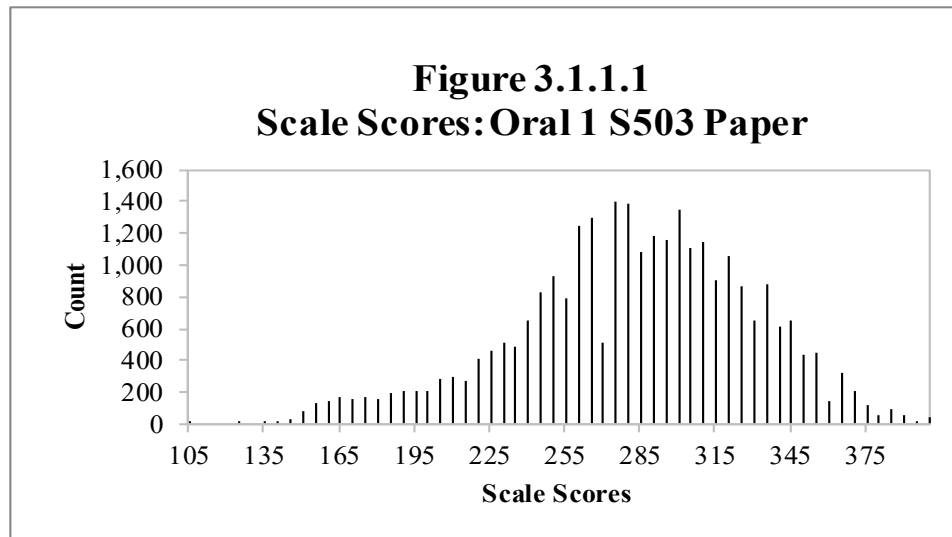


### 3.1.1.1 Grade 1

**Table 3.1.1.1**

Scale Score Descriptive Statistics: Oral 1 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	28,240	105	408	283.85	47.62

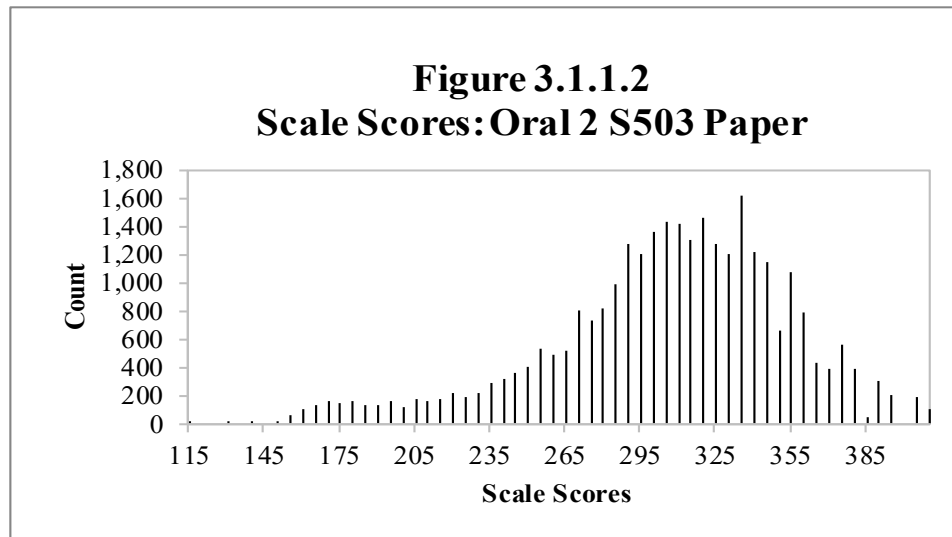


### 3.1.1.2 Grade 2

**Table 3.1.1.2**

Scale Score Descriptive Statistics: Oral 2 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	29,834	115	413	307.27	48.53

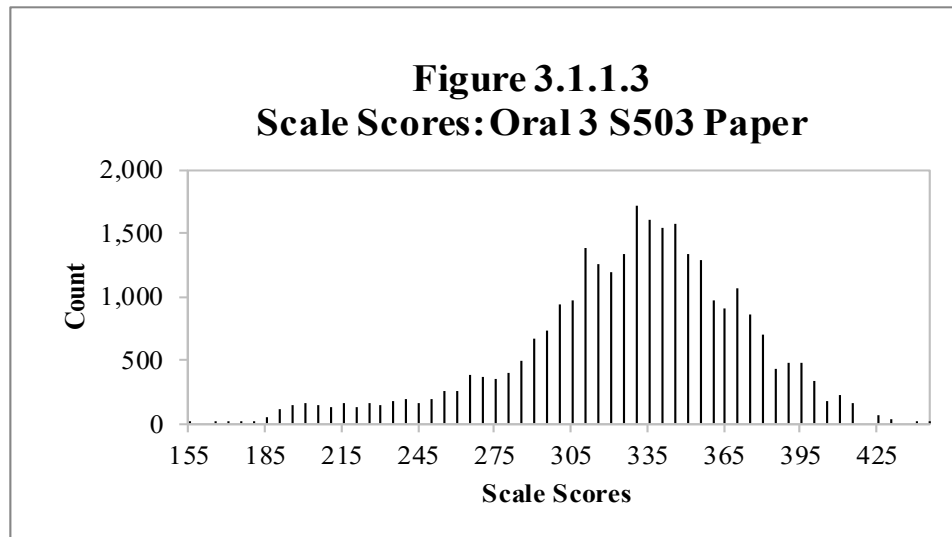


### 3.1.1.3 Grade 3

**Table 3.1.1.3**

Scale Score Descriptive Statistics: Oral 3 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	29,200	157	447	328.40	45.74

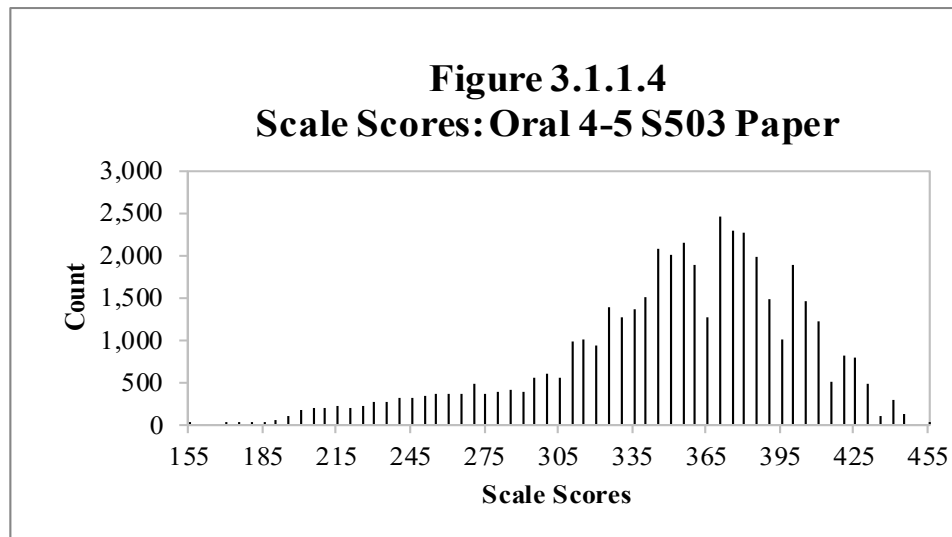


### 3.1.1.4 Grades 4-5

**Table 3.1.1.4**

Scale Score Descriptive Statistics: Oral 4-5 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	22,409	157	456	347.37	50.66
5	22,512	157	456	355.48	52.21
Total	44,921	157	456	351.43	51.60

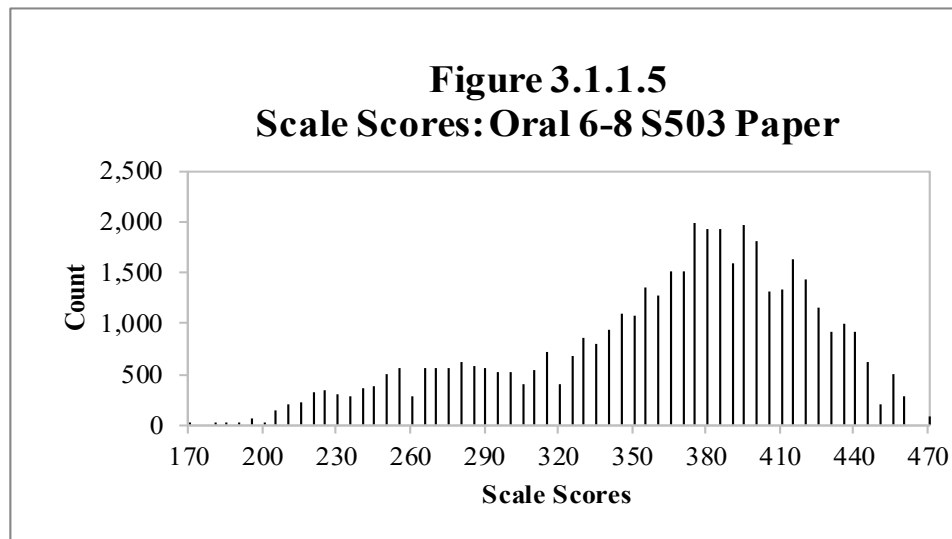


### 3.1.1.5 Grades 6-8

**Table 3.1.1.5**

Scale Score Descriptive Statistics: Oral 6-8 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	15,747	170	471	359.32	58.05
7	14,586	180	471	361.26	60.21
8	14,002	191	471	366.91	61.66
Total	44,335	170	471	362.36	60.00

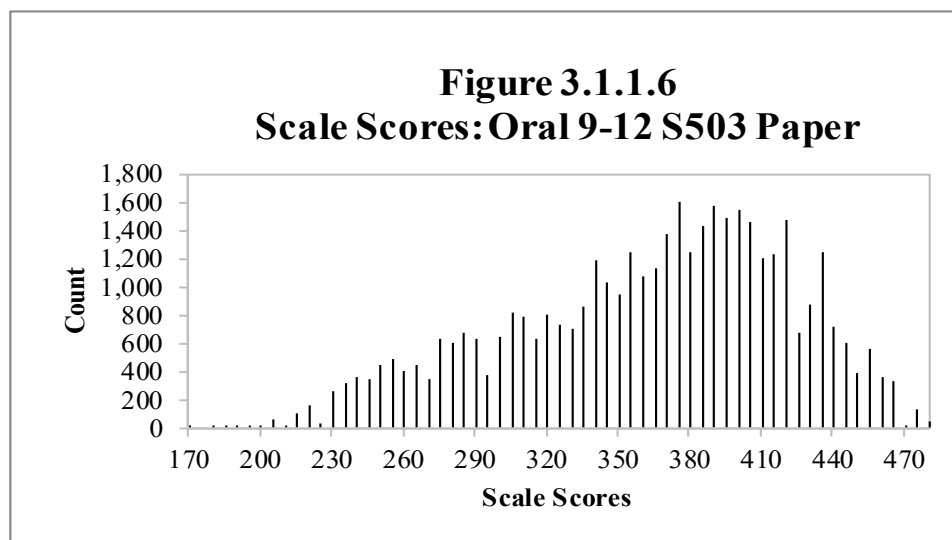


### 3.1.1.6 Grades 9-12

**Table 3.1.1.6**

Scale Score Descriptive Statistics: Oral 9-12 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	12,742	172	483	353.40	62.62
10	11,381	192	483	364.43	58.38
11	10,088	194	483	368.31	57.18
12	6,751	209	483	376.45	51.65
<b>Total</b>	<b>40,962</b>	<b>172</b>	<b>483</b>	<b>363.93</b>	<b>58.96</b>



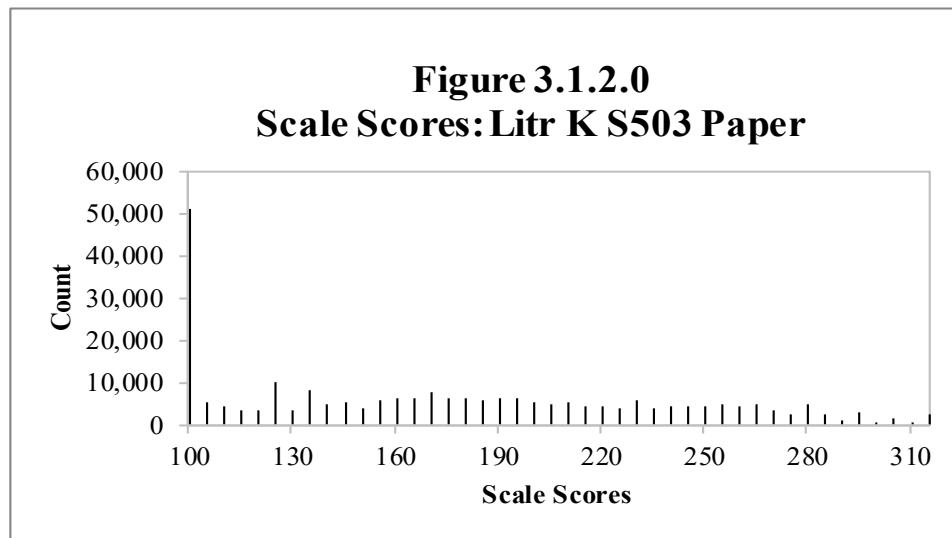
## 3.1.2 Literacy

### 3.1.2.0 Kindergarten

**Table 3.1.2.0**

Scale Score Descriptive Statistics: Litr K S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	251,486	100	315	176.87	62.63

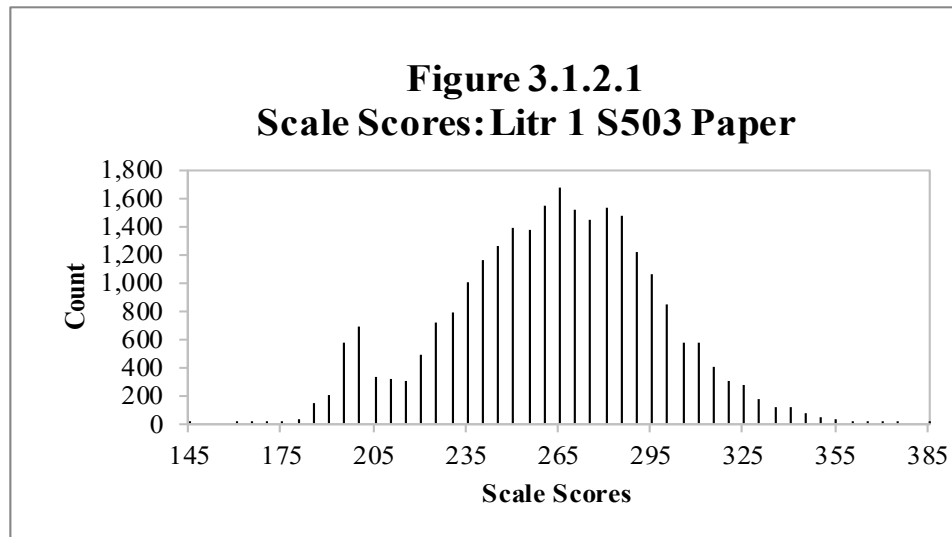


### 3.1.2.1 Grade 1

**Table 3.1.2.1**

Scale Score Descriptive Statistics: Litr 1 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	25,859	145	387	264.57	33.56



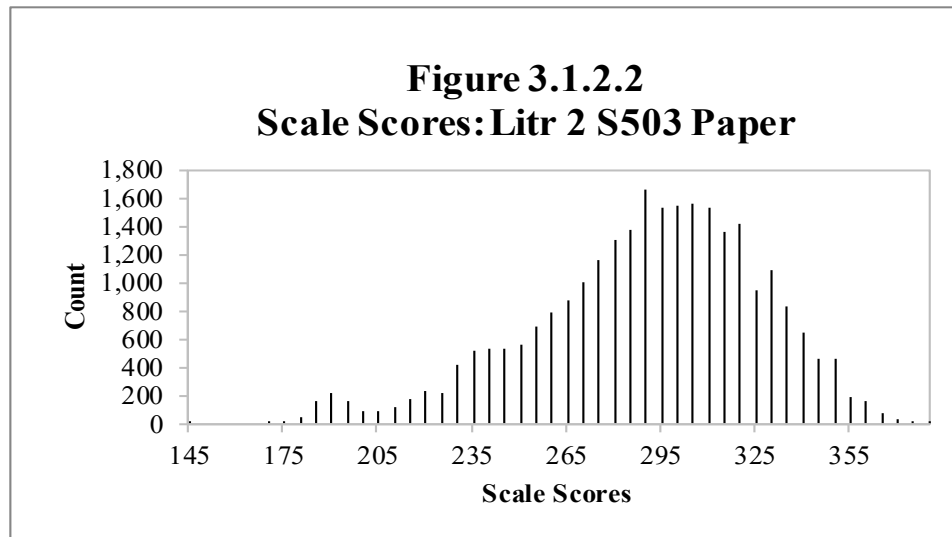


### 3.1.2.2 Grade 2

**Table 3.1.2.2**

Scale Score Descriptive Statistics: Litr 2 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	26,860	146	384	291.94	36.57

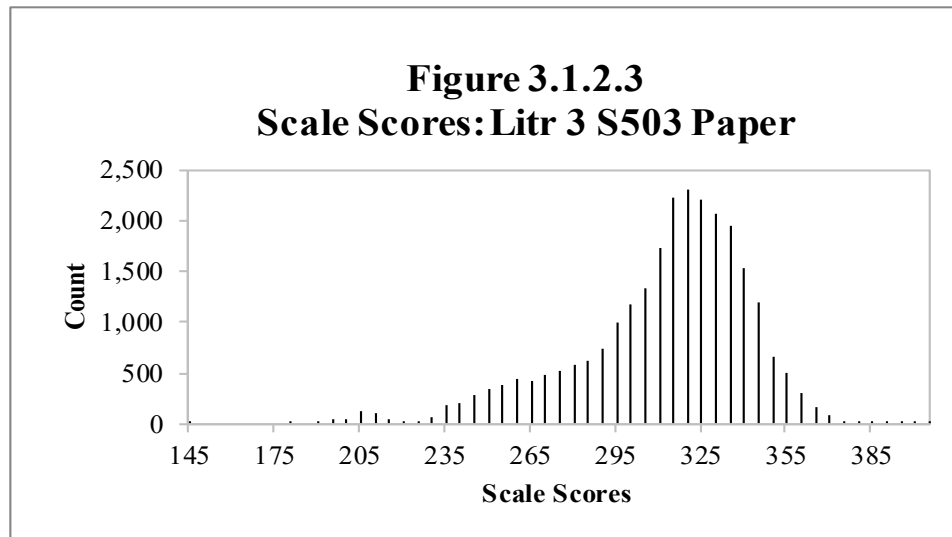


### 3.1.2.3 Grade 3

**Table 3.1.2.3**

Scale Score Descriptive Statistics: Litr 3 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	26,239	146	407	313.12	31.21

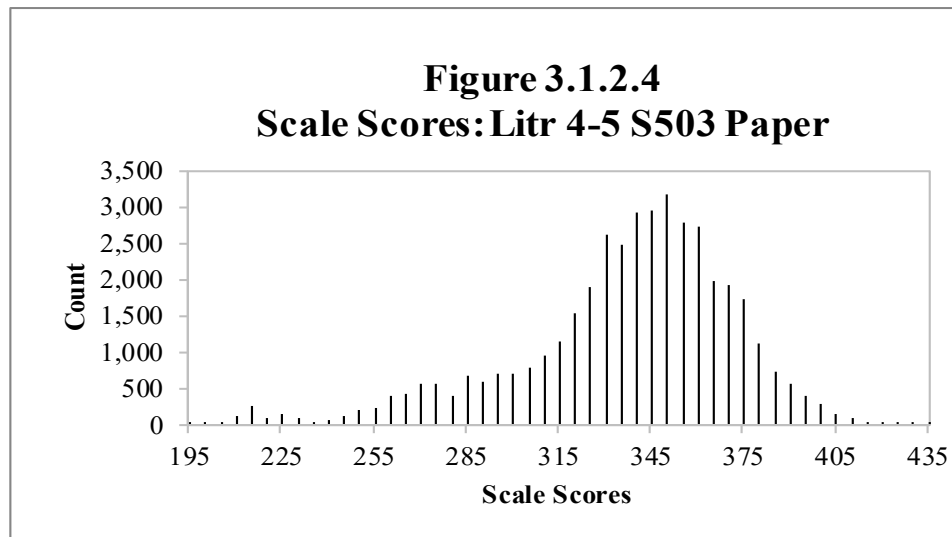


### 3.1.2.4 Grades 4-5

**Table 3.1.2.4**

Scale Score Descriptive Statistics: Litr 4-5 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	20,620	195	425	333.31	34.76
5	20,888	195	437	343.37	36.53
Total	41,508	195	437	338.37	36.02

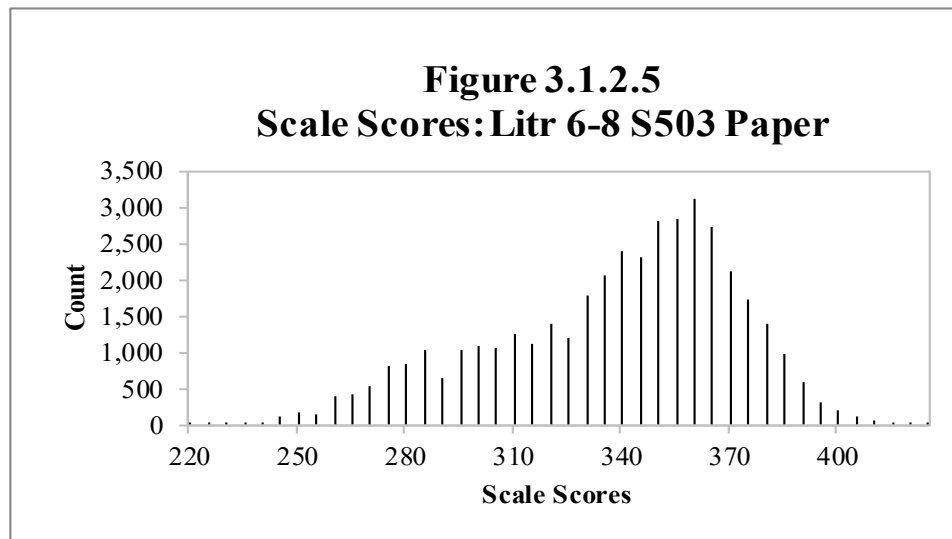


### 3.1.2.5 Grades 6-8

**Table 3.1.2.5**

Scale Score Descriptive Statistics: Litr 6-8 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	14,345	222	427	334.76	32.20
7	13,496	228	427	339.49	33.62
8	13,122	229	429	345.33	35.06
Total	40,963	222	429	339.71	33.88

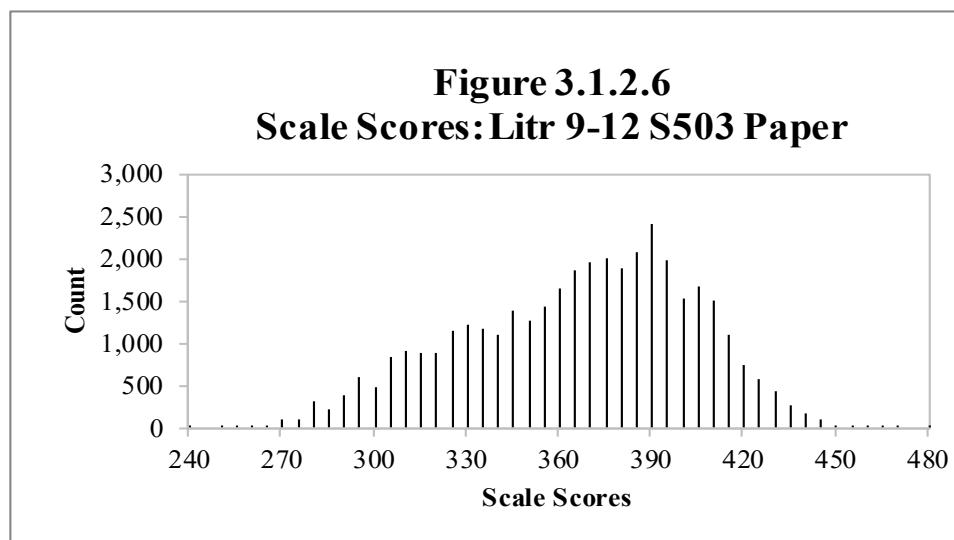


### 3.1.2.6 Grades 9-12

**Table 3.1.2.6**

Scale Score Descriptive Statistics: Litr 9-12 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	11,948	241	468	359.22	38.48
10	10,667	258	471	368.10	36.70
11	9,555	251	480	373.05	36.83
12	6,472	267	462	376.29	33.66
<b>Total</b>	<b>38,642</b>	<b>241</b>	<b>480</b>	<b>367.95</b>	<b>37.37</b>



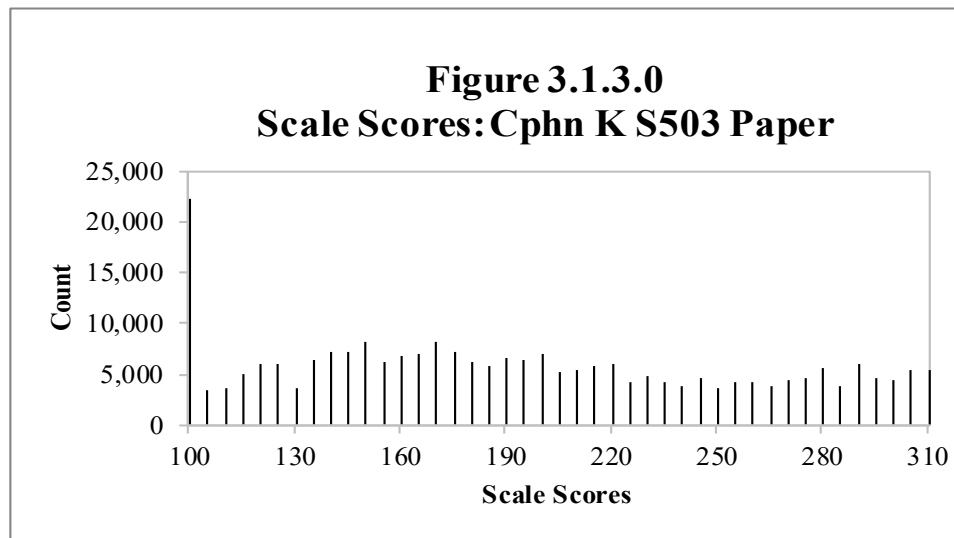
### 3.1.3 Comprehension

#### 3.1.3.0 Kindergarten

**Table 3.1.3.0**

Scale Score Descriptive Statistics: Cphn K S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	251,488	100	312	195.04	63.24

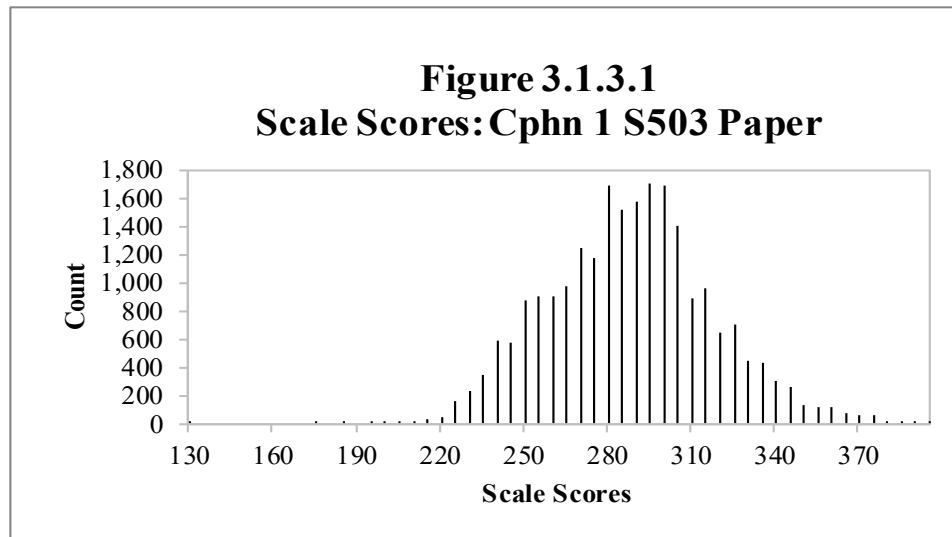


### 3.1.3.1 Grade 1

**Table 3.1.3.1**

Scale Score Descriptive Statistics: Cphn 1 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	22,913	130	395	289.92	29.57

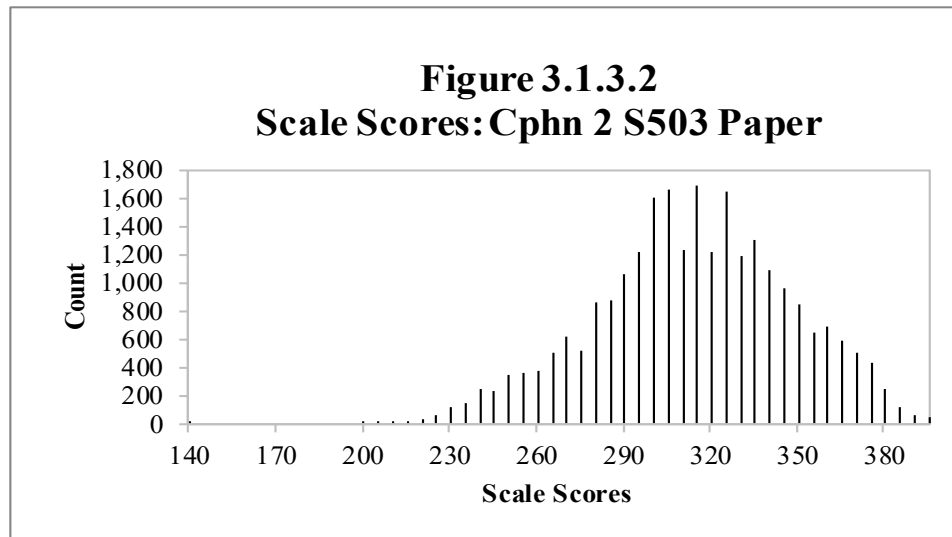


### 3.1.3.2 Grade 2

**Table 3.1.3.2**

Scale Score Descriptive Statistics: Cphn 2 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	25,383	144	395	315.55	33.73



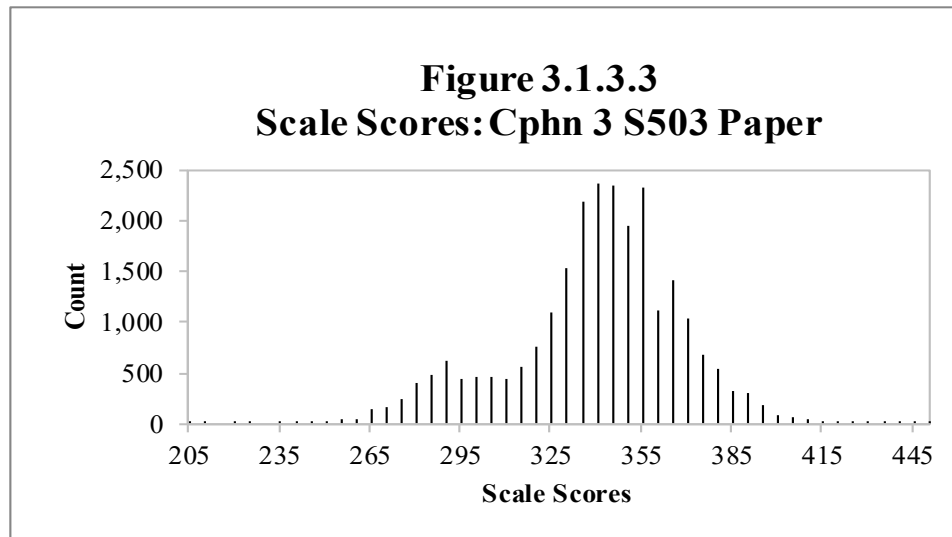


### 3.1.3.3 Grade 3

**Table 3.1.3.3**

Scale Score Descriptive Statistics: Cphn 3 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	25,022	209	454	341.62	28.17

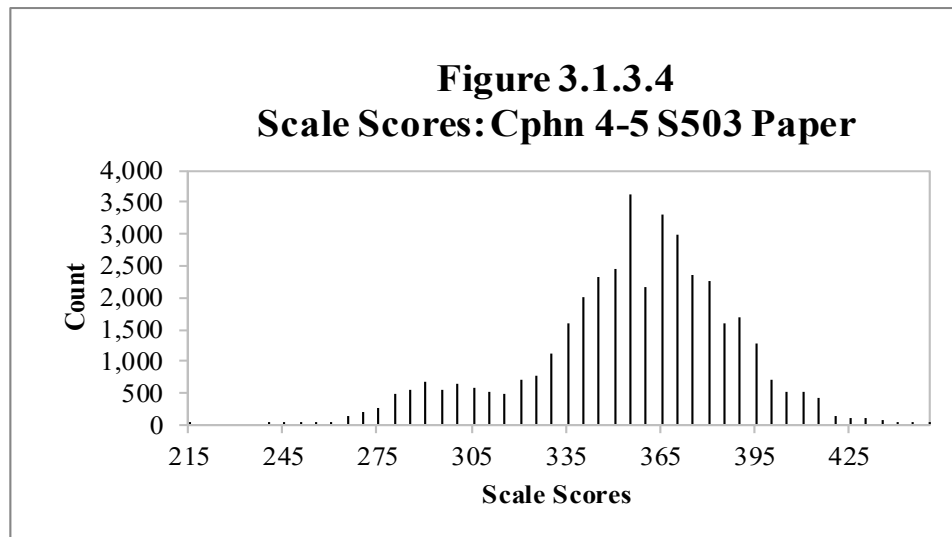


### 3.1.3.4 Grades 4-5

**Table 3.1.3.4**

Scale Score Descriptive Statistics: Cphn 4-5 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	19,905	217	454	352.92	31.04
5	20,299	241	454	361.14	32.99
Total	40,204	217	454	357.07	32.30



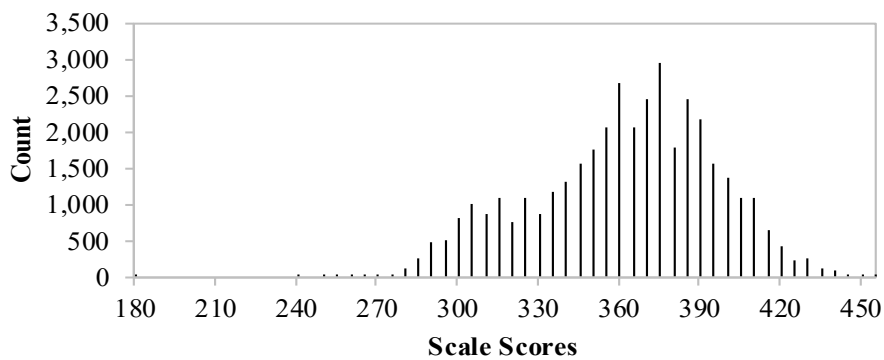
### 3.1.3.5 Grades 6-8

**Table 3.1.3.5**

Scale Score Descriptive Statistics: Cphn 6-8 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	13,801	180	456	358.64	31.43
7	12,989	259	456	363.38	33.33
8	12,599	244	456	369.75	35.73
Total	39,389	180	456	363.76	33.79

**Figure 3.1.3.5**  
**Scale Scores: Cphn 6-8 S503 Paper**

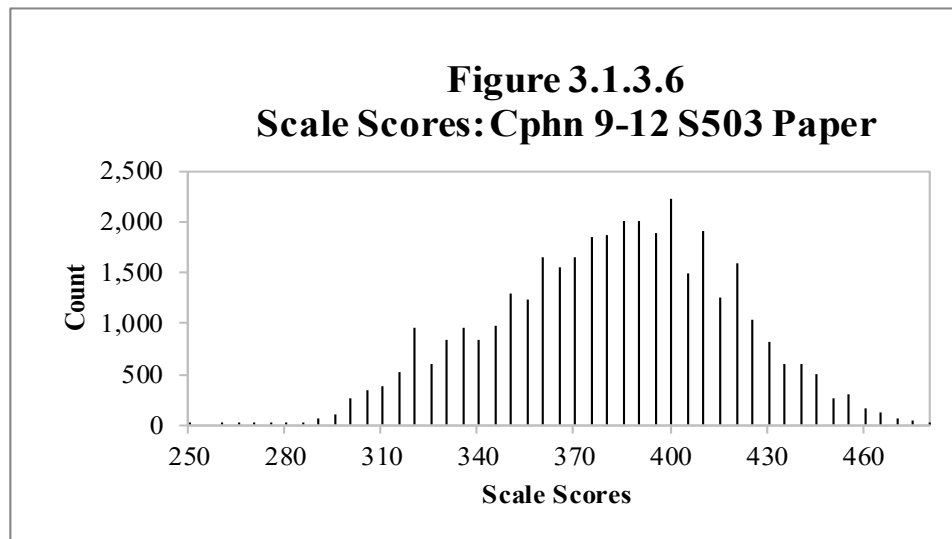


### 3.1.3.6 Grades 9-12

**Table 3.1.3.6**

Scale Score Descriptive Statistics: Cphn 9-12 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	11,388	253	482	375.18	37.18
10	10,224	268	482	383.51	36.08
11	9,166	263	482	387.88	36.31
12	6,226	288	482	391.10	33.16
<b>Total</b>	<b>37,004</b>	<b>253</b>	<b>482</b>	<b>383.31</b>	<b>36.50</b>



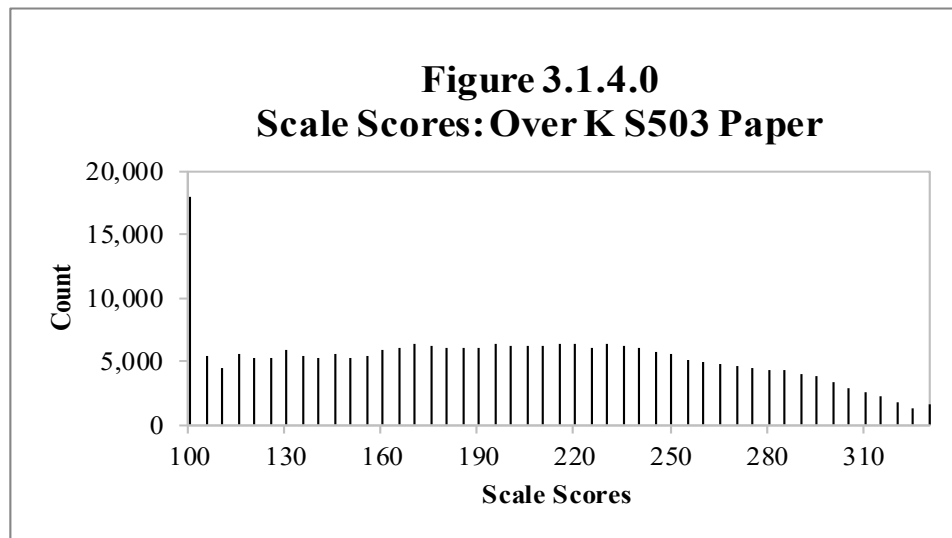
### 3.1.4 Overall

#### 3.1.4.0 Kindergarten

**Table 3.1.4.0**

Scale Score Descriptive Statistics: Over K S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	251,474	100	333	199.91	63.28

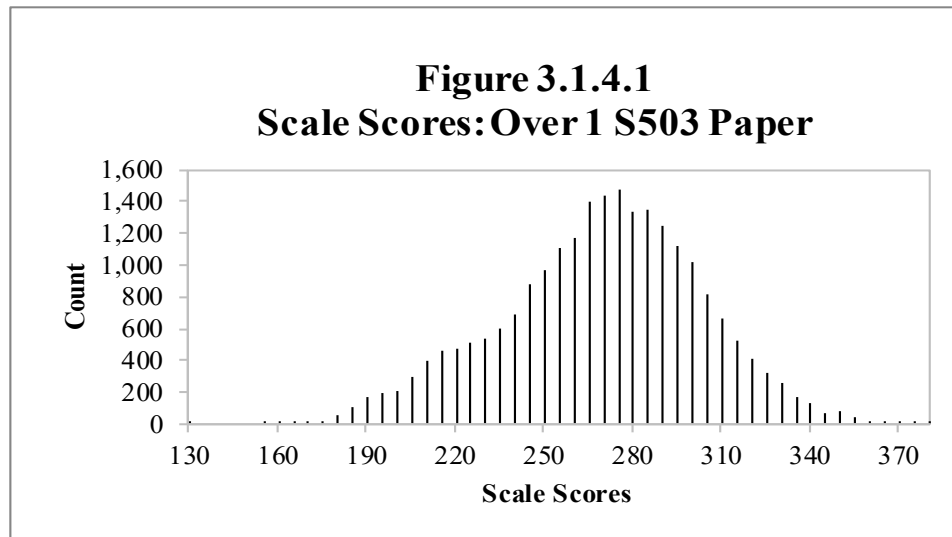


### 3.1.4.1 Grade 1

**Table 3.1.4.1**

Scale Score Descriptive Statistics: Over 1 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	22,785	133	384	271.13	33.91

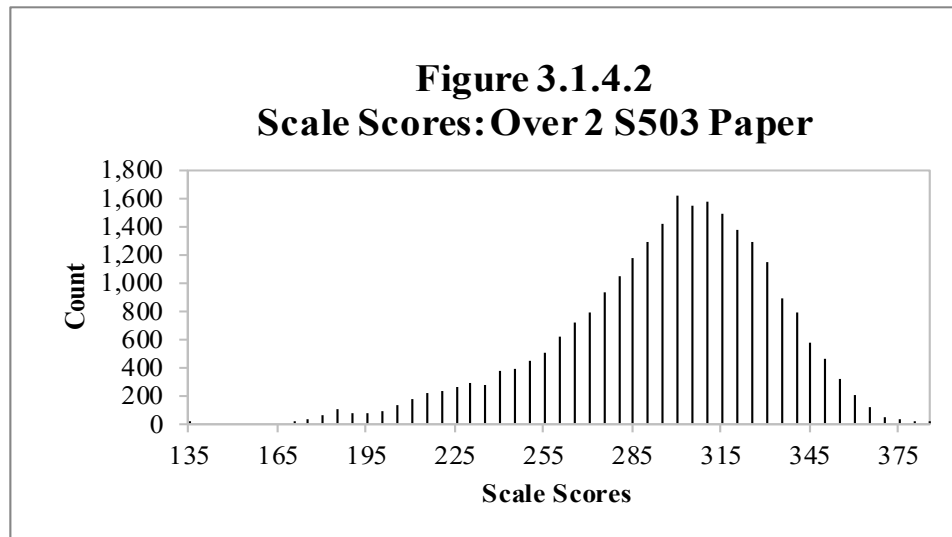


### 3.1.4.2 Grade 2

**Table 3.1.4.2**

Scale Score Descriptive Statistics: Over 2 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	25,221	136	387	297.08	36.58

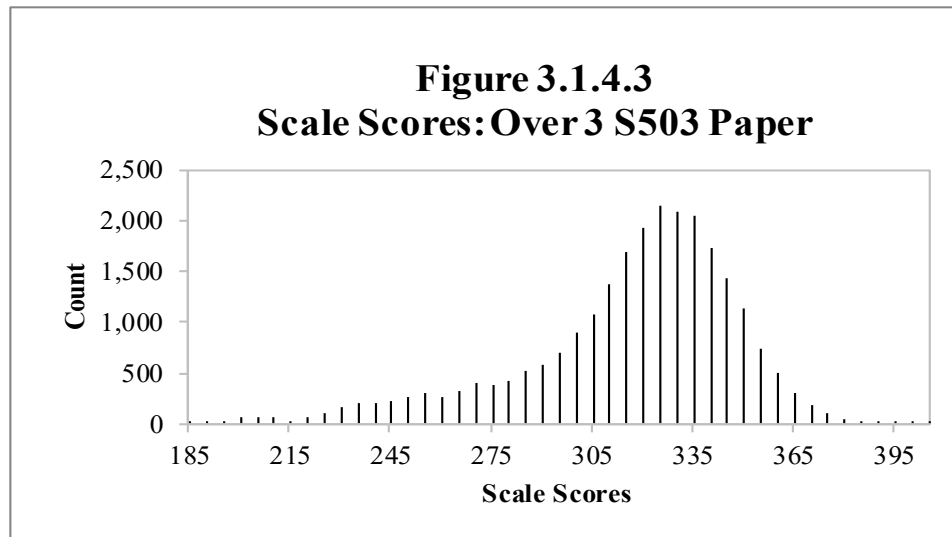


### 3.1.4.3 Grade 3

**Table 3.1.4.3**

Scale Score Descriptive Statistics: Over 3 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	24,867	188	406	318.00	32.51



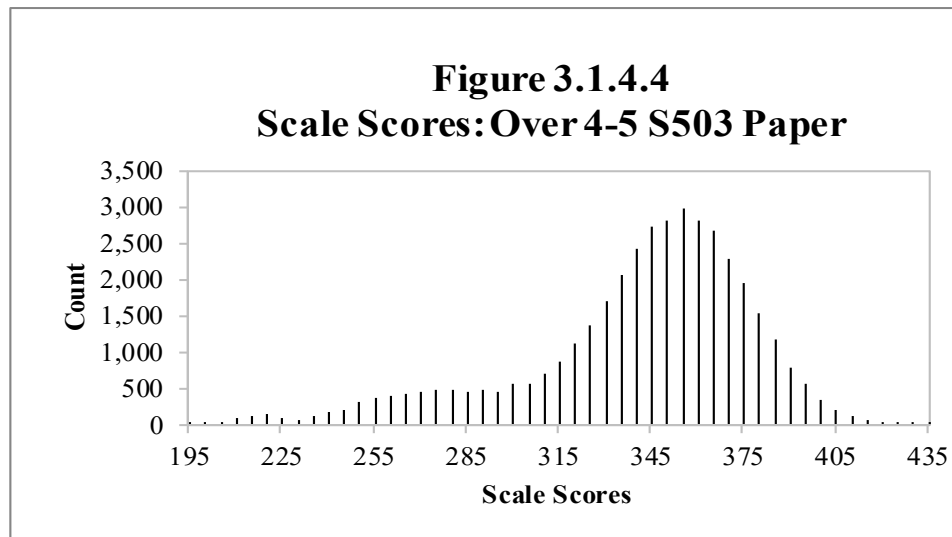


#### 3.1.4.4 Grades 4-5

**Table 3.1.4.4**

Scale Score Descriptive Statistics: Over 4-5 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	19,741	198	431	337.88	36.72
5	20,164	195	437	347.43	38.45
Total	39,905	195	437	342.71	37.91

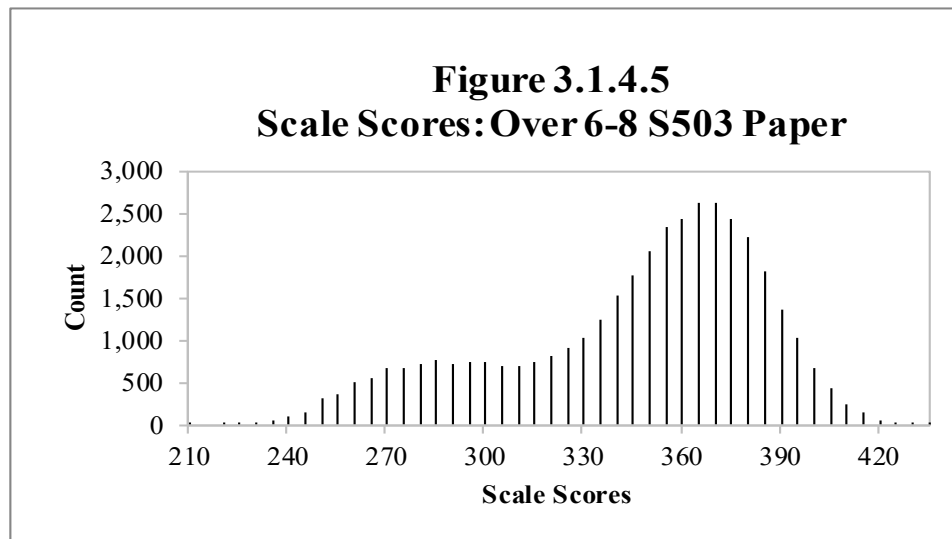


### 3.1.4.5 Grades 6-8

**Table 3.1.4.5**

Scale Score Descriptive Statistics: Over 6-8 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	13,695	212	437	342.10	37.90
7	12,900	222	433	346.14	39.56
8	12,513	227	439	352.08	41.18
Total	39,108	212	439	346.63	39.73

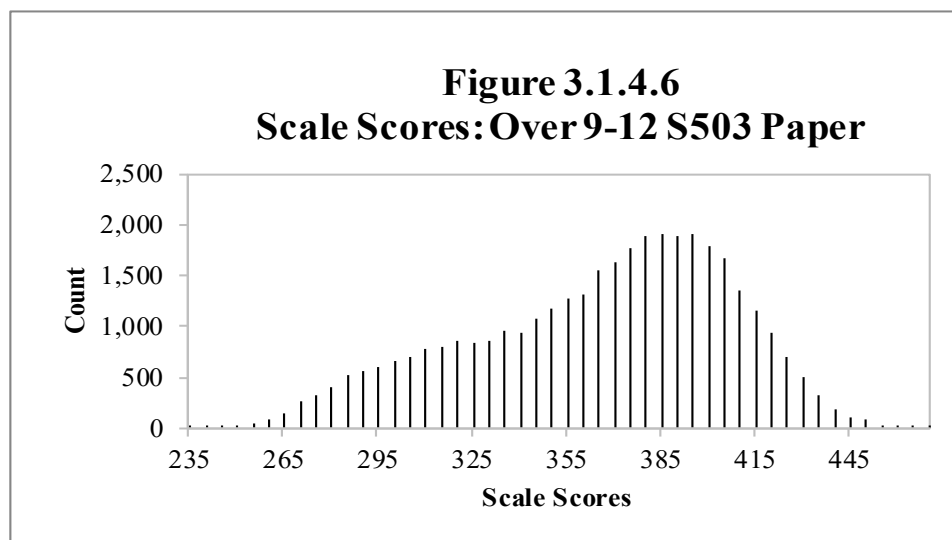


### 3.1.4.6 Grades 9-12

**Table 3.1.4.6**

Scale Score Descriptive Statistics: Over 9-12 S503 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	11,277	235	472	357.89	43.56
10	10,111	251	464	367.21	40.99
11	9,072	249	474	371.92	40.71
12	6,144	258	462	376.70	36.31
Total	36,604	235	474	367.10	41.57



### ***3.2 Proficiency Level Distribution for Composites***

Figures and tables in this section provide information on the proficiency level distribution for each of the composites for each grade-level cluster.

In each figure, the horizontal axis shows the six WIDA proficiency levels. The vertical axis shows the percentage of students. Each bar shows the percentage of students who were placed into each proficiency level in the domain being tested on this test form.

The tables in this section present, by grade and by total for the grade-level cluster:

- The WIDA proficiency level designation (1–6)
- The number of students (count) whose performance on the test form placed them into that proficiency level in the domain being tested
- The percentage of students, out of the total number of students taking the form, who were placed into that proficiency level in the domain being tested

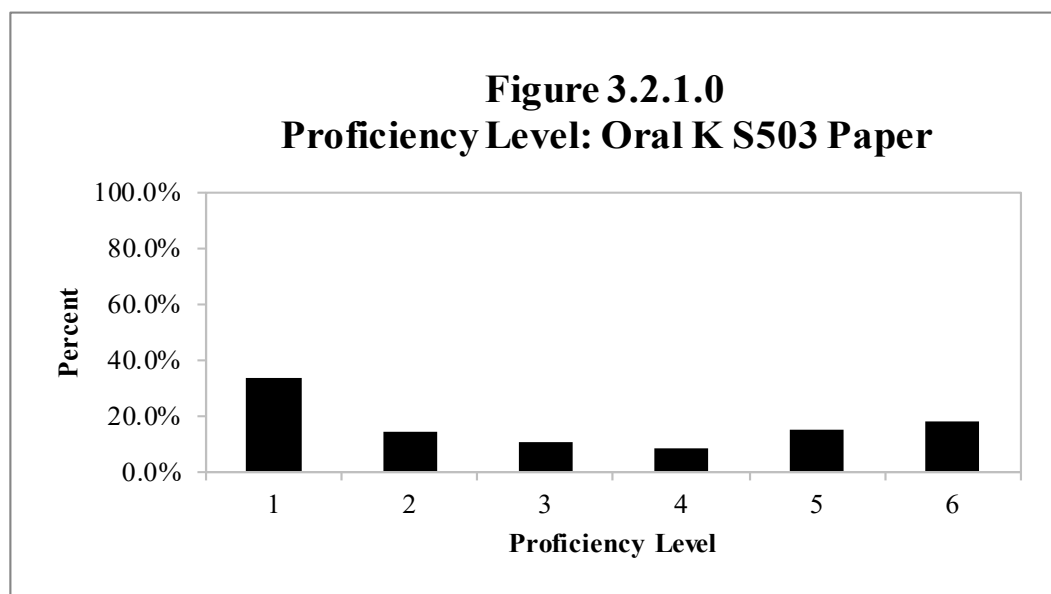
### 3.2.1 Oral

#### 3.2.1.0 Kindergarten

**Table 3.2.1.0**

Proficiency Level Distribution: Oral K S503 Paper

Level	Count	Percent
1	83,937	33.38%
2	35,494	14.11%
3	27,847	11.07%
4	21,509	8.55%
5	38,024	15.12%
6	44,670	17.76%
Total	251,481	100.00%

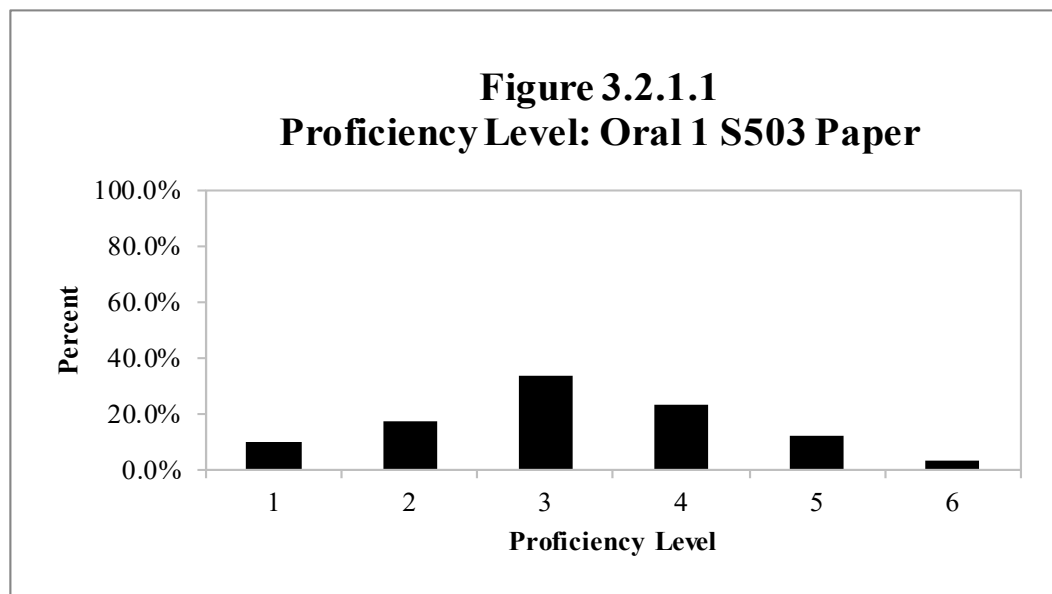


### 3.2.1.1 Grade 1

**Table 3.2.1.1**

Proficiency Level Distribution: Oral 1 S503 Paper

Level	Count	Percent
1	2,834	10.04%
2	4,965	17.58%
3	9,546	33.80%
4	6,581	23.30%
5	3,401	12.04%
6	913	3.23%
Total	28,240	100.00%

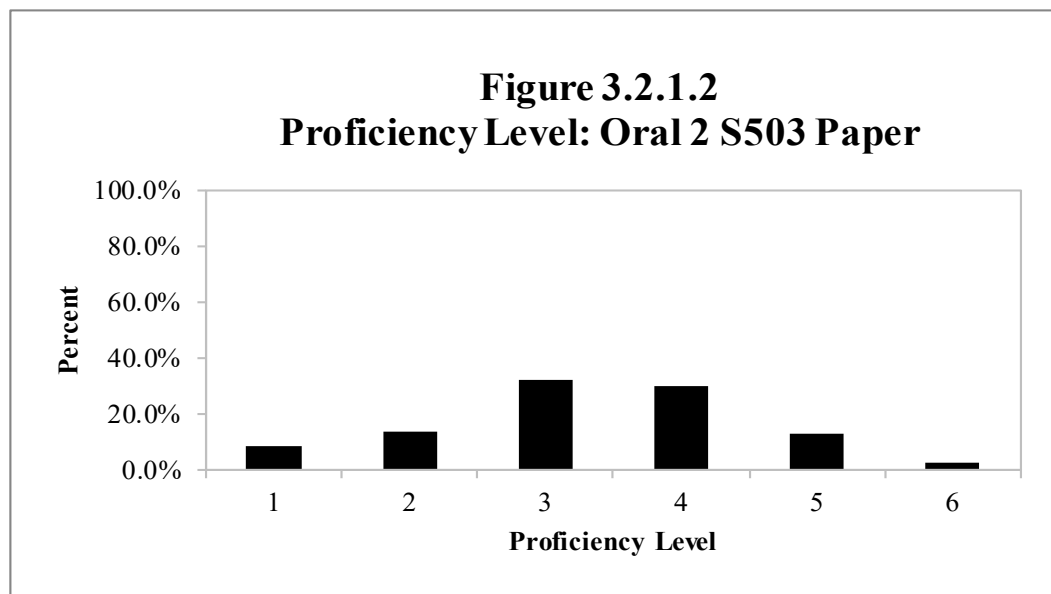


### 3.2.1.2 Grade 2

**Table 3.2.1.2**

Proficiency Level Distribution: Oral 2 S503 Paper

Level	Count	Percent
1	2,475	8.30%
2	4,027	13.50%
3	9,643	32.32%
4	8,981	30.10%
5	3,856	12.92%
6	852	2.86%
Total	29,834	100.00%

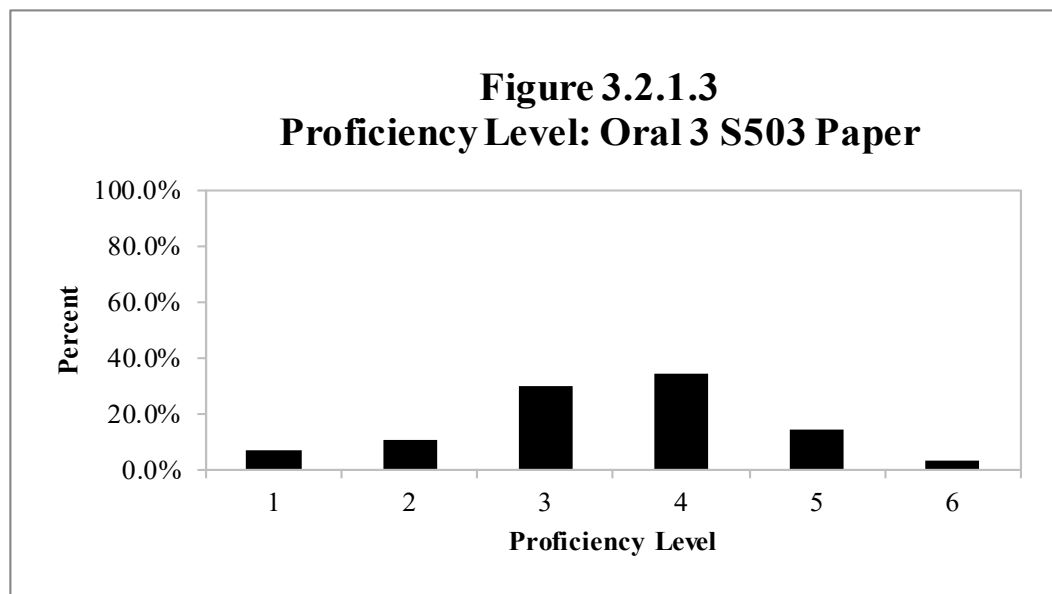


### 3.2.1.3 Grade 3

**Table 3.2.1.3**

Proficiency Level Distribution: Oral 3 S503 Paper

Level	Count	Percent
1	1,950	6.68%
2	3,181	10.89%
3	8,810	30.17%
4	9,995	34.23%
5	4,214	14.43%
6	1,050	3.60%
Total	29,200	100.00%



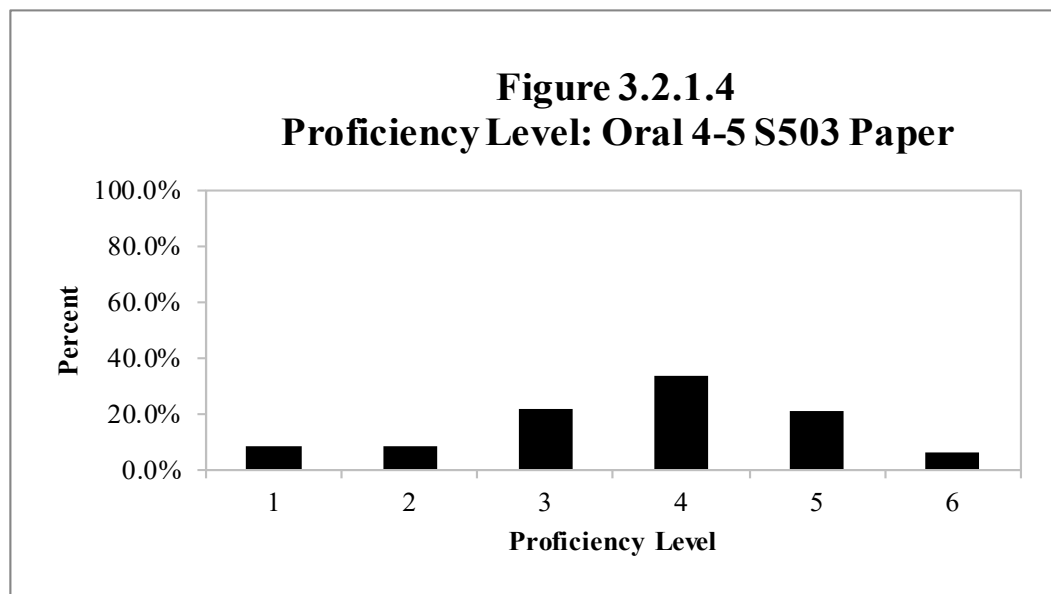


### 3.2.1.4 Grades 4-5

**Table 3.2.1.4**

Proficiency Level Distribution: Oral 4-5 S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	1,745	7.79%	2,025	9.00%	3,770	8.39%
2	1,990	8.88%	1,924	8.55%	3,914	8.71%
3	4,951	22.09%	4,976	22.10%	9,927	22.10%
4	7,422	33.12%	7,681	34.12%	15,103	33.62%
5	4,689	20.92%	4,653	20.67%	9,342	20.80%
6	1,612	7.19%	1,253	5.57%	2,865	6.38%
Total	22,409	100.00%	22,512	100.00%	44,921	100.00%

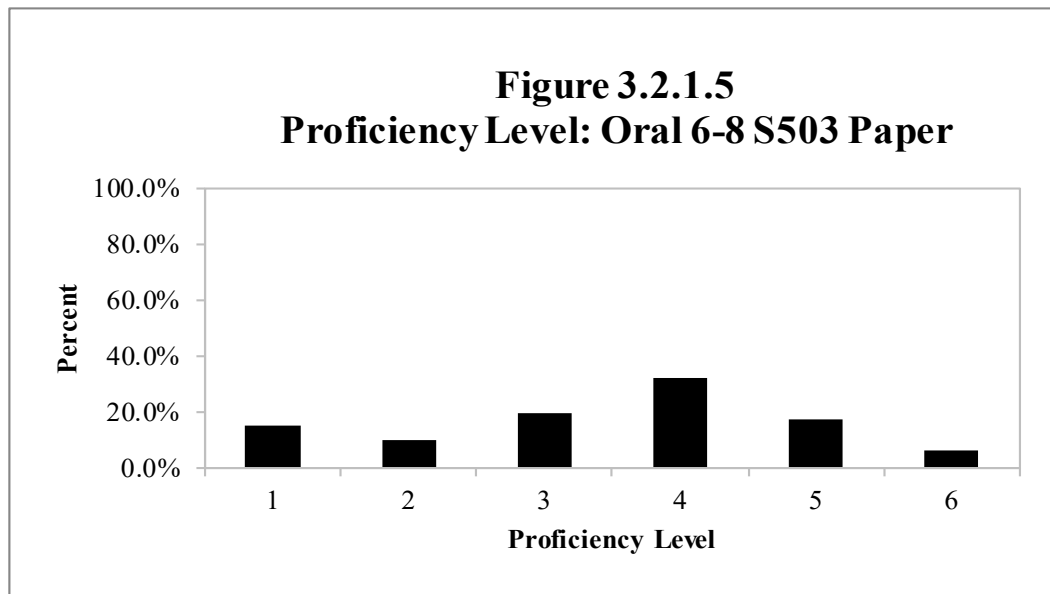


### 3.2.1.5 Grades 6-8

**Table 3.2.1.5**

Proficiency Level Distribution: Oral 6-8 S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,068	13.13%	2,332	15.99%	2,360	16.85%	6,760	15.25%
2	1,545	9.81%	1,439	9.87%	1,365	9.75%	4,349	9.81%
3	2,969	18.85%	2,989	20.49%	2,757	19.69%	8,715	19.66%
4	5,152	32.72%	4,573	31.35%	4,510	32.21%	14,235	32.11%
5	2,945	18.70%	2,436	16.70%	2,208	15.77%	7,589	17.12%
6	1,068	6.78%	817	5.60%	802	5.73%	2,687	6.06%
Total	15,747	100.00%	14,586	100.00%	14,002	100.00%	44,335	100.00%

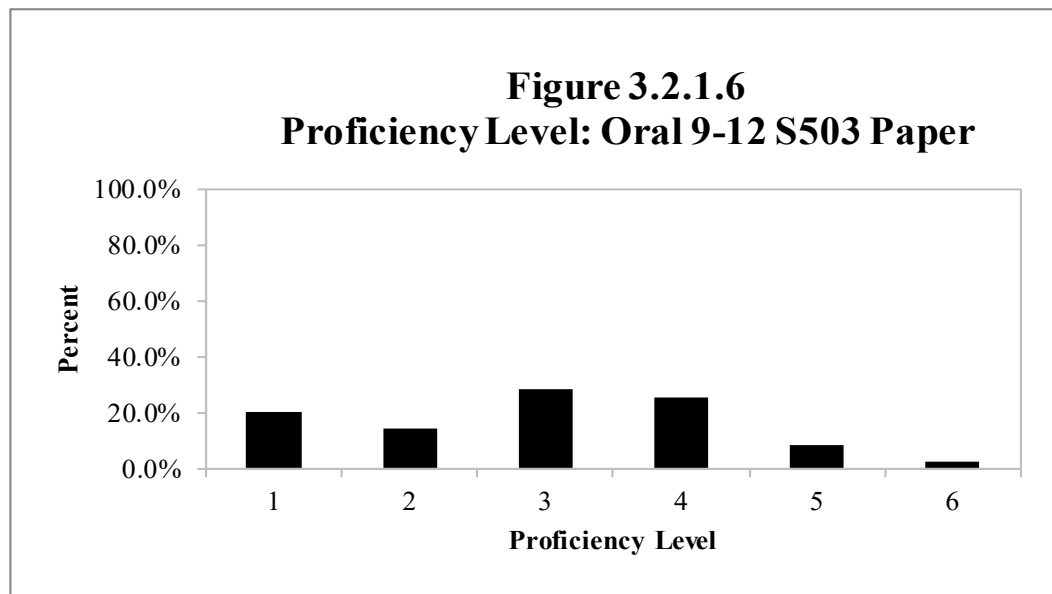


### 3.2.1.6 Grades 9-12

**Table 3.2.1.6**

Proficiency Level Distribution: Oral 9-12 S503 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	3,037	23.83%	2,279	20.02%	1,971	19.54%	1,011	14.98%	8,298	20.26%
2	1,851	14.53%	1,643	14.44%	1,503	14.90%	1,035	15.33%	6,032	14.73%
3	3,018	23.69%	3,113	27.35%	3,117	30.90%	2,358	34.93%	11,606	28.33%
4	3,330	26.13%	3,021	26.54%	2,312	22.92%	1,715	25.40%	10,378	25.34%
5	1,131	8.88%	1,017	8.94%	906	8.98%	526	7.79%	3,580	8.74%
6	375	2.94%	308	2.71%	279	2.77%	106	1.57%	1,068	2.61%
Total	12,742	100.00%	11,381	100.00%	10,088	100.00%	6,751	100.00%	40,962	100.00%



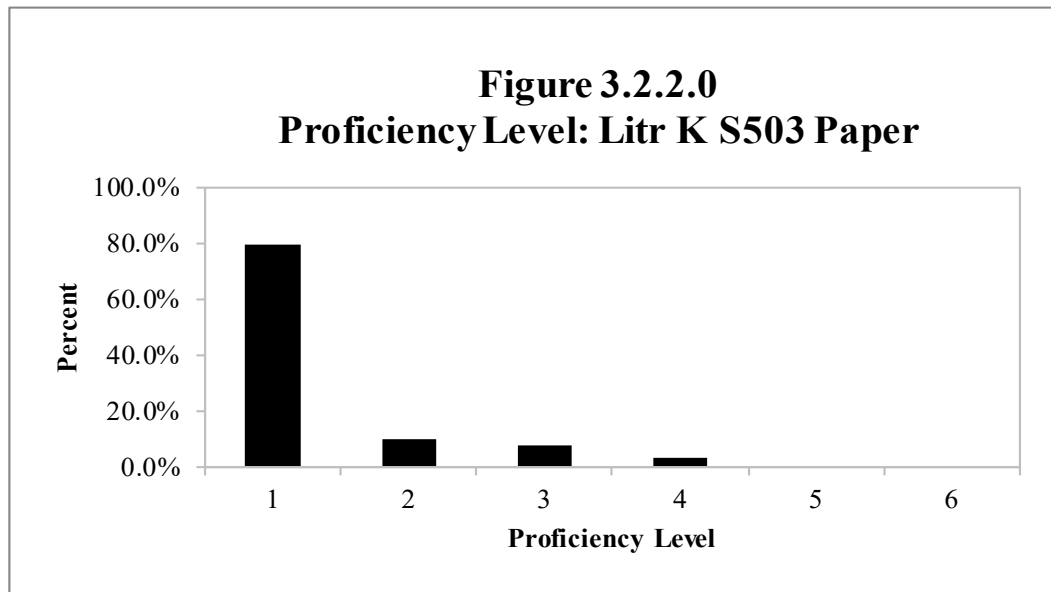
## 3.2.2 Literacy

### 3.2.2.0 Kindergarten

**Table 3.2.2.0**

Proficiency Level Distribution: Litr K S503 Paper

Level	Count	Percent
1	199,065	79.16%
2	24,468	9.73%
3	19,645	7.81%
4	8,308	3.30%
5	0	0.00%
6	0	0.00%
Total	251,486	100.00%

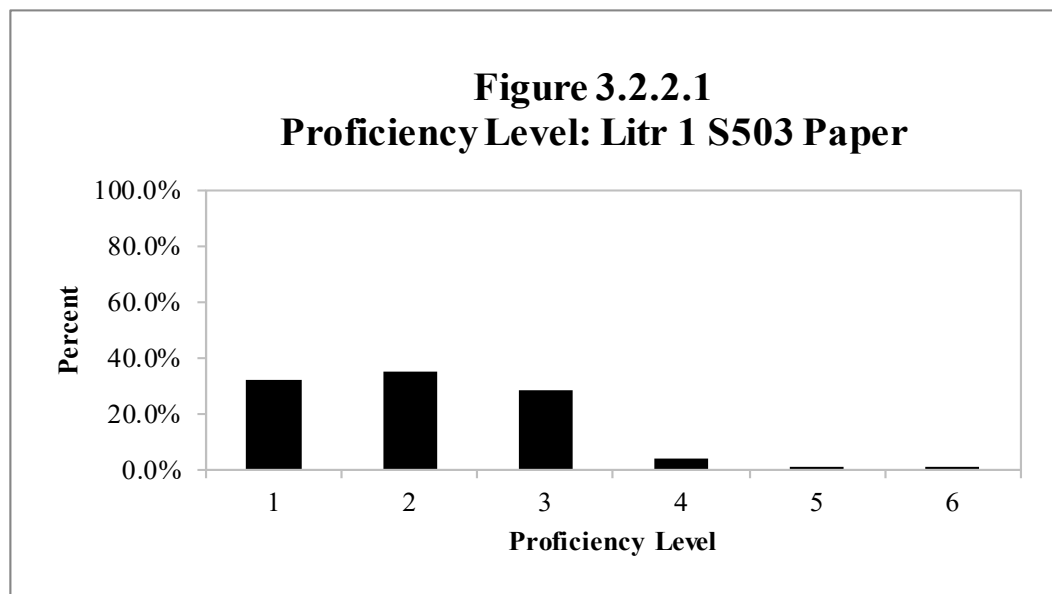


### 3.2.2.1 Grade 1

**Table 3.2.2.1**

Proficiency Level Distribution: Litr 1 S503 Paper

Level	Count	Percent
1	8,312	32.14%
2	9,099	35.19%
3	7,348	28.42%
4	974	3.77%
5	120	0.46%
6	6	0.02%
Total	25,859	100.00%

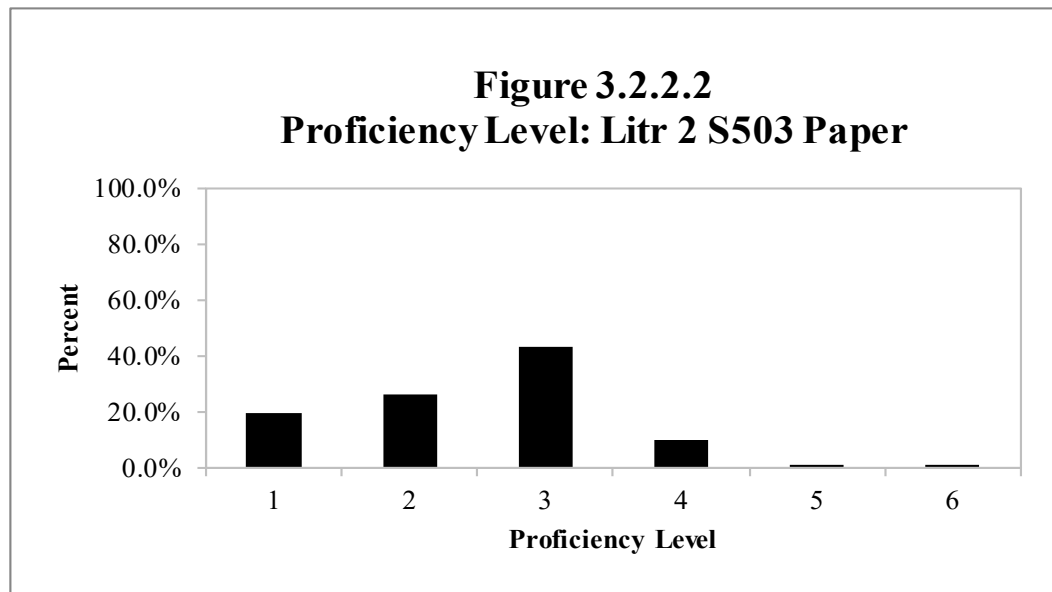


### 3.2.2.2 Grade 2

**Table 3.2.2.2**

Proficiency Level Distribution: Litr 2 S503 Paper

Level	Count	Percent
1	5,335	19.86%
2	7,048	26.24%
3	11,536	42.95%
4	2,738	10.19%
5	201	0.75%
6	2	0.01%
Total	26,860	100.00%

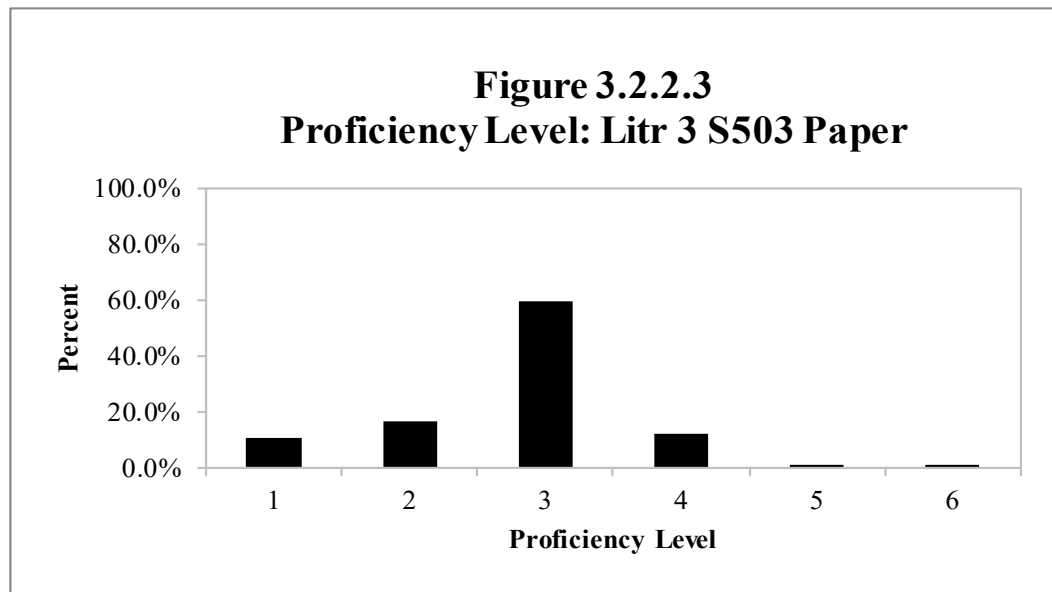


### 3.2.2.3 Grade 3

**Table 3.2.2.3**

Proficiency Level Distribution: Litr 3 S503 Paper

Level	Count	Percent
1	2,905	11.07%
2	4,423	16.86%
3	15,675	59.74%
4	3,133	11.94%
5	97	0.37%
6	6	0.02%
Total	26,239	100.00%



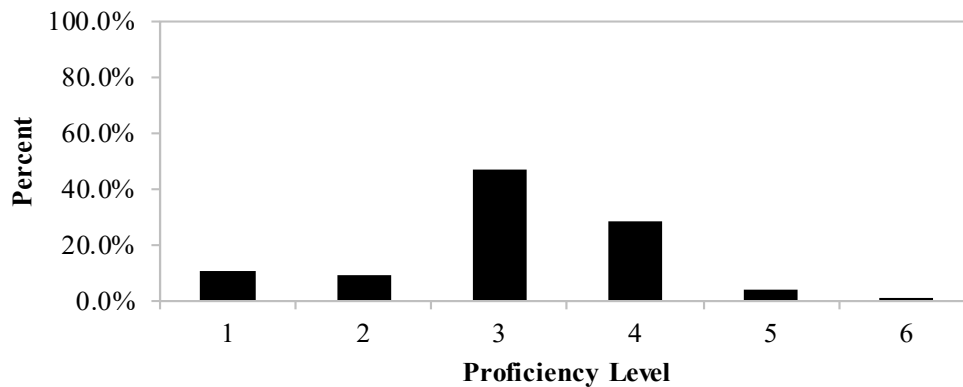
### 3.2.2.4 Grades 4-5

**Table 3.2.2.4**

Proficiency Level Distribution: Litr 4-5 S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	2,295	11.13%	2,126	10.18%	4,421	10.65%
2	1,876	9.10%	1,970	9.43%	3,846	9.27%
3	10,299	49.95%	9,178	43.94%	19,477	46.92%
4	5,458	26.47%	6,410	30.69%	11,868	28.59%
5	619	3.00%	1,080	5.17%	1,699	4.09%
6	73	0.35%	124	0.59%	197	0.47%
Total	20,620	100.00%	20,888	100.00%	41,508	100.00%

**Figure 3.2.2.4**  
**Proficiency Level: Litr 4-5 S503 Paper**



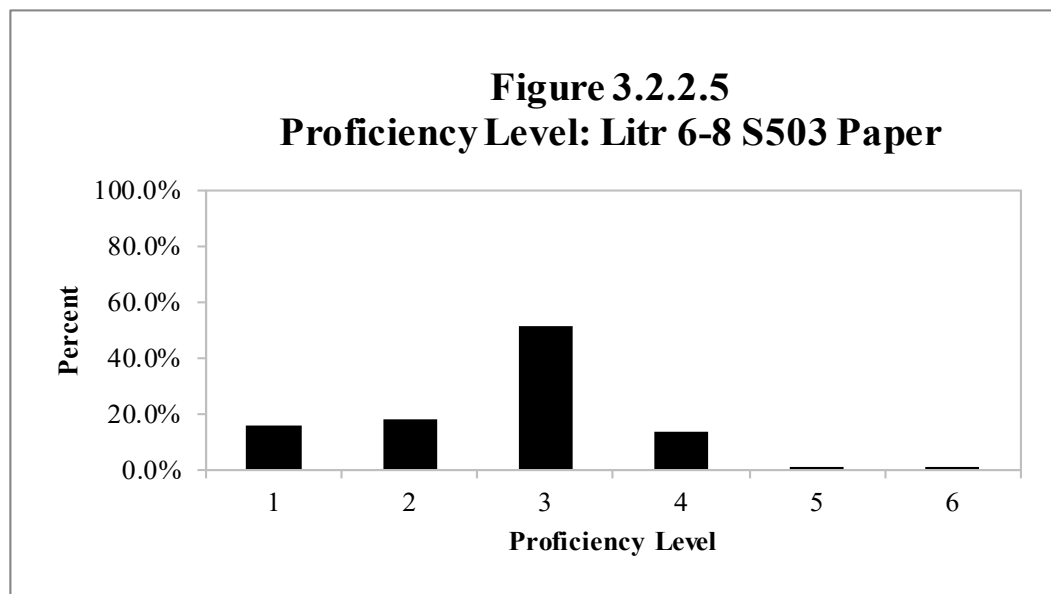


### 3.2.2.5 Grades 6-8

**Table 3.2.2.5**

Proficiency Level Distribution: Litr 6-8 S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,142	14.93%	2,202	16.32%	2,292	17.47%	6,636	16.20%
2	2,550	17.78%	2,531	18.75%	2,407	18.34%	7,488	18.28%
3	7,725	53.85%	6,983	51.74%	6,270	47.78%	20,978	51.21%
4	1,867	13.01%	1,721	12.75%	2,072	15.79%	5,660	13.82%
5	58	0.40%	59	0.44%	81	0.62%	198	0.48%
6	3	0.02%	0	0.00%	0	0.00%	3	0.01%
Total	14,345	100.00%	13,496	100.00%	13,122	100.00%	40,963	100.00%

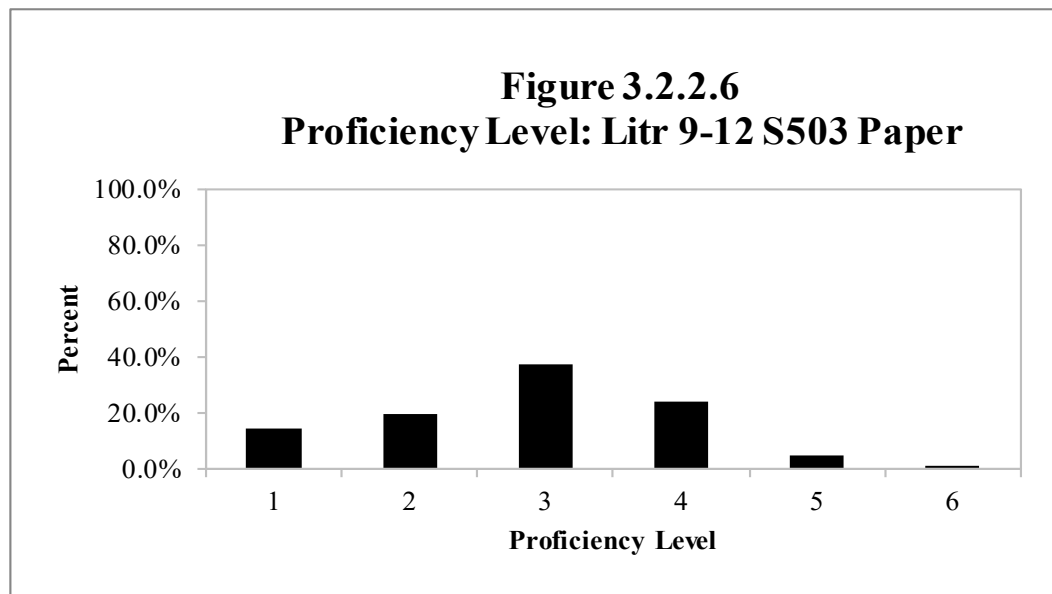


### 3.2.2.6 Grades 9-12

**Table 3.2.2.6**

Proficiency Level Distribution: Litr 9-12 S503 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,936	16.20%	1,381	12.95%	1,297	13.57%	870	13.44%	5,484	14.19%
2	2,313	19.36%	2,023	18.97%	1,827	19.12%	1,343	20.75%	7,506	19.42%
3	4,108	34.38%	3,980	37.31%	3,577	37.44%	2,711	41.89%	14,376	37.20%
4	2,982	24.96%	2,719	25.49%	2,330	24.39%	1,337	20.66%	9,368	24.24%
5	576	4.82%	549	5.15%	514	5.38%	211	3.26%	1,850	4.79%
6	33	0.28%	15	0.14%	10	0.10%	0	0.00%	58	0.15%
Total	11,948	100.00%	10,667	100.00%	9,555	100.00%	6,472	100.00%	38,642	100.00%



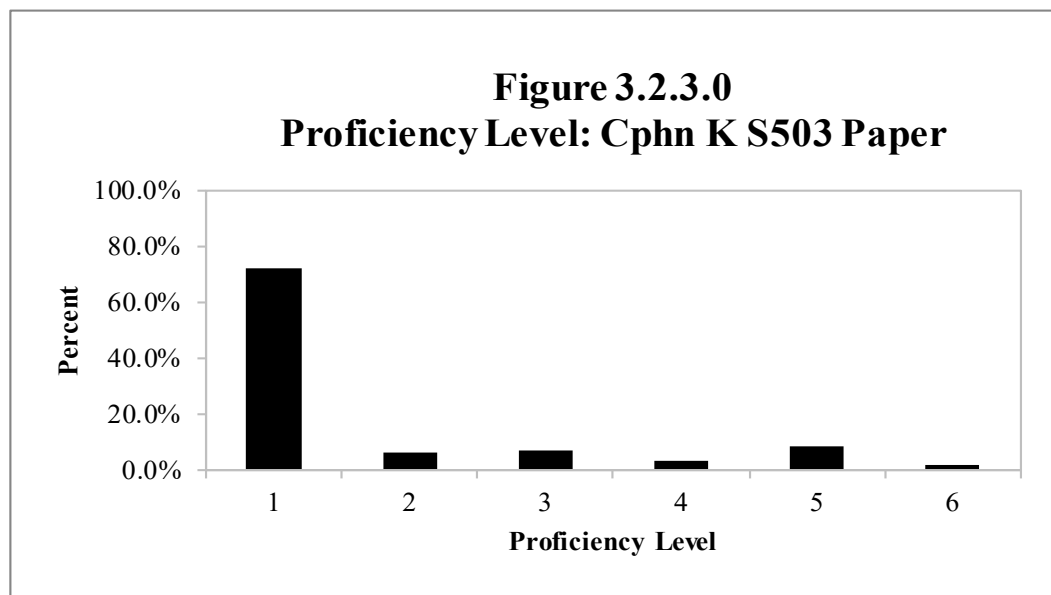
### 3.2.3 Comprehension

#### 3.2.3.0 Kindergarten

**Table 3.2.3.0**

Proficiency Level Distribution: Cphn K S503 Paper

Level	Count	Percent
1	180,528	71.78%
2	16,520	6.57%
3	18,297	7.28%
4	9,157	3.64%
5	21,630	8.60%
6	5,356	2.13%
Total	251,488	100.00%

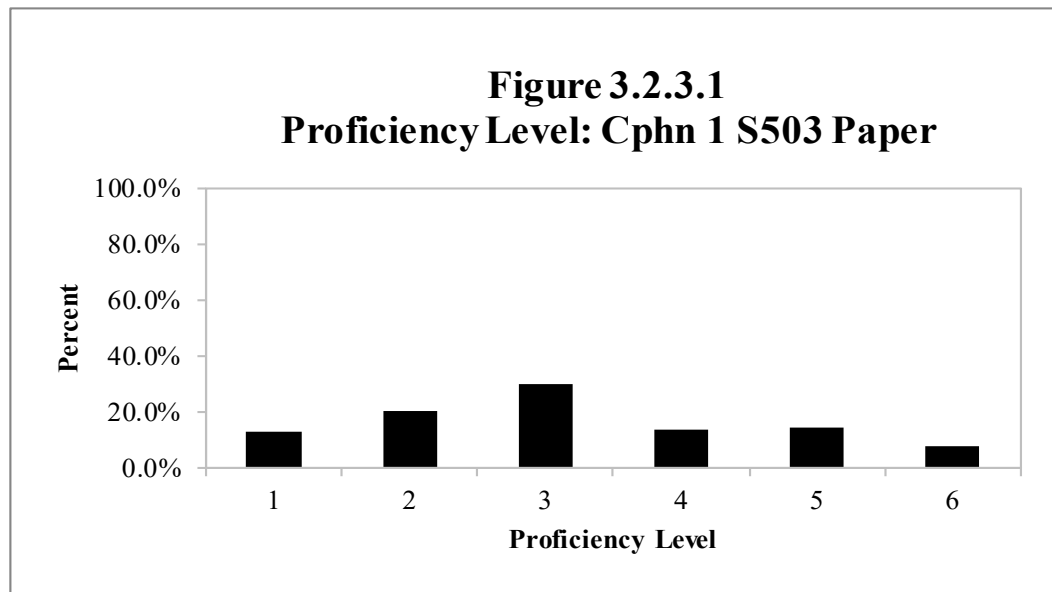


### 3.2.3.1 Grade 1

**Table 3.2.3.1**

Proficiency Level Distribution: Cphn 1 S503 Paper

Level	Count	Percent
1	3,006	13.12%
2	4,625	20.19%
3	6,926	30.23%
4	3,216	14.04%
5	3,300	14.40%
6	1,840	8.03%
Total	22,913	100.00%

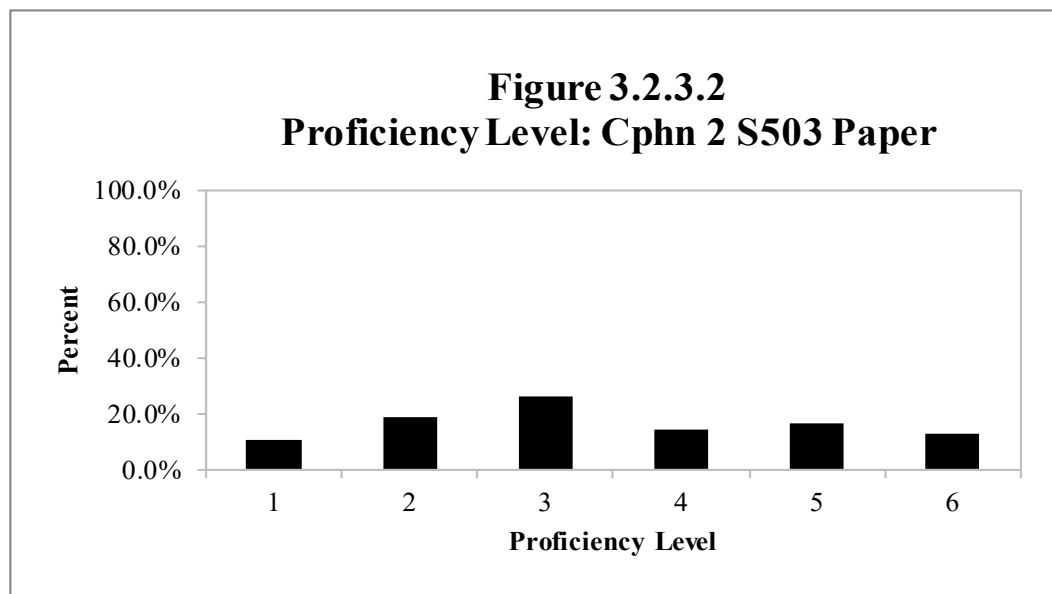


### 3.2.3.2 Grade 2

**Table 3.2.3.2**

Proficiency Level Distribution: Cphn 2 S503 Paper

Level	Count	Percent
1	2,781	10.96%
2	4,845	19.09%
3	6,646	26.18%
4	3,568	14.06%
5	4,209	16.58%
6	3,334	13.13%
Total	25,383	100.00%

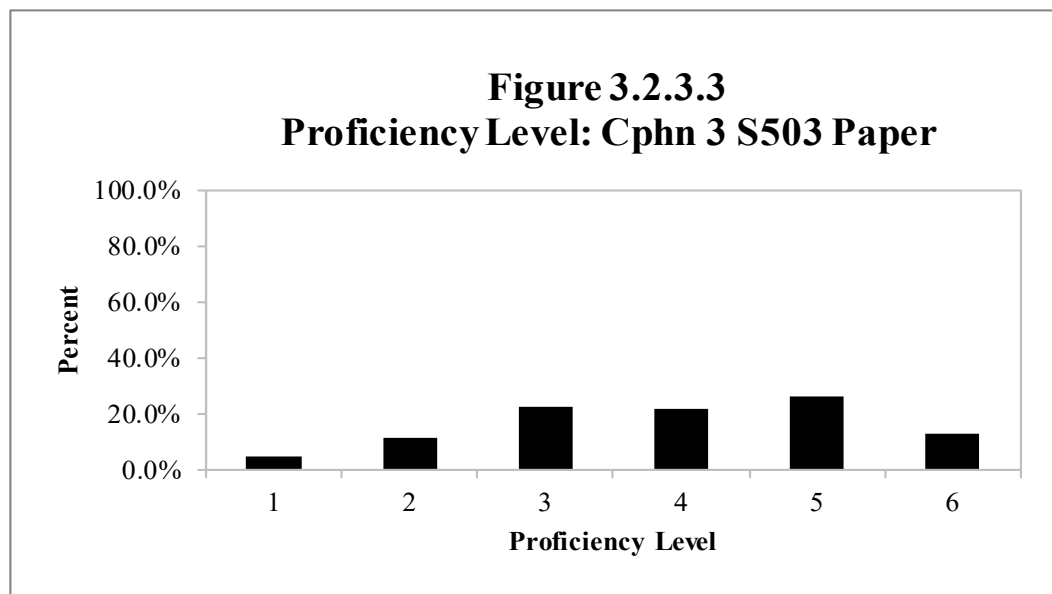


### 3.2.3.3 Grade 3

**Table 3.2.3.3**

Proficiency Level Distribution: Cphn 3 S503 Paper

Level	Count	Percent
1	1,239	4.95%
2	2,824	11.29%
3	5,731	22.90%
4	5,452	21.79%
5	6,590	26.34%
6	3,186	12.73%
Total	25,022	100.00%

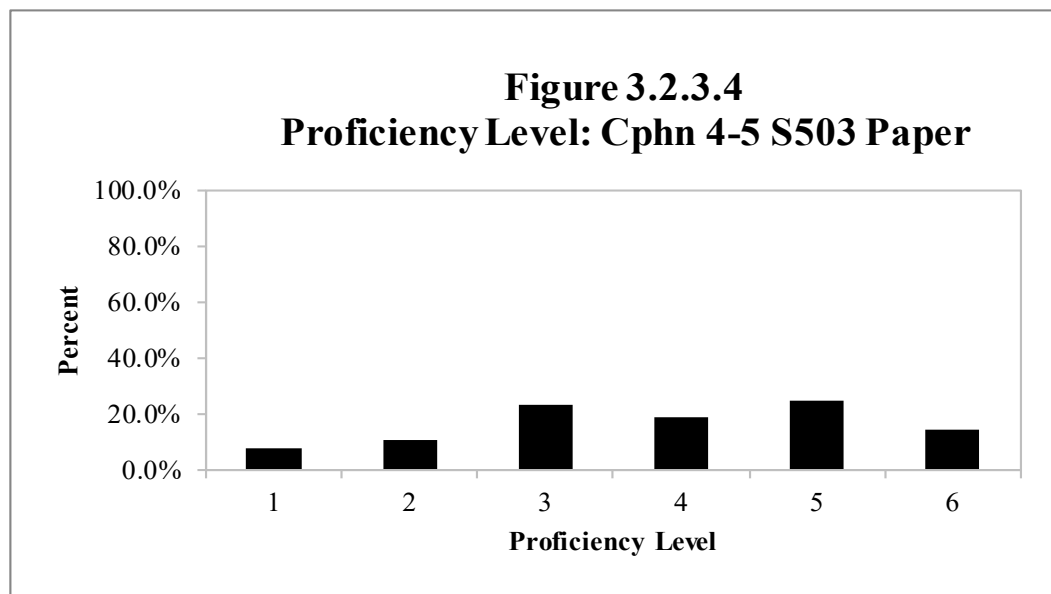


### 3.2.3.4 Grades 4-5

**Table 3.2.3.4**

Proficiency Level Distribution: Cphn 4-5 S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	1,414	7.10%	1,752	8.63%	3,166	7.87%
2	2,056	10.33%	2,110	10.39%	4,166	10.36%
3	4,525	22.73%	4,971	24.49%	9,496	23.62%
4	4,083	20.51%	3,561	17.54%	7,644	19.01%
5	5,070	25.47%	4,964	24.45%	10,034	24.96%
6	2,757	13.85%	2,941	14.49%	5,698	14.17%
Total	19,905	100.00%	20,299	100.00%	40,204	100.00%

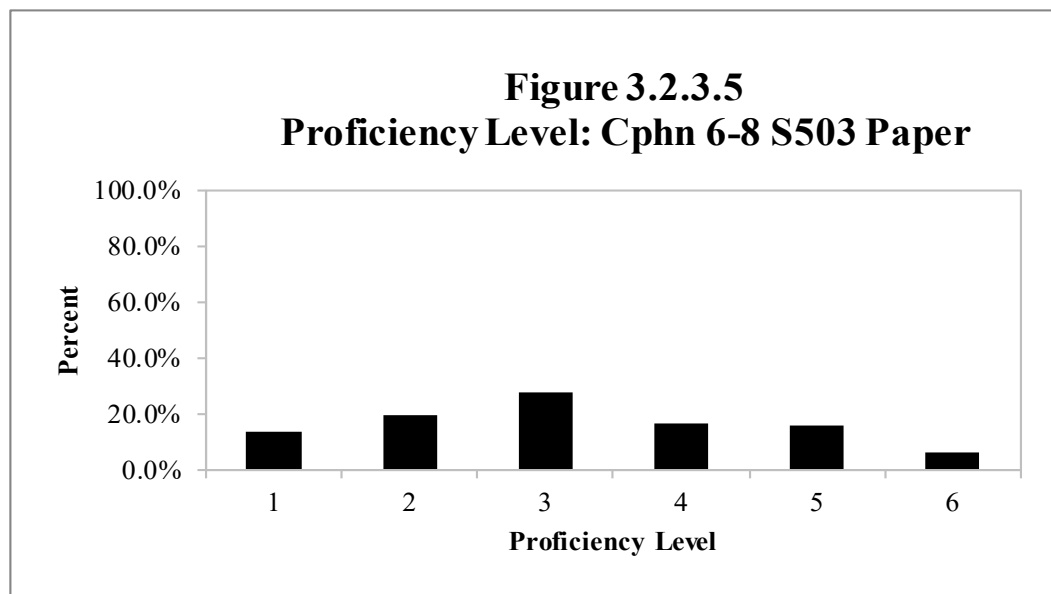


### 3.2.3.5 Grades 6-8

**Table 3.2.3.5**

Proficiency Level Distribution: Cphn 6-8 S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,586	11.49%	1,769	13.62%	1,916	15.21%	5,271	13.38%
2	2,648	19.19%	2,649	20.39%	2,451	19.45%	7,748	19.67%
3	3,990	28.91%	3,730	28.72%	3,201	25.41%	10,921	27.73%
4	2,533	18.35%	1,994	15.35%	1,907	15.14%	6,434	16.33%
5	2,249	16.30%	1,963	15.11%	2,189	17.37%	6,401	16.25%
6	795	5.76%	884	6.81%	935	7.42%	2,614	6.64%
Total	13,801	100.00%	12,989	100.00%	12,599	100.00%	39,389	100.00%



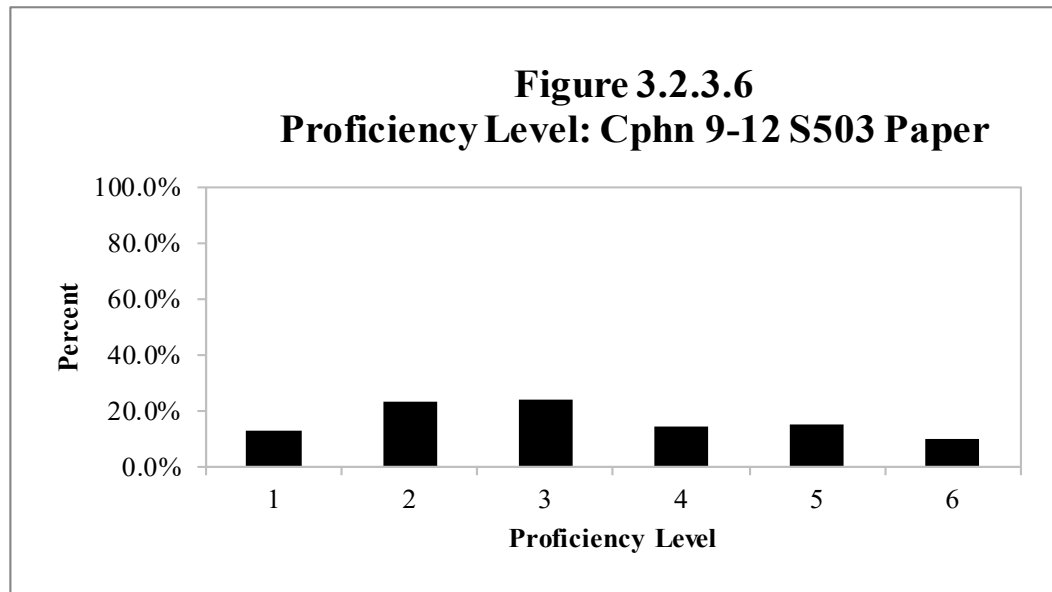


### 3.2.3.6 Grades 9-12

**Table 3.2.3.6**

Proficiency Level Distribution: Cphn 9-12 S503 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,735	15.24%	1,216	11.89%	1,204	13.14%	680	10.92%	4,835	13.07%
2	2,582	22.67%	2,334	22.83%	2,147	23.42%	1,577	25.33%	8,640	23.35%
3	2,556	22.44%	2,538	24.82%	2,133	23.27%	1,690	27.14%	8,917	24.10%
4	1,682	14.77%	1,511	14.78%	1,205	13.15%	956	15.35%	5,354	14.47%
5	1,757	15.43%	1,458	14.26%	1,539	16.79%	874	14.04%	5,628	15.21%
6	1,076	9.45%	1,167	11.41%	938	10.23%	449	7.21%	3,630	9.81%
Total	11,388	100.00%	10,224	100.00%	9,166	100.00%	6,226	100.00%	37,004	100.00%



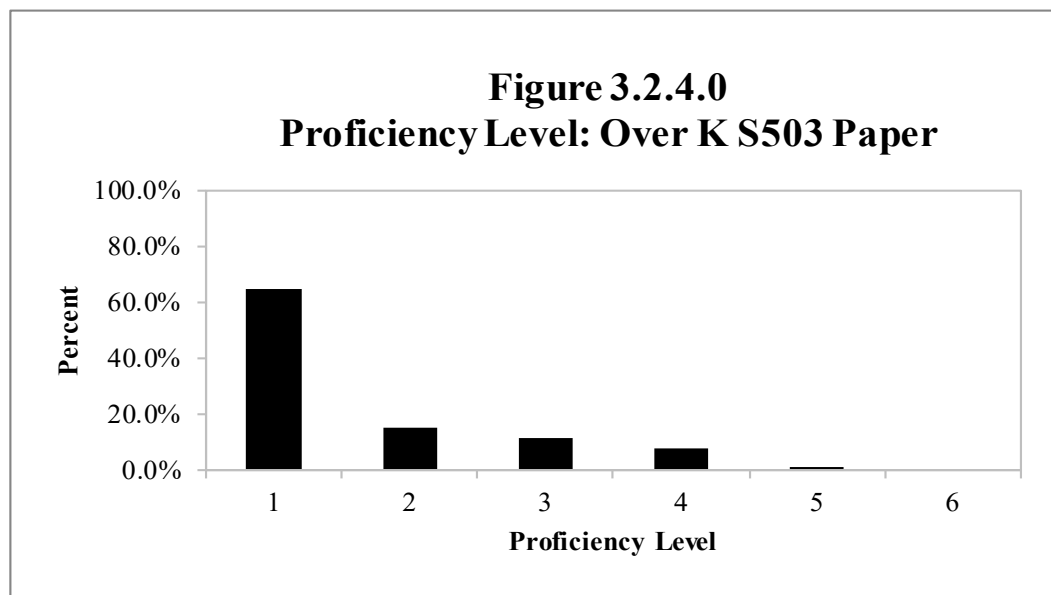
### 3.2.4 Overall

#### 3.2.4.0 Kindergarten

**Table 3.2.4.0**

Proficiency Level Distribution: Over K S503 Paper

Level	Count	Percent
1	162,955	64.80%
2	37,497	14.91%
3	29,327	11.66%
4	18,654	7.42%
5	3,041	1.21%
6	0	0.00%
Total	251,474	100.00%

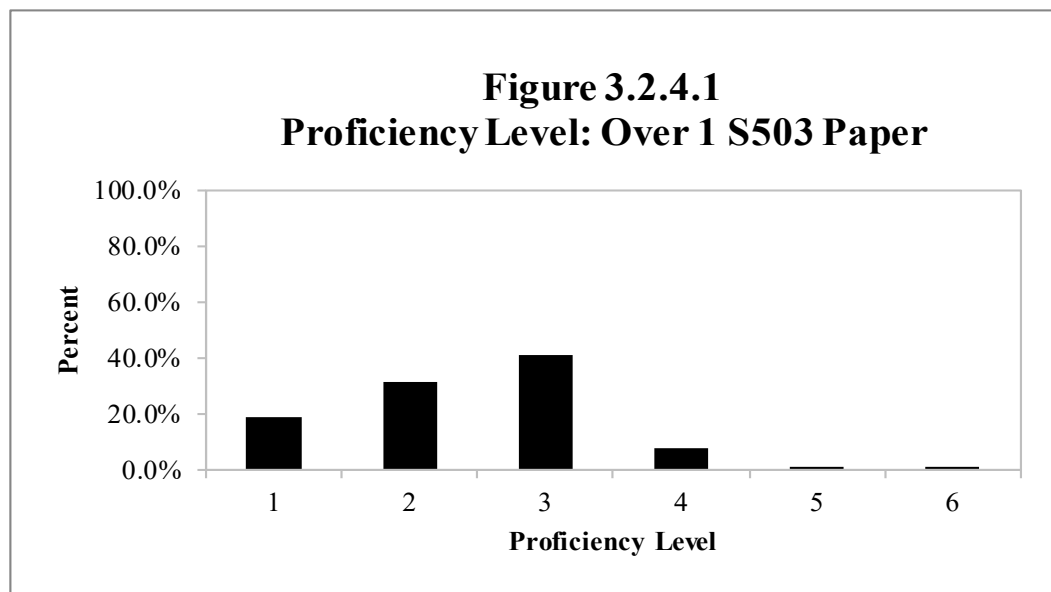


### 3.2.4.1 Grade 1

**Table 3.2.4.1**

Proficiency Level Distribution: Over 1 S503 Paper

Level	Count	Percent
1	4,317	18.95%
2	7,074	31.05%
3	9,354	41.05%
4	1,778	7.80%
5	244	1.07%
6	18	0.08%
Total	22,785	100.00%

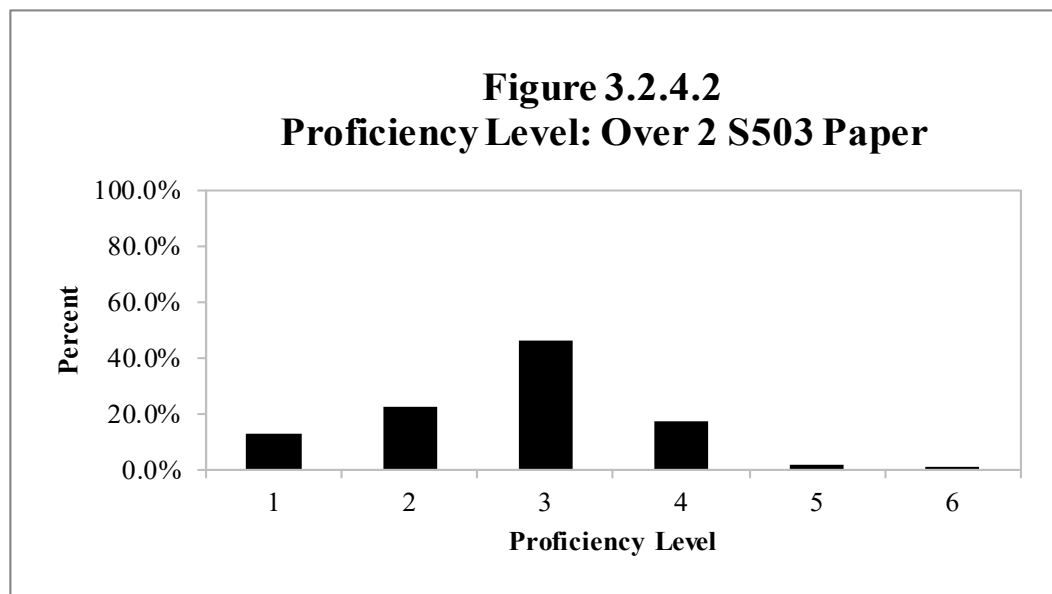


### 3.2.4.2 Grade 2

**Table 3.2.4.2**

Proficiency Level Distribution: Over 2 S503 Paper

Level	Count	Percent
1	3,168	12.56%
2	5,605	22.22%
3	11,605	46.01%
4	4,389	17.40%
5	449	1.78%
6	5	0.02%
Total	25,221	100.00%

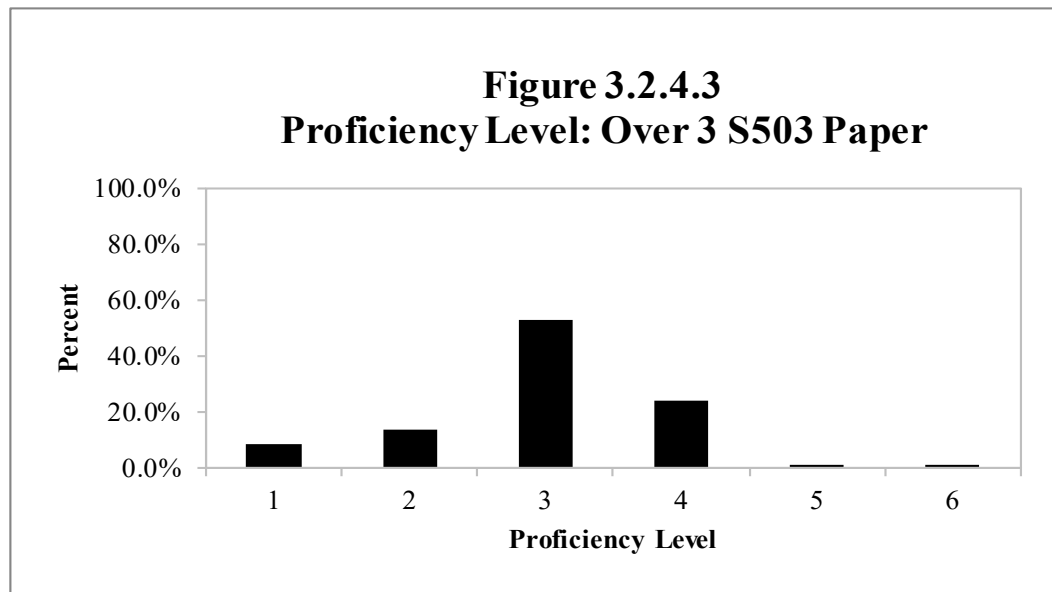


### 3.2.4.3 Grade 3

**Table 3.2.4.3**

Proficiency Level Distribution: Over 3 S503 Paper

Level	Count	Percent
1	2,072	8.33%
2	3,336	13.42%
3	13,213	53.13%
4	5,894	23.70%
5	342	1.38%
6	10	0.04%
Total	24,867	100.00%

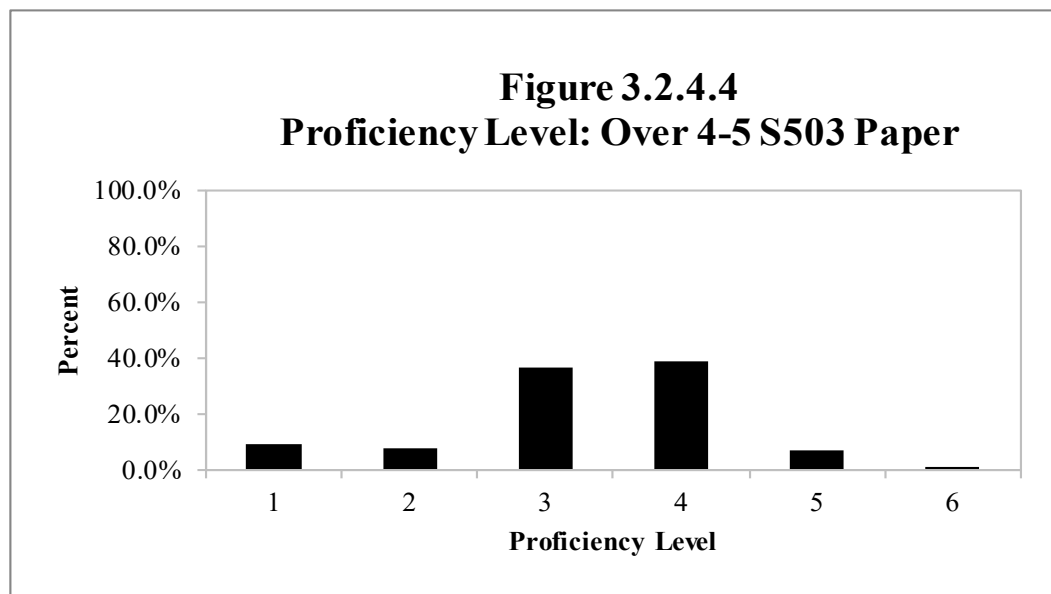


### 3.2.4.4 Grades 4-5

**Table 3.2.4.4**

Proficiency Level Distribution: Over 4-5 S503 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	1,849	9.37%	1,869	9.27%	3,718	9.32%
2	1,585	8.03%	1,612	7.99%	3,197	8.01%
3	7,572	38.36%	7,132	35.37%	14,704	36.85%
4	7,521	38.10%	7,900	39.18%	15,421	38.64%
5	1,134	5.74%	1,543	7.65%	2,677	6.71%
6	80	0.41%	108	0.54%	188	0.47%
Total	19,741	100.00%	20,164	100.00%	39,905	100.00%

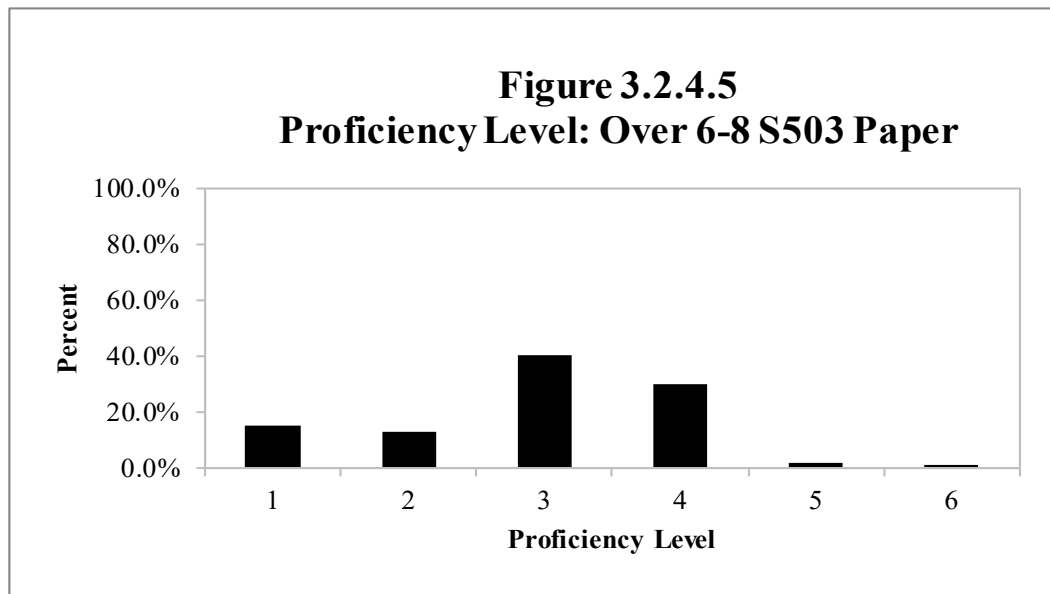


### 3.2.4.5 Grades 6-8

**Table 3.2.4.5**

Proficiency Level Distribution: Over 6-8 S503 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,885	13.76%	2,044	15.84%	2,117	16.92%	6,046	15.46%
2	1,721	12.57%	1,656	12.84%	1,615	12.91%	4,992	12.76%
3	5,661	41.34%	5,307	41.14%	4,745	37.92%	15,713	40.18%
4	4,192	30.61%	3,691	28.61%	3,798	30.35%	11,681	29.87%
5	227	1.66%	198	1.53%	236	1.89%	661	1.69%
6	9	0.07%	4	0.03%	2	0.02%	15	0.04%
Total	13,695	100.00%	12,900	100.00%	12,513	100.00%	39,108	100.00%

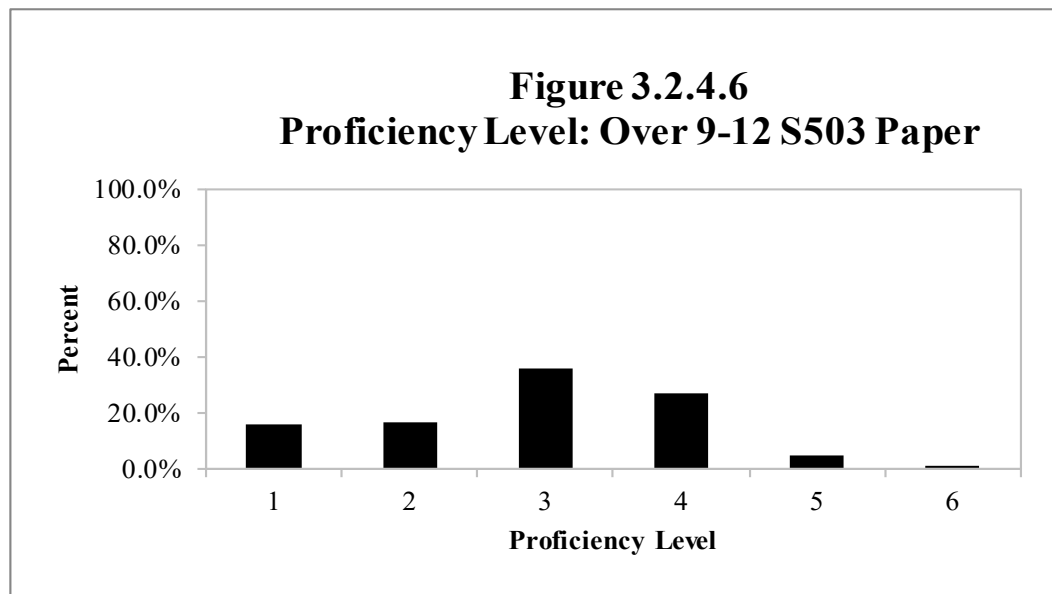


### 3.2.4.6 Grades 9-12

**Table 3.2.4.6**

Proficiency Level Distribution: Over 9-12 S503 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,064	18.30%	1,523	15.06%	1,411	15.55%	772	12.57%	5,770	15.76%
2	1,839	16.31%	1,624	16.06%	1,458	16.07%	1,137	18.51%	6,058	16.55%
3	3,701	32.82%	3,555	35.16%	3,334	36.75%	2,551	41.52%	13,141	35.90%
4	3,048	27.03%	2,872	28.40%	2,404	26.50%	1,487	24.20%	9,811	26.80%
5	591	5.24%	517	5.11%	455	5.02%	197	3.21%	1,760	4.81%
6	34	0.30%	20	0.20%	10	0.11%	0	0.00%	64	0.17%
Total	11,277	100.00%	10,111	100.00%	9,072	100.00%	6,144	100.00%	36,604	100.00%





## 4. Annual Updates of Validity Evidence

This section presents studies conducted as validity evidence for the WIDA ACCESS assessments. 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. Interpretations for specified uses begin by specifying the construct the test is intended to measure. Rather than referring to distinct types of validity, the 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 guidance to the U.S. Department of Education’s Assessment Peer Review Process” document (Department of Education, 2018 <https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf>) to support states’ use of ELP assessments for reviewing of validity evidence, as well as is linked to Assessment User Argument (AUA) to support the claims of validity of Online ACCESS assessment. WIDA structures its validity arguments using AUA model in lieu of the model highlighted in the *Standards for Educational and Psychological Testing*. AUA has similar topics; however, they are organized differently. Below is a short summary of each AUA claim. For the full AUA validity claims please refer to WIDA Assessment User Argument document.

**Claim 1 (Consequences):** With the use of ACCESS, the intended decisions will have beneficial consequences for stakeholders, in terms of using ACCESS and the decisions made based on ACCESS.

**Claim 2 (Decisions):** Decisions based on ACCESS test results are made by individuals, in a timely manner and affect a variety of stakeholders. Two types of decisions 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 and composites 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 (ELD) Standards; *generalizable* to the academic English language used in K–12 instructional settings, and *impartial* to all students.

**Claim 4 (Assessment records: Scores):** 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.

## **4.1.      *Standards***

### **4.1.1. 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 appropriate the content reflects the domain as we interpret test scores.

### **4.1.2. 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.

### **4.1.3. Internal Structure**

Validity related to internal structure indicates how test items/components agree with the construct we base for the score interpretation. The internal structure of the construct can be unidimensional or contain multidimensional components.

### **4.1.4. Relation to Other Variables**

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.

## 5 Reliability

In accordance with the *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014), when interpreting test scores, it is important to evaluate their reliability, as the interpretation of test scores depends on the assumption that students exhibit some degree of consistency in their scores across independent administrations of the same testing procedure. We expect that students mastering the domain will consistently perform well, and those who have not mastered the domain will consistently perform less well, regardless of the sample of items and tasks used to assess students. Furthermore, because we assume that all items and tasks on such a test measure some aspect of the domain of interest, we expect that students will perform consistently across different items and tasks measuring the same ability within the test. Therefore, it is important to evaluate the degree to which students' test scores are consistent across replications of the same testing condition.

However, different samples of performances from the same student are rarely identical. A student's responses to sets of test items or tasks vary from one sample of test items or tasks targeting the domain to another, and from one occasion to another, even under strictly controlled conditions. In addition, different raters may award different scores to the same student performance on a test task. These sources of variation are reflected in the students' scores. Therefore, it is important to evaluate the extent to which differences in students' test scores reflect true differences in the knowledge, skills, or ability being tested, rather than fluctuations due to chance.

The reliability of the test scores depends on how much the scores vary across replications of the testing procedure, and analyses of reliability depend on the types of variability likely to be of concern in the testing procedure. There are several ways to collect reliability data and to estimate reliability, some of which depend on the exact nature of the measurement, the intended use of the test scores, the assessment design, and the potential sources of measurement error that might contribute to inconsistency in students' scores across different test administrations.

The reliability information presented in this section is organized to be in compliance with Critical Element 4.1 of the Every Student Succeeds Act Peer Review requirements (U.S. Department of Education, 2018) and follows the guidelines of the *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014). We present information regarding the reliability of the domain scale scores first, followed by information about the reliability of the composite scale scores.

Policy makers in states and districts use ACCESS Listening, Reading, Writing, and Speaking tests to determine the English language proficiency of students based on their scores in each of the four domains. Therefore, the main concern in interpreting these scores is how consistent the scores would be over replications of the same testing procedure. We use **internal consistency reliability statistics** to address this question (Section 5.1).

Additionally, for the Writing and Speaking domains, because having different raters evaluate the same students' responses to tasks may result in inconsistent scoring, a potential source of variation of those scores is the rater. In Section 5.2, we report the **interrater agreement** rates that the raters achieved when evaluating students' responses to the Writing and Speaking tasks. We can use these statistics to determine how consistent the students' scores would have been if different raters had evaluated their responses. Since we use an item response theory (IRT)–based method to estimate students' **latent scores** (i.e., test scores based on variables that we cannot see or directly measure but which we can infer mathematically

through advanced statistical techniques by using students' scores on variables that we can observe), we also examine the amount of **measurement error** in students' scores using the **conditional standard error of measurement** (CSEM) (Section 5.3). Lastly, in Section 5.4, we evaluate the reliability of the classifications of students into WIDA proficiency levels based on their domain scores (the most important interpretation of the test scores) in terms of the **accuracy and consistency** of the classification decisions made. In each subsection, we present detailed descriptions of the methods, data sources, and procedures.

Policy makers in states and districts use ACCESS **composite scale scores** to describe the English language proficiency of students in the respective composites. Therefore, the most important concern in interpreting these scores is how consistent the scores would be over replications of the same testing procedure. We use internal consistency reliability statistics to address this question, and in Section 5.5 we provide the results. In addition, in Section 5.6, we examine the CSEM of these scores. Lastly, in Section 5.7, we evaluate the reliability of the classifications in terms of the accuracy and consistency of the decisions made about students' levels of English language proficiency based on their composite scale scores. In each subsection, we present detailed descriptions of the methods, data sources, and procedures.

### ***Internal Consistency Reliability Statistics***

One way to evaluate the consistency of students' test scores across test administrations is to examine how the students would have performed on alternate forms of the same test (i.e., **parallel test form reliability**). Given our assumption that the ability the test measures is constant for each student over two administrations of alternate forms, the more variation found across the two administrations, the more evidence for lower reliability. The **measurement error** represents the sources of inconsistency across the two administrations, taken together. We consider measurement error to be random and to occur by chance. For example, there may be some construct-irrelevant knowledge and/or skills that some items or tasks measure that affect students' scores but are not part of the ability that the test intends to measure.

Unless students take two alternate versions of the same test, we cannot calculate test score reliability directly. Thus, we usually estimate it from student responses to a single form of the test. Methods employed to estimate reliability using test scores from a single test administration are based on classical test theory and are referred to as estimates of **internal consistency**. An internal consistency reliability statistic is a useful estimate of alternate-forms reliability, providing an estimate of the consistency of students' performances across items and tasks within a test. The most common index of internal consistency reliability is **Cronbach's coefficient alpha** (Cronbach, 1951), which is a lower-bound estimate of test reliability. Conceptually, we think of Cronbach's coefficient alpha as the correlation obtained between performances on two halves of the same test if every possible way of dividing the test items and tasks in two were attempted. Because Cronbach's coefficient alpha is a correlation of students' performances on all possible pairs of test items and tasks, it may be low if some items or tasks are measuring something other than what most of the other items and tasks are measuring (and thus leading to inconsistent student performances). In this way, Cronbach's coefficient alpha expresses how well the items and tasks on a test appear to measure the same ability. The Cronbach's coefficient alpha of internal consistency ranges from 0 to 1. If students achieve their scores by a completely random process (i.e., their scores are not correlated or share no covariance), then the reliability estimate is very close to 0. On the other hand, if students' scores are perfectly consistent (i.e., their scores have high covariances), then the internal consistency coefficient will approach 1.

While there is no one set of criteria that the testing community uses when interpreting Cronbach's coefficient alpha values, from time to time, researchers have proposed various arbitrary criteria that one could apply. Initially, Cronbach (1951) argued that it was 'desirable' to have a high alpha value for an instrument that test developers were using to report individual scores since the scores on that instrument needed to be interpretable, and that would require a high alpha value. Later, Nunnally (1978) suggested that researchers should consider a value of 0.70 as an acceptable lower limit if they were engaged in the early stages of research (e.g., when developing a scale). Today, it has become common practice to cite Nunnally's suggested 0.70 criterion as a minimum acceptable lower limit for this value for all types of research. However, in so doing, researchers ignore Nunnally's more nuanced guidance: If researchers were engaged in basic research, Nunnally advised that they should use a higher cut-off value (i.e., 0.80 or higher), and those engaged in applied research should use a much higher cut-off value (0.90 or higher) (Lance et al., 2006). Since Nunnally's time, some researchers have suggested even more nuanced interpretations of various alpha values. For example, George and Mallery (2003) proposed the following interpretations: " $\geq 0.90$  – Excellent,  $\geq 0.80$  – Good,  $\geq 0.70$  – Acceptable,  $\geq 0.60$  – Questionable,  $\geq 0.50$  – Poor, and  $\leq 0.50$  – Unacceptable" (p. 231). Clearly, there is little consensus among the experts in their views of what the acceptable lower limit of the Cronbach's coefficient alpha value should be, or for that matter, how one should interpret various values. This lack of consensus led the authors of the *Standards for Educational and Psychological Measurement* (2014) to conclude, "The choice of [reliability/precision] estimation and the minimum acceptable level for any index remain a matter of professional judgment" (p. 41). For the purposes of this report then, WIDA has made the decision that within the domains of Listening, Reading, and Speaking, an alpha value of  $\geq 0.80$  is acceptable, while an alpha value of  $\geq 0.65$  is acceptable for the Writing domain.

Reliability statistics such as the Cronbach's coefficient alpha of internal consistency are affected by two factors: (1) the number of test items or tasks, and (2) the total number of score points students achieve. That is, all things being equal, the greater the number of items or tasks measuring the same ability there are on the test, the higher the internal consistency reliability statistics. Additionally, because reliability statistics refer to the consistency of scores *for a group of students*, the distribution of that specific group's ability measures affects these statistics. If the students in the group are nearly equal in the ability that the test measures (i.e., their scores are concentrated in the center of the ability distribution), small changes in their scores can easily change their relative positions in the group. Consequently, the internal consistency reliability statistics will be low. In this case, the statistic may be telling us more about the group of students tested than about the test itself. On the other hand, if the students in the group differ widely in the ability that the test measures (i.e., their scores are distributed across the ability continuum), small changes in their scores will not affect their relative positions in the group as much, and the internal consistency reliability statistics will be higher. Therefore, reliability can be as much a function of the performance of test items and tasks as of the performance of the sample of students tested. That is, the exact same test can produce widely disparate reliability indices based on the ability distribution of the group of students. This means, in turn, that when interpreting estimates of internal consistency, it is wise to keep in mind the specific set of test items and tasks and the distribution of ability measures in the group of students used in the estimation.

### ***Interrater Agreement***

The behavior of raters is a potential source of variance in students' scores for the productive domains of ACCESS (i.e., Writing and Speaking). ACCESS scoring procedures and rater training and quality control monitoring processes are described elsewhere in this report (see Part 1, Section 3.2.2). In Section 5.2, we report the **interrater agreement rates** for the scoring of students' responses to the Writing and Speaking

tasks. These values reflect how consistent the students' scores would be if different groups of raters scored their responses. Additionally, in this section of the report we present a detailed description of the methods, data sources, and procedures we used when calculating interrater agreement rates.

### ***Measurement Error***

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, 2018, p. 9). 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.

### ***Classification Accuracy and Consistency***

One of the main purposes of the WIDA ACCESS program is to identify the English language proficiency levels of students with respect to the WIDA ELD Standards. Because of the emphasis on the classification of student performance into six WIDA proficiency levels, it is important to know how consistently ACCESS scores do indeed classify students into those proficiency levels (American Educational Research Association et al., 2014). The questions that we want to answer are different from the questions that the reliability coefficient answers. Instead of looking at the reliability of a specific student score, we want to know the consistency of the decisions we make when we use students' test scores to classify them into a smaller number of proficiency levels. One way to approach this question is to estimate the degree to which the classification decisions we are making based on the students' **observed test scores** agree with the classification decisions we would make based on students' **theoretical true scores**. This estimate is

known as **decision accuracy**. A second way to approach this question is to estimate the degree to which the classification decisions we are making based on the students' test scores agree with the classification decisions we would make based on students' scores on an alternate form of the test. This estimate is known as **decision consistency**.

## 5.1 Reliability of the Domain Scores

Cronbach's coefficient alpha is widely used as an estimate of reliability, particularly for the internal consistency of test items and/or tasks, and this statistic is appropriate for calculating the reliabilities of students' scores from the administration of the fixed forms of the Writing and Speaking tests.

Conceptually, we can think of it as the correlation obtained between students' performances on two halves of the Writing or Speaking test if every possible way of dividing the test tasks in two were attempted. Thus, Cronbach's coefficient alpha may be low if some tasks are measuring something other than what the majority of the tasks are measuring. In this way, Cronbach's coefficient alpha expresses how well the tasks on a test appear to measure the same ability.

The formula for calculating Cronbach's coefficient alpha for the fixed forms of the Writing and Speaking tests 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  $i$

$\sigma_i^2$  = variance of score on item  $i$

$\sigma_t^2$  = variance of total score

For the Writing test, a slight modification was made in the estimation of Cronbach's alpha for tiered forms that have differential weighting across tasks. This modification is an attempt to take into account the different weighting of tasks when deriving students' ability measures for these tiered forms. For Writing tasks with a weight greater than one, students' responses to the tasks are replicated as a function of their weights. For example, the fourth task in Writing G1A is weighted three; therefore, students' response to this task is repeated three times when computing the Cronbach's alpha. This modification means that the number of pieces of information for Writing tasks that contribute to the estimation of the Cronbach's alpha for G1A is six, not four.

For the Kindergarten Writing domain, a stratified Cronbach's alpha is reported instead of Cronbach's alpha because the dichotomous and polytomous items are heterogeneous, with different true score variance. It is more appropriate to report stratified alpha (Feldt & Brennan, 1989), as this statistic was derived to measure the consistency in students' scores when the total score consists of heterogeneous parts. Stratified alpha is a weighted average of coefficient alphas for item sets with different maximum score points or "strata." Stratified alpha is a reliability estimate computed by dividing the test into parts (strata), computing Cronbach's alpha separately for each part, and using the results to estimate a reliability coefficient for the total score. (See Section 5.5 for more details regarding stratified Cronbach's alpha.) In computing the stratified Cronbach's alpha for Kindergarten Writing, each part that makes up the total score is treated as a stratum. In other words, two strata (dichotomous and polytomous) are entered into the computation. The stratified Cronbach's alpha is interpreted like other traditional internal consistency statistics such as Cronbach's coefficient alpha. Like Cronbach's alpha, stratified Cronbach's



alpha is an estimate of the proportion of the total variance of the observed composite score that can be explained by the variance of the true composite score.

Tables in this section also present the standard error of measurement (SEM), which provides a value for the errors of measurement in students' scores using classical test theory. It is a function of two statistics: the reliability estimate of the test and the (observed) standard deviation (SD) of the test scores in the student population, and it is on the raw score metric. It is calculated as

$$SEM = SD\sqrt{1 - reliability}$$

Since the SEM is an estimate of the standard deviation of the distribution of measurement errors, SEM can be used to create a band around a student's observed score. Under the assumption that the error of measurement follows a normal distribution, the student's true score would lie with a certain degree of probability within this band. Statistically speaking, then, there is an expectation that a student's true score has a 68% probability of falling within the band extending from the observed score minus 1 SEM to the observed score plus 1 SEM. Since SEMs are expressed on the raw score metric, it is wise to keep the range of the raw score points in mind when interpreting the SEM. Raw score statistics by domains are reported below.

In the tables below, we provide the number of tasks, Cronbach's alpha, and SEM for all students and for subgroups as required by the Every Student Succeeds Act Peer Review so that the reliability estimates of the subgroups can be compared with those computed based on all students. For these domains, the first table provides the Cronbach's alpha and SEM for all students. Each row in the table represents a specific grade cluster and test form. For each form for the receptive (Listening and Reading) and expressive (Speaking and Writing) skills, the numbers of students, numbers of tasks, Cronbach's alpha, and SEM are provided. The second table for each domain provides the same information for the population of female students and the population of male students. The third table provides information by ethnicity, for Hispanic and non-Hispanic test-takers, and the fourth table provides information for the population of students who have an individualized education plan (IEP).

**Kindergarten:** For the Kindergarten Listening test, the reliability for all students was 0.95, and reliability values across subgroups ranged from 0.94 to 0.96. For the Kindergarten Reading test, the reliability for all students was 0.95, and reliability values across subgroups ranged from 0.95 to 0.96. For the Kindergarten Writing test, the reliability for all students was 0.92, and reliability values across subgroups ranged from 0.91 to 0.93. For the Kindergarten Speaking test, the reliability for all students was 0.92, and reliability values across subgroups ranged from 0.90 to 0.92.

**Listening Tier A:** The Listening Tier A Cronbach's coefficient alphas computed for all students ranged from 0.65 to 0.79. The Listening Tier A Cronbach's alpha ranged from 0.65 to 0.79 for male students; 0.65 to 0.78 for female students; 0.64 to 0.78 for Hispanic students; 0.65 to 0.79 for non-Hispanic students; and 0.62 to 0.77 for students with an IEP.

**Listening Tier B/C:** The Listening Tier B/C Cronbach's coefficient alphas computed for all students ranged from 0.62 to 0.70. The Listening Tier B/C Cronbach's coefficient alphas ranged from 0.63 to 0.71 for male students; 0.61 to 0.69 for female students; 0.62 to 0.70 for Hispanic students; 0.63 to 0.71 for non-Hispanic students; and 0.59 to 0.74 for students with an IEP.

**Reading Tier A:** The Reading Tier A Cronbach's coefficient alphas computed for all students ranged from 0.79 to 0.83. The Reading Tier A Cronbach's coefficient alphas ranged from 0.78 to 0.82 for male students; 0.79 to 0.83 for female students; 0.78 to 0.82 for Hispanic students; 0.81 to 0.85 for non-Hispanic students; and 0.67 to 0.77 for students with an IEP.

**Reading Tier B/C:** The Reading Tier B/C Cronbach's coefficient alphas computed for all students ranged from 0.74 to 0.83. The Reading Tier B/C Cronbach's coefficient alphas ranged from 0.75 to 0.84 for male students; 0.72 to 0.83 for female students; 0.73 to 0.83 for Hispanic students; 0.76 to 0.84 for non-Hispanic students; and 0.67 to 0.79 for students with an IEP.

**Writing Tier A:** The Writing Tier A Cronbach's coefficient alphas computed for all students ranged from 0.87 to 0.92. The Writing Tier A Cronbach's coefficient alphas ranged from 0.87 to 0.92 for male students; 0.86 to 0.92 for female students; 0.87 to 0.92 for Hispanic students; 0.85 to 0.92 for non-Hispanic students; and 0.86 to 0.91 for students with an IEP.

**Writing Tier B/C:** The Writing Tier B/C Cronbach's coefficient alphas computed for all students ranged from 0.92 to 0.95. The Writing Tier B/C Cronbach's coefficient alphas ranged from 0.92 to 0.96 for male students; 0.91 to 0.95 for female students; 0.92 to 0.95 for Hispanic students; 0.91 to 0.96 for non-Hispanic students; and 0.93 to 0.96 for students with an IEP.

**Speaking Tier A:** The Speaking Tier A Cronbach's coefficient alphas computed for all students ranged from 0.89 to 0.92. Cronbach's coefficient alphas ranged from 0.90 to 0.91 for male students; 0.89 to 0.92 for female students; 0.90 to 0.92 for Hispanic students; 0.87 to 0.90 for non-Hispanic students; and 0.85 to 0.87 for students with an IEP.

**Speaking Tier B/C:** The Speaking Tier B/C Cronbach's coefficient alphas computed for all students ranged from 0.90 to 0.93. Cronbach's coefficient alphas ranged from 0.90 to 0.93 for male students; 0.90 to 0.93 for female students; 0.90 to 0.94 for Hispanic students; 0.90 to 0.93 for non-Hispanic students; and 0.90 to 0.93 for students with an IEP.

## 5.1.1 Listening

**Table 5.1.1.1**

Reliabilities of Domain Scores: List S503 Paper

Cluster	Tier	No. of Students	No. of Items	Cronbach's Alpha	SEM
K	-	251,493	30	0.95	1.85
1	A	23,816	18	0.79	1.61
	B/C	34,946	21	0.70	1.77
2	A	23,816	18	0.79	1.61
	B/C	34,946	21	0.70	1.77
3	A	15,613	18	0.74	1.86
	B/C	59,416	21	0.63	1.91
4-5	A	15,613	18	0.74	1.86
	B/C	59,416	21	0.63	1.91
6-8	A	12,527	18	0.71	1.85
	B/C	32,348	21	0.62	1.83
9-12	A	12,077	18	0.65	1.82
	B/C	29,473	21	0.69	1.97

Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

**Table 5.1.1.2**

Reliabilities of Domain Scores: List S503 Paper by Gender

Cluster	Tier	No. of Items	Female			Male		
			No. of Students	Cronbach's Alpha	SEM	No. of Students	Cronbach's Alpha	SEM
K	-	30	109,565	0.95	1.83	122,830	0.95	1.86
1	A	18	10,940	0.78	1.60	12,754	0.79	1.62
	B/C	21	16,967	0.69	1.75	17,927	0.71	1.78
2	A	18	10,940	0.78	1.60	12,754	0.79	1.62
	B/C	21	16,967	0.69	1.75	17,927	0.71	1.78
3	A	18	6,961	0.74	1.86	8,542	0.75	1.86
	B/C	21	27,523	0.62	1.91	31,765	0.64	1.90
4-5	A	18	6,961	0.74	1.86	8,542	0.75	1.86
	B/C	21	27,523	0.62	1.91	31,765	0.64	1.90
6-8	A	18	5,751	0.71	1.84	6,701	0.71	1.86
	B/C	21	14,875	0.61	1.82	17,420	0.63	1.84
9-12	A	18	5,512	0.65	1.81	6,469	0.65	1.83
	B/C	21	13,772	0.68	1.96	15,635	0.69	1.98

Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

**Table 5.1.1.3**

Reliabilities of Domain Scores: List S503 Paper by Ethnicity

Cluster	Tier	No. of Items	Hispanic			Other		
			No. of Students	Cronbach's Alpha	SEM	No. of Students	Cronbach's Alpha	SEM
K	-	30	170,210	0.95	1.88	72,343	0.94	1.77
1	A	18	19,258	0.78	1.62	4,398	0.79	1.57
	B/C	21	27,427	0.70	1.76	7,320	0.71	1.78
2	A	18	19,258	0.78	1.62	4,398	0.79	1.57
	B/C	21	27,427	0.70	1.76	7,320	0.71	1.78
3	A	18	12,736	0.74	1.87	2,721	0.76	1.82
	B/C	21	47,066	0.63	1.91	11,981	0.64	1.90
4-5	A	18	12,736	0.74	1.87	2,721	0.76	1.82
	B/C	21	47,066	0.63	1.91	11,981	0.64	1.90
6-8	A	18	10,558	0.70	1.86	1,865	0.75	1.80
	B/C	21	25,178	0.62	1.83	6,860	0.63	1.82
9-12	A	18	9,979	0.64	1.82	1,852	0.65	1.79
	B/C	21	22,364	0.69	1.97	6,908	0.69	1.97

Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

**Table 5.1.1.4**

Reliabilities of Domain Scores: List S503 Paper by IEP Status

Cluster	Tier	No. of Students	No. of Items	Cronbach's Alpha	SEM
K	-	20,570	30	0.96	1.90
1	A	2,835	18	0.77	1.67
	B/C	3,894	21	0.74	1.86
2	A	2,835	18	0.77	1.67
	B/C	3,894	21	0.74	1.86
3	A	1,713	18	0.68	1.87
	B/C	11,214	21	0.63	1.97
4-5	A	1,713	18	0.68	1.87
	B/C	11,214	21	0.63	1.97
6-8	A	512	18	0.62	1.89
	B/C	4,843	21	0.59	1.94
9-12	A	397	18	0.65	1.82
	B/C	2,072	21	0.63	2.04

Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

## 5.1.2 Reading

**Table 5.1.2.1**

Reliabilities of Domain Scores: Read S503 Paper

Cluster	Tier	No. of Students	No. of Items	Cronbach's Alpha	SEM
K	-	251,491	30	0.95	1.75
1	A	22,617	24	0.80	2.18
	B/C	30,412	27	0.83	2.32
2	A	22,617	24	0.80	2.18
	B/C	30,412	27	0.83	2.32
3	A	14,376	24	0.83	2.13
	B/C	53,728	27	0.74	2.38
4-5	A	14,376	24	0.83	2.13
	B/C	53,728	27	0.74	2.38
6-8	A	12,135	24	0.79	2.16
	B/C	29,061	27	0.77	2.36
9-12	A	11,987	24	0.80	2.10
	B/C	26,791	27	0.83	2.29

Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

**Table 5.1.2.2**

Reliabilities of Domain Scores: Read S503 Paper by Gender

Cluster	Tier	No. of Items	Female			Male		
			No. of Students	Cronbach's Alpha	SEM	No. of Students	Cronbach's Alpha	SEM
K	-	30	109,565	0.95	1.75	122,827	0.95	1.74
1	A	24	10,315	0.79	2.18	12,180	0.81	2.18
	B/C	27	14,781	0.83	2.32	15,582	0.84	2.31
2	A	24	10,315	0.79	2.18	12,180	0.81	2.18
	B/C	27	14,781	0.83	2.32	15,582	0.84	2.31
3	A	24	6,445	0.83	2.11	7,829	0.82	2.14
	B/C	27	24,999	0.72	2.38	28,619	0.75	2.37
4-5	A	24	6,445	0.83	2.11	7,829	0.82	2.14
	B/C	27	24,999	0.72	2.38	28,619	0.75	2.37
6-8	A	24	5,591	0.79	2.15	6,469	0.78	2.17
	B/C	27	13,481	0.76	2.35	15,530	0.78	2.36
9-12	A	24	5,469	0.80	2.09	6,410	0.81	2.11
	B/C	27	12,591	0.82	2.29	14,143	0.84	2.30

Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

**Table 5.1.2.3**

Reliabilities of Domain Scores: Read S503 Paper by Ethnicity

Cluster	Tier	No. of Items	Hispanic			Other		
			No. of Students	Cronbach's Alpha	SEM	No. of Students	Cronbach's Alpha	SEM
K	-	30	170,208	0.95	1.74	72,343	0.96	1.73
1	A	24	18,353	0.79	2.19	4,108	0.84	2.12
	B/C	27	23,836	0.83	2.32	6,399	0.84	2.30
2	A	24	18,353	0.79	2.19	4,108	0.84	2.12
	B/C	27	23,836	0.83	2.32	6,399	0.84	2.30
3	A	24	11,756	0.82	2.13	2,485	0.85	2.11
	B/C	27	42,641	0.73	2.38	10,765	0.76	2.37
4-5	A	24	11,756	0.82	2.13	2,485	0.85	2.11
	B/C	27	42,641	0.73	2.38	10,765	0.76	2.37
6-8	A	24	10,243	0.78	2.16	1,782	0.81	2.14
	B/C	27	22,572	0.77	2.36	6,202	0.79	2.34
9-12	A	24	9,912	0.80	2.11	1,825	0.81	2.07
	B/C	27	20,371	0.83	2.29	6,238	0.83	2.28

Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

**Table 5.1.2.4**

Reliabilities of Domain Scores: Read S503 Paper by IEP Status

Cluster	Tier	No. of Students	No. of Items	Cronbach's Alpha	SEM
K	-	20,570	30	0.96	1.73
1	A	2,747	24	0.76	2.22
	B/C	3,484	27	0.79	2.38
2	A	2,747	24	0.76	2.22
	B/C	3,484	27	0.79	2.38
3	A	1,549	24	0.72	2.21
	B/C	10,255	27	0.67	2.38
4-5	A	1,549	24	0.72	2.21
	B/C	10,255	27	0.67	2.38
6-8	A	507	24	0.67	2.21
	B/C	4,298	27	0.71	2.39
9-12	A	396	24	0.77	2.15
	B/C	1,829	27	0.79	2.35

Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

## 5.1.3 Writing

**Table 5.1.3.1**

Reliabilities of Domain Scores: Writ S503 Paper

Cluster	Tier	No. of Students	No. of Tasks	Total Possible Raw Score Points	Cronbach's Alpha*	SEM
K	-	251,486	6	0-17	0.92	1.15
1	A	18,607	4	0-40	0.89	1.99
	B/C	15,223	3	0-54	0.95	2.00
2	A	16,451	3	0-27	0.92	1.36
	B/C	47,784	3	0-54	0.95	1.72
3	A	16,451	3	0-27	0.92	1.36
	B/C	47,784	3	0-54	0.95	1.72
4-5	A	9,843	3	0-27	0.90	1.39
	B/C	37,668	3	0-54	0.92	2.01
6-8	A	13,461	3	0-27	0.89	1.40
	B/C	33,833	3	0-54	0.92	1.72
9-12	A	13,196	3	0-27	0.87	1.68
	B/C	30,972	3	0-54	0.92	2.26

\*Note that for Kindergarten, which includes both dichotomous and polytomous tasks in the Writing test, a stratified Cronbach's alpha is computed.

Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

**Table 5.1.3.2**

Reliabilities of Domain Scores: Writ S503 Paper by Gender

Cluster	Tier	No. of Tasks	Total Possible Raw Score Points	Female			Male		
				No. of Students	Cronbach's Alpha*	SEM	No. of Students	Cronbach's Alpha*	SEM
K	-	6	0-17	109,562	0.92	1.15	122,826	0.93	1.15
1	A	4	0-40	8,493	0.88	1.98	10,025	0.89	1.99
	B/C	3	0-54	7,290	0.95	2.01	7,913	0.96	1.99
2	A	3	0-27	7,249	0.92	1.36	9,106	0.92	1.35
	B/C	3	0-54	22,475	0.94	1.69	25,218	0.95	1.74
3	A	3	0-27	7,249	0.92	1.36	9,106	0.92	1.35
	B/C	3	0-54	22,475	0.94	1.69	25,218	0.95	1.74
4-5	A	3	0-27	4,379	0.90	1.40	5,385	0.90	1.38
	B/C	3	0-54	17,240	0.91	1.98	20,350	0.92	2.02
6-8	A	3	0-27	6,191	0.88	1.40	7,185	0.89	1.40
	B/C	3	0-54	15,516	0.91	1.67	18,262	0.92	1.75
9-12	A	3	0-27	5,968	0.86	1.69	7,105	0.87	1.68
	B/C	3	0-54	14,408	0.91	2.17	16,496	0.92	2.32

\*Note that for Kindergarten, which includes both dichotomous and polytomous tasks in the Writing test, a stratified Cronbach's alpha is computed.

Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

**Table 5.1.3.3**

Reliabilities of Domain Scores: Writ S503 Paper by Ethnicity

Cluster	Tier	No. of Tasks	Total Possible Raw Score Points	Hispanic			Other		
				No. of Students	Cronbach's Alpha*	SEM	No. of Students	Cronbach's Alpha*	SEM
K	-	6	0-17	170,206	0.91	1.14	72,340	0.93	1.17
1	A	4	0-40	15,081	0.89	1.99	3,414	0.88	1.99
	B/C	3	0-54	11,682	0.95	1.99	3,443	0.96	2.03
2	A	3	0-27	13,216	0.92	1.36	3,085	0.92	1.36
	B/C	3	0-54	37,866	0.95	1.71	9,643	0.95	1.76
3	A	3	0-27	13,216	0.92	1.36	3,085	0.92	1.36
	B/C	3	0-54	37,866	0.95	1.71	9,643	0.95	1.76
4-5	A	3	0-27	8,068	0.90	1.40	1,665	0.91	1.36
	B/C	3	0-54	29,661	0.92	2.00	7,770	0.91	2.04
6-8	A	3	0-27	11,320	0.88	1.40	2,019	0.89	1.42
	B/C	3	0-54	26,246	0.92	1.69	7,263	0.92	1.79
9-12	A	3	0-27	10,844	0.87	1.65	2,061	0.85	1.82
	B/C	3	0-54	23,439	0.92	2.22	7,320	0.91	2.36

\*Note that for Kindergarten, which includes both dichotomous and polytomous tasks in the Writing test, a stratified Cronbach's alpha is computed.

Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

**Table 5.1.3.4**

Reliabilities of Domain Scores: Writ S503 Paper by IEP Status

Cluster	Tier	No. of Students	No. of Tasks	Total Possible Raw Score Points	Cronbach's Alpha*	SEM
K	-	20,567	6	0-17	0.92	1.10
1	A	2,064	4	0-40	0.89	1.90
	B/C	1,549	3	0-54	0.96	1.93
2	A	2,296	3	0-27	0.91	1.33
	B/C	6,766	3	0-54	0.96	1.79
3	A	2,296	3	0-27	0.91	1.33
	B/C	6,766	3	0-54	0.96	1.79
4-5	A	900	3	0-27	0.88	1.42
	B/C	7,870	3	0-54	0.93	2.06
6-8	A	554	3	0-27	0.89	1.38
	B/C	5,068	3	0-54	0.93	1.85
9-12	A	433	3	0-27	0.86	1.71
	B/C	2,169	3	0-54	0.93	2.18

\*Note that for Kindergarten, which includes both dichotomous and polytomous tasks in the Writing test, a stratified Cronbach's alpha is computed.

Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.



## 5.1.4 Speaking

**Table 5.1.4.1**

Reliabilities of Domain Scores: Spek S503 Paper

Cluster	Tier	No. of Students	No. of Tasks	Total Possible Raw Score Points	Cronbach's Alpha	SEM
K	-	251,483	10	0-10	0.92	1.02
1	A	18,500	6	0-18	0.90	1.35
	B/C	15,135	6	0-24	0.91	1.41
2	A	16,341	6	0-18	0.92	1.35
	B/C	47,486	6	0-24	0.91	1.34
3	A	16,341	6	0-18	0.92	1.35
	B/C	47,486	6	0-24	0.91	1.34
4-5	A	9,753	6	0-18	0.91	1.40
	B/C	37,394	6	0-24	0.90	1.36
6-8	A	13,360	6	0-18	0.91	1.42
	B/C	33,607	6	0-24	0.92	1.36
9-12	A	13,035	6	0-18	0.89	1.44
	B/C	30,671	6	0-24	0.93	1.35

Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

**Table 5.1.4.2**

Reliabilities of Domain Scores: Spek S503 Paper by Gender

Cluster	Tier	No. of Tasks	Total Possible Raw Score Points	Female			Male		
				No. of Students	Cronbach's Alpha	SEM	No. of Students	Cronbach's Alpha	SEM
K	-	10	0-10	109,564	0.92	1.01	122,821	0.91	1.04
1	A	6	0-18	8,439	0.90	1.37	9,972	0.90	1.34
	B/C	6	0-24	7,250	0.91	1.41	7,865	0.91	1.41
2	A	6	0-18	7,205	0.92	1.36	9,041	0.91	1.34
	B/C	6	0-24	22,333	0.91	1.34	25,062	0.91	1.33
3	A	6	0-18	7,205	0.92	1.36	9,041	0.91	1.34
	B/C	6	0-24	22,333	0.91	1.34	25,062	0.91	1.33
4-5	A	6	0-18	4,336	0.91	1.42	5,338	0.91	1.38
	B/C	6	0-24	17,113	0.90	1.36	20,203	0.90	1.35
6-8	A	6	0-18	6,139	0.90	1.43	7,138	0.91	1.41
	B/C	6	0-24	15,393	0.92	1.38	18,159	0.92	1.34
9-12	A	6	0-18	5,890	0.89	1.47	7,024	0.90	1.42
	B/C	6	0-24	14,281	0.93	1.36	16,322	0.93	1.33

Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

**Table 5.1.4.3**

Reliabilities of Domain Scores: Spek S503 Paper by Ethnicity

Cluster	Tier	No. of Tasks	Total Possible Raw Score Points	Hispanic			Other		
				No. of Students	Cronbach's Alpha	SEM	No. of Students	Cronbach's Alpha	SEM
K	-	10	0-10	170,203	0.92	1.03	72,340	0.90	1.01
1	A	6	0-18	15,001	0.90	1.35	3,388	0.87	1.38
	B/C	6	0-24	11,622	0.91	1.40	3,416	0.91	1.43
2	A	6	0-18	13,131	0.92	1.34	3,061	0.89	1.38
	B/C	6	0-24	37,640	0.91	1.34	9,573	0.91	1.35
3	A	6	0-18	13,131	0.92	1.34	3,061	0.89	1.38
	B/C	6	0-24	37,640	0.91	1.34	9,573	0.91	1.35
4-5	A	6	0-18	7,999	0.92	1.38	1,645	0.89	1.43
	B/C	6	0-24	29,453	0.90	1.35	7,706	0.90	1.37
6-8	A	6	0-18	11,232	0.91	1.40	2,006	0.90	1.48
	B/C	6	0-24	26,087	0.92	1.36	7,197	0.92	1.36
9-12	A	6	0-18	10,714	0.90	1.43	2,037	0.88	1.48
	B/C	6	0-24	23,226	0.94	1.34	7,234	0.93	1.35

Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

**Table 5.1.4.4**

Reliabilities of Domain Scores: Spek S503 Paper by IEP Status

Cluster	Tier	No. of Students	No. of Tasks	Total Possible Raw Score Points	Cronbach's Alpha	SEM
K	-	20,564	10	0-10	0.90	1.04
1	A	2,057	6	0-18	0.87	1.28
	B/C	1,535	6	0-24	0.93	1.34
2	A	2,284	6	0-18	0.86	1.32
	B/C	6,725	6	0-24	0.91	1.34
3	A	2,284	6	0-18	0.86	1.32
	B/C	6,725	6	0-24	0.91	1.34
4-5	A	892	6	0-18	0.86	1.42
	B/C	7,824	6	0-24	0.90	1.36
6-8	A	550	6	0-18	0.87	1.43
	B/C	5,033	6	0-24	0.92	1.37
9-12	A	428	6	0-18	0.85	1.48
	B/C	2,151	6	0-24	0.93	1.43

Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

## 5.2. Interrater Agreement Rates

DRC raters score students' responses to the tasks included on the ACCESS Writing tests (except Kindergarten, which is scored by the Test Administrator). We describe the scoring of students' responses to these performance tasks in Section 3.2.2. DRC selects a sample of 20% of all responses scored, chosen at random during the operational scoring process, for double scoring. The tables below provide information on interrater agreement for a sample of 20% of task raters. These tables show, for each of the tasks, the percentage of agreement between two raters. The first column shows the task, and the second

column shows the number of responses that were double scored. DRC selects a sample of 20% of all responses scored, chosen at random during the operational scoring process. The next columns show the rates of agreement.

For Writing, the scoring rubric that the raters used defines six levels of performance ranging from 0 to 6, with the possibility of awarding a “plus” score between levels (e.g., 3, 3+, or 4 are all valid scores). We considered scores that matched or were contiguous as signifying **agreement** (%AG)—for example, if Rater 1 assigned a score of 3+ while Rater 2 assigned a score of 3, 3+, or 4. We considered scores that were one whole score point apart as **adjacent scores** (%AD)—for example, if Rater 1 assigned a score of 3+ while Rater 2 assigned a score of 2+ or 4+. Finally, if two raters assigned scores that were more than one whole score point apart, we considered those scores to be **nonadjacent scores** (%NA).

As the Speaking test is scored locally, it is not possible to provide interrater agreement data for Speaking. Section 3.2.3 in Part 1 of this report describes training procedures that local raters must complete before being certified to administer and score the Speaking test.

WIDA stipulates a minimum interrater agreement rate of 70%. DRC defines this “**agreement**” as being scored as adjacent agreement (AG) for Writing. See Section 3.2.2 for more detail about how WIDA and DRC used the agreement rates to ensure that DRC maintains sufficient quality control throughout the course of scoring.

For Writing, the lowest value for interrater agreement was 94%.

5.2.1 Listening

Interrater Agreement is not relevant for the domain of Listening, as all items are multiple choice items.

5.2.2 Reading

Interrater Agreement is not relevant for the domain of Reading, as all items are multiple choice items.

5.2.3 Writing

5.2.3.0 Kindergarten

Table 5.2.3.0  
Interrater Agreement: Writ K S503 Paper

Interrater Agreement	n/a
----------------------	-----

### 5.2.3.1 Grade 1

**Table 5.2.3.1.1**

Interrater Agreement: Writ 1 A S503 Paper

<b>Interrater Agreement</b>	<b>Task</b>	<b>No. in Sample</b>	<b>%AG</b>	<b>%AD</b>	<b>%NA</b>
	1	8,198	100	0	0
	2	11,754	100	0	0
	3	12,584	99	1	0
	4	11,396	98	2	0

**Table 5.2.3.1.2**

Interrater Agreement: Writ 1 B/C S503 Paper

<b>Interrater Agreement</b>	<b>Task</b>	<b>No. in Sample</b>	<b>%AG</b>	<b>%AD</b>	<b>%NA</b>
	1	8,336	97	3	0
	2	10,484	98	2	0
	3	9,184	99	1	0

### 5.2.3.2 Grade 2

**Table 5.2.3.2.1**

Interrater Agreement: Writ 2 A S503 Paper

<b>Interrater Agreement</b>	<b>Task</b>	<b>No. in Sample</b>	<b>%AG</b>	<b>%AD</b>	<b>%NA</b>
	1	6,778	98	2	0
	2	6,852	99	1	0
	3	7,632	99	1	0

Note: the test form is shared between 2A and 3A.

**Table 5.2.3.2.2**

Interrater Agreement: Writ 2 B/C S503 Paper

<b>Interrater Agreement</b>	<b>Task</b>	<b>No. in Sample</b>	<b>%AG</b>	<b>%AD</b>	<b>%NA</b>
	1	11,258	97	3	0
	2	11,752	98	2	0
	3	11,542	97	3	0

Note: the test form is shared between 2B/C and 3B/C.

*5.2.3.3 Grade 3***Table 5.2.3.3.1**

Interrater Agreement: Writ 3 A S503 Paper

<b>Interrater Agreement</b>	<b>Task</b>	<b>No. in Sample</b>	<b>%AG</b>	<b>%AD</b>	<b>%NA</b>
	1	4,622	98	2	0
	2	4,380	97	3	0
	3	5,034	99	1	0

Note: the test form is shared between 2A and 3A.

**Table 5.2.3.3.2**

Interrater Agreement: Writ 3 B/C S503 Paper

<b>Interrater Agreement</b>	<b>Task</b>	<b>No. in Sample</b>	<b>%AG</b>	<b>%AD</b>	<b>%NA</b>
	1	10,892	97	3	0
	2	11,254	97	3	0
	3	11,084	97	3	0

Note: the test form is shared between 2B/C and 3B/C.

*5.2.3.4 Grades 4–5***Table 5.2.3.4.1**

Interrater Agreement: Writ 4-5 A S503 Paper

<b>Interrater Agreement</b>	<b>Task</b>	<b>No. in Sample</b>	<b>%AG</b>	<b>%AD</b>	<b>%NA</b>
	1	7,574	98	0	2
	2	6,316	98	0	2
	3	7,438	99	0	1

**Table 5.2.3.4.2**

Interrater Agreement: Writ 4-5 B/C S503 Paper

<b>Interrater Agreement</b>	<b>Task</b>	<b>No. in Sample</b>	<b>%AG</b>	<b>%AD</b>	<b>%NA</b>
	1	16,506	96	4	0
	2	16,546	97	3	0
	3	16,702	96	4	0

*5.2.3.5 Grades 6–8***Table 5.2.3.5.1**

Interrater Agreement: Writ 6-8 A S503 Paper

<b>Interrater Agreement</b>	<b>Task</b>	<b>No. in Sample</b>	<b>%AG</b>	<b>%AD</b>	<b>%NA</b>
	1	7,618	98	2	0
	2	8,440	98	2	0
	3	6,744	97	2	0

**Table 5.2.3.5.2**

Interrater Agreement: Writ 6-8 B/C S503 Paper

<b>Interrater Agreement</b>	<b>Task</b>	<b>No. in Sample</b>	<b>%AG</b>	<b>%AD</b>	<b>%NA</b>
	1	14,532	98	2	0
	2	14,578	98	2	0
	3	14,792	98	2	0

*5.2.3.6 Grades 9–12***Table 5.2.3.6.1**

Interrater Agreement: Writ 9-12 A S503 Paper

<b>Interrater Agreement</b>	<b>Task</b>	<b>No. in Sample</b>	<b>%AG</b>	<b>%AD</b>	<b>%NA</b>
	1	8,542	97	3	0
	2	6,386	97	3	0
	3	10,428	98	2	0

**Table 5.2.3.6.2**

Interrater Agreement: Writ 9-12 B/C S503 Paper

<b>Interrater Agreement</b>	<b>Task</b>	<b>No. in Sample</b>	<b>%AG</b>	<b>%AD</b>	<b>%NA</b>
	1	13,872	96	4	0
	2	14,194	97	3	0
	3	14,604	94	6	0

### ***5.3 Conditional Standard Errors of Measurement at Cut Score***

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 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 ACCESS score scale (See Section 2.2).

When calculated using an IRT approach, CSEM values can vary across the scale scores. For example, in the Listening and Reading domains, if a student answers correctly either a very few or a very large number of items (i.e., scores at the extremes of the scale score distribution), the CSEM value will be larger than it would be if the student correctly answers a moderate number of items. Scale scores near the middle of the score distribution typically have lower CSEM values compared to scale scores near the extremes because many tests are comprised of a large proportion of moderately difficult items, which are well suited to measuring students of moderate proficiency.

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).

As a rule, lower CSEM values around scale scores at important decision points are desirable. Generally speaking, the most important decision points for the ACCESS scores are at the PL 3/4 and PL 4/5 cut points, although the approaches that WIDA states use to make decisions about ACCESS scores differ. As discussed in Section 5, all WIDA states use composite scale scores when making reclassification decisions, and no WIDA state uses a single domain scale score when making those decisions. Because each grade has its own set of cut points, we provide information for each grade within a grade-level cluster.

Since we scale ACCESS test scores using an IRT approach, CSEM values for the scale scores at the highest cut points are typically large. Use of this approach tends to produce larger CSEM values at the lower and the higher ends of the score scale. In addition, because students exit the EL program when they demonstrate that they are English language proficient, there are typically fewer students at the highest cut points than at those other cut points. Therefore, the CSEM values associated with the scale scores at the

highest cut points tend to be larger than those of the scale scores at the lower cut points since there are fewer students available for estimating the scores and the CSEM values for these scores.

For each domain, we present the values by tier. From these tables, it is possible to identify how well the different tiers are targeted for making decisions about students at the various proficiency level cuts.

In the tables below, the leftmost column shows the proficiency level cut (e.g., 1/2, which is the cut between PL 1 and PL 2). The second column shows the grade level. The third column shows the cut score in the scale score metric (e.g., 305). In the last column(s), the corresponding CSEM value for the scale score at each cut point are shown.



## 5.3.1 Listening

### 5.3.1.0 Kindergarten

**Table 5.3.1.0**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List K S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM
1/2	K	229	17.28
2/3	K	251	18.41
3/4	K	278	20.66
4/5	K	286	21.42
5/6	K	308	24.80

### 5.3.1.1 Grade 1

**Table 5.3.1.1**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 1 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	1	236	19.16	19.54
2/3	1	259	19.54	18.41
3/4	1	291	22.54	18.41
4/5	1	303	24.42	18.79
5/6	1	327	29.31	20.66

### 5.3.1.2 Grade 2

**Table 5.3.1.2**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 2 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	2	245	19.16	19.01
2/3	2	283	21.42	18.03
3/4	2	314	26.30	19.54
4/5	2	330	30.43	21.04
5/6	2	354	38.32	24.80

### 5.3.1.3 Grade 3

**Table 5.3.1.3**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 3 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	3	262	21.42	24.80
2/3	3	300	18.79	20.66
3/4	3	331	19.54	19.16
4/5	3	349	21.04	18.79
5/6	3	374	25.17	19.16

### 5.3.1.4 Grade 4-5

**Table 5.3.1.4**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 4-5 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	4	275	20.06	23.29
	5	285	19.54	22.17
2/3	4	313	18.79	19.91
	5	323	19.16	19.16
3/4	4	343	20.66	18.79
	5	354	21.79	18.79
4/5	4	363	22.92	18.79
	5	375	25.55	19.35
5/6	4	388	28.55	20.29
	5	401	32.31	21.42

5.3.1.5 Grade 6-8

**Table 5.3.1.5**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 6-8 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	6	294	20.29	21.42
	7	302	19.91	20.29
	8	308	19.91	19.91
2/3	6	332	19.91	18.41
	7	340	20.40	18.03
	8	347	21.04	17.83
3/4	6	363	22.54	18.03
	7	370	23.29	18.03
	8	377	24.42	18.41
4/5	6	385	25.92	18.79
	7	394	28.18	19.54
	8	402	30.06	20.39
5/6	6	411	33.06	21.79
	7	420	36.07	22.92
	8	427	39.07	24.42

### 5.3.1.6 Grade 9-12

**Table 5.3.1.6**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 9-12 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	9	314	20.66	21.42
	10	325	20.66	19.91
	11	335	20.66	19.16
	12	342	21.04	18.41
2/3	9	353	21.42	18.03
	10	358	21.80	18.03
	11	364	22.17	17.66
	12	368	22.54	17.66
3/4	9	383	24.80	18.03
	10	389	25.55	18.03
	11	394	26.68	18.41
	12	398	27.43	18.41
4/5	9	409	30.06	19.16
	10	415	31.94	19.91
	11	420	33.44	20.29
	12	426	35.32	21.04
5/6	9	434	38.32	21.99
	10	441	41.33	23.29
	11	447	44.33	24.42
	12	452	46.96	25.17

## 5.3.2 Reading

### 5.3.2.0 Kindergarten

**Table 5.3.2.0**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read K S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM
1/2	K	241	15.34
2/3	K	259	18.46
3/4	K	279	23.92
4/5	K	289	27.82
5/6	K	310	39.26

### 5.3.2.1 Grade 1

**Table 5.3.2.1**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 1 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	1	264	11.44	13.78
2/3	1	286	11.70	11.44
3/4	1	304	13.00	10.40
4/5	1	315	14.56	10.40
5/6	1	334	18.46	11.44

### 5.3.2.2 Grade 2

**Table 5.3.2.2**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 2 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	2	283	11.70	11.52
2/3	2	307	13.52	10.40
3/4	2	326	16.51	10.92
4/5	2	337	19.24	11.78
5/6	2	355	25.48	14.04

### 5.3.2.3 Grade 3

**Table 5.3.2.3**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 3 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	3	297	11.70	17.94
2/3	3	323	11.96	13.00
3/4	3	342	13.52	11.13
4/5	3	352	14.82	10.69
5/6	3	370	18.46	10.66

### 5.3.2.4 Grade 4-5

**Table 5.3.2.4**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 4-5 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	4	307	11.44	15.86
	5	316	11.70	14.04
2/3	4	335	12.74	11.70
	5	345	13.83	10.92
3/4	4	354	15.34	10.66
	5	364	17.16	10.48
4/5	4	364	17.16	10.48
	5	373	19.50	10.61
5/6	4	382	22.10	11.00
	5	391	25.48	11.70

### 5.3.2.5 Grade 6-8

**Table 5.3.2.5**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 6-8 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	6	323	11.60	13.52
	7	329	11.44	12.74
	8	335	11.70	11.96
2/3	6	353	12.48	10.66
	7	360	13.00	10.54
	8	366	13.52	10.40
3/4	6	373	14.56	10.53
	7	380	15.60	10.66
	8	386	16.67	11.04
4/5	6	382	16.12	10.80
	7	389	17.42	11.18
	8	395	18.98	11.70
5/6	6	399	20.02	12.22
	7	406	22.36	13.00
	8	412	24.70	13.78

### 5.3.2.6 Grade 9-12

**Table 5.3.2.6**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 9-12 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	9	340	11.83	14.51
	10	344	11.70	13.78
	11	348	11.70	13.26
	12	352	11.70	12.74
2/3	9	372	12.48	10.92
	10	377	12.74	10.66
	11	382	13.26	10.66
	12	386	13.78	10.40
3/4	9	392	14.82	10.40
	10	397	15.60	10.66
	11	402	16.38	10.66
	12	407	17.68	10.92
4/5	9	401	16.38	10.66
	10	406	17.42	10.92
	11	410	18.46	11.18
	12	414	19.50	11.41
5/6	9	418	20.54	11.70
	10	423	22.36	12.22
	11	427	23.66	12.74
	12	432	25.74	13.52



### 5.3.3 Writing

#### 5.3.3.0 Kindergarten

**Table 5.3.3.0**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ K S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM
1/2	K	234	18.97
2/3	K	271	21.15
3/4	K	311	31.41
4/5	K	367	43.22
5/6	K	389	52.55

#### 5.3.3.1 Grade 1

**Table 5.3.3.1**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 1 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	1	238	10.31	8.59
2/3	1	275	13.96	10.74
3/4	1	337	13.69	12.51
4/5	1	382	13.16	10.82
5/6	1	405	16.38	11.38

#### 5.3.3.2 Grade 2

**Table 5.3.3.2**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 2 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	2	242	11.81	8.46
2/3	2	279	16.27	11.28
3/4	2	341	17.18	12.35
4/5	2	388	15.31	10.87
5/6	2	411	18.26	12.35

### 5.3.3.3 Grade 3

**Table 5.3.3.3**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 3 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	3	247	12.08	8.59
2/3	3	283	16.65	11.44
3/4	3	346	17.18	12.08
4/5	3	394	15.65	11.01
5/6	3	418	20.41	13.69

### 5.3.3.4 Grade 4-5

**Table 5.3.3.4**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 4-5 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	4	266	11.81	9.93
	5	267	11.55	9.67
2/3	4	288	13.43	8.86
	5	293	14.04	8.86
3/4	4	351	17.99	12.35
	5	356	17.72	12.62
4/5	4	401	15.57	11.55
	5	407	15.31	11.33
5/6	4	425	15.57	11.01
	5	433	16.65	11.01

5.3.3.5 Grade 6-8

**Table 5.3.3.5**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 6-8 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	6	268	12.35	8.32
	7	273	12.62	8.32
	8	281	13.69	8.59
2/3	6	298	15.84	10.20
	7	305	16.65	11.01
	8	311	16.92	11.55
3/4	6	361	17.45	12.62
	7	367	17.18	12.35
	8	372	16.92	12.35
4/5	6	413	15.57	10.74
	7	419	16.11	10.74
	8	424	16.92	11.01
5/6	6	441	20.68	12.35
	7	450	23.90	14.23
	8	459	27.93	16.38

### 5.3.3.6 Grade 9-12

**Table 5.3.3.6**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 9-12 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	9	289	12.35	8.22
	10	298	12.08	8.59
	11	308	12.89	9.67
	12	318	14.23	10.77
2/3	9	319	14.23	11.01
	10	326	15.31	11.55
	11	335	16.38	12.08
	12	344	17.02	12.35
3/4	9	378	17.72	12.62
	10	385	17.72	12.35
	11	391	17.45	12.08
	12	398	17.18	11.81
4/5	9	430	15.47	10.74
	10	436	15.31	10.74
	11	441	15.57	11.01
	12	447	15.84	11.55
5/6	9	469	19.33	15.04
	10	479	22.29	17.72
	11	490	27.12	22.02
	12	501	33.03	27.12

## 5.3.4 Speaking

### 5.3.4.0 Kindergarten

**Table 5.3.4.0**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek K S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM
1/2	K	191	28.06
2/3	K	250	20.92
3/4	K	301	16.33
4/5	K	349	22.45
5/6	K	392	53.57

### 5.3.4.1 Grade 1

**Table 5.3.4.1**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 1 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	1	205	19.30	14.74
2/3	1	261	27.20	19.30
3/4	1	311	25.15	18.13
4/5	1	361	26.32	18.13
5/6	1	403	39.48	26.91

### 5.3.4.2 Grade 2

**Table 5.3.4.2**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 2 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	2	220	23.98	16.96
2/3	2	273	26.91	18.72
3/4	2	322	24.57	17.84
4/5	2	374	35.97	21.94
5/6	2	415	64.05	33.34

### 5.3.4.3 Grade 3

**Table 5.3.4.3**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 3 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	3	234	26.32	17.84
2/3	3	283	25.74	18.43
3/4	3	332	25.15	18.13
4/5	3	386	41.82	24.28
5/6	3	425	74.87	38.02

### 5.3.4.4 Grade 4-5

**Table 5.3.4.4**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 4-5 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	4	246	22.23	16.09
	5	258	23.98	16.38
2/3	4	293	26.29	18.25
	5	302	26.32	18.72
3/4	4	342	25.74	18.78
	5	350	26.03	18.72
4/5	4	397	35.10	20.18
	5	407	39.19	21.35
5/6	4	435	57.33	27.79
	5	443	64.64	30.71

### 5.3.4.5 Grade 6-8

**Table 5.3.4.5**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 6-8 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	6	268	21.18	15.50
	7	277	22.81	15.79
	8	284	24.28	16.38
2/3	6	310	27.79	18.43
	7	317	27.79	19.01
	8	323	27.79	19.30
3/4	6	360	24.57	18.43
	7	369	23.98	17.84
	8	377	23.98	17.55
4/5	6	417	29.54	18.72
	7	425	32.17	19.89
	8	433	35.10	21.06
5/6	6	451	44.46	24.86
	7	457	48.55	26.62
	8	463	53.23	28.96

### 5.3.4.6 Grade 9-12

**Table 5.3.4.6**

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 9-12 S503 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	9	290	25.45	17.26
	10	295	26.03	17.55
	11	299	26.62	18.13
	12	302	27.20	18.43
2/3	9	328	27.49	19.60
	10	333	27.20	19.60
	11	337	26.76	19.60
	12	340	26.62	19.30
3/4	9	385	24.57	17.26
	10	393	24.86	17.26
	11	400	25.74	17.67
	12	406	26.62	18.13
4/5	9	440	36.85	23.40
	10	446	40.07	24.86
	11	451	42.70	26.32
	12	455	45.04	27.79
5/6	9	468	54.40	32.76
	10	471	57.03	34.22
	11	474	59.67	35.68
	12	476	61.42	36.85

## 5.4 Accuracy and Consistency

One of the main purposes of the WIDA ACCESS program is to identify the English language proficiency level of students with respect to the WIDA ELD Standards. Because of the emphasis on the classification of student performance, a question of interest is how accurately and consistently ACCESS domain scale scores can classify students into the WIDA proficiency levels determined by the 2016 ACCESS standard-setting process (Cook & MacGregor, 2017). Test users can examine indices that report on the accuracy and consistency of these classifications and can use that information to judge the utility of WIDA's proficiency level categorization, while policy makers can use these indices to assist them when making decisions about ACCESS test design and score reporting (American Educational Research Association et al., 2014). The analyses we conduct to examine the accuracy and consistency of classifications utilize the methods that Livingston and Lewis (1995) and Young and Yoon (1998) outlined, as implemented in the software program BB-CLASS (Brennan, 2004; cf. also Lee, Hanson, & Brennan, 2002).

**Classification accuracy** is defined conceptually as the extent to which the proficiency classifications of students based on their observed raw scores or scale scores would agree with those made based on their true scores (Livingston, 2018; Livingston & Lewis, 1995). A student's true score is the average of the scores that the student would have received, averaging over some



set of prespecified factors or conditions (e., different versions of the test, different times of test administration). Therefore, the calculation of the true scores depends upon the particular factors over which one chooses to average (Livingston, 2018). We assume that true scores measure perfectly, but those scores are unknown. Therefore, to provide the best estimation of classification accuracy for WIDA, we use test data from one ACCESS administration to estimate students' true scale scores based on their domain scale scores and the parameters of the model used in estimating those true scale scores. We can then use the results from our analysis to estimate the percentages of the students who were accurately classified into each proficiency level.

**Classification consistency** is defined conceptually as the extent to which the proficiency classifications of students agree, given two independent administrations of the same or two parallel test forms. It is impractical to obtain repeated administrations of the same or parallel test forms because of cost, testing burden, and the effects of student memory and practice. However, it is possible to estimate the percentages of the students who would be consistently classified with the assumption that the same test is independently administered twice to the same group of students.

The approach that Livingston and Lewis (1995) took, which we implemented here, uses information about the reliability of the students' domain scale scores, the cut points, and the observed distribution of scores. Then, using a four-parameter beta distribution, we model the distribution of the true scale scores and of the domain scale scores on a parallel form. The Livingston and Lewis procedure requires that the reliability estimate of the students' scores on a test form be provided when calculating the classification consistency and accuracy indices. For Listening and Reading, we used the Rasch student separation reliability estimates by grade-level clusters in the procedure. Since the Writing and Speaking tests were tiered, we needed to produce a single reliability estimate across tiers to implement the Livingston and Lewis procedure. This is a weighted reliability estimate across tiers (see Section 5.1).

## 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). For example, an overall classification accuracy index of 0.774 means that we would classify 77% of the students into the same proficiency level according to their domain scale scores and their true scale scores. **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. For example, an overall classification consistency index of 0.664 means that we would classify 66% of the students into the same proficiency level if they took two parallel forms of the test. A classification consistency index is always lower than its corresponding classification accuracy index, because in classification consistency, a classification based on a student's performance on the administered test and a classification based on that student's performance on a parallel test are both subject to measurement error. In contrast, in classification accuracy, only the classification based on a student's performance on the administered test contains error while we assume that the classification based on that student's true scale score is free of measurement error.

## Marginal Classification Accuracy and Consistency

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 PL 4/PL 5 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.

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. For example, a classification accuracy index of 0.774 at the PL 4/PL 5 cut point means that we would classify 77% of the students in the same way using their domain scale scores or their true scale scores, either into the proficiency levels below the cut point (i.e., PL 1 to PL 4) or into the proficiency levels above the cut point (i.e., PL 5 to PL 6). 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. For example, a classification consistency index of 0.664 at the PL 4/PL 5 cut point means that we would classify 66% of the students in the same way if they took two parallel forms, either into the proficiency levels below the cut point (i.e., PL 1 to PL 4) or into the proficiency levels above the cut point (i.e., PL 5 to PL 6). 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, 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.

The interactions of a number of factors affect the calculation of classification accuracy and consistency: (1) the number of proficiency level cut points, (2) the magnitude of the test score reliability coefficient, (3) measurement accuracy for scale scores at the cut points, (4) the distances between adjacent cut points, (5) the locations of the cut points on the ability scale, and (6) the proportion of students' scale scores around a cut point (Ercikan & Julian, 2002; Lee et al., 2002). These factors are functions of the test design and, most importantly, the standard-setting decisions. The indices are lower when there is a greater number of proficiency levels, a lower test score reliability coefficient, and higher measurement accuracy of the scale scores at the cut points, as well as when the two adjacent cut points are closer, and when more students' domain scale scores are around a cut point. Furthermore, the numbers and types of items on a test affect the calculation of the test score reliability coefficient. The lower the test score reliability, the lower the classification accuracy and consistency indices would be. For example, the test score reliability coefficient for the ACCESS Online Writing domain raw scores would be lower than the test score reliability coefficients for similar tests that include more items or tasks since we estimate the test score reliability coefficient for ACCESS Online Writing domain raw scores based on students' performance on only two tasks. Therefore, the classification accuracy and consistency indices for the Writing domain might be lower than those for other domains.

For each test domain, we present three tables. The first reports indices that describe the overall accuracy and overall consistency of the proficiency level classifications for each grade level. The second reports

the marginal classification accuracy indices based on domain scale scores at the cut points for each grade level. The third reports the marginal classification consistency indices based on domain scale scores at the cut points for each grade level. If we could not estimate the overall and marginal classification accuracy and consistency indices because we classified fewer than 200 students into a given proficiency level, we combined the affected proficiency level and the proficiency level below it and placed 'N/A' in the table for the affected proficiency level.

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 ACCESS 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.

For Listening, as shown in Table 5.4.1.1, the overall classification accuracy indices ranged from 0.360 to 0.712 and the overall classification consistency indices ranged from 0.312 to 0.652. Grade 8 had the lowest overall classification accuracy and consistency indices for Listening.

For Reading, as shown in Table 5.4.2.1, the overall classification accuracy indices ranged from 0.423 to 0.854 and the overall classification consistency indices ranged from 0.329 to 0.834. Grade 3 had the lowest overall classification accuracy and consistency indices for Reading.

For Writing, as shown in Table 5.4.3.1, the overall classification accuracy indices ranged from 0.701 to 0.882, and the overall classification consistency indices ranged from 0.637 to 0.843. Grade 4 had the lowest overall classification accuracy and consistency indices for Writing.

For Speaking, as shown in Table 5.4.4.1, the overall classification accuracy indices ranged from 0.512 to 0.694 and the overall classification consistency indices ranged from 0.516 to 0.594. The lowest overall classification accuracy and consistency indices for Speaking was Kindergarten.

From an accountability perspective, the most important indices for test users and policy makers to examine are the marginal classification accuracy and consistency indices. To help them interpret our results, we report for each domain the range of the marginal classification accuracy and consistency indices 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.

For Listening, the marginal classification accuracy indices based on scale scores at the cut points ranged from 0.753 to 0.992 (Table 5.4.1.2) and the marginal classification consistency indices ranged from 0.711 to 0.986 (Table 5.4.1.3). Grade 8, at the PL 5/PL 6 and PL 4/PL 5 cut point, had the lowest classification accuracy and consistency indices for Listening. Note that Grade 8 was also identified as having the lowest overall classification accuracy and consistency indices in the Listening domain. The low marginal classification accuracy and consistency at the PL 5/PL 6 and PL 4/PL 5 cut score appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal classification accuracy and consistency for Grade 8 Listening were still in the .70's and .90's.

For Reading, the marginal classification accuracy indices based on scale scores at the cut points ranged from 0.779 to 0.973 (Table 5.4.2.2) and the marginal classification consistency indices scores ranged from

0.707 to 0.958 (Table 5.4.2.3). Grade 3, at PL 3/PL 4 cut point, had the lowest classification accuracy and consistency indices. Note that Grade 3 was also identified as having the lowest overall classification accuracy and consistency indices in the Reading domain. The low marginal classification consistency at the PL 3/PL 4 cut appeared to have contributed to its low overall classification consistency. However, it should be noted that the marginal classification accuracy and consistency for Grade 3 Reading were still in the .70's and .90's.

For Writing, the marginal classification accuracy indices based on scale scores at the cut points ranged from 0.813 to 0.991 (Table 5.4.3.2) and the marginal classification consistency indices ranged from 0.738 to 0.991 (Table 5.4.3.3). Grade 4, at the PL 3/PL 4 cut point, had the lowest classification accuracy and consistency indices for Writing. Note that Grade 4 was also identified as having the lowest overall classification accuracy and consistency indices in the Writing domain. The low marginal classification accuracy and consistency at the PL 3/PL 4 cut appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal classification accuracy and consistency for Grade 4 Writing were still in the .70's and .90's. For Speaking, the marginal classification accuracy indices based on scale scores at the cut points ranged from 0.754 to 0.991 (Table 5.4.4.2) and the marginal classification consistency indices ranged from 0.816 to 0.991 (Table 5.4.4.3). Kindergarten, at the PL 5/PL 6 cut point, had the lowest classification accuracy and consistency value for Speaking. Note that Kindergarten was also identified as having the lowest overall classification accuracy and consistency indices in the Speaking domain. The low marginal classification accuracy and consistency at the PL 5/PL 6 cut appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal classification accuracy and consistency for Kindergarten Speaking were still in the .70's and .90's.

When we compared the overall and marginal classification accuracy and consistency indices based on the domain scale scores for a particular grade, we saw that in many instances they told the same story (i.e., for a given grade, when the overall classification accuracy and consistency indices were low, then the marginal classification accuracy and consistency indices also tended to be low).

We observed that in the domains of Reading and Writing, the marginal classification accuracy and consistency indices for PL cut points in the middle of the proficiency level range (e.g., the PL 2/3 and PL 3/4 cut points) tended, on average, to be lower than the marginal classification accuracy and consistency indices for cut points at the lower and upper ends of the range, a finding that is consistent with findings from previous researchers (Ercikan & Julian, 2002; Lee et al., 2002). One possible reason might be that the cut points for the proficiency levels in the middle of the proficiency level range tend to be closer together than the cut points for the proficiency levels at the ends of that range. (Cut points tend to be closer to each other when there are a large number of proficiency levels.) We would expect marginal classification accuracy and consistency to vary for different ability levels due to variation in measurement accuracy. That is, the further away the students' domain scale scores are from the cut points, the smaller the classification errors would be, or the more accurate the classification decisions would be. With many proficiency levels, there are more student domain scale scores near the cut points than there would be if there were fewer proficiency levels. Therefore, the higher the number of proficiency levels, the higher the probability that we would misclassify students (Ercikan & Julian, 2002). Additionally, the intervals between cut points that are in the middle of the ACCESS proficiency level range are smaller than the intervals between cut points that are at the upper and lower ends of the proficiency level range. Consequently, the marginal classification accuracy and consistency indices based on the domain scale scores for the PL 2/3 and PL 3/4 cut points tend to be lower than for other cut points, as we might expect.

Although 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, the ranges of the classification accuracy and consistency indices for the ACCESS domains are very similar to those reported for similar testing programs such as ELPA21 (American Institutes of Research, 2018), with the exception of the Writing domain. Since the ACCESS Online Writing test consists of only two tasks, the test score reliability estimate may be lower than similar writing tests that include more tasks. The classification accuracy and consistency indices derived using the Livingston and Lewis (1995) procedure are affected by the magnitude of the test score reliability, which is lower when a test has fewer tasks. Also note that we would not expect the indices estimated for ACCESS domains to be exactly the same as those computed in other programs, because testing programs differ in their student populations, the numbers of proficiency levels, their test designs, their score distributions, and the methods used to compute classification accuracy and consistency indices. For example, compared to similar testing programs, students taking ACCESS represent a much larger and more diverse population. Additionally, the ACCESS testing program defines more proficiency levels than other similar testing programs, and the ACCESS test design is more complex. Therefore, it is difficult to compare the classification accuracy and consistency indices for ACCESS domains to those for other testing programs.

## 5.4.1 Listening

**Table 5.4.1.1**

Overall Accuracy and Consistency of Classification Indices: List S503 Paper

Grade	Accuracy	Consistency
K	0.712	0.652
1	0.492	0.404
2	0.474	0.390
3	0.471	0.387
4	0.484	0.389
5	0.474	0.382
6	0.420	0.337
7	0.391	0.313
8	0.360	0.312
9	0.421	0.338
10	0.430	0.340
11	0.428	0.339
12	0.440	0.342

**Table 5.4.1.2**

Marginal Classification Accuracy Indices Based on the Domain Scale Scores at the Cut Points: List S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.944	0.935	0.928	0.920	0.908
1	0.950	0.899	0.843	0.824	0.844
2	0.965	0.909	0.838	0.810	0.838
3	0.992	0.935	0.845	0.791	0.792
4	0.987	0.932	0.865	0.800	0.791
5	0.979	0.925	0.861	0.795	0.800
6	0.943	0.889	0.840	0.792	0.817
7	0.921	0.882	0.828	0.785	0.818
8	0.917	0.884	0.849	0.797	0.753
9	0.895	0.849	0.815	0.841	0.896
10	0.898	0.850	0.800	0.851	0.917
11	0.891	0.827	0.816	0.861	0.913
12	0.897	0.835	0.801	0.869	0.942

**Table 5.4.1.3**

Marginal Classification Consistency Indices Based on the Domain Scale Scores at the Cut Points: List S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.923	0.911	0.900	0.892	0.871
1	0.927	0.860	0.778	0.759	0.786
2	0.949	0.867	0.778	0.744	0.774
3	0.986	0.905	0.781	0.717	0.719
4	0.978	0.901	0.802	0.729	0.718
5	0.967	0.891	0.796	0.724	0.728
6	0.922	0.842	0.770	0.722	0.751
7	0.895	0.827	0.760	0.713	0.757
8	0.889	0.833	0.777	0.711	0.716
9	0.858	0.783	0.749	0.781	0.856
10	0.860	0.782	0.734	0.787	0.880
11	0.843	0.764	0.746	0.801	0.883
12	0.851	0.772	0.732	0.809	0.913

## 5.4.2 Reading

**Table 5.4.2.1**

Overall Accuracy and Consistency of Classification Indices: Read S503 Paper

Grade	Accuracy	Consistency
K	0.854	0.834
1	0.528	0.424
2	0.570	0.468
3	0.423	0.329
4	0.484	0.386
5	0.477	0.381
6	0.574	0.462
7	0.561	0.454
8	0.548	0.443
9	0.561	0.457
10	0.558	0.455
11	0.550	0.449
12	0.583	0.473

**Table 5.4.2.2**

Marginal Classification Accuracy Indices Based on the Domain Scale Scores at the Cut Points: Read S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.966	0.968	0.960	0.936	N/A
1	0.841	0.836	0.889	0.929	0.964
2	0.886	0.872	0.899	0.915	0.944
3	0.928	0.874	0.779	0.781	0.924
4	0.936	0.872	0.822	0.845	0.914
5	0.932	0.860	0.825	0.839	0.907
6	0.913	0.843	0.880	0.912	0.973
7	0.901	0.842	0.876	0.916	0.966
8	0.890	0.854	0.872	0.902	0.960
9	0.915	0.859	0.876	0.899	0.938
10	0.928	0.863	0.870	0.891	0.932
11	0.924	0.867	0.864	0.880	0.929
12	0.928	0.868	0.881	0.886	0.945

**Table 5.4.2.3**

Marginal Classification Consistency Indices Based on the Domain Scale Scores at the Cut Points: Read S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.949	0.952	0.945	0.931	N/A
1	0.780	0.778	0.844	0.893	0.948
2	0.842	0.823	0.856	0.879	0.921
3	0.904	0.805	0.707	0.716	0.868
4	0.911	0.815	0.761	0.783	0.878
5	0.902	0.803	0.763	0.778	0.866
6	0.875	0.783	0.833	0.877	0.958
7	0.861	0.783	0.829	0.876	0.950
8	0.847	0.795	0.827	0.863	0.940
9	0.879	0.806	0.828	0.856	0.911
10	0.897	0.811	0.820	0.846	0.902
11	0.894	0.814	0.815	0.834	0.895
12	0.897	0.815	0.832	0.845	0.916



### 5.4.3 Writing

**Table 5.4.3.1**

Overall Accuracy and Consistency of Classification Indices: Writ S503 Paper

Grade	Accuracy	Consistency
K	0.802	0.774
1	0.761	0.688
2	0.855	0.806
3	0.882	0.843
4	0.701	0.637
5	0.747	0.667
6	0.774	0.703
7	0.784	0.708
8	0.789	0.713
9	0.774	0.689
10	0.779	0.695
11	0.773	0.687
12	0.774	0.689

**Table 5.4.3.2**

Marginal Classification Accuracy Indices Based on the Domain Scale Scores at the Cut Points: Writ S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.923	0.878	0.973	N/A	N/A
1	0.928	0.847	0.990	N/A	N/A
2	0.959	0.917	0.980	N/A	N/A
3	0.972	0.939	0.972	N/A	N/A
4	0.972	0.955	0.813	0.991	N/A
5	0.977	0.956	0.835	0.977	N/A
6	0.961	0.938	0.875	N/A	N/A
7	0.956	0.931	0.897	N/A	N/A
8	0.955	0.936	0.897	N/A	N/A
9	0.949	0.928	0.896	N/A	N/A
10	0.954	0.928	0.896	N/A	N/A
11	0.949	0.926	0.896	N/A	N/A
12	0.946	0.924	0.902	N/A	N/A

**Table 5.4.3.3**

Marginal Classification Consistency Indices Based on the Domain Scale Scores at the Cut Points: Writ S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.889	0.872	0.970	N/A	N/A
1	0.897	0.796	0.990	N/A	N/A
2	0.940	0.887	0.979	N/A	N/A
3	0.959	0.916	0.969	N/A	N/A
4	0.960	0.937	0.738	0.991	N/A
5	0.966	0.938	0.777	0.977	N/A
6	0.944	0.912	0.843	N/A	N/A
7	0.938	0.903	0.863	N/A	N/A
8	0.938	0.907	0.861	N/A	N/A
9	0.928	0.900	0.854	N/A	N/A
10	0.934	0.899	0.854	N/A	N/A
11	0.928	0.895	0.855	N/A	N/A
12	0.924	0.892	0.862	N/A	N/A

## 5.4.4 Speaking

**Table 5.4.4.1**

Overall Accuracy and Consistency of Classification Indices: Spek S503 Paper

Grade	Accuracy	Consistency
K	0.512	0.516
1	0.694	0.588
2	0.674	0.570
3	0.666	0.559
4	0.643	0.538
5	0.629	0.528
6	0.627	0.532
7	0.650	0.553
8	0.643	0.548
9	0.677	0.594
10	0.674	0.593
11	0.669	0.589
12	0.666	0.591

**Table 5.4.4.2**

Marginal Classification Accuracy Indices Based on the Domain Scale Scores at the Cut Points: Spek S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.894	0.931	0.953	0.955	0.754
1	0.936	0.895	0.916	0.953	0.991
2	0.951	0.900	0.900	0.941	0.970
3	0.951	0.897	0.895	0.923	0.956
4	0.958	0.922	0.890	0.898	0.963
5	0.949	0.915	0.895	0.896	0.948
6	0.949	0.924	0.904	0.913	0.924
7	0.943	0.922	0.912	0.905	0.954
8	0.944	0.922	0.903	0.917	0.938
9	0.931	0.918	0.918	0.951	0.939
10	0.935	0.913	0.918	0.965	0.931
11	0.931	0.914	0.924	0.962	0.923
12	0.937	0.912	0.930	0.967	0.909

**Table 5.4.4.3**

Marginal Classification Consistency Indices Based on the Domain Scale Scores at the Cut Points: Spek S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.857	0.904	0.929	0.931	0.816
1	0.907	0.854	0.880	0.936	0.991
2	0.928	0.860	0.860	0.922	0.970
3	0.926	0.856	0.853	0.922	0.956
4	0.938	0.889	0.849	0.857	0.961
5	0.926	0.882	0.855	0.860	0.947
6	0.927	0.893	0.867	0.868	0.921
7	0.918	0.891	0.876	0.866	0.947
8	0.919	0.890	0.866	0.877	0.936
9	0.901	0.884	0.885	0.923	0.939
10	0.905	0.878	0.883	0.941	0.939
11	0.902	0.880	0.892	0.939	0.931
12	0.910	0.877	0.898	0.950	0.923

## 5.5 Reliabilities of Students' Composite Scale Scores

The reliabilities of the ACCESS composite scale scores indicate the consistency of those scores over replications of the testing procedure. Because the domains that make up the composites consist of different test items, and because items from different domains may measure different abilities (even though items within the domain are assumed to measure a single ability), a traditional internal consistency index such as Cronbach's coefficient alpha is not appropriate, since statisticians who devised such indices assumed that items in a test measure similar ability. It is more appropriate to report a stratified Cronbach's coefficient alpha (Feldt & Brennan, 1989), which measures consistency in students' composite scale scores when those scores are based on students' responses to sets of items that measure different abilities. A stratified alpha is a weighted average of Cronbach's coefficient alphas for item sets that differ in the maximum score points or "strata." Stratified alpha is a reliability estimate computed by dividing the test into components (strata), computing a Cronbach's coefficient alpha separately for the scale scores for each component, and then using the results to estimate a reliability coefficient for the composite scale scores.

In computing the stratified Cronbach's coefficient alphas for ACCESS composite scale scores, we treated each domain that makes up a composite as a separate component (or stratum). For example, when computing the stratified Cronbach's coefficient alphas for students' Literacy scale scores, we entered the variances of the students' scale scores for two components (i.e., Reading and Writing) and the weights of those two components. The stratified Cronbach's coefficient alpha is interpreted like other traditional internal consistency statistics such as Cronbach's coefficient alpha. Like Cronbach's coefficient alpha, a stratified Cronbach's coefficient alpha is an estimate of the proportion of the total variance in the students' composite scale scores that the variance in their true composite scale scores can explain.

Because of the differential weights applied to the ACCESS domains that contribute to the students' composite scale scores, the stratified Cronbach's coefficient alpha is weighted by the contribution that each domain makes to the students' composite scale scores (Kamata, Turhan, & Darandari, 2003; Kane & Case, 2004; Rudner, 2001). Specifically, the formula is

$$\alpha_c = 1 - \frac{\sum_{j=1}^k w_j^2 \sigma_j^2 (1 - \rho_j)}{\sigma_c^2}$$

where

$k$  = the number of components (domains)  $j$  that contribute to the composite

$w_j$  = the weight of component (domain)  $j$

$\sigma_j^2$  = the variance of the students' scale scores for component (domain)  $j$

$\sigma_c^2$  = the variance of the students' composite scale scores

$\rho_j$  = the reliability coefficient for students' scale scores for component (domain)  $j$ .

As is true for the Cronbach's coefficient alpha (see the explanation in Section 5), there is no one set of criteria that the testing community uses when interpreting stratified Cronbach's coefficient alpha values. There is little consensus among the experts in their views of what the acceptable lower limit of the stratified Cronbach's coefficient alpha value should be, or for that matter, how one should interpret various values. This lack of consensus led the authors of the *Standards for Educational and Psychological Measurement* (2014) to conclude, "The choice of [reliability/precision] estimation and the minimum acceptable level for any index remain a matter of professional judgment" (p. 41).

The tables report the stratified Cronbach's coefficient alphas for the students' scale scores for each of the four composites (Oral, Literacy, Comprehension, Overall). The first table for each composite provides stratified Cronbach's coefficient alphas for all students' composite scale scores. The second table for each composite provides the same information for the population of female students and for the population of male students. The third table provides information by ethnicity, for Hispanic and for non-Hispanic students, and the fourth table provides information for the population of students who have an IEP.

The first column of each table shows the grade-level clusters. The tables report the input values that we used to compute the stratified Cronbach's coefficient alphas (i.e., the number of components for each composite, each component's weight, and the variance of the students' scale scores for each component). See Chapter 3 for an explanation of the procedures we used to compute the composite scale scores.

For each grade-level cluster excluding Kindergarten, we derive a reliability coefficient across tiers for each domain. (The Kindergarten test is not tiered and so this step is not necessary.) To produce this coefficient, values for Cronbach's alpha for each of the tiers in the grade-level cluster (provided in Section 5.1) are weighted by the number of students who were administered the tier form, and a weighted average is expressed in the tables.

For each relevant domain component, we report the variance of the students' domain scale scores. We also report the variance of the students' composite scale scores. When we computed the variances of the students' domain scale scores and the variances of the students' composite scale scores, we included the students who had valid scores for all four domains.

Finally, the tables present the computed stratified Cronbach's coefficient alphas for students' scale scores for each composite, by grade-level cluster.

Additionally, we used the stratified Cronbach's coefficient alphas, presented in the tables in this section, to produce the **Accuracy and Consistency** classification tables for the composites (Section 5.7).

The stratified Cronbach's alpha of the Oral composite computed for all students ranged from 0.88 to 0.96. The stratified Cronbach's alpha of the Oral composite ranged from 0.89 to 0.96 for male students; from 0.88 to 0.96 for female students; from 0.89 to 0.96 for Hispanic students; from 0.88 to 0.95 for non-Hispanic students; and from 0.86 to 0.96 for students with an IEP.

The stratified Cronbach's alpha of the Literacy composite computed for all students ranged from 0.90 to 0.96. The stratified Cronbach's alpha of the Literacy composite ranged from 0.90 to 0.97 for male students; from 0.89 to 0.96 for female students; from 0.90 to 0.96 for Hispanic students; from 0.91 to 0.97 for non-Hispanic students; and from 0.88 to 0.97 for students with an IEP.

The stratified Cronbach's alpha of the Comprehension composite computed for all students ranged from 0.77 to 0.97. The stratified Cronbach's alpha of the Comprehension composite ranged from 0.78 to 0.97

for male students; from 0.75 to 0.96 for female students; from 0.76 to 0.96 for Hispanic students; from 0.79 to 0.97 for non-Hispanic students; and from 0.69 to 0.97 for students with an IEP.

The stratified Cronbach's alpha of the Overall composite computed for all students ranged from 0.94 to 0.98. The stratified Cronbach's alpha of the Overall composite ranged from 0.94 to 0.98 for male students; from 0.93 to 0.98 for female students; from 0.93 to 0.97 for Hispanic students; from 0.94 to 0.98 for non-Hispanic students; and from 0.92 to 0.98 for students with an IEP.

## 5.5.1 Oral

**Table 5.5.1.1**

Reliabilities of Composite Scale Scores: Oral S503 Paper

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.50	6628.21	0.95
	Speaking	0.50	10873.52	0.92
	Oral		7784.36	0.96
1	Listening	0.50	1680.44	0.72
	Speaking	0.50	4375.79	0.90
	Oral		2260.79	0.90
2	Listening	0.50	1798.27	0.70
	Speaking	0.50	4443.61	0.91
	Oral		2373.74	0.90
3	Listening	0.50	1377.91	0.60
	Speaking	0.50	4367.13	0.91
	Oral		2067.25	0.88
4-5	Listening	0.50	1763.85	0.64
	Speaking	0.50	5147.51	0.90
	Oral		2635.71	0.89
6-8	Listening	0.50	2531.68	0.65
	Speaking	0.50	6541.97	0.92
	Oral		3664.13	0.90
9-12	Listening	0.50	2418.58	0.67
	Speaking	0.50	6479.95	0.92
	Oral		3536.27	0.91

**Table 5.5.1.2**

Reliabilities of Composite Scale Scores: Oral S503 Paper by Gender

Cluster	Component	Weight	Female		Male	
			Variance	Reliability	Variance	Reliability
K	Listening	0.50	6410.96	0.95	6768.89	0.95
	Speaking	0.50	11047.54	0.92	10511.85	0.91
	Oral		7739.96	0.96	7694.61	0.96
1	Listening	0.50	1616.49	0.71	1723.58	0.73
	Speaking	0.50	4389.53	0.90	4315.20	0.90
	Oral		2230.61	0.90	2260.32	0.90
2	Listening	0.50	1753.76	0.69	1829.74	0.71
	Speaking	0.50	4510.62	0.91	4363.01	0.91
	Oral		2379.54	0.90	2353.73	0.90
3	Listening	0.50	1293.69	0.57	1445.63	0.61
	Speaking	0.50	4439.67	0.91	4280.74	0.91
	Oral		2056.35	0.88	2064.79	0.89
4-5	Listening	0.50	1680.84	0.63	1828.17	0.65
	Speaking	0.50	5180.14	0.90	5094.41	0.91
	Oral		2602.34	0.89	2650.11	0.89
6-8	Listening	0.50	2503.18	0.64	2544.18	0.66
	Speaking	0.50	6543.07	0.92	6502.26	0.92
	Oral		3652.30	0.90	3650.90	0.90
9-12	Listening	0.50	2318.88	0.67	2501.86	0.68
	Speaking	0.50	6420.92	0.92	6514.98	0.92
	Oral		3460.68	0.91	3593.05	0.91



**Table 5.5.1.3**

Reliabilities of Composite Scale Scores: Oral S503 Paper by Ethnicity

Cluster	Component	Weight	Hispanic		Other	
			Variance	Reliability	Variance	Reliability
K	Listening	0.50	6718.66	0.95	5582.67	0.94
	Speaking	0.50	10893.49	0.92	9746.28	0.90
	Oral		7838.96	0.96	6690.71	0.95
1	Listening	0.50	1675.22	0.72	1674.44	0.73
	Speaking	0.50	4418.83	0.91	4000.69	0.89
	Oral		2278.09	0.90	2098.22	0.89
2	Listening	0.50	1830.78	0.70	1649.21	0.71
	Speaking	0.50	4548.31	0.91	3912.09	0.90
	Oral		2438.80	0.90	2059.03	0.89
3	Listening	0.50	1372.17	0.59	1386.58	0.62
	Speaking	0.50	4472.43	0.91	3828.77	0.90
	Oral		2108.31	0.89	1847.77	0.88
4-5	Listening	0.50	1791.55	0.64	1610.52	0.65
	Speaking	0.50	5337.21	0.91	4167.06	0.89
	Oral		2731.23	0.89	2136.40	0.88
6-8	Listening	0.50	2602.25	0.64	2127.26	0.66
	Speaking	0.50	6803.11	0.92	5092.54	0.91
	Oral		3812.48	0.90	2827.91	0.90
9-12	Listening	0.50	2474.93	0.67	2038.83	0.68
	Speaking	0.50	6650.90	0.92	5291.85	0.92
	Oral		3645.87	0.91	2793.27	0.90

**Table 5.5.1.4**

Reliabilities of Composite Scale Scores: Oral S503 Paper by IEP Status

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.50	6958.23	0.96
	Speaking	0.50	8870.68	0.90
	Oral		7083.04	0.96
1	Listening	0.50	1690.20	0.74
	Speaking	0.50	3463.79	0.90
	Oral		1882.93	0.89
2	Listening	0.50	1690.24	0.74
	Speaking	0.50	3608.15	0.90
	Oral		1919.01	0.89
3	Listening	0.50	1173.66	0.57
	Speaking	0.50	3188.66	0.89
	Oral		1469.31	0.86
4-5	Listening	0.50	1331.41	0.62
	Speaking	0.50	3299.54	0.90
	Oral		1599.49	0.87
6-8	Listening	0.50	1444.73	0.60
	Speaking	0.50	3726.72	0.91
	Oral		1826.59	0.87
9-12	Listening	0.50	1553.86	0.63
	Speaking	0.50	4946.97	0.91
	Oral		2355.39	0.89

## 5.5.2 Literacy

**Table 5.5.2.1**

Reliabilities of Composite Scale Scores: Litr S503 Paper

Cluster	Component	Weight	Variance	Reliability
K	Reading	0.50	4506.44	0.95
	Writing	0.50	4611.20	0.92
	Literacy		3922.94	0.96
1	Reading	0.50	942.77	0.76
	Writing	0.50	2020.06	0.92
	Literacy		1099.33	0.91
2	Reading	0.50	1228.63	0.84
	Writing	0.50	1967.63	0.94
	Literacy		1303.00	0.94
3	Reading	0.50	798.89	0.66
	Writing	0.50	1659.81	0.94
	Literacy		952.32	0.90
4-5	Reading	0.50	993.45	0.77
	Writing	0.50	2109.74	0.91
	Literacy		1265.03	0.92
6-8	Reading	0.50	907.36	0.78
	Writing	0.50	1829.39	0.91
	Literacy		1128.65	0.92
9-12	Reading	0.50	1179.38	0.82
	Writing	0.50	2109.82	0.90
	Literacy		1367.31	0.92

**Table 5.5.2.2**

Reliabilities of Composite Scale Scores: LitR S503 Paper by Gender

Cluster	Component	Weight	Female		Male	
			Variance	Reliability	Variance	Reliability
K	Reading	0.50	4297.85	0.95	4650.39	0.95
	Writing	0.50	4527.92	0.92	4637.31	0.93
	Literacy		3787.12	0.96	3999.43	0.97
1	Reading	0.50	910.28	0.75	971.06	0.77
	Writing	0.50	1874.47	0.91	2111.11	0.92
	Literacy		1041.48	0.91	1137.85	0.91
2	Reading	0.50	1179.23	0.83	1270.18	0.84
	Writing	0.50	1828.48	0.94	2039.00	0.94
	Literacy		1237.94	0.94	1342.66	0.94
3	Reading	0.50	747.81	0.63	841.87	0.68
	Writing	0.50	1573.17	0.93	1673.48	0.94
	Literacy		906.18	0.89	974.53	0.90
4-5	Reading	0.50	933.54	0.75	1041.14	0.78
	Writing	0.50	2068.01	0.91	2062.04	0.91
	Literacy		1226.98	0.91	1269.90	0.92
6-8	Reading	0.50	848.30	0.77	954.20	0.78
	Writing	0.50	1799.12	0.90	1808.99	0.91
	Literacy		1085.41	0.91	1146.75	0.92
9-12	Reading	0.50	1140.53	0.81	1202.81	0.83
	Writing	0.50	2076.84	0.90	2096.52	0.90
	Literacy		1343.21	0.92	1364.79	0.92

**Table 5.5.2.3**

Reliabilities of Composite Scale Scores: LitR S503 Paper by Ethnicity

Cluster	Component	Weight	Hispanic		Other	
			Variance	Reliability	Variance	Reliability
K	Reading	0.50	3832.53	0.95	5109.56	0.96
	Writing	0.50	4125.79	0.91	4853.52	0.93
	Literacy		3354.86	0.96	4309.05	0.97
1	Reading	0.50	911.17	0.75	1027.20	0.80
	Writing	0.50	1992.93	0.92	2066.40	0.92
	Literacy		1071.12	0.91	1167.05	0.92
2	Reading	0.50	1225.36	0.83	1221.62	0.85
	Writing	0.50	1977.96	0.94	1867.26	0.94
	Literacy		1305.25	0.94	1260.98	0.94
3	Reading	0.50	795.90	0.65	798.39	0.70
	Writing	0.50	1665.65	0.94	1601.37	0.93
	Literacy		952.84	0.90	931.06	0.91
4-5	Reading	0.50	1000.12	0.76	941.59	0.79
	Writing	0.50	2162.05	0.91	1862.45	0.91
	Literacy		1291.38	0.92	1131.43	0.92
6-8	Reading	0.50	901.05	0.77	893.32	0.79
	Writing	0.50	1875.91	0.91	1563.38	0.91
	Literacy		1146.23	0.92	1001.70	0.92
9-12	Reading	0.50	1196.65	0.82	1039.69	0.83
	Writing	0.50	2151.26	0.90	1855.43	0.90
	Literacy		1398.73	0.92	1164.37	0.92

**Table 5.5.2.4**

Reliabilities of Composite Scale Scores: LitR S503 Paper by IEP Status

Cluster	Component	Weight	Variance	Reliability
K	Reading	0.50	4632.10	0.96
	Writing	0.50	3976.63	0.92
	Literacy		3584.57	0.97
1	Reading	0.50	773.37	0.69
	Writing	0.50	2410.81	0.92
	Literacy		1049.15	0.90
2	Reading	0.50	994.32	0.80
	Writing	0.50	2115.34	0.95
	Literacy		1197.35	0.94
3	Reading	0.50	666.75	0.55
	Writing	0.50	1667.00	0.94
	Literacy		853.47	0.88
4-5	Reading	0.50	677.01	0.70
	Writing	0.50	1626.35	0.92
	Literacy		864.83	0.90
6-8	Reading	0.50	593.48	0.71
	Writing	0.50	1280.20	0.92
	Literacy		704.07	0.90
9-12	Reading	0.50	835.59	0.79
	Writing	0.50	1683.91	0.92
	Literacy		965.30	0.92

### 5.5.3 Comprehension

**Table 5.5.3.1**

Reliabilities of Composite Scale Scores: Cphn S503 Paper

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.30	6628.21	0.95
	Reading	0.70	4506.44	0.95
	Comprehension		3999.38	0.97
1	Listening	0.30	1680.44	0.72
	Reading	0.70	942.77	0.76
	Comprehension		875.49	0.83
2	Listening	0.30	1798.27	0.70
	Reading	0.70	1228.63	0.84
	Comprehension		1137.96	0.87
3	Listening	0.30	1377.91	0.60
	Reading	0.70	798.89	0.66
	Comprehension		793.26	0.77
4-5	Listening	0.30	1763.85	0.64
	Reading	0.70	993.45	0.77
	Comprehension		1043.10	0.84
6-8	Listening	0.30	2531.68	0.65
	Reading	0.70	907.36	0.78
	Comprehension		1140.46	0.84
9-12	Listening	0.30	2418.58	0.67
	Reading	0.70	1179.38	0.82
	Comprehension		1332.28	0.87

**Table 5.5.3.2**

Reliabilities of Composite Scale Scores: Cphn S503 Paper by Gender

Cluster	Component	Weight	Female		Male	
			Variance	Reliability	Variance	Reliability
K	Listening	0.30	6410.96	0.95	6768.89	0.95
	Reading	0.70	4297.85	0.95	4650.39	0.95
	Comprehension		3846.73	0.96	4100.17	0.97
1	Listening	0.30	1616.49	0.71	1723.58	0.73
	Reading	0.70	910.28	0.75	971.06	0.77
	Comprehension		850.30	0.82	894.04	0.83
2	Listening	0.30	1753.76	0.69	1829.74	0.71
	Reading	0.70	1179.23	0.83	1270.18	0.84
	Comprehension		1099.18	0.87	1169.29	0.88
3	Listening	0.30	1293.69	0.57	1445.63	0.61
	Reading	0.70	747.81	0.63	841.87	0.68
	Comprehension		739.96	0.75	837.26	0.78
4-5	Listening	0.30	1680.84	0.63	1828.17	0.65
	Reading	0.70	933.54	0.75	1041.14	0.78
	Comprehension		983.42	0.83	1090.25	0.84
6-8	Listening	0.30	2503.18	0.64	2544.18	0.66
	Reading	0.70	848.30	0.77	954.20	0.78
	Comprehension		1087.41	0.84	1181.30	0.85
9-12	Listening	0.30	2318.88	0.67	2501.86	0.68
	Reading	0.70	1140.53	0.81	1202.81	0.83
	Comprehension		1287.21	0.86	1364.02	0.87



**Table 5.5.3.3**

Reliabilities of Composite Scale Scores: Cphn S503 Paper by Ethnicity

Cluster	Component	Weight	Hispanic		Other	
			Variance	Reliability	Variance	Reliability
K	Listening	0.30	6718.66	0.95	5582.67	0.94
	Reading	0.70	3832.53	0.95	5109.56	0.96
	Comprehension		3538.73	0.96	4159.40	0.97
1	Listening	0.30	1675.22	0.72	1674.44	0.73
	Reading	0.70	911.17	0.75	1027.20	0.80
	Comprehension		850.91	0.82	939.76	0.85
2	Listening	0.30	1830.78	0.70	1649.21	0.71
	Reading	0.70	1225.36	0.83	1221.62	0.85
	Comprehension		1142.92	0.87	1100.98	0.88
3	Listening	0.30	1372.17	0.59	1386.58	0.62
	Reading	0.70	795.90	0.65	798.39	0.70
	Comprehension		790.84	0.76	790.82	0.79
4-5	Listening	0.30	1791.55	0.64	1610.52	0.65
	Reading	0.70	1000.12	0.76	941.59	0.79
	Comprehension		1056.78	0.83	957.99	0.84
6-8	Listening	0.30	2602.25	0.64	2127.26	0.66
	Reading	0.70	901.05	0.77	893.32	0.79
	Comprehension		1150.83	0.84	1039.45	0.85
9-12	Listening	0.30	2474.93	0.67	2038.83	0.68
	Reading	0.70	1196.65	0.82	1039.69	0.83
	Comprehension		1358.98	0.87	1136.02	0.87

**Table 5.5.3.4**

Reliabilities of Composite Scale Scores: Cphn S503 Paper by IEP Status

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.30	6958.23	0.96
	Reading	0.70	4632.10	0.96
	Comprehension		3930.14	0.97
1	Listening	0.30	1690.20	0.74
	Reading	0.70	773.37	0.69
	Comprehension		724.05	0.78
2	Listening	0.30	1690.24	0.74
	Reading	0.70	994.32	0.80
	Comprehension		911.26	0.85
3	Listening	0.30	1173.66	0.57
	Reading	0.70	666.75	0.55
	Comprehension		625.04	0.69
4-5	Listening	0.30	1331.41	0.62
	Reading	0.70	677.01	0.70
	Comprehension		682.63	0.79
6-8	Listening	0.30	1444.73	0.60
	Reading	0.70	593.48	0.71
	Comprehension		655.15	0.79
9-12	Listening	0.30	1553.86	0.63
	Reading	0.70	835.59	0.79
	Comprehension		862.03	0.84

## 5.5.4 Overall

**Table 5.5.4.1**

Reliabilities of Composite Scale Scores: Over S503 Paper

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.15	6628.21	0.95
	Reading	0.35	4506.44	0.95
	Writing	0.35	4611.20	0.92
	Speaking	0.15	10873.52	0.92
	Overall Composite		4003.87	0.98
1	Listening	0.15	1680.44	0.72
	Reading	0.35	942.77	0.76
	Writing	0.35	2020.06	0.92
	Speaking	0.15	4375.79	0.90
	Overall Composite		1149.64	0.94
2	Listening	0.15	1798.27	0.70
	Reading	0.35	1228.63	0.84
	Writing	0.35	1967.63	0.94
	Speaking	0.15	4443.61	0.91
	Overall Composite		1338.00	0.96
3	Listening	0.15	1377.91	0.60
	Reading	0.35	798.89	0.66
	Writing	0.35	1659.81	0.94
	Speaking	0.15	4367.13	0.91
	Overall Composite		1056.59	0.94
4-5	Listening	0.15	1763.85	0.64
	Reading	0.35	993.45	0.77
	Writing	0.35	2109.74	0.91
	Speaking	0.15	5147.51	0.90
	Overall Composite		1436.85	0.95
6-8	Listening	0.15	2531.68	0.65
	Reading	0.35	907.36	0.78
	Writing	0.35	1829.39	0.91
	Speaking	0.15	6541.97	0.92
	Overall Composite		1578.66	0.95
9-12	Listening	0.15	2418.58	0.67
	Reading	0.35	1179.38	0.82
	Writing	0.35	2109.82	0.90
	Speaking	0.15	6479.95	0.92
	Overall Composite		1727.88	0.95

**Table 5.5.4.2**

Reliabilities of Composite Scale Scores: Over S503 Paper by Gender

Cluster	Component	Weight	Female		Male	
			Variance	Reliability	Variance	Reliability
K	Listening	0.15	6410.96	0.95	6768.89	0.95
	Reading	0.35	4297.85	0.95	4650.39	0.95
	Writing	0.35	4527.92	0.92	4637.31	0.93
	Speaking	0.15	11047.54	0.92	10511.85	0.91
	Overall Composite		3908.21	0.98	4038.24	0.98
1	Listening	0.15	1616.49	0.71	1723.58	0.73
	Reading	0.35	910.28	0.75	971.06	0.77
	Writing	0.35	1874.47	0.91	2111.11	0.92
	Speaking	0.15	4389.53	0.90	4315.20	0.90
	Overall Composite		1109.06	0.94	1169.73	0.94
2	Listening	0.15	1753.76	0.69	1829.74	0.71
	Reading	0.35	1179.23	0.83	1270.18	0.84
	Writing	0.35	1828.48	0.94	2039.00	0.94
	Speaking	0.15	4510.62	0.91	4363.01	0.91
	Overall Composite		1299.83	0.95	1356.05	0.96
3	Listening	0.15	1293.69	0.57	1445.63	0.61
	Reading	0.35	747.81	0.63	841.87	0.68
	Writing	0.35	1573.17	0.93	1673.48	0.94
	Speaking	0.15	4439.67	0.91	4280.74	0.91
	Overall Composite		1029.47	0.93	1066.94	0.94
4-5	Listening	0.15	1680.84	0.63	1828.17	0.65
	Reading	0.35	933.54	0.75	1041.14	0.78
	Writing	0.35	2068.01	0.91	2062.04	0.91
	Speaking	0.15	5180.14	0.90	5094.41	0.91
	Overall Composite		1409.58	0.95	1441.16	0.95
6-8	Listening	0.15	2503.18	0.64	2544.18	0.66
	Reading	0.35	848.30	0.77	954.20	0.78
	Writing	0.35	1799.12	0.90	1808.99	0.91
	Speaking	0.15	6543.07	0.92	6502.26	0.92
	Overall Composite		1551.68	0.95	1587.78	0.95
9-12	Listening	0.15	2318.88	0.67	2501.86	0.68
	Reading	0.35	1140.53	0.81	1202.81	0.83
	Writing	0.35	2076.84	0.90	2096.52	0.90
	Speaking	0.15	6420.92	0.92	6514.98	0.92
	Overall Composite		1708.32	0.95	1730.67	0.95

**Table 5.5.4.3**

Reliabilities of Composite Scale Scores: Over S503 Paper by Ethnicity

Cluster	Component	Weight	Hispanic		Other	
			Variance	Reliability	Variance	Reliability
K	Listening	0.15	6718.66	0.95	5582.67	0.94
	Reading	0.35	3832.53	0.95	5109.56	0.96
	Writing	0.35	4125.79	0.91	4853.52	0.93
	Speaking	0.15	10893.49	0.92	9746.28	0.90
	Overall Composite		3586.94	0.97	4039.24	0.98
1	Listening	0.15	1675.22	0.72	1674.44	0.73
	Reading	0.35	911.17	0.75	1027.20	0.80
	Writing	0.35	1992.93	0.92	2066.40	0.92
	Speaking	0.15	4418.83	0.91	4000.69	0.89
	Overall Composite		1130.55	0.94	1168.81	0.94
2	Listening	0.15	1830.78	0.70	1649.21	0.71
	Reading	0.35	1225.36	0.83	1221.62	0.85
	Writing	0.35	1977.96	0.94	1867.26	0.94
	Speaking	0.15	4548.31	0.91	3912.09	0.90
	Overall Composite		1353.94	0.96	1237.45	0.96
3	Listening	0.15	1372.17	0.59	1386.58	0.62
	Reading	0.35	795.90	0.65	798.39	0.70
	Writing	0.35	1665.65	0.94	1601.37	0.93
	Speaking	0.15	4472.43	0.91	3828.77	0.90
	Overall Composite		1066.09	0.93	990.64	0.94
4-5	Listening	0.15	1791.55	0.64	1610.52	0.65
	Reading	0.35	1000.12	0.76	941.59	0.79
	Writing	0.35	2162.05	0.91	1862.45	0.91
	Speaking	0.15	5337.21	0.91	4167.06	0.89
	Overall Composite		1481.64	0.95	1209.37	0.94
6-8	Listening	0.15	2602.25	0.64	2127.26	0.66
	Reading	0.35	901.05	0.77	893.32	0.79
	Writing	0.35	1875.91	0.91	1563.38	0.91
	Speaking	0.15	6803.11	0.92	5092.54	0.91
	Overall Composite		1627.42	0.95	1285.20	0.95
9-12	Listening	0.15	2474.93	0.67	2038.83	0.68
	Reading	0.35	1196.65	0.82	1039.69	0.83
	Writing	0.35	2151.26	0.90	1855.43	0.90
	Speaking	0.15	6650.90	0.92	5291.85	0.92
	Overall Composite		1781.40	0.95	1384.84	0.95

**Table 5.5.4.4**

Reliabilities of Composite Scale Scores: Over S503 Paper by IEP Status

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.15	6958.23	0.96
	Reading	0.35	4632.10	0.96
	Writing	0.35	3976.63	0.92
	Speaking	0.15	8870.68	0.90
	Overall Composite		3561.80	0.98
1	Listening	0.15	1690.20	0.74
	Reading	0.35	773.37	0.69
	Writing	0.35	2410.81	0.92
	Speaking	0.15	3463.79	0.90
	Overall Composite		1008.11	0.93
2	Listening	0.15	1690.24	0.74
	Reading	0.35	994.32	0.80
	Writing	0.35	2115.34	0.95
	Speaking	0.15	3608.15	0.90
	Overall Composite		1135.10	0.95
3	Listening	0.15	1173.66	0.57
	Reading	0.35	666.75	0.55
	Writing	0.35	1667.00	0.94
	Speaking	0.15	3188.66	0.89
	Overall Composite		821.75	0.92
4-5	Listening	0.15	1331.41	0.62
	Reading	0.35	677.01	0.70
	Writing	0.35	1626.35	0.92
	Speaking	0.15	3299.54	0.90
	Overall Composite		862.37	0.93
6-8	Listening	0.15	1444.73	0.60
	Reading	0.35	593.48	0.71
	Writing	0.35	1280.20	0.92
	Speaking	0.15	3726.72	0.91
	Overall Composite		808.73	0.93
9-12	Listening	0.15	1553.86	0.63
	Reading	0.35	835.59	0.79
	Writing	0.35	1683.91	0.92
	Speaking	0.15	4946.97	0.91
	Overall Composite		1114.51	0.94

## 5.6 Conditional Standard Error of Measurement for Composites

CSEMs for the four ACCESS composite scale scores provide test users with a benchmark indicating how free a student's composite scale score is from measurement errors at different WIDA proficiency levels. Due to the differential weights applied to different ACCESS domains (see the introduction to Section 3 for weighting conventions), WIDA estimates the CSEMs using a procedure that is based on IRT (Lord, 1980) and developed by Price, Lurie, Raju, Wilkins, and Zhu (2006). Price et al. (2006) extended the work by Lord (1980) and Kolen, Hanson, and Brennan (1992) in estimating the CSEMs of students' composite scale scores consisting of components. The basic premise of this procedure is that one can estimate empirically the CSEM for a student's weighted composite scale score using the IRT-based CSEMs for each student's component scale scores and the weights associated with the components. We used this method to estimate the CSEMs for ACCESS composite scale scores by treating the ACCESS domains as components.

We used a three-step process to derive the CSEM for each ACCESS composite scale score. We calculated a unique CSEM for each composite scale score by grade. Since this procedure relies on empirical student data, which are subject to year-to-year fluctuations, we used all population student data from all previous three ACCESS 2.0 series in our calculations to obtain more stable estimates than using data from just a single series.

**Step 1.** Since we calibrated ACCESS domains separately, measurement errors associated with each of the ACCESS domains, as expressed in the CSEM, were independent of each other. Therefore, we estimated the CSEM for a student's composite scale score  $x$ ,  $SEM_x$ , using the equation derived by Price et al. (2006):

$$SEM_x = \sqrt{W_1^2 SEM_1^2 + W_2^2 SEM_2^2 + W_3^2 SEM_3^2 + \dots + W_k^2 SEM_k^2}$$

Where  $SEM_i^2$  is the student's IRT-based score error variance or student's squared CSEM in ACCESS domain  $i$  and  $W_i$  is the weight applied to domain  $i$ , for  $i=1, \dots, k$ .

**Step 2.** Due to the differential weights applied to different ACCESS domains, two students with the same sum of weighted domain score, or composite, may obtain different CSEMs; therefore, we took an additional step to obtain a unique value for each composite score. Specifically, we estimated the expected value of the CSEM functions for a composite score using a regression approach, and we reported this expected value as the CSEM for that composite score.

**Step 3.** We applied a linear smoothing procedure to derive the CSEMs for composite scale scores that we did not observe in the data.

The figures in this section show graphically the CSEMs for various composite scale scores by grade level. The students' composite scale scores appear on the horizontal axis, and the corresponding CSEMs appear on the vertical axis. Each point in a figure represents a student in the dataset, showing the relationship between the CSEM and that student's composite scale score. We did not plot values for students who received the lowest possible scale scores for any ACCESS domains, as it is not possible to compute accurately the CSEM for these students' scale scores. For grade-level clusters with multiple grades, we use different colors in the figures to represent students in different grades.

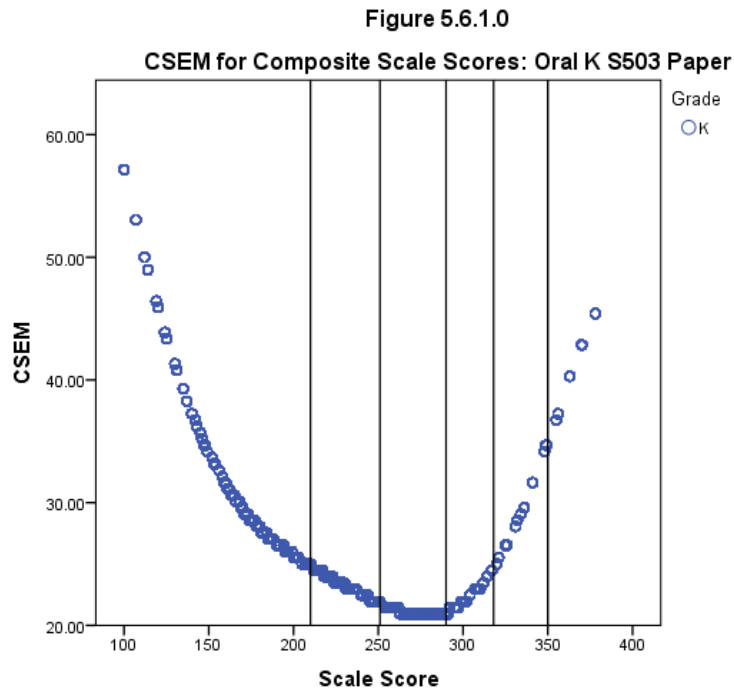
The five vertical lines in the figure indicate the five ACCESS composite scale score cut points for the highest grade in the grade-level cluster for the test form, dividing the figure into six sections representing the six WIDA proficiency levels.

Smaller CSEM values indicate less measurement error (i.e., greater accuracy in measurement). In general, these figures show that the CSEMs are smaller and fairly constant in the middle of the composite scale score range but larger and more variable for extreme low and high composite scale scores. This is to be expected, since we used an IRT approach when scaling ACCESS, which typically produces larger CSEMs for scale scores that are at the lower and the higher ends of the scale score range. In addition, because students exit the EL program when they demonstrate that they are English language proficient, the number of students whose composite scale scores are at the extreme high end of the score range is typically small, as compared to the number of students whose composite scale scores are in the middle of the score range. Therefore, the measurement errors associated with the composite scale scores at the extreme high end of the score range tend to be larger since the calculation of these scale scores is based on the test performances of fewer students.

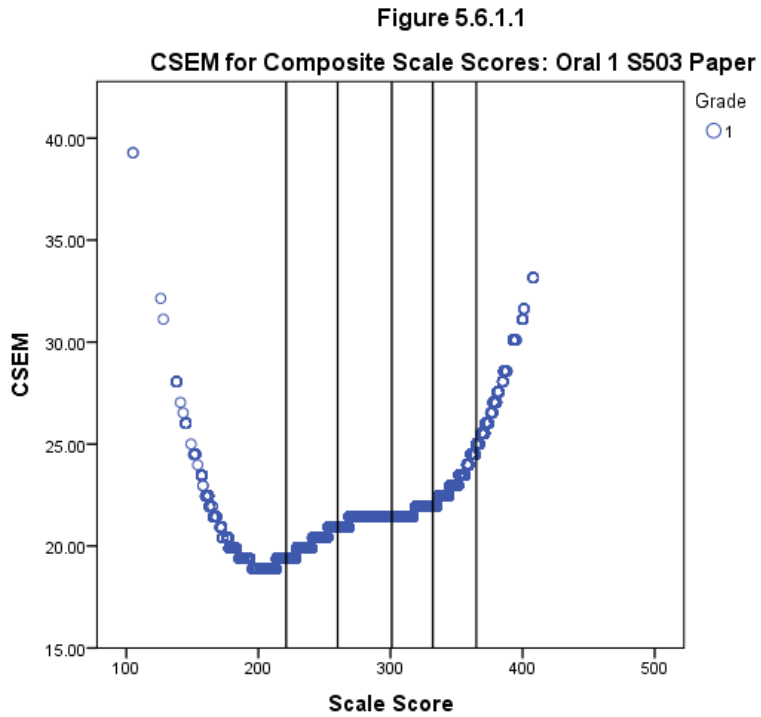


## 5.6.1 Oral

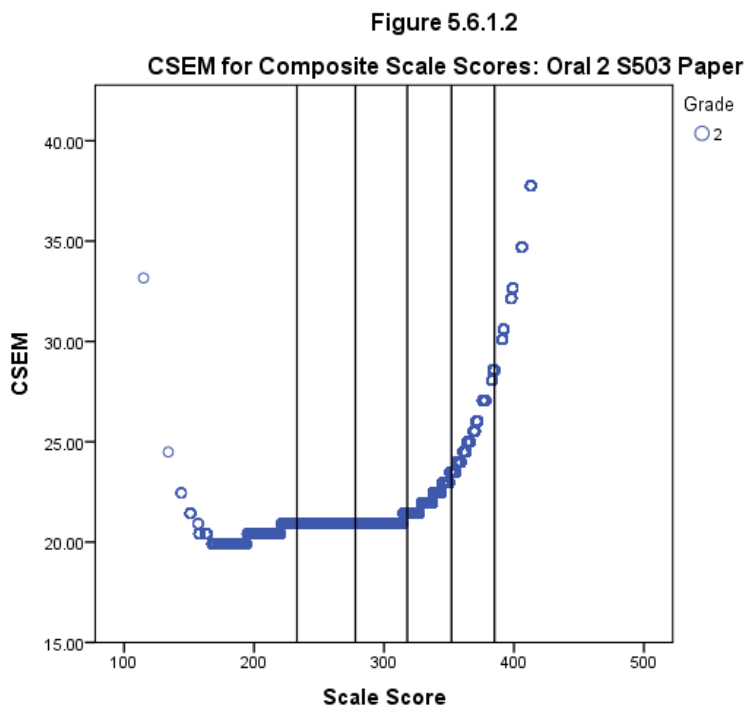
### 5.6.1.0 Kindergarten



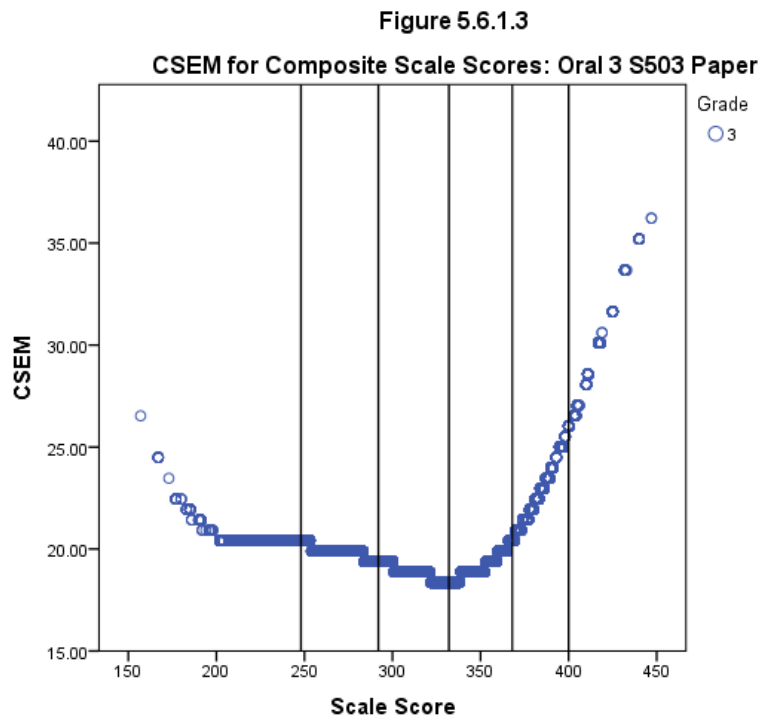
### 5.6.1.1 Grade 1



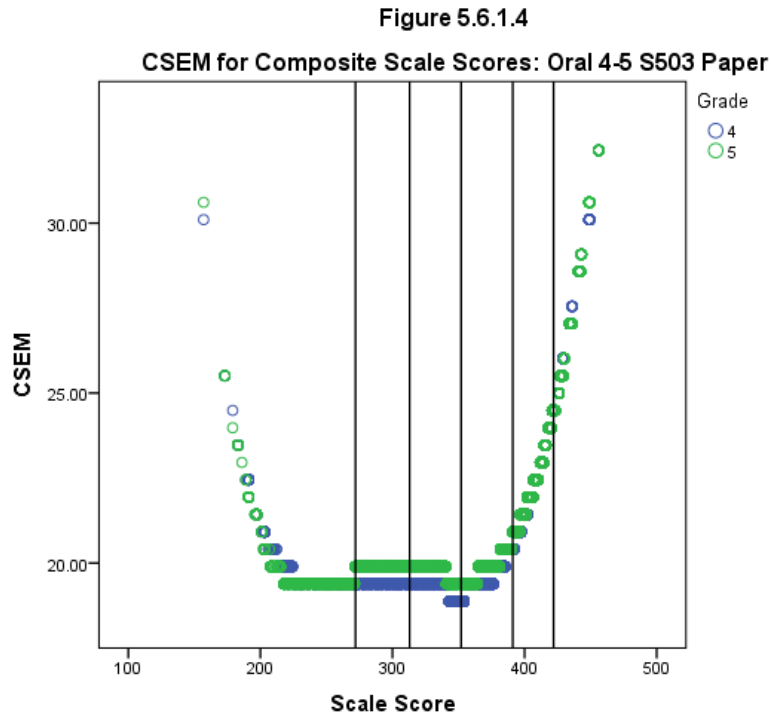
### 5.6.1.2 Grade 2



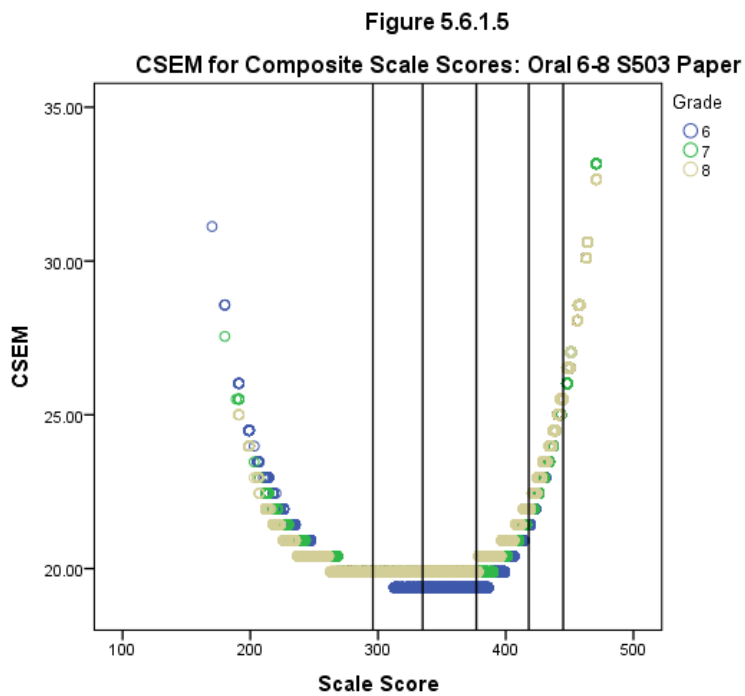
5.6.1.3 Grade 3



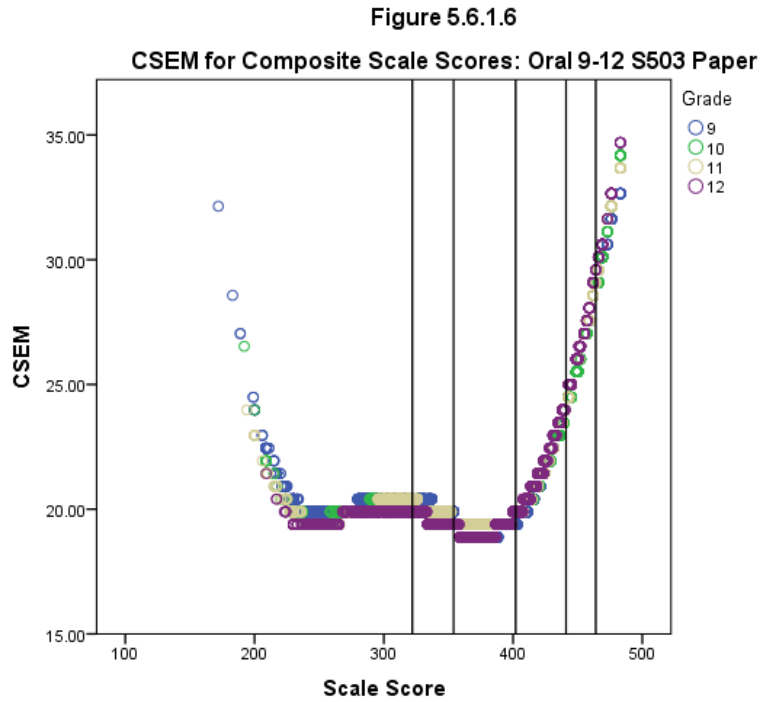
#### 5.6.1.4 Grades 4-5



#### 5.6.1.5 Grades 6-8

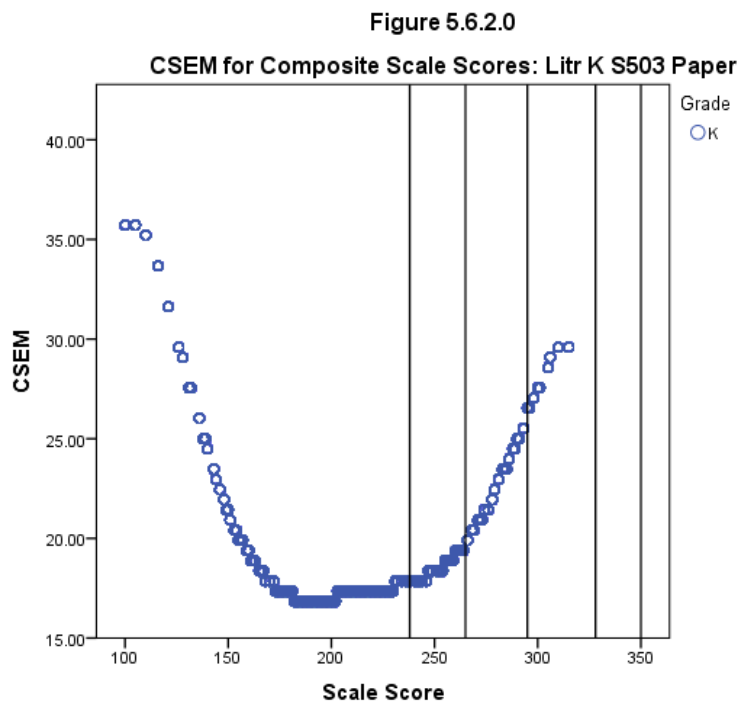


5.6.1.6 Grades 9-12

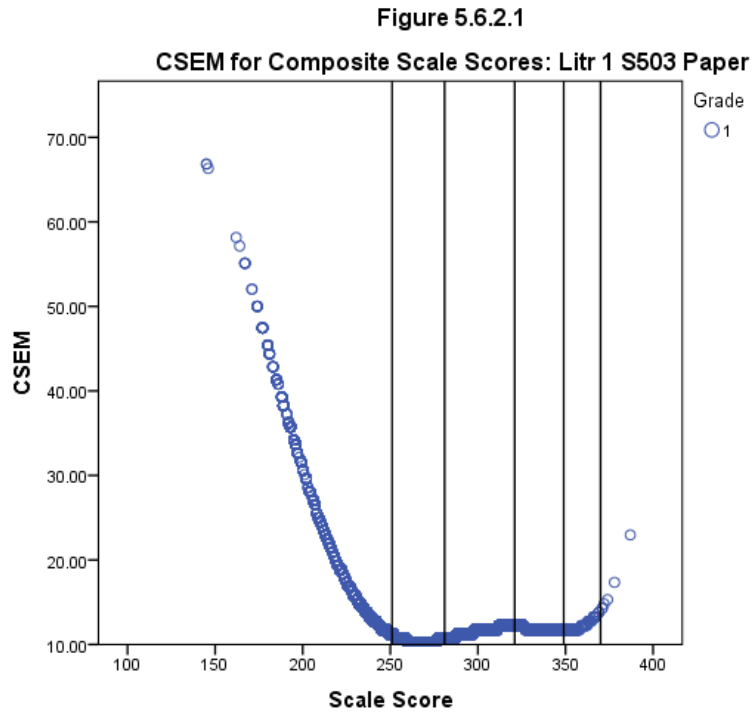


## 5.6.2 Literacy

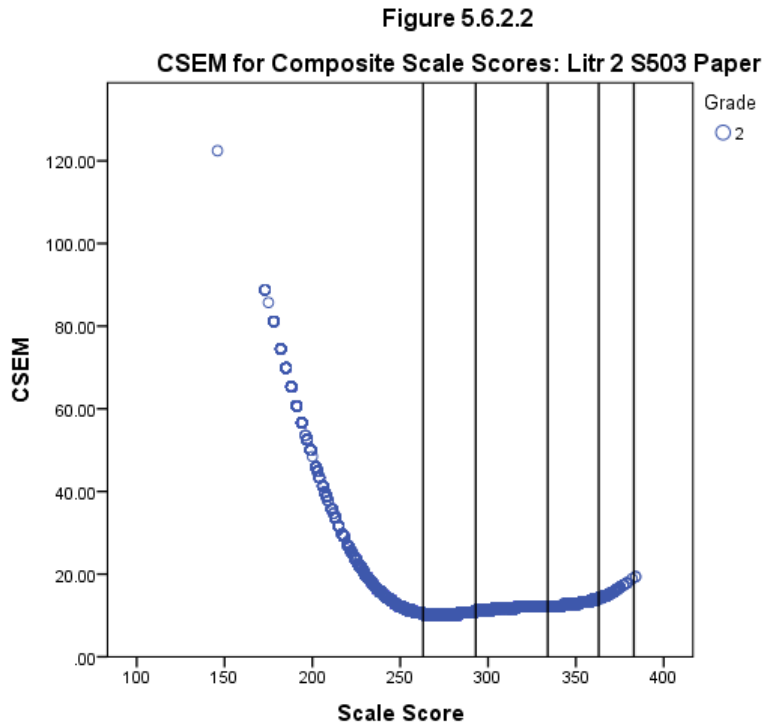
### 5.6.2.0 Kindergarten



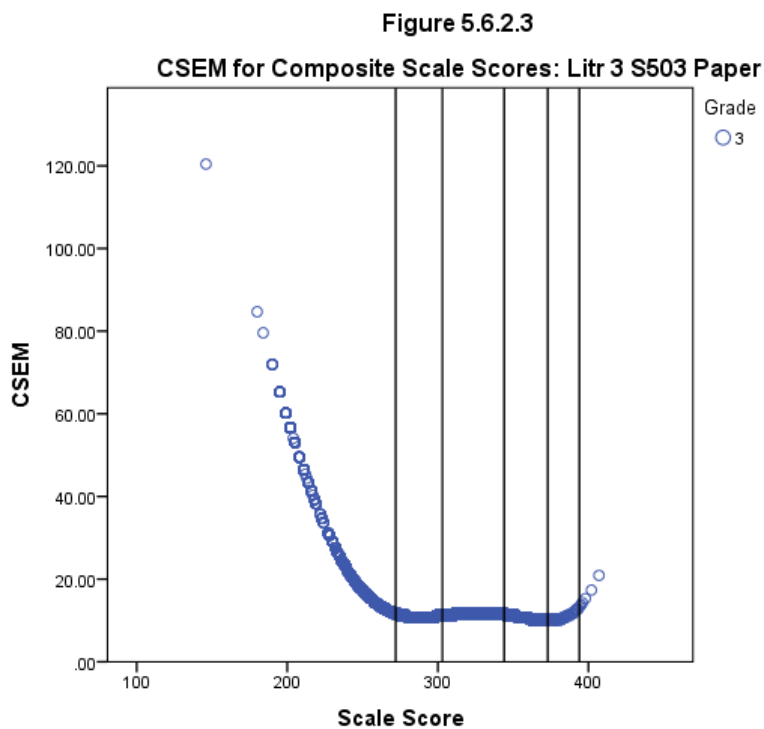
5.6.2.1 Grade 1



### 5.6.2.2 Grade 2

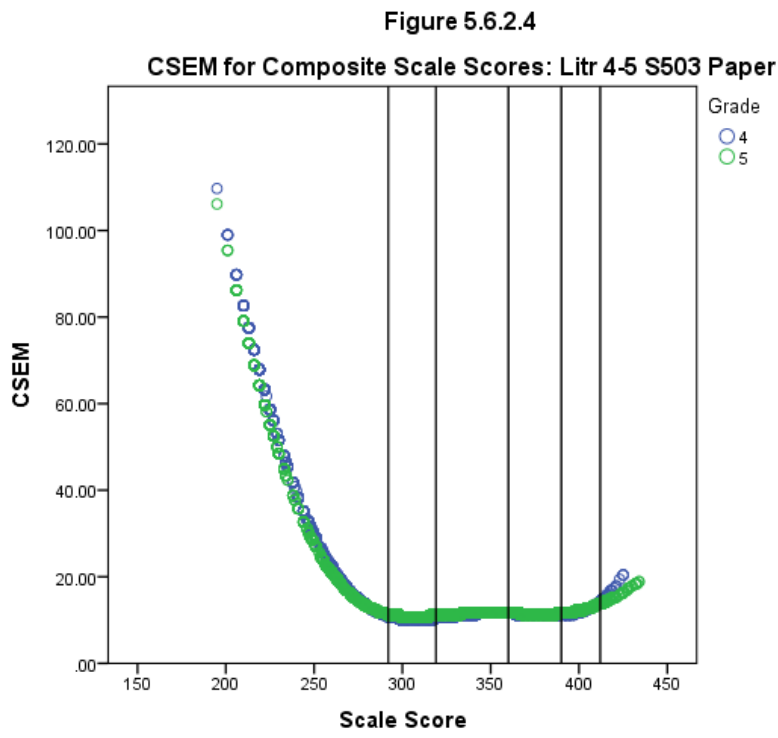


### 5.6.2.3 Grade 3

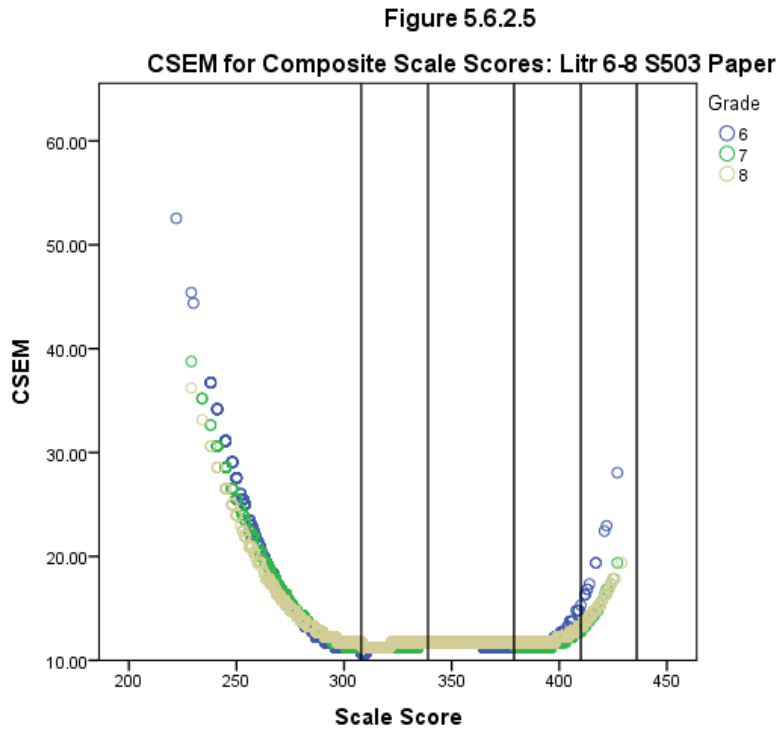




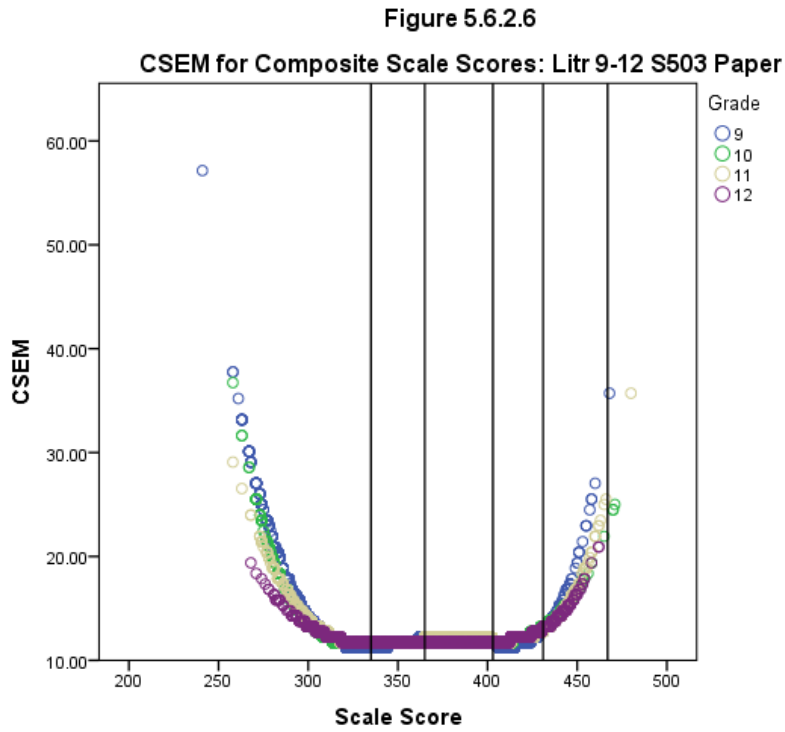
5.6.2.4 Grades 4-5



5.6.2.5 Grades 6-8

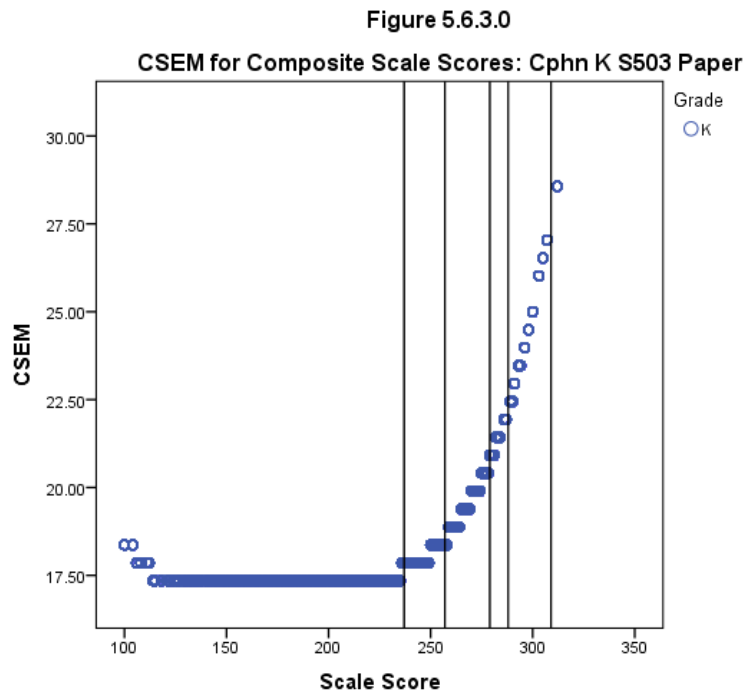


5.6.2.6 Grades 9-12

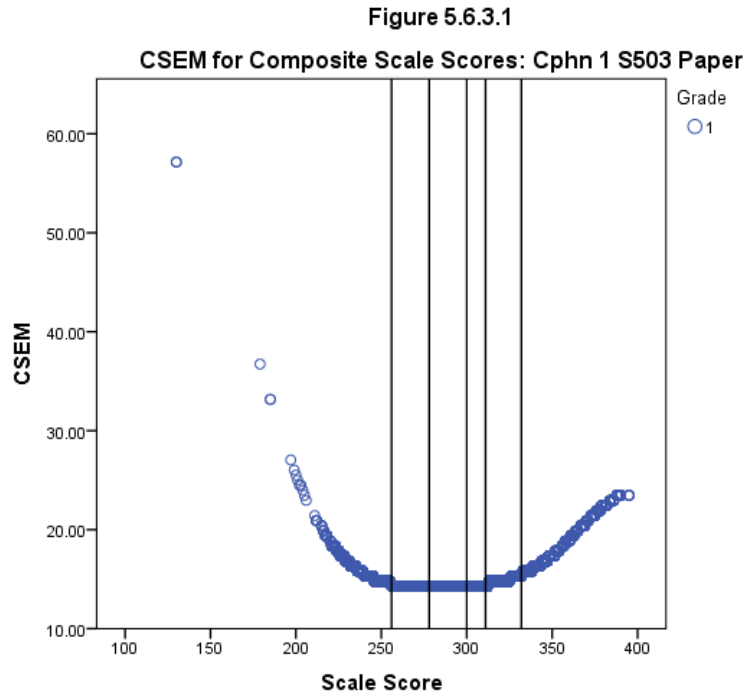


## 5.6.3 Comprehension

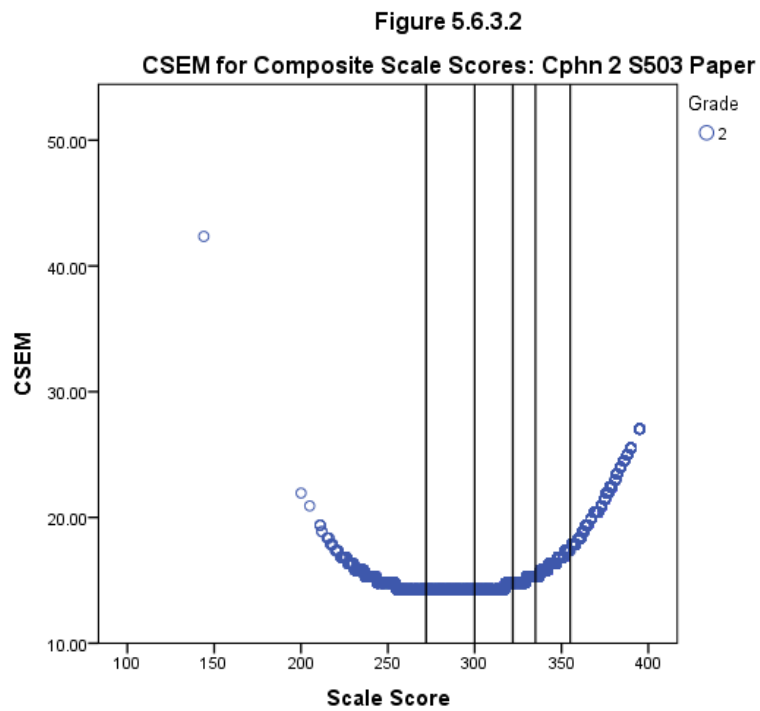
### 5.6.3.0 Kindergarten



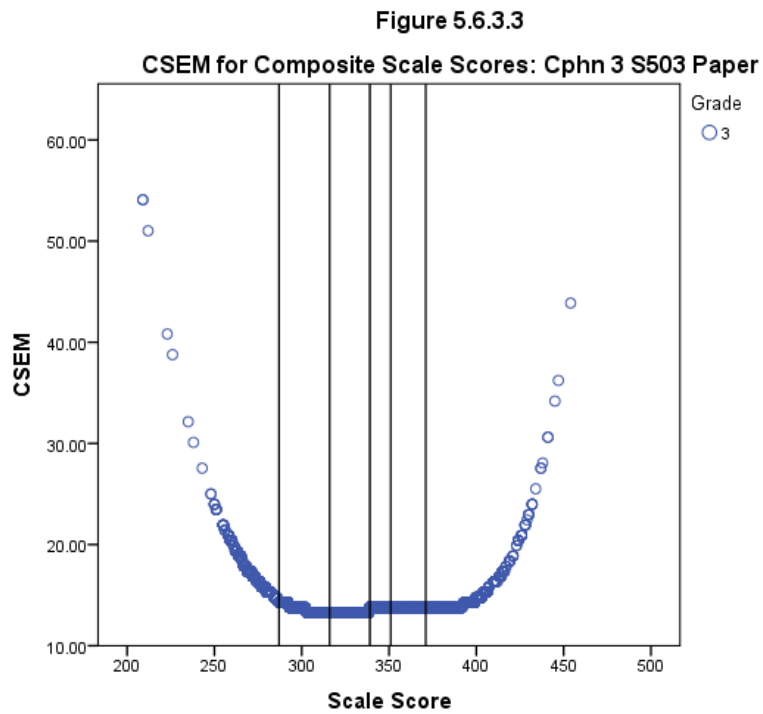
### 5.6.3.1 Grade 1



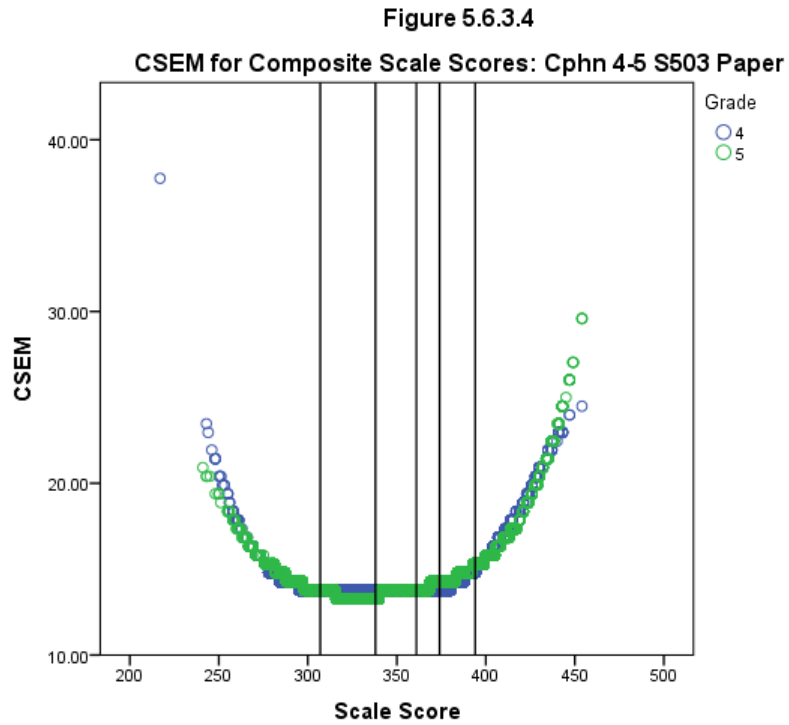
### 5.6.3.2 Grade 2



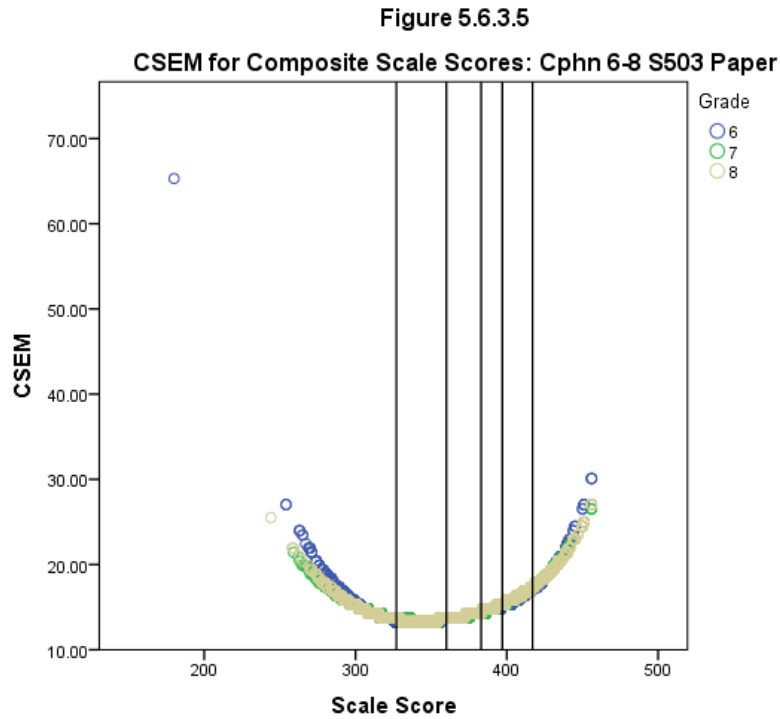
### 5.6.3.3 Grade 3



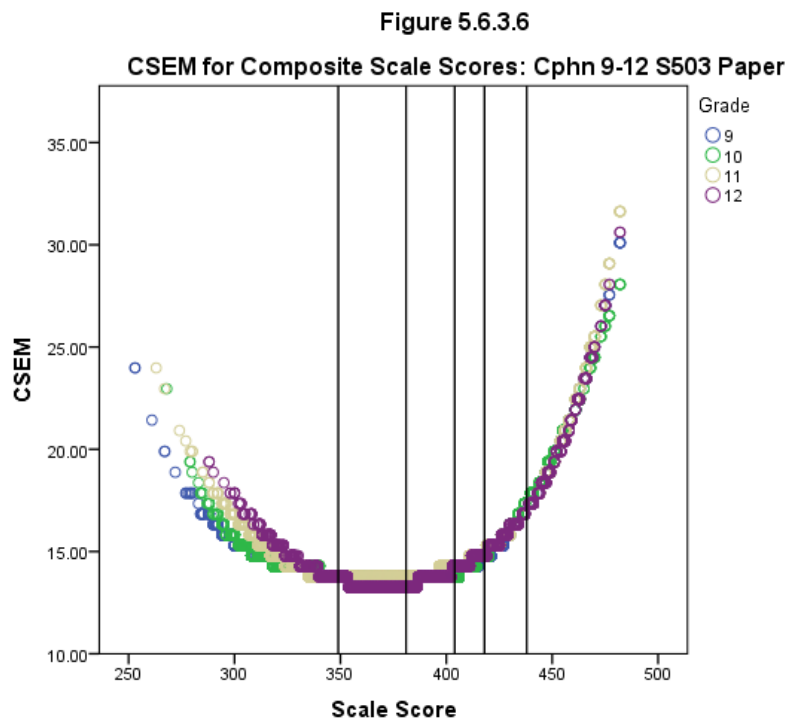
5.6.3.4 Grades 4-5



### 5.6.3.5 Grades 6-8



### 5.6.3.6 Grades 9-12

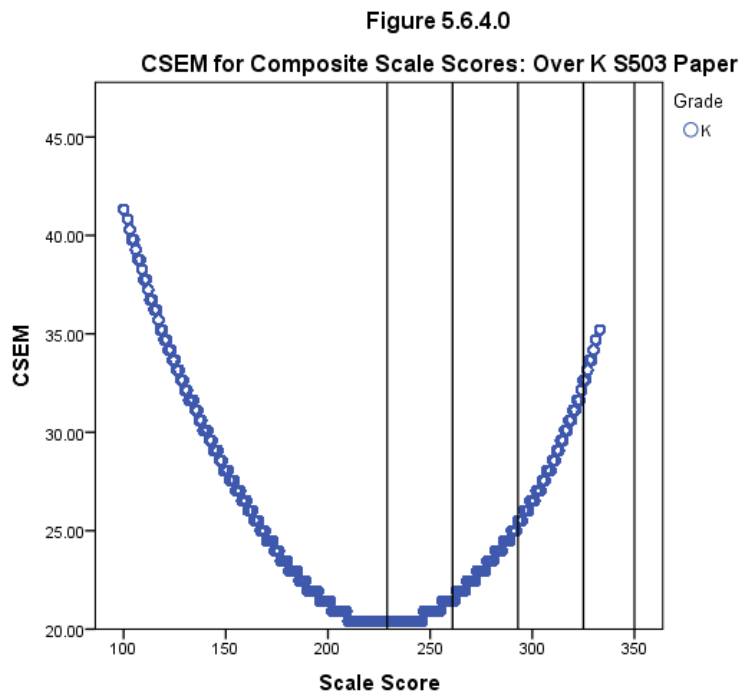




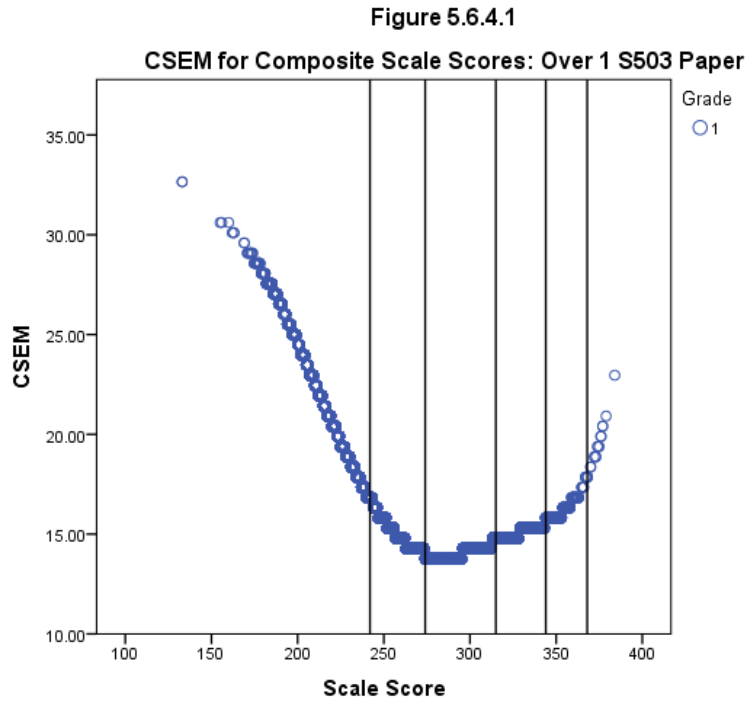


## 5.6.4 Overall

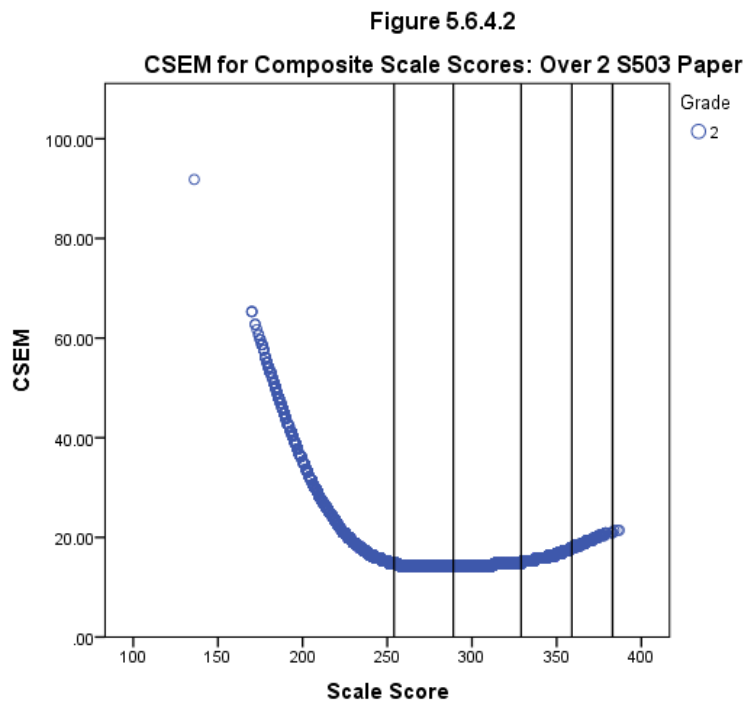
### 5.6.4.0 Kindergarten



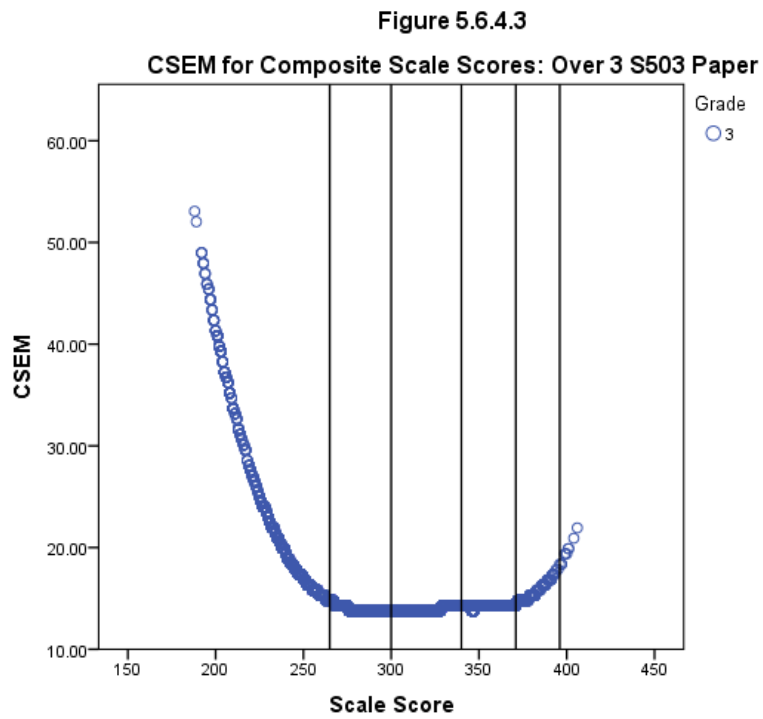
#### 5.6.4.1 Grade 1



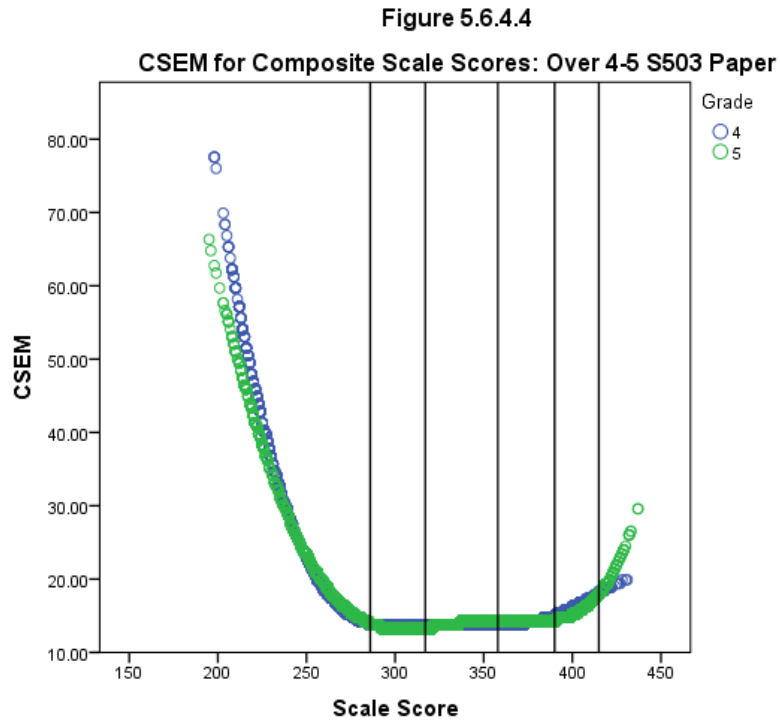
#### 5.6.4.2 Grade 2



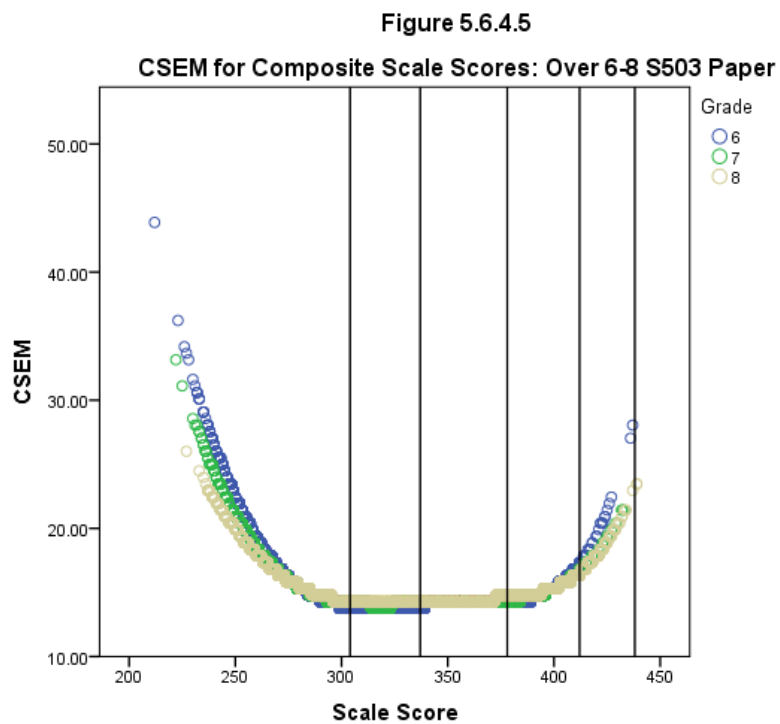
### 5.6.4.3 Grade 3



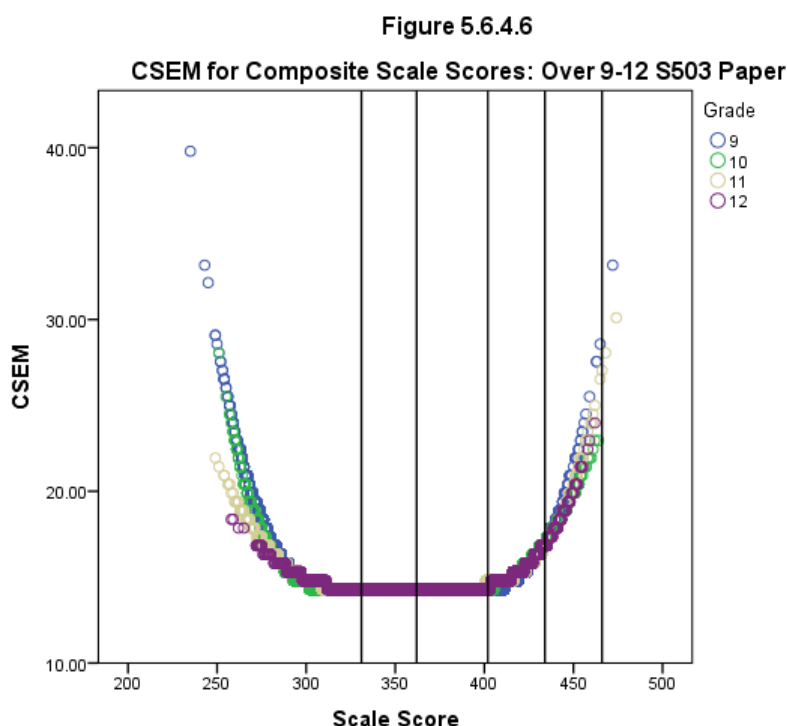
#### 5.6.4.4 Grades 4-5



#### 5.6.4.5 Grades 6-8



#### 5.6.4.6 Grades 9-12



### 5.7 Accuracy and Consistency of Composites

One of the main purposes of the WIDA ACCESS program is to identify the English language proficiency level of students with respect to the WIDA ELD Standards. Because of the emphasis on the classification of student performance, a question of interest is how accurately and consistently the ACCESS composite scale scores can classify students into WIDA proficiency categories determined by the 2016 ACCESS standard-setting process (Cook & MacGregor, 2017). Although states in the WIDA Consortium take into consideration one or more of the domain and composite scale scores when making accountability decisions, all WIDA Consortium states use the Overall composite scale score as the primary score when making classification decisions about students. Therefore, it is especially important to examine the accuracy and consistency of the classifications based on the Overall composite scale scores to help test users and policy makers judge the utility of this information and to make decisions about score reporting (American Educational Research Association et al., 2014). The analyses utilize the methods that Livingston and Lewis (1995) and Young and Yoon (1998) outlined, as implemented in the software program BB-CLASS (Brennan, 2004; cf. also Lee et al., 2002).

The method and descriptions of the classification accuracy and consistency indices reported in this section appear in detail in Section 5.4. The only substantive methodological difference between the estimation of the classification accuracy and consistency of the domain scale scores versus the composite scale scores is that to estimate the classification accuracy and consistency of the composite scale scores, we first estimate

the reliability of the composite scale scores using a stratified Cronbach's coefficient alpha, as described in Section 5.4.

For each composite, we present three tables. The first reports the overall accuracy and the overall consistency indices for each grade. The second reports the marginal classification accuracy indices based on the composite scale scores at the cut points for each grade. The third reports the marginal classification consistency indices based on the composite scale scores at the cut points for each grade.

If we could not estimate the overall and marginal classification accuracy and consistency indices because there were fewer than 200 students in the proficiency level, we collapsed the affected proficiency level with the level below it and placed 'N/A' in the table for the affected proficiency level.

As noted in Section 5.4, 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. To help test users and policy makers interpret the results from our analyses, we report for each composite the range of these indices, highlighting the grade with the lowest classification accuracy and consistency indices for that composite. Since overall accuracy and consistency indices are summaries of the degree of classification accuracy and consistency for the composite scale scores across all proficiency level cut points, we also examine the marginal classification accuracy and consistency indices for these grades to identify the specific source(s) of low classification accuracy and consistency.

For the Oral composite, as shown in Table 5.7.1.1, overall classification accuracy ranged from 0.629 to 0.731 and overall classification consistency ranged from 0.521 to 0.651 across grades. The lowest overall classification accuracy and consistency values were found for students in Grade 8.

For the Literacy composite, overall classification accuracy ranged from 0.752 to 0.890 and overall classification consistency ranged from 0.667 to 0.861 across grades, as shown in Table 5.7.2.1. The lowest overall classification accuracy and consistency values were found for students in Grade 5.

For the Comprehension composite, as shown in Table 5.7.3.1, overall classification accuracy ranged from 0.508 to 0.859 and overall classification consistency ranged from 0.400 to 0.819 across grades. The lowest overall classification accuracy and consistency values were found for students in Grade 3.

For the Overall composite, as shown in Table 5.7.4.1, overall classification accuracy ranged from 0.787 to 0.882 and overall classification consistency ranged from 0.716 to 0.840 across grades. The lowest overall classification accuracy and consistency values were found for students in Grade 5.

The results reveal that Grade 5 had the lowest overall classification accuracy and consistency indices for two out of the four composites (Literacy and Overall), while Grade 8 had the lowest overall classification accuracy and consistency indices for the Oral composite and Grade 3 had the lowest overall classification accuracy and consistency indices for the Comprehension composite.

From an accountability perspective, the most important indices for test users and policy makers to examine are the marginal classification accuracy and consistency indices. We report for each composite the range of the marginal classification accuracy and consistency indices for the composite scale scores across grades and then highlight the grade (and the cut point within that grade) that had the lowest marginal classification accuracy and the lowest consistency indices.

For the Oral composite, classification accuracy indices at the cut ranged from 0.875 to 0.976 (Table 5.7.1.2) and classification consistency at the cut ranged from 0.822 to 0.968 (Table 5.7.1.3). The lowest marginal classification accuracy and consistency values were found for students in Grade 5 at the PL 4/PL

5 cut. Grade 5, at the PL 4/5 cut point, had the lowest marginal classification accuracy and consistency indices. However, it should be noted that the marginal classification accuracy and consistency for the Grade 5 Oral composite are still in the .80's and 90's.

For the Literacy composite, classification accuracy indices at the cut ranged from 0.868 to 0.992 (Table 5.7.2.2) and classification consistency at the cut ranged from 0.816 to 0.993 (Table 5.7.2.3). Grade 4, at the PL 3/4 cut point, had the lowest marginal classification accuracy and consistency indices. However, it should be noted that the marginal classification accuracy and consistency for the Grade 4 Literacy composite are still in the .80's and .90's.

For the Comprehension composite, classification accuracy indices at the cut ranged from 0.819 to 0.986 (Table 5.7.3.2) and classification consistency at the cut ranged from 0.757 to 0.980 (Table 5.7.3.3). The lowest marginal classification accuracy and consistency values were found for students in Grade 3 at the PL 4/PL 5 cut. Note that Grade 3 was also identified as having the lowest overall classification accuracy and consistency in the Comprehension composite. The low marginal classification accuracy and consistency at the PL 4/PL 5 cut appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal classification accuracy and consistency for the Grade 3 Comprehension composite are still in the high .70's and .90's.

For the Overall composite, classification accuracy indices at the cut ranged from 0.876 to 0.992 (Table 5.7.4.2) and classification consistency at the cut ranged from 0.829 to 0.991 (Table 5.7.4.3). The lowest marginal classification accuracy and consistency values were found for students in Grade 3 at the PL 3/PL 4 cut. Note that Grade 3 was also identified as having the lowest marginal classification accuracy and consistency in the Comprehension composite. The low marginal classification accuracy and consistency at the PL 3/PL 4 cut appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal classification accuracy and consistency for the Grade 3 Overall composite are still in the .80's and .90's.

When we compared the overall and marginal classification accuracy and consistency indices for the composites for a particular grade, we saw that in many instances they told the same story (i.e., for a given grade, if the overall classification accuracy and consistency indices were low, then the marginal classification accuracy and consistency indices also tended to be low). This was especially true for Grade 3 for two of the four composites (Comprehension and Overall). Grade 3 had the lowest overall and marginal classification accuracy and consistency indices for the Comprehension composites.

Grade 3 had the lowest marginal classification accuracy and consistency in two of the four composites (Comprehension and Overall). Grade 4 had the lowest marginal classification accuracy and consistency in the Literacy composite. Grade 5 had the lowest overall and marginal classification accuracy and consistency in the Oral composite.

In addition, the lowest marginal classification accuracy and consistency of the composites occurred at the PL 3/PL 4 and PL 4/PL 5 cut points. This finding is consistent with previous research (Lee et al., 2002), in that classification accuracy and consistency at cut points in the middle of the proficiency level range are lower than those at the lower and upper ends.

A higher number of proficiency levels typically results in cut scores that are closer to each other than if a smaller number of proficiency levels is used. Classification accuracy and consistency are expected to vary for different ability levels due to variation in measurement accuracy. The further away the scores are from the cut scores, the smaller the classification errors would be or the more accurate the classification



decisions would be. When there is a large number of proficiency levels, more students are near the cut scores than there would be if there were fewer proficiency levels. Therefore, the higher the number of proficiency levels, the higher the probability that students are misclassified (Ercikan & Julian, 2002). Since ACCESS has six proficiency levels and PL 3 and PL 4 occupy relatively narrow ranges on the ability scale compared with other proficiency levels, the classification accuracy and consistency for the 3/4 and 4/5 cuts are lower than for other cuts.

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 that report composite scale scores. From an accountability perspective, the most important indices are the marginal classification accuracy and consistency indices. The marginal classification accuracy and consistency indices were at or above 0.800 for all four composites. Additionally, the marginal classification accuracy and consistency indices were at or above 0.838 for the Overall composite scale score, which is the primary score that WIDA Consortium states use when making accountability decisions.

## 5.7.1 Oral

**Table 5.7.1.1**

Overall Accuracy and Consistency of Classification Indices: Oral S503 Paper

Grade	Accuracy	Consistency
K	0.731	0.651
1	0.683	0.574
2	0.679	0.569
3	0.659	0.548
4	0.636	0.528
5	0.639	0.527
6	0.631	0.523
7	0.632	0.524
8	0.629	0.521
9	0.674	0.570
10	0.68	0.574
11	0.680	0.577
12	0.712	0.609

**Table 5.7.1.2**

Marginal Classification Accuracy Indices Based on the Composite Scale Scores at the Cut Points: Oral S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.946	0.942	0.948	0.943	0.941
1	0.961	0.919	0.896	0.934	0.972
2	0.971	0.932	0.887	0.916	0.971
3	0.976	0.942	0.884	0.889	0.964
4	0.974	0.950	0.900	0.881	0.928
5	0.971	0.947	0.897	0.875	0.944
6	0.962	0.945	0.903	0.883	0.932
7	0.956	0.939	0.900	0.885	0.944
8	0.955	0.940	0.901	0.881	0.943
9	0.939	0.924	0.905	0.929	0.971
10	0.943	0.922	0.902	0.935	0.973
11	0.941	0.915	0.905	0.941	0.972
12	0.950	0.919	0.905	0.936	N/A

**Table 5.7.1.3**

Marginal Classification Consistency Indices Based on the Composite Scale Scores at the Cut Points: Oral S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.924	0.920	0.925	0.920	0.916
1	0.945	0.885	0.855	0.904	0.967
2	0.959	0.901	0.844	0.877	0.968
3	0.967	0.914	0.840	0.841	0.958
4	0.964	0.926	0.860	0.831	0.912
5	0.960	0.922	0.858	0.822	0.927
6	0.947	0.920	0.864	0.835	0.911
7	0.939	0.911	0.860	0.837	0.924
8	0.936	0.912	0.862	0.833	0.921
9	0.913	0.891	0.868	0.897	0.960
10	0.919	0.889	0.864	0.905	0.965
11	0.917	0.880	0.867	0.915	0.967
12	0.928	0.886	0.866	0.914	N/A

## 5.7.2 Literacy

**Table 5.7.2.1**

Overall Accuracy and Consistency of Classification Indices: Litr S503 Paper

Grade	Accuracy	Consistency
K	0.890	0.861
1	0.799	0.719
2	0.813	0.739
3	0.784	0.704
4	0.759	0.678
5	0.752	0.667
6	0.802	0.722
7	0.798	0.717
8	0.789	0.705
9	0.762	0.669
10	0.767	0.675
11	0.765	0.672
12	0.782	0.696

**Table 5.7.2.2**

Marginal Classification Accuracy Indices Based on the Composite Scale Scores at the Cut Points: LitR S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.958	0.964	0.967	N/A	N/A
1	0.915	0.910	0.975	N/A	N/A
2	0.951	0.923	0.946	0.992	N/A
3	0.968	0.929	0.887	N/A	N/A
4	0.974	0.951	0.868	0.966	N/A
5	0.975	0.951	0.884	0.942	N/A
6	0.963	0.932	0.908	N/A	N/A
7	0.959	0.929	0.911	N/A	N/A
8	0.956	0.928	0.905	N/A	N/A
9	0.953	0.931	0.920	0.959	N/A
10	0.961	0.932	0.916	0.959	N/A
11	0.960	0.930	0.916	0.960	N/A
12	0.959	0.926	0.924	0.974	N/A

**Table 5.7.2.3**

Marginal Classification Consistency Indices Based on the Composite Scale Scores at the Cut Points: LitR S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.941	0.947	0.965	N/A	N/A
1	0.880	0.873	0.964	N/A	N/A
2	0.930	0.892	0.924	0.993	N/A
3	0.955	0.899	0.849	N/A	N/A
4	0.963	0.928	0.816	0.962	N/A
5	0.964	0.928	0.838	0.930	N/A
6	0.948	0.902	0.871	N/A	N/A
7	0.944	0.898	0.875	N/A	N/A
8	0.938	0.897	0.868	N/A	N/A
9	0.934	0.902	0.887	0.942	N/A
10	0.945	0.903	0.881	0.942	N/A
11	0.943	0.900	0.882	0.943	N/A
12	0.942	0.895	0.892	0.965	N/A

### 5.7.3 Comprehension

**Table 5.7.3.1**

Overall Accuracy and Consistency of Classification Indices: Cphn S503 Paper

Grade	Accuracy	Consistency
K	0.859	0.819
1	0.559	0.448
2	0.602	0.493
3	0.508	0.400
4	0.553	0.442
5	0.541	0.433
6	0.580	0.469
7	0.567	0.461
8	0.561	0.453
9	0.596	0.491
10	0.597	0.490
11	0.598	0.491
12	0.613	0.501

**Table 5.7.3.2**

Marginal Classification Accuracy Indices Based on the Composite Scale Scores at the Cut Points: Cphn S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.959	0.966	0.968	0.970	0.986
1	0.915	0.870	0.871	0.905	0.959
2	0.944	0.897	0.885	0.906	0.943
3	0.972	0.926	0.844	0.819	0.891
4	0.970	0.936	0.866	0.848	0.896
5	0.966	0.930	0.863	0.852	0.884
6	0.948	0.898	0.861	0.889	0.955
7	0.944	0.893	0.865	0.889	0.940
8	0.936	0.894	0.869	0.885	0.937
9	0.936	0.899	0.883	0.899	0.946
10	0.946	0.896	0.882	0.902	0.940
11	0.941	0.895	0.885	0.898	0.945
12	0.949	0.891	0.883	0.911	0.957

**Table 5.7.3.3**

Marginal Classification Consistency Indices Based on the Composite Scale Scores at the Cut Points: Cphn S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.941	0.950	0.956	0.958	0.980
1	0.883	0.817	0.824	0.865	0.940
2	0.921	0.855	0.841	0.866	0.919
3	0.963	0.889	0.784	0.757	0.844
4	0.960	0.903	0.816	0.793	0.852
5	0.954	0.895	0.813	0.795	0.840
6	0.930	0.854	0.812	0.844	0.934
7	0.923	0.848	0.815	0.843	0.917
8	0.912	0.849	0.820	0.839	0.910
9	0.913	0.857	0.840	0.860	0.921
10	0.924	0.853	0.838	0.862	0.915
11	0.918	0.852	0.840	0.860	0.919
12	0.927	0.847	0.839	0.873	0.938

## 5.7.4 Overall

**Table 5.7.4.1**

Overall Accuracy and Consistency of Classification Indices: Over S503 Paper

Grade	Accuracy	Consistency
K	0.882	0.840
1	0.828	0.759
2	0.834	0.771
3	0.797	0.724
4	0.804	0.738
5	0.787	0.716
6	0.832	0.768
7	0.840	0.775
8	0.823	0.755
9	0.801	0.722
10	0.808	0.732
11	0.809	0.733
12	0.849	0.787

**Table 5.7.4.2**

Marginal Classification Accuracy Indices Based on the Composite Scale Scores at the Cut Points: Over S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.957	0.964	0.974	0.988	N/A
1	0.950	0.922	0.965	0.992	N/A
2	0.971	0.941	0.940	0.982	N/A
3	0.981	0.954	0.876	0.986	N/A
4	0.983	0.966	0.917	0.939	N/A
5	0.982	0.966	0.921	0.918	N/A
6	0.975	0.957	0.918	0.983	N/A
7	0.972	0.954	0.915	N/A	N/A
8	0.970	0.954	0.919	0.981	N/A
9	0.964	0.949	0.933	0.955	N/A
10	0.968	0.949	0.933	0.958	N/A
11	0.968	0.948	0.932	0.962	N/A
12	0.970	0.945	0.934	N/A	N/A

**Table 5.7.4.3**

Marginal Classification Consistency Indices Based on the Composite Scale Scores at the Cut Points: Over S503 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.939	0.949	0.963	0.988	N/A
1	0.929	0.890	0.949	0.991	N/A
2	0.959	0.917	0.914	0.982	N/A
3	0.974	0.934	0.829	0.986	N/A
4	0.976	0.951	0.884	0.927	N/A
5	0.976	0.950	0.889	0.900	N/A
6	0.965	0.938	0.884	0.980	N/A
7	0.961	0.934	0.881	N/A	N/A
8	0.958	0.934	0.885	0.978	N/A
9	0.950	0.928	0.906	0.938	N/A
10	0.956	0.928	0.905	0.943	N/A
11	0.955	0.926	0.904	0.948	N/A
12	0.958	0.922	0.907	N/A	N/A

## 6 Quality Control

### 6.1 Content Development Quality Control

The Center for Applied Linguistics (CAL) utilizes educators and other consultants at a number of phases throughout the test-development cycle. These educators and consultants are recruited, vetted, and trained by CAL and/or WIDA and make crucial contributions to these phases of the test development cycle. The phases of development in which educators or consultants are involved, as well as the procedures and criteria for recruitment and training, are described below.

#### *Theme Generation*

During theme generation, CAL and WIDA recruit educators to generate raw ideas to be used in new item development. Educators with ESL or content-area expertise and two or more years of teaching experience in a WIDA state (in the grade cluster for which they will generate themes) are invited to participate. Recruitment also focuses on a geographical distribution of educators from across the consortium. Upon selection, educators participate in a short training that introduces the theme-generation process, along with how to understand the item specifications that they use to generate themes.

#### *Item Writing*

CAL recruits professional item writers to generate raw item/task content based on the ideas from theme generation. To recruit item writers, CAL has a standing announcement on its website asking prospective item writers to submit their resume and fill out a survey describing their past item-writing experience. CAL selects individuals with significant experience in writing items, both in large-scale assessment programs (ESL/EFL or ELA) and in other contexts (e.g., writing items for assessment-programs in university-based ESL programs).

Item writers undergo a 90-minute orientation prior to beginning item writing. This training focuses on the item specifications, the process and procedures, the item writing checklist, the acceptance criteria for the items, and the security protocols. Item writers also receive an item writing handbook, which formalizes the content of the orientation, along with assignment of themes to develop and the associated item specifications. After the orientation, CAL Language Testing Specialists and managers provide feedback to the item writers on the items, focusing on alignment with the item writing checklist and the item specifications. After completion of item writing for a given development cycle, item writers are evaluated by CAL staff for their compliance with the requirements and the quality of their items.

#### *Standards Expert Review*

After items have been drafted by item writers, CAL Language Testing Specialists review all of the raw content internally. This review focuses on determining which sets of items will move on to further development and which will be discontinued, based on criteria from an item review checklist. The Language Testing Specialists then do minor editing and formatting to the items to make sure that they are complete, with no stray comments or other editorial notes from previous drafts, and they produce a short questionnaire for each set of items that becomes part of Standards Expert review. The purpose of Standards Expert review is to ensure that the items are appropriate for the grade-level and intended



difficulty level in terms of both the content and the language, and the items have not drifted from their intended target between theme generation and item writing. The questionnaires produced by CAL's Language Testing Specialists guide the Standards Experts through the review process, asking questions specific to the purpose of this review.

Educators are recruited jointly by CAL and WIDA to serve as Standards Experts; educators with ESL or content-area expertise and two or more years of teaching experience in a WIDA state are invited to participate. Recruitment also focuses on a geographical distribution of educators from across the consortium. Standards Experts receive written instructions and a questionnaire to complete for each set of items they review.

### *Bias and Sensitivity and Content Review*

After Standards Expert Review has been completed, all items undergo an additional phase of review and revision internal to CAL, leading up to Bias & Sensitivity and Content Review. These are technically two separate reviews, although a single recruitment effort is conducted by WIDA, and the reviews occur consecutively in a single week (generally 3 days for Content review followed by 2 days for Bias & Sensitivity review). As with other reviews, educators for Content review must have at least 2 years of ESL teaching experience (with a preference for content-area experience as well). Recruitment also focuses on selecting educators with a variety of cultural and linguistic backgrounds and obtaining a geographical distribution of educators from across the consortium. Recruitment for Bias & Sensitivity review focuses on selecting educators with culturally and linguistically diverse backgrounds who have experience interacting with English learners from a range of cultural, regional, religious, linguistic, ethnic, and socioeconomic backgrounds.

At the beginning of both Bias & Sensitivity and Content review meetings, CAL and WIDA staff conduct an intensive training to orient the reviewers to the specific purpose of the review (Bias & Sensitivity or Content), how to use the review checklist and what to look for in the review, and the procedures and security protocols for the review. Then, the reviews are conducted in breakout groups by grade cluster (or combinations of grade clusters; for example, Bias & Sensitivity review of Grade 1 and Grades 2–3 is often combined). Although Bias & Sensitivity and Content reviews are generally held in -person, the reviews for the Writing domain occur virtually each year due to timeline constraints. For both the in-person and virtual contexts, CAL and WIDA facilitators are present in each breakout group to guide the educators in their reviews of the materials.

### *Writing Tryouts*

For the Writing domain, all tasks in the Writing domain are subject to tryouts in the field. The Writing tryouts only occur once the tasks have been through a thorough Bias & Sensitivity and Content review and subsequent revision. CAL and WIDA recruit educators who are willing to administer the Writing tasks to their students; these educators are classroom ESL or content teachers who work with ELLs. All students who participate are required to have parent/guardian consent.

Once the students complete the Writing tasks, both the students and educators fill out questionnaires. Student questionnaires focus on whether the students understood the task, their engagement with the task, and their ability to complete the task; educator surveys ask the teachers to evaluate the effectiveness of

the task input, the appropriateness of the task, the comparability of the task with other classroom-based writing tasks, and the ability of the students to complete the task.

CAL provides the teachers with a number of documents outlining the procedures for administering the tasks, recording student responses to the tasks, recording student and teacher responses to the questionnaires, and protecting the personally identifiable information of the students. CAL staff are also available throughout the tryouts process to answer any questions the teachers might have. Following the Writing tryouts, CAL specialists review the writing responses both qualitatively and quantitatively, providing WIDA with a report on how the Writing tasks performed.

## ***6.2 Test Administration Quality Control***

This section describes how WIDA monitors test administration to ensure standardized test administration procedures are implemented with fidelity across districts and schools. To support standardized administrations, WIDA provides test administrators with a series of resources, such as a Test Administration Manual, a training course, and a Test Administration Script for each assessment.

### *Qualifications of Test Administrators*

Before, during, and after a state's testing window, educators hold 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. The Test Administrator administers and monitors the test. They are also responsible for managing student data prior to, during, and after testing.

WIDA has worked directly with each state education agency to develop the ACCESS for ELLs Checklist for the school year. This list highlights all tasks that need to be completed before, during, and after testing within a school or district and outlines which tasks are assigned to Test Coordinators at the district and school level and to Test Administrators. It also provides additional guidance that a state expects test administrators to follow as they prepare for and administer the ACCESS for ELLs suite of assessments.

Test administrators are responsible for reviewing each state's checklist in detail prior to completing any training and for working with the district or school Test Coordinator to complete these tasks. The state's checklist can be found in the training course and on each state's WIDA webpage at [www.wida.us/membership/states](http://www.wida.us/membership/states).

The training course within the WIDA Secure Portal (<https://www.grow.wida.us/>) is where educators can access both training to become certified to administer ACCESS for ELLs as well as additional materials and resources to assist administrators and coordinators before, during, and after each state's testing window. WIDA user accounts provide access to the training course and Facilitator Toolkit within the WIDA Secure Portal. Educators must pass an administration quiz at the end of the training with a score of 80% or higher. WIDA recommends taking the quiz immediately after completing the training. There is no limit to the number of times educators can attempt the quiz. Once individuals pass an administration quiz, training certificates within the WIDA Secure Portal are updated to reflect their status as a certified test administrator for that component of the assessment suite.

### *Paper Testing (for Writing Grades 1–3)*

Depending on state, district, and school policy, not all Test Administrators will be responsible for initially labeling and/or bubbling booklets. However, it is the responsibility of all Test Administrators and Test Coordinators to ensure that correct and complete information is either labeled or bubbled in each student booklet. Each state's ACCESS for ELLs checklist has more information on who is responsible for each task related to materials management in the state.

To ensure all booklets have the detailed and necessary information needed to score, all Test Administrators must adhere to the following:

- Prior to administration
  - Review labels and/or bubbled information to ensure all student information is accurate.
  - Complete labeling or bubbling if needed.
- During administration
  - Distribute the test booklets, as applicable, to the correct students.
  - Verify that students have been given their assigned booklet.
- Immediately following administration
  - Collect all material from all students.
  - Review student test booklets once more for any errors or discrepancies in student information.
  - Confirm all necessary fields are completed and all necessary labels are correctly adhered to student test booklets.
  - Ensure all booklets are in proper condition to be returned, with no loose or damaged pages.
  - Return test materials to a Test Coordinator or store the booklets in a secure area until they can be handed over to a Test Coordinator.

Failure to address incorrect, missing, or incomplete booklet information and labels may result in late reporting or no student score. In addition, the WIDA Consortium's national research agenda relies on complete and accurate student demographic data to inform the field and benefit English language learners.

When preparing test materials for return to DRC, test administrators need to confirm that any booklet that contains student response information has either a Pre-ID Label or a District/School Label with bubbled student information. If a booklet is unused, there is no need to place any labels on the booklet. Placing a label on a booklet will cause it to be processed (and either scored, if the label is a Pre-ID or School/District label, or not scored, if it is a Do Not Process label).

### ***6.3 Rater Quality Control***

#### *Rater Training*

Students who take the ACCESS for ELLs Paper Speaking test have their spoken responses scored by the Test Administrator who administered the Speaking test. Another term for this Test Administrator is *rater*. Raters must be trained and certified, so we can be confident that they interpret students' spoken language consistently and fairly, and that the scores are reported according to the WIDA English language proficiency standards. WIDA provides several different types of resources to support raters' training and reliability.

Students who take ACCESS for ELLs Paper have their spoken responses scored in real time by the Test Administrator who administers the Speaking test. It is important that the individual who scores the spoken responses is trained and certified.

WIDA provides a series of training modules in the Secure Portal on the WIDA website. ACCESS for ELLs Speaking test raters should complete three core modules:

1. Overview and Test Structure
2. Speaking Assessment Scoring Practice
3. Speaking Assessment Recommended Practice

WIDA strongly recommends that all new raters complete all three of these modules. These modules provide a comprehensive introduction to the ACCESS for ELLs Speaking test and the opportunity to learn how to score students' spoken English reliably using the ACCESS for ELLs Speaking Scoring Scale.

In addition to the modules described above, WIDA also releases supplemental training materials each year to refamiliarize experienced raters with the Speaking Scoring Scale and introduce new Speaking tasks and sample responses for the coming year. These materials, called Supplemental Training for the Speaking Assessment, reflect the Speaking tasks that will appear on the test in the current year. WIDA recommends that all raters (new and experienced) engage with these supplementary materials at the start of each scoring season. Reading and reviewing these materials will help raters maintain their reliability from year to year and contribute to the fairness of test scores awarded to all students.

#### *Rater Certification*

After completing the training modules described in the section above, new raters should take the relevant certification quiz. WIDA provides two quizzes: one for raters who will evaluate students in Grades 1–5 and another for raters who will evaluate students in Grades 6–12. Raters should take the appropriate quiz.

The purpose of the quiz is to ensure that raters have internalized the Speaking Scoring Scale and can apply it consistently. Only raters who pass the quiz(zes) should administer and score the ACCESS for ELLs Paper Speaking test.

#### *Checklist for Rater Training, Monitoring, and Recertification*

- ☐ ☐ New raters complete all Speaking assessment training
- ☐ ☐ New raters take and pass the appropriate certification quizzes
- ☐ ☐ All raters recertify at the start of each testing season (review new materials, retake quiz)
- ☐ ☐ Only certified raters administer and score the ACCESS for ELLs 2.0 Speaking test

- ☐ ☐ Raters do not evaluate their own students, if at all possible
- ☐ ☐ Rater reliability and/or score point distributions are monitored regularly

For more information on Writing rater QC, please refer to section 3.2.2.

## ***6.4 Score Reporting Quality Control***

WIDA conducts an annual score reporting quality control process to (1) verify the accuracy of paper-based test scores (i.e., ACCESS for ELLs Paper, Kindergarten ACCESS for ELLs, and Alternate ACCESS) and (2) verify the accuracy of all score reports (the Individual Student Report, the Student Roster Report, the School Frequency Report, the District Frequency Report, and the State Frequency Report) for both ACCESS (Online, Paper, and Kindergarten) and Alternate ACCESS.

The Score Reporting quality control is conducted at DRC's offices in Maple Grove, Minnesota. The team generally includes five state education agency representatives, one CAL employee, and four WIDA employees. This team examines data from three districts: a primary district, for quality control of all score reports; a secondary district, for quality control of State Frequency Reports only; and a tertiary district for quality control of paper-based tests only.

After an introductory presentation, which includes details of the quality control processes undertaken by DRC and WIDA and instructions on using the data entry tools, panelists begin by confirming the scoring of ACCESS Paper. Using the information in the State Student Response file, panelists enter the grade level, grade level cluster, tier, the Listening and Reading responses, and the Speaking and Writing scores into the data entry tool. The tool then calculates the student's raw scores and, using a series of look-ups, the student's scale score, proficiency level score, and confidence bands for all domains and composites. Panelists check student scores on the Individual Student Reports against those calculations. Any discrepancies are brought to the attention of the WIDA facilitator who investigates and, if there seems to be an issue with the report (rather than the data entry or data entry tool), discusses the issue further with DRC.

The panelists follow a similar process with the Kindergarten ACCESS tests, but with the raw scores for these tests copied directly from the response booklets.

After checking the paper-based tests, panelists turn their attention to the score reports. Panelists first check both the demographic information and the student scores in the Individual Student Reports against the information in the Student Roster Reports. Again, any discrepancies are brought to the attention of the facilitator, who investigates and discusses the issue with DRC if necessary. Panelists use the verified Individual Student Reports to check the Student Roster Report. Once the Student Roster Report is verified, panelists use it to check the State Frequency Report; they then use the verified State Frequency Reports to check the District Frequency Report. Finally, panelists check the State Frequency Reports against verified District Frequency Reports from the primary district along with District Frequency Reports from the secondary district.

## 6.5 Data Forensic Quality Control

### Caveon Data Forensic Analysis Results

WIDA hired Caveon to perform data forensic analysis during the 2021–2022 test administration cycle to examine whether ACCESS data has been compromised or has evidence of item exposure.

Caveon security statistics are based on mathematical models, where the test response data are used to create a baseline model of normal or “typical” test taking among that population. Individuals or groups are then compared to the baseline, and observations that are significantly different from the baseline are flagged as anomalous. Caveon’s statistics are designed to be robust but also conservative regarding which and how many individuals or groups are flagged as anomalous, thereby reducing the chances of false-positive detections.

Data forensics analysis was performed after the administration window for the following administrations:

- December 2021 through Spring 2022 online multistage adaptive test administrations, Listening and Reading domains
- December 2021 through Spring 2022 paper fixed-form administrations, Listening and Reading domains

The analysis utilized several of Caveon’s security statistics to detect evidence of whether the assessment instrument has been compromised through disclosure of the content. This analysis attempted to understand where and when disclosure of the test content may have occurred and what items and forms may have been affected. Results of this analysis enable WIDA to take specific actions to limit the impact of disclosed content. Such actions may include

- Republishing or reworking items or forms
- Rotating disclosed items to limit their exposure
- Designing a republication or rotation strategy for future items and forms

Caveon security statistics were computed for each individual test instance. These data were aggregated or summarized at the group level. The aggregated statistics were compared against the population model.

#### *Analysis of Tests*

Caveon aggregated the data according to individual test forms using the security statistics to determine whether rates of detections by the security statistics were higher for certain test forms. For fixed-form paper tests, two forms—A and B/C—were analyzed. For the multistage adaptive test, there is a finite number of ways a student could progress through the test. Caveon analyzed each pathway as a separate form. Higher rates of security detections for a specific form of the test suggest that compromise of the form may have occurred.

#### *Analysis of Items*

*Item security:* In this portion of the analysis, the security of the items was evaluated using aberrance statistics. Aberrance statistics detect test-taking behaviors such as answering difficult items correctly but answering easy items incorrectly, or unusual patterns in the time taken to answer test items. In the absence

of security issues, aberrant test taking is expected to be the result of poor or uneven test preparation, illness or other physical malady, mental and emotional distractions, and so forth. These factors usually result in lower levels of test performance. When aberrance is associated with higher performance, however, test fraud may have occurred, such as pre-knowledge of test content. By applying aberrance measures and comparing the performance between aberrant and nonaberrant test instances on individual items, inferences can be made about item security.

*Item performance changes:* Analysis of item performance changes tracks individual item performance rates over time. The item performance shifts are measured within the context of the item response theory model and adjusted for varying test-taker performance levels. This means that detected performance shifts are invariant to fluctuations in the test-taker population. When performance shifts indicate the item has become significantly easier, the item may have been disclosed. Items with significant performance shifts become candidates for revision or replacement. Item performance shifts were detected with a granularity of 1 week, where Monday to Sunday represents 1 week.

### *Analysis of Groups*

*Analysis by week:* This analysis aggregates the data according to the week in which the test was taken to identify whether security threats and pass rates appeared to be more prevalent at certain times during the testing window. Increases in scores or security detections during certain periods of time suggest the content may have been disclosed at some point prior to that time. This analysis also includes a form-date grouping to determine if increasing security threats are associated with a particular form of the test. This analysis is performed for online and paper tests, where relevant test date data are provided.

*Analysis of WIDA jurisdictions:* Caveon analyzed WIDA member jurisdictions (states and districts) to determine whether rates of detections by the security statistics were higher for certain jurisdictions. This analysis is intended to detect whether compromise at the state or member jurisdiction level potentially occurred. This analysis is performed for online and paper tests.

*Analysis of administration mode:* Caveon aggregates the data according to administration mode (i.e., online versus paper) to determine if security threats are associated with the mode of testing.

### *Other Analyses*

*Analysis of mean score over time* was used to identify whether mean scores increased over time during the testing window. Increases in scores over time suggest the content may have been disclosed during the testing window.

### *Findings of Data Forensic Analyses*

Generally, no major data forensic anomalies were observed across WIDA states. There were some general findings and a few minor localized anomalies. States where these anomalies occurred were notified.

## References

- Allen, N. L., Carlson, J. E., & Zalanak, C. A. (1999). *The NAEP 1996 technical report*. Washington, DC: National Center for Education Statistics.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Institutes of Research. (2018). *ELPA21 technical report, part I—summative assessment*. Washington DC: Author.
- Andrich, D. A. (1978). A rating scale formulation for ordered response categories. *Psychometrika*, 43, 561–573.
- Baker, F. B., & Kim, S.-H. (2017). *The Basics of Item Response Theory Using R*. Springer International Publishing AG.
- Brennan, R. (2004). *Linking with equivalent group or single group design (LEGS)* (Version 2.0) [Computer software]. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment.
- Chapman, M., Montee, M., & Musser, S. (2022). ACCESS Speaking Educator Perceptions and Instructional Practices Study (*Unpublished WIDA Internal Report*). Madison, WI: WIDA, University of Wisconsin-Madison.
- Center for Applied Linguistics. (2016). *ACCESS for ELLs® Series 400 Listening and Reading scale maintenance: Technical brief*. Washington, DC: Author.
- Center for Applied Linguistics. (2017). *ACCESS for ELLs® 2.0 Speaking and Writing score scale reconstruction: Technical brief*. Washington, DC: Author.
- Center for Applied Linguistics. (2019). *Maintaining the ACCESS for ELLs Online Writing scale: Preparations for the Series 501 redesign: Technical brief*. Washington, DC: Author.
- Cook, H. G., & MacGregor, D. (2017). *The ACCESS for ELLs 2.0 2016 Standard-setting study* (Technical Report). Madison, WI: Board of Regents of the University of Wisconsin System.
- Crabtree, A. R. (2016). *Psychometric properties of technology-enhanced item formats: An evaluation of construct validity and technical characteristics*. Unpublished doctoral dissertation, University of Iowa, Iowa City, IA. doi:10.17077/etd.922fbj4d
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- Department of Education, (2018). *A State's Guide to the U.S. Department of Education's Assessment Peer Review Process*. U.S. Department of Education.
- Elementary and Secondary Education Act of 1965, amended 2015. 20 USC §6301-8961.
- Engelhard, G., Jr., & Wind, S. A. (2018). *Invariant measurement with raters and rating scales: Rasch models for rater-mediated assessments*. Routledge/Taylor & Francis Group.



- Ercikan, K. & Julian, M. (2002). Classification accuracy of assigning student performance to proficiency levels: Guidelines for assessment design. *Applied Measurement in Education*, 15(3), 269–294.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105–146). New York, NY: Macmillan.
- Gottlieb, M. (2004). *English language proficiency standards for English language learners in kindergarten through grade 12: Framework for large-scale state and classroom assessment*. Madison, WI: WIDA Consortium.
- Kamata, A., Turhan, A., & Darandari, E. (2003, April). *Estimating reliability for multidimensional composite scale scores*. Paper presented at the annual meeting of the 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, 221–240.
- Kenyon, D. M. (2006). *Development and field test of ACCESS for ELLs®* (WIDA Consortium Technical Report No. 1). Washington, DC: Center for Applied Linguistics.
- Kenyon, D. M., Ryu, J. R., & MacGregor, D. (2013). *Setting grade level cut scores for ACCESS for ELLs®* (WIDA Consortium Technical Report No. 4). Washington, DC: Center for Applied Linguistics.
- Kim, A., Kondo, A., Blair, A., Mancilla, L., Chapman, & M., Wilmes, C. (2016). *Interpretation and Use of K–12 Language Proficiency Assessment Score Reports: Perspectives of Educators and Parents* WCER Working Paper No. 2016-8.
- Kim, A., S., Baghastani, S., Macgregor, D., & Ho, P. (2022). Supporting K-12 Educators' language Assessment Literacy via Resources Informed by Validation (*Unpublished WIDA Internal Report*). Madison, WI: WIDA, University of Wisconsin-Madison.
- Kim, A., S., Yumsek, M., Kemp, J., Chapman, M., & Cook, H. (2022). Universal tools Activation in English Language Proficiency Assessments; A Comparison of grades 1-12 English Learner with and without Disabilities (*Unpublished WIDA Internal Report*). Madison, WI: WIDA, University of Wisconsin-Madison.
- Kolen, M. J., Hanson, B. A., & Brennan, R. L. (1992). Conditional standard errors of measurement. *Journal of Educational Measurement*, 29, 285–307.
- Lee, W., Hanson, B. A., & Brennan, R. L. (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement*, 26, 412–432.
- Linacre, J. M. (1994). Sample size and item calibrations stability. *Rasch Measurement Transactions*, 7(4), 328.

- Linacre, J. M. (1999). Relating Cronbach and Rasch reliabilities. *Rasch Measurement Transactions*, 13(2), 696. Retrieved from <http://www.rasch.org/rmt/rmt132i.htm>
- Linacre, J. M. (2002a). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Linacre, J. M. (2002b, Autumn). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16(2), 878. Retrieved from <http://www.rasch.org/rmt/rmt162f.htm>
- Linacre, J. M. (2004). Optimizing rating scale category effectiveness. In E. V. Smith Jr. & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 258–278). Maple Grove, MN: JAM Press.
- Linacre, J. M. (2006). Winsteps Rasch analysis (Version 3.60.1) [Computer software]. Retrieved from <http://www.winsteps.com>
- Linacre, J. M. (2020). *Reliability and separation of measures*. Retrieved from <https://www.winsteps.com/winman/reliability.htm>
- Linacre, J. M. (n.d.). Displacement measures. Retrieved from <http://www.winsteps.com/winman/displacement.htm>
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179–197.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- MacGregor, D., & Ma, Y. (2022). *Speaking Contrasting Group Study (Unpublished WIDA Internal Report)*. Madison, WI: WIDA, University of Wisconsin-Madison.
- Mantel, N., & Haenszel, W. (1959). Statistical aspect of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719–748.
- Meyer, J. P. (2018). jMetrik [Computer software]. Retrieved from <http://itemanalysis.com/jmetrik-download/>
- Min, S., & Bishop, K. (2022). *Evaluation of ACCESS Online Multistage Test Design (Unpublished WIDA Internal Report)*. Madison, WI: WIDA, University of Wisconsin-Madison.
- Muraki, E. (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement*, 17(4), 151–363.
- National Center on Educational Outcomes. (2021). *Universal design of assessments*. Retrieved from [https://nceo.info/Assessments/universal\\_design#:~:text=Universal%20design%20principles%20include%20careful,of%20content%20and%20skills%20tested](https://nceo.info/Assessments/universal_design#:~:text=Universal%20design%20principles%20include%20careful,of%20content%20and%20skills%20tested)

- Price, L. R., Lurie, A., Raju, N., Wilkins, C., & Zhu, J. (2006). Conditional standard errors of measurement for composite scores on the Wechsler Preschool and Primary Scale of Intelligence – Third Edition. *Psychological Reports, 98*(1), 237–252.
- Reise, S. P. (1999). Personality Measurement Issues Viewed Through the Eyes of IRT. In S. E. Embretson, & S. L. Hershberger (Eds.). *The New Rules of Measurement: What Every Psychologist and Educator Should Know* (pp. 219-240). Psychology Press.
- Rudner, L. (2001, Spring). Informed test component weighting. *Educational Measurement: Issues and Practice, 20*(1), 16–19.
- Sahakyan, N., (2020). “Generating alternate overall composite scale scores for English Learners with disabilities who are missing domain scores in the ACCESS for ELLs assessment”. WIDA Technical Report. September 2020.
- Sireci, S. G., & Zenisky, A. L. (2006). Innovative item formats in computer-based testing: In pursuit of improved construct representation. In T. M. Haladyna & S. M. Downing (Eds.), *Handbook of test development* (pp. 329–347). Mahwah, NJ: Routledge.
- Stahl, J. A., & Muckle, T. (2007). Investigating drift displacement in Rasch item calibrations. *Rasch Measurement Transactions, 21*(3), 1126–1127.
- Thissen, D. (2000). Reliability and measurement precision. In H. Wainer, N. Dorans, D. Eignor, R. Flaugher, B. Green, R. Mislevy, L. Steinberg, & D. Thissen (Eds.), *Computerized adaptive testing: A primer* (2nd ed., pp. 159–184). Hillsdale, NJ: Lawrence Erlbaum Associates.
- U.S. Department of Education. (2018). *A state’s guide to the U.S. Department of Education’s assessment peer review process*. Retrieved from [https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf?utm\\_content=&utm\\_medium=email&utm\\_name=&utm\\_source=govdelivery&utm\\_term=](https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf?utm_content=&utm_medium=email&utm_name=&utm_source=govdelivery&utm_term=)
- WIDA Consortium. (2007). *English language proficiency standards and resource guide, 2007 edition, prekindergarten through grade 12*. Madison, WI: Board of Regents of the University of Wisconsin System.
- WIDA Consortium. (2012). *2012 amplification of the English language development standards kindergarten–grade 12*. Madison, WI: Board of Regents of the University of Wisconsin System.
- WIDA Consortium. (2022). *ACCESS for ELLs Interpretive Guide for Score Reports*. Madison, WI: Board of Regents of the University of Wisconsin System.
- WIDA Consortium. (2021). *ACCESS for ELLs Test administrator manual*. Madison, WI: Board of Regents of the University of Wisconsin System.
- WIDA Consortium. (2021). *ACCESS for ELLs District and School Test Coordinator manual*. Madison, WI: Board of Regents of the University of Wisconsin System.
- WIDA Consortium. (2021). *Test Policy Handbook*. Madison, WI: Board of Regents of the University of Wisconsin System.

- WIDA Consortium. (2020). *WIDA consortium English Language Proficiency Assessment for grades 1-12 Test and Item Design Plan ACCESS for ELLs Online Annual Summative and WIDA Screener Online*. Madison, WI: Board of Regents of the University of Wisconsin System.
- Wright, B.D. & Douglas, G.A. (1975). *Best test design and self-tailored testing*. Research memorandum, Statistical Laboratory, Department of Education, University of Chicago.
- Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago, IL: MESA Press.
- Young, M. J., & Yoon, B. (1998, April). *Estimating the consistency and accuracy of classifications in a standards-referenced assessment* (CSE Technical Report 475). Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education and Information Studies.
- Zieky, M. (1993). DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337–347). Hillsdale, NJ: Erlbaum.
- Zwick, R., & Bridgeman, B. (2014). Evaluating validity, fairness, and differential item functioning in multistage testing. In Y. Duanli, A. A. von Davier, & C. Lewis (Eds.), *Computer multistage testing: Theory and applications* (pp. 271–284). Hoboken, NJ: CRC Press.
- Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 30, 233–251.
- Zwick, R., Thayer, D. T., & Wingersky, M. (1993). *A simulation study of methods for addressing differential item functioning in computer-adaptive tests* [ETS Research Report RR-93-11]. Princeton, NJ: Educational Testing Service. Retrieved from <https://onlinelibrary.wiley.com/doi/epdf/10.1002/j.2333-8504.1993.tb01522.x>

# Acknowledgments

We would like to extend our appreciation to the many CAL and WIDA staff members who have supported this work, including the following:

From CAL:

Tanya Bitterman, M.A.  
Sofia Buitrago, MS.  
Yage (Leah) Guo, Ph.D.  
Michele Kawood, M.S.Ed.  
Justin Kelly, Ph.D.  
Dorry M. Kenyon, Ph.D.  
Reshmi Kumpakha, M.A.  
Jung-Jung Lee, M.Sc.  
Samantha Musser, M.A.  
Yoon Ah Song, Ph.D.  
Jasmine Tsai, M.Ed.  
Frank Wucinski, M.A.  
Shu Jing Yen, Ph.D.  
Xin Yu, M.A.

From WIDA:

Kyoungwon Bishop, Ph.D.  
Yutong Zhang