



**Annual Technical Report for
ACCESS for ELLs Paper
English Language Proficiency Test
Series 502, 2020–2021 Administration**

Annual Technical Report No. 17B

Prepared by:

Center for Applied Linguistics

Language Assessment Division
Psychometrics and Quantitative Research Team

March 2022

The WIDA ACCESS for ELLs Technical Advisory Committee

This report has been reviewed by the WIDA ACCESS for ELLs Technical Advisory Committee (TAC), which includes 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

Executive Summary

This is the 17th annual technical report on the ACCESS for ELLs English Language Proficiency test and the 5th report on the ACCESS for ELLs assessment delivered in Paper format since the online assessment was launched.

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 2020–2021 ACCESS for ELLs 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 Series 502 Paper was administered in the school year 2020–2021 in 35 states, the Bureau of Indian Education, the Department of Defense Education Activity, the District of Columbia, and Northern Mariana Islands for a total of 39 state entities (henceforth “states”).

The ACCESS Series 502 Paper data set used in this report included the results of 376,246 students as of September 2021. The final number of students who participated in the ACCESS Series 502 Paper tests is 392,805. The grade with the most students participating was Kindergarten, with 163,557 students, while the grade with the fewest students was Grade 12, with 6,170 students. Of the participating WIDA states, Florida has the largest number of students, with 229,511, while the District of Columbia had the fewest, with 35 students.

During the 2020–2021 testing year, many states suspended in-person schooling due to the COVID-19 public health emergency. Based on a comparison with prior years' numbers of participating students, WIDA believes that 25% fewer students participated in ACCESS Series

502 testing than the ACCESS Series 501 testing. Further detail on the impact of COVID-19 is contained in the ACCESS 2020–2021 *Year in Review Report*.

ACCESS for ELLs Series 502 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 17B) provides technical information pertaining to ACCESS for ELLs Series 502 Paper. A second report (WIDA ACCESS Technical Report 17A) provides technical information for the ACCESS for ELLs Series 502 Online 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	1-1
1.3. The WIDA Proficiency Levels	1-3
1.4. Language Domains	1-4
1.5. Grade-Level Clusters	1-5
1.6. Tiers.....	1-6
2. Test Development.....	2-1
2.1. Item and Task Design	2-1
2.1.1. Listening Items	2-1
2.1.2. Reading Items	2-2
2.1.3. Writing Tasks	2-2
2.1.4. Speaking Tasks.....	2-3
2.2. Test Design	2-3
2.2.1. Listening.....	2-4
2.2.2. Reading.....	2-5
2.2.3. Writing.....	2-6
2.2.4. Speaking	2-8
2.3. Test Construction.....	2-10
2.3.1. Item Development	2-10
2.3.2. Field Testing and Item Selection	2-13
2.4. Kindergarten	2-15
2.4.1. Test Design.....	2-15
2.4.2. Test Construction.....	2-16
2.4.3. Item and Task Design.....	2-16
3. Test Administration.....	3-1
3.1. Test Delivery	3-1
3.2. Operational Administration.....	3-1
3.2.1. Listening Test Administration	3-2
3.2.2. Reading Test Administration.....	3-2
3.2.3. Writing Test Administration.....	3-3
3.2.4. Speaking Test Administration.....	3-3
3.2.5. Test Administrator Training.....	3-5
3.2.6. Test Security	3-5

3.3	Fairness and Accessibility	3-6
3.3.1	Support Provided to All ELLs	3-6
3.3.2	Support Provided to ELLs with IEPs or 504 Plans	3-6
4.	Scoring Procedures.....	4-1
4.1.	Multiple Choice Scoring: Listening and Reading	4-1
4.2.	Scoring Writing	4-1
4.3.	Writing Scoring Scale.....	4-6
4.4.	Speaking Scoring Scale	4-10
5.	Summary of Score Reports.....	5-13
5.1.	Individual Student Report.....	5-13
5.2.	Other Reports.....	5-16

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 English language learners.

ACCESS for ELLs is offered in two formats: ACCESS Paper, described in this report, and ACCESS Online, 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 the cases of some test items and 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 Paper 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, as long as each of the Standards is measured on the assessment in some capacity (although ACCESS for ELLs 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.

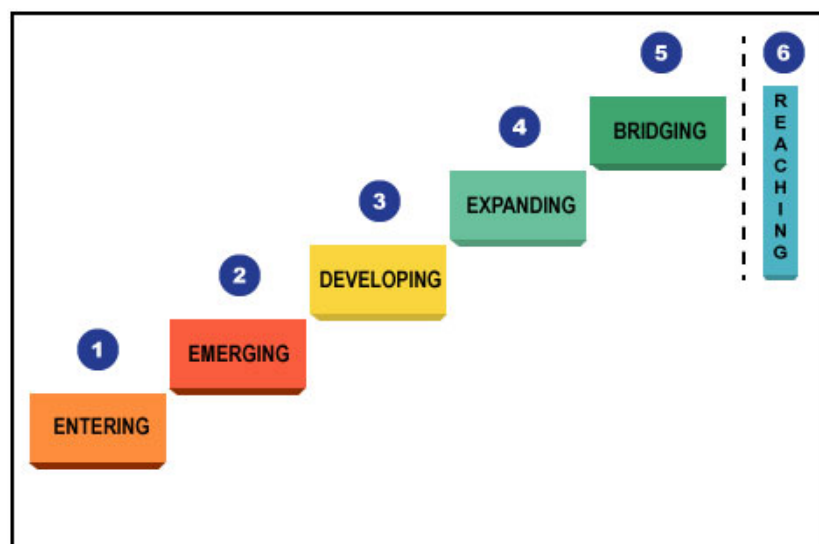


Figure 1. The language proficiency levels of the WIDA ELD Standards.

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: (a) vocabulary usage at the word/phrase level; (b) language forms and conventions at the sentence level; and (c) 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' understanding or demonstration of oral interaction and 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 PLs 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, that 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, 2021). 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 Paper is as follows: K, 1, 2, 3, 4–5, 6–8, 9–12.

In the lower grades (Grades 1–5), test forms may be shared across grade-level clusters. As described in Section 2.2.1 below, the Listening and Reading tests were developed prior to the launch of the 2016 operational administration, which represented the shift to the new cluster structure of ACCESS Online. Earlier ACCESS tests had a cluster structure that differs from that of the current ACCESS items in newer development, in the lower grades. The Speaking and Writing tests were developed using the ACCESS Online cluster structure. ACCESS Paper clusters, therefore, bridge the cluster structure of the older ACCESS assessments and ACCESS Online. For example, the Cluster 2 tests in the domains of Reading and Listening are the same test forms as the Cluster 1 tests. The Cluster 2 tests in the domains of Speaking and Writing are the same test forms as the Cluster 3 tests in these domains. Table 1 details the grade-level cluster structure of ACCESS Paper and the shared forms across clusters.

Table 1
ACCESS Paper Grade-Level Clusters and Shared Forms Across Clusters

ACCESS Paper Grade-level Clusters	Shared Test Forms (Listening and Reading)	Shared Test Forms (Speaking and Writing)	Grade	
K	K	K	K	
1	Cluster 1 and Cluster 2	Cluster 1	1	
2		Cluster 2 and Cluster 3	2	
3	Cluster 3 and Cluster 4–5		Cluster 3	3
4–5		Cluster 4–5	Cluster 4–5	4
			5	
6–8	Cluster 6–8	Cluster 6–8	6	
			7	
			8	
9–12	Cluster 9–12	Cluster 9–12	9	
			10	
			11	
			12	

Note that in our analyses of student participation in the assessment (Part 2, Chapter 1), analysis is conducted by cluster (K, 1, 2, 3, 4–5, 6–8, 9–12). In our analyses of test forms (Part 2, Chapter 2), analysis is conducted at the form level (i.e., in Listening and Reading, a single analysis is conducted for the Cluster 1 and Cluster 2 form). Test form level analyses are presented for each cluster that the form appears in; if a table of results pertains to more than one cluster, it is repeated in each cluster.

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. 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 that allow 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.

Paper ACCESS test forms are constructed at either Tier A (for students at beginning levels of English proficiency) or Tier B/C (for students at higher proficiency levels). Each Grade 1–12 test-taker takes either the Tier A form or the Tier B/C form. The Kindergarten assessment is not tiered.

In Listening and Reading, Tier A has items and tasks designed to allow students at the lowest language proficiency levels (PLs 1 and 2) to meet the WIDA ELD Standards at their language proficiency levels, and it includes some items targeted to PL 3. Tier B/C tests include items constructed to target PLs 2 (Emerging) through 5 (Bridging).

In the domain of Writing, Tier A forms include tasks written to elicit language up to PL 3, and Tier B/C forms include tasks written to elicit language up to PL 4 or PL 5. In the domain of Speaking, 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.

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 Sample Items page on WIDA’s website.

When the task models for ACCESS Paper were first developed, CAL and WIDA accounted for 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 and WIDA collaborated to design the item and task layout on the page to be maximally readable/legible and to contain 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 materials, and to include procedures for accommodation (such as human reader of item stimuli). The ways in which the CAL TD team ensures fairness by adhering to principles of UDA in item development are described in Section 2.3.1 below.

Note that this section applies to ACCESS Paper Grades 1–12. For detail on the item and task design for Kindergarten, see Section 2.4 below and the technical report on the development of the Kindergarten static form (MacGregor, Yen, & Yu, 2009).

2.1.1. Listening Items

All Listening items are multiple choice and are designed to be group administered. They 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 illustrations; 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.

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. *Folders* group together three test items that are written around a common theme, with each item targeting a progressively higher proficiency level.

In ACCESS Paper, the Listening tests have a Tier A and a Tier B/C form for each grade-level cluster; students are placed into the tier based on a decision made at the school or district level as local EL teachers judge students’ abilities based on their classroom performance.

Listening items are developed so that each folder appears on a 2-page spread in a test booklet, although some folders go onto a third page. Scripts containing the item orientation, stimulus, and

question stem are audio recorded with professional voice actors and produced by a professional recording studio. Audio playback of test item content is done via audio CD, and explicit instructions on starting and pausing the CD are provided in the Test Administrator’s Script and the Test Administrator Manual.

Listening items are centrally scored by Data Recognition Corporation (DRC) via an automated process.

2.1.2. Reading Items

All Reading items are multiple choice and are designed to be group administered. They are similar in format to Listening items. Reading items are selected response items, with one key and either two or three distractors, depending on grade-level cluster and targeted proficiency level. For Grades 1 and 2, all items have a key and two distractors. For Grades 3, 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.

The stimulus for Reading items is written text, and answer choices primarily are also written text, though for Grades 1–12 response options for items targeting PLs 1, 2, and 3 may be illustrations rather than text. As with Listening items, Reading items are grouped into thematic folders of three test items each. In ACCESS Paper, the Reading tests have a Tier A and a Tier B/C form for each grade-level cluster; students are placed into the tier based on a decision made at the school or district level.

Reading items are centrally scored by DRC via an automated process.

2.1.3. Writing Tasks

All Writing tasks are constructed response tasks and are designed to be group administered. Students write responses by hand in paper booklets.

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. As described in Section 2.2.3 below, DRC raters score students’ written responses to these tasks using the entire breadth of the scoring scale; 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 4 or 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 performances may extend above or below the PL 5 range.

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 written 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 Grades 1–5, a word box may be provided, depending on the grade level, targeted proficiency level, and task.

For all grade clusters and tiers, the Writing test is group administered by a live Test Administrator. The Test Administrator reads instructions aloud from the Test Administrator’s Script and monitors student progress through the test. For all grade clusters and tiers, the students hand-write their answers in the same test booklet containing the Listening and Reading tests.

2.1.4. Speaking Tasks

The Speaking test is administered individually to each test-taker. The test is media delivered. Students listen to an audio recording of the test input while following along in a test booklet.

Stimuli on the Speaking test include graphics, audio, and text, presented in a test booklet as a series of “speech bubbles” from the perspective of the Virtual Test Administrator (VTA) and virtual model student. All text is multimodal, presented both in the test booklet and read aloud on the audio CD. Scripts containing the task content are audio recorded with professional voice actors and produced by a professional recording studio. Audio playback of test item content is done via audio CD, and explicit instructions on starting and pausing the CD are provided in the Test Administrator’s Script and the Test Administrator Manual.

The CD audio stimuli are presented in terms of a 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 to the model student respond to a parallel task. The purpose of the model is to demonstrate task expectations to both students and to the Test Administrator, who scores the Speaking test. Students respond orally to the tasks, with their responses scored immediately by the Test Administrator using a scoring scale. The Test Administrator records scores on the Speaking test in the same booklet the student used for the Listening, Reading, and Writing tests.

2.2. Test Design

This section describes how ACCESS Paper is assembled to ensure that the evidence collected is (a) sufficient to make the required decisions based on the test results, and (b) appropriate for the student’s level of proficiency. This section provides information on the test design for the two forms of Paper ACCESS (Tier A and Tier B/C) and the design of each form. Note that this section applies to ACCESS Paper Grades 1–12. For detail on Kindergarten, see Section 2.4 below and the technical report on the development of the Kindergarten static form (MacGregor, Kenyon, Gibson, & Evans, 2009).

2.2.1. Listening

For the ACCESS Listening test, Table 2 shows, for each test form, the number of items, the targeted range of WIDA proficiency levels, the item types, the response format, and the scoring procedure.

Table 2
Number and Types of Items on the Listening Test

Grade-Level Cluster	Tier	Number of Items	Targeted PL Range	Item Types	Response Formats	Scoring Procedures
1	A	18	PL1–PL4	Multiple choice	Dichotomous selected response	Machine scored
1	B/C	21	PL2–PL5			
2	A	18	PL1–PL4	Multiple choice	Dichotomous selected response	Machine scored
2	B/C	21	PL2–PL5			
3	A	18	PL1–PL4	Multiple choice	Dichotomous selected response	Machine scored
3	B/C	21	PL2–PL5			
4–5	A	18	PL1–PL4	Multiple choice	Dichotomous selected response	Machine scored
4–5	B/C	21	PL2–PL5			
6–8	A	18	PL1–PL4	Multiple choice	Dichotomous selected response	Machine scored
6–8	B/C	21	PL2–PL5			
9–12	A	18	PL1–PL4	Multiple choice	Dichotomous selected response	Machine scored
9–12	B/C	21	PL2–PL5			

Figure 2 presents the Listening test design, showing the distribution of folders by Standard for each tier. In this figure, each small gray box represents an item.

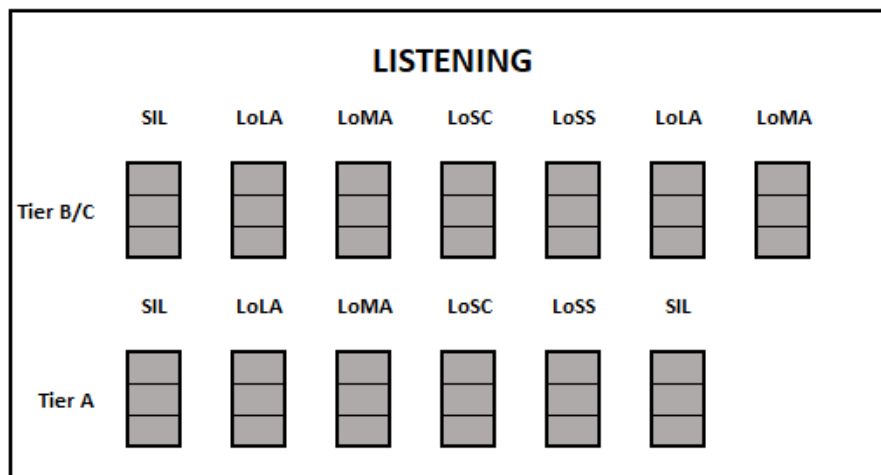


Figure 2. Distribution of items by Standard for each tier of the Listening test.

Note that the test design is slightly different between Tier A and Tier B/C. Tier B/C students, who potentially may be reclassified by the assessment, take a slightly longer test and take two folders each assessing the Language of Language Arts and the Language of Mathematics Standards. Tier A students receive a second folder assessing the Social and Instructional Language Standard, under the assumption that less proficient students will find this Standard more accessible.

Although timing guidance is provided to Test Administrators in the Test Administrator Manual, the Listening subtest is untimed.

2.2.2. Reading

For the ACCESS Reading test, Table 3 shows, for each test form, the number of items, the targeted range of WIDA proficiency levels, the item types, the response format, and the scoring procedure.

Table 3
Number and Types of Items on the Reading Test

Grade-Level Cluster	Tier	Number of Items	Targeted PL Range	Item Types	Response Formats	Scoring Procedures
1	A	24	PL1–PL4	Multiple choice	Dichotomous selected response	Machine scored
1	B/C	27	PL2–PL5			
2	A	24	PL1–PL4	Multiple choice	Dichotomous selected response	Machine scored
2	B/C	27	PL2–PL5			
3	A	24	PL1–PL4	Multiple choice	Dichotomous selected response	Machine scored
3	B/C	27	PL2–PL5			
4–5	A	24	PL1–PL4	Multiple choice	Dichotomous selected response	Machine scored
4–5	B/C	27	PL2–PL5			
6–8	A	24	PL1–PL4	Multiple choice	Dichotomous selected response	Machine scored
6–8	B/C	27	PL2–PL5			
9–12	A	24	PL1–PL4	Multiple choice	Dichotomous selected response	Machine scored
9–12	B/C	27	PL2–PL5			

Figure 3 presents the Reading test design, showing the distribution of folders by Standard for each tier. In this figure, each small gray box represents an item.

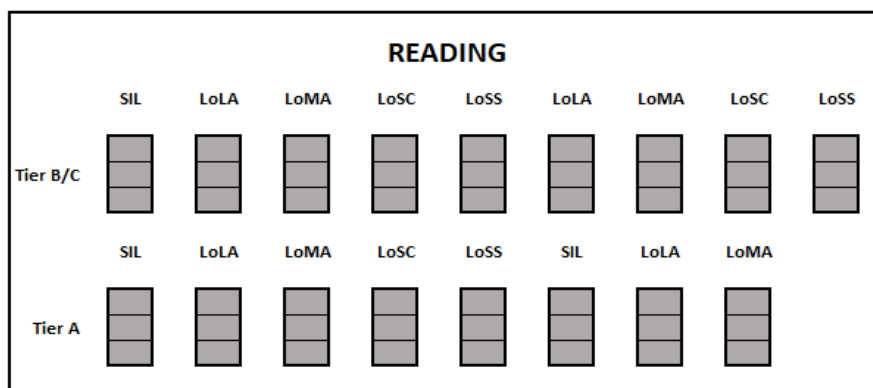


Figure 3. Distribution of items by Standard for each tier of the Reading test.

As with Listening, the Reading Tier A test is shorter and focuses on Standards deemed more accessible for lower-proficiency students.

Although timing guidance is provided to Test Administrators in the Test Administrator Manual, the Reading subtest is untimed.

2.2.3. Writing

For the ACCESS Writing test, Table 4 shows, for each test form, the number of tasks, the targeted range of WIDA proficiency levels, the task types, the response format, and the scoring procedure.

Table 4
Number and Types of Items on the Writing Test

Grade-Level Cluster	Tier	Number of Tasks	Targeted PL Range	Task Types	Response Formats	Scoring Procedures
1	A	4	PL1–PL3	Writing constructed response	Polytomous constructed response; handwritten in test booklet	Human scored: centrally scored by DRC
1	B/C	3	PL2–PL5			
2	A	3	PL1–PL3	Writing constructed response	Polytomous constructed response; handwritten in test booklet	Human scored: centrally scored by DRC
2	B/C	3	PL2–PL5			
3	A	3	PL1–PL3	Writing constructed response	Polytomous constructed response; handwritten in test booklet	Human scored: centrally scored by DRC
3	B/C	3	PL2–PL5			
4–5	A	3	PL1–PL3	Writing constructed response	Polytomous constructed response; handwritten in test booklet	Human scored: centrally scored by DRC
4–5	B/C	3	PL2–PL5			
6–8	A	3	PL1–PL3	Writing constructed response	Polytomous constructed response; handwritten in test booklet	Human scored: centrally scored by DRC
6–8	B/C	3	PL2–PL5			
9–12	A	3	PL1–PL4			

9–12	B/C	3	PL2–PL5	Writing constructed response	Polytomous constructed response; handwritten in test booklet	Human scored: centrally scored by DRC
------	-----	---	---------	------------------------------	--	---------------------------------------

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, some students will score throughout the proficiency range.

Except for Grade 1 Tier A, both tiers consist of three tasks. Grade 1 Tier A has four tasks, designed specifically to allow beginning writers at this grade to demonstrate their ability in the domain of Writing. Figures 4 and 5 present the Writing test design, showing the distribution of tasks for each tier. In these figures, each colored box represents a task. The number in the box represents the targeted proficiency level of the task.

Although timing guidance is provided to Test Administrators in the Test Administrator Manual, the Writing subtest is untimed.

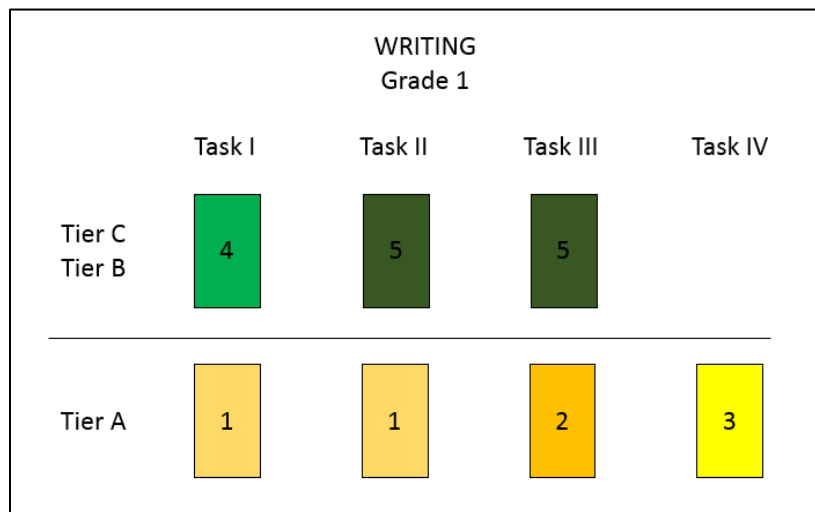


Figure 4. Distribution of tasks by targeted proficiency level for each tier of the Grade 1 Writing test.

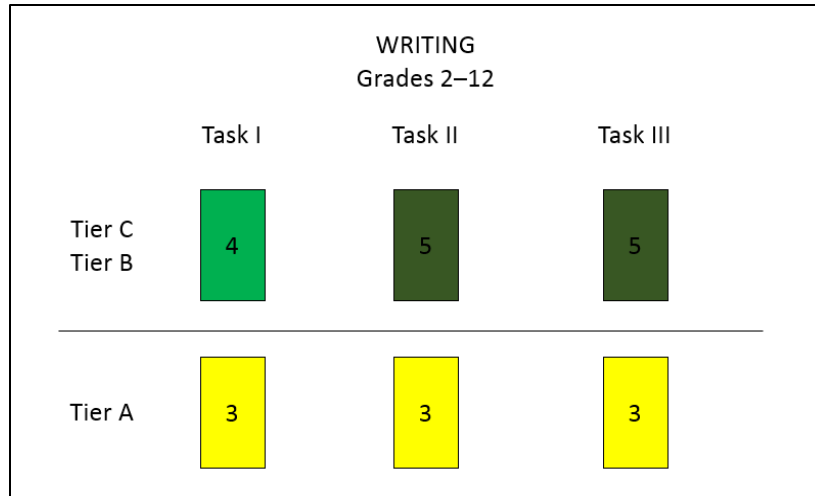


Figure 5. Distribution of tasks by targeted proficiency level for each tier of the Grades 2–12 Writing test.

2.2.4. Speaking

For the ACCESS Speaking test, Table 5 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 5
Number and Types of Items on the Speaking Test

Grade-Level Cluster	Tier	Number of Tasks	Targeted PL Range	Task Types	Response Formats	Scoring Procedures
1	A	6	PL1–PL3	Speaking constructed response	Polytomous constructed response	Human scored; scored by Test Administrator
1	B/C	6	PL3–PL5			
2	A	6	PL1–PL3	Speaking constructed response	Polytomous constructed response	Human scored; scored by Test Administrator
2	B/C	6	PL3–PL5			
3	A	6	PL1–PL3	Speaking constructed response	Polytomous constructed response	Human scored; scored by Test Administrator
3	B/C	6	PL3–PL5			
4–5	A	6	PL1–PL3	Speaking constructed response	Polytomous constructed response	Human scored; scored by Test Administrator
4–5	B/C	6	PL3–PL5			
6–8	A	6	PL1–PL3	Speaking constructed response	Polytomous constructed response	Human scored; scored by Test Administrator
6–8	B/C	6	PL3–PL5			
9–12	A	6	PL1–PL3			

9–12	B/C	6	PL3–PL5	Speaking constructed response	Polytomous constructed response	Human scored; scored by Test Administrator
------	-----	---	---------	-------------------------------	---------------------------------	--

Figure 6 shows the format of the Speaking test. The Speaking test includes tasks that target language elicitation at three proficiency levels: 1, 3, and 5. The tasks are grouped into thematic folders, which are aligned to one or two of the WIDA Standards. These folders are generally presented in the same order as the folders in the Listening and Reading tests; folders aligned to SIL are presented first, then folders aligned to LoLA, and then folders aligned to LoMa.

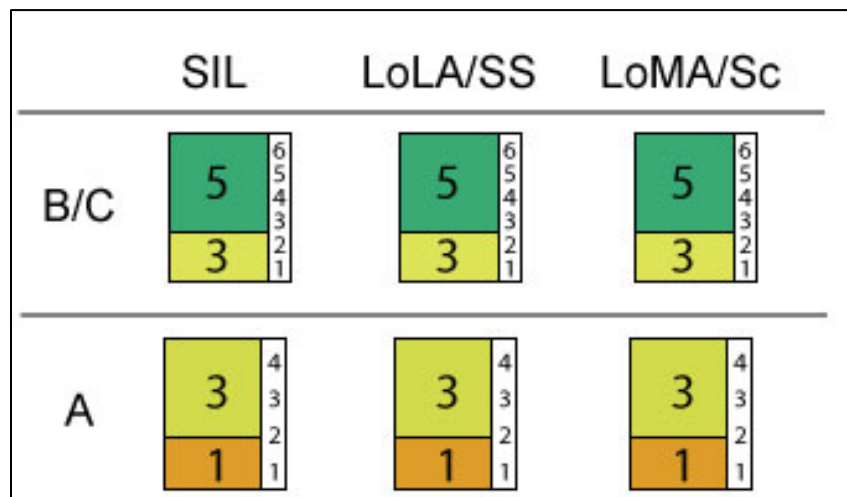


Figure 6. Distribution of tasks for each tier of the Speaking test.

As shown in Figure 6, the Speaking test includes two tiers. 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. For example, the Tier A and Tier B/C folders that address Social and Instructional Language in each grade cluster make up a single thematic panel, with the PL 3 tasks shared across tiered folders in a panel. In other words, within a Social and Instructional Language panel, the same PL 3 task appears on both the Tier A and the Tier B/C form.

Although timing guidance is provided to Test Administrators in the Test Administrator Manual, the Speaking subtest is untimed.

2.3. Test Construction

2.3.1. Item Development

ACCESS Paper Series 502 is one of two static rotating Paper test forms. The ACCESS testing program transitioned in 2016 from an entirely paper-based program to the launch of ACCESS in both Online and Paper formats.

The CAL TD team developed the Listening and Reading items for ACCESS Paper prior to the launch of ACCESS Online, when ACCESS was entirely paper based. The CAL TD team also developed most ACCESS Paper Writing tasks for ACCESS when it was entirely paper based; however, a small subset of Writing tasks on ACCESS Paper Series 502 were developed as online tasks that were subsequently reformatted for administration as paper-based tasks. The CAL TD team developed the Speaking tasks and field tested them as ACCESS Online tasks before being reformatted for administration as ACCESS Paper tasks.

The general process of item writing and editing, and of item Content and Bias and Sensitivity reviews, remains similar across these transitions. For ACCESS Paper items, trained item writers worked from item specifications to draft items within the thematic folder design. After item writing was complete, the CAL TD team reviewed the folders, using a standard checklist, to determine which would undergo further development and which would be retired. Folders then went to their first external review, the Standards Expert review.

During the Standards Expert review, educators provided feedback about the overall grade-level appropriateness of the language and content of the items to ensure that no drift, in terms of grade-level appropriateness of the content or the language, occurred between the content generated during item writing and what was intended in the specifications. CAL recruited educators with ESL and content-area expertise to serve as Standards Experts and provided synchronous training on how to conduct the review and complete the review questionnaire. CAL Language Testing Specialists prepared a short questionnaire with open-ended questions about each folder and sent the questionnaires and folders to the Standards Experts.

Subsequent to the Standards Expert review, all content proceeded through a rigorous folder refinement stage internal to CAL. Folder refinement included numerous steps, including additional research and sourcing/fact-checking, meticulous review against a comprehensive, industry-standard item development checklist with peer review that other Language Testing Specialists carried out, as well as review by the Test Development Manager and the Director of Test Development and successive rounds of revision before sign-off. During this stage, all aspects of the items were scrutinized: the WIDA proficiency level of the stimulus, the graphic support, the question stems and response options (for the Listening and Reading tests), and the task prompts (for the Speaking and Writing tests). The CAL TD team also conducted mock administrations. During this phase, Language Testing Specialists produced other ancillary materials, such as Test Administrator scripts. Upon sign-off, the CAL TD team worked with the

CAL Production team to generate the graphics used on the test. Once the graphics had been generated, they were inserted into the folders, and layout review and fact-checking were conducted (with Test Development Manager sign-off) to ensure that the items were ready for external Content Review and Bias and Sensitivity Review.

Content Review and Bias and Sensitivity Review are external reviews that educators and WIDA staff carry out on ACCESS items. WIDA assembles 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 involves several criteria in the selection process which differ slightly between content review and bias and sensitivity review.

Content reviews occur by grade cluster (G1, G2-3, G4-5, G6-8, and G9-12) and the educators who are recruited to review a particular grade cluster's content (4 reviewers per grade cluster) have experience teaching English language learners and are either currently teaching that grade cluster or have extensive experience teaching that grade cluster. Further criteria are used to try to ensure a good balance within the panels. These criteria include recruiting at least one educator within each panel with experience in each of the following areas: ELA, Science, Math, Social Studies, Special Education. Additionally, during the recruitment process, WIDA seeks to ensure diversity and balance across a) consortium states, b) locale (rural/suburban/urban), c) educator background, and d) years of teaching experience. CAL and WIDA first train the Content Review Panel on the procedures and scope of the review. The panelists are introduced to the test layout, are instructed on the logistics of the review, and are trained on the use of the review checklist. The panel then reviews each item and task to determine whether the content is accessible and relevant to students in the targeted grade-level cluster and at the targeted WIDA proficiency level, and that each item or task 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 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. Bias and sensitivity reviews occur by grade groupings (e.g., G1-3, G4-5, G6-8, and G9-12) and the educators who are recruited to review a particular grade cluster's content (5 or 6 reviewers per grade grouping) are educators or administrators who have experience teaching English language learners and are either currently teaching the grades within their group or have experience teaching those grades. Further criteria are used to ensure balance within the panels, as a variety of perspectives is crucial for the bias and sensitivity reviews. These criteria include recruiting at least one educator per panel with experience in Special Education. Additionally, during the recruitment process, WIDA seeks to ensure diversity and balance across a) consortium states, b) locale (rural/suburban/urban), c) educator background, and d) years of teaching experience. CAL and WIDA conduct training for all new and returning reviewers before any items are reviewed. CAL and WIDA staff facilitate the synchronous reviews and take extensive

notes to capture all feedback during the reviews. WIDA TD staff also conducts 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 provides written feedback on the materials.

Once the CAL TD team compiled all Content Review and Bias and Sensitivity Review feedback from educators and from WIDA, CAL Language Testing Specialists worked to implement the feedback, with CAL Test Development Manager sign-off as a final step. Graphics were subsequently revised by the CAL Test Production team accordingly. The input and feedback from educators at various stages in the item development process served as evidence that each item was appropriate for the age and grade-level cluster for which it was intended.

Tasks in the domain of Writing and Speaking underwent one additional step: two rounds of small-scale tryouts with educators and students. These tryouts allowed CAL to evaluate whether the Speaking and Writing tasks would effectively elicit language at the targeted WIDA proficiency levels. In the initial round of tryouts, members of the CAL TD team recruited schools to permit CAL staff to administer the tasks to, and conduct cognitive labs with, students with consent to participate. The tasks were then revised and subject to a second round of tryouts, this time conducted by classroom teachers with their students, who were also recruited by CAL and WIDA to participate. CAL Language Testing Specialists used the results, including student responses, cognitive lab observations of students, and student and teacher feedback, to inform final revisions to the tasks prior to field testing.

After the CAL Language Testing Specialists completed edits from the Content Review and Bias and Sensitivity Review (and tryout edits for Speaking and Writing), they then prepared the folders for final production. Additionally, they produced audio recording scripts for professional audio recording, arranged for recording the audio files, completed extensive quality control checks for both content and technical specifications of the audio (e.g., file types, recording quality, and compression levels), conducted final layout reviews, and performed key checks for the Listening and Reading tests. WIDA signed off on all materials prior to administration. Items and tasks that reached this point then went through field testing and test assembly processes, described in the next subsection by domain.

Throughout item development, the CAL TD team focused on issues of fairness. First, the team applied principles of Universal Design of Assessments (UDA) during item development. At the item specification level, the CAL TD team aimed to precisely define the construct that each item or task was intended to measure. For the linguistic content of items, several principles for UDA were built into the item development checklists and were specifically reviewed for by CAL's TD managers and external reviewers (including WIDA staff and outside educators during Standards Expert review and Bias and Sensitivity and Content reviews), including:

- Accessible, nonbiased items
- Amenability to accommodations

- Simple, clear, and intuitive instructions and procedures
- Maximum readability and comprehensibility
- Maximum legibility

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

Note that this section applies to ACCESS Paper Grades 1–12. For detail on Kindergarten, see Section 2.4 below and the technical report on the development of the Kindergarten static form (MacGregor et al., 2009).

2.3.2. Field Testing and Item Selection

2.3.2.1. Listening and Reading

The Listening and Reading items for ACCESS Paper were created by the CAL TD team prior to the launch of ACCESS Online, when ACCESS was entirely paper based. ACCESS was first field tested in 2004, and from 2004 to 2014, development continued for ACCESS, culminating in Series 303, operational in 2014–2015. For further detail on this original field test and on the processes for ongoing item development from 2004 to 2014, see Section 2.3.1 above, along with the ACCESS for ELLs Technical Reports, particularly ACCESS for ELLs Technical Report No. 1, *Development and Field Test of ACCESS for ELLs* (Kenyon, 2006), and *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test, Series 303* (CAL, 2016b).

In all grade clusters, the Tier A Listening and Reading forms are static forms, which were constructed prior to the launch of ACCESS Online.

In all grade clusters, the operational Tier B/C forms in Listening and Reading forms for Series 502 are identical to those administered in Series 403. These forms are composed of items that were previously operational in Series 400 and 401 and that were developed, as described in Section 2.3.1 above, during the development cycles when ACCESS was entirely paper based. Beginning with Series 403, to streamline operational administration, CAL and WIDA decided to combine ACCESS Paper Listening and Reading Tier B and Tier C tests to create a new Tier B/C test in Listening and in Reading for each grade-level cluster.

To select these new forms, the pool of Listening and Reading Paper Tier B and Tier C items that were administered to the Series 401 and Series 400 populations was recalibrated using the population data (see Part 2, Section 2.7 for more information on the recalibration). CAL and WIDA conducted a forms selection meeting in early 2018, prior to the operational administration of Series 403. Staff from WIDA and CAL reviewed the pool of items in Series 401 and 400 Listening and Reading Tier B and Tier C and selected two new static Tier B/C forms for each grade-level cluster in Listening and Reading—one for use in Series 403 and the other for use in

Series 501, with alternating administrations henceforth. Forms were selected to maintain the coverage of WIDA ELD Standards as called for in the test design and to ensure inclusion of items of sufficient difficulty to measure students in the Tier C range.

2.3.2.2. *Writing*

There are two static rotating forms for ACCESS Paper Writing. The first of these is composed of the same set of items, across all grade-level clusters and tiers, as the test used the first year of ACCESS Online. The second form is composed of the same set of items, across all grade-level clusters and tiers, as the test used the second year of ACCESS Online.

Tasks on the first of the two rotating static forms were used operationally prior to the launch of ACCESS Online and were re-field tested in the Online mode for the first year of ACCESS Online. Tasks selected for use in the first ACCESS Online operational test were then reformatted for presentation in the first of the Paper static forms.

The second rotating static form uses continuing tasks from the first form, as well as tasks newly field tested for the second year of ACCESS Online and then reformatted for Paper presentation. For further detail on this field test, see the Series 401 Online ACCESS technical report (CAL, 2018).

ACCESS Paper 502 is the second of the two rotating static forms.

2.3.2.3. *Speaking*

The Speaking test for ACCESS Paper is likewise one of two static rotating forms. The first of these forms is composed of the same set of items, across all grade-level clusters and tiers, as the second year of the ACCESS Online Speaking test; the second form is composed of the same set of items, across all grade-level clusters and tiers, as the third year of the ACCESS Online Speaking test. Speaking tasks have some differences in presentation between Online and Paper. In addition, the Paper test does not include the Speaking tier Pre-A, which is included in the Online test.¹

Tasks for these two rotating forms were field tested during the initial ACCESS Online field test, as well as through embedded field testing during the first and second years of the ACCESS Online assessments. These Speaking tasks went through both quantitative and qualitative analyses following the field test to determine their appropriateness for inclusion in the next year's operational test. After field testing, the Speaking tasks were then produced in the paper-based format.

¹ Students with very low ability levels in the Listening and Reading domains are routed to the Pre-A tier in ACCESS Online Speaking. The purpose of the Pre-A tier is to reduce the affective impact of the test on these students. As the Paper test is not adaptive, there is no way to route these students to Pre-A for Paper.

2.4. Kindergarten

Kindergarten ACCESS for ELLs is a static form and is not refreshed from year to year.

2.4.1. Test Design

CAL and WIDA designed Kindergarten ACCESS for ELLs to be engaging for very young children, and the test design was informed by consultation with kindergarten teachers and a panel of early childhood assessment experts. The test design incorporates a high-interest, age-appropriate storybook format, using child-friendly graphics, and includes manipulatives for students to demonstrate comprehension. The test is built on two thematic texts in a storybook format, one narrative and one expository. The storybook is read aloud by the Test Administrator. There are Listening, Speaking, Reading, and Writing assessment tasks related to each text. To minimize testing times and to ensure that students are presented with assessment tasks appropriate to their abilities, the test includes stopping rules (designed to ensure that children of beginning proficiency are not overchallenged) and skipping rules (designed so that children of more advanced proficiency can skip forward to more challenging tasks).

The test is administered one-on-one by trained Test Administrators, who mark responses in the Student Response Booklet.

Table 6 provides, for each domain, the number of items, the targeted range of WIDA proficiency levels, the item types, the response format, and the scoring procedure.

Table 6
Number and Types of Items on Kindergarten ACCESS

Domain	Number of Items	Targeted PL Range	Item Types	Response Formats	Scoring Procedures
Listening	30	P1–P5	Dichotomous	Student points to picture or manipulates cards	Administrator records response (correct/incorrect) in Student Response Booklet
Speaking	10	P1–P5	Dichotomous	Oral response	Administrator records response (correct/incorrect) in Student Response Booklet
Writing	6	P1–P5	Dichotomous and Polytomous	Student handwrites in booklet	Administrator records response (correct/incorrect) for dichotomous tasks and rates responses and records rating for polytomous tasks
Reading	30	P1–P5	Dichotomous	Student reads aloud or matches picture cards with text cards	Administrator records response (correct/incorrect) in Student Response Booklet

2.4.2. Test Construction

Field testing for Kindergarten ACCESS was conducted in 2008. A full description of item development, field testing, final forms selection, and initial standard setting for Kindergarten ACCESS can be found in the technical brief *Development and Field Test of Kindergarten ACCESS for ELLs* (MacGregor et al., 2009). Cut scores for Kindergarten were most recently updated in the 2016 ACCESS standard setting (Cook & MacGregor, 2017); see Part 2, Section 2.1 for more information.

2.4.3. Item and Task Design

As noted above, the Kindergarten ACCESS test is composed of two thematic texts. The items and tasks are designed to build upon the content of these texts.

In the domain of Listening, the Test Administrator reads the prompt aloud to the student, and the student responds by either pointing to an item in a picture or manipulating a picture card. The Test Administrator records the response (correct or incorrect) in the Student Response Booklet.

Students respond to Writing tasks in the Student Response Booklet. The initial Writing tasks for each thematic text are dichotomously scored by the Test Administrator. The Test Administrator Script indicates the level required for a task to meet expectations and to be scored correct. The Test Administrator scores the final Writing task in each thematic text section using a rating scale. The Test Administrator rates the student's Writing on a scale of 0 to 6.

The Test Administrator reads the Speaking tasks aloud, and students respond orally. Tasks are dichotomously scored by the Test Administrator. The Test Administrator Script indicates the level required for a task to meet expectations and to be scored correct.

To administer Reading tasks, Test Administrators ask students to identify letters or read text. Students respond by manipulating picture cards or by pointing at pictures. Students may also read aloud. The Test Administrator records the response (correct or incorrect) in the Student Response Booklet.

The items on Kindergarten ACCESS were developed to collectively assess all five WIDA Standards in all domains across the proficiency levels, as shown in Table 7. To keep the test an appropriate length for the population, it was not possible to assess each Standard at each proficiency level in each domain. Therefore, tasks were distributed by Standard across the proficiency levels and domains to achieve appropriate coverage.

Although the average time per test is provided to Test Administrators in the Test Administrator Manual, Kindergarten ACCESS is untimed.

Student Response Booklets are centrally scanned at DRC.

Table 7

Number of Items by WIDA Standard and Targeted Proficiency Level on Kindergarten ACCESS

Listening										
WIDA Standard	Narrative Storyline					Expository Storyline				
	Number of items at targeted PL range					Number of items at targeted PL range				
	1	2	3	4	5	1	2	3	4	5
SI	3		3		3	3				
LA									3	
MA							3			
SC										
SS		3		3				3		3
Speaking										
WIDA Standard	Narrative Storyline					Expository Storyline				
	Number of items at targeted PL range					Number of items at targeted PL range				
	1	2	3	4	5	1	2	3	4	5
SI			3						3	
LA					3					
MA							3	3		3
SC						3				
SS	3	3		3						
Writing										
WIDA Standard	Narrative Storyline					Expository Storyline				
	Number of items at targeted PL range					Number of items at targeted PL range				
	1	2-5				1	2	3	4/5	
SI	1					1				
LA										
MA							3			
SC										
SS								4		
IT (SIL, LoLA, LoSS)		1							1	
Reading										
WIDA Standard	Narrative Storyline					Expository Storyline				
	Number of items at targeted PL range					Number of items at targeted PL range				
	1	2	3	4	5	1	2	3	4	5
SI	3				3	3			3	
LA										
MA										
SC		3	3				3	3		3

SS				3						
----	--	--	--	---	--	--	--	--	--	--

3. Test Administration

3.1. Test Delivery

Administration of ACCESS Paper typically takes place between December and April of the academic year, with testing windows determined at the state level. During the 2019–2020 school year, many states extended their testing windows due to the COVID-19 pandemic. The domain tests may be administered in any order. The test may be administered in several sessions within 1 day or over a series of days.

The Listening and Reading tests may be group or individually administered. Students are administered the Listening and Reading test forms using paper test booklets, and students record their answers directly in the test booklets. For the Listening test, the audio stimuli are played aloud via an audio CD.

The Writing test may be group or individually administered. Students are administered the Writing test via paper test booklets. Students record their responses directly in the test booklet.

The Speaking test is individually administered. Students listen to an audio recording and follow along in an accompanying test booklet. Each task also includes a model student response, which serves as an exemplar to the student and as a benchmark to the Test Administrator who scores the task. All audio stimuli are presented via audio CD.

3.2. Operational Administration

Before, during, and after your 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, and Test Administrators. The Test Administrator administers and monitors the test and is responsible for managing student data prior to, during, and after testing.

The training course within the WIDA Secure Portal (<https://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 your state’s testing window. Training courses include test preparation and administration tutorials and online administration quiz.

The roles of the test administrator and technology coordinator are critical for the proper administration of the assessments as proper training and familiarity with ACCESS for ELLs administration requirements is key to the validity of the test and the appropriate interpretations of test scores.

For detailed guidelines about training and test administration, please refer to the *ACCESS for ELLs Test administrator manual* and the *ACCESS for ELLs District and School Test Coordinator Manual*.

3.2.1. Listening Test Administration

The ACCESS for ELLs Paper Listening test is media delivered. Listening test items are delivered via CD.

3.2.1.1. Listening Test Materials

Test materials include the following items:

- Test Administrator’s Script
- Student Test Booklet(s)
- Listening and Speaking Test CD (a separate CD for each grade-level cluster and tiered test form). In the rare event that a student requires a human reader as an accommodation, the Human Reader Accommodation Script is required to administer the Listening section individually for that student.
- At least one sharpened number 2 pencil for each student to mark responses
- Speakers
- A CD player or desktop/laptop computer (to play the CD)

3.2.1.2. Organization and Timing of the Listening Test

The Listening test is designed to take approximately 25 to 40 minutes, depending on the grade-level cluster and tier. The test administration time does not include time for convening students, taking attendance, distributing, and collecting test materials, explaining test directions, or completing practice items. The length of test items increases with students’ language proficiency and grade level. For example, the Tier B/C Listening test takes longer to administer than the Tier A Listening test, and the Listening test for Grades 9–12 may take slightly longer than the test for Grades 4–5.

3.2.2 Reading Test Administration

The ACCESS for ELLs Reading test is completed within Student Test Booklets after a scripted introduction by the Test Administrator.

3.3.2.1. Reading Test Materials

Reading test materials include the following items:

- Test Administrator’s Script

- Student Test Booklet(s)
- At least one sharpened number 2 pencil for each student to mark responses

3.2.2.1 Organization and Timing of the Reading Test

The Reading test is designed to take no more than 35 to 45 minutes. The test administration time does not include time for convening students, taking attendance, distributing, and collecting test materials, explaining test directions, or completing practice items.

3.2.3 Writing Test Administration

Students respond to a set of tasks, writing their responses in their Student Test Booklets.

3.2.3.1. Writing Test Materials

Writing test materials include the following items:

- Test Administrator’s Script
- Student Test Booklet(s)
- At least one sharpened number 2 pencil for each student to write responses
- Scratch paper

3.2.3.2 Organization and Timing of the Writing Test

There are three tasks (Parts A, B, and C) on each Tier (Tiers A and B/C) of the Writing test for all grade levels except Tier A for Grade 1, which contains four tasks. For grade-level clusters 2, 3, 4–5, 6–8, and 9–12, the Tier A Writing tests have recommended guidelines for Parts A, B, and C of 15 minutes each, with up to 5 additional minutes for each part if needed for students to finish writing, for a total of 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.

3.2.4 Speaking Test Administration

The ACCESS for ELLs Speaking test is an individually administered test that standardizes test administration across students. Speaking test items are media delivered. Speaking test audio is provided on the same CD as the Listening test. The Speaking test provides ELLs with the opportunity to demonstrate their academic English language proficiency in Speaking across the WIDA ELD Standards through a set of constructed response tasks. The Speaking test is tiered. Students will either take the Tier A form or the Tier B/C form; both are included in the same Speaking Test Booklet.

3.2.4.1 Audio Format of the Speaking Test

The Speaking test is multimodal. The student hears audio input and sees the input as text in the Speaking Test Booklet. This presentation format supports the student in understanding test input. Media delivery of the Speaking test means that an audio recording will guide the student through the Speaking test. The audio recording includes two voices: a model student and a Virtual Test Administrator.

Each task on the Speaking test is preceded by a model student task and response. The questions posed to the model student are at the same proficiency level as the tasks to which the student will respond, allowing the model student to demonstrate the expected language use at a given proficiency level. In most cases the model questions are designed to be parallel to but not exactly the same as the examinee questions. The model student also has an important function in scoring since the scoring scale is designed to evaluate student responses relative to the model student's response.

The Virtual Test Administrator guides the student through the test and asks the student questions designed to elicit language at targeted proficiency levels. While the Virtual Test Administrator will instruct and guide the student through the Speaking test, the administrator may also need to assist the student in navigating test materials (e.g., turning the page when prompted). The Speaking test includes standardized, built-in response time for every task. The amount of time varies according to the grade-level cluster, tier, and proficiency level of the task and ranges from 15 to 50 seconds in Grades 1–3 and from 15 to 45 seconds in Grades 4–12. Students may not require the entire time allotted. After the response time has ended, the test audio will automatically continue to the next Speaking task.

3.2.4.2 Speaking Test Materials

Speaking test materials include the following items:

- Test Administrator's Script
- Speaking Test Booklet (contains test graphics and prompts)
- Student Test Booklet (contains Speaking test scoring sheet and scoring scale)
- Listening and Speaking test CD (a separate CD for each grade-level cluster and tiered test form). In the rare event that a student requires a human reader as an accommodation, the Recording Script is required to administer the Speaking section.
- A CD player or desktop/laptop computer (to play the CD)
- Speakers

3.2.4.3 Organization and Timing of the Speaking Test

Speaking tasks on the Speaking test are contained within three parts: A, B, and C. As in other domains of ACCESS for ELLs, tasks on the Speaking test are grouped thematically. Each part addresses one or more of the WIDA ELD Standards and contains two tasks. In all, the Speaking test contains six individual tasks across the three parts. Each task is associated with a proficiency

level (1, 3, or 5) and includes one or two questions to which the student responds. Student questions are indicated by a blue speech bubble in the test booklet.

The Speaking test is designed to take approximately 15 to 35 minutes per student, but the actual time will depend on the grade-level cluster and tier of the test administered. Note that the approximate test administration time does not include setting up the test session or explaining test directions. An additional 10 minutes should be allocated to set up the Speaking test.

3.2.5 Test Administrator Training

To prepare individuals to serve as Test Administrators, Test Administrator training for ACCESS Series 502 Paper is conducted through online training modules hosted on the WIDA website. Three certifications are offered to participants: a group test administration certification pertaining to the Listening, Reading, and Writing portions of ACCESS; a certification for the Speaking test; and a certification for Kindergarten ACCESS. To receive any of the three certifications, participants must complete the relevant online course and pass a qualifying exam after completing the course.

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 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 accessibility to the assessments for all ELLs. (Please see the 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 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.

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, 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 Paper and Kindergarten ACCESS to address their individual accessibility needs. Universal tools do not affect the construct being measured on the assessment. Audio aids, color contrast, color overlay, highlighters, colored pencils or crayons, line guide or tracking tool, low-vision aids or magnification devices, sticky notes, and scratch paper are the universal tools used in the ACCESS Paper administration.

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 who have 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. WIDA is planning on studying the efficacy of accommodations.

Accessibility features include tools that are available to all ELLs taking ACCESS for ELLs. Accessibility features are provided to ELLs by Test Administrators for paper-based tests. All accessibility features are available to all ELLs during testing; specific designation is not required prior to testing to make them available to the student. Features available during paper-based test administration include the following:

- Audio amplification device (provided by student)
- Highlighter, colored pencils, or crayons
- Place marker (blank)
- Low-vision aids or magnification device
- Color overlay
- Equipment or technology that the student uses for other tests and schoolwork, e.g., adapted pencil (altered size or grip), slant board, wedge, etc.
- Scratch/blank paper (submit with test or dispose 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 test developers at CAL.

WIDA also 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.

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. For students with visual impairments, the Large Print Test is used, where the font size 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.

4. Scoring Procedures

4.1. Multiple Choice Scoring: Listening and Reading

Listening and Reading items are scored dichotomously, as correct or incorrect. Students mark their answers directly in their test booklets, and each page is scanned into an electronic database. Scale scores for each domain are calculated based on the items that are administered to the test-taker and the number 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 Writing

Trained raters score the performance-based tasks in the domain of Writing. DRC retains many raters from year to year; the return rater rate was approximately 60% in 2021, and, overall, most raters scoring the performance-based tasks 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 holds recruiting events, after which applications for rater positions are screened by DRC’s recruiting staff and likely candidates are personally interviewed by DRC staff. As part of the hiring process, DRC requires 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 gives preference to candidates with previous experience scoring large-scale assessments and degrees emphasizing expertise in English language arts. The rater pool consisted of educators, writers, editors, and other professionals with content-specific backgrounds. While DRC valued these individuals for their content-specific knowledge, they were required to set aside their own biases about student performance and accept the scoring standards outlined in the training for scoring the ACCESS for ELLs.

Prior to scoring live student responses, the raters undergo thorough training and qualifying. Training is task specific to ensure that raters understand the nuances of each unique Writing task. Team leaders, who are selected by DRC based on prior performance as raters and for their leadership skills, are assigned to small groups of raters, typically 7 to 10 raters per team. The team leaders are responsible for monitoring the performance of their team members and providing ongoing feedback to support accurate scoring. DRC promotes scoring directors, who earn their positions by demonstrating quality work as raters and as team leaders on previous projects, from within. Scoring directors are responsible for a specific set of tasks within a single domain. The scoring directors train and oversee the teams of raters assigned to these tasks. What follows are general scoring procedures utilized by DRC.

Rater Training and Qualifying

- DRC assigns each rater a unique ID number and password.
- The scoring director provides detailed directions for use of DRC’s computerized scoring system and remote communication tools.

- The scoring director trains the raters using task-specific anchor sets and training sets.
- Raters must demonstrate scoring proficiency by scoring at least 70% agreement on a qualifying set before scoring live responses.
- Once raters are qualified, DRC provides further training for their grade-level cluster and on the specific tasks for which they will rate responses.
- Once raters have trained, qualified, and begun live scoring, DRC uses calibration sets (of which there are two types, recalibration sets and validation sets, which we explain below) to keep the raters calibrated on the actual tasks they are scoring.

Calculating Score Agreement for Score Monitoring

- DRC’s handscoring system generates handscoring reports, detailing agreement rates for each rater and item. These reports are customized based on input and direction from WIDA. The reports are automatically generated overnight throughout the course of handscoring and may also be run on demand. DRC provides weekly interrater reliability reports to WIDA throughout the handscoring process to ensure that DRC maintains sufficient quality control throughout the course of scoring.
- For Writing, we define **agreement** as two adjacent scores, reported as %AG. (See Section 3.2.3 for a description of the Writing Scoring Scale.) For example, using the Writing Scoring Scale, we consider scores of 2 and 2+ as agreement, as well as scores of 2 and 2 or scores of 2+ and 3. However, we consider scores of 2 and 3 on the Writing Scoring Scale as **adjacent**, while we consider scores of 2 and 3+ as **nonadjacent**.
- Speaking tasks are locally scored for ACCESS Paper, so an explanation for DRC’s rating procedures is not applicable here.
- WIDA stipulates a minimum interrater agreement rate of 70% for Writing.

Routing Responses to Ensure “Blind” Second Ratings

- The DRC scoring system routes and reroutes responses to raters until enough raters perform the prescribed number of ratings for all responses.
- Raters do not see the scores of the other raters and do not know if they are the first or second rater.
- The purpose of the first and second ratings is to monitor interrater reliability by comparing the scores given by two separate raters to the same response. When calculating final scores, the first score given is the score of record.

Monitoring Scoring (Quality Control)

Ongoing quality control checks and procedures help monitor and maintain the quality of the scoring sessions. DRC’s handscoring reports are automatically generated overnight and are also available on demand to monitor progress and maintain handscoring quality control. DRC provides WIDA with access to these reports on a regular basis throughout the scoring process to provide assurance that the quality control metrics meet or exceed expectations.

- During the handscoring process, the scoring directors communicate 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 conduct routine read-behinds to observe, in real time, raters’ performance. Team leaders utilize live, scored responses to provide ongoing feedback and, if necessary, retraining for raters.
- The scoring system randomly selects at least 20% of tasks for two raters to independently score, for the purpose of monitoring interrater reliability. Raters are not aware that another rater may have previously scored a task.
- The DRC system generates interrater reliability reports daily to monitor how often each rater’s scores match other raters’ scores, and scoring leaders continually monitor individual statistics compared to the group average. If the agreement rates for a rater falls below 70%, supervisors increase monitoring and retraining activities with the rater. If the rater fails to demonstrate improved reliability, the rater is released from scoring the item.
- Since the interrater agreement rates were all at or above 70%, the target stipulated by WIDA, the focus turned to raters with lower-than-average agreement rates—even if their agreement was at or above 70%. Even when all agreement rates are at or above 70%, scoring supervisors continue to seek opportunities to increase reliability by providing ongoing feedback and retraining to the raters based on the specific performance of each rater as evidenced by the quality control reports and observations made when reviewing scores given by raters to tasks.
- Responses can be retrieved on demand (e.g., specific grade-level clusters, specific students) should the need arise during or after the scoring process. If needed, responses can be rescored based on task- or response-level information, such as task number, date, score value assigned, or rater ID.
- DRC employs the use of both recalibration sets and validity responses to monitor handscoring quality control. DRC, CAL, and WIDA developed these recalibration sets and validity responses together. CAL developed an initial pool of responses for use as recalibration and validity 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 raters score a task, they take one recalibration set of five responses. The recalibration sets did not differ from rater to rater. For example, a recalibration set was specified for the first day that a rater scored a specific task; every rater who scored that task took this same recalibration set on the first day that they scored that task. After the raters took the recalibration sets, 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 scoring a

task, DRC used validity responses to continue monitoring rater performance. DRC seeded the validity responses into operational scoring so that the raters did not know which responses were operational and which were validity responses. Reports generated daily compared the scores given by each rater to the “true” score for each validity response. When a rater was working on a task, DRC dealt the validity responses to that rater in a random order. Each validity response was dealt to multiple raters over the course of the project (i.e., given enough time, every rater working on a task would score every validity response for that task), but the validity responses were not dealt in the same order to each rater.

Handling Unusual Responses

The following processes were in place to manage specific types of “unusual” responses:

- **Scoring questions.** If raters 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 raters forwarded the response to team leaders for assistance. The team leaders then reviewed the response and applied the proper score. If anything about the response and the rater’s question indicated that the rater needed any clarifications about the scoring guidelines, the team leaders met with raters to review the response and to explain how to score it based on the 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 constitutes a nonscorable response to ensure consistency of nonscorable codes, and this information was provided from CAL to DRC along with other item-specific training materials that were used to train DRC’s raters. During scoring, when raters applied a nonscorable code (except for Blank), the response was automatically forwarded to a handscoring supervisor for review and approval. If the handscoring supervisors had any questions about the application of nonscore codes to specific responses, DRC contacted WIDA and CAL representatives for further review and discussion.
- **Alerts.** To handle possible alert papers (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 raters flagged a response with the alert status, the system automatically routed the response to handscoring supervisors for review. When the handscoring supervisors concurred with the “alert” status of the response, the system then passed the response on 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 would flag the response with a “request original” status. When a rater flagged a response as “request original,” it was automatically forwarded 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 request was forwarded to staff in DRC’s Operations Services, who located the original student response so that it could be reviewed by handscoring supervisors to score the response.

Remote Scoring Procedures due to the COVID-19 Pandemic

Prior to 2020, all WIDA handscoring was conducted in DRC’s handscoring centers. 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 done in the physical scoring locations. The platform, content, and expectations for quality remained the same, and interactive technology and content training and discussions were conducted live (virtually). The differences came with the method through which DRC delivered training (online) and in the modes of communication used (web screen sharing, webcast, video chat, and chat). DRC equipped scoring leaders with a variety of tools to ensure 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 training via Moodle Learning Management System, which mirrors aspects of the scoring room and provides 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, the system required them 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). Printing was disabled for raters in ScoreBoard to protect the security of the student responses, test questions, and training materials. Restrictions built into ScoreBoard defined the hours during the day 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.
- **Content training with Moodle.** DRC provided content training remotely as an interactive, comprehensive, hands-on experience. For Writing training, scoring directors trained groups of raters by screensharing PDFs of training materials. Each training example was viewed individually, with supervisors directing scorers to relevant text. 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-behinds. DRC's handscoring system allowed scoring supervisors to determine specific read-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% agreement and failed to improve after retraining and feedback, DRC removed the rater from the project and assigned the responses to be redealt and rescored.

4.3. Writing Scoring Scale

The Writing Scoring Scale has six whole score points that range from 1 to 6. For responses that fall in between the whole score points, “plus” score points are available (e.g., a response that falls between 3 and 4 is scored as 3+). 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 7 shows the Writing Scoring Scale.

ACCESS for ELLS 2.0 Writing Scoring Scale, Grades 1–12

5+	<p>Score Point 6</p> <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>
4+	<p>Score Point 5</p> <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>
3+	<p>Score Point 4</p> <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>
2+	<p>Score Point 3</p> <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>
1+	<p>Score Point 2</p> <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>
	<p>Score Point 1</p> <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</i> <i>S: Sentence Level</i> <i>W: Word/Phrase Level</i></p>

Figure 7. 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, a whole score point is awarded. If there is clear evidence that one or two descriptors from an adjacent score point are a better fit, the rater awards a plus score point between the two applicable whole score points.

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, as defined below.

Nonscorable: The response is blank; consists only of verbatim copied text; consists only of text that is completely off task; is entirely in a language other than English; or appears to have been plagiarized from an outside source during testing.

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. These responses are scored in their entirety using the scoring scale; however, the maximum holistic score for a completely off-topic response is 2+.

Partially off-task response: The response contains both off-task and on-task writing. These responses are scored 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). These responses are scored in their entirety using the scoring scale.

Both nonscorable and completely off-task responses are scored as 0. Completely off-topic responses receive a maximum score of 2+. Partially off-topic responses are scored in their entirety, while partially off-task responses are scored by ignoring the off-task portion of the response and scoring only the on-task portion.

To calculate a raw score for the Writing test, raters' scores for each Writing task are converted to whole numbers ranging from 0 to 9, as shown in Table 8.

Table 8
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

On Tier A tests, for all grade-level clusters except for Grade 1, the scores from the three tasks are added to calculate a total raw score, which can range from 0 to 27. For the Grade 1 Tier A test, there are four Writing tasks. The first two of these tasks use a modified version of the scoring scale and have score ranges of 0 to 1 and 0 to 3, respectively. The third and fourth tasks use the full scoring scale from 0 to 9; additionally, the last task is weighted as 3. Therefore, the possible final raw scores for Grade 1 Tier A range from 0 to 40.

On Tier B/C tests for all grade-level clusters, results from the different tasks are given different weights. These weights are specified to reflect intended amounts of time that a student should spend on each task. The first task is given a weight of 1, the second task is given a weight of 2, and the third task is given a weight of 3. Thus, for example, a student with raw scores of 5, 6, and 7 on the three tasks would have a total raw score of 38 ($[1 * 5] + [2 * 6] + [3 * 7]$), while a student with raw scores of 7, 6, and 5 on the three tasks would have a total raw score of 34 ($[1 * 7] + [2 * 6] + [3 * 5]$). Raw scores on the Tier B/C tests can range from 0 to 54.

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. The Writing Scoring Scale was designed specifically as a scoring tool and is not appropriate for any other purposes.

4.4. Speaking Scoring Scale

The Speaking test is scored using a scoring scale that is designed to evaluate student responses relative to the model student's response. (See Section 2.2.4 above for more information about the role of the model student in the design of the Speaking tasks.) As part of test administration, the Test Administrators hear the model student response before each student response, which

supports them in assigning an appropriate score relative to the model response. Speaking responses are immediately scored by the administrator while the test is administered. After listening to the student's responses, the administrator assigns a score.

The Speaking Test is the only portion of ACCESS Paper that is scored locally. Test Administrators must complete the relevant virtual ACCESS Paper Test Administrator training module for the Speaking test and pass the accompanying quiz (either Grades 1–5 or Grades 6–12). The training focuses on developing the Test Administrators' ability to score the test reliably. Separate training materials are available that address test administration and monitoring procedures. To help ensure that Test Administrators reliably score the test, they are trained on the Speaking Scoring Scale. Training materials are available for each grade-level cluster, and raters listen to anchor samples and view score justifications that provide detailed explanations for scores based on the scoring scale. Practice samples are also available so that raters can practice assigning scores. The course includes both required training material for each grade-level cluster as well as optional training material. Raters are required to complete training sections for each grade-level cluster they will administer and score. However, if raters will score more than three grade-level clusters, they may complete rater training for only three. The quizzes include 12 Speaking rating tasks in which raters listen to and assign a score to a task response. The pass rate for the quiz is 80% correct.

The Speaking Scoring Scale defines five score points: *Exemplary*, *Strong*, *Adequate*, *Attempted*, and *No Response (in English)*. The *No Response* score point only applies if the examinee refuses to respond, or if the examinee responds in a language other than English.

These score points are applied based on the proficiency level expectations of each task, that is, the level of language proficiency that each task is designed to elicit. These expectations are exemplified by the model student response (see Section 2.2.4). In this way, the model response serves as a scoring benchmark. Raters listen to the model response and score test-taker responses relative to the model. A score of *Exemplary* means that the student response demonstrates English language use that is equal to or beyond the English language use illustrated by the model student's response.

Figure 8 shows the Speaking Scoring Scale.

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

Figure 8. Speaking Scoring Scale.

The Speaking Scoring Scale includes descriptors for overall language use, response sophistication, language delivery, and word choice. As stated above, the scale is applied relative to the proficiency level demands of the task. 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.

To calculate a raw score for the Speaking test, the five score points are converted to whole numbers, as shown in Table 9. To calculate a total raw score, the raw scores for each task are added together; additionally, 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 the B/C test would easily achieve a score of *Adequate and Above* on these tasks. Thus, on the Pre-A test, scores can range from 0 to 6; on the A test, from 0 to 18; and on the B/C test, from 6 to 30.

Table 9

Rating to Raw Score Conversion (Speaking)

Rating	Raw score
No Response (in English)	0
Attempted	1
Adequate/Adequate and Above	2
Strong	3
Exemplary	4

Speaking tasks are scored 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. The Speaking Scoring Scale was designed specifically for test scoring use and is not intended for classroom purposes.

5. Summary of Score Reports

5.1. Individual Student Report

Score reports (district, school, and student level reports) are made available in the WIDA Assessment Management System (AMS) as soon as they are available for each state and printed reports are shipped to school districts and schools at the same time or shortly thereafter. Score reports are available for states to identify students' language performance and properly determine language support for ELLs. Each state and school district determines when and how students individual score reports are provide to students' parents or guardians. Communication about student score reports and resources that districts use to support interpretation is a local decision. WIDA provides resources that schools, districts and states may use to aid in score interpretation. (See below.) How that material is used 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 terms to be challenging, which suggests the need for describing terms with more clarity in score reports. WIDA plans to evaluate current score reports through focus groups to identify how

improvements can be made to help educators, families, and students to better understand score information.

The Individual Student Report (Figure 9) contains detailed information about the performance of a single student within Grades K–12. Its primary users are students, parents/guardians, teachers, and school teams. It describes the language needed to access content and succeed in school, one indicator of a student’s English language proficiency.







ACCESS for ELLs 2.0*
English Language Proficiency Test

Sample Student
 Birth Date: mm/dd/yyyy | Grade: sample grade
 Tier: sample tier
 District ID: XXXXXXXXXXXXXXXXXX | State ID: XXXXXXXXXXXXXXXXXX
 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
Listening 	4.0						368					
Speaking 	2.2						320					
Reading 	3.4						356					
Writing 	3.5						355					
Oral Language 50% Listening + 50% Speaking	3.2						344					
Literacy 50% Reading + 50% Writing	3.5						356					
Comprehension 70% Reading + 30% Listening	3.7						360					
Overall* 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...
Listening	4	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"> • Exchange information and ideas with others • Connect people and events based on oral information • Apply key information about processes or concepts presented orally • Identify positions or points of view on issues in oral discussions
Speaking	2	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"> • Share about what, when, or where something happened • Compare objects, people, pictures, events • Describe steps in cycles or processes • Express opinions
Reading	3	understand written language related to common topics in school and can participate in class discussions, for example: <ul style="list-style-type: none"> • Classify main ideas and examples in written information • Identify main information that tells who, what, when or where something happened • Identify steps in written processes and procedures • Recognize language related to claims and supporting evidence
Writing	3	communicate in writing in English using language related to common topics in school, for example: <ul style="list-style-type: none"> • Describe familiar issues and events • Create stories or short narratives • Describe processes and procedures with some details • Give opinions with reasons in a few short sentences

Figure 9. Individual Student Report.

The score report includes four domain scores (Listening, Speaking, Reading, and Writing) and four composite scores (Oral Language, Literacy, Comprehension, and Overall). Each composite score is represented by a label, a breakdown of how individual domains are used to calculate it, and a visual display of the results. Composition of single domain scores in composite scores is presented in the individual student report. 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 requests (Sahakyan, N., (2020)).

The proficiency level is presented both graphically and as a whole number followed by a decimal. The shaded bar of the graph reflects the exact position of the student's performance on the 6-point English Language Proficiency Scale. The whole number reflects a student's English language proficiency level (1–Entering, 2–Emerging, 3–Developing, 4–Expanding, 5–Bridging, and 6–Reaching) in accord 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 associated confidence band. The confidence band reflects the standard error of measurement of 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 is equal to the 95% probability level.

If a student does not complete one or more of the language domains, NA (not available) is inserted in that language domain as well as in all applicable composite scores, including the overall score. Students with identical overall scores may have very different profiles in terms of their Listening, Speaking, Reading, and Writing.

The second part of the Student Report provides information about the individual student's proficiency levels as whole numbers and describes what students at the reported proficiency level may typically be expected to be able to do in English. For example, if the student received a proficiency level score of 2 for Speaking, the report will include a description of the type of spoken language the student may be expected to be able to produce.

When interpreting scores, the following points should be kept in mind by all stakeholders:

- 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 the individual test.
 - No single 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 from different domains should not be compared. 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 scores of 350 in Speaking at grade 3. For performance comparisons across domains, proficiency levels should be used.
 - Either scale scores or proficiency level scores can be used to compare test scores 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, identical to those 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 attain each proficiency level within a given population.

Part 2:
Technical Results

Contents

1	Annual Test Results	1-1
1.1	Participation.....	1-1
1.1.1	Grade-Level Cluster	1-1
1.1.2	Grade	1-4
1.1.3	Tier	1-8
1.2	Scale Score Results.....	1-12
1.2.1	Mean Scale Score Across Domain and Composite Score by Cluster	1-12
1.2.2	Mean Scale Score Across Domain and Composite Score by Grade	1-17
1.2.3	Correlations	1-27
1.3	Proficiency Level Results	1-31
1.3.1	Domains.....	1-32
1.3.2	Composites	1-48
2	Analysis of Domains.....	2-1
2.1	Complete Item or Task Analysis and Summary	2-5
2.1.1	Listening	2-8
2.1.2	Reading.....	2-21
2.1.3	Writing.....	2-34
2.1.4	Speaking	2-47
2.2	DIF Analysis and Summary.....	2-54
2.2.1	Listening	2-57
2.2.2	Reading.....	2-62
2.2.3	Writing.....	2-66
2.2.4	Speaking	2-70
2.3	Raw Score Distribution.....	2-74
2.3.1	Listening	2-75
2.3.2	Reading.....	2-88
2.3.3	Writing.....	2-101
2.3.4	Speaking	2-114
2.4	Scale Score Distribution	2-127
2.4.1	Listening	2-128
2.4.2	Reading.....	2-147
2.4.3	Writing.....	2-166
2.4.4	Speaking	2-185
2.5	Proficiency Level Distribution	2-204
2.5.1	Listening	2-205

2.5.2	Reading.....	2-224
2.5.3	Writing.....	2-243
2.5.4	Speaking.....	2-262
2.6	Raw Score to Scale Score to Proficiency Level Conversion.....	2-281
2.6.1	Listening.....	2-282
2.6.2	Reading.....	2-289
2.6.3	Writing.....	2-302
2.6.4	Speaking.....	2-317
2.7	Equating and Recalibration Summary.....	2-330
2.8	Test Characteristic Curve.....	2-332
2.8.1	Listening.....	2-334
2.8.2	Reading.....	2-347
2.8.3	Writing.....	2-360
2.8.4	Speaking.....	2-373
2.9	Test Information Function.....	2-386
2.9.1	Listening.....	2-389
2.9.2	Reading.....	2-399
2.9.3	Writing.....	2-409
2.9.4	Speaking.....	2-419
3	Analyses of Composite Scores.....	3-1
3.1	Scale Score Distribution for Composites.....	3-1
3.1.1	Oral.....	3-2
3.1.2	Literacy.....	3-9
3.1.3	Comprehension.....	3-16
3.1.4	Overall.....	3-23
3.2	Proficiency Level Distribution for Composites.....	3-30
3.2.1	Oral.....	3-31
3.2.2	Literacy.....	3-38
3.2.3	Comprehension.....	3-45
3.2.4	Overall.....	3-52
4	Annual Updates of Validity Evidence.....	4-1
4.1	Standards.....	4-2
4.1.1	Test Content.....	4-2
4.1.2	Response Processes.....	4-2
4.1.3	Internal Structure.....	4-2
4.1.4	Relation to Other Variables.....	4-2

5	Reliability.....	5-1
5.1	Reliability of the Domain Scores.....	5-6
5.1.1	Listening.....	5-9
5.1.2	Reading.....	5-11
5.1.3	Writing.....	5-13
5.1.4	Speaking.....	5-15
5.2	Interrater Agreement Rates.....	5-17
5.2.1	Listening.....	5-17
5.2.2	Reading.....	5-17
5.2.3	Writing.....	5-18
5.3	Conditional Standard Errors of Measurement at Cut Score	5-22
5.3.1	Listening.....	5-24
5.3.2	Reading.....	5-28
5.3.3	Writing.....	5-32
5.3.4	Speaking.....	5-36
5.4	Accuracy and Consistency of Domains.....	5-40
5.4.1	Listening.....	5-46
5.4.2	Reading.....	5-47
5.4.3	Writing.....	5-49
5.4.4	Speaking.....	5-50
5.5	Reliabilities of Students' Composite Scale Scores.....	5-52
5.5.1	Oral.....	5-55
5.5.2	Literacy.....	5-59
5.5.3	Comprehension.....	5-63
5.5.4	Overall	5-67
5.6	Conditional Standard Error of Measurement for Composites	5-71
5.6.1	Oral.....	5-73
5.6.2	Literacy.....	5-76
5.6.3	Comprehension.....	5-80
5.6.4	Overall	5-83
5.7	Accuracy and Consistency of Composites	5-87
5.7.1	Oral.....	5-91
5.7.2	Literacy.....	5-93
5.7.3	Comprehension.....	5-95
5.7.4	Overall	5-97
6	Quality Control	6-1
6.1	Content Development Quality Control.....	6-1

6.2 Test Administration Quality Control	6-3
6.3 Rater Quality Control	6-5
6.4 Score Reporting Quality Control	6-6
6.5 Data Forensic Quality Control.....	6-7

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 are labeled 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 19 students were excluded from the analyses due to mismatches in students' tiers across domains. In addition, 8,998 students taking Paper ACCESS tests in Colorado used equated scores to the Online ACCESS tests; therefore, their score analyses were not included in this 502 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 39 WIDA states and U.S. territories that participated in the operational testing program of ACCESS Paper in 2020–2021 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 39 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 61% of the students who take Paper ACCESS. Georgia, Illinois, and North Carolina 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; and MP, Northern Mariana Islands.

Table 1.1.1.1
Participation by Grade-Level Cluster by State, S502 Paper

State	Cluster							Total
	K	1	2	3	4-5	6-8	9-12	
AK	417	1	2	4	22	31	96	573
AL	3,374	.	.	3	3	3	2	3,385
BI	103	33	42	42	52	50	.	322
CO	7,707	276	220	192	289	208	106	8,998
DC	35	35
DD	647	4	6	10	12	6	.	685
DE	1,293	.	.	.	6	1	1	1,301
FL	27,668	28,016	28,519	25,351	43,986	39,964	36,007	229,511
GA	12,743	1,426	1,457	1,329	27	25	16	17,023
HI	1,333	.	5	1	1	3	5	1,348
ID	1,787	4	3	1	6	8	10	1,819
IL	14,053	63	101	94	157	117	29	14,614
IN	7,131	25	24	22	27	21	7	7,257
KY	3,596	4	8	4	5	5	6	3,628
MA	8,322	21	26	40	44	23	12	8,488
MD	819	1	.	1	7	3	3	834
ME	253	.	1	.	1	2	.	257
MI	6,784	93	103	101	180	263	181	7,705
MN	5,819	28	32	33	58	37	22	6,029
MO	3,642	9	8	5	5	9	2	3,680
MP	41	41
MT	153	153
NC	9,229	9	15	15	23	16	9	9,316
ND	337	1	2	3	2	12	6	363
NH	361	27	30	34	40	35	19	546
NJ	7,962	69	67	45	47	20	20	8,230
NM	264	264
NV	4,389	4,389
OK	5,829	12	18	19	55	77	35	6,045
PA	3,229	193	207	128	208	44	55	4,064
RI	1,147	2	.	3	1	5	5	1,163
SC	3,322	8	4	12	22	10	1	3,379
SD	763	31	31	25	60	35	.	945
TN	5,771	8	8	7	5	5	4	5,808
UT	3,562	1	.	.	1	1	2	3,567
VA	5,690	164	123	96	131	42	13	6,259
VT	139	1	1	1	1	1	.	144
WI	3,572	29	42	36	36	41	11	3,767
WY	271	4	.	4	8	9	15	311
Total	163,557	30,563	31,105	27,661	45,528	41,132	36,700	376,246

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 46% female and 51% male in Clusters 1–3 and 44% female and 52% male in Clusters 4–12. The Hispanic ethnicity percentage was about 76% in all clusters except Kindergarten, which was 46%.

Table 1.1.1.2
Participation by Grade-Level Cluster by Gender, S502 Paper

Cluster		Gender			Total
		F	M	Missing	
K	Count	75,984	85,163	2,410	163,557
	% within Cluster	46.5%	52.1%	1.5%	100.0%
1	Count	14,310	16,200	53	30,563
	% within Cluster	46.8%	53.0%	0.2%	100.0%
2	Count	14,395	16,631	79	31,105
	% within Cluster	46.3%	53.5%	0.3%	100.0%
3	Count	12,538	15,046	77	27,661
	% within Cluster	45.3%	54.4%	0.3%	100.0%
4–5	Count	20,754	24,661	113	45,528
	% within Cluster	45.6%	54.2%	0.3%	100.0%
6–8	Count	18,631	22,376	125	41,132
	% within Cluster	45.3%	54.4%	0.3%	100.0%
9–12	Count	16,858	19,726	116	36,700
	% within Cluster	45.9%	53.8%	0.3%	100.0%
Total	Count	173,470	199,803	2,973	376,246
	% within Cluster	46.1%	53.1%	0.8%	100%

Table 1.1.1.3

Participation by Grade-Level Cluster by Ethnicity, S502 Paper

Cluster		Ethnicity			Total
		Hispanic	Non-Hispanic	Unknown	
K	Count	107,593	46,013	9,951	163,557
	% within Cluster	65.8%	28.1%	6.1%	100.0%
1	Count	24,173	6,099	291	30,563
	% within Cluster	79.1%	20.0%	1.0%	100.0%
2	Count	24,502	6,278	325	31,105
	% within Cluster	78.8%	20.2%	1.0%	100.0%
3	Count	21,868	5,541	252	27,661
	% within Cluster	79.1%	20.0%	0.9%	100.0%
4–5	Count	36,184	8,922	422	45,528
	% within Cluster	79.5%	19.6%	0.9%	100.0%
6–8	Count	32,759	8,072	301	41,132
	% within Cluster	79.6%	19.6%	0.7%	100.0%
9–12	Count	28,587	7,708	405	36,700
	% within Cluster	77.9%	21.0%	1.1%	100.0%
Total	Count	275,666	88,633	11,947	376,246
	% within Cluster	73.3%	23.6%	3.2%	100.0%

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 45% of the Paper ACCESS population.

Table 1.1.2.1
Participation by Grade by State, S502 Paper

State	Grade												Total	
	K	1	2	3	4	5	6	7	8	9	10	11		12
AK	417	1	2	4	13	9	15	9	7	23	26	21	26	573
AL	3,374	.	.	3	1	2	1	1	1	1	.	1	.	3,385
BI	103	33	42	42	24	28	23	16	11	322
CO	7,707	276	220	192	159	130	65	81	62	12	29	37	28	8,998
DC	35	35
DD	647	4	6	10	8	4	6	685
DE	1,293	.	.	.	4	2	1	1	1,301
FL	27,668	28,016	28,519	25,351	24,890	19,096	14,635	13,393	11,936	11,046	10,138	8,798	6,025	229,511
GA	12,743	1,426	1,457	1,329	19	8	10	8	7	5	5	2	4	17,023
HI	1,333	.	5	1	.	1	1	.	2	2	1	1	1	1,348
ID	1,787	4	3	1	4	2	2	4	2	8	.	2	.	1,819
IL	14,053	63	101	94	83	74	42	47	28	9	7	9	4	14,614
IN	7,131	25	24	22	16	11	7	9	5	1	.	4	2	7,257
KY	3,596	4	8	4	3	2	3	1	1	3	3	.	.	3,628
MA	8,322	21	26	40	24	20	10	8	5	3	3	4	2	8,488
MD	819	1	.	1	4	3	1	.	2	2	1	.	.	834
ME	253	.	1	.	1	.	.	.	2	257
MI	6,784	93	103	101	95	85	94	80	89	44	56	44	37	7,705
MN	5,819	28	32	33	37	21	16	12	9	10	3	4	5	6,029
MO	3,642	9	8	5	3	2	2	6	1	.	.	1	1	3,680
MP	41	41
MT	153	153
NC	9,229	9	15	15	17	6	6	5	5	4	3	1	1	9,316
ND	337	1	2	3	1	1	4	6	2	1	3	2	.	363
NH	361	27	30	34	16	24	9	14	12	8	4	4	3	546
NJ	7,962	69	67	45	29	18	7	3	10	7	5	5	3	8,230
NM	264	264
NV	4,389	4,389
OK	5,829	12	18	19	35	20	18	36	23	14	11	7	3	6,045
PA	3,229	193	207	128	106	102	17	13	14	17	11	14	13	4,064
RI	1,147	2	.	3	1	.	.	2	3	.	5	.	.	1,163
SC	3,322	8	4	12	9	13	5	5	.	1	.	.	.	3,379
SD	763	31	31	25	34	26	10	17	8	945
TN	5,771	8	8	7	2	3	3	1	1	1	1	1	1	5,808
UT	3,562	1	.	.	1	.	.	.	1	.	.	1	1	3,567
VA	5,690	164	123	96	91	40	21	10	11	2	5	2	4	6,259
VT	139	1	1	1	1	.	.	1	144
WI	3,572	29	42	36	19	17	19	12	10	4	1	5	1	3,767
WY	271	4	.	4	5	3	1	3	5	4	5	2	4	311
Total	163,557	30,563	31,105	27,661	25,755	19,773	15,054	13,803	12,275	11,232	10,326	8,972	6,170	376,246

Table 1.1.2.2

Participation by Grade by Gender, S502 Paper

Grade		Gender			Total
		F	M	Missing	
K	Count	75,984	85,163	2,410	163,557
	% within Grade	46.46%	52.07%	1.47%	100.0%
1	Count	14,310	16,200	53	30,563
	% within Grade	46.82%	53.01%	0.17%	100.0%
2	Count	14,395	16,631	79	31,105
	% within Grade	46.28%	53.47%	0.25%	100.0%
3	Count	12,538	15,046	77	27,661
	% within Grade	45.33%	54.39%	0.28%	100.0%
4	Count	11,634	14,047	74	25,755
	% within Grade	45.17%	54.54%	0.29%	100.0%
5	Count	9,120	10,614	39	19,773
	% within Grade	46.12%	53.68%	0.20%	100.0%
6	Count	6,826	8,149	79	15,054
	% within Grade	45.34%	54.13%	0.52%	100.0%
7	Count	6,247	7,535	21	13,803
	% within Grade	45.26%	54.59%	0.15%	100.0%
8	Count	5,558	6,692	25	12,275
	% within Grade	45.28%	54.52%	0.20%	100.0%
9	Count	5,109	6,040	83	11,232
	% within Grade	45.49%	53.77%	0.74%	100.0%
10	Count	4,703	5,611	12	10,326
	% within Grade	45.55%	54.34%	0.12%	100.0%
11	Count	4,149	4,808	15	8,972
	% within Grade	46.24%	53.59%	0.17%	100.0%
12	Count	2,897	3,267	6	6,170
	% within Grade	46.95	52.95	0.10	100.0
Total	Count	173,470	199,803	2,973	376,246
	% within Grade	46.11%	53.10%	0.79%	100.0%

Table 1.1.2.3

Participation by Grade by Ethnicity, S502 Paper

Grade		Ethnicity			Total
		Hispanic	Non-Hispanic	Unknown	
K	Count	107,593	46,013	9,951	163,557
	% within Grade	65.78%	28.13%	6.08%	100.0%
1	Count	24,173	6,099	291	30,563
	% within Grade	79.09%	19.96%	0.95%	100.0%
2	Count	24,502	6,278	325	31,105
	% within Grade	78.77%	20.18%	1.04%	100.0%
3	Count	21,868	5,541	252	27,661
	% within Grade	79.06%	20.03%	0.91%	100.0%
4	Count	20,499	5,009	247	25,755
	% within Grade	79.59%	19.45%	0.96%	100.0%
5	Count	15,685	3,913	175	19,773
	% within Grade	79.33%	19.79%	0.89%	100.0%
6	Count	11,979	2,924	151	15,054
	% within Grade	79.57%	19.42%	1.00%	100.0%
7	Count	10,952	2,768	83	13,803
	% within Grade	79.35%	20.05%	0.60%	100.0%
8	Count	9,828	2,380	67	12,275
	% within Grade	80.07%	19.39%	0.55%	100.0%
9	Count	8,863	2,170	199	11,232
	% within Grade	78.91%	19.32%	1.77%	100.0%
10	Count	8,117	2,137	72	10,326
	% within Grade	78.61	20.70	0.70	100.0
11	Count	7,000	1,898	74	8,972
	% within Grade	78.02%	21.15%	0.82%	100.0%
12	Count	4,607	1,503	60	6,170
	% within Grade	74.67%	24.36%	0.97%	100.0%
Total	Count	275,666	88,633	11,947	376,246
	% within Grade	73.27%	23.56%	3.18%	100.0%

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, 49% of students were in Tier A and 51% in Tier B/C. In Grade 2, 23% of students were in Tier A and 77% in Tier B/C. In Grade 3 and Grades 4–5, 20% were in Tier A and 80% in Tier B/C. In Grades 6–8 and 9–12, there were about 25% of students in Tier A and 75% 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, S502 Paper

Cluster			Domain			
			Listening	Reading	Speaking	Writing
K	Tier	-	163,540	163,536	163,510	163,536
1	Tier	A	15,251	15,255	15,255	15,253
		B	15,299	15,304	15,300	15,306
	Total		30,550	30,559	30,555	30,559
2	Tier	A	7,231	7,233	7,233	7,234
		B	23,864	23,866	23,865	23,867
	Total		31,095	31,099	31,098	31,101
3	Tier	A	4,973	4,974	4,973	4,974
		B	22,679	22,680	22,681	22,683
	Total		27,652	27,654	27,654	27,657
4–5	Tier	A	6,693	6,693	6,693	6,693
		B	38,833	38,834	38,831	38,832
	Total		45,526	45,527	45,524	45,525
6–8	Tier	A	9,000	9,001	9,000	9,001
		B	32,125	32,125	32,128	32,127
	Total		41,125	41,126	41,128	41,128
9–12	Tier	A	8,575	8,576	8,572	8,576
		B	28,120	28,119	28,113	28,118
	Total		36,695	36,695	36,685	36,694

Table 1.1.3.2
Participation by Grade by Tier by Domain, S502 Paper

Cluster			Domain			
			Listening	Reading	Speaking	Writing
K	Tier	-	163,540	163,536	163,510	163,536
1	Tier	A	15,251	15,255	15,255	15,253
		B	15,299	15,304	15,300	15,306
	Total		30,550	30,559	30,555	30,559
2	Tier	A	7,231	7,233	7,233	7,234
		B	23,864	23,866	23,865	23,867
	Total		31,095	31,099	31,098	31,101
3	Tier	A	4,973	4,974	4,973	4,974
		B	22,679	22,680	22,681	22,683
	Total		27,652	27,654	27,654	27,657
4	Tier	A	3,564	3,564	3,564	3,564
		B	22,189	22,190	22,188	22,189
	Total		25,753	25,754	25,752	25,753
5	Tier	A	3,129	3,129	3,129	3,129
		B	16,644	16,644	16,643	16,643
	Total		19,773	19,773	19,772	19,772
6	Tier	A	3,088	3,089	3,088	3,089
		B	11,962	11,961	11,963	11,963
	Total		15,050	15,050	15,051	15,052
7	Tier	A	2,928	2,928	2,928	2,928
		B	10,874	10,874	10,875	10,874
	Total		13,802	13,802	13,803	13,802
8	Tier	A	2,984	2,984	2,984	2,984
		B	9,289	9,290	9,290	9,290
	Total		12,273	12,274	12,274	12,274
9	Tier	A	2,809	2,809	2,809	2,809
		B	8,420	8,420	8,420	8,420
	Total		11,229	11,229	11,229	11,229
10	Tier	A	2,545	2,546	2,546	2,546
		B	7,780	7,779	7,779	7,779
	Total		10,325	10,325	10,325	10,325
11	Tier	A	1,975	1,975	1,972	1,975
		B	6,997	6,997	6,992	6,996
	Total		8,972	8,972	8,964	8,971
12	Tier	A	1,246	1,246	1,245	1,246
		B	4,923	4,923	4,922	4,923
	Total		6,169	6,169	6,167	6,169

Table 1.1.3.3

Participation by Grade-Level Cluster by Tier by Gender, S502 Paper

Cluster	Tier		Gender			Total
			F	M	Missing	
K	-	Count	75,984	85,163	2,410	163,557
		% within Tier	46.5%	52.1%	1.5%	100.0%
1	A	Count	6,962	8,259	35	15,256
		% within Tier	45.6%	54.1%	0.2%	100.0%
	BC	Count	7,348	7,941	18	15,307
		% within Tier	48.0%	51.9%	0.1%	100.0%
2	A	Count	3,105	4,102	28	7,235
		% within Tier	42.9%	56.7%	0.4%	100.0%
	BC	Count	11,290	12,529	51	23,870
		% within Tier	47.3%	52.5%	0.2%	100.0%
3	A	Count	2,144	2,807	24	4,975
		% within Tier	43.1%	56.4%	0.5%	100.0%
	BC	Count	10,394	12,239	53	22,686
		% within Tier	45.8%	54.0%	0.2%	100.0%
4-5	A	Count	3,046	3,612	35	6,693
		% within Tier	45.5%	54.0%	0.5%	100.0%
	BC	Count	17,708	21,049	78	38,835
		% within Tier	45.6%	54.2%	0.2%	100.0%
6-8	A	Count	4,021	4,950	31	9,002
		% within Tier	44.7%	55.0%	0.3%	100.0%
	BC	Count	14,610	17,426	94	32,130
		% within Tier	45.5%	54.2%	0.3%	100.0%
9-12	A	Count	3,935	4,602	39	8,576
		% within Tier	45.9%	53.7%	0.5%	100.0%
	BC	Count	12,923	15,124	77	28,124
		% within Tier	46.0%	53.8%	0.3%	100.0%

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 4% to 5% 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, S502 Paper

Cluster	Tier		Ethnicity			Total
			Hispanic	Other	Unknown	
K	-	Count	107,593	46,013	9,951	163,557
		% within Tier	65.8%	28.1%	6.1%	100.0%
1	A	Count	12,306	2,744	206	15,256
		% within Tier	80.7%	18.0%	1.4%	100.0%
	BC	Count	11,867	3,355	85	15,307
		% within Tier	77.5%	21.9%	0.6%	100.0%
2	A	Count	5,752	1,364	119	7,235
		% within Tier	79.5%	18.9%	1.6%	100.0%
	BC	Count	18,750	4,914	206	23,870
		% within Tier	78.6%	20.6%	0.9%	100.0%
3	A	Count	4,020	879	76	4,975
		% within Tier	80.8%	17.7%	1.5%	100.0%
	BC	Count	17,848	4,662	176	22,686
		% within Tier	78.7%	20.6%	0.8%	100.0%
4-5	A	Count	5,549	1,045	99	6,693
		% within Tier	82.9%	15.6%	1.5%	100.0%
	BC	Count	30,635	7,877	323	38,835
		% within Tier	78.9%	20.3%	0.8%	100.0%
6-8	A	Count	7,495	1,421	86	9,002
		% within Tier	83.3%	15.8%	1.0%	100.0%
	BC	Count	25,264	6,651	215	32,130
		% within Tier	78.6%	20.7%	0.7%	100.0%
9-12	A	Count	6,953	1,490	133	8,576
		% within Tier	81.1%	17.4%	1.6%	100.0%
	BC	Count	21,634	6,218	272	28,124
		% within Tier	76.9%	22.1%	1.0%	100.0%

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 Listening, Reading, Writing, and then Speaking. Clusters 6–8 and 9–12 showed the highest average scale scores in all single domains across all clusters.

Table 1.2.1.1

Mean Scale Scores by Grade-Level Cluster, S502 Paper

Cluster		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
K	Mean	257.35	173.92	183.31	261.36	259.62	178.83	198.94	202.85
	N	155,560	155,552	155,550	155,526	155,524	155,549	155,549	155,512
1	Mean	304.97	284.39	242.69	269.7	289.41	264.47	291.07	272.66
	N	25,440	23,493	30,277	30,100	25,299	23,491	20,703	20,595
2	Mean	332.2	308.7	276.4	287.6	310.8	293.4	316.1	299
	N	28,612	25,902	30,872	30,686	28,446	25,898	24,478	24,343
3	Mean	354.77	333.03	295.16	299.32	327.7	314.55	339.61	318.49
	N	25,871	22,806	27,457	27,299	25,723	22,799	21,869	21,742
4–5	Mean	376.54	350.91	330.79	338.93	358.45	341.58	358.87	346.82
	N	43,615	39,648	45,231	45,039	43,428	39,644	38,588	38,416
6–8	Mean	381.03	358.71	330.43	354.68	368.73	345.36	365.78	352.69
	N	39,157	36,862	40,903	40,610	38,870	36,857	35,673	35,418
9–12	Mean	382.83	383.66	358.07	355.39	370.17	371.51	383.87	371.43
	N	34,871	32,808	36,568	36,179	34,491	32,796	31,665	31,312

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, S502 Paper

Cluster	Gender		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
K	F	Mean	264.19	174.88	187.69	271.3	268	181.51	201.66	207.23
		N	72,127	72,125	72,124	72,107	72,105	72,124	72,123	72,100
	M	Mean	251.54	173.39	179.8	252.78	252.42	176.8	196.83	199.28
		N	81,035	81,029	81,029	81,022	81,022	81,028	81,028	81,015
	Missing	Mean	248.01	163.04	169.98	252.43	250.47	166.7	188.52	191.63
		N	2,398	2,398	2,397	2,397	2,397	2,397	2,398	2,397
1	F	Mean	307.51	284.56	247.84	274.36	292.85	266.88	291.8	275.19
		N	12,192	11,181	14,187	14,087	12,109	11,180	9,991	9,931
	M	Mean	302.67	284.28	238.18	265.66	286.32	262.3	290.44	270.34
		N	13,202	12,274	16,037	15,960	13,144	12,273	10,677	10,629
	Missing	Mean	288.78	272.89	228.75	247.23	267.7	253.79	277.77	257.4
		N	46	38	53	53	46	38	35	35
2	F	Mean	335.08	310.13	283.29	290.63	313.79	297.53	317.94	302.72
		N	13,404	12,122	14,302	14,215	13,322	12,121	11,548	11,487
	M	Mean	329.66	307.59	270.49	285.1	308.36	289.84	314.65	295.81
		N	15,135	13,718	16,491	16,392	15,051	13,715	12,870	12,796
	Missing	Mean	329	300	268	260	295	285	310	289
		N	73	62	79	79	73	62	60	60
3	F	Mean	354.66	333.37	302.56	301.18	328.42	318.34	339.72	321.17
		N	11,830	10,488	12,448	12,374	11,762	10,486	10,110	10,054
	M	Mean	354.92	332.77	289.05	297.91	327.18	311.35	339.57	316.24
		N	13,968	12,251	14,932	14,850	13,889	12,246	11,694	11,624
	Missing	Mean	345.58	325.34	285.3	270.27	309.29	304.94	331.74	306.7
		N	73	67	77	75	72	67	65	64
4-5	F	Mean	376.59	351.24	337.17	340.9	359.48	344.92	359.12	349.45
		N	19,935	18,226	20,622	20,526	19,841	18,225	17,777	17,694
	M	Mean	376.55	350.7	325.49	337.43	357.67	338.81	358.71	344.64
		N	23,574	21,321	24,496	24,401	23,482	21,318	20,716	20,628
	Missing	Mean	364.41	338.19	313.79	307.48	338.31	326.24	346.83	329.99
		N	106	101	113	112	105	101	95	94
6-8	F	Mean	381.39	360.65	336.78	355.87	369.51	349.5	367.25	355.8
		N	17,791	16,903	18,536	18,406	17,665	16,901	16,392	16,277
	M	Mean	380.75	357.07	325.17	353.71	368.11	341.85	364.56	350.06
		N	21,245	19,849	22,242	22,080	21,085	19,846	19,174	19,035
	Missing	Mean	376.01	356.12	324.52	351.29	363.82	341.77	361.59	347.7
		N	121	110	125	124	120	110	107	106
9-12	F	Mean	382.02	385.98	363.27	354.79	369.45	375.23	385.24	373.82
		N	16,106	15,263	16,805	16,612	15,914	15,259	14,780	14,600
	M	Mean	383.56	381.65	353.63	355.96	370.82	368.25	382.68	369.34
		N	18,658	17,439	19,647	19,451	18,470	17,431	16,784	16,611
	Missing	Mean	377.09	381.46	357.53	345.57	364.1	371.02	381.13	369.81
		N	107	106	116	116	107	106	101	101

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

Table 1.2.1.3

Mean Scale Scores by Grade-Level Cluster by Ethnicity, S502 Paper

Cluster	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
K	Non-Hispanic Asian	Mean	281.46	218.00	225.76	287.50	284.75	222.13	237.02	240.68
		N	18,340	18,340	18,339	18,336	18,336	18,339	18,340	18,336
	Non-Hispanic Pacific Islander	Mean	245.58	147.64	159.18	249.98	248.04	153.59	177.03	181.73
		N	939	939	939	939	939	939	939	939
	Non-Hispanic Black	Mean	267.09	183.14	187.21	282.19	274.90	185.39	208.32	212.03
		N	8,623	8,624	8,624	8,623	8,622	8,624	8,623	8,622
	Hispanic (of Any Race)	Mean	250.37	163.84	173.45	252.83	251.86	168.85	189.80	193.54
		N	101,604	101,595	101,594	101,582	101,582	101,594	101,594	101,571
	Non-Hispanic American Indian	Mean	261.74	167.51	176.04	269.38	265.85	171.97	195.76	199.93
		N	660	660	660	660	660	660	660	660
	Non-Hispanic Multiracial	Mean	284.54	199.88	203.54	289.80	287.44	201.94	225.26	227.36
		N	844	844	844	844	844	844	844	844
	Non-Hispanic White	Mean	273.43	190.18	202.95	280.46	277.21	196.80	215.14	220.71
		N	14,642	14,642	14,643	14,635	14,634	14,642	14,641	14,633
	Unknown	Mean	250.55	164.33	174.44	252.20	251.63	169.58	190.18	194.00
		N	9,908	9,908	9,907	9,907	9,907	9,907	9,908	9,907
1	Non-Hispanic Asian	Mean	310.53	296.37	258.14	278.60	296.57	277.82	301.61	284.31
		N	1,108	1,038	1,292	1,288	1,105	1,038	930	927
	Non-Hispanic Pacific Islander	Mean	316.47	277.53	243.75	277.33	310.87	262.63	292.92	287.23
		N	15	19	24	24	15	19	13	13
	Non-Hispanic Black	Mean	299.65	284.37	233.22	271.12	288.10	260.24	289.50	269.97
		N	1,848	1,747	2,365	2,343	1,832	1,746	1,462	1,450
	Hispanic (of Any Race)	Mean	304.92	283.42	242.20	268.10	288.52	263.65	290.36	271.73
		N	20,221	18,610	23,949	23,809	20,108	18,609	16,452	16,368
	Non-Hispanic American Indian	Mean	296.16	284.04	239.24	266.34	280.22	262.82	288.63	269.00
		N	97	84	108	108	97	84	76	76
	Non-Hispanic Multiracial	Mean	317.84	290.60	251.11	278.98	303.40	272.42	301.05	284.07
		N	131	133	168	167	130	133	109	108
	Non-Hispanic White	Mean	309.85	288.63	250.75	281.49	298.04	270.88	295.33	279.66
		N	1,747	1,603	2,080	2,070	1,739	1,603	1,411	1,403
	Unknown	Mean	286.71	277.68	230.09	261.48	275.47	254.74	280.32	261.55
		N	273	259	291	291	273	259	250	250

Cluster	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
2	Non-Hispanic Asian	Mean	336.92	323.45	292.61	295.61	317.56	308.95	328.12	312.21
		N	1,221	1,144	1,313	1,307	1,215	1,144	1,084	1,080
	Non-Hispanic Pacific Islander	Mean	332.50	312.26	292.48	290.28	318.08	303.21	319.06	308.78
		N	26	19	29	29	26	19	18	18
	Non-Hispanic Black	Mean	327.10	304.94	266.42	286.73	308.15	286.26	311.94	293.29
		N	2,264	2,033	2,569	2,553	2,253	2,032	1,854	1,845
	Hispanic (of Any Race)	Mean	332.01	307.55	275.88	286.34	310.06	292.50	315.23	298.08
		N	22,622	20,454	24,310	24,164	22,491	20,451	19,387	19,281
	Non-Hispanic American Indian	Mean	331.77	300.16	268.43	273.65	304.83	283.61	310.46	291.95
		N	104	85	113	112	103	85	79	78
	Non-Hispanic Multiracial	Mean	337.23	319.06	282.10	298.27	320.00	303.61	326.47	310.72
		N	171	156	186	182	167	156	145	141
	Non-Hispanic White	Mean	338.77	318.26	285.73	299.88	320.54	303.08	325.05	308.92
		N	1,891	1,720	2,027	2,016	1,880	1,720	1,624	1,614
Unknown	Mean	321.88	303.41	269.92	278.30	301.32	287.98	309.49	292.43	
	N	313	291	325	323	311	291	287	286	
3	Non-Hispanic Asian	Mean	364.61	344.76	309.52	309.86	337.66	327.44	350.73	330.46
		N	1,051	923	1,113	1,108	1,046	923	887	884
	Non-Hispanic Pacific Islander	Mean	351.18	333.48	290.28	298.56	321.82	311.43	336.63	310.47
		N	28	21	32	32	28	21	19	19
	Non-Hispanic Black	Mean	352.98	331.64	289.23	301.71	328.22	311.23	338.27	316.63
		N	2,256	1,956	2,463	2,442	2,240	1,955	1,835	1,821
	Hispanic (of Any Race)	Mean	354.05	332.30	294.48	297.35	326.35	313.80	338.85	317.53
		N	20,509	18,090	21,707	21,587	20,393	18,084	17,378	17,278
	Non-Hispanic American Indian	Mean	347.52	330.16	291.28	291.33	319.54	310.94	335.21	313.00
		N	122	113	130	130	122	113	106	106
	Non-Hispanic Multiracial	Mean	367.04	340.81	305.94	317.67	342.57	322.35	348.56	328.14
		N	134	121	139	138	133	121	117	116
	Non-Hispanic White	Mean	360.48	337.37	303.38	314.55	338.10	320.90	344.53	326.07
		N	1,531	1,352	1,621	1,613	1,524	1,352	1,303	1,297
Unknown	Mean	350.80	326.88	292.58	294.80	324.69	310.36	334.56	315.17	
	N	240	230	252	249	237	230	224	221	

Cluster	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
4-5	Non-Hispanic Asian	Mean	388.49	364.96	341.89	349.81	369.49	354.23	372.33	359.08
		N	1,252	1,164	1,300	1,293	1,247	1,164	1,135	1,131
	Non-Hispanic Pacific Islander	Mean	383.14	359.91	332.94	340.60	364.80	348.56	367.43	353.74
		N	49	43	52	52	49	43	42	42
	Non-Hispanic Black	Mean	375.64	349.15	327.46	342.62	359.65	339.27	357.55	345.73
		N	4,244	3,745	4,454	4,430	4,220	3,744	3,619	3,595
	Hispanic (of Any Race)	Mean	375.78	350.22	330.50	337.33	357.28	341.06	358.12	346.05
		N	34,720	31,575	35,933	35,796	34,584	31,572	30,773	30,650
	Non-Hispanic American Indian	Mean	371.06	347.51	328.65	326.94	349.51	338.19	354.52	341.01
		N	218	200	220	220	218	200	198	198
	Non-Hispanic Multiracial	Mean	379.12	353.68	333.23	345.03	362.70	344.21	360.97	349.56
		N	162	154	168	168	162	154	150	150
	Non-Hispanic White	Mean	383.84	357.09	336.45	351.75	368.52	347.46	365.49	354.16
		N	2,572	2,383	2,682	2,660	2,552	2,383	2,304	2,285
Unknown	Mean	368.60	344.06	319.68	325.92	350.67	332.56	352.20	339.35	
	N	398	384	422	420	396	384	367	365	
6-8	Non-Hispanic Asian	Mean	390.41	368.42	340.91	360.95	376.65	355.79	375.58	362.55
		N	1,127	1,074	1,182	1,169	1,116	1,074	1,040	1,029
	Non-Hispanic Pacific Islander	Mean	385.24	362.70	333.33	360.64	373.12	349.45	369.02	356.40
		N	59	53	60	59	58	53	52	52
	Non-Hispanic Black	Mean	385.23	358.60	327.20	361.84	374.17	343.81	367.01	353.24
		N	3,595	3,310	3,832	3,791	3,557	3,309	3,160	3,134
	Hispanic (of Any Race)	Mean	379.57	357.89	330.03	352.67	367.00	344.70	364.76	351.70
		N	31,279	29,477	32,563	32,349	31,066	29,474	28,592	28,396
	Non-Hispanic American Indian	Mean	374.49	353.13	320.79	344.94	360.52	337.44	360.14	344.87
		N	157	149	163	163	157	149	144	144
	Non-Hispanic Multiracial	Mean	390.01	366.73	336.71	363.27	378.19	353.00	374.68	361.75
		N	166	161	177	176	165	161	152	151
	Non-Hispanic White	Mean	390.33	364.92	337.21	368.98	380.88	351.92	373.00	361.08
		N	2,491	2,374	2,625	2,606	2,471	2,373	2,282	2,264
Unknown	Mean	367.04	353.43	314.72	331.49	350.73	335.41	357.90	340.46	
	N	283	264	301	297	280	264	251	248	

Cluster	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
9–12	Non-Hispanic Asian	Mean	394.79	392.04	369.76	366.52	381.35	381.87	393.32	381.94
		N	1,168	1,102	1,207	1,191	1,154	1,101	1,079	1,064
	Non-Hispanic Pacific Islander	Mean	383.15	379.13	361.49	357.97	371.71	370.48	379.46	369.49
		N	66	61	69	69	66	61	59	59
	Non-Hispanic Black	Mean	380.44	380.53	351.79	362.16	372.03	366.91	380.93	368.72
		N	3,742	3,352	3,998	3,951	3,692	3,350	3,184	3,142
	Hispanic (of Any Race)	Mean	381.87	383.20	357.98	352.98	368.54	371.18	383.23	370.70
		N	27,198	25,739	28,483	28,186	26,913	25,732	24,864	24,596
	Non-Hispanic American Indian	Mean	373.12	373.52	346.19	331.51	352.70	359.73	373.88	357.36
		N	102	94	106	106	102	94	90	90
	Non-Hispanic Multiracial	Mean	396.86	391.77	370.09	380.69	389.21	382.11	394.30	385.23
		N	148	132	152	151	147	132	129	129
	Non-Hispanic White	Mean	393.56	391.11	364.92	369.86	382.73	378.42	392.37	380.18
		N	2,071	1,967	2,148	2,124	2,045	1,965	1,916	1,892
	Unknown	Mean	376.31	379.97	353.28	345.08	362.43	367.90	379.39	366.80
		N	376	361	405	401	372	361	344	340

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. Listening domain peaked at Grade 8. Reading and Writing domains had the highest mean scale scores in Grade 11. Speaking had the highest mean scale score in Grade 12. 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 in the Listening domain.

Table 1.2.2.1
Mean Scale Scores by Grade, S502 Paper

Grade		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
K	Mean	257.35	173.92	183.31	261.36	259.62	178.83	198.94	202.85
	N	155,560	155,552	155,550	155,526	155,524	155,549	155,549	155,512
1	Mean	304.97	284.39	242.69	269.70	289.41	264.47	291.07	272.66
	N	25,440	23,493	30,277	30,100	25,299	23,491	20,703	20,595
2	Mean	332.20	308.76	276.41	287.60	310.87	293.43	316.19	299.06
	N	28,612	25,902	30,872	30,686	28,446	25,898	24,478	24,343
3	Mean	354.77	333.03	295.16	299.32	327.70	314.55	339.61	318.49
	N	25,871	22,806	27,457	27,299	25,723	22,799	21,869	21,742
4	Mean	372.08	346.70	324.99	335.93	354.75	336.51	354.49	342.05
	N	24,580	22,113	25,591	25,461	24,452	22,109	21,457	21,338
5	Mean	382.30	356.24	338.34	342.83	363.22	347.98	364.37	352.78
	N	19,035	17,535	19,640	19,578	18,976	17,535	17,131	17,078
6	Mean	376.58	352.89	325.76	352.10	365.09	340.08	360.19	347.80
	N	14,343	13,266	14,982	14,872	14,236	13,265	12,829	12,731
7	Mean	382.40	359.33	331.54	355.26	369.77	346.31	366.70	353.69
	N	13,122	12,344	13,714	13,610	13,020	12,341	11,954	11,868
8	Mean	384.95	364.89	334.91	357.20	372.04	350.54	371.37	357.35
	N	11,692	11,252	12,207	12,128	11,614	11,251	10,890	10,819
9	Mean	380.78	379.83	354.06	354.70	368.93	367.61	380.54	368.34
	N	10,616	9,838	11,208	11,096	10,512	9,837	9,450	9,355
10	Mean	381.26	382.75	356.76	350.68	366.96	370.39	382.76	369.62
	N	9,836	9,255	10,291	10,199	9,745	9,250	8,948	8,860
11	Mean	386.17	387.72	361.68	358.42	373.34	375.25	387.66	375.00
	N	8,540	8,113	8,931	8,823	8,433	8,109	7,846	7,745
12	Mean	384.28	386.03	362.34	360.17	373.17	374.77	386.01	374.66
	N	5,879	5,602	6,138	6,061	5,801	5,600	5,421	5,352

Table 1.2.2.2
Mean Scale Scores by Grade by Gender, S502 Paper

Grade	Gender		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
K	F	Mean	264.19	174.88	187.69	271.30	268.00	181.51	201.66	207.23
		N	72,127	72,125	72,124	72,107	72,105	72,124	72,123	72,100
	M	Mean	251.54	173.39	179.80	252.78	252.42	176.80	196.83	199.28
		N	81,035	81,029	81,029	81,022	81,022	81,028	81,028	81,015
	Missing	Mean	248.01	163.04	169.98	252.43	250.47	166.70	188.52	191.63
		N	2,398	2,398	2,397	2,397	2,397	2,397	2,398	2,397
1	F	Mean	307.51	284.56	247.84	274.36	292.85	266.88	291.80	275.19
		N	12,192	11,181	14,187	14,087	12,109	11,180	9,991	9,931
	M	Mean	302.67	284.28	238.18	265.66	286.32	262.30	290.44	270.34
		N	13,202	12,274	16,037	15,960	13,144	12,273	10,677	10,629
	Missing	Mean	288.78	272.89	228.75	247.23	267.70	253.79	277.77	257.40
		N	46	38	53	53	46	38	35	35
2	F	Mean	335.08	310.13	283.29	290.63	313.79	297.53	317.94	302.72
		N	13,404	12,122	14,302	14,215	13,322	12,121	11,548	11,487
	M	Mean	329.66	307.59	270.49	285.10	308.36	289.84	314.65	295.81
		N	15,135	13,718	16,491	16,392	15,051	13,715	12,870	12,796
	Missing	Mean	329.00	300.45	268.91	260.99	295.42	285.53	310.73	289.73
		N	73	62	79	79	73	62	60	60
3	F	Mean	354.66	333.37	302.56	301.18	328.42	318.34	339.72	321.17
		N	11,830	10,488	12,448	12,374	11,762	10,486	10,110	10,054
	M	Mean	354.92	332.77	289.05	297.91	327.18	311.35	339.57	316.24
		N	13,968	12,251	14,932	14,850	13,889	12,246	11,694	11,624
	Missing	Mean	345.58	325.34	285.30	270.27	309.29	304.94	331.74	306.70
		N	73	67	77	75	72	67	65	64
4	F	Mean	372.00	346.93	331.03	338.35	355.89	339.68	354.64	344.61
		N	11,136	10,101	11,561	11,492	11,068	10,100	9,814	9,753
	M	Mean	372.22	346.57	320.07	334.11	353.91	333.91	354.42	339.96
		N	13,375	11,946	13,956	13,896	13,316	11,943	11,581	11,524
	Missing	Mean	357.81	334.70	308.22	302.00	333.13	321.58	342.45	325.20
		N	69	66	74	73	68	66	62	61
5	F	Mean	382.40	356.60	345.01	344.13	364.01	351.43	364.65	355.38
		N	8,799	8,125	9,061	9,034	8,773	8,125	7,963	7,941
	M	Mean	382.23	355.96	332.66	341.81	362.60	345.04	364.16	350.56
		N	10,199	9,375	10,540	10,505	10,166	9,375	9,135	9,104
	Missing	Mean	376.70	344.77	324.36	317.74	347.84	335.03	355.06	338.85
		N	37	35	39	39	37	35	33	33

Grade	Gender		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
6	F	Mean	376.71	354.39	332.22	353.07	365.61	344.07	361.27	350.71
		N	6,522	6,082	6,797	6,746	6,473	6,082	5,895	5,849
	M	Mean	376.47	351.58	320.37	351.29	364.67	336.67	359.25	345.32
		N	7,744	7,115	8,106	8,047	7,686	7,114	6,867	6,815
	Missing	Mean	375.51	355.26	322.85	351.29	362.61	339.45	360.81	345.51
		N	77	69	79	79	77	69	67	67
7	F	Mean	382.77	361.41	338.45	356.72	370.68	350.68	368.30	357.09
		N	5,961	5,685	6,211	6,166	5,918	5,683	5,519	5,479
	M	Mean	382.11	357.55	325.82	354.10	369.04	342.58	365.35	350.80
		N	7,141	6,641	7,482	7,423	7,082	6,640	6,418	6,372
	Missing	Mean	377.80	353.83	323.29	337.86	358.25	341.44	359.65	344.88
		N	20	18	21	21	20	18	17	17
8	F	Mean	385.58	367.21	340.51	358.34	372.99	354.61	373.15	360.40
		N	5,308	5,136	5,528	5,494	5,274	5,136	4,978	4,949
	M	Mean	384.45	362.95	330.28	356.24	371.24	347.11	369.88	354.78
		N	6,360	6,093	6,654	6,610	6,317	6,092	5,889	5,848
	Missing	Mean	376.13	360.48	330.84	363.04	372.70	349.00	365.30	356.55
		N	24	23	25	24	23	23	23	22
9	F	Mean	380.52	382.56	360.34	355.15	369.01	371.98	382.39	371.40
		N	4,853	4,542	5,104	5,049	4,801	4,542	4,373	4,324
	M	Mean	381.05	377.46	348.70	354.48	368.93	363.79	378.93	365.67
		N	5,688	5,220	6,021	5,964	5,636	5,219	5,006	4,960
	Missing	Mean	377.53	378.89	356.49	343.83	363.83	368.82	379.54	368.24
		N	75	76	83	83	75	76	71	71
10	F	Mean	380.19	385.15	361.82	349.34	365.76	374.16	384.16	371.89
		N	4,494	4,283	4,689	4,643	4,450	4,283	4,151	4,110
	M	Mean	382.17	380.67	352.52	351.82	367.98	367.12	381.54	367.64
		N	5,331	4,962	5,590	5,544	5,284	4,957	4,787	4,740
	Missing	Mean	372.82	389.70	349.58	348.42	365.18	376.20	386.80	375.70
		N	11	10	12	12	11	10	10	10
11	F	Mean	385.37	390.16	366.60	358.44	372.92	378.95	389.13	377.53
		N	3,982	3,781	4,130	4,074	3,924	3,778	3,672	3,617
	M	Mean	386.94	385.64	357.48	358.49	373.78	372.07	386.43	372.84
		N	4,543	4,318	4,786	4,734	4,494	4,317	4,160	4,114
	Missing	Mean	365.13	372.29	350.33	333.73	349.73	360.86	369.36	356.29
		N	15	14	15	15	15	14	14	14
12	F	Mean	382.77	387.22	366.05	357.83	371.24	377.25	386.26	375.77
		N	2,777	2,657	2,882	2,846	2,739	2,656	2,584	2,549
	M	Mean	385.59	384.88	358.97	362.18	374.85	372.45	385.71	373.57
		N	3,096	2,939	3,250	3,209	3,056	2,938	2,831	2,797
	Missing	Mean	409.33	421.67	405.83	393.50	401.50	414.00	418.00	410.17
		N	6	6	6	6	6	6	6	6

Table 1.2.2.3
Mean Scale Scores by Grade by Ethnicity, S502 Paper

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
K	Non-Hispanic Asian	Mean	281.46	218.00	225.76	287.50	284.75	222.13	237.02	240.68
		N	18,340	18,340	18,339	18,336	18,336	18,339	18,340	18,336
	Non-Hispanic Pacific Islander	Mean	245.58	147.64	159.18	249.98	248.04	153.59	177.03	181.73
		N	939	939	939	939	939	939	939	939
	Non-Hispanic Black	Mean	267.09	183.14	187.21	282.19	274.90	185.39	208.32	212.03
		N	8,623	8,624	8,624	8,623	8,622	8,624	8,623	8,622
	Hispanic (of Any Race)	Mean	250.37	163.84	173.45	252.83	251.86	168.85	189.80	193.54
		N	101,604	101,595	101,594	101,582	101,582	101,594	101,594	101,571
	Non-Hispanic American Indian	Mean	261.74	167.51	176.04	269.38	265.85	171.97	195.76	199.93
		N	660	660	660	660	660	660	660	660
	Non-Hispanic Multiracial	Mean	284.54	199.88	203.54	289.80	287.44	201.94	225.26	227.36
		N	844	844	844	844	844	844	844	844
	Non-Hispanic White	Mean	273.43	190.18	202.95	280.46	277.21	196.80	215.14	220.71
		N	14,642	14,642	14,643	14,635	14,634	14,642	14,641	14,633
	Unknown	Mean	250.55	164.33	174.44	252.20	251.63	169.58	190.18	194.00
		N	9,908	9,908	9,907	9,907	9,907	9,907	9,908	9,907
1	Non-Hispanic Asian	Mean	310.53	296.37	258.14	278.60	296.57	277.82	301.61	284.31
		N	1,108	1,038	1,292	1,288	1,105	1,038	930	927
	Non-Hispanic Pacific Islander	Mean	316.47	277.53	243.75	277.33	310.87	262.63	292.92	287.23
		N	15	19	24	24	15	19	13	13
	Non-Hispanic Black	Mean	299.65	284.37	233.22	271.12	288.10	260.24	289.50	269.97
		N	1,848	1,747	2,365	2,343	1,832	1,746	1,462	1,450
	Hispanic (of Any Race)	Mean	304.92	283.42	242.20	268.10	288.52	263.65	290.36	271.73
		N	20,221	18,610	23,949	23,809	20,108	18,609	16,452	16,368
	Non-Hispanic American Indian	Mean	296.16	284.04	239.24	266.34	280.22	262.82	288.63	269.00
		N	97	84	108	108	97	84	76	76
	Non-Hispanic Multiracial	Mean	317.84	290.60	251.11	278.98	303.40	272.42	301.05	284.07
		N	131	133	168	167	130	133	109	108
	Non-Hispanic White	Mean	309.85	288.63	250.75	281.49	298.04	270.88	295.33	279.66
		N	1,747	1,603	2,080	2,070	1,739	1,603	1,411	1,403
	Unknown	Mean	286.71	277.68	230.09	261.48	275.47	254.74	280.32	261.55
		N	273	259	291	291	273	259	250	250

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall	
2	Non-Hispanic Asian	Mean	336.92	323.45	292.61	295.61	317.56	308.95	328.12	312.21	
		N	1,221	1,144	1,313	1,307	1,215	1,144	1,084	1,080	
	Non-Hispanic Pacific Islander	Mean	332.50	312.26	292.48	290.28	318.08	303.21	319.06	308.78	
		N	26	19	29	29	26	19	18	18	
	Non-Hispanic Black	Mean	327.10	304.94	266.42	286.73	308.15	286.26	311.94	293.29	
		N	2,264	2,033	2,569	2,553	2,253	2,032	1,854	1,845	
	Hispanic (of Any Race)	Mean	332.01	307.55	275.88	286.34	310.06	292.50	315.23	298.08	
		N	22,622	20,454	24,310	24,164	22,491	20,451	19,387	19,281	
	Non-Hispanic American Indian	Mean	331.77	300.16	268.43	273.65	304.83	283.61	310.46	291.95	
		N	104	85	113	112	103	85	79	78	
	Non-Hispanic Multiracial	Mean	337.23	319.06	282.10	298.27	320.00	303.61	326.47	310.72	
		N	171	156	186	182	167	156	145	141	
	Non-Hispanic White	Mean	338.77	318.26	285.73	299.88	320.54	303.08	325.05	308.92	
		N	1,891	1,720	2,027	2,016	1,880	1,720	1,624	1,614	
	Unknown	Mean	321.88	303.41	269.92	278.30	301.32	287.98	309.49	292.43	
		N	313	291	325	323	311	291	287	286	
	3	Non-Hispanic Asian	Mean	364.61	344.76	309.52	309.86	337.66	327.44	350.73	330.46
			N	1,051	923	1,113	1,108	1,046	923	887	884
Non-Hispanic Pacific Islander		Mean	351.18	333.48	290.28	298.56	321.82	311.43	336.63	310.47	
		N	28	21	32	32	28	21	19	19	
Non-Hispanic Black		Mean	352.98	331.64	289.23	301.71	328.22	311.23	338.27	316.63	
		N	2,256	1,956	2,463	2,442	2,240	1,955	1,835	1,821	
Hispanic (of Any Race)		Mean	354.05	332.30	294.48	297.35	326.35	313.80	338.85	317.53	
		N	20,509	18,090	21,707	21,587	20,393	18,084	17,378	17,278	
Non-Hispanic American Indian		Mean	347.52	330.16	291.28	291.33	319.54	310.94	335.21	313.00	
		N	122	113	130	130	122	113	106	106	
Non-Hispanic Multiracial		Mean	367.04	340.81	305.94	317.67	342.57	322.35	348.56	328.14	
		N	134	121	139	138	133	121	117	116	
Non-Hispanic White		Mean	360.48	337.37	303.38	314.55	338.10	320.90	344.53	326.07	
		N	1,531	1,352	1,621	1,613	1,524	1,352	1,303	1,297	
Unknown		Mean	350.80	326.88	292.58	294.80	324.69	310.36	334.56	315.17	
		N	240	230	252	249	237	230	224	221	

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
4	Non-Hispanic Asian	Mean	381.74	359.54	334.70	343.69	363.07	347.97	366.26	352.48
		N	680	618	711	705	675	618	599	595
	Non-Hispanic Pacific Islander	Mean	376.41	354.60	326.24	337.83	360.70	341.52	362.16	347.96
		N	27	25	29	29	27	25	25	25
	Non-Hispanic Black	Mean	370.61	344.14	320.47	337.36	354.54	333.05	352.43	339.74
		N	2,407	2,112	2,541	2,523	2,389	2,111	2,028	2,010
	Hispanic (of Any Race)	Mean	371.53	346.18	324.88	334.77	353.90	336.17	353.93	341.50
		N	19,613	17,651	20,364	20,272	19,521	17,648	17,163	17,078
	Non-Hispanic American Indian	Mean	363.93	342.23	322.07	319.53	342.49	332.58	348.80	334.98
		N	100	93	102	102	100	93	91	91
	Non-Hispanic Multiracial	Mean	377.17	350.23	330.30	343.37	361.44	341.22	357.97	347.17
		N	102	97	106	106	102	97	95	95
	Non-Hispanic White	Mean	378.96	352.82	331.73	349.48	364.93	343.03	360.90	349.86
		N	1,418	1,296	1,491	1,479	1,407	1,296	1,244	1,234
	Unknown	Mean	364.31	339.85	310.32	316.80	344.32	325.44	348.15	332.65
		N	233	221	247	245	231	221	212	210
5	Non-Hispanic Asian	Mean	396.51	371.09	350.56	357.15	377.06	361.32	379.11	366.40
		N	572	546	589	588	572	546	536	536
	Non-Hispanic Pacific Islander	Mean	391.41	367.28	341.39	344.09	369.82	358.33	375.18	362.24
		N	22	18	23	23	22	18	17	17
	Non-Hispanic Black	Mean	382.23	355.62	336.74	349.59	366.32	347.30	364.08	353.33
		N	1,837	1,633	1,913	1,907	1,831	1,633	1,591	1,585
	Hispanic (of Any Race)	Mean	381.31	355.34	337.86	340.66	361.67	347.26	363.41	351.77
		N	15,107	13,924	15,569	15,524	15,063	13,924	13,610	13,572
	Non-Hispanic American Indian	Mean	377.09	352.10	334.33	333.35	355.47	343.06	359.38	346.13
		N	118	107	118	118	118	107	107	107
	Non-Hispanic Multiracial	Mean	382.45	359.54	338.23	347.87	364.83	349.32	366.15	353.69
		N	60	57	62	62	60	57	55	55
	Non-Hispanic White	Mean	389.84	362.17	342.37	354.60	372.94	352.73	370.88	359.20
		N	1,154	1,087	1,191	1,181	1,145	1,087	1,060	1,051
	Unknown	Mean	374.65	349.76	332.90	338.67	359.57	342.20	357.75	348.42
		N	165	163	175	175	165	163	155	155

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
6	Non-Hispanic Asian	Mean	384.03	361.80	334.88	355.09	370.33	349.89	368.80	356.36
		N	397	363	413	409	393	363	355	351
	Non-Hispanic Pacific Islander	Mean	379.42	354.82	331.95	344.63	362.37	345.29	360.71	350.00
		N	19	17	19	19	19	17	17	17
	Non-Hispanic Black	Mean	381.15	353.35	322.28	358.10	370.41	338.79	361.90	348.46
		N	1,322	1,197	1,410	1,396	1,309	1,197	1,140	1,130
	Hispanic (of Any Race)	Mean	375.35	352.21	325.51	350.63	363.75	339.56	359.37	347.05
		N	11,427	10,604	11,917	11,836	11,347	10,603	10,265	10,189
	Non-Hispanic American Indian	Mean	369.68	347.47	318.10	345.56	357.98	332.34	354.15	339.11
		N	62	53	63	63	62	53	53	53
	Non-Hispanic Multiracial	Mean	381.00	358.71	328.50	356.56	369.80	345.49	365.98	354.12
		N	65	59	66	66	65	59	58	58
	Non-Hispanic White	Mean	384.46	357.17	333.00	364.60	375.25	345.61	365.29	354.33
		N	906	843	943	934	897	843	816	809
	Unknown	Mean	362.01	350.59	309.28	327.30	345.65	331.46	354.09	336.31
		N	145	130	151	149	144	130	125	124
7	Non-Hispanic Asian	Mean	390.49	367.69	342.37	359.74	376.24	355.84	375.16	362.39
		N	395	381	418	413	392	381	366	363
	Non-Hispanic Pacific Islander	Mean	386.11	359.56	328.89	369.78	378.28	343.63	367.56	352.63
		N	18	16	18	18	18	16	16	16
	Non-Hispanic Black	Mean	386.57	359.67	328.27	362.08	374.79	345.00	367.98	354.17
		N	1,228	1,116	1,302	1,285	1,210	1,115	1,075	1,064
	Hispanic (of Any Race)	Mean	380.79	358.46	331.19	353.27	367.94	345.66	365.59	352.65
		N	10,457	9,843	10,880	10,804	10,382	9,841	9,557	9,490
	Non-Hispanic American Indian	Mean	370.06	353.57	312.05	329.11	351.12	333.96	359.63	341.31
		N	52	53	56	56	52	53	49	49
	Non-Hispanic Multiracial	Mean	392.86	367.26	336.57	353.50	373.47	353.78	376.24	361.33
		N	49	50	54	54	49	50	46	46
	Non-Hispanic White	Mean	393.85	366.07	337.91	371.19	384.41	353.02	375.29	363.38
		N	847	813	903	898	842	813	779	775
	Unknown	Mean	368.67	351.14	315.23	329.82	351.71	334.06	356.59	339.03
		N	76	72	83	82	75	72	66	65

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
8	Non-Hispanic Asian	Mean	397.88	376.55	346.26	369.30	384.65	362.22	383.62	369.63
		N	335	330	351	347	331	330	319	315
	Non-Hispanic Pacific Islander	Mean	389.55	371.90	337.96	367.00	378.43	357.65	377.68	365.32
		N	22	20	23	22	21	20	19	19
	Non-Hispanic Black	Mean	388.82	363.70	332.16	366.29	378.19	348.51	372.06	357.93
		N	1,045	997	1,120	1,110	1,038	997	945	940
	Hispanic (of Any Race)	Mean	383.34	363.94	334.28	354.48	369.91	349.71	370.17	356.09
		N	9,395	9,030	9,766	9,709	9,337	9,030	8,770	8,717
	Non-Hispanic American Indian	Mean	386.79	359.58	335.77	364.23	375.53	348.00	368.29	356.29
		N	43	43	44	44	43	43	42	42
	Non-Hispanic Multiracial	Mean	398.60	375.33	346.35	380.61	393.41	360.77	383.69	371.57
		N	52	52	57	56	51	52	48	47
	Non-Hispanic White	Mean	393.51	372.70	341.49	371.71	383.72	358.10	379.56	366.48
		N	738	718	779	774	732	717	687	680
	Unknown	Mean	376.82	362.03	326.36	343.03	361.49	345.24	367.28	350.76
		N	62	62	67	66	61	62	60	59
9	Non-Hispanic Asian	Mean	390.08	388.83	363.56	368.07	380.21	377.72	389.74	378.63
		N	320	299	333	329	317	299	294	291
	Non-Hispanic Pacific Islander	Mean	379.44	368.31	351.11	354.61	367.22	358.00	370.44	359.56
		N	18	16	18	18	18	16	16	16
	Non-Hispanic Black	Mean	382.84	378.52	351.31	366.11	375.01	365.65	380.08	368.64
		N	1,026	882	1,096	1,084	1,013	881	839	826
	Hispanic (of Any Race)	Mean	379.78	379.15	353.73	352.24	367.27	367.04	379.74	367.45
		N	8,392	7,825	8,848	8,759	8,312	7,825	7,518	7,447
	Non-Hispanic American Indian	Mean	373.26	375.26	340.26	328.84	351.32	358.78	375.96	357.00
		N	31	27	31	31	31	27	27	27
	Non-Hispanic Multiracial	Mean	407.90	392.73	369.00	382.04	396.04	381.52	398.26	386.98
		N	50	48	52	52	50	48	46	46
	Non-Hispanic White	Mean	390.19	388.05	361.69	370.37	381.07	375.20	389.28	377.41
		N	598	566	631	626	592	566	545	539
	Unknown	Mean	362.31	372.79	342.01	326.08	346.73	359.45	370.53	356.73
		N	181	175	199	197	179	175	165	163

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall	
10	Non-Hispanic Asian	Mean	391.54	390.30	368.49	360.01	375.49	379.81	390.91	378.04	
		N	307	294	318	314	303	293	285	280	
	Non-Hispanic Pacific Islander	Mean	381.54	379.45	362.56	357.52	371.71	373.91	380.68	374.18	
		N	24	22	25	25	24	22	22	22	
	Non-Hispanic Black	Mean	380.88	380.73	350.93	360.59	371.38	366.63	381.31	368.15	
		N	988	858	1,056	1,042	974	858	813	802	
	Hispanic (of Any Race)	Mean	379.93	382.16	356.46	347.83	364.96	369.91	381.94	368.72	
		N	7,739	7,346	8,087	8,021	7,673	7,342	7,110	7,045	
	Non-Hispanic American Indian	Mean	356.33	359.74	336.35	316.52	339.10	346.11	358.96	344.27	
		N	30	27	31	31	30	27	26	26	
	Non-Hispanic Multiracial	Mean	378.67	382.07	368.92	371.49	373.89	376.52	381.79	377.66	
		N	36	29	37	37	36	29	29	29	
	Non-Hispanic White	Mean	393.44	389.83	363.83	363.93	379.84	377.19	391.38	378.14	
		N	643	616	665	658	637	616	602	596	
	Unknown	Mean	387.78	385.71	358.89	365.31	378.54	373.27	387.03	376.18	
		N	69	63	72	71	68	63	61	60	
	11	Non-Hispanic Asian	Mean	400.36	395.98	373.80	369.17	386.07	386.28	397.87	386.99
			N	302	280	307	303	298	280	278	274
Non-Hispanic Pacific Islander		Mean	381.06	386.83	364.74	348.53	365.35	375.06	384.63	369.06	
		N	17	18	19	19	17	18	16	16	
Non-Hispanic Black		Mean	380.18	382.67	352.86	363.11	372.21	368.59	382.28	369.83	
		N	947	878	1,014	1,008	941	878	837	833	
Hispanic (of Any Race)		Mean	385.60	387.57	361.92	356.15	372.01	375.19	387.36	374.55	
		N	6,668	6,368	6,967	6,878	6,582	6,366	6,158	6,074	
Non-Hispanic American Indian		Mean	385.85	379.21	356.05	336.45	361.15	369.00	381.41	367.12	
		N	20	19	22	22	20	19	17	17	
Non-Hispanic Multiracial		Mean	400.35	400.45	377.80	400.91	402.36	392.10	403.04	396.43	
		N	34	29	35	34	33	29	28	28	
Non-Hispanic White		Mean	396.46	395.06	367.94	371.22	384.40	381.98	395.80	383.22	
		N	484	456	493	486	475	454	450	442	
Unknown		Mean	383.88	381.28	361.61	367.93	375.97	371.06	381.06	370.82	
		N	68	65	74	73	67	65	62	61	

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre- hension	Overall
12	Non-Hispanic Asian	Mean	398.22	393.65	374.67	369.53	384.42	384.55	395.47	385.03
		N	239	229	249	245	236	229	222	219
	Non-Hispanic Pacific Islander	Mean	403.29	384.60	375.57	393.86	398.71	378.80	386.40	382.00
		N	7	5	7	7	7	5	5	5
	Non-Hispanic Black	Mean	377.03	380.15	352.21	357.77	368.70	366.72	379.90	368.15
		N	781	734	832	817	764	733	695	681
	Hispanic (of Any Race)	Mean	383.64	385.94	362.87	358.70	372.03	375.00	385.70	374.33
		N	4,399	4,200	4,581	4,528	4,346	4,199	4,078	4,030
	Non-Hispanic American	Mean	384.76	383.86	358.55	351.45	366.10	370.10	384.05	366.55
		N	21	21	22	22	21	21	20	20
	Non-Hispanic Multiracial	Mean	396.29	391.12	364.00	365.79	381.21	378.27	391.85	378.54
		N	28	26	28	28	28	26	26	26
	Non-Hispanic White	Mean	395.57	393.26	368.45	378.14	388.67	381.33	394.71	384.54
		N	346	329	359	354	341	329	319	315
	Unknown	Mean	397.47	393.93	373.67	355.77	376.33	384.02	395.30	381.70
		N	58	58	60	60	58	58	56	56

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 the earlier grades of K, 1, and 2, the correlation between Listening and Speaking, and the correlation between Reading and Writing were pronounced. In Grades 3 to 12, the highest correlations were between Listening and Reading and between Reading and Writing.

Table 1.2.3.1
Correlations Among Scale Scores: K, S502 Paper

Domains	Correlations and N counts	Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	0.504	0.541	0.793
	N	155,560	155,549	155,547	155,524
Reading	Pearson Correlation		1	0.735	0.458
	N		155,552	155,549	155,515
Writing	Pearson Correlation			1	0.509
	N			155,550	155,515
Speaking	Pearson Correlation				1
	N				155,526

Table 1.2.3.2
Correlations Among Scale Scores: Grade 1, S502 Paper

Domains	Correlations and N counts	Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	0.475	0.464	0.506
	N	25,440	20,703	25,438	25,299
Reading	Pearson Correlation		1	0.448	0.417
	N		23,493	23,491	23,371
Writing	Pearson Correlation			1	0.447
	N			30,277	30,097
Speaking	Pearson Correlation				1
	N				30,100

Table 1.2.3.3
Correlations Among Scale Scores: Grade 2, S502 Paper

Domains	Correlations and N counts	Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	0.542	0.500	0.479
	N	28,612	24,478	28,609	28,446
Reading	Pearson Correlation		1	0.623	0.452
	N		25,902	25,898	25,756
Writing	Pearson Correlation			1	0.481
	N			30,872	30,681
Speaking	Pearson Correlation				1
	N				30,686

Table 1.2.3.4

Correlations Among Scale Scores: Grade 3, S502 Paper

Domains	Correlations and N counts	Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	0.617	0.497	0.477
	N	25,871	21,869	25,865	25,723
Reading	Pearson Correlation		1	0.568	0.485
	N		22,806	22,799	22,679
Writing	Pearson Correlation			1	0.505
	N			27,457	27,292
Speaking	Pearson Correlation				1
	N				27,299

Table 1.2.3.5

Correlations Among Scale Scores: Grades 4–5, S502 Paper

Domains	Correlations and N counts	Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	0.692	0.573	0.556
	N	43,615	38,588	43,611	43,428
Reading	Pearson Correlation		1	0.643	0.560
	N		39,648	39,644	39,472
Writing	Pearson Correlation			1	0.577
	N			45,231	45,036
Speaking	Pearson Correlation				1
	N				45,039

Table 1.2.3.6

Correlations Among Scale Scores: Grades 6–8, S502 Paper

Domains	Correlations and N counts	Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	0.711	0.677	0.639
	N	39,157	35,673	39,150	38,870
Reading	Pearson Correlation		1	0.652	0.567
	N		36,862	36,857	36,600
Writing	Pearson Correlation			1	0.641
	N			40,903	40,598
Speaking	Pearson Correlation				1
	N				40,610

Table 1.2.3.7

Correlations Among Scale Scores: Grades 9–12, S502 Paper

Domains	Correlations and N counts	Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	0.725	0.637	0.647
	N	34,871	31,665	34,854	34,491
Reading	Pearson Correlation		1	0.666	0.590
	N		32,808	32,796	32,446
Writing	Pearson Correlation			1	0.626
	N			36,568	36,161
Speaking	Pearson Correlation				1
	N				36,179

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. The percentages of PL 5 and PL 6 in Tier B/C in Listening were as follows: K, 44%; Grade 1, 64%; Grade 2, 67%; Grade 3, 73%; Grades 4–5, 71%; Grades 6–8, 53%; and Grades 9–12, 31%. The percentages of PL 5 and PL 6 in Tier B/C in the Reading domain were as follows: K, 15%; Grade 1, 20%; Grade 2, 25%; Grade 3, 24%; Grades 4–5, 31%; Grades 6–8, 19%; and Grades 9–12, 28%. For the Writing domain, less than 1% were in PL 5 and PL 6. In the Speaking domain, 36% of Kindergarten students reached PL 5 and PL 6. In Grades 1–12 in Tier B/C, percentages in PL 5 and P6 were low but increased to Grades 4–5 and then decreased (Grade 1, 10%; Grade 2, 9%; Grade 3, 8%; Grades 4–5, 22%; Grades 6–8, 21%; Grades 9–12, 17%).

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, S502 Paper

Cluster	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	49,595	15,477	13,189	8,412	21,288	47,599	155,560
1	A	1,016	2,022	2,237	1,420	3,377	2,209	12,281
	BC	170	472	1,794	2,307	3,019	5,397	13,159
2	A	664	1,352	1,410	956	2,054	.	6,436
	BC	141	1,190	3,809	2,247	5,979	8,810	22,176
3	A	151	1,258	1,395	730	551	413	4,498
	BC	14	528	3,097	2,049	8,207	7,478	21,373
4-5	A	458	1,883	1,789	989	586	536	6,241
	BC	32	801	3,966	6,031	14,241	12,303	37,374
6-8	A	2,599	3,238	1,273	673	518	113	8,414
	BC	56	1,103	4,423	8,760	8,263	8,138	30,743
9-12	A	4,173	2,452	838	346	100	.	7,909
	BC	479	2,885	7,170	8,087	4,877	3,464	26,962

Table 1.3.1.1.1.2

Proficiency Level by Cluster (Percent): Listening, S502 Paper

Cluster	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	31.9%	10.0%	8.5%	5.4%	13.7%	30.6%	100.0%
1	A	8.3%	16.5%	18.2%	11.6%	27.5%	18.0%	100.0%
	BC	1.3%	3.6%	13.6%	17.5%	22.9%	41.0%	100.0%
2	A	10.3%	21.0%	21.9%	14.9%	31.9%	.	100.0%
	BC	0.6%	5.4%	17.2%	10.1%	27.0%	39.7%	100.0%
3	A	3.4%	28.0%	31.0%	16.2%	12.3%	9.2%	100.0%
	BC	0.1%	2.5%	14.5%	9.6%	38.4%	35.0%	100.0%
4-5	A	7.3%	30.2%	28.7%	15.9%	9.4%	8.6%	100.0%
	BC	0.1%	2.1%	10.6%	16.1%	38.1%	32.9%	100.0%
6-8	A	30.9%	38.5%	15.1%	8.0%	6.2%	1.3%	100.0%
	BC	0.2%	3.6%	14.4%	28.5%	26.9%	26.5%	100.0%
9-12	A	52.8%	31.0%	10.6%	4.4%	1.3%	.	100.0%
	BC	1.8%	10.7%	26.6%	30.0%	18.1%	12.9%	100.0%

1.3.1.1.2 By Grade by Tier

Table 1.3.1.1.2.1
Proficiency Level by Grade (Count): Listening, S502 Paper

Grade	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	49,595	15,477	13,189	8,412	21,288	47,599	155,560
1	A	1,016	2,022	2,237	1,420	3,377	2,209	12,281
	BC	170	472	1,794	2,307	3,019	5,397	13,159
2	A	664	1,352	1,410	956	2,054	.	6,436
	BC	141	1,190	3,809	2,247	5,979	8,810	22,176
3	A	151	1,258	1,395	730	551	413	4,498
	BC	14	528	3,097	2,049	8,207	7,478	21,373
4	A	187	933	965	540	399	268	3,292
	BC	25	443	2,423	3,611	7,985	6,801	21,288
5	A	271	950	824	449	187	268	2,949
	BC	7	358	1,543	2,420	6,256	5,502	16,086
6	A	674	1,279	417	302	198	42	2,912
	BC	16	341	1,407	3,620	3,042	3,005	11,431
7	A	954	837	554	124	169	71	2,709
	BC	20	456	1,424	3,298	2,815	2,400	10,413
8	A	971	1,122	302	247	151	.	2,793
	BC	20	306	1,592	1,842	2,406	2,733	8,899
9	A	986	1,170	282	76	62	.	2,576
	BC	64	506	1,588	2,762	1,689	1,431	8,040
10	A	1,261	708	251	120	13	.	2,353
	BC	65	909	1,892	2,409	1,348	860	7,483
11	A	1,112	381	196	119	21	.	1,829
	BC	123	713	2,267	1,440	1,231	937	6,711
12	A	814	193	109	31	4	.	1,151
	BC	227	757	1,423	1,476	609	236	4,728

Table 1.3.1.1.2.2

Proficiency Level by Grade (Percent): Listening, S502 Paper

Grade	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	31.9%	10.0%	8.5%	5.4%	13.7%	30.6%	100.0%
1	A	8.3%	16.5%	18.2%	11.6%	27.5%	18.0%	100.0%
	BC	1.3%	3.6%	13.6%	17.5%	22.9%	41.0%	100.0%
2	A	10.3%	21.0%	21.9%	14.9%	31.9%	.	100.0%
	BC	0.6%	5.4%	17.2%	10.1%	27.0%	39.7%	100.0%
3	A	3.4%	28.0%	31.0%	16.2%	12.3%	9.2%	100.0%
	BC	0.1%	2.5%	14.5%	9.6%	38.4%	35.0%	100.0%
4	A	5.7%	28.3%	29.3%	16.4%	12.1%	8.1%	100.0%
	BC	0.1%	2.1%	11.4%	17.0%	37.5%	32.0%	100.0%
5	A	9.2%	32.2%	27.9%	15.2%	6.3%	9.1%	100.0%
	BC	0.0%	2.2%	9.6%	15.0%	38.9%	34.2%	100.0%
6	A	23.2%	43.9%	14.3%	10.4%	6.8%	1.4%	100.0%
	BC	0.1%	3.0%	12.3%	31.7%	26.6%	26.3%	100.0%
7	A	35.2%	30.9%	20.5%	4.6%	6.2%	2.6%	100.0%
	BC	0.2%	4.4%	13.7%	31.7%	27.0%	23.1%	100.0%
8	A	34.8%	40.2%	10.8%	8.8%	5.4%	.	100.0%
	BC	0.2%	3.4%	17.9%	20.7%	27.0%	30.7%	100.0%
9	A	38.3%	45.4%	11.0%	3.0%	2.4%	.	100.0%
	BC	0.8%	6.3%	19.8%	34.4%	21.0%	17.8%	100.0%
10	A	53.6%	30.1%	10.7%	5.1%	0.6%	.	100.0%
	BC	0.9%	12.2%	25.3%	32.2%	18.0%	11.5%	100.0%
11	A	60.8%	20.8%	10.7%	6.5%	1.2%	.	100.0%
	BC	1.8%	10.6%	33.8%	21.5%	18.3%	14.0%	100.0%
12	A	70.7%	16.8%	9.5%	2.7%	0.4%	.	100.0%
	BC	4.8%	16.0%	30.1%	31.2%	12.9%	5.0%	100.0%

1.3.1.1.3 By Grade

Table 1.3.1.1.3.1

Proficiency Level by Grade (Count): Listening, S502 Paper

Grade	Listening Proficiency Range						Total
	1	2	3	4	5	6	
K	49,595	15,477	13,189	8,412	21,288	47,599	155,560
1	1,186	2,494	4,031	3,727	6,396	7,606	25,440
2	805	2,542	5,219	3,203	8,033	8,810	28,612
3	165	1,786	4,492	2,779	8,758	7,891	25,871
4	212	1,376	3,388	4,151	8,384	7,069	24,580
5	278	1,308	2,367	2,869	6,443	5,770	19,035
6	690	1,620	1,824	3,922	3,240	3,047	14,343
7	974	1,293	1,978	3,422	2,984	2,471	13,122
8	991	1,428	1,894	2,089	2,557	2,733	11,692
9	1,050	1,676	1,870	2,838	1,751	1,431	10,616
10	1,326	1,617	2,143	2,529	1,361	860	9,836
11	1,235	1,094	2,463	1,559	1,252	937	8,540
12	1,041	950	1,532	1,507	613	236	5,879

Table 1.3.1.1.3.2

Proficiency Level by Grade (Percent): Listening, S502 Paper

Grade	Listening Proficiency Range						Total
	1	2	3	4	5	6	
K	31.9%	10.0%	8.5%	5.4%	13.7%	30.6%	100.0%
1	4.7%	9.8%	15.9%	14.7%	25.1%	29.9%	100.0%
2	2.8%	8.9%	18.2%	11.2%	28.1%	30.8%	100.0%
3	0.6%	6.9%	17.4%	10.7%	33.9%	30.5%	100.0%
4	0.9	5.6%	13.8%	16.9%	34.1%	28.8%	100.0%
5	1.5%	6.9%	12.4%	15.1%	33.9%	30.3%	100.0%
6	4.8%	11.3%	12.7%	27.3%	22.6%	21.2%	100.0%
7	7.4%	9.9%	15.1%	26.1%	22.7%	18.8%	100.0%
8	8.5%	12.2%	16.2%	17.9%	21.9%	23.4%	100.0%
9	9.9%	15.8%	17.6%	26.7%	16.5%	13.5%	100.0%
10	13.5%	16.4%	21.8%	25.7%	13.8%	8.7%	100.0%
11	14.5%	12.8%	28.8%	18.3%	14.7%	11.0%	100.0%
12	17.7%	16.2%	26.1%	25.6%	10.4%	4.0%	100.0%

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, S502 Paper

Cluster	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	118,572	3,901	10,504	7,552	15,023	.	155,552
1	A	5,525	3,937	1,320	466	328	292	11,868
	BC	192	3,072	4,844	1,182	1,281	1,054	11,625
2	A	3,299	1,569	680	192	318	56	6,114
	BC	1,587	6,171	5,139	1,988	2,673	2,230	19,788
3	A	2,278	1,205	452	96	162	56	4,249
	BC	128	2,277	8,517	3,268	3,004	1,363	18,557
4-5	A	3,367	1,517	476	230	296	35	5,921
	BC	401	6,113	11,219	5,533	6,474	3,987	33,727
6-8	A	4,199	2,782	632	158	200	123	8,094
	BC	1,121	10,460	8,417	3,221	3,695	1,854	28,768
9-12	A	3,184	3,068	904	265	339	139	7,899
	BC	576	7,289	6,984	3,147	3,741	3,172	24,909

Table 1.3.1.2.1.2

Proficiency Level by Cluster (Percent): Reading, S502 Paper

Cluster	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	76.2%	2.5%	6.8%	4.9%	9.7%	.	100.0%
1	A	46.6%	33.2%	11.1%	3.9%	2.8%	2.5%	100.0%
	BC	1.7%	26.4%	41.7%	10.2%	11.0%	9.1%	100.0%
2	A	54.0%	25.7%	11.1%	3.1%	5.2%	0.9%	100.0%
	BC	8.0%	31.2%	26.0%	10.1%	13.5%	11.3%	100.0%
3	A	53.6%	28.4%	10.6%	2.3%	3.8%	1.3%	100.0%
	BC	0.7%	12.3%	45.9%	17.6%	16.2%	7.3%	100.0%
4-5	A	56.9%	25.6%	8.0%	3.9%	5.0%	0.6%	100.0%
	BC	1.2%	18.1%	33.3%	16.4%	19.2%	11.8%	100.0%
6-8	A	51.9%	34.4%	7.8%	2.0%	2.5%	1.5%	100.0%
	BC	3.9%	36.4%	29.3%	11.2%	12.8%	6.4%	100.0%
9-12	A	40.3%	38.8%	11.4%	3.4%	4.3%	1.8%	100.0%
	BC	2.3%	29.3%	28.0%	12.6%	15.0%	12.7%	100.0%

1.3.1.2.2. By Grade by Tier

Table 1.3.1.2.2.1

Proficiency Level by Grade (Count): Reading, S502 Paper

Grade	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	118,572	3,901	10,504	7,552	15,023	.	155,552
1	A	5,525	3,937	1,320	466	328	292	11,868
	BC	192	3,072	4,844	1,182	1,281	1,054	11,625
2	A	3,299	1,569	680	192	318	56	6,114
	BC	1,587	6,171	5,139	1,988	2,673	2,230	19,788
3	A	2,278	1,205	452	96	162	56	4,249
	BC	128	2,277	8,517	3,268	3,004	1,363	18,557
4	A	1,712	841	250	109	170	35	3,117
	BC	186	3,316	6,488	3,708	3,329	1,969	18,996
5	A	1,655	676	226	121	126	.	2,804
	BC	215	2,797	4,731	1,825	3,145	2,018	14,731
6	A	1,244	1,095	228	51	78	44	2,740
	BC	280	4,267	3,053	1,125	1,337	464	10,526
7	A	1,393	843	214	49	77	41	2,617
	BC	408	3,318	3,184	1,066	1,093	658	9,727
8	A	1,562	844	190	58	45	38	2,737
	BC	433	2,875	2,180	1,030	1,265	732	8,515
9	A	1,034	910	345	83	136	50	2,558
	BC	84	1,835	1,854	1,401	1,111	995	7,280
10	A	913	960	289	62	67	57	2,348
	BC	82	1,994	2,076	782	993	980	6,907
11	A	731	737	177	63	115	21	1,844
	BC	153	1,843	1,776	707	997	793	6,269
12	A	506	461	93	57	21	11	1,149
	BC	257	1,617	1,278	257	640	404	4,453

Table 1.3.1.2.2.2

Proficiency Level by Grade (Percent): Reading, S502 Paper

Grade	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	76.2%	2.5%	6.8%	4.9%	9.7%	.	100.0%
1	A	46.6%	33.2%	11.1%	3.9%	2.8%	2.5%	100.0%
	BC	1.7%	26.4%	41.7%	10.2%	11.0%	9.1%	100.0%
2	A	54.0%	25.7%	11.1%	3.1%	5.2%	0.9%	100.0%
	BC	8.0%	31.2%	26.0%	10.1%	13.5%	11.3%	100.0%
3	A	53.6%	28.4%	10.6%	2.3%	3.8%	1.3%	100.0%
	BC	0.7%	12.3%	45.9%	17.6%	16.2%	7.3%	100.0%
4	A	54.9%	27.0%	8.0%	3.5%	5.5%	1.1%	100.0%
	BC	1.0%	17.5%	34.2%	19.5%	17.5%	10.4	100.0%
5	A	59.0%	24.1%	8.1%	4.3%	4.5%	.	100.0%
	BC	1.5%	19.0%	32.1%	12.4%	21.4%	13.7%	100.0%
6	A	45.4%	40.0%	8.3%	1.9%	2.9%	1.6%	100.0%
	BC	2.7%	40.5%	29.0%	10.7%	12.7%	4.4%	100.0%
7	A	53.2%	32.2%	8.2%	1.9%	2.9%	1.6%	100.0%
	BC	4.2%	34.1%	32.7%	11.0%	11.2%	6.8%	100.0%
8	A	57.1%	30.8%	6.9%	2.1%	1.6%	1.4%	100.0%
	BC	5.1%	33.8%	25.6%	12.1%	14.9%	8.6%	100.0%
9	A	40.4%	35.6%	13.5%	3.2%	5.3%	2.0%	100.0%
	BC	1.2%	25.2%	25.5%	19.2%	15.3%	13.7%	100.0%
10	A	38.9%	40.9%	12.3%	2.6%	2.9%	2.4%	100.0%
	BC	1.2%	28.9%	30.1%	11.3%	14.4%	14.2%	100.0%
11	A	39.6%	40.0%	9.6%	3.4%	6.2%	1.1%	100.0%
	BC	2.4%	29.4%	28.3%	11.3%	15.9%	12.7%	100.0%
12	A	44.0%	40.1%	8.1%	5.0%	1.8%	1.0%	100.0%
	BC	5.8%	36.3%	28.7%	5.8%	14.4%	9.1%	100.0%

1.3.1.2.3. By Grade

Table 1.3.1.2.3.1

Proficiency Level by Grade (Count): Reading, S502 Paper

Grade	Reading Proficiency Range						Total
	1	2	3	4	5	6	
K	118,572	3,901	10,504	7,552	15,023	.	155,552
1	5,717	7,009	6,164	1,648	1,609	1,346	23,493
2	4,886	7,740	5,819	2,180	2,991	2,286	25,902
3	2,406	3,482	8,969	3,364	3,166	1,419	22,806
4	1,898	4,157	6,738	3,817	3,499	2,004	22,113
5	1,870	3,473	4,957	1,946	3,271	2,018	17,535
6	1,524	5,362	3,281	1,176	1,415	508	13,266
7	1,801	4,161	3,398	1,115	1,170	699	12,344
8	1,995	3,719	2,370	1,088	1,310	770	11,252
9	1,118	2,745	2,199	1,484	1,247	1,045	9,838
10	995	2,954	2,365	844	1,060	1,037	9,255
11	884	2,580	1,953	770	1,112	814	8,113
12	763	2,078	1,371	314	661	415	5,602

Table 1.3.1.2.3.2

Proficiency Level by Grade (Percent): Reading, S502 Paper

Grade	Reading Proficiency Range						Total
	1	2	3	4	5	6	
K	76.2%	2.5%	6.8%	4.9%	9.7%	.	100.0%
1	24.3%	29.8%	26.2%	7.0%	6.9%	5.7%	100.0%
2	18.9%	29.9%	22.5%	8.4%	11.6%	8.8%	100.0%
3	10.6%	15.3%	39.3%	14.8%	13.9%	6.2%	100.0%
4	8.6%	18.8%	30.5%	17.3%	15.8%	9.1%	100.0%
5	10.7%	19.8%	28.3%	11.1%	18.7%	11.5%	100.0%
6	11.5%	40.4%	24.7%	8.9%	10.7%	3.8%	100.0%
7	14.6%	33.7%	27.5%	9.0%	9.5%	5.7%	100.0%
8	17.7%	33.1%	21.1%	9.7%	11.6%	6.8%	100.0%
9	11.4%	27.9%	22.4%	15.1%	12.7%	10.6%	100.0%
10	10.8%	31.9%	25.6%	9.1%	11.5%	11.2%	100.0%
11	10.9%	31.8%	24.1%	9.5%	13.7%	10.0%	100.0%
12	13.6%	37.1%	24.5%	5.6%	11.8%	7.4%	100.0%

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, S502 Paper

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	112,219	22,511	16,136	4,684	.	.	155,550
1	A	7,650	6,880	545	.	.	.	15,075
	BC	3,443	6,277	5,299	179	3	1	15,202
2	A	3,055	2,327	1,774	3	.	.	7,159
	BC	2,362	6,077	13,752	1,508	14	.	23,713
3	A	1,929	1,917	1,069	8	.	.	4,923
	BC	1,227	3,371	15,468	2,451	16	1	22,534
4-5	A	2,022	1,727	2,858	38	.	.	6,645
	BC	659	1,572	23,779	12,192	361	23	38,586
6-8	A	3,816	3,192	1,934	20	.	.	8,962
	BC	1,034	2,262	20,342	8,267	36	.	31,941
9-12	A	2,573	2,640	2,923	416	2	.	8,554
	BC	1,235	1,979	15,184	9,467	149	.	28,014

Table 1.3.1.3.1.2

Proficiency Level by Cluster (Percent): Writing, S502 Paper

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	72.1%	14.5%	10.4%	3.0%	.	.	100.0%
1	A	50.8%	45.6%	3.6%	.	.	.	100.0%
	BC	22.7%	41.3%	34.9%	1.2%	0.0%	0.0%	100.0%
2	A	42.7%	32.5%	24.8%	0.0%	.	.	100.0%
	BC	10.0%	25.6%	58.0%	6.4%	0.1%	.	100.0%
3	A	39.2%	38.9%	21.7%	0.2%	.	.	100.0%
	BC	5.5%	15.0%	68.6%	10.9%	0.1%	0.0%	100.0%
4-5	A	30.4%	26.0%	43.0%	0.6%	.	.	100.0%
	BC	1.7%	4.1%	61.6%	31.6%	0.9%	0.1%	100.0%
6-8	A	42.6%	35.6%	21.6%	0.2%	.	.	100.0%
	BC	3.2%	7.1%	63.7%	25.9%	0.1%	.	100.0%
9-12	A	30.1%	30.9%	34.2%	4.9%	0.0%	.	100.0%
	BC	4.4%	7.1%	54.2%	33.8%	0.5%	.	100.0%

1.3.1.3.2 By Grade by Tier

Table 1.3.1.3.2.1

Proficiency Level by Grade (Count): Writing, S502 Paper

Grade	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	112,219	22,511	16,136	4,684	.	.	155,550
1	A	7,650	6,880	545	.	.	.	15,075
	BC	3,443	6,277	5,299	179	3	1	15,202
2	A	3,055	2,327	1,774	3	.	.	7,159
	BC	2,362	6,077	13,752	1,508	14	.	23,713
3	A	1,929	1,917	1,069	8	.	.	4,923
	BC	1,227	3,371	15,468	2,451	16	1	22,534
4	A	1,173	913	1,425	20	.	.	3,531
	BC	437	1,015	14,755	5,666	182	5	22,060
5	A	849	814	1,433	18	.	.	3,114
	BC	222	557	9,024	6,526	179	18	16,526
6	A	1,149	1,113	807	8	.	.	3,077
	BC	331	1,048	7,354	3,155	17	.	11,905
7	A	1,206	1,176	523	8	.	.	2,913
	BC	324	802	6,955	2,711	9	.	10,801
8	A	1,461	903	604	4	.	.	2,972
	BC	379	412	6,033	2,401	10	.	9,235
9	A	678	866	1,036	220	2	.	2,802
	BC	224	407	4,372	3,326	77	.	8,406
10	A	708	908	811	112	.	.	2,539
	BC	260	525	4,103	2,820	44	.	7,752
11	A	658	638	622	54	.	.	1,972
	BC	356	617	3,576	2,389	21	.	6,959
12	A	529	228	454	30	.	.	1,241
	BC	395	430	3,133	932	7	.	4,897

Table 1.3.1.3.2.2

Proficiency Level by Grade (Percent): Writing, S502 Paper

Grade	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	72.1%	14.5%	10.4%	3.0%	.	.	100.0%
1	A	50.8%	45.6%	3.6%	.	.	.	100.0%
	BC	22.7%	41.3%	34.9%	1.2%	0.0%	0.0%	100.0%
2	A	42.7%	32.5%	24.8%	0.0%	.	.	100.0%
	BC	10.0%	25.6%	58.0%	6.4%	0.1%	.	100.0%
3	A	39.2%	38.9%	21.7%	0.2%	.	.	100.0%
	BC	5.5%	15.0%	68.6%	10.9%	0.1%	0.0%	100.0%
4	A	33.2%	25.9%	40.4%	0.6%	.	.	100.0%
	BC	2.0%	4.6%	66.9%	25.7%	0.8%	0.0%	100.0%
5	A	27.3%	26.1%	46.0%	0.6%	.	.	100.0%
	BC	1.3%	3.4%	54.6%	39.5%	1.1%	0.1%	100.0%
6	A	37.3%	36.2%	26.2%	0.3%	.	.	100.0%
	BC	2.8%	8.8%	61.8%	26.5%	0.1%	.	100.0%
7	A	41.4%	40.4%	18.0%	0.3%	.	.	100.0%
	BC	3.0%	7.4%	64.4%	25.1%	0.1%	.	100.0%
8	A	49.2%	30.4%	20.3%	0.1%	.	.	100.0%
	BC	4.1%	4.5%	65.3%	26.0%	0.1%	.	100.0%
9	A	24.2%	30.9%	37.0%	7.9%	0.1%	.	100.0%
	BC	2.7%	4.8%	52.0%	39.6%	0.9%	.	100.0%
10	A	27.9%	35.8%	31.9%	4.4%	.	.	100.0%
	BC	3.4%	6.8%	52.9%	36.4%	0.6%	.	100.0%
11	A	33.4%	32.4%	31.5%	2.7%	.	.	100.0%
	BC	5.1%	8.9%	51.4%	34.3%	0.3%	.	100.0%
12	A	42.6%	18.4%	36.6%	2.4%	.	.	100.0%
	BC	8.1%	8.8%	64.0%	19.0%	0.1%	.	100.0%

1.3.1.3.3 By Grade

Table 1.3.1.3.3.1
Proficiency Level by Grade (Count): Writing, S502 Paper

Grade	Writing Proficiency Range						Total
	1	2	3	4	5	6	
K	112,219	22,511	16,136	4,684	.	.	155,550
1	11,093	13,157	5,844	179	3	1	30,277
2	5,417	8,404	15,526	1,511	14	.	30,872
3	3,156	5,288	16,537	2,459	16	1	27,457
4	1,610	1,928	16,180	5,686	182	5	25,591
5	1,071	1,371	10,457	6,544	179	18	19,640
6	1,480	2,161	8,161	3,163	17	.	14,982
7	1,530	1,978	7,478	2,719	9	.	13,714
8	1,840	1,315	6,637	2,405	10	.	12,207
9	902	1,273	5,408	3,546	79	.	11,208
10	968	1,433	4,914	2,932	44	.	10,291
11	1,014	1,255	4,198	2,443	21	.	8,931
12	924	658	3,587	962	7	.	6,138

Table 1.3.1.3.3.2
Proficiency Level by Grade (Percent): Writing, S502 Paper

Grade	Writing Proficiency Range						Total
	1	2	3	4	5	6	
K	72.1%	14.5%	10.4%	3.0%	.	.	100.0%
1	36.6%	43.5%	19.3%	0.6%	0.0%	0.0%	100.0%
2	17.6%	27.2%	50.3%	4.9%	0.1%	.	100.0%
3	11.5%	19.3%	60.2%	9.0%	0.1%	0.0%	100.0%
4	6.3%	7.5%	63.2%	22.2%	0.7%	0.0%	100.0%
5	5.5%	7.0%	53.2%	33.3%	0.9%	0.1%	100.0%
6	9.9%	14.4%	54.5%	21.1%	0.1%	.	100.0%
7	11.2%	14.4%	54.5%	19.8%	0.1%	.	100.0%
8	15.1%	10.8%	54.4%	19.7%	0.1%	.	100.0%
9	8.1%	11.4%	48.3%	31.6%	0.7%	.	100.0%
10	9.4%	13.9%	47.8%	28.5%	0.4%	.	100.0%
11	11.4%	14.1%	47.0%	27.4%	0.2	.	100.0%
12	15.1%	10.7%	58.4%	15.7%	0.1%	.	100.0%

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, S502 Paper

Cluster	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	41,877	33,694	11,426	12,643	16,311	39,575	155,526
1	A	4,330	5,364	3,148	1,656	503	.	15,001
	BC	448	3,435	5,180	4,473	1,199	364	15,099
2	A	2,377	1,837	2,230	496	158	.	7,098
	BC	1,423	5,077	9,719	5,191	1,494	684	23,588
3	A	2,282	1,299	855	450	.	.	4,886
	BC	1,203	4,689	9,805	4,833	902	981	22,413
4-5	A	3,467	1,588	861	550	135	.	6,601
	BC	1,044	3,824	10,255	14,769	5,940	2,606	38,438
6-8	A	4,731	1,591	1,521	747	224	65	8,879
	BC	1,178	3,993	8,221	11,522	4,105	2,712	31,731
9-12	A	5,463	984	1,467	459	52	.	8,425
	BC	2,694	4,024	9,503	6,909	1,977	2,647	27,754

Table 1.3.1.4.1.2

Proficiency Level by Cluster (Percent): Speaking, S502 Paper

Cluster	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	26.9%	21.7%	7.4%	8.1%	10.5%	25.5%	100.0%
1	A	28.9%	35.8%	21.0%	11.0%	3.4%	.	100.0%
	BC	3.0%	22.8%	34.3%	29.6%	7.9%	2.4%	100.0%
2	A	33.5%	25.9%	31.4%	7.0%	2.2%	.	100.0%
	BC	6.0%	21.5%	41.2%	22.0%	6.3%	2.9%	100.0%
3	A	46.7%	26.6%	17.5%	9.2%	.	.	100.0%
	BC	5.4%	20.9%	43.8%	21.6%	4.0%	4.4%	100.0%
4-5	A	52.5%	24.1%	13.0%	8.3%	2.1%	.	100.0%
	BC	2.7%	10.0%	26.7%	38.4%	15.5%	6.8%	100.0%
6-8	A	53.3%	17.9%	17.1%	8.4%	2.5%	0.7	100.0%
	BC	3.7%	12.6%	25.9%	36.3%	12.9%	8.6%	100.0%
9-12	A	64.8%	11.7%	17.4%	5.5%	0.6%	.	100.0%
	BC	9.7%	14.5%	34.2%	24.9%	7.1%	9.5%	100.0%

1.3.1.4.2 By Grade by Tier

Table 1.3.1.4.2.1

Proficiency Level by Grade (Count): Speaking, S502 Paper

Grade	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	41,877	33,694	11,426	12,643	16,311	39,575	155,526
1	A	4,330	5,364	3,148	1,656	503	.	15,001
	BC	448	3,435	5,180	4,473	1,199	364	15,099
2	A	2,377	1,837	2,230	496	158	.	7,098
	BC	1,423	5,077	9,719	5,191	1,494	684	23,588
3	A	2,282	1,299	855	450	.	.	4,886
	BC	1,203	4,689	9,805	4,833	902	981	22,413
4	A	1,629	1,033	495	255	93	.	3,505
	BC	565	2,252	5,828	8,518	3,184	1,609	21,956
5	A	1,838	555	366	295	42	.	3,096
	BC	479	1,572	4,427	6,251	2,756	997	16,482
6	A	1,488	663	485	270	101	33	3,040
	BC	265	1,679	3,272	4,056	1,619	941	11,832
7	A	1,439	606	471	300	46	32	2,894
	BC	432	1,195	2,622	4,404	1,026	1,037	10,716
8	A	1,804	322	565	177	77	.	2,945
	BC	481	1,119	2,327	3,062	1,460	734	9,183
9	A	1,838	261	487	116	52	.	2,754
	BC	564	1,254	2,522	2,223	942	837	8,342
10	A	1,710	243	423	139	.	.	2,515
	BC	848	1,042	2,372	2,375	386	661	7,684
11	A	1,197	225	368	147	.	.	1,937
	BC	667	951	2,664	1,502	404	698	6,886
12	A	718	255	189	57	.	.	1,219
	BC	615	777	1,945	809	245	451	4,842

Table 1.3.1.4.2.2

Proficiency Level by Grade (Percent): Speaking, S502 Paper

Grade	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	26.9%	21.7%	7.4%	8.1%	10.5%	25.5%	100.0%
1	A	28.9%	35.8%	21.0%	11.0%	3.4%	.	100.0%
	BC	3.0%	22.8%	34.3%	29.6%	7.9%	2.4%	100.0%
2	A	33.5%	25.9%	31.4%	7.0%	2.2%	.	100.0%
	BC	6.0%	21.5%	41.2%	22.0%	6.3%	2.9%	100.0%
3	A	46.7%	26.6%	17.5%	9.2%	.	.	100.0%
	BC	5.4%	20.9%	43.8%	21.6%	4.0%	4.4%	100.0%
4	A	46.5%	29.5%	14.1%	7.3%	2.7%	.	100.0%
	BC	2.6%	10.3%	26.5%	38.8%	14.5%	7.3%	100.0%
5	A	59.4%	17.9%	11.8%	9.5%	1.4%	.	100.0%
	BC	2.9%	9.5%	26.9%	37.9%	16.7%	6.1%	100.0%
6	A	49.0%	21.8%	16.0%	8.9%	3.3%	1.1%	100.0%
	BC	2.2%	14.2%	27.7%	34.3%	13.7%	8.0%	100.0%
7	A	49.7%	20.9%	16.3%	10.4%	1.6%	1.1%	100.0%
	BC	4.0%	11.2%	24.5%	41.1%	9.6%	9.7%	100.0%
8	A	61.3%	10.9%	19.2%	6.0%	2.6%	.	100.0%
	BC	5.2%	12.2%	25.3%	33.3%	15.9%	8.0%	100.0%
9	A	66.7%	9.5%	17.7%	4.2%	1.9%	.	100.0%
	BC	6.8%	15.0%	30.2%	26.7%	11.3%	10.0%	100.0%
10	A	68.0%	9.7%	16.8%	5.5%	.	.	100.0%
	BC	11.0%	13.6%	30.9%	30.9%	5.0%	8.6%	100.0%
11	A	61.8%	11.6%	19.0%	7.6%	.	.	100.0%
	BC	9.7%	13.8%	38.7%	21.8%	5.9%	10.1%	100.0%
12	A	58.9%	20.9%	15.5%	4.7%	.	.	100.0%
	BC	12.7%	16.1%	40.2%	16.7%	5.1%	9.3%	100.0%

1.3.1.4.3 By Grade

Table 1.3.1.4.3.1
Proficiency Level by Grade (Count): Speaking, S502 Paper

Grade	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
K	41,877	33,694	11,426	12,643	16,311	39,575	155,526
1	4,778	8,799	8,328	6,129	1,702	364	30,100
2	3,800	6,914	11,949	5,687	1,652	684	30,686
3	3,485	5,988	10,660	5,283	902	981	27,299
4	2,194	3,285	6,323	8,773	3,277	1,609	25,461
5	2,317	2,127	4,793	6,546	2,798	997	19,578
6	1,753	2,342	3,757	4,326	1,720	974	14,872
7	1,871	1,801	3,093	4,704	1,072	1,069	13,610
8	2,285	1,441	2,892	3,239	1,537	734	12,128
9	2,402	1,515	3,009	2,339	994	837	11,096
10	2,558	1,285	2,795	2,514	386	661	10,199
11	1,864	1,176	3,032	1,649	404	698	8,823
12	1,333	1,032	2,134	866	245	451	6,061

Table 1.3.1.4.3.2
Proficiency Level by Grade (Percent): Speaking, S502 Paper

Grade	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
K	26.9%	21.7%	7.4%	8.1%	10.5%	25.5%	100.0%
1	15.9%	29.2%	27.7%	20.4%	5.7%	1.2%	100.0%
2	12.4%	22.5%	38.9%	18.5%	5.4%	2.2%	100.0%
3	12.8%	21.9%	39.1%	19.4%	3.3%	3.6%	100.0%
4	8.6%	12.9%	24.8%	34.5%	12.9%	6.3%	100.0%
5	11.8%	10.9%	24.5%	33.4%	14.3%	5.1%	100.0%
6	11.8%	15.8%	25.3%	29.1%	11.6%	6.6%	100.0%
7	13.8%	13.2%	22.7%	34.6%	7.9%	7.9%	100.0%
8	18.8%	11.9%	23.9%	26.7%	12.7%	6.1%	100.0%
9	21.7%	13.7%	27.1%	21.1%	9.0%	7.5%	100.0%
10	25.1%	12.6%	27.4%	24.7%	3.8%	6.5%	100.0%
11	21.1%	13.3%	34.4%	18.7%	4.6%	7.9%	100.0%
12	22.0%	17.0%	35.2%	14.3%	4.0%	7.4%	100.0%

1.3.2 Composites

Performance of composites is observed in their percentage in PL5 and 6 in grades: Comprehension (12-42%), Oral (7-34%), Overall (2-11%), and Literacy (0-7.4%). 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, S502 Paper

Cluster	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	47,532	23,374	18,118	13,885	23,997	28,618	155,524
1	A	1,743	2,990	4,177	2,368	917	29	12,224
	BC	129	1,236	4,154	4,008	2,861	687	13,075
2	A	1,362	1,626	2,192	1,020	184	.	6,384
	BC	180	2,106	7,474	8,051	3,321	930	22,062
3	A	1,314	1,352	1,129	543	121	9	4,468
	BC	74	1,514	7,523	8,264	3,178	702	21,255
4-5	A	2,214	1,756	1,213	771	222	25	6,201
	BC	140	1,468	7,619	14,762	9,785	3,453	37,227
6-8	A	3,560	2,479	1,339	723	201	26	8,328
	BC	229	1,736	7,168	12,184	6,633	2,592	30,542
9-12	A	4,762	1,582	1,092	328	28	.	7,792
	BC	955	3,270	8,928	9,235	3,272	1,039	26,699

Table 1.3.2.1.1.2

Proficiency Level by Cluster (Percent): Oral, S502 Paper

Cluster	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	30.6%	15.0%	11.7%	8.9%	15.4%	18.4%	100.0%
1	A	14.3%	24.5%	34.2%	19.4%	7.5%	0.2%	100.0%
	BC	1.0%	9.5%	31.8%	30.7%	21.9%	5.3%	100.0%
2	A	21.3%	25.5%	34.3%	16.0%	2.9%	.	100.0%
	BC	0.8%	9.6%	33.9%	36.5%	15.1%	4.2%	100.0%
3	A	29.4%	30.3%	25.3%	12.2%	2.7%	0.2%	100.0%
	BC	0.4%	7.1%	35.4%	38.9%	15.0%	3.3%	100.0%
4-5	A	35.7%	28.3%	19.6%	12.4%	3.6%	0.4%	100.0%
	BC	0.4%	3.9%	20.5%	39.7%	26.3%	9.3%	100.0%
6-8	A	42.8%	29.8%	16.1%	8.7%	2.4%	0.3%	100.0%
	BC	0.8%	5.7%	23.5%	39.9%	21.7%	8.5%	100.0%
9-12	A	61.1%	20.3%	14.0%	4.2%	0.4%	.	100.0%
	BC	3.6%	12.3%	33.4%	34.6%	12.3%	3.9%	100.0%

1.3.2.1.2 By Grade by Tier

Table 1.3.2.1.2.1

Proficiency Level by Grade (Count): Oral, S502 Paper

Grade	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	47,532	23,374	18,118	13,885	23,997	28,618	155,524
1	A	1,743	2,990	4,177	2,368	917	29	12,224
	BC	129	1,236	4,154	4,008	2,861	687	13,075
2	A	1,362	1,626	2,192	1,020	184	.	6,384
	BC	180	2,106	7,474	8,051	3,321	930	22,062
3	A	1,314	1,352	1,129	543	121	9	4,468
	BC	74	1,514	7,523	8,264	3,178	702	21,255
4	A	1,023	972	709	408	131	25	3,268
	BC	57	852	4,682	8,231	5,411	1,951	21,184
5	A	1,191	784	504	363	91	.	2,933
	BC	83	616	2,937	6,531	4,374	1,502	16,043
6	A	1,023	942	532	273	91	15	2,876
	BC	58	513	2,689	4,549	2,560	991	11,360
7	A	1,173	792	417	224	71	11	2,688
	BC	90	634	2,425	4,187	2,199	797	10,332
8	A	1,364	745	390	226	39	.	2,764
	BC	81	589	2,054	3,448	1,874	804	8,850
9	A	1,416	614	366	121	17	.	2,534
	BC	155	752	2,288	3,081	1,321	381	7,978
10	A	1,501	414	310	103	3	.	2,331
	BC	247	984	2,497	2,459	918	309	7,414
11	A	1,131	333	256	69	8	.	1,797
	BC	282	889	2,248	2,250	734	233	6,636
12	A	714	221	160	35	.	.	1,130
	BC	271	645	1,895	1,445	299	116	4,671

Table 1.3.2.1.2.2

Proficiency Level by Grade (Percent): Oral, S502 Paper

Grade	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	30.6%	15.0%	11.7%	8.9%	15.4%	18.4%	100.0%
1	A	14.3%	24.5%	34.2%	19.4%	7.5%	0.2%	100.0%
	BC	1.0%	9.5%	31.8%	30.7%	21.9%	5.3%	100.0%
2	A	21.3%	25.5%	34.3%	16.0%	2.9%	.	100.0%
	BC	0.8%	9.6%	33.9%	36.5%	15.1%	4.2%	100.0%
3	A	29.4%	30.3%	25.3%	12.2%	2.7%	0.2%	100.0%
	BC	0.4%	7.1%	35.4%	38.9%	15.0%	3.3%	100.0%
4	A	31.3%	29.7%	21.7%	12.5%	4.0%	0.8%	100.0%
	BC	0.3%	4.0%	22.1%	38.9%	25.5%	9.2%	100.0%
5	A	40.6%	26.7%	17.2%	12.4%	3.1%	.	100.0%
	BC	0.5%	3.8%	18.3%	40.7%	27.3%	9.4%	100.0%
6	A	35.6%	32.8%	18.5%	9.5%	3.2%	0.5%	100.0%
	BC	0.5%	4.5%	23.7%	40.0%	22.5%	8.7%	100.0%
7	A	43.6%	29.5%	15.5%	8.3%	2.6%	0.4%	100.0%
	BC	0.9%	6.1%	23.5%	40.5%	21.3%	7.7%	100.0%
8	A	49.4%	27.0%	14.1%	8.2%	1.4%	.	100.0%
	BC	0.9%	6.7%	23.2%	39.0%	21.2%	9.1%	100.0%
9	A	55.9%	24.2%	14.4%	4.8%	0.7%	.	100.0%
	BC	1.9%	9.4%	28.7%	38.6%	16.6%	4.8%	100.0%
10	A	64.4%	17.8%	13.3%	4.4%	0.1%	.	100.0%
	BC	3.3%	13.3%	33.7%	33.2%	12.4%	4.2%	100.0%
11	A	62.9%	18.5%	14.3%	3.8%	0.5%	.	100.0%
	BC	4.3%	13.4%	33.9%	33.9%	11.1%	3.5%	100.0%
12	A	63.2%	19.6%	14.2%	3.1%	.	.	100.0%
	BC	5.8%	13.8%	40.6%	30.9%	6.4%	2.5%	100.0%

1.3.2.1.3 By Grade

Table 1.3.2.1.3.1
Proficiency Level by Grade (Count): Oral, S502 Paper

Grade	Oral Language Proficiency Range						Total
	1	2	3	4	5	6	
K	47,532	23,374	18,118	13,885	23,997	28,618	155,524
1	1,872	4,226	8,331	6,376	3,778	716	25,299
2	1,542	3,732	9,666	9,071	3,505	930	28,446
3	1,388	2,866	8,652	8,807	3,299	711	25,723
4	1,080	1,824	5,391	8,639	5,542	1,976	24,452
5	1,274	1,400	3,441	6,894	4,465	1,502	18,976
6	1,081	1,455	3,221	4,822	2,651	1,006	14,236
7	1,263	1,426	2,842	4,411	2,270	808	13,020
8	1,445	1,334	2,444	3,674	1,913	804	11,614
9	1,571	1,366	2,654	3,202	1,338	381	10,512
10	1,748	1,398	2,807	2,562	921	309	9,745
11	1,413	1,222	2,504	2,319	742	233	8,433
12	985	866	2,055	1,480	299	116	5,801

Table 1.3.2.1.3.2
Proficiency Level by Grade (Percent): Oral, S502 Paper

Grade	Oral Language Proficiency Range						Total
	1	2	3	4	5	6	
K	30.6%	15.0%	11.7%	8.9%	15.4%	18.4%	100.0%
1	7.4%	16.7%	32.9%	25.2%	14.9%	2.8%	100.0%
2	5.4%	13.1%	34.0%	31.9%	12.3%	3.3%	100.0%
3	5.4%	11.1%	33.6%	34.2%	12.8%	2.8%	100.0%
4	4.4%	7.5%	22.1%	35.3%	22.7%	8.1%	100.0%
5	6.7%	7.4%	18.1%	36.3%	23.5%	7.9%	100.0%
6	7.6%	10.2%	22.6%	33.9%	18.6%	7.1%	100.0%
7	9.7%	11.0%	21.8%	33.9%	17.4%	6.2%	100.0%
8	12.4%	11.5%	21.0%	31.6%	16.5%	6.9%	100.0%
9	14.9%	13.0%	25.3%	30.5%	12.7%	3.6%	100.0%
10	17.9%	14.4%	28.8%	26.3%	9.5%	3.2%	100.0%
11	16.8%	14.5%	29.7%	27.5%	8.8%	2.8%	100.0%
12	17.0%	14.9%	35.4%	25.5%	5.2%	2.0%	100.0%

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, S502 Paper

Cluster	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	119,579	15,871	13,896	6,203	.	.	155,549
1	A	5,545	5,047	1,262	13	.	.	11,867
	BC	1,537	4,318	5,045	630	83	11	11,624
2	A	2,942	1,911	1,218	41	.	.	6,112
	BC	1,563	5,630	9,701	2,658	221	13	19,786
3	A	1,882	1,555	758	53	1	.	4,249
	BC	341	2,384	12,686	2,950	173	16	18,550
4-5	A	2,530	1,862	1,436	92	1	.	5,921
	BC	368	1,589	18,272	11,450	1,808	236	33,723
6-8	A	3,804	3,004	1,206	76	2	.	8,092
	BC	496	3,551	17,439	6,751	518	10	28,765
9-12	A	2,489	3,040	1,955	394	21	.	7,899
	BC	519	3,137	12,221	7,698	1,295	27	24,897

Table 1.3.2.2.1.2

Proficiency Level by Cluster (Percent): Literacy, S502 Paper

Cluster	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	76.9%	10.2%	8.9%	4.0%	.	.	100.0%
1	A	46.7%	42.5%	10.6%	0.1%	.	.	100.0%
	BC	13.2%	37.2%	43.4%	5.4%	0.7%	0.1%	100.0%
2	A	48.1%	31.3%	19.9%	0.7%	.	.	100.0%
	BC	7.9%	28.5%	49.0%	13.4%	1.1%	0.1%	100.0%
3	A	44.3%	36.6%	17.8%	1.3%	0.0%	.	100.0%
	BC	1.8%	12.9%	68.4%	15.9%	0.9%	0.1%	100.0%
4-5	A	42.7%	31.5%	24.3%	1.6%	0.0%	.	100.0%
	BC	1.1%	4.7%	54.2%	34.0%	5.4%	0.7%	100.0%
6-8	A	47.0%	37.1%	14.9%	0.9%	0.0%	.	100.0%
	BC	1.7%	12.3%	60.6%	23.5%	1.8%	0.0%	100.0%
9-12	A	31.5%	38.5%	24.8%	5.0%	0.3%	.	100.0%
	BC	2.1%	12.6%	49.1%	30.9%	5.2%	0.1%	100.0%

1.3.2.2.2 By Grade by Tier

Table 1.3.2.2.1
Proficiency Level by Grade (Count): Literacy, S502 Paper

Grade	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	119,579	15,871	13,896	6,203	.	.	155,549
1	A	5,545	5,047	1,262	13	.	.	11,867
	BC	1,537	4,318	5,045	630	83	11	11,624
2	A	2,942	1,911	1,218	41	.	.	6,112
	BC	1,563	5,630	9,701	2,658	221	13	19,786
3	A	1,882	1,555	758	53	1	.	4,249
	BC	341	2,384	12,686	2,950	173	16	18,550
4	A	1,334	947	788	47	1	.	3,117
	BC	254	967	11,244	5,794	645	88	18,992
5	A	1,196	915	648	45	.	.	2,804
	BC	114	622	7,028	5,656	1,163	148	14,731
6	A	1,109	1,107	495	28	1	.	2,740
	BC	151	1,301	6,646	2,276	148	3	10,525
7	A	1,200	1,008	382	25	1	.	2,616
	BC	150	1,234	5,930	2,237	169	5	9,725
8	A	1,495	889	329	23	.	.	2,736
	BC	195	1,016	4,863	2,238	201	2	8,515
9	A	665	1,025	701	155	12	.	2,558
	BC	76	611	3,507	2,639	429	17	7,279
10	A	763	891	560	127	7	.	2,348
	BC	80	788	3,381	2,216	430	7	6,902
11	A	633	686	434	89	2	.	1,844
	BC	166	868	2,969	1,919	340	3	6,265
12	A	428	438	260	23	.	.	1,149
	BC	197	870	2,364	924	96	.	4,451

Table 1.3.2.2.2.2

Proficiency Level by Grade (Percent): Literacy, S502 Paper

Grade	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	76.9%	10.2%	8.9%	4.0%	.	.	100.0%
1	A	46.7%	42.5%	10.6%	0.1%	.	.	100.0%
	BC	13.2%	37.2%	43.4%	5.4%	0.7%	0.1%	100.0%
2	A	48.1%	31.3%	19.9%	0.7%	.	.	100.0%
	BC	7.9%	28.5%	49.0%	13.4%	1.1%	0.1%	100.0%
3	A	44.3%	36.6%	17.8%	1.3%	0.0%	.	100.0%
	BC	1.8%	12.9%	68.4%	15.9%	0.9%	0.1%	100.0%
4	A	42.8%	30.4%	25.3%	1.5%	0.0%	.	100.0%
	BC	1.3%	5.1%	59.2%	30.5%	3.4%	0.5%	100.0%
5	A	42.7%	32.6%	23.1%	1.6%	.	.	100.0%
	BC	0.8%	4.2%	47.7%	38.4%	7.9%	1.0%	100.0%
6	A	40.5%	40.4%	18.1%	1.0%	0.0%	.	100.0%
	BC	1.4%	12.4%	63.1%	21.6%	1.4%	0.0%	100.0%
7	A	45.9%	38.5%	14.6%	1.0%	0.0%	.	100.0%
	BC	1.5%	12.7%	61.0%	23.0%	1.7%	0.1%	100.0%
8	A	54.6%	32.5%	12.0%	0.8%	.	.	100.0%
	BC	2.3%	11.9%	57.1%	26.3%	2.4%	0.0%	100.0%
9	A	26.0%	40.1%	27.4%	6.1%	0.5%	.	100.0%
	BC	1.0%	8.4%	48.2%	36.3%	5.9%	0.2%	100.0%
10	A	32.5%	38.0%	23.9%	5.4%	0.3%	.	100.0%
	BC	1.2%	11.4%	49.0%	32.1%	6.2%	0.1%	100.0%
11	A	34.3%	37.2%	23.5%	4.8%	0.1%	.	100.0%
	BC	2.7%	13.9%	47.4%	30.6%	5.4%	0.1%	100.0%
12	A	37.3%	38.1%	22.6%	2.0%	.	.	100.0%
	BC	4.4%	19.6%	53.1%	20.8%	2.2%	.	100.0%

1.3.2.2.3 By Grade

Table 1.3.2.2.3.1
Proficiency Level by Grade (Count): Literacy, S502 Paper

Grade	Literacy Proficiency Range						Total
	1	2	3	4	5	6	
K	119,579	15,871	13,896	6,203	.	.	155,549
1	7,082	9,365	6,307	643	83	11	23,491
2	4,505	7,541	10,919	2,699	221	13	25,898
3	2,223	3,939	13,444	3,003	174	16	22,799
4	1,588	1,914	12,032	5,841	646	88	22,109
5	1,310	1,537	7,676	5,701	1,163	148	17,535
6	1,260	2,408	7,141	2,304	149	3	13,265
7	1,350	2,242	6,312	2,262	170	5	12,341
8	1,690	1,905	5,192	2,261	201	2	11,251
9	741	1,636	4,208	2,794	441	17	9,837
10	843	1,679	3,941	2,343	437	7	9,250
11	799	1,554	3,403	2,008	342	3	8,109
12	625	1,308	2,624	947	96	.	5,600

Table 1.3.2.2.3.2
Proficiency Level by Grade (Percent): Literacy, S502 Paper

Grade	Literacy Proficiency Range						Total
	1	2	3	4	5	6	
K	76.9%	10.2%	8.9%	4.0%	.	.	100.0%
1	30.2%	39.9%	26.9%	2.7%	0.4%	0.1%	100.0%
2	17.4%	29.1%	42.2%	10.4%	0.9%	0.1%	100.0%
3	9.8%	17.3%	59.0%	13.2%	0.8%	0.1%	100.0%
4	7.2%	8.7%	54.4%	26.4%	2.9%	0.4%	100.0%
5	7.5%	8.8%	43.8%	32.5%	6.6%	0.8%	100.0%
6	9.5%	18.2%	53.8%	17.4%	1.1%	0.0%	100.0%
7	10.9%	18.2%	51.2%	18.3%	1.4%	0.0%	100.0%
8	15.0%	16.9%	46.2%	20.1%	1.8%	0.0%	100.0%
9	7.5%	16.6%	42.8%	28.4%	4.5%	0.2%	100.0%
10	9.1%	18.2%	42.6%	25.3%	4.7%	0.1%	100.0%
11	9.9%	19.2%	42.0%	24.8%	4.2%	0.0%	100.0%
12	11.2%	23.4%	46.9%	16.9%	1.7%	.	100.0%

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, S502 Paper

Cluster	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	107,919	10,214	12,020	6,305	15,170	3,921	155,549
1	A	1,785	3,781	3,174	762	530	257	10,289
	BC	20	699	4,104	2,017	2,305	1,269	10,414
2	A	1,605	2,112	1,069	396	444	26	5,652
	BC	167	3,380	5,852	3,039	3,938	2,450	18,826
3	A	971	1,908	565	193	207	125	3,969
	BC	11	452	5,464	5,094	5,075	1,804	17,900
4-5	A	2,107	1,996	752	317	385	78	5,635
	BC	21	1,822	9,002	7,168	9,308	5,632	32,953
6-8	A	3,523	2,763	858	275	239	67	7,725
	BC	127	4,566	9,588	5,869	5,332	2,466	27,948
9-12	A	3,322	2,822	794	287	198	13	7,436
	BC	238	4,636	7,533	4,651	4,420	2,751	24,229

Table 1.3.2.3.1.2

Proficiency Level by Cluster (Percent): Comprehension, S502 Paper

Cluster	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	69.4%	6.6%	7.7%	4.1%	9.8%	2.5%	100.0%
1	A	17.4%	36.8%	30.9%	7.4%	5.2%	2.5%	100.0%
	BC	0.2%	6.7%	39.4%	19.4%	22.1%	12.2%	100.0%
2	A	28.4%	37.4%	18.9%	7.0%	7.9%	0.5%	100.0%
	BC	0.9%	18.0%	31.1%	16.1%	20.9%	13.0%	100.0%
3	A	24.5%	48.1%	14.2%	4.9%	5.2%	3.2%	100.0%
	BC	0.1%	2.5%	30.5%	28.5%	28.4%	10.1%	100.0%
4-5	A	37.4%	35.4%	13.4%	5.6%	6.8%	1.4%	100.0%
	BC	0.1%	5.5%	27.3%	21.8%	28.3%	17.1%	100.0%
6-8	A	45.6%	35.8%	11.1%	3.6%	3.1%	0.9%	100.0%
	BC	0.5%	16.3%	34.3%	21.0%	19.1%	8.8%	100.0%
9-12	A	44.7%	38.0%	10.7%	3.9%	2.7%	0.2%	100.0%
	BC	1.0%	19.1%	31.1%	19.2%	18.2%	11.4%	100.0%

1.3.2.3.2 By Grade by Tier

Table 1.3.2.3.2.1

Proficiency Level by Grade (Count): Comprehension, S502 Paper

Grade	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	107,919	10,214	12,020	6,305	15,170	3,921	155,549
1	A	1,785	3,781	3,174	762	530	257	10,289
	BC	20	699	4,104	2,017	2,305	1,269	10,414
2	A	1,605	2,112	1,069	396	444	26	5,652
	BC	167	3,380	5,852	3,039	3,938	2,450	18,826
3	A	971	1,908	565	193	207	125	3,969
	BC	11	452	5,464	5,094	5,075	1,804	17,900
4	A	985	1,106	415	157	227	62	2,952
	BC	8	817	5,585	4,106	5,225	2,764	18,505
5	A	1,122	890	337	160	158	16	2,683
	BC	13	1,005	3,417	3,062	4,083	2,868	14,448
6	A	928	1,136	362	99	73	32	2,630
	BC	17	1,512	3,862	2,284	1,769	755	10,199
7	A	1,192	825	268	92	78	28	2,483
	BC	46	1,643	3,236	1,963	1,707	876	9,471
8	A	1,403	802	228	84	88	7	2,612
	BC	64	1,411	2,490	1,622	1,856	835	8,278
9	A	865	1,065	267	108	74	13	2,392
	BC	12	892	2,162	1,612	1,561	819	7,058
10	A	1,029	821	229	81	59	.	2,219
	BC	34	1,144	2,240	1,287	1,144	880	6,729
11	A	855	587	184	64	51	.	1,741
	BC	78	1,307	1,826	1,004	1,110	780	6,105
12	A	573	349	114	34	14	.	1,084
	BC	114	1,293	1,305	748	605	272	4,337

Table 1.3.2.3.2.2

Proficiency Level by Grade (Percent): Comprehension, S502 Paper

Grade	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	69.4%	6.6%	7.7%	4.1%	9.8%	2.5%	100.0%
1	A	17.4%	36.8%	30.9%	7.4%	5.2%	2.5%	100.0%
	BC	0.2%	6.7%	39.4%	19.4%	22.1%	12.2%	100.0%
2	A	28.4%	37.4%	18.9%	7.0%	7.9%	0.5%	100.0%
	BC	0.9%	18.0%	31.1%	16.1%	20.9%	13.0%	100.0%
3	A	24.5%	48.1%	14.2%	4.9%	5.2%	3.2%	100.0%
	BC	0.1%	2.5%	30.5%	28.5%	28.4%	10.1%	100.0%
4	A	33.4%	37.5%	14.1%	5.3%	7.7%	2.1%	100.0%
	BC	0.0%	4.4%	30.2%	22.2%	28.2%	14.9%	100.0%
5	A	41.8%	33.2%	12.6%	6.0%	5.9%	0.6%	100.0%
	BC	0.1%	7.0%	23.7%	21.2%	28.3%	19.9%	100.0%
6	A	35.3%	43.2%	13.8%	3.8%	2.8%	1.2%	100.0%
	BC	0.2%	14.8%	37.9%	22.4%	17.3%	7.4%	100.0%
7	A	48.0%	33.2%	10.8%	3.7%	3.1%	1.1%	100.0%
	BC	0.5%	17.4%	34.2%	20.7%	18.0%	9.3%	100.0%
8	A	53.7%	30.7%	8.7%	3.2%	3.4%	0.3%	100.0%
	BC	0.8%	17.1%	30.1%	19.6%	22.4%	10.1%	100.0%
9	A	36.2%	44.5%	11.2%	4.5%	3.1%	0.5%	100.0%
	BC	0.2%	12.6%	30.6%	22.8%	22.1%	11.6%	100.0%
10	A	46.4%	37.0%	10.3%	3.7%	2.7%	.	100.0%
	BC	0.5%	17.0%	33.3%	19.1%	17.0%	13.1%	100.0%
11	A	49.1%	33.7%	10.6%	3.7%	2.9%	.	100.0%
	BC	1.3%	21.4%	29.9%	16.5%	18.2%	12.8%	100.0%
12	A	52.9%	32.2%	10.5%	3.1%	1.3%	.	100.0%
	BC	2.6%	29.8%	30.1%	17.3%	14.0%	6.3%	100.0%

1.3.2.3.3 By Grade

Table 1.3.2.3.3.1

Proficiency Level by Grade (Count): Comprehension, S502 Paper

Grade	Comprehension Proficiency Range						Total
	1	2	3	4	5	6	
K	107,919	10,214	12,020	6,305	15,170	3,921	155,549
1	1,805	4,480	7,278	2,779	2,835	1,526	20,703
2	1,772	5,492	6,921	3,435	4,382	2,476	24,478
3	982	2,360	6,029	5,287	5,282	1,929	21,869
4	993	1,923	6,000	4,263	5,452	2,826	21,457
5	1,135	1,895	3,754	3,222	4,241	2,884	17,131
6	945	2,648	4,224	2,383	1,842	787	12,829
7	1,238	2,468	3,504	2,055	1,785	904	11,954
8	1,467	2,213	2,718	1,706	1,944	842	10,890
9	877	1,957	2,429	1,720	1,635	832	9,450
10	1,063	1,965	2,469	1,368	1,203	880	8,948
11	933	1,894	2,010	1,068	1,161	780	7,846
12	687	1,642	1,419	782	619	272	5,421

Table 1.3.2.3.3.2

Proficiency Level by Grade (Percent): Comprehension, S502 Paper

Grade	Comprehension Proficiency Range						Total
	1	2	3	4	5	6	
K	69.4%	6.6%	7.7%	4.1%	9.8%	2.5%	100.0%
1	8.7%	21.6%	35.2%	13.4%	13.7%	7.4%	100.0%
2	7.2%	22.4%	28.3%	14.0%	17.9%	10.1%	100.0%
3	4.5%	10.8%	27.6%	24.2%	24.2%	8.8%	100.0%
4	4.6%	9.0%	28.0%	19.9%	25.4%	13.2%	100.0%
5	6.6%	11.1%	21.9%	18.8%	24.8%	16.8%	100.0%
6	7.4%	20.6%	32.9%	18.6%	14.4%	6.1%	100.0%
7	10.4%	20.7%	29.3%	17.2%	14.9%	7.6%	100.0%
8	13.5%	20.3%	25.0%	15.7%	17.9%	7.7%	100.0%
9	9.3%	20.7%	25.7%	18.2%	17.3%	8.8%	100.0%
10	11.9%	22.0%	27.6%	15.3%	13.4%	9.8%	100.0%
11	11.9%	24.1%	25.6%	13.6%	14.8%	9.9%	100.0%
12	12.7%	30.3%	26.2%	14.4%	11.4%	5.0%	100.0%

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, S502 Paper

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	98,427	22,820	19,100	12,926	2,239	.	155,512
1	A	2,652	4,610	2,889	94	1	.	10,246
	BC	693	2,033	6,084	1,342	179	18	10,349
2	A	1,881	2,094	1,530	107	.	.	5,612
	BC	501	3,629	10,200	3,957	435	9	18,731
3	A	1,439	1,538	846	115	2	.	3,940
	BC	143	1,371	11,204	4,701	363	20	17,802
4-5	A	2,205	1,746	1,394	247	5	.	5,597
	BC	251	909	12,920	15,477	2,989	273	32,819
6-8	A	3,419	2,621	1,384	222	3	.	7,649
	BC	230	1,965	13,073	11,340	1,121	40	27,769
9-12	A	3,189	2,347	1,489	291	9	.	7,325
	BC	392	2,566	10,987	8,605	1,404	33	23,987

Table 1.3.2.4.1.2

Proficiency Level by Grade-Level Cluster (Percent): Overall, S502 Paper

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	63.3%	14.7%	12.3%	8.3%	1.4%	.	100.0%
1	A	25.9%	45.0%	28.2%	0.9%	0.0%	.	100.0%
	BC	6.7%	19.6%	58.8%	13.0%	1.7%	0.2%	100.0%
2	A	33.5%	37.3%	27.3%	1.9%	.	.	100.0%
	BC	2.7%	19.4%	54.5%	21.1%	2.3%	0.1%	100.0%
3	A	36.5%	39.0%	21.5%	2.9%	0.1%	.	100.0%
	BC	0.8%	7.7%	62.9%	26.4%	2.0%	0.1%	100.0%
4-5	A	39.4%	31.2%	24.9%	4.4%	0.1%	.	100.0%
	BC	0.8%	2.8%	39.4%	47.2%	9.1%	0.8%	100.0%
6-8	A	44.7%	34.3%	18.1%	2.9%	0.0%	.	100.0%
	BC	0.8%	7.1%	47.1%	40.8%	4.0%	0.1%	100.0%
9-12	A	43.5%	32.0%	20.3%	4.0%	0.1%	.	100.0%
	BC	1.6%	10.7%	45.8%	35.9%	5.9%	0.1%	100.0%

1.3.2.4.2 By Grade by Tier

Table 1.3.2.4.2.1
Proficiency Level by Grade (Count): Overall, S502 Paper

Grade	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	98,427	22,820	19,100	12,926	2,239	.	155,512
1	A	2,652	4,610	2,889	94	1	.	10,246
	BC	693	2,033	6,084	1,342	179	18	10,349
2	A	1,881	2,094	1,530	107	.	.	5,612
	BC	501	3,629	10,200	3,957	435	9	18,731
3	A	1,439	1,538	846	115	2	.	3,940
	BC	143	1,371	11,204	4,701	363	20	17,802
4	A	1,104	925	753	142	5	.	2,929
	BC	165	520	8,059	8,244	1,298	123	18,409
5	A	1,101	821	641	105	.	.	2,668
	BC	86	389	4,861	7,233	1,691	150	14,410
6	A	935	1,026	559	76	2	.	2,598
	BC	60	650	5,030	4,058	320	15	10,133
7	A	1,133	810	434	88	1	.	2,466
	BC	77	711	4,422	3,804	375	13	9,402
8	A	1,351	785	391	58	.	.	2,585
	BC	93	604	3,621	3,478	426	12	8,234
9	A	869	817	535	124	7	.	2,352
	BC	52	485	2,928	3,000	518	20	7,003
10	A	1,025	643	442	86	2	.	2,198
	BC	78	691	3,022	2,405	455	11	6,662
11	A	789	536	324	62	.	.	1,711
	BC	124	696	2,793	2,075	344	2	6,034
12	A	506	351	188	19	.	.	1,064
	BC	138	694	2,244	1,125	87	.	4,288

Table 1.3.2.4.2.2

Proficiency Level by Grade (Percent): Overall, S502 Paper

Grade	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	63.3%	14.7%	12.3%	8.3%	1.4%	.	100.0%
1	A	25.9%	45.0%	28.2%	0.9%	0.0%	.	100.0%
	BC	6.7%	19.6%	58.8%	13.0%	1.7%	0.2%	100.0%
2	A	33.5%	37.3%	27.3%	1.9%	.	.	100.0%
	BC	2.7%	19.4%	54.5%	21.1%	2.3%	0.1%	100.0%
3	A	36.5%	39.0%	21.5%	2.9%	0.1%	.	100.0%
	BC	0.8%	7.7%	62.9%	26.4%	2.0%	0.1%	100.0%
4	A	37.7%	31.6%	25.7%	4.9%	0.2%	.	100.0%
	BC	0.9%	2.8%	43.8%	44.8%	7.1%	0.7%	100.0%
5	A	41.3%	30.8%	24.0%	3.9%	.	.	100.0%
	BC	0.6%	2.7%	33.7%	50.2%	11.7	1.0%	100.0%
6	A	36.0%	39.5%	21.5%	2.9%	0.1%	.	100.0%
	BC	0.6%	6.4%	49.6%	40.1%	3.2%	0.2%	100.0%
7	A	45.9%	32.9%	17.6%	3.6%	0.0%	.	100.0%
	BC	0.8%	7.6%	47.0%	40.5%	4.0%	0.1%	100.0%
8	A	52.3%	30.4%	15.1%	2.2%	.	.	100.0%
	BC	1.1%	7.3%	44.0%	42.2%	5.2%	0.2%	100.0%
9	A	37.0%	34.7%	22.8%	5.3%	0.3%	.	100.0%
	BC	0.7%	6.9%	41.8%	42.8%	7.4%	0.3%	100.0%
10	A	46.6%	29.3%	20.1%	3.9%	0.1%	.	100.0%
	BC	1.2%	10.4%	45.4%	36.1%	6.8%	0.2%	100.0%
11	A	46.1%	31.3%	18.9%	3.6%	.	.	100.0%
	BC	2.1%	11.5%	46.3%	34.4%	5.7%	0.0%	100.0%
12	A	47.6%	33.0%	17.7%	1.8%	.	.	100.0%
	BC	3.2%	16.2%	52.3%	26.2%	2.0%	.	100.0%

1.3.2.4.3 By Grade

Table 1.3.2.4.3.1
Proficiency Level by Grade (Count): Overall, S502 Paper

Grade	Overall Proficiency Range						Total
	1	2	3	4	5	6	
K	98,427	22,820	19,100	12,926	2,239	.	155,512
1	3,345	6,643	8,973	1,436	180	18	20,595
2	2,382	5,723	11,730	4,064	435	9	24,343
3	1,582	2,909	12,050	4,816	365	20	21,742
4	1,269	1,445	8,812	8,386	1,303	123	21,338
5	1,187	1,210	5,502	7,338	1,691	150	17,078
6	995	1,676	5,589	4,134	322	15	12,731
7	1,210	1,521	4,856	3,892	376	13	11,868
8	1,444	1,389	4,012	3,536	426	12	10,819
9	921	1,302	3,463	3,124	525	20	9,355
10	1,103	1,334	3,464	2,491	457	11	8,860
11	913	1,232	3,117	2,137	344	2	7,745
12	644	1,045	2,432	1,144	87	.	5,352

Table 1.3.2.4.3.2
Proficiency Level by Grade (Percent): Overall, S502 Paper

Grade	Overall Proficiency Range						Total
	1	2	3	4	5	6	
K	63.3%	14.7%	12.3%	8.3%	1.4%	.	100.0%
1	16.2%	32.3%	43.6%	7.0%	0.9%	0.1%	100.0%
2	9.8%	23.5%	48.2%	16.7%	1.8%	0.0%	100.0%
3	7.3%	13.4%	55.4%	22.2%	1.7%	0.1%	100.0%
4	6.0%	6.8%	41.3%	39.3%	6.1%	0.6%	100.0%
5	7.0%	7.1%	32.2%	43.0%	9.9%	0.9%	100.0%
6	7.8%	13.2%	43.9%	32.5%	2.5%	0.1%	100.0%
7	10.2%	12.8%	40.9%	32.8%	3.2%	0.1%	100.0%
8	13.4%	12.8%	37.1%	32.7%	3.9%	0.1%	100.0%
9	9.9%	13.9%	37.0%	33.4%	5.6%	0.2%	100.0%
10	12.5%	15.1%	39.1%	28.1%	5.2%	0.1%	100.0%
11	11.8%	15.9%	40.3%	27.6%	4.4%	0.0%	100.0%
12	12.0%	19.5%	45.4%	21.4%	1.6%	.	100.0%

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)

β_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 4 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 to 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 CAL, 2017).

For Speaking, we model PL 1 tasks as a group on a 0 to 2 scale, and PL 3 and PL 5 tasks as a group on a 0 to 4 scale (see Part 1, Section 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 (CAL, 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:

- Listening: (Ability Measure in Logits * 37.571) + 316.637
- Reading: (Ability Measure in Logits * 26.000) + 323.272
- Writing: (Ability Measure in Logits * 26.851) + 303.332
- Speaking: (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 (CAL, 2016a). In the domains of Writing and Speaking, we conducted a study in the summer of 2016 to reconstruct the logit scale (see CAL, 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:

- Listening: (Ability Measure in Logits * 37.571) + 316.637
- Reading: (Ability Measure in Logits * 26.000) + 323.272
- Writing: (Ability Measure in Logits * 31.097) + 317.068
- Speaking: (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 the Center for Applied Linguistics (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 PL 4 and PL 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, 3.8 for a Grade 5 student, 3.1 for a Grade 8 student, and 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* (Linacre, 2006). When speaking of the measure of student ability, we use the term *ability measure* (rather than *theta* used commonly when discussing models based on item response theory). When speaking of the measure of how hard an item is, we use the term *item difficulty measure* (rather than *b parameter* used commonly when discussing models based on item response theory). *Step measures* refer to the calibration of the steps in the Rasch rating scale model previously presented. All three measures (ability, difficulty, and step) are expressed in terms of Rasch logits, which then are converted into scores on the ACCESS score scale for reporting purposes.

Fit statistics for the Rasch model are calculated by comparing the observed empirical data with the data that the Rasch model would be expected to produce if the data fit the model perfectly. Outfit mean square statistics for items and tasks are influenced by outlier responses for machine-scored dichotomous items or outlier ratings for rater-scored performance tasks. For example, a difficult item that some low-ability students get correct—for reasons unknown—will have a high outfit mean square statistic. Similarly, an easy item that some high-ability students get wrong will also have a high outfit mean square statistic. Infit mean square statistics are influenced by unexpected patterns of students' responses and ratings on items and tasks that are roughly targeted for them and generally indicate a more serious measurement problem. The expectation for both statistics is 1.00, and values near 1.00 are not of great concern. Values less than 1.00 indicate that the response and rating patterns are too predictable and thus redundant 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 degrade¹ the measurement system.”
- Values between 1.5 and 2.0 are “unproductive for construction of measurement, but not degrading.”

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

- 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.

Because we followed conservative guidelines in the development of ACCESS for ELLs, the vast majority of dichotomous items on the test forms have mean square fit statistics in the range of 0.5 to 1.5; thus, they fit the range that is “productive for measurement” according to the guidelines above.

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. In order 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 dichotomously scored items (Listening and Reading), the next column shows the p value (percentage of correct answers on that item). 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 most items and tasks provide good measurement for students around the ability range 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. Four items in the Listening domain, 11 items in the Reading domain, one task 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.

2.2 DIF Analysis and Summary

Differential item functioning (DIF) analysis investigates 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. We exclude students for whom gender or ethnicity² 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. 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 one group is favored; a negative result indicates that the other 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:

- 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

² 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.

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 this, the absolute value is taken in JMetrik (Meyer, 2018), thereby restricting the range of the rescaled effect size (standardized P-DIF*) 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 ≥ 0.05 but <0.10
- CC (strong DIF) when the Cochran-Mantel-Haenszel chi-square statistic is significant and standardized P-DIF* is ≥ 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 versus female or Hispanic versus non-Hispanic ethnicities. 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.

Panel members were drawn from CAL staff members who have expertise in instruction and/or professional development for English learner students. The panel included a mix of women and men and included CAL staff who have a language other than English as a first language, with attention paid to ensuring representation of individuals from Spanish-language backgrounds and non-Spanish-language backgrounds. The facilitator asked the panel to discuss the item and come to consensus on whether the item demonstrates bias against a particular group and is appropriate to place on the operational test. The facilitator does not disclose to the panel which subgroup the DIF analysis indicates is favored by the item.

One item showed a C-level DIF. The item, on the Grades 9–12 Listening Tier A test, showed a C-level DIF favoring non-Hispanic students. However, the panel concluded that the item showed bias in favor of Spanish-speaking students because the English idiom “hang out” is borrowed into some Spanish dialects, although the panel noted that this did not apply to all Spanish speakers. As mentioned above, DIF analysis is done using data from the first used form. Please see section 2.7, below, for further explanation of the operational history of static forms.

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.

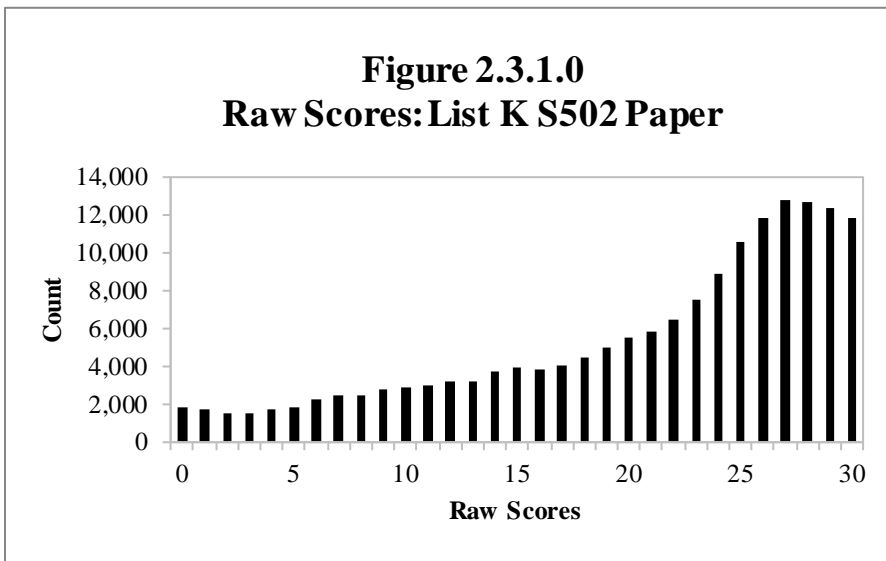
2.3.1 Listening

2.3.1.0 Kindergarten

Table 2.3.1.0

Raw Score Descriptive Statistics: List K S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	163,226	0	30	20.83	7.95
Total	163,226	0	30	20.83	7.95



2.3.1.1 Grade 1

Table 2.3.1.1.1

Raw Score Descriptive Statistics: List 1 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	12,416	1	18	13.51	3.17
Total	12,416	1	18	13.51	3.17

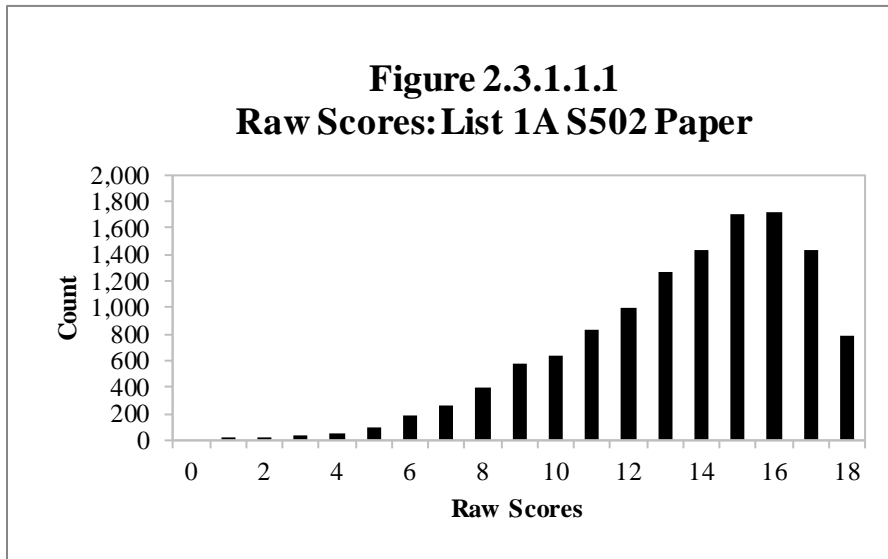
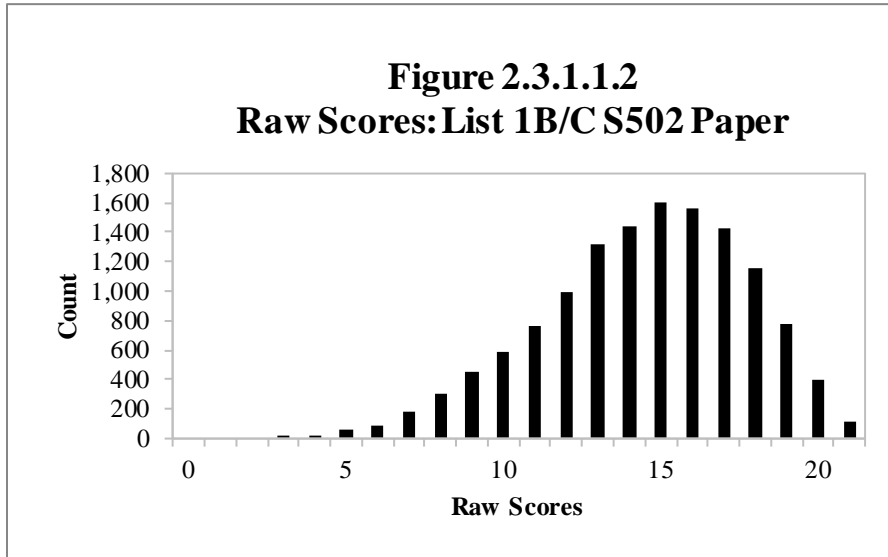


Table 2.3.1.1.2

Raw Score Descriptive Statistics: List 1 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	13,243	3	21	14.44	3.27
Total	13,243	3	21	14.44	3.27



2.3.1.2 Grade 2

Table 2.3.1.2.1

Raw Score Descriptive Statistics: List 2 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	6,498	2	18	14.37	3.21
Total	6,498	2	18	14.37	3.21

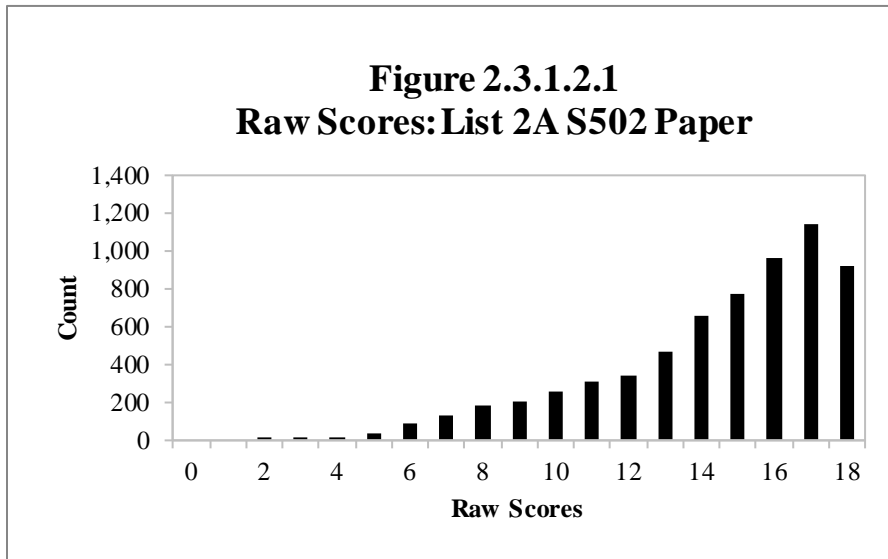
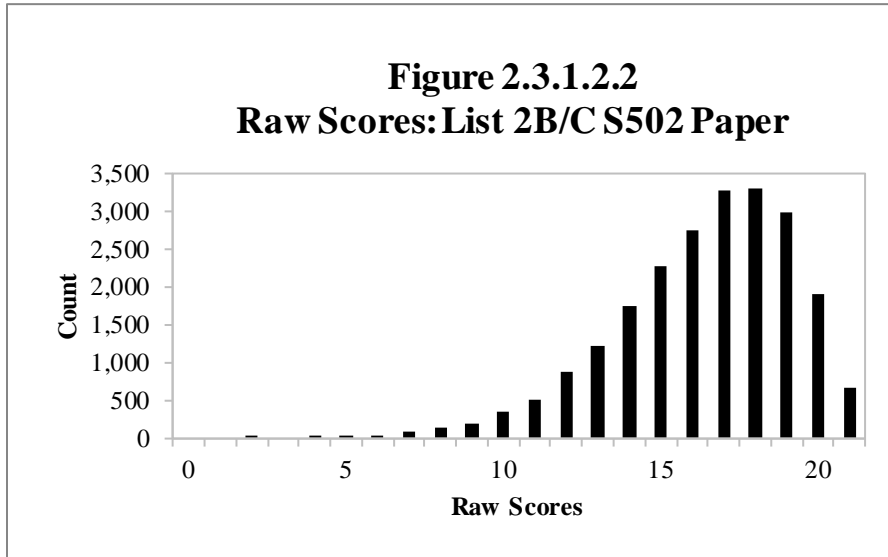


Table 2.3.1.2.2

Raw Score Descriptive Statistics: List 2 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	22,315	2	21	16.37	2.83
Total	22,315	2	21	16.37	2.83



2.3.1.3 Grade 3

Table 2.3.1.3.1

Raw Score Descriptive Statistics: List 3 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	4,546	1	18	10.43	3.49
Total	4,546	1	18	10.43	3.49

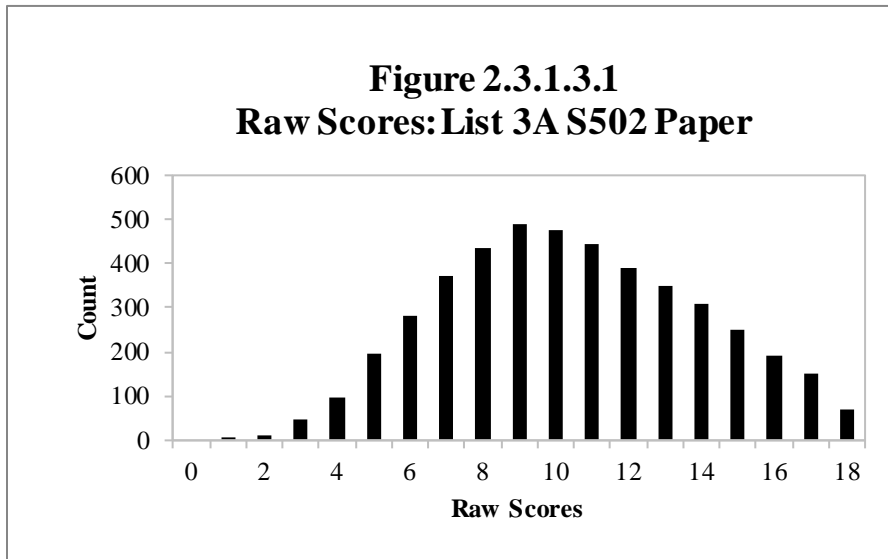
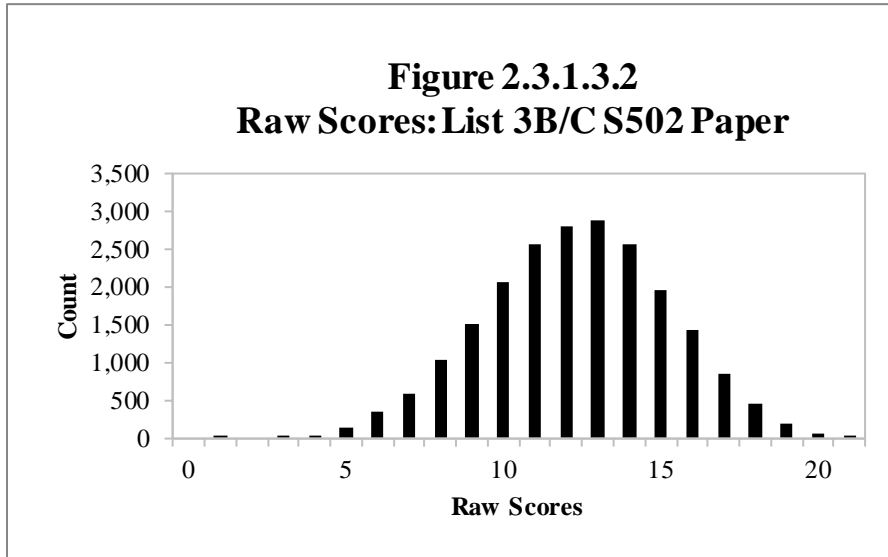


Table 2.3.1.3.2

Raw Score Descriptive Statistics: List 3 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	21,505	1	21	12.31	2.91
Total	21,505	1	21	12.31	2.91

Figure 2.3.1.3.2
Raw Scores: List 3B/C S502 Paper



2.3.1.4 Grades 4–5

Table 2.3.1.4.1

Raw Score Descriptive Statistics: List 4-5 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	3,322	1	18	11.14	3.59
5	2,964	1	18	11.38	3.59
Total	6,286	1	18	11.25	3.59

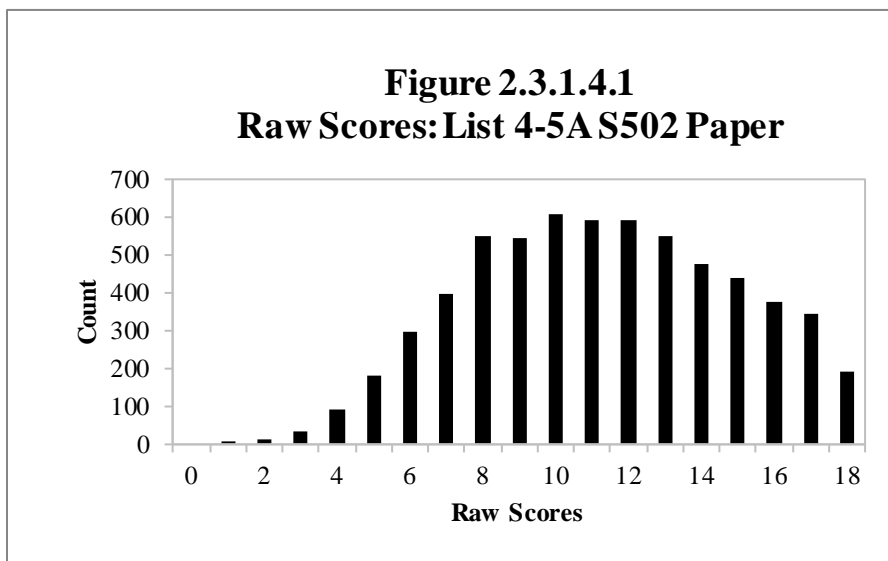
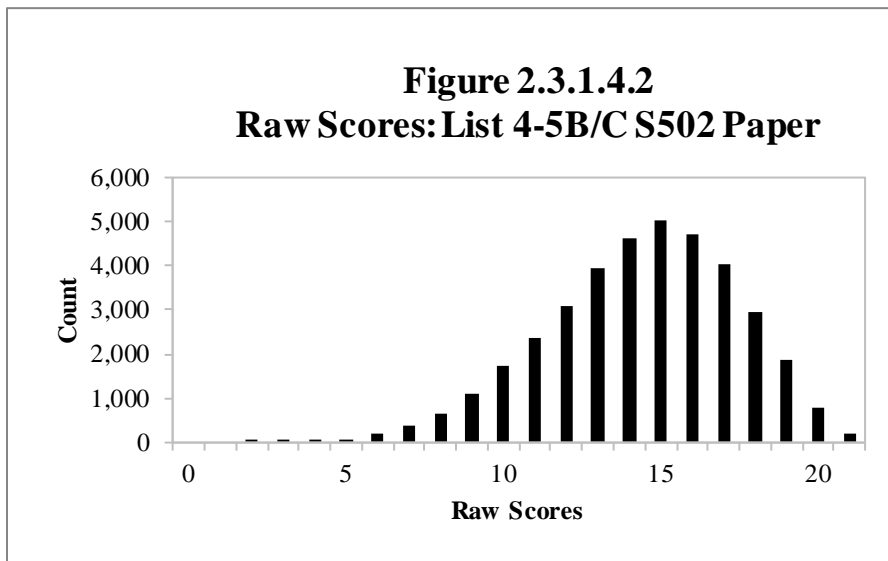


Table 2.3.1.4.2

Raw Score Descriptive Statistics: List 4-5 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	21,406	2	21	13.91	2.98
5	16,195	3	21	15.06	2.91
Total	37,601	2	21	14.41	3.00

Figure 2.3.1.4.2
Raw Scores: List 4-5B/C S502 Paper



2.3.1.5 Grades 6–8

Table 2.3.1.5.1

Raw Score Descriptive Statistics: List 6-8 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,923	2	18	10.17	3.37
7	2,720	1	18	10.20	3.47
8	2,799	2	18	10.23	3.46
Total	8,442	1	18	10.20	3.43

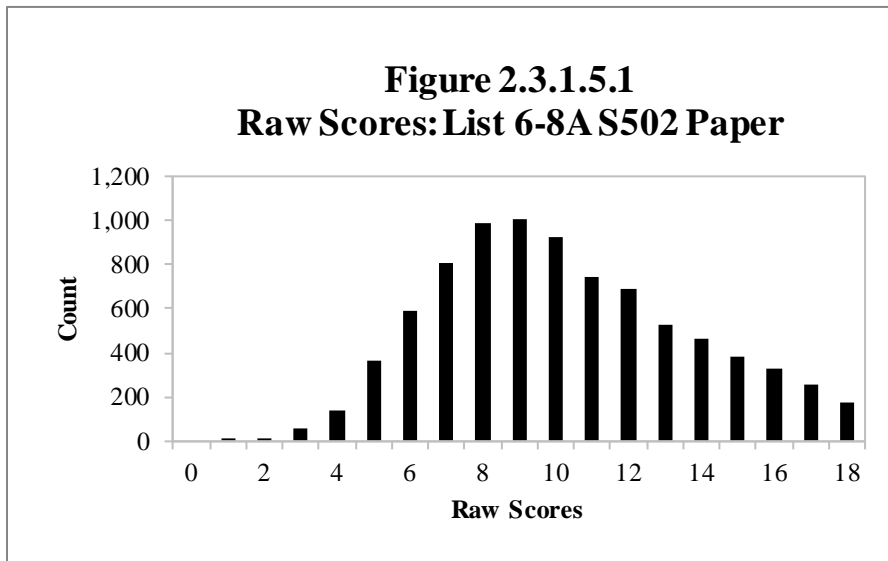
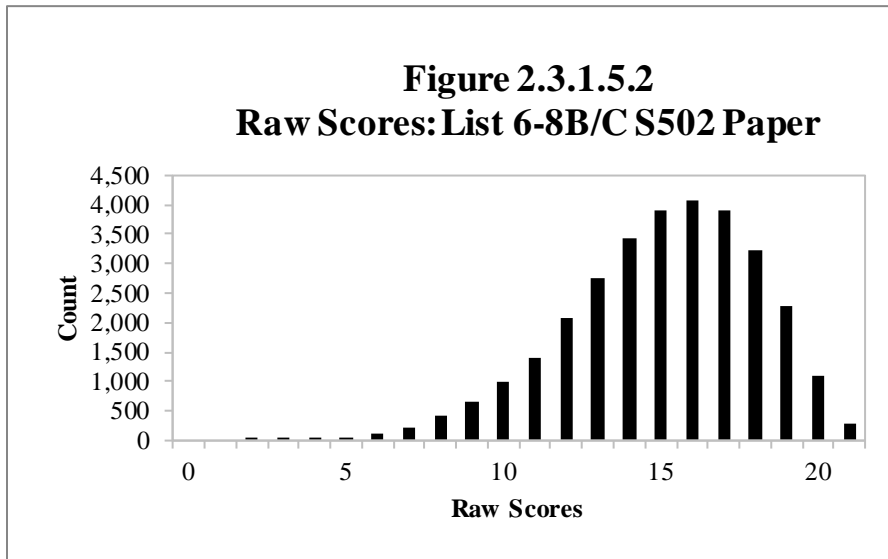


Table 2.3.1.5.2

Raw Score Descriptive Statistics: List 6-8 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	11,485	2	21	14.48	2.93
7	10,478	3	21	15.14	2.98
8	8,953	2	21	15.68	3.00
Total	30,916	2	21	15.05	3.01



2.3.1.6 Grades 9–12

Table 2.3.1.6.1

Raw Score Descriptive Statistics: List 9-12 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	2,579	0	18	10.48	2.98
10	2,360	2	18	10.47	2.92
11	1,832	3	18	10.87	2.88
12	1,156	4	18	10.97	2.78
Total	7,927	0	18	10.64	2.92

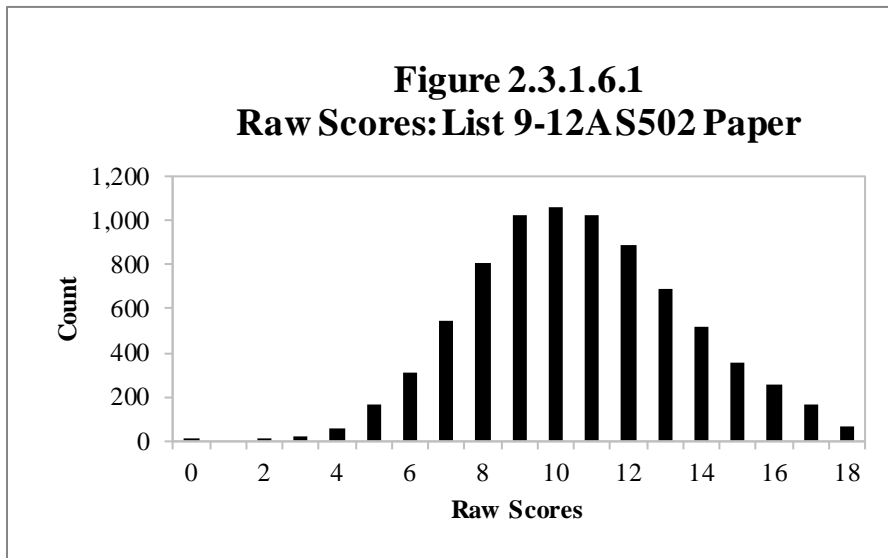
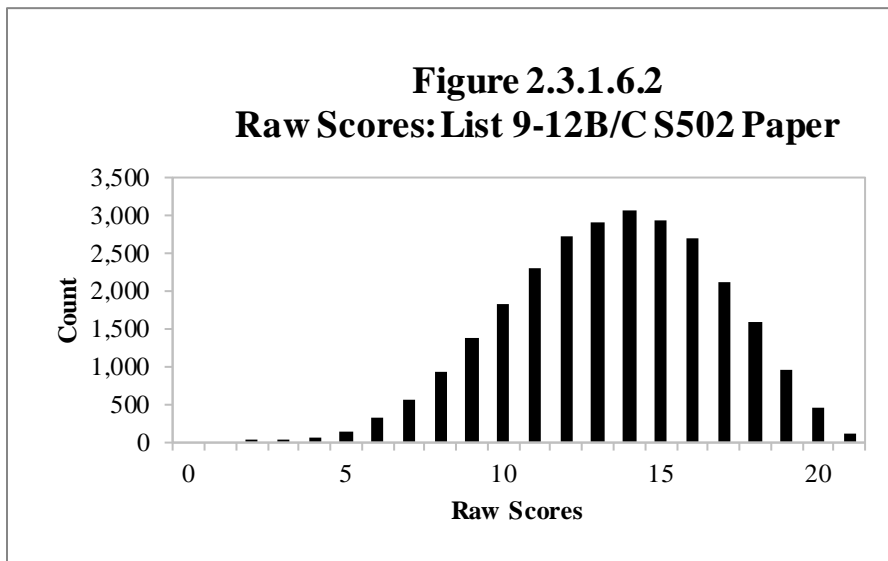


Table 2.3.1.6.2

Raw Score Descriptive Statistics: List 9-12 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	8,049	2	21	13.44	3.21
10	7,505	2	21	13.44	3.35
11	6,745	2	21	13.66	3.41
12	4,751	2	21	13.27	3.34
Total	27,050	2	21	13.47	3.32



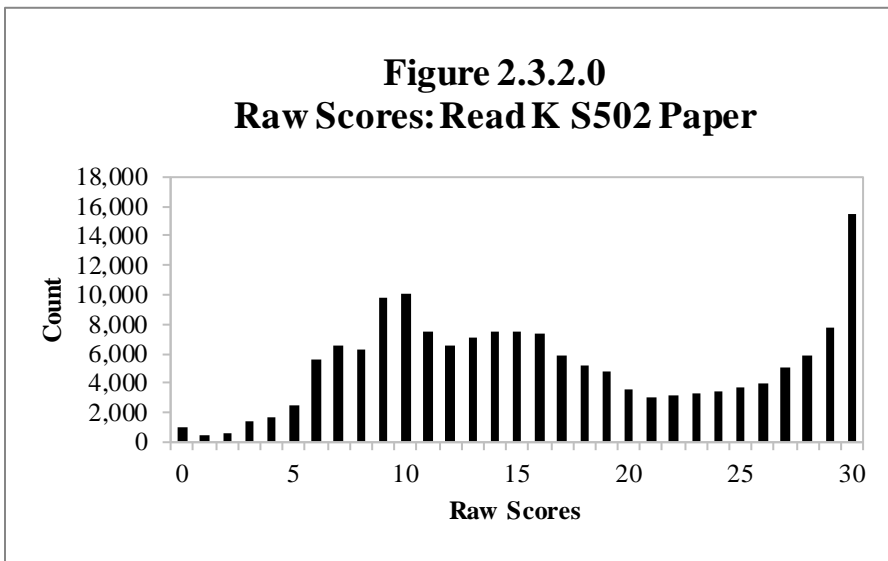
2.3.2 Reading

2.3.2.0 Kindergarten

Table 2.3.2.0

Raw Score Descriptive Statistics: Read K S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	163,218	0	30	16.91	8.21
Total	163,218	0	30	16.91	8.21



2.3.2.1 Grade 1

Table 2.3.2.1.1

Raw Score Descriptive Statistics: Read 1 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	12,012	0	24	10.53	3.98
Total	12,012	0	24	10.53	3.98

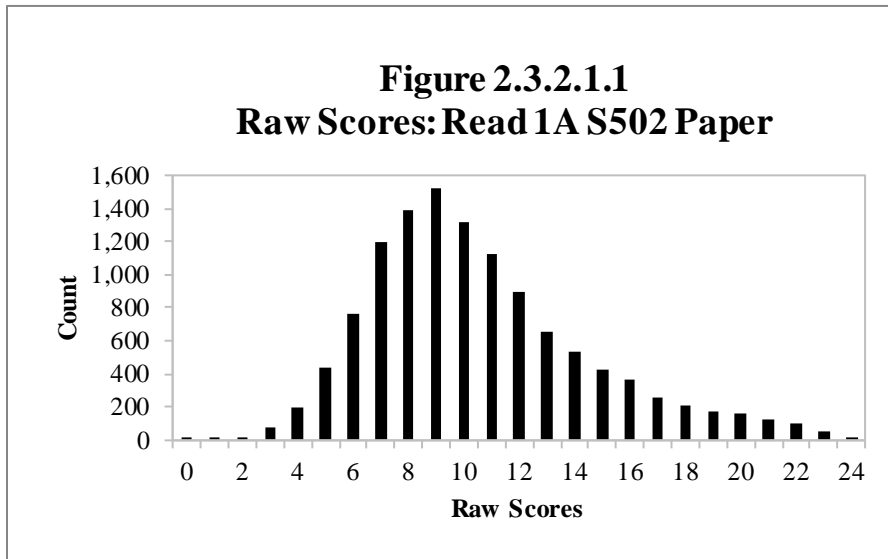
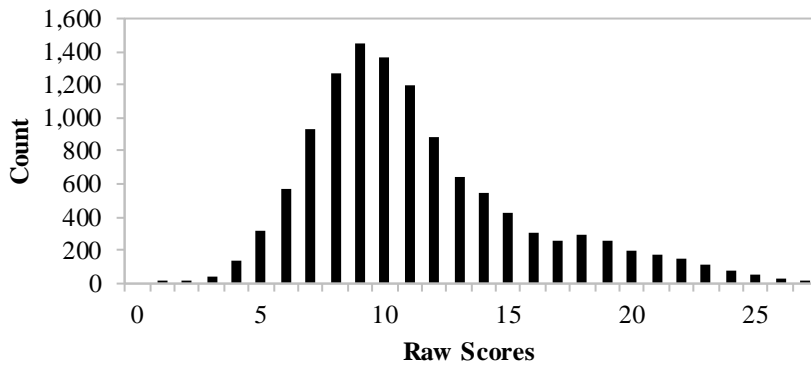


Table 2.3.2.1.2

Raw Score Descriptive Statistics: Read 1 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	11,702	1	27	11.30	4.46
Total	11,702	1	27	11.30	4.46

Figure 2.3.2.1.2
Raw Scores: Read 1B/C S502 Paper



2.3.2.2 Grade 2

Table 2.3.2.2.1

Raw Score Descriptive Statistics: Read 2 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	6,181	1	24	12.75	4.94
Total	6,181	1	24	12.75	4.94

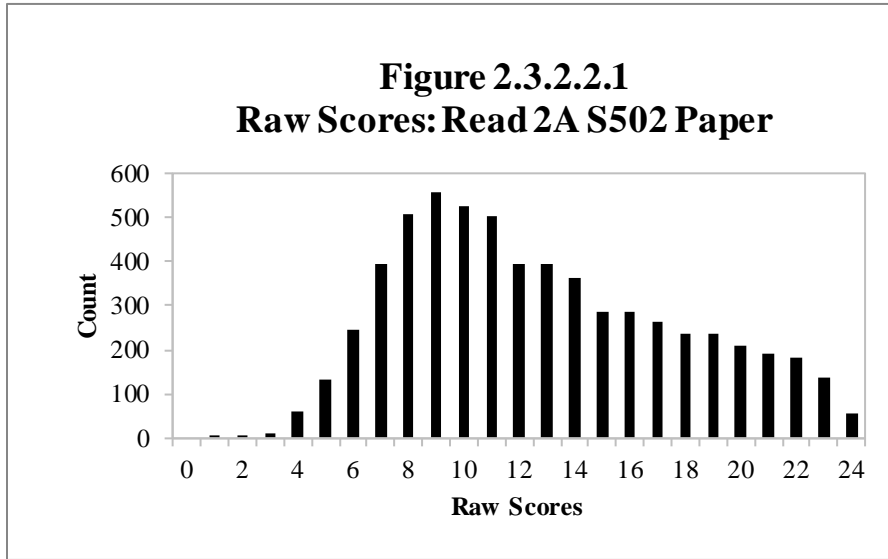
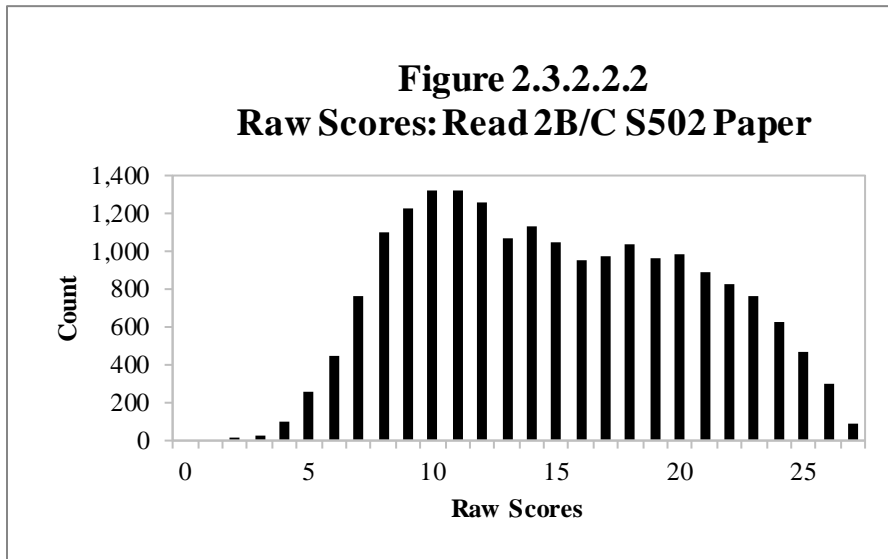


Table 2.3.2.2.2

Raw Score Descriptive Statistics: Read 2 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	19,910	2	27	14.92	5.57
Total	19,910	2	27	14.92	5.57

Figure 2.3.2.2.2
Raw Scores: Read 2B/C S502 Paper



2.3.2.3 Grade 3

Table 2.3.2.3.1

Raw Score Descriptive Statistics: Read 3 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	4,291	2	24	11.16	4.56
Total	4,291	2	24	11.16	4.56

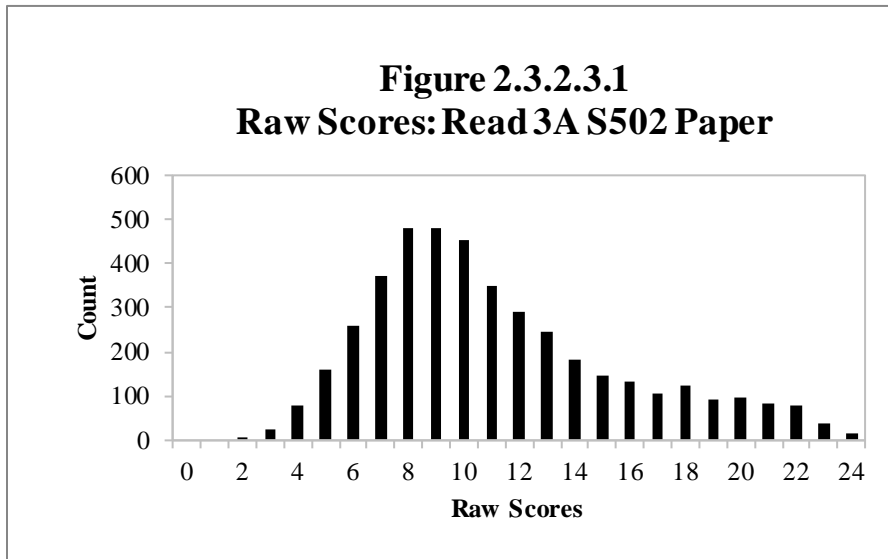
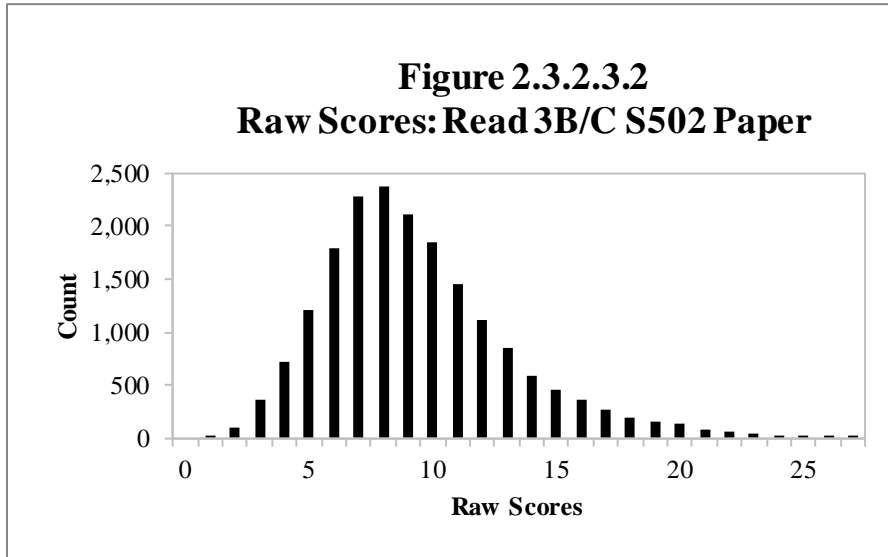


Table 2.3.2.3.2

Raw Score Descriptive Statistics: Read 3 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	18,665	1	27	9.33	3.82
Total	18,665	1	27	9.33	3.82



2.3.2.4 Grades 4–5

Table 2.3.2.4.1

Raw Score Descriptive Statistics: Read 4-5 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	3,144	2	24	12.76	5.06
5	2,819	2	24	13.59	5.06
Total	5,963	2	24	13.15	5.08

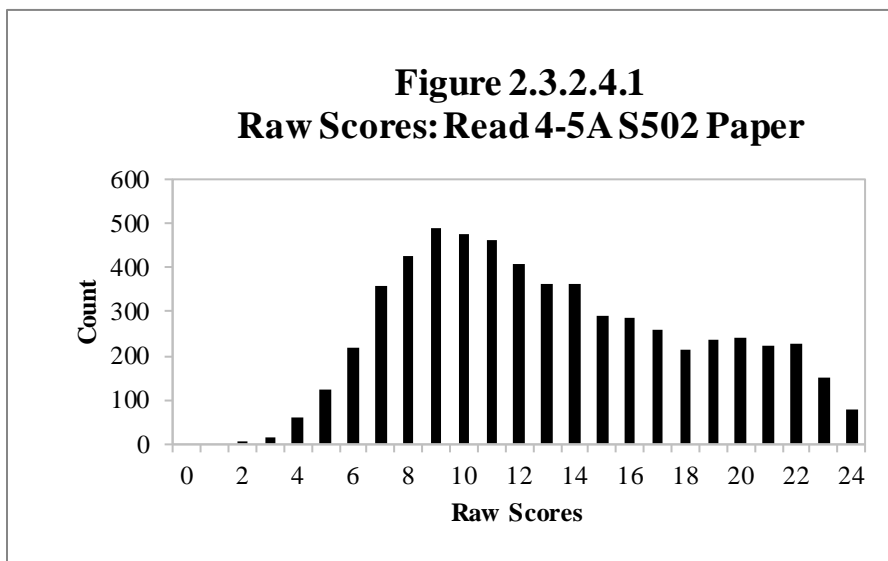
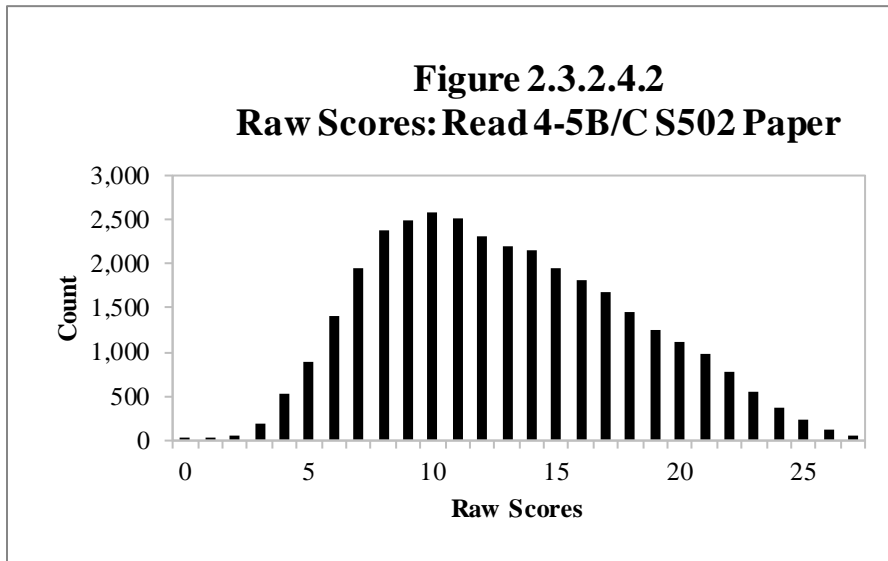


Table 2.3.2.4.2

Raw Score Descriptive Statistics: Read 4-5 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	19,115	0	27	11.87	4.75
5	14,832	1	27	14.26	5.19
Total	33,947	0	27	12.92	5.09

Figure 2.3.2.4.2
Raw Scores: Read 4-5B/C S502 Paper



2.3.2.5 Grades 6–8

Table 2.3.2.5.1

Raw Score Descriptive Statistics: Read 6-8 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,751	1	24	10.67	4.31
7	2,627	2	24	11.06	4.62
8	2,743	1	24	11.39	4.66
Total	8,121	1	24	11.04	4.54

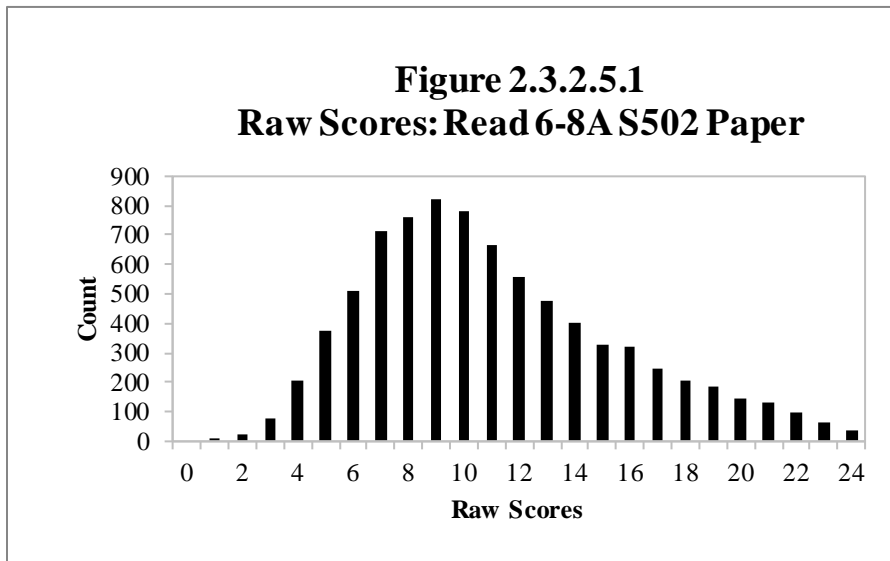
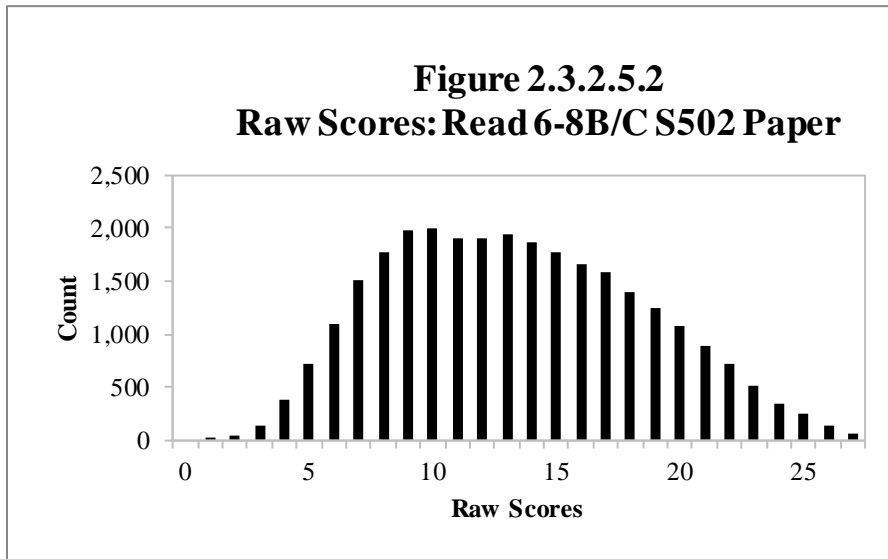


Table 2.3.2.5.2

Raw Score Descriptive Statistics: Read 6-8 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	10,577	1	27	11.86	4.58
7	9,789	1	27	13.46	5.07
8	8,569	2	27	15.13	5.33
Total	28,935	1	27	13.37	5.15

Figure 2.3.2.5.2
Raw Scores: Read 6-8B/C S502 Paper



2.3.2.6 Grades 9–12

Table 2.3.2.6.1

Raw Score Descriptive Statistics: Read 9-12 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	2,560	2	24	13.12	4.53
10	2,355	2	24	13.29	4.53
11	1,846	3	24	14.19	4.57
12	1,154	1	24	14.58	4.37
Total	7,915	1	24	13.63	4.55

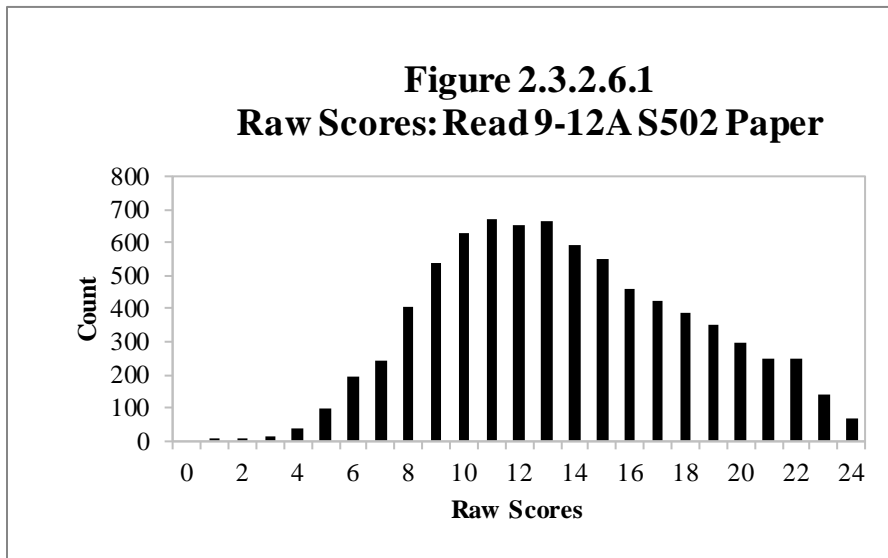
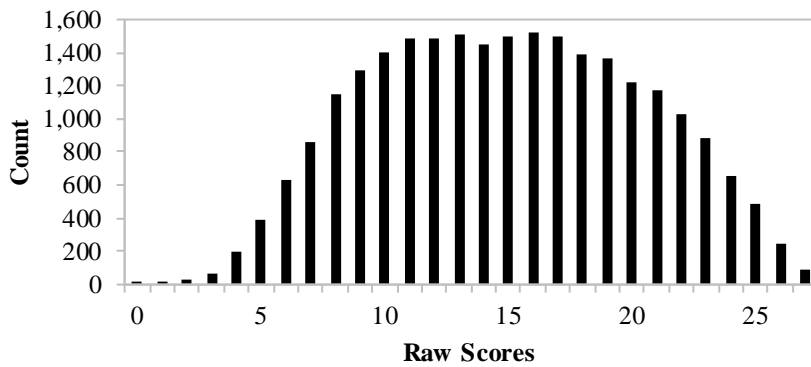


Table 2.3.2.6.2

Raw Score Descriptive Statistics: Read 9-12 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	7,289	2	27	14.35	5.17
10	6,928	0	27	14.95	5.42
11	6,302	1	27	15.63	5.55
12	4,476	2	27	14.92	5.45
Total	24,995	0	27	14.94	5.41

Figure 2.3.2.6.2
Raw Scores: Read 9-12B/C S502 Paper



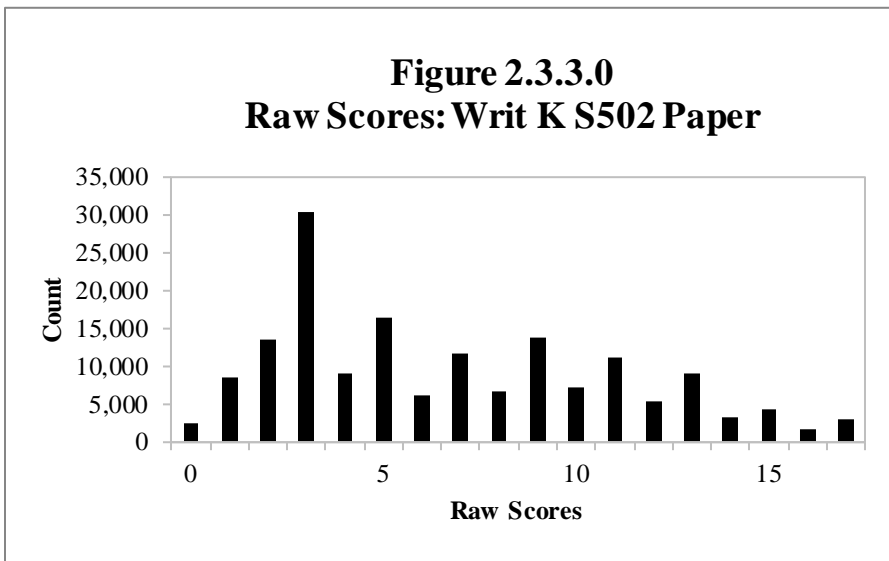
2.3.3 Writing

2.3.3.0 Kindergarten

Table 2.3.3.0

Raw Score Descriptive Statistics: Writ K S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	163,216	0	17	6.77	4.31
Total	163,216	0	17	6.77	4.31



2.3.3.1 Grade 1

Table 2.3.3.1.1

Raw Score Descriptive Statistics: Writ 1 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,252	0	25	10.68	5.91
Total	15,252	0	25	10.68	5.91

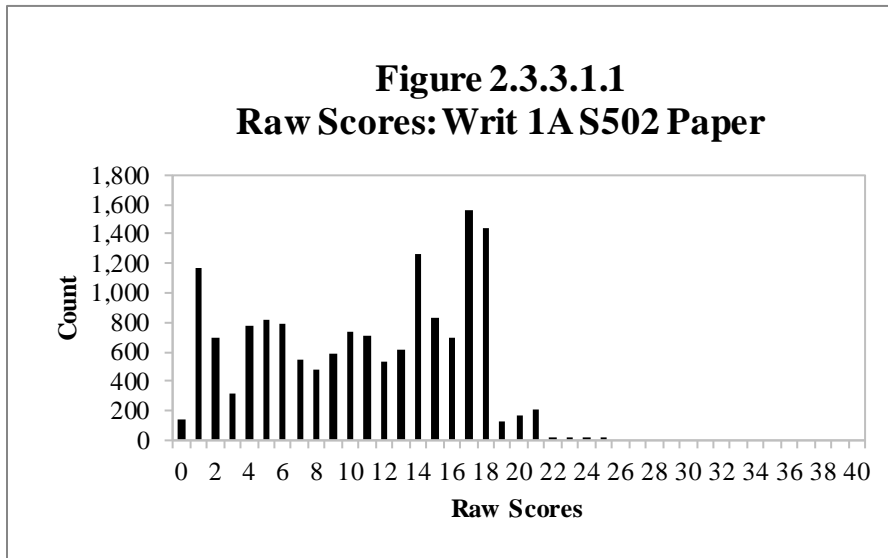
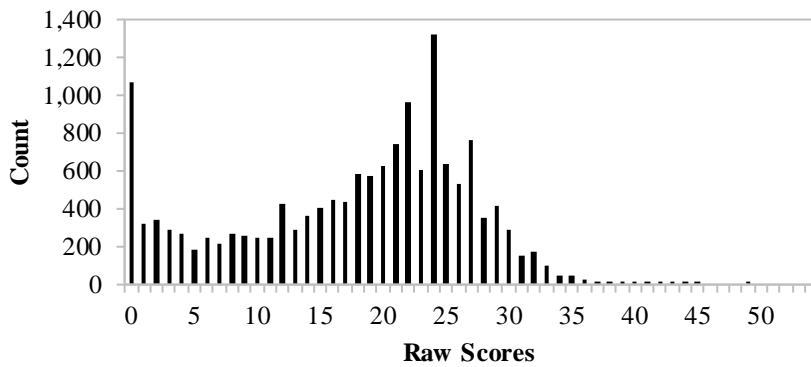


Table 2.3.3.1.2

Raw Score Descriptive Statistics: Writ 1 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,300	0	49	17.52	9.32
Total	15,300	0	49	17.52	9.32

Figure 2.3.3.1.2
Raw Scores: Writ 1B/C S502 Paper



2.3.3.2 Grade 2

Table 2.3.3.2.1

Raw Score Descriptive Statistics: Writ 2 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	7,231	0	20	7.31	4.80
Total	7,231	0	20	7.31	4.80

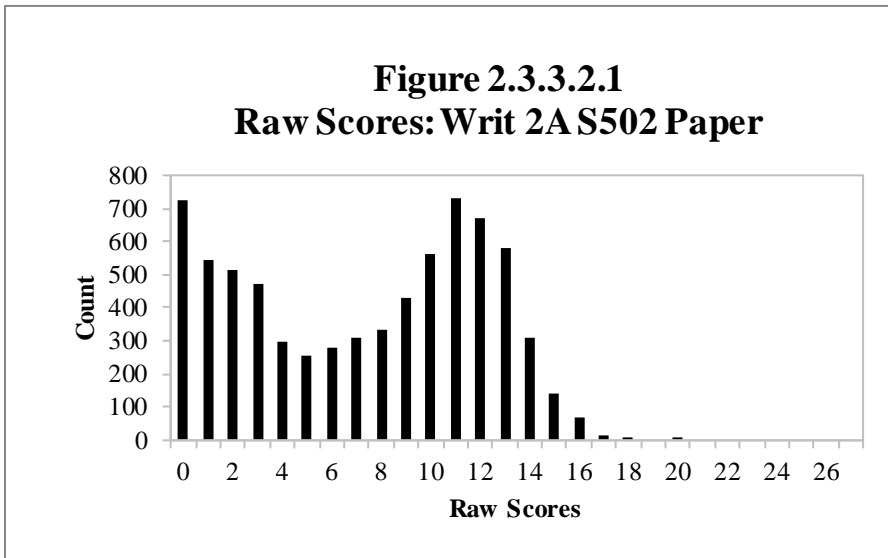
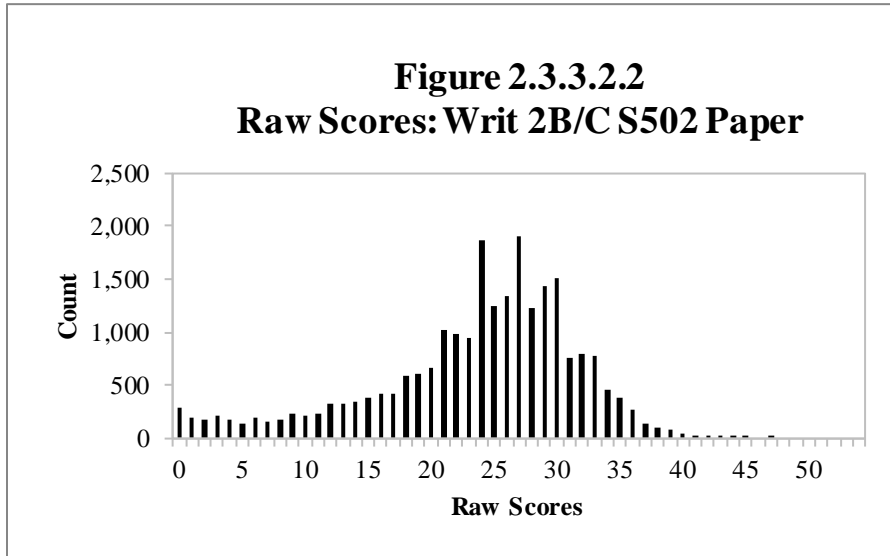


Table 2.3.3.2.2

Raw Score Descriptive Statistics: Writ 2 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	23,861	0	47	23.49	8.23
Total	23,861	0	47	23.49	8.23



2.3.3.3 Grade 3

Table 2.3.3.3.1

Raw Score Descriptive Statistics: Writ 3 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	4,974	0	21	8.28	4.84
Total	4,974	0	21	8.28	4.84

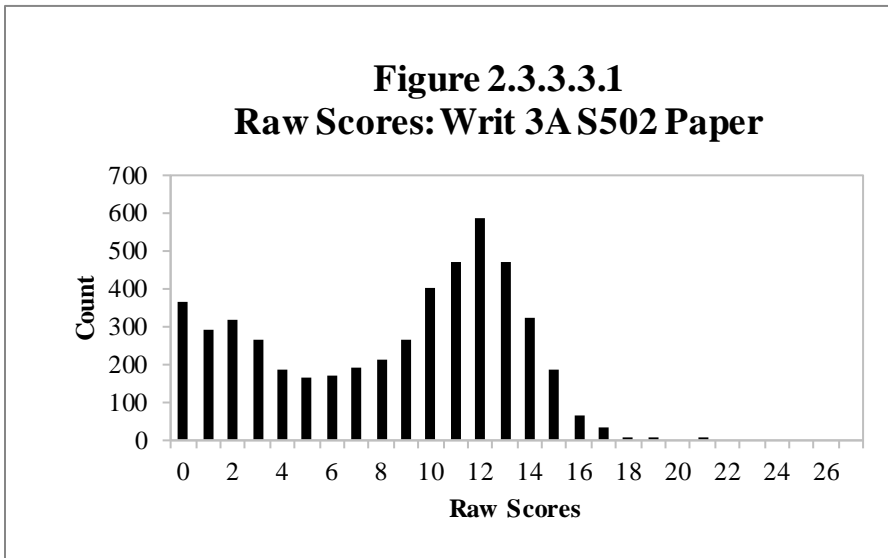
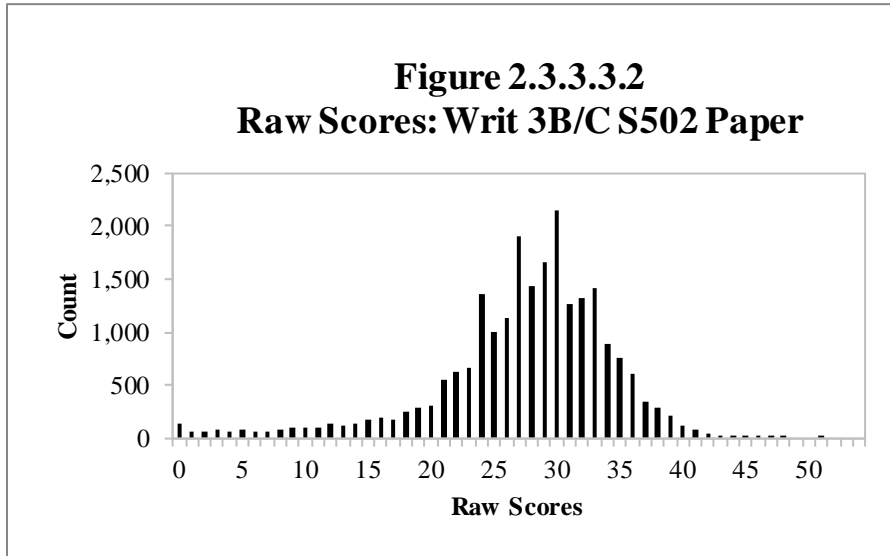


Table 2.3.3.3.2

Raw Score Descriptive Statistics: Writ 3 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	22,675	0	51	27.28	7.15
Total	22,675	0	51	27.28	7.15



2.3.3.4 Grades 4–5

Table 2.3.3.4.1

Raw Score Descriptive Statistics: Writ 4-5 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	3,564	0	19	7.44	4.34
5	3,129	0	19	8.17	4.26
Total	6,693	0	19	7.78	4.32

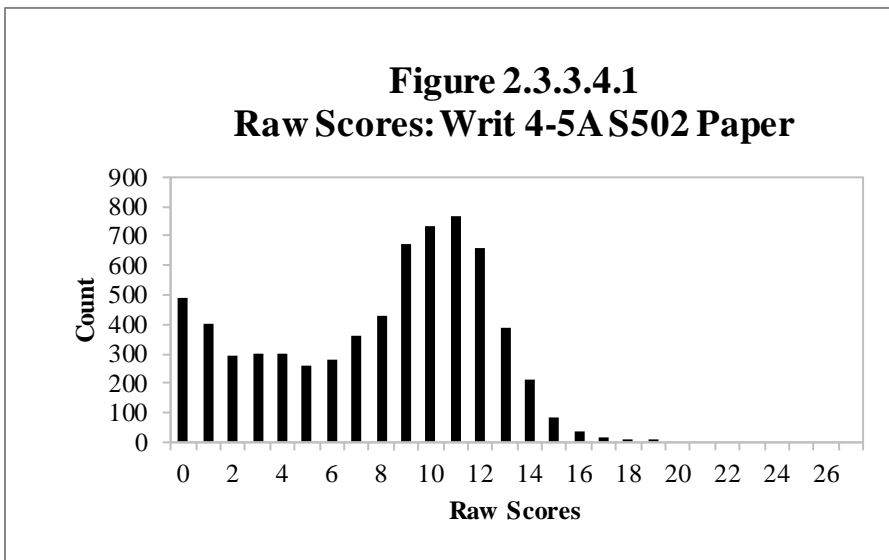
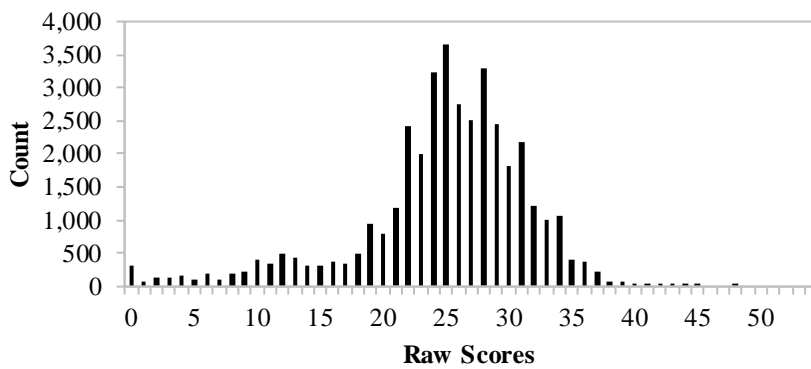


Table 2.3.3.4.2

Raw Score Descriptive Statistics: Writ 4-5 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	22,186	0	43	23.35	6.81
5	16,641	0	48	26.59	6.45
Total	38,827	0	48	24.74	6.85

Figure 2.3.3.4.2
Raw Scores: Writ 4-5B/C S502 Paper



2.3.3.5 Grades 6–8

Table 2.3.3.5.1

Raw Score Descriptive Statistics: Writ 6-8 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	3,088	0	20	8.49	4.08
7	2,925	0	20	8.85	4.00
8	2,979	0	20	9.13	3.91
Total	8,992	0	20	8.82	4.01

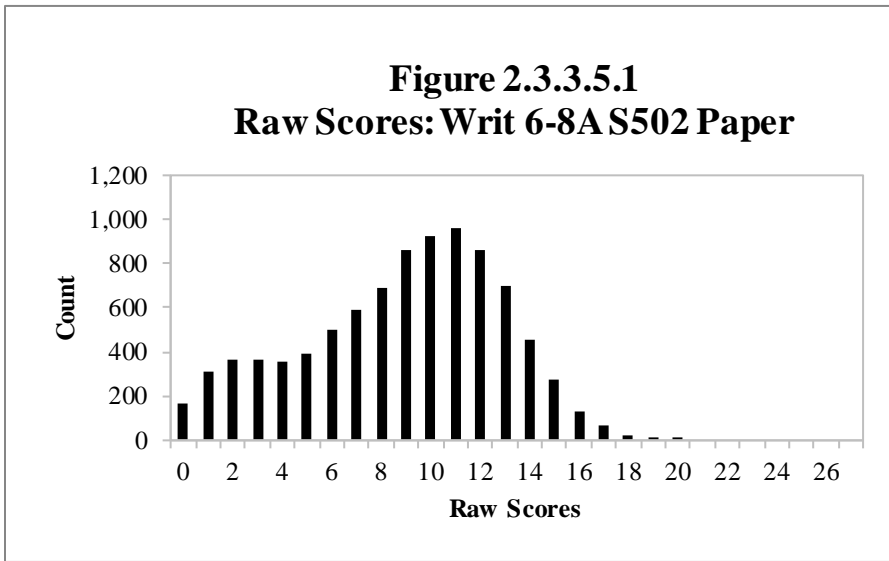
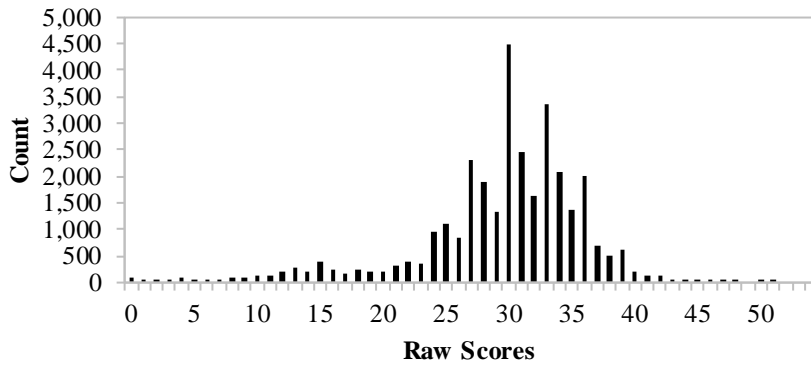


Table 2.3.3.5.2

Raw Score Descriptive Statistics: Writ 6-8 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	11,959	0	47	28.19	6.82
7	10,870	0	50	29.55	6.57
8	9,290	0	51	30.79	6.35
Total	32,119	0	51	29.40	6.69

Figure 2.3.3.5.2
Raw Scores: Writ 6-8B/C S502 Paper



2.3.3.6 Grades 9–12

Table 2.3.3.6.1

Raw Score Descriptive Statistics: Writ 9-12 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	2,805	0	22	9.29	4.69
10	2,546	0	21	9.44	4.59
11	1,975	0	22	10.25	4.39
12	1,246	0	21	10.74	4.22
Total	8,572	0	22	9.77	4.56

Figure 2.3.3.6.1
Raw Scores: Writ 9-12A S502 Paper

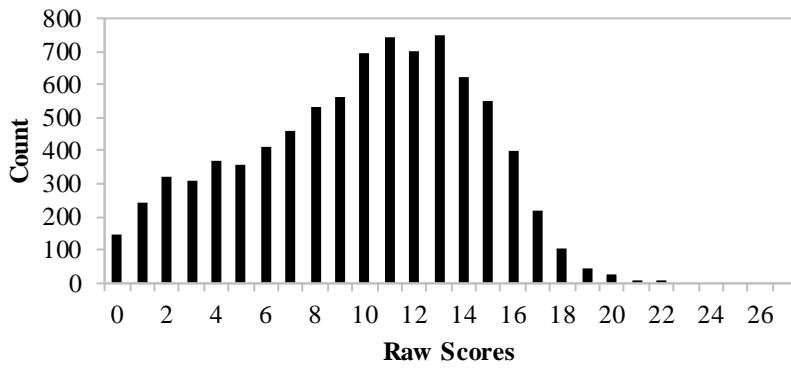
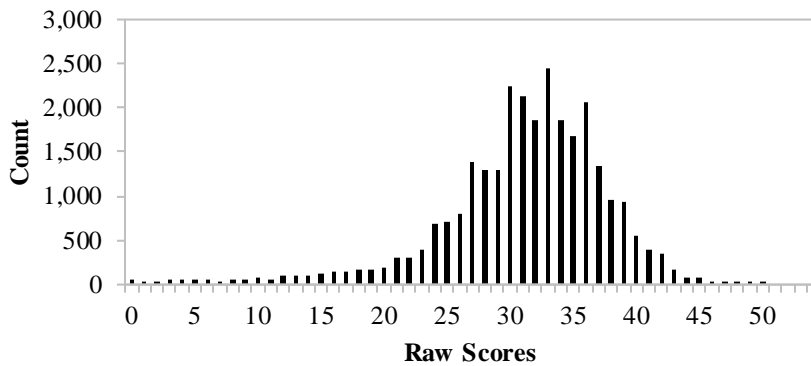


Table 2.3.3.6.2

Raw Score Descriptive Statistics: Writ 9-12 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	8,415	0	50	30.46	6.66
10	7,774	0	50	31.00	6.72
11	6,993	0	49	31.52	6.84
12	4,920	0	48	31.28	6.95
Total	28,102	0	50	31.02	6.78

Figure 2.3.3.6.2
Raw Scores: Writ 9-12B/C S502 Paper



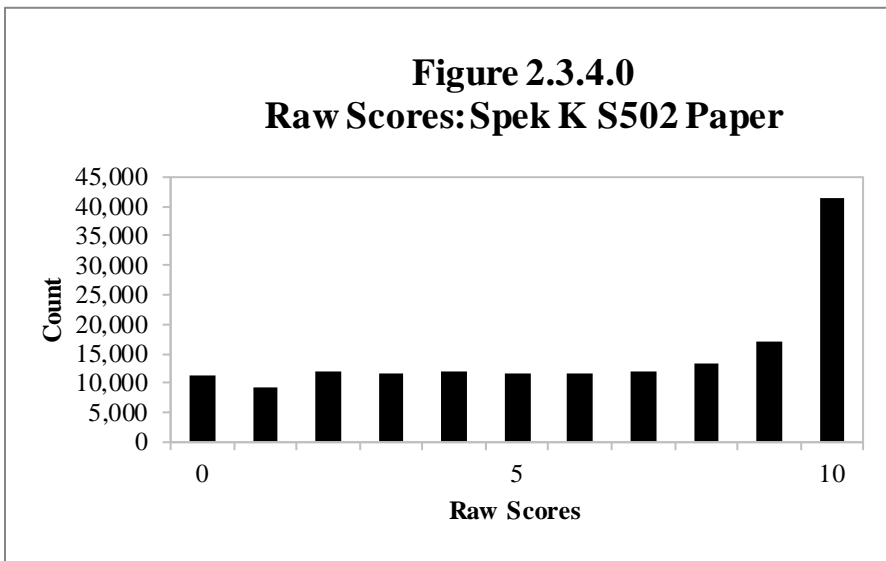
2.3.4 Speaking

2.3.4.0 Kindergarten

Table 2.3.4.0

Raw Score Descriptive Statistics: Spek K S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	163,192	0	10	6.15	3.38
Total	163,192	0	10	6.15	3.38



2.3.4.1 Grade 1

Table 2.3.4.1.1

Raw Score Descriptive Statistics: Spek 1 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,178	0	18	10.87	3.83
Total	15,178	0	18	10.87	3.83

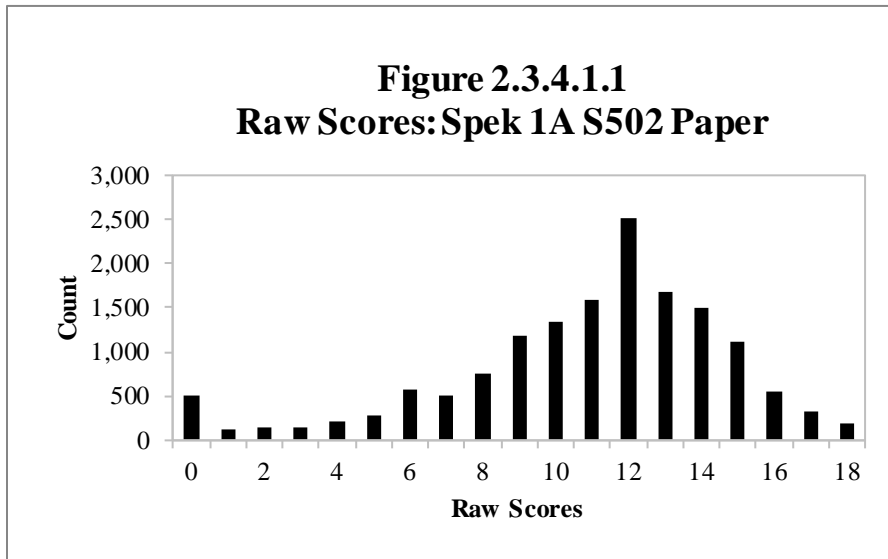
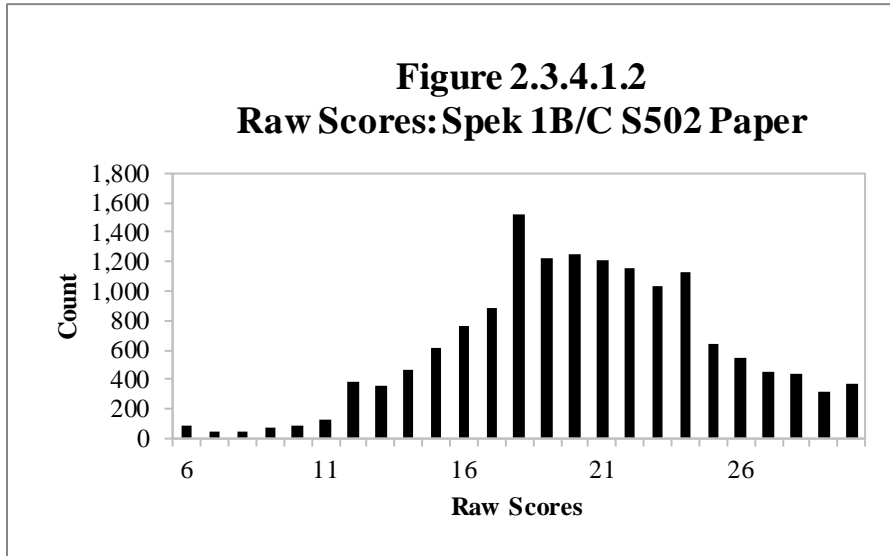


Table 2.3.4.1.2

Raw Score Descriptive Statistics: Spek 1 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,197	6	30	20.29	4.71
Total	15,197	6	30	20.29	4.71



2.3.4.2 Grade 2

Table 2.3.4.2.1

Raw Score Descriptive Statistics: Spek 2 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	7,170	0	18	11.17	4.09
Total	7,170	0	18	11.17	4.09

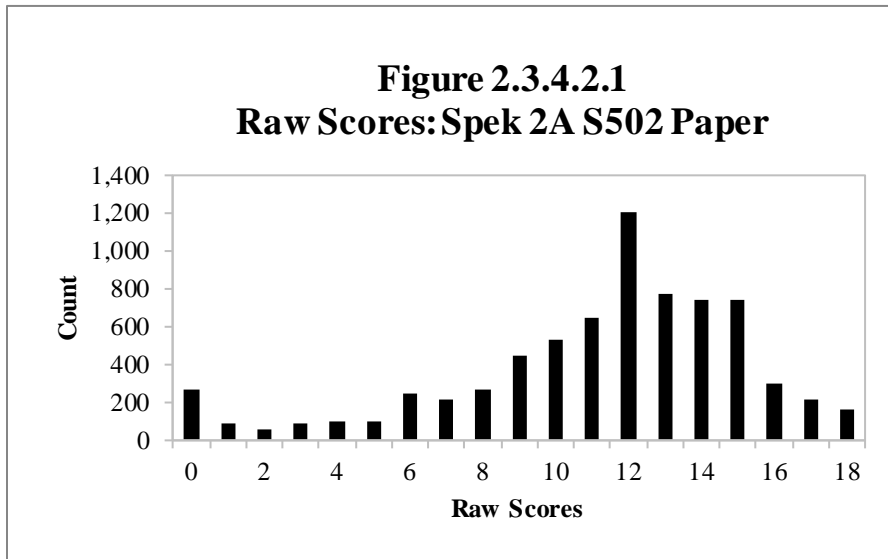
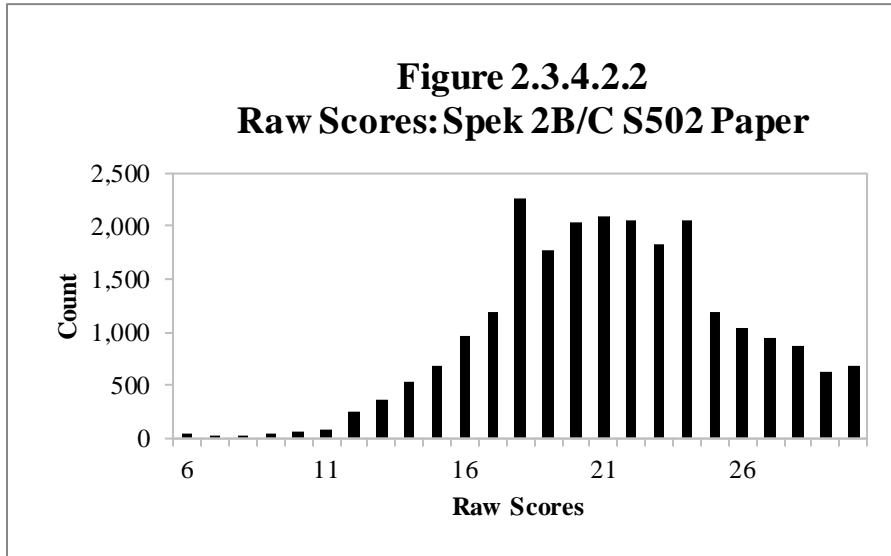


Table 2.3.4.2.2

Raw Score Descriptive Statistics: Spek 2 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	23,736	6	30	21.27	4.36
Total	23,736	6	30	21.27	4.36



2.3.4.3 Grade 3

Table 2.3.4.3.1

Raw Score Descriptive Statistics: Spek 3 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	4,937	0	18	10.73	4.34
Total	4,937	0	18	10.73	4.34

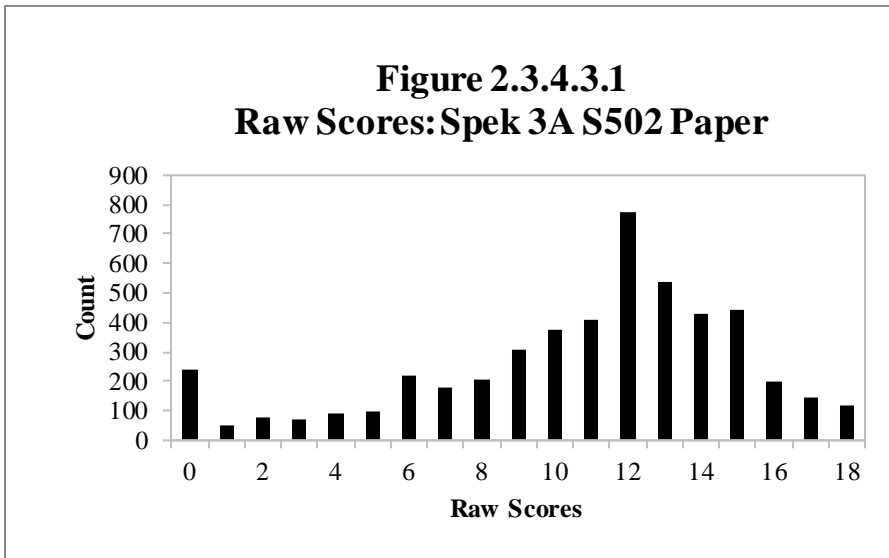
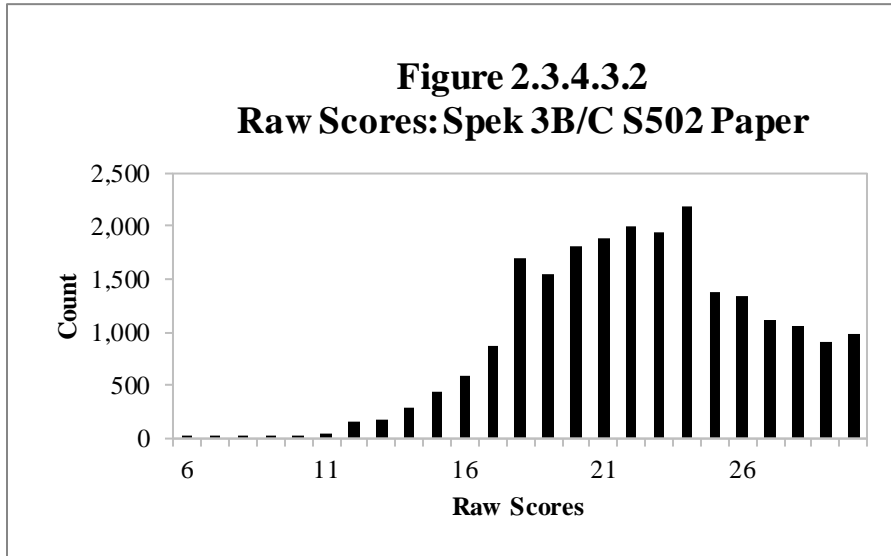


Table 2.3.4.3.2

Raw Score Descriptive Statistics: Spek 3 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	22,554	6	30	22.30	4.25
Total	22,554	6	30	22.30	4.25



2.3.4.4 Grades 4–5

Table 2.3.4.4.1

Raw Score Descriptive Statistics: Spek 4-5 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	3,538	0	18	9.33	4.30
5	3,111	0	18	9.29	4.25
Total	6,649	0	18	9.31	4.28

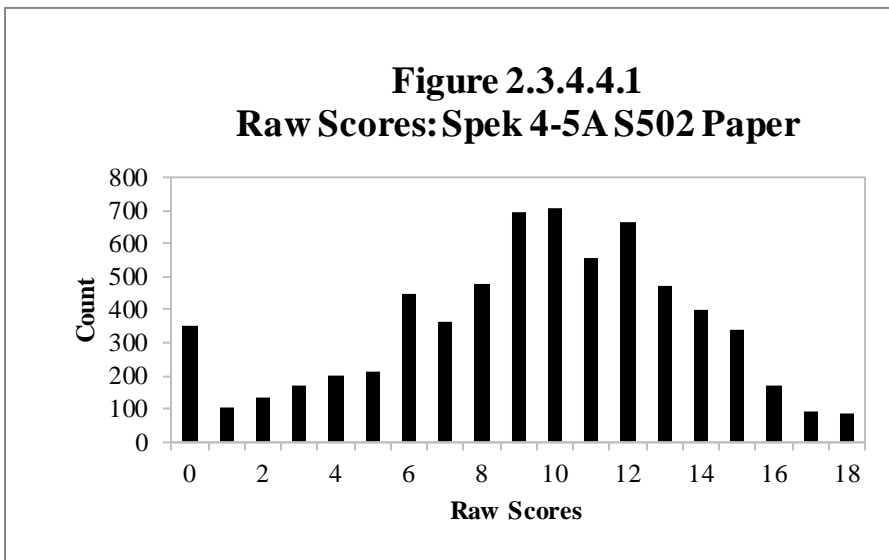
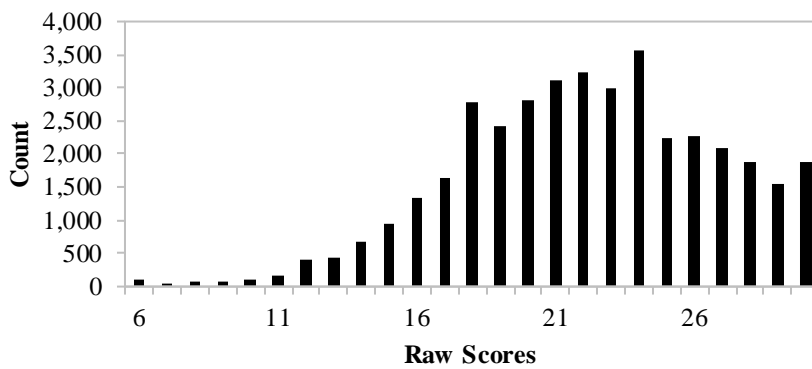


Table 2.3.4.4.2

Raw Score Descriptive Statistics: Spek 4-5 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	22,080	6	30	21.68	4.55
5	16,597	6	30	22.62	4.56
Total	38,677	6	30	22.09	4.58

Figure 2.3.4.4.2
Raw Scores: Spek 4-5B/C S502 Paper



2.3.4.5 Grades 6–8

Table 2.3.4.5.1

Raw Score Descriptive Statistics: Spek 6-8 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	3,051	0	18	9.24	4.03
7	2,906	0	18	9.26	4.00
8	2,952	0	18	9.27	3.95
Total	8,909	0	18	9.26	4.00

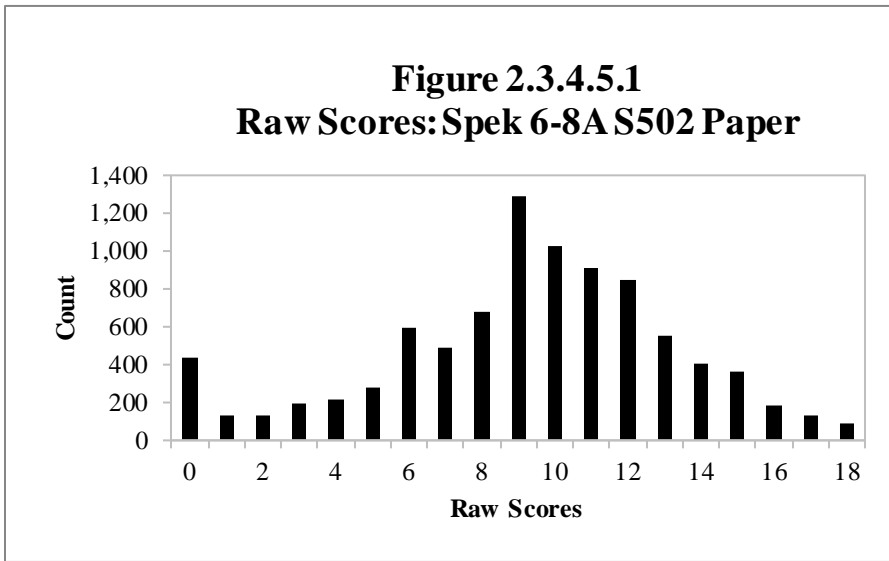
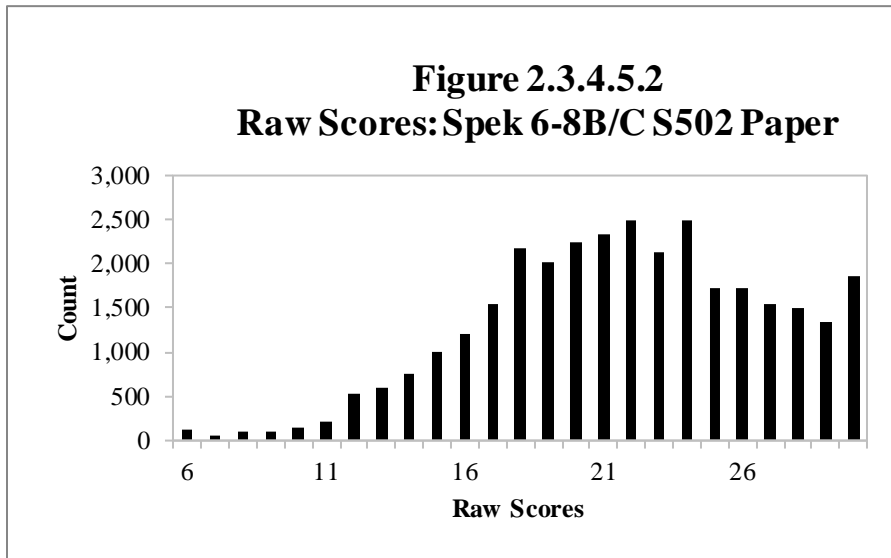


Table 2.3.4.5.2

Raw Score Descriptive Statistics: Spek 6-8 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	11,886	6	30	21.26	4.90
7	10,784	6	30	21.70	5.00
8	9,236	6	30	22.28	5.08
Total	31,906	6	30	21.70	5.00



2.3.4.6 Grades 9–12

Table 2.3.4.6.1

Raw Score Descriptive Statistics: Spek 9-12 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	2,756	0	18	9.36	4.29
10	2,522	0	18	9.31	4.20
11	1,940	0	18	9.95	4.13
12	1,224	0	18	10.38	3.78
Total	8,442	0	18	9.63	4.17

Figure 2.3.4.6.1
Raw Scores: Spek 9-12A S502 Paper

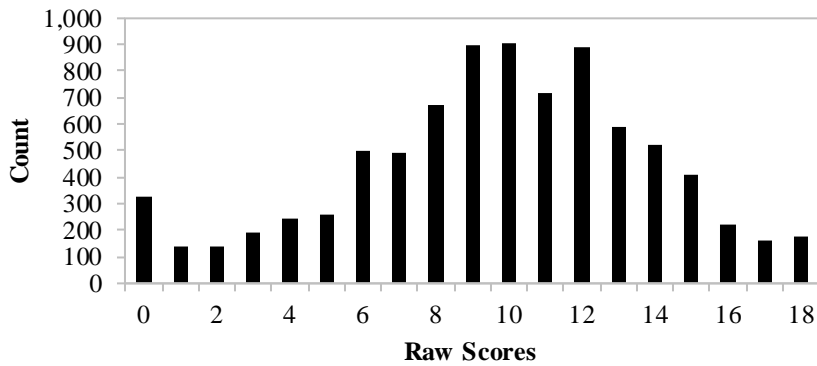
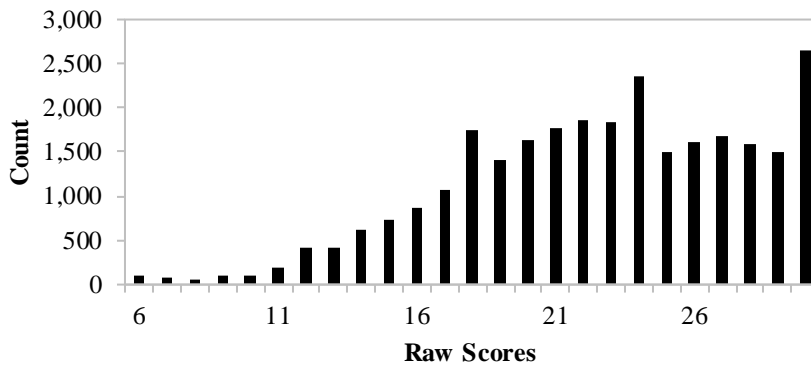


Table 2.3.4.6.2

Raw Score Descriptive Statistics: Spek 9-12 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	8,351	6	30	22.71	5.14
10	7,706	6	30	22.29	5.25
11	6,918	6	30	22.58	5.20
12	4,865	6	30	22.47	5.08
Total	27,840	6	30	22.52	5.18

Figure 2.3.4.6.2
Raw Scores: Spek 9-12B/C S502 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.

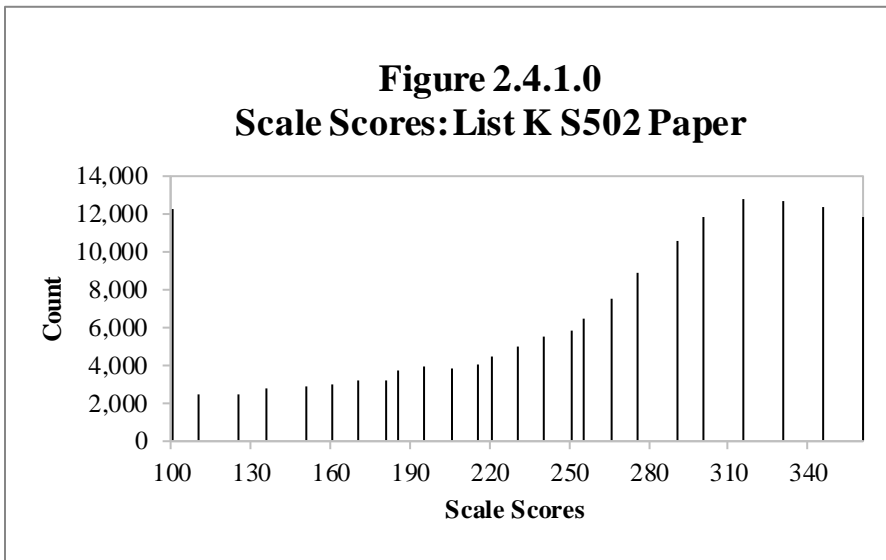
2.4.1 Listening

2.4.1.0 Kindergarten

Table 2.4.1.0

Scale Score Descriptive Statistics: List K S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	163,226	100	363	257.14	77.97
Total	163,226	100	363	257.14	77.97



2.4.1.1. Grade 1

Table 2.4.1.1.1

Scale Score Descriptive Statistics: List 1 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	12,281	121	352	290.47	38.89
Total	12,281	121	352	290.47	38.89

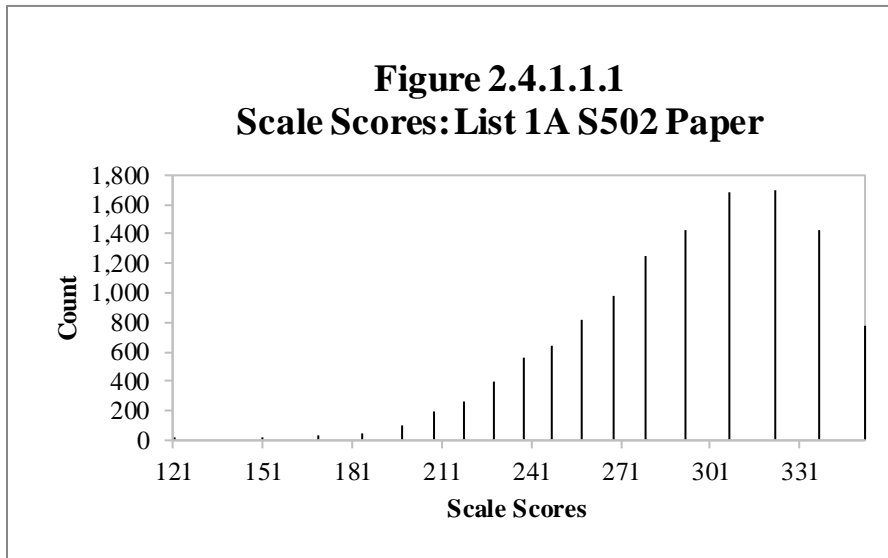


Table 2.4.1.1.2

Scale Score Descriptive Statistics: List 1 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	13,159	194	405	318.50	36.22
Total	13,159	194	405	318.50	36.22

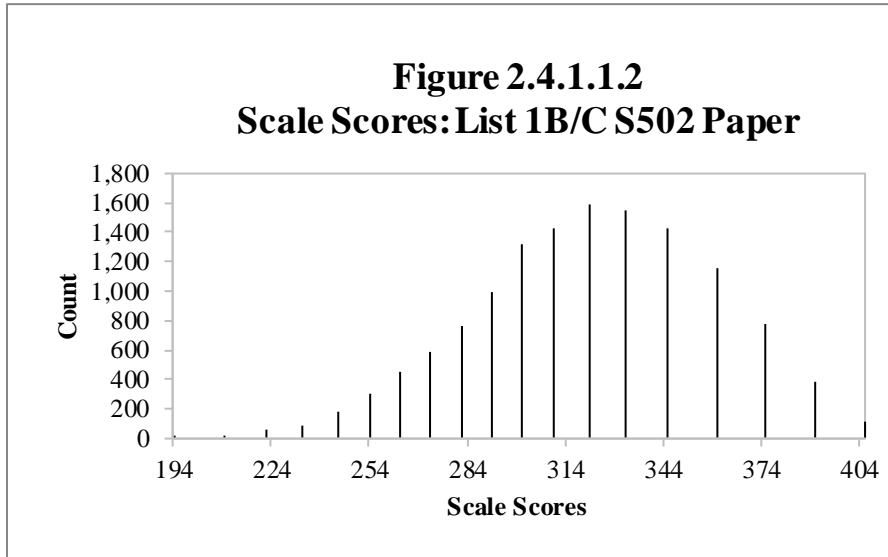
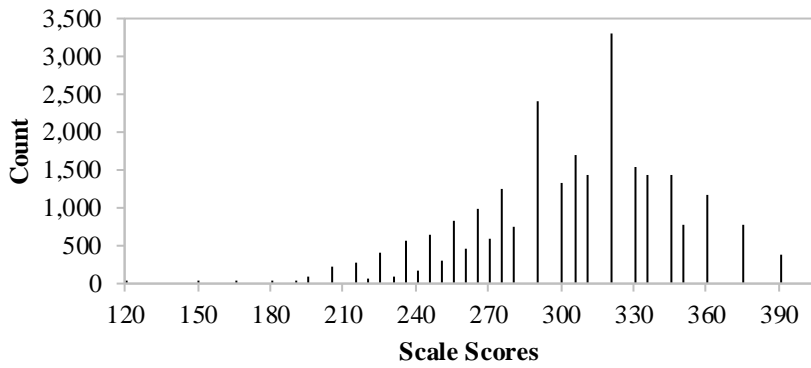


Table 2.4.1.1.3

Scale Score Descriptive Statistics: List 1 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	25,440	121	405	304.97	40.06
Total	25,440	121	405	304.97	40.06

Figure 2.4.1.1.3
Scale Scores: List 1 S502 Paper



2.4.1.2 Grade 2

Table 2.4.1.2.1

Scale Score Descriptive Statistics: List 2 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	6,436	150	352	301.98	40.36
Total	6,436	150	352	301.98	40.36

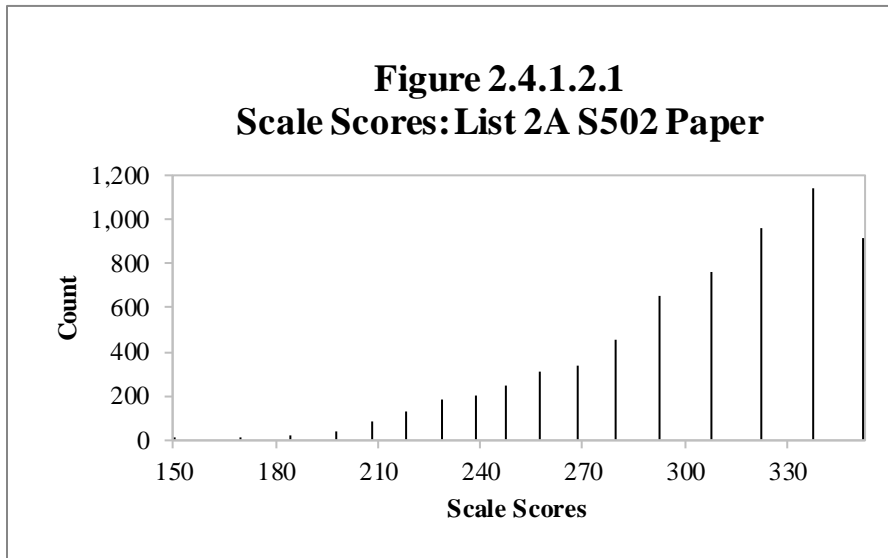


Table 2.4.1.2.2

Scale Score Descriptive Statistics: List 2 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	22,176	175	405	340.97	34.32
Total	22,176	175	405	340.97	34.32

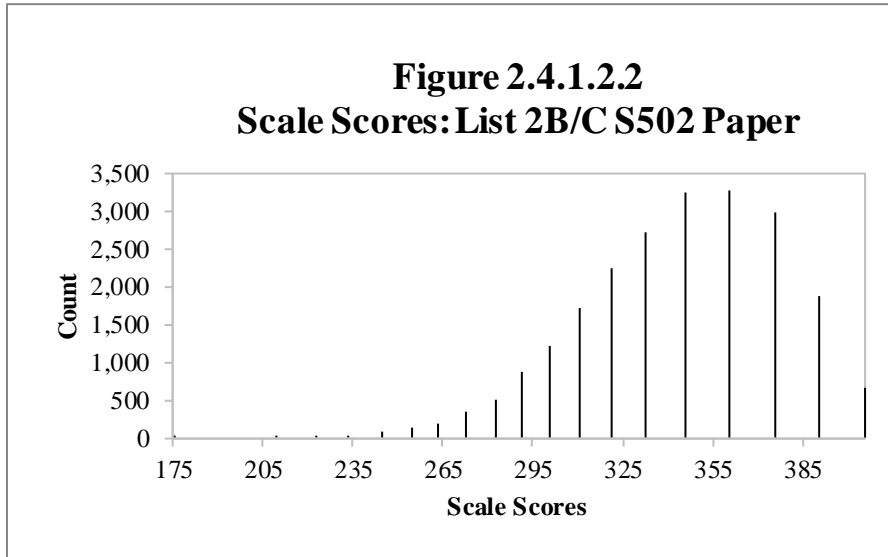
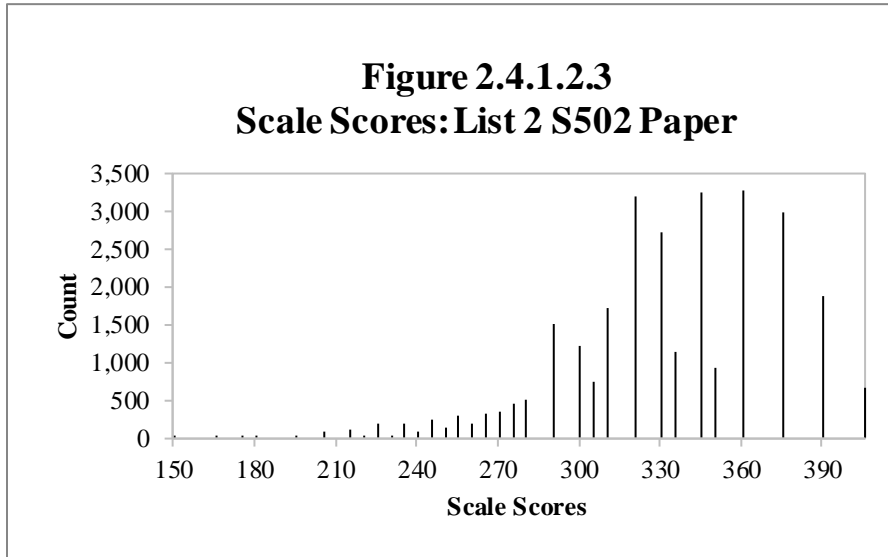


Table 2.4.1.2.3

Scale Score Descriptive Statistics: List 2 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	28,612	150	405	332.20	39.30
Total	28,612	150	405	332.20	39.30



2.4.1.3 Grade 3

Table 2.4.1.3.1

Scale Score Descriptive Statistics: List 3 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	4,498	184	416	320.63	38.47
Total	4,498	184	416	320.63	38.47

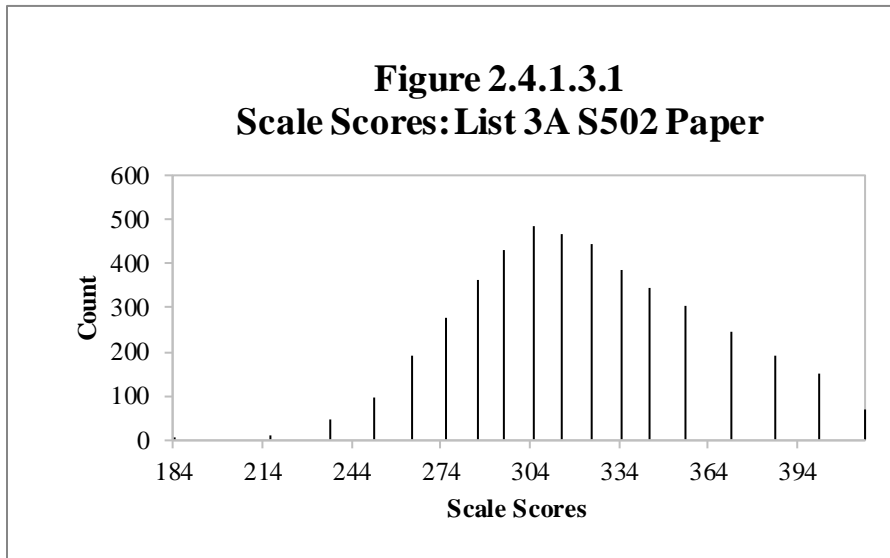


Table 2.4.1.3.2

Scale Score Descriptive Statistics: List 3 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	21,373	201	471	361.96	29.58
Total	21,373	201	471	361.96	29.58

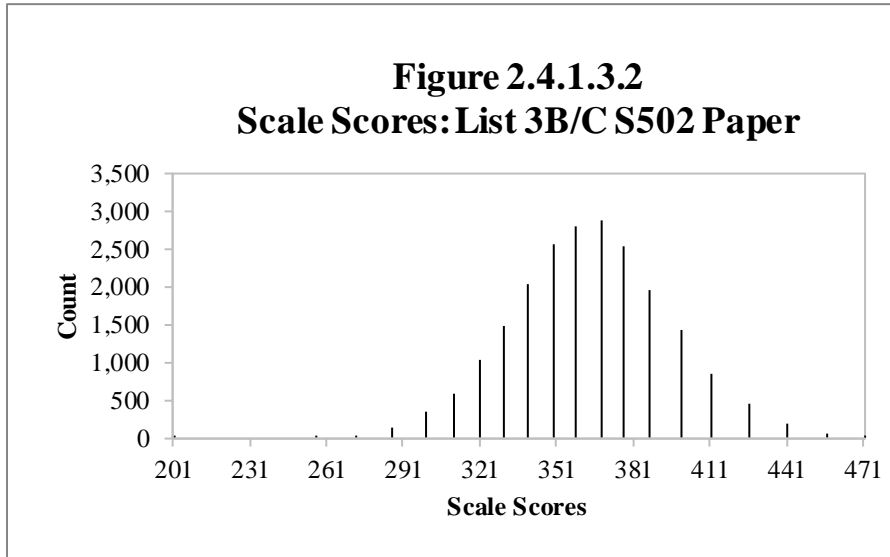
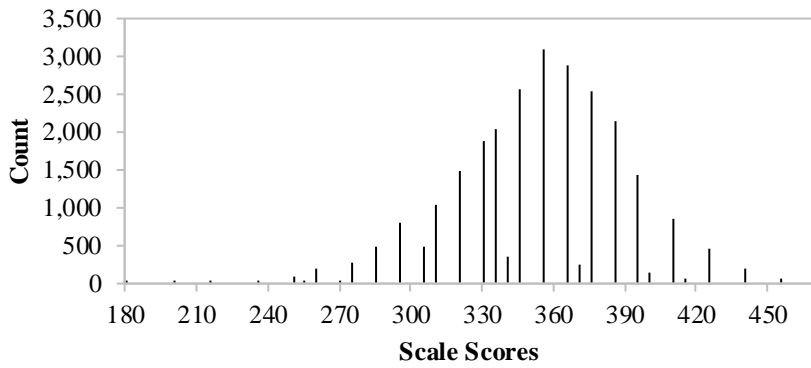


Table 2.4.1.3.3

Scale Score Descriptive Statistics: List 3 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	25,871	184	471	354.77	35.01
Total	25,871	184	471	354.77	35.01

Figure 2.4.1.3.3
Scale Scores: List 3 S502 Paper



2.4.1.4. Grades 4–5

Table 2.4.1.4.1

Scale Score Descriptive Statistics: List 4-5 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	3,292	184	416	328.56	40.35
5	2,949	184	416	331.27	40.54
Total	6,241	184	416	329.84	40.46

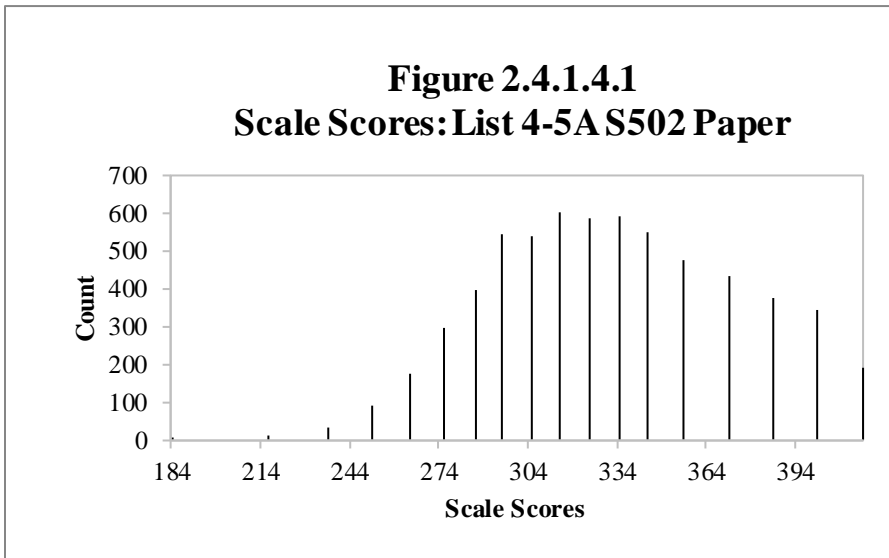


Table 2.4.1.4.2

Scale Score Descriptive Statistics: List 4-5 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	21,288	234	471	378.81	31.89
5	16,086	256	471	391.66	32.75
Total	37,374	234	471	384.34	32.88

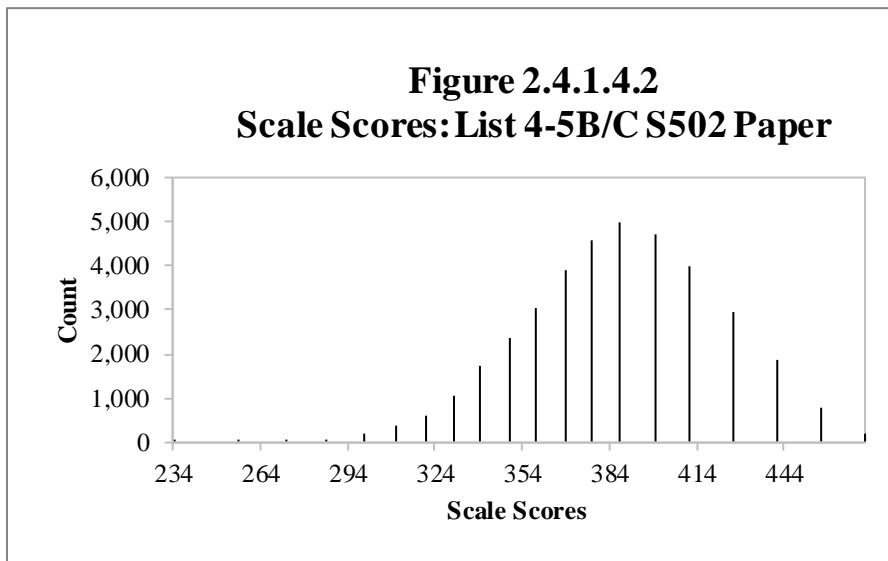
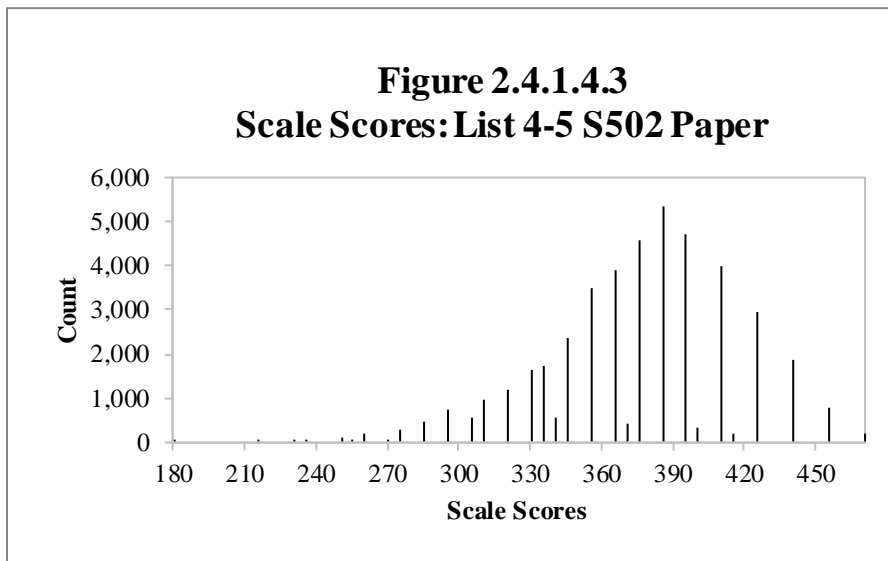


Table 2.4.1.4.3

Scale Score Descriptive Statistics: List 4-5 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	24,580	184	471	372.08	37.30
5	19,035	184	471	382.30	40.48
Total	43,615	184	471	376.54	39.05

Figure 2.4.1.4.3
Scale Scores: List 4-5 S502 Paper



2.4.1.5. Grades 6–8

Table 2.4.1.5.1

Scale Score Descriptive Statistics: List 6-8 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,912	212	424	322.29	39.54
7	2,709	181	424	322.73	41.03
8	2,793	212	424	323.09	40.87
Total	8,414	181	424	322.70	40.46

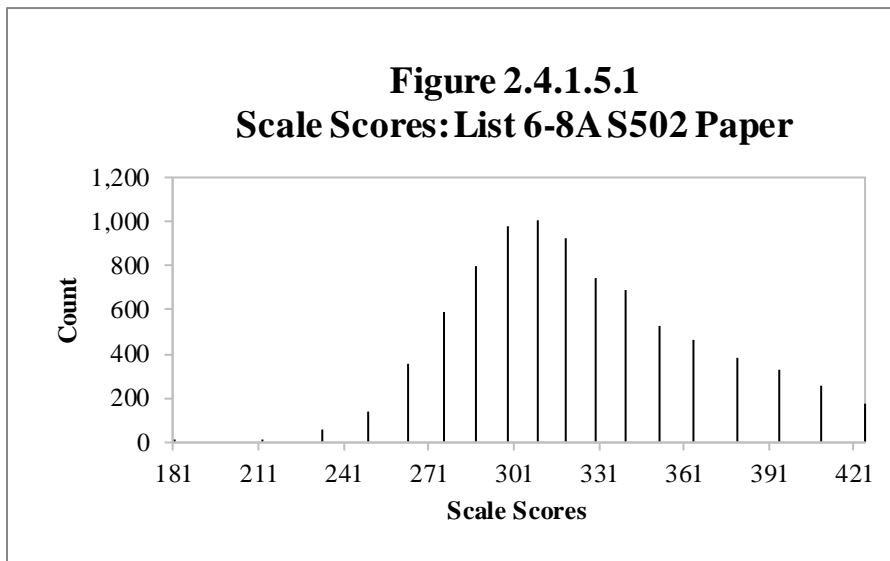


Table 2.4.1.5.2

Scale Score Descriptive Statistics: List 6-8 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	11,431	255	476	390.41	31.25
7	10,413	274	476	397.93	32.82
8	8,899	255	476	404.36	33.98
Total	30,743	255	476	396.99	33.08

Figure 2.4.1.5.2
Scale Scores: List 6-8B/C S502 Paper

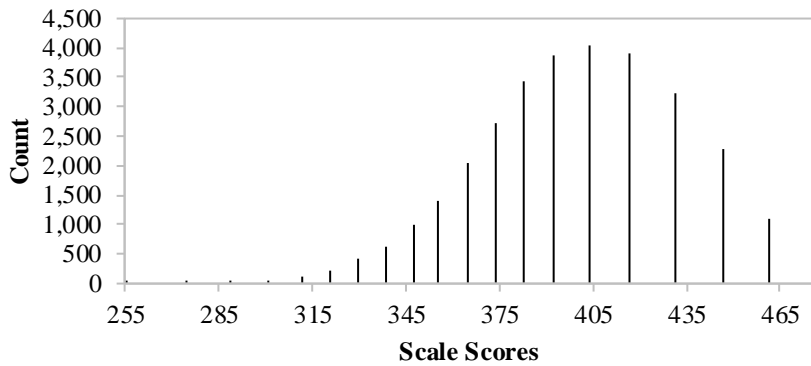
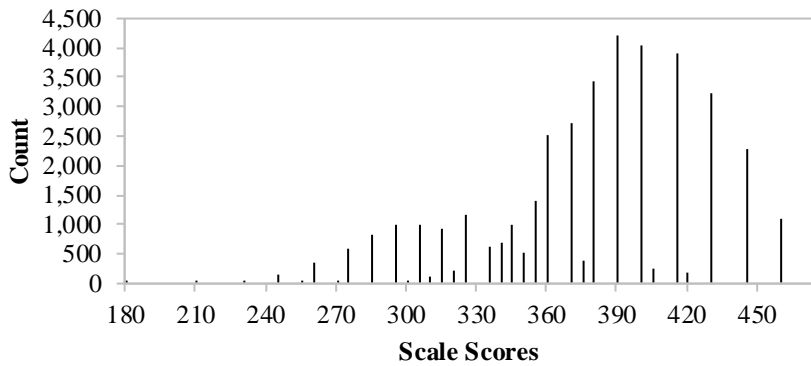


Table 2.4.1.5.3

Scale Score Descriptive Statistics: List 6-8 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	14,343	212	476	376.58	42.97
7	13,122	181	476	382.40	46.13
8	11,692	212	476	384.95	49.78
Total	39,157	181	476	381.03	46.28

Figure 2.4.1.5.3
Scale Scores: List 6-8 S502 Paper



2.4.1.6. Grades 9–12

Table 2.4.1.6.1

Scale Score Descriptive Statistics: List 9-12 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	2,576	148	428	322.01	37.14
10	2,353	201	428	321.94	36.44
11	1,829	223	428	326.79	36.20
12	1,151	240	428	328.15	34.74
Total	7,909	148	428	323.99	36.46

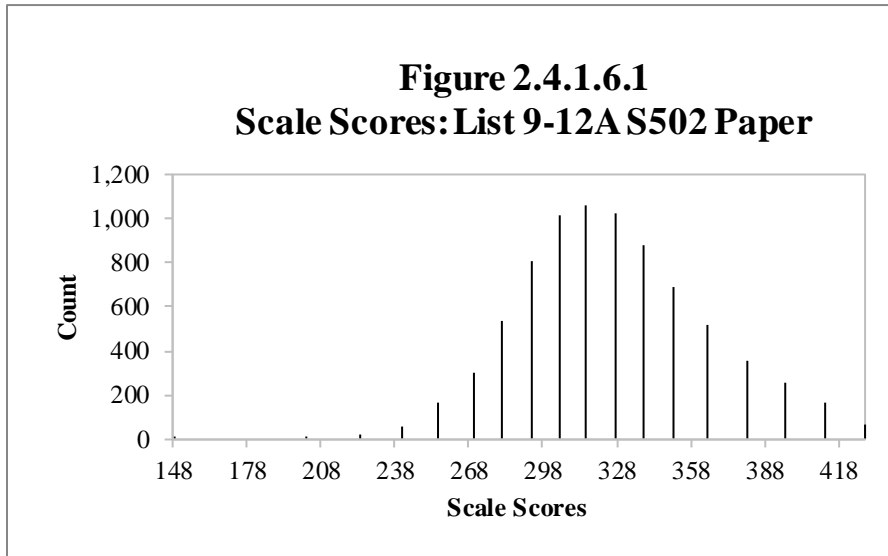


Table 2.4.1.6.2

Scale Score Descriptive Statistics: List 9-12 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	8,040	263	493	399.62	33.69
10	7,483	263	493	399.91	35.32
11	6,711	263	493	402.35	36.28
12	4,728	263	493	397.95	35.11
Total	26,962	263	493	400.09	35.08

Figure 2.4.1.6.2
Scale Scores: List 9-12B/C S502 Paper

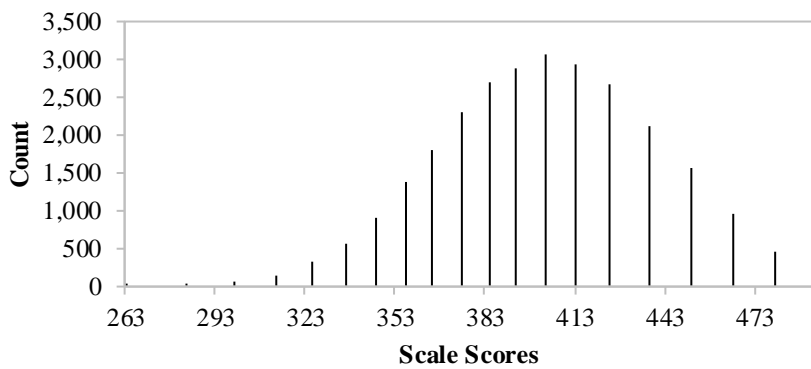
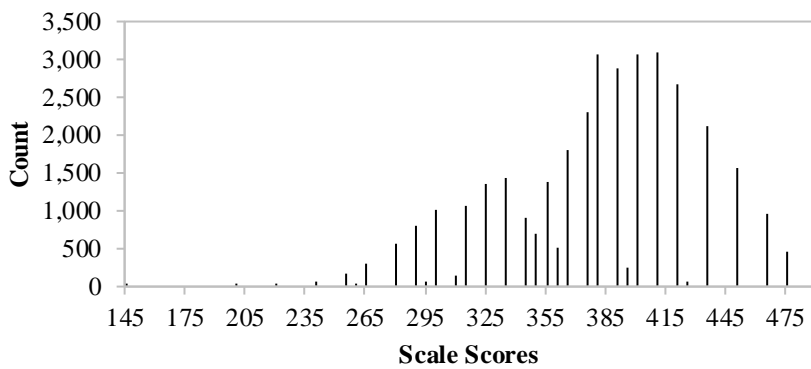


Table 2.4.1.6.3

Scale Score Descriptive Statistics: List 9-12 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	10,616	148	493	380.78	47.97
10	9,836	201	493	381.26	48.71
11	8,540	223	493	386.17	47.70
12	5,879	240	493	384.28	44.66
Total	34,871	148	493	382.83	47.63

Figure 2.4.1.6.3
Scale Scores: List 9-12 S502 Paper



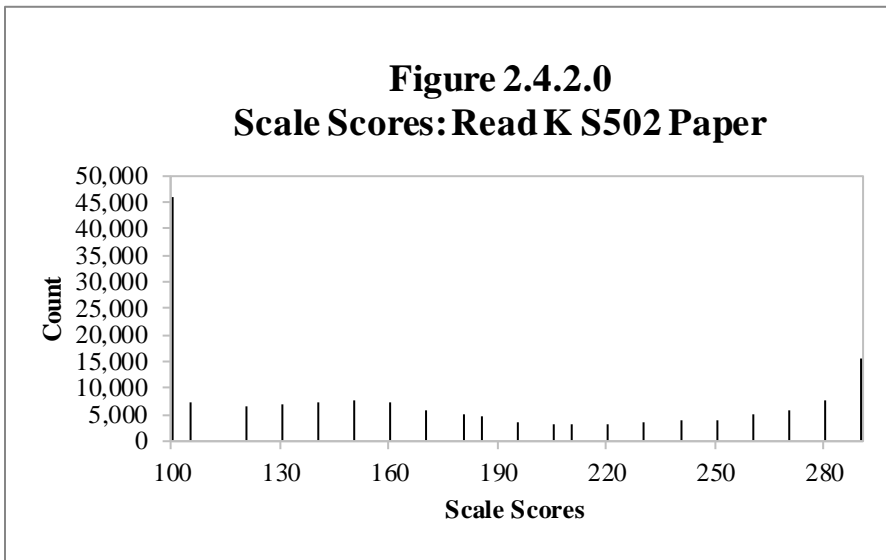
2.4.2 Reading

2.4.2.0 Kindergarten

Table 2.4.2.0

Scale Score Descriptive Statistics: Read K S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	163,218	100	290	173.15	69.14
Total	163,218	100	290	173.15	69.14



2.4.2.1 Grade 1

Table 2.4.2.1.1

Scale Score Descriptive Statistics: Read 1 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	11,868	141	361	270.21	22.77
Total	11,868	141	361	270.21	22.77

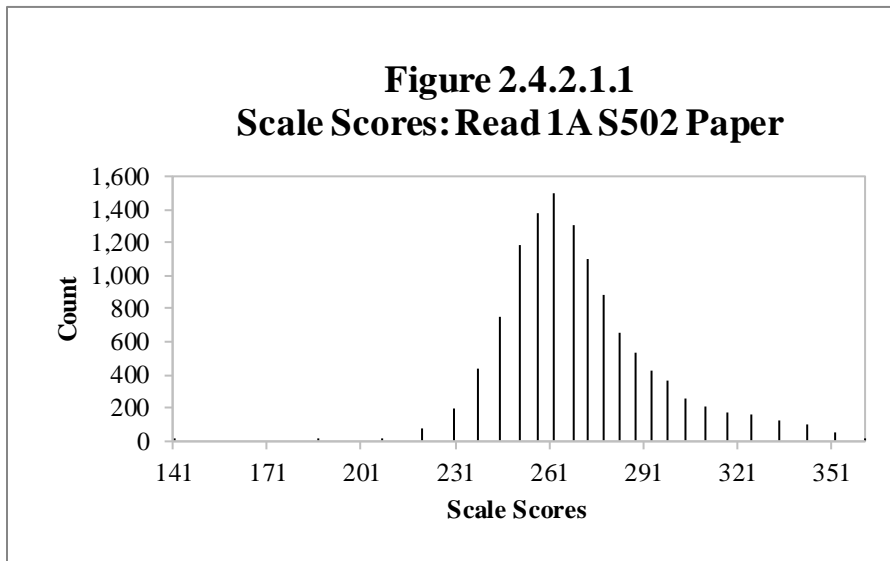


Table 2.4.2.1.2

Scale Score Descriptive Statistics: Read 1 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	11,625	216	394	298.87	22.01
Total	11,625	216	394	298.87	22.01

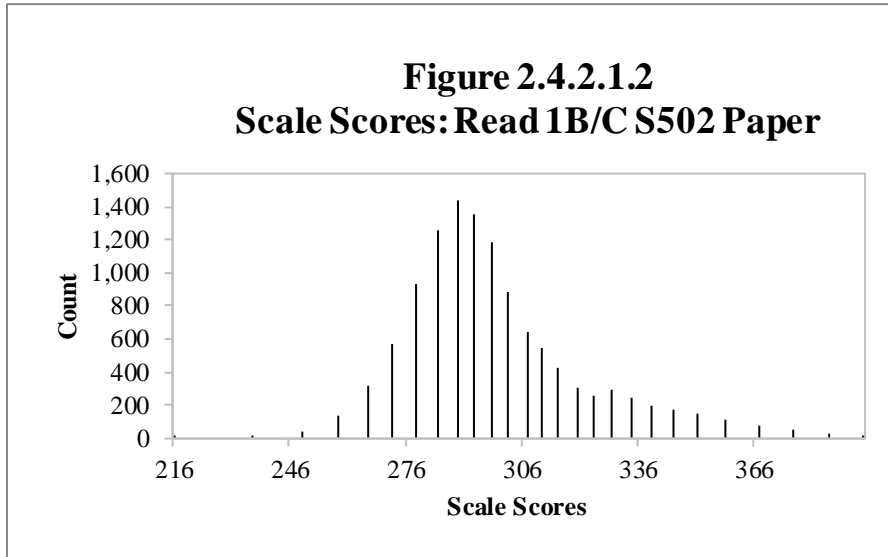
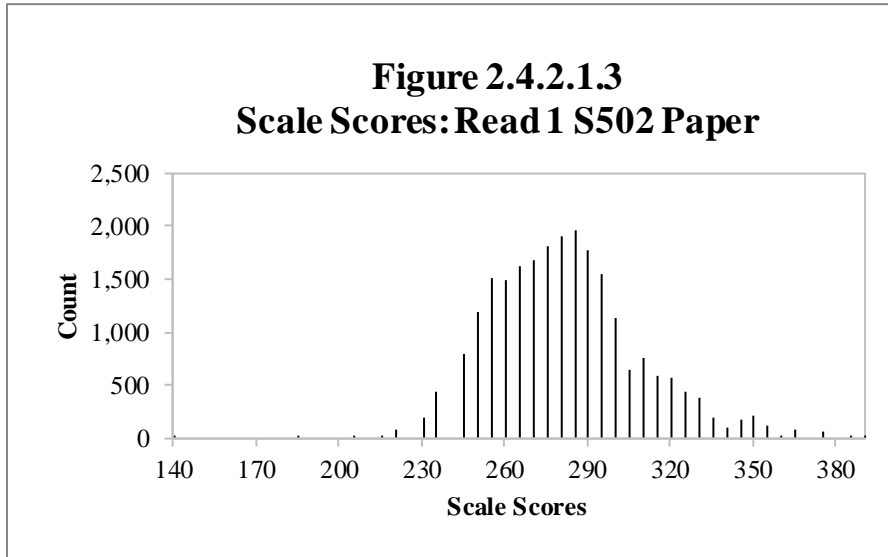


Table 2.4.2.1.3

Scale Score Descriptive Statistics: Read 1 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	23,493	141	394	284.39	26.59
Total	23,493	141	394	284.39	26.59



2.4.2.2 Grade 2

Table 2.4.2.2.1

Scale Score Descriptive Statistics: Read 2 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	6,114	187	361	283.08	29.22
Total	6,114	187	361	283.08	29.22

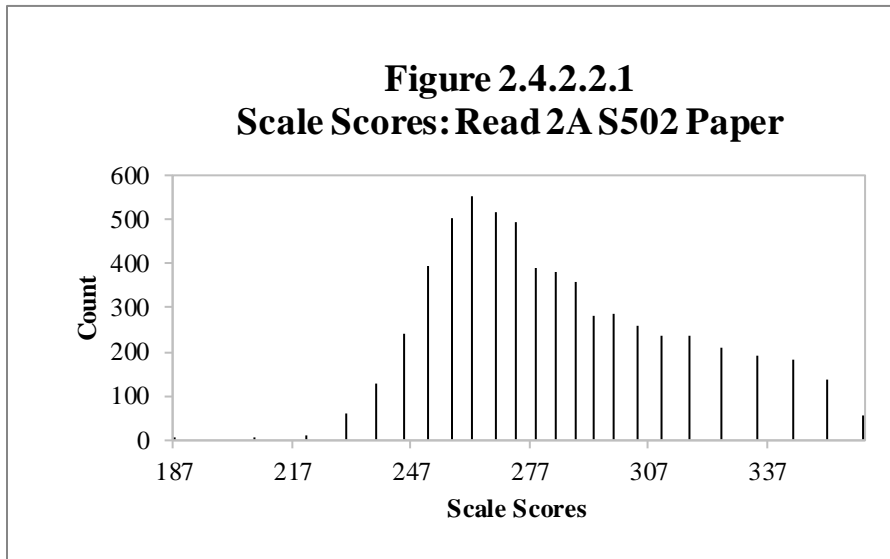


Table 2.4.2.2.2

Scale Score Descriptive Statistics: Read 2 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	19,788	236	394	316.69	28.40
Total	19,788	236	394	316.69	28.40

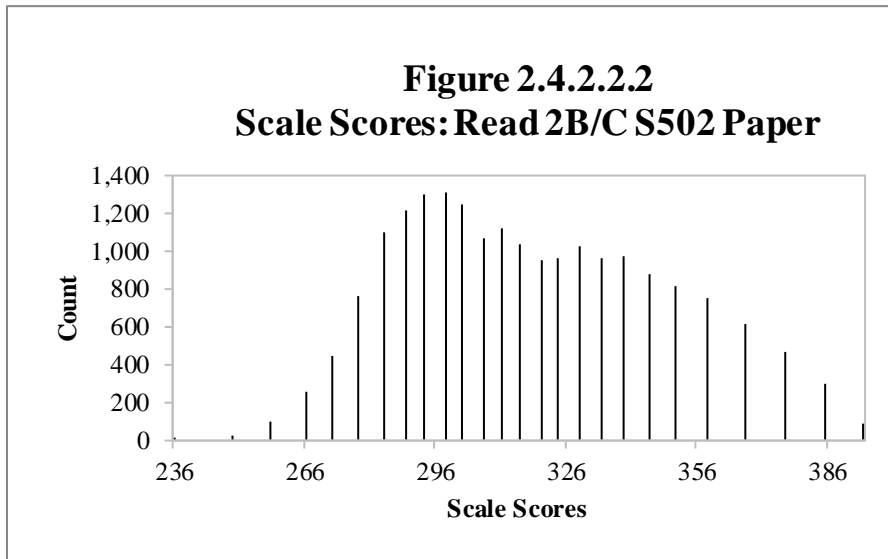
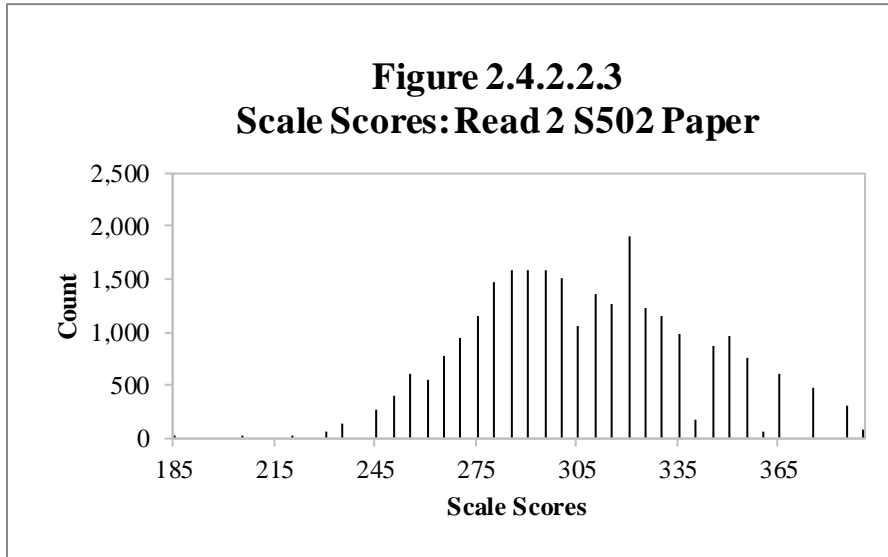


Table 2.4.2.2.3

Scale Score Descriptive Statistics: Read 2 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	25,902	187	394	308.76	31.96
Total	25,902	187	394	308.76	31.96



2.4.2.3 Grade 3

Table 2.4.2.3.1

Scale Score Descriptive Statistics: Read 3 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	4,249	233	385	298.47	26.13
Total	4,249	233	385	298.47	26.13

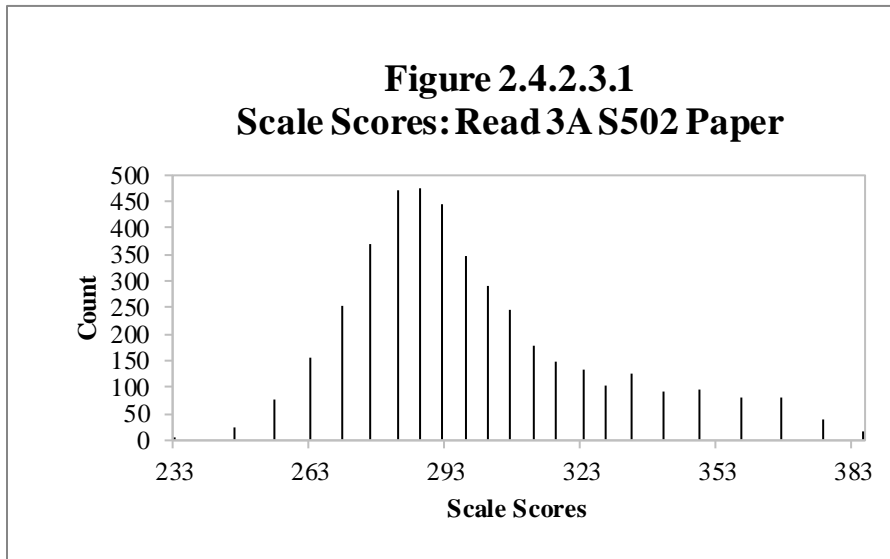


Table 2.4.2.3.2

Scale Score Descriptive Statistics: Read 3 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	18,557	271	445	340.94	18.99
Total	18,557	271	445	340.94	18.99

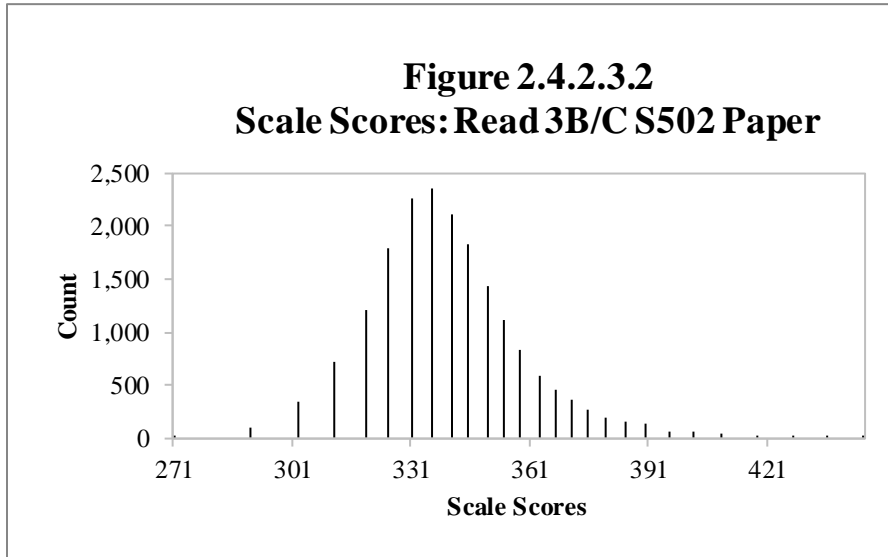
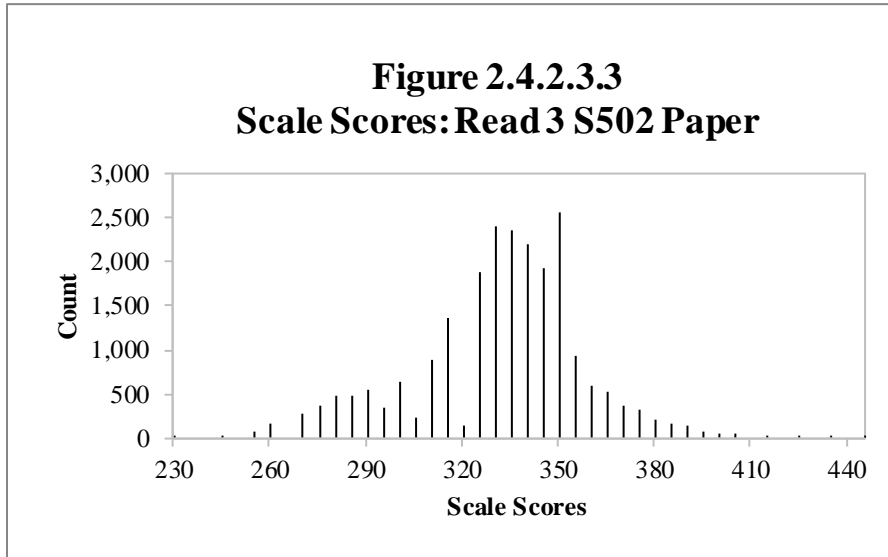


Table 2.4.2.3.3

Scale Score Descriptive Statistics: Read 3 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	22,806	233	445	333.03	26.35
Total	22,806	233	445	333.03	26.35



2.4.2.4 Grades 4–5

Table 2.4.2.4.1

Scale Score Descriptive Statistics: Read 4-5 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	3,117	233	385	307.68	29.74
5	2,804	233	385	312.40	30.08
Total	5,921	233	385	309.91	29.99

Figure 2.4.2.4.1
Scale Scores: Read 4-5A S502 Paper

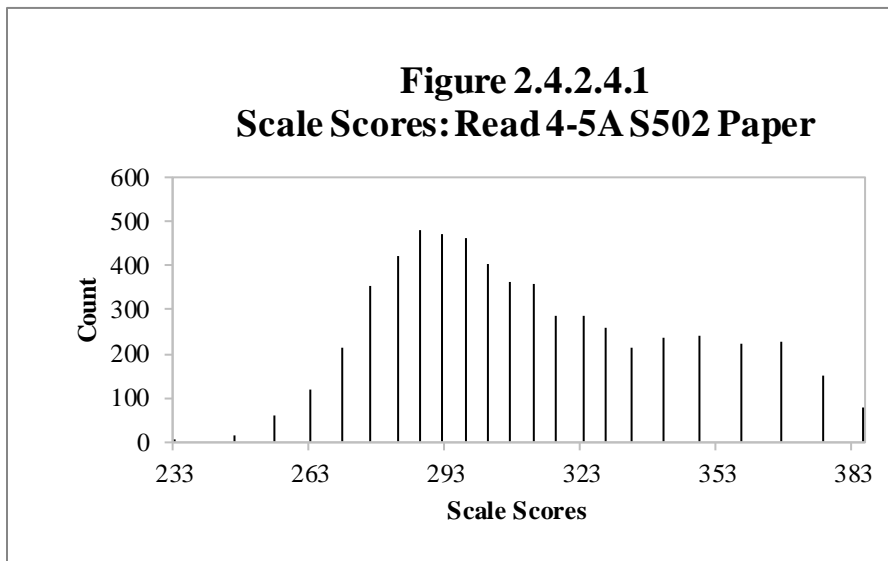


Table 2.4.2.4.2

Scale Score Descriptive Statistics: Read 4-5 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	18,996	175	445	353.10	22.85
5	14,731	271	445	364.58	25.36
Total	33,727	175	445	358.11	24.65

Figure 2.4.2.4.2
Scale Scores: Read 4-5B/C S502 Paper

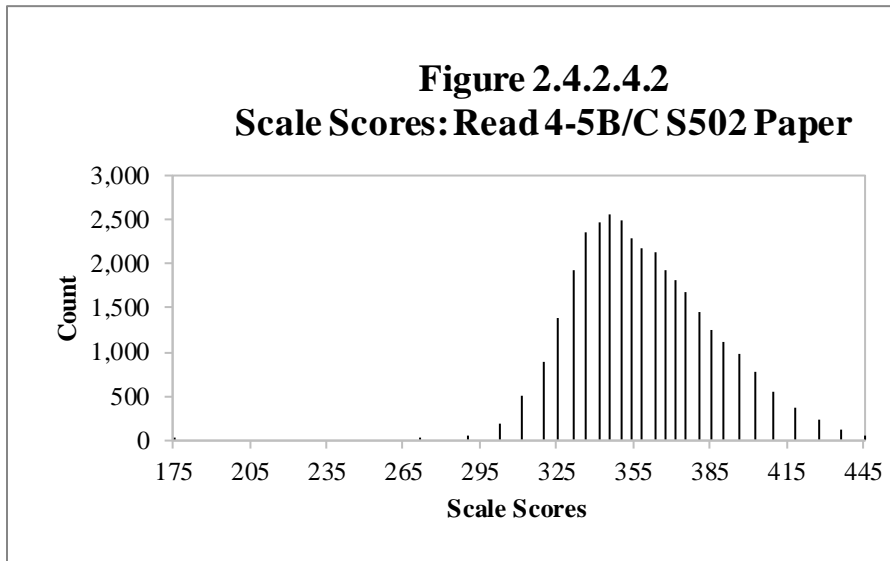
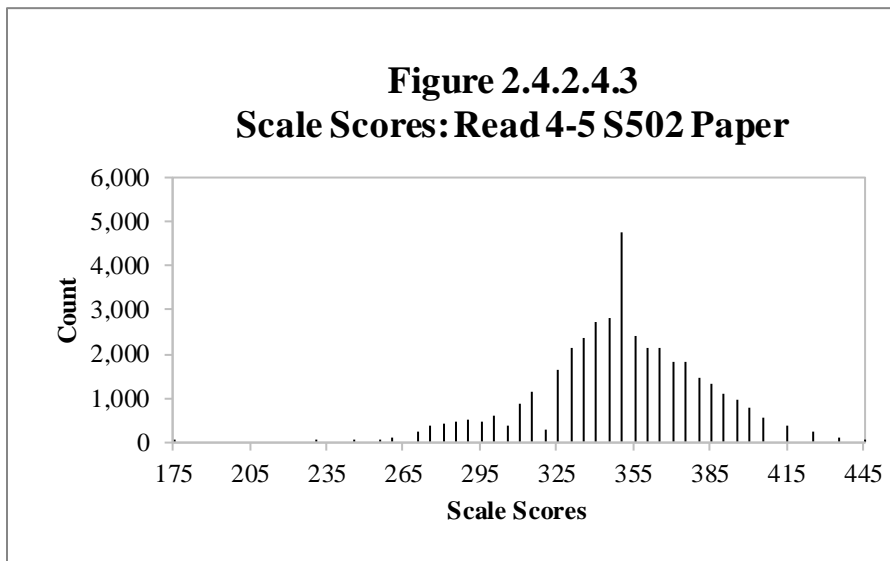


Table 2.4.2.4.3

Scale Score Descriptive Statistics: Read 4-5 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	22,113	175	445	346.70	28.69
5	17,535	233	445	356.24	32.42
Total	39,648	175	445	350.91	30.76

Figure 2.4.2.4.3
Scale Scores: Read 4-5 S502 Paper



2.4.2.5 Grades 6–8

Table 2.4.2.5.1

Scale Score Descriptive Statistics: Read 6-8 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,740	245	422	327.92	25.08
7	2,617	265	422	330.39	27.18
8	2,737	245	422	332.17	27.59
Total	8,094	245	422	330.16	26.69

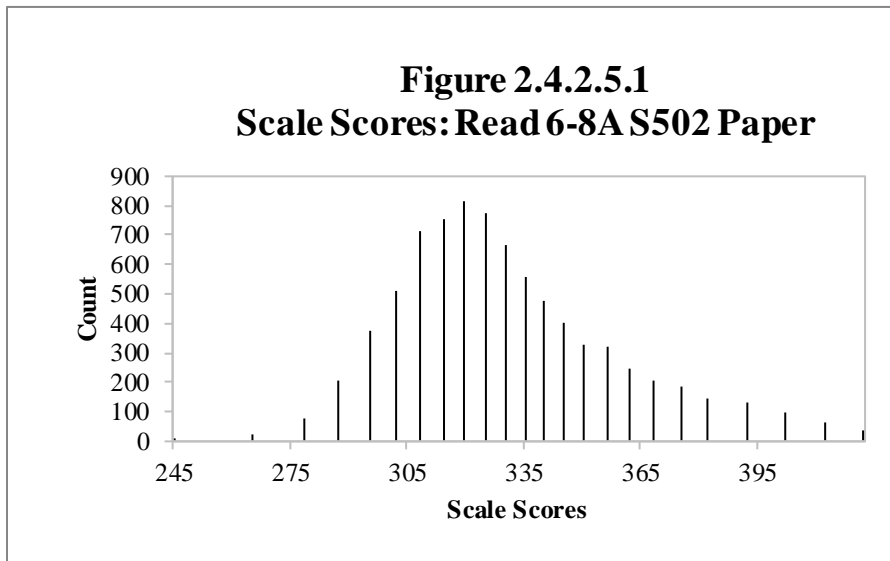


Table 2.4.2.5.2

Scale Score Descriptive Statistics: Read 6-8 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	10,526	277	452	359.39	22.13
7	9,727	277	452	367.11	24.85
8	8,515	296	452	375.41	26.78
Total	28,768	277	452	366.74	25.34

Figure 2.4.2.5.2
Scale Scores: Read 6-8B/C S502 Paper

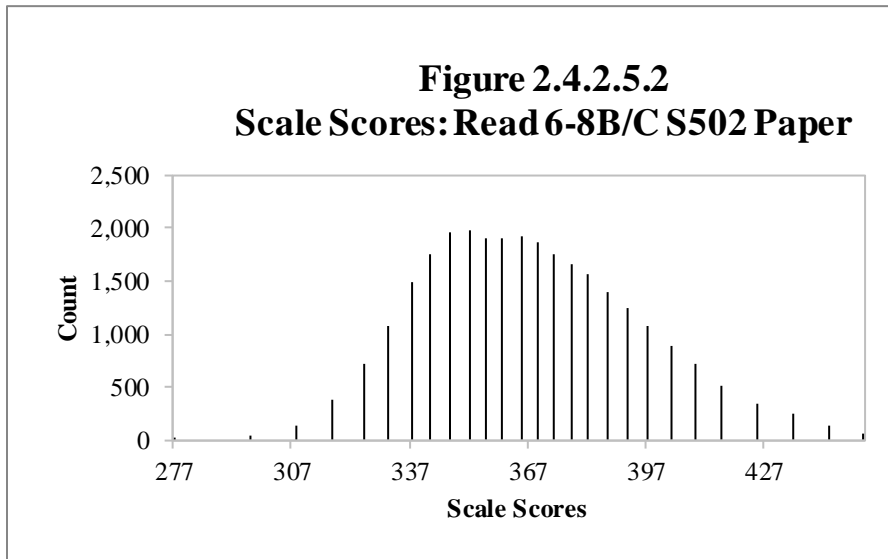
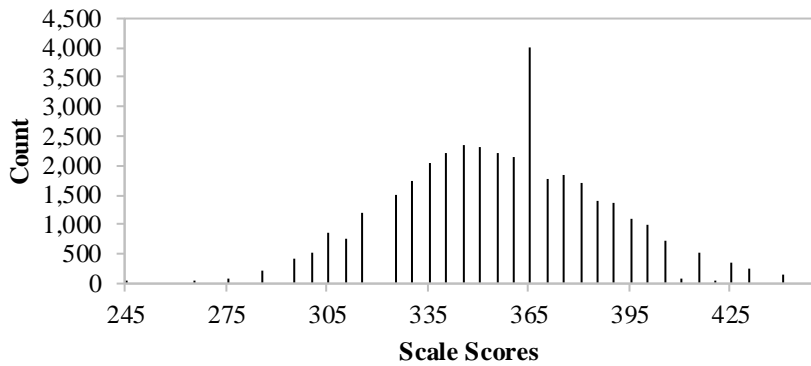


Table 2.4.2.5.3

Scale Score Descriptive Statistics: Read 6-8 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	13,266	245	452	352.89	26.09
7	12,344	265	452	359.33	29.47
8	11,252	245	452	364.89	32.74
Total	36,862	245	452	358.71	29.78

Figure 2.4.2.5.3
Scale Scores: Read 6-8 S502 Paper



2.4.2.6 Grades 9–12

Table 2.4.2.6.1

Scale Score Descriptive Statistics: Read 9-12 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	2,558	272	433	351.69	27.61
10	2,348	272	433	352.84	27.88
11	1,844	285	433	358.30	28.56
12	1,149	252	433	360.47	27.55
Total	7,899	252	433	354.85	28.11

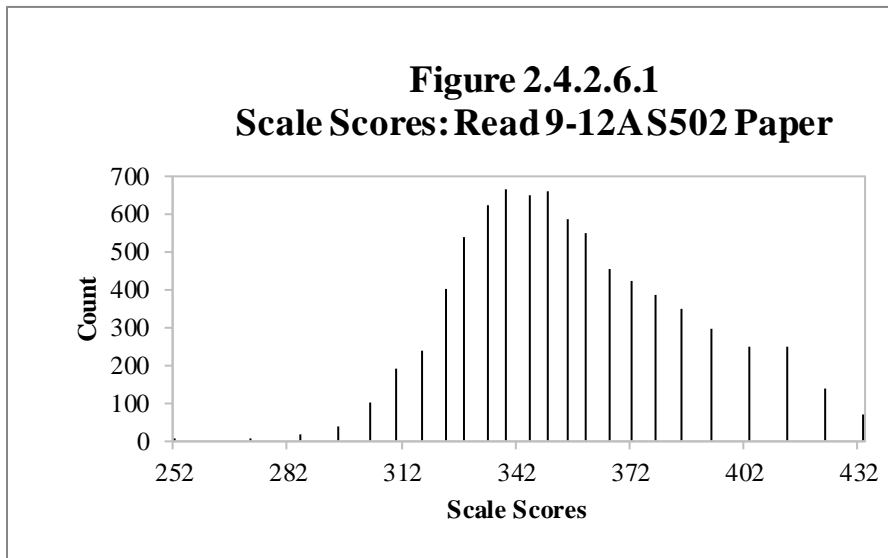


Table 2.4.2.6.2

Scale Score Descriptive Statistics: Read 9-12 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	7,280	316	470	389.71	25.07
10	6,907	233	470	392.92	26.91
11	6,269	297	470	396.38	27.89
12	4,453	316	470	392.62	26.88
Total	24,909	233	470	392.80	26.75

Figure 2.4.2.6.2
Scale Scores: Read 9-12B/C S502 Paper

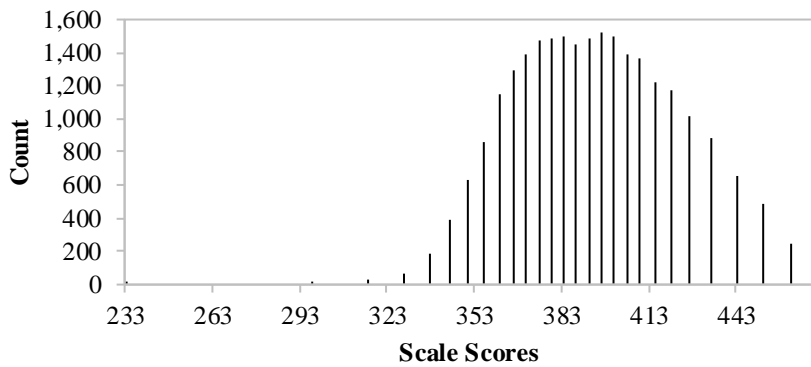
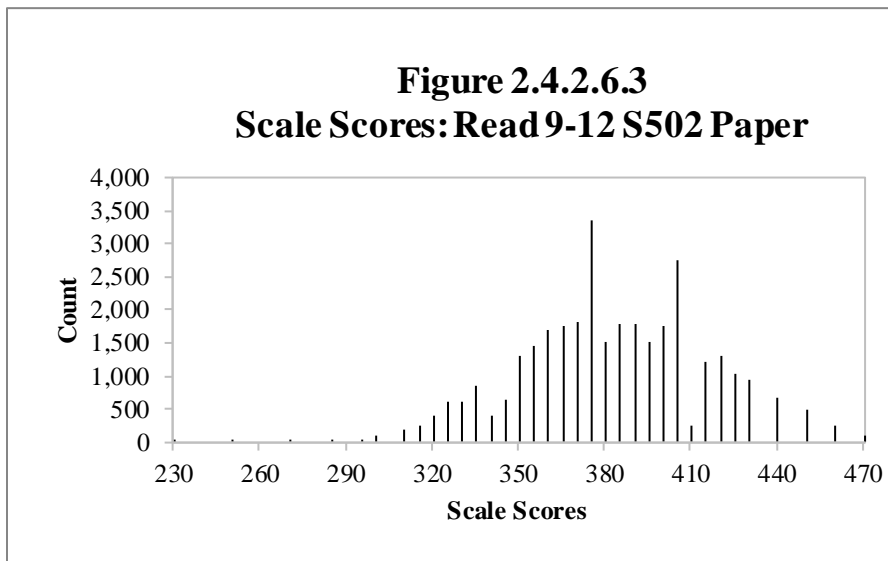


Table 2.4.2.6.3

Scale Score Descriptive Statistics: Read 9-12 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	9,838	272	470	379.83	30.68
10	9,255	233	470	382.75	32.28
11	8,113	285	470	387.72	32.26
12	5,602	252	470	386.03	29.97
Total	32,808	233	470	383.66	31.57



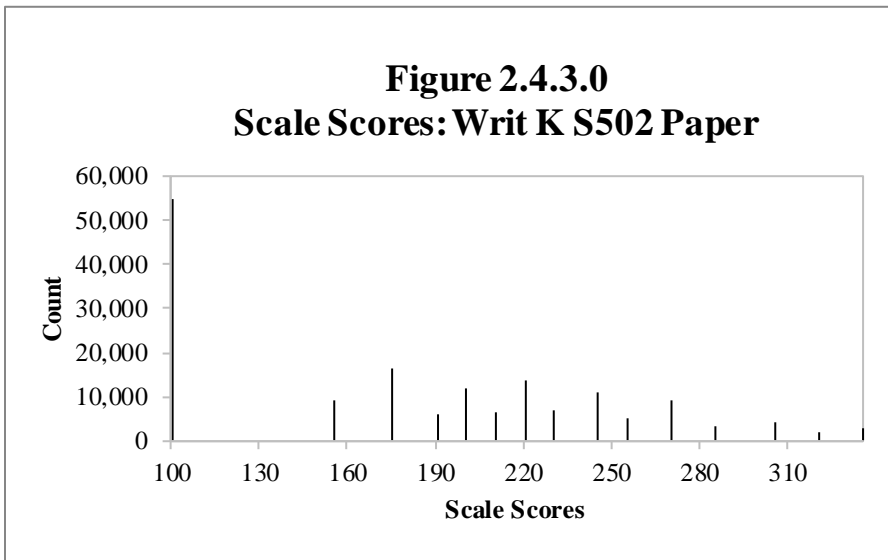
2.4.3 Writing

2.4.3.0 Kindergarten

Table 2.4.3.0

Scale Score Descriptive Statistics: Writ K S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	163,216	100	339	182.86	69.48
Total	163,216	100	339	182.86	69.48



2.4.3.1 Grade 1

Table 2.4.3.1.1

Scale Score Descriptive Statistics: Writ 1 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,252	111	321	229.41	36.92
Total	15,252	111	321	229.41	36.92

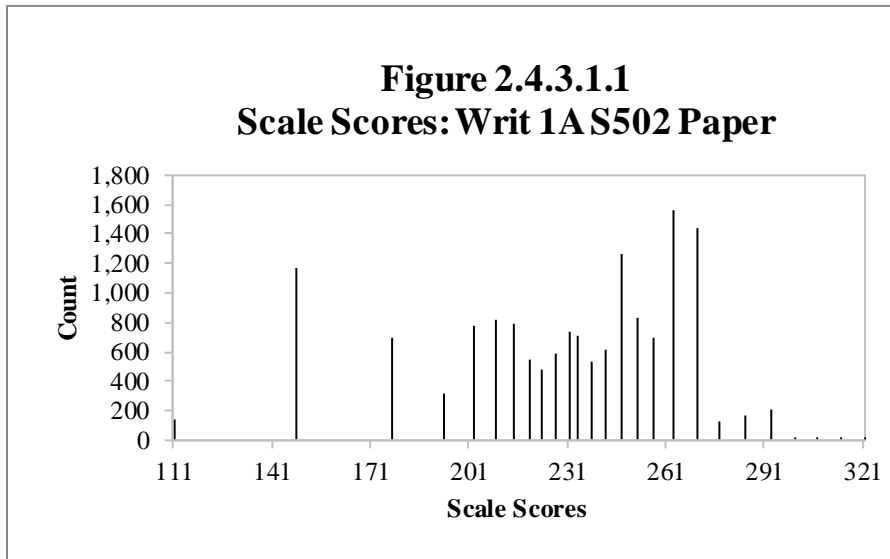


Table 2.4.3.1.2

Scale Score Descriptive Statistics: Writ 1 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,300	111	413	255.76	50.43
Total	15,300	111	413	255.76	50.43

Figure 2.4.3.1.2
Scale Scores: Writ 1B/C S502 Paper

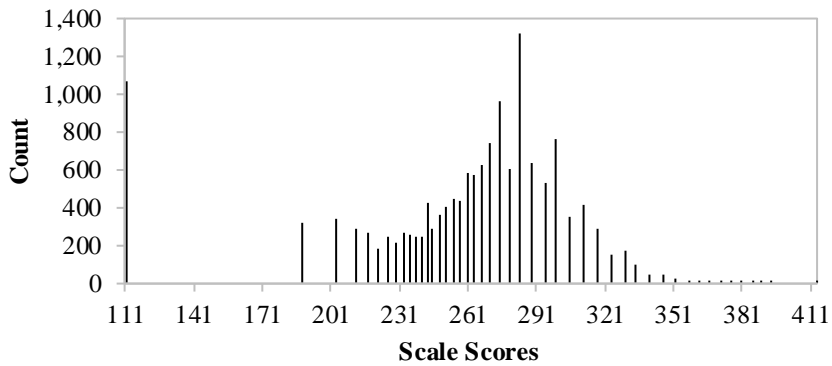
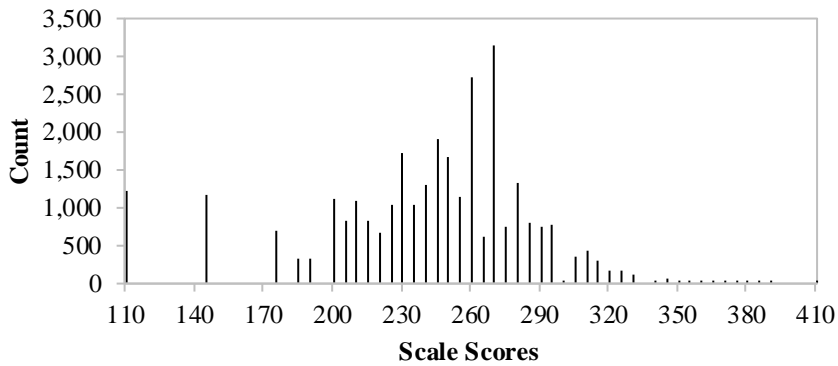


Table 2.4.3.1.3

Scale Score Descriptive Statistics: Writ 1 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	30,552	111	413	242.61	46.12
Total	30,552	111	413	242.61	46.12

Figure 2.4.3.1.3
Scale Scores: Writ 1 S502 Paper



2.4.3.2 Grade 2

Table 2.4.3.2.1

Scale Score Descriptive Statistics: Writ 2 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	7,231	133	367	241.30	46.85
Total	7,231	133	367	241.30	46.85

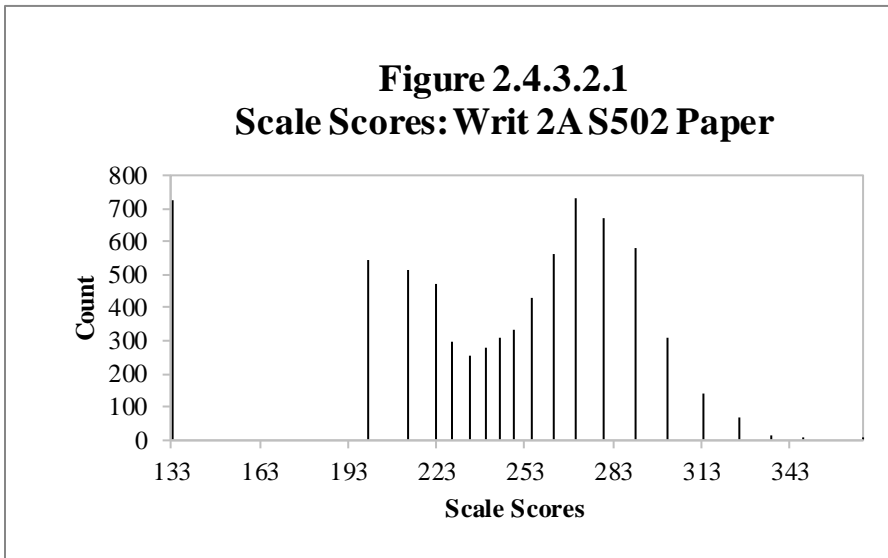


Table 2.4.3.2.2

Scale Score Descriptive Statistics: Writ 2 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	23,861	133	404	287.02	38.04
Total	23,861	133	404	287.02	38.04

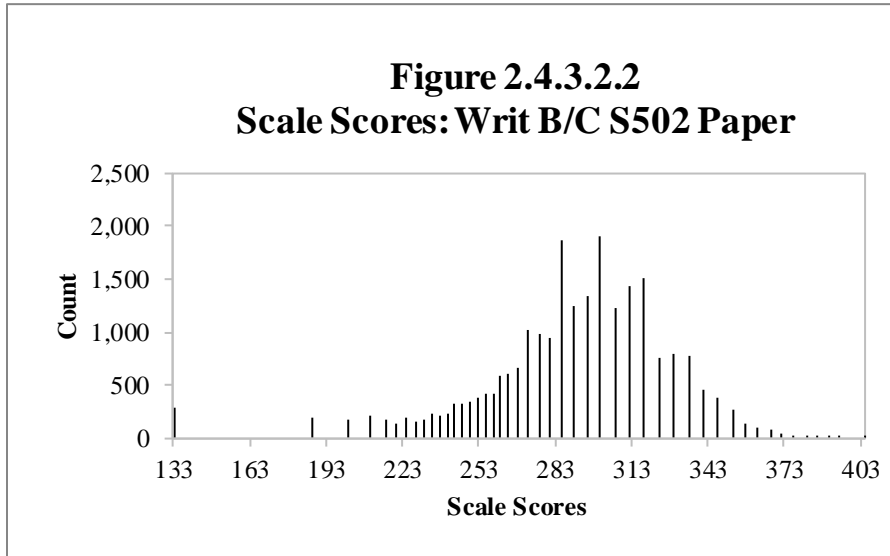
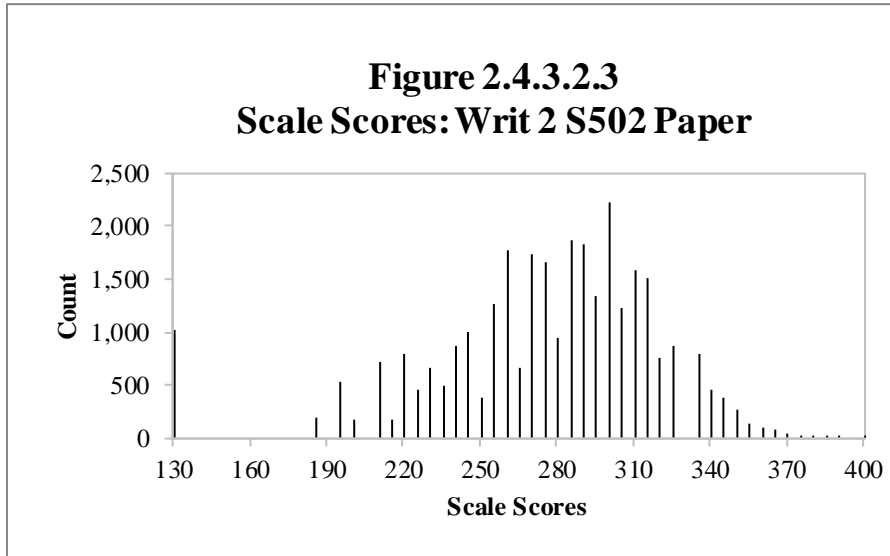


Table 2.4.3.2.3

Scale Score Descriptive Statistics: Writ 2 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	31,092	133	404	276.39	44.65
Total	31,092	133	404	276.39	44.65



2.4.3.3 Grade 3

Table 2.4.3.3.1

Scale Score Descriptive Statistics: Writ 3 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	4,974	133	376	250.29	45.64
Total	4,974	133	376	250.29	45.64

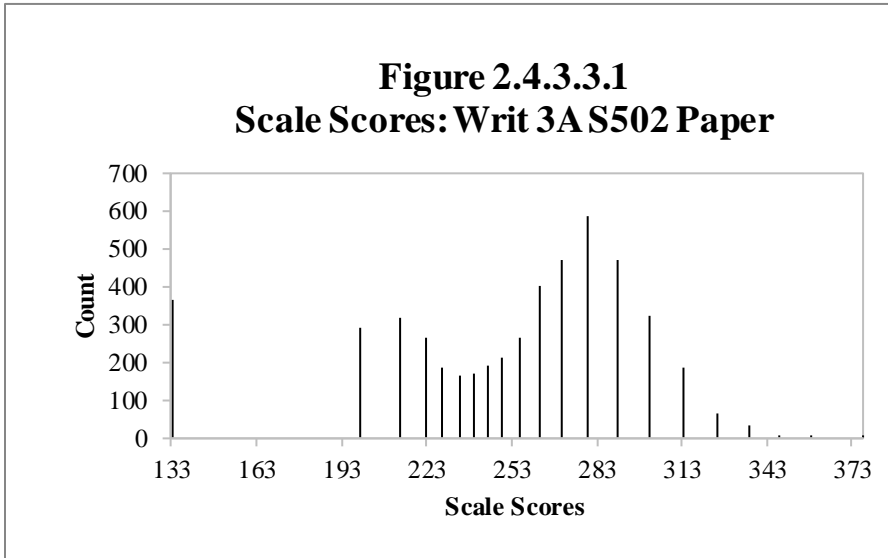


Table 2.4.3.3.2

Scale Score Descriptive Statistics: Writ 3 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	22,675	133	430	305.01	34.74
Total	22,675	133	430	305.01	34.74

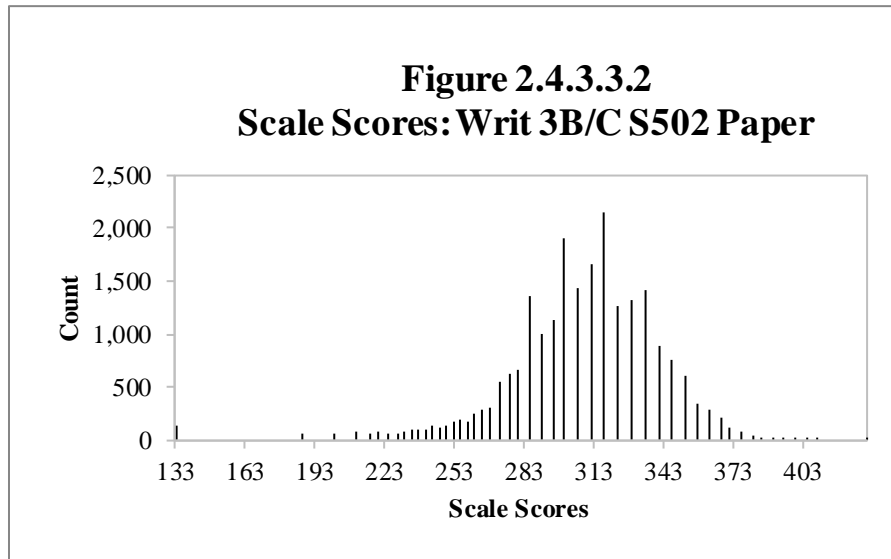
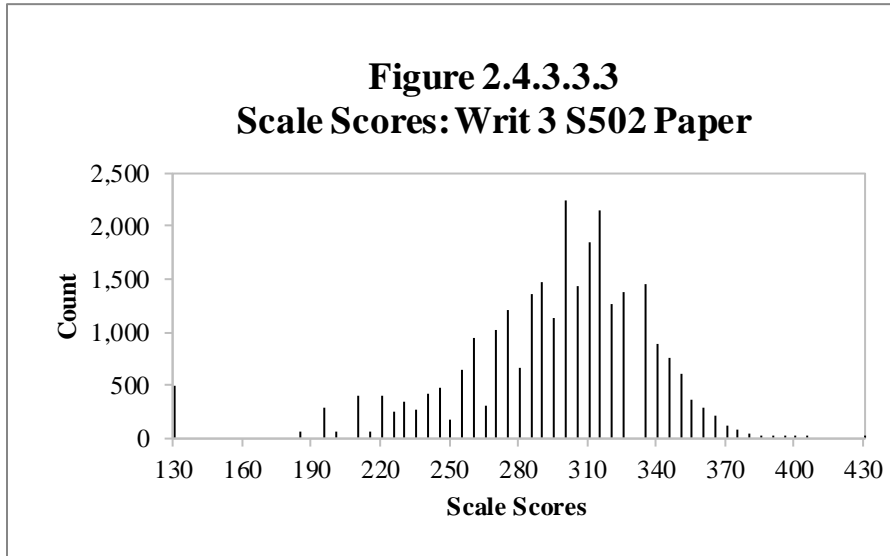


Table 2.4.3.3.3

Scale Score Descriptive Statistics: Writ 3 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	27,649	133	430	295.17	42.50
Total	27,649	133	430	295.17	42.50



2.4.3.4 Grades 4–5

Table 2.4.3.4.1

Scale Score Descriptive Statistics: Writ 4-5 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	3,531	155	388	272.68	44.08
5	3,114	155	388	279.63	41.40
Total	6,645	155	388	275.94	42.98

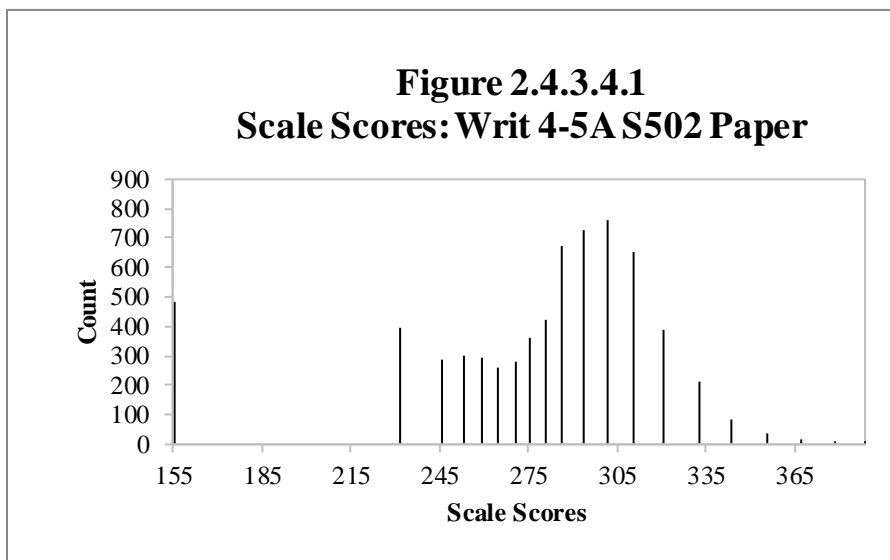


Table 2.4.3.4.2

Scale Score Descriptive Statistics: Writ 4-5 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	22,060	155	434	333.36	32.32
5	16,526	155	457	349.40	32.02
Total	38,586	155	457	340.23	33.15

Figure 2.4.3.4.2
Scale Scores: Writ 4-5B/C S502 Paper

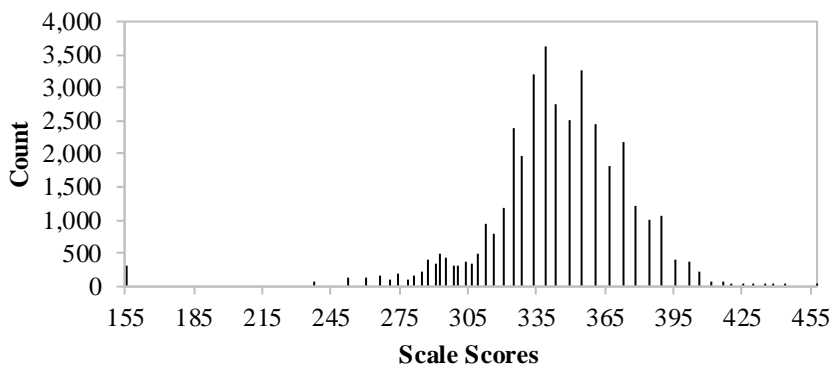
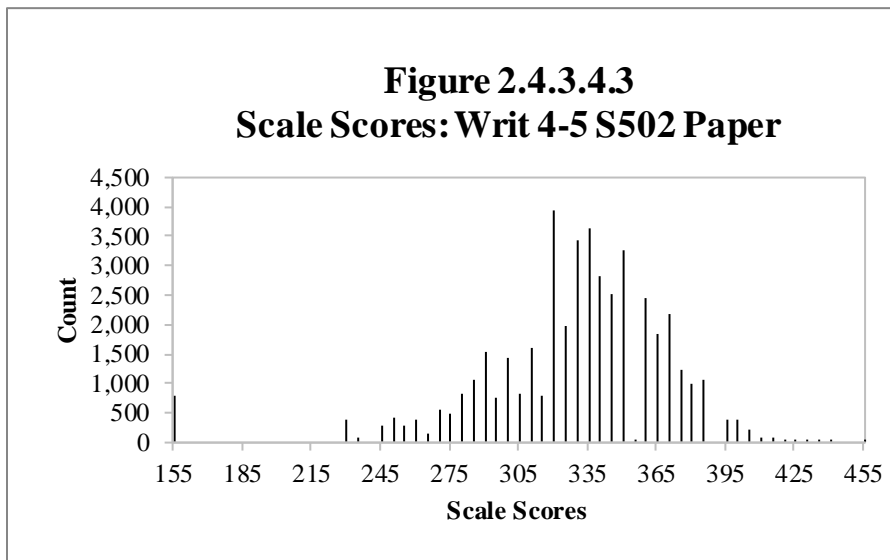


Table 2.4.3.4.3

Scale Score Descriptive Statistics: Writ 4-5 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	25,591	155	434	324.99	40.08
5	19,640	155	457	338.34	42.24
Total	45,231	155	457	330.79	41.56

Figure 2.4.3.4.3
Scale Scores: Writ 4-5 S502 Paper



2.4.3.5 Grades 6–8

Table 2.4.3.5.1

Scale Score Descriptive Statistics: Writ 6-8 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	3,077	188	389	276.78	31.74
7	2,913	188	389	279.50	31.46
8	2,972	188	389	281.86	30.65
Total	8,962	188	389	279.35	31.36

Figure 2.4.3.5.1
Scale Scores: Writ 6-8A S502 Paper

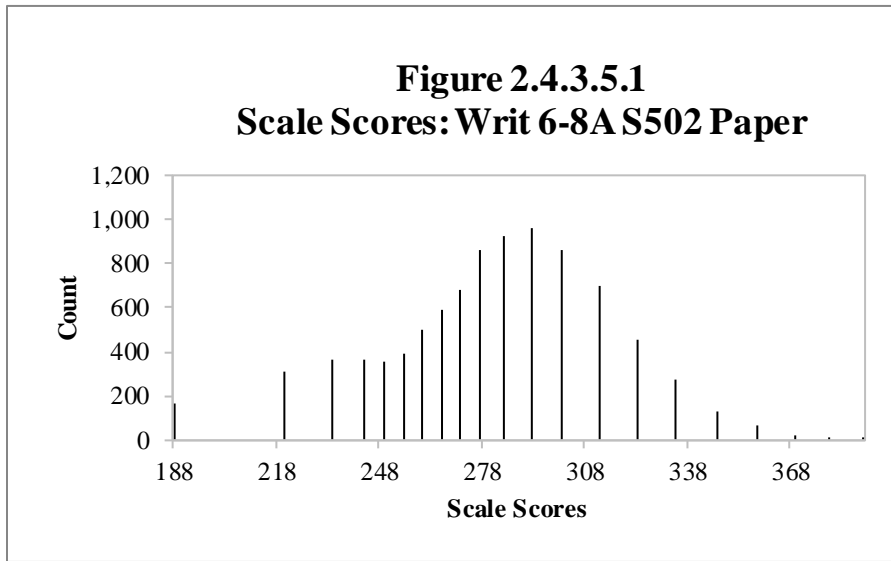


Table 2.4.3.5.2

Scale Score Descriptive Statistics: Writ 6-8 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	11,905	188	431	338.42	32.61
7	10,801	188	448	345.57	31.96
8	9,235	188	456	351.98	31.50
Total	31,941	188	456	344.76	32.54

Figure 2.4.3.5.2
Scale Scores: Writ 6-8B/C S502 Paper

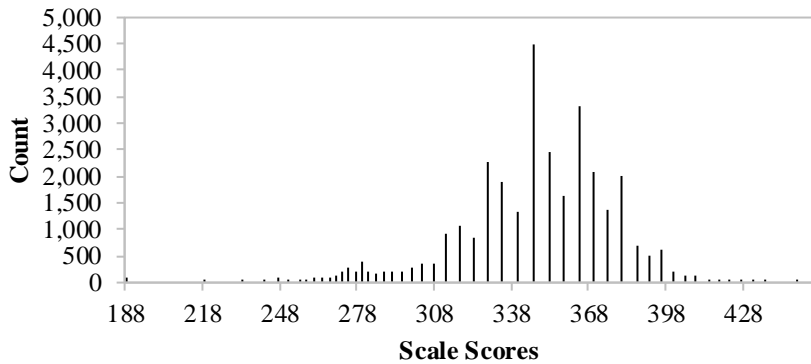
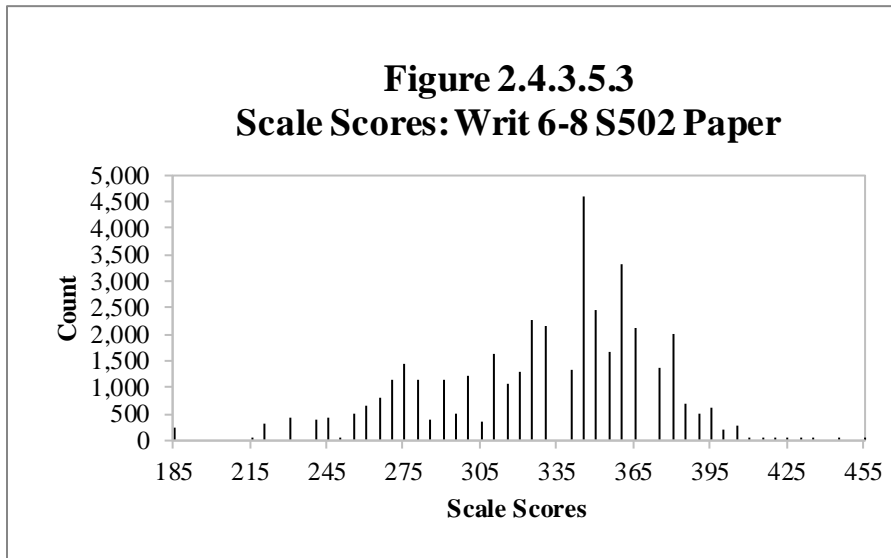


Table 2.4.3.5.3

Scale Score Descriptive Statistics: Writ 6-8 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	14,982	188	431	325.76	40.89
7	13,714	188	448	331.54	41.77
8	12,207	188	456	334.91	43.41
Total	40,903	188	456	330.43	42.12

Figure 2.4.3.5.3
Scale Scores: Writ 6-8 S502 Paper



2.4.3.6 Grades 9–12

Table 2.4.3.6.1

Scale Score Descriptive Statistics: Writ 9-12 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	2,802	232	439	317.37	37.82
10	2,539	232	430	318.75	37.25
11	1,972	232	439	325.10	36.32
12	1,241	232	430	328.77	35.25
Total	8,554	232	439	321.21	37.18

Figure 2.4.3.6.1
Scale Scores: Writ 9-12A S502 Paper

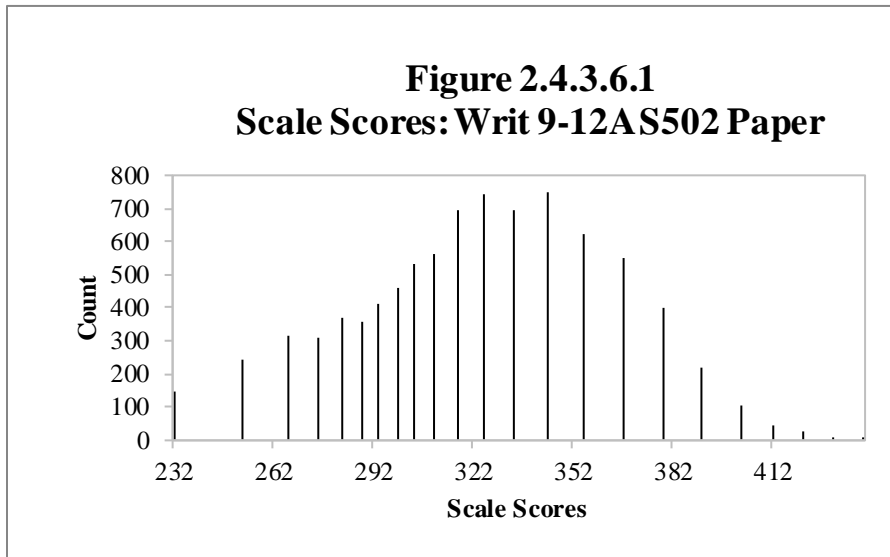


Table 2.4.3.6.2

Scale Score Descriptive Statistics: Writ 9-12 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	8,406	232	464	366.29	32.87
10	7,752	232	464	369.21	33.51
11	6,959	232	458	372.05	34.09
12	4,897	232	453	370.85	34.41
Total	28,014	232	464	369.32	33.69

Figure 2.4.3.6.2
Scale Scores: Writ 9-12B/C S502 Paper

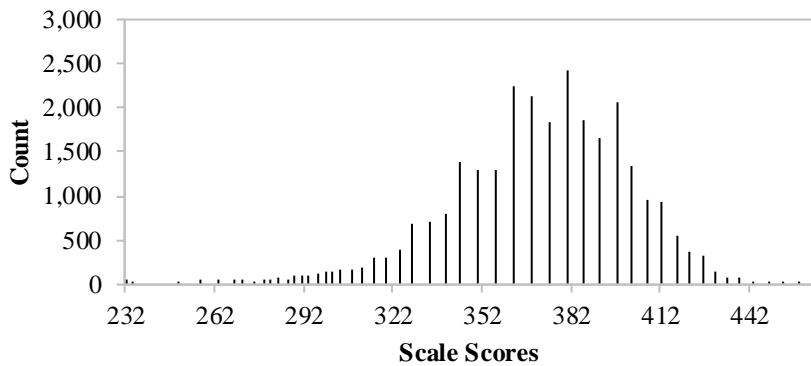
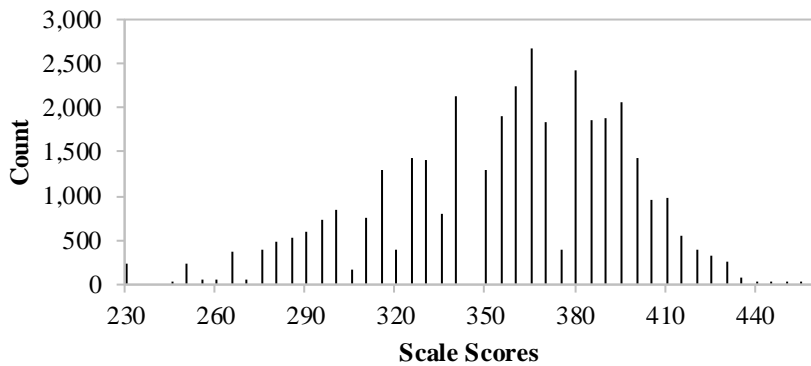


Table 2.4.3.6.3

Scale Score Descriptive Statistics: Writ 9-12 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	11,208	232	464	354.06	40.20
10	10,291	232	464	356.76	40.76
11	8,931	232	458	361.68	39.69
12	6,138	232	453	362.34	38.49
Total	36,568	232	464	358.07	40.10

Figure 2.4.3.6.3
Scale Scores: Writ 9-12 S502 Paper



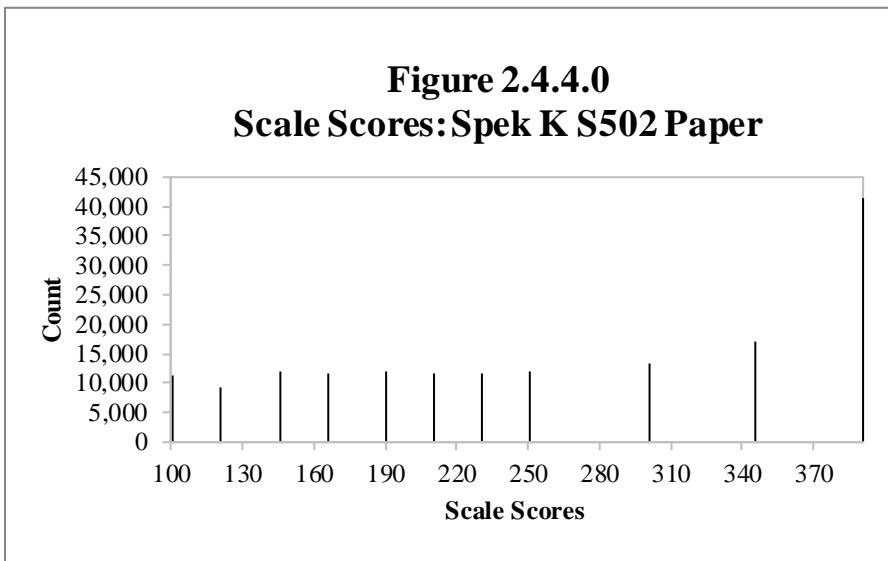
2.4.4 Speaking

2.4.4.0 Kindergarten

Table 2.4.4.0

Scale Score Descriptive Statistics: Spek K S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	163,192	100	392	261.26	101.64
Total	163,192	100	392	261.26	101.64



2.4.4.1 Grade 1

Table 2.4.4.1.1

Scale Score Descriptive Statistics: Spek 1 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,001	106	391	246.09	65.57
Total	15,001	106	391	246.09	65.57

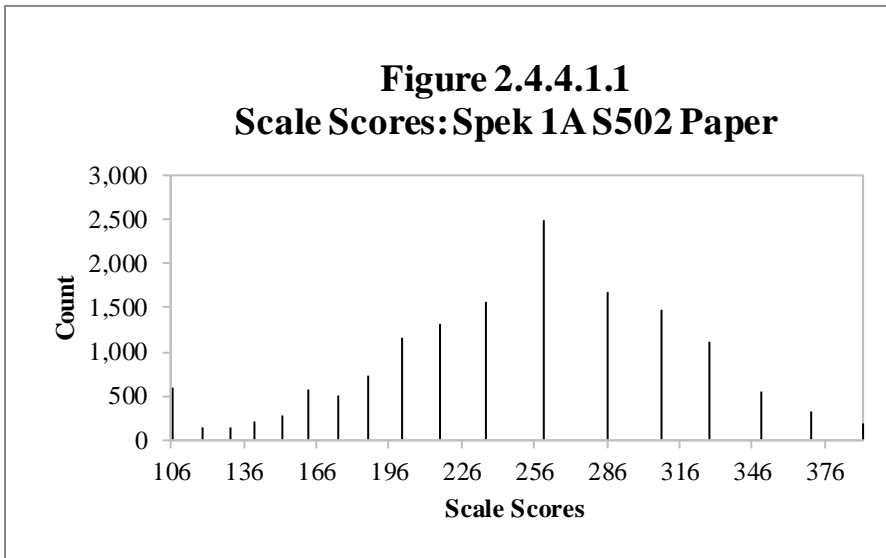


Table 2.4.4.1.2

Scale Score Descriptive Statistics: Spek 1 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,099	106	407	293.16	53.49
Total	15,099	106	407	293.16	53.49

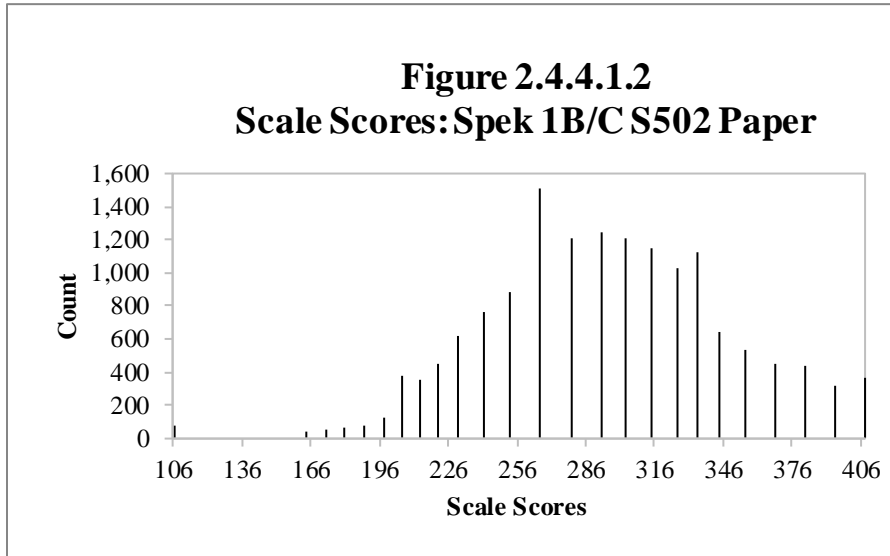
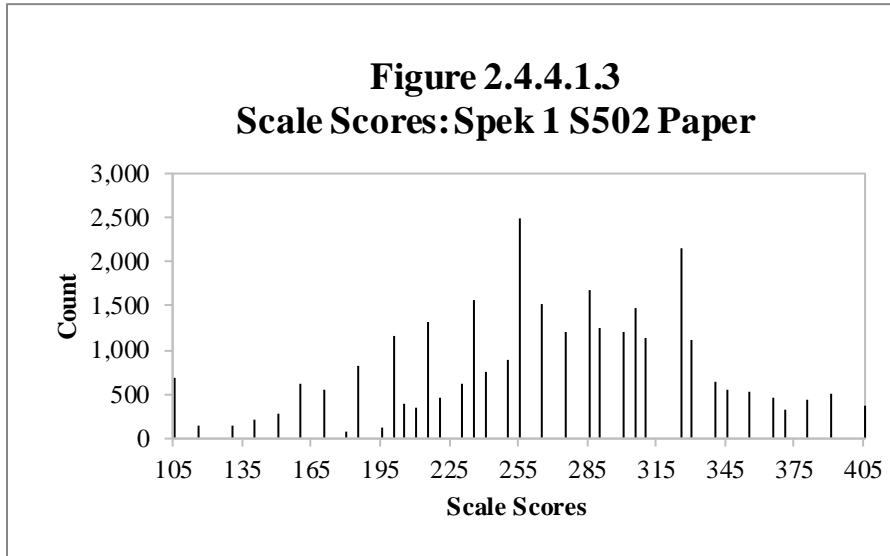


Table 2.4.4.1.3

Scale Score Descriptive Statistics: Spek 1 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	30,100	106	407	269.70	64.28
Total	30,100	106	407	269.70	64.28



2.4.4.2 Grade 2

Table 2.4.4.2.1

Scale Score Descriptive Statistics: Spek 2 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	7,098	118	383	241.13	70.50
Total	7,098	118	383	241.13	70.50

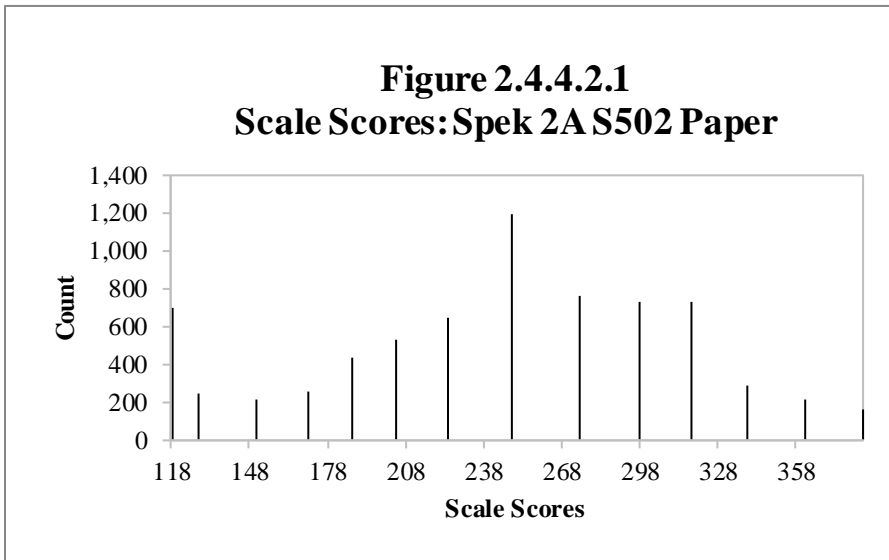


Table 2.4.4.2.2

Scale Score Descriptive Statistics: Spek 2 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	23,588	118	425	301.58	51.81
Total	23,588	118	425	301.58	51.81

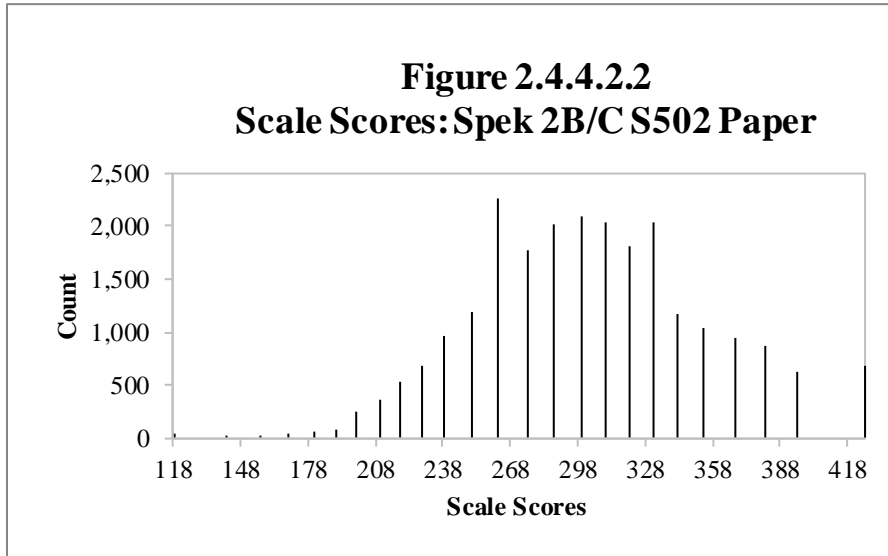
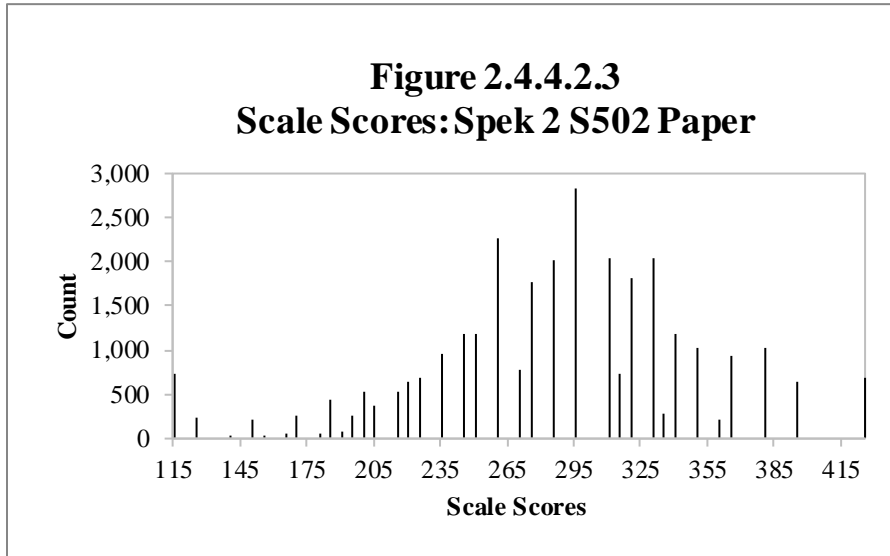


Table 2.4.4.2.3

Scale Score Descriptive Statistics: Spek 2 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	30,686	118	425	287.60	62.15
Total	30,686	118	425	287.60	62.15



2.4.4.3 Grade 3

Table 2.4.4.3.1

Scale Score Descriptive Statistics: Spek 3 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	4,886	118	383	233.74	73.49
Total	4,886	118	383	233.74	73.49

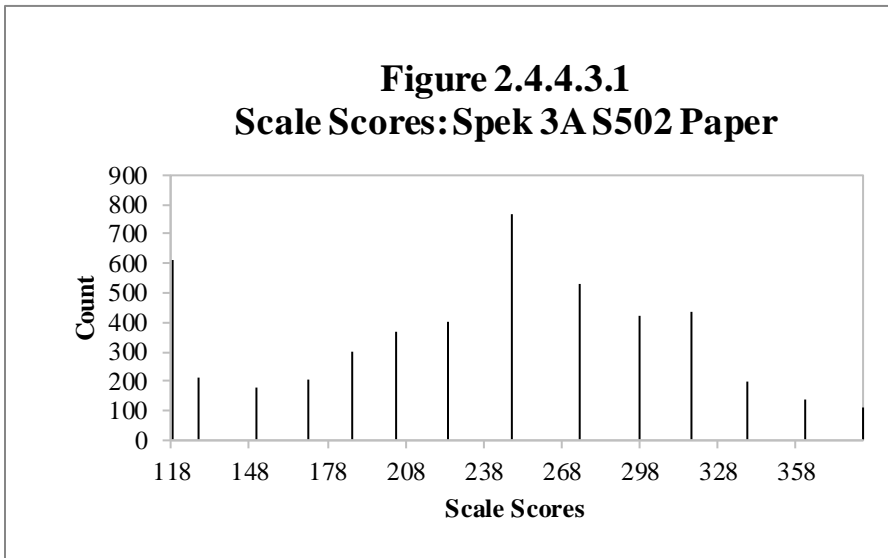


Table 2.4.4.3.2

Scale Score Descriptive Statistics: Spek 3 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	22,413	118	425	313.61	51.16
Total	22,413	118	425	313.61	51.16

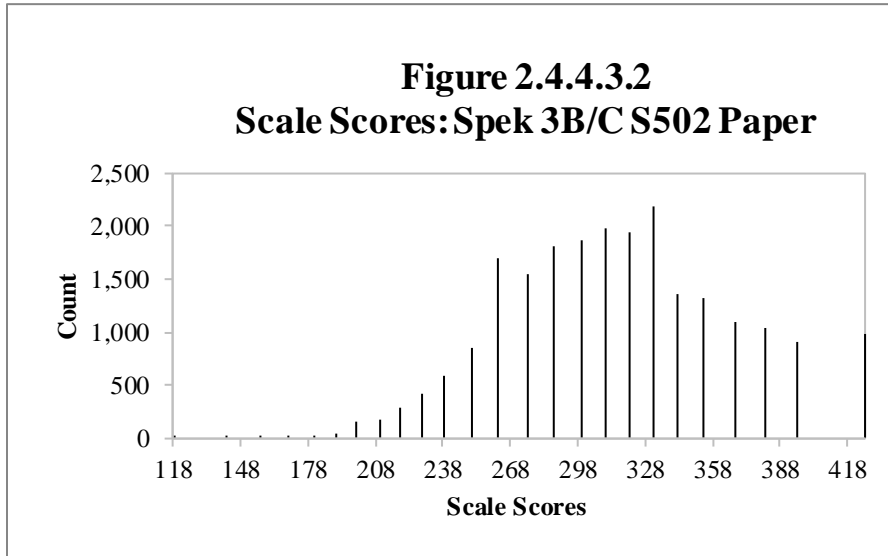
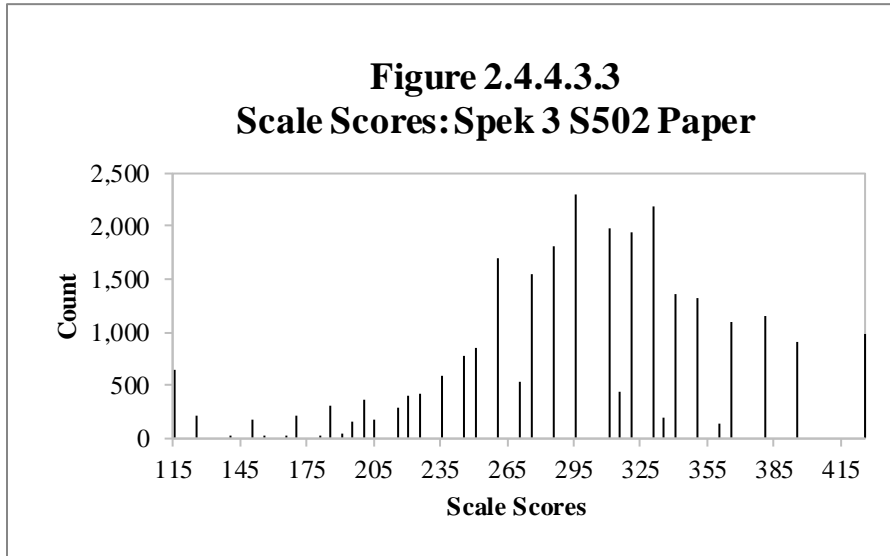


Table 2.4.4.3.3

Scale Score Descriptive Statistics: Spek 3 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	27,299	118	425	299.32	63.66
Total	27,299	118	425	299.32	63.66



2.4.4.4 Grades 4-5

Table 2.4.4.1

Scale Score Descriptive Statistics: Spek 4-5 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	3,505	130	423	248.49	73.32
5	3,096	130	423	247.29	73.38
Total	6,601	130	423	247.93	73.35

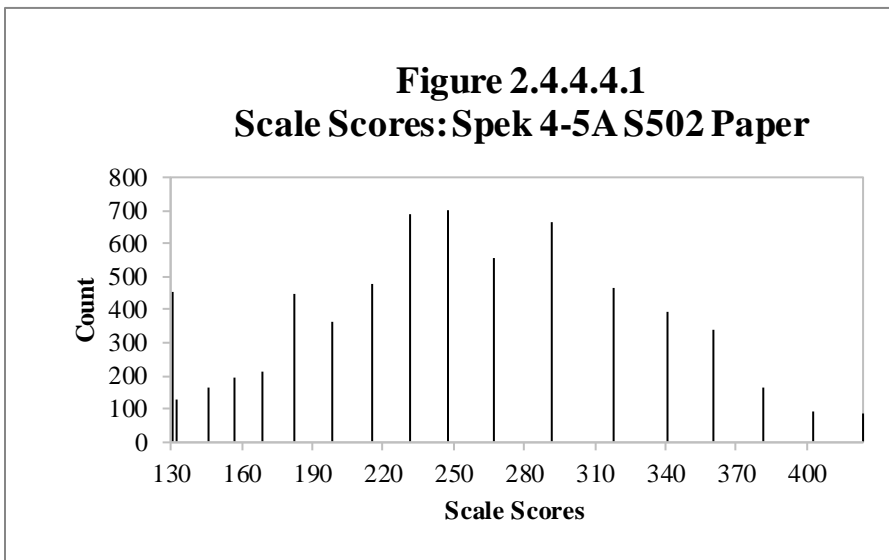


Table 2.4.4.2

Scale Score Descriptive Statistics: Spek 4-5 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	21,956	130	450	349.89	52.77
5	16,482	130	450	360.78	53.27
Total	38,438	130	450	354.56	53.26

Figure 2.4.4.2
Scale Scores: Spek 4-5B/C S502 Paper

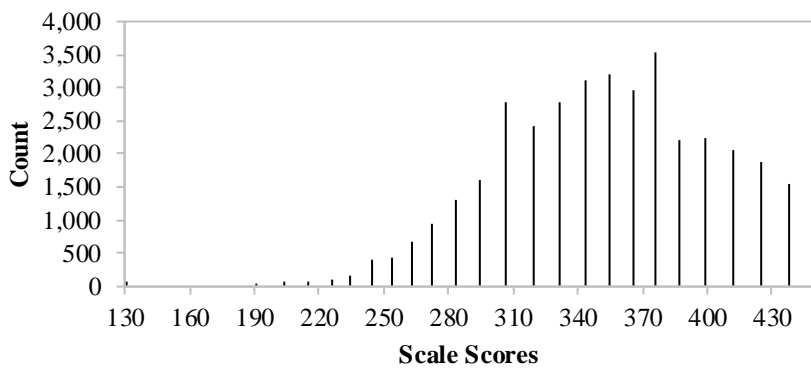
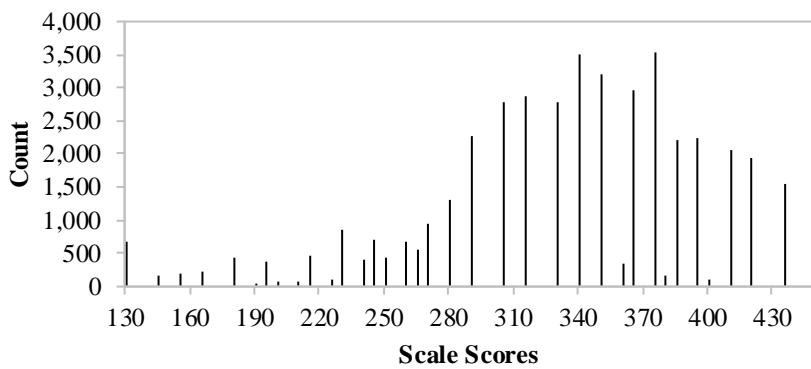


Table 2.4.4.3

Scale Score Descriptive Statistics: Spek 4-5 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	25,461	130	450	335.93	66.04
5	19,578	130	450	342.83	70.39
Total	45,039	130	450	338.93	68.05

Figure 2.4.4.3
Scale Scores: Spek 4-5 S502 Paper



2.4.4.5 Grades 6–8

Table 2.4.4.5.1

Scale Score Descriptive Statistics: Spek 6-8 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	3,040	148	459	278.22	72.49
7	2,894	148	459	278.39	72.83
8	2,945	148	459	278.19	71.54
Total	8,879	148	459	278.26	72.28

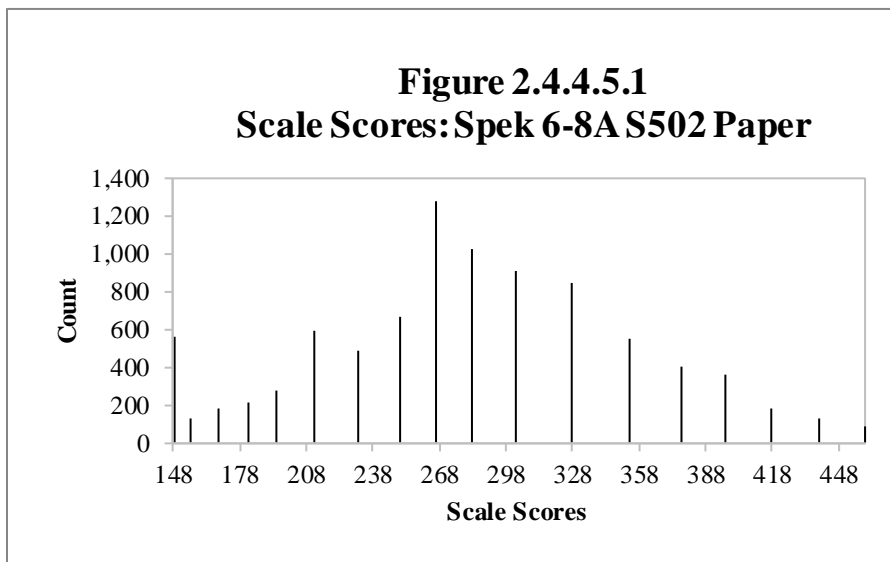


Table 2.4.4.5.2

Scale Score Descriptive Statistics: Spek 6-8 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	11,832	148	471	371.08	56.06
7	10,716	148	471	376.02	57.16
8	9,183	148	471	382.54	58.06
Total	31,731	148	471	376.07	57.20

Figure 2.4.4.5.2
Scale Scores: Spek 6-8B/C S502 Paper

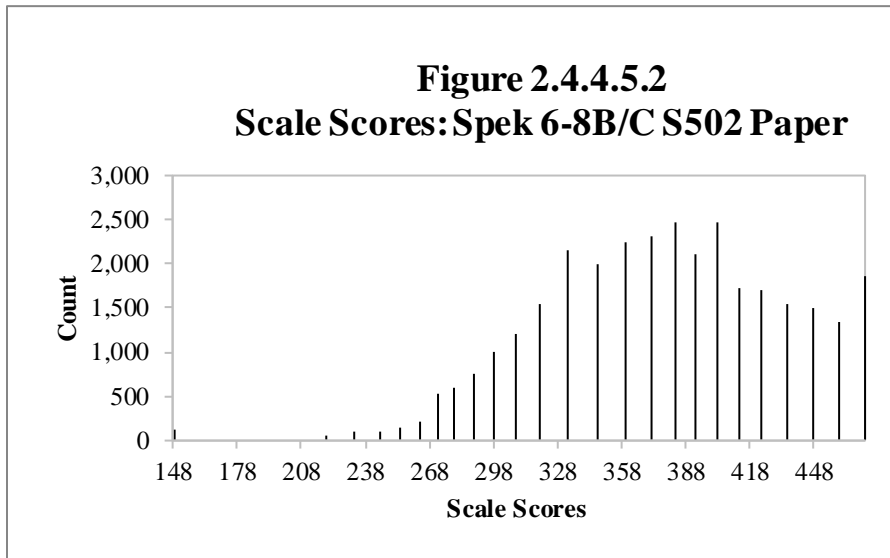
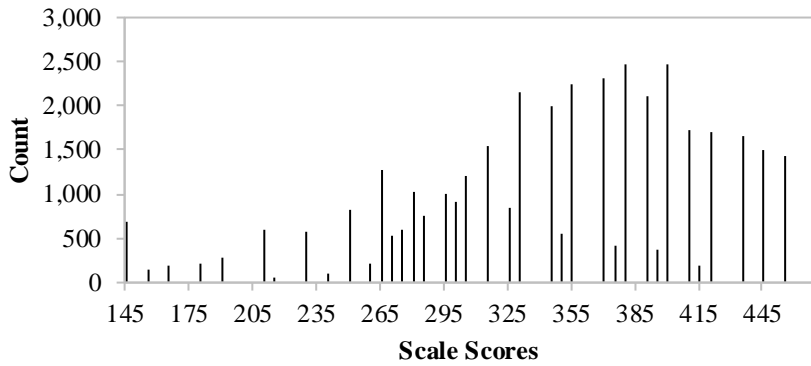


Table 2.4.4.5.3

Scale Score Descriptive Statistics: Spek 6-8 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	14,872	148	471	352.10	70.54
7	13,610	148	471	355.26	72.78
8	12,128	148	471	357.20	76.14
Total	40,610	148	471	354.68	73.03

Figure 2.4.4.5.3
Scale Scores: Spek 6-8 S502 Paper



2.4.4.6 Grades 9–12

Table 2.4.4.6.1

Scale Score Descriptive Statistics: Spek 9-12 A S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	2,754	172	445	274.75	69.43
10	2,515	172	445	273.17	68.39
11	1,937	172	445	283.72	69.10
12	1,219	172	445	289.49	66.61
Total	8,425	172	445	278.47	68.90

Figure 2.4.4.6.1
Scale Scores: Spek 9-12A S502 Paper

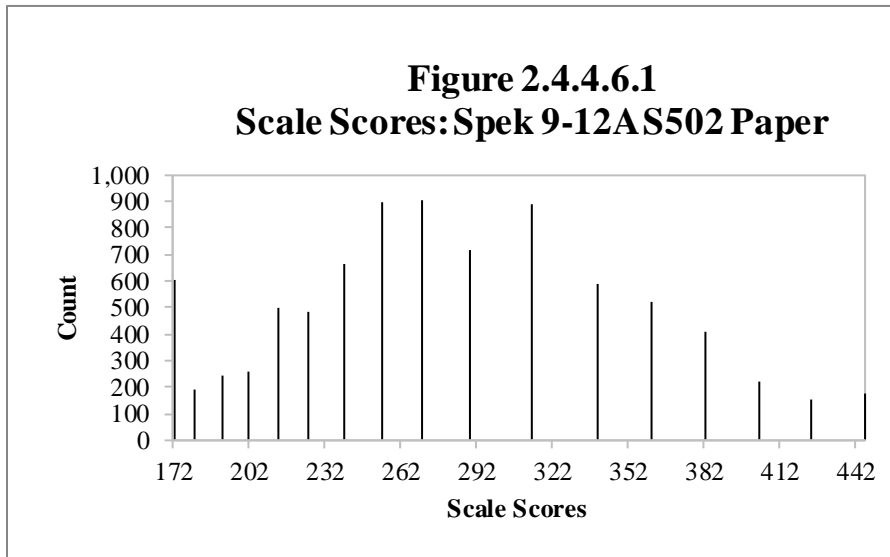


Table 2.4.4.6.2

Scale Score Descriptive Statistics: Spek 9-12 B/C S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	8,342	172	476	381.10	60.50
10	7,684	172	476	376.05	61.51
11	6,886	172	476	379.44	61.29
12	4,842	172	476	377.96	59.94
Total	27,754	172	476	378.74	60.91

Figure 2.4.4.6.2
Scale Scores: Spek 9-12B/C S502 Paper

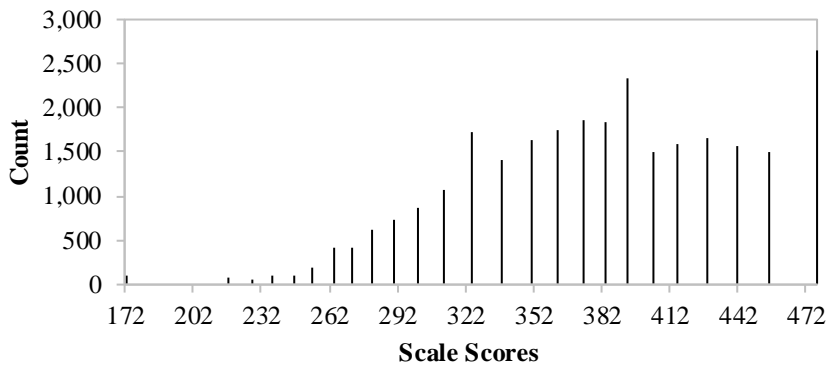
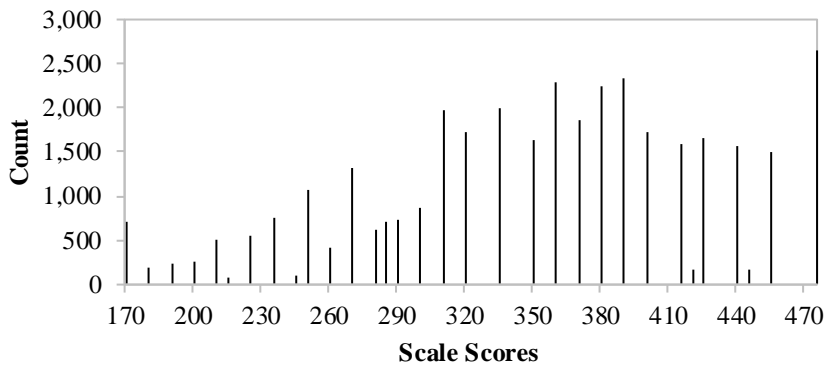


Table 2.4.4.6.3

Scale Score Descriptive Statistics: Spek 9-12 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	11,096	172	476	354.70	77.84
10	10,199	172	476	350.68	77.27
11	8,823	172	476	358.42	74.50
12	6,061	172	476	360.17	70.85
Total	36,179	172	476	355.39	75.81

Figure 2.4.4.6.3
Scale Scores: Spek 9-12 S502 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

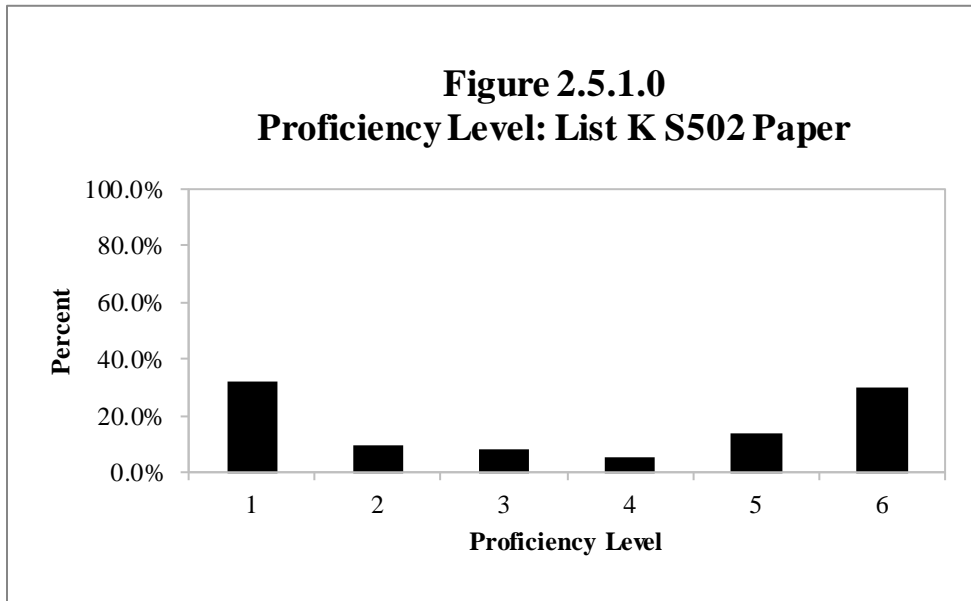
2.5.1 Listening

2.5.1.0 Kindergarten

Table 2.5.1.0

Proficiency Level Distribution: List K S502 Paper

Level	Grade K		Total	
	Count	Percent	Count	Percent
1	52,136	31.94%	52,136	31.94%
2	16,276	9.97%	16,276	9.97%
3	13,921	8.53%	13,921	8.53%
4	8,845	5.42%	8,845	5.42%
5	22,398	13.72%	22,398	13.72%
6	49,650	30.42%	49,650	30.42%
Total	163,226	100.00%	163,226	100.00%



2.5.1.1 Grade 1

Table 2.5.1.1.1

Proficiency Level Distribution: List 1 A S502 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	1,016	8.27%	1,016	8.27%
2	2,022	16.46%	2,022	16.46%
3	2,237	18.22%	2,237	18.22%
4	1,420	11.56%	1,420	11.56%
5	3,377	27.50%	3,377	27.50%
6	2,209	17.99%	2,209	17.99%
Total	12,281	100.00%	12,281	100.00%

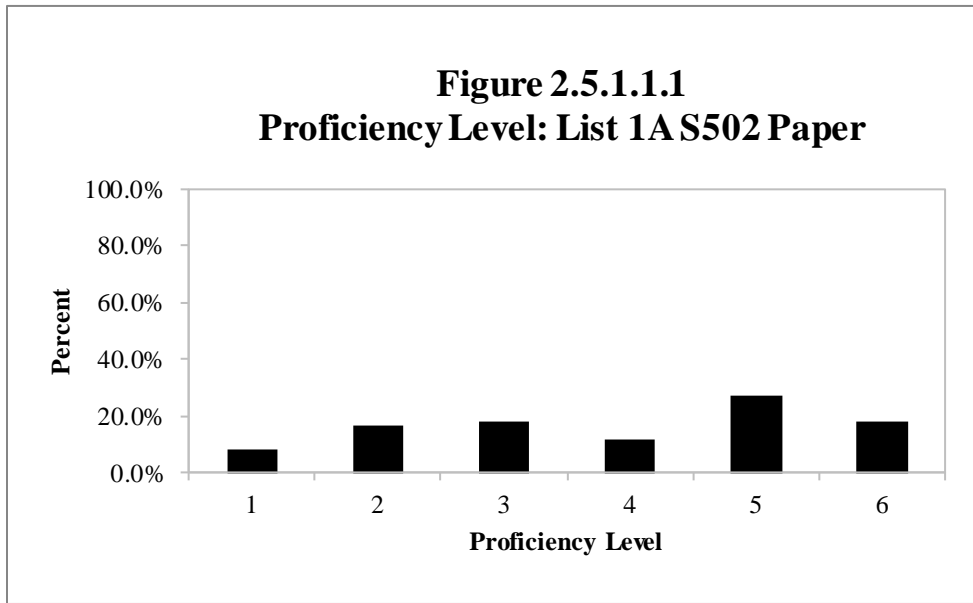


Table 2.5.1.1.2

Proficiency Level Distribution: List 1 B/C S502 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	170	1.29%	170	1.29%
2	472	3.59%	472	3.59%
3	1,794	13.63%	1,794	13.63%
4	2,307	17.53%	2,307	17.53%
5	3,019	22.94%	3,019	22.94%
6	5,397	41.01%	5,397	41.01%
Total	13,159	100.00%	13,159	100.00%

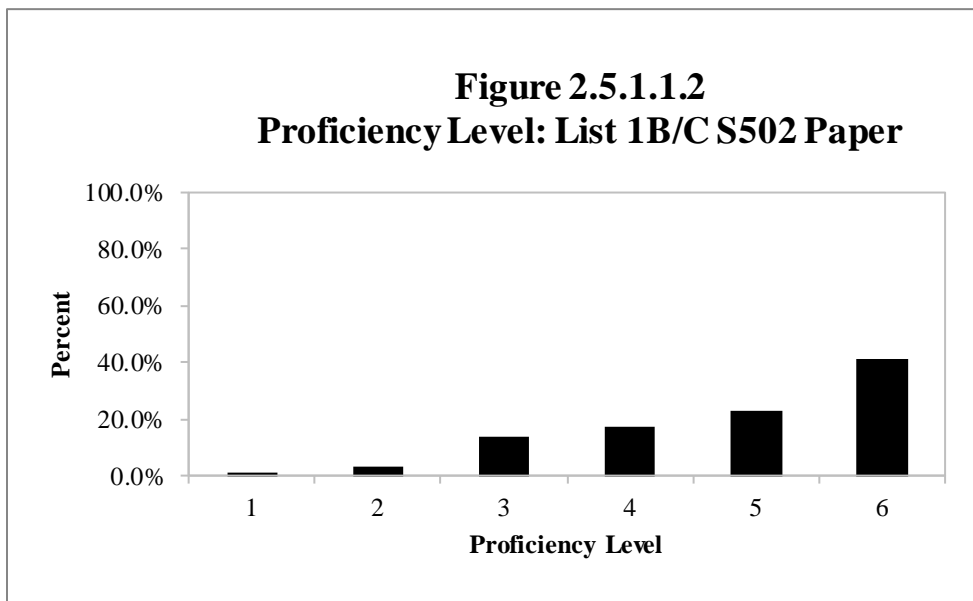
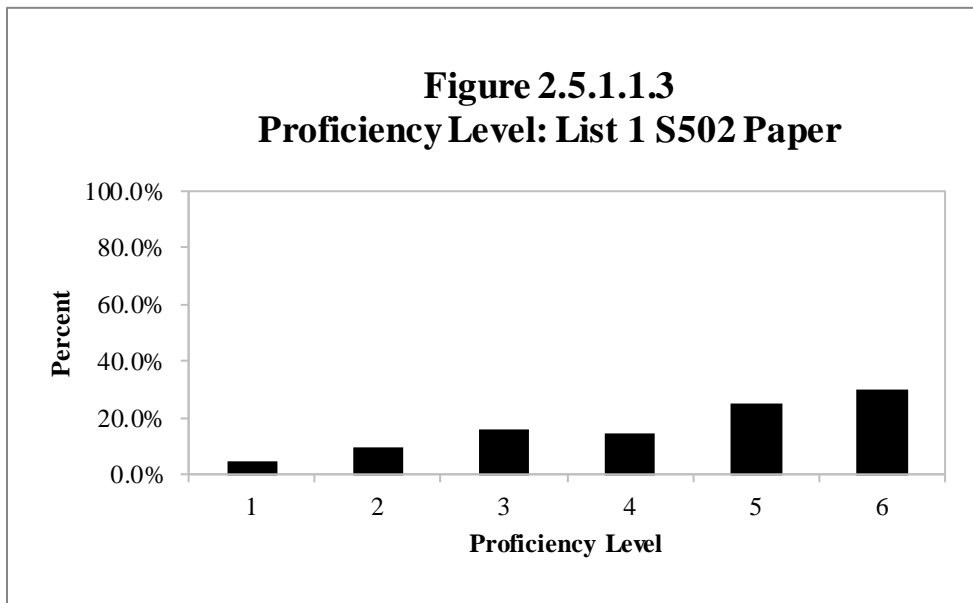


Table 2.5.1.1.3

Proficiency Level Distribution: List 1 S502 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	1,186	4.66%	1,186	4.66%
2	2,494	9.80%	2,494	9.80%
3	4,031	15.85%	4,031	15.85%
4	3,727	14.65%	3,727	14.65%
5	6,396	25.14%	6,396	25.14%
6	7,606	29.90%	7,606	29.90%
Total	25,440	100.00%	25,440	100.00%



2.5.1.2 Grade 2

Table 2.5.1.2.1

Proficiency Level Distribution: List 2 A S502 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	664	10.32%	664	10.32%
2	1,352	21.01%	1,352	21.01%
3	1,410	21.91%	1,410	21.91%
4	956	14.85%	956	14.85%
5	2,054	31.91%	2,054	31.91%
6	0	0.00%	0	0.00%
Total	6,436	100.00%	6,436	100.00%

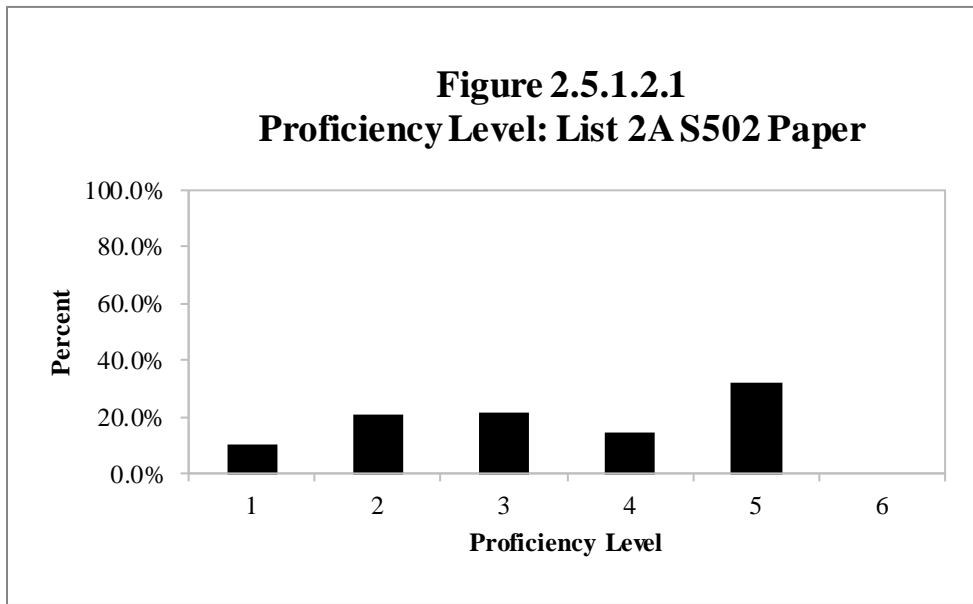


Table 2.5.1.2.2

Proficiency Level Distribution: List 2 B/C S502 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	141	0.64%	141	0.64%
2	1,190	5.37%	1,190	5.37%
3	3,809	17.18%	3,809	17.18%
4	2,247	10.13%	2,247	10.13%
5	5,979	26.96%	5,979	26.96%
6	8,810	39.73%	8,810	39.73%
Total	22,176	100.00%	22,176	100.00%

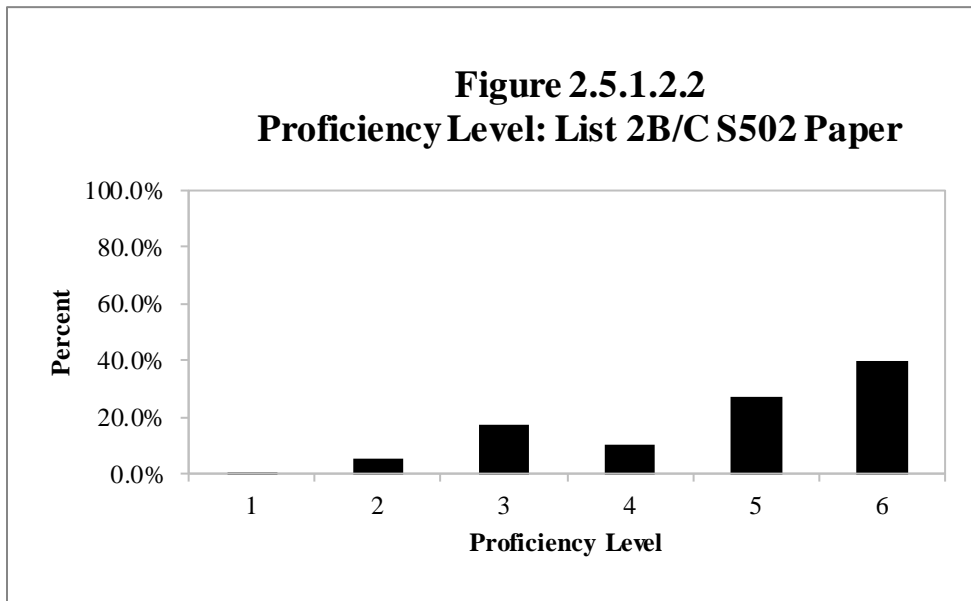
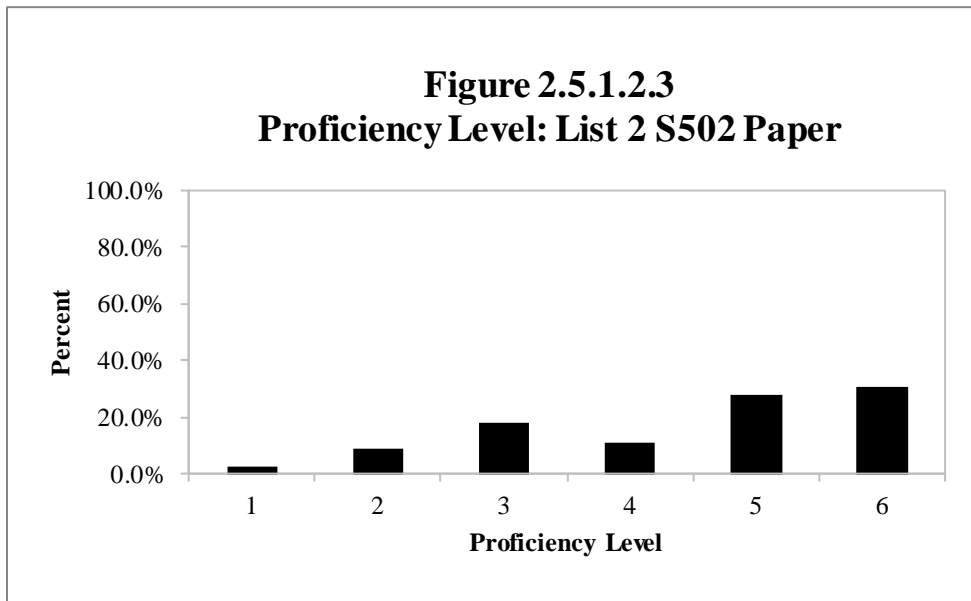


Table 2.5.1.2.3

Proficiency Level Distribution: List 2 S502 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	805	2.81%	805	2.81%
2	2,542	8.88%	2,542	8.88%
3	5,219	18.24%	5,219	18.24%
4	3,203	11.19%	3,203	11.19%
5	8,033	28.08%	8,033	28.08%
6	8,810	30.79%	8,810	30.79%
Total	28,612	100.00%	28,612	100.00%

Figure 2.5.1.2.3
Proficiency Level: List 2 S502 Paper



2.5.1.3 Grade 3

Table 2.5.1.3.1

Proficiency Level Distribution: List 3 A S502 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	151	3.36%	151	3.36%
2	1,258	27.97%	1,258	27.97%
3	1,395	31.01%	1,395	31.01%
4	730	16.23%	730	16.23%
5	551	12.25%	551	12.25%
6	413	9.18%	413	9.18%
Total	4,498	100.00%	4,498	100.00%

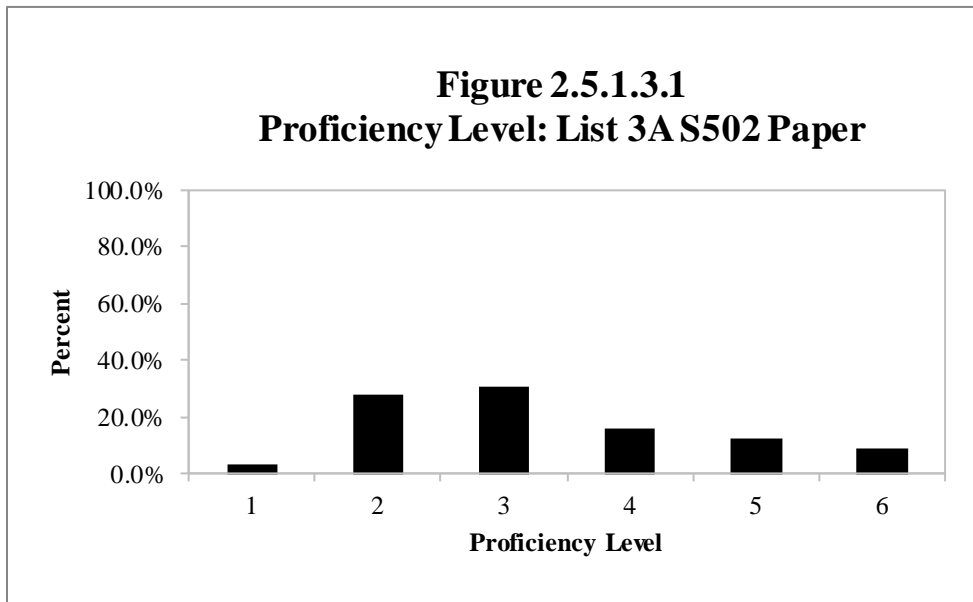


Table 2.5.1.3.2

Proficiency Level Distribution: List 3 B/C S502 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	14	0.07%	14	0.07%
2	528	2.47%	528	2.47%
3	3,097	14.49%	3,097	14.49%
4	2,049	9.59%	2,049	9.59%
5	8,207	38.40%	8,207	38.40%
6	7,478	34.99%	7,478	34.99%
Total	21,373	100.00%	21,373	100.00%

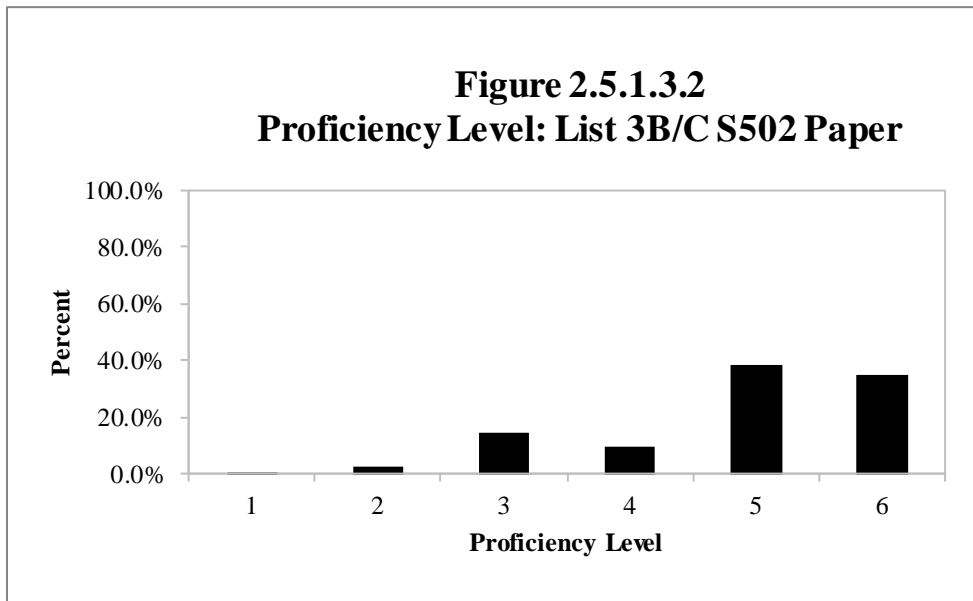
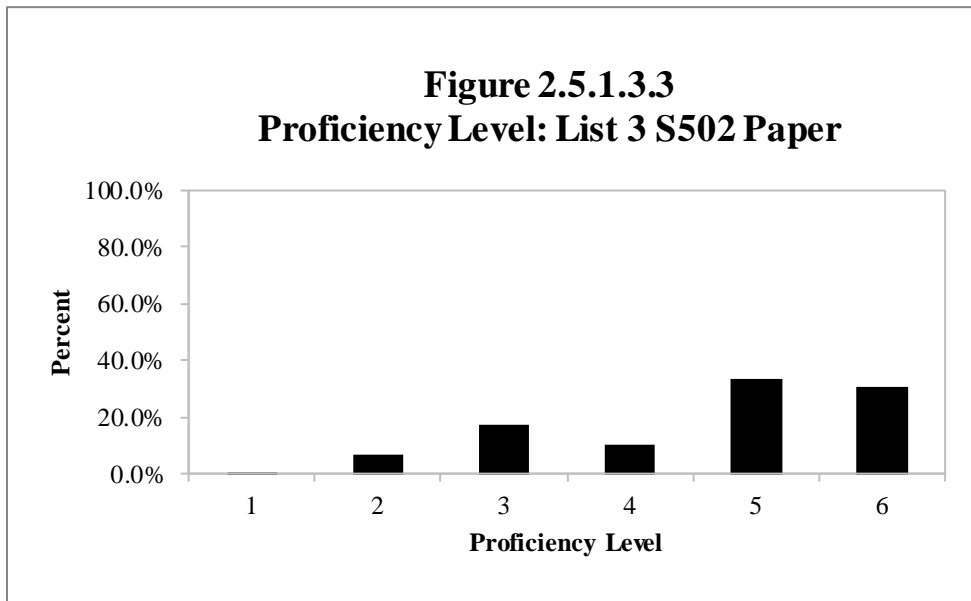


Table 2.5.1.3.3

Proficiency Level Distribution: List 3 S502 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	165	0.64%	165	0.64%
2	1,786	6.90%	1,786	6.90%
3	4,492	17.36%	4,492	17.36%
4	2,779	10.74%	2,779	10.74%
5	8,758	33.85%	8,758	33.85%
6	7,891	30.50%	7,891	30.50%
Total	25,871	100.00%	25,871	100.00%

Figure 2.5.1.3.3
Proficiency Level: List 3 S502 Paper



2.5.1.4 Grades 4–5

Table 2.5.1.4.1

Proficiency Level Distribution: List 4-5 A S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	187	5.68%	271	9.19%	458	7.34%
2	933	28.34%	950	32.21%	1,883	30.17%
3	965	29.31%	824	27.94%	1,789	28.67%
4	540	16.40%	449	15.23%	989	15.85%
5	399	12.12%	187	6.34%	586	9.39%
6	268	8.14%	268	9.09%	536	8.59%
Total	3,292	100.00%	2,949	100.00%	6,241	100.00%

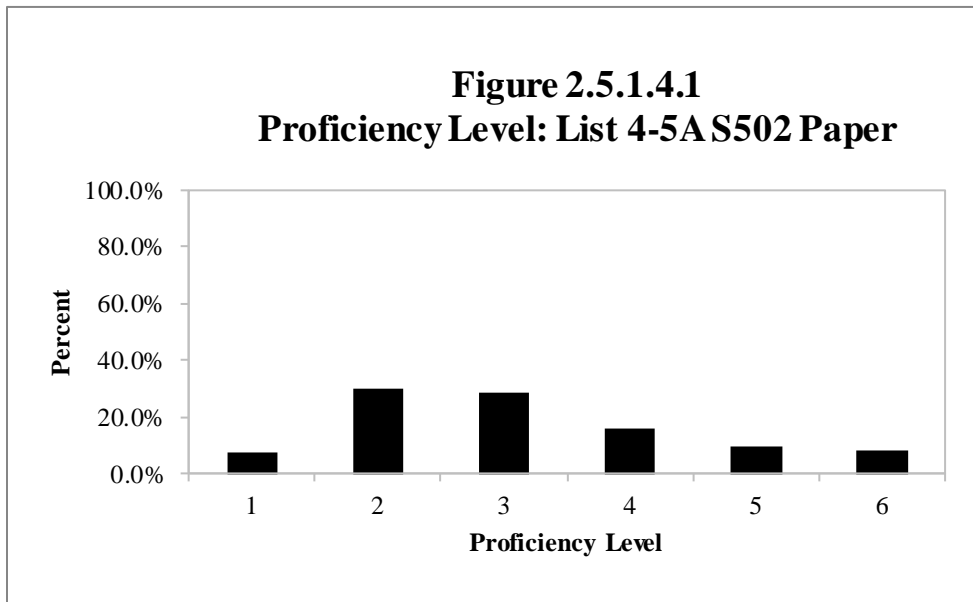


Table 2.5.1.4.2

Proficiency Level Distribution: List 4-5 B/C S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	25	0.12%	7	0.04%	32	0.09%
2	443	2.08%	358	2.23%	801	2.14%
3	2,423	11.38%	1,543	9.59%	3,966	10.61%
4	3,611	16.96%	2,420	15.04%	6,031	16.14%
5	7,985	37.51%	6,256	38.89%	14,241	38.10%
6	6,801	31.95%	5,502	34.20%	12,303	32.92%
Total	21,288	100.00%	16,086	100.00%	37,374	100.00%

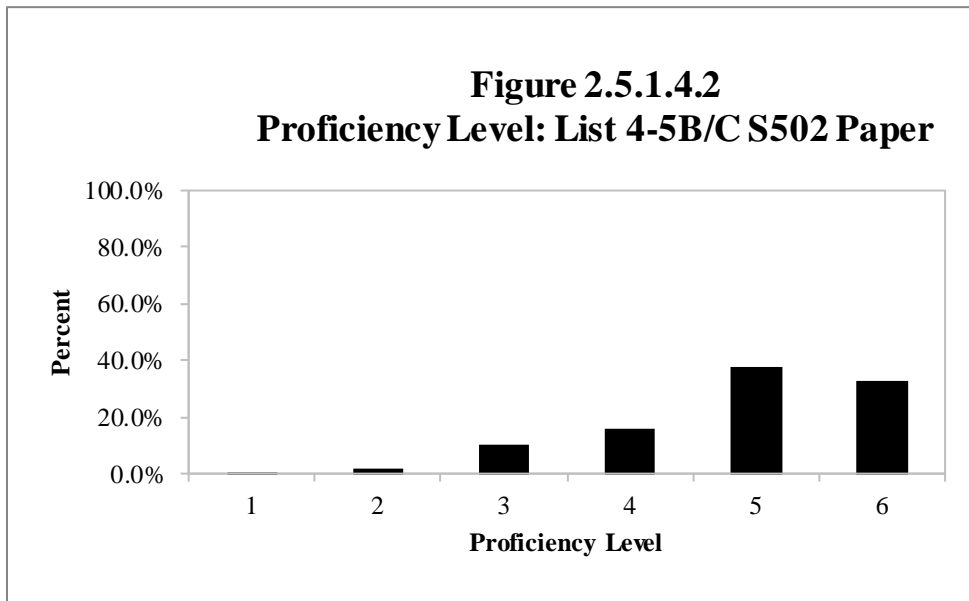
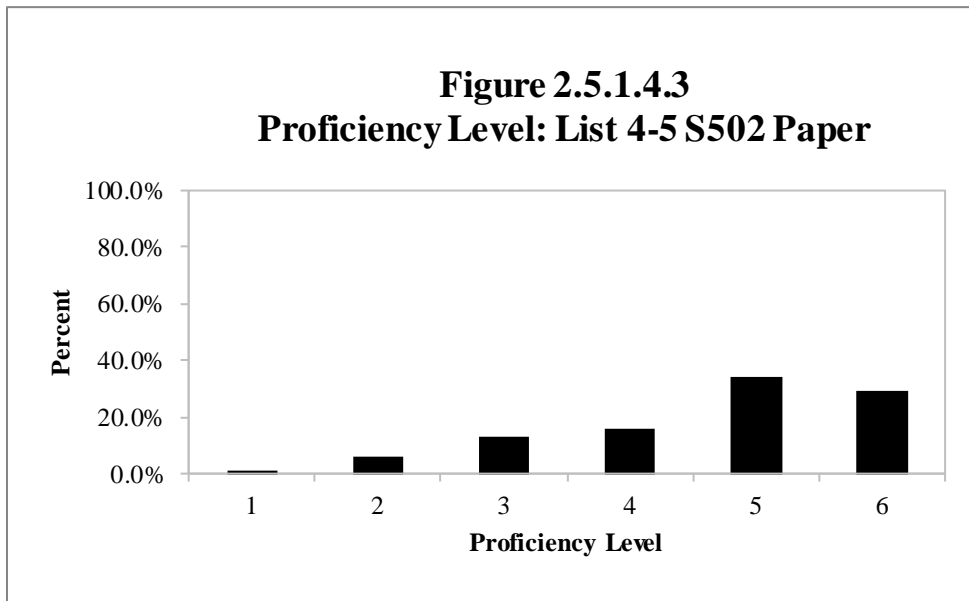


Table 2.5.1.4.3

Proficiency Level Distribution: List 4-5 S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	212	0.86%	278	1.46%	490	1.12%
2	1,376	5.60%	1,308	6.87%	2,684	6.15%
3	3,388	13.78%	2,367	12.43%	5,755	13.20%
4	4,151	16.89%	2,869	15.07%	7,020	16.10%
5	8,384	34.11%	6,443	33.85%	14,827	34.00%
6	7,069	28.76%	5,770	30.31%	12,839	29.44%
Total	24,580	100.00%	19,035	100.00%	43,615	100.00%



2.5.1.5 Grades 6–8

Table 2.5.1.5.1

Proficiency Level Distribution: List 6-8 A S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	674	23.15%	954	35.22%	971	34.77%	2,599	30.89%
2	1,279	43.92%	837	30.90%	1,122	40.17%	3,238	38.48%
3	417	14.32%	554	20.45%	302	10.81%	1,273	15.13%
4	302	10.37%	124	4.58%	247	8.84%	673	8.00%
5	198	6.80%	169	6.24%	151	5.41%	518	6.16%
6	42	1.44%	71	2.62%	0	0.00%	113	1.34%
Total	2,912	100.00%	2,709	100.00%	2,793	100.00%	8,414	100.00%

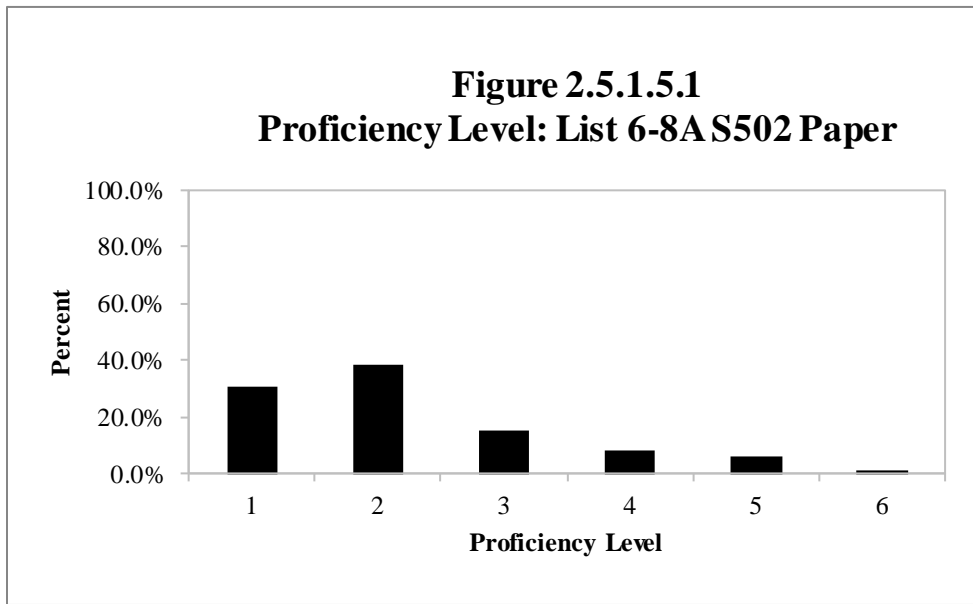


Table 2.5.1.5.2

Proficiency Level Distribution: List 6-8 B/C S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	16	0.14%	20	0.19%	20	0.22%	56	0.18%
2	341	2.98%	456	4.38%	306	3.44%	1,103	3.59%
3	1,407	12.31%	1,424	13.68%	1,592	17.89%	4,423	14.39%
4	3,620	31.67%	3,298	31.67%	1,842	20.70%	8,760	28.49%
5	3,042	26.61%	2,815	27.03%	2,406	27.04%	8,263	26.88%
6	3,005	26.29%	2,400	23.05%	2,733	30.71%	8,138	26.47%
Total	11,431	100.00%	10,413	100.00%	8,899	100.00%	30,743	100.00%

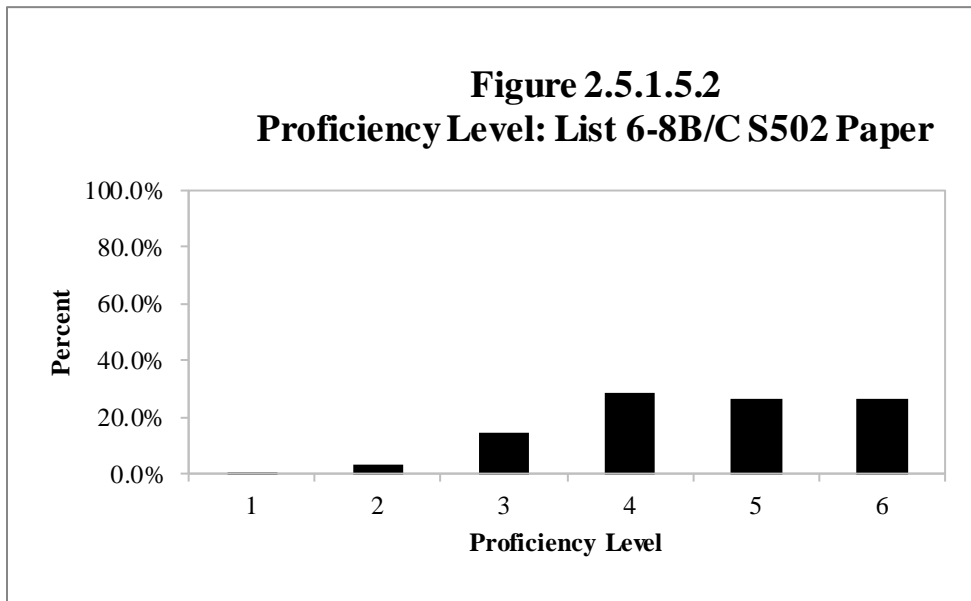
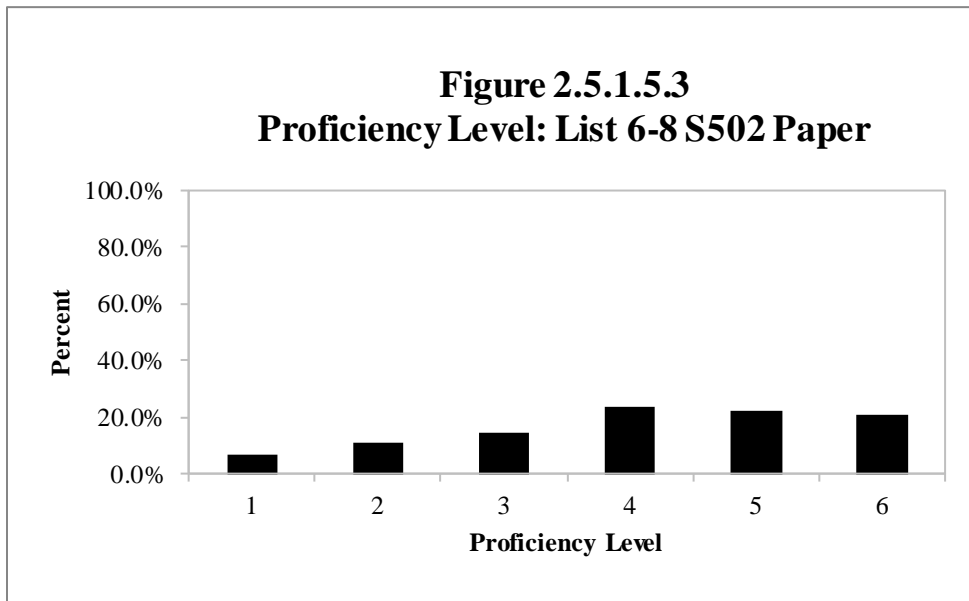


Table 2.5.1.5.3

Proficiency Level Distribution: List 6-8 S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	690	4.81%	974	7.42%	991	8.48%	2,655	6.78%
2	1,620	11.29%	1,293	9.85%	1,428	12.21%	4,341	11.09%
3	1,824	12.72%	1,978	15.07%	1,894	16.20%	5,696	14.55%
4	3,922	27.34%	3,422	26.08%	2,089	17.87%	9,433	24.09%
5	3,240	22.59%	2,984	22.74%	2,557	21.87%	8,781	22.43%
6	3,047	21.24%	2,471	18.83%	2,733	23.37%	8,251	21.07%
Total	14,343	100.00%	13,122	100.00%	11,692	100.00%	39,157	100.00%



2.5.1.6 Grades 9–12

Table 2.5.1.6.1

Proficiency Level Distribution: List 9-12 A S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	986	38.28%	1,261	53.59%	1,112	60.80%	814	70.72%	4,173	52.76%
2	1,170	45.42%	708	30.09%	381	20.83%	193	16.77%	2,452	31.00%
3	282	10.95%	251	10.67%	196	10.72%	109	9.47%	838	10.60%
4	76	2.95%	120	5.10%	119	6.51%	31	2.69%	346	4.37%
5	62	2.41%	13	0.55%	21	1.15%	4	0.35%	100	1.26%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	2,576	100.00%	2,353	100.00%	1,829	100.00%	1,151	100.00%	7,909	100.00%

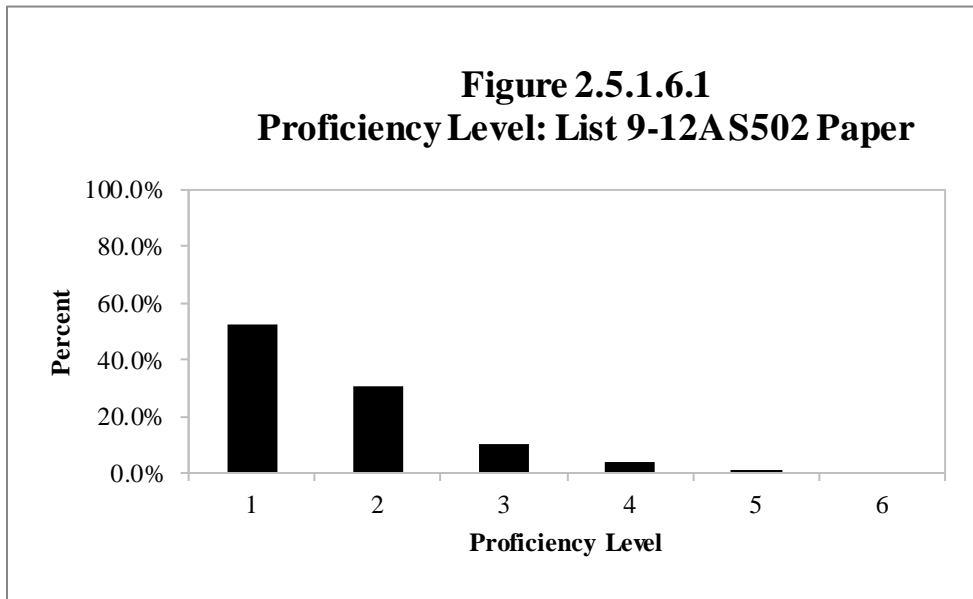


Table 2.5.1.6.2

Proficiency Level Distribution: List 9-12 B/C S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	64	0.80%	65	0.87%	123	1.83%	227	4.80%	479	1.78%
2	506	6.29%	909	12.15%	713	10.62%	757	16.01%	2,885	10.70%
3	1,588	19.75%	1,892	25.28%	2,267	33.78%	1,423	30.10%	7,170	26.59%
4	2,762	34.35%	2,409	32.19%	1,440	21.46%	1,476	31.22%	8,087	29.99%
5	1,689	21.01%	1,348	18.01%	1,231	18.34%	609	12.88%	4,877	18.09%
6	1,431	17.80%	860	11.49%	937	13.96%	236	4.99%	3,464	12.85%
Total	8,040	100.00%	7,483	100.00%	6,711	100.00%	4,728	100.00%	26,962	100.00%

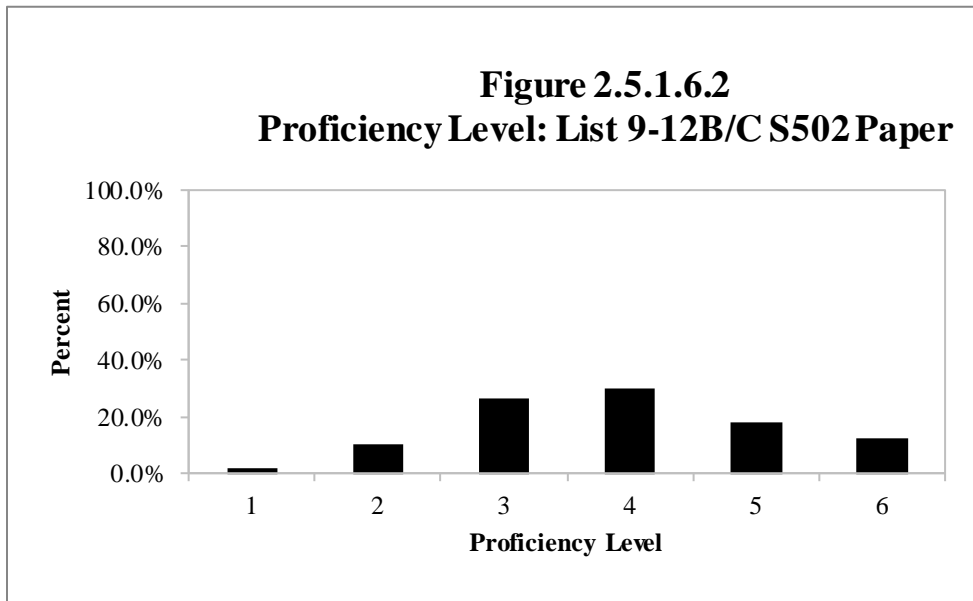
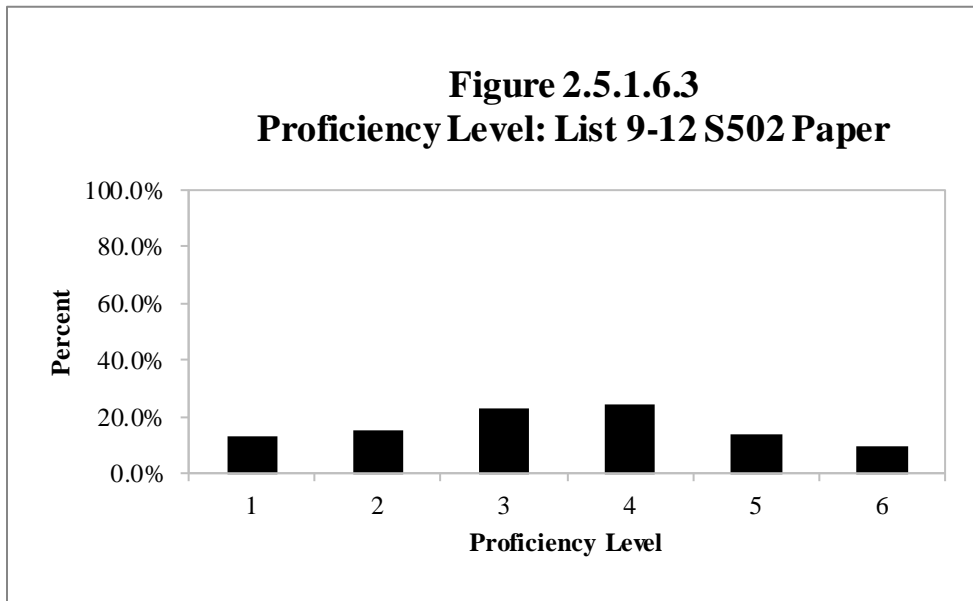


Table 2.5.1.6.3

Proficiency Level Distribution: List 9-12 S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,050	9.89%	1,326	13.48%	1,235	14.46%	1,041	17.71%	4,652	13.34%
2	1,676	15.79%	1,617	16.44%	1,094	12.81%	950	16.16%	5,337	15.30%
3	1,870	17.61%	2,143	21.79%	2,463	28.84%	1,532	26.06%	8,008	22.96%
4	2,838	26.73%	2,529	25.71%	1,559	18.26%	1,507	25.63%	8,433	24.18%
5	1,751	16.49%	1,361	13.84%	1,252	14.66%	613	10.43%	4,977	14.27%
6	1,431	13.48%	860	8.74%	937	10.97%	236	4.01%	3,464	9.93%
Total	10,616	100.00%	9,836	100.00%	8,540	100.00%	5,879	100.00%	34,871	100.00%



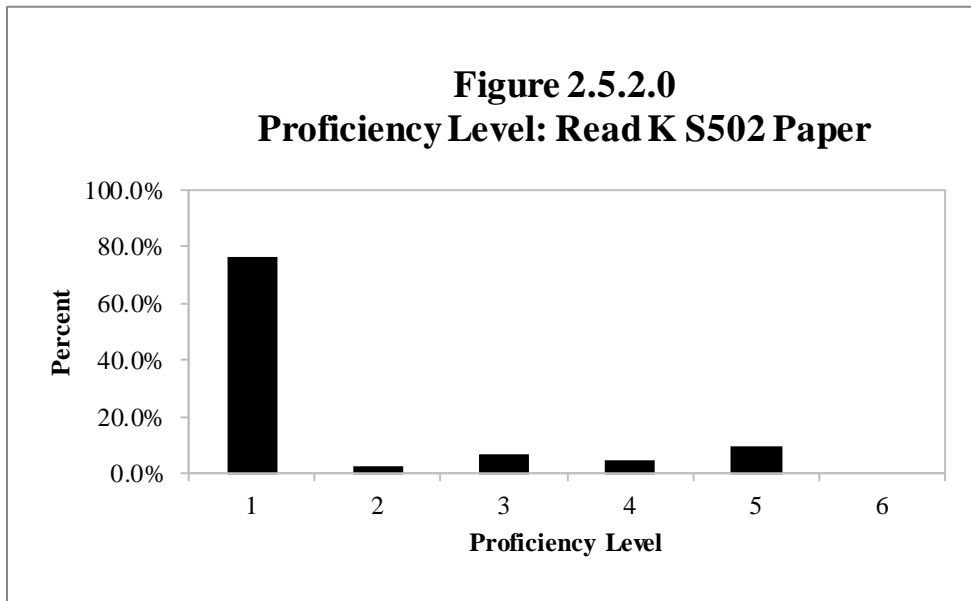
2.5.2 Reading

2.5.2.0 Kindergarten

Table 2.5.2.0

Proficiency Level Distribution: Read K S502 Paper

Level	Grade K		Total	
	Count	Percent	Count	Percent
1	125,146	76.67%	125,146	76.67%
2	4,037	2.47%	4,037	2.47%
3	10,851	6.65%	10,851	6.65%
4	7,770	4.76%	7,770	4.76%
5	15,414	9.44%	15,414	9.44%
6	0	0.00%	0	0.00%
Total	163,218	100.00%	163,218	100.00%



2.5.2.1 Grade 1

Table 2.5.2.1.1

Proficiency Level Distribution: Read 1 A S502 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	5,525	46.55%	5,525	46.55%
2	3,937	33.17%	3,937	33.17%
3	1,320	11.12%	1,320	11.12%
4	466	3.93%	466	3.93%
5	328	2.76%	328	2.76%
6	292	2.46%	292	2.46%
Total	11,868	100.00%	11,868	100.00%

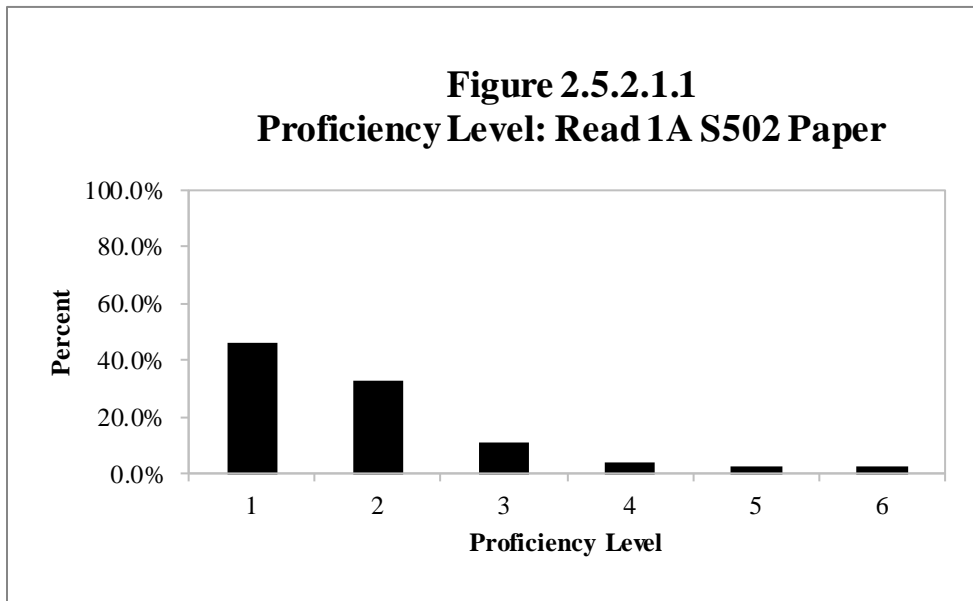


Table 2.5.2.1.2

Proficiency Level Distribution: Read 1 B/C S502 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	192	1.65%	192	1.65%
2	3,072	26.43%	3,072	26.43%
3	4,844	41.67%	4,844	41.67%
4	1,182	10.17%	1,182	10.17%
5	1,281	11.02%	1,281	11.02%
6	1,054	9.07%	1,054	9.07%
Total	11,625	100.00%	11,625	100.00%

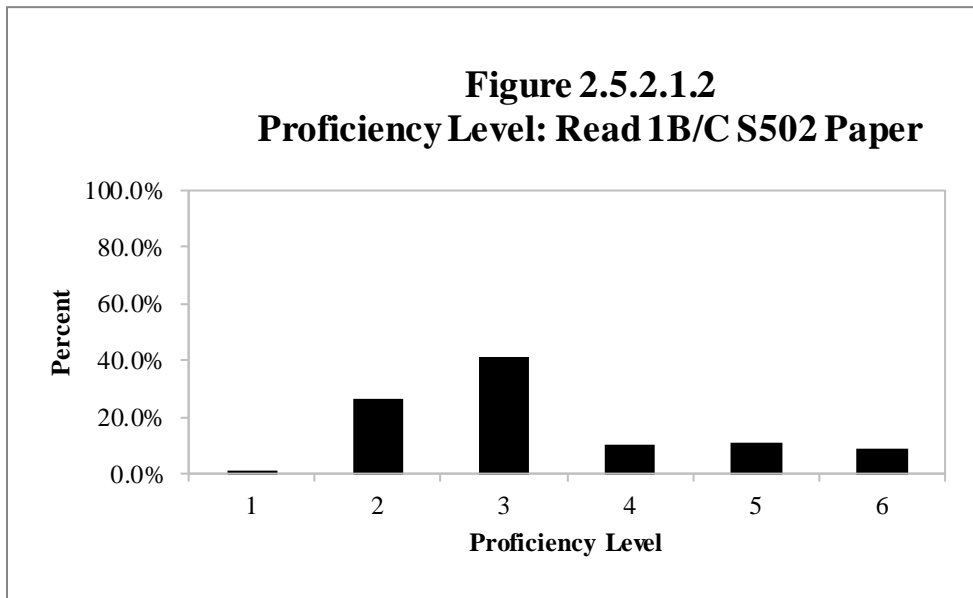
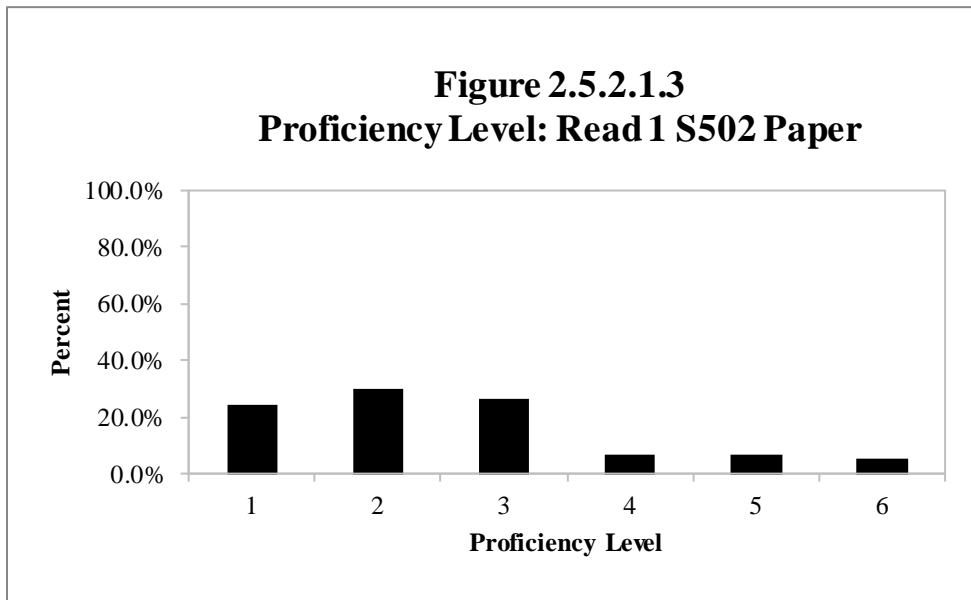


Table 2.5.2.1.3

Proficiency Level Distribution: Read 1 S502 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	5,717	24.33%	5,717	24.33%
2	7,009	29.83%	7,009	29.83%
3	6,164	26.24%	6,164	26.24%
4	1,648	7.01%	1,648	7.01%
5	1,609	6.85%	1,609	6.85%
6	1,346	5.73%	1,346	5.73%
Total	23,493	100.00%	23,493	100.00%

Figure 2.5.2.1.3
Proficiency Level: Read 1 S502 Paper



2.5.2.2 Grade 2

Table 2.5.2.2.1

Proficiency Level Distribution: Read 2 A S502 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	3,299	53.96%	3,299	53.96%
2	1,569	25.66%	1,569	25.66%
3	680	11.12%	680	11.12%
4	192	3.14%	192	3.14%
5	318	5.20%	318	5.20%
6	56	0.92%	56	0.92%
Total	6,114	100.00%	6,114	100.00%

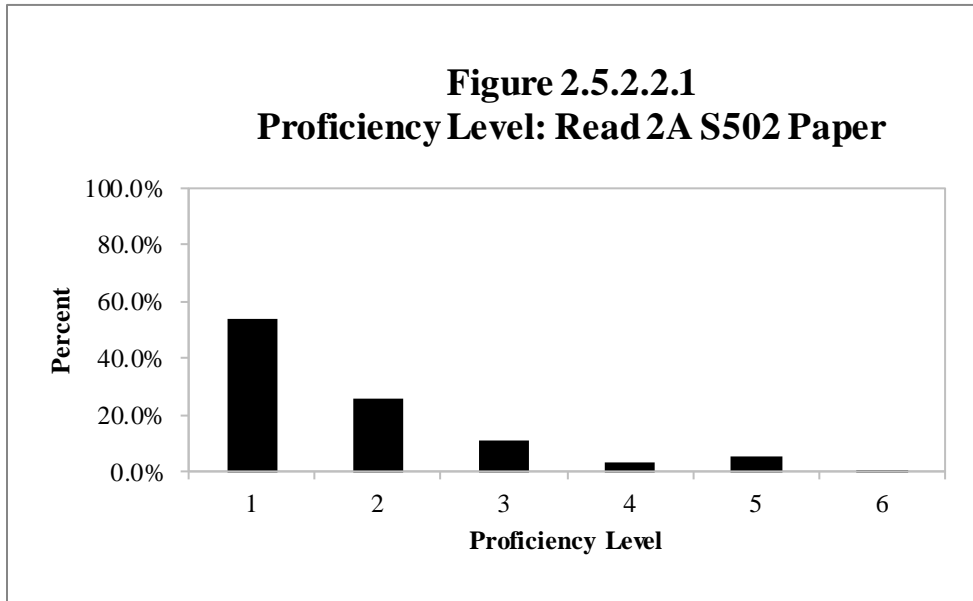


Table 2.5.2.2.2

Proficiency Level Distribution: Read 2 B/C S502 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	1,587	8.02%	1,587	8.02%
2	6,171	31.19%	6,171	31.19%
3	5,139	25.97%	5,139	25.97%
4	1,988	10.05%	1,988	10.05%
5	2,673	13.51%	2,673	13.51%
6	2,230	11.27%	2,230	11.27%
Total	19,788	100.00%	19,788	100.00%

Figure 2.5.2.2.2
Proficiency Level: Read 2B/C S502 Paper

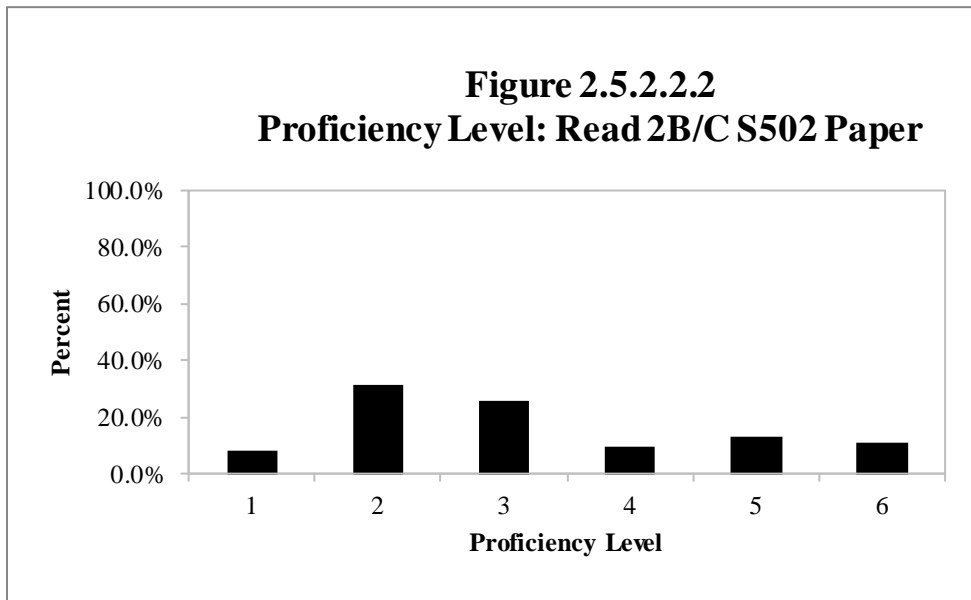
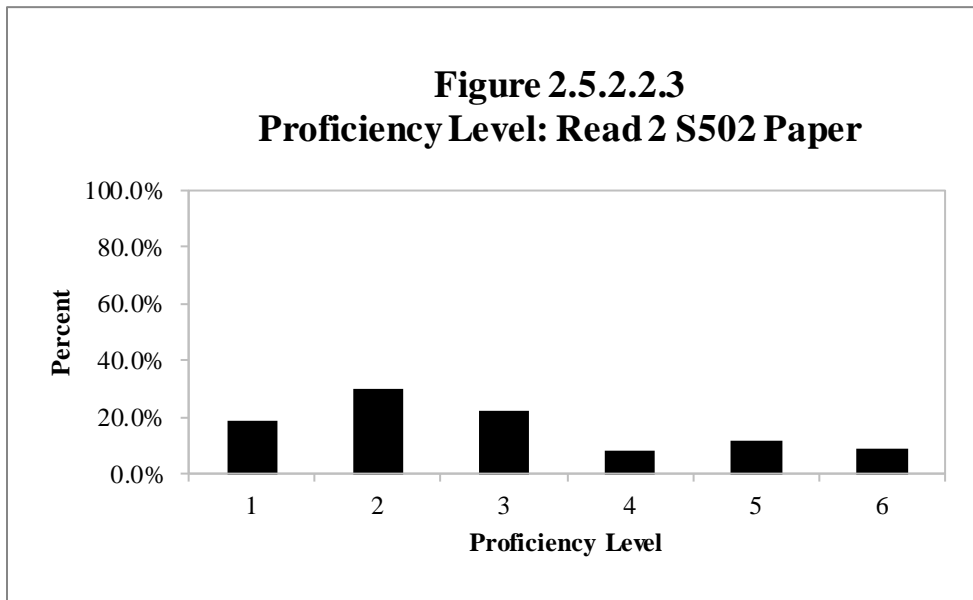


Table 2.5.2.2.3

Proficiency Level Distribution: Read 2 S502 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	4,886	18.86%	4,886	18.86%
2	7,740	29.88%	7,740	29.88%
3	5,819	22.47%	5,819	22.47%
4	2,180	8.42%	2,180	8.42%
5	2,991	11.55%	2,991	11.55%
6	2,286	8.83%	2,286	8.83%
Total	25,902	100.00%	25,902	100.00%



2.5.2.3 Grade 3

Table 2.5.2.3.1

Proficiency Level Distribution: Read 3 A S502 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	2,278	53.61%	2,278	53.61%
2	1,205	28.36%	1,205	28.36%
3	452	10.64%	452	10.64%
4	96	2.26%	96	2.26%
5	162	3.81%	162	3.81%
6	56	1.32%	56	1.32%
Total	4,249	100.00%	4,249	100.00%

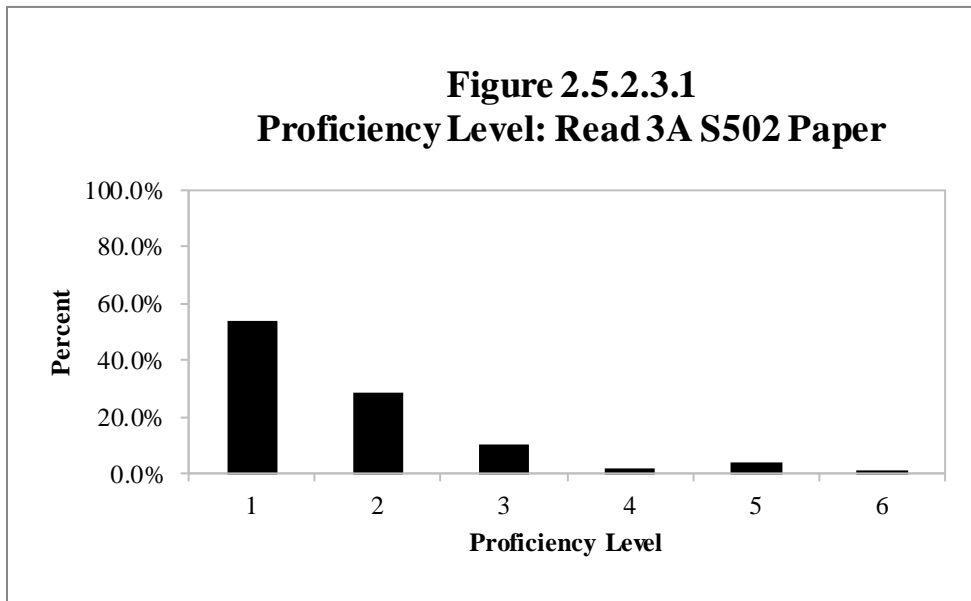


Table 2.5.2.3.2

Proficiency Level Distribution: Read 3 B/C S502 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	128	0.69%	128	0.69%
2	2,277	12.27%	2,277	12.27%
3	8,517	45.90%	8,517	45.90%
4	3,268	17.61%	3,268	17.61%
5	3,004	16.19%	3,004	16.19%
6	1,363	7.34%	1,363	7.34%
Total	18,557	100.00%	18,557	100.00%

Figure 2.5.2.3.2
Proficiency Level: Read 3B/C S502 Paper

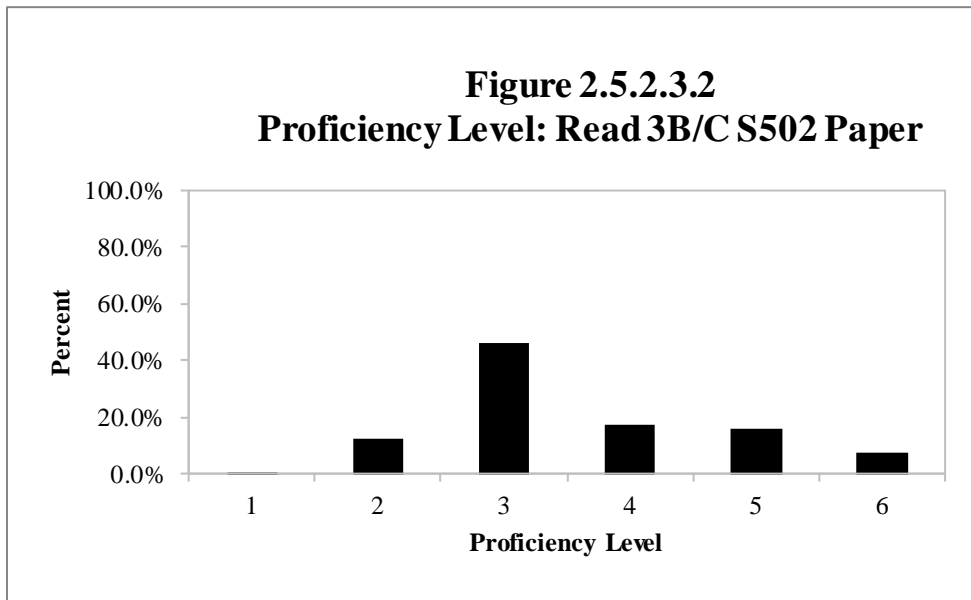
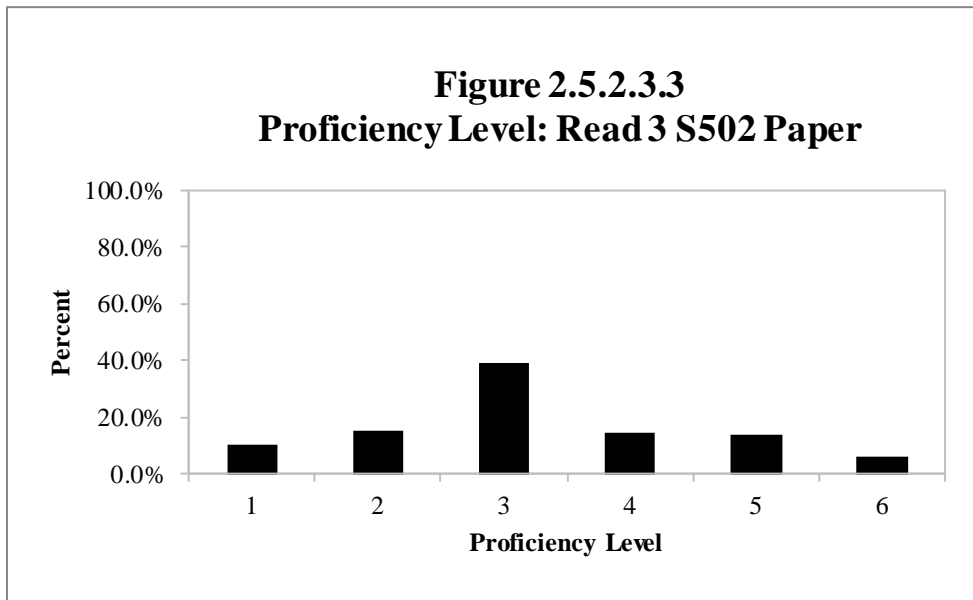


Table 2.5.2.3.3

Proficiency Level Distribution: Read 3 S502 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	2,406	10.55%	2,406	10.55%
2	3,482	15.27%	3,482	15.27%
3	8,969	39.33%	8,969	39.33%
4	3,364	14.75%	3,364	14.75%
5	3,166	13.88%	3,166	13.88%
6	1,419	6.22%	1,419	6.22%
Total	22,806	100.00%	22,806	100.00%



2.5.2.4 Grades 4–5

Table 2.5.2.4.1

Proficiency Level Distribution: Read 4-5 A S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	1,712	54.92%	1,655	59.02%	3,367	56.87%
2	841	26.98%	676	24.11%	1,517	25.62%
3	250	8.02%	226	8.06%	476	8.04%
4	109	3.50%	121	4.32%	230	3.88%
5	170	5.45%	126	4.49%	296	5.00%
6	35	1.12%	0	0.00%	35	0.59%
Total	3,117	100.00%	2,804	100.00%	5,921	100.00%

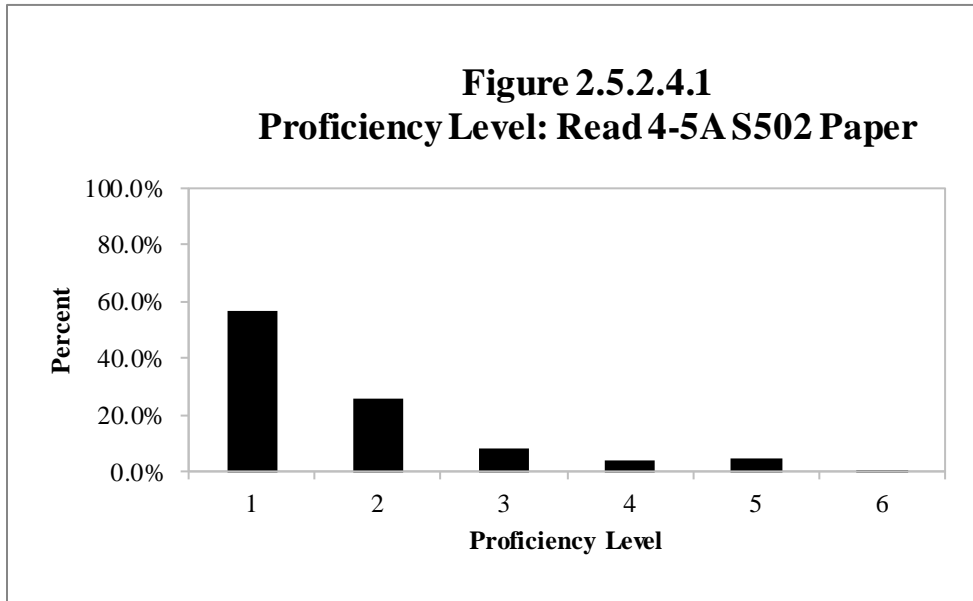


Table 2.5.2.4.2

Proficiency Level Distribution: Read 4-5 B/C S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	186	0.98%	215	1.46%	401	1.19%
2	3,316	17.46%	2,797	18.99%	6,113	18.12%
3	6,488	34.15%	4,731	32.12%	11,219	33.26%
4	3,708	19.52%	1,825	12.39%	5,533	16.41%
5	3,329	17.52%	3,145	21.35%	6,474	19.20%
6	1,969	10.37%	2,018	13.70%	3,987	11.82%
Total	18,996	100.00%	14,731	100.00%	33,727	100.00%

Figure 2.5.2.4.2
Proficiency Level: Read 4-5B/C S502 Paper

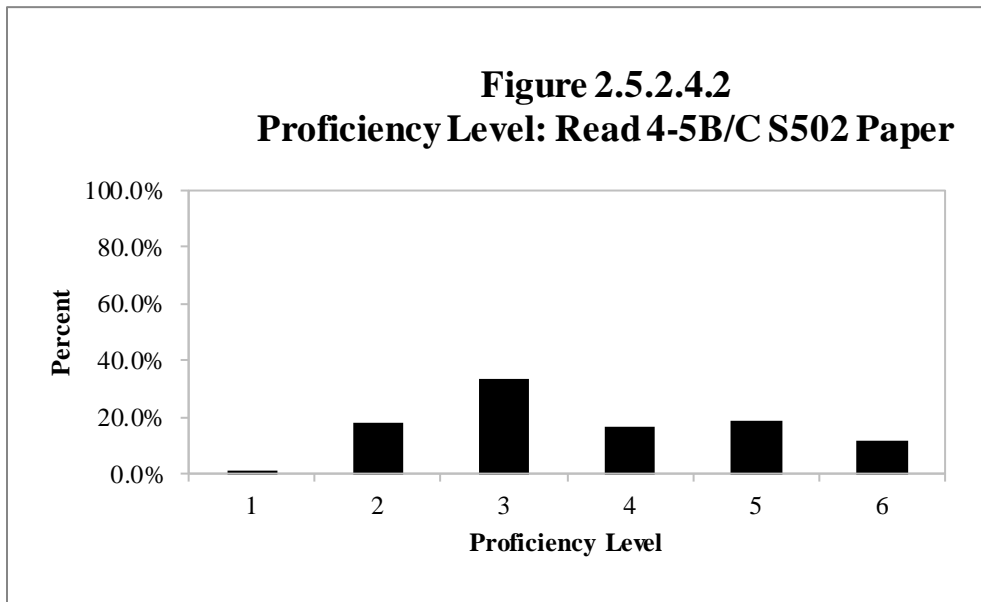
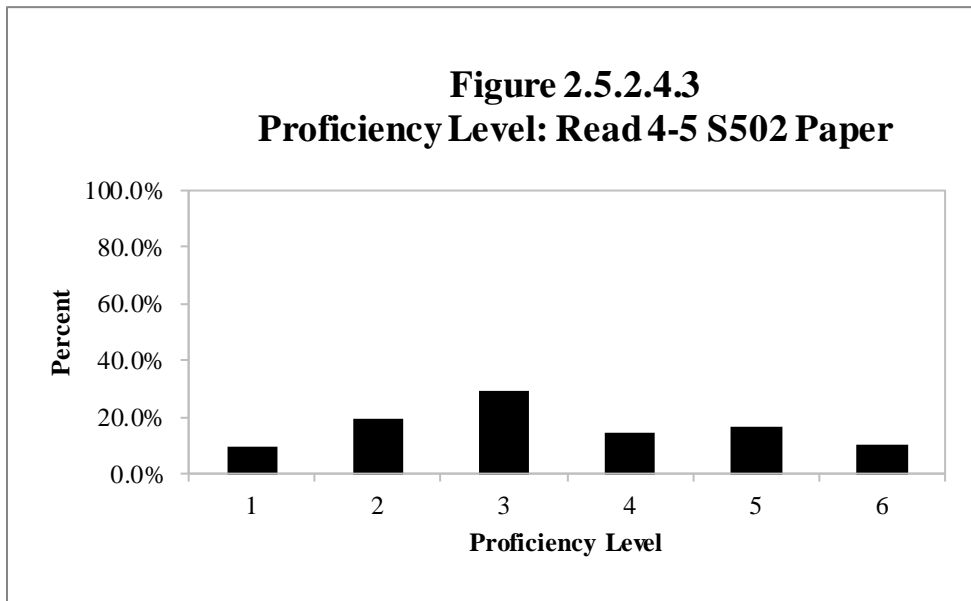


Table 2.5.2.4.3

Proficiency Level Distribution: Read 4-5 S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	1,898	8.58%	1,870	10.66%	3,768	9.50%
2	4,157	18.80%	3,473	19.81%	7,630	19.24%
3	6,738	30.47%	4,957	28.27%	11,695	29.50%
4	3,817	17.26%	1,946	11.10%	5,763	14.54%
5	3,499	15.82%	3,271	18.65%	6,770	17.08%
6	2,004	9.06%	2,018	11.51%	4,022	10.14%
Total	22,113	100.00%	17,535	100.00%	39,648	100.00%

Figure 2.5.2.4.3
Proficiency Level: Read 4-5 S502 Paper



2.5.2.5 Grades 6–8

Table 2.5.2.5.1

Proficiency Level Distribution: Read 6-8 A S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,244	45.40%	1,393	53.23%	1,562	57.07%	4,199	51.88%
2	1,095	39.96%	843	32.21%	844	30.84%	2,782	34.37%
3	228	8.32%	214	8.18%	190	6.94%	632	7.81%
4	51	1.86%	49	1.87%	58	2.12%	158	1.95%
5	78	2.85%	77	2.94%	45	1.64%	200	2.47%
6	44	1.61%	41	1.57%	38	1.39%	123	1.52%
Total	2,740	100.00%	2,617	100.00%	2,737	100.00%	8,094	100.00%

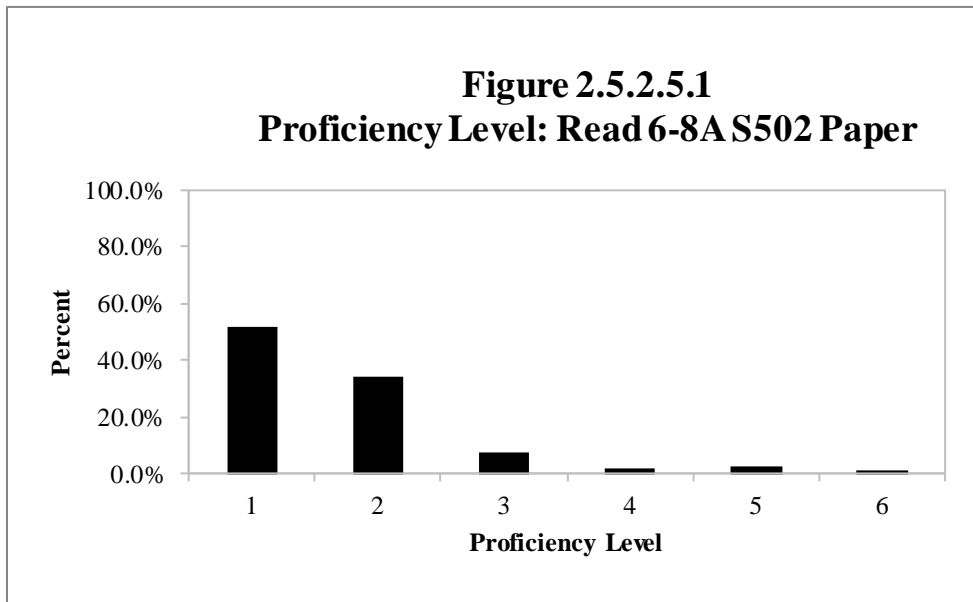


Table 2.5.2.5.2

Proficiency Level Distribution: Read 6-8 B/C S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	280	2.66%	408	4.19%	433	5.09%	1,121	3.90%
2	4,267	40.54%	3,318	34.11%	2,875	33.76%	10,460	36.36%
3	3,053	29.00%	3,184	32.73%	2,180	25.60%	8,417	29.26%
4	1,125	10.69%	1,066	10.96%	1,030	12.10%	3,221	11.20%
5	1,337	12.70%	1,093	11.24%	1,265	14.86%	3,695	12.84%
6	464	4.41%	658	6.76%	732	8.60%	1,854	6.44%
Total	10,526	100.00%	9,727	100.00%	8,515	100.00%	28,768	100.00%

Figure 2.5.2.5.2
Proficiency Level: Read 6-8B/C S502 Paper

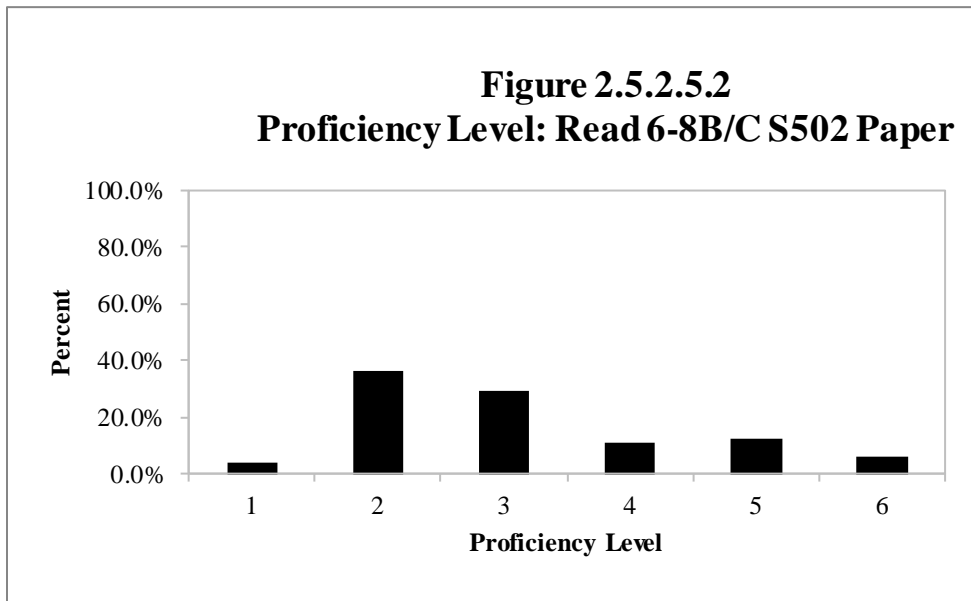
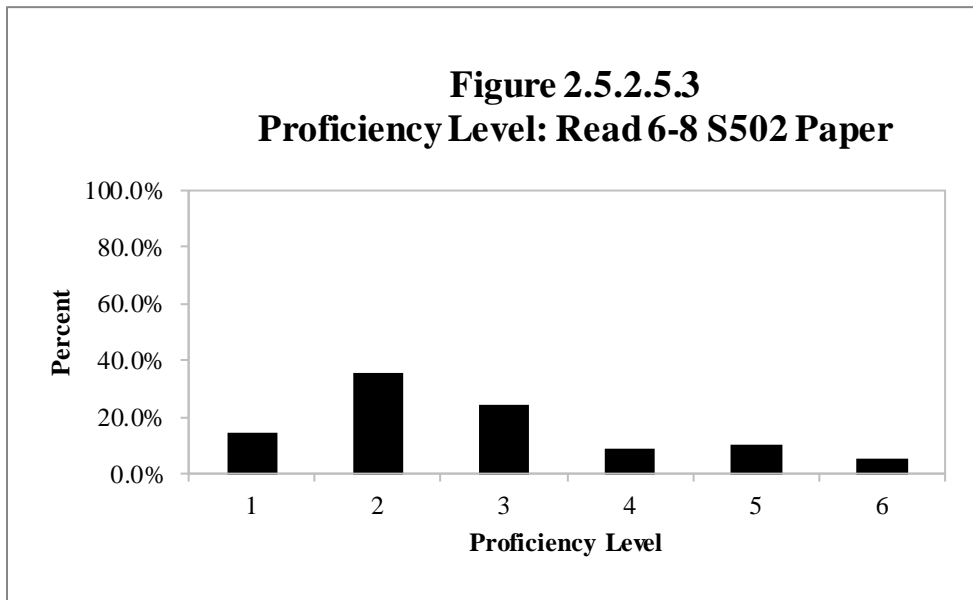


Table 2.5.2.5.3

Proficiency Level Distribution: Read 6-8 S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,524	11.49%	1,801	14.59%	1,995	17.73%	5,320	14.43%
2	5,362	40.42%	4,161	33.71%	3,719	33.05%	13,242	35.92%
3	3,281	24.73%	3,398	27.53%	2,370	21.06%	9,049	24.55%
4	1,176	8.86%	1,115	9.03%	1,088	9.67%	3,379	9.17%
5	1,415	10.67%	1,170	9.48%	1,310	11.64%	3,895	10.57%
6	508	3.83%	699	5.66%	770	6.84%	1,977	5.36%
Total	13,266	100.00%	12,344	100.00%	11,252	100.00%	36,862	100.00%



2.5.2.6 Grades 9–12

Table 2.5.2.6.1

Proficiency Level Distribution: Read 9-12 A S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,034	40.42%	913	38.88%	731	39.64%	506	44.04%	3,184	40.31%
2	910	35.57%	960	40.89%	737	39.97%	461	40.12%	3,068	38.84%
3	345	13.49%	289	12.31%	177	9.60%	93	8.09%	904	11.44%
4	83	3.24%	62	2.64%	63	3.42%	57	4.96%	265	3.35%
5	136	5.32%	67	2.85%	115	6.24%	21	1.83%	339	4.29%
6	50	1.95%	57	2.43%	21	1.14%	11	0.96%	139	1.76%
Total	2,558	100.00%	2,348	100.00%	1,844	100.00%	1,149	100.00%	7,899	100.00%

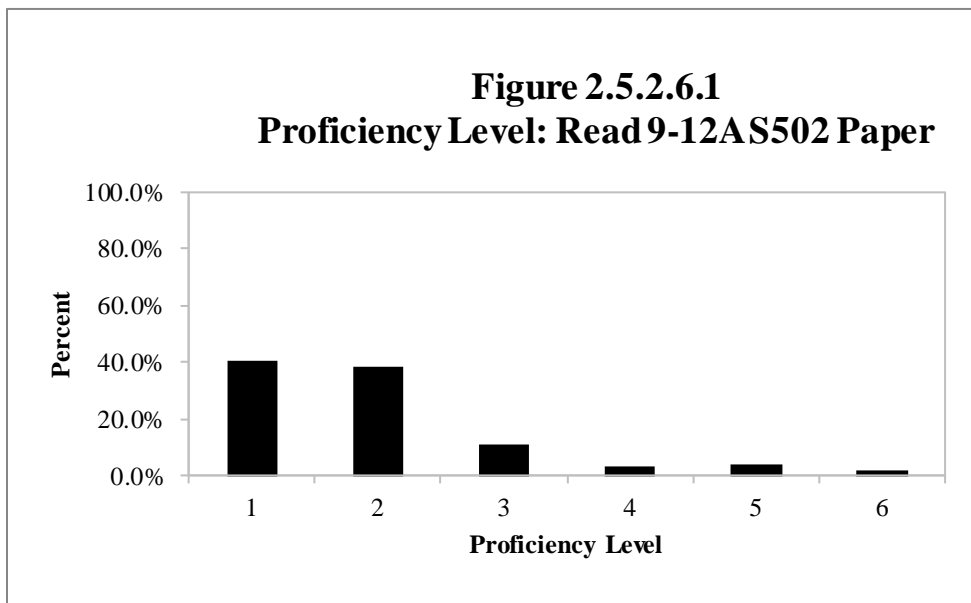


Table 2.5.2.6.2

Proficiency Level Distribution: Read 9-12 B/C S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	84	1.15%	82	1.19%	153	2.44%	257	5.77%	576	2.31%
2	1,835	25.21%	1,994	28.87%	1,843	29.40%	1,617	36.31%	7,289	29.26%
3	1,854	25.47%	2,076	30.06%	1,776	28.33%	1,278	28.70%	6,984	28.04%
4	1,401	19.24%	782	11.32%	707	11.28%	257	5.77%	3,147	12.63%
5	1,111	15.26%	993	14.38%	997	15.90%	640	14.37%	3,741	15.02%
6	995	13.67%	980	14.19%	793	12.65%	404	9.07%	3,172	12.73%
Total	7,280	100.00%	6,907	100.00%	6,269	100.00%	4,453	100.00%	24,909	100.00%

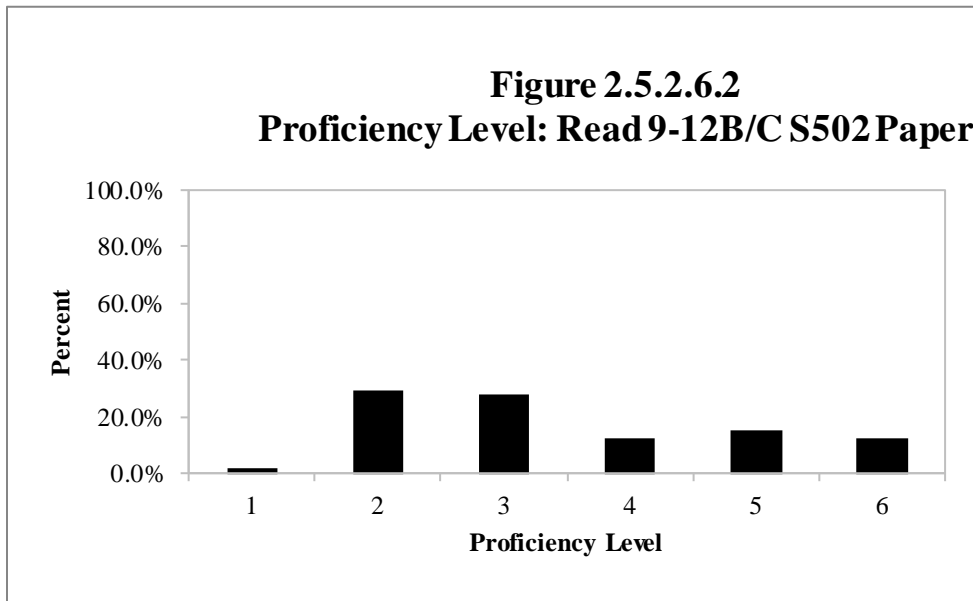
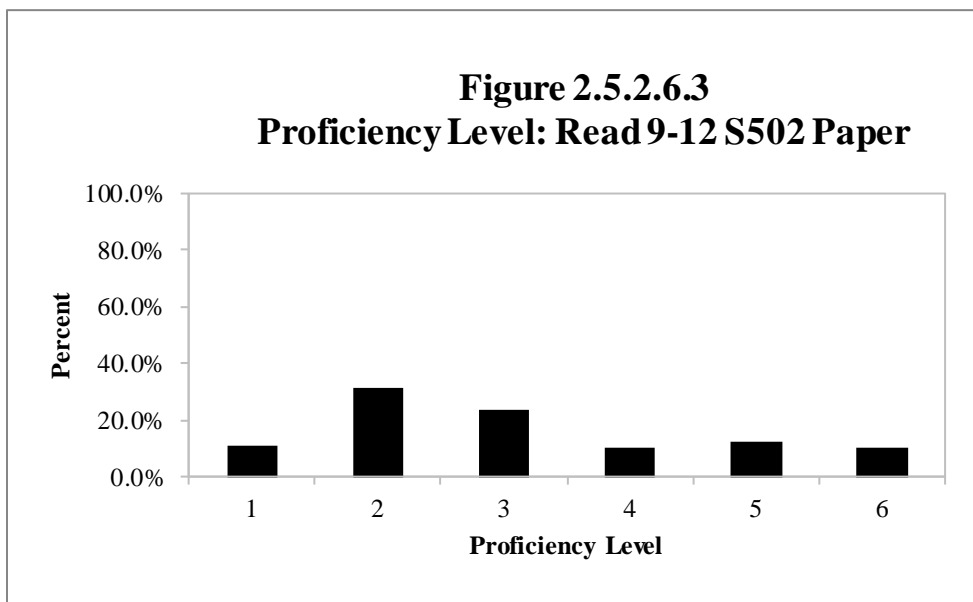


Table 2.5.2.6.3

Proficiency Level Distribution: Read 9-12 S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,118	11.36%	995	10.75%	884	10.90%	763	13.62%	3,760	11.46%
2	2,745	27.90%	2,954	31.92%	2,580	31.80%	2,078	37.09%	10,357	31.57%
3	2,199	22.35%	2,365	25.55%	1,953	24.07%	1,371	24.47%	7,888	24.04%
4	1,484	15.08%	844	9.12%	770	9.49%	314	5.61%	3,412	10.40%
5	1,247	12.68%	1,060	11.45%	1,112	13.71%	661	11.80%	4,080	12.44%
6	1,045	10.62%	1,037	11.20%	814	10.03%	415	7.41%	3,311	10.09%
Total	9,838	100.00%	9,255	100.00%	8,113	100.00%	5,602	100.00%	32,808	100.00%



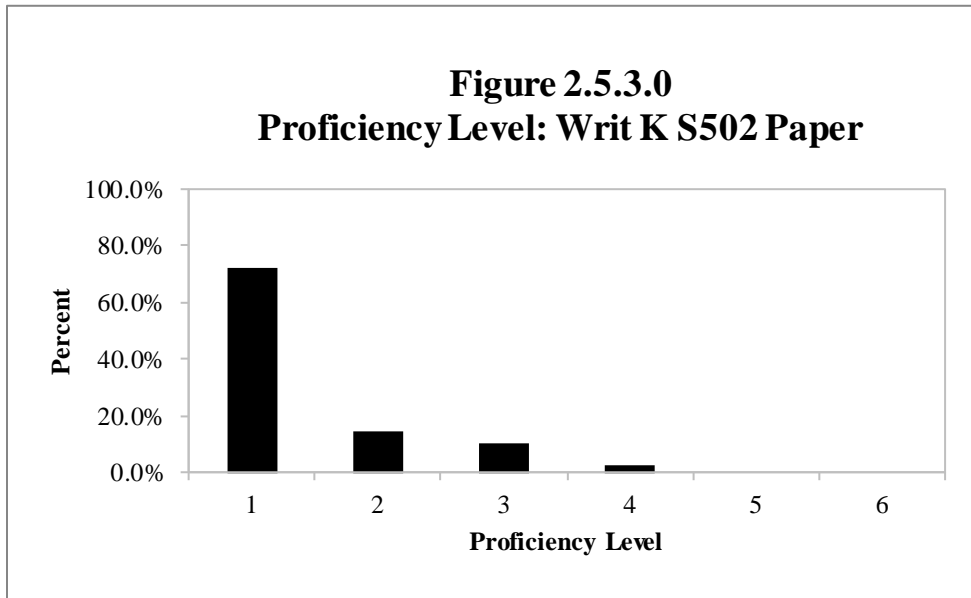
2.5.3 Writing

2.5.3.0 Kindergarten

Table 2.5.3.0

Proficiency Level Distribution: Writ K S502 Paper

Level	Grade K		Total	
	Count	Percent	Count	Percent
1	118,227	72.44%	118,227	72.44%
2	23,477	14.38%	23,477	14.38%
3	16,733	10.25%	16,733	10.25%
4	4,779	2.93%	4,779	2.93%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	163,216	100.00%	163,216	100.00%



2.5.3.1 Grade 1

Table 2.5.3.1.1

Proficiency Level Distribution: Writ 1 A S502 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	7,758	50.87%	7,758	50.87%
2	6,947	45.55%	6,947	45.55%
3	547	3.59%	547	3.59%
4	0	0.00%	0	0.00%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	15,252	100.00%	15,252	100.00%

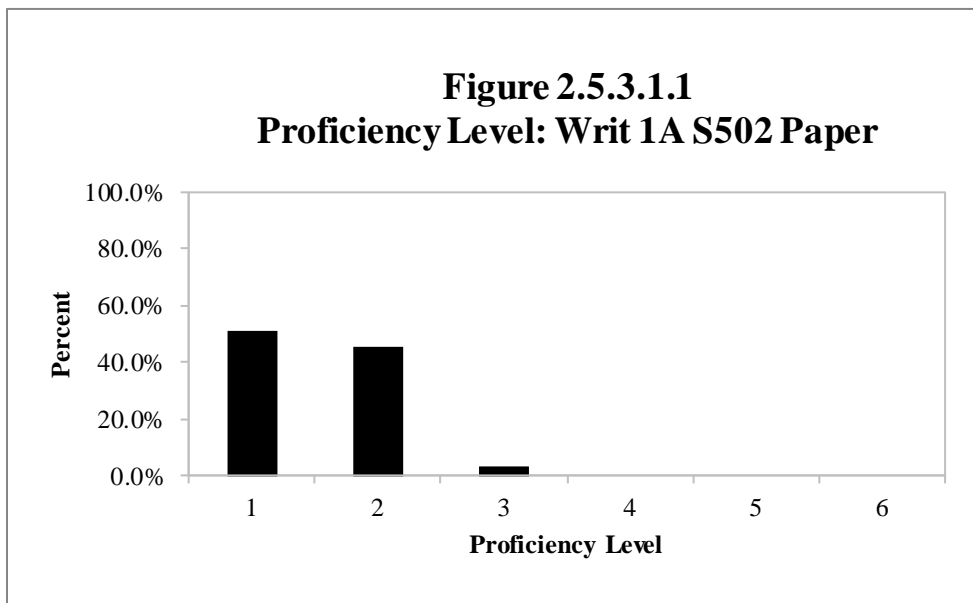


Table 2.5.3.1.2

Proficiency Level Distribution: Writ 1 B/C S502 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	3,459	22.61%	3,459	22.61%
2	6,326	41.35%	6,326	41.35%
3	5,330	34.84%	5,330	34.84%
4	181	1.18%	181	1.18%
5	3	0.02%	3	0.02%
6	1	0.01%	1	0.01%
Total	15,300	100.00%	15,300	100.00%

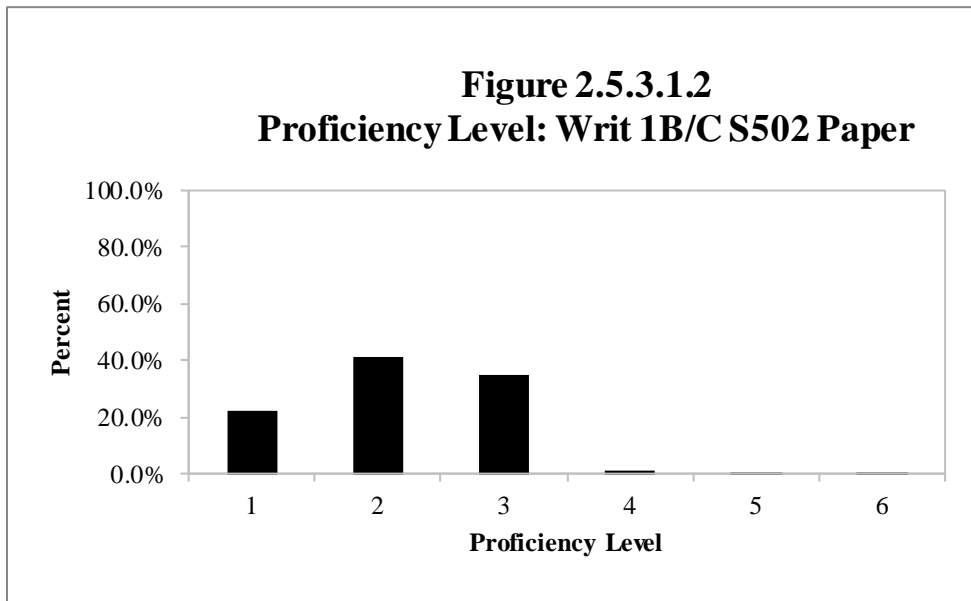
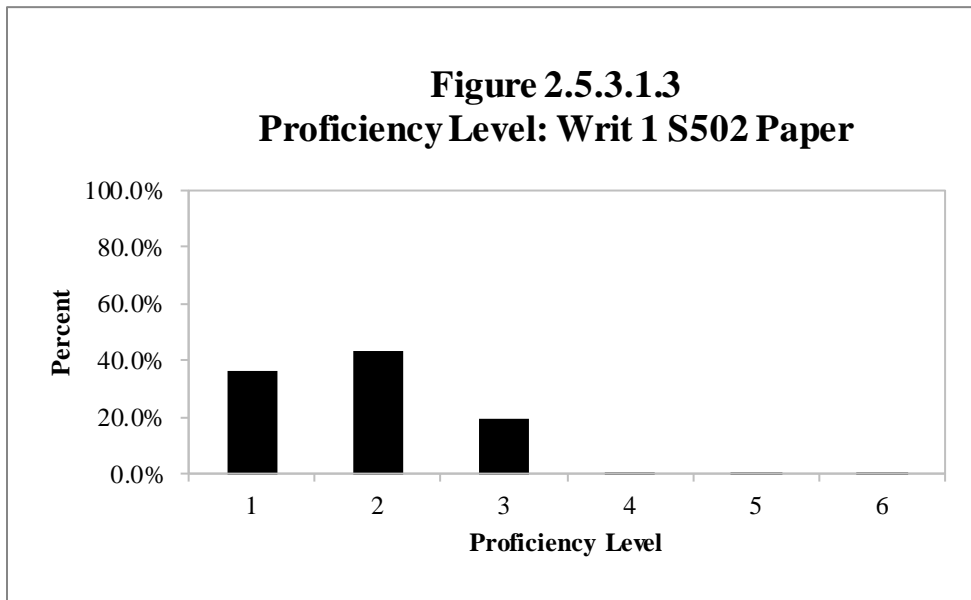


Table 2.5.3.1.3

Proficiency Level Distribution: Writ 1 S502 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	11,217	36.71%	11,217	36.71%
2	13,273	43.44%	13,273	43.44%
3	5,877	19.24%	5,877	19.24%
4	181	0.59%	181	0.59%
5	3	0.01%	3	0.01%
6	1	0.00%	1	0.00%
Total	30,552	100.00%	30,552	100.00%

Figure 2.5.3.1.3
Proficiency Level: Writ 1 S502 Paper



2.5.3.2 Grade 2

Table 2.5.3.2.1

Proficiency Level Distribution: Writ 2 A S502 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	3,080	42.59%	3,080	42.59%
2	2,364	32.69%	2,364	32.69%
3	1,784	24.67%	1,784	24.67%
4	3	0.04%	3	0.04%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	7,231	100.00%	7,231	100.00%

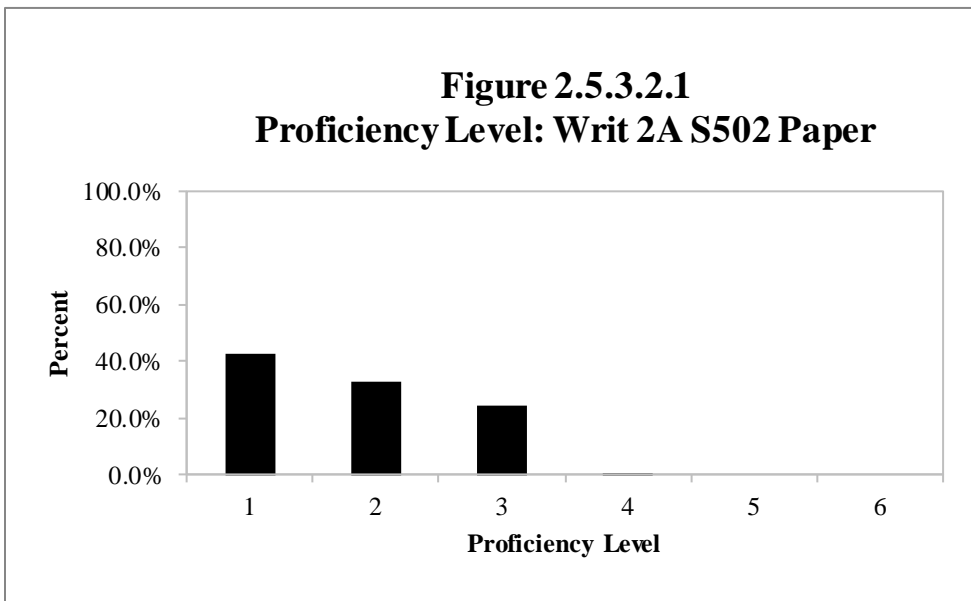


Table 2.5.3.2.2

Proficiency Level Distribution: Writ 2 B/C S502 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	2,381	9.98%	2,381	9.98%
2	6,107	25.59%	6,107	25.59%
3	13,842	58.01%	13,842	58.01%
4	1,517	6.36%	1,517	6.36%
5	14	0.06%	14	0.06%
6	0	0.00%	0	0.00%
Total	23,861	100.00%	23,861	100.00%

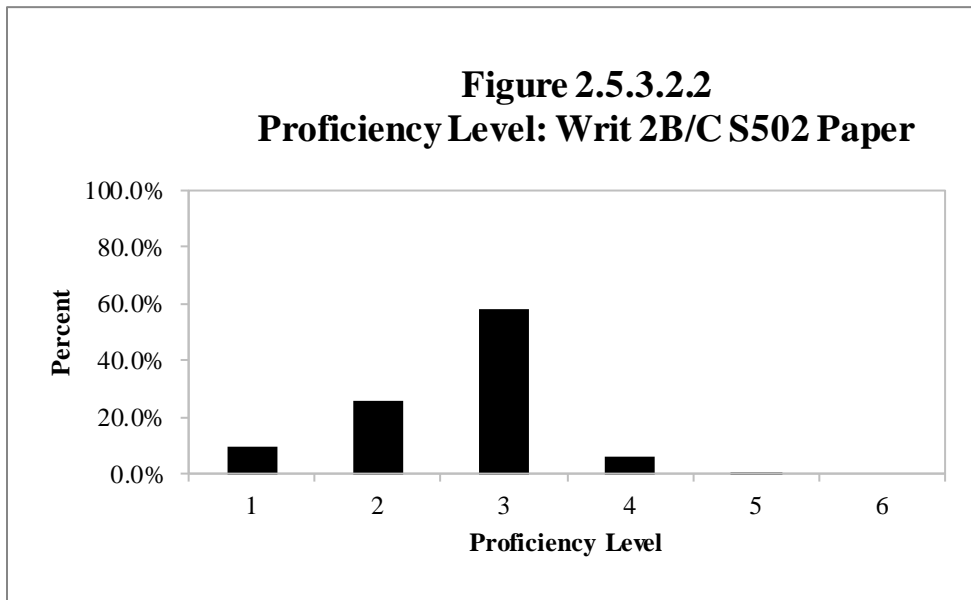
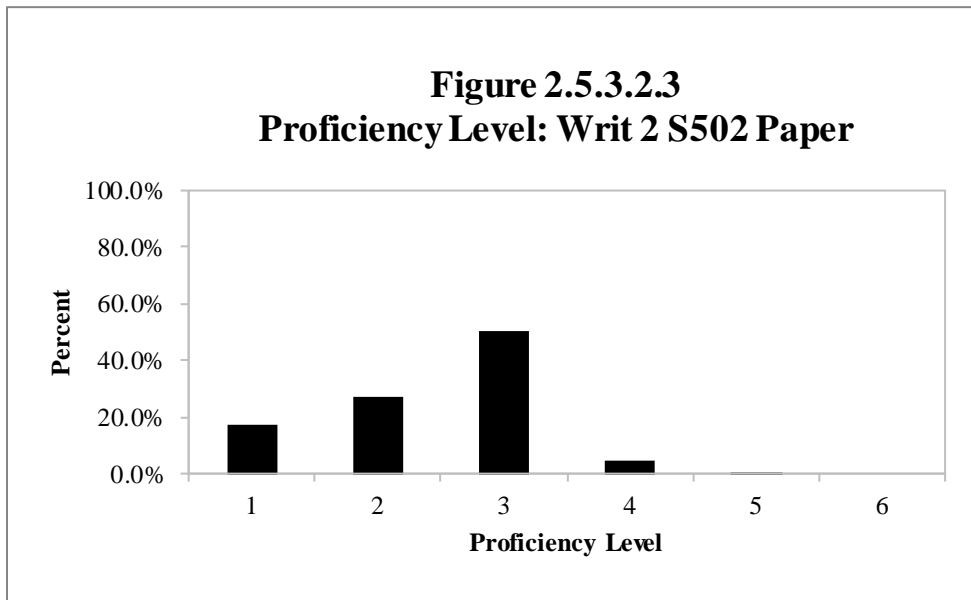


Table 2.5.3.2.3

Proficiency Level Distribution: Writ 2 S502 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	5,461	17.56%	5,461	17.56%
2	8,471	27.24%	8,471	27.24%
3	15,626	50.26%	15,626	50.26%
4	1,520	4.89%	1,520	4.89%
5	14	0.05%	14	0.05%
6	0	0.00%	0	0.00%
Total	31,092	100.00%	31,092	100.00%

Figure 2.5.3.2.3
Proficiency Level: Writ 2 S502 Paper



2.5.3.3 Grade 3

Table 2.5.3.3.1

Proficiency Level Distribution: Writ 3 A S502 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	1,951	39.22%	1,951	39.22%
2	1,935	38.90%	1,935	38.90%
3	1,080	21.71%	1,080	21.71%
4	8	0.16%	8	0.16%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	4,974	100.00%	4,974	100.00%

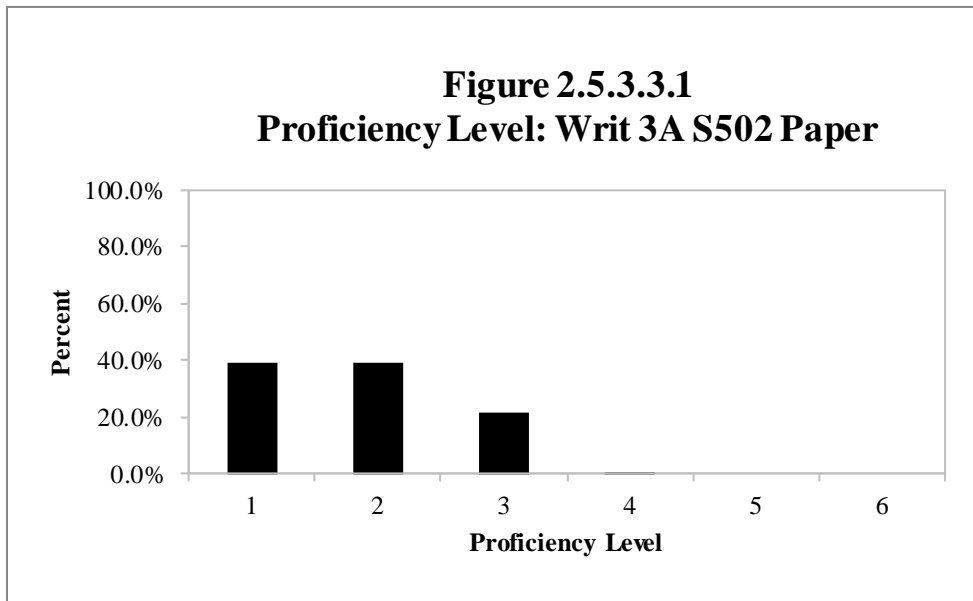


Table 2.5.3.3.2

Proficiency Level Distribution: Writ 3 B/C S502 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	1,237	5.46%	1,237	5.46%
2	3,390	14.95%	3,390	14.95%
3	15,557	68.61%	15,557	68.61%
4	2,474	10.91%	2,474	10.91%
5	16	0.07%	16	0.07%
6	1	0.00%	1	0.00%
Total	22,675	100.00%	22,675	100.00%

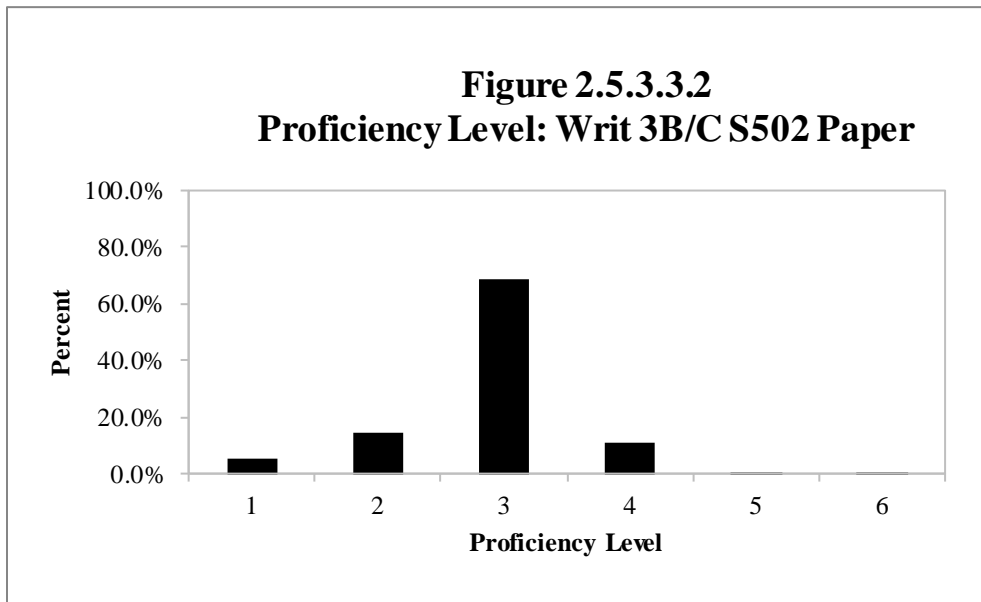
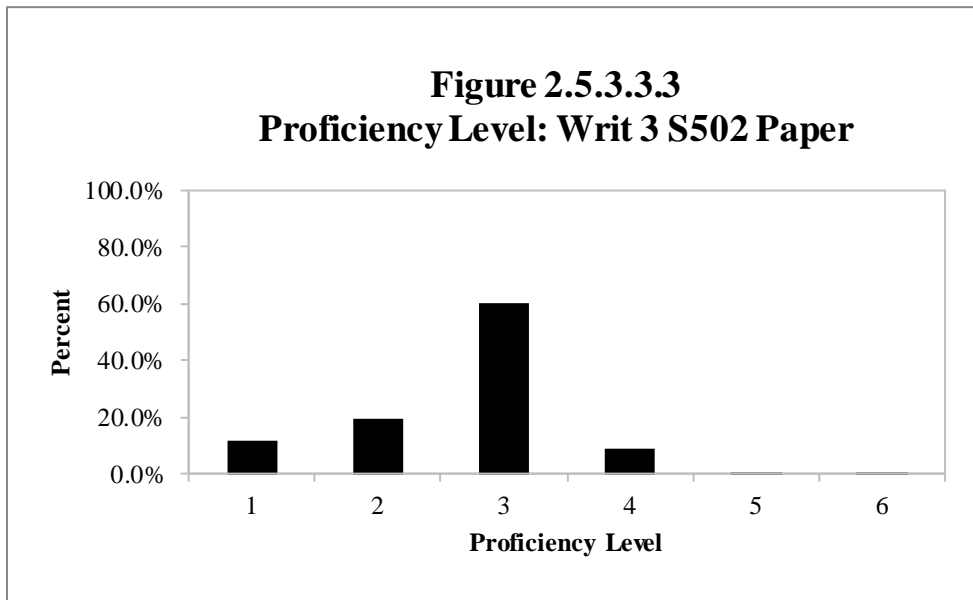


Table 2.5.3.3.3

Proficiency Level Distribution: Writ 3 S502 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	3,188	11.53%	3,188	11.53%
2	5,325	19.26%	5,325	19.26%
3	16,637	60.17%	16,637	60.17%
4	2,482	8.98%	2,482	8.98%
5	16	0.06%	16	0.06%
6	1	0.00%	1	0.00%
Total	27,649	100.00%	27,649	100.00%

Figure 2.5.3.3.3
Proficiency Level: Writ 3 S502 Paper



2.5.3.4 Grades 4–5

Table 2.5.3.4.1

Proficiency Level Distribution: Writ 4-5 A S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	1,173	33.22%	849	27.26%	2,022	30.43%
2	913	25.86%	814	26.14%	1,727	25.99%
3	1,425	40.36%	1,433	46.02%	2,858	43.01%
4	20	0.57%	18	0.58%	38	0.57%
5	0	0.00%	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%	0	0.00%
Total	3,531	100.00%	3,114	100.00%	6,645	100.00%

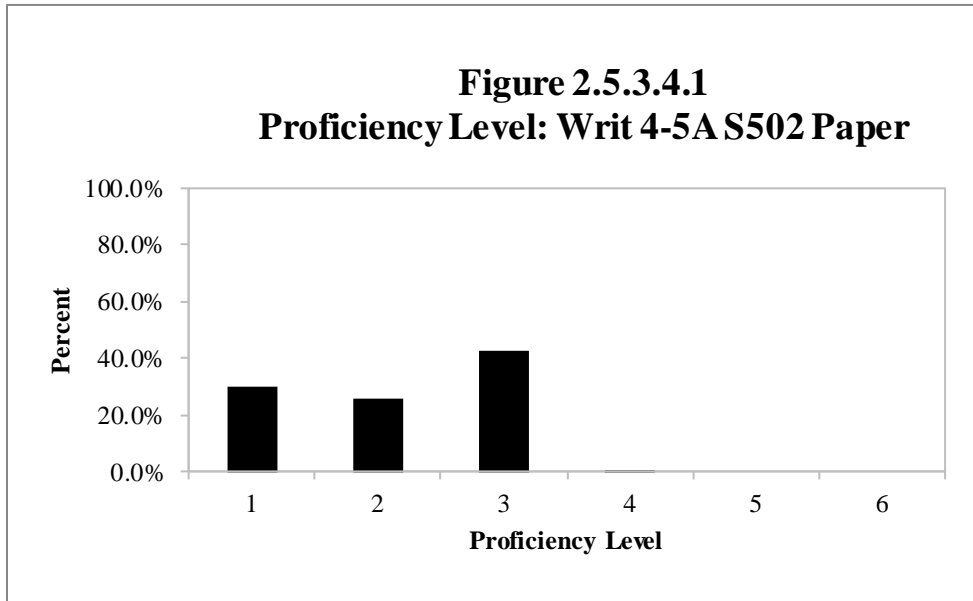


Table 2.5.3.4.2

Proficiency Level Distribution: Writ 4-5 B/C S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	437	1.98%	222	1.34%	659	1.71%
2	1,015	4.60%	557	3.37%	1,572	4.07%
3	14,755	66.89%	9,024	54.60%	23,779	61.63%
4	5,666	25.68%	6,526	39.49%	12,192	31.60%
5	182	0.83%	179	1.08%	361	0.94%
6	5	0.02%	18	0.11%	23	0.06%
Total	22,060	100.00%	16,526	100.00%	38,586	100.00%

Figure 2.5.3.4.2
Proficiency Level: Writ 4-5B/C S502 Paper

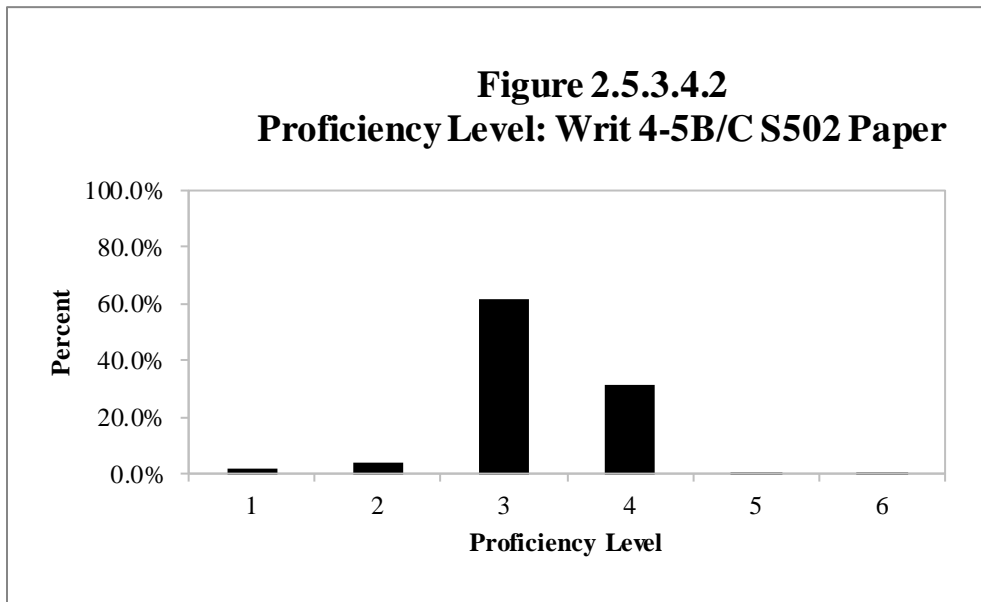
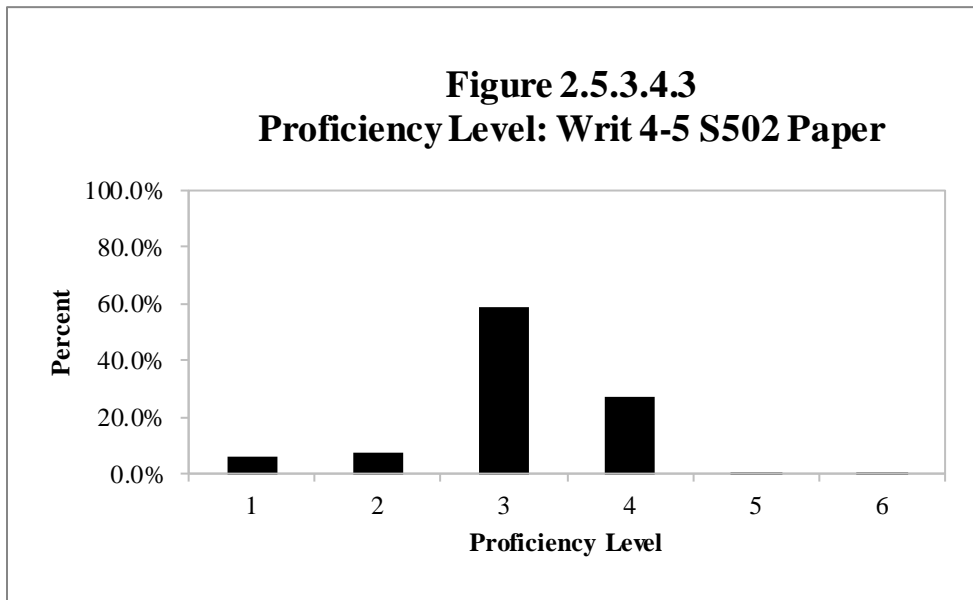


Table 2.5.3.4.3

Proficiency Level Distribution: Writ 4-5 S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	1,610	6.29%	1,071	5.45%	2,681	5.93%
2	1,928	7.53%	1,371	6.98%	3,299	7.29%
3	16,180	63.23%	10,457	53.24%	26,637	58.89%
4	5,686	22.22%	6,544	33.32%	12,230	27.04%
5	182	0.71%	179	0.91%	361	0.80%
6	5	0.02%	18	0.09%	23	0.05%
Total	25,591	100.00%	19,640	100.00%	45,231	100.00%



2.5.3.5 Grades 6–8

Table 2.5.3.5.1

Proficiency Level Distribution: Writ 6-8 A S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,149	37.34%	1,206	41.40%	1,461	49.16%	3,816	42.58%
2	1,113	36.17%	1,176	40.37%	903	30.38%	3,192	35.62%
3	807	26.23%	523	17.95%	604	20.32%	1,934	21.58%
4	8	0.26%	8	0.27%	4	0.13%	20	0.22%
5	0	0.00%	0	0.00%	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	3,077	100.00%	2,913	100.00%	2,972	100.00%	8,962	100.00%

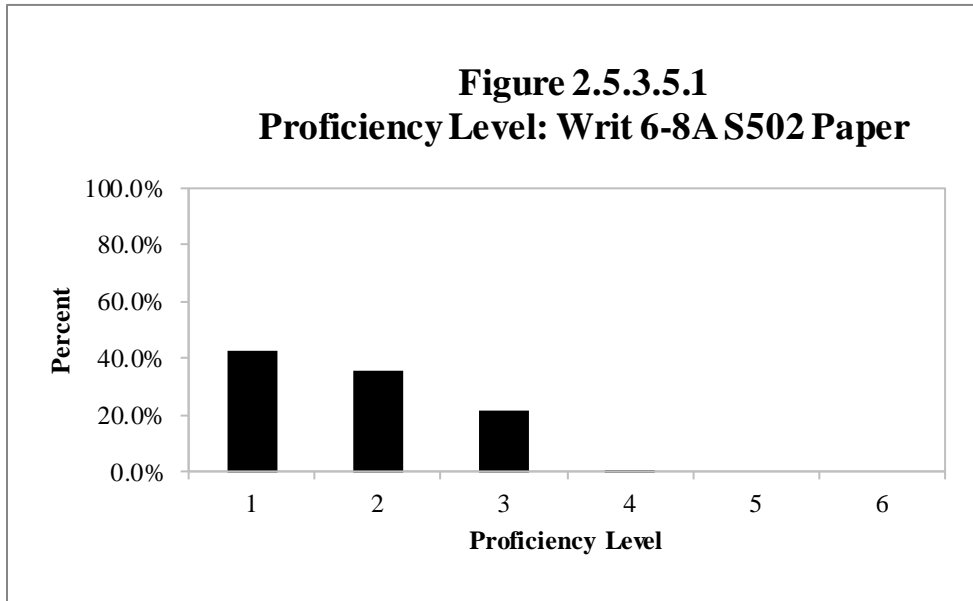


Table 2.5.3.5.2

Proficiency Level Distribution: Writ 6-8 B/C S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	331	2.78%	324	3.00%	379	4.10%	1,034	3.24%
2	1,048	8.80%	802	7.43%	412	4.46%	2,262	7.08%
3	7,354	61.77%	6,955	64.39%	6,033	65.33%	20,342	63.69%
4	3,155	26.50%	2,711	25.10%	2,401	26.00%	8,267	25.88%
5	17	0.14%	9	0.08%	10	0.11%	36	0.11%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	11,905	100.00%	10,801	100.00%	9,235	100.00%	31,941	100.00%

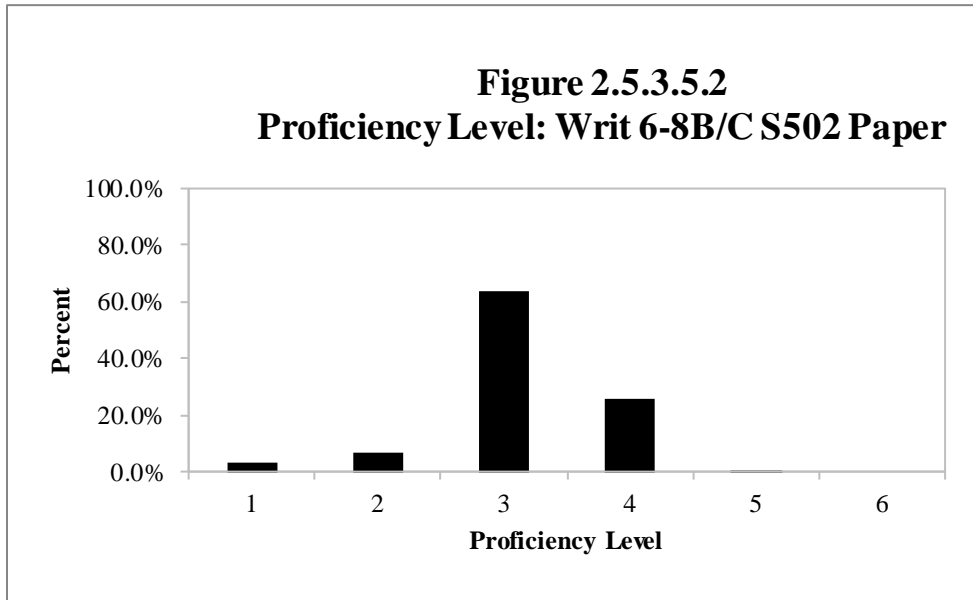
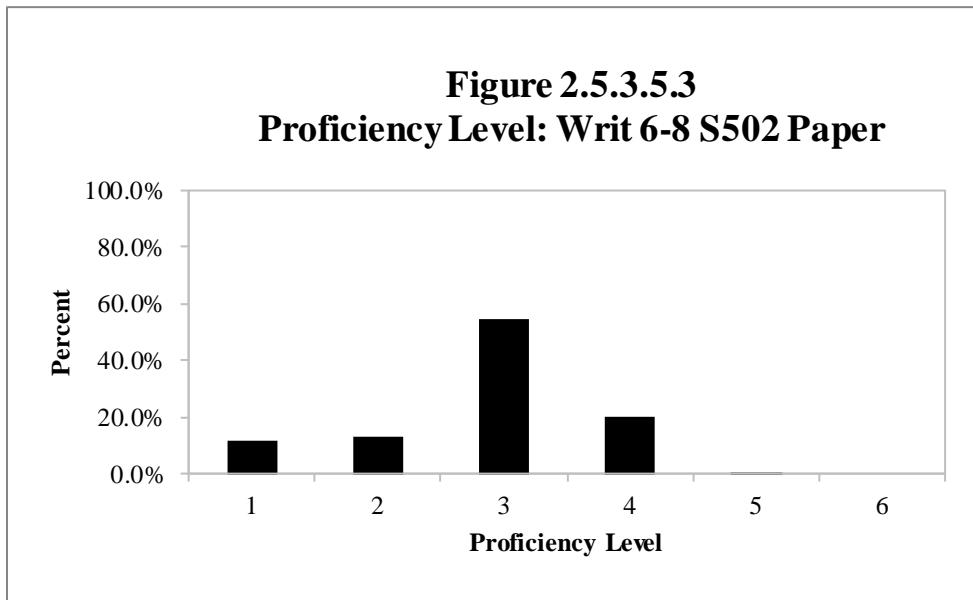


Table 2.5.3.5.3

Proficiency Level Distribution: Writ 6-8 S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,480	9.88%	1,530	11.16%	1,840	15.07%	4,850	11.86%
2	2,161	14.42%	1,978	14.42%	1,315	10.77%	5,454	13.33%
3	8,161	54.47%	7,478	54.53%	6,637	54.37%	22,276	54.46%
4	3,163	21.11%	2,719	19.83%	2,405	19.70%	8,287	20.26%
5	17	0.11%	9	0.07%	10	0.08%	36	0.09%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	14,982	100.00%	13,714	100.00%	12,207	100.00%	40,903	100.00%



2.5.3.6 Grades 9–12

Table 2.5.3.6.1

Proficiency Level Distribution: Writ 9-12 A S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	678	24.20%	708	27.88%	658	33.37%	529	42.63%	2,573	30.08%
2	866	30.91%	908	35.76%	638	32.35%	228	18.37%	2,640	30.86%
3	1,036	36.97%	811	31.94%	622	31.54%	454	36.58%	2,923	34.17%
4	220	7.85%	112	4.41%	54	2.74%	30	2.42%	416	4.86%
5	2	0.07%	0	0.00%	0	0.00%	0	0.00%	2	0.02%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	2,802	100.00%	2,539	100.00%	1,972	100.00%	1,241	100.00%	8,554	100.00%

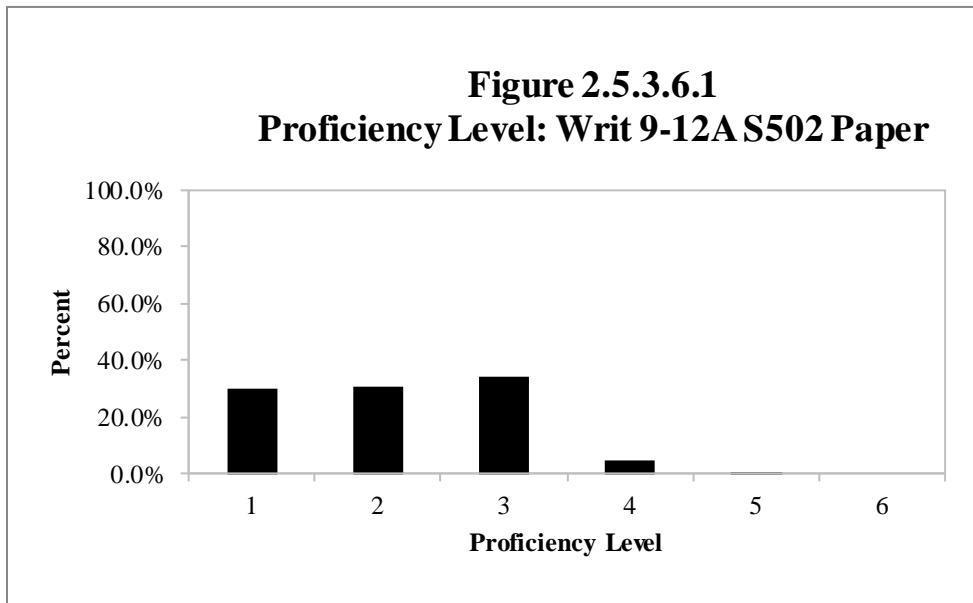


Table 2.5.3.6.2

Proficiency Level Distribution: Writ 9-12 B/C S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	224	2.66%	260	3.35%	356	5.12%	395	8.07%	1,235	4.41%
2	407	4.84%	525	6.77%	617	8.87%	430	8.78%	1,979	7.06%
3	4,372	52.01%	4,103	52.93%	3,576	51.39%	3,133	63.98%	15,184	54.20%
4	3,326	39.57%	2,820	36.38%	2,389	34.33%	932	19.03%	9,467	33.79%
5	77	0.92%	44	0.57%	21	0.30%	7	0.14%	149	0.53%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	8,406	100.00%	7,752	100.00%	6,959	100.00%	4,897	100.00%	28,014	100.00%

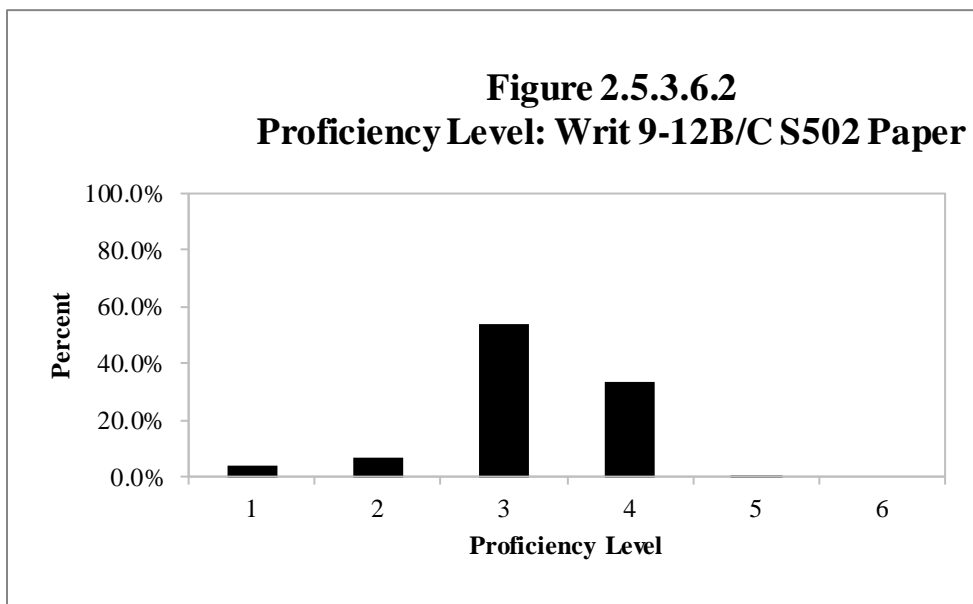
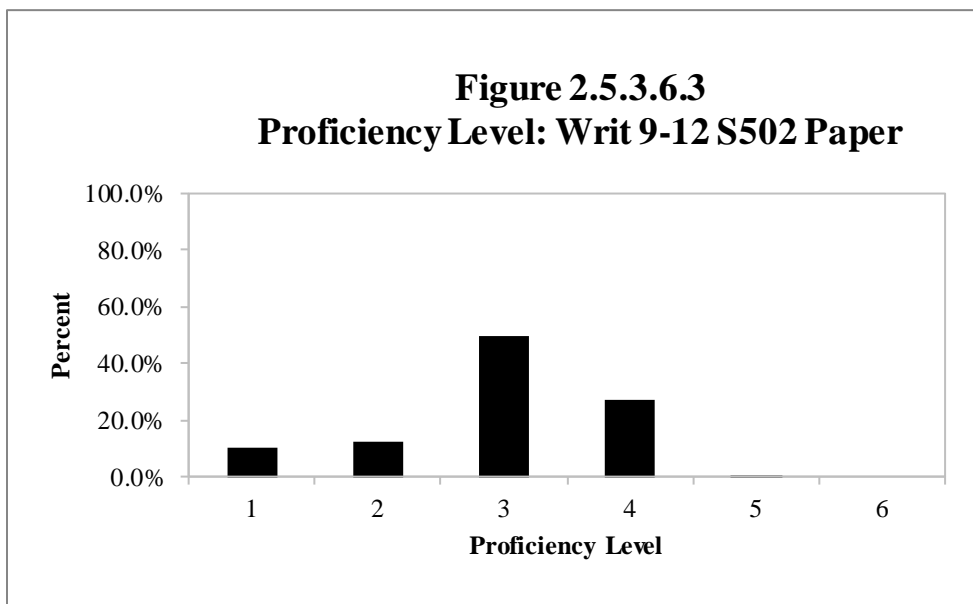


Table 2.5.3.6.3

Proficiency Level Distribution: Writ 9-12 S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	902	8.05%	968	9.41%	1,014	11.35%	924	15.05%	3,808	10.41%
2	1,273	11.36%	1,433	13.92%	1,255	14.05%	658	10.72%	4,619	12.63%
3	5,408	48.25%	4,914	47.75%	4,198	47.00%	3,587	58.44%	18,107	49.52%
4	3,546	31.64%	2,932	28.49%	2,443	27.35%	962	15.67%	9,883	27.03%
5	79	0.70%	44	0.43%	21	0.24%	7	0.11%	151	0.41%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	11,208	100.00%	10,291	100.00%	8,931	100.00%	6,138	100.00%	36,568	100.00%



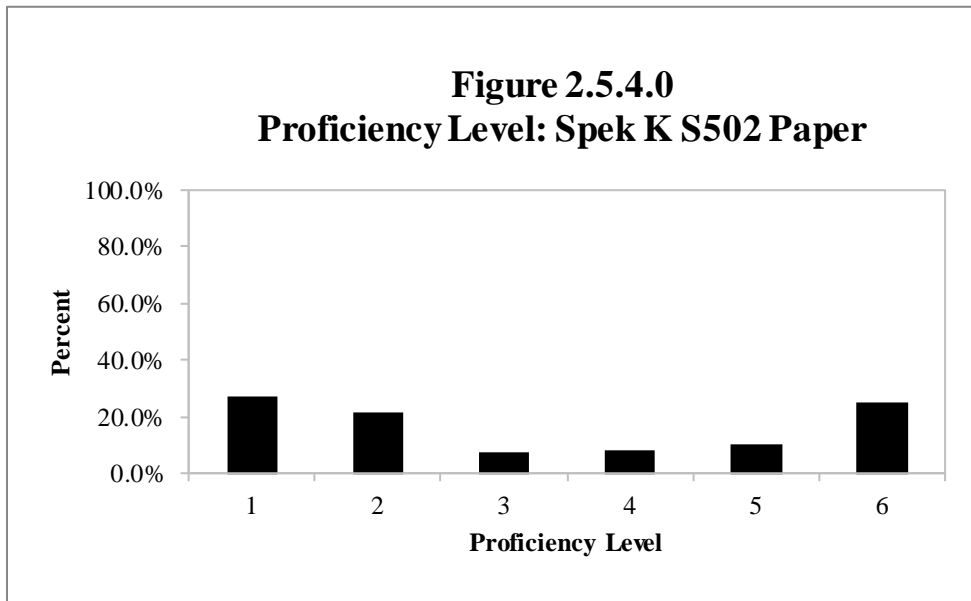
2.5.4 Speaking

2.5.4.0 Kindergarten

Table 2.5.4.0

Proficiency Level Distribution: Spek K S502 Paper

Level	Grade K		Total	
	Count	Percent	Count	Percent
1	43,995	26.96%	43,995	26.96%
2	35,336	21.65%	35,336	21.65%
3	12,007	7.36%	12,007	7.36%
4	13,276	8.14%	13,276	8.14%
5	17,132	10.50%	17,132	10.50%
6	41,446	25.40%	41,446	25.40%
Total	163,192	100.00%	163,192	100.00%



2.5.4.1. Grade 1

Table 2.5.4.1.1

Proficiency Level Distribution: Spek 1 A S502 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	4,330	28.86%	4,330	28.86%
2	5,364	35.76%	5,364	35.76%
3	3,148	20.99%	3,148	20.99%
4	1,656	11.04%	1,656	11.04%
5	503	3.35%	503	3.35%
6	0	0.00%	0	0.00%
Total	15,001	100.00%	15,001	100.00%

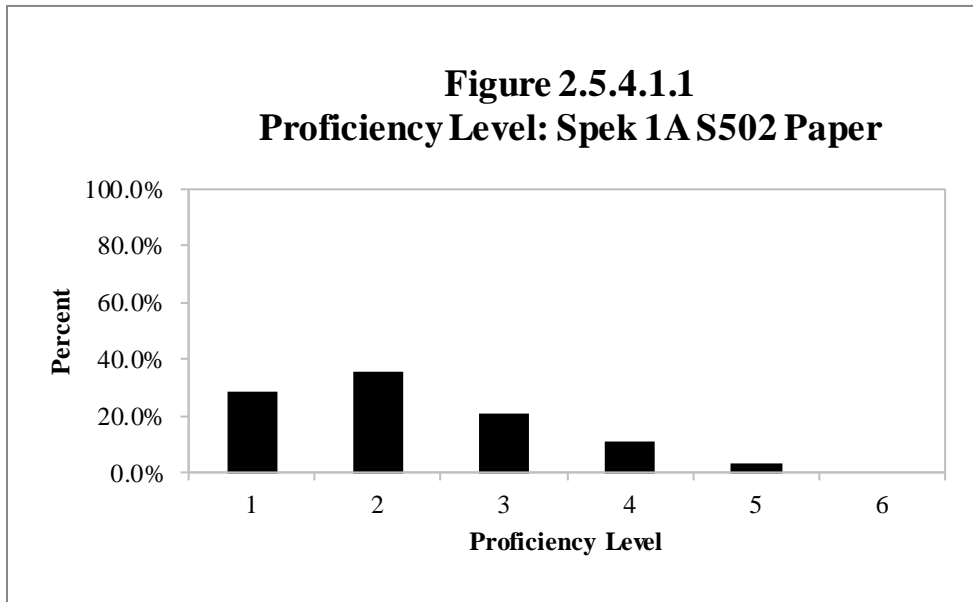


Table 2.5.4.1.2

Proficiency Level Distribution: Spek 1 B/C S502 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	448	2.97%	448	2.97%
2	3,435	22.75%	3,435	22.75%
3	5,180	34.31%	5,180	34.31%
4	4,473	29.62%	4,473	29.62%
5	1,199	7.94%	1,199	7.94%
6	364	2.41%	364	2.41%
Total	15,099	100.00%	15,099	100.00%

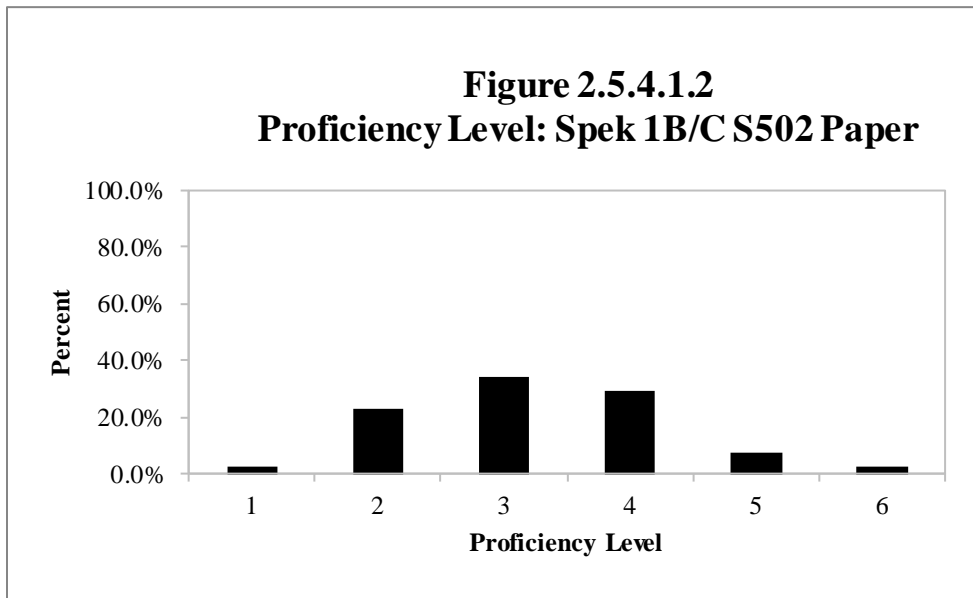
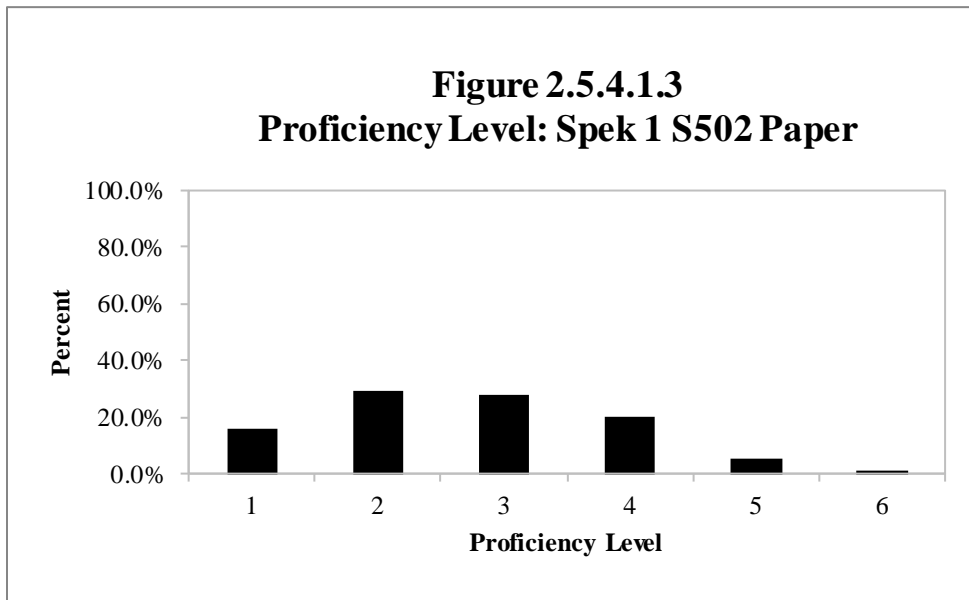


Table 2.5.4.1.3

Proficiency Level Distribution: Spek 1 S502 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	4,778	15.87%	4,778	15.87%
2	8,799	29.23%	8,799	29.23%
3	8,328	27.67%	8,328	27.67%
4	6,129	20.36%	6,129	20.36%
5	1,702	5.65%	1,702	5.65%
6	364	1.21%	364	1.21%
Total	30,100	100.00%	30,100	100.00%

Figure 2.5.4.1.3
Proficiency Level: Spek 1 S502 Paper



2.5.4.2 Grade 2

Table 2.5.4.2.1

Proficiency Level Distribution: Spek 2 A S502 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	2,377	33.49%	2,377	33.49%
2	1,837	25.88%	1,837	25.88%
3	2,230	31.42%	2,230	31.42%
4	496	6.99%	496	6.99%
5	158	2.23%	158	2.23%
6	0	0.00%	0	0.00%
Total	7,098	100.00%	7,098	100.00%

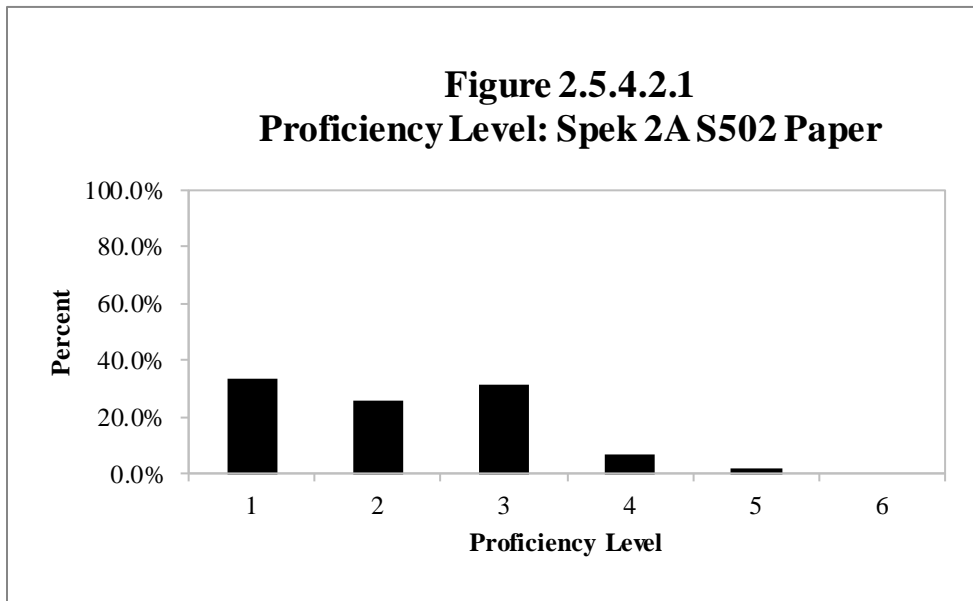


Table 2.5.4.2.2

Proficiency Level Distribution: Spek 2 B/C S502 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	1,423	6.03%	1,423	6.03%
2	5,077	21.52%	5,077	21.52%
3	9,719	41.20%	9,719	41.20%
4	5,191	22.01%	5,191	22.01%
5	1,494	6.33%	1,494	6.33%
6	684	2.90%	684	2.90%
Total	23,588	100.00%	23,588	100.00%

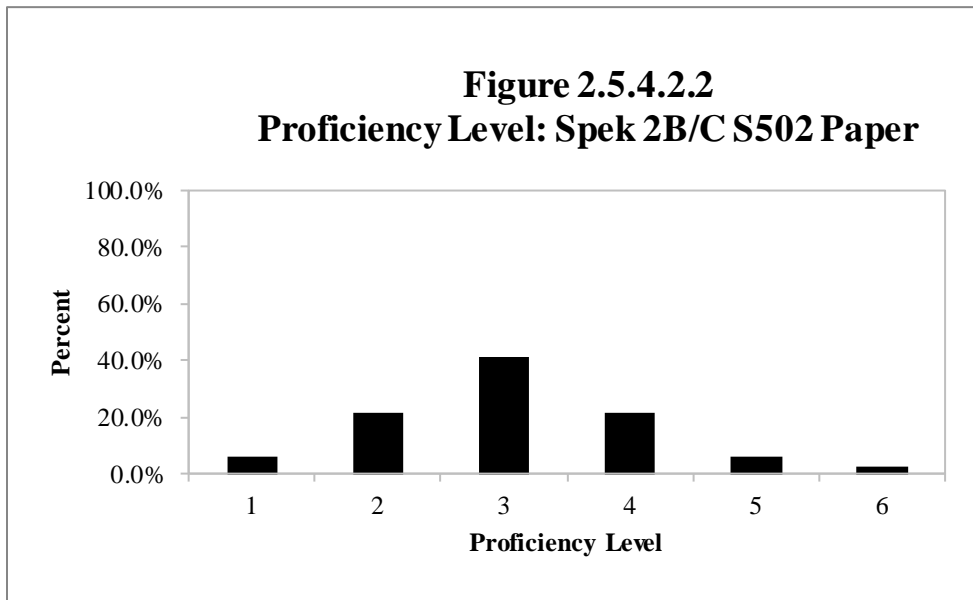
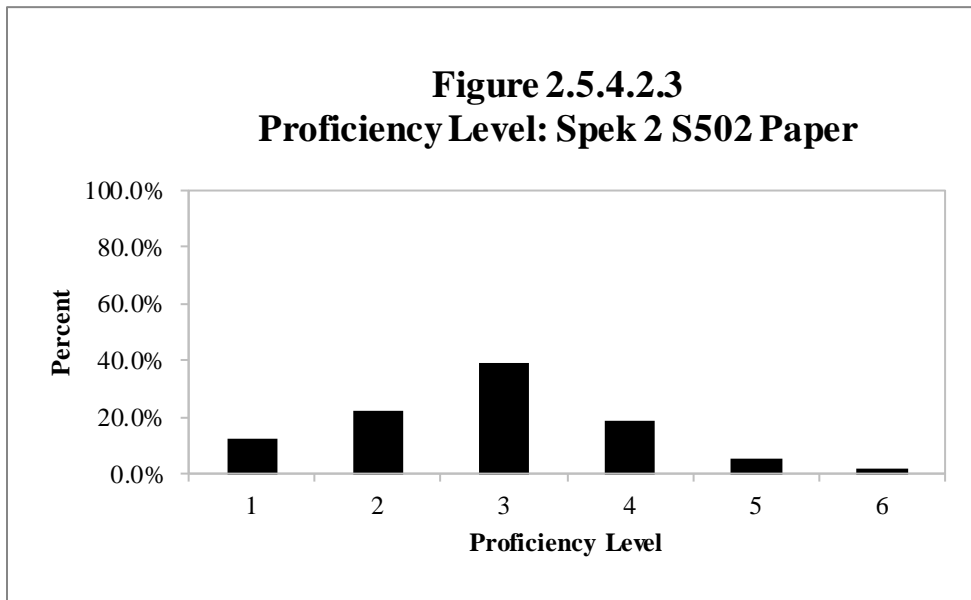


Table 2.5.4.2.3

Proficiency Level Distribution: Spek 2 S502 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	3,800	12.38%	3,800	12.38%
2	6,914	22.53%	6,914	22.53%
3	11,949	38.94%	11,949	38.94%
4	5,687	18.53%	5,687	18.53%
5	1,652	5.38%	1,652	5.38%
6	684	2.23%	684	2.23%
Total	30,686	100.00%	30,686	100.00%

Figure 2.5.4.2.3
Proficiency Level: Spek 2 S502 Paper



2.5.4.3. Grade 3

Table 2.5.4.3.1

Proficiency Level Distribution: Spek 3 A S502 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	2,282	46.70%	2,282	46.70%
2	1,299	26.59%	1,299	26.59%
3	855	17.50%	855	17.50%
4	450	9.21%	450	9.21%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	4,886	100.00%	4,886	100.00%

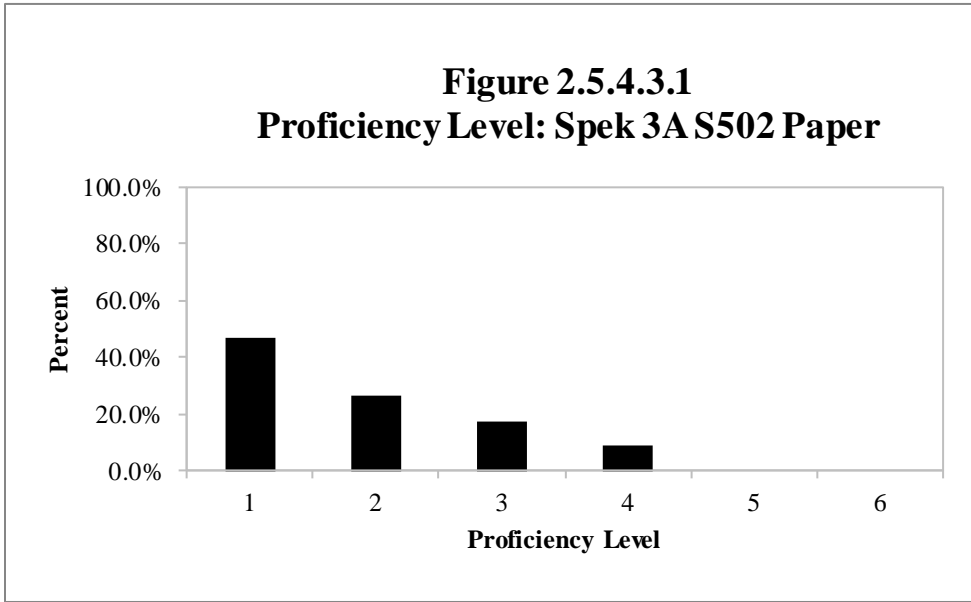


Table 2.5.4.3.2

Proficiency Level Distribution: Spek 3 B/C S502 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	1,203	5.37%	1,203	5.37%
2	4,689	20.92%	4,689	20.92%
3	9,805	43.75%	9,805	43.75%
4	4,833	21.56%	4,833	21.56%
5	902	4.02%	902	4.02%
6	981	4.38%	981	4.38%
Total	22,413	100.00%	22,413	100.00%

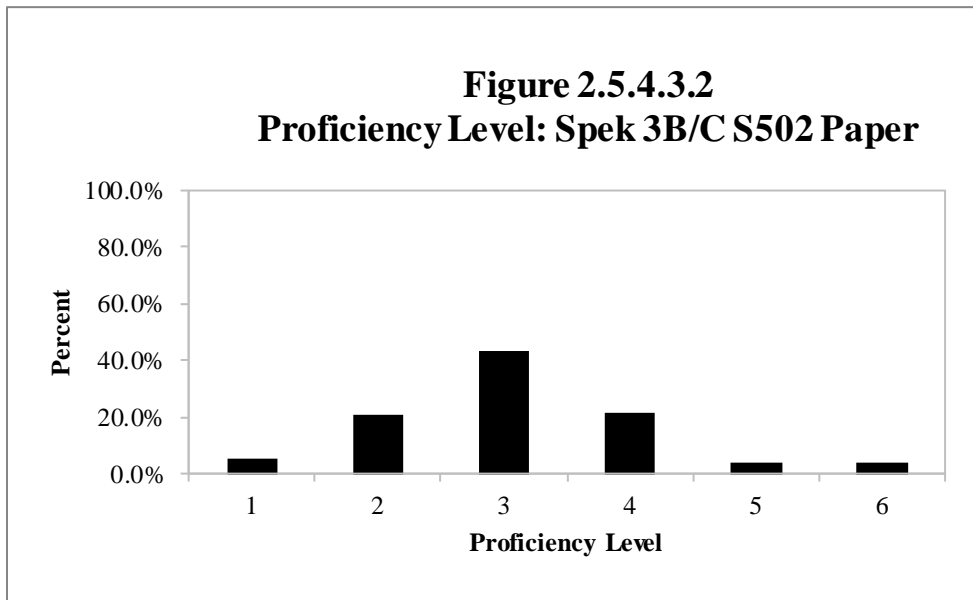
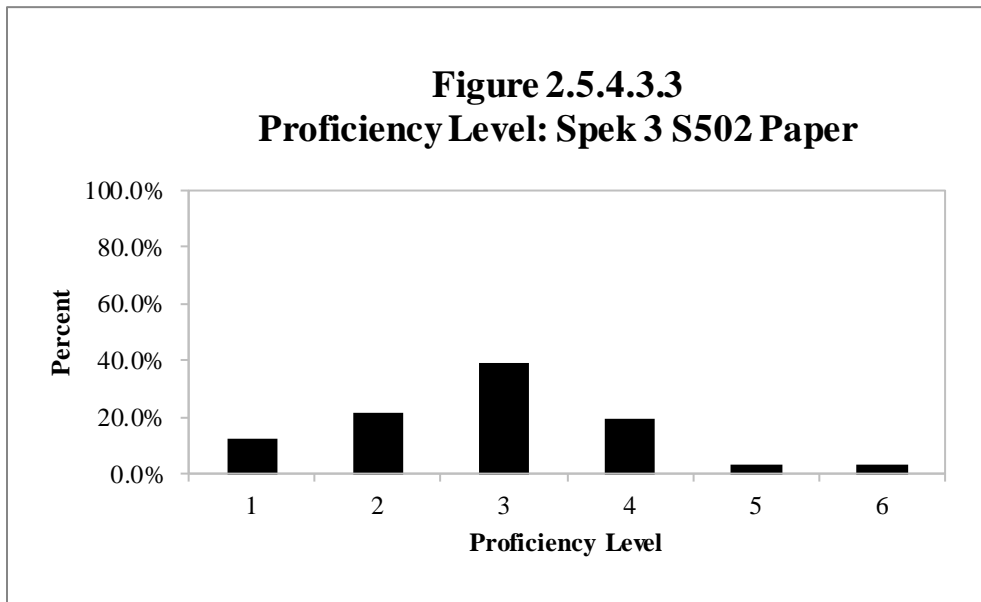


Table 2.5.4.3.3

Proficiency Level Distribution: Spek 3 S502 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	3,485	12.77%	3,485	12.77%
2	5,988	21.93%	5,988	21.93%
3	10,660	39.05%	10,660	39.05%
4	5,283	19.35%	5,283	19.35%
5	902	3.30%	902	3.30%
6	981	3.59%	981	3.59%
Total	27,299	100.00%	27,299	100.00%



2.5.4.4 Grades 4–5

Table 2.5.4.4.1

Proficiency Level Distribution: Spek 4-5 A S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	1,629	46.48%	1,838	59.37%	3,467	52.52%
2	1,033	29.47%	555	17.93%	1,588	24.06%
3	495	14.12%	366	11.82%	861	13.04%
4	255	7.28%	295	9.53%	550	8.33%
5	93	2.65%	42	1.36%	135	2.05%
6	0	0.00%	0	0.00%	0	0.00%
Total	3,505	100.00%	3,096	100.00%	6,601	100.00%

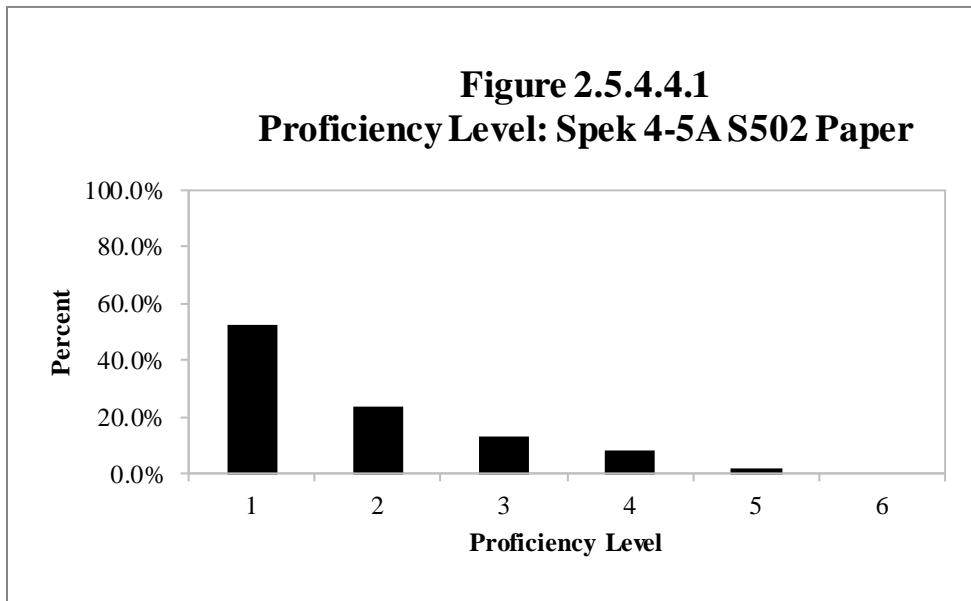


Table 2.5.4.4.2

Proficiency Level Distribution: Spek 4-5 B/C S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	565	2.57%	479	2.91%	1,044	2.72%
2	2,252	10.26%	1,572	9.54%	3,824	9.95%
3	5,828	26.54%	4,427	26.86%	10,255	26.68%
4	8,518	38.80%	6,251	37.93%	14,769	38.42%
5	3,184	14.50%	2,756	16.72%	5,940	15.45%
6	1,609	7.33%	997	6.05%	2,606	6.78%
Total	21,956	100.00%	16,482	100.00%	38,438	100.00%

Figure 2.5.4.4.2
Proficiency Level: Spek 4-5B/C S502 Paper

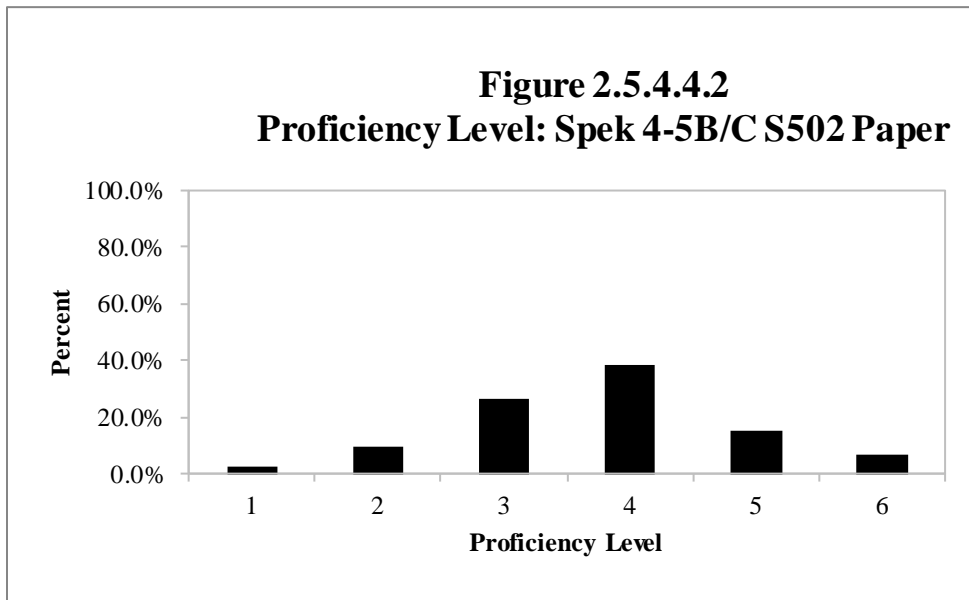
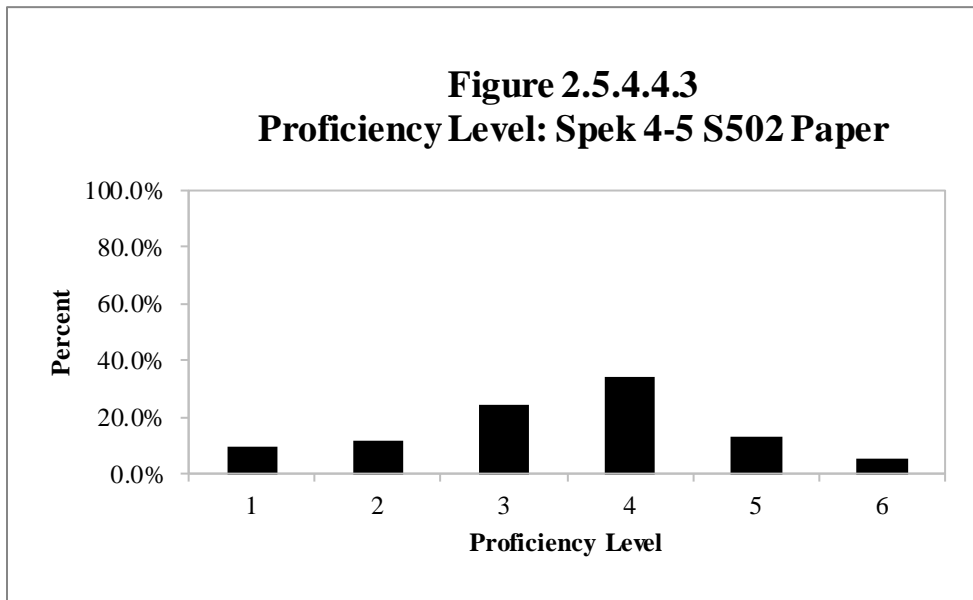


Table 2.5.4.4.3

Proficiency Level Distribution: Spek 4-5 S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	2,194	8.62%	2,317	11.83%	4,511	10.02%
2	3,285	12.90%	2,127	10.86%	5,412	12.02%
3	6,323	24.83%	4,793	24.48%	11,116	24.68%
4	8,773	34.46%	6,546	33.44%	15,319	34.01%
5	3,277	12.87%	2,798	14.29%	6,075	13.49%
6	1,609	6.32%	997	5.09%	2,606	5.79%
Total	25,461	100.00%	19,578	100.00%	45,039	100.00%



2.5.4.5 Grades 6–8

Table 2.5.4.5.1

Proficiency Level Distribution: Spek 6-8 A S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,488	48.95%	1,439	49.72%	1,804	61.26%	4,731	53.28%
2	663	21.81%	606	20.94%	322	10.93%	1,591	17.92%
3	485	15.95%	471	16.28%	565	19.19%	1,521	17.13%
4	270	8.88%	300	10.37%	177	6.01%	747	8.41%
5	101	3.32%	46	1.59%	77	2.61%	224	2.52%
6	33	1.09%	32	1.11%	0	0.00%	65	0.73%
Total	3,040	100.00%	2,894	100.00%	2,945	100.00%	8,879	100.00%

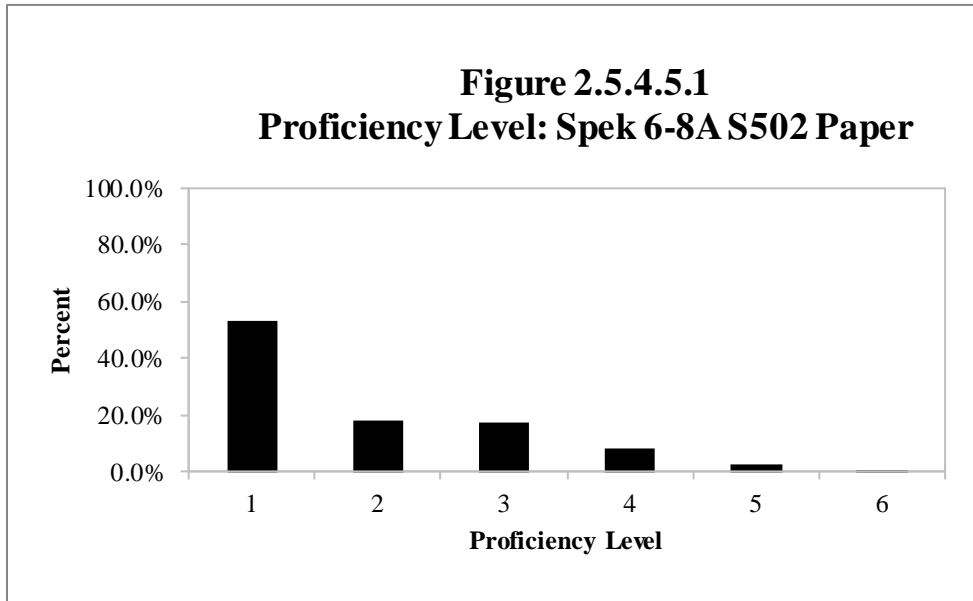


Table 2.5.4.5.2

Proficiency Level Distribution: Spek 6-8 B/C S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	265	2.24%	432	4.03%	481	5.24%	1,178	3.71%
2	1,679	14.19%	1,195	11.15%	1,119	12.19%	3,993	12.58%
3	3,272	27.65%	2,622	24.47%	2,327	25.34%	8,221	25.91%
4	4,056	34.28%	4,404	41.10%	3,062	33.34%	11,522	36.31%
5	1,619	13.68%	1,026	9.57%	1,460	15.90%	4,105	12.94%
6	941	7.95%	1,037	9.68%	734	7.99%	2,712	8.55%
Total	11,832	100.00%	10,716	100.00%	9,183	100.00%	31,731	100.00%

Figure 2.5.4.5.2
Proficiency Level: Spek 6-8B/C S502 Paper

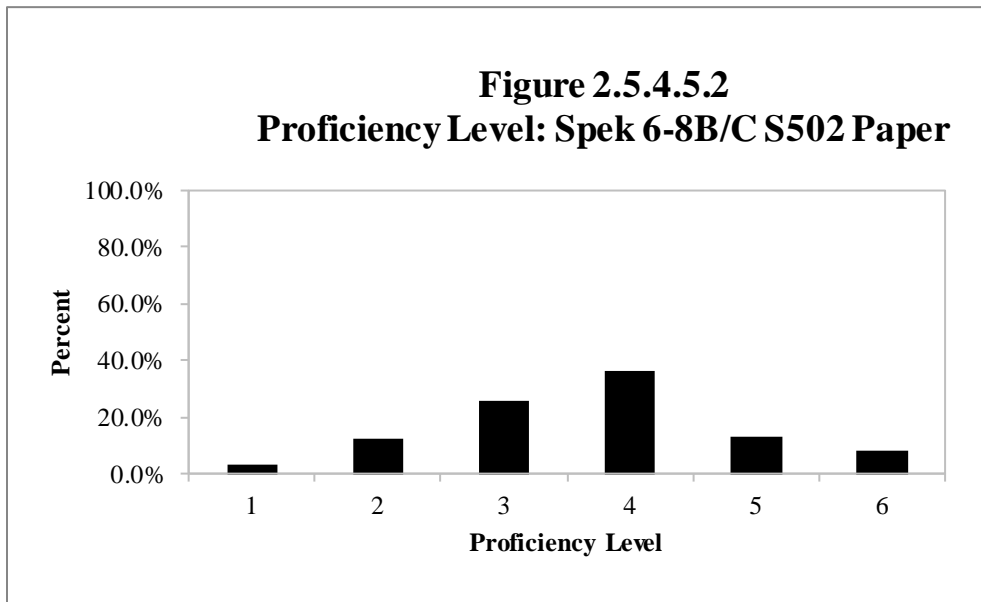
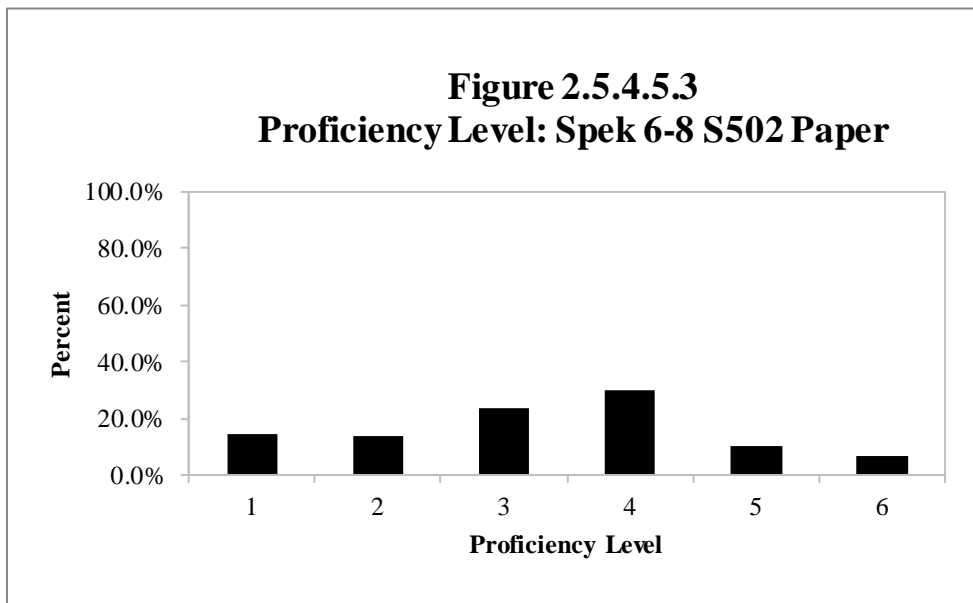


Table 2.5.4.5.3

Proficiency Level Distribution: Spek 6-8 S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,753	11.79%	1,871	13.75%	2,285	18.84%	5,909	14.55%
2	2,342	15.75%	1,801	13.23%	1,441	11.88%	5,584	13.75%
3	3,757	25.26%	3,093	22.73%	2,892	23.85%	9,742	23.99%
4	4,326	29.09%	4,704	34.56%	3,239	26.71%	12,269	30.21%
5	1,720	11.57%	1,072	7.88%	1,537	12.67%	4,329	10.66%
6	974	6.55%	1,069	7.85%	734	6.05%	2,777	6.84%
Total	14,872	100.00%	13,610	100.00%	12,128	100.00%	40,610	100.00%



2.5.4.6 Grades 9–12

Table 2.5.4.6.1

Proficiency Level Distribution: Spek 9-12 A S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,838	66.74%	1,710	67.99%	1,197	61.80%	718	58.90%	5,463	64.84%
2	261	9.48%	243	9.66%	225	11.62%	255	20.92%	984	11.68%
3	487	17.68%	423	16.82%	368	19.00%	189	15.50%	1,467	17.41%
4	116	4.21%	139	5.53%	147	7.59%	57	4.68%	459	5.45%
5	52	1.89%	0	0.00%	0	0.00%	0	0.00%	52	0.62%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	2,754	100.00%	2,515	100.00%	1,937	100.00%	1,219	100.00%	8,425	100.00%

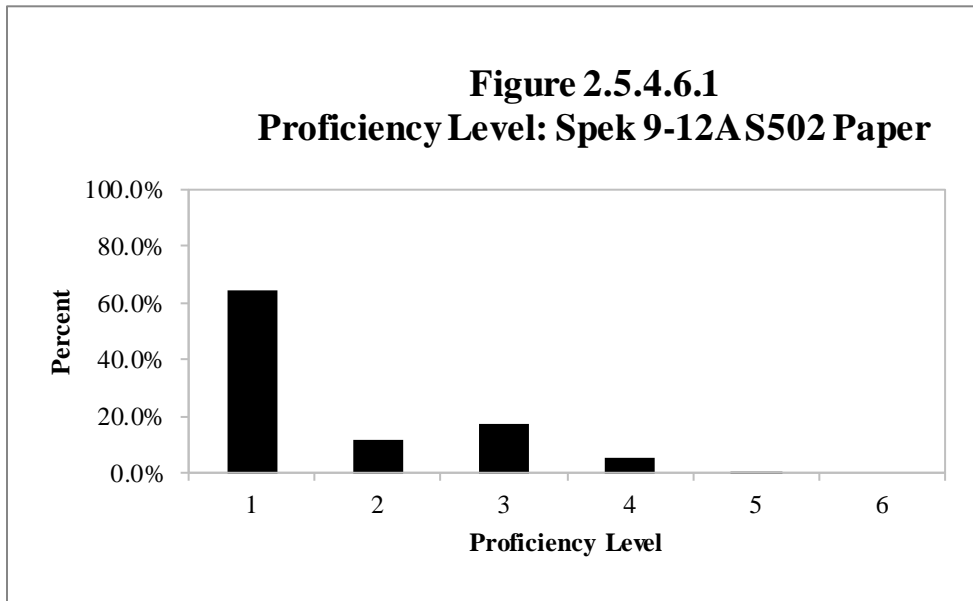


Table 2.5.4.6.2

Proficiency Level Distribution: Spek 9-12 B/C S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	564	6.76%	848	11.04%	667	9.69%	615	12.70%	2,694	9.71%
2	1,254	15.03%	1,042	13.56%	951	13.81%	777	16.05%	4,024	14.50%
3	2,522	30.23%	2,372	30.87%	2,664	38.69%	1,945	40.17%	9,503	34.24%
4	2,223	26.65%	2,375	30.91%	1,502	21.81%	809	16.71%	6,909	24.89%
5	942	11.29%	386	5.02%	404	5.87%	245	5.06%	1,977	7.12%
6	837	10.03%	661	8.60%	698	10.14%	451	9.31%	2,647	9.54%
Total	8,342	100.00%	7,684	100.00%	6,886	100.00%	4,842	100.00%	27,754	100.00%

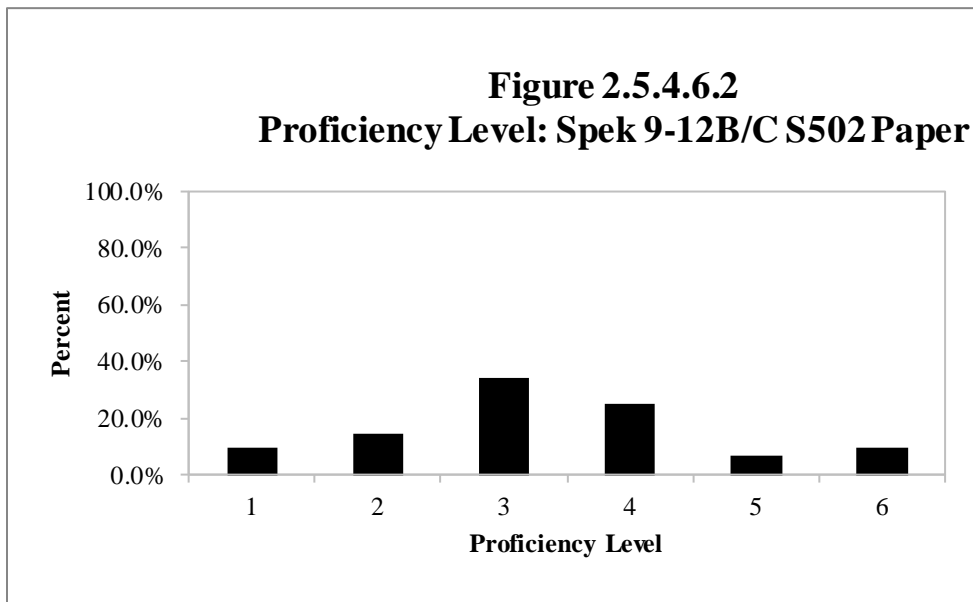
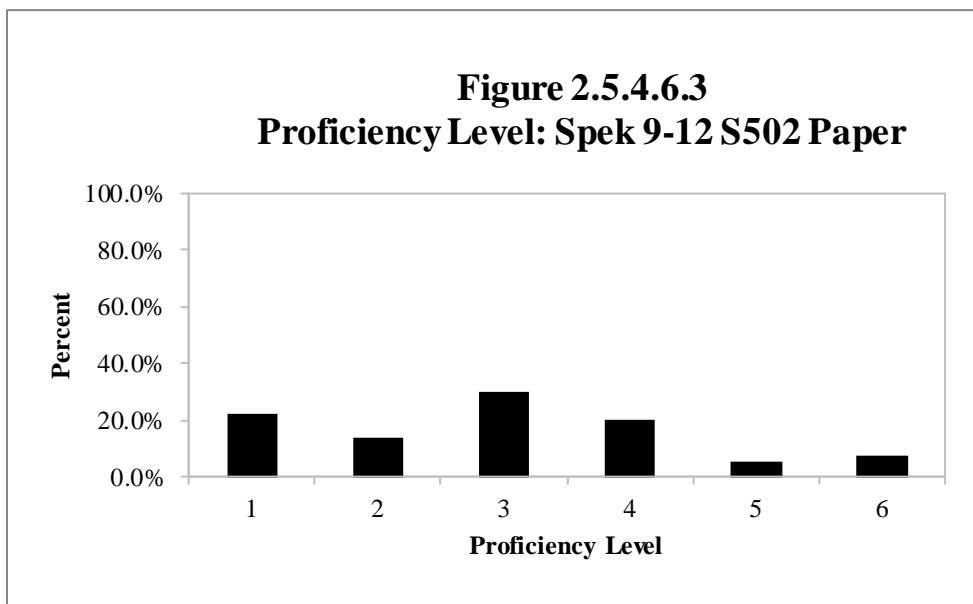


Table 2.5.4.6.3

Proficiency Level Distribution: Spek 9-12 S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,402	21.65%	2,558	25.08%	1,864	21.13%	1,333	21.99%	8,157	22.55%
2	1,515	13.65%	1,285	12.60%	1,176	13.33%	1,032	17.03%	5,008	13.84%
3	3,009	27.12%	2,795	27.40%	3,032	34.36%	2,134	35.21%	10,970	30.32%
4	2,339	21.08%	2,514	24.65%	1,649	18.69%	866	14.29%	7,368	20.37%
5	994	8.96%	386	3.78%	404	4.58%	245	4.04%	2,029	5.61%
6	837	7.54%	661	6.48%	698	7.91%	451	7.44%	2,647	7.32%
Total	11,096	100.00%	10,199	100.00%	8,823	100.00%	6,061	100.00%	36,179	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 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. This is the confidence band as reported on students' score reports. 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 S502 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 S502 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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	104	125	1.0
1	145	77	1.3
2	175	57	1.5
3	194	49	1.6
4	209	44	1.7
5	222	41	1.8
6	233	39	1.9
7	244	38	2.3
8	254	37	2.7
9	263	36	3.1
10	272	36	3.4
11	282	36	3.7
12	291	36	4.0
13	300	37	4.7
14	310	38	5.2
15	321	39	5.7
16	332	41	6.0
17	345	44	6.0
18	360	49	6.0
19	375	55	6.0
20	390	63	6.0
21	405	73	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 S502 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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G2
0	112	113	1.0
1	145	77	1.2
2	175	57	1.4
3	194	49	1.6
4	209	44	1.7
5	222	41	1.8
6	233	39	1.9
7	244	38	1.9
8	254	37	2.2
9	263	36	2.4
10	272	36	2.7
11	282	36	2.9
12	291	36	3.2
13	300	37	3.5
14	310	38	3.8
15	321	39	4.4
16	332	41	5.0
17	345	44	5.6
18	360	49	6.0
19	375	55	6.0
20	390	63	6.0
21	405	73	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 S502 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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G3
0	112	225	1.0
1	201	80	1.5
2	234	60	1.8
3	256	51	1.9
4	272	46	2.2
5	286	43	2.6
6	299	41	2.9
7	310	39	3.3
8	320	38	3.6
9	330	37	3.9
10	339	36	4.4
11	349	36	5.0
12	358	36	5.3
13	368	37	5.7
14	377	37	6.0
15	387	39	6.0
16	399	40	6.0
17	411	43	6.0
18	426	48	6.0
19	441	55	6.0
20	456	64	6.0
21	471	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 S502 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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5
0	120	203	1.0	1.0
1	201	80	1.5	1.4
2	234	60	1.7	1.6
3	256	51	1.8	1.8
4	272	46	1.9	1.9
5	286	43	2.2	2.0
6	299	41	2.6	2.3
7	310	39	2.9	2.6
8	320	38	3.2	2.9
9	330	37	3.5	3.2
10	339	36	3.8	3.5
11	349	36	4.3	3.8
12	358	36	4.7	4.1
13	368	37	5.2	4.6
14	377	37	5.5	5.0
15	387	39	5.9	5.4
16	399	40	6.0	5.9
17	411	43	6.0	6.0
18	426	48	6.0	6.0
19	441	55	6.0	6.0
20	456	64	6.0	6.0
21	471	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 S502 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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
0	132	248	1.0	1.0	1.0
1	226	76	1.6	1.5	1.5
2	255	56	1.7	1.7	1.6
3	274	47	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	329	35	2.9	2.7	2.5
9	338	35	3.1	2.9	2.7
10	347	35	3.4	3.2	3.0
11	355	35	3.7	3.5	3.2
12	364	35	4.0	3.8	3.5
13	373	36	4.4	4.1	3.8
14	382	37	4.8	4.5	4.2
15	392	38	5.2	4.9	4.6
16	403	40	5.6	5.3	5.0
17	416	44	6.0	5.8	5.5
18	431	49	6.0	6.0	6.0
19	446	55	6.0	6.0	6.0
20	461	64	6.0	6.0	6.0
21	476	74	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	148	98	1.0	1.0	1.0	1.0
1	169	79	1.1	1.1	1.1	1.1
2	201	60	1.3	1.3	1.2	1.2
3	223	52	1.4	1.4	1.4	1.3
4	240	47	1.5	1.5	1.5	1.4
5	255	44	1.6	1.6	1.5	1.5
6	269	42	1.7	1.6	1.6	1.6
7	281	41	1.8	1.7	1.7	1.6
8	293	40	1.8	1.8	1.7	1.7
9	304	40	1.9	1.8	1.8	1.8
10	315	40	2.0	1.9	1.8	1.8
11	327	40	2.3	2.0	1.9	1.9
12	338	41	2.6	2.3	2.1	1.9
13	350	43	2.9	2.7	2.5	2.3
14	364	45	3.3	3.1	3.0	2.8
15	380	50	3.9	3.7	3.5	3.4
16	396	56	4.5	4.2	4.0	3.9
17	412	64	5.1	4.8	4.6	4.5
18	428	76	5.7	5.5	5.2	5.0

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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	148	214	1.0	1.0	1.0	1.0
1	232	78	1.5	1.4	1.4	1.4
2	263	58	1.7	1.6	1.6	1.5
3	283	50	1.8	1.7	1.7	1.6
4	299	45	1.9	1.8	1.8	1.7
5	313	42	1.9	1.9	1.8	1.8
6	325	40	2.2	2.0	1.9	1.9
7	336	39	2.5	2.3	2.0	1.9
8	346	37	2.8	2.6	2.3	2.1
9	356	37	3.1	2.9	2.7	2.5
10	365	36	3.4	3.2	3.0	2.8
11	375	36	3.7	3.5	3.3	3.2
12	384	36	4.0	3.8	3.6	3.5
13	393	37	4.3	4.1	3.9	3.8
14	403	37	4.7	4.5	4.3	4.1
15	413	39	5.1	4.9	4.7	4.5
16	424	41	5.6	5.3	5.1	4.9
17	437	43	6.0	5.8	5.6	5.4
18	451	48	6.0	6.0	6.0	5.9
19	465	54	6.0	6.0	6.0	6.0
20	479	61	6.0	6.0	6.0	6.0
21	493	71	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 S502 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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	141	120	1.0
1	187	53	1.3
2	207	39	1.5
3	220	33	1.6
4	230	29	1.7
5	238	27	1.7
6	245	25	1.8
7	251	24	1.8
8	257	23	1.9
9	262	23	1.9
10	268	22	2.1
11	273	22	2.4
12	278	22	2.6
13	283	22	2.8
14	288	22	3.1
15	293	22	3.3
16	298	23	3.6
17	304	24	4.0
18	310	25	4.5
19	317	26	5.1
20	325	28	5.5
21	334	32	6.0
22	343	36	6.0
23	352	42	6.0
24	361	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	141	204	1.0
1	216	53	1.6
2	236	38	1.7
3	249	32	1.8
4	258	29	1.9
5	266	26	2.0
6	272	24	2.3
7	278	23	2.6
8	284	22	2.9
9	289	22	3.1
10	293	21	3.3
11	298	21	3.6
12	302	20	3.8
13	307	20	4.2
14	311	20	4.6
15	315	20	5.0
16	320	20	5.2
17	324	21	5.4
18	329	21	5.7
19	334	22	6.0
20	339	23	6.0
21	345	24	6.0
22	351	26	6.0
23	358	28	6.0
24	367	31	6.0
25	376	36	6.0
26	385	41	6.0
27	394	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G2
0	158	87	1.0
1	187	53	1.2
2	207	39	1.4
3	220	33	1.5
4	230	29	1.6
5	238	27	1.6
6	245	25	1.7
7	251	24	1.7
8	257	23	1.8
9	262	23	1.8
10	268	22	1.8
11	273	22	1.9
12	278	22	1.9
13	283	22	2.0
14	288	22	2.2
15	293	22	2.4
16	298	23	2.6
17	304	24	2.8
18	310	25	3.1
19	317	26	3.5
20	325	28	3.9
21	334	32	4.7
22	343	36	5.3
23	352	42	5.8
24	361	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 2.6.2.2.2

Raw Score to Scale Score to Proficiency Level Conversion: Read 2 B/C S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G2
0	158	149	1.0
1	216	53	1.4
2	236	38	1.6
3	249	32	1.7
4	258	29	1.8
5	266	26	1.8
6	272	24	1.9
7	278	23	1.9
8	284	22	2.0
9	289	22	2.2
10	293	21	2.4
11	298	21	2.6
12	302	20	2.7
13	307	20	3.0
14	311	20	3.2
15	315	20	3.4
16	320	20	3.6
17	324	21	3.8
18	329	21	4.2
19	334	22	4.7
20	339	23	5.1
21	345	24	5.4
22	351	26	5.7
23	358	28	6.0
24	367	31	6.0
25	376	36	6.0
26	385	41	6.0
27	394	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G3
0	158	144	1.0
1	213	52	1.3
2	233	38	1.5
3	246	32	1.6
4	255	29	1.6
5	263	26	1.7
6	270	25	1.8
7	276	24	1.8
8	282	23	1.8
9	287	22	1.9
10	292	22	1.9
11	297	22	2.0
12	302	22	2.1
13	307	22	2.3
14	312	22	2.5
15	317	22	2.7
16	323	23	3.0
17	328	24	3.2
18	334	25	3.5
19	341	26	3.9
20	349	28	4.7
21	358	32	5.3
22	367	36	5.8
23	376	41	6.0
24	385	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G3
0	158	438	1.0
1	271	52	1.8
2	290	38	1.9
3	302	31	2.1
4	311	28	2.5
5	319	26	2.8
6	325	24	3.1
7	331	23	3.4
8	336	22	3.6
9	341	21	3.9
10	345	21	4.3
11	350	20	4.8
12	354	20	5.1
13	358	20	5.3
14	363	20	5.6
15	367	20	5.8
16	371	20	6.0
17	375	21	6.0
18	380	21	6.0
19	385	22	6.0
20	390	23	6.0
21	396	24	6.0
22	402	25	6.0
23	409	28	6.0
24	418	31	6.0
25	427	36	6.0
26	436	41	6.0
27	445	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 S502 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	233	38	1.4	1.4
3	246	32	1.5	1.5
4	255	29	1.6	1.5
5	263	26	1.6	1.6
6	270	25	1.7	1.6
7	276	24	1.7	1.7
8	282	23	1.8	1.7
9	287	22	1.8	1.7
10	292	22	1.8	1.8
11	297	22	1.9	1.8
12	302	22	1.9	1.9
13	307	22	2.0	1.9
14	312	22	2.1	1.9
15	317	22	2.3	2.0
16	323	23	2.5	2.2
17	328	24	2.7	2.4
18	334	25	2.9	2.6
19	341	26	3.3	2.8
20	349	28	3.7	3.2
21	358	32	4.4	3.6
22	367	36	5.1	4.3
23	376	41	5.6	5.1
24	385	48	6.0	5.6

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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5
0	175	315	1.0	1.0
1	271	52	1.7	1.6
2	290	38	1.8	1.8
3	302	31	1.9	1.9
4	311	28	2.1	1.9
5	319	26	2.4	2.1
6	325	24	2.6	2.3
7	331	23	2.8	2.5
8	336	22	3.0	2.6
9	341	21	3.3	2.8
10	345	21	3.5	3.0
11	350	20	3.7	3.2
12	354	20	4.0	3.4
13	358	20	4.4	3.6
14	363	20	4.9	3.9
15	367	20	5.1	4.3
16	371	20	5.3	4.7
17	375	21	5.6	5.1
18	380	21	5.8	5.3
19	385	22	6.0	5.6
20	390	23	6.0	5.9
21	396	24	6.0	6.0
22	402	25	6.0	6.0
23	409	28	6.0	6.0
24	418	31	6.0	6.0
25	427	36	6.0	6.0
26	436	41	6.0	6.0
27	445	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
0	200	119	1.1	1.0	1.0
1	245	52	1.4	1.3	1.3
2	265	38	1.5	1.5	1.4
3	278	32	1.6	1.6	1.5
4	287	29	1.7	1.6	1.6
5	295	27	1.8	1.7	1.7
6	302	25	1.8	1.8	1.7
7	308	24	1.8	1.8	1.8
8	314	23	1.9	1.8	1.8
9	319	23	1.9	1.9	1.8
10	325	22	2.0	1.9	1.9
11	330	22	2.2	2.0	1.9
12	335	22	2.4	2.1	2.0
13	340	22	2.5	2.3	2.1
14	345	22	2.7	2.5	2.3
15	350	23	2.9	2.6	2.4
16	356	23	3.1	2.8	2.6
17	362	24	3.4	3.1	2.8
18	368	25	3.7	3.4	3.1
19	375	26	4.2	3.7	3.4
20	382	29	5.0	4.2	3.8
21	392	32	5.5	5.1	4.6
22	402	37	6.0	5.7	5.4
23	412	43	6.0	6.0	6.0
24	422	50	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
0	200	219	1.1	1.0	1.0
1	277	52	1.6	1.6	1.5
2	296	38	1.8	1.7	1.7
3	308	31	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	378	20	4.5	3.9	3.6
17	382	21	5.0	4.2	3.8
18	387	21	5.2	4.7	4.1
19	392	22	5.5	5.1	4.6
20	397	23	5.8	5.4	5.1
21	403	24	6.0	5.8	5.4
22	409	26	6.0	6.0	5.8
23	416	28	6.0	6.0	6.0
24	425	31	6.0	6.0	6.0
25	434	36	6.0	6.0	6.0
26	443	41	6.0	6.0	6.0
27	452	48	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	233	73	1.1	1.1	1.0	1.0
1	252	53	1.3	1.2	1.2	1.1
2	272	39	1.4	1.4	1.3	1.3
3	285	33	1.5	1.5	1.4	1.4
4	295	29	1.6	1.6	1.5	1.5
5	303	27	1.7	1.6	1.6	1.5
6	310	26	1.7	1.7	1.6	1.6
7	317	24	1.8	1.7	1.7	1.7
8	323	24	1.8	1.8	1.7	1.7
9	328	23	1.9	1.8	1.8	1.7
10	334	23	1.9	1.9	1.8	1.8
11	339	22	1.9	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	355	23	2.4	2.3	2.2	2.0
15	360	23	2.6	2.4	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	378	25	3.3	3.0	2.8	2.7
19	385	27	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	50	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	233	171	1.1	1.1	1.0	1.0
1	297	52	1.6	1.6	1.5	1.5
2	316	37	1.8	1.7	1.7	1.6
3	328	31	1.9	1.8	1.8	1.7
4	337	28	1.9	1.9	1.9	1.8
5	344	25	2.1	2.0	1.9	1.9
6	350	24	2.3	2.1	2.0	1.9
7	356	23	2.5	2.3	2.2	2.1
8	361	22	2.6	2.5	2.3	2.2
9	366	21	2.8	2.6	2.5	2.4
10	370	21	2.9	2.7	2.6	2.5
11	375	20	3.1	2.9	2.7	2.6
12	379	20	3.3	3.1	2.9	2.7
13	383	20	3.5	3.3	3.0	2.9
14	387	20	3.7	3.5	3.2	3.0
15	392	20	4.0	3.7	3.5	3.2
16	396	20	4.4	3.9	3.7	3.4
17	400	21	4.8	4.3	3.9	3.6
18	405	21	5.2	4.8	4.3	3.9
19	409	22	5.4	5.1	4.8	4.2
20	415	23	5.8	5.5	5.2	5.0
21	420	24	6.0	5.8	5.5	5.3
22	426	25	6.0	6.0	5.9	5.6
23	434	28	6.0	6.0	6.0	6.0
24	443	31	6.0	6.0	6.0	6.0
25	452	36	6.0	6.0	6.0	6.0
26	461	42	6.0	6.0	6.0	6.0
27	470	48	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for K
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 S502 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 S502 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	247	1.0	34	340	24	4.0
1	188	47	1.6	35	346	23	4.2
2	203	32	1.7	36	351	23	4.3
3	211	26	1.7	37	357	23	4.4
4	217	22	1.8	38	362	22	4.5
5	221	20	1.8	39	366	22	4.6
6	225	19	1.8	40	371	21	4.7
7	229	18	1.9	41	376	21	4.8
8	232	17	1.9	42	380	21	4.9
9	235	17	1.9	43	385	21	5.1
10	238	16	2.0	44	389	21	5.3
11	240	16	2.0	45	393	21	5.4
12	243	16	2.1	46	398	21	5.6
13	245	16	2.1	47	403	22	5.9
14	248	16	2.2	48	408	23	6.0
15	251	16	2.3	49	413	24	6.0
16	254	17	2.4	50	420	26	6.0
17	257	17	2.5	51	428	30	6.0
18	260	18	2.5	52	438	36	6.0
19	263	18	2.6	53	457	51	6.0
20	266	19	2.7	54	488	95	6.0
21	270	20	2.8				
22	274	20	2.9				
23	279	21	3.0				
24	283	22	3.1				
25	288	23	3.2				
26	294	23	3.3				
27	299	24	3.3				
28	305	24	3.4				
29	311	24	3.5				
30	317	24	3.6				
31	323	24	3.7				
32	329	24	3.8				
33	334	24	3.9				

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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G2
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 S502 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	158	1.0	34	341	24	4.0
1	187	47	1.5	35	346	23	4.1
2	201	32	1.6	36	352	23	4.2
3	210	26	1.7	37	357	22	4.3
4	216	23	1.7	38	362	22	4.4
5	220	21	1.8	39	367	22	4.5
6	224	19	1.8	40	371	21	4.6
7	228	18	1.8	41	376	21	4.7
8	231	18	1.9	42	381	21	4.8
9	234	17	1.9	43	385	21	4.9
10	237	17	1.9	44	390	21	5.0
11	240	17	1.9	45	394	21	5.2
12	243	17	2.0	46	399	22	5.4
13	246	17	2.1	47	404	22	5.6
14	249	17	2.1	48	409	23	5.9
15	252	17	2.2	49	415	25	6.0
16	255	17	2.3	50	422	27	6.0
17	258	17	2.4	51	430	30	6.0
18	261	18	2.5	52	441	36	6.0
19	264	18	2.5	53	459	51	6.0
20	268	19	2.7	54	490	95	6.0
21	272	19	2.8				
22	276	20	2.9				
23	280	21	3.0				
24	285	22	3.0				
25	290	22	3.1				
26	295	23	3.2				
27	300	23	3.3				
28	306	24	3.4				
29	312	24	3.5				
30	317	24	3.6				
31	323	24	3.7				
32	329	24	3.8				
33	335	24	3.9				

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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G3
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 S502 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	158	1.0	34	341	24	3.9
1	187	47	1.4	35	346	23	4.0
2	201	32	1.5	36	352	23	4.1
3	210	26	1.6	37	357	22	4.2
4	216	23	1.7	38	362	22	4.3
5	220	21	1.7	39	367	22	4.4
6	224	19	1.7	40	371	21	4.5
7	228	18	1.8	41	376	21	4.6
8	231	18	1.8	42	381	21	4.7
9	234	17	1.8	43	385	21	4.8
10	237	17	1.9	44	390	21	4.9
11	240	17	1.9	45	394	21	5.0
12	243	17	1.9	46	399	22	5.2
13	246	17	1.9	47	404	22	5.4
14	249	17	2.0	48	409	23	5.6
15	252	17	2.1	49	415	25	5.8
16	255	17	2.2	50	422	27	6.0
17	258	17	2.3	51	430	30	6.0
18	261	18	2.3	52	441	36	6.0
19	264	18	2.4	53	459	51	6.0
20	268	19	2.5	54	490	95	6.0
21	272	19	2.6				
22	276	20	2.8				
23	280	21	2.9				
24	285	22	3.0				
25	290	22	3.1				
26	295	23	3.1				
27	300	23	3.2				
28	306	24	3.3				
29	312	24	3.4				
30	317	24	3.5				
31	323	24	3.6				
32	329	24	3.7				
33	335	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5
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 S502 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	272	1.0	1.0	34	389	24	4.7	4.6
1	237	47	1.7	1.7	35	395	23	4.8	4.7
2	252	32	1.8	1.8	36	401	23	5.0	4.8
3	260	26	1.9	1.9	37	406	23	5.2	4.9
4	266	23	2.0	1.9	38	411	22	5.4	5.1
5	270	20	2.1	2.1	39	416	22	5.6	5.3
6	274	19	2.3	2.2	40	420	21	5.7	5.5
7	278	18	2.5	2.4	41	425	21	6.0	5.6
8	281	17	2.6	2.5	42	429	21	6.0	5.8
9	284	17	2.8	2.6	43	434	21	6.0	6.0
10	287	16	2.9	2.7	44	438	21	6.0	6.0
11	290	16	3.0	2.8	45	443	21	6.0	6.0
12	292	16	3.0	2.9	46	447	21	6.0	6.0
13	295	16	3.1	3.0	47	452	22	6.0	6.0
14	298	16	3.1	3.0	48	457	23	6.0	6.0
15	300	16	3.1	3.1	49	463	24	6.0	6.0
16	303	17	3.2	3.1	50	469	26	6.0	6.0
17	306	17	3.2	3.2	51	477	30	6.0	6.0
18	309	18	3.3	3.2	52	488	36	6.0	6.0
19	312	18	3.3	3.3	53	506	51	6.0	6.0
20	316	19	3.4	3.3	54	538	95	6.0	6.0
21	320	20	3.5	3.4					
22	324	20	3.5	3.4					
23	328	21	3.6	3.5					
24	333	22	3.7	3.6					
25	338	23	3.7	3.7					
26	343	23	3.8	3.7					
27	349	24	3.9	3.8					
28	354	24	4.0	3.9					
29	360	24	4.1	4.0					
30	366	24	4.3	4.1					
31	372	24	4.4	4.3					
32	378	24	4.5	4.4					
33	384	24	4.6	4.5					

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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
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	364	24	4.0	3.9	3.8

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
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	396	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	414	21	5.0	4.9	4.8
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 S502 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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	232	51	1.3	1.2	1.1	1.0
1	234	47	1.3	1.2	1.1	1.0
2	249	32	1.5	1.4	1.3	1.1
3	257	26	1.6	1.5	1.4	1.2
4	263	22	1.7	1.6	1.4	1.3
5	268	20	1.7	1.6	1.5	1.4
6	271	18	1.8	1.6	1.5	1.4
7	275	17	1.8	1.7	1.6	1.5
8	278	17	1.8	1.7	1.6	1.5
9	280	16	1.9	1.7	1.6	1.5
10	283	16	1.9	1.8	1.7	1.5
11	286	16	1.9	1.8	1.7	1.6
12	288	16	1.9	1.8	1.7	1.6
13	291	16	2.0	1.9	1.8	1.6
14	293	16	2.1	1.9	1.8	1.7
15	296	16	2.2	1.9	1.8	1.7
16	299	16	2.3	2.0	1.8	1.7
17	301	17	2.4	2.1	1.9	1.8
18	304	17	2.5	2.2	1.9	1.8
19	308	18	2.6	2.3	2.0	1.8
20	311	19	2.7	2.4	2.1	1.9
21	315	20	2.8	2.6	2.2	1.9
22	319	21	3.0	2.7	2.4	2.0
23	324	21	3.0	2.9	2.5	2.2
24	328	22	3.1	3.0	2.7	2.3
25	334	23	3.2	3.1	2.9	2.6
26	339	23	3.3	3.2	3.0	2.8
27	344	24	3.4	3.3	3.1	3.0
28	350	24	3.5	3.4	3.2	3.1
29	356	24	3.6	3.5	3.3	3.2
30	362	24	3.7	3.6	3.4	3.3
31	368	24	3.8	3.7	3.5	3.4
32	374	24	3.9	3.8	3.6	3.5
33	380	24	4.0	3.9	3.8	3.6

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
34	386	24	4.1	4.0	3.9	3.7
35	391	23	4.2	4.1	4.0	3.8
36	397	23	4.3	4.2	4.1	3.9
37	402	23	4.4	4.3	4.2	4.0
38	407	22	4.5	4.4	4.3	4.1
39	412	22	4.6	4.5	4.4	4.2
40	417	21	4.7	4.6	4.5	4.3
41	421	21	4.8	4.7	4.6	4.4
42	426	21	4.9	4.8	4.7	4.5
43	430	21	5.0	4.8	4.7	4.6
44	434	21	5.1	4.9	4.8	4.7
45	438	21	5.2	5.0	4.9	4.8
46	443	21	5.3	5.1	5.0	4.9
47	448	22	5.4	5.2	5.1	5.0
48	453	23	5.5	5.3	5.2	5.1
49	458	24	5.7	5.5	5.3	5.2
50	464	26	5.8	5.6	5.4	5.3
51	472	30	6.0	5.8	5.6	5.4
52	483	36	6.0	6.0	5.8	5.6
53	501	51	6.0	6.0	6.0	6.0
54	533	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for K
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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G1
0	106	47	1.0
1	106	47	1.0
2	118	38	1.1
3	130	34	1.2
4	140	33	1.3
5	151	34	1.4
6	162	36	1.5
7	174	37	1.6
8	187	38	1.8
9	201	40	1.9
10	216	42	2.1
11	235	49	2.5
12	259	55	2.9
13	286	52	3.5
14	308	48	3.9
15	328	47	4.3
16	349	50	4.7
17	370	59	5.2
18	391	75	5.7

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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G1
6	106	45	1.0
7	163	31	1.5
8	172	31	1.6
9	180	30	1.7
10	189	30	1.8
11	197	29	1.9
12	205	29	2.0
13	213	30	2.1
14	221	31	2.2
15	230	32	2.4
16	241	34	2.6
17	252	37	2.8
18	265	38	3.0
19	279	38	3.3
20	292	37	3.6
21	303	35	3.8
22	314	34	4.0
23	325	33	4.2
24	334	33	4.4
25	344	34	4.6
26	355	36	4.8
27	368	38	5.1
28	381	44	5.4
29	394	51	5.7
30	407	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G2
0	118	44	1.0
1	118	44	1.0
2	118	44	1.0
3	118	44	1.0
4	118	44	1.0
5	118	44	1.0
6	128	48	1.1
7	150	48	1.3
8	170	45	1.5
9	187	43	1.6
10	204	45	1.8
11	224	49	2.0
12	248	54	2.5
13	274	52	3.0
14	297	48	3.4
15	317	48	3.8
16	339	51	4.3
17	361	60	4.7
18	383	77	5.2

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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G2
6	118	41	1.0
7	141	43	1.2
8	156	39	1.4
9	169	36	1.5
10	180	33	1.6
11	190	32	1.7
12	199	32	1.8
13	209	32	1.8
14	218	32	1.9
15	228	33	2.1
16	238	34	2.3
17	250	36	2.5
18	262	37	2.7
19	275	37	3.0
20	287	36	3.2
21	299	35	3.5
22	310	34	3.7
23	320	34	3.9
24	331	34	4.1
25	342	35	4.3
26	353	37	4.5
27	367	40	4.8
28	381	45	5.1
29	395	52	5.5
30	425	78	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G3
0	118	44	1.0
1	118	44	1.0
2	118	44	1.0
3	118	44	1.0
4	118	44	1.0
5	118	44	1.0
6	128	48	1.0
7	150	48	1.2
8	170	45	1.4
9	187	43	1.5
10	204	45	1.7
11	224	49	1.9
12	248	54	2.2
13	274	52	2.8
14	297	48	3.2
15	317	48	3.6
16	339	51	4.1
17	361	60	4.5
18	383	77	4.9

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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G3
6	118	41	1.0
7	141	43	1.1
8	156	39	1.3
9	169	36	1.4
10	180	33	1.5
11	190	32	1.6
12	199	32	1.6
13	209	32	1.7
14	218	32	1.8
15	228	33	1.9
16	238	34	2.0
17	250	36	2.3
18	262	37	2.5
19	275	37	2.8
20	287	36	3.0
21	299	35	3.3
22	310	34	3.5
23	320	34	3.7
24	331	34	3.9
25	342	35	4.1
26	353	37	4.3
27	367	40	4.6
28	381	45	4.9
29	395	52	5.2
30	425	78	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5
0	130	40	1.0	1.0
1	130	40	1.0	1.0
2	132	39	1.0	1.0
3	145	36	1.1	1.1
4	156	35	1.2	1.2
5	168	37	1.3	1.2
6	182	41	1.4	1.4
7	198	42	1.6	1.5
8	215	42	1.7	1.6
9	231	42	1.8	1.7
10	247	44	2.0	1.9
11	266	49	2.4	2.1
12	291	55	2.9	2.7
13	317	52	3.4	3.3
14	340	48	3.9	3.7
15	360	47	4.3	4.1
16	381	51	4.7	4.5
17	402	59	5.1	4.9
18	423	75	5.6	5.4

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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G4	PL for G5
6	130	40	1.0	1.0
7	190	38	1.5	1.4
8	203	36	1.6	1.5
9	214	34	1.7	1.6
10	225	33	1.8	1.7
11	234	32	1.9	1.8
12	244	32	1.9	1.8
13	253	32	2.1	1.9
14	262	32	2.3	2.0
15	272	33	2.5	2.3
16	283	34	2.7	2.5
17	294	36	3.0	2.8
18	306	37	3.2	3.0
19	319	37	3.5	3.3
20	331	36	3.7	3.6
21	343	35	4.0	3.8
22	354	34	4.2	4.0
23	365	34	4.4	4.2
24	375	34	4.6	4.4
25	386	35	4.8	4.6
26	398	37	5.0	4.8
27	411	40	5.3	5.1
28	424	44	5.7	5.4
29	437	51	6.0	5.8
30	450	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
0	148	44	1.0	1.0	1.0
1	148	44	1.0	1.0	1.0
2	155	40	1.1	1.0	1.0
3	168	36	1.2	1.1	1.1
4	181	37	1.3	1.2	1.2
5	194	40	1.4	1.3	1.3
6	211	46	1.5	1.5	1.4
7	231	47	1.7	1.6	1.6
8	250	43	1.8	1.8	1.7
9	266	42	1.9	1.9	1.8
10	282	44	2.3	2.1	1.9
11	302	49	2.8	2.6	2.4
12	327	55	3.3	3.1	3.0
13	353	52	3.8	3.6	3.5
14	376	48	4.2	4.1	3.9
15	396	47	4.6	4.4	4.3
16	417	50	5.0	4.8	4.7
17	438	59	5.6	5.4	5.1
18	459	75	6.0	6.0	5.8

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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G6	PL for G7	PL for G8
6	148	43	1.0	1.0	1.0
7	219	40	1.6	1.5	1.5
8	232	37	1.7	1.6	1.6
9	244	34	1.8	1.7	1.7
10	254	32	1.8	1.8	1.7
11	263	30	1.9	1.8	1.8
12	271	30	2.0	1.9	1.9
13	279	30	2.2	2.0	1.9
14	288	31	2.4	2.2	2.1
15	297	33	2.6	2.5	2.3
16	308	34	2.9	2.7	2.6
17	319	37	3.1	3.0	2.8
18	332	39	3.4	3.2	3.1
19	346	38	3.7	3.5	3.4
20	359	37	3.9	3.8	3.6
21	371	35	4.1	4.0	3.8
22	382	34	4.3	4.2	4.0
23	392	33	4.5	4.4	4.2
24	402	33	4.7	4.5	4.4
25	412	34	4.9	4.7	4.6
26	423	35	5.1	4.9	4.8
27	435	38	5.5	5.3	5.0
28	447	43	5.8	5.6	5.4
29	459	49	6.0	6.0	5.8
30	471	57	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	172	37	1.1	1.0	1.0	1.0
1	172	37	1.1	1.0	1.0	1.0
2	172	37	1.1	1.0	1.0	1.0
3	180	34	1.1	1.1	1.1	1.0
4	191	33	1.2	1.2	1.1	1.1
5	201	34	1.3	1.3	1.2	1.2
6	213	36	1.4	1.3	1.3	1.3
7	225	38	1.5	1.4	1.4	1.4
8	239	40	1.6	1.5	1.5	1.5
9	254	41	1.7	1.6	1.6	1.6
10	270	43	1.8	1.8	1.7	1.7
11	289	48	1.9	1.9	1.9	1.9
12	313	54	2.6	2.4	2.3	2.2
13	339	52	3.1	3.1	3.0	2.9
14	361	48	3.5	3.4	3.3	3.3
15	382	47	3.9	3.8	3.7	3.6
16	403	51	4.3	4.1	4.0	3.9
17	424	59	4.7	4.5	4.4	4.3
18	445	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 S502 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
6	172	36	1.1	1.0	1.0	1.0
7	217	33	1.4	1.4	1.3	1.3
8	227	33	1.5	1.4	1.4	1.4
9	236	32	1.6	1.5	1.5	1.4
10	246	31	1.6	1.6	1.6	1.5
11	254	30	1.7	1.6	1.6	1.6
12	263	30	1.8	1.7	1.7	1.7
13	271	30	1.8	1.8	1.7	1.7
14	280	32	1.9	1.8	1.8	1.8
15	290	33	2.0	1.9	1.9	1.9
16	300	34	2.2	2.1	2.0	1.9
17	312	37	2.5	2.4	2.3	2.2
18	324	38	2.8	2.7	2.6	2.5
19	337	38	3.1	3.0	3.0	2.9
20	350	37	3.3	3.2	3.2	3.1
21	362	35	3.5	3.4	3.3	3.3
22	373	34	3.7	3.6	3.5	3.5
23	383	33	3.9	3.8	3.7	3.6
24	393	33	4.1	4.0	3.8	3.8
25	404	34	4.3	4.2	4.0	3.9
26	415	36	4.5	4.4	4.2	4.1
27	428	39	4.7	4.6	4.5	4.4
28	441	44	5.0	4.9	4.8	4.7
29	455	52	5.5	5.3	5.1	5.0
30	476	68	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 501 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 501 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

For ACCESS Paper Listening and Reading Grades 1–12 Tier A, all forms have been used in prior years. For ACCESS Paper Listening and Reading Grades 1–12 Tier B/C, we used Series 403, which was created using the Series 302 and Series 303 Tier B and Tier C item pools (see Part 1, Section 2.3). Table 2.7.1 summarizes the sources of Listening and Reading forms for Paper Series 502.

Table 2.7.1

Sources of Series 502 Paper Listening and Reading Forms

	Listening		Reading	
Tier A	Years previously used: Series 501 Paper Series 403 Paper Series 402 Paper Series 401 Paper Series 400 Paper Series 303	2019-20 2018-19 2017-18 2016-17 2015-16 2014-15	Years previously used: Series 403 Paper Series 401 Paper Series 303	2018-19 2016-17 2014-15
Tier B/C	Series 403		Series 403	

The S403 Tier B/C forms were drawn from the pool of Series 302 and 303 ACCESS. These forms were operational in 2013–2014 and 2014–2015, which were the 2 years prior to the launch of ACCESS Online. To mitigate concerns that there might be systematic differences between the population of students who took ACCESS 302 and 303 and the population of students who currently take Paper ACCESS, we conducted a series of recalibration studies using Series 400 and Series 401 Paper population data to refine Series 302 and Series 303 Listening and Reading item parameters.

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 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 501 Paper Reading Grades 1–12 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.

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 302.

For Listening and Reading Tier B/C, we used the refined parameters derived from the 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.

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

	Writing		Speaking	
Tier A	Years previously used: Series 403 Paper	2018-19	Years previously used: Series 403 Paper	2018-19
	Series 401 Paper	2016-17	Series 401 Paper	2016-17
Tier B/C	Years previously used: Series 403 Paper	2018-19	Years previously used: Series 403 Paper	2018-19
	Series 401 Paper	2016-17	Series 401 Paper	2016-17

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.

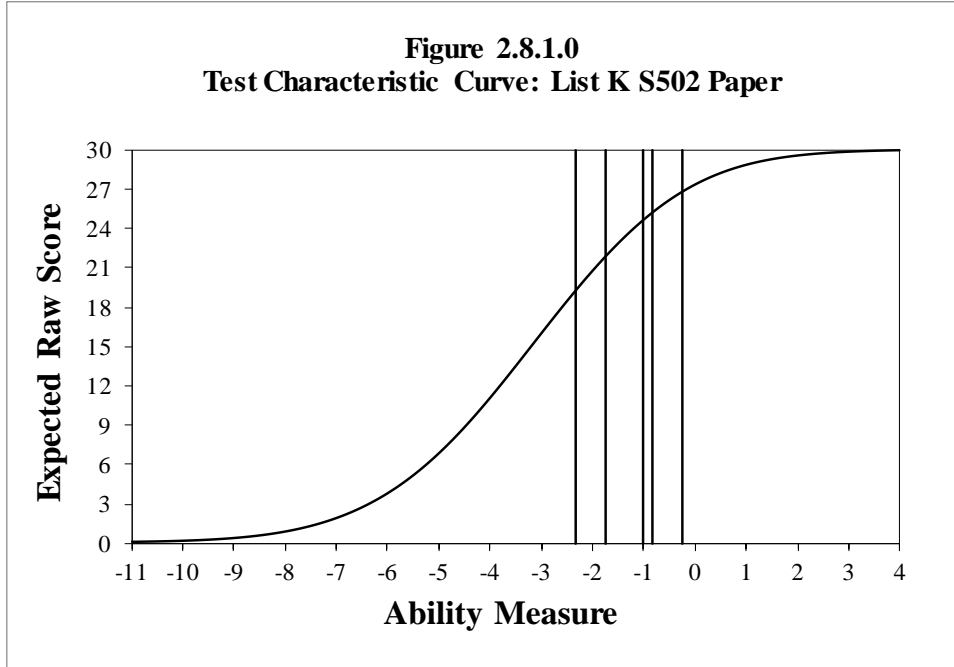
There are five vertical lines in each of the TCC plots indicating the five cut scores for the highest grade in the grade-level cluster for the test form, dividing the figure into six sections for each of the WIDA proficiency levels (PLs 1–6) for the domain being tested. (Note that for Kindergarten and 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 into 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.

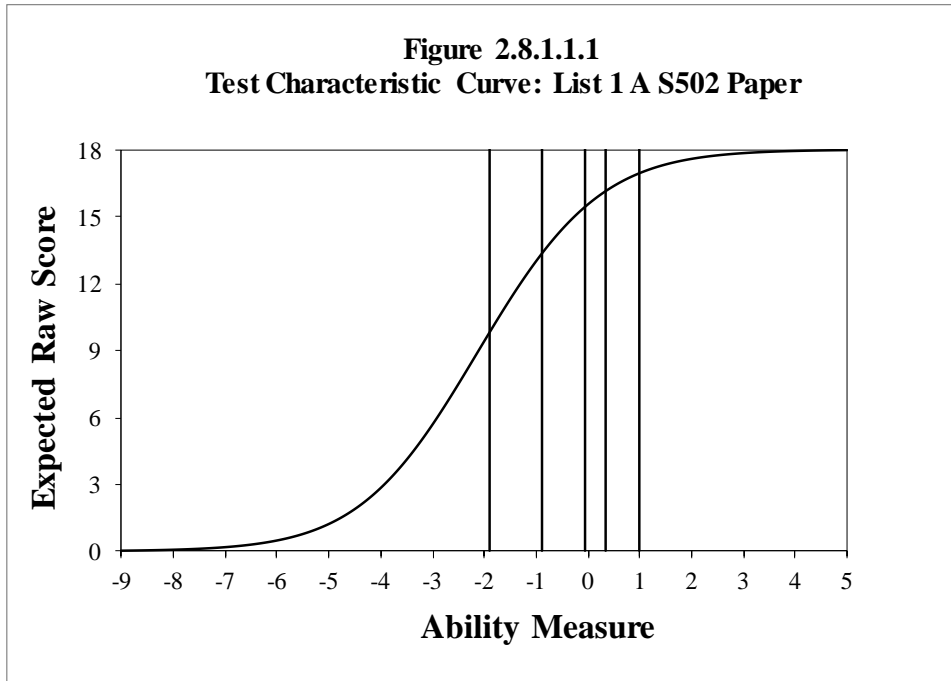
In addition to the TCC by tier, TCCs across tier for the grade cluster are plotted on the same graph. Since each tier has different numbers of expected raw score points, it is not appropriate to compare the expected raw score points for the same proficiency measure between tiers. It is, however, informative to compare where the slopes are the steepest, which corresponds to the ability range where the best measurement information is provided. For example, the across-tier TCC for Listening Grade 1 showed that the Listening Grade 1 Tier A form provides the best measurement at around an ability measure of -1.0, or around PL 3, while the Listening Grade 1 Tier B/C form provides the best measurement at a higher proficiency level (an ability measure of 0.3 or around PL 5), as expected. In addition, it is informative to compare the area under the curve for the TCC of each tier form. For example, the Grade 1 Tier A curve covers an area of lower ability range than the Grade 1 Tier B/C curve. As expected, there is also considerable overlap between the areas covered by the two forms.

2.8.1 Listening

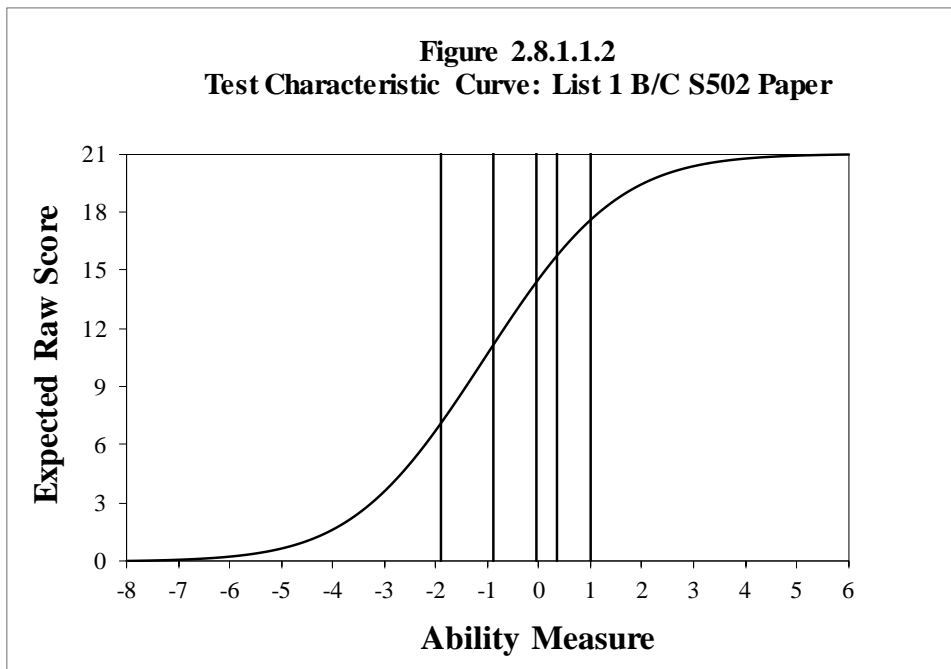
2.8.1.0 Kindergarten



2.8.1.1 Grade 1

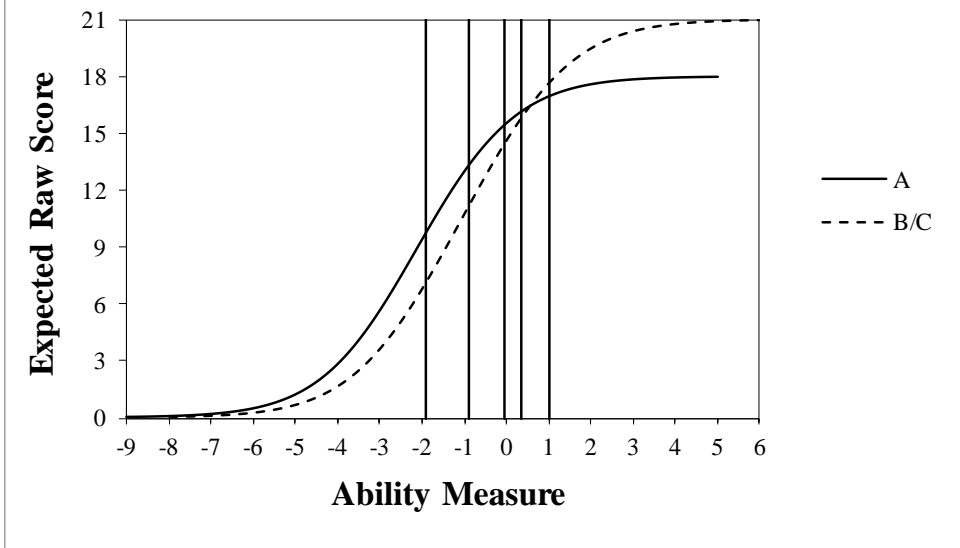


Note: The test form is shared between 1A and 2A.



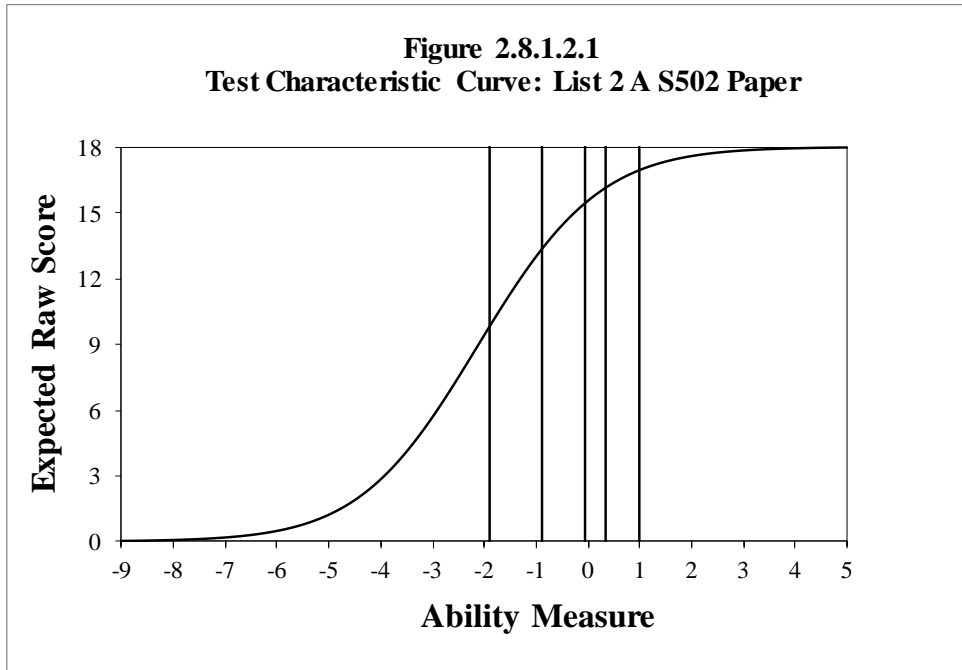
Note: The test form is shared between 1B/C and 2B/C.

Figure 2.8.1.1.3
Test Characteristic Curve: List 1 S502 Paper

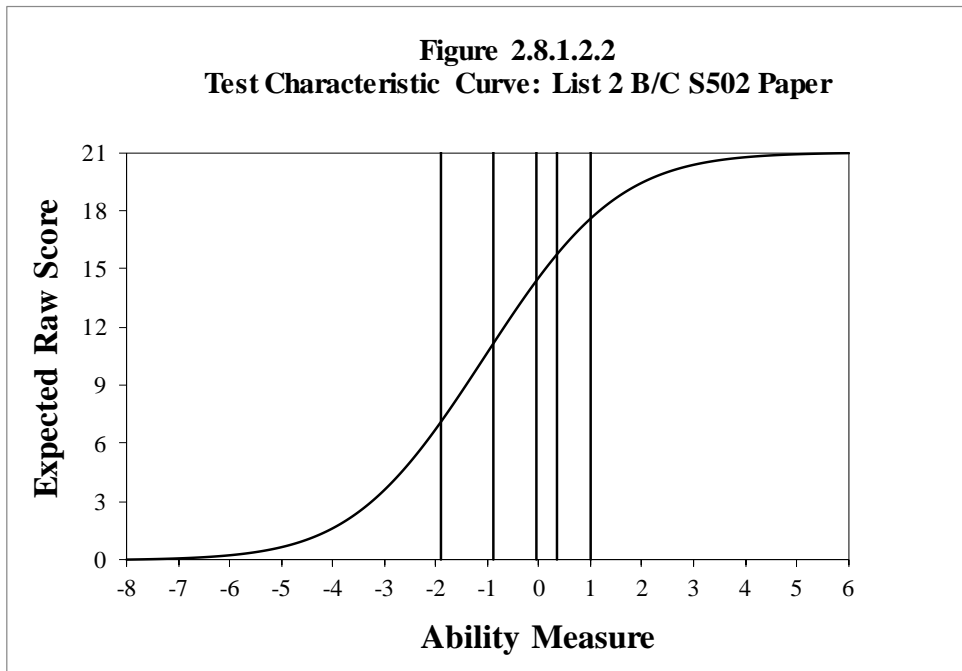


Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

2.8.1.2 Grade 2

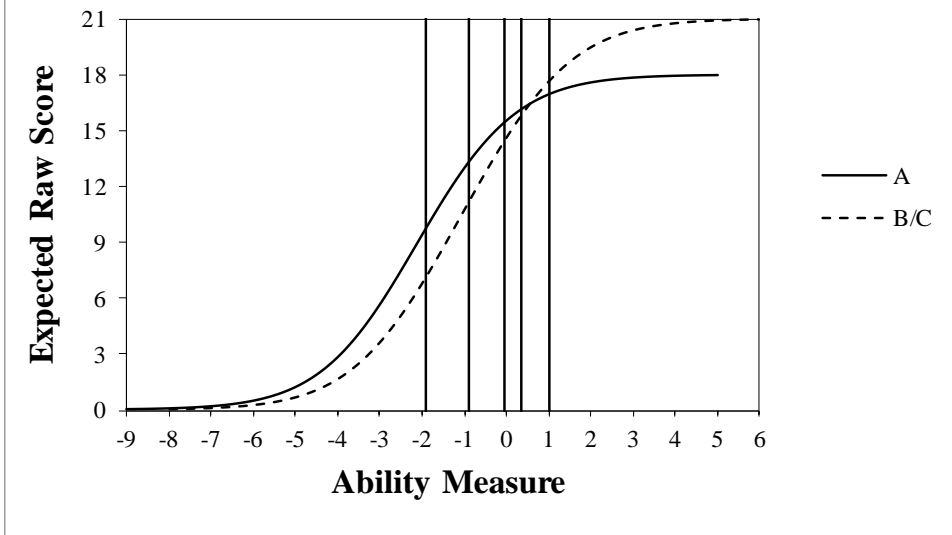


Note: The test form is shared between 1A and 2A.



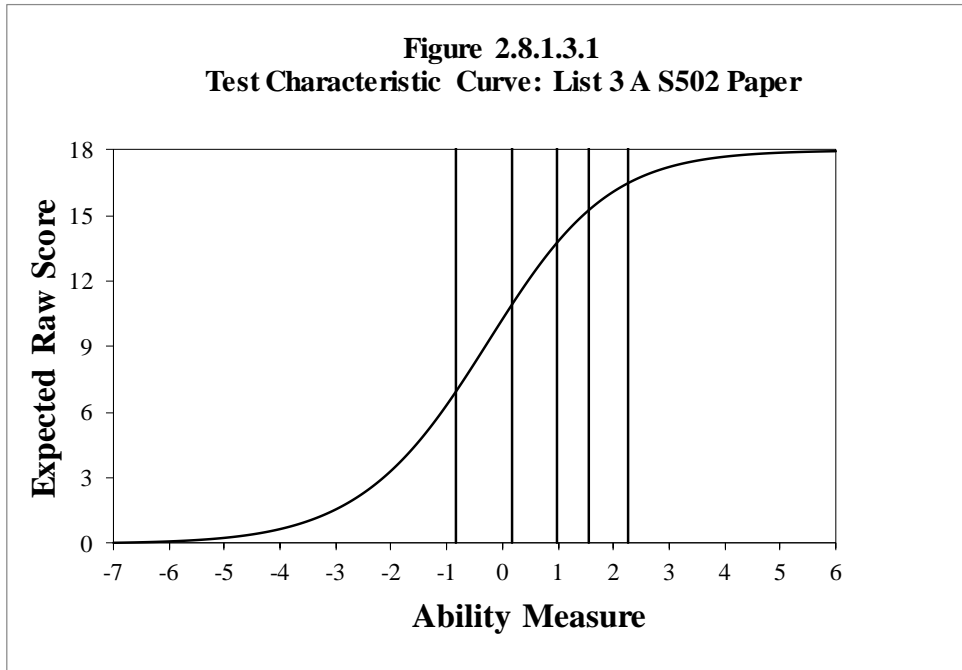
Note: The test form is shared between 1B/C and 2B/C.

Figure 2.8.1.2.3
Test Characteristic Curve: List 2 S502 Paper

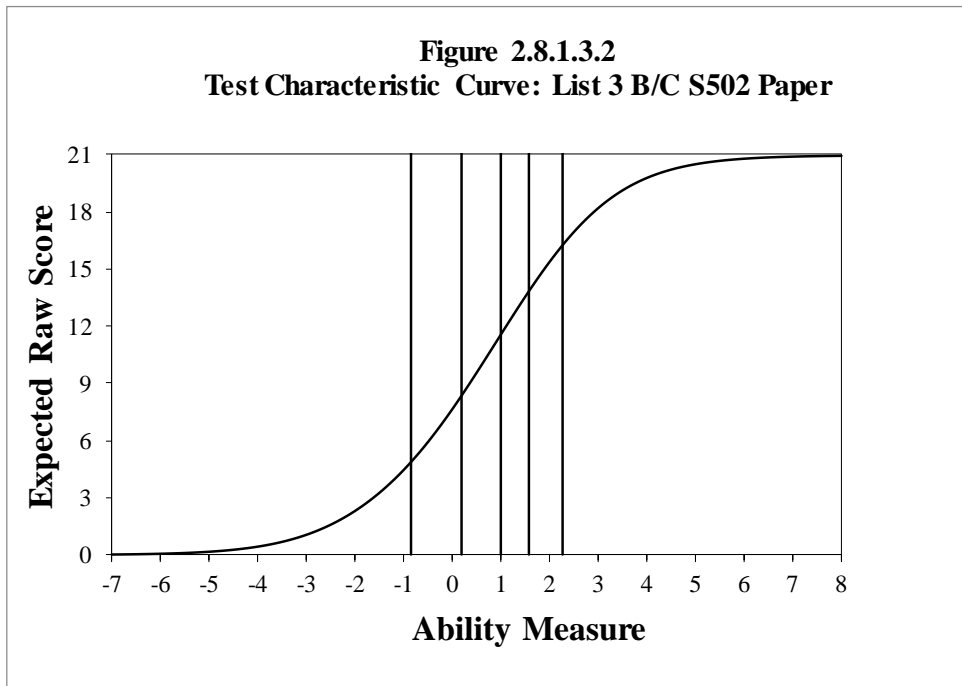


Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

2.8.1.3 Grade 3

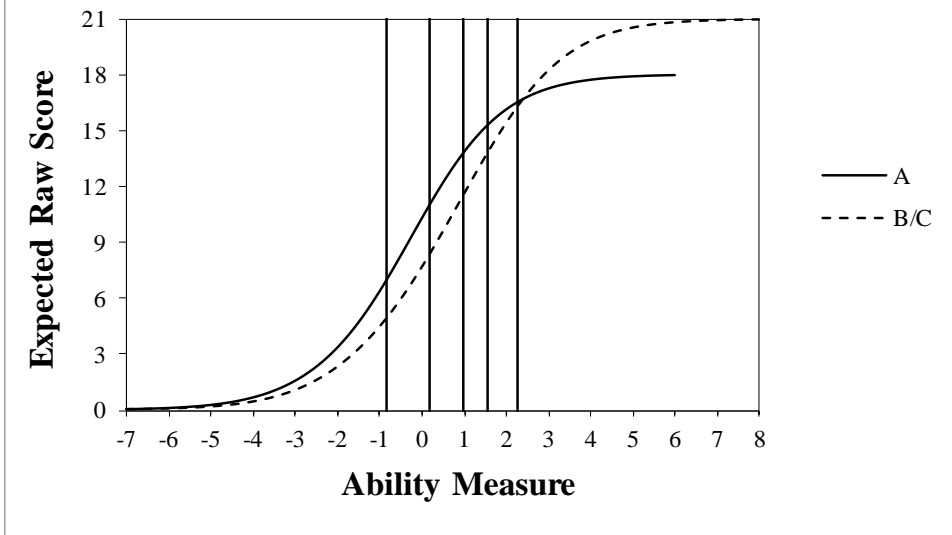


Note: The test form is shared between 3A and 4–5A.



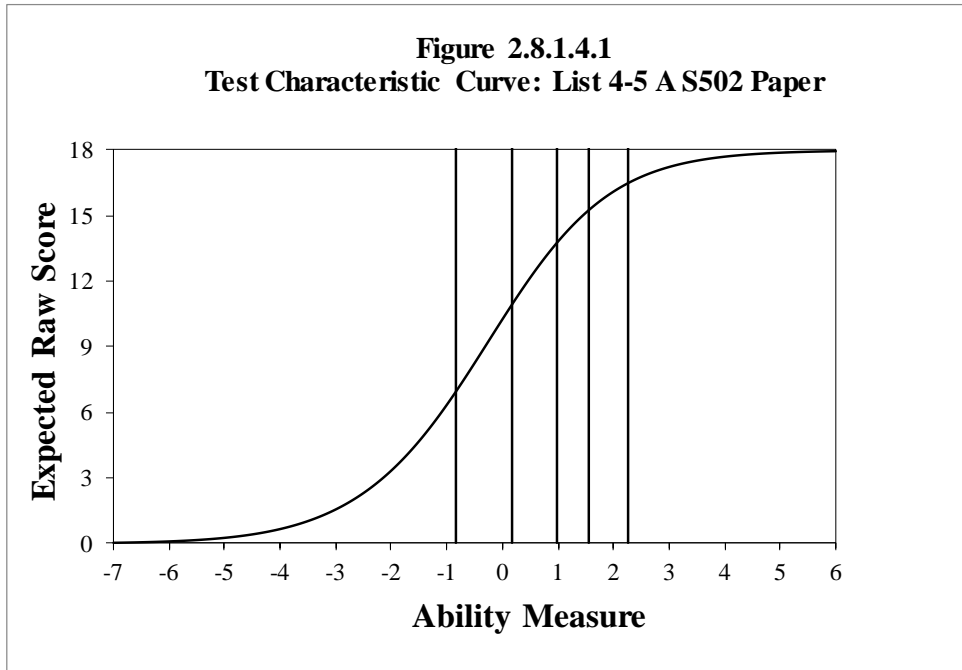
Note: The test form is shared between 3B/C and 4–5B/C.

Figure 2.8.1.3.3
Test Characteristic Curve: List 3 S502 Paper

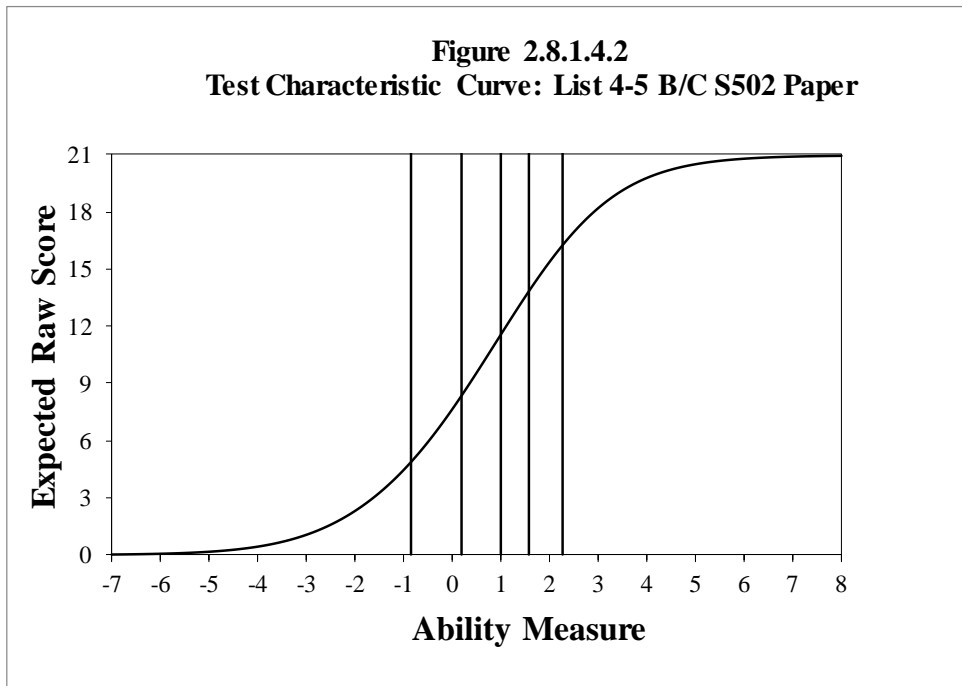


Note: The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

2.8.1.4 Grades 4–5

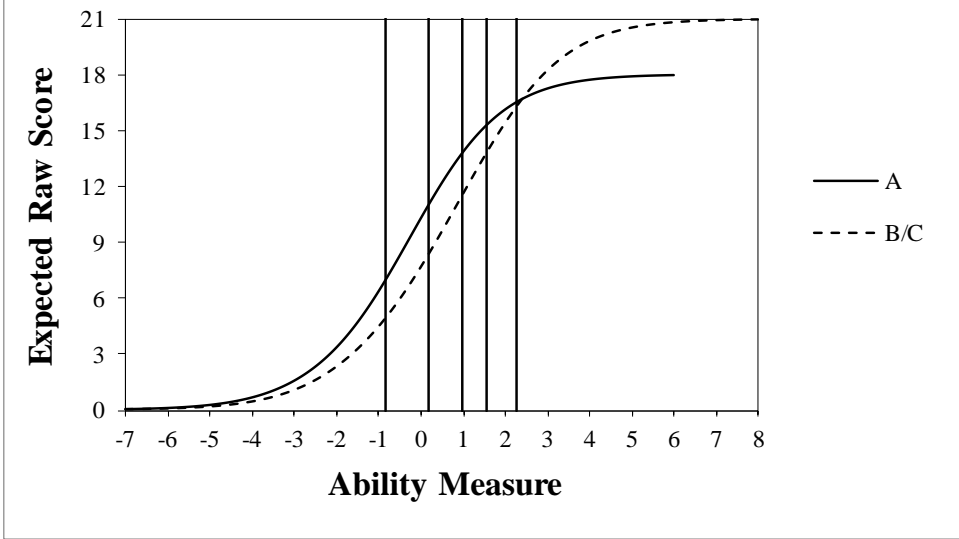


Note: The test form is shared between 3A and 4–5A.



Note: The test form is shared between 3B/C and 4–5B/C.

Figure 2.8.1.4.3
Test Characteristic Curve: List 4-5 S502 Paper



Note: The test form is shared between 3A and 4-5A, 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 S502 Paper

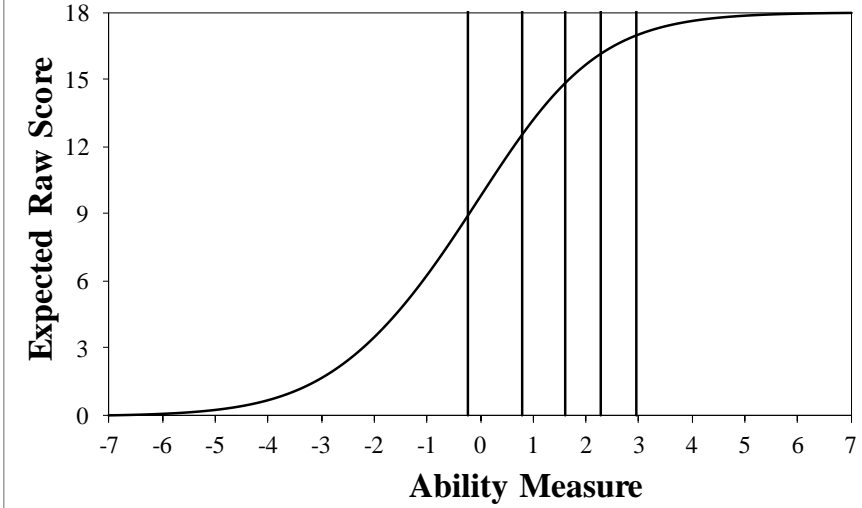


Figure 2.8.1.5.2
Test Characteristic Curve: List 6-8 B/C S502 Paper

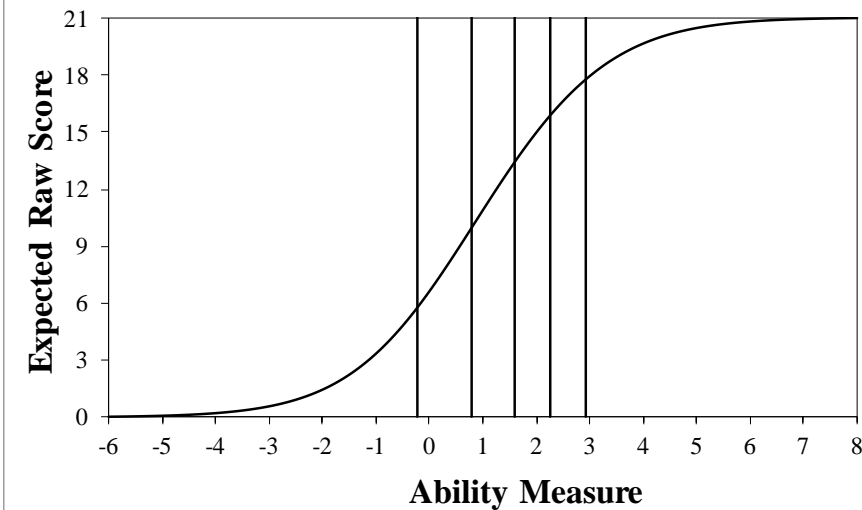
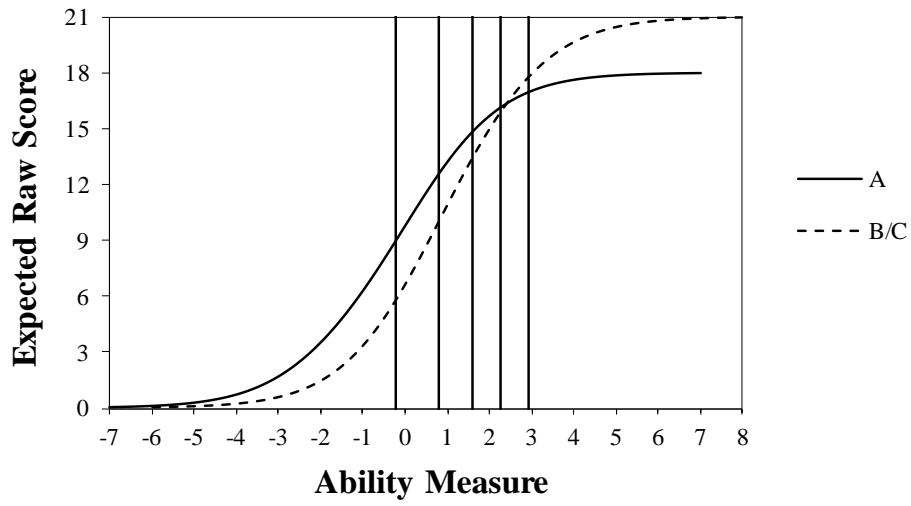


Figure 2.8.1.5.3
Test Characteristic Curve: List 6-8 S502 Paper



2.8.1.6 Grades 9–12

Figure 2.8.1.6.1
Test Characteristic Curve: List 9-12 A S502 Paper

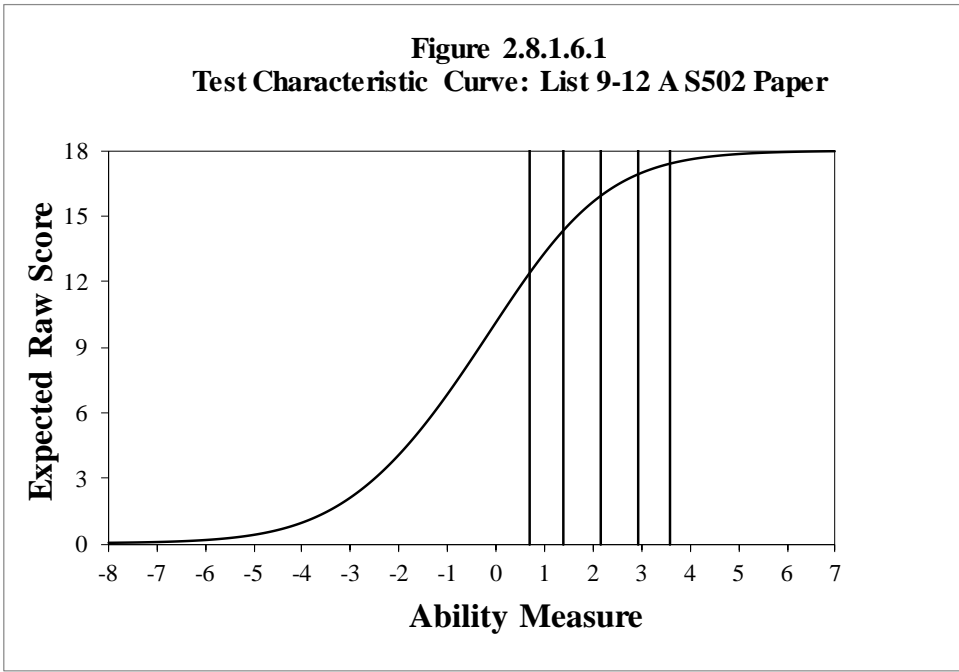


Figure 2.8.1.6.2
Test Characteristic Curve: List 9-12 B/C S502 Paper

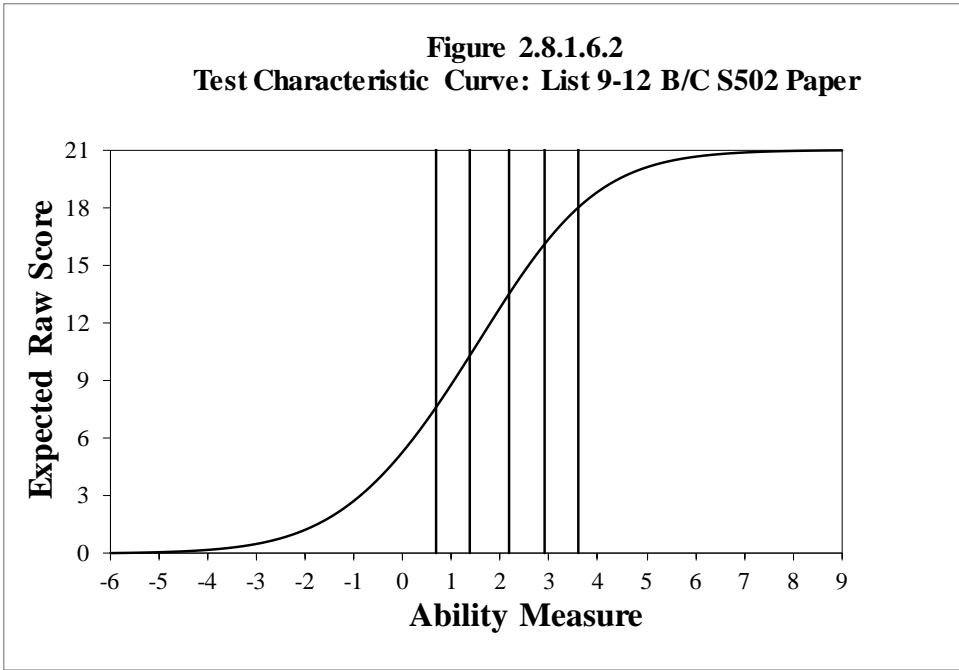
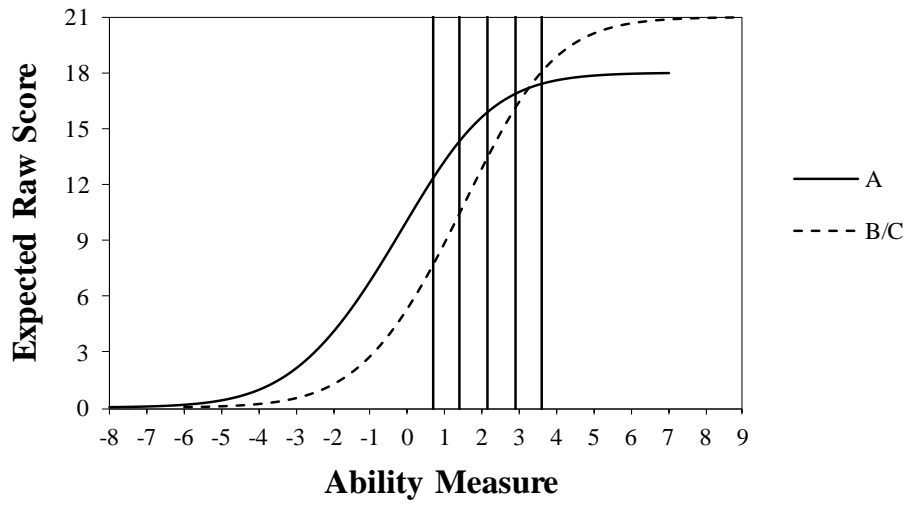
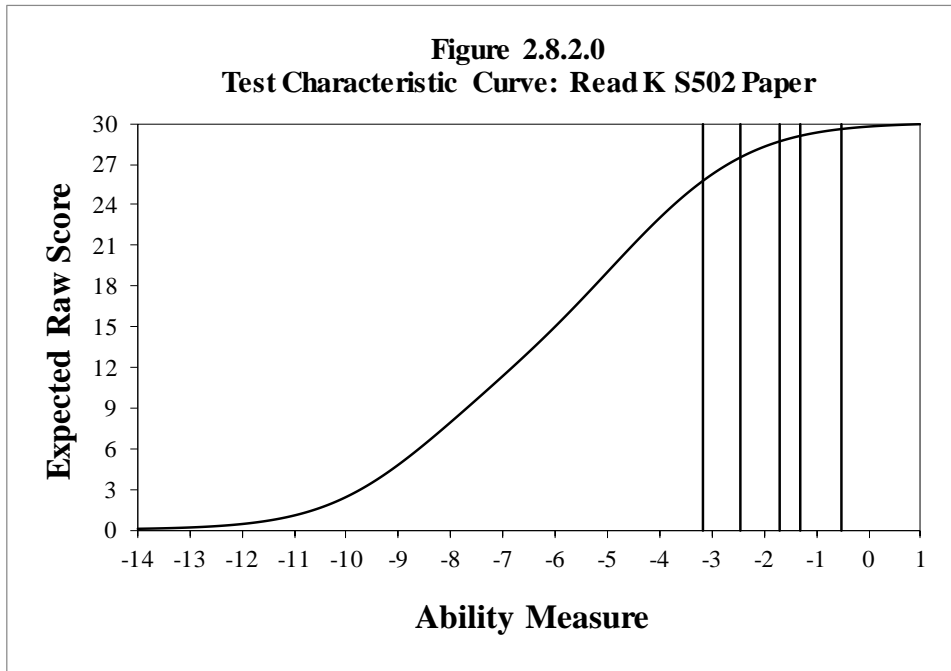


Figure 2.8.1.6.3
Test Characteristic Curve: List 9-12 S502 Paper

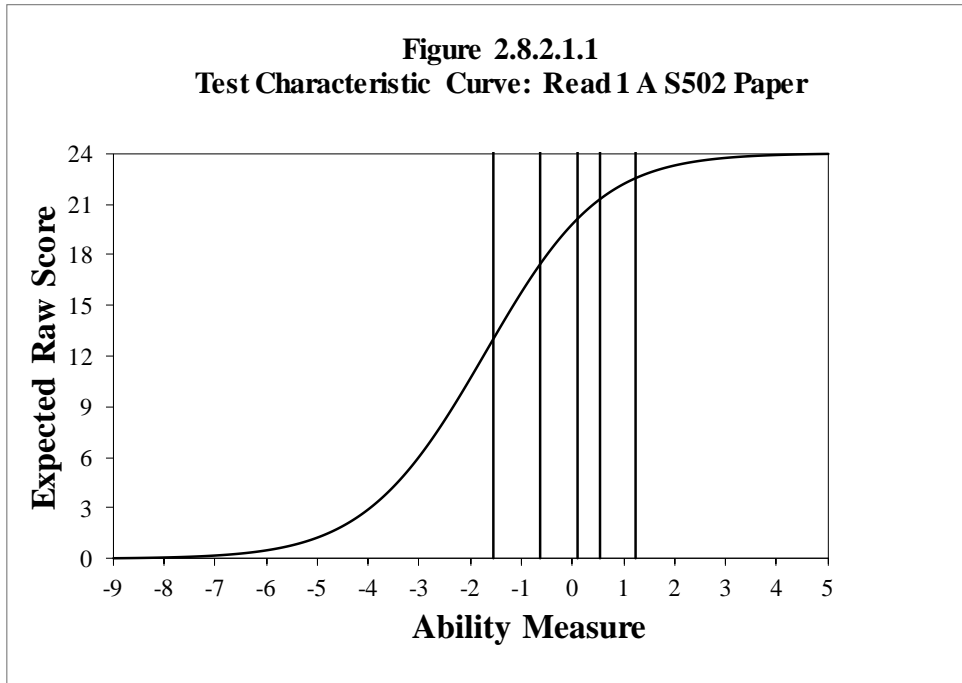


2.8.2 Reading

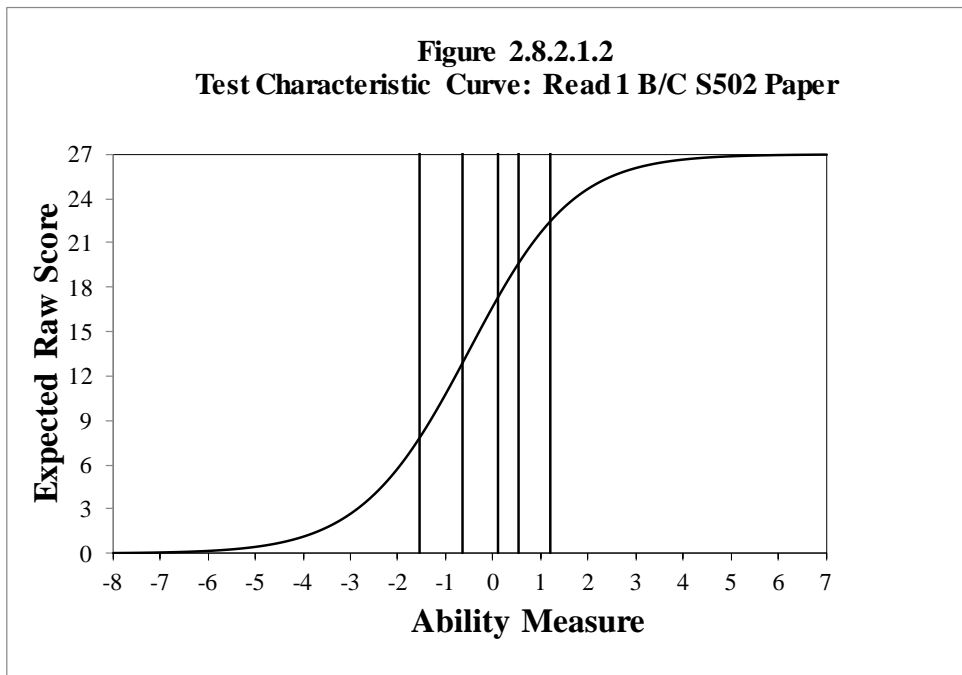
2.8.2.0 Kindergarten



2.8.2.1 Grade 1

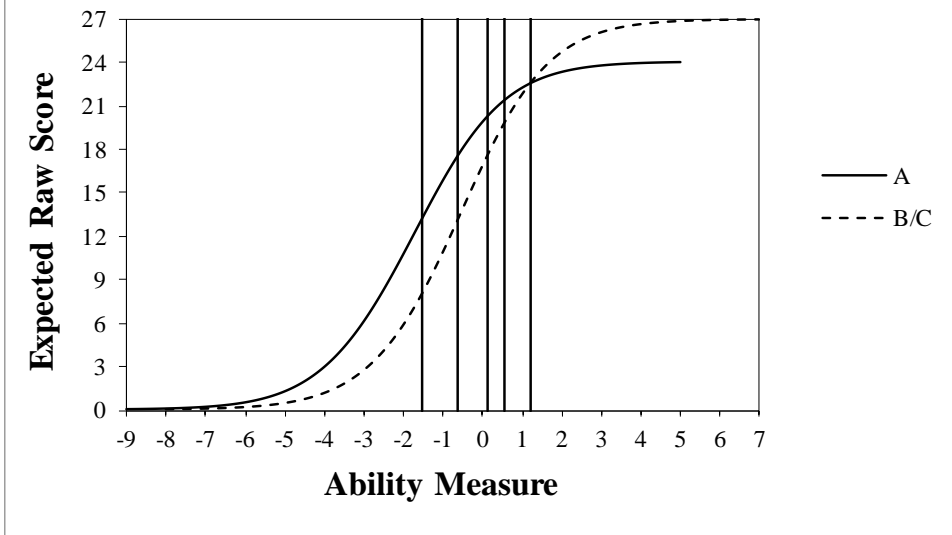


Note: The test form is shared between 1A and 2A.



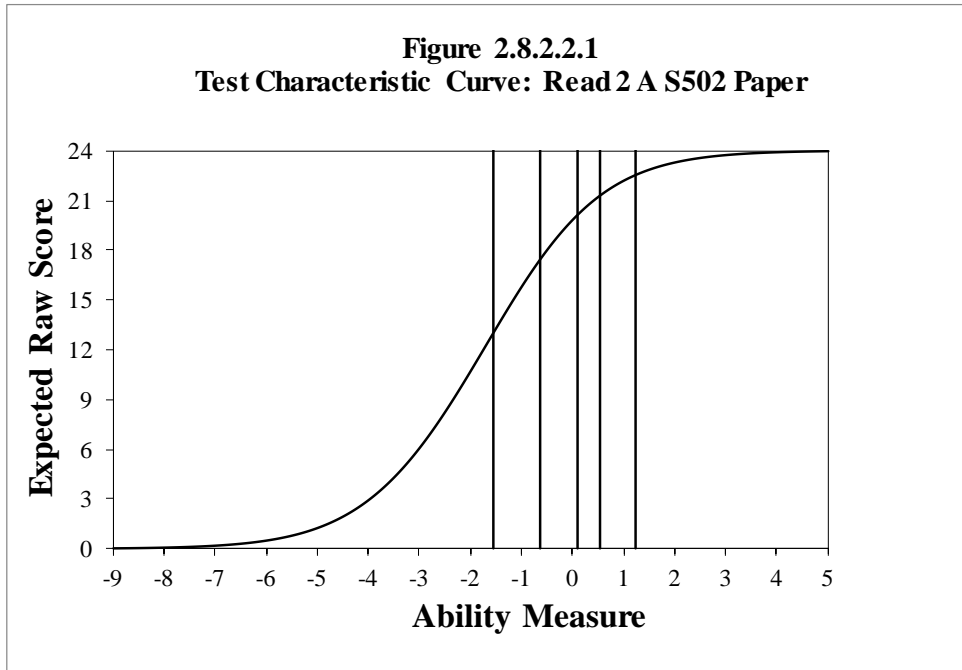
Note: The test form is shared between 1B/C and 2B/C.

Figure 2.8.2.1.3
Test Characteristic Curve: Read 1 S502 Paper

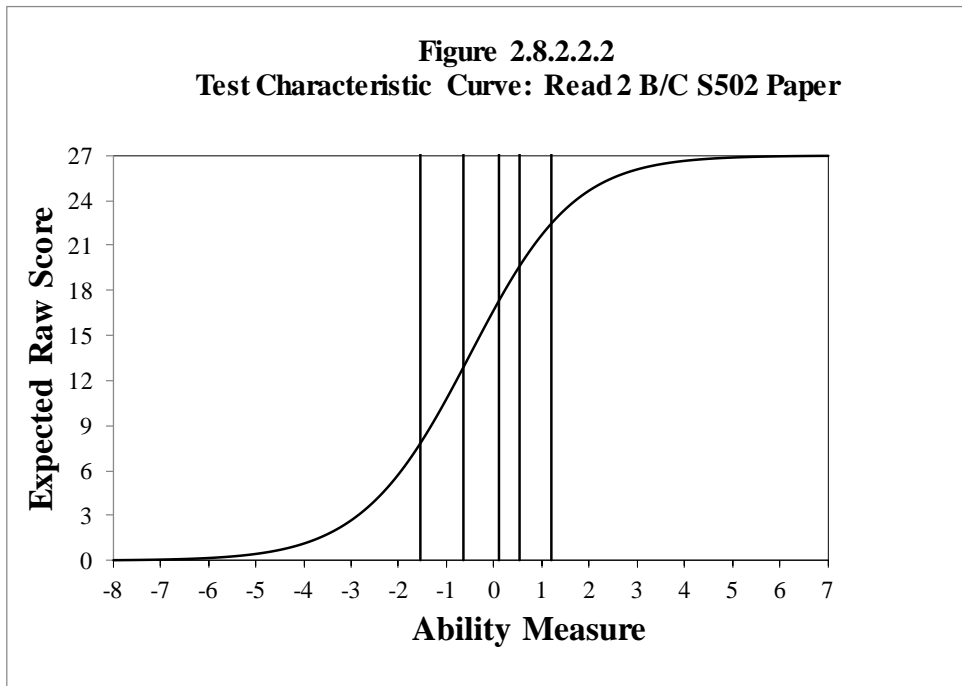


Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

2.8.2.2 Grade 2

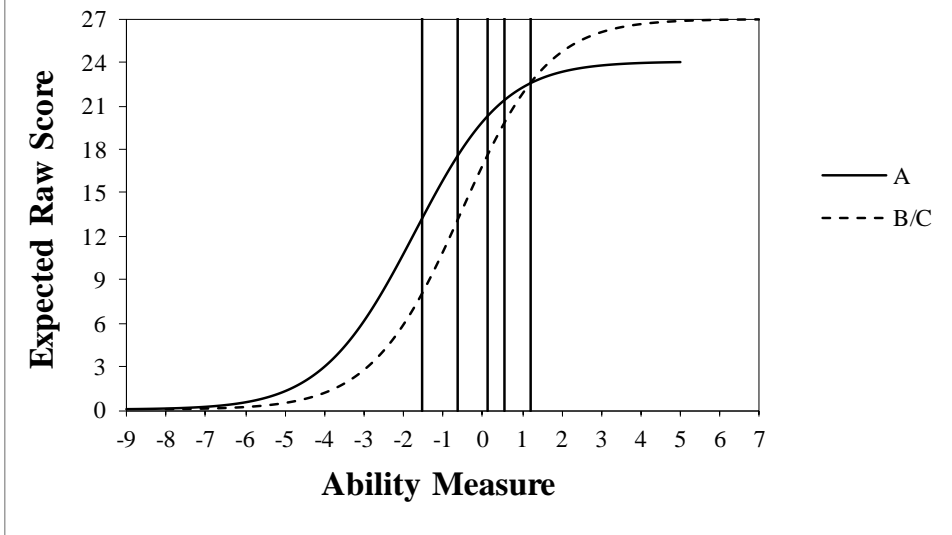


Note: The test form is shared between 1A and 2A.



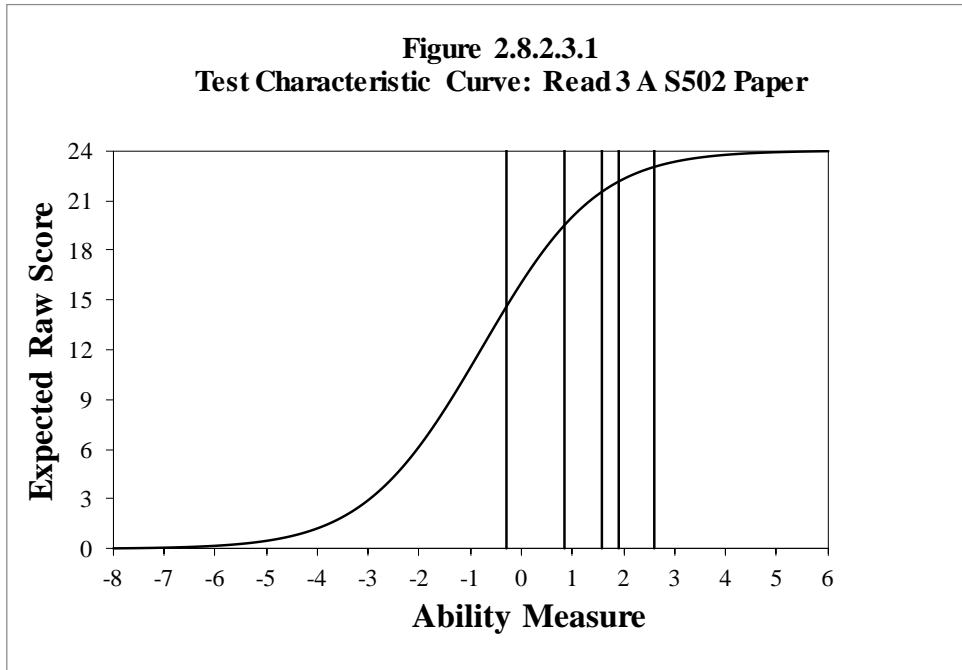
Note: The test form is shared between 1B/C and 2B/C.

Figure 2.8.2.2.3
Test Characteristic Curve: Read 2 S502 Paper

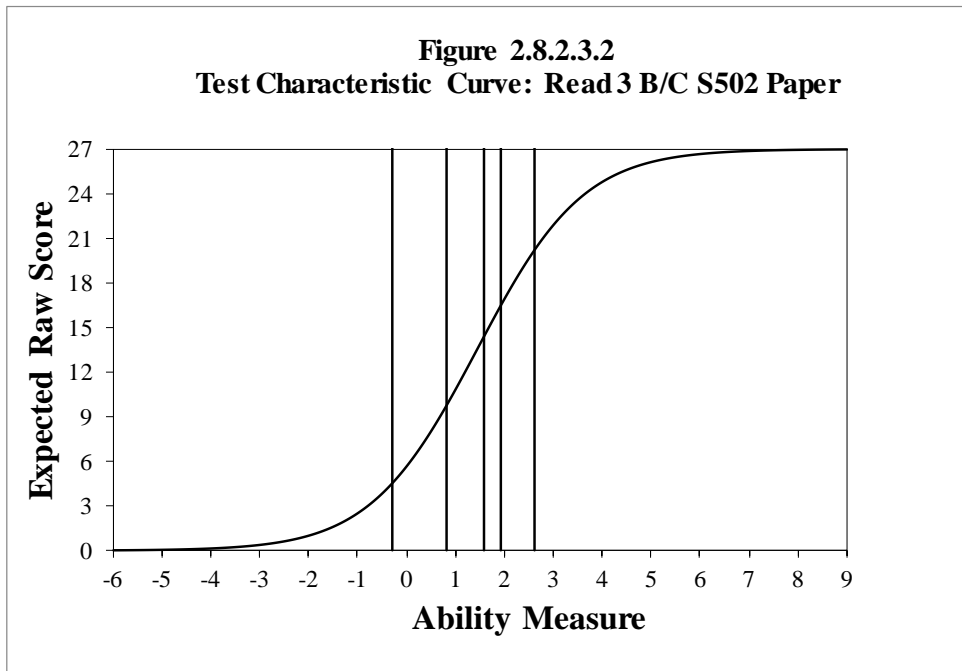


Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

2.8.2.3 Grade 3

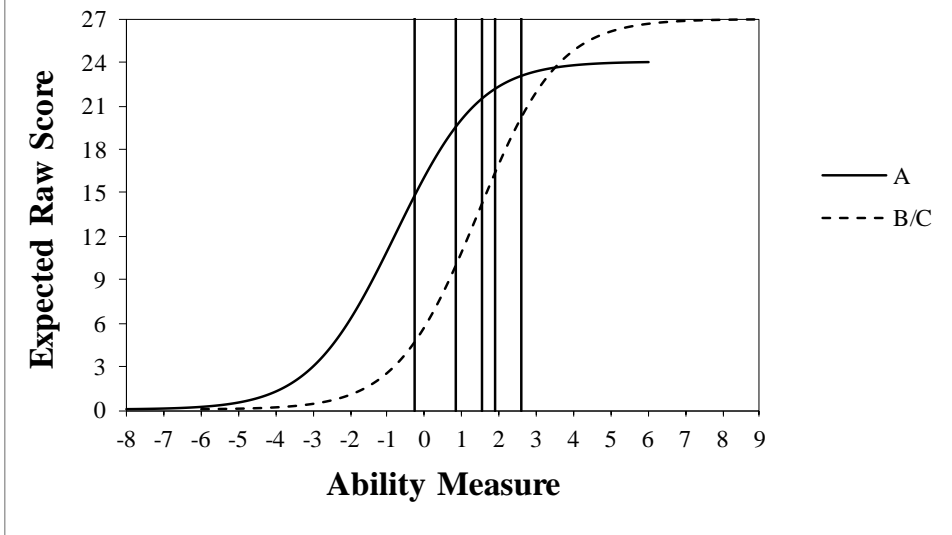


Note: The test form is shared between 3A and 4–5A.



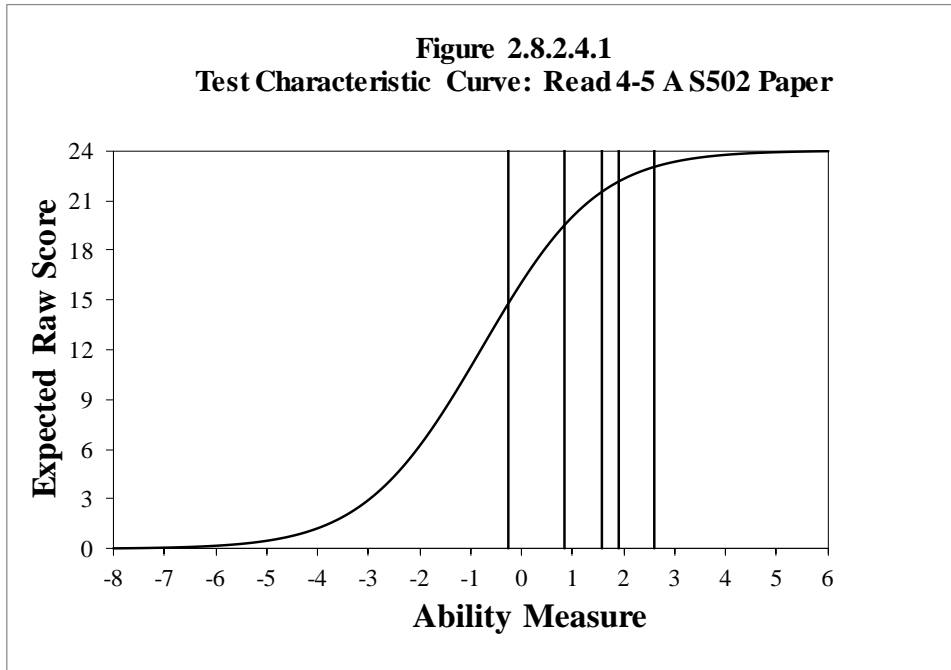
Note: The test form is shared between 3B/C and 4–5B/C.

Figure 2.8.2.3.3
Test Characteristic Curve: Read 3 S502 Paper

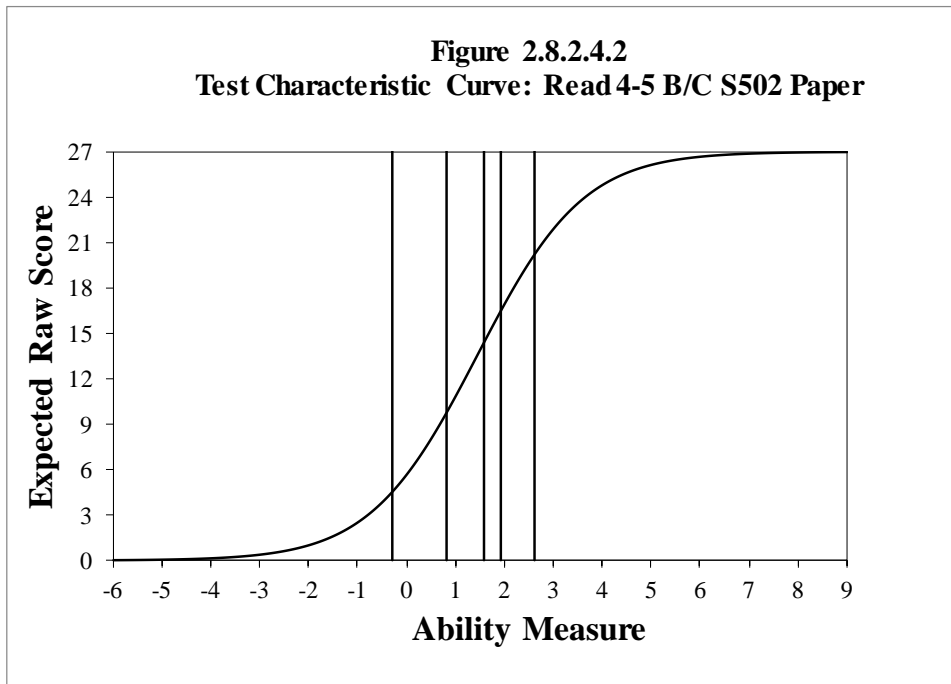


Note: The test form is shared between 3A and 4–5A, 3B/C and 4–5B/C.

2.8.2.4 Grades 4–5

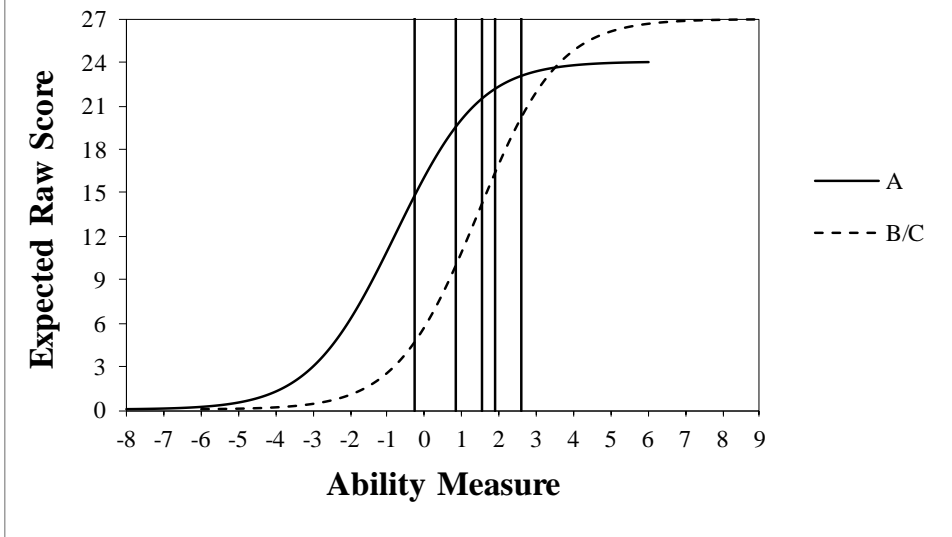


Note: The test form is shared between 3A and 4–5A.



Note: The test form is shared between 3B/C and 4–5B/C.

Figure 2.8.2.4.3
Test Characteristic Curve: Read 4-5 S502 Paper



Note: The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

2.8.2.5 Grades 6–8

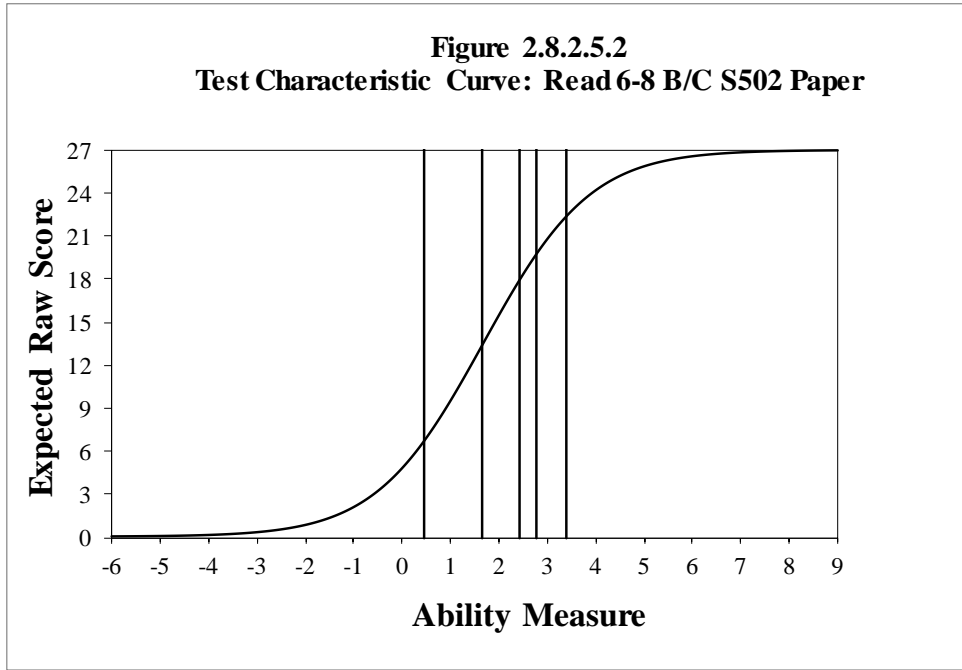
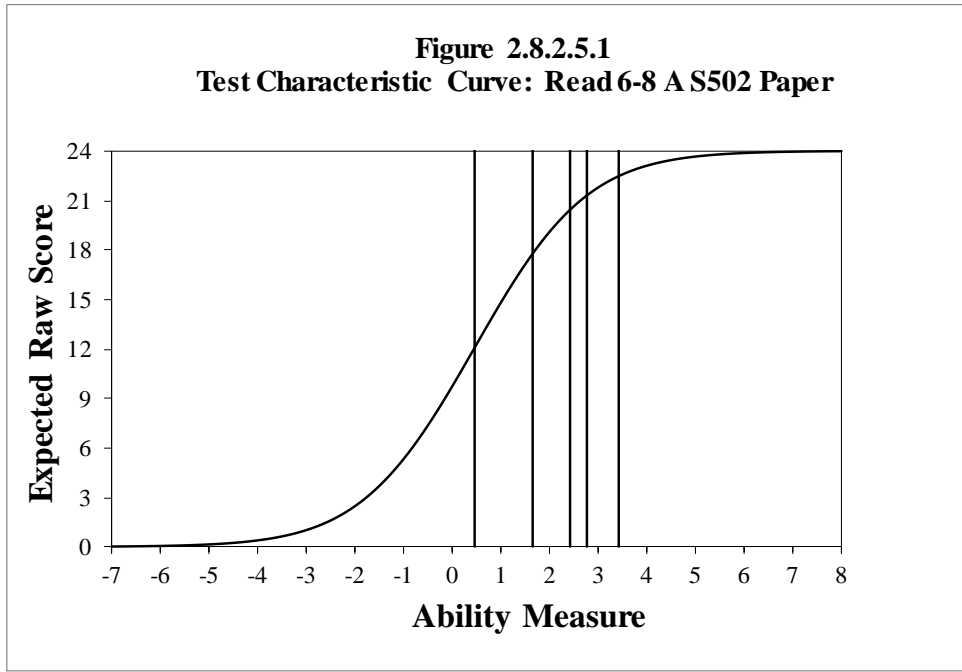
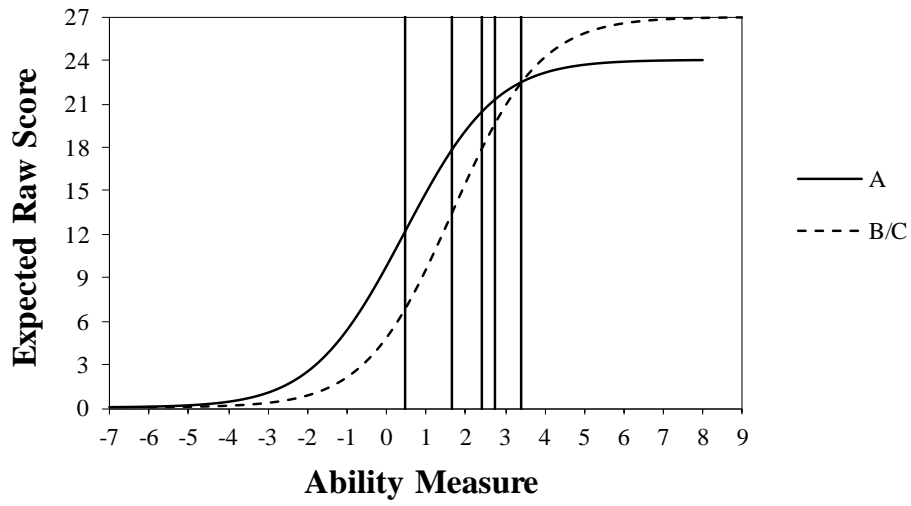


Figure 2.8.2.5.3
Test Characteristic Curve: Read 6-8 S502 Paper



2.8.2.6 Grades 9–12

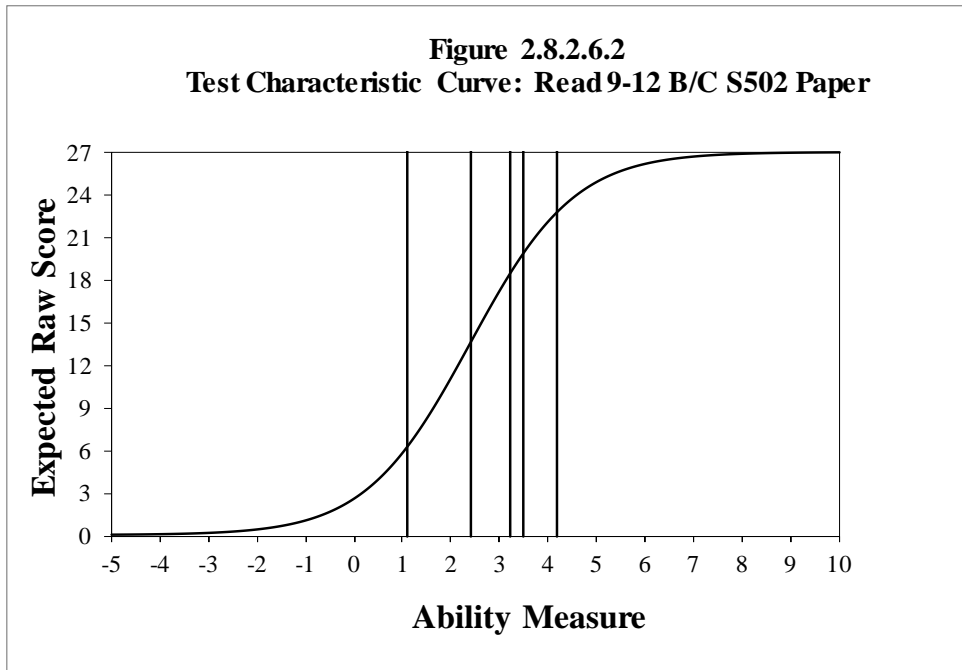
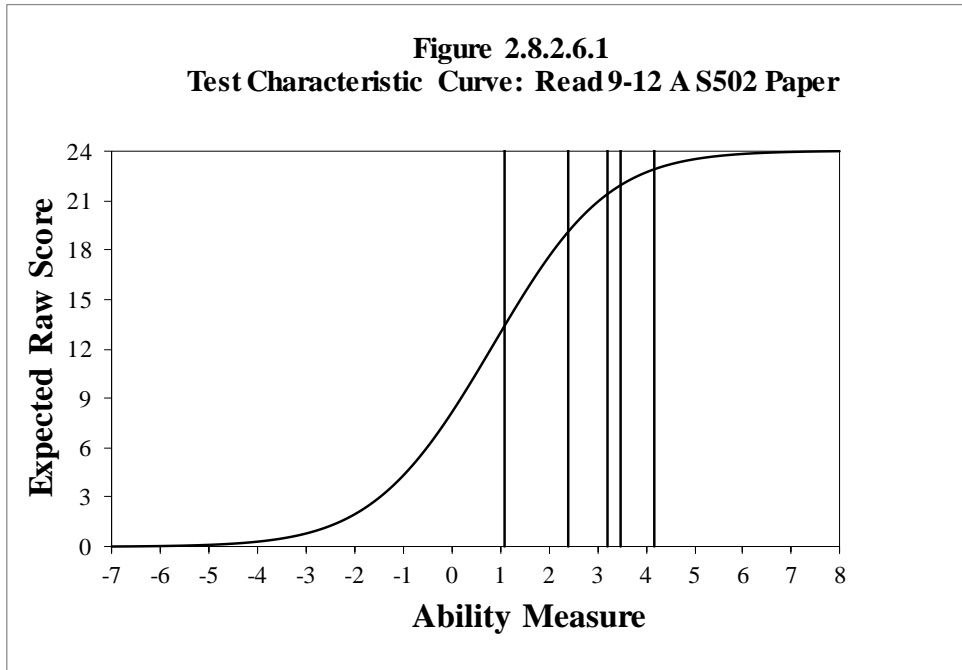
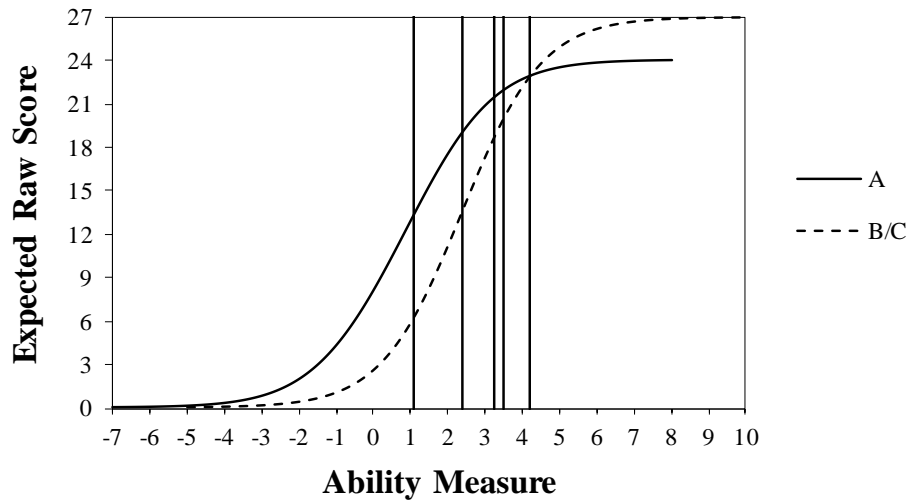
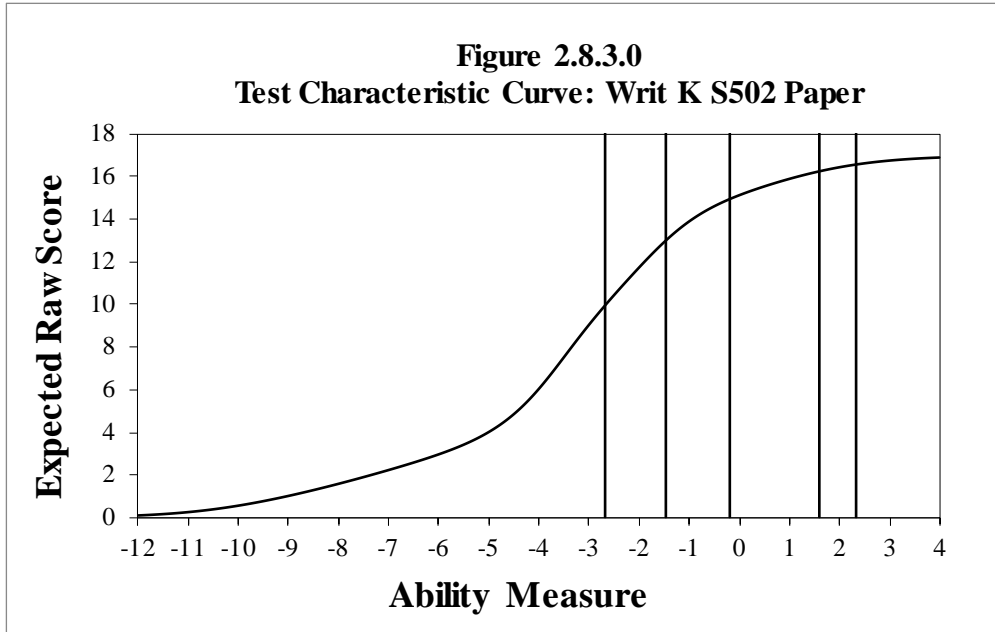


Figure 2.8.2.6.3
Test Characteristic Curve: Read 9-12 S502 Paper



2.8.3 Writing

2.8.3.0 Kindergarten



2.8.3.1 Grade 1

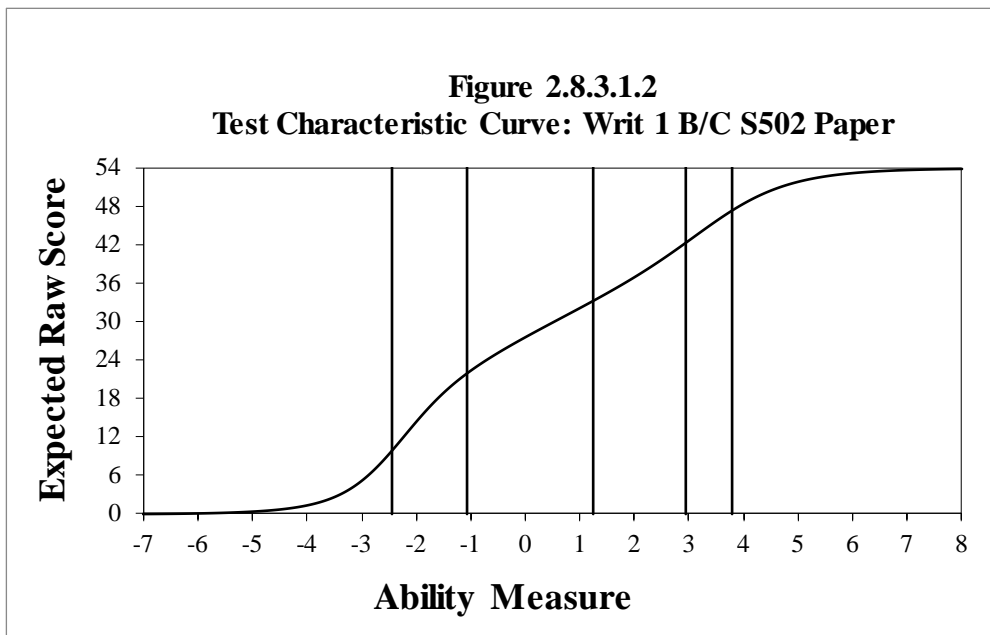
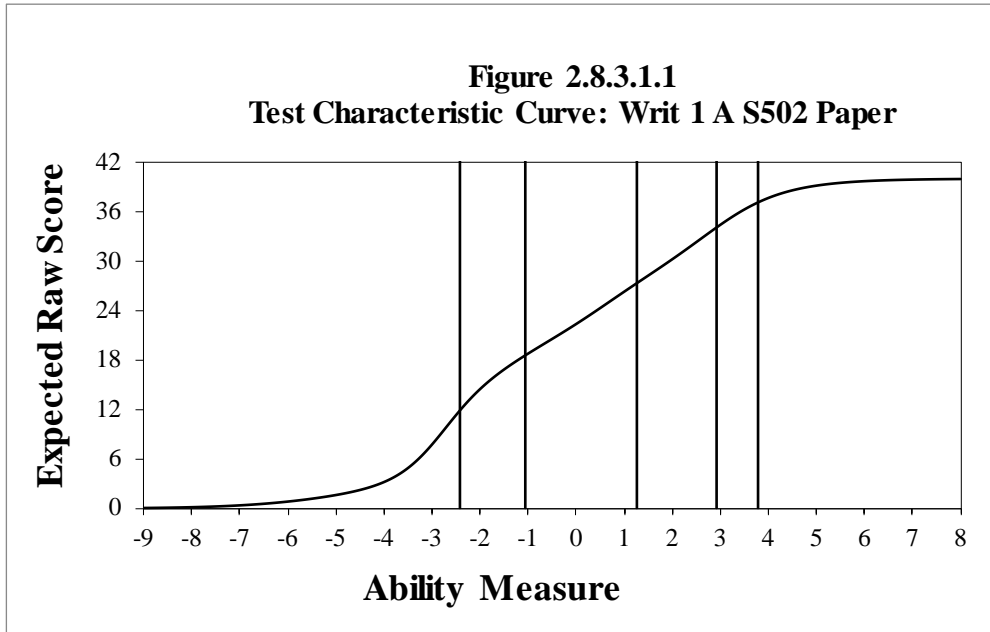
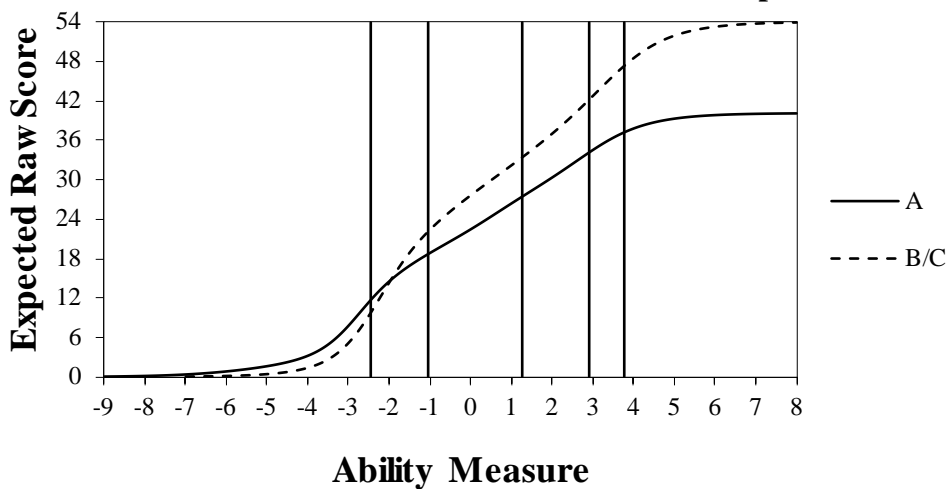
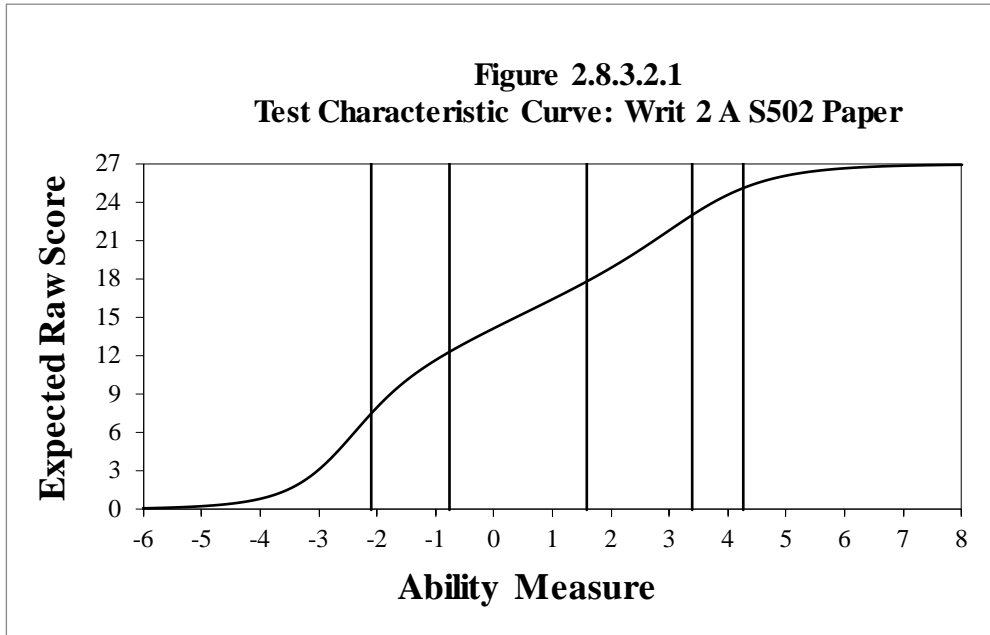


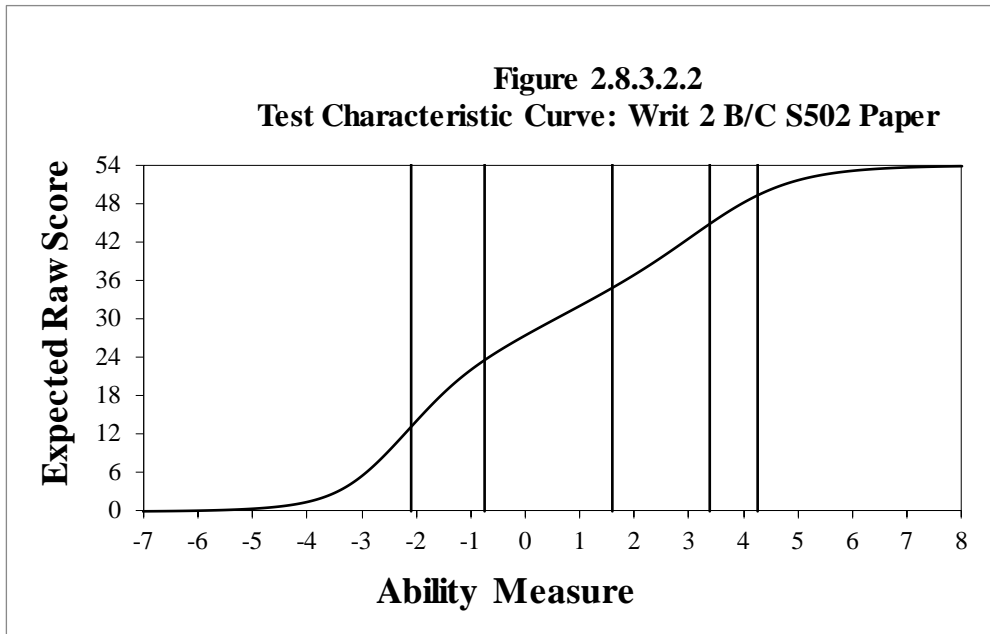
Figure 2.8.3.1.3
Test Characteristic Curve: Writ 1 S502 Paper



2.8.3.2 Grade 2

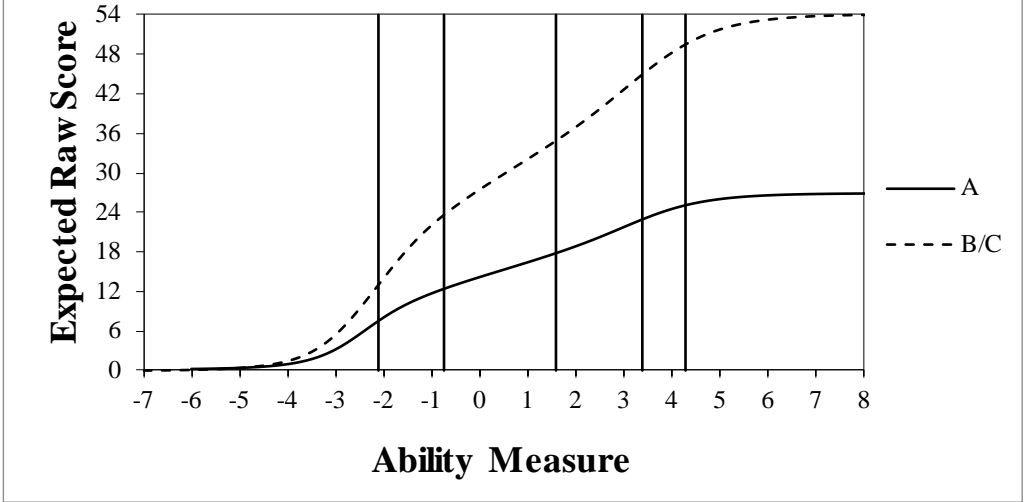


Note: The test form is shared between 2A and 3A.



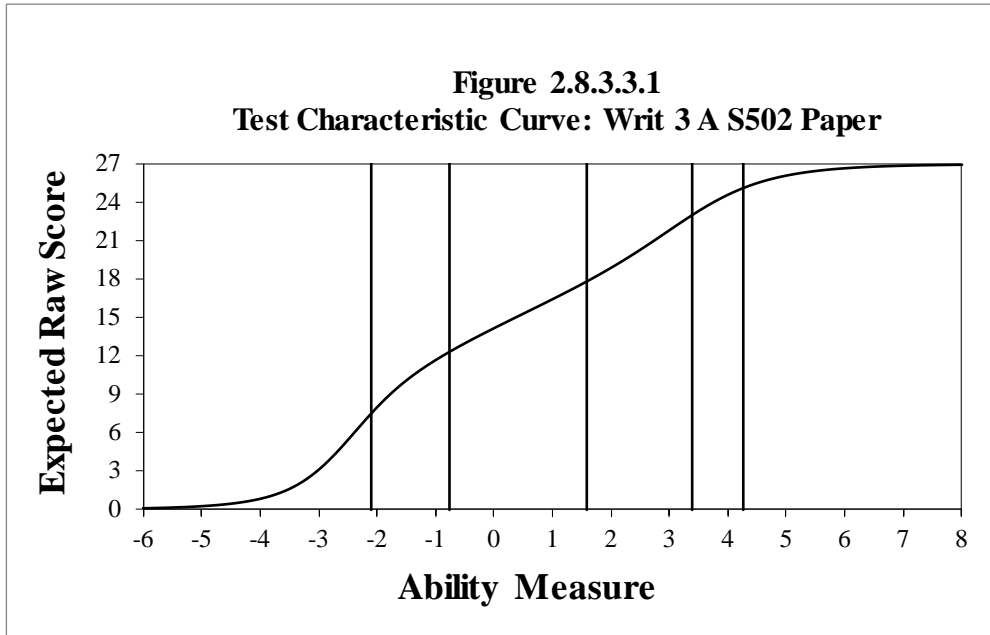
Note: The test form is shared between 2B/C and 3B/C.

Figure 2.8.3.2.3
Test Characteristic Curve: Writ 2 S502 Paper

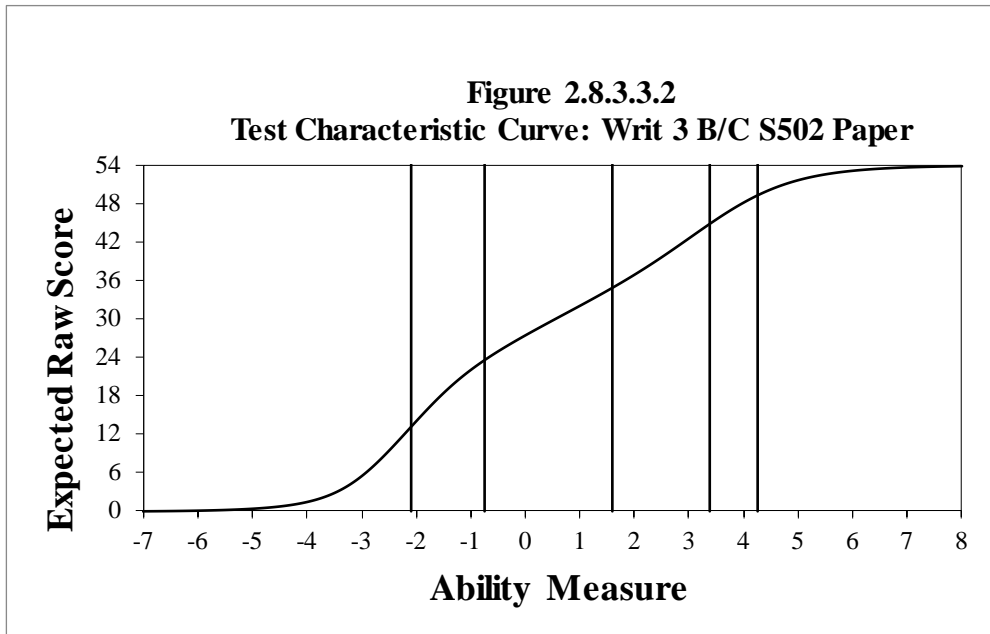


Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

2.8.3.3 Grade 3

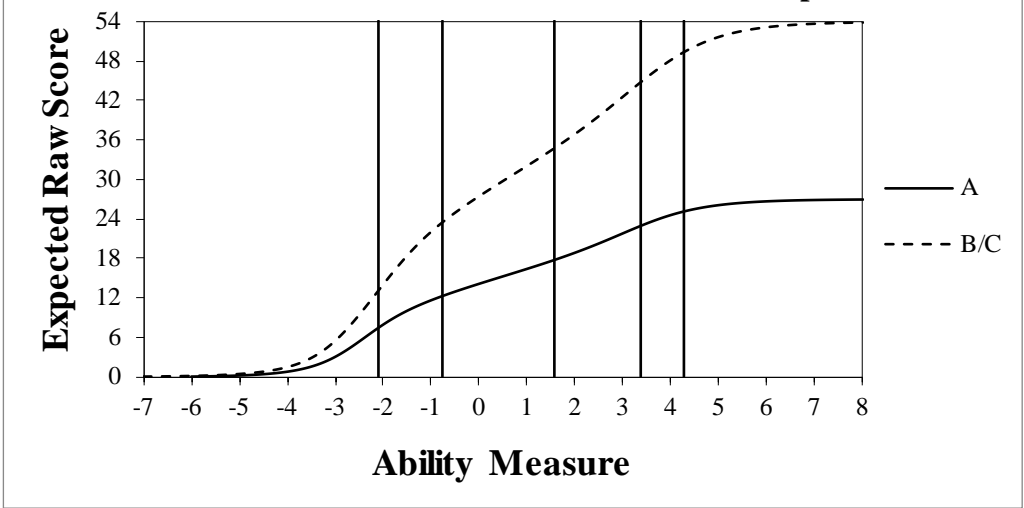


Note: The test form is shared between 2A and 3A.



Note: The test form is shared between 2B/C and 3B/C.

Figure 2.8.3.3.3
Test Characteristic Curve: Writ 3 S502 Paper



Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

2.8.3.4 Grades 4–5

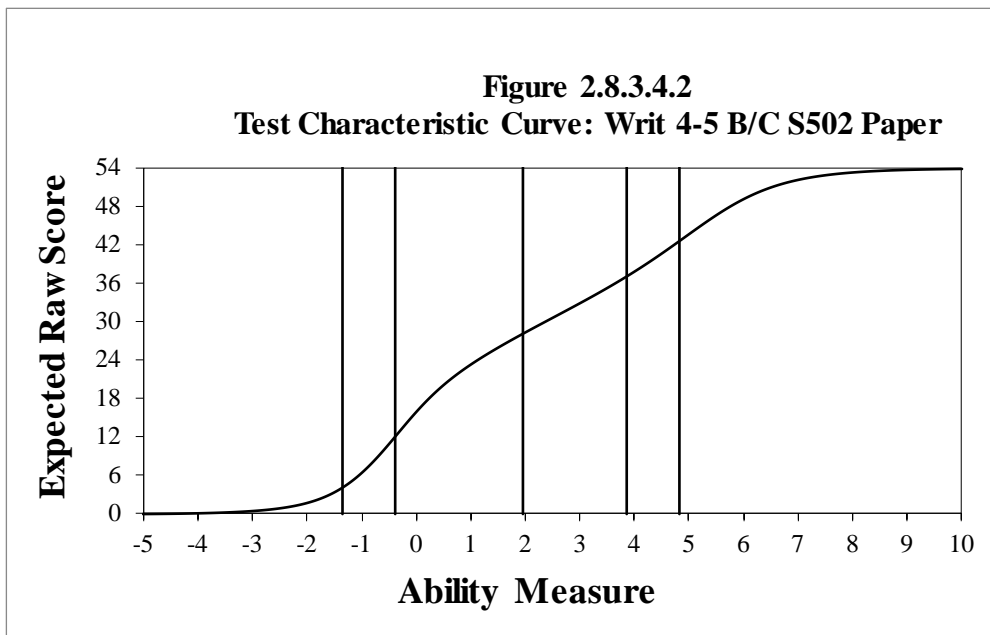
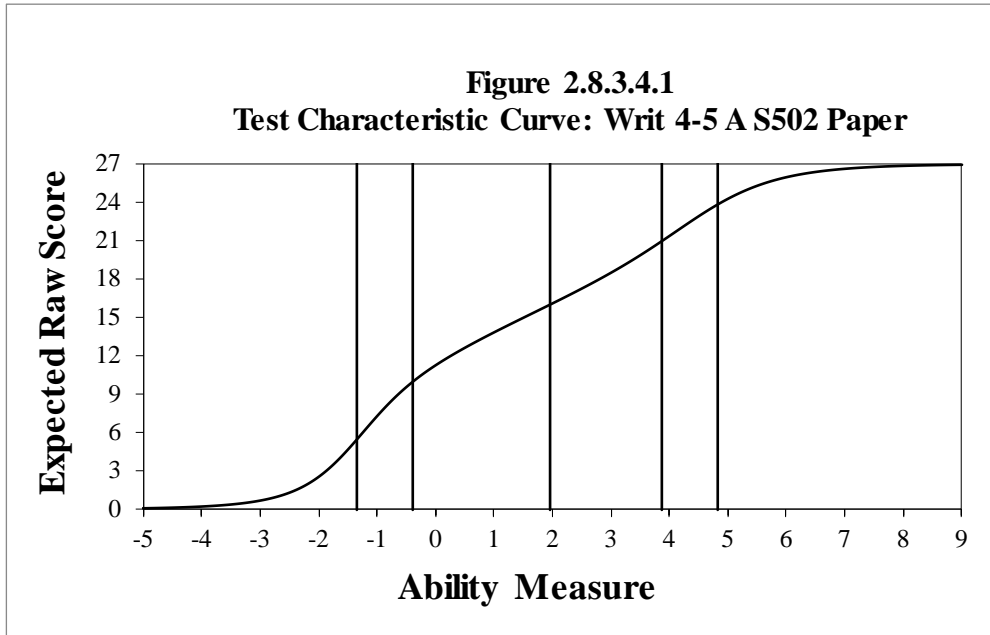
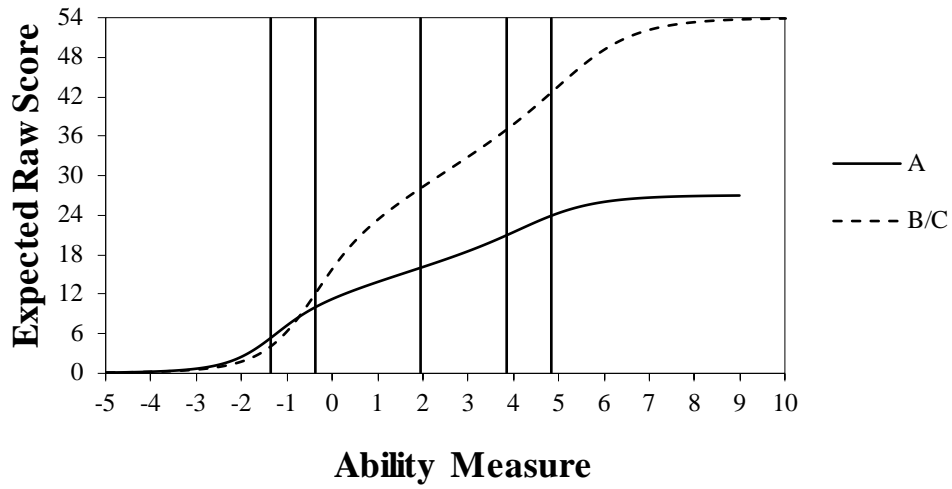


Figure 2.8.3.4.3
Test Characteristic Curve: Writ 4-5 S502 Paper



2.8.3.5 Grades 6–8

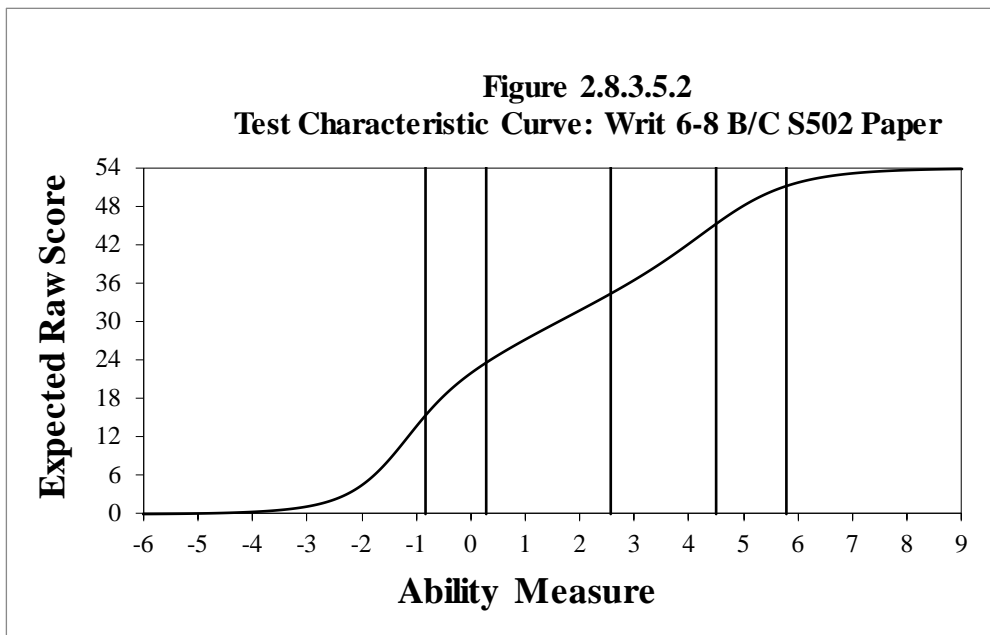
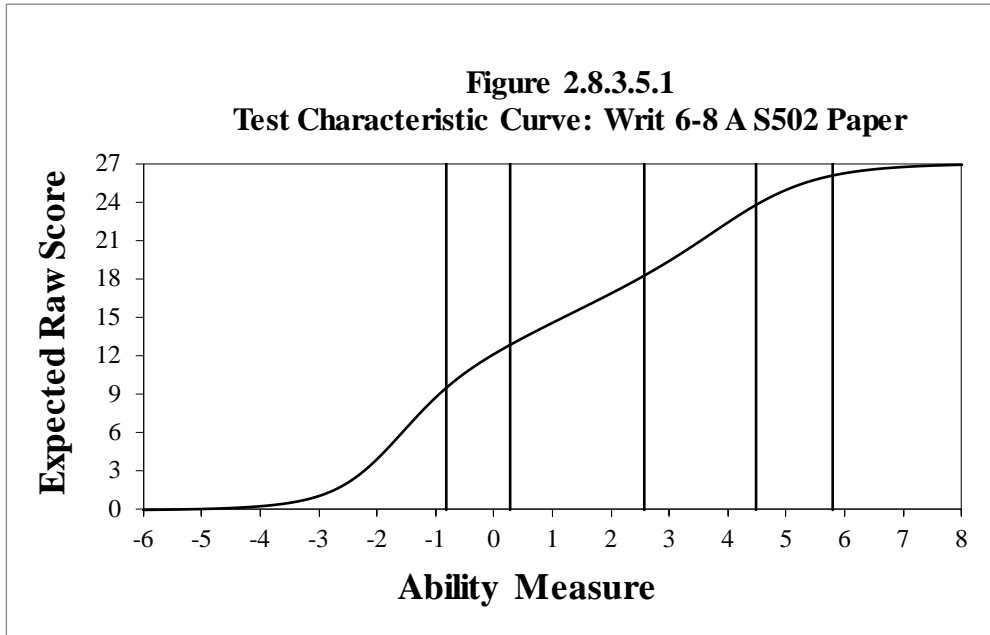
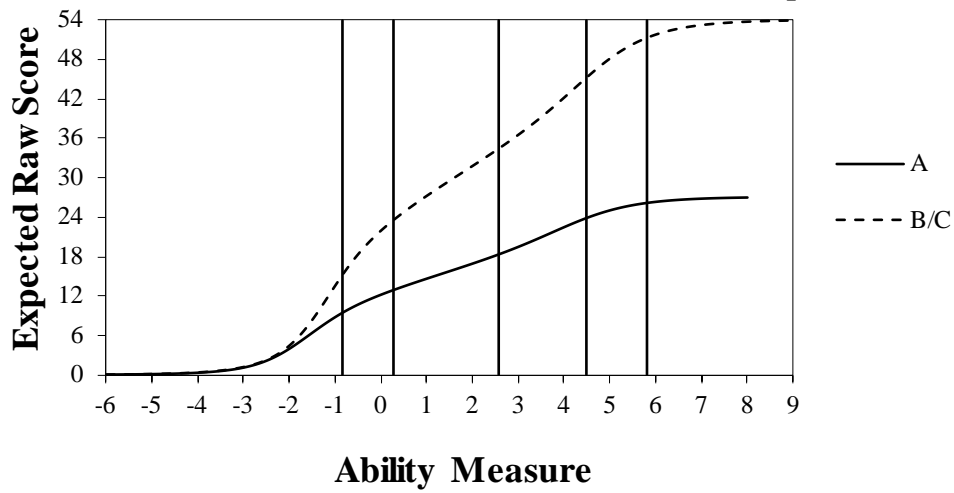


Figure 2.8.3.5.3
Test Characteristic Curve: Writ 6-8 S502 Paper



2.8.3.6 Grades 9–12

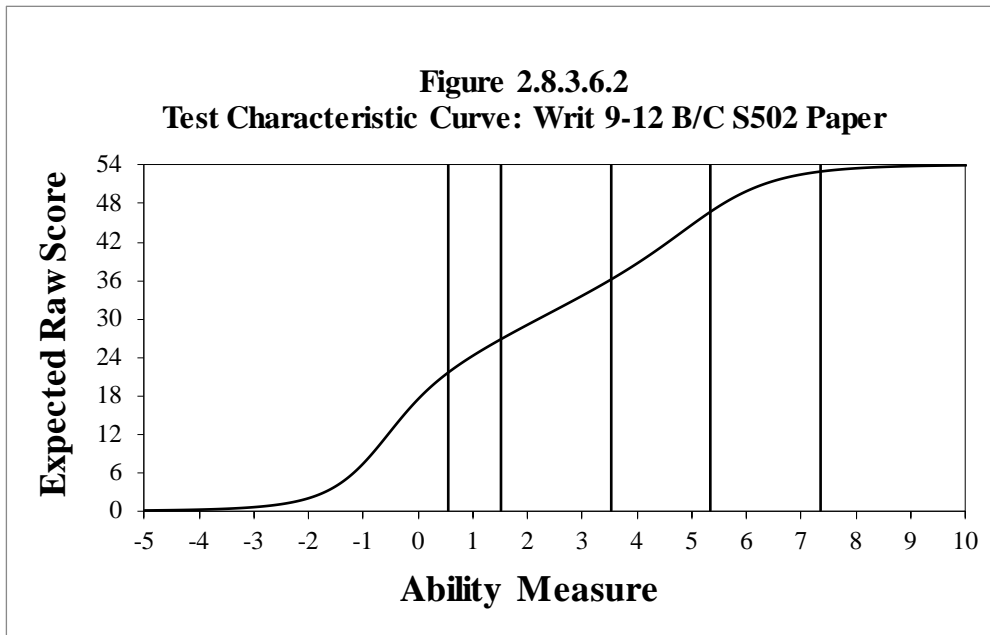
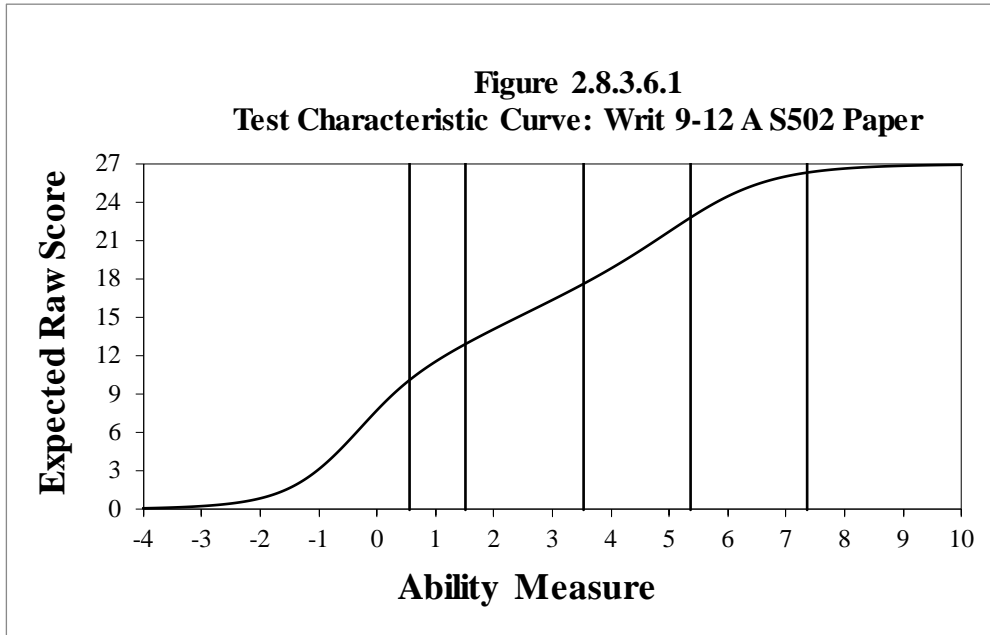
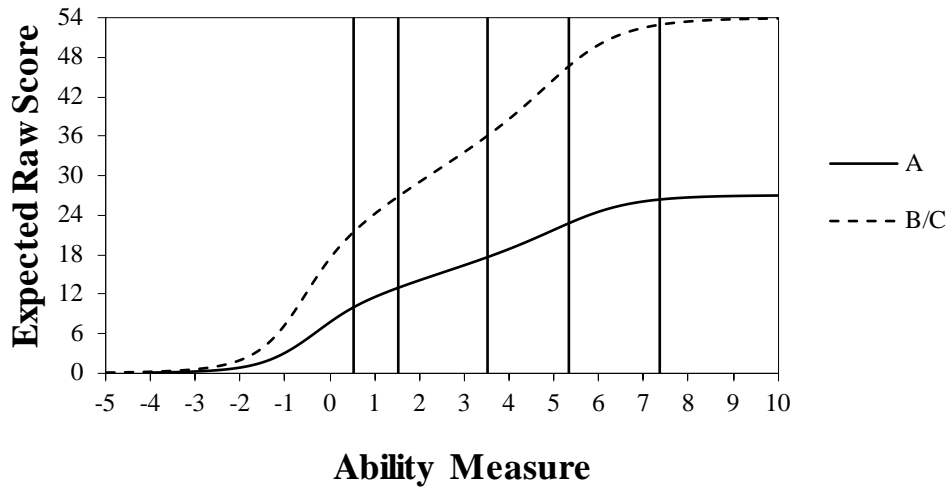
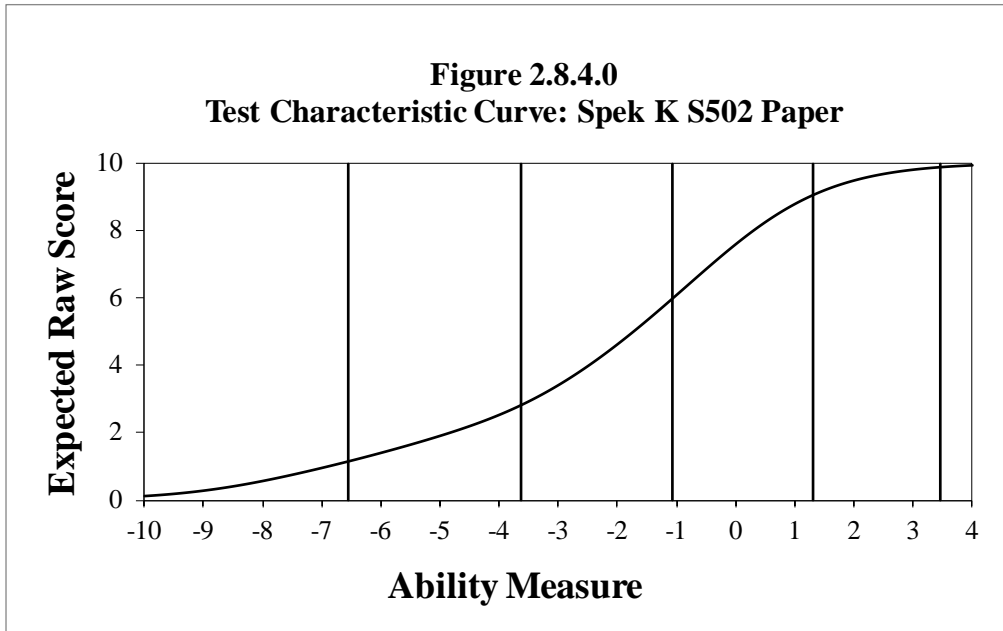


Figure 2.8.3.6.3
Test Characteristic Curve: Writ 9-12 S502 Paper



2.8.4 Speaking

2.8.4.0 Kindergarten



2.8.4.1 Grade 1

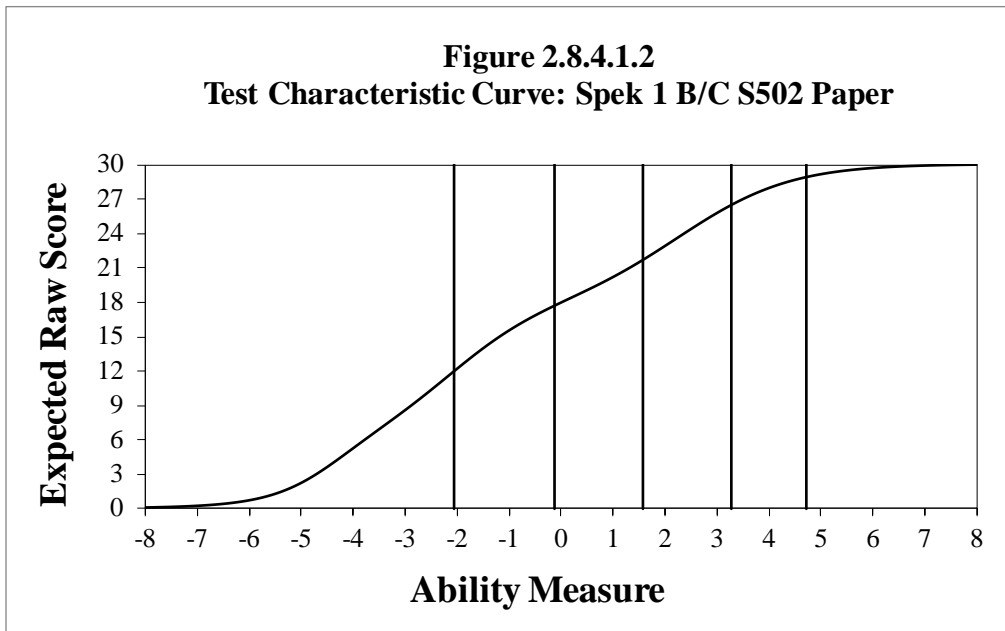
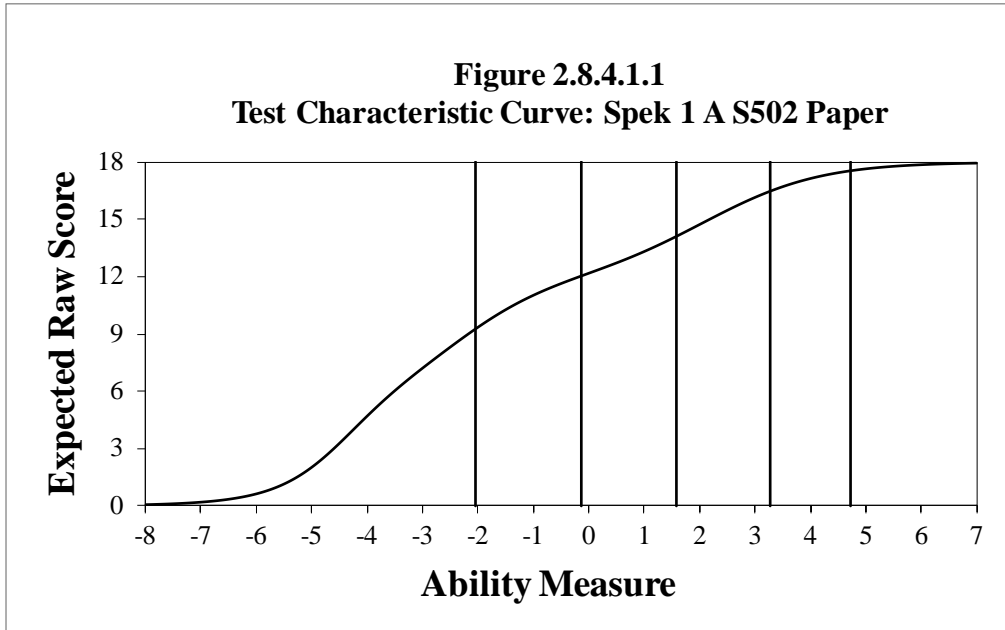
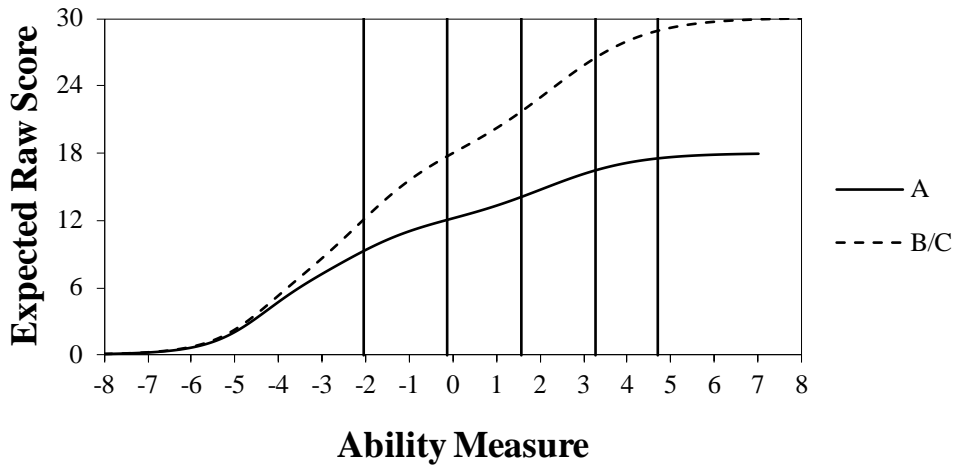
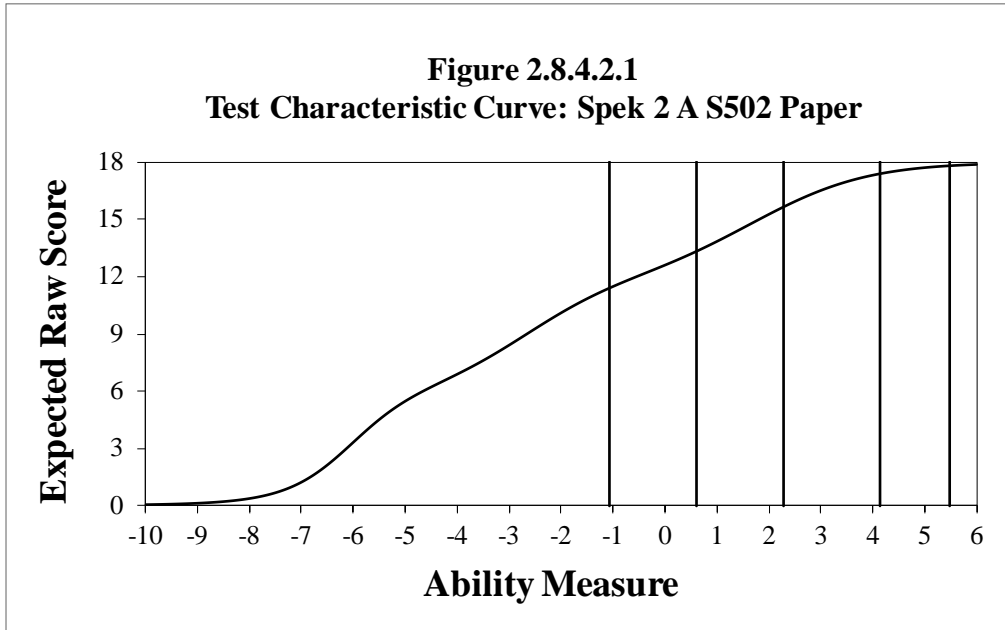


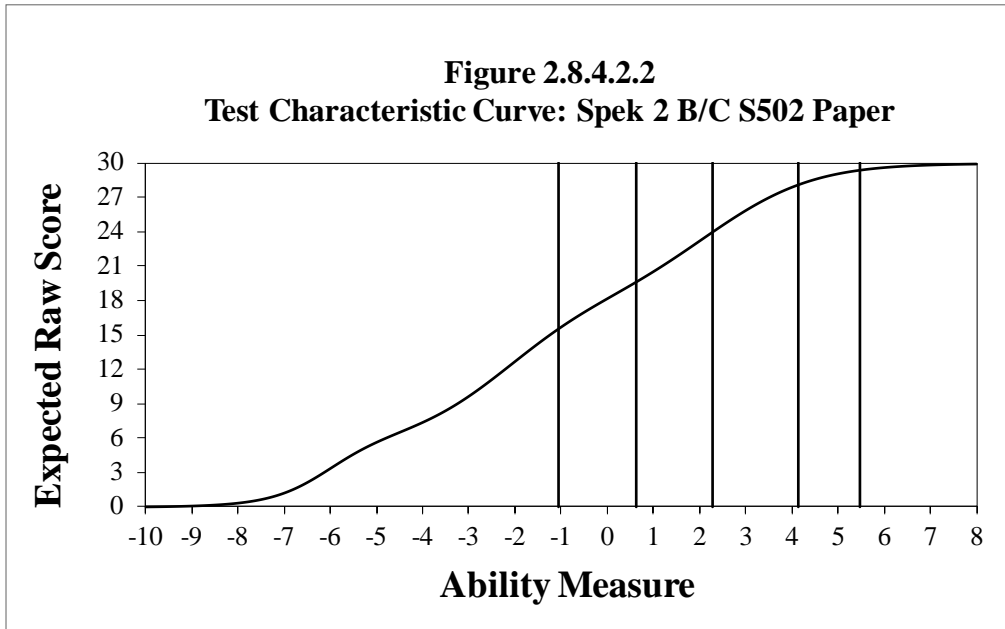
Figure 2.8.4.1.3
Test Characteristic Curve: Spek 1 S502 Paper



2.8.4.2 Grade 2

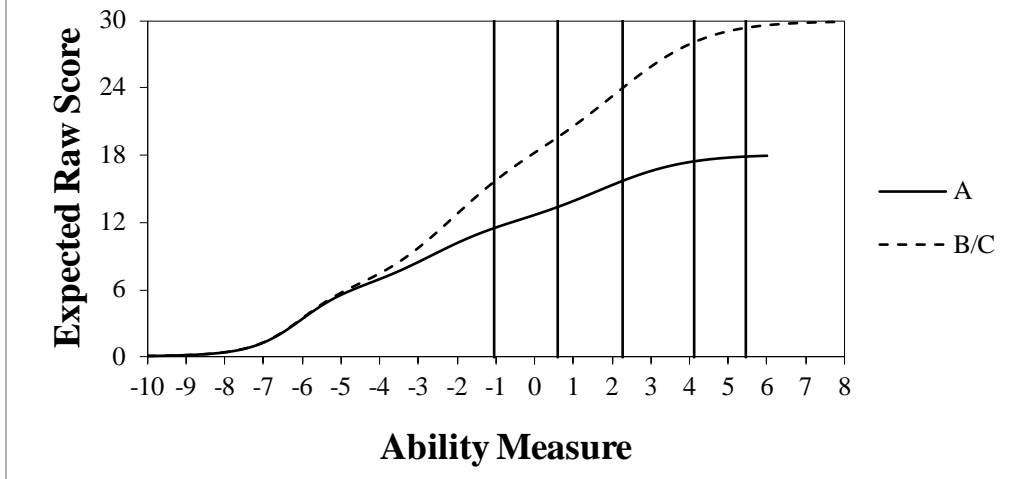


Note: The test form is shared between 2A and 3A.



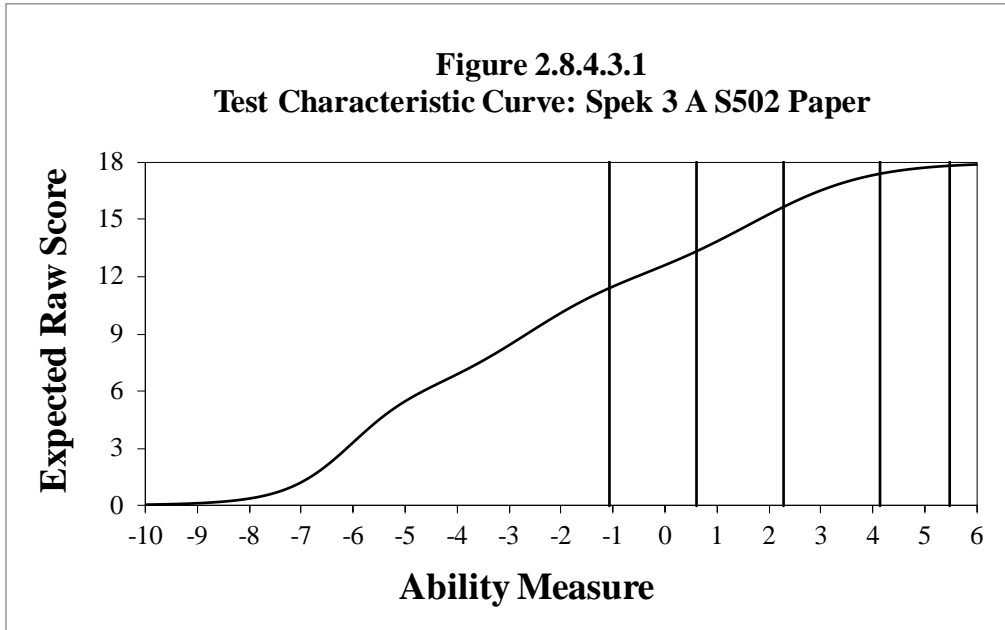
Note: The test form is shared between 2B/C and 3B/C.

Figure 2.8.4.2.3
Test Characteristic Curve: Spek 2 S502 Paper

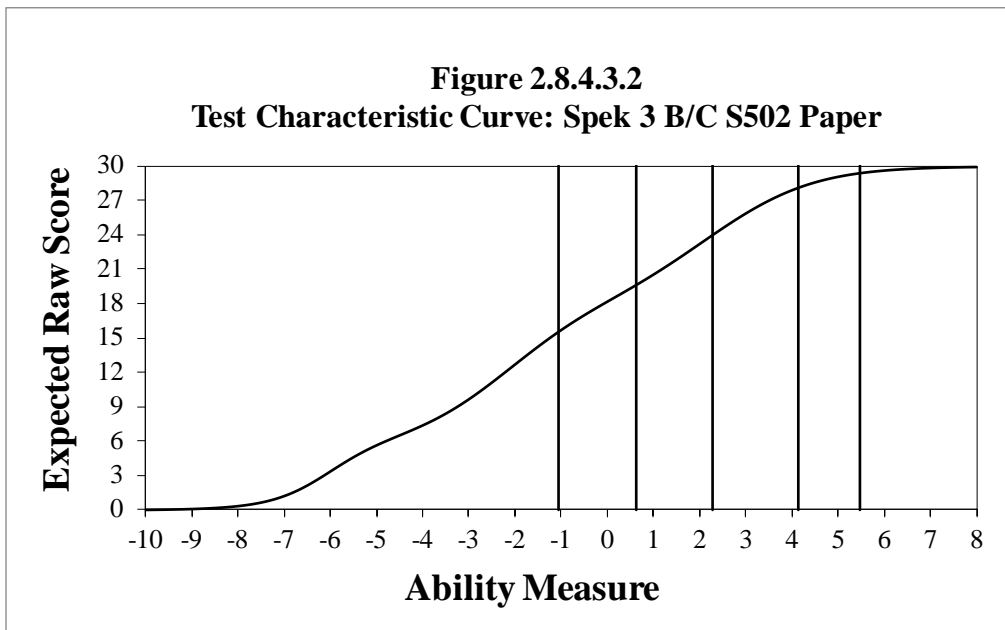


Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

2.8.4.3 Grade 3

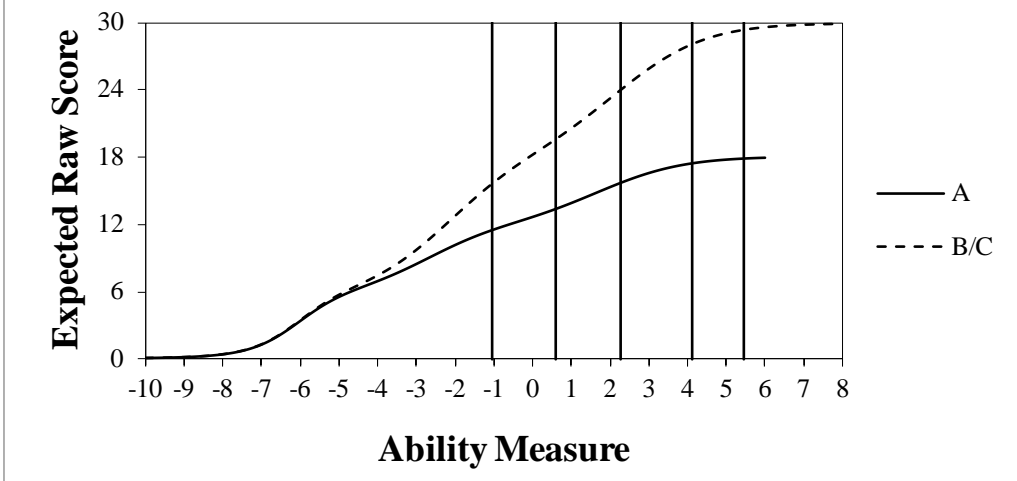


Note: The test form is shared between 2A and 3A.



Note: The test form is shared between 2B/C and 3B/C.

Figure 2.8.4.3.3
Test Characteristic Curve: Spek 3 S502 Paper



Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

2.8.4.4 Grades 4–5

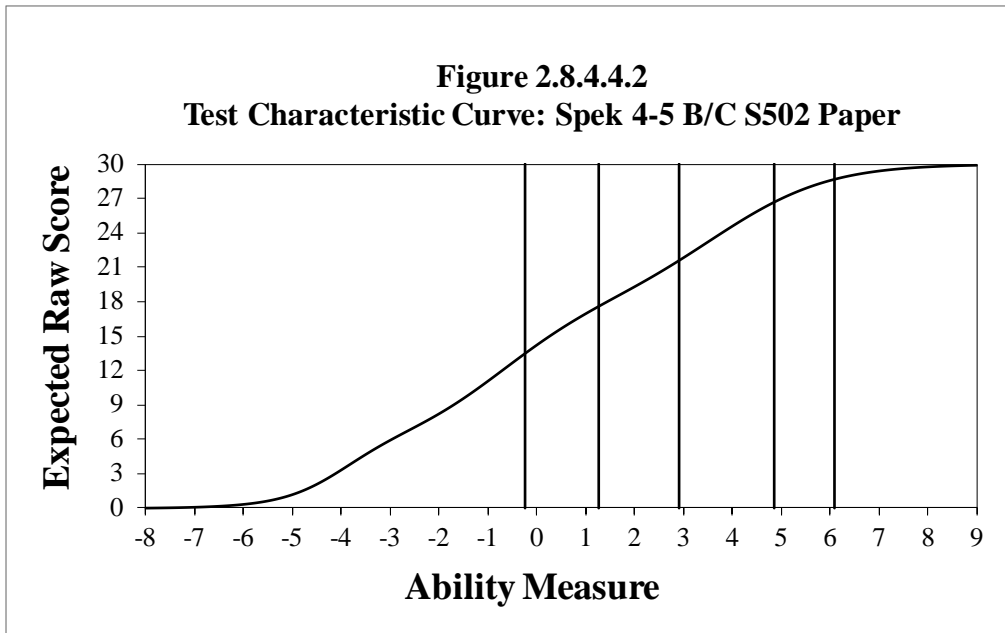
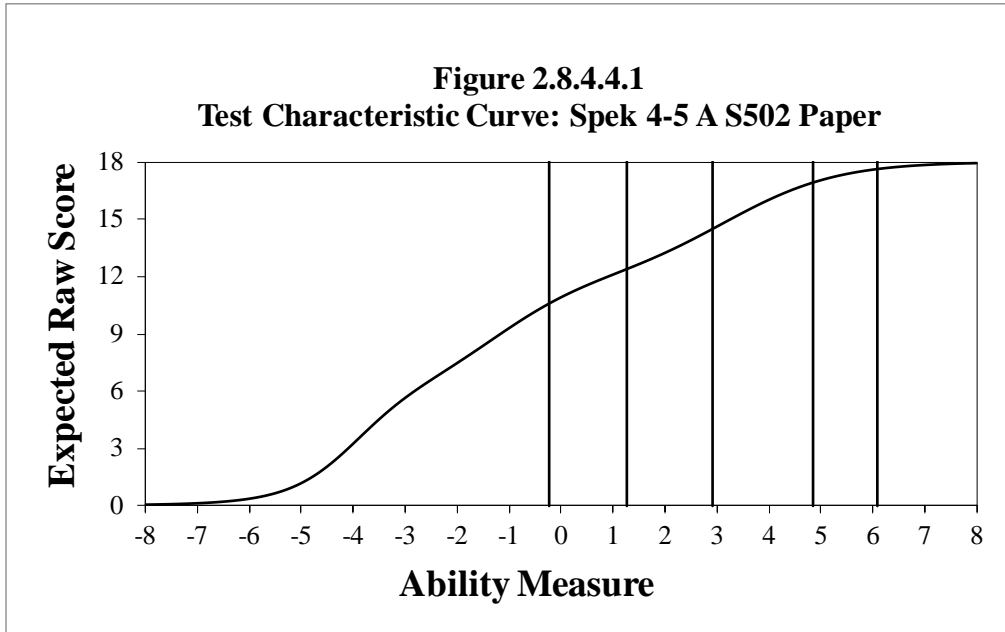
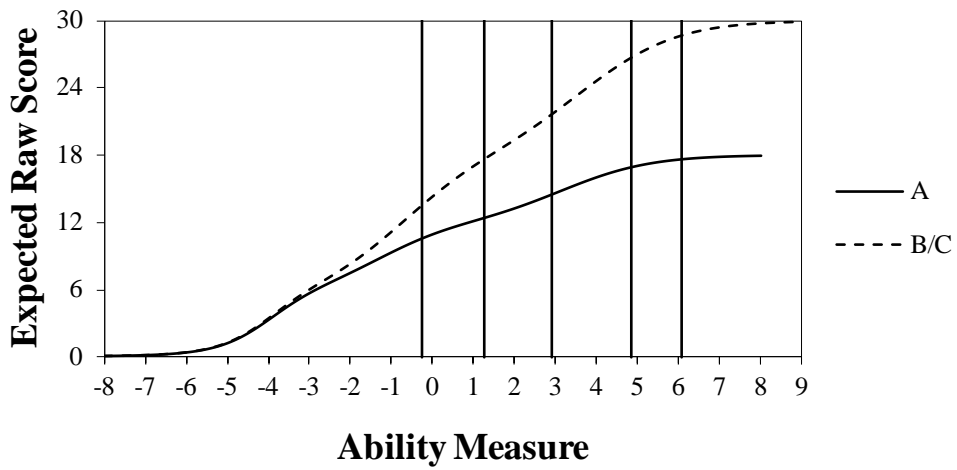


Figure 2.8.4.4.3
Test Characteristic Curve: Spek 4-5 S502 Paper



2.8.4.5 Grades 6–8

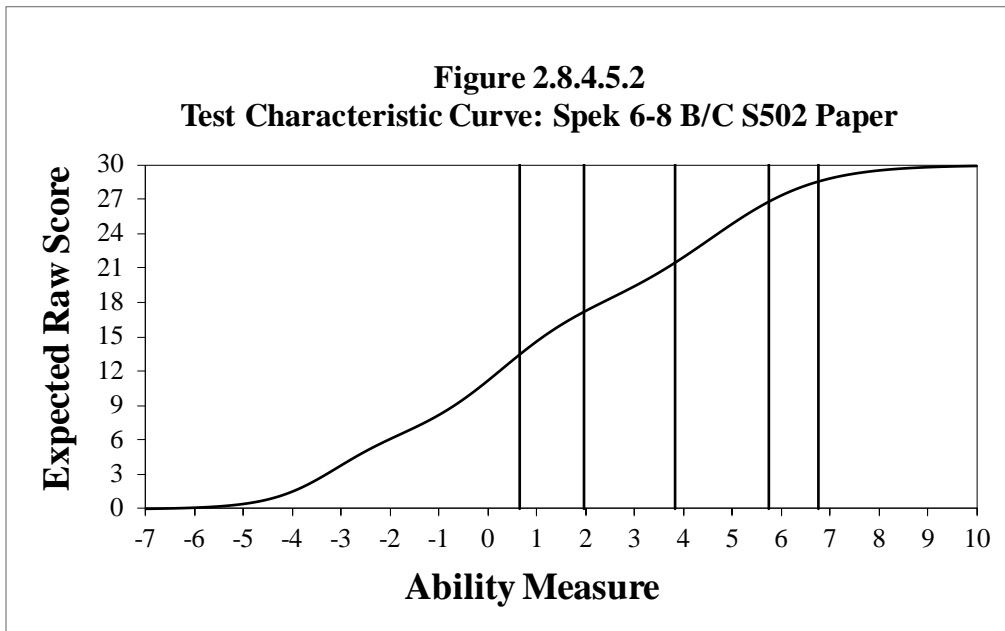
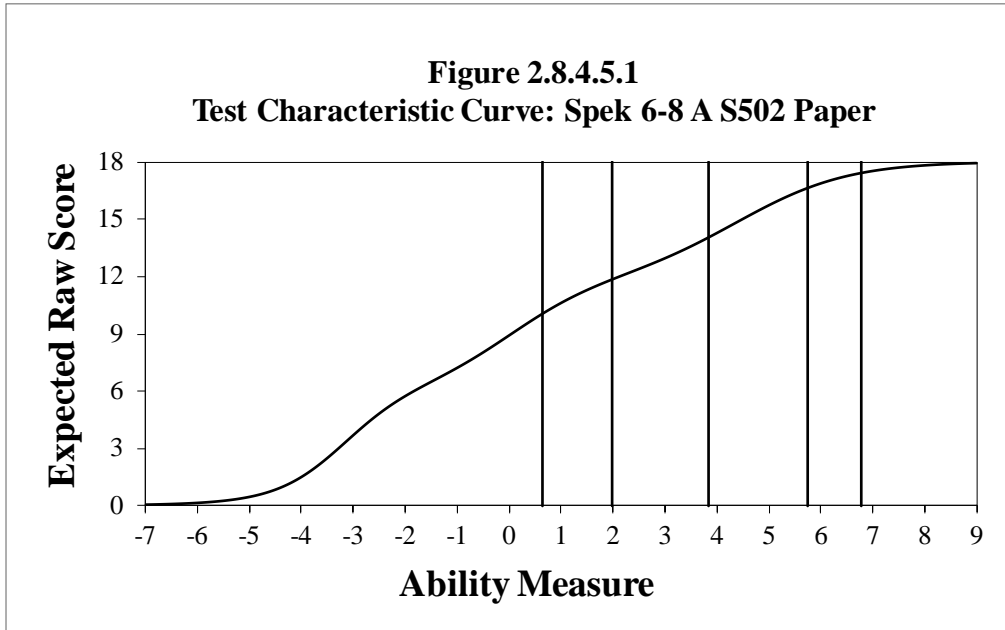
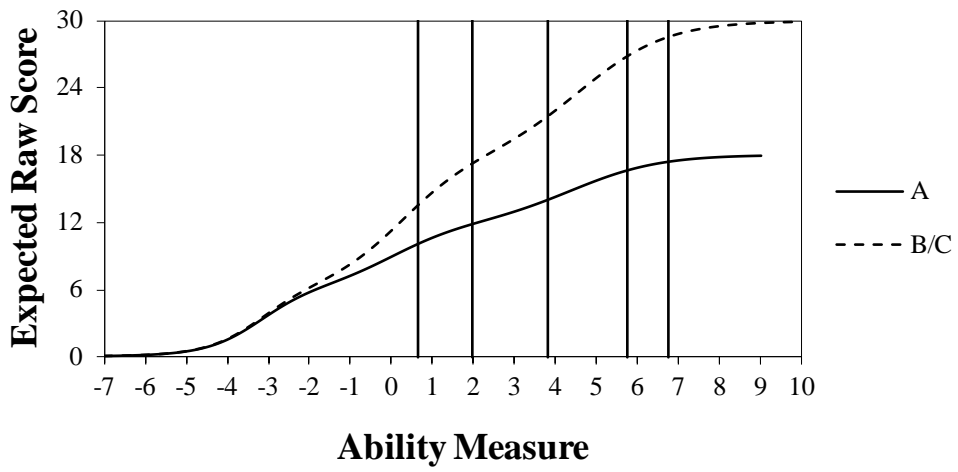


Figure 2.8.4.5.3
Test Characteristic Curve: Spek 6-8 S502 Paper



2.8.4.6 Grades 9–12

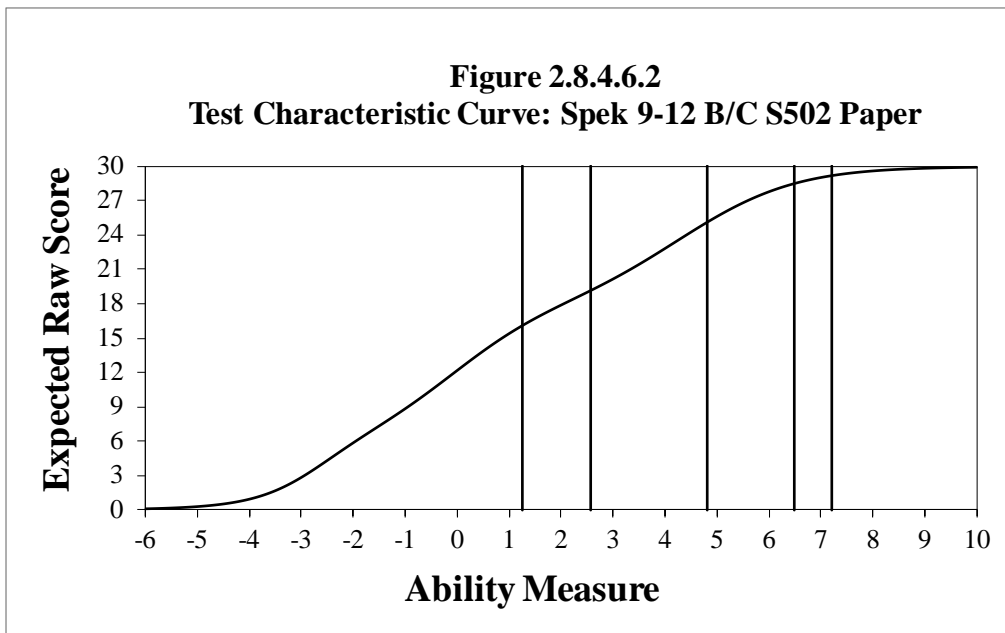
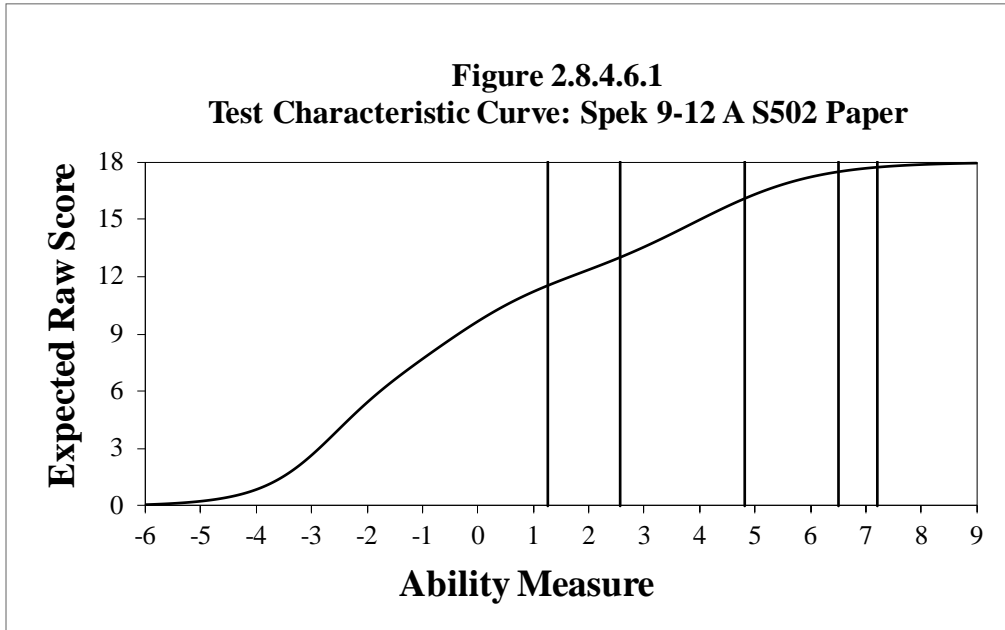
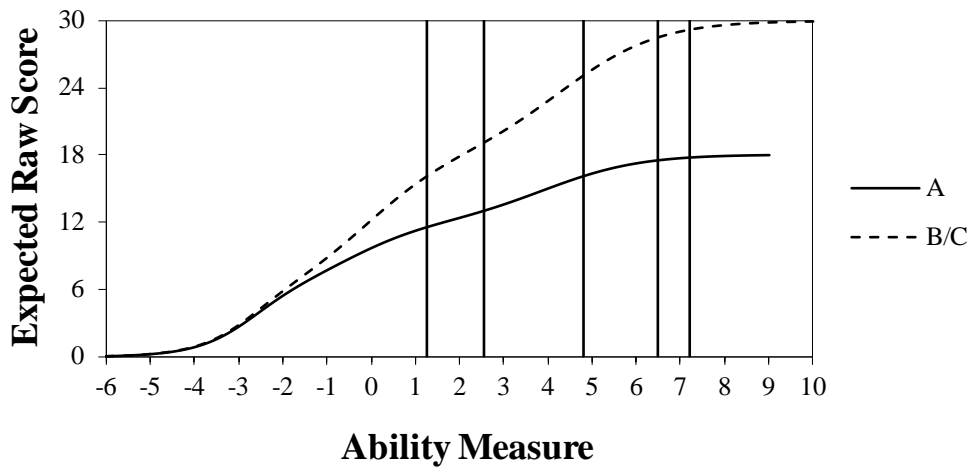


Figure 2.8.4.6.3
Test Characteristic Curve: Spek 9-12 S502 Paper



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 reasonably 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 reasonably 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 characteristic functions on the test form (Lord, 1980), the TIF depends on the item information functions (Lord, 1980) of the items 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, the item response theory model used, and the values of the

item parameters. With some exceptions, there is a general pattern to the shape of test information curves. Test information curves peak at the area where the test provides higher discrimination and better measurement as compared to other areas 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 is usually unimodal. The values of the item category parameters, in addition to factors mentioned earlier, affect the shape of the information curves for Writing and Speaking tests, which consist of polytomous tasks. Since polytomous tasks have more score categories than multiple-choice items and measure a wider range of values on the proficiency scale, adjacent category boundaries are sometimes far apart as a result. In this situation, a test information curve will have a dip in the area between the adjacent category boundaries indicating the loss of information in this ability range. Therefore, the shape of a test information curve for ACCESS Writing and Speaking tests may not be unimodal and instead may have one or more peaks. This is consistent with other tests with polytomous items, such as the National Assessment of Educational Progress Writing assessment (Muraki, 1993).

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.

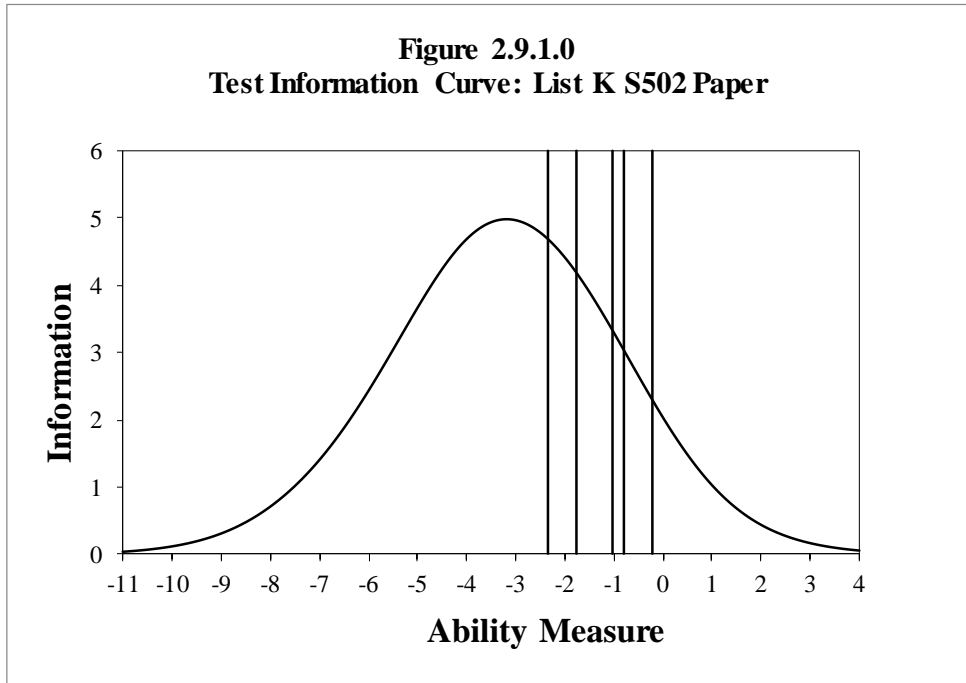
The parameter values for the individual categories on the scoring tools that raters use to evaluate students' responses to the tasks, in addition to the factors mentioned earlier, affect the shape of the test information curves for the Writing and Speaking tests. Accordingly, some refer to these test information curves as "category information functions" (Engelhard & Wind, 2018). The rating 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 rating 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 scale to assign scores (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).

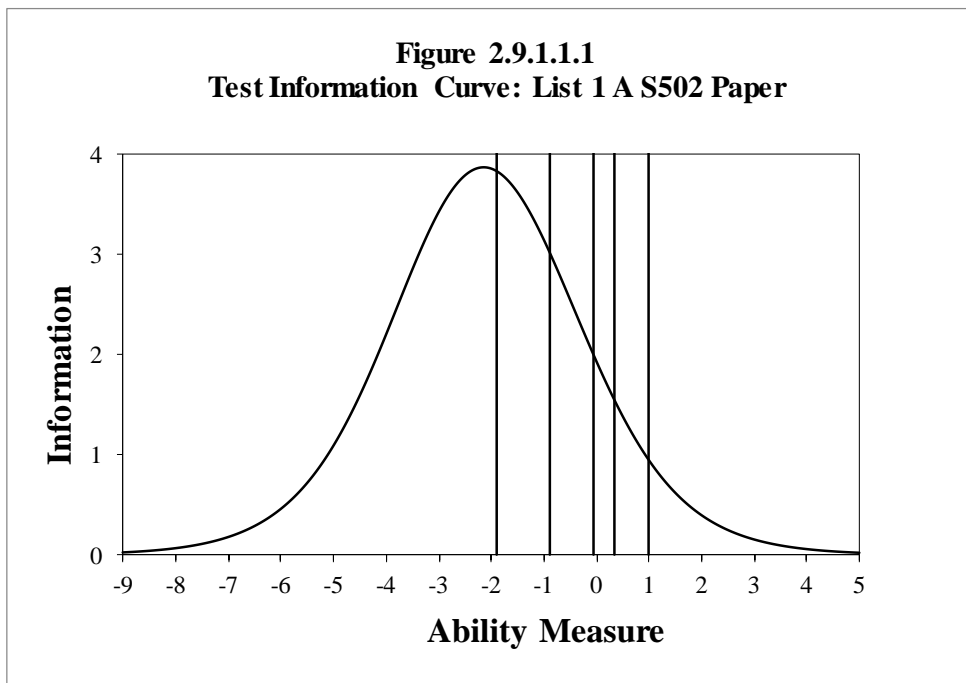
In addition to the TIF graphs by tier, we provide plots of the TIFs across tiers, by grade cluster, on the same graph. It is informative to compare the ability ranges where the curves peak (where the best measurement information is provided) across tiers. For example, the TIF across tiers for Listening Grade 1 shows that the Listening Grade 1 Tier A form provides the most information right below PL 2, while the Listening Grade 1 Tier B/C form provides the most information at a higher proficiency level (right below PL 3), as expected. In addition, the plot shows that the Listening Grade 1 Tier A form provides more information than the B/C form before the PL 2 cut, while the B/C form provides more information than the Tier A form after the PL 2 cut.

2.9.1 Listening

2.9.1.0 Kindergarten

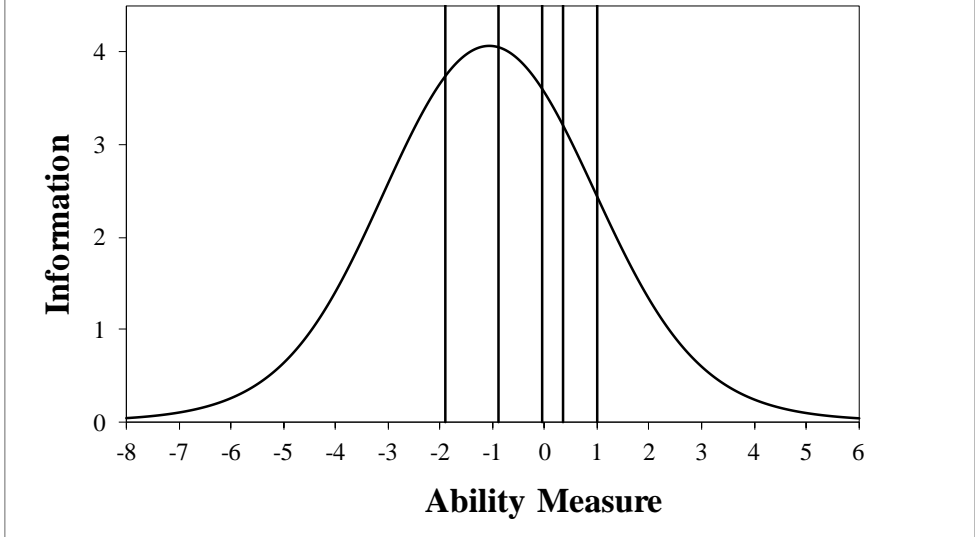


2.9.1.1 Grade 1



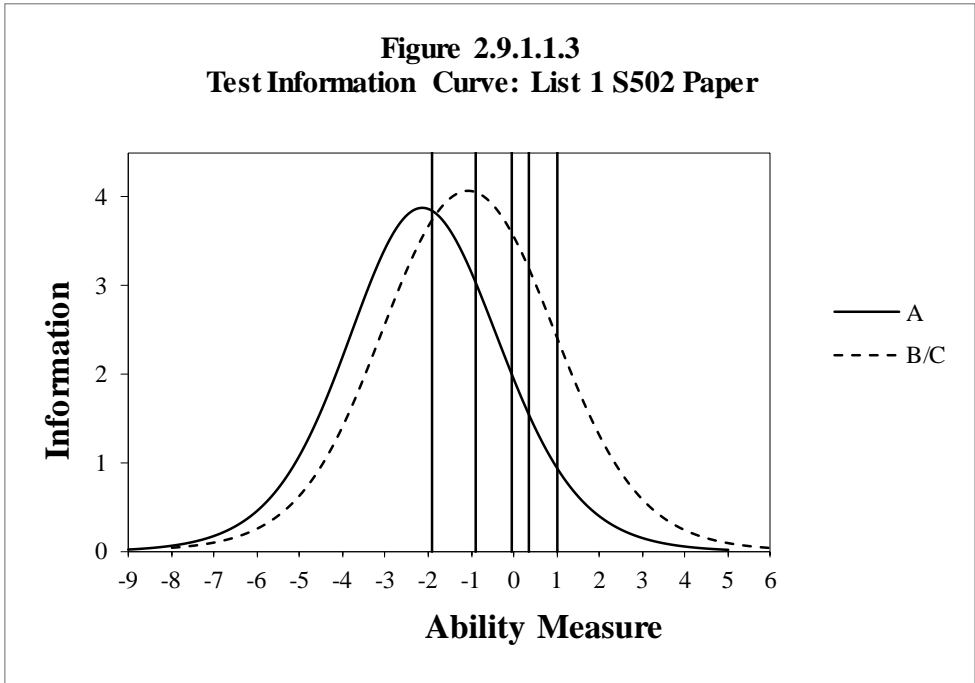
Note: The test form is shared between 1A and 2A.

Figure 2.9.1.1.2
Test Information Curve: List 1 B/C S502 Paper



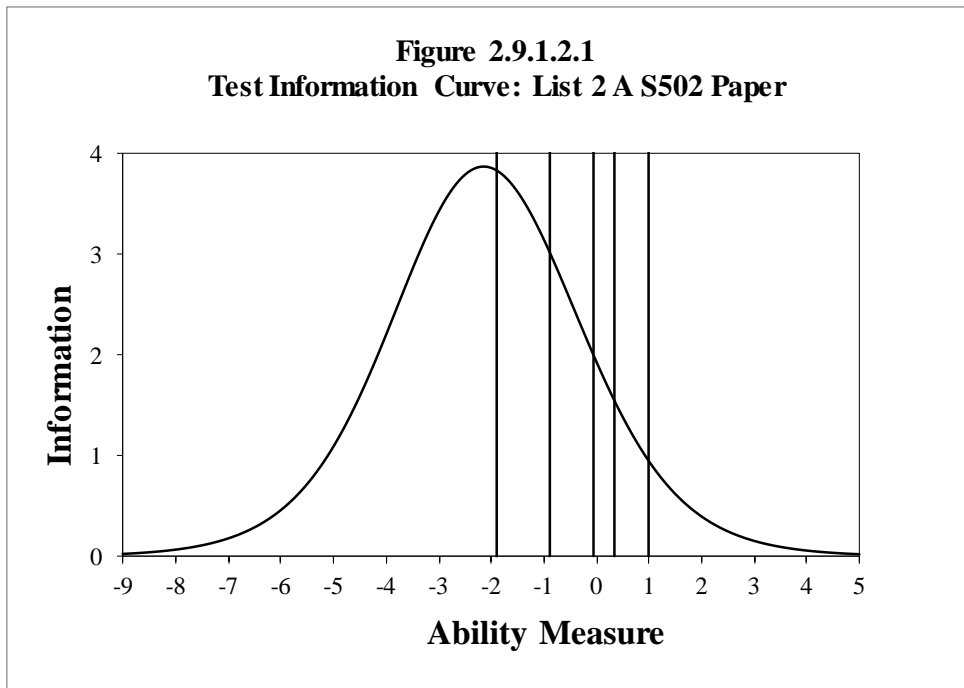
Note: The test form is shared between 1B/C and 2B/C.

Figure 2.9.1.1.3
Test Information Curve: List 1 S502 Paper

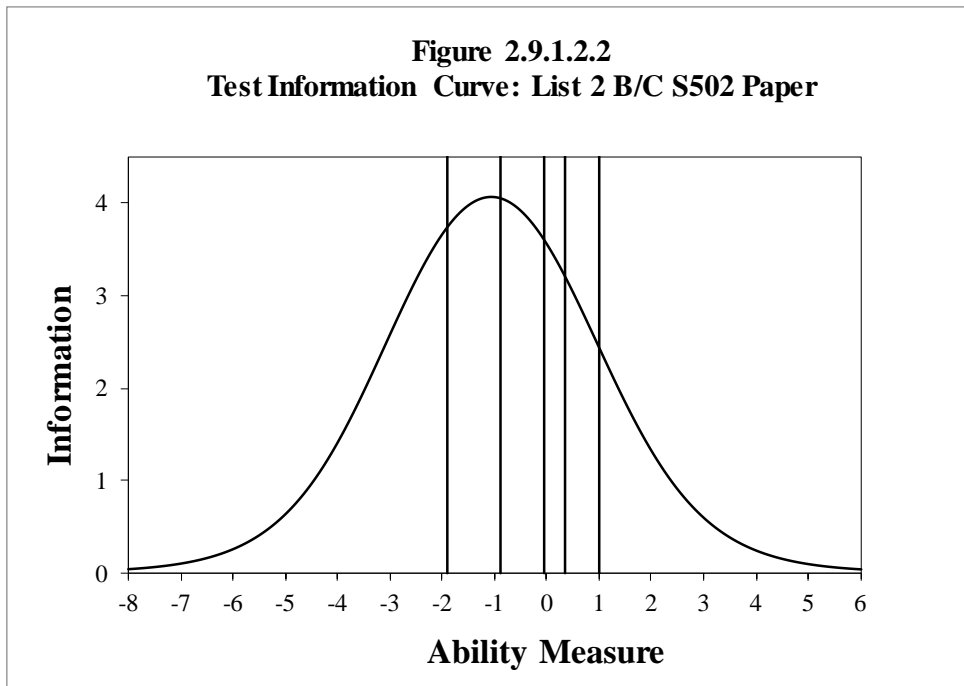


Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

2.9.1.2 Grade 2

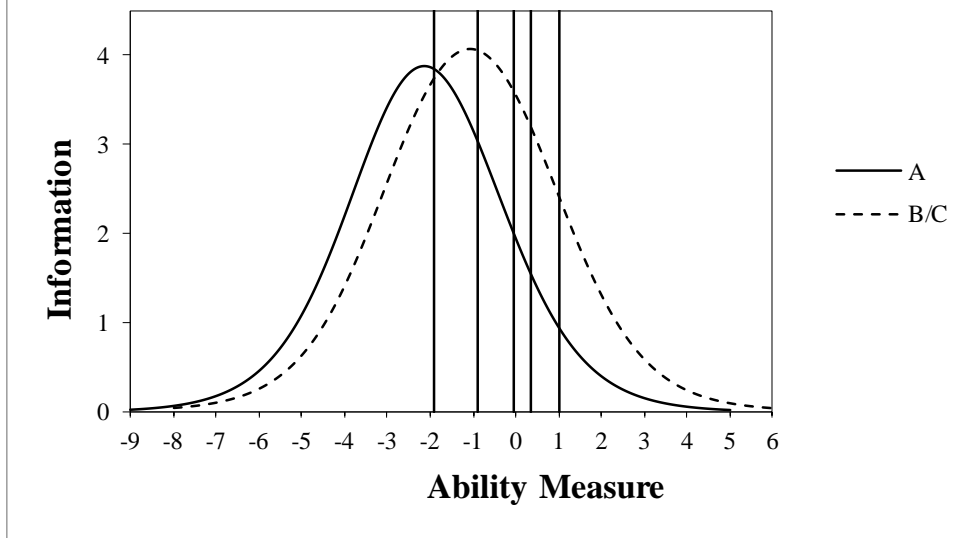


Note: The test form is shared between 1A and 2A.



Note: The test form is shared between 1B/C and 2B/C.

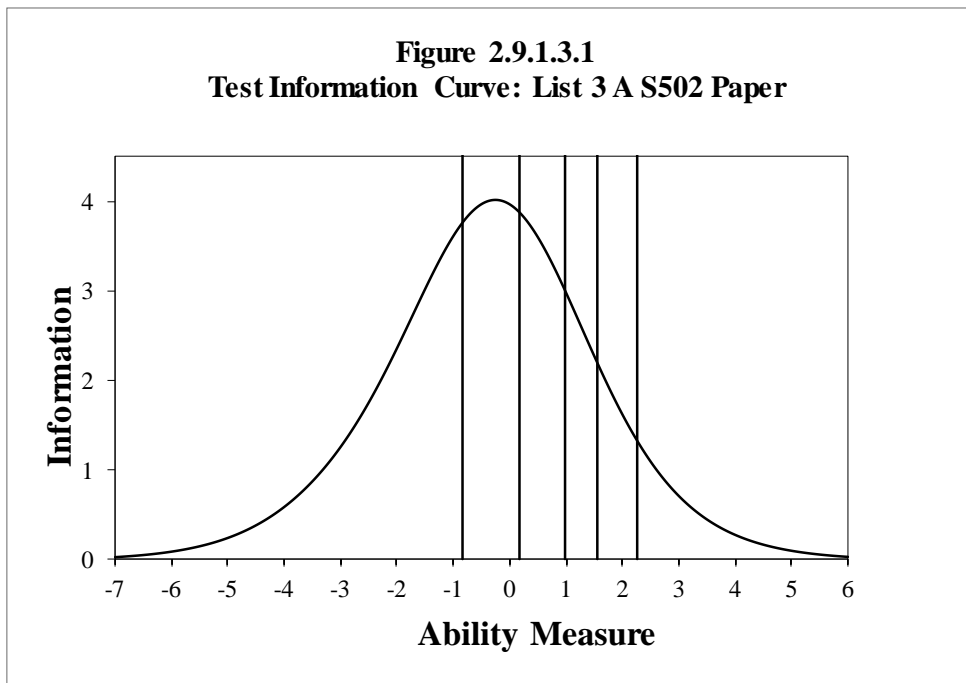
Figure 2.9.1.2.3
Test Information Curve: List 2 S502 Paper



Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

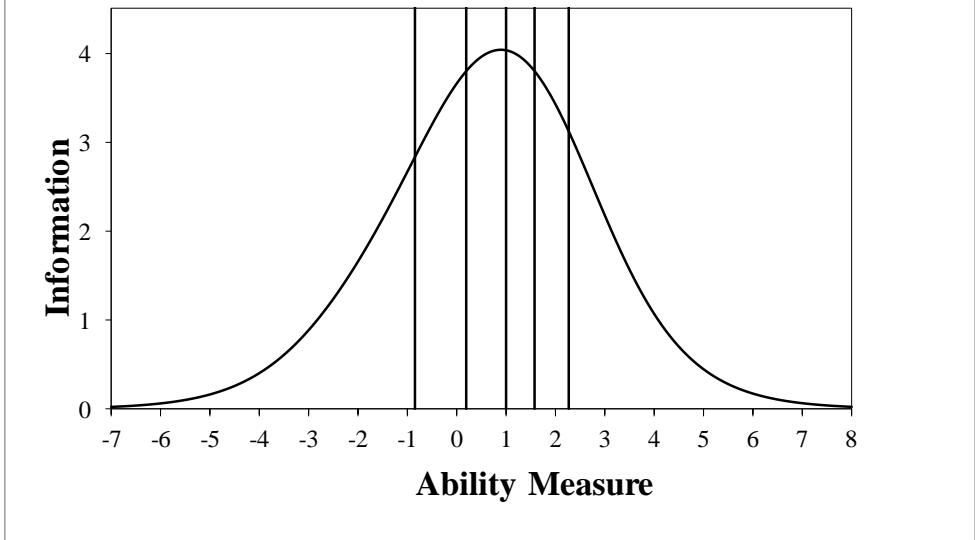
2.9.1.3 Grade 3

Figure 2.9.1.3.1
Test Information Curve: List 3 A S502 Paper



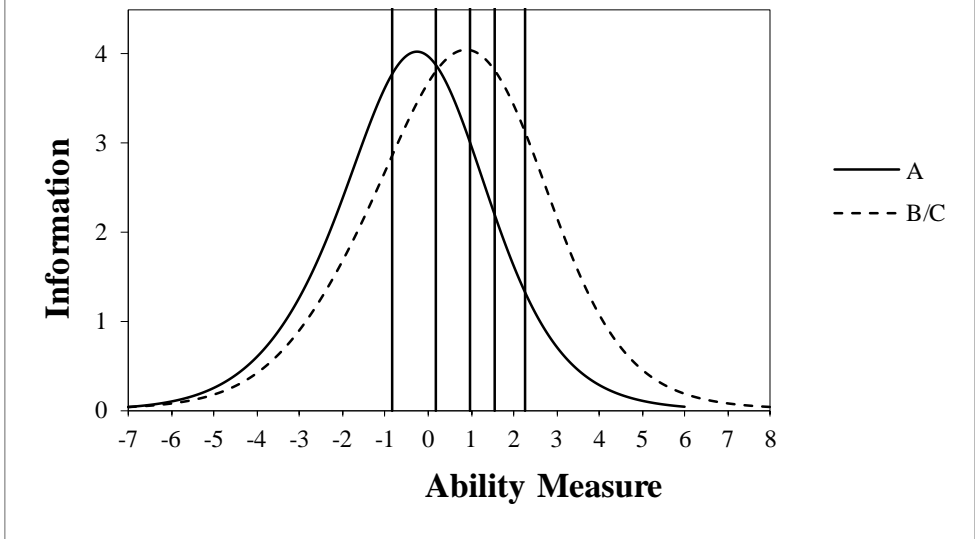
Note: The test form is shared between 3A and 4–5A.

Figure 2.9.1.3.2
Test Information Curve: List 3 B/C S502 Paper



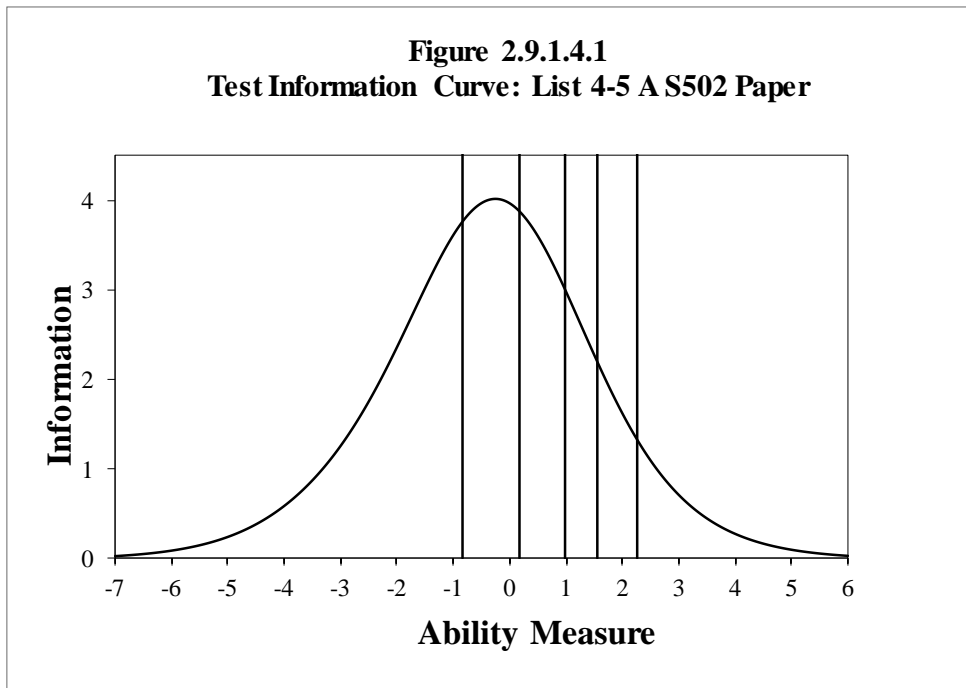
Note: The test form is shared between 3B/C and 4–5B/C.

Figure 2.9.1.3.3
Test Information Curve: List 3 S502 Paper

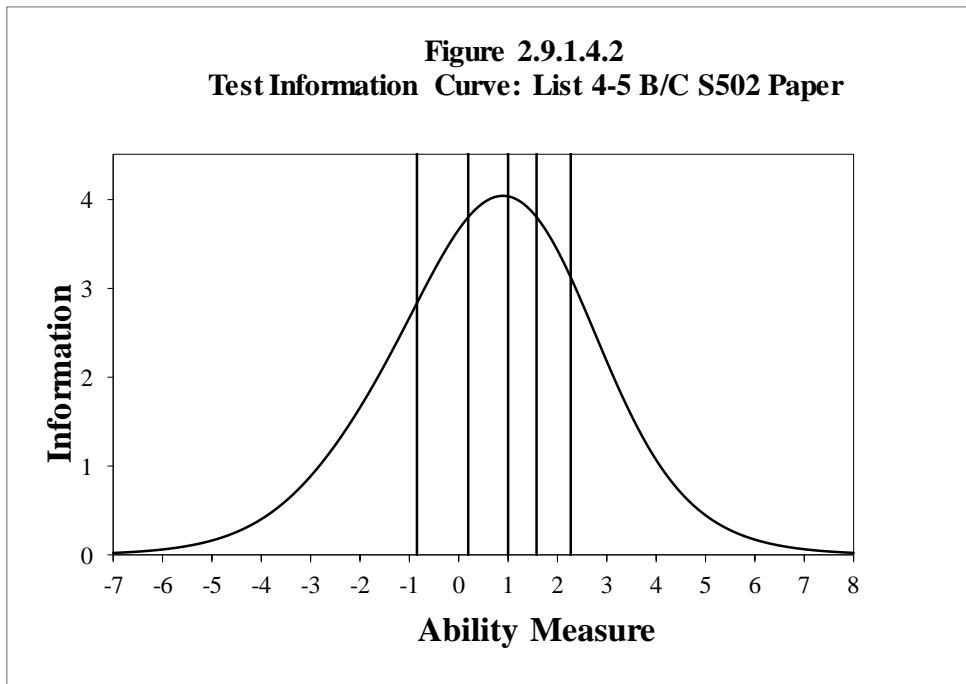


Note: The test form is shared between 3A and 4–5A, 3B/C and 4–5B/C.

2.9.1.4 Grades 4–5

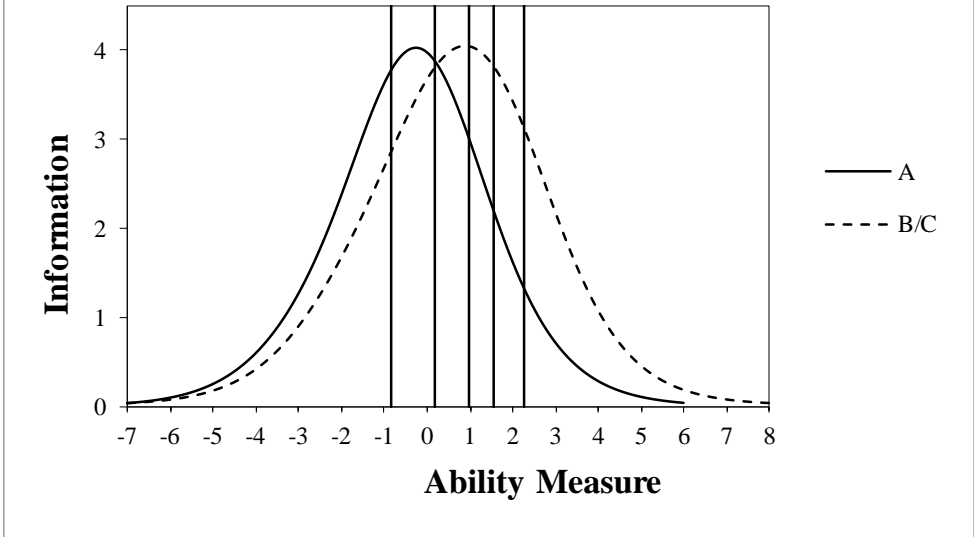


Note: The test form is shared between 3A and 4–5A.



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 S502 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 S502 Paper

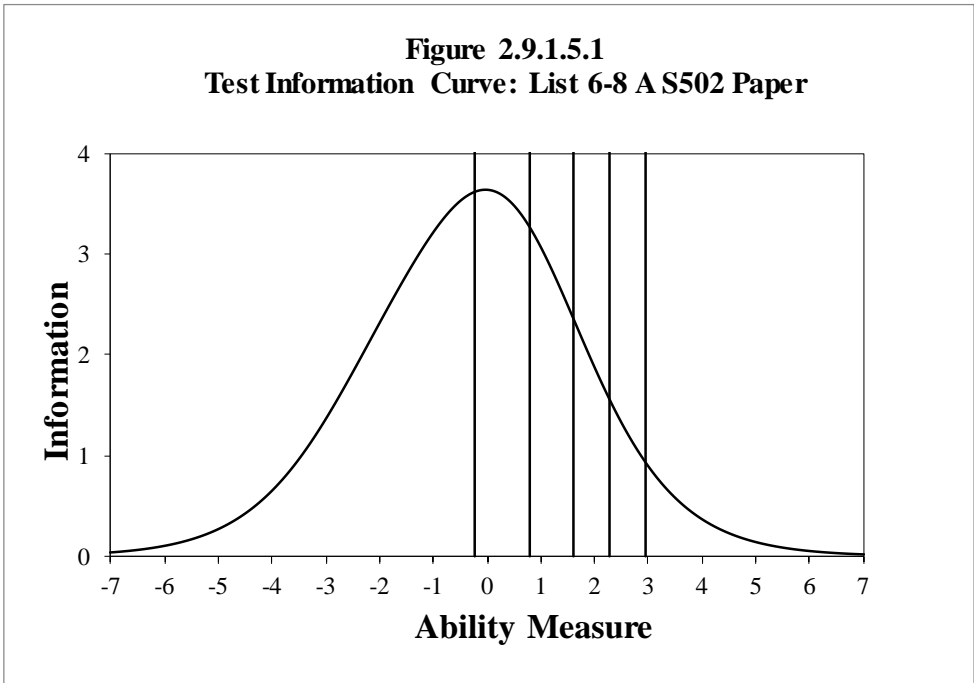


Figure 2.9.1.5.2
Test Information Curve: List 6-8 B/C S502 Paper

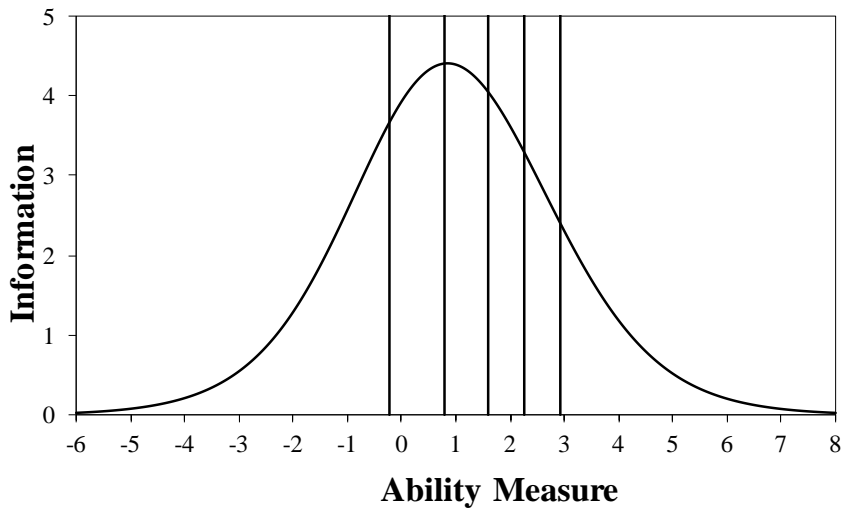
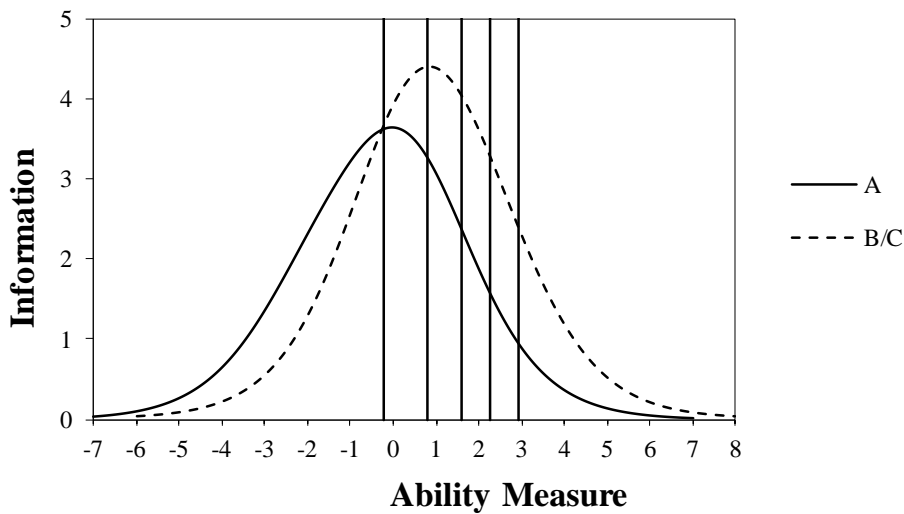


Figure 2.9.1.5.3
Test Information Curve: List 6-8 S502 Paper



2.9.1.6 Grades 9–12

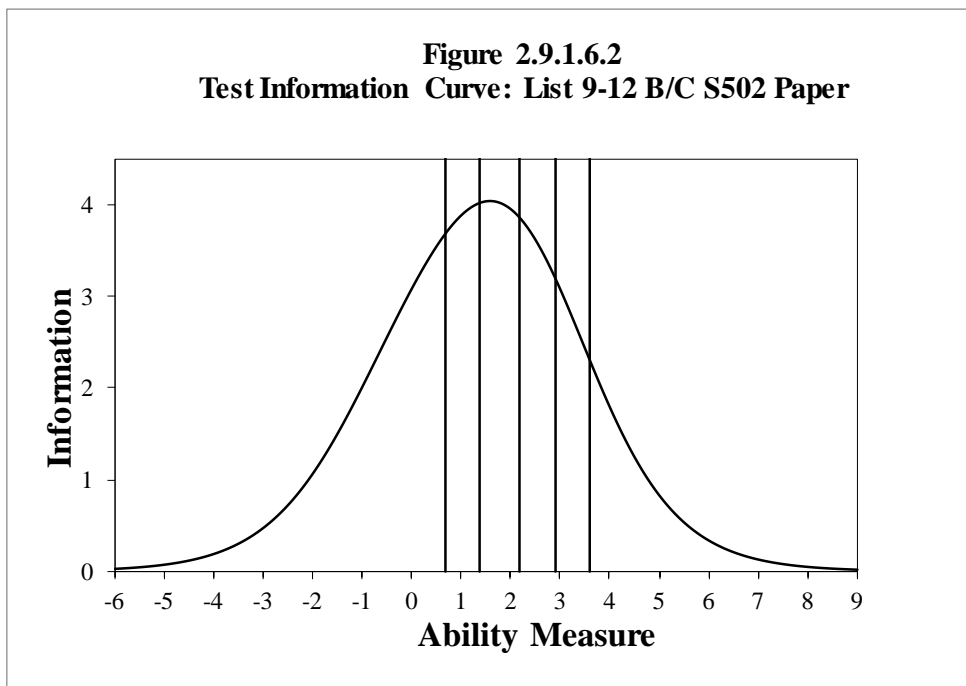
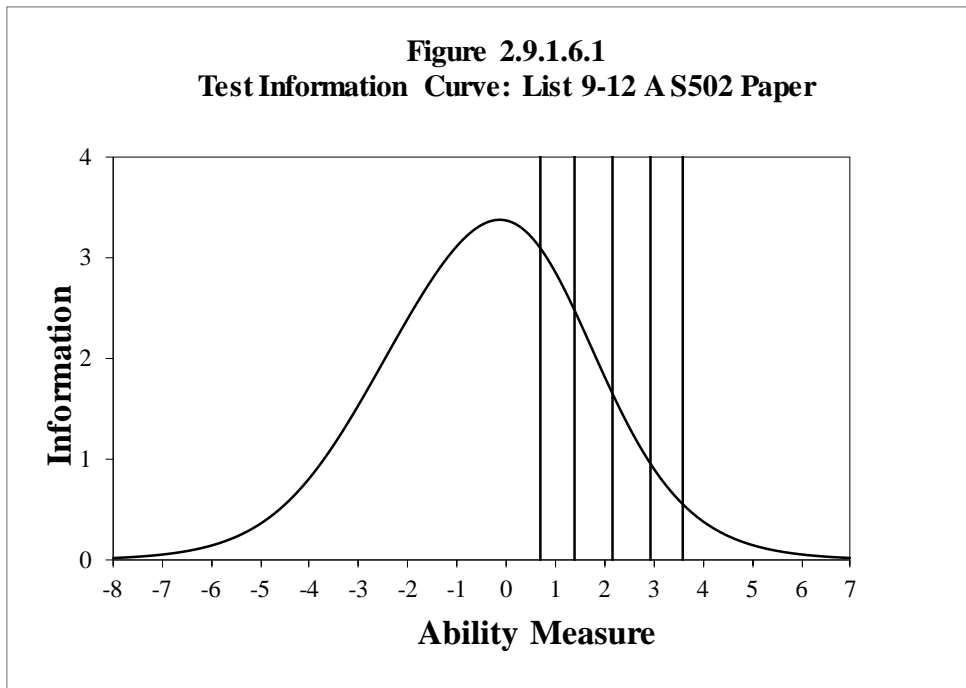
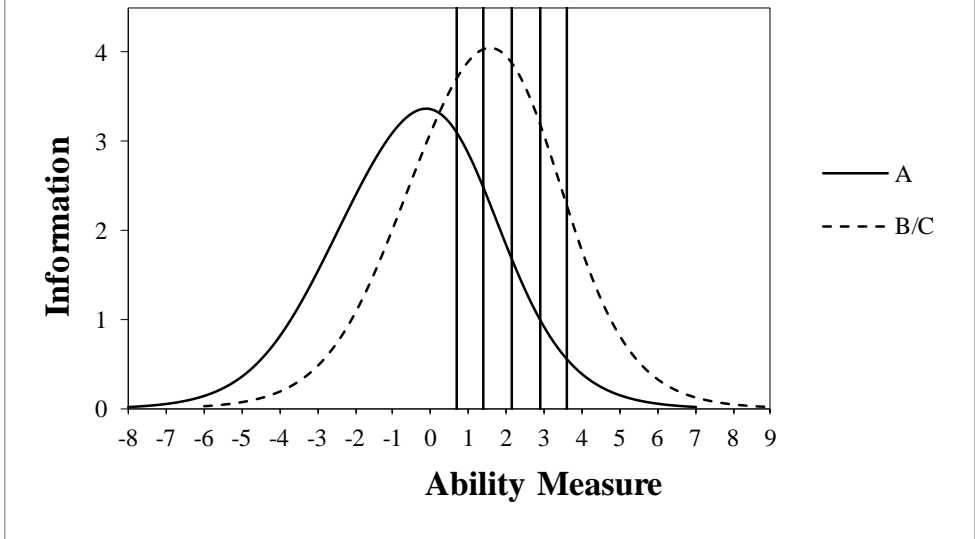
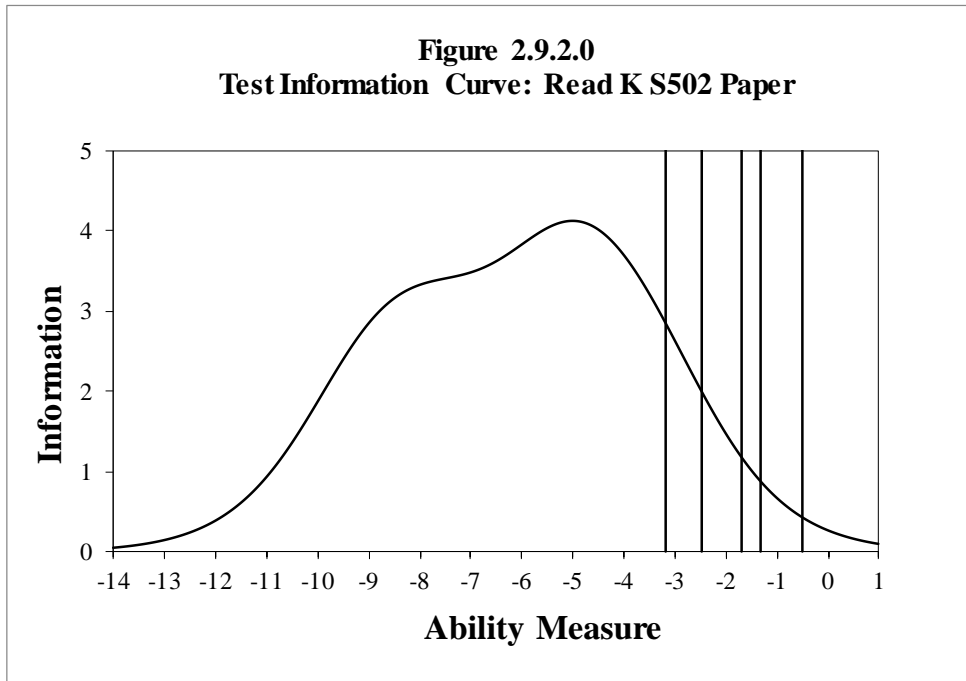


Figure 2.9.1.6.3
Test Information Curve: List 9-12 S502 Paper

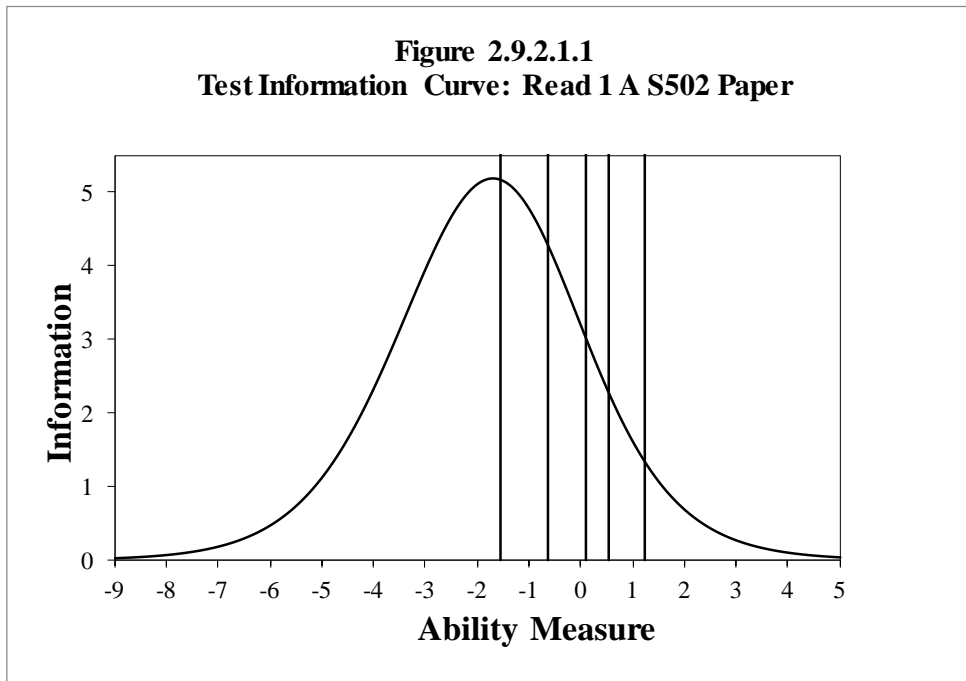


2.9.2 Reading

2.9.2.0 Kindergarten

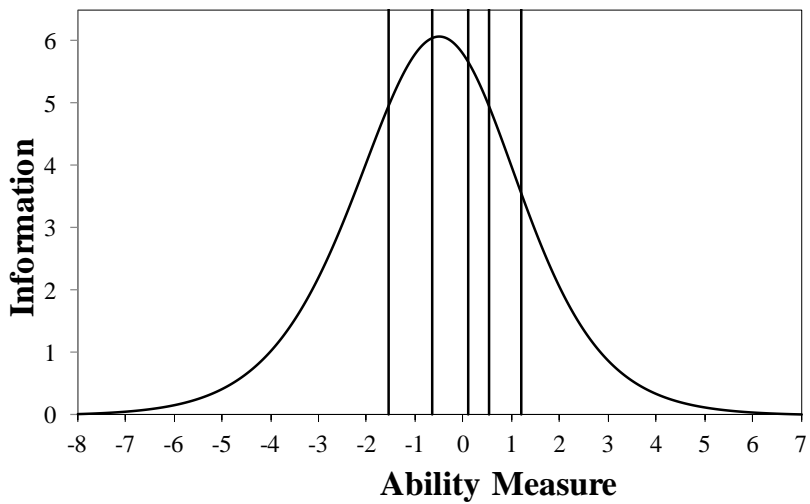


2.9.2.1 Grade 1



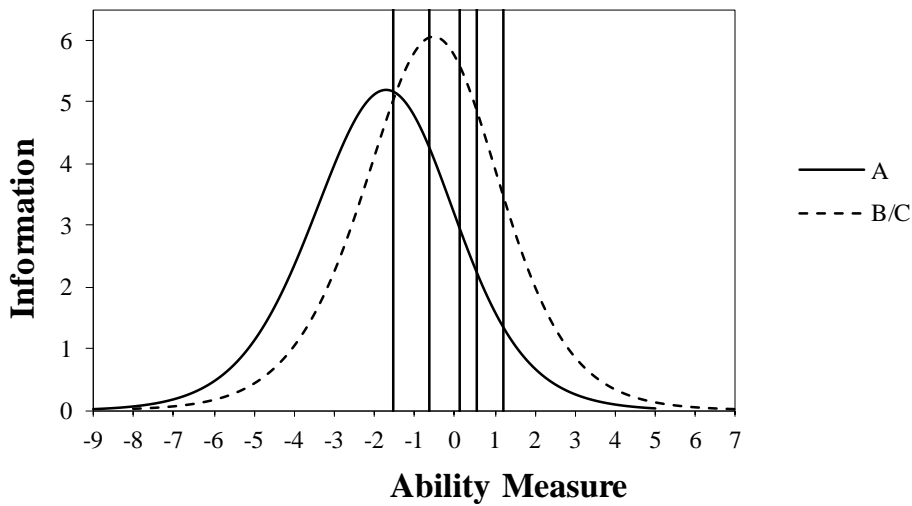
Note: The test form is shared between 1A and 2A.

Figure 2.9.2.1.2
Test Information Curve: Read 1 B/C S502 Paper



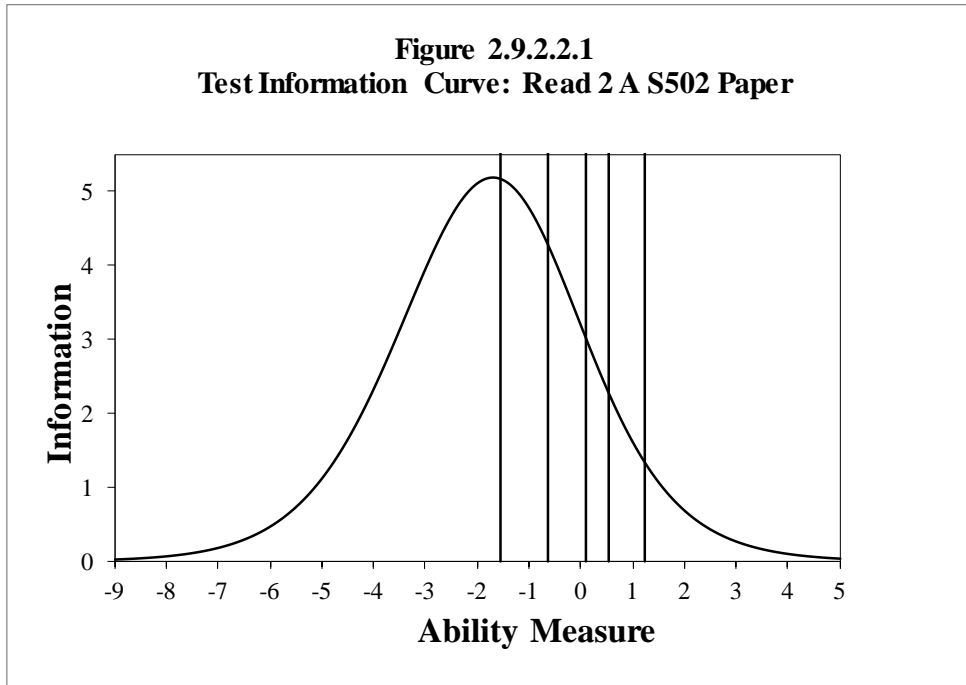
Note: The test form is shared between 1B/C and 2B/C.

Figure 2.9.2.1.3
Test Information Curve: Read 1 S502 Paper

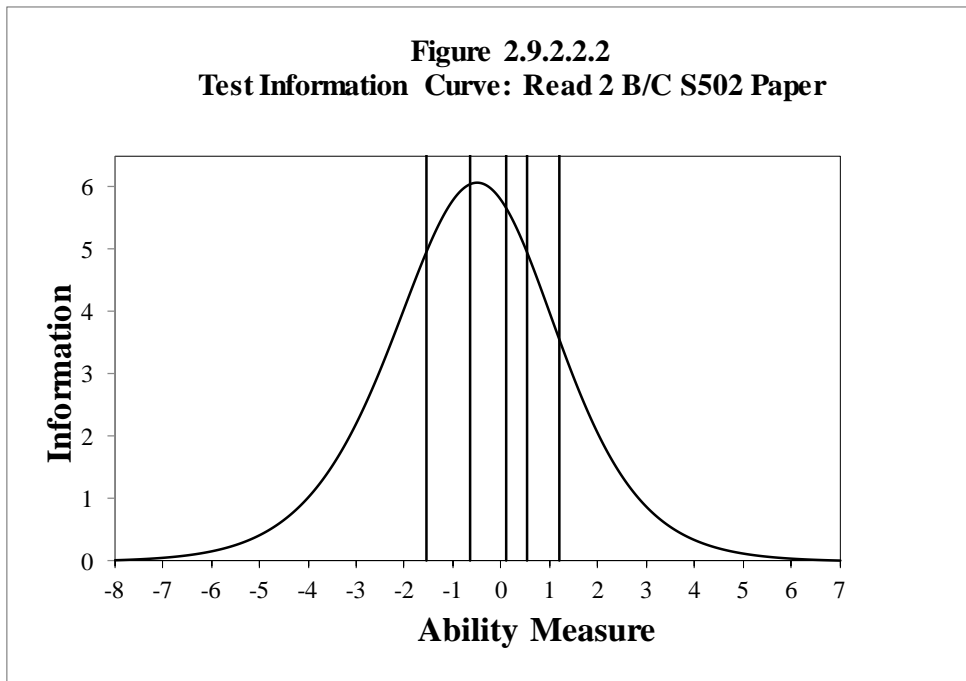


Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

2.9.2.2 Grade 2

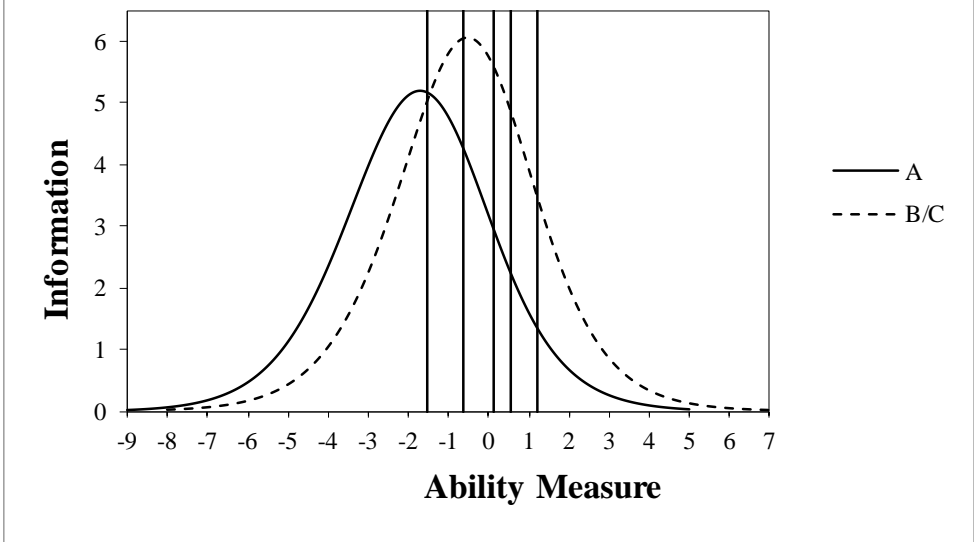


Note: The test form is shared between 1A and 2A.



Note: The test form is shared between 1B/C and 2B/C.

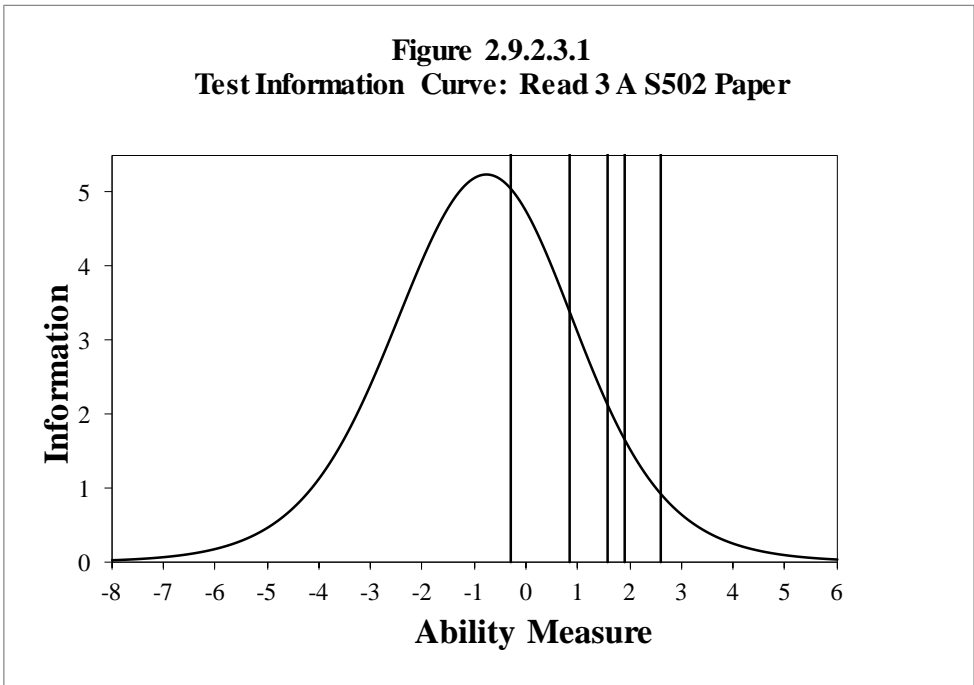
Figure 2.9.2.2.3
Test Information Curve: Read 2 S502 Paper



Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

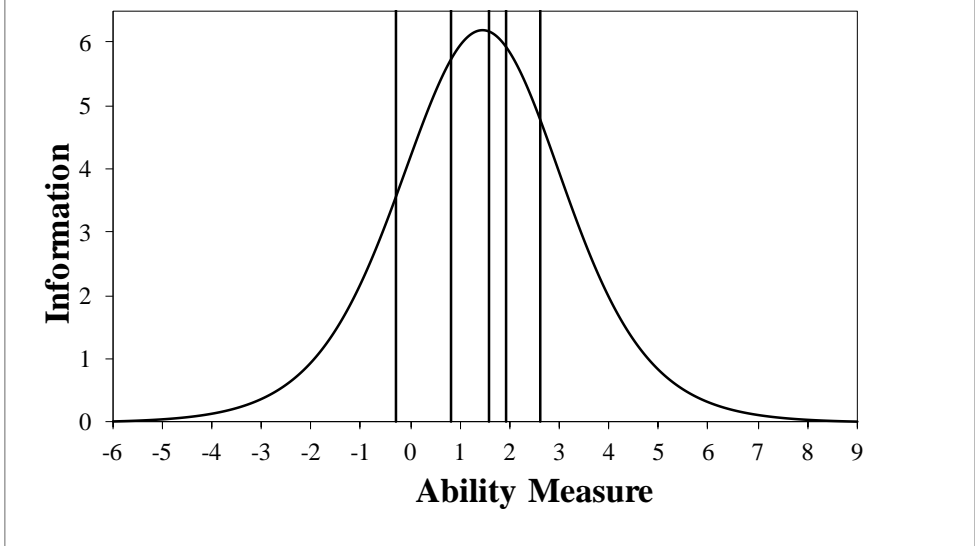
2.9.2.3 Grade 3

Figure 2.9.2.3.1
Test Information Curve: Read 3 A S502 Paper



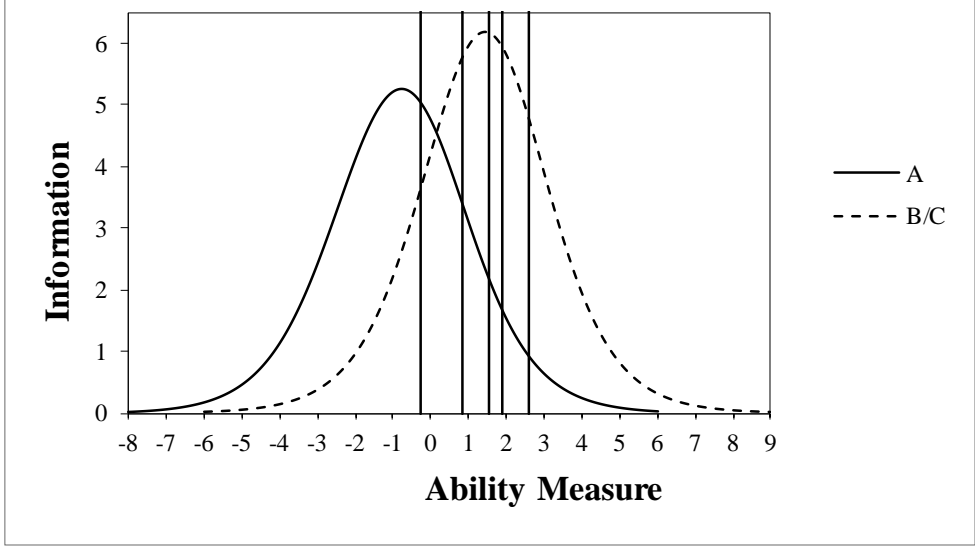
Note: The test form is shared between 3A and 4–5A.

Figure 2.9.2.3.2
Test Information Curve: Read 3 B/C S502 Paper



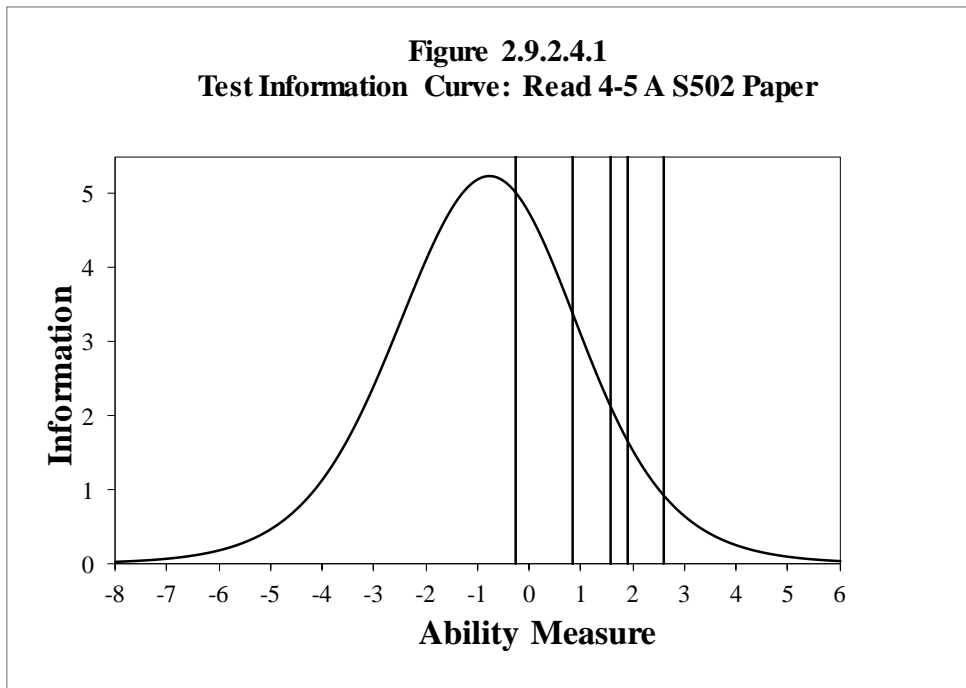
Note: The test form is shared between 3B/C and 4–5B/C.

Figure 2.9.2.3.3
Test Information Curve: Read 3 S502 Paper

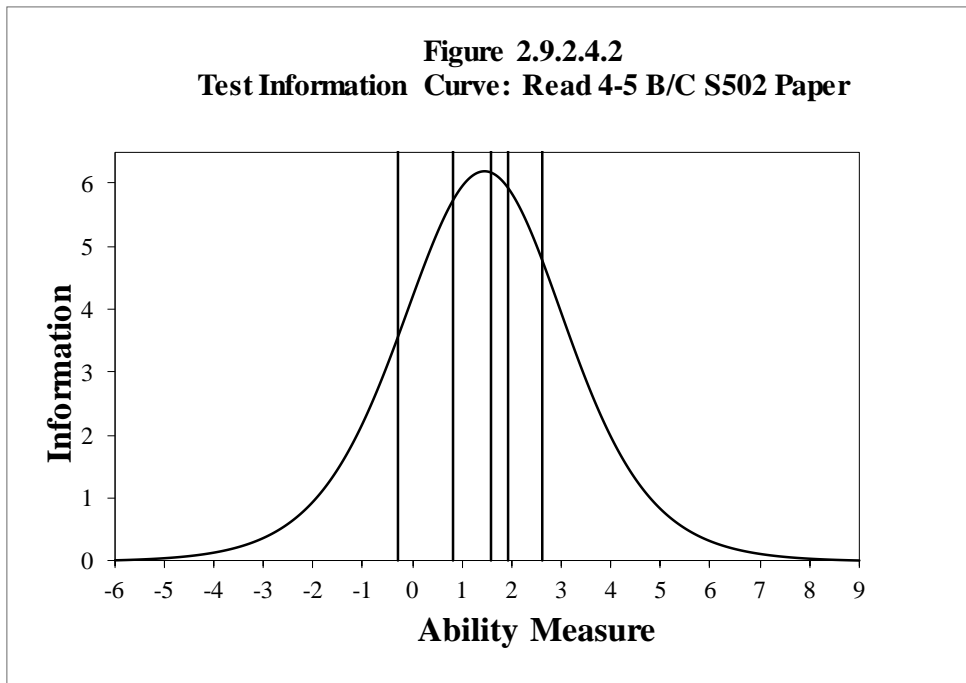


Note: The test form is shared between 3A and 4–5A, 3B/C and 4–5B/C.

2.9.2.4 Grades 4–5

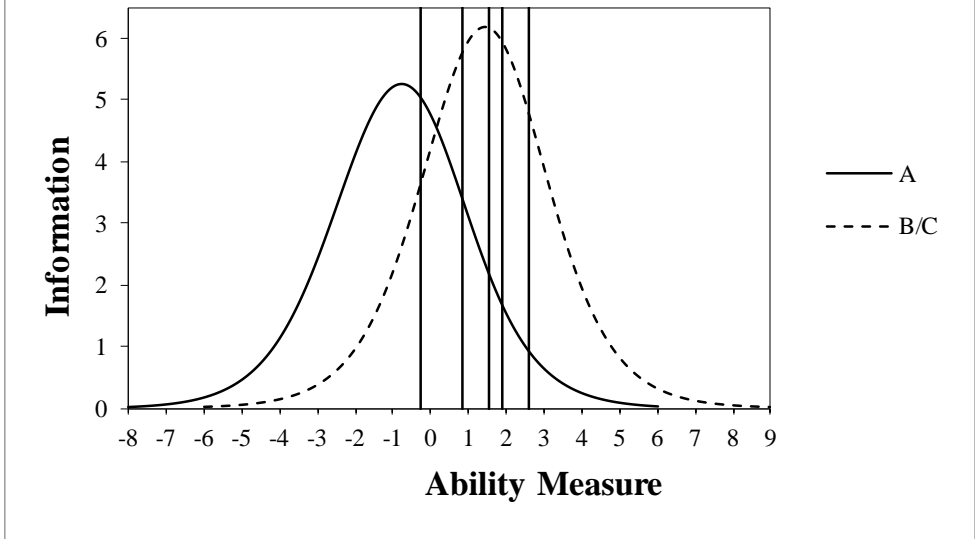


Note: The test form is shared between 3A and 4–5A.



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 S502 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 S502 Paper

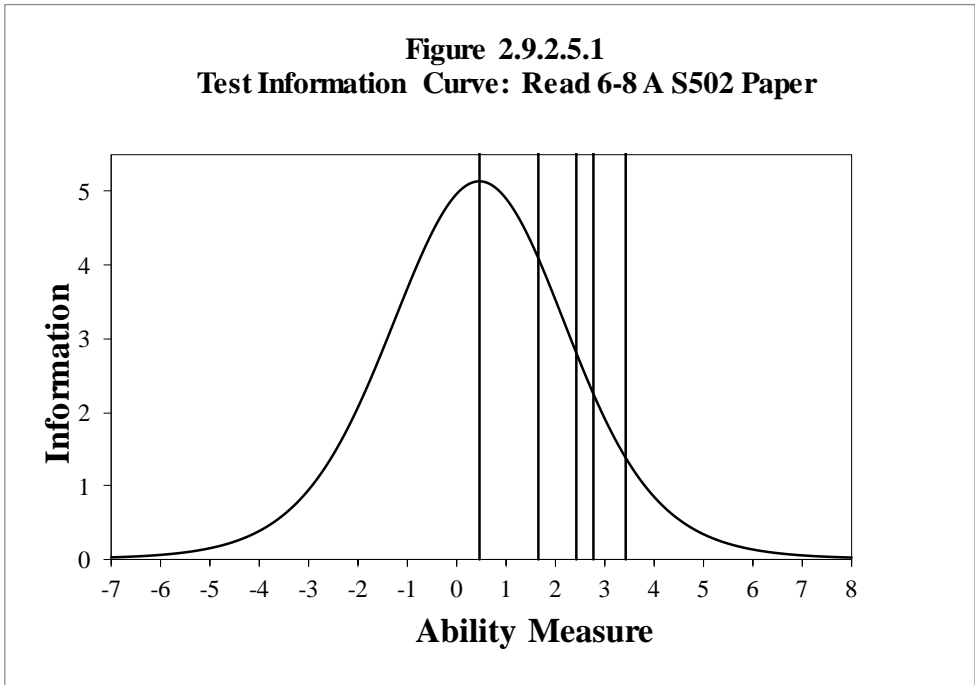


Figure 2.9.2.5.2
Test Information Curve: Read 6-8 B/C S502 Paper

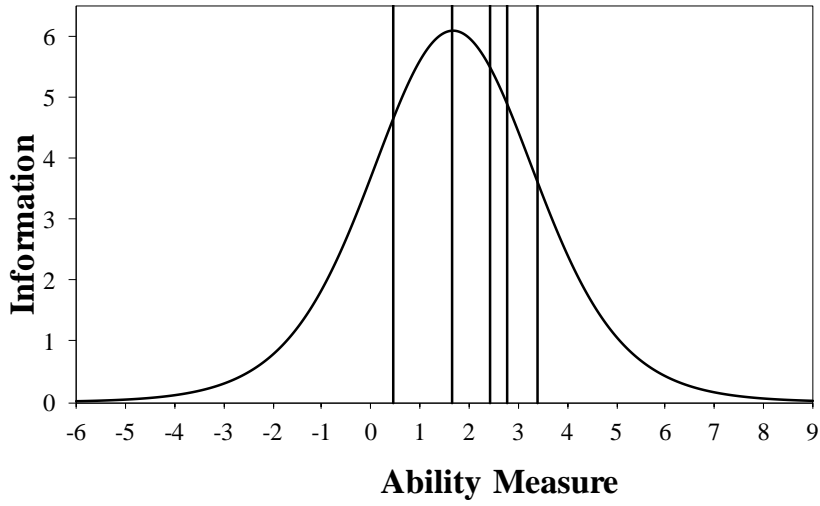
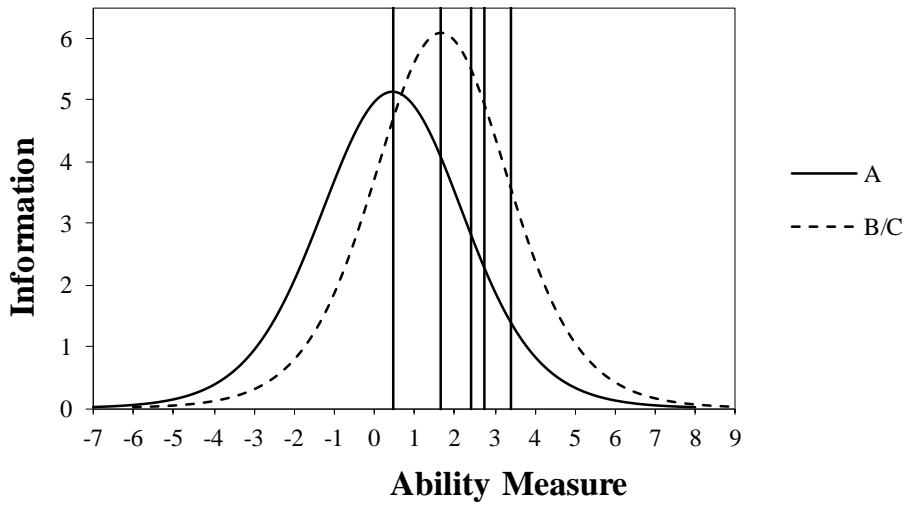


Figure 2.9.2.5.3
Test Information Curve: Read 6-8 S502 Paper



2.9.2.6 Grades 9–12

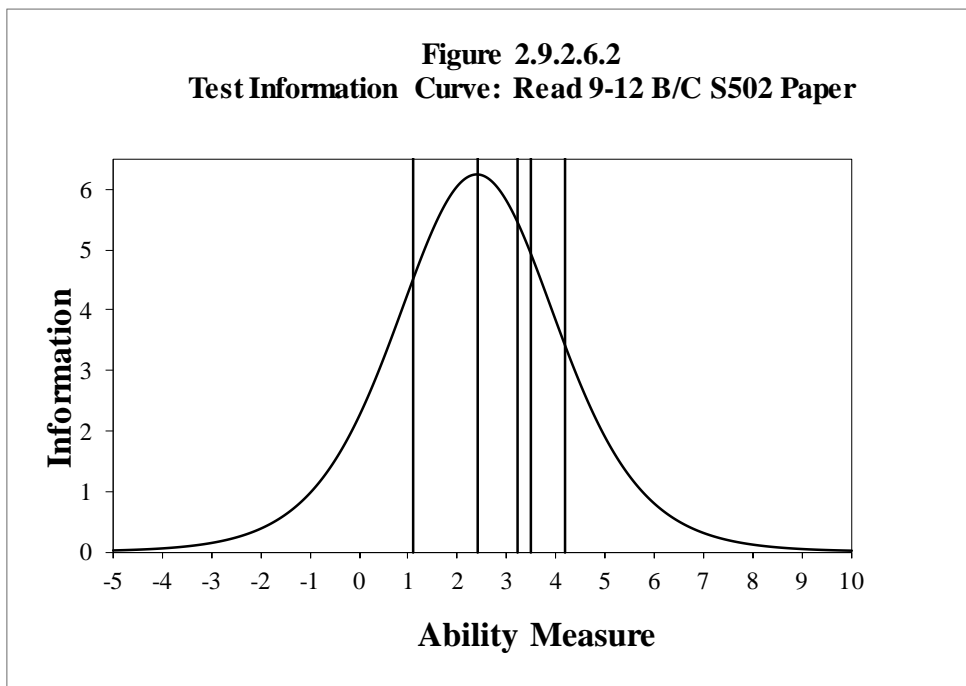
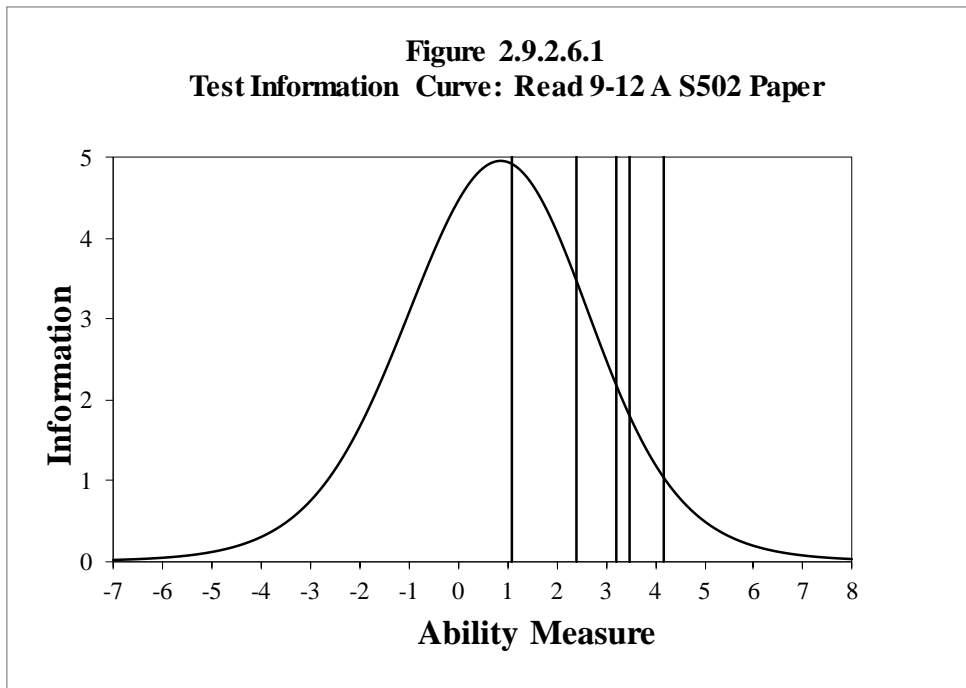
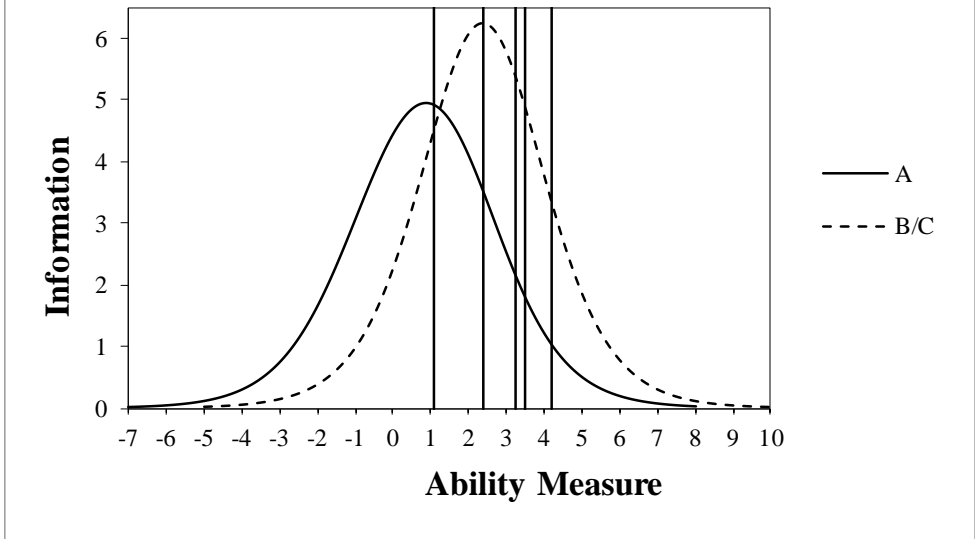
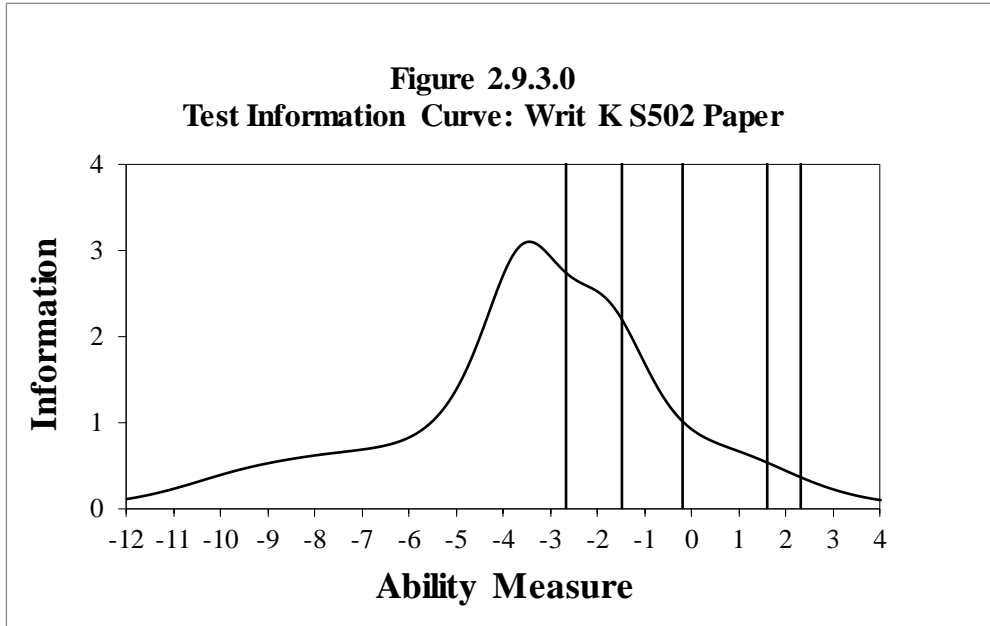


Figure 2.9.2.6.3
Test Information Curve: Read 9-12 S502 Paper



2.9.3 Writing

2.9.3.0 Kindergarten



2.9.3.1 Grade 1

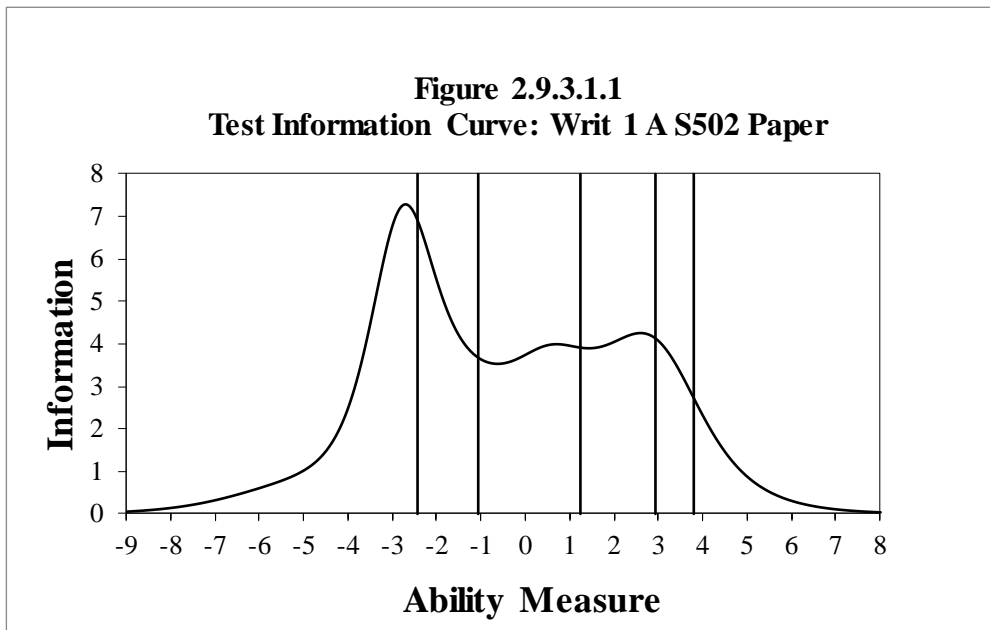


Figure 2.9.3.1.2
Test Information Curve: Writ 1 B/C S502 Paper

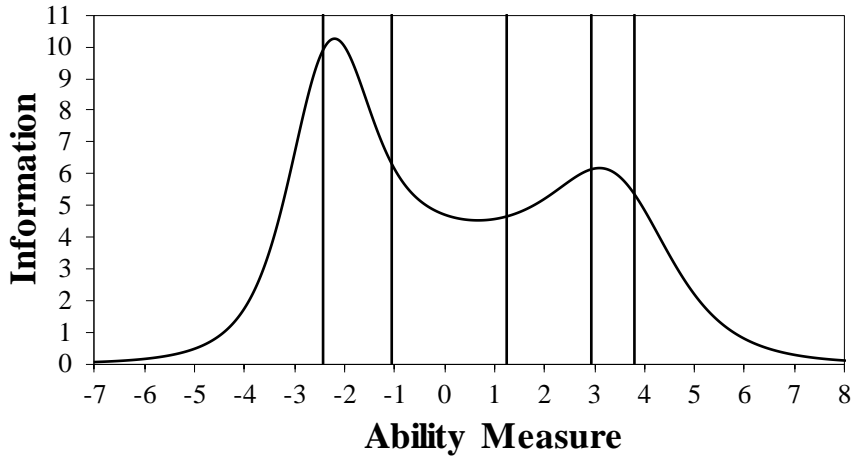
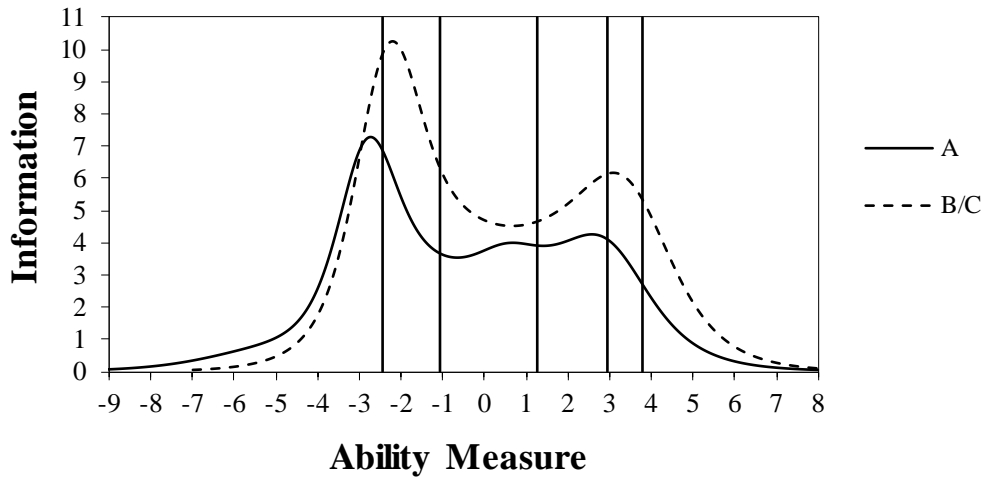
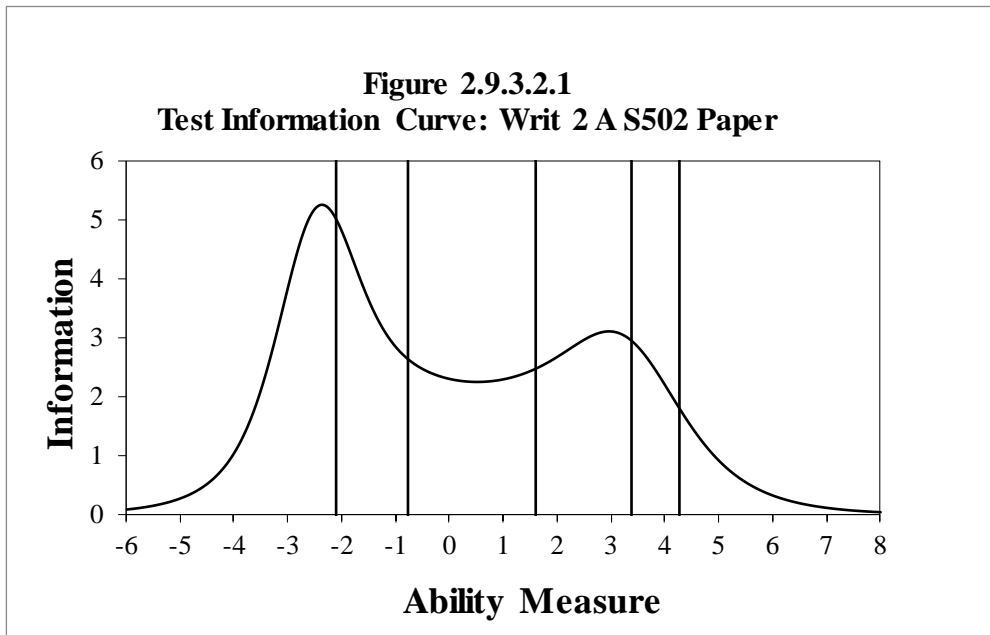


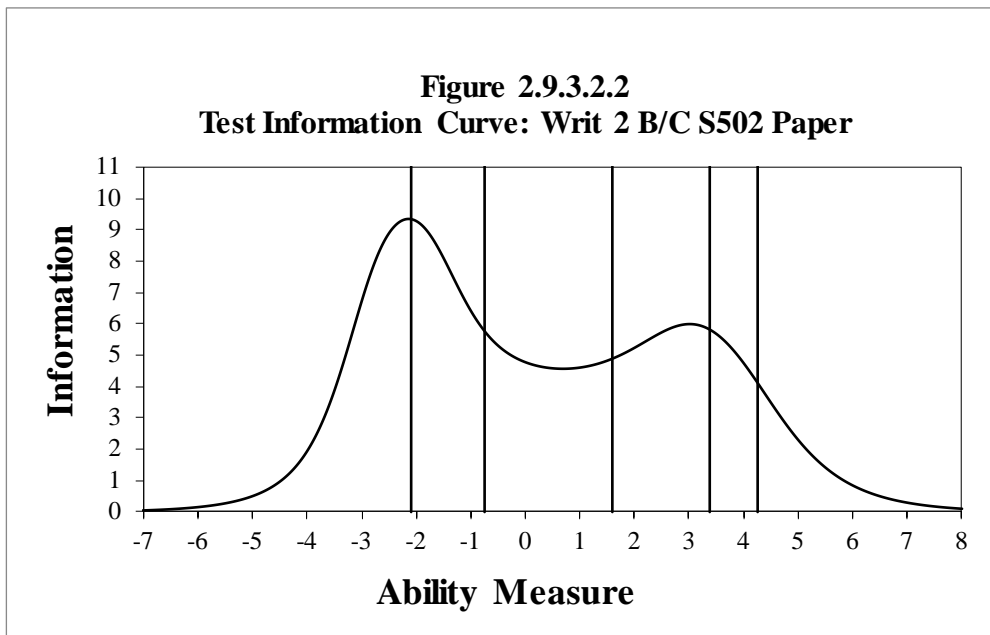
Figure 2.9.3.1.3
Test Information Curve: Writ 1 S502 Paper



2.9.3.2 Grade 2

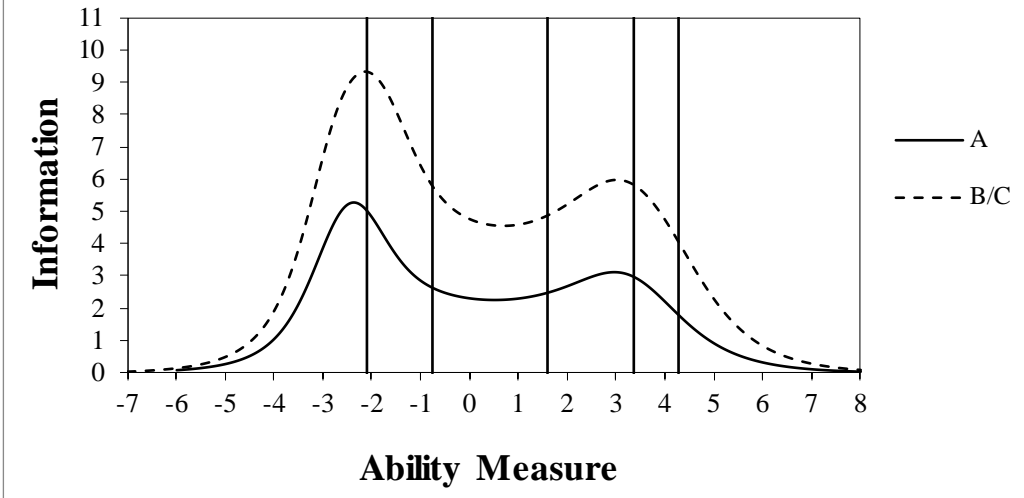


Note: The test form is shared between 2A and 3A.



Note: The test form is shared between 2B/C and 3B/C.

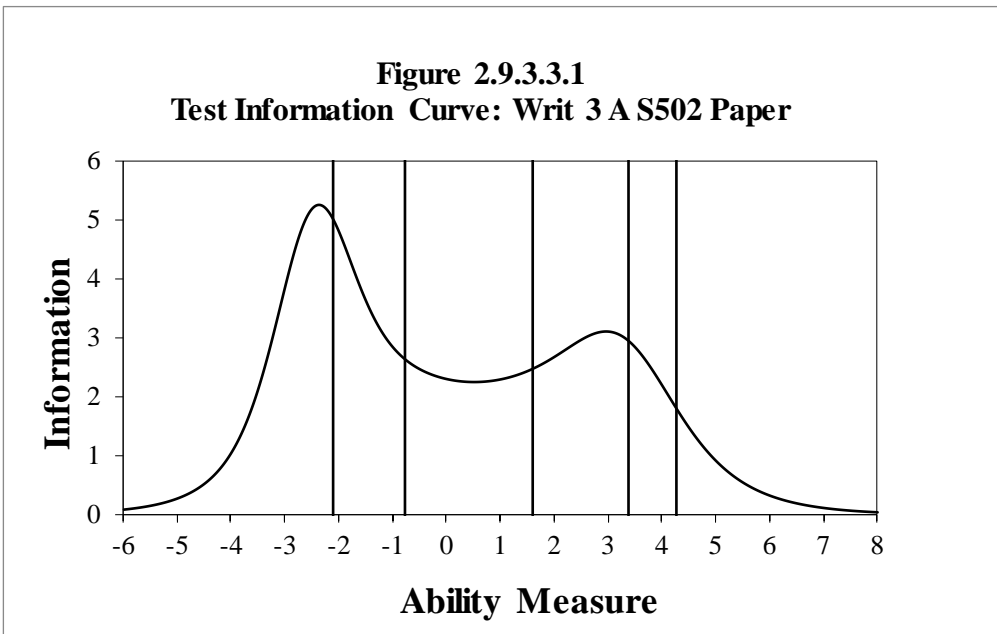
Figure 2.9.3.2.3
Test Information Curve: Writ 2 S502 Paper



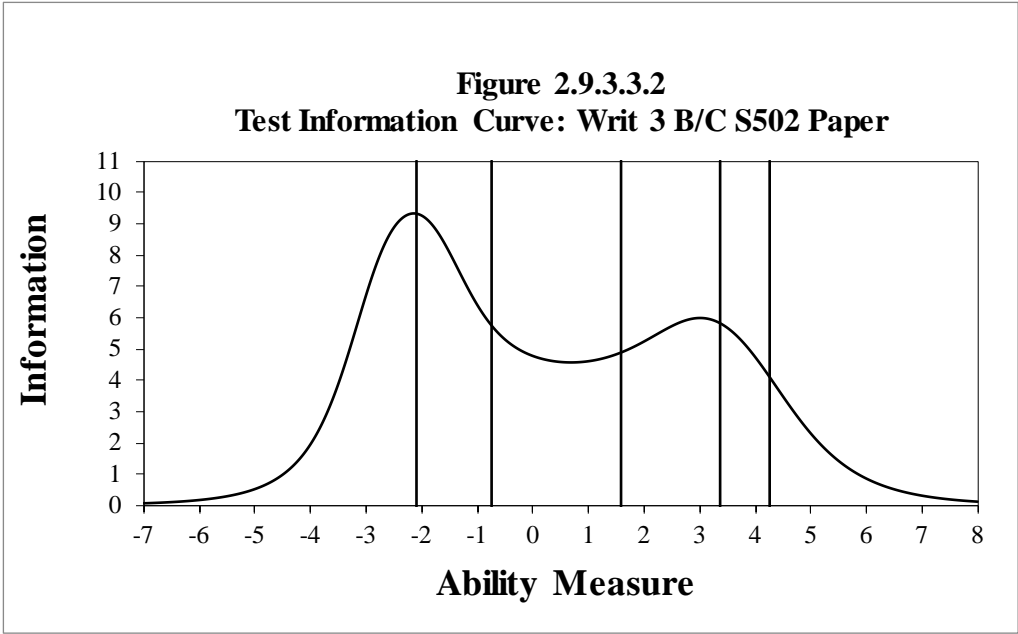
Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

2.9.3.3 Grade 3

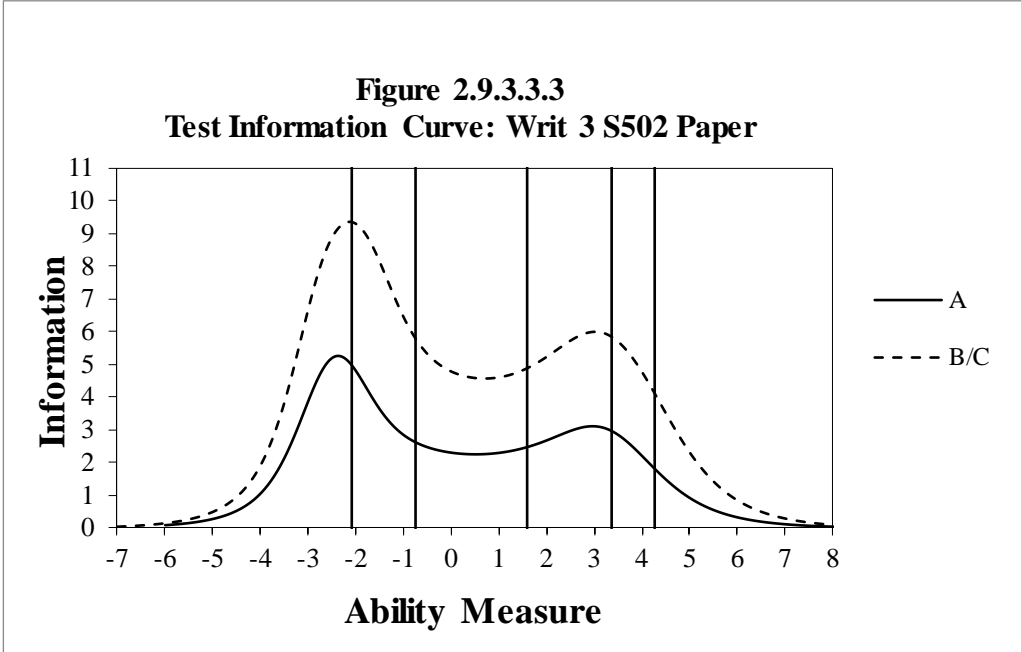
Figure 2.9.3.3.1
Test Information Curve: Writ 3 A S502 Paper



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.

2.9.3.4 Grades 4–5

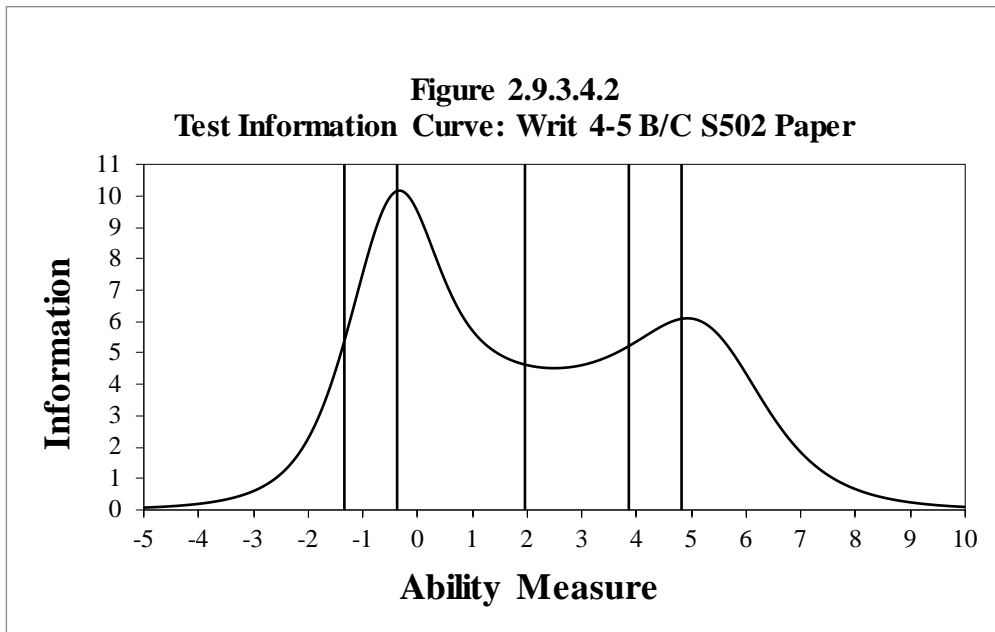
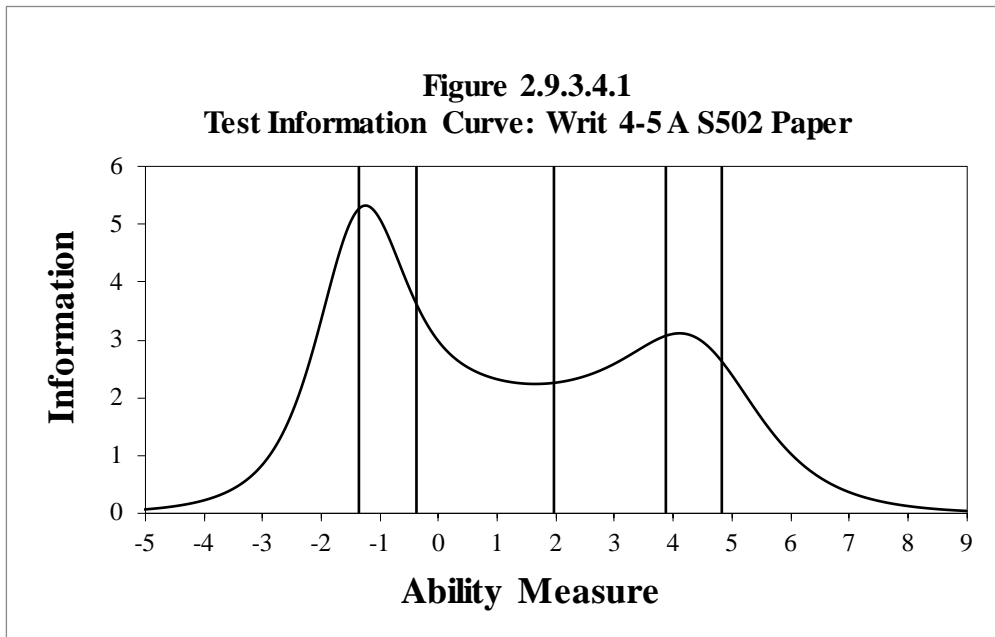
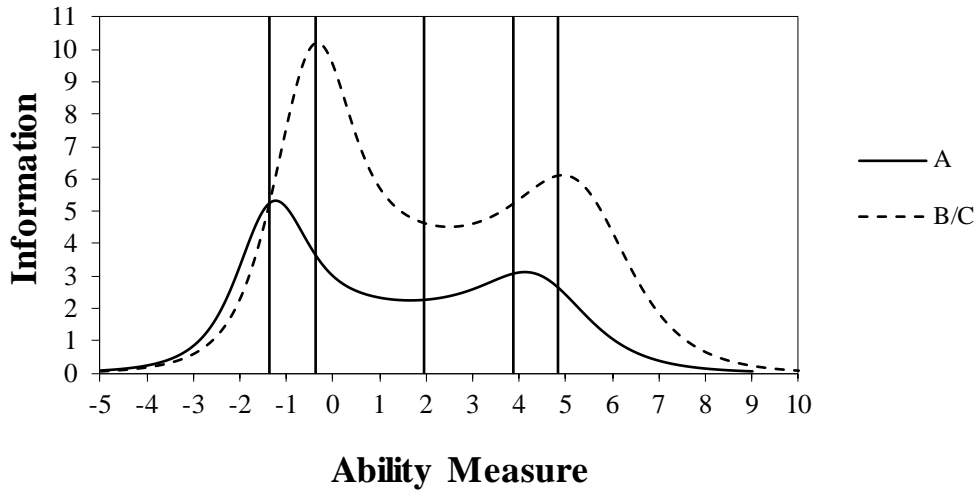


Figure 2.9.3.4.3
Test Information Curve: Writ 4-5 S502 Paper



2.9.3.5 Grades 6–8

Figure 2.9.3.5.1
Test Information Curve: Writ 6-8 A S502 Paper

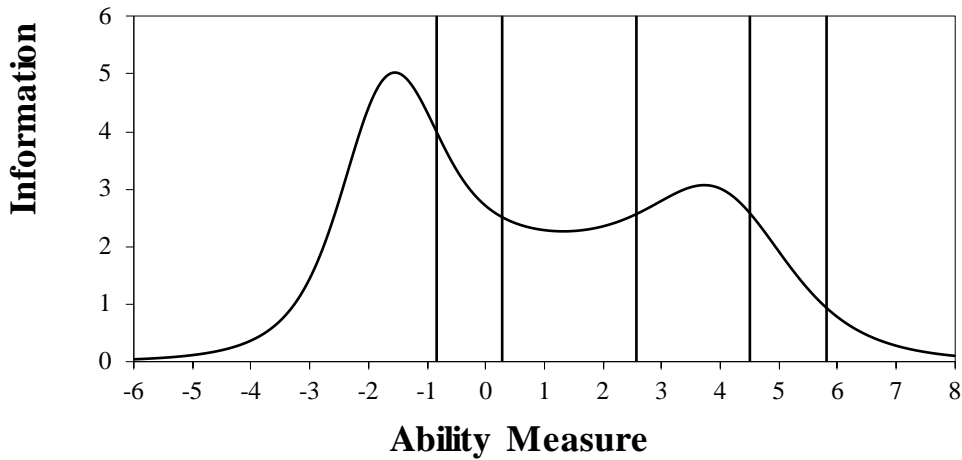


Figure 2.9.3.5.2
Test Information Curve: Writ 6-8 B/C S502 Paper

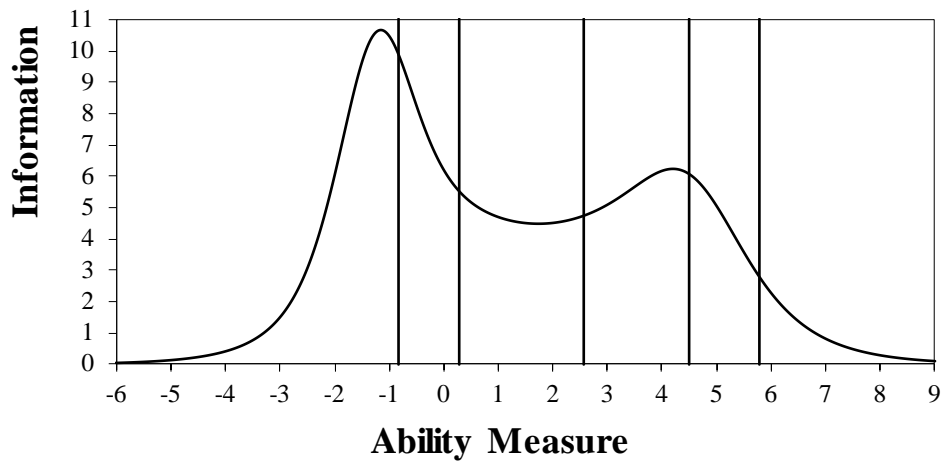
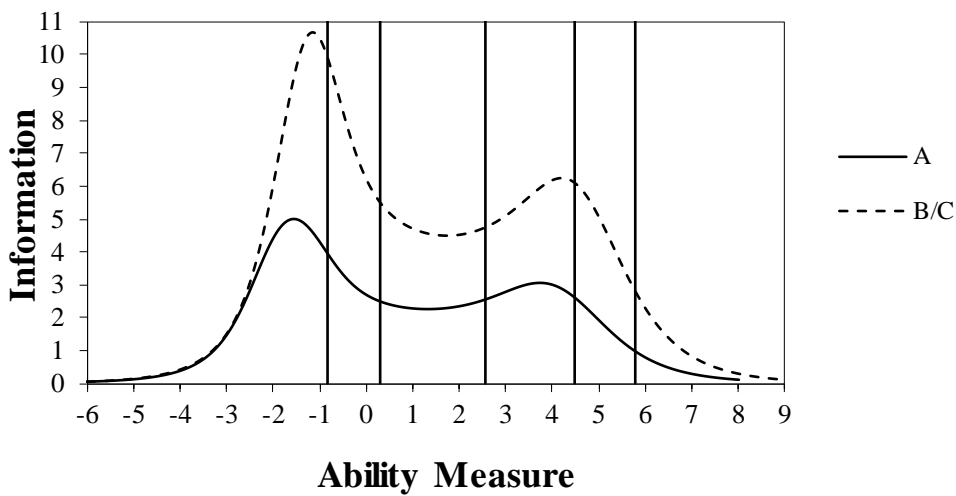


Figure 2.9.3.5.3
Test Information Curve: Writ 6-8 S502 Paper



2.9.3.6 Grades 9–12

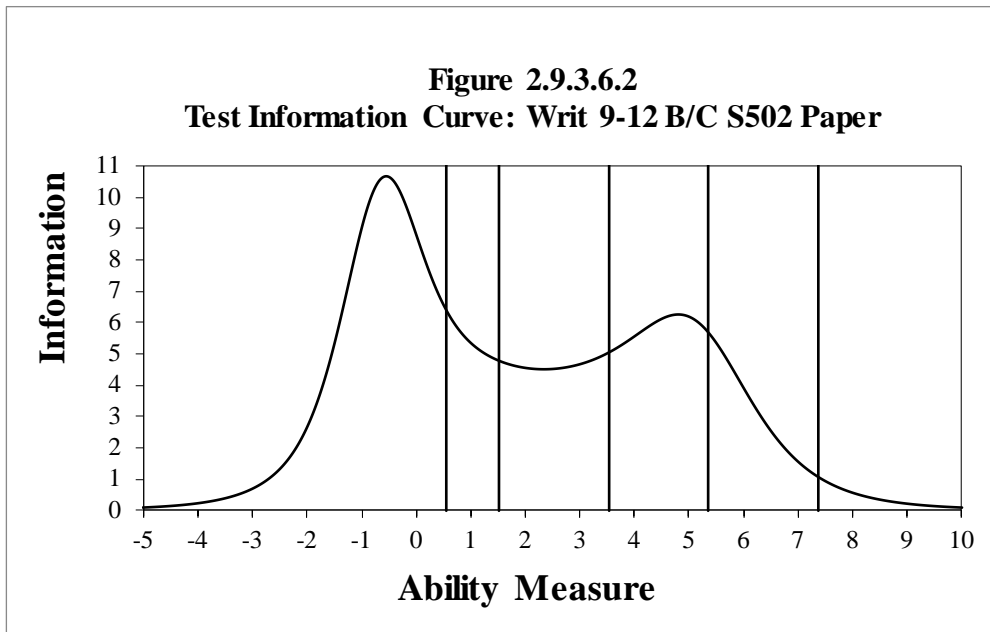
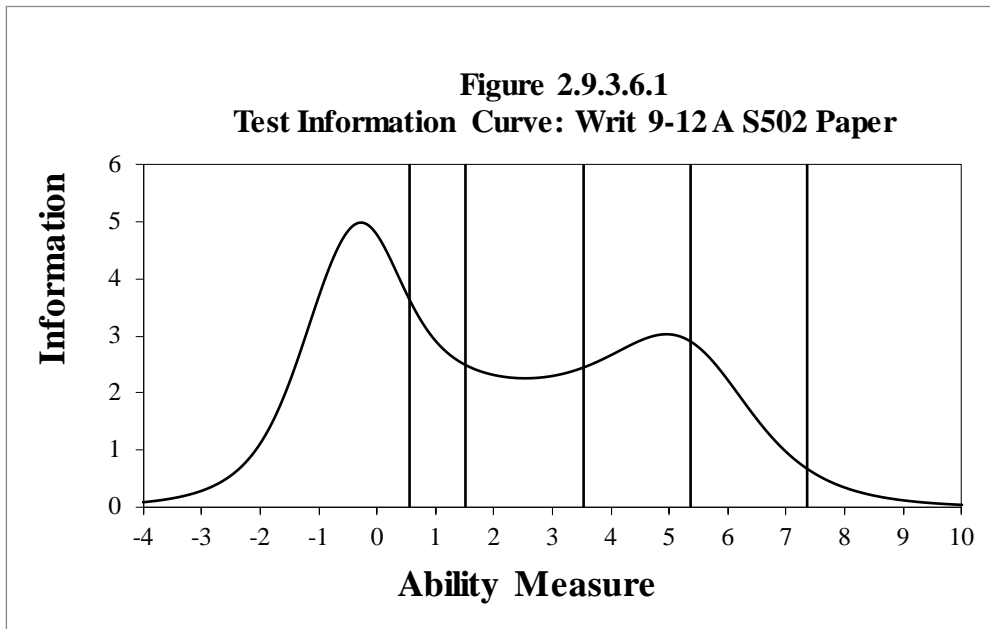
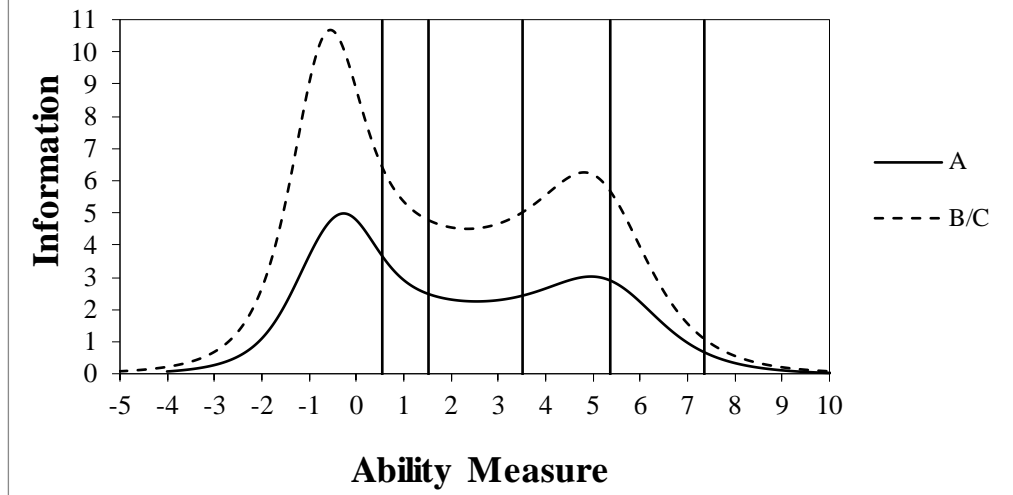
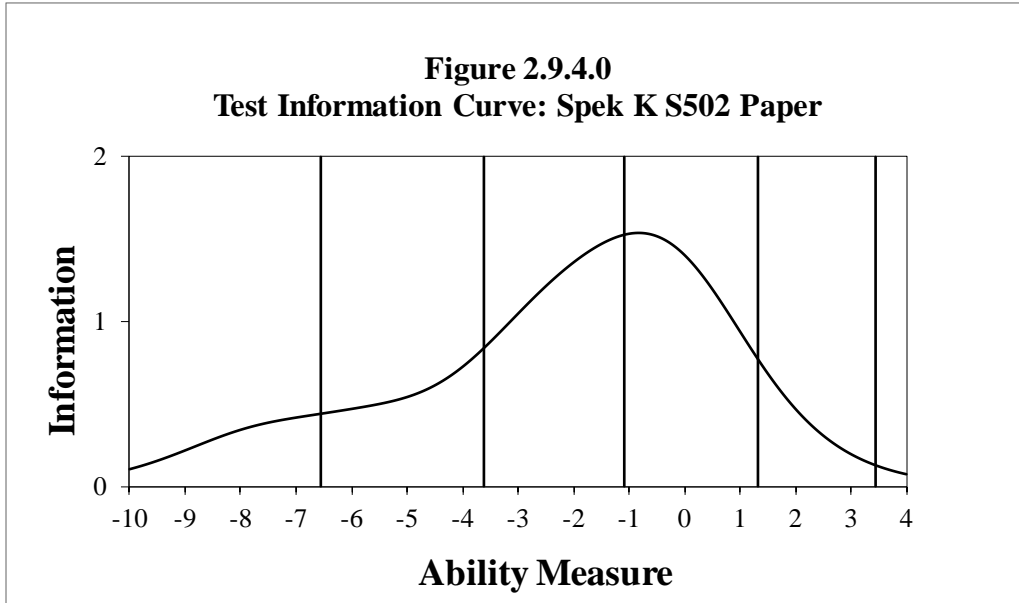


Figure 2.9.3.6.3
Test Information Curve: Writ 9-12 S502 Paper



2.9.4 Speaking

2.9.4.0 Kindergarten



2.9.4.1 Grade 1

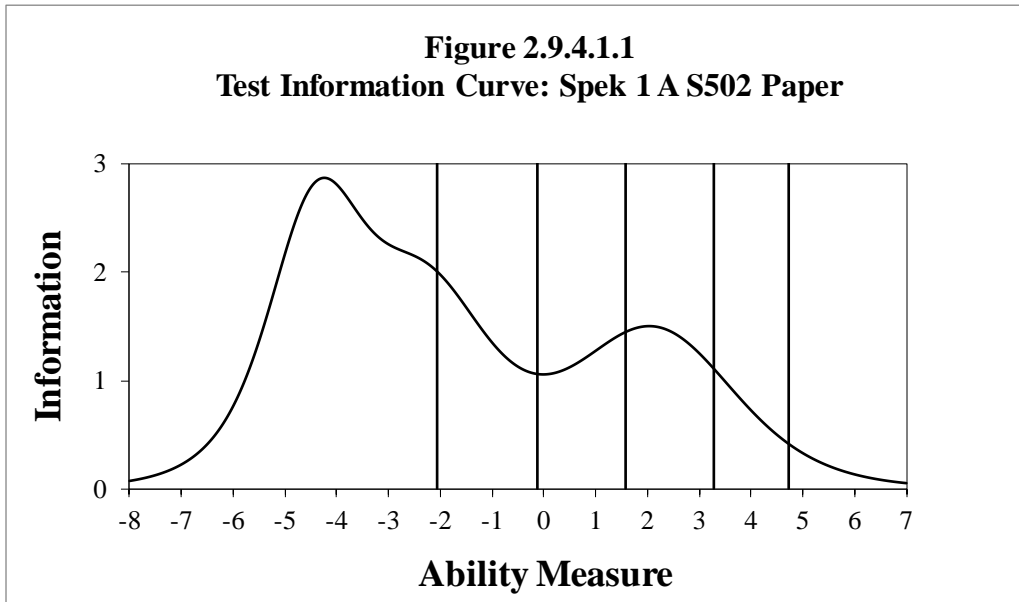


Figure 2.9.4.1.2
Test Information Curve: Spek 1 B/C S502 Paper

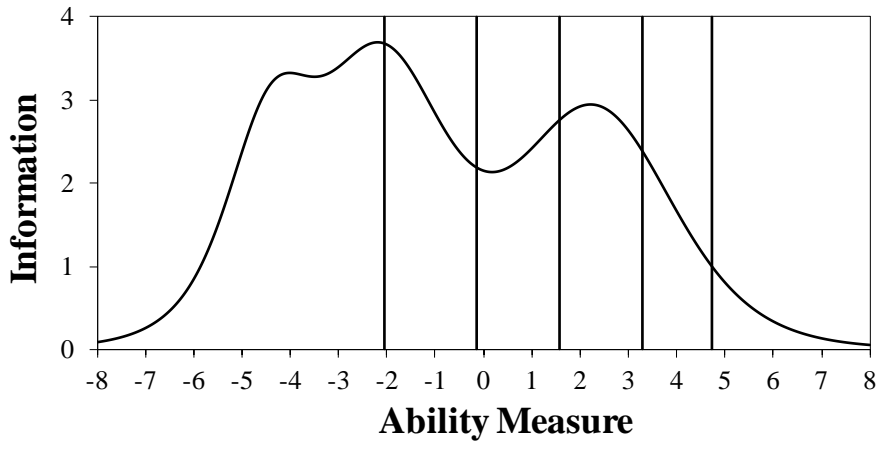
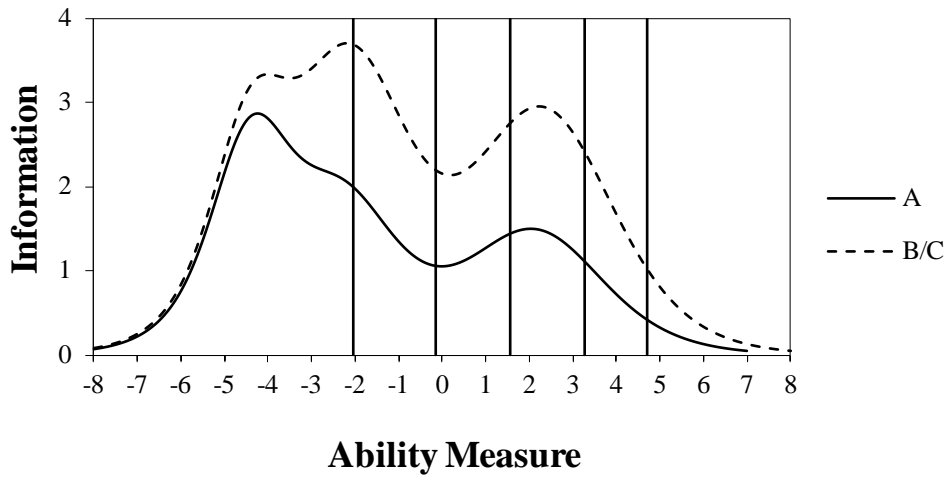
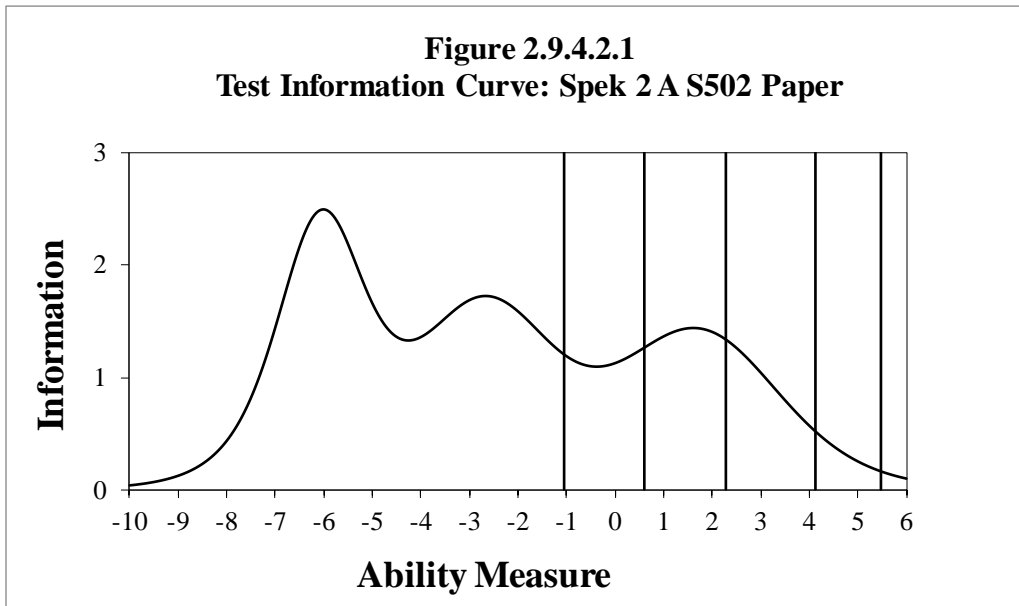


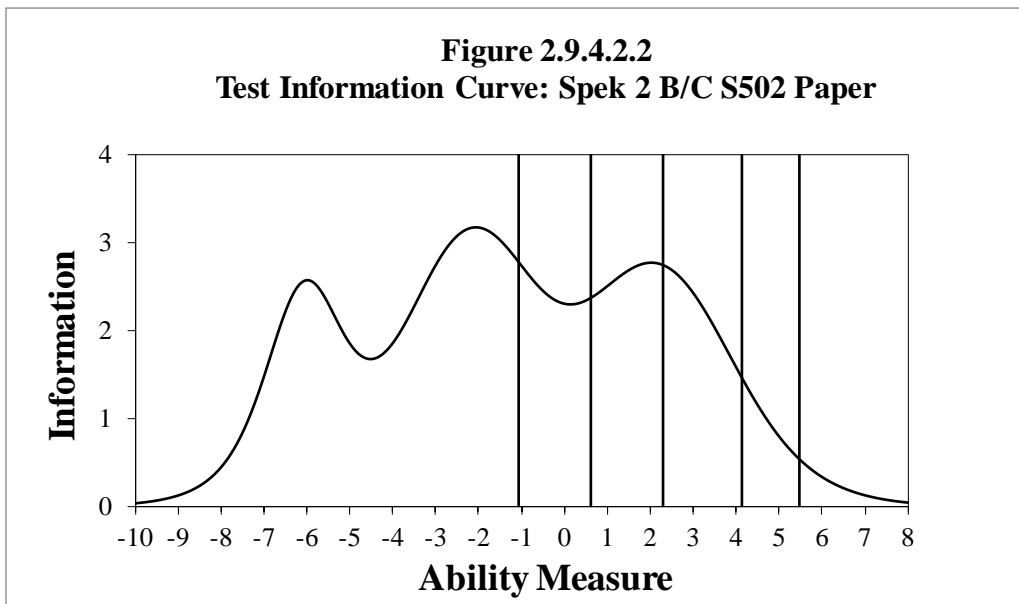
Figure 2.9.4.1.3
Test Information Curve: Spek 1 S502 Paper



2.9.4.2 Grade 2

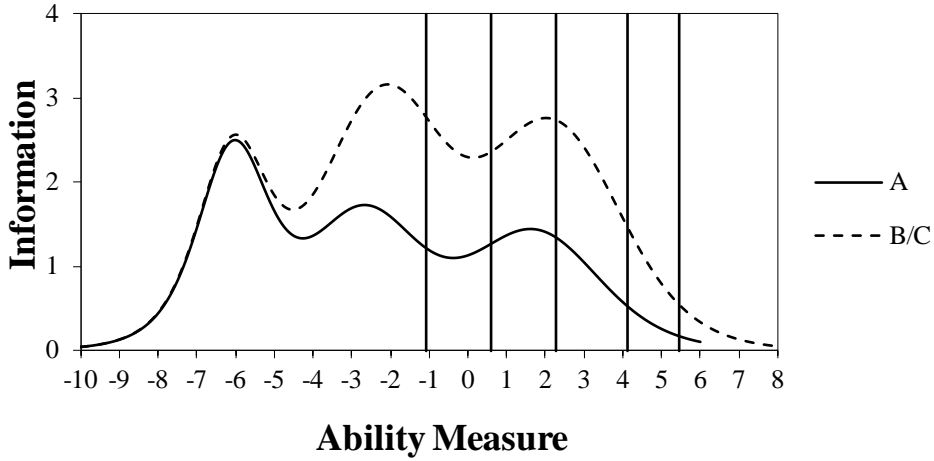


Note: The test form is shared between 2A and 3A.



Note: The test form is shared between 2B/C and 3B/C.

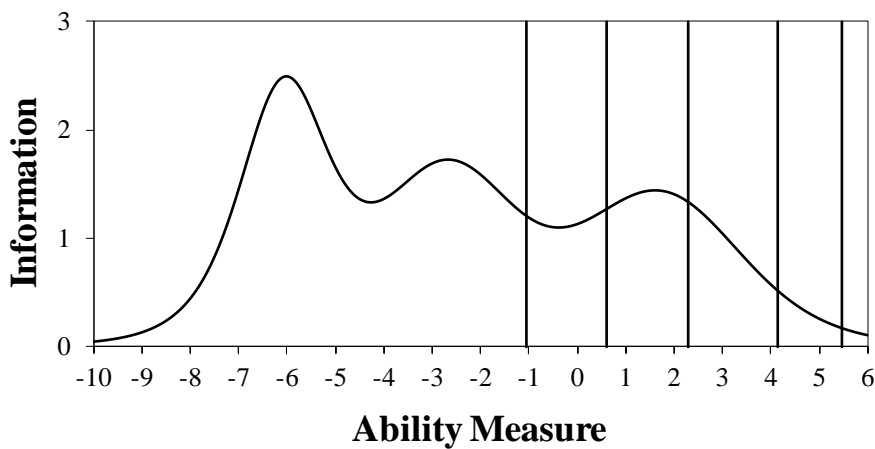
Figure 2.9.4.2.3
Test Information Curve: Spek 2 S502 Paper



Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

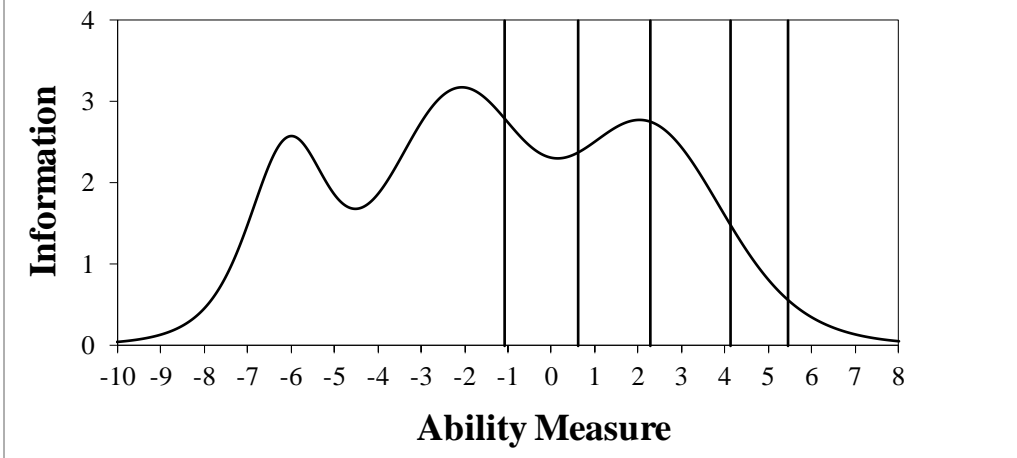
2.9.4.3 Grade 3

Figure 2.9.4.3.1
Test Information Curve: Spek 3 A S502 Paper



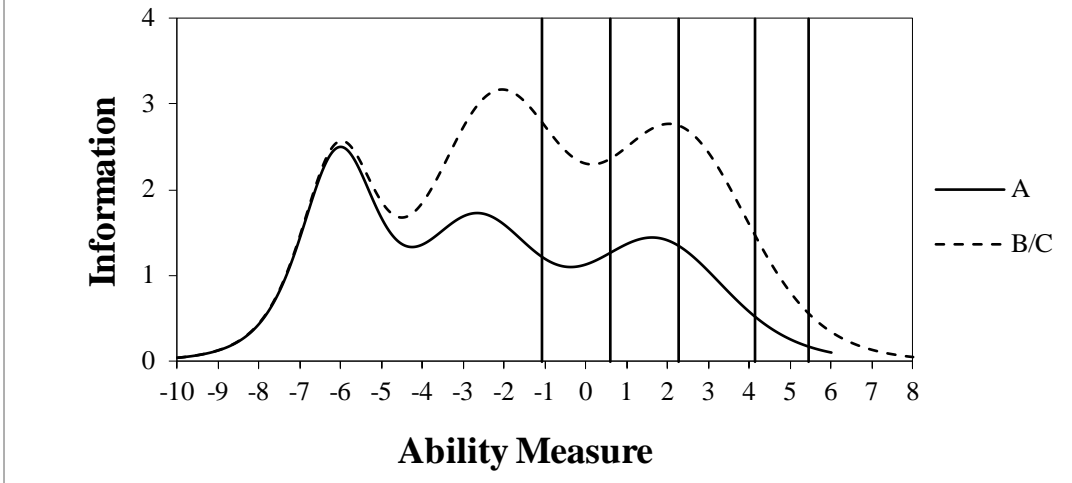
Note: The test form is shared between 2A and 3A.

Figure 2.9.4.3.2
Test Information Curve: Spek 3 B/C S502 Paper



Note: The test form is shared between 2B/C and 3B/C.

Figure 2.9.4.3.3
Test Information Curve: Spek 3 S502 Paper



Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

2.9.4.4 Grades 4–5

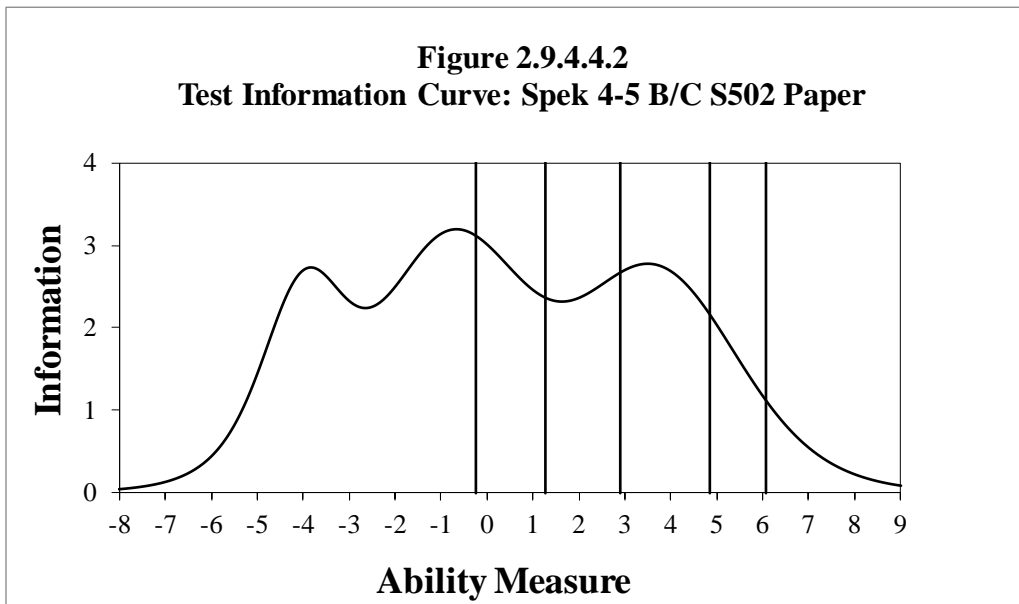
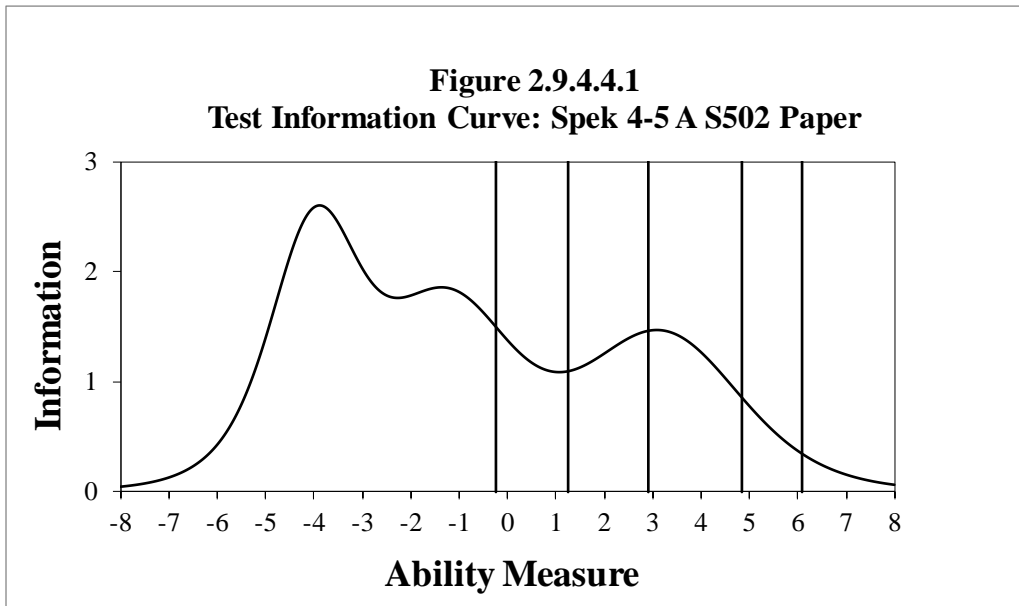
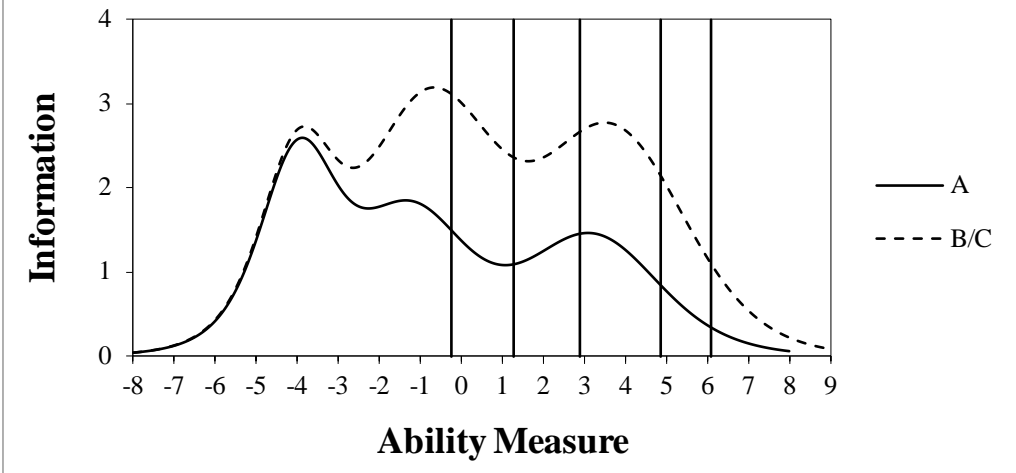


Figure 2.9.4.4.3
Test Information Curve: Spek 4-5 S502 Paper



2.9.4.5 Grades 6–8

Figure 2.9.4.5.1
Test Information Curve: Spek 6-8 A S502 Paper

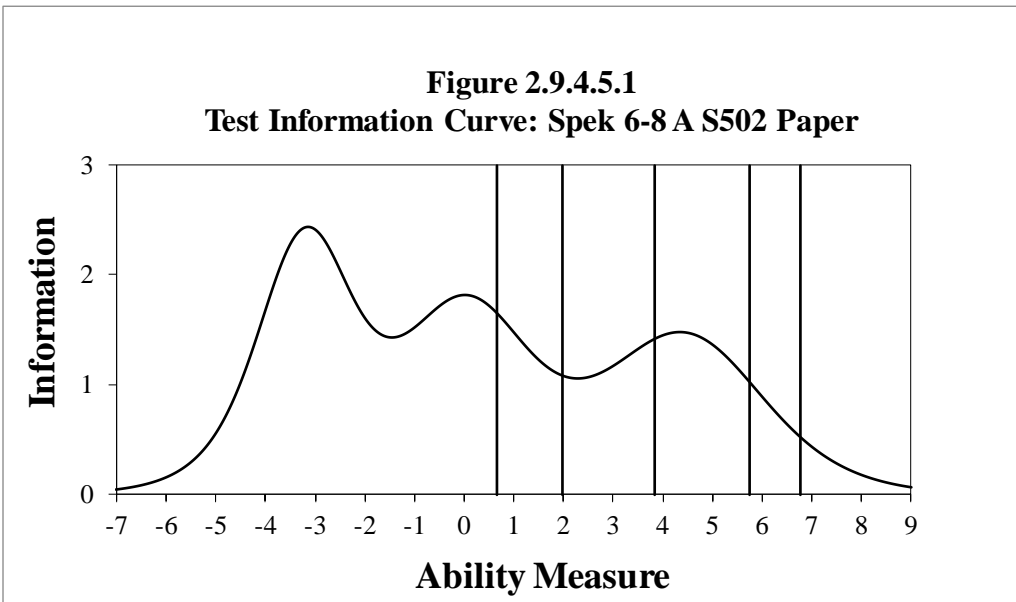


Figure 2.9.4.5.2
Test Information Curve: Spek 6-8 B/C S502 Paper

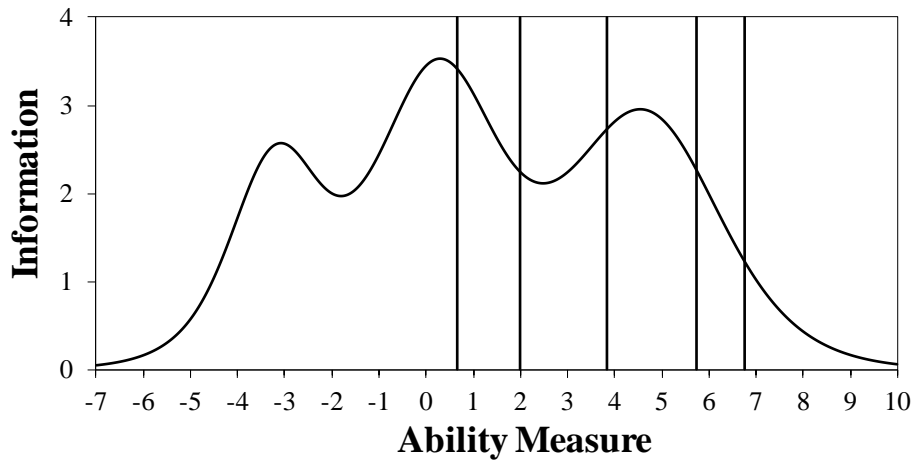
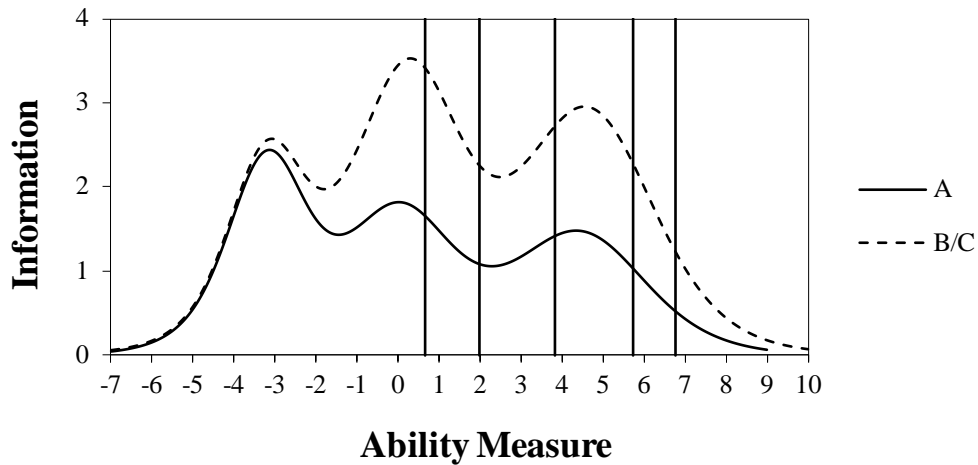


Figure 2.9.4.5.3
Test Information Curve: Spek 6-8 S502 Paper



2.9.4.6 Grades 9–12

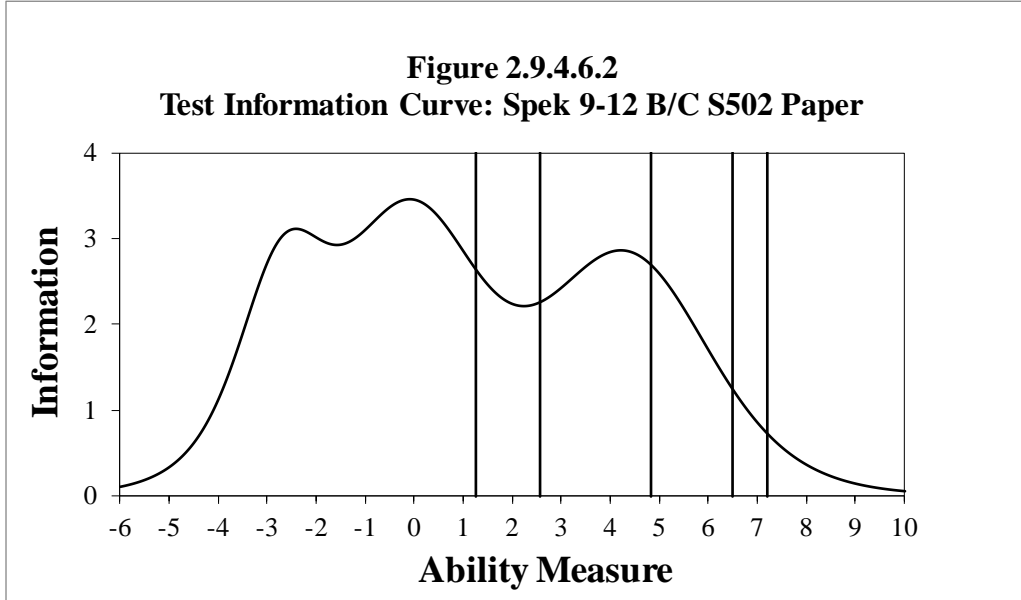
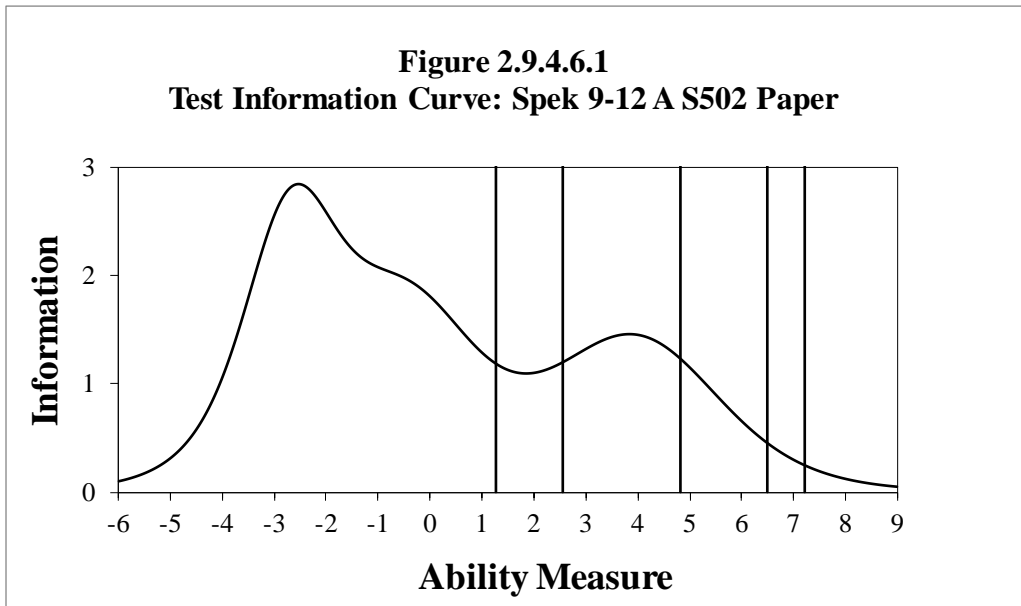
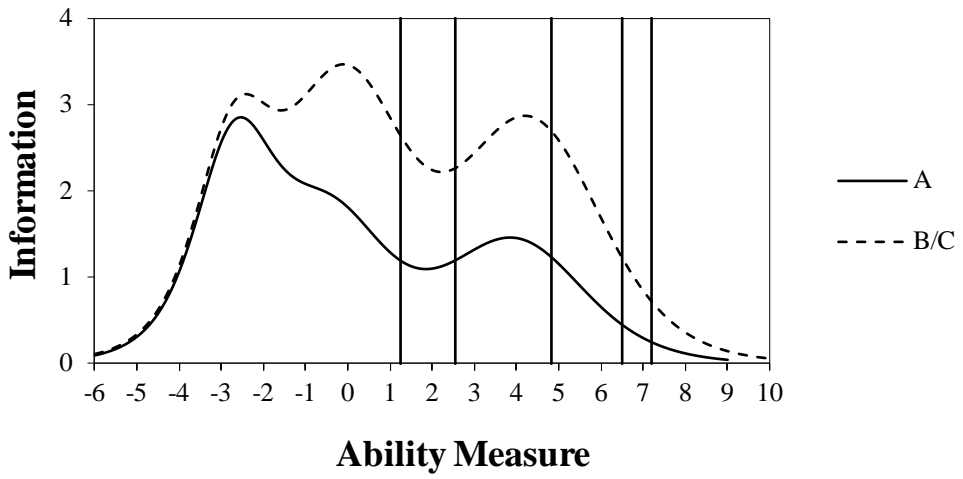


Figure 2.9.4.6.3
Test Information Curve: Spek 9-12 S502 Paper



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

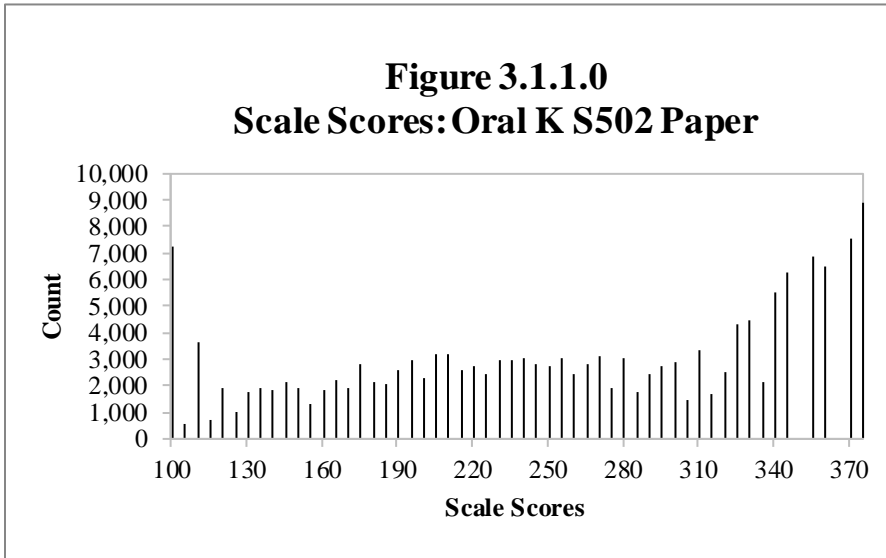
3.1.1 Oral

3.1.1.0 Kindergarten

Table 3.1.1.0

Scale Score Descriptive Statistics: Oral K S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	163,190	100	378	259.45	85.13

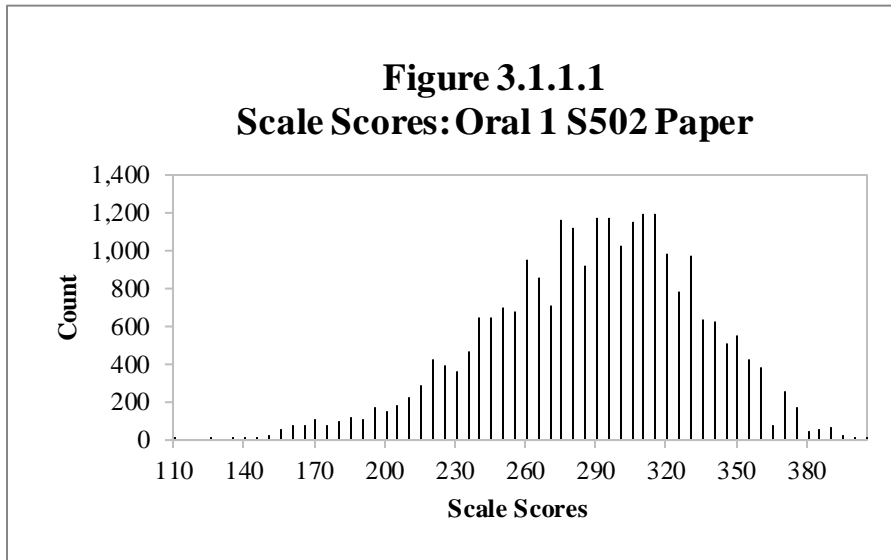


3.1.1.1 Grade 1

Table 3.1.1.1

Scale Score Descriptive Statistics: Oral 1 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	25,299	114	406	289.41	45.08

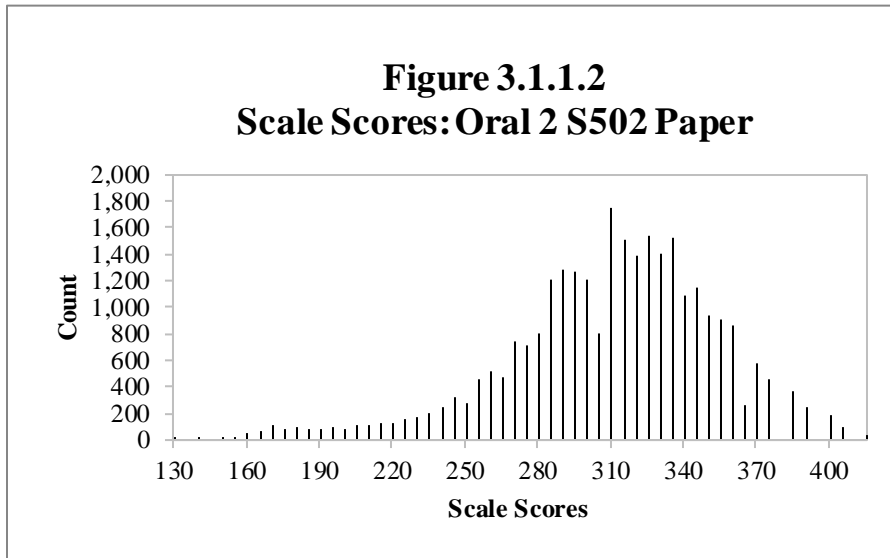


3.1.1.2 Grade 2

Table 3.1.1.2

Scale Score Descriptive Statistics: Oral 2 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	28,446	134	415	310.87	43.67

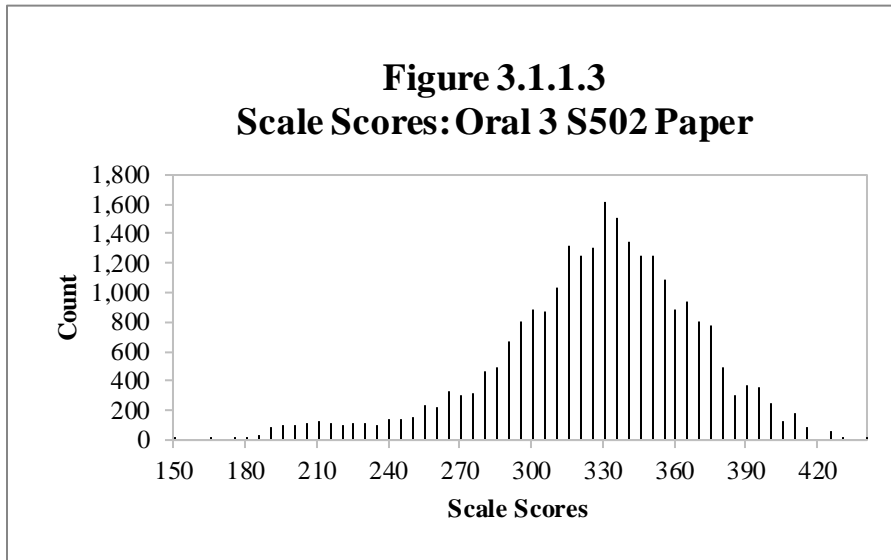


3.1.1.3 Grade 3

Table 3.1.1.3

Scale Score Descriptive Statistics: Oral 3 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	25,723	151	441	327.70	42.70

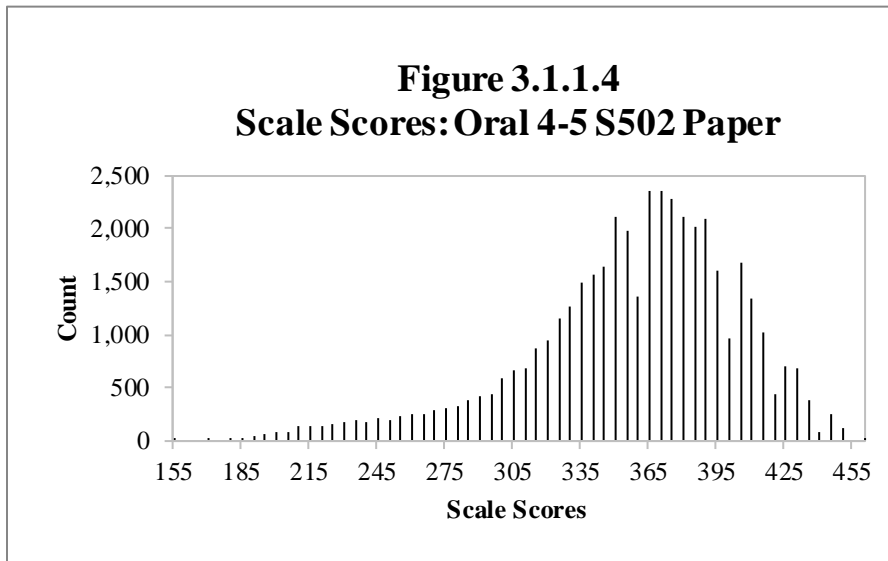


3.1.1.4 Grades 4–5

Table 3.1.1.4

Scale Score Descriptive Statistics: Oral 4-5 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	24,452	173	461	354.75	45.22
5	18,976	157	461	363.22	49.59
Total	43,428	157	461	358.45	47.36

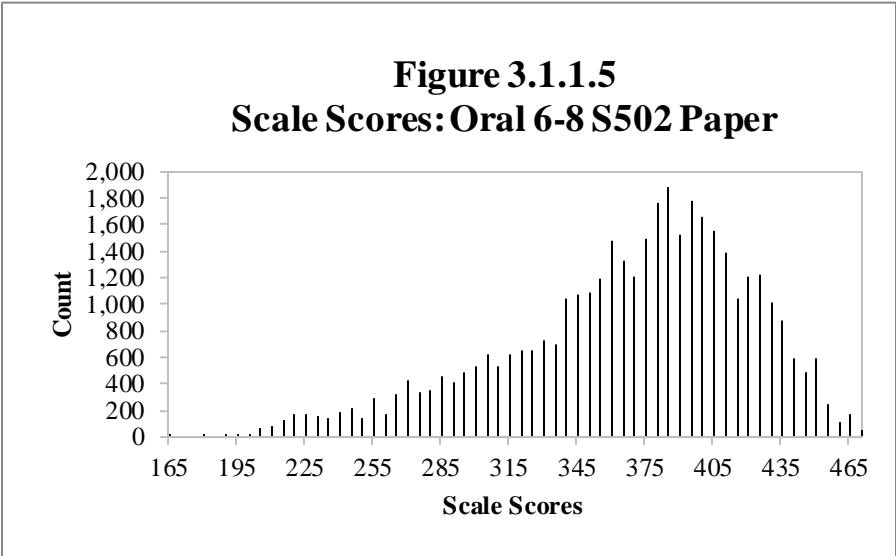


3.1.1.5 Grades 6–8

Table 3.1.1.5

Scale Score Descriptive Statistics: Oral 6-8 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	14,236	180	474	365.09	50.88
7	13,020	165	474	369.77	53.64
8	11,614	180	474	372.04	57.63
Total	38,870	165	474	368.73	53.97



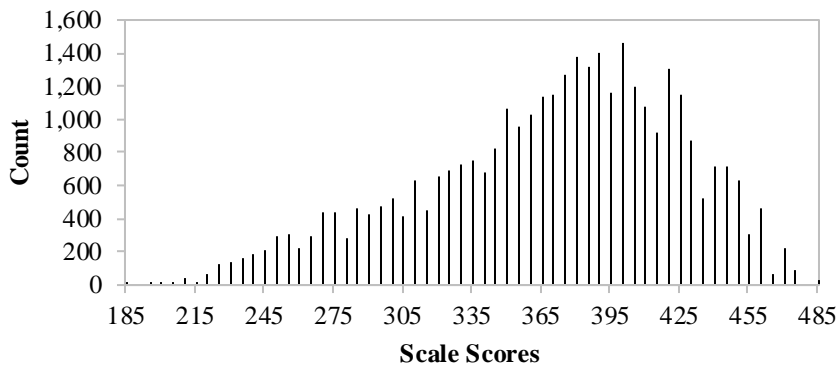
3.1.1.6 Grades 9–12

Table 3.1.1.6

Scale Score Descriptive Statistics: Oral 9-12 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	10,512	187	485	368.93	57.27
10	9,745	198	485	366.96	57.48
11	8,433	198	485	373.34	55.16
12	5,801	214	485	373.17	51.58
Total	34,491	187	485	370.17	55.96

Figure 3.1.1.6
Scale Scores: Oral 9-12 S502 Paper



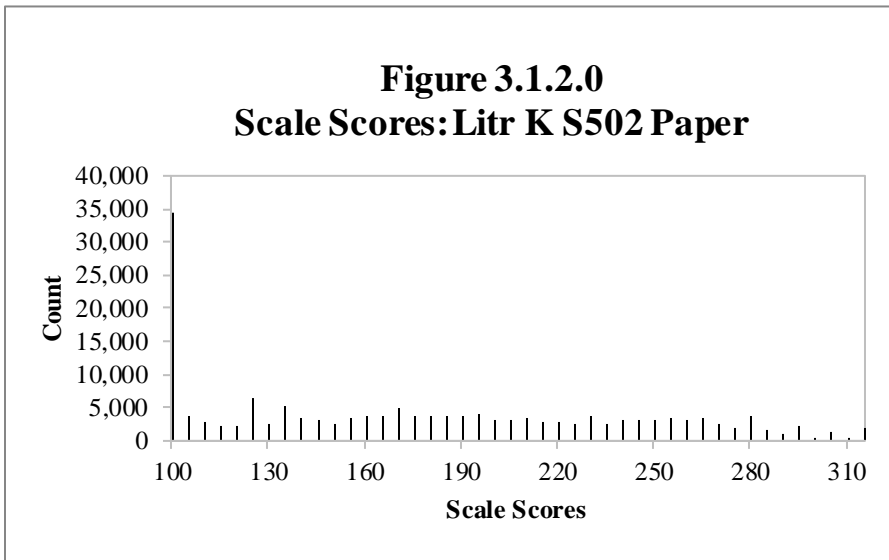
3.1.2 Literacy

3.1.2.0 Kindergarten

Table 3.1.2.0

Scale Score Descriptive Statistics: Litr K S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	163,215	100	315	178.22	64.60

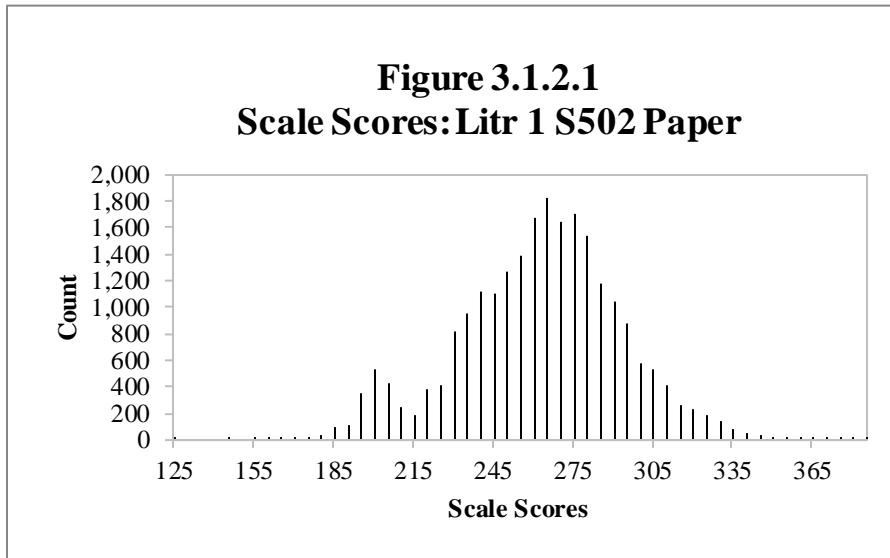


3.1.2.1 Grade 1

Table 3.1.2.1

Scale Score Descriptive Statistics: Litr 1 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	23,491	126	387	264.47	30.93

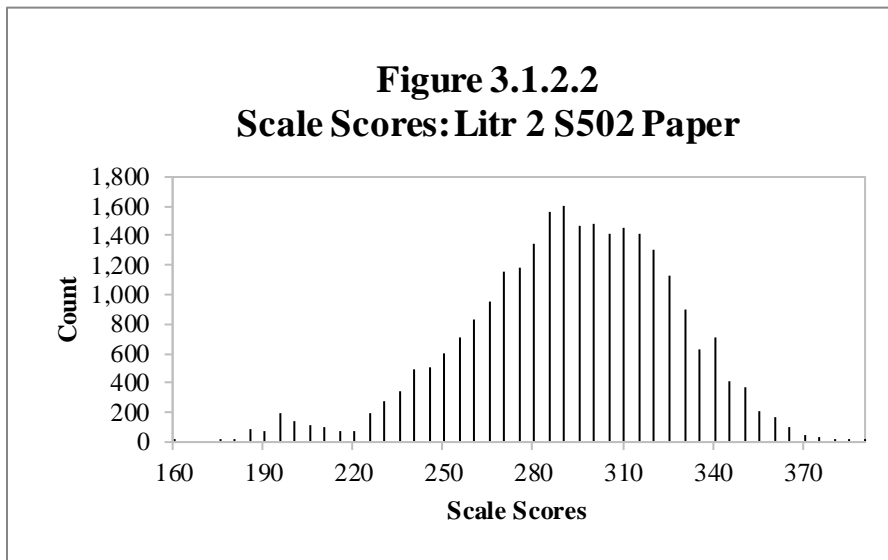


3.1.2.2 Grade 2

Table 3.1.2.2

Scale Score Descriptive Statistics: Litr 2 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	25,898	160	394	293.43	34.46

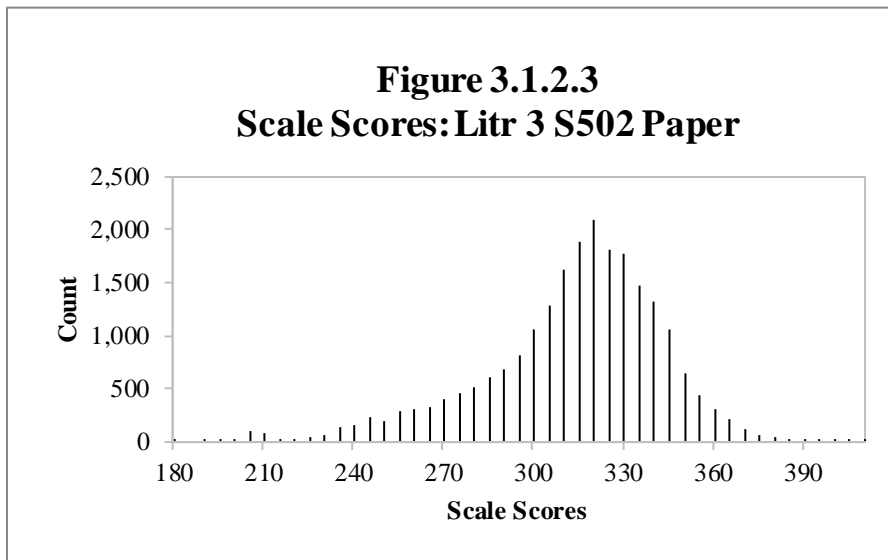


3.1.2.3 Grade 3

Table 3.1.2.3

Scale Score Descriptive Statistics: Litr 3 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	22,799	183	413	314.55	30.67

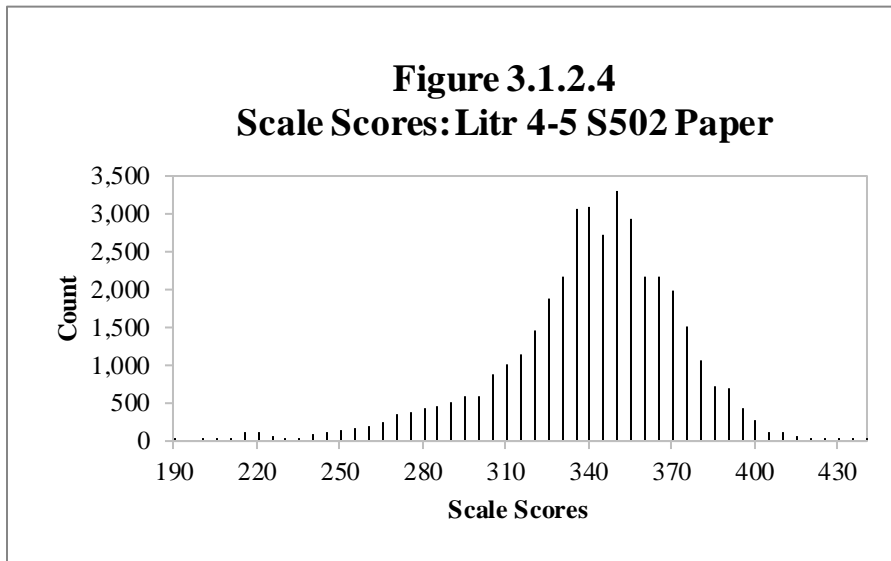


3.1.2.4 Grades 4–5

Table 3.1.2.4

Scale Score Descriptive Statistics: Litr 4-5 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	22,109	194	431	336.51	30.83
5	17,535	205	442	347.98	33.97
Total	39,644	194	442	341.58	32.75

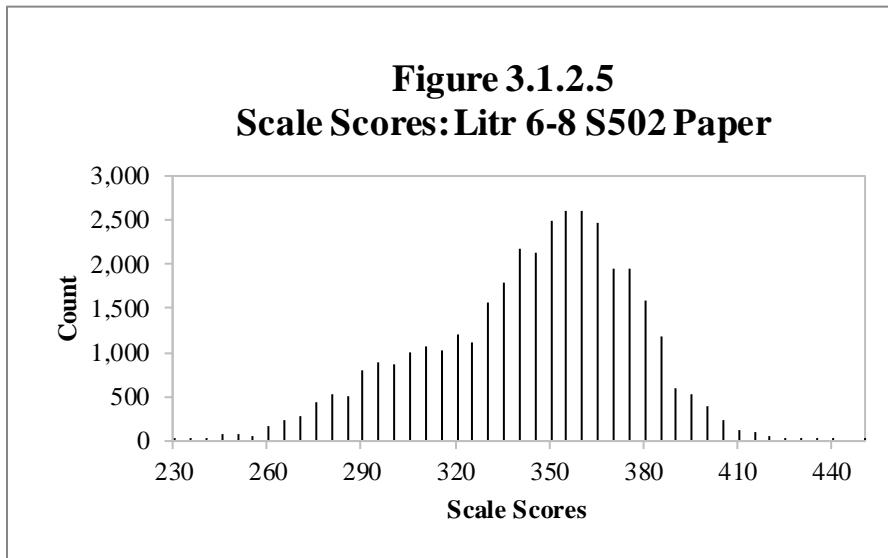


3.1.2.5 Grades 6–8

Table 3.1.2.5

Scale Score Descriptive Statistics: Litr 6-8 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	13,265	233	429	340.08	29.97
7	12,341	233	441	346.31	32.21
8	11,251	233	454	350.54	35.11
Total	36,857	233	454	345.36	32.64

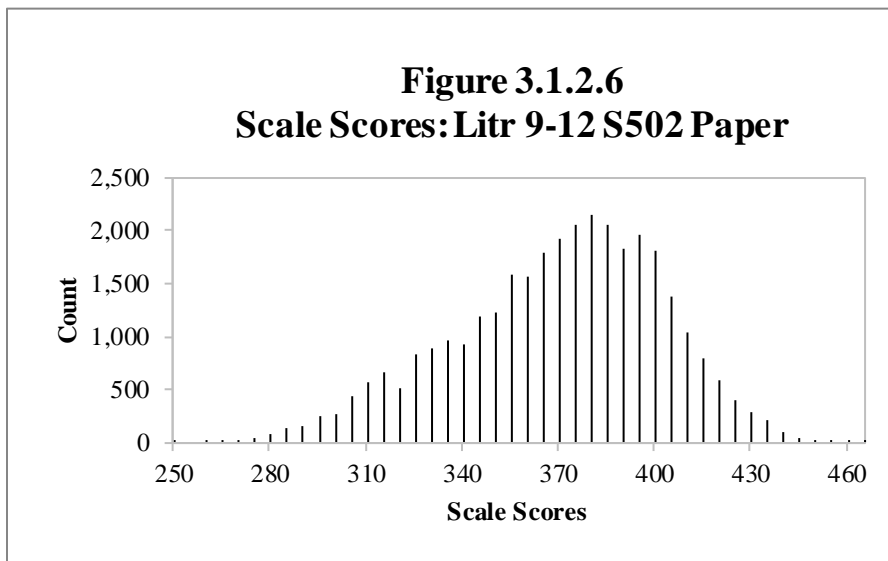


3.1.2.6 Grades 9–12

Table 3.1.2.6

Scale Score Descriptive Statistics: Litr 9-12 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	9,837	252	467	367.61	32.44
10	9,250	252	458	370.39	33.54
11	8,109	268	462	375.25	32.88
12	5,600	271	454	374.77	30.52
Total	32,796	252	467	371.51	32.70



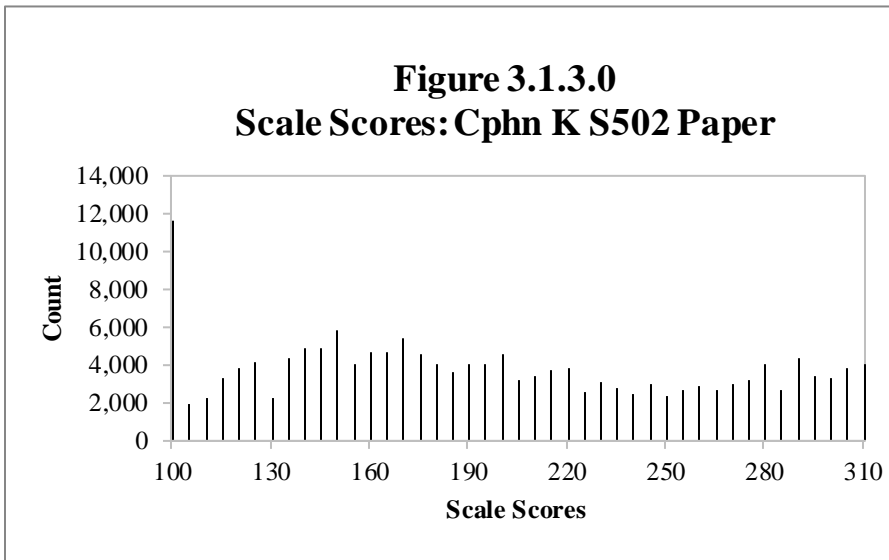
3.1.3 Comprehension

3.1.3.0 Kindergarten

Table 3.1.3.0

Scale Score Descriptive Statistics: Cphn K S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	163,215	100	312	198.34	63.44

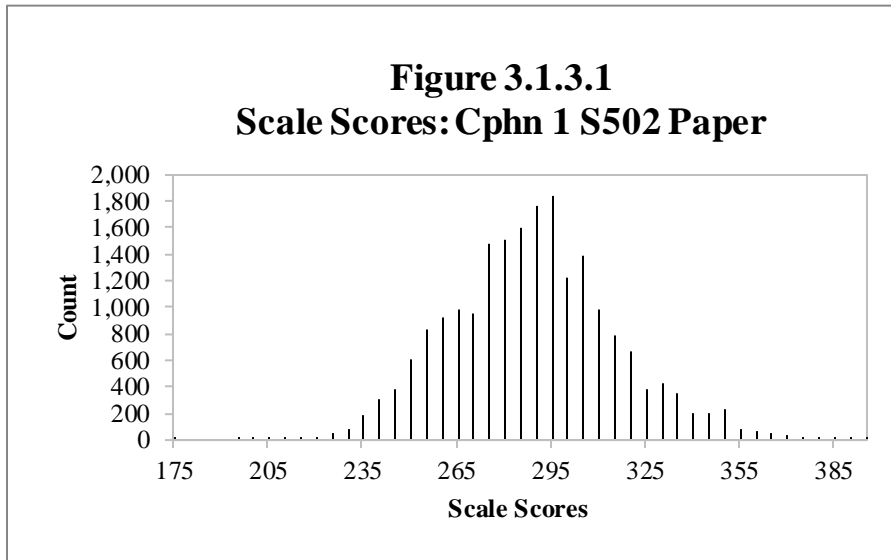


3.1.3.1 Grade 1

Table 3.1.3.1

Scale Score Descriptive Statistics: Cphn 1 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	20,703	179	397	291.07	26.58

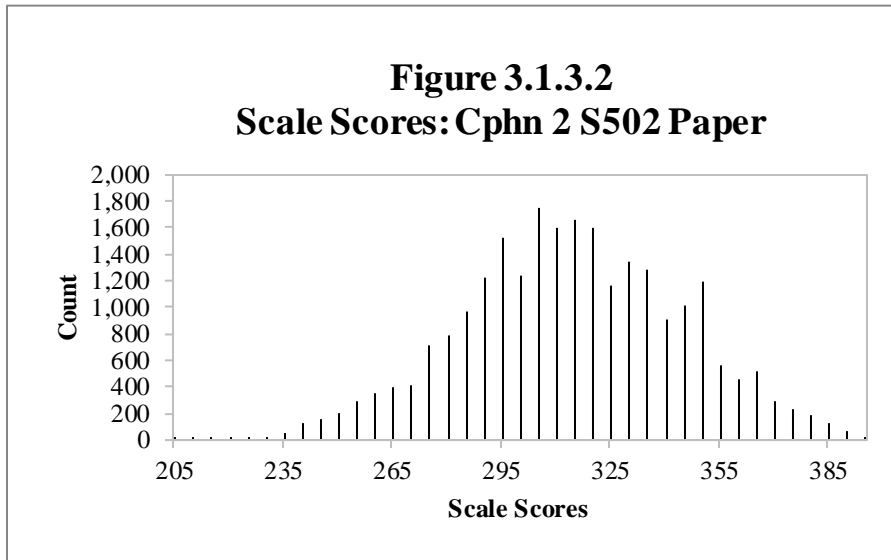


3.1.3.2 Grade 2

Table 3.1.3.2

Scale Score Descriptive Statistics: Cphn 2 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	24,478	209	397	316.19	30.38

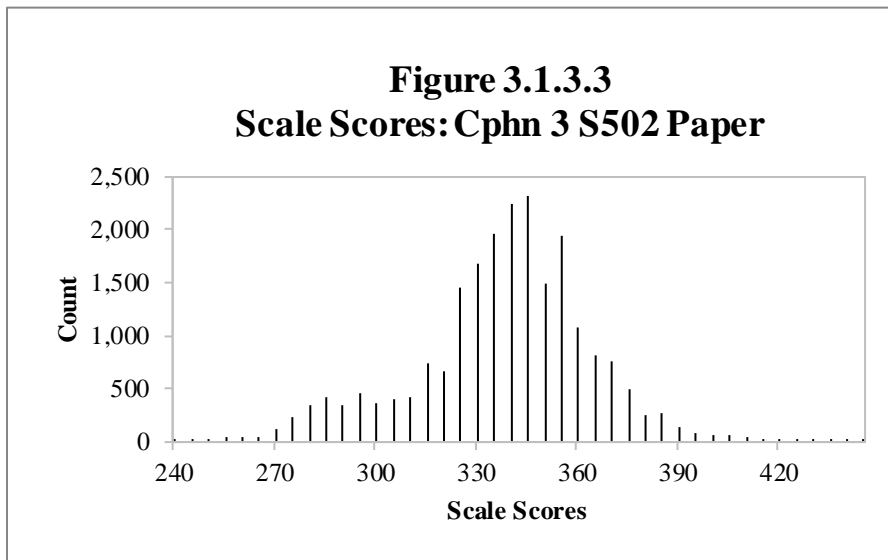


3.1.3.3 Grade 3

Table 3.1.3.3

Scale Score Descriptive Statistics: Cphn 3 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	21,869	242	448	339.61	26.19

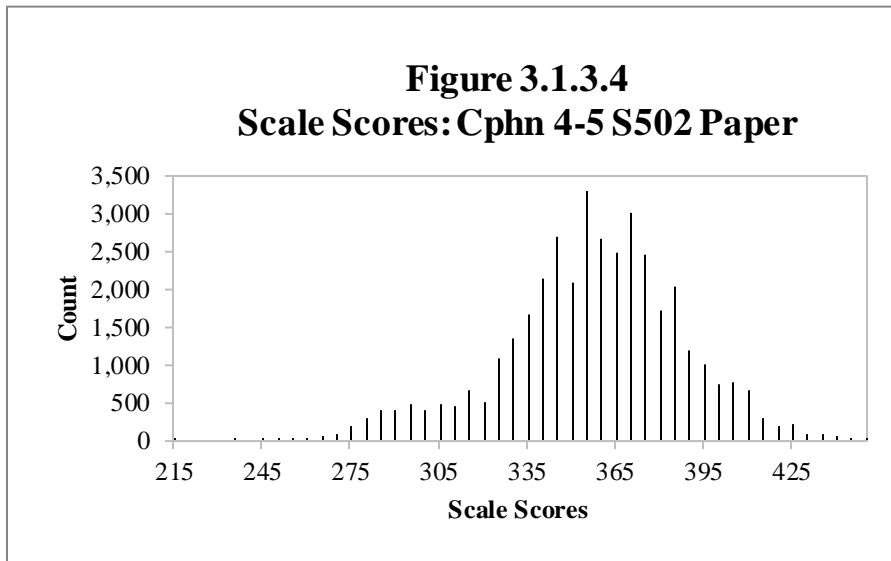


3.1.3.4 Grades 4–5

Table 3.1.3.4

Scale Score Descriptive Statistics: Cphn 4-5 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	21,457	218	453	354.49	28.67
5	17,131	249	453	364.37	32.41
Total	38,588	218	453	358.87	30.78

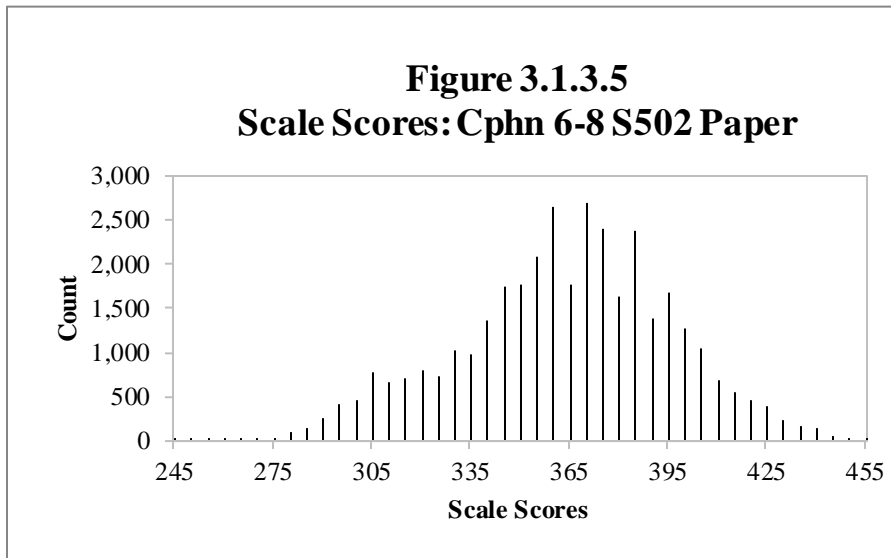


3.1.3.5 Grades 6–8

Table 3.1.3.5

Scale Score Descriptive Statistics: Cphn 6-8 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	12,829	246	455	360.19	28.54
7	11,954	258	459	366.70	31.89
8	10,890	254	459	371.37	35.40
Total	35,673	246	459	365.78	32.21

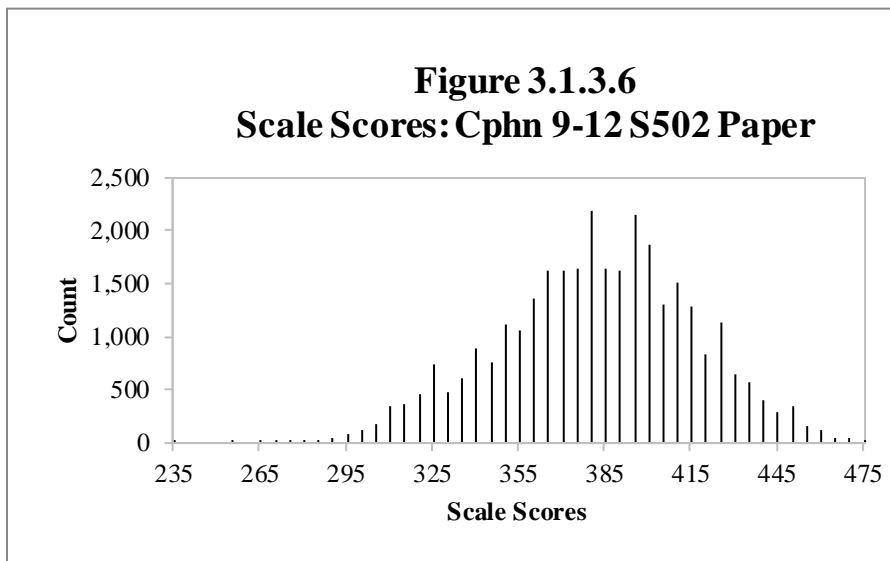


3.1.3.6 Grades 9–12

Table 3.1.3.6

Scale Score Descriptive Statistics: Cphn 9-12 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	9,450	235	473	380.54	33.48
10	8,948	266	477	382.76	34.87
11	7,846	287	477	387.66	34.48
12	5,421	257	473	386.01	31.59
Total	31,665	235	477	383.87	33.93



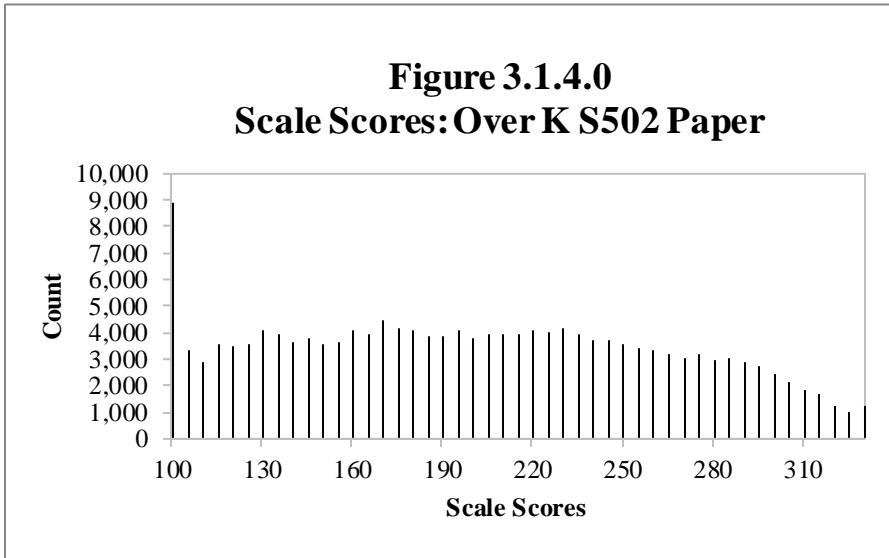
3.1.4 Overall

3.1.4.0 Kindergarten

Table 3.1.4.0

Scale Score Descriptive Statistics: Over K S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	163,178	100	333	202.38	63.23

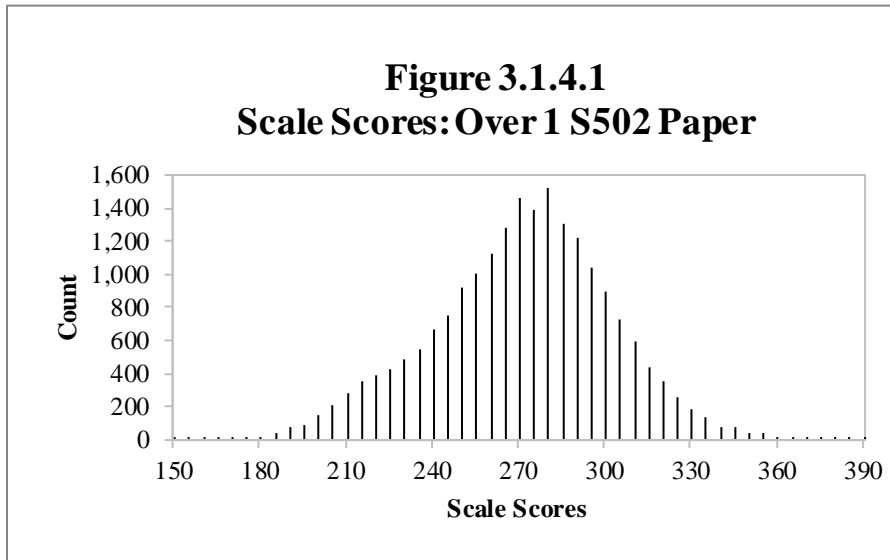


3.1.4.1 Grade 1

Table 3.1.4.1

Scale Score Descriptive Statistics: Over 1 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	20,595	153	391	272.66	31.25

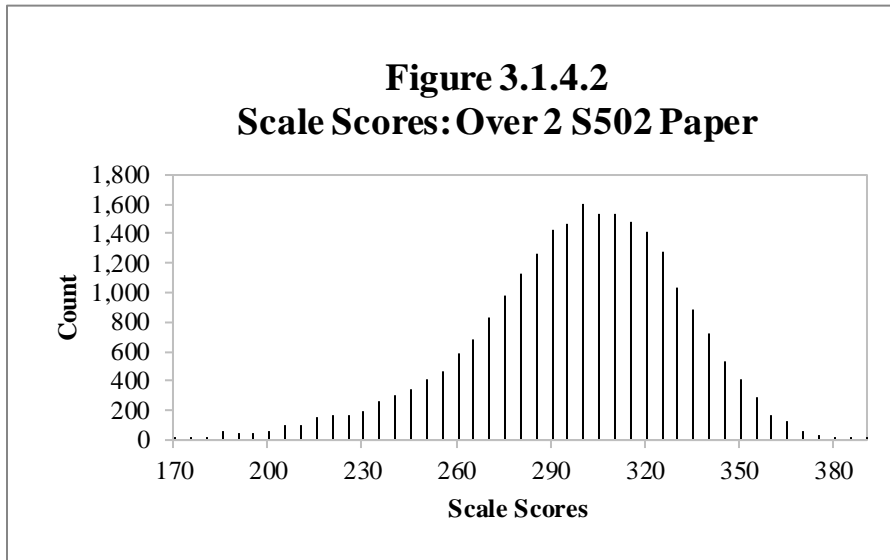


3.1.4.2 Grade 2

Table 3.1.4.2

Scale Score Descriptive Statistics: Over 2 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	24,343	171	390	299.06	33.58

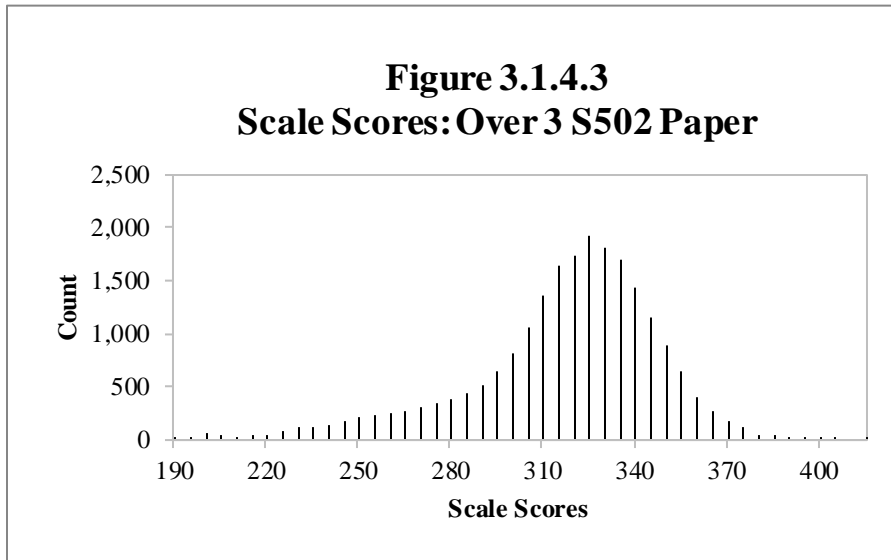


3.1.4.3 Grade 3

Table 3.1.4.3

Scale Score Descriptive Statistics: Over 3 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	21,742	190	416	318.49	31.28

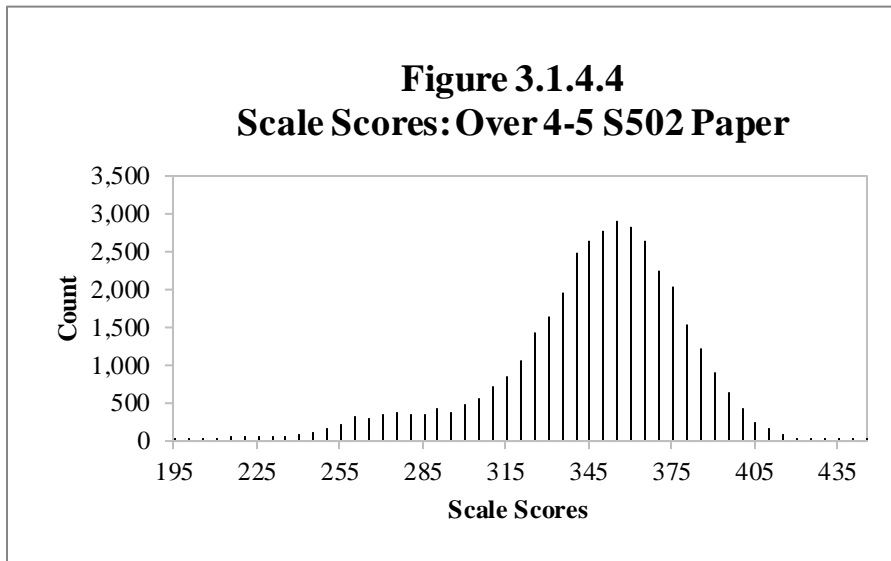


3.1.4.4 Grades 4–5

Table 3.1.4.4

Scale Score Descriptive Statistics: Over 4-5 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	21,338	198	440	342.05	32.41
5	17,078	201	446	352.78	36.10
Total	38,416	198	446	346.82	34.51

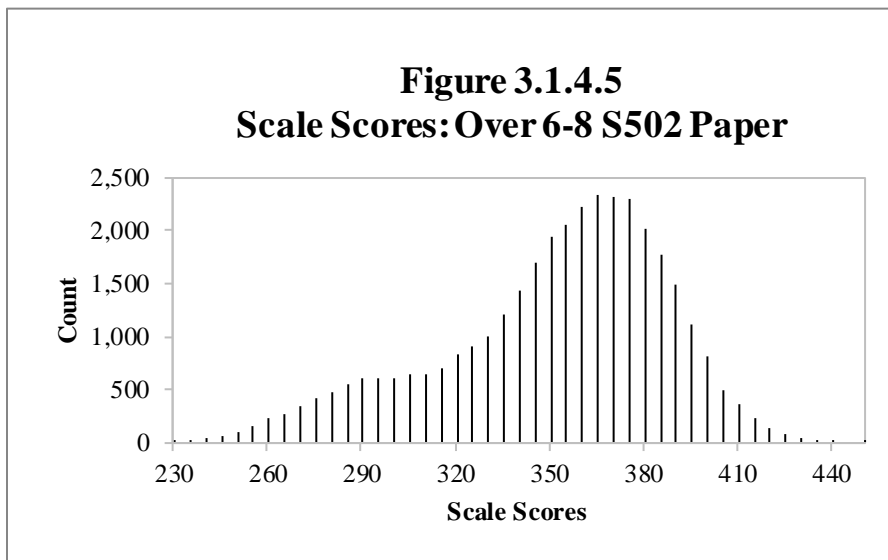


3.1.4.5 Grades 6–8

Table 3.1.4.5

Scale Score Descriptive Statistics: Over 6-8 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	12,731	230	435	347.80	33.62
7	11,868	230	443	353.69	36.23
8	10,819	233	451	357.35	39.72
Total	35,418	230	451	352.69	36.65

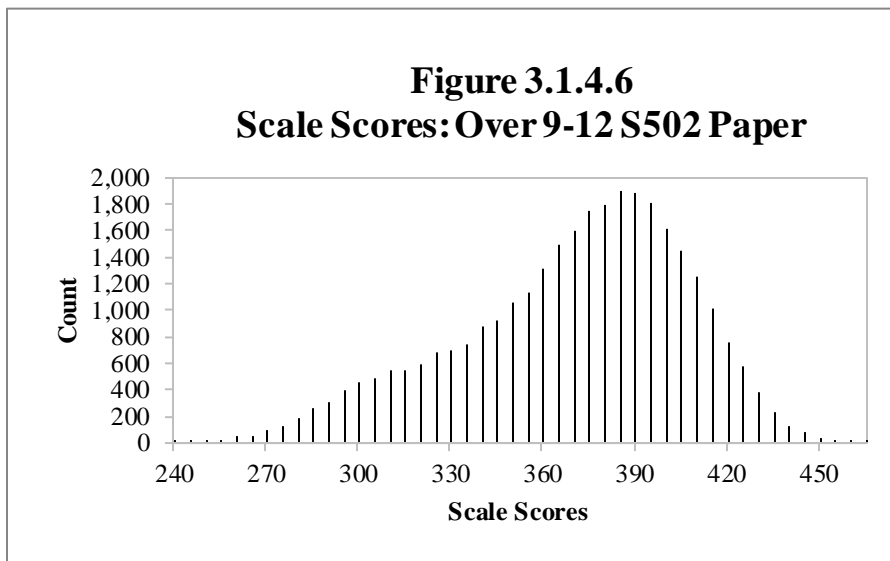


3.1.4.6 Grades 9–12

Table 3.1.4.6

Scale Score Descriptive Statistics: Over 9-12 S502 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	9,355	242	461	368.34	37.49
10	8,860	247	465	369.62	38.48
11	7,745	255	464	375.00	37.14
12	5,352	261	457	374.66	34.04
Total	31,312	242	465	371.43	37.24



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

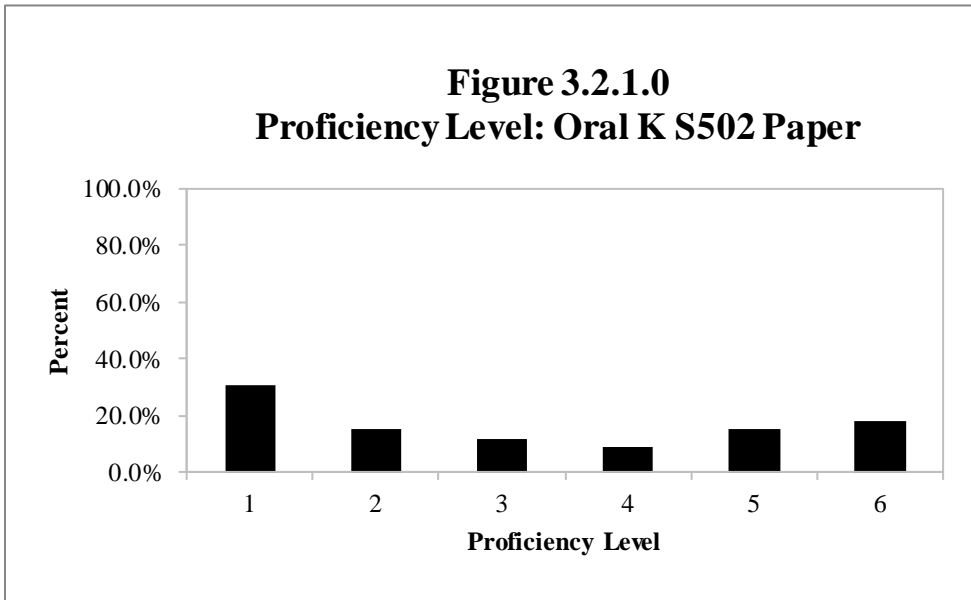
3.2.1 Oral

3.2.1.0 Kindergarten

Table 3.2.1.0

Proficiency Level Distribution: Oral K S502 Paper

Level	Count	Percent
1	49,934	30.6%
2	24,535	15.0%
3	19,036	11.7%
4	14,617	9.0%
5	25,236	15.5%
6	29,832	18.3%
Total	163,190	100.0%

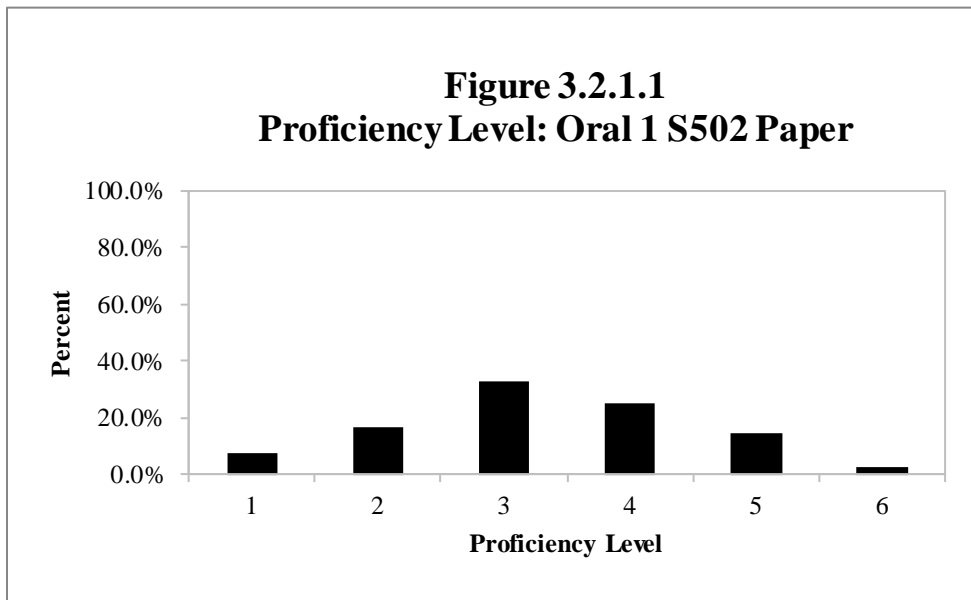


3.2.1.1 Grade 1

Table 3.2.1.1

Proficiency Level Distribution: Oral 1 S502 Paper

Level	Count	Percent
1	1,872	7.4%
2	4,226	16.7%
3	8,331	32.9%
4	6,376	25.2%
5	3,778	14.9%
6	716	2.8%
Total	25,299	100.0%

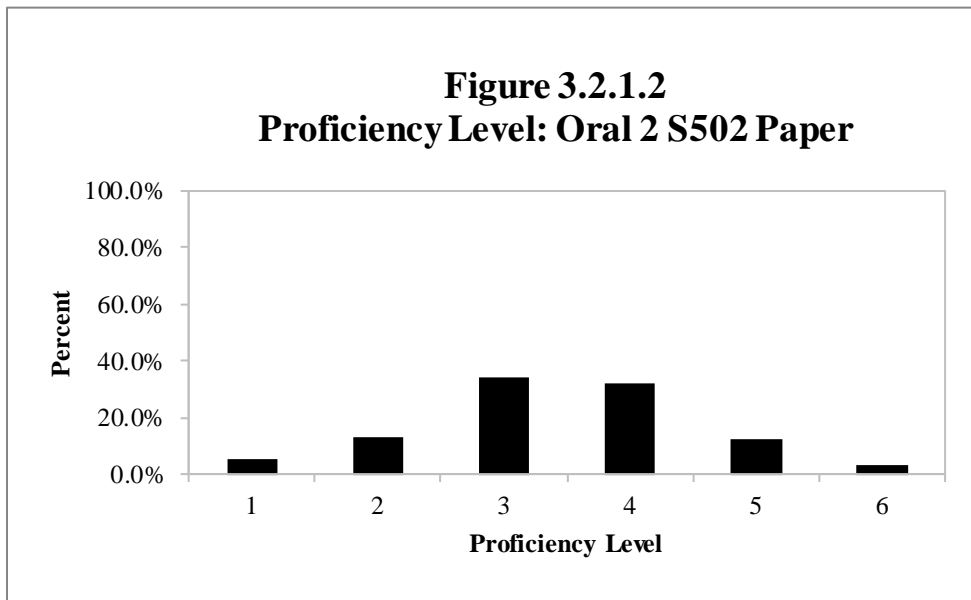


3.2.1.2 Grade 2

Table 3.2.1.2

Proficiency Level Distribution: Oral 2 S502 Paper

Level	Count	Percent
1	1,542	5.4%
2	3,732	13.1%
3	9,666	34.0%
4	9,071	31.9%
5	3,505	12.3%
6	930	3.3%
Total	28,446	100.0%

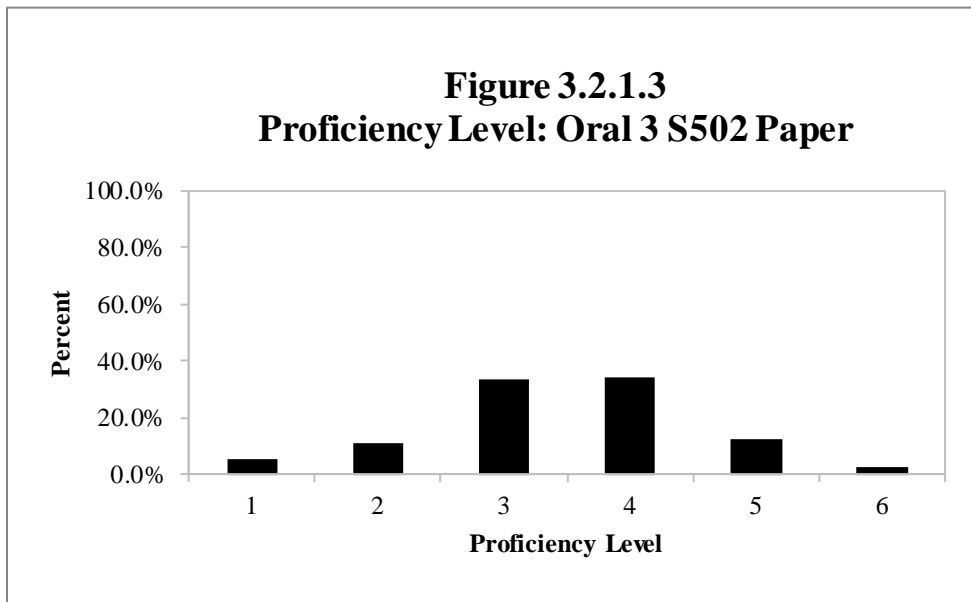


3.2.1.3 Grade 3

Table 3.2.1.3

Proficiency Level Distribution: Oral 3 S502 Paper

Level	Count	Percent
1	1,388	5.4%
2	2,866	11.1%
3	8,652	33.6%
4	8,807	34.2%
5	3,299	12.8%
6	711	2.8%
Total	25,723	100.0%

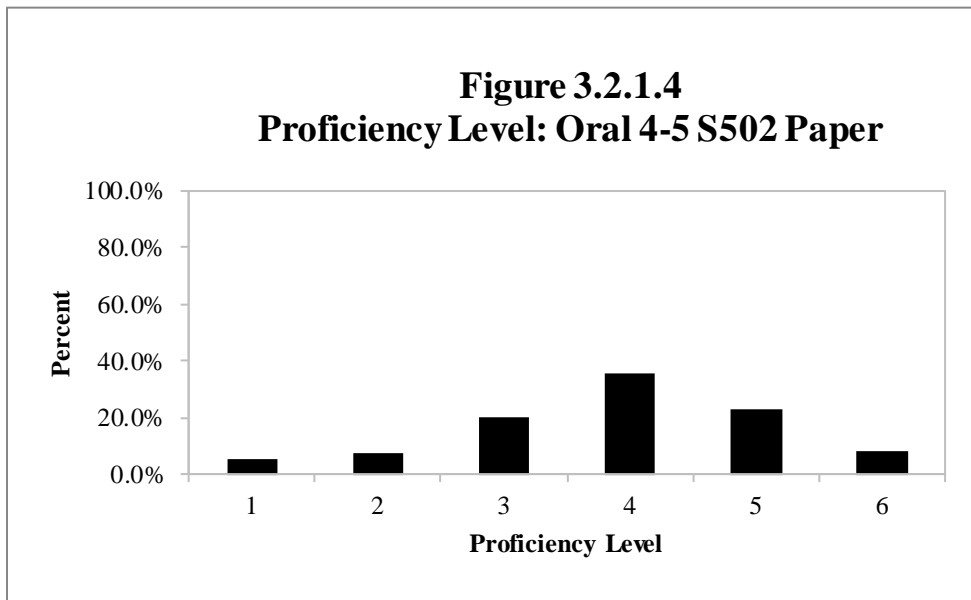


3.2.1.4 Grades 4–5

Table 3.2.1.4

Proficiency Level Distribution: Oral 4-5 S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	1,080	4.4%	1,274	6.7%	2,354	5.4%
2	1,824	7.5%	1,400	7.4%	3,224	7.4%
3	5,391	22.0%	3,441	18.1%	8,832	20.3%
4	8,639	35.3%	6,894	36.3%	15,533	35.8%
5	5,542	22.7%	4,465	23.5%	10,007	23.0%
6	1,976	8.1%	1,502	7.9%	3,478	8.0%
Total	24,452	100.0%	18,976	100.0%	43,428	100.0%

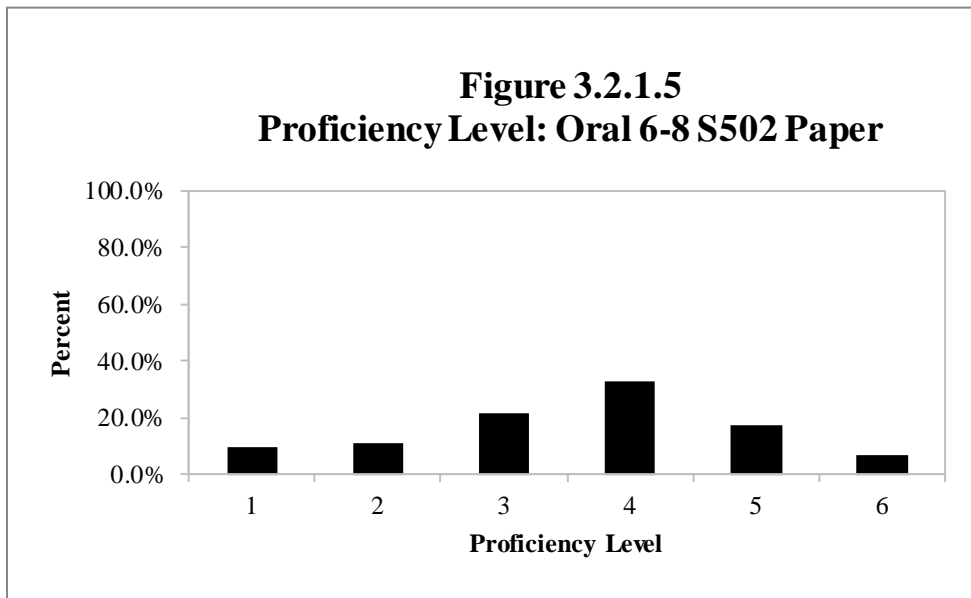


3.2.1.5 Grades 6–8

Table 3.2.1.5

Proficiency Level Distribution: Oral 6-8 S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,081	7.6%	1,263	9.7%	1,445	12.4%	3,789	9.7%
2	1,455	10.2%	1,426	11.0%	1,334	11.5%	4,215	10.8%
3	3,221	22.6%	2,842	21.8%	2,444	21.0%	8,507	21.9%
4	4,822	33.9%	4,411	33.9%	3,674	31.6%	12,907	33.2%
5	2,651	18.6%	2,270	17.4%	1,913	16.5%	6,834	17.6%
6	1,006	7.1%	808	6.2%	804	6.9%	2,618	6.7%
Total	14,236	100.0%	13,020	100.0%	11,614	100.0%	38,870	100.0%

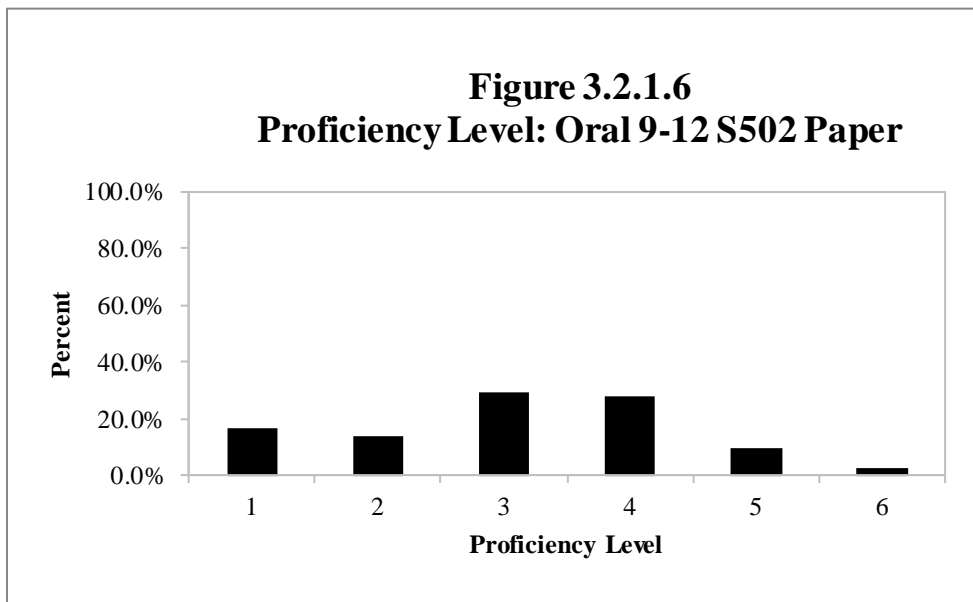


3.2.1.6 Grades 9–12

Table 3.2.1.6

Proficiency Level Distribution: Oral 9-12 S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,571	14.9%	1,748	17.9%	1,413	16.8%	985	17.0%	5,717	16.6%
2	1,366	13.0%	1,398	14.3%	1,222	14.5%	866	14.9%	4,852	14.1%
3	2,654	25.2%	2,807	28.8%	2,504	29.7%	2,055	35.4%	10,020	29.1%
4	3,202	30.5%	2,562	26.3%	2,319	27.5%	1,480	25.5%	9,563	27.7%
5	1,338	12.7%	921	9.5%	742	8.8%	299	5.2%	3,300	9.6%
6	381	3.6%	309	3.2%	233	2.8%	116	2.0%	1,039	3.0%
Total	10,512	100.0%	9,745	100.0%	8,433	100.0%	5,801	100.0%	34,491	100.0%



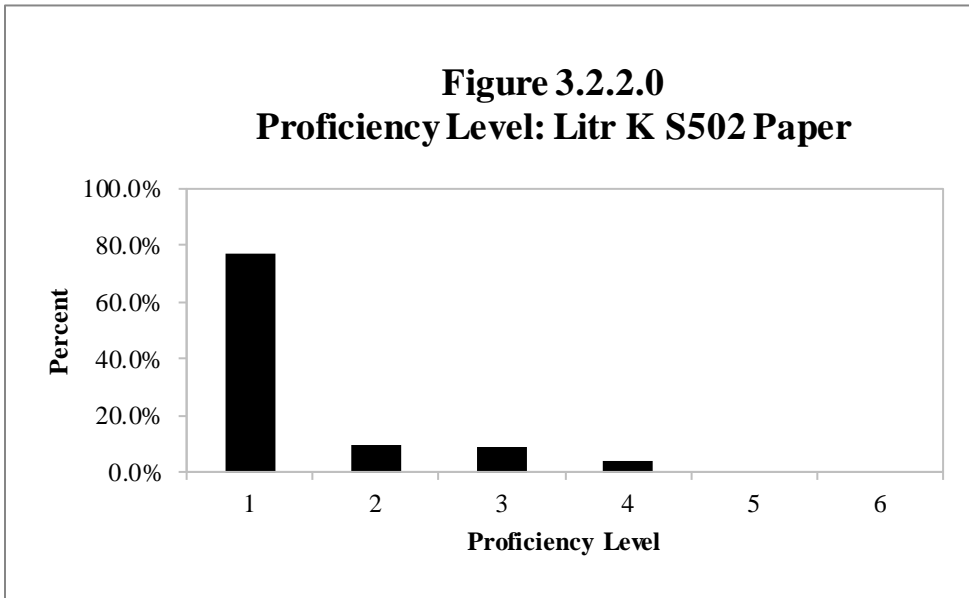
3.2.2 Literacy

3.2.2.0 Kindergarten

Table 3.2.2.0

Proficiency Level Distribution: Litr K S502 Paper

Level	Count	Percent
1	126,099	77.3%
2	16,447	10.1%
3	14,331	8.8%
4	6,338	3.9%
5	0	0.0%
6	0	0.0%
Total	163,215	100.0%

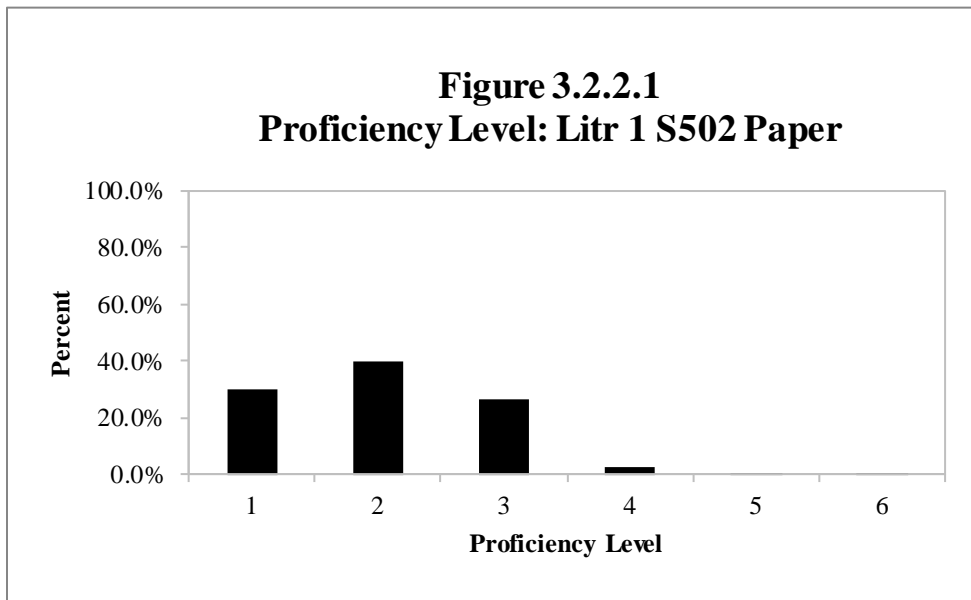


3.2.2.1 Grade 1

Table 3.2.2.1

Proficiency Level Distribution: Litr 1 S502 Paper

Level	Count	Percent
1	7,082	30.1%
2	9,365	39.9%
3	6,307	26.8%
4	643	2.7%
5	83	0.4%
6	11	0.0%
Total	23,491	100.0%

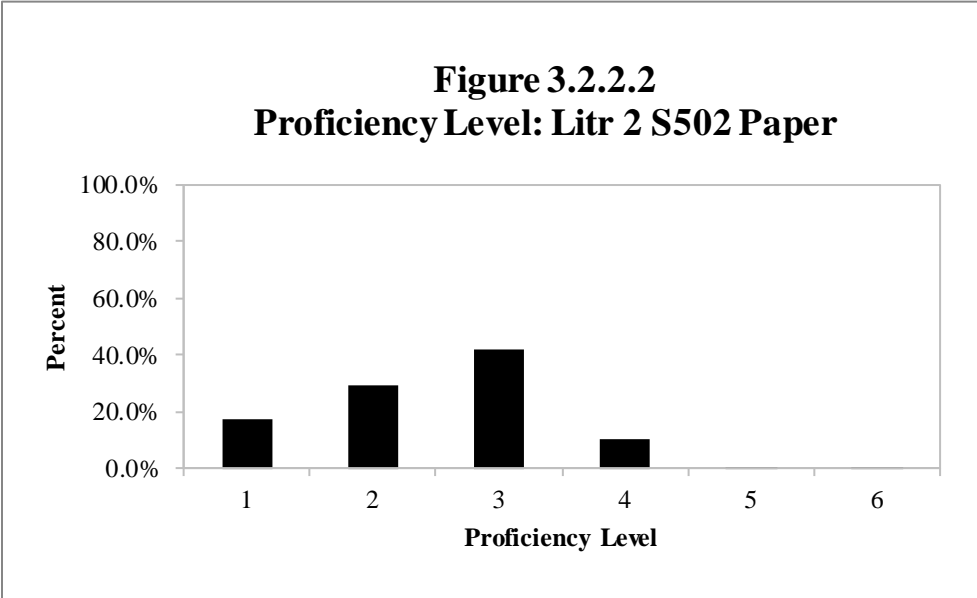


3.2.2.2 Grade 2

Table 3.2.2.2

Proficiency Level Distribution: Litr 2 S502 Paper

Level	Count	Percent
1	4,505	17.4%
2	7,541	29.1%
3	10,919	42.2%
4	2,699	10.4%
5	221	0.9%
6	13	0.1%
Total	25,898	100.0%

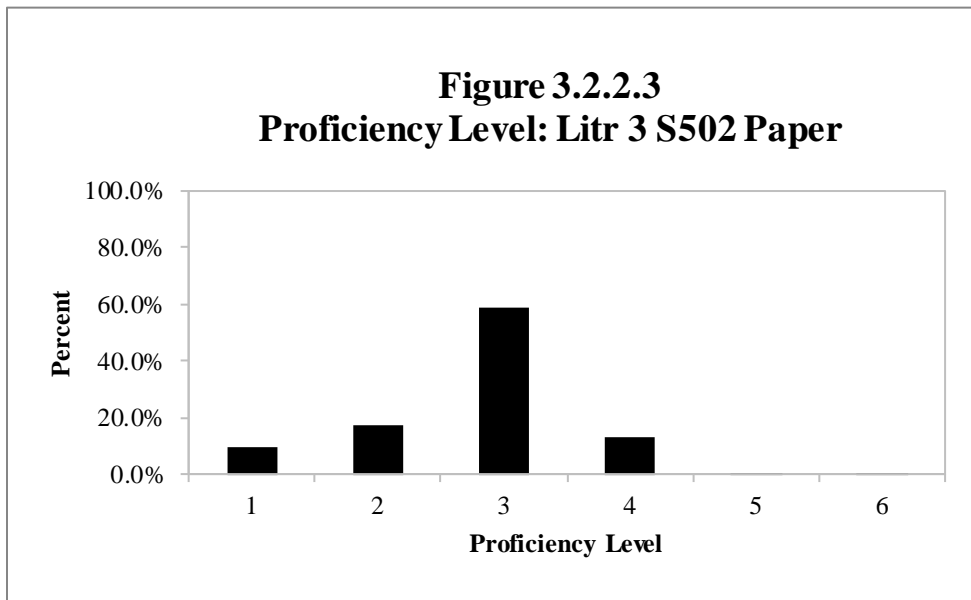


3.2.2.3 Grade 3

Table 3.2.2.3

Proficiency Level Distribution: Litr 3 S502 Paper

Level	Count	Percent
1	2,223	9.8%
2	3,939	17.3%
3	13,444	59.0%
4	3,003	13.2%
5	174	0.8%
6	16	0.1%
Total	22,799	100.0%

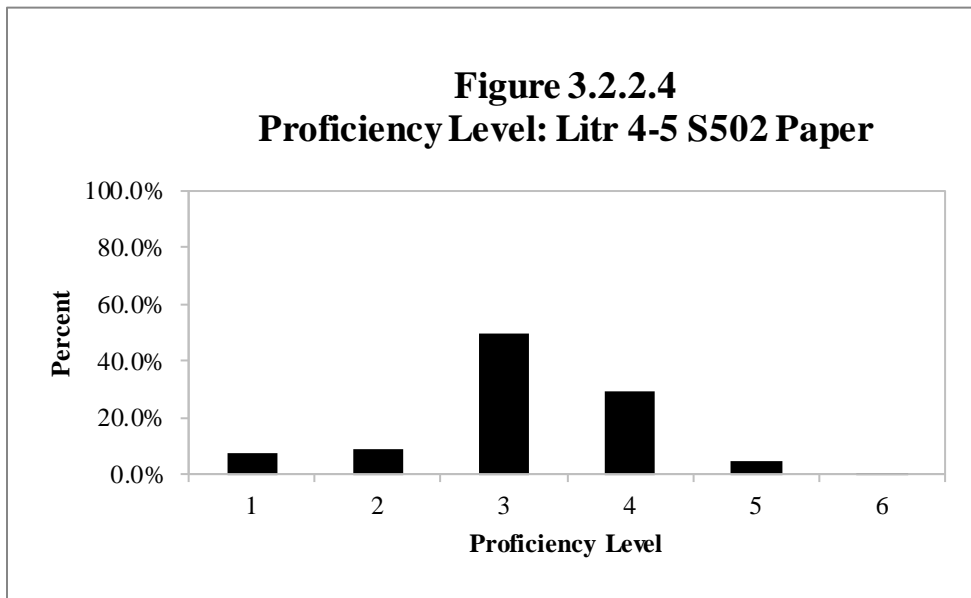


3.2.2.4 Grades 4–5

Table 3.2.2.4

Proficiency Level Distribution: Litr 4-5 S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	1,588	7.2%	1,310	7.5%	2,898	7.3%
2	1,914	8.7%	1,537	8.8%	3,451	8.7%
3	12,032	54.4%	7,676	43.8%	19,708	49.7%
4	5,841	26.4%	5,701	32.5%	11,542	29.1%
5	646	2.9%	1,163	6.6%	1,809	4.6%
6	88	0.4%	148	0.8%	236	0.6%
Total	22,109	100.0%	17,535	100.0%	39,644	100.0%

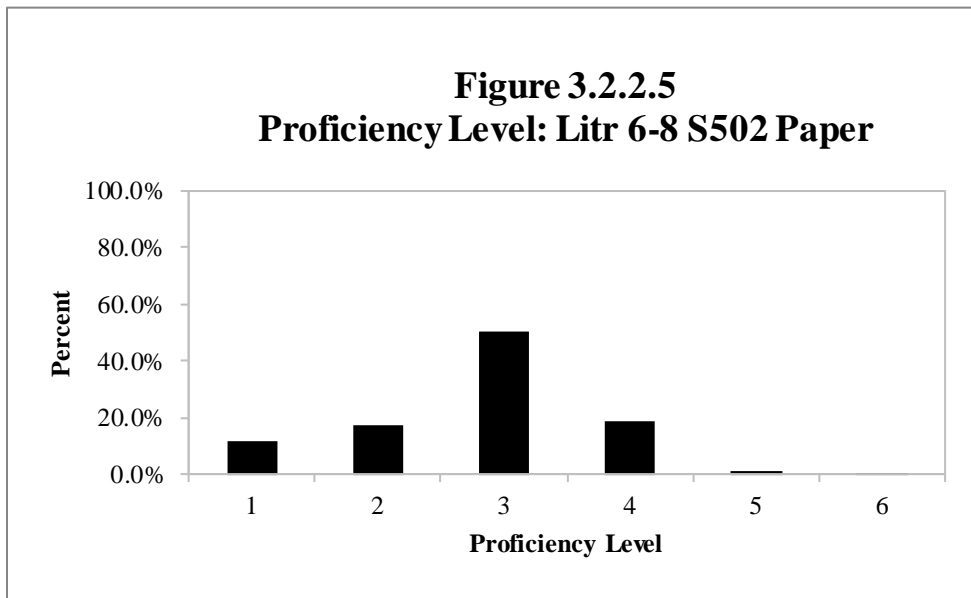


3.2.2.5 Grades 6–8

Table 3.2.2.5

Proficiency Level Distribution: Litr 6-8 S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,260	9.5%	1,350	10.9%	1,690	15.0%	4,300	11.7%
2	2,408	18.2%	2,242	18.2%	1,905	16.9%	6,555	17.8%
3	7,141	53.8%	6,312	51.1%	5,192	46.1%	18,645	50.6%
4	2,304	17.4%	2,262	18.3%	2,261	20.1%	6,827	18.5%
5	149	1.1%	170	1.4%	201	1.8%	520	1.4%
6	3	0.0%	5	0.0%	2	0.0%	10	0.0%
Total	13,265	100.0%	12,341	100.0%	11,251	100.0%	36,857	100.0%

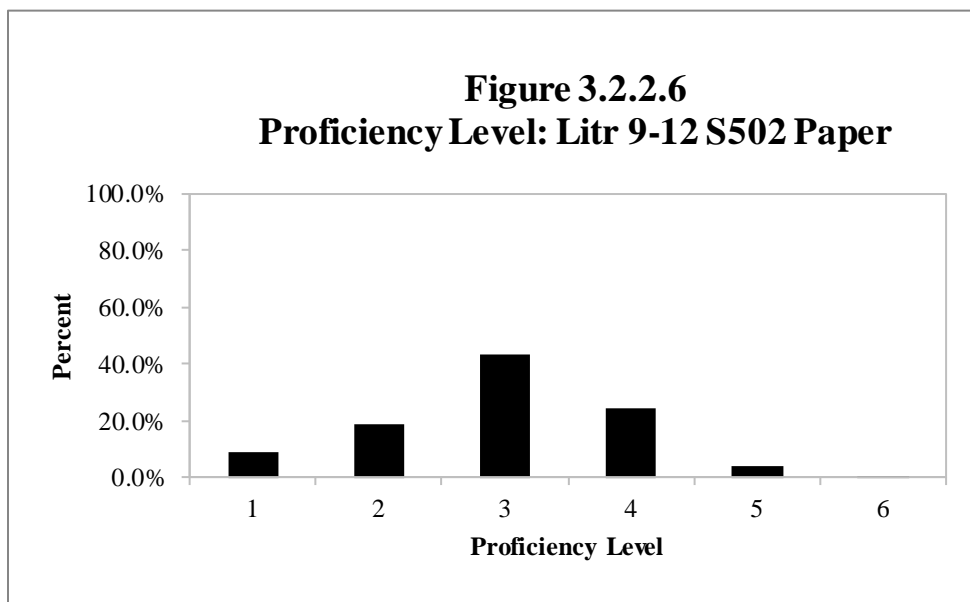


3.2.2.6 Grades 9–12

Table 3.2.2.6

Proficiency Level Distribution: Litr 9-12 S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	741	7.5%	843	9.1%	799	9.9%	625	11.2%	3,008	9.2%
2	1,636	16.6%	1,679	18.2%	1,554	19.2%	1,308	23.4%	6,177	18.8%
3	4,208	42.8%	3,941	42.6%	3,403	42.0%	2,624	46.9%	14,176	43.2%
4	2,794	28.4%	2,343	25.3%	2,008	24.8%	947	16.9%	8,092	24.7%
5	441	4.5%	437	4.7%	342	4.2%	96	1.7%	1,316	4.0%
6	17	0.2%	7	0.1%	3	0.0%	0	0.0%	27	0.1%
Total	9,837	100.0%	9,250	100.0%	8,109	100.0%	5,600	100.0%	32,796	100.0%



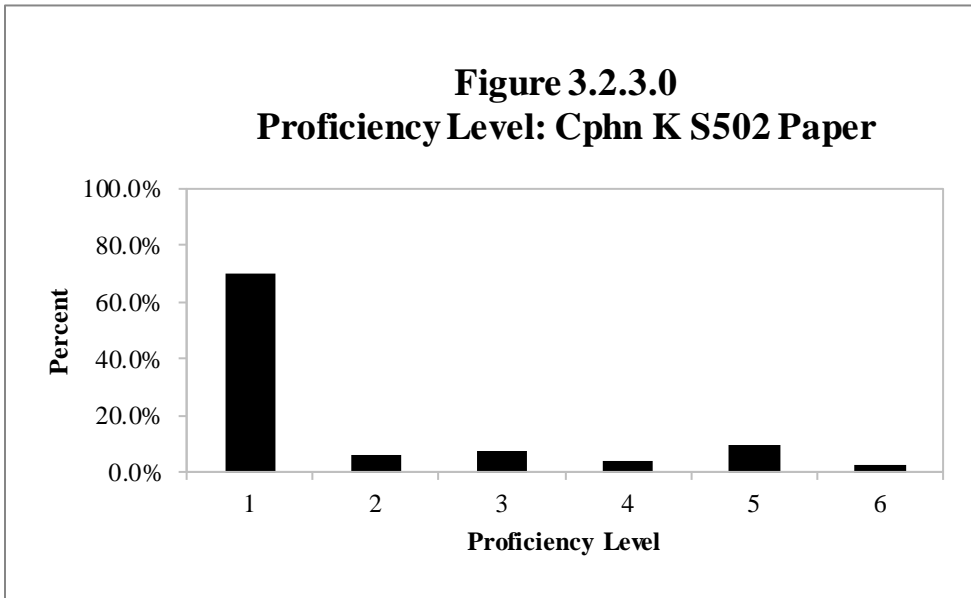
3.2.3 Comprehension

3.2.3.0 Kindergarten

Table 3.2.3.0

Proficiency Level Distribution: Cphn K S502 Paper

Level	Count	Percent
1	114,015	69.9%
2	10,623	6.5%
3	12,456	7.6%
4	6,500	4.0%
5	15,608	9.6%
6	4,013	2.5%
Total	163,215	100.0%

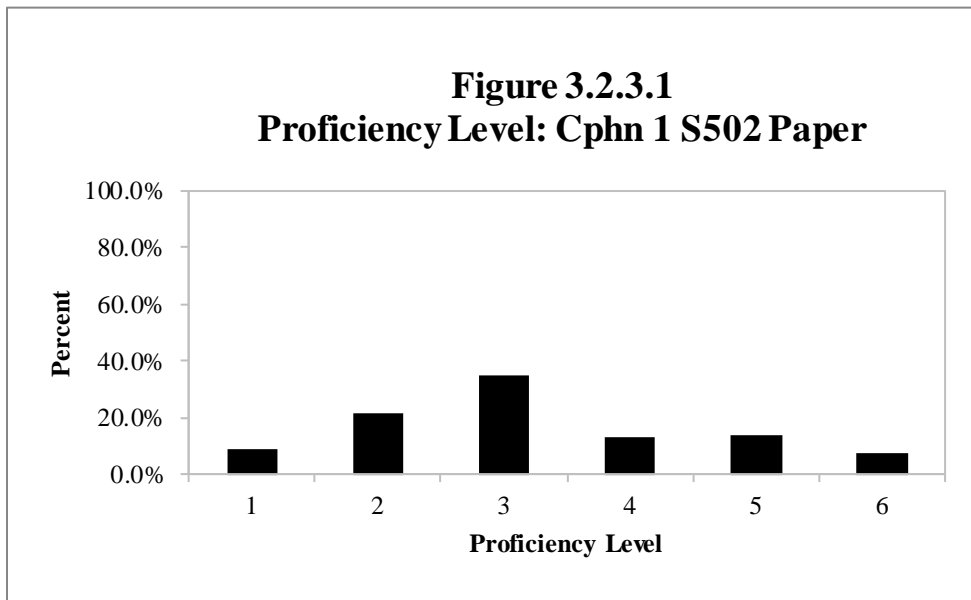


3.2.3.1 Grade 1

Table 3.2.3.1

Proficiency Level Distribution: Cphn 1 S502 Paper

Level	Count	Percent
1	1,805	8.7%
2	4,480	21.6%
3	7,278	35.2%
4	2,779	13.4%
5	2,835	13.7%
6	1,526	7.4%
Total	20,703	100.0%

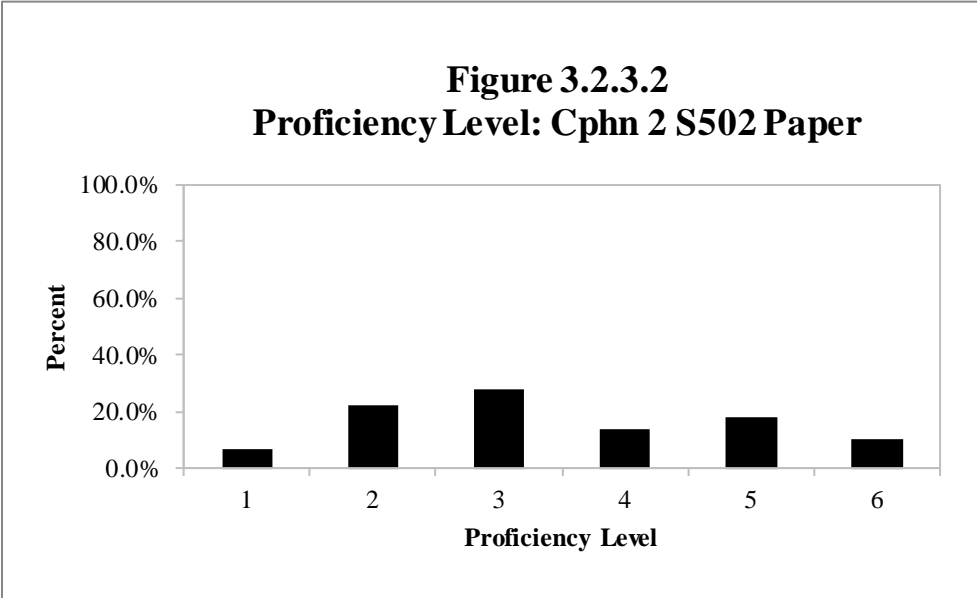


3.2.3.2 Grade 2

Table 3.2.3.2

Proficiency Level Distribution: Cphn 2 S502 Paper

Level	Count	Percent
1	1,772	7.2%
2	5,492	22.4%
3	6,921	28.3%
4	3,435	14.0%
5	4,382	17.9%
6	2,476	10.1%
Total	24,478	100.0%

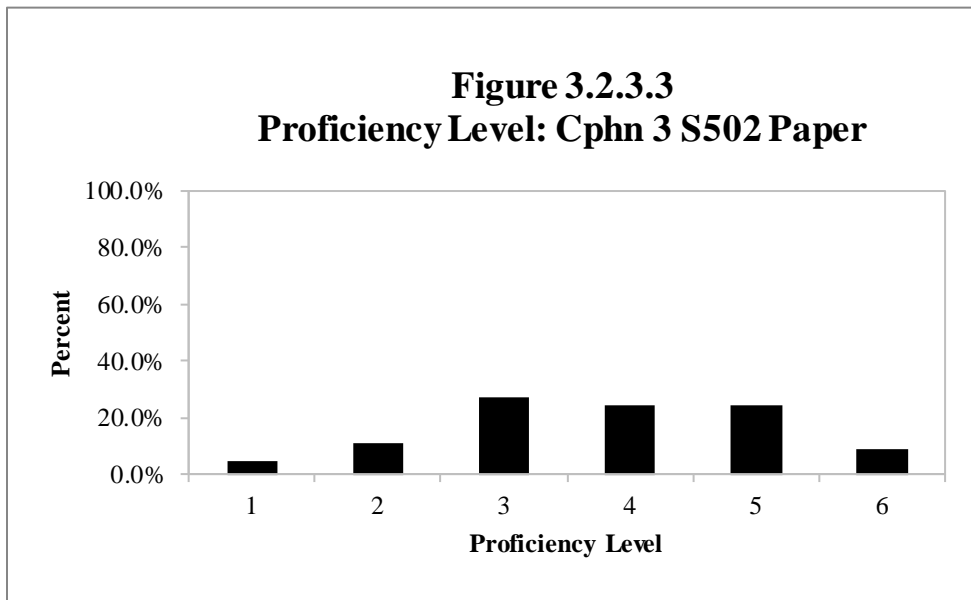


3.2.3.3 Grade 3

Table 3.2.3.3

Proficiency Level Distribution: Cphn 3 S502 Paper

Level	Count	Percent
1	982	4.5%
2	2,360	10.8%
3	6,029	27.6%
4	5,287	24.2%
5	5,282	24.2%
6	1,929	8.8%
Total	21,869	100.0%

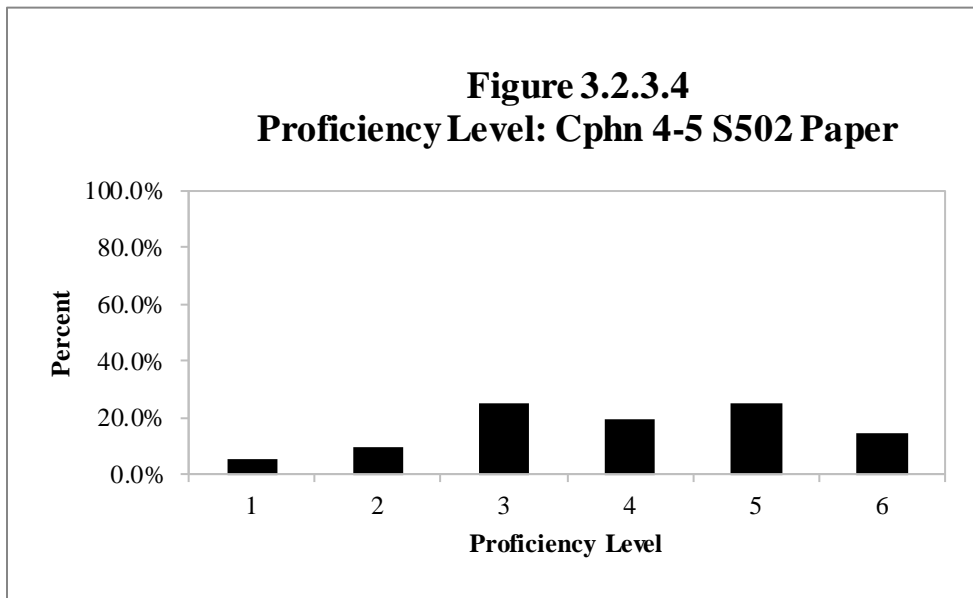


3.2.3.4 Grades 4–5

Table 3.2.3.4

Proficiency Level Distribution: Cphn 4-5 S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	993	4.6%	1,135	6.6%	2,128	5.5%
2	1,923	9.0%	1,895	11.1%	3,818	9.9%
3	6,000	28.0%	3,754	21.9%	9,754	25.3%
4	4,263	19.9%	3,222	18.8%	7,485	19.4%
5	5,452	25.4%	4,241	24.8%	9,693	25.1%
6	2,826	13.2%	2,884	16.8%	5,710	14.8%
Total	21,457	100.0%	17,131	100.0%	38,588	100.0%

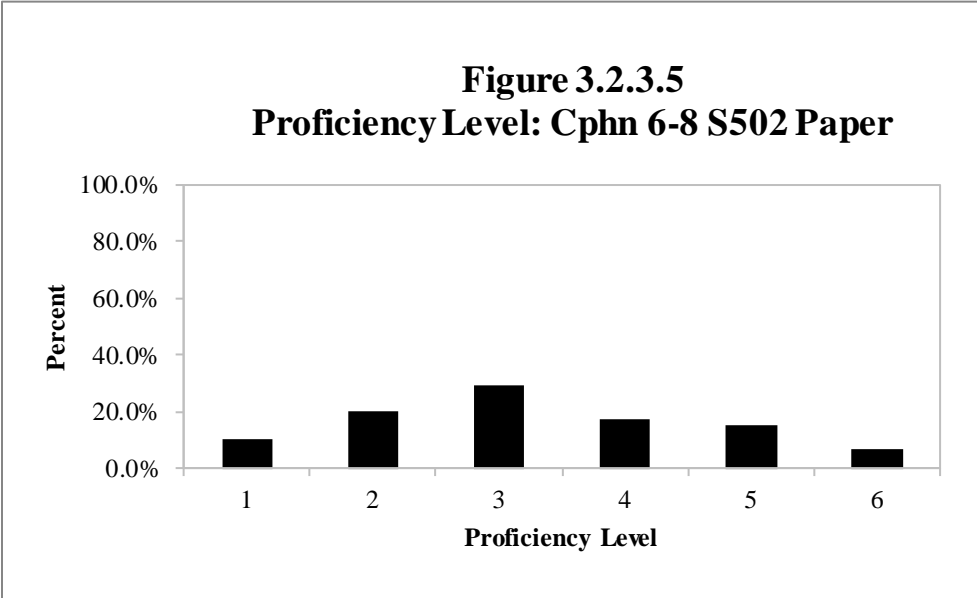


3.2.3.5 Grades 6–8

Table 3.2.3.5

Proficiency Level Distribution: Cphn 6-8 S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	945	7.4%	1,238	10.4%	1,467	13.5%	3,650	10.2%
2	2,648	20.6%	2,468	20.6%	2,213	20.3%	7,329	20.5%
3	4,224	32.9%	3,504	29.3%	2,718	25.0%	10,446	29.3%
4	2,383	18.6%	2,055	17.2%	1,706	15.7%	6,144	17.2%
5	1,842	14.4%	1,785	14.9%	1,944	17.9%	5,571	15.6%
6	787	6.1%	904	7.6%	842	7.7%	2,533	7.1%
Total	12,829	100.0%	11,954	100.0%	10,890	100.0%	35,673	100.0%

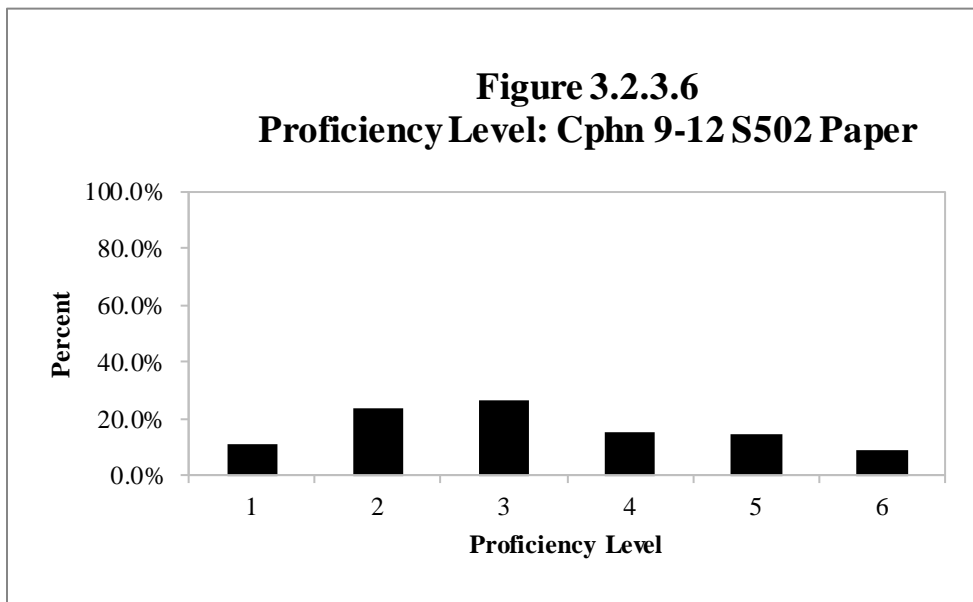


3.2.3.6 Grades 9–12

Table 3.2.3.6

Proficiency Level Distribution: Cphn 9-12 S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	877	9.3%	1,063	11.9%	933	11.9%	687	12.7%	3,560	11.2%
2	1,957	20.7%	1,965	22.0%	1,894	24.1%	1,642	30.3%	7,458	23.6%
3	2,429	25.7%	2,469	27.6%	2,010	25.6%	1,419	26.2%	8,327	26.3%
4	1,720	18.2%	1,368	15.3%	1,068	13.6%	782	14.4%	4,938	15.6%
5	1,635	17.3%	1,203	13.4%	1,161	14.8%	619	11.4%	4,618	14.6%
6	832	8.8%	880	9.8%	780	9.9%	272	5.0%	2,764	8.7%
Total	9,450	100.0%	8,948	100.0%	7,846	100.0%	5,421	100.0%	31,665	100.0%



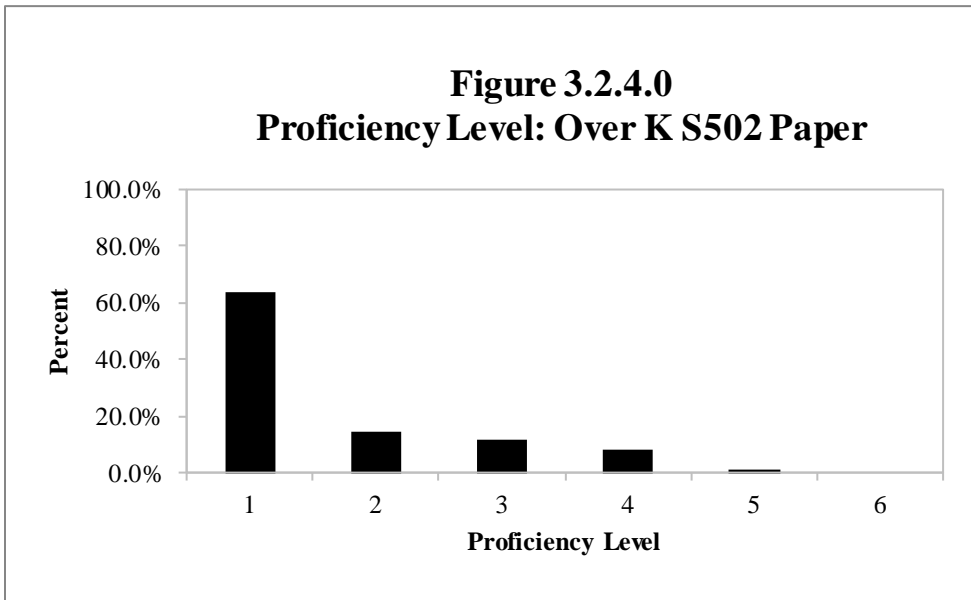
3.2.4 Overall

3.2.4.0. Kindergarten

Table 3.2.4.0

Proficiency Level Distribution: Over K S502 Paper

Level	Count	Percent
1	103,863	63.7%
2	23,916	14.7%
3	19,808	12.1%
4	13,320	8.2%
5	2,271	1.4%
6	0	0.0%
Total	163,178	100.0%

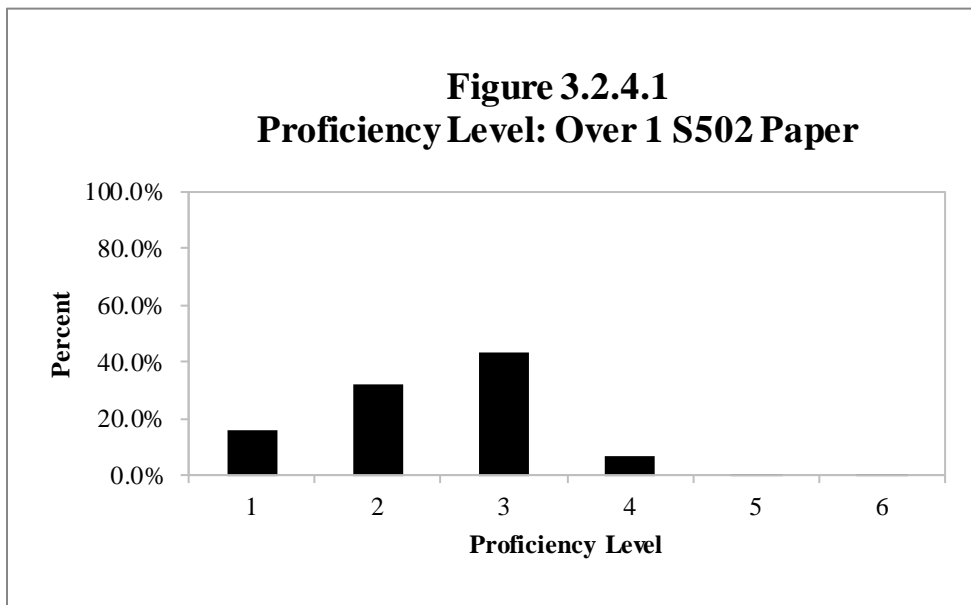


3.2.4.1 Grade 1

Table 3.2.4.1

Proficiency Level Distribution: Over 1 S502 Paper

Level	Count	Percent
1	3,345	16.2%
2	6,643	32.3%
3	8,973	43.6%
4	1,436	7.0%
5	180	0.9%
6	18	0.1%
Total	20,595	100.0%

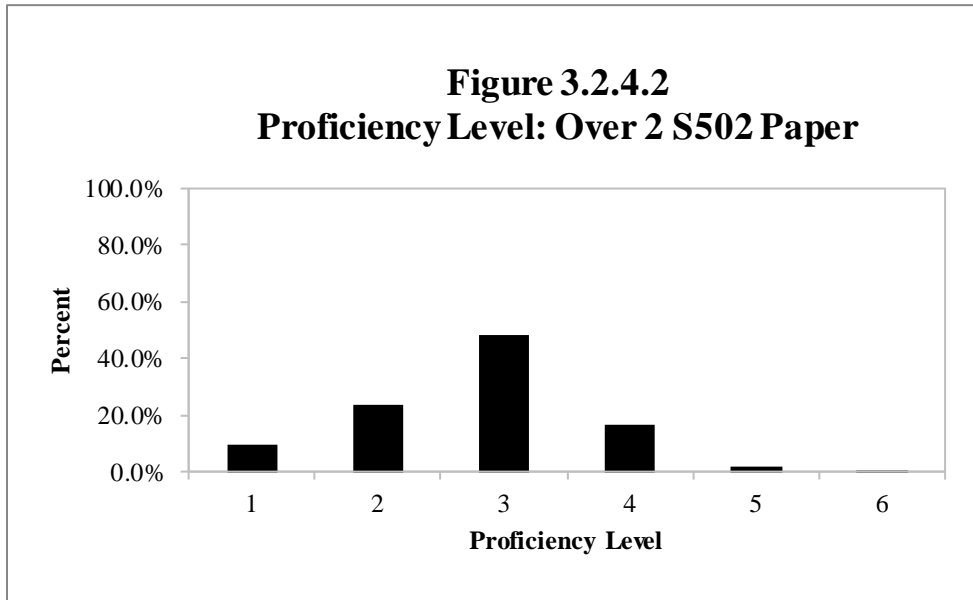


3.2.4.2 Grade 2

Table 3.2.4.2

Proficiency Level Distribution: Over 2 S502 Paper

Level	Count	Percent
1	2,382	9.8%
2	5,723	23.5%
3	11,730	48.2%
4	4,064	16.7%
5	435	1.8%
6	9	0.0%
Total	24,343	100.0%

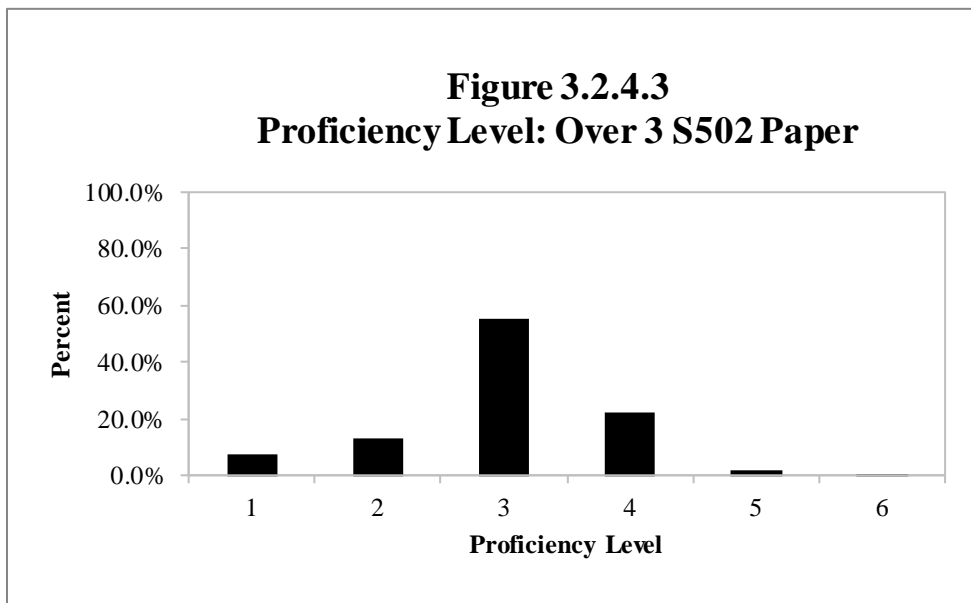


3.2.4.3 Grade 3

Table 3.2.4.3

Proficiency Level Distribution: Over 3 S502 Paper

Level	Count	Percent
1	1,582	7.3%
2	2,909	13.4%
3	12,050	55.4%
4	4,816	22.2%
5	365	1.7%
6	20	0.1%
Total	21,742	100.0%

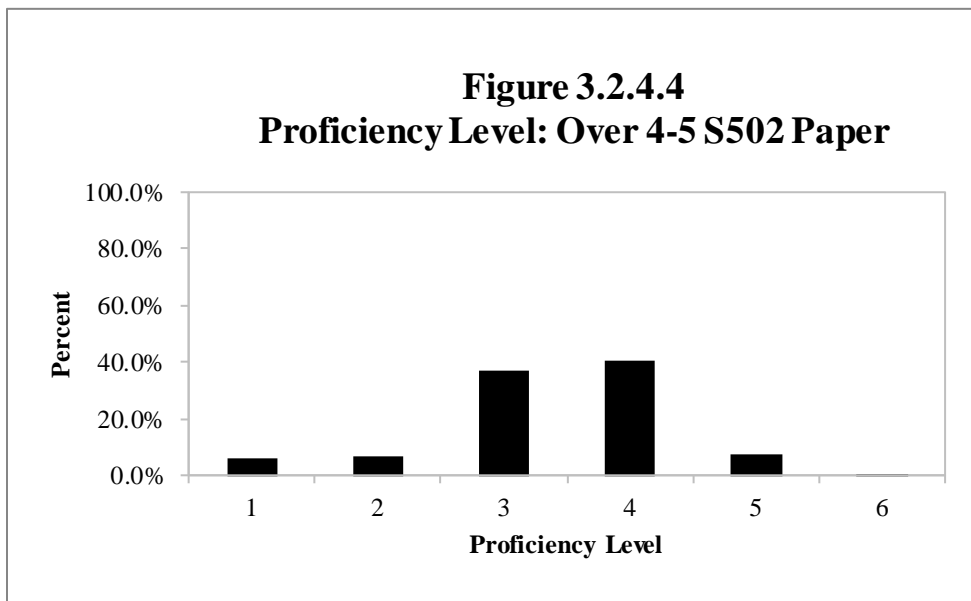


3.2.4.4 Grades 4–6

Table 3.2.4.4

Proficiency Level Distribution: Over 4-5 S502 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	1,269	5.9%	1,187	7.0%	2,456	6.4%
2	1,445	6.8%	1,210	7.1%	2,655	6.9%
3	8,812	41.3%	5,502	32.2%	14,314	37.3%
4	8,386	39.3%	7,338	43.0%	15,724	40.9%
5	1,303	6.1%	1,691	9.9%	2,994	7.8%
6	123	0.6%	150	0.9%	273	0.7%
Total	21,338	100.0%	17,078	100.0%	38,416	100.0%

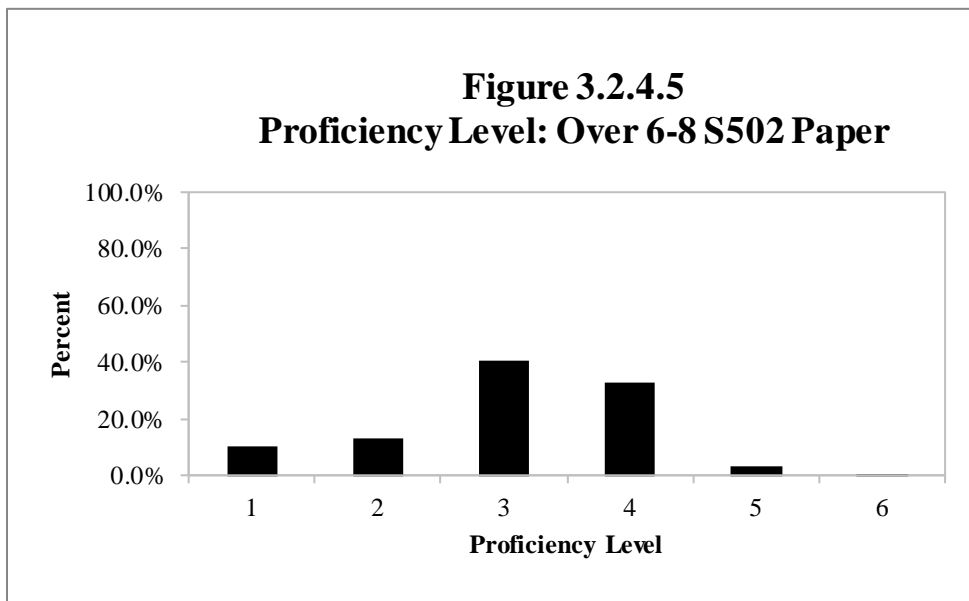


3.2.4.5 Grades 6–8

Table 3.2.4.5

Proficiency Level Distribution: Over 6-8 S502 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	995	7.8%	1,210	10.2%	1,444	13.3%	3,649	10.3%
2	1,676	13.2%	1,521	12.8%	1,389	12.8%	4,586	12.9%
3	5,589	43.9%	4,856	40.9%	4,012	37.1%	14,457	40.8%
4	4,134	32.5%	3,892	32.8%	3,536	32.7%	11,562	32.6%
5	322	2.5%	376	3.2%	426	3.9%	1,124	3.2%
6	15	0.1%	13	0.1%	12	0.1%	40	0.1%
Total	12,731	100.0%	11,868	100.0%	10,819	100.0%	35,418	100.0%

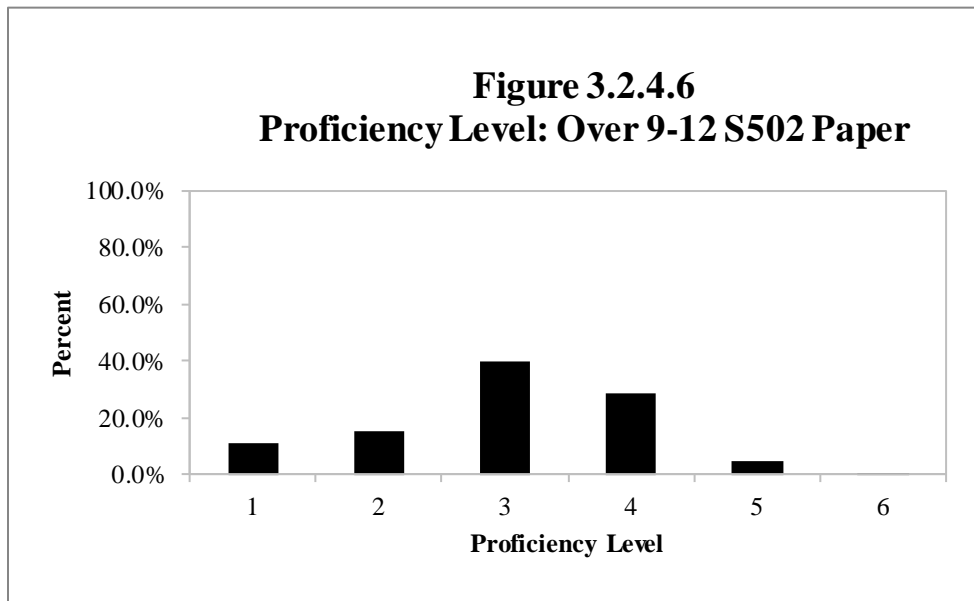


3.2.4.6 Grades 9–12

Table 3.2.4.6

Proficiency Level Distribution: Over 9-12 S502 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	921	9.8%	1,103	12.4%	913	11.8%	644	12.0%	3,581	11.4%
2	1,302	13.9%	1,334	15.1%	1,232	15.9%	1,045	19.5%	4,913	15.7%
3	3,463	37.0%	3,464	39.1%	3,117	40.2%	2,432	45.4%	12,476	39.8%
4	3,124	33.4%	2,491	28.1%	2,137	27.6%	1,144	21.4%	8,896	28.4%
5	525	5.6%	457	5.2%	344	4.4%	87	1.6%	1,413	4.5%
6	20	0.2%	11	0.1%	2	0.0%	0	0.0%	33	0.1%
Total	9,355	100.0%	8,860	100.0%	7,745	100.0%	5,352	100.0%	31,312	100.0%



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 the document *A State's Guidance to the U.S. Department of Education's Assessment Peer Review Process* (U.S. 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, and is linked to the Assessment User Argument (AUA) to support the claims of validity of the Online ACCESS assessment. WIDA structures its validity arguments using the AUA model in lieu of the model highlighted in the *Standards for Educational and Psychological Testing*. AUA has similar topics; however, they are organized differently. Below is a short summary of each AUA claim. For the full AUA validity claims, please refer to the WIDA AUA 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 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 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

Important validity evidence can be obtained from an analysis of the relationship between the content of a test and the construct it is intended to measure. Test content refers to the themes, wording, and format of the items, tasks, or questions on a test. Administration and scoring may also be relevant to content-based evidence. Evidence based on test content can include logical or empirical analyses of the adequacy with which the test content represents the content domain and of the relevance of the content domain to the proposed interpretation of test scores. Evidence based on test content can also come from expert judgment of the relationship between parts of the test and content.

4.1.2 Response Processes

Theoretical and empirical analyses of the response processes of test-takers can provide evidence concerning the fit between the construct and the detailed nature of the performance or response actually engaged in by test-takers. Evidence based on response processes generally comes from analysis of individual responses. Evidence of response processes can contribute to answering questions about differences in meaning or interpretation of test scores across relevant subgroups of test-takers. Studies of response processes are not limited to the test-taker. Assessment often relies on observers or judges to record and/or evaluate test-takers' performance or products.

4.1.3 Internal Structure

Analyses of the internal structure of a test can indicate the degree to which the relationships among the test items and test components conform to the construct on which the proposed test score interpretations are based. The conceptual framework for a test may imply a single dimension of behavior, or it may posit several components that are each expected to be homogeneous.

4.1.4 Relation to Other Variables

In many cases, the intended interpretation for a given use implies that the construct should be related to some other variables, and as a result, analysis of the relationship of the scores to variables external to the test provides another important source of validity evidence. Evidence about the relation to other variables is also used to investigate questions of differential prediction for subgroups. In the test-criterion relationship, the fundamental question is the accuracy with which test scores predict criterion performance. Historically, two designs, often called predictive and concurrent, have been differentiated for evaluating test-criterion relationships. A predictive study indicates the strength of the relationship between test scores and criterion scores that are obtained at a later time. A concurrent study obtains test scores and criterion information at about 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. Because students' scores reflect these sources of variation, 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.

We organized the reliability information presented in this section 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 to follow 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 raw 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. We report the **interrater agreement** rates that the raters achieved when evaluating students' responses to the Writing and Speaking tasks in Section 5.2. 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 have provided the results in Section 5.5. In addition, we examine the CSEM of these scores in Section 5.6. 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 good 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.

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). We describe ACCESS scoring procedures and rater training and quality control monitoring processes elsewhere in this report (see Part 1, Section 4). We report the **interrater agreement rates** for scoring students' responses to the

Writing and Speaking tasks in Section 5.2. These values reflect how consistent the students' scores would be if different groups of raters scored their responses, while we present a detailed description of the methods, data sources, and procedures in this section.

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 measure 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 **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 of the internal consistency of test items. Conceptually, we can think of this as the correlation obtained between performances on two halves of the test if every possible way of dividing the test tasks in two were attempted. Thus, Cronbach's alpha may be low if some items are measuring something other than what most of the items are measuring. In this way, Cronbach's alpha expresses how well the items and tasks on a test appear to measure the same ability.

The formula for Cronbach's alpha is

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_t^2} \right]$$

where

n = number of items i

σ_i^2 = variance of score on item i

σ_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 0.95. 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 0.93, and reliability values across subgroups ranged from 0.92 to 0.94. For the Kindergarten Speaking test, the reliability for all students was 0.91, and reliability values across subgroups ranged from 0.89 to 0.91.

Listening Tier A: The Listening Tier A Cronbach's coefficient alphas computed for all students ranged from 0.62 to 0.75. The Listening Tier A Cronbach's alpha ranged from 0.64 to 0.77 for male students; 0.60 to 0.74 for female students; 0.62 to 0.75 for Hispanic students; 0.64 to 0.76 for non-Hispanic students; and 0.61 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 0.61 to 0.67. The Listening Tier B/C Cronbach's coefficient alphas ranged from 0.62 to 0.67 for male students; 0.61 to 0.66 for female students; 0.61 to 0.66 for Hispanic students; 0.62 to 0.69 for non-Hispanic students; and 0.58 to 0.70 for students with an IEP.

Reading Tier A: The Reading Tier A Cronbach's coefficient alphas computed for all students ranged from 0.76 to 0.81. The Reading Tier A Cronbach's coefficient alphas ranged from 0.76 to 0.80 for male students; 0.76 to 0.81 for female students; 0.74 to 0.80 for Hispanic students; 0.78 to 0.84 for non-Hispanic students; and 0.67 to 0.71 for students with an IEP.

Reading Tier B/C: The Reading Tier B/C Cronbach's coefficient alphas computed for all students ranged from 0.77 to 0.82. The Reading Tier B/C Cronbach's coefficient alphas ranged from 0.78 to 0.82 for male students; 0.77 to 0.82 for female students; 0.77 to 0.81 for Hispanic students; 0.79 to 0.84 for non-Hispanic students; and 0.71 to 0.78 for students with an IEP.

Writing Tier A: The Writing Tier A Cronbach's coefficient alphas computed for all students ranged from 0.85 to 0.93. The Writing Tier A Cronbach's coefficient alphas ranged from 0.86 to 0.92 for male students; 0.83 to 0.92 for female students; 0.86 to 0.92 for Hispanic students; 0.83 to 0.93 for non-Hispanic students; and 0.84 to 0.92 for students with an IEP.

Writing Tier B/C: The Writing Tier B/C Cronbach's coefficient alphas computed for all students ranged from 0.91 to 0.96. 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; 0.91 to 0.95 for Hispanic students; 0.91 to 0.96 for non-Hispanic students; and 0.92 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 0.87 to 0.90. Cronbach's coefficient alphas ranged from 0.87 to 0.89 for male students; 0.86 to 0.91 for female students; 0.87 to 0.90 for Hispanic students; 0.85 to 0.88 for non-Hispanic students; and 0.83 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 0.91 to 0.93. Cronbach's coefficient alphas ranged from 0.91 to 0.93 for male students; 0.91 to 0.93 for female students; 0.91 to 0.93 for Hispanic students; 0.91 to 0.92 for non-Hispanic students; and 0.91 to 0.92 for students with an IEP.

5.1.1 Listening

Table 5.1.1.1

Reliabilities of Domain Scores: List S502 Paper

Cluster	Tier	No. of Students	No. of Items	Cronbach's Alpha	SEM
K	-	163,226	30	0.95	1.86
1	A	18,914	18	0.75	1.59
	B/C	35,558	21	0.67	1.82
2	A	18,914	18	0.75	1.59
	B/C	35,558	21	0.67	1.82
3	A	10,832	18	0.73	1.87
	B/C	59,106	21	0.62	1.92
4-5	A	10,832	18	0.73	1.87
	B/C	59,106	21	0.62	1.92
6-8	A	8,442	18	0.71	1.86
	B/C	30,916	21	0.61	1.87
9-12	A	7,927	18	0.62	1.80
	B/C	27,050	21	0.65	1.95

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 S502 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	75,832	0.94	1.84	84,992	0.95	1.89
1	A	18	8,636	0.74	1.58	10,223	0.77	1.60
	B/C	21	17,134	0.66	1.79	18,360	0.67	1.84
2	A	18	8,636	0.74	1.58	10,223	0.77	1.60
	B/C	21	17,134	0.66	1.79	18,360	0.67	1.84
3	A	18	4,858	0.73	1.85	5,923	0.72	1.87
	B/C	21	27,112	0.61	1.92	31,866	0.63	1.92
4-5	A	18	4,858	0.73	1.85	5,923	0.72	1.87
	B/C	21	27,112	0.61	1.92	31,866	0.63	1.92
6-8	A	18	3,785	0.70	1.85	4,626	0.71	1.87
	B/C	21	14,093	0.61	1.86	16,733	0.62	1.87
9-12	A	18	3,650	0.60	1.80	4,244	0.64	1.80
	B/C	21	12,502	0.64	1.95	14,474	0.67	1.95

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 S502 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	109,190	0.95	1.89	47,257	0.94	1.80
1	A	18	15,354	0.75	1.59	3,471	0.76	1.58
	B/C	21	27,994	0.66	1.81	7,445	0.69	1.82
2	A	18	15,354	0.75	1.59	3,471	0.76	1.58
	B/C	21	27,994	0.66	1.81	7,445	0.69	1.82
3	A	18	8,958	0.72	1.87	1,790	0.74	1.82
	B/C	21	46,760	0.62	1.92	12,143	0.64	1.92
4-5	A	18	8,958	0.72	1.87	1,790	0.74	1.82
	B/C	21	46,760	0.62	1.92	12,143	0.64	1.92
6-8	A	18	7,049	0.70	1.87	1,324	0.73	1.81
	B/C	21	24,435	0.61	1.87	6,326	0.62	1.85
9-12	A	18	6,438	0.62	1.79	1,384	0.64	1.81
	B/C	21	20,878	0.66	1.95	6,037	0.65	1.96

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 S502 Paper by IEP Status

Cluster	Tier	No. of Students	No. of Items	Cronbach's Alpha	SEM
K	-	13,493	30	0.95	1.91
1	A	2,422	18	0.77	1.68
	B/C	3,909	21	0.70	1.90
2	A	2,422	18	0.77	1.68
	B/C	3,909	21	0.70	1.90
3	A	1,219	18	0.67	1.88
	B/C	11,193	21	0.61	1.98
4-5	A	1,219	18	0.67	1.88
	B/C	11,193	21	0.61	1.98
6-8	A	461	18	0.68	1.86
	B/C	4,460	21	0.58	1.96
9-12	A	356	18	0.61	1.80
	B/C	1,817	21	0.61	2.00

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 S502 Paper

Cluster	Tier	No. of Students	No. of Items	Cronbach's Alpha	SEM
K	-	163,218	30	0.95	1.74
1	A	18,193	24	0.76	2.20
	B/C	31,612	27	0.82	2.33
2	A	18,193	24	0.76	2.20
	B/C	31,612	27	0.82	2.33
3	A	10,254	24	0.81	2.19
	B/C	52,612	27	0.77	2.37
4-5	A	10,254	24	0.81	2.19
	B/C	52,612	27	0.77	2.37
6-8	A	8,121	24	0.77	2.18
	B/C	28,935	27	0.79	2.35
9-12	A	7,915	24	0.79	2.11
	B/C	24,995	27	0.81	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.

Table 5.1.2.2

Reliabilities of Domain Scores: Read S502 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	75,830	0.95	1.75	84,986	0.96	1.74
1	A	24	8,293	0.76	2.20	9,851	0.76	2.20
	B/C	27	15,185	0.82	2.32	16,376	0.82	2.33
2	A	24	8,293	0.76	2.20	9,851	0.76	2.20
	B/C	27	15,185	0.82	2.32	16,376	0.82	2.33
3	A	24	4,618	0.81	2.17	5,583	0.80	2.20
	B/C	27	24,280	0.77	2.37	28,217	0.78	2.37
4-5	A	24	4,618	0.81	2.17	5,583	0.80	2.20
	B/C	27	24,280	0.77	2.37	28,217	0.78	2.37
6-8	A	24	3,649	0.78	2.17	4,443	0.76	2.18
	B/C	27	13,341	0.79	2.33	15,513	0.80	2.35
9-12	A	24	3,658	0.78	2.09	4,223	0.78	2.12
	B/C	27	11,651	0.80	2.34	13,272	0.82	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.

Table 5.1.2.3

Reliabilities of Domain Scores: Read S502 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	109,181	0.95	1.74	47,258	0.96	1.71
1	A	24	14,741	0.74	2.21	3,374	0.79	2.18
	B/C	27	24,812	0.81	2.33	6,706	0.84	2.30
2	A	24	14,741	0.74	2.21	3,374	0.79	2.18
	B/C	27	24,812	0.81	2.33	6,706	0.84	2.30
3	A	24	8,478	0.80	2.19	1,684	0.84	2.14
	B/C	27	41,648	0.77	2.37	10,793	0.79	2.37
4-5	A	24	8,478	0.80	2.19	1,684	0.84	2.14
	B/C	27	41,648	0.77	2.37	10,793	0.79	2.37
6-8	A	24	6,793	0.76	2.18	1,264	0.80	2.15
	B/C	27	22,888	0.79	2.35	5,908	0.80	2.33
9-12	A	24	6,453	0.79	2.11	1,355	0.78	2.10
	B/C	27	19,400	0.81	2.34	5,468	0.81	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.

Table 5.1.2.4

Reliabilities of Domain Scores: Read S502 Paper by IEP Status

Cluster	Tier	No. of Students	No. of Items	Cronbach's Alpha	SEM
K	-	13,493	30	0.96	1.73
1	A	2,338	24	0.67	2.24
	B/C	3,540	27	0.78	2.36
2	A	2,338	24	0.67	2.24
	B/C	3,540	27	0.78	2.36
3	A	1,110	24	0.71	2.24
	B/C	10,114	27	0.72	2.37
4-5	A	1,110	24	0.71	2.24
	B/C	10,114	27	0.72	2.37
6-8	A	443	24	0.67	2.22
	B/C	4,156	27	0.71	2.38
9-12	A	358	24	0.71	2.17
	B/C	1,686	27	0.75	2.41

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 S502 Paper

Cluster	Tier	No. of Students	No. of Tasks	Total Possible Raw Score Points	Cronbach's Alpha*	SEM
K	-	163,216	6	0-17	0.93	1.16
1	A	15,252	4	0-40	0.89	2.00
	B/C	15,300	3	0-54	0.96	1.96
2	A	12,205	3	0-27	0.93	1.32
	B/C	46,536	3	0-54	0.94	1.87
3	A	12,205	3	0-27	0.93	1.32
	B/C	46,536	3	0-54	0.94	1.87
4-5	A	6,693	3	0-27	0.90	1.37
	B/C	38,827	3	0-54	0.92	1.99
6-8	A	8,992	3	0-27	0.87	1.46
	B/C	32,119	3	0-54	0.91	1.96
9-12	A	8,572	3	0-27	0.85	1.76
	B/C	28,102	3	0-54	0.92	1.95

*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 S502 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	75,829	0.93	1.17	84,986	0.93	1.16
1	A	4	0-40	6,961	0.88	1.98	8,256	0.89	2.00
	B/C	3	0-54	7,345	0.95	1.96	7,937	0.96	1.96
2	A	3	0-27	5,247	0.92	1.33	6,906	0.92	1.32
	B/C	3	0-54	21,675	0.94	1.85	24,757	0.95	1.89
3	A	3	0-27	5,247	0.92	1.33	6,906	0.92	1.32
	B/C	3	0-54	21,675	0.94	1.85	24,757	0.95	1.89
4-5	A	3	0-27	3,046	0.90	1.35	3,612	0.89	1.39
	B/C	3	0-54	17,705	0.91	1.95	21,044	0.92	2.02
6-8	A	3	0-27	4,017	0.86	1.46	4,944	0.87	1.46
	B/C	3	0-54	14,608	0.91	1.89	17,417	0.92	2.00
9-12	A	3	0-27	3,935	0.83	1.80	4,598	0.86	1.73
	B/C	3	0-54	12,916	0.91	1.88	15,109	0.92	2.00

*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 S502 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	109,180	0.92	1.15	47,258	0.94	1.19
1	A	4	0-40	12,370	0.89	2.00	2,816	0.88	2.01
	B/C	3	0-54	11,878	0.95	1.95	3,377	0.96	1.99
2	A	3	0-27	9,828	0.92	1.32	2,295	0.93	1.34
	B/C	3	0-54	36,651	0.94	1.87	9,718	0.95	1.87
3	A	3	0-27	9,828	0.92	1.32	2,295	0.93	1.34
	B/C	3	0-54	36,651	0.94	1.87	9,718	0.95	1.87
4-5	A	3	0-27	5,567	0.90	1.38	1,062	0.90	1.33
	B/C	3	0-54	30,682	0.91	1.99	8,011	0.92	2.00
6-8	A	3	0-27	7,498	0.87	1.45	1,419	0.86	1.47
	B/C	3	0-54	25,276	0.91	1.93	6,680	0.91	2.05
9-12	A	3	0-27	6,955	0.86	1.72	1,495	0.83	1.92
	B/C	3	0-54	21,646	0.92	1.92	6,313	0.92	2.05

*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 S502 Paper by IEP Status

Cluster	Tier	No. of Students	No. of Tasks	Total Possible Raw Score Points	Cronbach's Alpha*	SEM
K	-	13,493	6	0-17	0.92	1.11
1	A	1,864	4	0-40	0.89	1.94
	B/C	1,520	3	0-54	0.96	1.90
2	A	1,759	3	0-27	0.92	1.29
	B/C	6,543	3	0-54	0.95	1.92
3	A	1,759	3	0-27	0.92	1.29
	B/C	6,543	3	0-54	0.95	1.92
4-5	A	621	3	0-27	0.90	1.31
	B/C	8,017	3	0-54	0.92	2.09
6-8	A	489	3	0-27	0.87	1.46
	B/C	4,640	3	0-54	0.92	2.10
9-12	A	384	3	0-27	0.84	1.79
	B/C	1,903	3	0-54	0.92	1.92

*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 S502 Paper

Cluster	Tier	No. of Students	No. of Tasks	Total Possible Raw Score Points	Cronbach's Alpha	SEM
K	-	163,192	10	0-10	0.91	1.04
1	A	15,178	6	0-18	0.88	1.35
	B/C	15,197	6	0-24	0.92	1.37
2	A	12,107	6	0-18	0.90	1.33
	B/C	46,290	6	0-24	0.91	1.31
3	A	12,107	6	0-18	0.90	1.33
	B/C	46,290	6	0-24	0.91	1.31
4-5	A	6,649	6	0-18	0.89	1.43
	B/C	38,677	6	0-24	0.91	1.37
6-8	A	8,909	6	0-18	0.87	1.42
	B/C	31,906	6	0-24	0.91	1.47
9-12	A	8,442	6	0-18	0.88	1.47
	B/C	27,840	6	0-24	0.93	1.40

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 S502 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	75,812	0.91	1.02	84,979	0.90	1.05
1	A	6	0-18	6,922	0.88	1.35	8,221	0.87	1.35
	B/C	6	0-24	7,284	0.91	1.38	7,895	0.92	1.36
2	A	6	0-18	5,207	0.91	1.32	6,850	0.89	1.33
	B/C	6	0-24	21,554	0.91	1.30	24,632	0.91	1.31
3	A	6	0-18	5,207	0.91	1.32	6,850	0.89	1.33
	B/C	6	0-24	21,554	0.91	1.30	24,632	0.91	1.31
4-5	A	6	0-18	3,025	0.89	1.45	3,589	0.89	1.41
	B/C	6	0-24	17,630	0.91	1.37	20,970	0.91	1.37
6-8	A	6	0-18	3,981	0.86	1.44	4,898	0.88	1.41
	B/C	6	0-24	14,512	0.91	1.47	17,300	0.91	1.47
9-12	A	6	0-18	3,871	0.87	1.48	4,532	0.88	1.46
	B/C	6	0-24	12,787	0.93	1.41	14,976	0.93	1.38

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 S502 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	109,168	0.91	1.04	47,246	0.89	1.02
1	A	6	0-18	12,312	0.88	1.35	2,800	0.86	1.36
	B/C	6	0-24	11,796	0.91	1.36	3,356	0.92	1.40
2	A	6	0-18	9,754	0.90	1.32	2,273	0.88	1.35
	B/C	6	0-24	36,459	0.91	1.30	9,667	0.91	1.32
3	A	6	0-18	9,754	0.90	1.32	2,273	0.88	1.35
	B/C	6	0-24	36,459	0.91	1.30	9,667	0.91	1.32
4-5	A	6	0-18	5,532	0.89	1.42	1,054	0.85	1.46
	B/C	6	0-24	30,578	0.91	1.37	7,966	0.91	1.40
6-8	A	6	0-18	7,433	0.87	1.42	1,403	0.86	1.43
	B/C	6	0-24	25,124	0.91	1.47	6,621	0.91	1.48
9-12	A	6	0-18	6,849	0.88	1.46	1,474	0.86	1.48
	B/C	6	0-24	21,452	0.93	1.40	6,245	0.92	1.38

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 S502 Paper by IEP Status

Cluster	Tier	No. of Students	No. of Tasks	Total Possible Raw Score Points	Cronbach's Alpha	SEM
K	-	13,490	10	0-10	0.90	1.05
1	A	1,850	6	0-18	0.87	1.30
	B/C	1,509	6	0-24	0.92	1.36
2	A	1,748	6	0-18	0.87	1.30
	B/C	6,514	6	0-24	0.91	1.31
3	A	1,748	6	0-18	0.87	1.30
	B/C	6,514	6	0-24	0.91	1.31
4-5	A	619	6	0-18	0.86	1.36
	B/C	7,995	6	0-24	0.91	1.40
6-8	A	486	6	0-18	0.83	1.38
	B/C	4,600	6	0-24	0.91	1.49
9-12	A	378	6	0-18	0.87	1.45
	B/C	1,881	6	0-24	0.92	1.44

Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

5.2 Interrater Agreement Rates

For the Writing tests (except Kindergarten, which is scored by the Test Administrator), 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 96%.

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 Listening, 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 S502 Paper

Interrater Agreement	n/a
-----------------------------	-----

5.2.3.1 Grade 1

Table 5.2.3.1.1

Interrater Agreement: Writ 1 A S502 Paper

Interrater Agreement	Task	No. in Sample	% AG	% AD	% NA
	1	7,360	100	0	0
	2	9,882	100	0	0
	3	11,228	100	0	0
	4	9,730	99	1	0

Table 5.2.3.1.2

Interrater Agreement: Writ 1 B/C S502 Paper

Interrater Agreement	Task	No. in Sample	% AG	% AD	% NA
	1	8,868	99	1	0
	2	11,254	99	1	0
	3	9,398	99	1	0

5.2.3.2 Grade 2

Table 5.2.3.2.1

Interrater Agreement: Writ 2 A S502 Paper

Interrater Agreement	Task	No. in Sample	% AG	% AD	% NA
	1	4,410	98	2	0
	2	4,902	99	1	0
	3	5,484	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 S502 Paper

Interrater Agreement	Task	No. in Sample	% AG	% AD	% NA
	1	11,658	97	3	0
	2	11,790	97	3	0
	3	11,994	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 S502 Paper

Interrater Agreement	Task	No. in Sample	% AG	% AD	% NA
	1	2,934	99	1	0
	2	3,004	98	2	0
	3	3,352	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 S502 Paper

Interrater Agreement	Task	No. in Sample	% AG	% AD	% NA
	1	10,190	97	3	0
	2	10,252	97	3	0
	3	10,366	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 S502 Paper

Interrater Agreement	Task	No. in Sample	% AG	% AD	% NA
	1	4,280	98	2	0
	2	3,602	99	1	0
	3	4,236	99	1	0

Table 5.2.3.4.2

Interrater Agreement: Writ 4-5 B/C S502 Paper

Interrater Agreement	Task	No. in Sample	% AG	% AD	% NA
	1	17,362	97	3	0
	2	17,472	98	2	0
	3	18,782	97	3	0

5.2.3.5 Grades 6–8

Table 5.2.3.5.1

Interrater Agreement: Writ 6-8 A S502 Paper

Interrater Agreement	Task	No. in Sample	% AG	% AD	% NA
	1	4,762	98	2	0
	2	5,352	98	2	0
	3	4,286	97	3	0

Table 5.2.3.5.2

Interrater Agreement: Writ 6-8 B/C S502 Paper

Interrater Agreement	Task	No. in Sample	% AG	% AD	% NA
	1	14,380	99	1	0
	2	14,686	98	2	0
	3	16,116	98	2	0

5.2.3.6 Grades 9–12

Table 5.2.3.6.1

Interrater Agreement: Writ 9-12 A S502 Paper

Interrater Agreement	Task	No. in Sample	% AG	% AD	% NA
	1	5,240	96	4	0
	2	4,328	97	3	0
	3	6,220	97	3	0

Table 5.2.3.6.2

Interrater Agreement: Writ 9-12 B/C S502 Paper

Interrater Agreement	Task	No. in Sample	% AG	% AD	% NA
	1	13,534	97	3	0
	2	13,590	97	3	0
	3	13,958	96	4	0

5.3 Conditional Standard Errors of Measurement at Cut Score

The tables in this section present information on the CSEM of scale scores at the most important points at which decisions are made 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 score at that point on the underlying latent ability scale. We first computed the CSEM on the theta metric, which is the square root of the inverse of the Test Information Function, and then linearly transformed the values to the ACCESS scale score metric using the multiplicative constant of the linear equation for the domain (See Section 2.2). The CSEM value based on IRT can vary across test scores. For example, in the Listening and Reading domain, if a student gets either a very few or a very large number of items correct (i.e., scores at the extremes of the score distribution), the CSEM will be greater in value than it would be if the student gets a moderate number of items correct. Scores near the middle of the score distribution typically have lower CSEM compared to the extremes because many tests are comprised of a large proportion of moderately difficult items which are suited to measuring students of moderate proficiency. The CSEM can be used to construct the error band quantifying uncertainty in a student’s score. An approximate 68% confidence interval for a scale score is given by one CSEM below the scale score and one CSEM above the scale score. 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 WIDA states vary in how decisions about the ACCESS scores are made. As discussed in Section 5, all WIDA states use composite scale scores in making reclassification decision and no WIDA state uses a single domain scale score in making reclassification decision. Because the cut points depend on the grade level, we provide information for each grade level within a grade-level cluster.

Since ACCESS test scores were scaled using the IRT method, CSEM values for the scale scores at the highest cut points are typically high. The IRT method is known to produce higher CSEMs 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, the numbers of students at the highest cut points are typically smaller than at other cut points. Therefore, the measurement errors associated with the scale scores at the highest cut points tend to be higher than those of the scale scores at the lower cut points since there are fewer students available in estimating the scores and the measurement errors 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. For example, Tier A is intended for students at the lowest end of the language proficiency continuum. Therefore, the CSEMs of the Tier A student scale scores are expected to

be lower at the 1/2 and the 2/3 proficiency level cut points as compared to those at the 4/5 cut point. These tables provide comparable information on how well the two-tier forms are targeted to provide the most accurate measure to place their intended examinees into the language proficiency levels that they target.

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 is given for each cut score in the scale score metric.

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 S502 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 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2 □	1	236	19.16	19.91
2/3 □	1	259	19.54	18.79
3/4 □	1	291	22.54	18.79
4/5 □	1	303	24.42	19.16
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 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2 □	2	245	19.16	19.54
2/3 □	2	283	21.42	18.79
3/4 □	2	314	26.30	19.91
4/5 □	2	330	30.43	21.04
5/6 □	2	354	38.32	24.05

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 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2□	3	262	21.42	25.55
2/3□	3	300	18.79	20.66
3/4□	3	331	19.54	19.16
4/5□	3	349	21.04	18.67
5/6□	3	374	25.17	19.16

5.3.1.4 Grades 4–5

Table 5.3.1.4

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 4-5 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2□	4	275	20.06	23.67
	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	19.16
5/6□	4	388	28.55	19.91
	5	401	32.31	21.04

5.3.1.5 Grades 6–8

Table 5.3.1.5

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 6-8 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2 □	6	294	20.29	21.04
	7	302	19.91	20.29
	8	308	19.91	19.54
2/3 □	6	332	19.91	18.03
	7	340	20.40	18.03
	8	347	21.04	17.92
3/4 □	6	363	22.54	18.03
	7	370	23.29	18.41
	8	377	24.42	18.79
4/5 □	6	385	25.92	19.16
	7	394	28.18	19.91
	8	402	30.06	20.66
5/6 □	6	411	33.06	21.79
	7	420	36.07	23.29
	8	427	39.07	24.42

5.3.1.6 Grades 9–12

Table 5.3.1.6

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 9-12 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2 □	9	314	20.66	21.79
	10	325	20.66	20.70
	11	335	21.04	19.91
	12	342	21.42	19.54
2/3 □	9	353	22.17	19.16
	10	358	22.54	18.79
	11	364	23.26	18.79
	12	368	23.67	18.79
3/4 □	9	383	25.92	18.79
	10	389	27.43	18.79
	11	394	28.55	19.16
	12	398	29.31	19.16
4/5 □	9	409	32.31	19.54
	10	415	34.19	20.29
	11	420	36.07	20.66
	12	426	37.95	21.04
5/6 □	9	434	41.33	22.17
	10	441	44.71	22.92
	11	447	47.72	24.05
	12	452	50.72	24.80

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 S502 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 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2□	1	264	11.70	13.78
2/3□	1	286	11.44	11.44
3/4□	1	304	12.35	10.66
4/5□	1	315	13.52	10.61
5/6□	1	334	16.54	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 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2□	2	283	11.41	11.70
2/3□	2	307	12.74	10.58
3/4□	2	326	15.08	10.92
4/5□	2	337	17.42	11.70
5/6□	2	355	22.62	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 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2□	3	297	11.39	17.68
2/3□	3	323	11.91	12.74
3/4□	3	342	13.78	10.92
4/5□	3	352	15.34	10.66
5/6□	3	370	19.50	10.66

5.3.2.4 Grades 4–5

Table 5.3.2.4

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 4-5 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2□	4	307	11.36	15.34
	5	316	11.70	13.78
2/3□	4	335	13.00	11.44
	5	345	14.30	10.82
3/4□	4	354	15.86	10.53
	5	364	17.94	10.40
4/5□	4	364	17.94	10.40
	5	373	20.54	10.66
5/6□	4	382	23.40	11.18
	5	391	27.30	11.96

5.3.2.5 Grades 6–8

Table 5.3.2.5

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 6-8 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2 □	6	323	11.70	13.52
	7	329	11.44	12.74
	8	335	11.47	11.96
2/3 □	6	353	11.96	10.92
	7	360	12.48	10.58
	8	366	12.74	10.66
3/4 □	6	373	13.52	10.58
	7	380	14.56	10.92
	8	386	15.60	11.18
4/5 □	6	382	14.87	10.87
	7	389	16.12	11.18
	8	395	17.42	11.70
5/6 □	6	399	18.20	12.22
	7	406	20.28	13.00
	8	412	22.10	13.78

5.3.2.6 Grades 9–12

Table 5.3.2.6

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 9-12 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2 □	9	340	11.70	13.78
	10	344	11.70	13.21
	11	348	11.70	12.74
	12	352	11.70	12.22
2/3 □	9	372	12.58	10.66
	10	377	13.00	10.40
	11	382	13.52	10.40
	12	386	14.04	10.40
3/4 □	9	392	14.82	10.45
	10	397	15.60	10.66
	11	402	16.64	10.92
	12	407	17.68	11.18
4/5 □	9	401	16.38	10.66
	10	406	17.42	11.18
	11	410	18.46	11.44
	12	414	19.50	11.70
5/6 □	9	418	20.54	12.22
	10	423	22.36	12.74
	11	427	23.66	13.26
	12	432	25.74	14.04

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 S502 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 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2□	1	238	10.31	8.54
2/3□	1	275	13.96	10.74
3/4□	1	337	13.69	12.35
4/5□	1	382	13.16	10.74
5/6□	1	405	16.38	11.81

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 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2□	2	242	11.81	8.86
2/3□	2	279	16.27	11.01
3/4□	2	341	17.18	12.30
4/5□	2	388	15.31	11.01
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 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2□	3	247	12.08	8.86
2/3□	3	283	16.65	11.28
3/4□	3	346	17.18	12.14
4/5□	3	394	15.65	11.14
5/6□	3	418	20.41	13.43

5.3.3.4 Grades 4–5

Table 5.3.3.4

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 4-5 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2□	4	266	11.81	11.79
	5	267	11.55	11.55
2/3□	4	288	13.43	8.59
	5	293	14.04	8.32
3/4□	4	351	17.99	12.35
	5	356	17.72	12.35
4/5□	4	401	15.57	11.98
	5	407	15.31	11.81
5/6□	4	425	15.57	11.01
	5	433	16.65	10.74

5.3.3.5 Grades 6–8

Table 5.3.3.5

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 6-8 S502 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 Grades 9–12

Table 5.3.3.6

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 9-12 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2 □	9	289	12.35	8.32
	10	298	12.08	8.59
	11	308	12.89	9.53
	12	318	14.23	10.74
2/3 □	9	319	14.23	10.77
	10	326	15.31	11.28
	11	335	16.38	11.81
	12	344	17.02	12.32
3/4 □	9	378	17.72	12.62
	10	385	17.72	12.35
	11	391	17.45	12.22
	12	398	17.18	12.08
4/5 □	9	430	15.47	10.74
	10	436	15.31	10.74
	11	441	15.57	11.01
	12	447	15.84	11.28
5/6 □	9	469	19.33	14.77
	10	479	22.29	17.45
	11	490	27.12	21.48
	12	501	33.03	26.42

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 S502 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 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2 □	1	205	20.77	15.21
2/3 □	1	261	28.37	19.89
3/4 □	1	311	24.28	17.55
4/5 □	1	361	28.08	19.01
5/6 □	1	403	45.63	29.25

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 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2 □	2	220	24.86	16.67
2/3 □	2	273	26.91	19.30
3/4 □	2	322	24.57	17.55
4/5 □	2	374	35.39	21.64
5/6 □	2	415	62.30	34.80

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 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2□	3	234	26.62	17.55
2/3□	3	283	26.03	19.01
3/4□	3	332	25.45	17.55
4/5□	3	386	40.95	24.28
5/6□	3	425	72.83	40.07

5.3.4.4 Grades 4–5

Table 5.3.4.4

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 4-5 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2□	4	246	22.52	16.38
	5	258	23.98	16.67
2/3□	4	293	28.08	18.72
	5	302	28.08	19.01
3/4□	4	342	24.57	18.13
	5	350	24.28	17.84
4/5□	4	397	29.25	19.01
	5	407	31.88	20.18
5/6□	4	435	45.04	25.74
	5	443	50.60	28.08

5.3.4.5 Grades 6–8

Table 5.3.4.5

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 6-8 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2 □	6	268	21.64	15.79
	7	277	22.23	15.50
	8	284	22.81	15.79
2/3 □	6	310	26.62	18.13
	7	317	27.49	19.01
	8	323	28.08	19.60
3/4 □	6	360	26.03	19.01
	7	369	25.15	18.43
	8	377	24.57	17.84
4/5 □	6	417	25.74	17.84
	7	425	27.20	18.43
	8	433	28.96	19.60
5/6 □	6	451	35.10	23.11
	7	457	37.73	24.86
	8	463	40.95	26.62

5.3.4.6 Grades 9–12

Table 5.3.4.6

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 9-12 S502 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2 □	9	290	25.15	16.96
	10	295	26.03	17.55
	11	299	26.62	17.84
	12	302	26.91	18.13
2/3 □	9	328	27.79	19.60
	10	333	27.20	19.60
	11	337	26.91	19.60
	12	340	26.62	19.30
3/4 □	9	385	24.28	17.26
	10	393	24.86	17.26
	11	400	25.74	17.55
	12	406	26.32	17.84
4/5 □	9	440	36.27	22.52
	10	446	38.90	23.98
	11	451	41.82	25.15
	12	455	43.87	26.62
5/6 □	9	468	52.94	31.29
	10	471	55.57	32.46
	11	474	57.91	33.93
	12	476	59.67	35.10

5.4 Accuracy and Consistency of Domains

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 WIDA proficiency categories determined by the 2016 ACCESS Standard setting process (Cook & MacGregor, 2017). The accuracy and consistency of these classifications can be useful for test users to judge the utility of this information and for policy makers to make decisions about test design and score reporting (American Educational Research Association et al., 2014). The analyses utilize the methods outlined by Livingston and Lewis (1995) and Young and Yoon (1998), as implemented in the software program BB-CLASS (Brennan, 2004; cf. also Lee, Hanson, & Brennan, 2002).

Classification accuracy is defined conceptually as the extent to which the proficiency classifications of students based on the observed test scores would agree with those made based on their true scores (Livingston, 2018; Livingston & Lewis, 1995). True scores are assumed to be measured perfectly but are unknown. Therefore, to provide the best estimation of classification accuracy, we use test data from one test administration to estimate the true scores based on observed scores and the parameters of the model used in estimating the true scores. It is then possible 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 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 taken by Livingston and Lewis (1995) and implemented here uses information about the reliability of the students' test scores, the cut scores, and the observed distribution of scores. Then, using a four-parameter beta distribution, the distribution of the true scores and of scores on a parallel form is modeled. The Livingston and Lewis procedure requires that the reliability estimate of the test form be provided in estimating the classification consistency and accuracy statistics. For Listening and Reading, the Rasch student reliability estimates by grade-level clusters were used in the procedure. Since the Writing and Speaking tests were tiered, it was necessary to produce a single reliability estimate across tiers for 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 who would be classified into the same language proficiency level by both their observed test scores and their true scores. For example, an overall classification accuracy index of 0.774 means that 77% of the students would be classified into the correct proficiency level across all six proficiency levels according to their observed and true scores. Overall classification consistency indicates the percentage of all students who would be classified into the same language proficiency level by both the administered test and by a parallel test. For example, an overall classification consistency index of 0.664 means that 66% of the students would be classified into the same proficiency level if two parallel forms were administered. Classification consistency indices are always lower than the corresponding classification accuracy indices, because in classification consistency, classification based on students' performance on the administered test and classification based on students' performance on a parallel test are both subject to measurement error. In contrast, in classification accuracy, only the classification based on students' performance on the administered test contains error, while classification based on students' true score is assumed to be free of measurement error.

Marginal Classification Accuracy and Consistency

Overall classification accuracy and consistency indices indicate the degree to which students are accurately and consistently classified in the same WIDA proficiency levels, but not the degree to which students are accurately or consistently classified 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 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 scale scores at the cut points report the percentage of students who are accurately placed into proficiency levels above and below each cut point based on their scale scores. For example, a classification accuracy index of 0.774 at the PL 4/PL 5 cut point means that 77% of the students would be classified in the same way if they were classified according to their observed scale score and their true scale score, 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 scale scores at the cut points report the percentage of students classified consistently above and below each cut point based on their scale scores. For example, a classification consistency index of 0.664 at the PL 4/PL 5 cut point means that 66% of the students would be classified in the same way if two parallel forms were administered, 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' scale scores at the cut points than the overall classification accuracy and consistency (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 calculation of classification accuracy and consistency indices is affected by the interactions of a number of factors: (1) the number of proficiency level cut points, (2) the magnitude of the test score reliability coefficient, (3) the 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 (Lee et al., 2002; Ercikan & Julian, 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 a 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' 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 the test score reliability coefficient for ACCESS Online Writing domain raw scores is estimated based on only two tasks. Therefore, the classification accuracy and consistency indices for the Writing domain might be lower than those of other domains as a result.

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 scale scores at the cut points for each grade level. The third reports the marginal classification consistency indices based on 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 fewer than 200 students were classified 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 very 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, overall classification accuracy ranged from 0.375 to 0.698 and overall classification consistency ranged from 0.311 to 0.637. The lowest overall classification accuracy indices for Listening were at Grade 8. The lowest overall classification consistency index for Listening was at Grade 7.

For Reading, as shown in Table 5.4.2.1, overall classification accuracy ranged from 0.433 to 0.841 and overall classification consistency ranged from 0.338 to 0.820. The lowest overall classification accuracy and consistency index for Reading was Grade 3.

For Writing, as shown in Table 5.4.3.1, overall classification accuracy ranged from 0.706 to 0.836 and overall classification consistency ranged from 0.645 to 0.788. The lowest overall classification accuracy and consistency index for Writing was Grade 4.

For Speaking, as shown in Table 5.4.4.1, overall classification accuracy ranged from 0.487 to 0.687 and overall classification consistency ranged from 0.490 to 0.590. The lowest overall classification accuracy and consistency index for Speaking was Kindergarten.

These results suggest that the grade level with the lowest classification accuracy and consistency tends to vary across the four domains.

From an accountability perspective, the most important information for test users and policy makers to examine is the marginal classification accuracy and consistency. We summarize the range of the marginal classification accuracy and consistency of domains across grades, by domain, and highlight the grade level with the lowest marginal classification accuracy and with the lowest consistency, by domain, for test users and policy makers.

For Listening, classification accuracy indices at the cut scores ranged from 0.766 to 0.994 (Table 5.4.1.2) and classification consistency at the cut scores ranged from 0.693 to 0.989 (Table 5.4.1.3). The lowest classification accuracy and consistency indices for Listening were Grade 6 and Grade 7 at the PL 4/PL 5 cut level. The low marginal classification consistency at the PL 4/PL 5 cut score appeared to have contributed to its low overall classification consistency.

For Reading, classification accuracy indices at the cut scores ranged from 0.760 to 0.971 (Table 5.4.2.2) and classification consistency at the cut scores ranged from 0.692 to 0.957 (Table 5.4.2.3). The lowest classification accuracy and consistency value for Reading was Grade 3 at the PL 3/PL 4 cut. Note that Grade 3 was also identified as having the lowest overall classification consistency 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.

For Writing, classification accuracy indices at the cut scores ranged from 0.762 to 0.983 (Table 5.4.3.2) and classification consistency at the cut scores ranged from 0.720 to 0.975 (Table 5.4.3.3). The lowest classification accuracy and consistency indices for Writing was Grade 4 at the PL 3/PL 4 cut. Note that Grade 4 was also identified as having the lowest overall classification consistency 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.

For Speaking, classification accuracy indices at the cut scores ranged from 0.746 to 0.988 (Table 5.4.4.2) and classification consistency at the cut scores ranged from 0.810 to 0.988 (Table 5.4.4.3). The lowest classification accuracy and consistency value for Speaking was Kindergarten at the PL 5/PL 6 cut. Note that Kindergarten was also identified as having the lowest overall classification accuracy and consistency 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 .80's.

The grades with the lowest overall classification accuracy and consistency were the same grades with the lowest marginal classification accuracy and consistency for three domains, Reading (Grade 3), Writing (Grade 4), and Speaking (Kindergarten). In Listening, Grade 8 had the lowest overall classification accuracy and consistency, and Grade 6 and Grade 7 had the lowest marginal classification accuracy and consistency.

We observed that the lowest marginal classification accuracy and consistency for three domains (Listening, Reading, and Writing) 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; Ercikan & Julian, 2002), in that classification accuracy and consistency at cut points in the middle of the proficiency level range are lower than those in the lower and upper ends.

Having 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 are many 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 would be misclassified (Ercikan & Julian, 2002). Since ACCESS has six proficiency levels and PL 3 and PL 4 occupy relatively narrow ranges on the ability scale as compared to other proficiency levels, the classification accuracy and consistency for the 3/4 and 4/5 cuts are lower than for other cuts.

The lowest marginal classification accuracy and consistency of the Speaking domain (Kindergarten) occurred at the PL 5/PL 6 cut point, which is the highest cut point on the proficiency scale. Extreme cuts tend to have larger measurement error. Thus, among the many factors mentioned earlier that affect the magnitude of classification accuracy and consistency, a large standard error at the PL 5/PL 6 cut point may have contributed to the lower classification accuracy and consistency at this cut point.

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 the English Language Proficiency Assessment for the 21st Century (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 that of 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 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 S502 Paper

Grade	Accuracy	Consistency
K	0.698	0.637
1	0.475	0.390
2	0.478	0.389
3	0.659	0.576
4	0.492	0.393
5	0.480	0.383
6	0.406	0.321
7	0.395	0.311
8	0.375	0.311
9	0.402	0.316
10	0.404	0.314
11	0.406	0.322
12	0.435	0.335

Table 5.4.1.2

Classification Accuracy Indices at Cut Score Level: List S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.941	0.931	0.923	0.916	0.907
1	0.960	0.903	0.835	0.805	0.834
2	0.976	0.918	0.839	0.806	0.810
3	0.994	0.937	0.846	0.801	N/A
4	0.992	0.948	0.865	0.792	0.805
5	0.987	0.937	0.866	0.794	0.791
6	0.963	0.905	0.837	0.766	0.814
7	0.945	0.897	0.820	0.766	0.827
8	0.935	0.888	0.834	0.789	0.778
9	0.929	0.858	0.799	0.805	0.881
10	0.906	0.835	0.784	0.829	0.918
11	0.903	0.839	0.792	0.837	0.901
12	0.885	0.821	0.787	0.868	0.960

Table 5.4.1.3

Classification Consistency Indices at Cut Score Level: List S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.917	0.904	0.893	0.886	0.870
1	0.939	0.866	0.768	0.740	0.770
2	0.965	0.880	0.775	0.735	0.741
3	0.989	0.911	0.781	0.715	N/A
4	0.987	0.920	0.804	0.721	0.730
5	0.979	0.906	0.802	0.721	0.716
6	0.948	0.861	0.764	0.693	0.748
7	0.924	0.844	0.749	0.693	0.758
8	0.908	0.837	0.762	0.708	0.728
9	0.899	0.795	0.727	0.737	0.831
10	0.865	0.767	0.712	0.760	0.870
11	0.857	0.773	0.720	0.768	0.866
12	0.834	0.751	0.716	0.804	0.927

5.4.2 Reading

Table 5.4.2.1

Overall Accuracy and Consistency of Classification Indices: Read S502 Paper

Grade	Accuracy	Consistency
K	0.841	0.820
1	0.494	0.395
2	0.567	0.457
3	0.433	0.338
4	0.512	0.404
5	0.502	0.398
6	0.598	0.485
7	0.566	0.456
8	0.547	0.440
9	0.538	0.436
10	0.559	0.450
11	0.552	0.445
12	0.600	0.490

Table 5.4.2.2

Classification Accuracy Indices at Cut Score Level: Read S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.965	0.966	0.956	0.926	N/A
1	0.805	0.815	0.903	0.941	0.969
2	0.884	0.860	0.899	0.922	0.956
3	0.940	0.853	0.760	0.815	0.938
4	0.950	0.869	0.825	0.863	0.938
5	0.942	0.863	0.831	0.850	0.915
6	0.924	0.850	0.891	0.918	0.971
7	0.915	0.843	0.876	0.913	0.961
8	0.898	0.851	0.873	0.900	0.953
9	0.932	0.868	0.852	0.881	0.938
10	0.922	0.851	0.875	0.902	0.942
11	0.926	0.855	0.864	0.887	0.941
12	0.908	0.861	0.902	0.918	0.958

Table 5.4.2.3

Classification Consistency Indices at Cut Score Level: Read S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.948	0.950	0.940	0.920	N/A
1	0.747	0.749	0.857	0.909	0.957
2	0.836	0.808	0.854	0.887	0.937
3	0.914	0.781	0.692	0.744	0.897
4	0.928	0.813	0.768	0.806	0.907
5	0.914	0.808	0.773	0.792	0.872
6	0.888	0.791	0.847	0.886	0.957
7	0.877	0.784	0.828	0.872	0.942
8	0.854	0.792	0.825	0.859	0.929
9	0.905	0.812	0.803	0.834	0.908
10	0.887	0.796	0.826	0.858	0.916
11	0.894	0.799	0.814	0.842	0.910
12	0.871	0.807	0.859	0.882	0.940

5.4.3 Writing

Table 5.4.3.1

Overall Accuracy and Consistency of Classification Indices: Writ S502 Paper

Grade	Accuracy	Consistency
K	0.808	0.774
1	0.754	0.687
2	0.836	0.788
3	0.826	0.778
4	0.706	0.645
5	0.781	0.704
6	0.788	0.707
7	0.786	0.705
8	0.789	0.708
9	0.806	0.727
10	0.795	0.715
11	0.788	0.707
12	0.795	0.714

Table 5.4.3.2

Classification Accuracy Indices at Cut Score Level: Writ S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.928	0.890	0.971	N/A	N/A
1	0.932	0.822	N/A	N/A	N/A
2	0.961	0.925	0.951	N/A	N/A
3	0.973	0.944	0.910	N/A	N/A
4	0.981	0.963	0.762	N/A	N/A
5	0.983	0.963	0.834	N/A	N/A
6	0.966	0.937	0.884	N/A	N/A
7	0.966	0.938	0.883	N/A	N/A
8	0.961	0.941	0.887	N/A	N/A
9	0.973	0.947	0.885	N/A	N/A
10	0.966	0.937	0.893	N/A	N/A
11	0.961	0.933	0.893	N/A	N/A
12	0.958	0.933	0.902	N/A	N/A

Table 5.4.3.3

Classification Consistency Indices at Cut Score Level: Writ S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.896	0.868	0.966	N/A	N/A
1	0.902	0.782	N/A	N/A	N/A
2	0.943	0.897	0.949	N/A	N/A
3	0.960	0.922	0.896	N/A	N/A
4	0.972	0.950	0.720	N/A	N/A
5	0.975	0.948	0.779	N/A	N/A
6	0.952	0.911	0.841	N/A	N/A
7	0.951	0.912	0.839	N/A	N/A
8	0.946	0.914	0.843	N/A	N/A
9	0.962	0.925	0.839	N/A	N/A
10	0.951	0.911	0.850	N/A	N/A
11	0.944	0.906	0.851	N/A	N/A
12	0.941	0.903	0.862	N/A	N/A

5.4.4 Speaking

Table 5.4.4.1

Overall Accuracy and Consistency of Classification Indices: Spek S502 Paper

Grade	Accuracy	Consistency
K	0.487	0.490
1	0.687	0.580
2	0.686	0.583
3	0.680	0.579
4	0.652	0.548
5	0.643	0.540
6	0.633	0.532
7	0.624	0.527
8	0.637	0.538
9	0.638	0.548
10	0.665	0.575
11	0.659	0.577
12	0.671	0.590

Table 5.4.4.2

Classification Accuracy Indices at Cut Score Level: Spek S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.882	0.928	0.950	0.953	0.746
1	0.938	0.895	0.911	0.951	0.988
2	0.955	0.899	0.901	0.942	0.978
3	0.951	0.896	0.897	0.931	0.964
4	0.967	0.931	0.901	0.911	0.937
5	0.961	0.930	0.901	0.891	0.949
6	0.952	0.913	0.902	0.925	0.935
7	0.950	0.921	0.897	0.924	0.921
8	0.942	0.920	0.906	0.917	0.939
9	0.936	0.913	0.914	0.936	0.925
10	0.934	0.918	0.909	0.953	0.935
11	0.937	0.912	0.916	0.961	0.921
12	0.929	0.905	0.931	0.969	0.926

Table 5.4.4.3

Classification Consistency Indices at Cut Score Level: Spek S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.843	0.901	0.926	0.928	0.810
1	0.910	0.854	0.873	0.933	0.988
2	0.934	0.858	0.860	0.930	0.978
3	0.930	0.857	0.857	0.932	0.966
4	0.951	0.903	0.861	0.867	0.935
5	0.943	0.901	0.863	0.844	0.946
6	0.930	0.880	0.862	0.888	0.932
7	0.928	0.890	0.856	0.881	0.921
8	0.917	0.888	0.868	0.879	0.933
9	0.908	0.879	0.878	0.905	0.925
10	0.906	0.884	0.873	0.923	0.938
11	0.910	0.876	0.880	0.938	0.932
12	0.899	0.868	0.899	0.953	0.936

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 can be explained by the variance in their true composite scale scores.

Because of the differential weights applied to the ACCESS domains that contribute to the composites, the stratified Cronbach's alpha coefficient is weighted by the contribution of each domain score into the composite (Rudner, 2001; Kamata, Turhan, & Darandari, 2003; Kane & Case, 2004). 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 = number of components j

w_j = weight of component j

σ_j^2 = variance of component j

σ_c^2 = variance of composite

ρ_j = reliability coefficient of component j

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 0.88 to 0.95 for male students; from 0.88 to 0.95 for female students; from 0.88 to 0.96 for Hispanic students; from 0.87 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 0.97. The stratified Cronbach's alpha of the Literacy composite ranged from 0.90 to 0.97 for male students; from 0.90 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 0.89 to 0.97 for students with an IEP.

The stratified Cronbach's alpha of the Comprehension composite computed for all students ranged from 0.75 to 0.97. The stratified Cronbach's alpha of the Comprehension composite ranged from 0.76 to 0.97 for male students; from 0.74 to 0.96 for female students; from 0.74 to

0.96 for Hispanic students; from 0.78 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 0.93 to 0.98. The stratified Cronbach's alpha of the Overall composite ranged from 0.93 to 0.98 for male students; from 0.93 to 0.97 for female students; from 0.93 to 0.97 for Hispanic students; from 0.94 to 0.98 for non-Hispanic students; and from 0.93 to 0.98 for students with an IEP.

5.5.1 Oral

Table 5.5.1.1

Reliabilities of Composite Scale Scores: Oral S502 Paper

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.50	6079.58	0.95
	Speaking	0.50	10329.95	0.91
	Oral		7247.00	0.96
1	Listening	0.50	1598.73	0.69
	Speaking	0.50	3969.69	0.90
	Oral		2033.29	0.89
2	Listening	0.50	1553.58	0.66
	Speaking	0.50	3781.73	0.91
	Oral		1919.44	0.88
3	Listening	0.50	1231.57	0.56
	Speaking	0.50	3988.13	0.91
	Oral		1837.89	0.88
4-5	Listening	0.50	1535.68	0.63
	Speaking	0.50	4543.84	0.91
	Oral		2261.08	0.89
6-8	Listening	0.50	2137.28	0.63
	Speaking	0.50	5207.62	0.90
	Oral		2901.37	0.89
9-12	Listening	0.50	2309.33	0.65
	Speaking	0.50	5683.97	0.92
	Oral		3176.29	0.90

Table 5.5.1.2

Reliabilities of Composite Scale Scores: Oral S502 Paper by Gender

Cluster	Component	Weight	Female		Male	
			Variance	Reliability	Variance	Reliability
K	Listening	0.50	5796.53	0.94	6221.58	0.95
	Speaking	0.50	10443.23	0.91	10046.00	0.90
	Oral		7145.37	0.95	7194.59	0.95
1	Listening	0.50	1559.13	0.68	1622.81	0.70
	Speaking	0.50	4039.54	0.89	3865.46	0.90
	Oral		2039.07	0.89	2003.55	0.89
2	Listening	0.50	1509.71	0.65	1580.37	0.66
	Speaking	0.50	3762.52	0.91	3782.93	0.90
	Oral		1891.37	0.88	1931.67	0.88
3	Listening	0.50	1189.43	0.55	1267.58	0.57
	Speaking	0.50	4004.99	0.91	3959.13	0.91
	Oral		1821.53	0.88	1845.63	0.88
4-5	Listening	0.50	1471.94	0.62	1586.76	0.64
	Speaking	0.50	4568.86	0.91	4507.89	0.91
	Oral		2232.85	0.89	2277.71	0.89
6-8	Listening	0.50	2106.94	0.63	2160.10	0.64
	Speaking	0.50	5156.50	0.90	5250.39	0.91
	Oral		2863.19	0.89	2932.16	0.89
9-12	Listening	0.50	2211.56	0.63	2395.86	0.66
	Speaking	0.50	5745.54	0.91	5631.26	0.92
	Oral		3156.36	0.90	3195.48	0.90

Table 5.5.1.3

Reliabilities of Composite Scale Scores: Oral S502 Paper by Ethnicity

Cluster	Component	Weight	Hispanic		Other	
			Variance	Reliability	Variance	Reliability
K	Listening	0.50	6182.06	0.95	5301.64	0.94
	Speaking	0.50	10452.42	0.91	9365.14	0.89
	Oral		7354.05	0.96	6384.68	0.95
1	Listening	0.50	1582.53	0.69	1650.25	0.70
	Speaking	0.50	3980.66	0.90	3795.43	0.89
	Oral		2023.32	0.89	2020.38	0.89
2	Listening	0.50	1560.12	0.66	1517.35	0.67
	Speaking	0.50	3822.59	0.91	3529.52	0.90
	Oral		1941.06	0.88	1798.29	0.88
3	Listening	0.50	1214.77	0.54	1276.56	0.60
	Speaking	0.50	4072.19	0.91	3471.29	0.90
	Oral		1859.80	0.88	1662.98	0.87
4-5	Listening	0.50	1559.80	0.63	1403.16	0.64
	Speaking	0.50	4727.28	0.91	3678.53	0.90
	Oral		2349.49	0.89	1829.27	0.88
6-8	Listening	0.50	2192.84	0.63	1827.43	0.64
	Speaking	0.50	5396.17	0.91	4270.44	0.90
	Oral		3007.06	0.89	2350.38	0.88
9-12	Listening	0.50	2347.86	0.65	2115.65	0.65
	Speaking	0.50	5879.28	0.92	4773.72	0.91
	Oral		3282.96	0.90	2673.80	0.89

Table 5.5.1.4

Reliabilities of Composite Scale Scores: Oral S502 Paper by IEP Status

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.50	6807.88	0.95
	Speaking	0.50	8883.13	0.90
	Oral		7019.85	0.96
1	Listening	0.50	1742.81	0.72
	Speaking	0.50	3638.05	0.89
	Oral		1990.09	0.89
2	Listening	0.50	1682.78	0.69
	Speaking	0.50	3437.56	0.89
	Oral		1829.17	0.88
3	Listening	0.50	1093.14	0.52
	Speaking	0.50	3401.61	0.90
	Oral		1528.78	0.86
4-5	Listening	0.50	1217.91	0.61
	Speaking	0.50	3335.35	0.90
	Oral		1566.31	0.87
6-8	Listening	0.50	1336.51	0.59
	Speaking	0.50	3658.91	0.90
	Oral		1749.83	0.87
9-12	Listening	0.50	1805.04	0.61
	Speaking	0.50	4974.53	0.91
	Oral		2522.71	0.89

5.5.2 Literacy

Table 5.5.2.1

Reliabilities of Composite Scale Scores: Litr S502 Paper

Cluster	Component	Weight	Variance	Reliability
K	Reading	0.50	4780.51	0.95
	Writing	0.50	4828.10	0.93
	Literacy			4173.59
1	Reading	0.50	713.73	0.70
	Writing	0.50	1936.06	0.92
	Literacy			929.10
2	Reading	0.50	1014.89	0.83
	Writing	0.50	1900.15	0.94
	Literacy			1159.09
3	Reading	0.50	686.34	0.64
	Writing	0.50	1761.29	0.94
	Literacy			921.12
4-5	Reading	0.50	938.27	0.79
	Writing	0.50	1671.83	0.91
	Literacy			1053.16
6-8	Reading	0.50	883.72	0.79
	Writing	0.50	1717.71	0.90
	Literacy			1051.60
9-12	Reading	0.50	987.04	0.81
	Writing	0.50	1562.82	0.90
	Literacy			1050.27

Table 5.5.2.2

Reliabilities of Composite Scale Scores: Litr S502 Paper by Gender

Cluster	Component	Weight	Female		Male	
			Variance	Reliability	Variance	Reliability
K	Reading	0.50	4615.26	0.95	4926.68	0.96
	Writing	0.50	4771.75	0.93	4843.27	0.93
	Literacy		4072.73	0.96	4248.16	0.97
1	Reading	0.50	689.33	0.70	736.33	0.71
	Writing	0.50	1728.94	0.92	2093.20	0.92
	Literacy		858.62	0.90	985.52	0.90
2	Reading	0.50	997.02	0.83	1027.91	0.82
	Writing	0.50	1736.45	0.94	1972.72	0.94
	Literacy		1095.71	0.94	1189.45	0.94
3	Reading	0.50	628.68	0.62	735.29	0.66
	Writing	0.50	1648.72	0.93	1780.52	0.94
	Literacy		855.23	0.90	956.83	0.91
4-5	Reading	0.50	890.60	0.78	976.07	0.79
	Writing	0.50	1568.46	0.91	1691.25	0.91
	Literacy		1002.09	0.92	1075.31	0.92
6-8	Reading	0.50	862.60	0.79	896.43	0.79
	Writing	0.50	1652.03	0.90	1711.93	0.91
	Literacy		1018.38	0.91	1053.75	0.92
9-12	Reading	0.50	955.75	0.80	1005.66	0.81
	Writing	0.50	1546.15	0.90	1534.89	0.90
	Literacy		1033.69	0.91	1042.38	0.92

Table 5.5.2.3

Reliabilities of Composite Scale Scores: Litr S502 Paper by Ethnicity

Cluster	Component	Weight	Hispanic		Other	
			Variance	Reliability	Variance	Reliability
K	Reading	0.50	4141.89	0.95	5377.44	0.96
	Writing	0.50	4295.21	0.92	5194.61	0.94
	Literacy			3593.26	0.96	4634.94
1	Reading	0.50	692.10	0.69	776.41	0.74
	Writing	0.50	1895.40	0.92	2054.55	0.92
	Literacy			900.57	0.90	1015.58
2	Reading	0.50	993.87	0.82	1059.31	0.84
	Writing	0.50	1880.49	0.94	1961.02	0.94
	Literacy			1138.25	0.93	1215.05
3	Reading	0.50	678.34	0.63	695.20	0.69
	Writing	0.50	1770.01	0.93	1694.31	0.94
	Literacy			919.98	0.90	898.81
4-5	Reading	0.50	945.65	0.78	885.35	0.80
	Writing	0.50	1705.54	0.91	1486.56	0.92
	Literacy			1073.65	0.92	938.02
6-8	Reading	0.50	888.37	0.78	846.21	0.80
	Writing	0.50	1758.30	0.90	1523.17	0.91
	Literacy			1072.60	0.92	944.07
9-12	Reading	0.50	1007.00	0.81	902.78	0.80
	Writing	0.50	1578.31	0.90	1488.78	0.90
	Literacy			1070.77	0.92	961.23

Table 5.5.2.4

Reliabilities of Composite Scale Scores: Litr S502 Paper by IEP Status

Cluster	Component	Weight	Variance	Reliability
K	Reading	0.50	4796.55	0.96
	Writing	0.50	4141.93	0.92
	Literacy		3768.49	0.97
1	Reading	0.50	573.25	0.60
	Writing	0.50	2255.76	0.92
	Literacy		921.67	0.89
2	Reading	0.50	879.67	0.78
	Writing	0.50	2171.74	0.94
	Literacy		1164.69	0.93
3	Reading	0.50	621.93	0.56
	Writing	0.50	1920.88	0.95
	Literacy		918.17	0.90
4-5	Reading	0.50	670.30	0.74
	Writing	0.50	1504.79	0.92
	Literacy		810.67	0.91
6-8	Reading	0.50	547.37	0.71
	Writing	0.50	1395.04	0.92
	Literacy		709.46	0.90
9-12	Reading	0.50	696.84	0.75
	Writing	0.50	1379.57	0.91
	Literacy		784.96	0.90

5.5.3 Comprehension

Table 5.5.3.1

Reliabilities of Composite Scale Scores: Cphn S502 Paper

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.30	6079.58	0.95
	Reading	0.70	4780.51	0.95
	Comprehension			4024.82
1	Listening	0.30	1598.73	0.69
	Reading	0.70	713.73	0.70
	Comprehension			706.30
2	Listening	0.30	1553.58	0.66
	Reading	0.70	1014.89	0.83
	Comprehension			922.67
3	Listening	0.30	1231.57	0.56
	Reading	0.70	686.34	0.64
	Comprehension			685.41
4-5	Listening	0.30	1535.68	0.63
	Reading	0.70	938.27	0.79
	Comprehension			947.43
6-8	Listening	0.30	2137.28	0.63
	Reading	0.70	883.72	0.79
	Comprehension			1035.39
9-12	Listening	0.30	2309.33	0.65
	Reading	0.70	987.04	0.81
	Comprehension			1149.88

Table 5.5.3.2

Reliabilities of Composite Scale Scores: Cphn S502 Paper by Gender

Cluster	Component	Weight	Female		Male	
			Variance	Reliability	Variance	Reliability
K	Listening	0.30	5796.53	0.94	6221.58	0.95
	Reading	0.70	4615.26	0.95	4926.68	0.96
	Comprehension			3885.59	0.96	4131.61
1	Listening	0.30	1559.13	0.68	1622.81	0.70
	Reading	0.70	689.33	0.70	736.33	0.71
	Comprehension			688.87	0.79	720.79
2	Listening	0.30	1509.71	0.65	1580.37	0.66
	Reading	0.70	997.02	0.83	1027.91	0.82
	Comprehension			906.15	0.86	932.47
3	Listening	0.30	1189.43	0.55	1267.58	0.57
	Reading	0.70	628.68	0.62	735.29	0.66
	Comprehension			633.58	0.74	729.24
4-5	Listening	0.30	1471.94	0.62	1586.76	0.64
	Reading	0.70	890.60	0.78	976.07	0.79
	Comprehension			899.09	0.84	985.73
6-8	Listening	0.30	2106.94	0.63	2160.10	0.64
	Reading	0.70	862.60	0.79	896.43	0.79
	Comprehension			1018.29	0.84	1046.05
9-12	Listening	0.30	2211.56	0.63	2395.86	0.66
	Reading	0.70	955.75	0.80	1005.66	0.81
	Comprehension			1110.03	0.85	1182.12

Table 5.5.3.3

Reliabilities of Composite Scale Scores: Cphn S502 Paper by Ethnicity

Cluster	Component	Weight	Hispanic		Other	
			Variance	Reliability	Variance	Reliability
K	Listening	0.30	6182.06	0.95	5301.64	0.94
	Reading	0.70	4141.89	0.95	5377.44	0.96
	Comprehension			3607.58	0.96	4236.00
1	Listening	0.30	1582.53	0.69	1650.25	0.70
	Reading	0.70	692.10	0.69	776.41	0.74
	Comprehension			686.59	0.78	769.27
2	Listening	0.30	1560.12	0.66	1517.35	0.67
	Reading	0.70	993.87	0.82	1059.31	0.84
	Comprehension			906.74	0.85	959.98
3	Listening	0.30	1214.77	0.54	1276.56	0.60
	Reading	0.70	678.34	0.63	695.20	0.69
	Comprehension			676.13	0.74	698.55
4-5	Listening	0.30	1559.80	0.63	1403.16	0.64
	Reading	0.70	945.65	0.78	885.35	0.80
	Comprehension			960.56	0.84	868.43
6-8	Listening	0.30	2192.84	0.63	1827.43	0.64
	Reading	0.70	888.37	0.78	846.21	0.80
	Comprehension			1050.90	0.84	939.61
9-12	Listening	0.30	2347.86	0.65	2115.65	0.65
	Reading	0.70	1007.00	0.81	902.78	0.80
	Comprehension			1173.62	0.85	1043.30

Table 5.5.3.4

Reliabilities of Composite Scale Scores: Cphn S502 Paper by IEP Status

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.30	6807.88	0.95
	Reading	0.70	4796.55	0.96
	Comprehension		4027.80	0.97
1	Listening	0.30	1742.81	0.72
	Reading	0.70	573.25	0.60
	Comprehension		614.75	0.75
2	Listening	0.30	1682.78	0.69
	Reading	0.70	879.67	0.78
	Comprehension		843.81	0.83
3	Listening	0.30	1093.14	0.52
	Reading	0.70	621.93	0.56
	Comprehension		593.01	0.69
4-5	Listening	0.30	1217.91	0.61
	Reading	0.70	670.30	0.74
	Comprehension		658.26	0.80
6-8	Listening	0.30	1336.51	0.59
	Reading	0.70	547.37	0.71
	Comprehension		597.12	0.79
9-12	Listening	0.30	1805.04	0.61
	Reading	0.70	696.84	0.75
	Comprehension		808.90	0.81

5.5.4 Overall

Table 5.5.4.1

Reliabilities of Composite Scale Scores: Over S502 Paper

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.15	6079.58	0.95
	Reading	0.35	4780.51	0.95
	Writing	0.35	4828.10	0.93
	Speaking	0.15	10329.95	0.91
	Overall Composite		3997.73	0.98
1	Listening	0.15	1598.73	0.69
	Reading	0.35	713.73	0.70
	Writing	0.35	1936.06	0.92
	Speaking	0.15	3969.69	0.90
	Overall Composite		976.39	0.93
2	Listening	0.15	1553.58	0.66
	Reading	0.35	1014.89	0.83
	Writing	0.35	1900.15	0.94
	Speaking	0.15	3781.73	0.91
	Overall Composite		1127.76	0.95
3	Listening	0.15	1231.57	0.56
	Reading	0.35	686.34	0.64
	Writing	0.35	1761.29	0.94
	Speaking	0.15	3988.13	0.91
	Overall Composite		978.53	0.93
4-5	Listening	0.15	1535.68	0.63
	Reading	0.35	938.27	0.79
	Writing	0.35	1671.83	0.91
	Speaking	0.15	4543.84	0.91
	Overall Composite		1191.14	0.95
6-8	Listening	0.15	2137.28	0.63
	Reading	0.35	883.72	0.79
	Writing	0.35	1717.71	0.90
	Speaking	0.15	5207.62	0.90
	Overall Composite		1343.46	0.95
9-12	Listening	0.15	2309.33	0.65
	Reading	0.35	987.04	0.81
	Writing	0.35	1562.82	0.90
	Speaking	0.15	5683.97	0.92
	Overall Composite		1386.73	0.95

Table 5.5.4.2

Reliabilities of Composite Scale Scores: Over S502 Paper by Gender

Cluster	Component	Weight	Female		Male	
			Variance	Reliability	Variance	Reliability
K	Listening	0.15	5796.53	0.94	6221.58	0.95
	Reading	0.35	4615.26	0.95	4926.68	0.96
	Writing	0.35	4771.75	0.93	4843.27	0.93
	Speaking	0.15	10443.23	0.91	10046.00	0.90
	Overall Composite		3910.10	0.97	4037.26	0.98
1	Listening	0.15	1559.13	0.68	1622.81	0.70
	Reading	0.35	689.33	0.70	736.33	0.71
	Writing	0.35	1728.94	0.92	2093.20	0.92
	Speaking	0.15	4039.54	0.89	3865.46	0.90
	Overall Composite		930.75	0.93	1005.84	0.93
2	Listening	0.15	1509.71	0.65	1580.37	0.66
	Reading	0.35	997.02	0.83	1027.91	0.82
	Writing	0.35	1736.45	0.94	1972.72	0.94
	Speaking	0.15	3762.52	0.91	3782.93	0.90
	Overall Composite		1081.61	0.95	1147.15	0.95
3	Listening	0.15	1189.43	0.55	1267.58	0.57
	Reading	0.35	628.68	0.62	735.29	0.66
	Writing	0.35	1648.72	0.93	1780.52	0.94
	Speaking	0.15	4004.99	0.91	3959.13	0.91
	Overall Composite		930.93	0.93	1005.79	0.94
4-5	Listening	0.15	1471.94	0.62	1586.76	0.64
	Reading	0.35	890.60	0.78	976.07	0.79
	Writing	0.35	1568.46	0.91	1691.25	0.91
	Speaking	0.15	4568.86	0.91	4507.89	0.91
	Overall Composite		1152.50	0.94	1208.49	0.95
6-8	Listening	0.15	2106.94	0.63	2160.10	0.64
	Reading	0.35	862.60	0.79	896.43	0.79
	Writing	0.35	1652.03	0.90	1711.93	0.91
	Speaking	0.15	5156.50	0.90	5250.39	0.91
	Overall Composite		1318.65	0.94	1349.02	0.95
9-12	Listening	0.15	2211.56	0.63	2395.86	0.66
	Reading	0.35	955.75	0.80	1005.66	0.81
	Writing	0.35	1546.15	0.90	1534.89	0.90
	Speaking	0.15	5745.54	0.91	5631.26	0.92
	Overall Composite		1384.07	0.95	1380.29	0.95

Table 5.5.4.3

Reliabilities of Composite Scale Scores: Over S502 Paper by Ethnicity

Cluster	Component	Weight	Hispanic		Other	
			Variance	Reliability	Variance	Reliability
K	Listening	0.15	6182.06	0.95	5301.64	0.94
	Reading	0.35	4141.89	0.95	5377.44	0.96
	Writing	0.35	4295.21	0.92	5194.61	0.94
	Speaking	0.15	10452.42	0.91	9365.14	0.89
	Overall Composite		3603.09	0.97	4124.98	0.98
1	Listening	0.15	1582.53	0.69	1650.25	0.70
	Reading	0.35	692.10	0.69	776.41	0.74
	Writing	0.35	1895.40	0.92	2054.55	0.92
	Speaking	0.15	3980.66	0.90	3795.43	0.89
	Overall Composite		953.26	0.93	1036.45	0.94
2	Listening	0.15	1560.12	0.66	1517.35	0.67
	Reading	0.35	993.87	0.82	1059.31	0.84
	Writing	0.35	1880.49	0.94	1961.02	0.94
	Speaking	0.15	3822.59	0.91	3529.52	0.90
	Overall Composite		1116.71	0.95	1141.39	0.95
3	Listening	0.15	1214.77	0.54	1276.56	0.60
	Reading	0.35	678.34	0.63	695.20	0.69
	Writing	0.35	1770.01	0.93	1694.31	0.94
	Speaking	0.15	4072.19	0.91	3471.29	0.90
	Overall Composite		984.44	0.93	912.82	0.94
4-5	Listening	0.15	1559.80	0.63	1403.16	0.64
	Reading	0.35	945.65	0.78	885.35	0.80
	Writing	0.35	1705.54	0.91	1486.56	0.92
	Speaking	0.15	4727.28	0.91	3678.53	0.90
	Overall Composite		1228.80	0.95	996.28	0.94
6-8	Listening	0.15	2192.84	0.63	1827.43	0.64
	Reading	0.35	888.37	0.78	846.21	0.80
	Writing	0.35	1758.30	0.90	1523.17	0.91
	Speaking	0.15	5396.17	0.91	4270.44	0.90
	Overall Composite		1383.51	0.95	1136.33	0.94
9-12	Listening	0.15	2347.86	0.65	2115.65	0.65
	Reading	0.35	1007.00	0.81	902.78	0.80
	Writing	0.35	1578.31	0.90	1488.78	0.90
	Speaking	0.15	5879.28	0.92	4773.72	0.91
	Overall Composite		1430.61	0.95	1193.49	0.94

Table 5.5.4.4

Reliabilities of Composite Scale Scores: Over S502 Paper by IEP Status

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.15	6807.88	0.95
	Reading	0.35	4796.55	0.96
	Writing	0.35	4141.93	0.92
	Speaking	0.15	8883.13	0.90
	Overall Composite		3649.55	0.98
1	Listening	0.15	1742.81	0.72
	Reading	0.35	573.25	0.60
	Writing	0.35	2255.76	0.92
	Speaking	0.15	3638.05	0.89
	Overall Composite		954.69	0.93
2	Listening	0.15	1682.78	0.69
	Reading	0.35	879.67	0.78
	Writing	0.35	2171.74	0.94
	Speaking	0.15	3437.56	0.89
	Overall Composite		1100.03	0.95
3	Listening	0.15	1093.14	0.52
	Reading	0.35	621.93	0.56
	Writing	0.35	1920.88	0.95
	Speaking	0.15	3401.61	0.90
	Overall Composite		891.35	0.93
4-5	Listening	0.15	1217.91	0.61
	Reading	0.35	670.30	0.74
	Writing	0.35	1504.79	0.92
	Speaking	0.15	3335.35	0.90
	Overall Composite		827.21	0.93
6-8	Listening	0.15	1336.51	0.59
	Reading	0.35	547.37	0.71
	Writing	0.35	1395.04	0.92
	Speaking	0.15	3658.91	0.90
	Overall Composite		790.57	0.93
9-12	Listening	0.15	1805.04	0.61
	Reading	0.35	696.84	0.75
	Writing	0.35	1379.57	0.91
	Speaking	0.15	4974.53	0.91
	Overall Composite		1008.60	0.94

5.6 Conditional Standard Error of Measurement for Composites

CSEMs for the four ACCESS composites provide test users a benchmark of how free the 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 estimated the CSEMs using a procedure 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' scale scores for a composite consisting of components. The basic premise of this procedure is that the student-level CSEM for a weighted composite scale score can be estimated empirically using the IRT-based CSEMs for each student's component scale score and the weights associated with the components. We used this method to estimate the CSEM for ACCESS composites by treating the ACCESS domains as components.

We use a three-step process to derive the CSEM for ACCESS composites. We conduct the derivation by grade and composite to obtain a unique CSEM for each composite score by grade. Since this procedure relies on empirical student data, which are subject to year-to-year fluctuation, we use all population student data from all previous three ACCESS series in the derivation to obtain more stable estimates than using only data from a single series.

Step 1. Since we calibrated ACCESS domains separately, measurement errors associated with each of the ACCESS domains, as expressed in the conditional errors of measurement, are independent of each other. Therefore, the CSEM for a student with composite score x , SEM_x , can be estimated 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 were not observed in the data.

The figures in this section show graphically the CSEMs for various composite scores by grade level. Figures show the relationship between the students' composite scores on the horizontal axis and conditional measurement errors on the vertical axis. Each point in the figures represents a student in the dataset, expressing both the student's CSEM and that student's scale score for

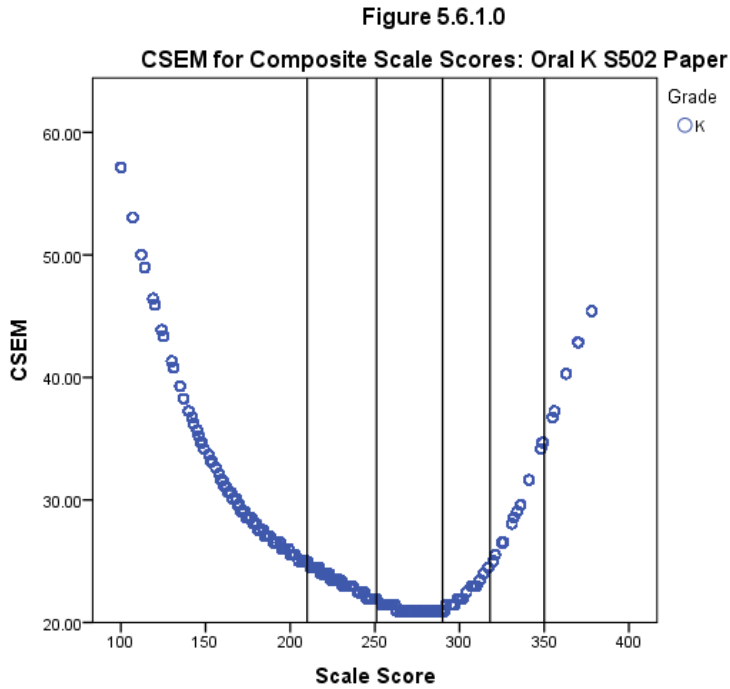
the given composite score. We do not plot values for students who received the lowest possible scores on any ACCESS domains, as it is not possible to compute accurately the conditional measurement errors for these students. For grade-level clusters with multiple grades, we use different colors in the figures to represent students in different grades.

Five vertical lines in the figure indicate the five ACCESS cut points 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 (1–6) for the composites.

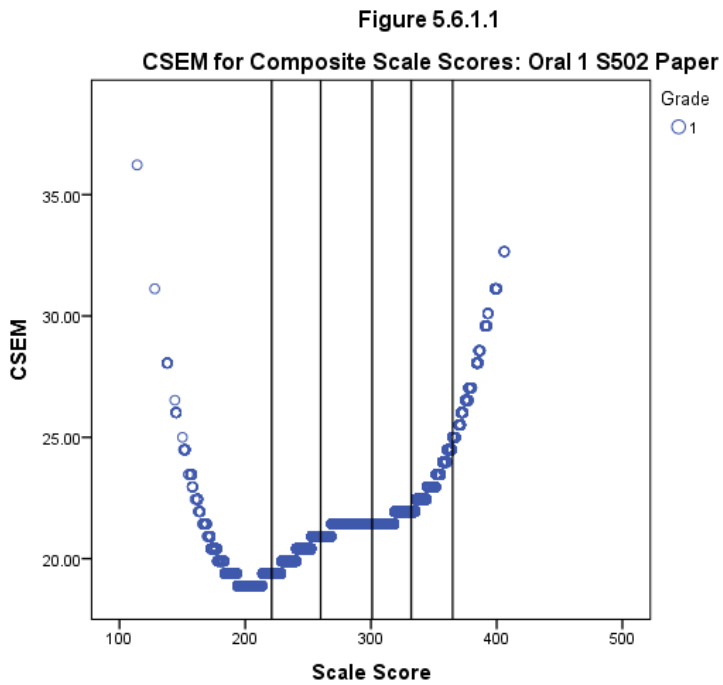
Low CSEM values indicate less measurement error (i.e., greater accuracy) in measurement. In general, these figures show that the CSEMs are lower and fairly constant in the middle of the composite scale score range and higher and more variable for extreme low and high composite scale scores. This is to be expected, as ACCESS test items and scores were scaled using the IRT method, which is known to produce higher CSEMs at the lower and the higher end of the scale score range. In addition, because students exit the EL program when they demonstrate that they are English language proficient, the numbers of students at the extreme high composite scale score range are typically small as compared to those at the middle composite scale score range. Therefore, the measurement errors associated with the scale scores at the extreme high composite scale score range tend to be higher since there are fewer students available in estimating the scores and the measurement errors for these scores.

5.6.1 Oral

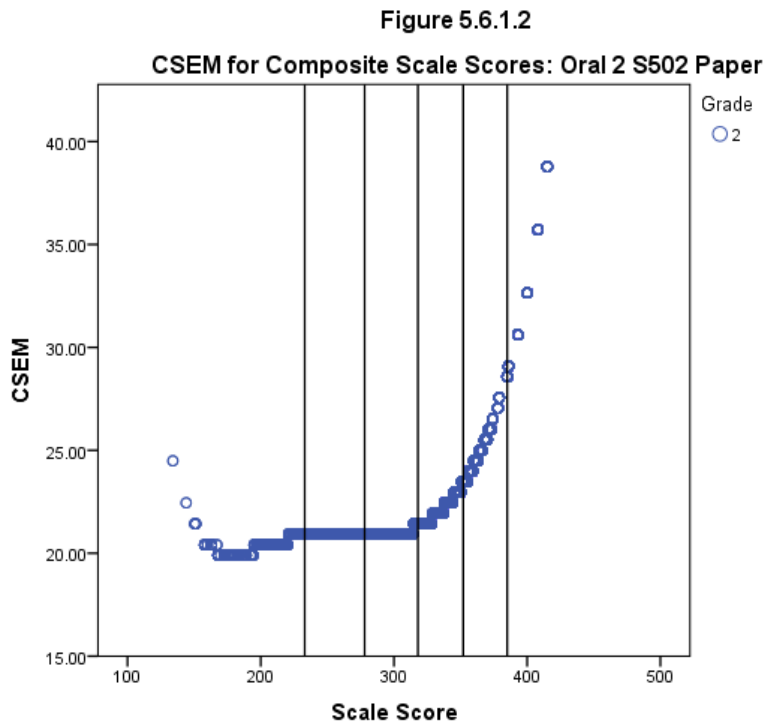
5.6.1.0 Kindergarten



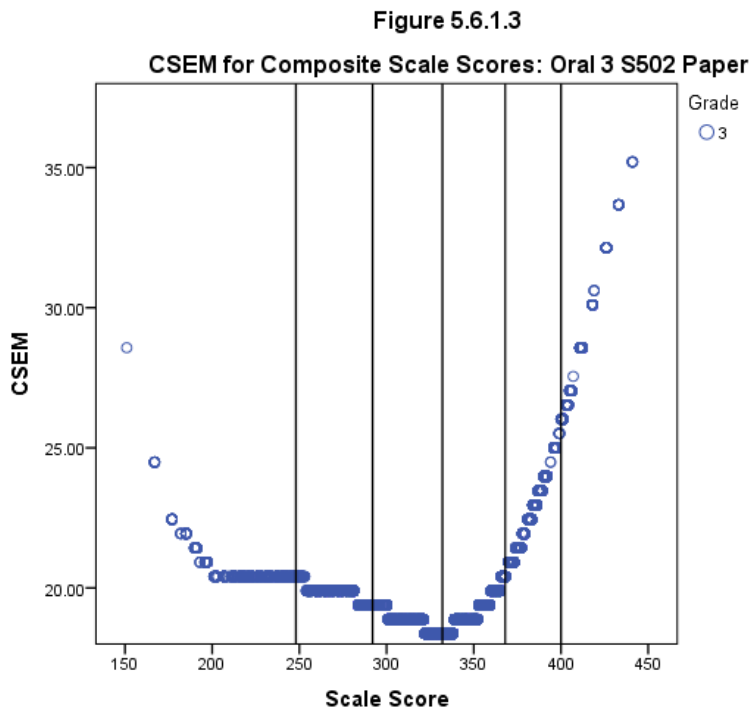
5.6.1.1 Grade 1



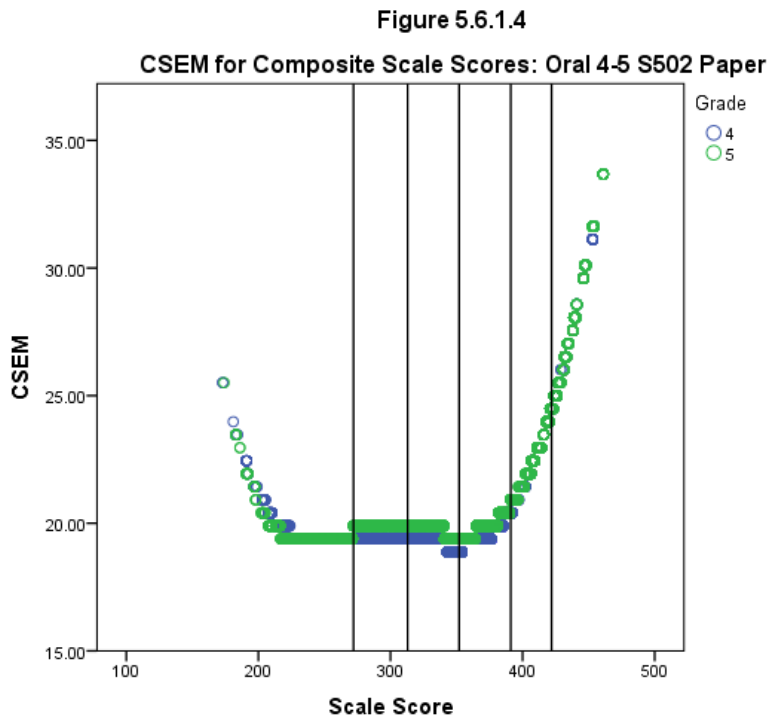
5.6.1.2 Grade 2



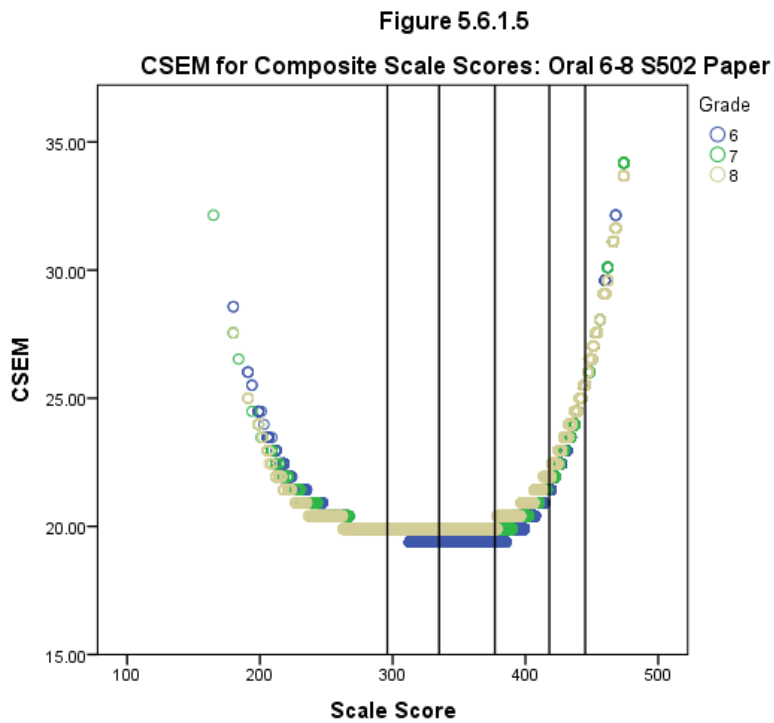
5.6.1.3 Grade 3



5.6.1.4 Grades 4–5

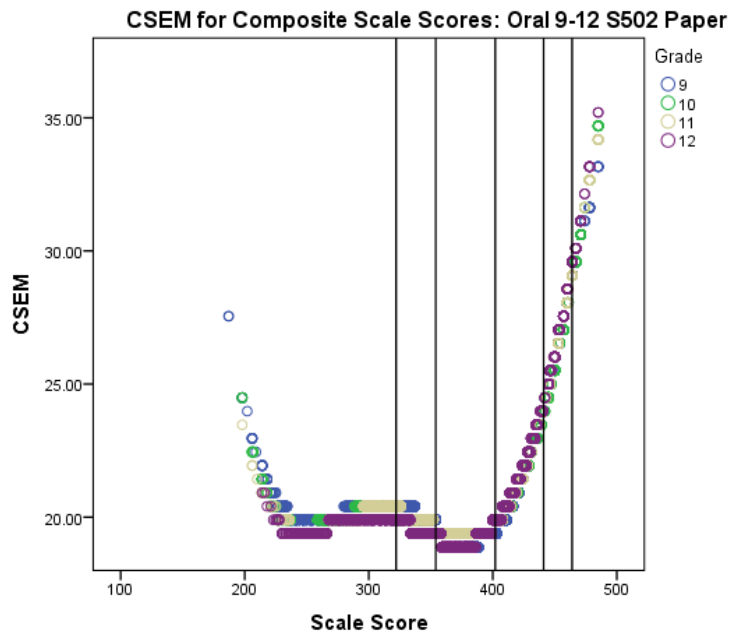


5.6.1.5 Grades 6–8



5.6.1.6 Grades 9–12

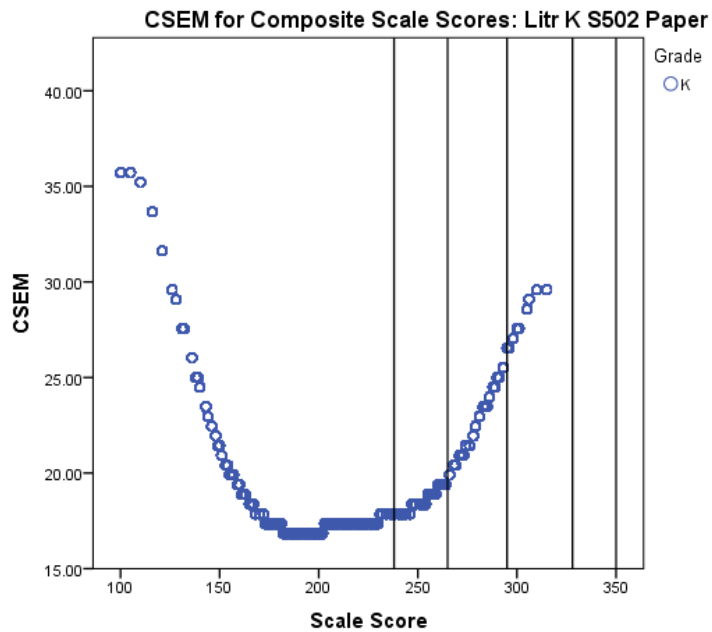
Figure 5.6.1.6



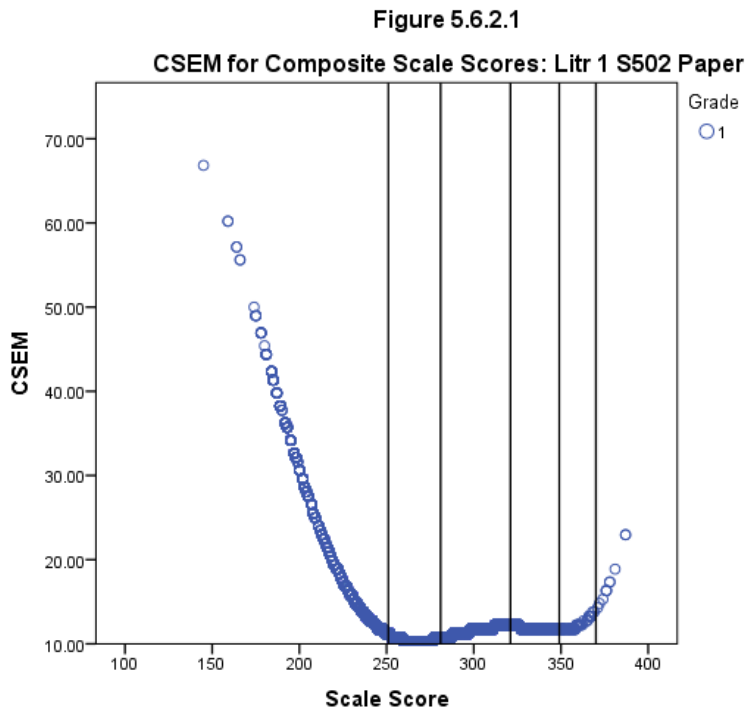
5.6.2 Literacy

5.6.2.0 Kindergarten

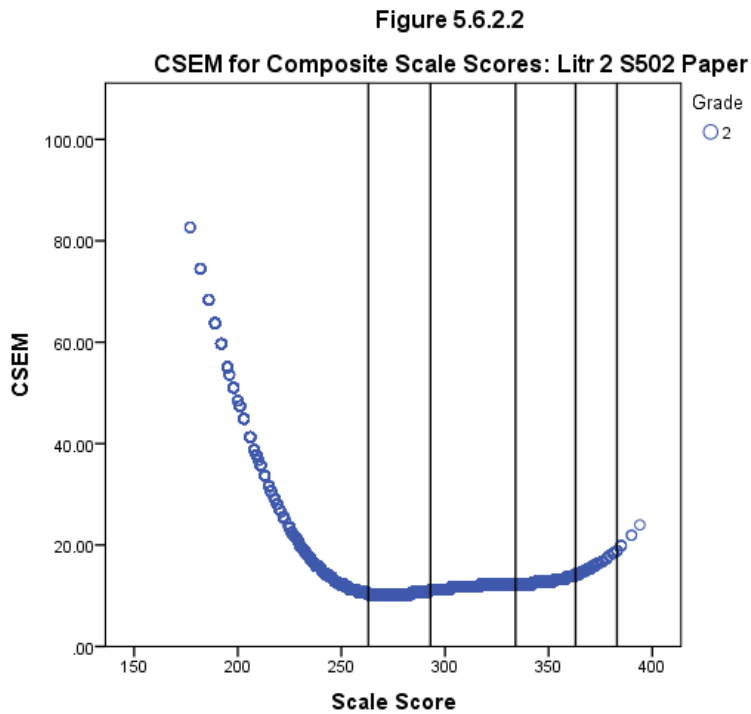
Figure 5.6.2.0



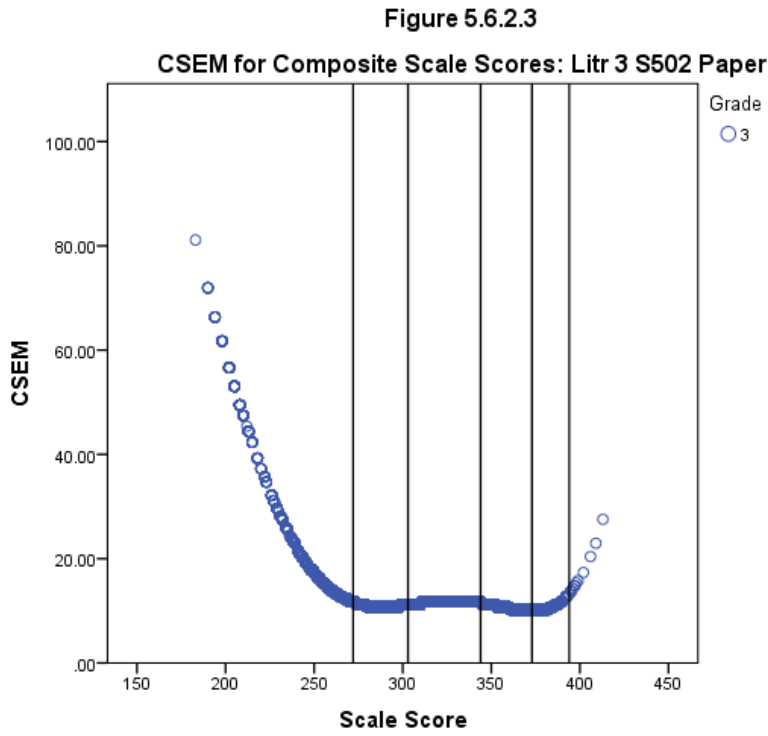
5.6.2.1 Grade 1



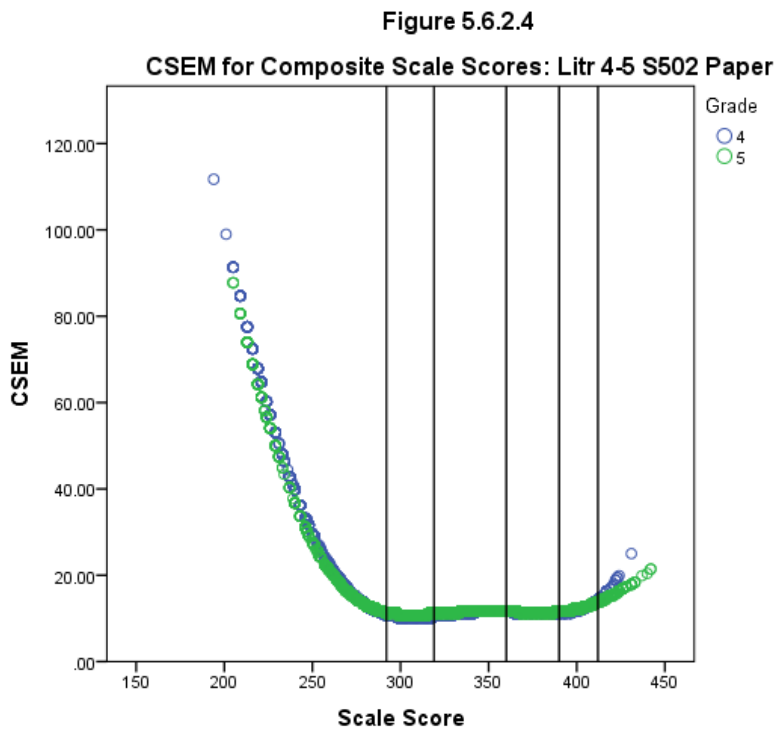
5.6.2.2 Grade 2



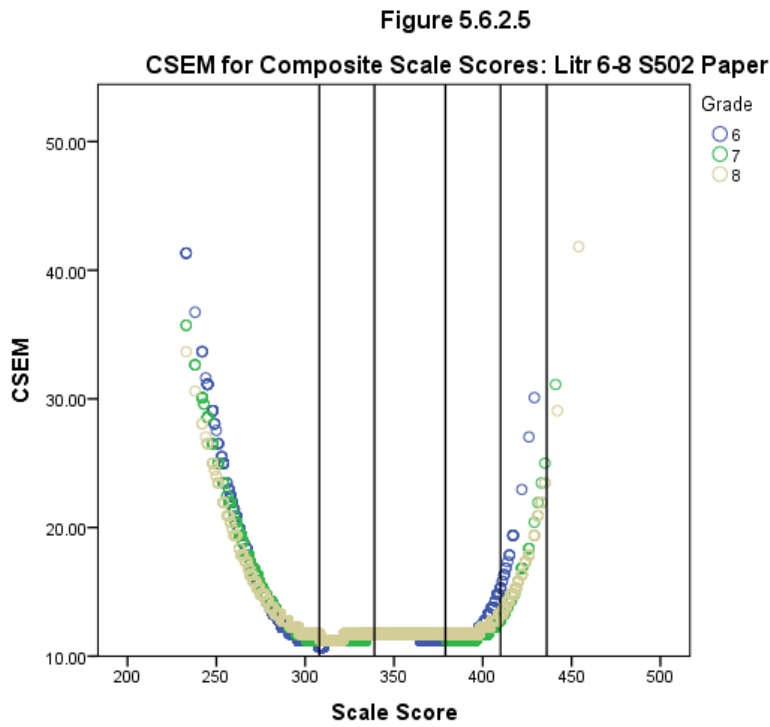
5.6.2.3 Grade 3



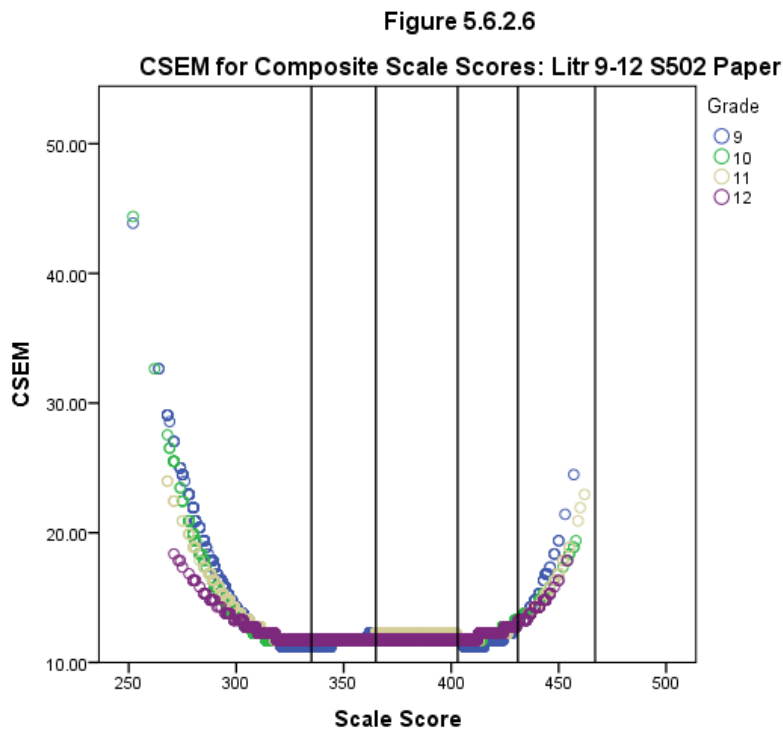
5.6.2.4 Grades 4–5



5.6.2.5 Grades 6–8

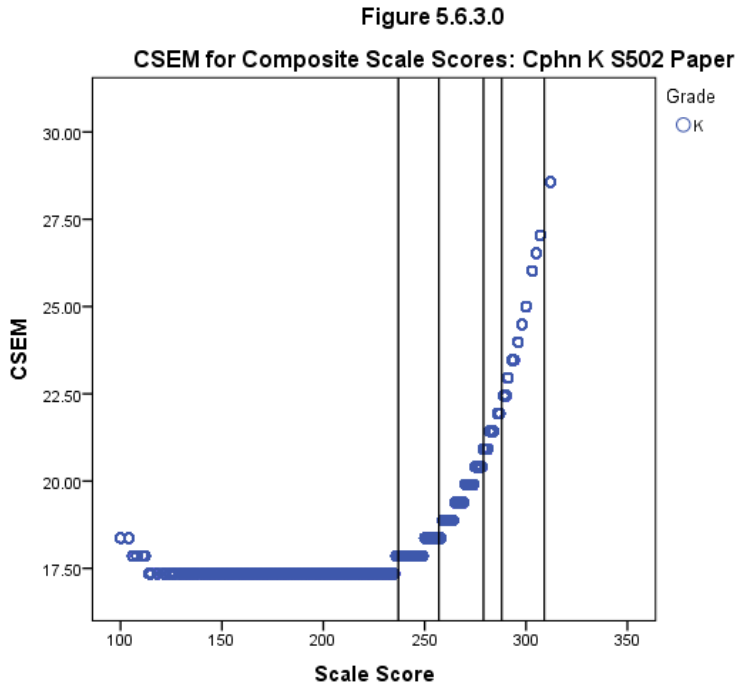


5.6.2.6 Grades 9–12

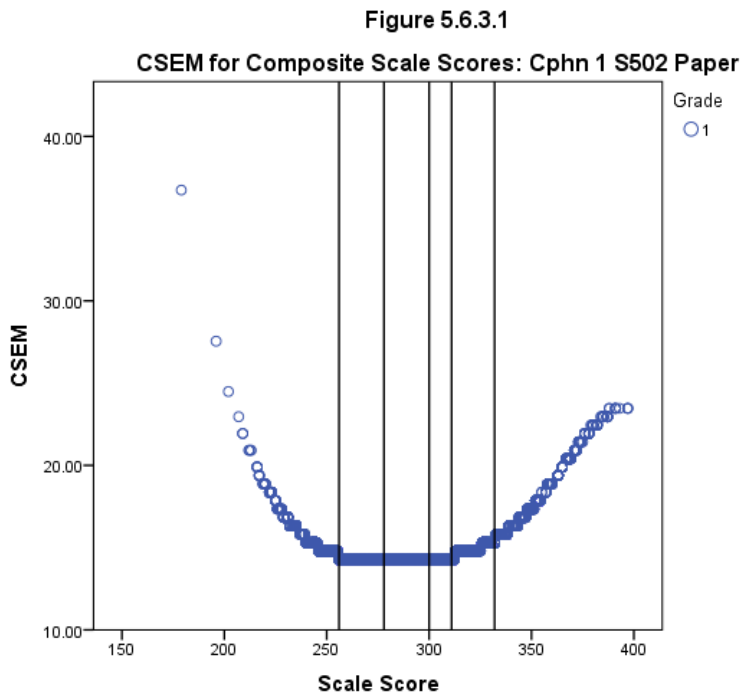


5.6.3 Comprehension

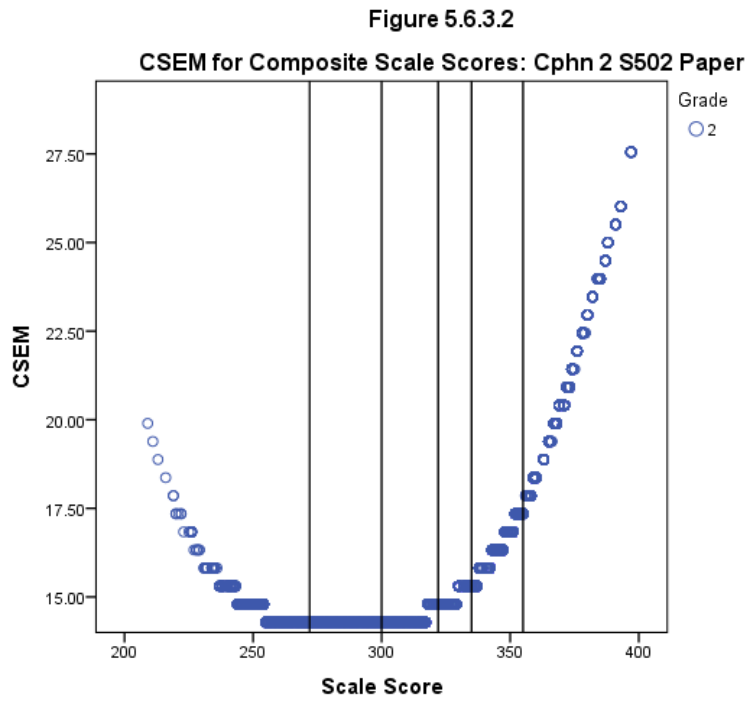
5.6.3.0 Kindergarten



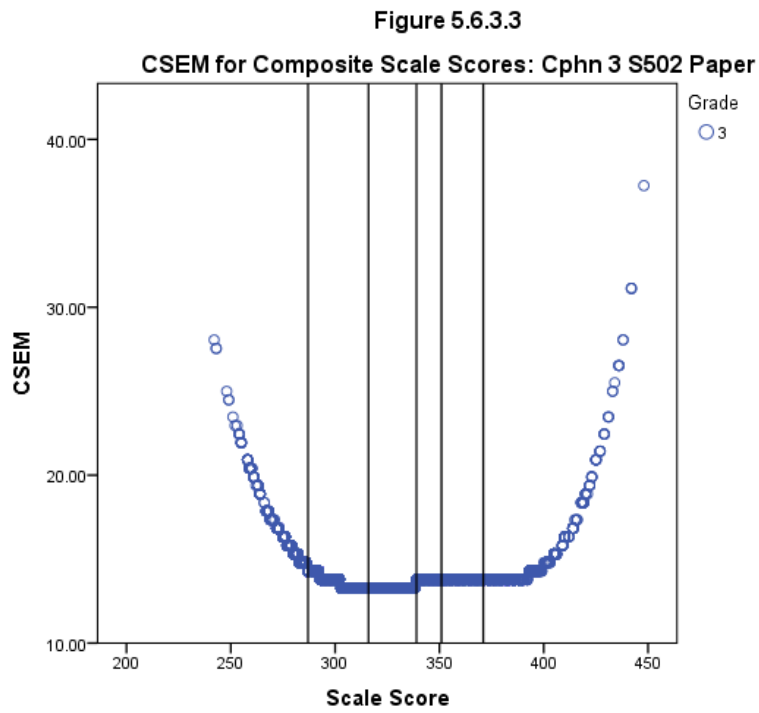
5.6.3.1 Grade 1



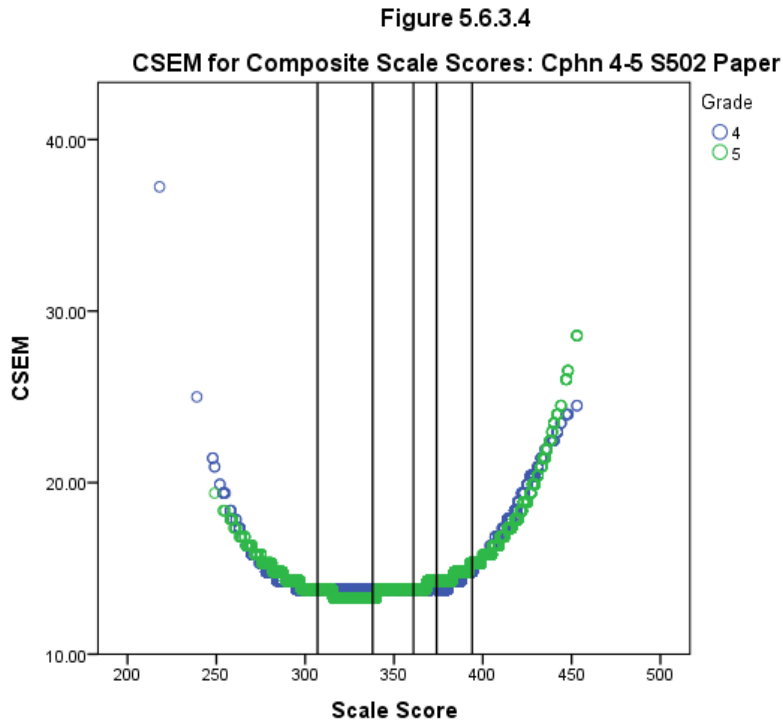
5.6.3.2 Grade 2



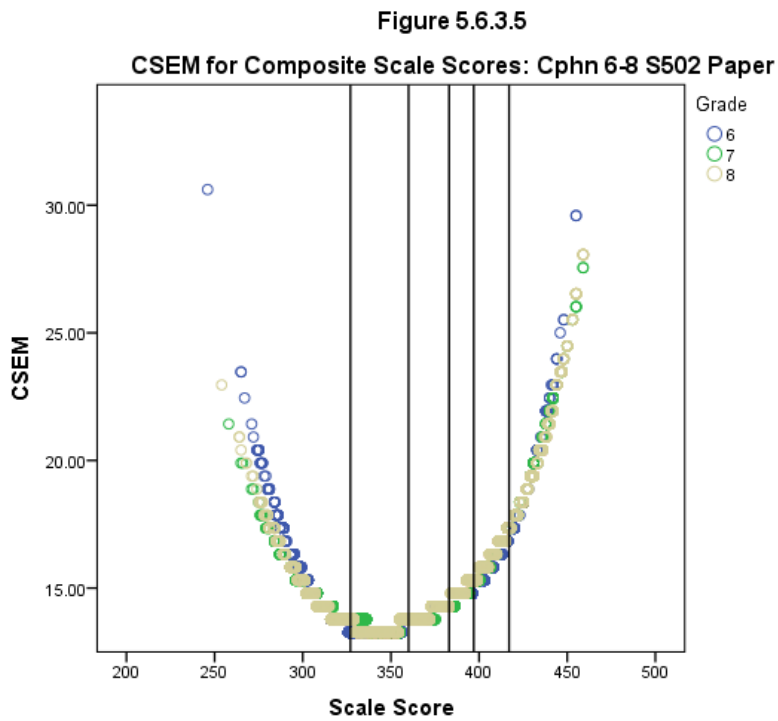
5.6.3.3 Grade 3



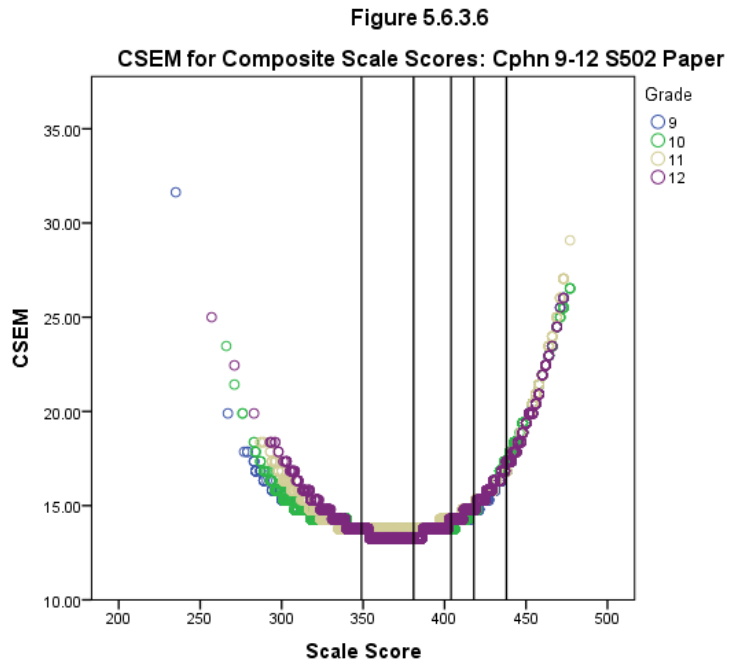
5.6.3.4 Grades 4–5



5.6.3.5 Grades 6–8

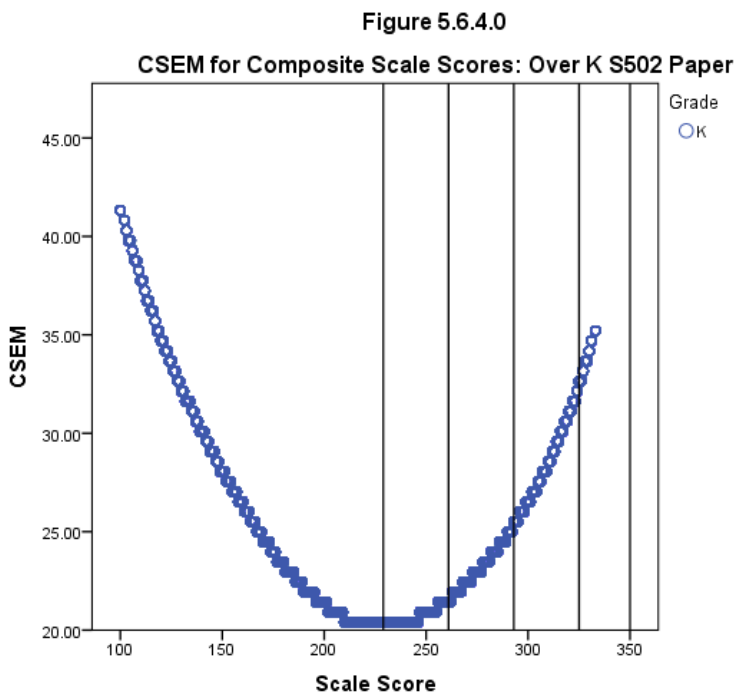


5.6.3.6 Grades 9–12

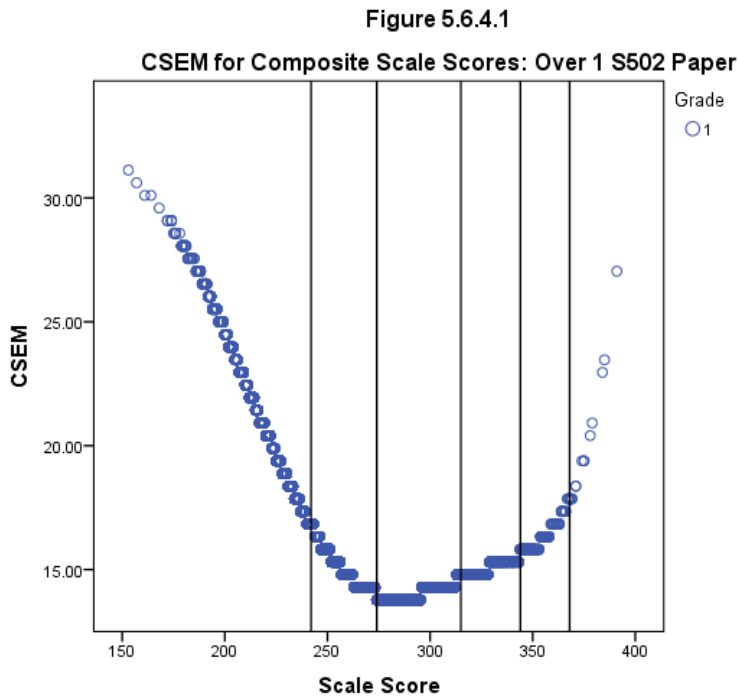


5.6.4 Overall

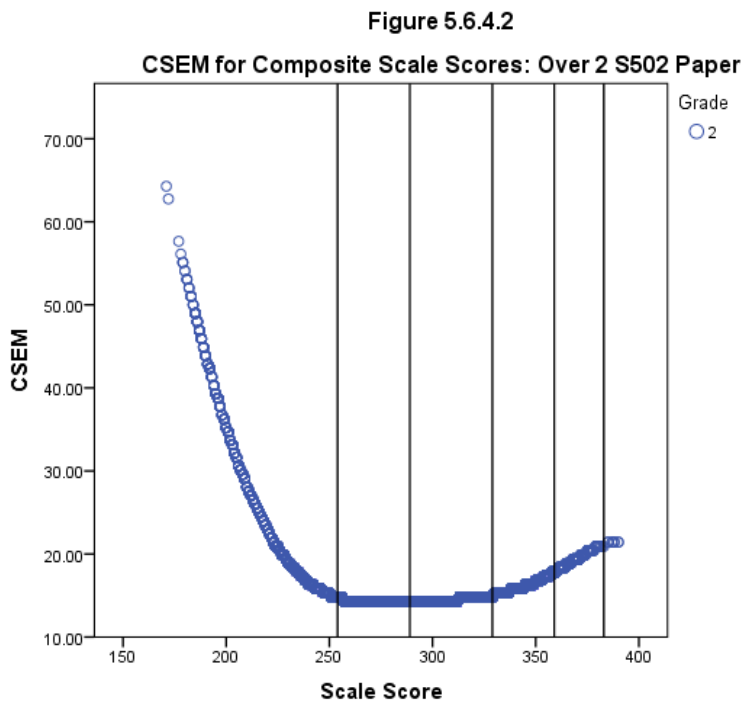
5.6.4.0 Kindergarten



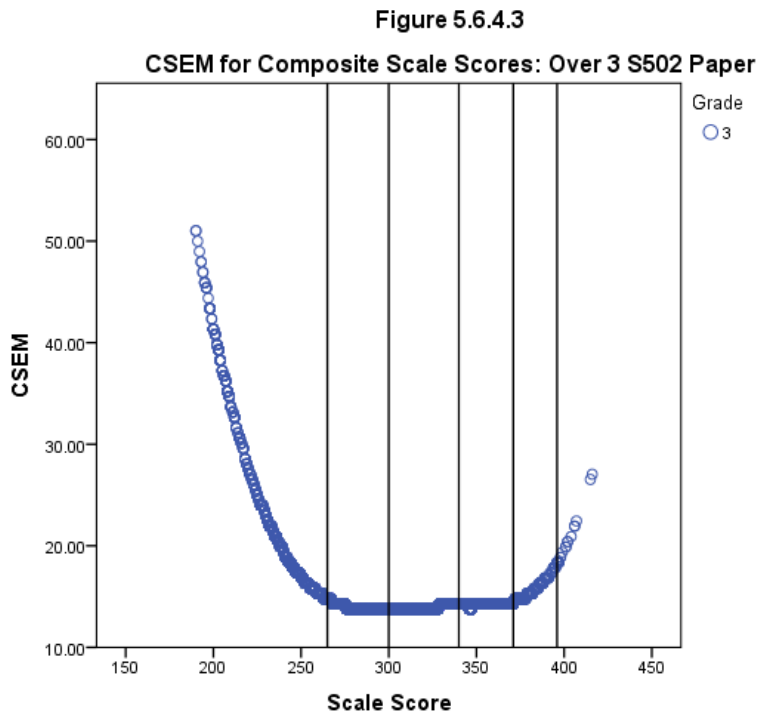
5.6.4.1 Grade 1



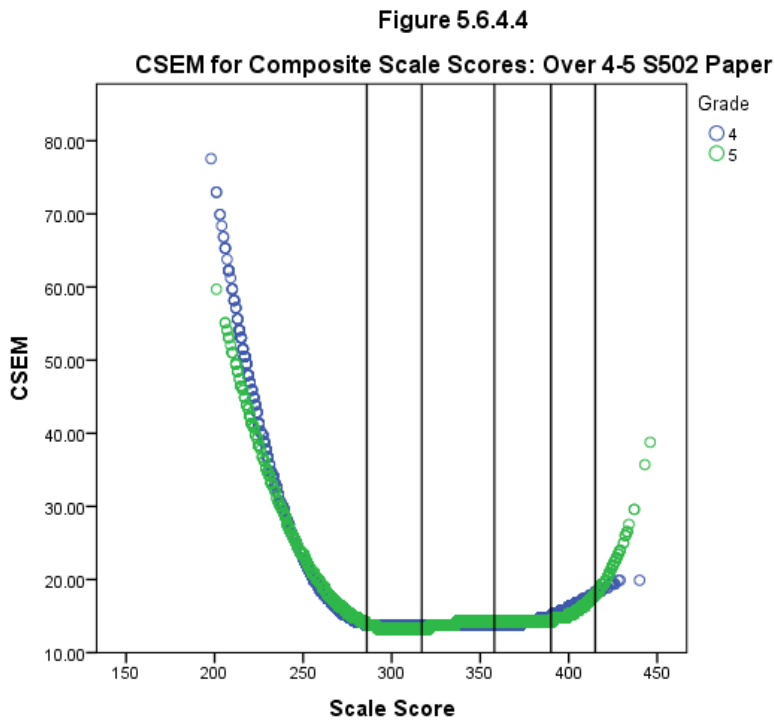
5.6.4.2 Grade 2



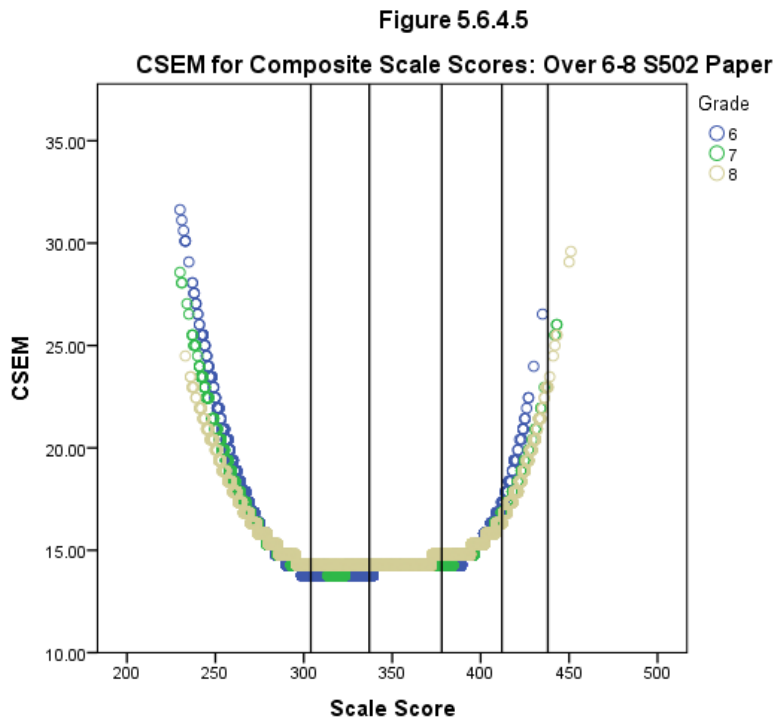
5.6.4.3 Grade 3



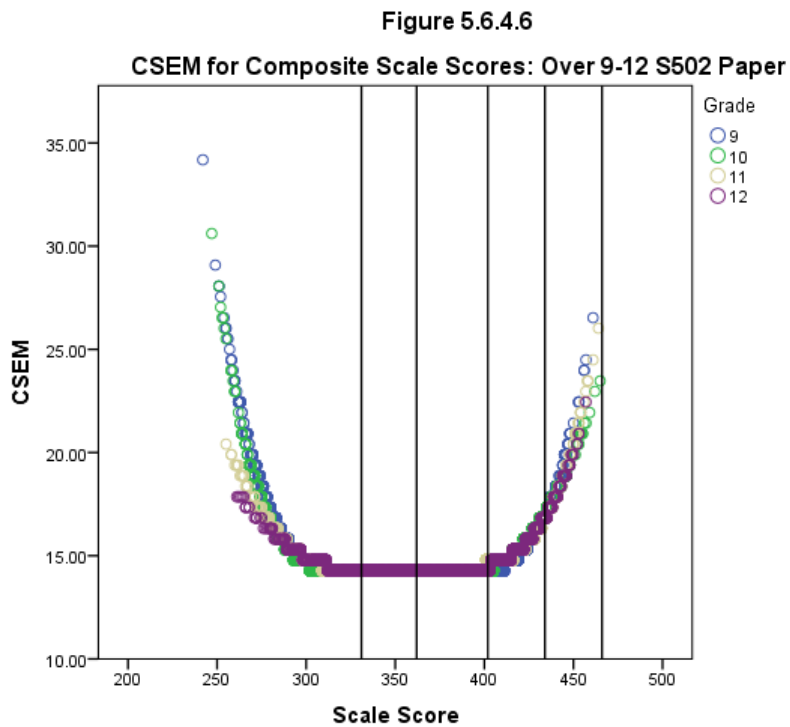
5.6.4.4 Grades 4–5



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 incorporate one or more of the domains and composite scores in making accountability decisions, all WIDA Consortium states use the **Overall composite scale score** as the primary score in 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 outlined by Livingston and Lewis (1995) and Young and Yoon (1998), 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 classification accuracy and consistency of the domains versus composites is that to estimate classification accuracy and consistency of the composites, we first estimate the reliability of the composite scores using a stratified Cronbach's coefficient alpha, as described in Section 5.4.

For each test domain, 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 scale scores at the cut points for each grade. The third reports the marginal classification consistency indices based on the 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 very 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 the range of these indices, by each composite, highlighting the grade with the lowest classification accuracy and consistency indices for each 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 0.614 to 0.721 and overall classification consistency ranged from 0.507 to 0.638 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 0.755 to 0.880 and overall classification consistency ranged from 0.676 to 0.851 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 0.510 to 0.853 and overall classification consistency ranged from 0.400 to 0.811 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 0.776 to 0.880 and overall classification consistency ranged from 0.708 to 0.838 across grades. The lowest overall classification accuracy and consistency values were found for students in Grade 5.

The results suggest that the grade level with the lowest overall classification accuracy and the lowest overall classification consistency tends to vary across these two indices and across the four composites.

The range of the marginal classification accuracy and consistency of composites are summarized and compared across grades by domains. In addition, the grade level with the lowest marginal classification accuracy and consistency of the composites is highlighted so that the test users and policy makers can use caution when making classification decisions in these grades at the specific cuts in the composites.

For the Oral composite, classification accuracy indices at the cut ranged from 0.872 to 0.983 (Table 5.7.1.2) and classification consistency at the cut ranged from 0.819 to 0.977 (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. Additionally, Grade 5 was identified as having the lowest overall classification accuracy and consistency for the Literacy and the Overall 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 5 Oral composite are still in the .80's.

For the Literacy composite, classification accuracy indices at the cut ranged from 0.875 to 0.991 (Table 5.7.2.2) and classification consistency at the cut ranged from 0.831 to 0.991 (Table

5.7.2.3). The lowest marginal classification accuracy and consistency values were found for students in Grade 3 at the PL 3/PL 4 cut for classification accuracy and in Grade 4 at the PL 3/PL 4 cut for classification consistency. Note that Grade 3 was also identified as having the lowest overall classification accuracy and second lowest overall classification 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 and Grade 4 Literacy composite are still in the .80's.

For the Comprehension composite, classification accuracy indices at the cut ranged from 0.815 to 0.984 (Table 5.7.3.2) and classification consistency at the cut ranged from 0.752 to 0.977 (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 low .80's.

For the Overall composite, classification accuracy indices at the cut ranged from 0.883 to 0.988 (Table 5.7.4.2) and classification consistency at the cut ranged from 0.838 to 0.986 (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. Additionally, Grade 3 was also identified as having the lowest marginal classification accuracy and second lowest marginal classification consistency in the Literacy 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.

Grade 3 had the lowest marginal classification accuracy and consistency in two of the four composites (Comprehension and Overall). Grade 3 also had the lowest marginal classification accuracy in the Literacy composite. Grade 4 had the lowest marginal classification 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.

There has been very little guidance for the ideal or expected levels of decision consistency and accuracy needed for educational assessments that use composite scores. From an accountability perspective, the most important information for test users and policy makers to examine is the marginal classification accuracy and consistency. The marginal classification accuracy and consistency indices were at or above 0.800 for all composites except for the Comprehension composite. The lowest marginal classification consistency for the Comprehension composite was 0.752 for Grade 3. Additionally, the marginal classification accuracy and consistency indices were at or above 0.838 for the Overall composite, where the major accountability decisions are being made.

5.7.1 Oral

Table 5.7.1.1

Overall Accuracy and Consistency of Classification Indices: Oral S502 Paper

Grade	Accuracy	Consistency
K	0.721	0.638
1	0.675	0.562
2	0.672	0.561
3	0.673	0.561
4	0.652	0.548
5	0.638	0.529
6	0.634	0.526
7	0.631	0.520
8	0.614	0.507
9	0.650	0.539
10	0.657	0.549
11	0.668	0.560
12	0.704	0.600

Table 5.7.1.2

Classification Accuracy Indices at Cut Score Level: Oral S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.943	0.939	0.945	0.940	0.942
1	0.967	0.920	0.891	0.921	0.975
2	0.978	0.933	0.881	0.911	0.967
3	0.978	0.939	0.878	0.901	0.972
4	0.983	0.959	0.905	0.883	0.919
5	0.977	0.957	0.908	0.872	0.921
6	0.972	0.944	0.898	0.886	0.929
7	0.964	0.938	0.895	0.890	0.938
8	0.957	0.932	0.895	0.890	0.931
9	0.953	0.927	0.894	0.905	0.964
10	0.945	0.917	0.895	0.924	0.968
11	0.945	0.917	0.896	0.931	0.972
12	0.945	0.916	0.898	0.941	N/A

Table 5.7.1.3

Classification Consistency Indices at Cut Score Level: Oral S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.921	0.914	0.921	0.916	0.918
1	0.952	0.887	0.847	0.887	0.967
2	0.968	0.903	0.835	0.868	0.965
3	0.970	0.910	0.832	0.858	0.969
4	0.977	0.940	0.868	0.834	0.906
5	0.969	0.937	0.871	0.819	0.901
6	0.961	0.920	0.858	0.837	0.914
7	0.950	0.910	0.853	0.843	0.922
8	0.939	0.903	0.854	0.843	0.913
9	0.934	0.896	0.853	0.865	0.949
10	0.921	0.882	0.854	0.891	0.959
11	0.922	0.883	0.854	0.899	0.966
12	0.922	0.881	0.857	0.920	N/A

5.7.2 Literacy

Table 5.7.2.1

Overall Accuracy and Consistency of Classification Indices: Litr S502 Paper

Grade	Accuracy	Consistency
K	0.880	0.851
1	0.799	0.719
2	0.812	0.738
3	0.780	0.700
4	0.782	0.707
5	0.755	0.676
6	0.815	0.738
7	0.808	0.729
8	0.781	0.696
9	0.782	0.698
10	0.778	0.689
11	0.775	0.687
12	0.809	0.730

Table 5.7.2.2

Classification Accuracy Indices at Cut Score Level: Litr S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.957	0.961	0.961	N/A	N/A
1	0.911	0.908	0.981	N/A	N/A
2	0.952	0.920	0.949	0.991	N/A
3	0.972	0.933	0.875	N/A	N/A
4	0.981	0.956	0.878	0.967	N/A
5	0.980	0.956	0.895	0.925	N/A
6	0.971	0.933	0.911	N/A	N/A
7	0.968	0.933	0.907	N/A	N/A
8	0.961	0.931	0.909	0.982	N/A
9	0.974	0.939	0.910	0.959	N/A
10	0.968	0.931	0.914	0.966	N/A
11	0.966	0.930	0.914	0.965	N/A
12	0.962	0.920	0.927	N/A	N/A

Table 5.7.2.3

Classification Consistency Indices at Cut Score Level: Litr S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.940	0.943	0.958	N/A	N/A
1	0.874	0.870	0.973	N/A	N/A
2	0.931	0.888	0.927	0.991	N/A
3	0.960	0.904	0.836	N/A	N/A
4	0.973	0.935	0.831	0.964	N/A
5	0.972	0.935	0.853	0.912	N/A
6	0.959	0.905	0.875	N/A	N/A
7	0.956	0.903	0.870	N/A	N/A
8	0.945	0.900	0.871	0.977	N/A
9	0.964	0.913	0.874	0.947	N/A
10	0.955	0.901	0.879	0.953	N/A
11	0.953	0.900	0.879	0.953	N/A
12	0.946	0.887	0.897	N/A	N/A

5.7.3 Comprehension

Table 5.7.3.1

Overall Accuracy and Consistency of Classification Indices: Cphn S502 Paper

Grade	Accuracy	Consistency
K	0.853	0.811
1	0.548	0.435
2	0.598	0.485
3	0.510	0.400
4	0.586	0.472
5	0.559	0.449
6	0.598	0.486
7	0.578	0.467
8	0.563	0.454
9	0.585	0.474
10	0.583	0.475
11	0.585	0.477
12	0.620	0.508

Table 5.7.3.2

Classification Accuracy Indices at Cut Score Level: Cphn S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.959	0.965	0.966	0.967	0.984
1	0.927	0.854	0.863	0.904	0.959
2	0.955	0.889	0.881	0.901	0.947
3	0.974	0.928	0.827	0.815	0.915
4	0.977	0.939	0.862	0.862	0.918
5	0.970	0.933	0.870	0.855	0.891
6	0.963	0.897	0.861	0.899	0.956
7	0.952	0.895	0.864	0.893	0.944
8	0.941	0.892	0.867	0.881	0.942
9	0.954	0.899	0.868	0.889	0.948
10	0.944	0.889	0.873	0.901	0.944
11	0.943	0.888	0.877	0.900	0.941
12	0.934	0.880	0.893	0.926	0.970

Table 5.7.3.3

Classification Consistency Indices at Cut Score Level: Cphn S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.941	0.950	0.953	0.955	0.977
1	0.899	0.796	0.810	0.861	0.941
2	0.934	0.845	0.833	0.860	0.924
3	0.967	0.888	0.766	0.752	0.872
4	0.971	0.906	0.811	0.809	0.881
5	0.960	0.900	0.822	0.802	0.845
6	0.949	0.852	0.811	0.854	0.939
7	0.934	0.850	0.815	0.847	0.920
8	0.918	0.847	0.817	0.836	0.913
9	0.937	0.855	0.819	0.845	0.922
10	0.922	0.843	0.825	0.859	0.920
11	0.920	0.842	0.831	0.857	0.916
12	0.906	0.833	0.851	0.894	0.957

5.7.4 Overall

Table 5.7.4.1

Overall Accuracy and Consistency of Classification Indices: Over S502 Paper

Grade	Accuracy	Consistency
K	0.880	0.838
1	0.833	0.764
2	0.833	0.770
3	0.802	0.731
4	0.813	0.754
5	0.776	0.708
6	0.832	0.771
7	0.821	0.755
8	0.806	0.733
9	0.805	0.733
10	0.806	0.731
11	0.808	0.737
12	0.843	0.778

Table 5.7.4.2

Classification Accuracy Indices at Cut Score Level: Over S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.956	0.963	0.975	0.986	N/A
1	0.953	0.918	0.963	N/A	N/A
2	0.974	0.940	0.937	0.982	N/A
3	0.983	0.954	0.883	0.982	N/A
4	0.988	0.972	0.920	0.933	N/A
5	0.986	0.971	0.927	0.892	N/A
6	0.981	0.958	0.920	0.974	N/A
7	0.978	0.956	0.920	0.967	N/A
8	0.973	0.954	0.921	0.960	N/A
9	0.977	0.955	0.926	0.948	N/A
10	0.971	0.949	0.928	0.958	N/A
11	0.972	0.948	0.929	0.960	N/A
12	0.970	0.942	0.932	N/A	N/A

Table 5.7.4.3

Classification Consistency Indices at Cut Score Level: Over S502 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.939	0.948	0.965	0.986	N/A
1	0.933	0.885	0.946	N/A	N/A
2	0.963	0.915	0.910	0.981	N/A
3	0.976	0.933	0.838	0.982	N/A
4	0.984	0.958	0.889	0.923	N/A
5	0.981	0.958	0.898	0.871	N/A
6	0.974	0.939	0.887	0.970	N/A
7	0.969	0.937	0.887	0.962	N/A
8	0.962	0.933	0.888	0.948	N/A
9	0.968	0.936	0.896	0.933	N/A
10	0.960	0.927	0.899	0.945	N/A
11	0.960	0.926	0.900	0.951	N/A
12	0.958	0.917	0.904	N/A	N/A

6 Quality Control

6.1 Content Development Quality Control

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 and 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 and 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 and 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 and 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 and 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 and Sensitivity review of Grade 1 and Grades 2–3 is often combined). Although Bias and 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

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 and 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 tryout 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 and is 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.

The training course within the WIDA Secure Portal (<https://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 upcoming 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(es) 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 quality control, please refer to Section 4.2.

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.³ 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 Report to check the District Frequency Report. Finally, panelists check the State Frequency Report against verified District Frequency Reports from the primary district along with District Frequency Reports from the secondary district.

6.5 Data Forensic Quality Control

Paper Booklet Issue

During a routine Web Patrol by the vendor Caveon, ACCESS for ELLs materials from the 2019–2020 administration were discovered on eBay. WIDA contacted the eBay seller to have the seller take down the materials and return them to WIDA. The materials consisted of test packets for Grades 1, 2, and 3. All materials were taken down immediately after seller notification, and all packets were sealed. Accordingly, WIDA assumes that no materials were exposed. The state

³ Due to the COVID-19 pandemic, the 2021 Score Reporting quality control was conducted online, with only WIDA and DRC employees participating.

where these materials were originally sent is investigating this incident to determine how custody of these materials was lost and will write a report. As yet, WIDA has not received a copy of that report.

Caveon Data Forensic Analysis Results

WIDA hired Caveon to perform data forensic analysis during the 2020–2021 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 2020 through Spring 2021 Online multistage adaptive test administrations, Listening and Reading domains
- December 2020 through Spring 2021 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 may 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 IRT 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.
- 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.
- Center for Applied Linguistics. (2016a). *ACCESS for ELLs® Series 400 Listening and Reading scale maintenance: Technical brief*. Washington, DC: Author.
- Center for Applied Linguistics. (2016b). *Annual technical report for ACCESS for ELLs® English Language Proficiency Test, Series 303, 2014–2015 administration* [WIDA Consortium Annual Technical Report No. 11]. 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. (2018). *Annual technical report for ACCESS for ELLs Paper Series 401*. Washington, DC: Author.
- Center for Applied Linguistics. (2019). *Annual technical report for ACCESS for ELLs® English Language Proficiency Test, Series 402 Paper, 2017–2018 administration* (WIDA Consortium Annual Technical Report No. 14B). 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.
- 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: 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*. 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., Chapman, & M., Kondo, A., & Wilmes, C. (2020). Examining the assessment literacy Required for interpreting score reports: A focus on educators of K-12 English learners, *Language Testing, Vol. 37*(1) 54-75.
- 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. (2002). 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>
- Livingston, S. A. (2018). *Reliability—basic concepts* [ETS Research Memorandum No. RM-18-01]. Princeton, NJ: ETS.
- 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., Kenyon, D., Gibson, S., and Evans, E. 2009. *Development and Field Test of Kindergarten ACCESS for ELLs®*. Madison, WI: WIDA Consortium.
- MacGregor, D., Yen, S., & Yu, X. (2021). Using multistage testing to enhance measurement of an English language proficiency test. *Language Assessment Quarterly*. doi: 10.1080/15434303.2021.1988953
- 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/>
- Muraki, E. (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement*, 17, 4, 351–363.
- 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
- 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.
- 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.

- 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=
- Waller, N. G. (n.d.). *EZDIF: A computer program for detecting uniform and nonuniform differential item functioning with the Mantel-Haenszel and logistic regression procedures* [Computer software]. Davis, CA: University of California – Davis.
- Waller, N. G. (1998). EZDIF: Detection of Uniform and Nonuniform Differential Item Functioning with the Mantel-Haenszel and Logistic Regression Procedures. *Applied Psychological Measurement*, 22, 391.
- WIDA Consortium. (2007). *English Language Proficiency Standards and resource guide, 2007 edition, Pre-Kindergarten 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. (2021). *Individual student report 2021*. Retrieved from <https://wida.wisc.edu/sites/default/files/resource/ACCESS-Sample-Individual-Score-Report-English.pdf>
- 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., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 30, 233–251.

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.
Yage (Leah) Guo, Ph.D.
Michele Kawood, M.S.Ed.
Justin Kelly, Ph.D.
Dorry M. Kenyon, Ph.D.
Jung-Jung Lee, M.Sc.
Isabella De Leon, B.S.
Erin Shaw-Meadow, M.Sc.
Samantha Musser, M.A.
Rachel Myers, M.S.
Yoon Ah Song, Ph.D.
Alice Tsai, M.S.
Frank Wucinski, M.A.
Shu Jing Yen, Ph.D.
Xin Yu, M.A.

From WIDA:

Anna Rhoad-Drogalis, MS.
Kyoungwon Bishop, Ph.D.
Sakine Göçer Sahin, Ph.D.