WIDA™

UNIVERSITY OF WISCONSIN–MADISON

Annual Technical Report for ACCESS for ELLs Paper English Language Proficiency Test Series 601, 2022–2023 Administration

Annual Technical Report No. 19B

Prepared by:

Center for Applied Linguistics

Language Assessment Division
Psychometrics and Quantitative Research Team

Nov 2023

The WIDA ACCESS for ELLs Technical Advisory Committee

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

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

Executive Summary

This is the 19th annual technical report on the ACCESS for ELLs English Language Proficiency Test and the seventh report on the paper-and-pencil version of the ACCESS for ELLs assessment (ACCESS Paper) 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 2022–2023 ACCESS assessment.

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

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

ACCESS for ELLs Paper Series 601 was administered in school year 2022–2023 in 36 states, the Bureau of Indian Education; the District of Columbia; the Department of Defense Education Activity; the Northern Mariana Islands; and the Virgin Islands for a total of 41 state entities (henceforth "states").

The Series 601 ACCESS Paper data set used in this report included the results of 508,753 students as of September 2022. The final number of students who participated in the Series 503 ACCESS Paper tests is 518,545. The grade with the largest number of students in this report was Kindergarten, with 257,011 students, while Grade 12 had the fewest, with 8,112 students. Of the participating WIDA states, Florida had the most students (291,285), while the Northern Mariana Islands (MP) had the fewest, with 53 students.

Based on a comparison with prior years' numbers of participating students, 4% more students participated in ACCESS Series 601 testing than the ACCESS Series 503 testing.

ACCESS for ELLs Series 601 was offered in two administrative formats, an online format (Grades 1–12) and a paper-and-pencil format (Kindergarten–Grade 12). The current report (WIDA ACCESS Technical Report 19B) provides technical information pertaining to ACCESS for ELLs Series 601 Paper. A second report (WIDA ACCESS Technical Report 19A) provides technical information for the ACCESS for ELLs Series 601 Online assessment.

Part 1:

Purpose, Design, Implementation

Contents

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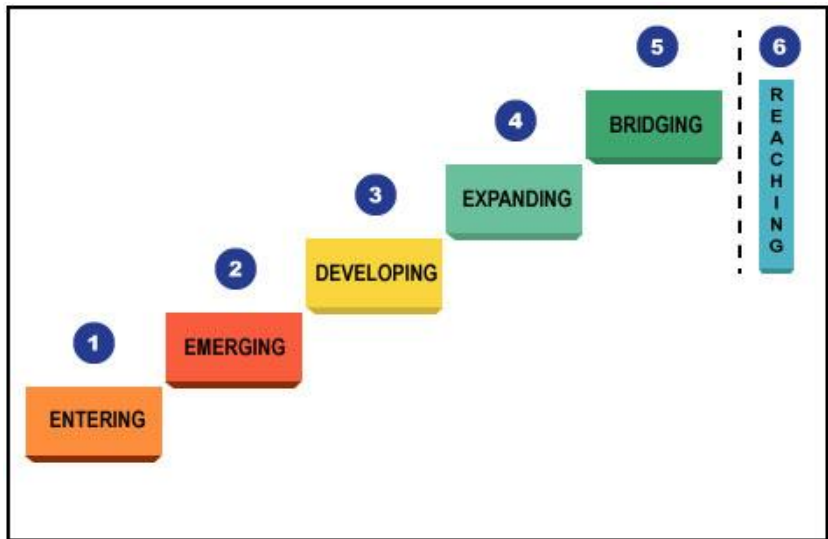
Figure 1. The language proficiency levels of the WIDA ELD Standards

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 refer 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, it 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 differed 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		3
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.

ACCESS Paper 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, https://wida.wisc.edu/assess/access/preparing-students/practice

When the task models for ACCESS Paper were first developed, CAL and WIDA addressed issues of fairness by ensuring that principles of Universal Design of Assessments (UDA) (National Center on Educational Outcomes, 2021) were adhered to in this design phase. 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, in addition to the process by which bias and sensitivity review panels evaluate items and tasks to ensure accessibility and fairness for all students, are described in Section 2.3.1 below.

Note that this section applies to ACCESS Paper Grades 1–12. For details 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. Students select their answer by filling in the oval below the response option with

a pencil in the test booklet. They can change their answer by erasing the filled-in oval. A screenshot of a sample Listening item is provided in Figure 2.
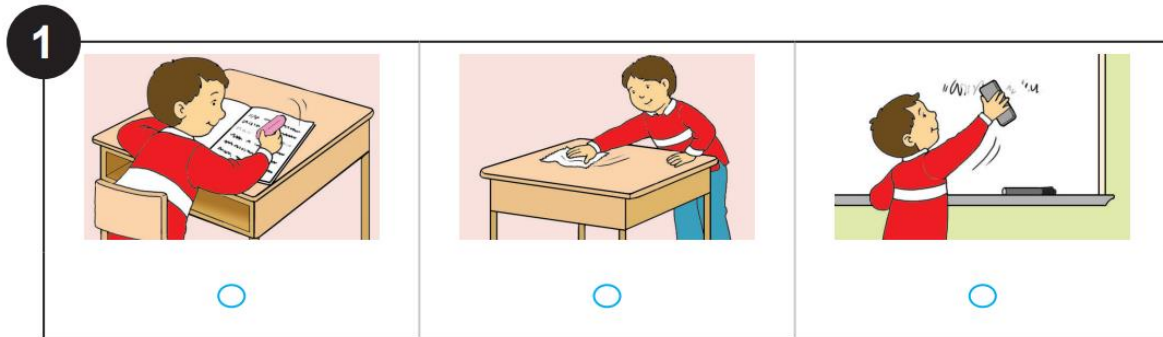
Figure 2. Item layout for the ACCESS Paper Listening Subtest

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 and fourth 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 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 and question stems for Reading items are 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. A screenshot of a sample Reading item is provided in Figure 3.

13

Mr. Sosa's students asked all the students in the school, "Do you play on a school sports team?" The class recorded how many students answered yes.

What information did the students collect to answer their question?

○ The total number of students who play sports

○ The total number of students of each gender

○ The total number of students in each grade

Figure 3. Item layout for the ACCESS Paper Reading Subtest

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-level clusters and tiers, the Writing test is group administered by a live Test Administrator. The Test Administrator reads instructions aloud from the Test Administrator Script and monitors student progress through the test. For all grade-level clusters and tiers, students hand-write their answers in the same test booklet containing the Listening and Reading tests. Figure 4 provides an example of the task layout for the Writing test. Figure 5 provides an example of the accompanying script.

Figure 4. Example task in test booklet for the ACCESS Paper Writing Subtest

Figure 5. Example script for the ACCESS Paper Writing Subtest

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 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 details 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			

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

Figure 6 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 6. Distribution of items by Standard for each tier of the Listening test
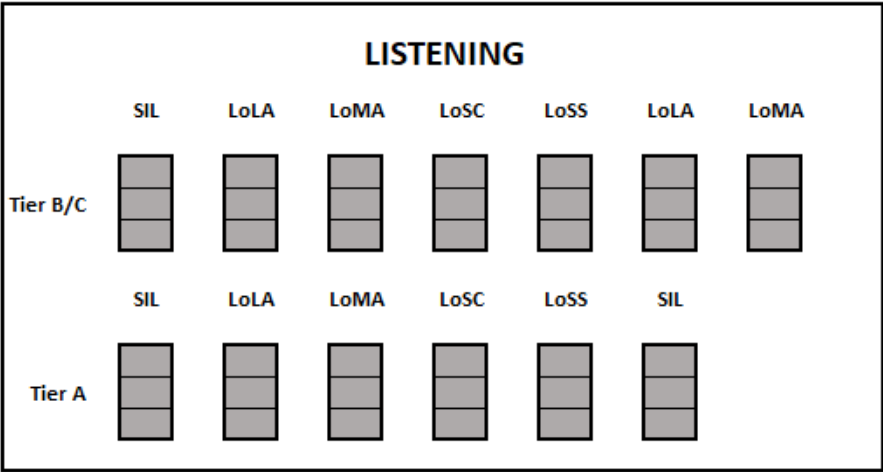
Note that the test design is slightly different between Tier A and Tier B/C. Tier B/C students, who potentially may be reclassified by the assessment, take a slightly longer test and take two folders each assessing the Language of Language Arts and the Language of Mathematics Standards. Tier A students 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 7 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 7. Distribution of items by Standard for each tier of the Reading test
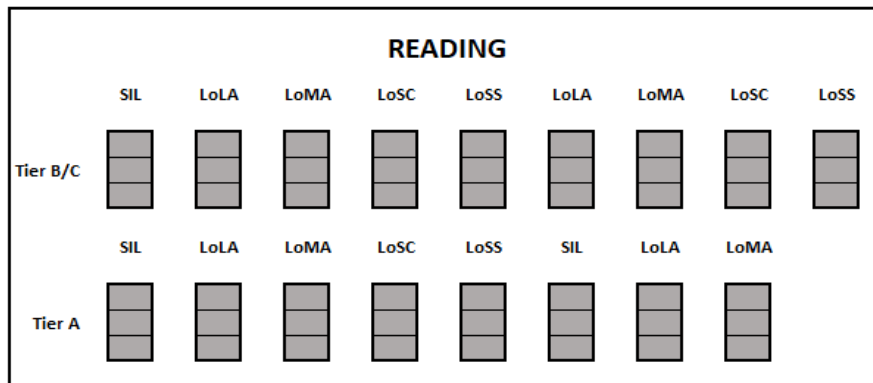
As with Listening, the Reading Tier A test is shorter and focuses on Standards deemed more accessible for lower-proficiency students.

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

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. Figure 8 and Figure 9 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 8. Distribution of tasks by targeted proficiency level for each tier of the Grade 1 Writing test
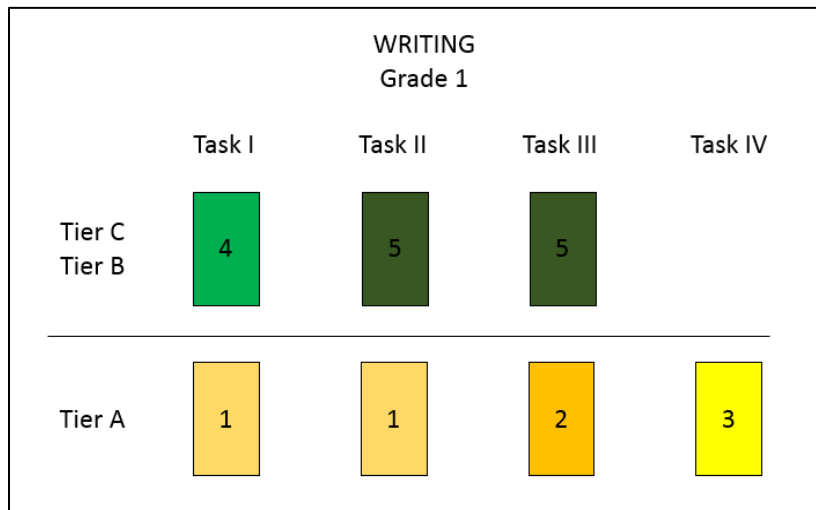
Figure 9. Distribution of tasks by targeted proficiency level for each tier of the Grades 2–12 Writing test
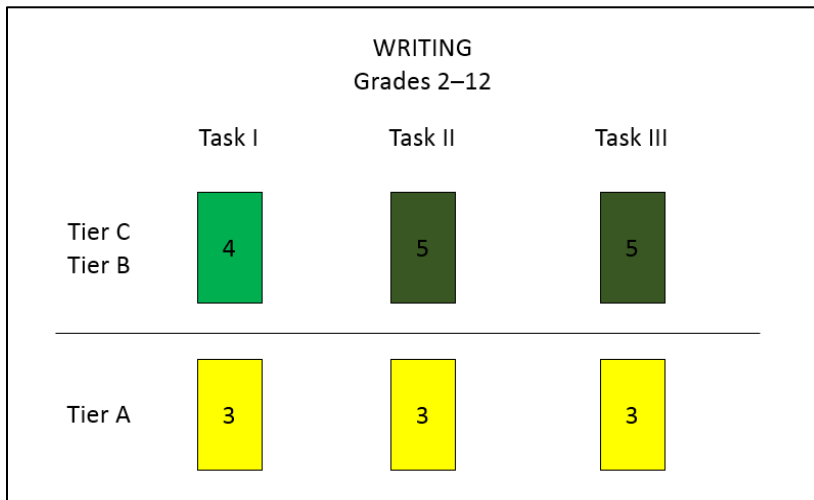
2.2.4 Speaking

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			

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

Figure 10 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 10. Distribution of tasks for each tier of the Speaking test
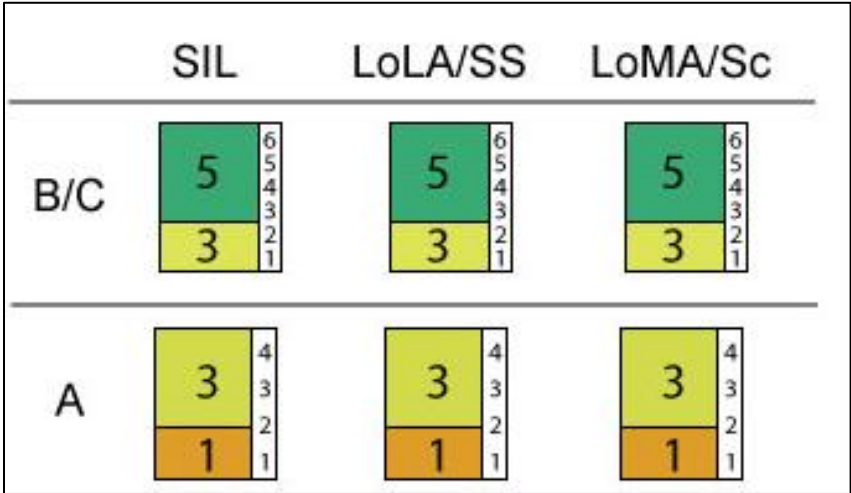
As shown in Figure 10, the Speaking test includes two tiers. Tier A includes tasks that target elicitation of language at PLs 1 and 3. Tier B/C includes tasks that target elicitation of language at PLs 3 and 5.

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-level 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 and Task Development

ACCESS Paper Series 503 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 503 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 and tasks, trained item writers worked from item/task specifications to draft items and tasks 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 and tasks 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 and tasks 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 and tasks 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 and tasks. 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 employs several criteria when recruiting educators to perform these tasks. The criteria used to recruit educators to conduct Content Reviews differ somewhat from the criteria used to recruit educators to conduct Bias and Sensitivity Reviews. Educators conduct Content Reviews by grade-level cluster (G1, G2-3, G4-5, G6-8, and G9-12). The educators who are recruited to review a particular grade-level cluster's content (4 reviewers per grade-level cluster) have experience teaching English language learners and are either currently teaching that grade-level cluster or have extensive prior experience teaching students who are in that grade-level cluster. Additionally, educators serving on each panel represent different content areas. WIDA TD staff seek to ensure that each panel includes at least one educator who has teaching experience in each of the following content areas: ELA, Science, Math, Social Studies, and Special Education. Additionally, during the recruitment process, WIDA TD staff seek to ensure diversity and balance across a) consortium states, b) school locale (rural/suburban/urban), and c) years of teaching experience. The CAL TD team and WIDA TD staff first train the Content Review Panel on the procedures and scope of the review. The panelists are introduced to the test layout, instructed on the logistics of the review, and trained to use the review checklist. The panel members then individually review each item and task, followed by a collective discussion of each item and task to determine (1) whether the content is accessible and relevant to students

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

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

The CAL Language Testing Specialists compile all Content Review and Bias and Sensitivity Review feedback from educators and from WIDA TD staff, and then work to implement the feedback, with the CAL Test Development Manager sign-off as a final step. The CAL Test Production team then revises the graphics. The input and feedback from educators at various stages in the item/task development process served as evidence that each item or task 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 the CAL TD team to evaluate whether each Speaking and Writing task would effectively elicit language at its 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 subjected 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/task 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 and tasks, several principles for UDA were built into the item development checklists and were specifically reviewed 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:

- Inclusive assessment population
- Precisely defined constructs
- Accessible, nonbiased items and tasks
- Amenability to accommodations

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

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

Note that this section applies to ACCESS Paper Grades 1–12. For details 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-level clusters, the Tier A Listening and Reading forms are static forms, which were constructed prior to the launch of ACCESS Online.

In all grade-level clusters, the operational Tier B/C forms in Listening and Reading forms for Series 601 are identical to those administered in Series 502. 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 details 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 Test Administrator reads the story aloud. 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 as "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. The Test Administrator dichotomously scores the tasks. The Test Administrator Script indicates the level required for a task to meet expectations and to be scored as "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										
	Narrative Storyline					Expository Storyline					

WIDA Standard	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. 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 a state's testing window, there are various roles that educators hold to ensure all tasks are carried out for successful test administration. These roles include Test Coordinators at the district and school level 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 Test Administrator Manual and the District and School Test Coordinator Manual contain more information related to responsibilities and required training for the various roles. These manuals can be found on the WIDA Secure Portal (https://portal.wida.us/).

The training course within the WIDA Secure Portal (https://portal.wida.us/) 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 a state's testing window. Training courses include test preparation and administration tutorials and an 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.

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 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.

3.2.2.1 Reading Test Materials

Reading test materials include the following items:

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

3.2.2.2 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 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 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 503 Paper is conducted through online training modules hosted within the WIDA Secure Portal (https://portal.wida.us/). 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 ([P:\Projects\Psychometrics\ACCESS 2.0\Operational\S603 2025\K ACCESS S603 2024-25\S603 K 2024-25 Functional Rules](P:\Projects\Psychometrics\ACCESS 2.0\Operational\S603 2025\K ACCESS S603 2024-25\S603 K 2024-25 Functional Rules)) 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 to the universal design principles throughout test and item design in the *WIDA Consortium English*

Language Proficiency Assessment for Grades 1-12 Test and Item Design Plan ACCESS for ELLs Paper Annual Summative Assessment and WIDA Screener Paper.

3.3.1 Support Provided to All ELLs

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

Administrative considerations include adaptive and specialized equipment or furniture, alternative microphone, familiar Test Administrator, frequent or additional supervised breaks, individual or small group setting, monitoring of the placement of responses in the test booklet, 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 accessibility needs. Audio aids, 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. A Test administrator locally delivers the accommodations.

WIDA is planning to study the efficacy of accommodations. More information regarding accommodations is provided in the ACCESS for ELLs Accessibility and Accommodations manual.

WIDA also offers Braille Test for ELLs and Large Print Test. The Braille test is delivered via paper-and-pencil, and the translation and graphics are provided in either contracted or uncontracted Braille for Tier B (Grades 1–12). This test is used to provide access to the test for ELLs who are blind. The Large Print Test is used for students with visual impairments. The font size on the large print paper test is increased to 18 point.

Universal tools are also available to all ELLs taking ACCESS for ELLs. 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.

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 student 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 students' responses to 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 for the ACCESS for ELLs were experienced DRC raters. DRC drew together this pool of experienced raters to staff the scoring pool for ACCESS for ELLs. To complete the rater staffing, DRC held recruiting events, after which DRC's recruiting staff screened applications for rater positions. DRC staff then personally interviewed likely candidates. As part of the hiring process, DRC required each candidate to provide an on-demand writing sample, an on-demand math sample, references, and proof of a 4-year college degree. In this screening process, DRC gave preference to candidates who had previous experience scoring students' responses to tasks included in large-scale assessments and candidates with degrees in English language arts. The rater pool consisted of educators, writers, editors, and other professionals with content-specific backgrounds. 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. DRC selects team leaders based on their prior performance as raters and for their leadership skills, are assigned to small groups of raters, typically 7 to 10 raters on each 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.

Preparing Rater Training Materials for Writing tasks

https://sea.wida.us/system/files/documents/Assessment/Technical-Reports/ACCESS-ATR-Series-502-Paper.pdf *stopped here*

CAL test development staff produce materials that DRC uses to train their raters to score ACCESS Writing responses. CAL test development staff members who are trained on the Writing Scoring Scale ("Expert Raters") prepared these rater training materials for Writing tasks when they were originally developed for ACCESS Online. Given that the ACCESS Paper Writing test is not refreshed annually, the rater training materials carry over from year to year.

To prepare the Writing rater training materials, the Expert Rater began by reviewing the storyboard for the task (graphics, text, audio script) and by reviewing the anchor responses for an existing task targeting the same grade level cluster, proficiency level, and WIDA ELD standard, in order to internalize the task input and expectations as well as become calibrated to how the Scoring Scale has previously been applied to a similar task. The Expert Rater also reviewed documented criteria for anchor responses and score explanations.

Next, the Expert Rater reviewed field test responses in DRC ScoreBoard and identified approximately 5–10 responses per score point. For each response reviewed, the Expert Rater determined the most appropriate score and recorded any recommendations for potential anchor responses, any questions, or any other observations.

Following the Expert Rater's initial review of responses, the Writing Test Development Manager (TD Manager) reviewed the responses selected. The TD Manager confirmed or revised the scores, recording notes and feedback, and finalized the selection of one anchor response per score point. Anchor responses are typical responses for the grade level cluster and the task, in terms of both the linguistic characteristics and the content of the response. They are clear examples of the score point with both the Expert Rater and the TD Manager agreeing on the score. For tasks with primarily handwritten responses, the handwriting must also be generally legible to facilitate internalization of the linguistic characteristics by raters.

Once anchor responses were finalized, the Expert Rater wrote score explanations for each anchor. Score explanations refer to each dimension of language described in the Scoring Scale descriptors and provide additional explanation with direct quotes from the response to justify why the score point was awarded.

Finally, the TD Manager reviewed the score explanations to check that they met the required criteria. The TD Manager also selected 20 responses from the initial review to be used as training samples, and reviewed and revised any accompanying score notes, as necessary. The 20 training samples were selected so that the full range of observed score points are included in the set, and so that the most commonly observed score points for the grade level cluster and tier are well-represented. The TD Manager also reviewed all notes from the anchor and training sample selection process and, when necessary, compiled any task-specific scoring guidance to be used by raters.

The anchors, explanations, training samples, training sample notes, and any task-specific scoring guidance were then provided to WIDA for review. CAL staff updated the materials as requested by WIDA and delivered the materials to DRC for field test scoring.

Following field test scoring and operational item selection, CAL added to the rater training materials for each task that was selected for the operational test. This primarily consisted of selecting and annotating additional training samples, so that a minimum of 30 samples were provided for operational rater training. In some cases, additional anchor responses were also added to the anchor set, when an appropriate anchor response for the highest observed score point was not found while preparing for field test scoring but could be identified once a larger pool of scored responses was available.

Rater Training and Qualifying

- DRC assigned each rater a unique ID number and password.
- The scoring director conducted a team leader training session before training the raters. This session followed the same procedures as rater training but was more rigorous and in-depth due to the extra responsibilities required of team leaders. During team leader training, all WIDA materials were reviewed and discussed. To facilitate scoring consistency, it was imperative that all team leaders imparted the same rationale for each response. Once the team leaders were qualified, leadership responsibilities were reviewed, and team assignments were given.

- Rater training began with the scoring director going through the ACCESS for ELLs PowerPoint presentation provided by CAL. The PowerPoint gave scorers a good overview of ACCESS for ELLs and the WIDA scoring process.

- Rater training continued with the scoring director providing an intensive review of the ACCESS for ELLs Scoring Scale, the model student response for Speaking items, and task-specific anchor sets created by CAL. The anchor set contained a collection of student responses that were used to exemplify each possible score point. Each response included a scoring annotation that explained the scoring rationale. Scorers used the ACCESS for ELLs Scoring Scale, the model student response for Speaking, and the anchor sets as primary references during scoring.

- Next, raters practiced by independently scoring responses in training sets. Training sets were created by DRC scoring directors from responses approved by WIDA and CAL. The responses were selected to show raters the range of each score point (e.g., high, mid, and low 2s). This process helped raters recognize the various ways that a student could respond in order to earn each score point outlined and defined in the scoring guidelines. After each training set was taken, the scoring director led a thorough discussion of the responses.

- Once the scoring scale, anchor sets, and training sets were thoroughly discussed, each rater was required to demonstrate understanding of the scoring criteria by qualifying (i.e., scoring with acceptable agreement to the true scores) on at least one of the qualifying sets. Raters who failed to achieve at least 70 percent exact agreement on the first qualifying set were given additional training, either individually or in a small group setting. Raters who did not perform at the required level of agreement by the end of the qualifying process were not allowed to score any student responses. These individuals were removed from the pool of potential raters in DRC's imaging system and released from the project. Qualifying sets were created by DRC scoring directors from responses approved by WIDA and CAL.

- Throughout training, the scoring director provided detailed directions for use of DRC's computerized scoring system and remote communication tools for raters.

- Once raters were trained, qualified, and began live scoring, DRC used recalibration sets and validity responses to keep the raters calibrated on the tasks they were scoring. Recalibration sets were pre-scored sets of responses that were approved by WIDA and

CAL and were used to help refocus raters on WIDA scoring guidelines. Validity responses were also approved by WIDA and CAL and were responses that were pre-scored and used to ensure raters were adhering to WIDA scoring criteria. Recalibration and validity are explained in greater detail below.

Calculating Score Agreement for Score Monitoring

- DRC's handscoring system generated handscoring reports, detailing agreement rates for each rater and task. The reports were automatically generated overnight throughout the course of handscoring and could also be run on demand. DRC provided weekly interrater reliability reports to WIDA throughout the handscoring process to ensure that DRC maintained sufficient quality control throughout the course of scoring.

- For Writing, DRC defines **agreement** as two adjacent scores, reported as %AG. (See Section 4.3 or a description of the Writing Scoring Scale.) For example, using the Writing Scoring Scale, DRC considers scores of 2 and 2+ as agreement, as well as scores of 2 and 2 or scores of 2+ and 3. However, DRC considers scores of 2 and 3 on the Writing Scoring Scale as **adjacent**, while considering scores of 2 and 3+ as **nonadjacent**.

- For Speaking, DRC defines **agreement** as two scores that are exactly the same, reported as %EX. (See Section 4.4 for a description of the Speaking Scoring Scale.) Unlike in Writing, where DRC considers two adjacent scores as "Agreement," raters scoring responses to Speaking tasks must demonstrate Exact Agreement (EX) in order to be considered in "agreement."

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

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

- The purpose of the first and second scores was to monitor interrater reliability by comparing the scores that two separate raters assigned to the same response. When calculating final scores, the first score assigned was the score of record.

Monitoring Scoring (Quality Control)

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

- During the handscoring process, the scoring directors communicated regularly with their team leaders to review the statistics generated from the previous day's work, including interrater reliability, score point distributions, and validity reports.

- Throughout handscoring, team leaders conducted routine read- and listen-behinds to observe, in real time, the raters' performance. Team leaders utilized live, scored responses to provide ongoing feedback and, if necessary, retraining for raters.

- The DRC system generated interrater reliability reports daily to monitor how often each rater's scores matched other raters' scores, and scoring leaders continually monitored individual rater statistics, comparing them to the group average. If the agreement rate for a rater fell below 70%, supervisors increased monitoring and retraining activities with the rater. If the rater failed to demonstrate improved reliability, DRC released the rater from scoring responses to that task.

- Since the interrater agreement rates were all at or above 70%, the target that WIDA stipulated, the focus turned to raters with lower-than-average agreement rates—even if their agreement rate was at or above 70%. Even when all agreement rates were at or above 70%, scoring supervisors continued to seek opportunities to increase reliability by providing ongoing feedback and retraining raters based on the specific performance of

each rater, as evidenced by the quality control reports and observations made when reviewing scores that a rater assigned.

- DRC can retrieve students' responses on demand (e.g., specific grade-level clusters, specific students) should the need arise during or after the scoring process.

- If needed, DRC can re-score a student's response to a task based on task- or response-level information, such as task number, date, score assigned, or rater ID.

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

Handling Unusual Responses

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

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

- **Non score codes.** Unusual or aberrant responses for which raters could not assign a score based on the scoring guidelines received a nonscorable code (e.g., Writing responses that are entirely blank or consist entirely of scribbles or pictures). DRC's handscoring team collaborated with WIDA and CAL to define what specifically constituted a nonscorable response to ensure consistency when applying nonscorable codes, and CAL provided this information to DRC along with other task-specific training materials that DRC then used to train its raters. During scoring, when raters assigned a nonscorable code (except for Blank), DRC's imaging system automatically forwarded the response to a handscoring supervisor for review and approval. If the handscoring supervisor had any questions about the application of non-score codes to specific responses, the supervisor contacted WIDA and CAL representatives for further review and discussion.

- **Alerts**. To handle possible alert responses (i.e., student responses indicating potential issues related to the student's safety and/or well-being that may require attention at the local level, as well as potential plagiarism and potential teacher interference), DRC's imaging system gave raters the ability to alert questionable student responses. When a rater flagged a response with the alert status, the imaging system automatically routed the response to handscoring supervisors for review. The states are notified within 24 hours. If the response was related to the student's safety and/or well-being, and the handscoring supervisors concurred with the alert, it was then forwarded to WIDA's project management team who provided the response to the appropriate local education agency.

- **Request for originals.** When a rater came across a scanned student response that was difficult to read (for example, having some partially erased text), the rater flagged the response with a "request original" status. If a rater flagged a response as "request original," DRC's imaging system automatically forwarded the response to a handscoring supervisor. If the handscoring supervisor agreed that the original student response needed to be reviewed to properly apply the scoring guidelines, the supervisor forwarded the

request to staff in DRC's Operations Services, who located the original student response so the handscoring supervisor could review the response and score it.

Remote Scoring Procedures After the COVID-19 Pandemic

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

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

- **System tools—scoring, training, chat.** ScoreBoard is DRC's secure, web-based scoring application that is designed to be used in a distributed environment. The platform is used within DRC's scoring centers and in remote locations (e.g., in a rater's home). Integrated training resources provide the capability to securely maintain digital training materials within the scoring platform itself.

- DRC conducted live, interactive rater training using the Moodle Learning Management System, which mirrored aspects of the scoring room and provided a versatile platform for training. It also served as a place to share files of important documents, including daily scoring statistics and platform user guides. Through embedded communication tools, scoring directors, assistant scoring directors, and team leaders facilitated group and one-on-one training sessions and discussions using audio and video.

- To facilitate instant communication between supervisors and raters, DRC utilized a chat tool called Zulip in conjunction with ScoreBoard and Moodle. Zulip provided a tool for raters to directly ask supervisors questions about responses and allowed supervisors to

direct individuals or groups of raters to join Moodle training rooms for important discussions and retraining.

- **Security**. Security is essential to the handscoring process. When users logged into ScoreBoard, they were required to read and accept the security policy before they were allowed to access the project. DRC also required raters to read and sign nondisclosure agreements. During training and large-group discussions, trainers continuously emphasized what security means, the importance of maintaining security, and how all staff accomplish this. In the remote environment, DRC could give these security reminders daily. DRC requires raters working remotely to work in a private environment away from other people (including family members). Raters working in ScoreBoard were not allowed to print from their computers in order to protect the security of the student responses, test questions, and training materials. Restrictions built into ScoreBoard defined the hours during the day that raters were able to log into the system, ensuring that raters were only scoring responses while supervisors were in place to monitor handscoring and answer any questions.

- **Rater training with Moodle**. DRC conducted rater training remotely as an interactive, comprehensive, hands-on experience. For Writing training, scoring directors trained groups of raters by screensharing PDFs of training materials. Raters individually viewed each training example, with supervisors directing raters to relevant text.

- For Speaking training, scoring directors trained groups of raters by playing the responses aloud over Moodle during live, remote training sessions.

- As with site-based training sessions, supervisors guided the discussion, and raters posed questions to supervisors. The scoring director directed the team leaders and raters to take training and qualifying sets, following the same training flow as they would in the scoring facility.

- **Quality control.** DRC utilized its robust quality control processes and handscoring metrics for all scoring sessions. Scored responses were monitored with second reads, and team leaders conducted read- and listen-behinds. DRC's handscoring system allowed scoring supervisors to determine specific read- and listen-behind rates (frequency of monitoring) for each rater. Any retraining and/or conversations needed because of the monitoring were held in one-on-one video chat sessions. Handscoring quality reports were available daily and on demand for handscoring supervisors and DRC's project

leadership, and DRC also provided WIDA staffing with handscoring reports. If a rater fell below 70% exact agreement and failed to improve after retraining and feedback, DRC removed the rater from the project and assigned the responses to other raters to score.

4.3   Writing Scoring Scale

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

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

Figure 11 shows the Writing Scoring Scale.

Figure 11. Writing Scoring Scale

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

Score Point 6
- D: Sophisticated organization of text that clearly demonstrates an overall sense of unity throughout, tailored to context (e.g., purpose, situation, and audience)
- S: Purposeful use of a variety of sentence structures that are essentially error-free
- W: Precise use of vocabulary with just the right word in just the right place

5+

Score Point 5
- D: Strong organization of text that supports an overall sense of unity, appropriate to context (e.g., purpose, situation, and audience)
- S: A variety of sentence structures with very few grammatical errors
- W: A wide range of vocabulary, used appropriately and with ease

4+

Score Point 4
- D: Organized text that presents a clear progression of ideas, demonstrating an awareness of context (e.g., purpose, situation, and audience)
- S: Complex and some simple sentence structures, containing occasional grammatical errors that don't generally interfere with comprehensibility
- W: A variety of vocabulary beyond the stimulus and prompt, generally conveying the intended meaning

3+

Score Point 3
- D: Text that shows developing organization including the use of elaboration and detail, though the progression of ideas may not always be clear
- S: Simple and some complex sentence structures, whose meaning may be obscured by noticeable grammatical errors
- W: Some vocabulary beyond the stimulus and prompt, although usage is noticeably awkward at times

2+

Score Point 2
- 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)
- S: Simple sentence structures; meaning is frequently obscured by noticeable grammatical errors when attempting beyond simple sentences
- W: Vocabulary primarily drawn from the stimulus and prompt

1+

Score Point 1
- D: Minimal text that represents an idea or ideas
- S: Primarily words, chunks of language, and short phrases rather than complete sentences
- W: Distinguishable English words that are often limited to high frequency words or reformulated expressions from the stimulus and prompt

D: Discourse Level	S: Sentence Level	W: Word/Phrase Level

When assigning a score, a rater makes an initial judgment about which whole score point (1–6) best describes a response and then determines whether the three descriptors for that whole score point suit that response. If all three descriptors suit the response, the rater assigns the score associated with that score point (e.g., if all three descriptors for score point 3 are appropriate, the

rater would assign a score of 3). However, if there is clear evidence that one or two descriptors from an adjacent score point are a better fit, the rater would assign a plus score between the two applicable whole score points (e.g., if two descriptors for score point 3 seem to fit, but one descriptor for score point 4 is a better fit than the associated descriptor for score point 3, the rater would assign a score of 3+).

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

- **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. More information on how plagiarized responses are handled by DRC is provided in Section 4.2, Handling Unusual Responses, above.

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

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

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

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

Each student responds to three (or four, for Grade 1, Tier A) Writing tasks. One rater assigns a score to each response. To calculate a student's total raw score, the scores that the raters assigned 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 the intended amount 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. CAL and WIDA designed the ACCESS Writing Scoring Scale for trained raters to use to evaluate students' responses to ACCESS writing tasks; thus, it is not appropriate for any other purposes.

4.4   Speaking Scoring Scale

The Speaking 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. Although the ACCESS Paper Test Administrator training module for the Speaking test was created by the WIDA instructional design team, CAL test development staff provided the anchor samples, score justifications, and practice samples to be used in the training module. These materials were identified and created in a manner analogous to the Writing rater training materials, as described in Section 4.2 above. 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 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 12 shows the Speaking Scoring Scale.

Figure 12. 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	• 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	• 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	• 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	• Language use does not support an adequate response • Comprehensibility may be compromised • Word choice may not be fully adequate
No response (in English)	• Does not respond (in English)

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**Error! Reference source not found.**. 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 Tier A test, scores can range from 0 to 18; and on the Tier B/C test, from 6 to 30.

Table 9 Rating to Raw Score Conversion (Speaking)

Score	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 WIDA ships printed reports 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' parents or guardians will receive individual score reports. 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 links below.) How these stakeholders use the material to communicate assessment results is determined locally.

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

WIDA offers several online resources to help communicate test score information to educators, families, and students. (See ACCESS for ELLs Score and Reports https://wida.wisc.edu/assess/access/scores-reports; Family Engagement https://wida.wisc.edu/teach/learners/engagement.) WIDA also provides a post-testing Q & A webinar about score interpretation (https://portal.wida.us/webinar/detail/702b69ef-0265-eb11-a2dd-0050568beee8).

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

The Individual Student Report (Figure 13) contains detailed information about the performance of a single student in Grades K–12. Its primary users are students, parents/guardians, teachers, and school teams. It provides information about the language needed to access content and succeed in school, one indicator of a student's English language proficiency.

 Figure 13. Individual Student Report

Individual Student Report 20XX

WIDA ACCESS for ELLs 2.0®
English Language Proficiency Test

Sample Student
Birth Date: mm/dd/yyyy | Grade: sample grade
Tier: sample tier
District ID: XXXXXXXXXXXXXXX | State ID: XXXXXXXXXXXXXX
School: sample school
District: sample district
State: sample state

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
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: • 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: • 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: • 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: • 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

The Individual Student Report includes four language domain scores (Listening, Speaking, Reading, and Writing) and four language domain composite scores (Oral Language, Literacy, Comprehension, and Overall), as shown in the first table of the score report. In the first column of the last four rows of that table, test users can see how WIDA uses a student's domain scores to calculate each composite score (e.g., for Oral Language, WIDA calculates the composite score based on a student's performance on the Listening and Speaking tests, with scores on each of those tests contributing equally to the composite score). For students who are unable to complete

all four domains due to their disabilities, WIDA provides states methods to compute alternative composite scores based on their available domain scores upon request (Sahakyan, N.,2020).

The proficiency level that a student attained in each language domain is presented both graphically and as a whole number followed by a decimal. A scale score represents a student's performance that has been put on a standardized scale. Students' performance relies on the number of items and item difficulties they respond to correctly. Scale scores allow comparison across different forms and grades. In ACCESS, the scale score ranges between 100-600 in all grades. These are interpretive scores that are based on, but separate from, scale scores. The shaded bar of the graph describes a student's performance in terms of the 6-level English Language Proficiency Scale. The whole number indicates a student's English language proficiency level (1–Entering, 2–Emerging, 3–Developing, 4–Expanding, 5–Bridging, and 6–Reaching) in accordance with the WIDA ELD Standards. ELLs who attain Level 6, Reaching, have moved through the entire second language continuum, as defined by the test and the WIDA ELD Standards.

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

To the right of the proficiency level is the reported scale score and associated confidence band. The confidence band for each domain and composite reflects the standard error of measurement for the scale score, a statistical calculation of a student's likelihood of scoring within a particular range of scores if he or she were to take the same test repeatedly without any change in ability. For ACCESS scale scores, the confidence band reflects a 95% probability level.

The second table in the Individual Student Report provides information about the student's proficiency levels expressed as whole numbers. The third column of the table describes what that student should generally be able to do in each of the four language domains, given his or her level of proficiency. For example, as shown in Figure 10, this student received a proficiency level score of 2 for Speaking, which suggests that the student should generally be able to "communicate ideas and information orally in English using language that contains short sentences and everyday words and phrases."

- If a student was not tested in one (or more) of the language domains, a code of NA (Not Available) will appear in the score report for the impacted language domain(s) and for all

composite scores that are calculated using those domain scores. For these students, WIDA provides states with information about statistical methods that can be used to compute alternative composite scores based on a student's available domain scores (Sahakyan, 2020).When interpreting scores, test users are cautioned to keep in mind these points: the report provides information on English proficiency. It does not provide information on a student's academic achievement or knowledge of content areas.

- Students do not typically acquire proficiency in Listening, Speaking, Reading, and Writing at the same pace. Generally,
  - Oral language (L+S) is acquired faster than literacy (R+W).
  - Receptive language (L+R) is acquired faster than productive language (S+W).
  - Writing is usually the last domain to be mastered.
- The students' foundation in their home or primary language is a predictor of their English language development. Those who have strong literacy backgrounds in their native language will most likely acquire literacy in English at a quicker pace than students who do not.
- The Overall score is helpful as a summary of other scores and is used because a single number may be needed for reference. However, it is important to remember that it is compensatory, averaged using weights; a particularly high score in one domain may effectively offset a low score in another domain and vice versa. Similar Overall scores can mask very different performances on individual tests.
- No single scale score or language proficiency level, including the Overall score (composite), should be used as the sole determiner for making decisions regarding a student's English language proficiency. School work and local assessment throughout the school year also provide evidence of a student's English language development.
- Scale scores can be used to make comparisons across grade levels, but not across domains. Each domain has its own score scale, so scale scores should not be used for comparing performance across domains. For example, a scale score of 350 in Listening at grade 3 is not equivalent to a scale score of 350 in Speaking at grade 3. For performance comparisons across domains, proficiency levels should be used.
- Either scale scores or proficiency levels can be used to compare test performance from different years, although it is easier to see changes when examining scale scores.

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

5.2 Other Reports

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

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

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

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

Part 2:

Technical Results

Contents

1.   Annual Test Results

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 21 students were excluded from the analyses due to mismatches in students' tiers across domains. In addition, 9,771 students taking ACCESS Paper tests in Colorado used equated scores to the ACCESS Online tests; therefore, their score analyses were not included in this 601 Paper Annual Technical Report. For the equated scoring procedure, please refer to the WIDA mode-adjustment procedure report.

1.1   Participation

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

1.1.1 Grade-Level Cluster

Table 1.1.1.1 shows participation across the 41 WIDA states and U.S. territories that participated in the operational testing program of ACCESS Paper in 2022–2023 by grade level. The rows provide data for the number of students in that grade-level cluster who took the test by state, with the final row showing the total number of participants across all 41 states and territories. Some

states' sample sizes are small except for Kindergarten, which is only in Paper form, since most students take the Online form of the tests. The biggest state was Florida, which constitutes about 56% of the students who take Paper ACCESS. Illinois, Georgia, and Virginia were the next largest states. The full names of acronyms of U.S. territories are the following: BI, Bureau of Indian Education; DC, District of Columbia; DD, Department of Defense Education Activity; MP, Northern Mariana Islands; and VI, Virgin Islands.

Table 1.1.1.1

Participation by Grade-Level Cluster by State, S601 Paper

State	Cluster							Total
	K	1	2	3	4–5	6–8	9–12	
AK	702	3	6	8	32	58	78	887
AL	4159	.	.	.	1	4	.	4164
BI	507	129	120	86	176	215	31	1264
CO	8895	174	131	130	157	168	116	9771
DC	951	1	1	.	1	2	1	957
DD	559	.	4	10	8	.	2	583
DE	1590	2	2	2	2	4	.	1602
FL	37621	37337	34155	34222	45955	52416	49579	291285
GA	15273	128	129	13	29	20	27	15619
HI	1380	1	4	.	3	3	6	1397
ID	1656	4	9	9	13	8	17	1716
IL	22540	109	110	83	203	157	118	23320
IN	8293	14	19	9	34	17	18	8404
KY	4715	2	2	5	13	2	4	4743
MA	11776	33	49	43	46	32	23	12002
MD	10610	7	5	12	40	15	13	10702
ME	494	2	2	1	4	2	3	508
MI	9963	50	33	24	70	47	39	10226
MN	7965	62	44	48	68	49	43	8279
MO	4536	8	9	13	25	10	3	4604
MP	53	.	.	.	.	.	.	53
MT	183	.	.	.	.	.	.	183
NC	14896	8	10	9	20	12	10	14965
ND	349	2	.	.	3	6	3	363
NH	513	3	.	4	3	2	.	525
NJ	13689	123	80	41	56	47	28	14064
NM	3890	1	.	1	2	6	12	3912
NV	5390	.	.	.	.	1	1	5392
OK	4798	13	17	12	36	38	51	4965
PA	8620	190	80	83	117	121	138	9349
RI	1518	1	1	.	4	2	7	1533
SC	4439	6	7	8	16	8	2	4486
SD	886	21	16	15	18	24	.	980
TN	7792	11	3	4	10	2	4	7826
UT	4073	1	2	.	1	1	2	4080
VA	13504	1065	197	148	188	131	135	15368
VI	80	.	.	.	.	.	.	80
VT	144	1	1	.	.	1	.	147
WA	13343	10	17	7	19	15	3	13414
WI	4422	10	19	23	45	41	7	4567
WY	244	3	2	2	1	5	3	260
Total	257011	39535	35286	35075	47419	53692	50527	518545

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 44% female and 49.5% male in general. The Hispanic ethnicity percentage was about 79% in all clusters except kindergarten, which was 65%.

Table 1.1.1.2

Participation by Grade-Level Cluster by Gender, S601 Paper

Cluster		Gender			Total
		F	M	Missing	
K	Count	106604	118227	32180	257011
	% within Grade	41.48%	46.0%	12.52%	100.0%
1	Count	18599	20821	115	39535
	% within Grade	47.04%	52.66%	0.29%	100.0%
2	Count	16692	18516	78	35286
	% within Grade	47.3%	52.47%	0.22%	100.0%
3	Count	16297	18697	81	35075
	% within Grade	46.46%	53.31%	0.23%	100.0%
4-5	Count	21918	25397	104	47419
	% within Grade	46.22%	53.56%	0.22%	100.0%
6-8	Count	24861	28685	146	53692
	% within Grade	46.3%	53.43%	0.27%	100.0%
9-12	Count	23524	26768	235	50527
	% within Grade	46.56%	52.98%	0.47%	100.0%
Total	Count	228495	257111	32939	518545
	% within Grade	44.06%	49.58%	6.35%	100.0%

Table 1.1.1.3

Participation by Grade-Level Cluster by Ethnicity, S601 Paper

Cluster		Ethnicity			Total
		Hispanic	Non-Hispanic	Unknown	
K	Count	167985	73184	15842	257011
	% within Grade	65.36%	28.48%	6.16%	100.0%
1	Count	31225	7935	375	39535
	% within Grade	78.98%	20.07%	0.95%	100.0%
2	Count	28192	6837	257	35286
	% within Grade	79.9%	19.38%	0.73%	100.0%
3	Count	28224	6597	254	35075
	% within Grade	80.47%	18.81%	0.72%	100.0%
4-5	Count	37788	9216	415	47419
	% within Grade	79.69%	19.44%	0.88%	100.0%
6-8	Count	43153	9976	563	53692
	% within Grade	80.37%	18.58%	1.05%	100.0%
9-12	Count	40176	9661	690	50527
	% within Grade	79.51%	19.12%	1.37%	100.0%
Total	Count	376743	123406	18396	518545
	% within Grade	72.65%	23.8%	3.55%	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 49.5% of the Paper ACCESS population.

Table 1.1.2.1

Participation by Grade by State, S601 Paper

State	Grade													Total
	K	1	2	3	4	5	6	7	8	9	10	11	12	
AK	702	3	6	8	13	19	20	19	19	25	25	14	14	887
AL	4159	.	.	.	1	.	1	1	2	.	.	.	.	4164
BI	507	129	120	86	90	86	73	80	62	4	7	11	9	1264
CO	8895	174	131	130	95	62	44	55	69	9	27	43	37	9771
DC	951	1	1	.	1	.	1	.	1	.	1	.	.	957
DD	559	.	4	10	5	3	.	.	.	1	.	.	1	583
DE	1590	2	2	2	.	2	2	2	.	.	.	.	.	1602
FL	37621	37337	34155	34222	24037	21918	19338	16978	16100	16097	13805	11769	7908	291285
GA	15273	128	129	13	14	15	9	7	4	15	5	5	2	15619
HI	1380	1	4	.	2	1	1	.	2	1	2	2	1	1397
ID	1656	4	9	9	8	5	4	3	1	11	4	.	2	1716
IL	22540	109	110	83	102	101	64	53	40	34	33	26	25	23320
IN	8293	14	19	9	20	14	8	5	4	6	5	4	3	8404
KY	4715	2	2	5	6	7	1	.	1	1	1	1	1	4743
MA	11776	33	49	43	20	26	18	9	5	7	1	5	10	12002
MD	10610	7	5	12	18	22	5	4	6	6	3	2	2	10702
ME	494	2	2	1	3	1	2	.	.	1	.	1	1	508

MI	9963	50	33	24	41	29	17	18	12	11	7	8	13	10226
MN	7965	62	44	48	26	42	24	9	16	10	16	9	8	8279
MO	4536	8	9	13	14	11	3	2	5	2	1	.	.	4604
MP	53	.	.	.	.	.	.	.	.	.	.	.	.	53
MT	183	.	.	.	.	.	.	.	.	.	.	.	.	183
NC	14896	8	10	9	6	14	5	4	3	4	3	2	1	14965
ND	349	2	.	.	1	2	2	1	3	.	2	.	1	363
NH	513	3	.	4	3	.	1	1	.	.	.	.	.	525
NJ	13689	123	80	41	30	26	18	17	12	6	8	8	6	14064
NM	3890	1	.	1	1	1	1	1	4	3	2	5	2	3912
NV	5390	.	.	.	.	.	.	1	.	.	1	.	.	5392
OK	4798	13	17	12	20	16	9	14	15	23	17	7	4	4965
PA	8620	190	80	83	65	52	40	45	36	42	40	22	34	9349
RI	1518	1	1	.	2	2	.	.	2	4	.	1	2	1533
SC	4439	6	7	8	8	8	2	2	4	.	.	1	1	4486
SD	886	21	16	15	10	8	14	8	2	.	.	.	.	980
TN	7792	11	3	4	7	3	.	.	2	.	1	2	1	7826
UT	4073	1	2	.	.	1	.	1	.	.	.	1	1	4080
VA	13504	1065	197	148	114	74	45	39	47	59	40	16	20	15368
VI	80	.	.	.	.	.	.	.	.	.	.	.	.	80
VT	144	1	1	.	.	.	1	.	.	.	.	.	.	147
WA	13343	10	17	7	9	10	2	8	5	.	.	2	1	13414
WI	4422	10	19	23	25	20	18	13	10	3	3	.	1	4567
WY	244	3	2	2	.	1	1	3	1	2	1	.	.	260

Total	257011	39535	35286	35075	24817	22602	19794	17403	16495	16387	14061	11967	8112	518545

Table 1.1.2.2

Participation by Grade by Gender, S601 Paper

Grade		Gender			Total
		F	M	Missing	
K	Count	106604	118227	32180	257011
	% within Grade	41.48%	46.0%	12.52%	100.0%
1	Count	18599	20821	115	39535
	% within Grade	47.04%	52.66%	0.29%	100.0%
2	Count	16692	18516	78	35286
	% within Grade	47.3%	52.47%	0.22%	100.0%
3	Count	16297	18697	81	35075
	% within Grade	46.46%	53.31%	0.23%	100.0%
4	Count	11513	13241	63	24817
	% within Grade	46.39%	53.35%	0.25%	100.0%
5	Count	10405	12156	41	22602
	% within Grade	46.04%	53.78%	0.18%	100.0%
6	Count	8992	10725	77	19794
	% within Grade	45.43%	54.18%	0.39%	100.0%
7	Count	8195	9169	39	17403
	% within Grade	47.09%	52.69%	0.22%	100.0%
8	Count	7674	8791	30	16495
	% within Grade	46.52%	53.29%	0.18%	100.0%
9	Count	7517	8705	165	16387
	% within Grade	45.87%	53.12%	1.01%	100.0%
10	Count	6480	7555	26	14061
	% within Grade	46.08%	53.73%	0.18%	100.0%
11	Count	5667	6274	26	11967
	% within Grade	47.36%	52.43%	0.22%	100.0%
12	Count	3860	4234	18	8112
	% within Grade	47.58%	52.19%	0.22%	100.0%
Total	Count	228495	257111	32939	518545
	% within Grade	44.06%	49.58%	6.35%	100.0%

Table 1.1.2.3

Participation by Grade by Ethnicity, S601 Paper

Grade		Ethnicity			Total
		Hispanic	Non-Hispanic	Unknown	
K	Count	167985	73184	15842	257011
	% within Grade	65.36%	28.48%	6.16%	100.0%
1	Count	31225	7935	375	39535
	% within Grade	78.98%	20.07%	0.95%	100.0%
2	Count	28192	6837	257	35286
	% within Grade	79.9%	19.38%	0.73%	100.0%
3	Count	28224	6597	254	35075
	% within Grade	80.47%	18.81%	0.72%	100.0%
4	Count	19665	4920	232	24817
	% within Grade	79.24%	19.83%	0.93%	100.0%
5	Count	18123	4296	183	22602
	% within Grade	80.18%	19.01%	0.81%	100.0%
6	Count	15803	3764	227	19794
	% within Grade	79.84%	19.02%	1.15%	100.0%
7	Count	14027	3190	186	17403
	% within Grade	80.6%	18.33%	1.07%	100.0%
8	Count	13323	3022	150	16495
	% within Grade	80.77%	18.32%	0.91%	100.0%
9	Count	13123	2892	372	16387

	% within Grade	80.08%	17.65%	2.27%	100.0%
10	Count	11430	2503	128	14061
	% within Grade	81.29%	17.8%	0.91%	100.0%
11	Count	9404	2448	115	11967
	% within Grade	78.58%	20.46%	0.96%	100.0%
12	Count	6219	1818	75	8112
	% within Grade	76.66%	22.41%	0.92%	100.0%
Total	Count	376743	123406	18396	518545
	% within Grade	72.65%	23.8%	3.55%	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, 60% of students were in Tier A and 40% in Tier B/C. In Grade 2, 35% of students were in Tier A and 65% in Tier B/C. In Grade 3 and Grades 4–5, 28-30% were in Tier A and 70-72% in Tier B/C. In Grades 6–8 and 9–12, there were about 38% of students in Tier A and 62% 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, S601 Paper

Cluster			Domain			
			Listening	Reading	Speaking	Writing
K	Tier	-	257005	257000	257005	257000
1	Tier	A	23687	23693	23692	23692
		B	15826	15830	15833	15831
	Total		39513	39523	39525	39523
2	Tier	A	12290	12293	12292	12293
		B	22981	22981	22981	22983
	Total		35271	35274	35273	35276
3	Tier	A	9884	9886	9885	9886
		B	25187	25184	25188	25188
	Total		35071	35070	35073	35074
4–5	Tier	A	14447	14449	14449	14445
		B	32965	32965	32965	32965
	Total		47412	47414	47414	47410
6–8	Tier	A	20626	20627	20623	20622
		B	33062	33061	33057	33058
	Total		53688	53688	53680	53680
9–12	Tier	A	19130	19131	19129	19130
		B	31378	31379	31390	31377
	Total		50508	50510	50519	50507

Table 1.1.3.2

Participation by Grade by Tier by Domain, S601 Paper

Cluster			Domain			
			Listening	Reading	Speaking	Writing
K	Tier	-	257005	257000	257005	257000
1	Tier	A	23687	23693	23692	23692
		B	15826	15830	15833	15831
	Total		39513	39523	39525	39523
2	Tier	A	12290	12293	12292	12293
		B	22981	22981	22981	22983
	Total		35271	35274	35273	35276
3	Tier	A	9884	9886	9885	9886
		B	25187	25184	25188	25188
	Total		35071	35070	35073	35074
4	Tier	A	7216	7217	7217	7216
		B	17598	17598	17599	17598
	Total		24814	24815	24816	24814
5	Tier	A	7231	7232	7232	7229
		B	15367	15367	15366	15367
	Total		22598	22599	22598	22596
6	Tier	A	6999	6999	6998	6996
		B	12794	12794	12792	12790
	Total		19793	19793	19790	19786
	Tier	A	6978	6978	6976	6978

7		B	10425	10424	10423	10424
	Total		17403	17402	17399	17402
8	Tier	A	6649	6650	6649	6648
		B	9843	9843	9842	9844
	Total		16492	16493	16491	16492
9	Tier	A	7203	7203	7202	7202
		B	9179	9179	9183	9179
	Total		16382	16382	16385	16381
10	Tier	A	5518	5518	5517	5518
		B	8540	8541	8542	8541
	Total		14058	14059	14059	14059
11	Tier	A	4150	4151	4151	4151
		B	7812	7811	7813	7809
	Total		11962	11962	11964	11960
12	Tier	A	2259	2259	2259	2259
		B	5847	5848	5852	5848
	Total		8106	8107	8111	8107

Table 1.1.3.3

Participation by Grade-Level Cluster by Tier by Gender, S601 Paper

Cluster	Tier		Gender			Total
			F	M	Missing	
K	-	Count	106604	118227	32180	257011
		% within Tier	41.48%	46.00%	12.52%	100.00%
1	A	Count	11029	12594	79	23702
		% within Tier	46.53%	53.13%	0.33%	100.00%
	BC	Count	7570	8227	36	15833
		% within Tier	47.81%	51.96%	0.23%	100.00%
2	A	Count	5610	6633	56	12299
		% within Tier	45.61%	53.93%	0.46%	100.00%
	BC	Count	11082	11883	22	22987
		% within Tier	48.21%	51.69%	0.10%	100.00%
3	A	Count	4566	5269	51	9886
		% within Tier	46.19%	53.30%	0.52%	100.00%
	BC	Count	11731	13428	30	25189
		% within Tier	46.57%	53.31%	0.12%	100.00%
4–5	A	Count	6764	7625	62	14451
		% within Tier	46.81%	52.76%	0.43%	100.00%
	BC	Count	15154	17772	42	32968
		% within Tier	45.97%	53.91%	0.13%	100.00%
6–8	A	Count	9733	10805	89	20627
		% within Tier	47.19%	52.38%	0.43%	100.00%
	BC	Count	15128	17880	57	33065
		% within Tier	45.75%	54.08%	0.17%	100.00%
9–12	A	Count	9045	9976	112	19133
		% within Tier	47.27%	52.14%	0.59%	100.00%
	BC	Count	14479	16792	123	31394
		% within Tier	46.12%	53.49%	0.39%	100.00%

Table 1.1.3.4 presents percentages of Hispanic and other ethnic groups in tiers. Overall, the percentages of Hispanic students in Tier A were 3% to 6% 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, S601 Paper

Cluster	Tier		Ethnicity			Total
			Hispanic	Other	Unknown	
K	-	Count	167985	73184	15842	257011
		% within Tier	65.36%	28.48%	6.16%	100.00%
1	A	Count	19040	4400	262	23702
		% within Tier	80.33%	18.56%	1.11%	100.00%
	BC	Count	12185	3535	113	15833
		% within Tier	76.96%	22.33%	0.71%	100.00%
2	A	Count	9995	2159	145	12299
		% within Tier	81.27%	17.55%	1.18%	100.00%
	BC	Count	18197	4678	112	22987
		% within Tier	79.16%	20.35%	0.49%	100.00%
3	A	Count	8054	1718	114	9886
		% within Tier	81.47%	17.38%	1.15%	100.00%
	BC	Count	20170	4879	140	25189
		% within Tier	80.07%	19.37%	0.56%	100.00%
4–5	A	Count	11940	2336	175	14451
		% within Tier	82.62%	16.16%	1.21%	100.00%
	BC	Count	25848	6880	240	32968
		% within Tier	78.40%	20.87%	0.73%	100.00%
		Count	17353	3036	238	20627

6–8	A	% within Tier	84.13%	14.72%	1.15%	100.00%
	BC	Count	25800	6940	325	33065
		% within Tier	78.03%	20.99%	0.98%	100.00%
9–12	A	Count	15882	2920	331	19133
		% within Tier	83.01%	15.26%	1.73%	100.00%
	BC	Count	24294	6741	359	31394
		% within Tier	77.38%	21.47%	1.14%	100.00%

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 was Listening, Reading, Speaking, and then Writing in clusters of 1, 2 and 6–8, whereas the order for cluster 3 and 4-5 was Listening, Reading, Writing and Speaking in descending order. Kindergarten had the average scale scores in the order of Speaking, Listening, Writing, and then Reading.

Table 1.2.1.1

Mean Scale Scores by Grade-Level Cluster, S601 Paper

Cluster		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
K	Mean	249.44	172.29	187.05	250.29	250.12	179.89	195.43	200.75
	N	247975	247965	247962	247956	247955	247957	247963	247940
1	Mean	300.19	282.52	239.87	259.33	281.99	262.51	288.49	269.08
	N	33452	30322	39328	39030	33201	30310	27121	26915
2	Mean	324.81	305.2	268.99	273.9	300.84	288.38	311.79	292.83
	N	32294	28753	35127	34890	32069	28745	27122	26929
3	Mean	351.56	329.76	286.96	286.42	319.88	309.16	336.58	312.64
	N	32625	28890	34937	34715	32413	28887	27514	27343
4–5	Mean	367.44	342.78	319.56	316.78	343.29	332.07	350.64	335.82
	N	45204	41229	47248	46984	44953	41223	39945	39734
6–8	Mean	366.14	349.91	312	329.91	349.23	331.74	355.2	337.35
	N	50554	47944	53490	53061	50131	47918	46030	45617
9–12	Mean	374.3	377.27	343.38	336.22	356.39	360.88	376.73	359.77
	N	47369	45149	50365	49933	46932	45129	43192	42781

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, S601 Paper

Cluster	Gender		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
K	F	Mean	254.93	171.66	189.72	257.62	256.53	180.91	196.63	203.38
		N	106546	106543	106542	106542	106542	106542	106543	106538
	M	Mean	243.46	171.94	183.75	241.68	242.83	178.06	193.39	197.28
		N	118129	118123	118121	118119	118118	118116	118121	118108
	Missing	Mean	250.79	171.73	187.9	255.22	253.27	180.04	195.43	201.79
		N	32172	32171	32171	32167	32167	32171	32171	32166
1	F	Mean	302.37	282.82	244.36	262.77	284.51	264.72	289.27	271.25
		N	16111	14518	18588	18463	16001	14514	13137	13046
	M	Mean	298.36	282.31	236.05	256.34	279.79	260.63	287.84	267.18
		N	17403	15867	20799	20626	17262	15859	14049	13933
	Missing	Mean	282.18	274.71	209.79	232.37	261.04	242.57	278.51	249.13
		N	95	92	115	114	94	92	79	79
2	F	Mean	327.13	305.98	274.44	274.96	302.26	291.46	312.93	295.23
		N	15554	13826	16682	16565	15446	13824	13170	13075
	M	Mean	322.59	304.53	264.21	273.05	299.51	285.62	310.72	290.59
		N	16794	14984	18497	18378	16677	14977	14007	13908
	Missing	Mean	308.03	289.75	247.92	234.17	276.36	270.78	295.2	273.5
		N	71	64	78	77	70	64	61	60
3	F	Mean	350.87	329.33	292.87	286.26	319.36	311.76	336.01	314.17
		N	15319	13640	16295	16199	15232	13640	13060	12990
	M	Mean	352.16	330.16	282.01	286.64	320.37	306.97	337.09	311.35
		N	17352	15288	18691	18564	17226	15285	14491	14389
	Missing	Mean	324	305.57	252.67	233.8	280.93	278.97	310.55	278.45
		N	76	67	81	81	76	67	65	65
4–5	F	Mean	366.65	342.26	324.54	316.16	342.5	334.08	349.92	336.82
		N	21061	19356	21911	21776	20934	19354	18808	18698
	M	Mean	368.41	343.3	315.44	317.42	344.17	330.39	351.39	335.07
		N	24200	21935	25388	25260	24075	21929	21204	21100
	Missing	Mean	335.53	319.34	287.78	274.98	305.62	305.26	324.14	305.62
		N	99	82	104	104	99	82	77	77
6–8	F	Mean	365.39	351.26	317.29	328.7	348.28	335.04	355.94	339.4
		N	23522	22535	24844	24667	23338	22521	21654	21469
	M	Mean	366.93	348.61	307.55	330.92	350.1	328.83	354.53	335.55
		N	27050	25422	28668	28416	26812	25410	24390	24163
	Missing	Mean	349.21	341.19	295.02	320.39	336.38	317.6	342.58	323.09
		N	137	134	146	143	134	134	127	124
9–12	F	Mean	373.61	379.17	347.44	334.23	355.11	363.87	377.88	361.51
		N	22180	21302	23498	23300	21977	21294	20424	20238
	M	Mean	375.05	375.58	339.97	338.15	357.66	358.29	375.75	358.3

	N	25081	23738	26748	26513	24845	23726	22667	22440
Missing	Mean	365.56	370.26	329	316.45	343.96	350.62	369.14	349.13
	N	216	208	235	234	216	208	196	196

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

Table 1.2.1.3

Mean Scale Scores by Grade-Level Cluster by Ethnicity, S601 Paper

Cluster	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
K	Hispanic (of any Race)	Mean	241.18	162.39	177.7	240.06	240.87	170.26	186.02	191.24
		N	161094	161088	161087	161081	161080	161082	161086	161068
	Non-Hispanic American Indian	Mean	266.1	173.86	183.09	267.99	267.3	178.7	201.52	205.08
		N	1589	1589	1589	1589	1589	1589	1589	1589
	Non-Hispanic Asian	Mean	281.33	212.87	225.83	284.63	283.25	219.6	233.39	238.47
		N	29067	29066	29066	29064	29064	29066	29066	29064
	Non-Hispanic Black	Mean	261.07	188.06	195.67	274.64	268.12	192.09	209.97	214.69
		N	13954	13953	13952	13953	13953	13952	13953	13952
	Non-Hispanic Multiracial	Mean	274.76	197.8	208.26	279.65	277.47	203.27	220.88	225.3
		N	1130	1130	1130	1130	1130	1130	1130	1130
	Non-Hispanic Pacific Islander	Mean	254.37	163.9	180.15	261.07	257.98	172.24	191.04	197.75
		N	1512	1512	1512	1512	1512	1512	1512	1512
	Non-Hispanic White	Mean	263.54	186.01	203.15	267.36	265.71	194.81	209.26	215.86
		N	23820	23818	23817	23819	23819	23817	23818	23817
	Unknown	Mean	239.49	162.83	178.71	239.31	239.65	170.97	185.82	191.38
		N	15809	15809	15809	15808	15808	15809	15809	15808
1	Hispanic (of any Race)	Mean	299.35	281.36	238.93	256.92	280.37	261.29	287.42	267.76
		N	26474	23976	31059	30819	26271	23964	21449	21282
	Non-Hispanic American Indian	Mean	303.46	280.45	243.92	262.37	285.63	264.61	288.01	271.16
		N	109	102	126	126	109	102	94	94
	Non-Hispanic Asian	Mean	309.16	295.51	257.94	276.13	294.35	277.91	300.33	283.46
		N	1291	1181	1468	1452	1275	1181	1084	1069
	Non-Hispanic Black	Mean	299.59	284.02	235.24	270.46	287.58	262.03	289.3	270.53
		N	2324	2119	2856	2838	2312	2119	1851	1842
	Non-Hispanic Multiracial	Mean	312.99	291.64	251.21	283.4	300.28	271.77	297.88	279.9
		N	209	191	250	247	206	191	168	165
	Non-Hispanic Pacific Islander	Mean	298.95	282.21	240.14	258.04	278.32	261.97	286.1	266.14
		N	42	34	49	48	41	34	29	28
	Non-Hispanic White	Mean	304.62	285.96	245.91	265.17	286.96	267.7	292.27	273.95
		N	2678	2418	3145	3127	2663	2418	2170	2160
	Unknown	Mean	291.55	280.95	222.48	242.73	271.81	254.56	284.8	261.08
		N	325	301	375	373	324	301	276	275

Cluster	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
2	Hispanic (of any Race)	Mean	324.05	304.24	268.25	271.29	299.18	287.5	310.87	291.69
		N	25838	23100	28070	27882	25660	23094	21812	21657
	Non-Hispanic American Indian	Mean	328.92	310	280.36	276.25	303.71	295.84	317.87	300.01
		N	104	85	107	107	103	85	82	81
	Non-Hispanic Asian	Mean	336.14	321.21	287.89	290.06	314.13	305.56	326.43	309.11
		N	1139	1032	1243	1233	1130	1032	966	958
	Non-Hispanic Black	Mean	323.97	301.54	262.59	284.32	305.66	283.26	308.94	290.88
		N	2323	1993	2587	2566	2305	1992	1854	1843
	Non-Hispanic Multiracial	Mean	334.7	315.74	281.5	304.48	319.83	298.44	321.37	304.23
		N	186	160	197	196	185	160	153	152
	Non-Hispanic Pacific Islander	Mean	316.15	292.33	255.33	254.54	288.89	277.88	301.81	283.71
		N	39	33	42	41	38	33	32	31
	Non-Hispanic White	Mean	329.33	311.38	274.81	284.44	308.23	294.69	317.66	299.6
		N	2430	2149	2624	2609	2414	2148	2030	2015
	Unknown	Mean	307.37	295.21	250.66	247.73	280.92	275.37	299.44	277.33
		N	235	201	257	256	234	201	193	192
3	Hispanic (of any Race)	Mean	350.76	329.18	286.51	284.66	318.59	308.66	335.93	311.9
		N	26321	23357	28110	27938	26157	23356	22302	22169
	Non-Hispanic American Indian	Mean	354.72	334.49	288.24	289.66	322.22	312.82	341.35	316.03
		N	120	99	125	124	119	99	96	96
	Non-Hispanic Asian	Mean	359.97	339.53	299.9	300.78	331.11	320.47	345.62	323.46
		N	937	819	996	989	931	818	785	781
	Non-Hispanic Black	Mean	353.61	329.54	283.8	294.24	325.31	307.3	337.11	313.09
		N	2637	2325	2886	2864	2616	2324	2165	2149
	Non-Hispanic Multiracial	Mean	359.75	336.35	300.66	307.19	333.9	318.66	343.92	323.08
		N	166	138	182	181	165	138	132	131
	Non-Hispanic Pacific Islander	Mean	346.26	330.46	291.9	281.05	312.42	311.43	335.11	311.51
		N	38	37	40	40	38	37	35	35
	Non-Hispanic White	Mean	355.37	332.82	291.07	293.07	324.96	312.86	339.96	316.62
		N	2174	1919	2344	2329	2158	1919	1816	1802
	Unknown	Mean	342.44	323.46	272.64	259	301.79	299.08	329.79	301.54
		N	232	196	254	250	229	196	183	180

Cluster	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
4-5	Hispanic (of any Race)	Mean	365.95	341.67	318.66	313.06	340.69	331.05	349.4	334.34
		N	36091	33075	37646	37437	35891	33069	32076	31906
	Non-Hispanic American Indian	Mean	374.59	344.99	322.8	319.26	348.82	335.1	354.26	339.45
		N	179	163	189	189	179	163	159	159
	Non-Hispanic Asian	Mean	381.77	357.16	332.55	338.37	360.89	345.59	365.23	350.72
		N	1304	1186	1366	1353	1292	1186	1148	1138
	Non-Hispanic Black	Mean	372.59	344.8	320.76	335.37	354.98	333.76	353.47	340.37
		N	3921	3416	4127	4104	3898	3416	3298	3280
	Non-Hispanic Multiracial	Mean	384	356.01	332.26	356.7	370.98	344.83	364.38	352.83
		N	202	170	211	209	202	170	167	167
	Non-Hispanic Pacific Islander	Mean	363.61	339.98	324.94	328.99	347.21	332.13	345.42	335.34
		N	74	60	77	75	72	60	59	58
	Non-Hispanic White	Mean	372.7	347.79	323.58	327	351.4	336.75	356.05	341.91
		N	3040	2819	3218	3203	3026	2819	2711	2699
	Unknown	Mean	353.56	332.1	306.05	294.98	326.33	320.39	339.27	322.76
		N	393	340	414	414	393	340	327	327
6-8	Hispanic (of any Race)	Mean	363.64	348.66	310.75	325.4	345.79	330.43	353.55	335.37
		N	40694	38674	42977	42626	40344	38651	37159	36814
	Non-Hispanic American Indian	Mean	383.03	354.83	318.63	345.11	365.43	337.65	363.82	346.82
		N	243	226	259	258	242	226	219	218
	Non-Hispanic Asian	Mean	388.71	365.35	331.49	362.88	376.88	349.21	372.74	357.66
		N	1167	1117	1235	1229	1161	1117	1076	1070
	Non-Hispanic Black	Mean	376.78	351.37	313.94	349.96	364.17	333.87	359.51	343.4
		N	4336	3988	4648	4612	4305	3986	3793	3765
	Non-Hispanic Multiracial	Mean	384.53	361.78	328.09	363.44	373.81	346.18	369.23	355.11
		N	194	183	204	204	193	183	178	177
	Non-Hispanic Pacific Islander	Mean	370.49	345.4	315.43	336.82	358.97	331.58	354.3	340.85
		N	67	67	74	73	67	67	61	61
	Non-Hispanic White	Mean	373.49	356.68	317.02	345.35	360.41	337.62	362.4	345.06
		N	3330	3191	3530	3505	3305	3190	3064	3041
	Unknown	Mean	360.25	350.81	308.05	318.81	341.75	331.36	354.34	336.11
		N	523	498	563	554	514	498	480	471

Cluster	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
9-12	Hispanic (of any Race)	Mean	372.02	376.24	342.19	331.62	352.89	359.73	375.31	357.88
		N	37758	36132	40037	39703	37421	36116	34618	34292
	Non-Hispanic American Indian	Mean	376.43	376.15	334.49	332.02	355.45	355.91	376.05	355.57
		N	127	122	134	134	127	122	116	116
	Non-Hispanic Asian	Mean	399.23	391	363.74	369.33	385.69	378.09	394.21	381.04
		N	1254	1211	1321	1310	1243	1210	1162	1151
	Non-Hispanic Black	Mean	378.91	376.89	342.84	355.2	368.48	360.7	377.85	363.32
		N	4219	3818	4564	4517	4169	3817	3613	3575
	Non-Hispanic Multiracial	Mean	396.71	387.03	357.07	370.27	384.23	371.97	389.98	375.39
		N	201	194	210	207	198	194	189	186
	Non-Hispanic Pacific Islander	Mean	384.3	377.19	350.61	351.47	370.27	364.21	380.29	366.16
		N	66	63	72	68	62	63	59	56
	Non-Hispanic White	Mean	385.76	385.49	352.15	353.94	371.05	369.32	385.98	370.23
		N	3123	3019	3337	3310	3094	3017	2885	2858
	Unknown	Mean	364.89	369.82	331.08	317.9	344.12	351.27	368.91	350.14
		N	621	590	690	684	618	590	550	547

1.2.2 Mean Scale Score Across Domain and Composite Score by Grade

This section shows the mean scale scores broken down by grade rather than by grade-level cluster. Tables are provided for the total student population, for the student population by gender, and for the student population by race and ethnicity. Table 1.2.2.1 shows the increment of scale scores by grade. Across domains, mean scale scores increased with grades. Mean scale scores increased as grades increased, with slight decreases in grade 6 in Listening, Reading, and Writing, and grade 9 in Listening. In the Speaking domain there was a slight decrease in grades 7 and 9. Table 1.2.2.2 exhibits student performance by gender. Female student groups mostly scored higher than male groups throughout grades, except for a few grades across domains.

Table 1.2.2.1

Mean Scale Scores by Grade, S601 Paper

Grade		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
K	Mean	249.44	172.29	187.05	250.29	250.12	179.89	195.43	200.75
	N	247975	247965	247962	247956	247955	247957	247963	247940
1	Mean	300.19	282.52	239.87	259.33	281.99	262.51	288.49	269.08
	N	33452	30322	39328	39030	33201	30310	27121	26915
2	Mean	324.81	305.2	268.99	273.9	300.84	288.38	311.79	292.83
	N	32294	28753	35127	34890	32069	28745	27122	26929
3	Mean	351.56	329.76	286.96	286.42	319.88	309.16	336.58	312.64
	N	32625	28890	34937	34715	32413	28887	27514	27343
4	Mean	364.7	340.4	315.58	316.29	341.64	328.85	348.04	332.96
	N	23571	21240	24713	24557	23422	21236	20536	20417
5	Mean	370.42	345.31	323.92	317.31	345.08	335.49	353.39	338.86
	N	21633	19989	22535	22427	21531	19987	19409	19317
6	Mean	363.19	344.74	308.21	328.72	347.07	327.23	350.55	333.41
	N	18674	17539	19736	19580	18525	17531	16835	16694
7	Mean	365.46	349.69	311.62	327.66	347.85	331.38	354.82	336.67
	N	16407	15554	17343	17202	16269	15550	14961	14828
8	Mean	370.41	356.23	316.98	333.72	353.29	337.44	361.1	342.71
	N	15473	14851	16411	16279	15337	14837	14234	14095
9	Mean	367.87	371.2	335.96	325.69	348.25	354	370.53	352.5
	N	15306	14505	16361	16211	15153	14497	13825	13683
10	Mean	371.77	375.8	341.06	330.17	351.91	358.96	374.86	356.97
	N	13195	12603	14030	13906	13076	12603	12065	11955
11	Mean	379.69	382.42	349.83	344.45	363.15	366.78	381.99	365.98
	N	11213	10655	11911	11815	11116	10649	10195	10102
12	Mean	383.62	384.26	352.92	355.92	370.45	369.17	384.46	369.73
	N	7655	7386	8063	8001	7587	7380	7107	7041

Table 1.2.2.2

Mean Scale Scores by Grade by Gender, S601 Paper

Grade	Gender		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
K	F	Mean	255.24	172.25	190.05	257.88	256.82	181.38	197.13	203.8
		N	102334	102331	102330	102330	102330	102330	102331	102326
	M	Mean	243.82	172.49	184.1	242.05	243.19	178.51	193.88	197.71
		N	113473	113467	113465	113463	113462	113460	113465	113452
	Missing	Mean	250.79	171.73	187.9	255.23	253.28	180.04	195.44	201.8
		N	32168	32167	32167	32163	32163	32167	32167	32162
1	F	Mean	302.31	282.82	244.31	262.77	284.49	264.69	289.25	271.23
		N	16039	14449	18512	18387	15929	14445	13072	12981
	M	Mean	298.33	282.3	236.07	256.4	279.8	260.64	287.83	267.19
		N	17318	15781	20701	20529	17178	15773	13970	13855
	Missing	Mean	282.18	274.71	209.79	232.37	261.04	242.57	278.51	249.13
		N	95	92	115	114	94	92	79	79
2	F	Mean	327.19	306.01	274.44	275	302.31	291.48	312.97	295.27
		N	15502	13773	16627	16511	15395	13771	13120	13026
	M	Mean	322.68	304.53	264.16	273.08	299.58	285.6	310.75	290.61
		N	16721	14916	18422	18302	16604	14910	13941	13843
	Missing	Mean	308.03	289.75	247.92	234.17	276.36	270.78	295.2	273.5
		N	71	64	78	77	70	64	61	60
3	F	Mean	350.94	329.39	292.85	286.36	319.45	311.78	336.08	314.22
		N	15261	13589	16234	16139	15175	13589	13011	12942
	M	Mean	352.22	330.2	281.97	286.7	320.42	306.97	337.14	311.37
		N	17288	15234	18622	18495	17162	15231	14438	14336
	Missing	Mean	324	305.57	252.67	233.8	280.93	278.97	310.55	278.45
		N	76	67	81	81	76	67	65	65
4	F	Mean	363.46	339.28	319.79	315.54	340.57	330.21	346.78	333.46
		N	11011	10006	11469	11387	10932	10005	9713	9646
	M	Mean	365.94	341.52	312.07	317.2	342.78	327.76	349.29	332.65
		N	12501	11187	13181	13107	12431	11184	10780	10728
	Missing	Mean	332.39	314.02	281.73	262.43	298.49	297.3	318.49	297
		N	59	47	63	63	59	47	43	43
5	F	Mean	369.99	345.49	329.71	316.88	344.55	338.22	353.25	340.38
		N	9987	9291	10379	10325	9939	9291	9036	8994
	M	Mean	370.9	345.22	319.06	317.75	345.64	333.17	353.57	337.59
		N	11606	10663	12115	12061	11552	10661	10339	10289
	Missing	Mean	340.15	326.49	297.07	294.27	316.12	315.94	331.29	316.53
		N	40	35	41	41	40	35	34	34

Grade	Gender		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
6	F	Mean	361.98	345.69	313.93	327	345.58	330.51	350.82	335.22
		N	8497	8049	8965	8901	8429	8044	7740	7674
	M	Mean	364.28	343.94	303.5	330.19	348.38	324.49	350.35	331.93
		N	10106	9419	10694	10603	10026	9416	9029	8955
	Missing	Mean	353.2	344.38	295.84	324.68	340.81	319.37	345.14	324.89
		N	71	71	77	76	70	71	66	65
7	F	Mean	364.85	351.22	316.44	326.62	347.08	334.64	355.8	338.82
		N	7743	7446	8172	8103	7676	7445	7169	7104
	M	Mean	366.1	348.34	307.39	328.67	348.61	328.45	353.99	334.77
		N	8627	8072	9132	9061	8557	8069	7758	7691
	Missing	Mean	346.14	335.78	291.08	310.32	330.19	312.17	337.68	317.24
		N	37	36	39	38	36	36	34	33
8	F	Mean	369.88	357.93	322.09	333.1	352.8	340.77	362.13	344.93
		N	7216	6978	7636	7592	7167	6970	6684	6630
	M	Mean	370.98	354.78	312.59	334.3	353.8	334.53	360.26	340.79
		N	8228	7846	8745	8658	8142	7840	7523	7439
	Missing	Mean	343.38	340.04	298.03	322.34	333.25	320.22	342.52	326
		N	29	27	30	29	28	27	27	26
9	F	Mean	367.19	373.28	340.71	324.42	347.29	357.41	371.8	354.6
		N	7035	6719	7505	7432	6960	6716	6406	6339
	M	Mean	368.5	369.37	332.03	327.07	349.2	351.07	369.4	350.72
		N	8122	7644	8691	8615	8044	7639	7288	7213
	Missing	Mean	366.05	371.54	327.28	310.35	341.89	350.5	370.77	349.08
		N	149	142	165	164	149	142	131	131
10	F	Mean	372.01	378.1	345.46	329.04	351.61	362.33	376.56	359.32
		N	6100	5875	6467	6418	6054	5875	5641	5597
	M	Mean	371.6	373.82	337.33	331.18	352.2	356.05	373.39	354.93
		N	7069	6702	7537	7462	6996	6702	6398	6332
	Missing	Mean	364.38	368.73	330.31	319.58	342.23	349.81	367.46	347.23
		N	26	26	26	26	26	26	26	26
11	F	Mean	378.41	383.76	352.83	341.76	361.29	368.95	382.63	367.01
		N	5339	5121	5646	5603	5292	5117	4905	4861
	M	Mean	380.91	381.26	347.16	346.87	364.87	364.84	381.47	365.07
		N	5850	5510	6239	6186	5800	5508	5267	5218
	Missing	Mean	368.62	363.92	338.12	349.65	358.79	351.42	363.74	352.52
		N	24	24	26	26	24	24	23	23
12	F	Mean	381.43	385.58	355.79	350.99	366.81	371.28	384.66	370.04
		N	3669	3552	3842	3810	3635	3551	3438	3408
	M	Mean	385.74	383.08	350.4	360.57	373.92	367.28	384.35	369.54
		N	3969	3818	4203	4173	3935	3813	3653	3617
	Missing	Mean	358.76	371	329.67	319.61	343.82	351.81	366.25	347.81
		N	17	16	18	18	17	16	16	16

Table 1.2.2.3

Mean Scale Scores by Grade by Ethnicity, S601 Paper

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
0	Hispanic (of any Race)	Mean	241.18	162.39	177.7	240.06	240.87	170.26	186.02	191.24
		N	161094	161088	161087	161081	161080	161082	161086	161068
	Non-Hispanic American Indian	Mean	266.1	173.86	183.09	267.99	267.3	178.7	201.52	205.08
		N	1589	1589	1589	1589	1589	1589	1589	1589
	Non-Hispanic Asian	Mean	281.33	212.87	225.83	284.63	283.25	219.6	233.39	238.47
		N	29067	29066	29066	29064	29064	29066	29066	29064
	Non-Hispanic Black	Mean	261.07	188.06	195.67	274.64	268.12	192.09	209.97	214.69
		N	13954	13953	13952	13953	13953	13952	13953	13952
	Non-Hispanic Multiracial	Mean	274.76	197.8	208.26	279.65	277.47	203.27	220.88	225.3
		N	1130	1130	1130	1130	1130	1130	1130	1130
	Non-Hispanic Pacific Islander	Mean	254.37	163.9	180.15	261.07	257.98	172.24	191.04	197.75
		N	1512	1512	1512	1512	1512	1512	1512	1512
	Non-Hispanic White	Mean	263.54	186.01	203.15	267.36	265.71	194.81	209.26	215.86
		N	23820	23818	23817	23819	23819	23817	23818	23817
	Unknown	Mean	239.49	162.83	178.71	239.31	239.65	170.97	185.82	191.38
		N	15809	15809	15809	15808	15808	15809	15809	15808
1	Hispanic (of any Race)	Mean	299.35	281.36	238.93	256.92	280.37	261.29	287.42	267.76
		N	26474	23976	31059	30819	26271	23964	21449	21282
	Non-Hispanic American Indian	Mean	303.46	280.45	243.92	262.37	285.63	264.61	288.01	271.16
		N	109	102	126	126	109	102	94	94
	Non-Hispanic Asian	Mean	309.16	295.51	257.94	276.13	294.35	277.91	300.33	283.46
		N	1291	1181	1468	1452	1275	1181	1084	1069
	Non-Hispanic Black	Mean	299.59	284.02	235.24	270.46	287.58	262.03	289.3	270.53
		N	2324	2119	2856	2838	2312	2119	1851	1842
	Non-Hispanic Multiracial	Mean	312.99	291.64	251.21	283.4	300.28	271.77	297.88	279.9
		N	209	191	250	247	206	191	168	165
	Non-Hispanic Pacific Islander	Mean	298.95	282.21	240.14	258.04	278.32	261.97	286.1	266.14
		N	42	34	49	48	41	34	29	28
	Non-Hispanic White	Mean	304.62	285.96	245.91	265.17	286.96	267.7	292.27	273.95
		N	2678	2418	3145	3127	2663	2418	2170	2160
	Unknown	Mean	291.55	280.95	222.48	242.73	271.81	254.56	284.8	261.08
		N	325	301	375	373	324	301	276	275

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
2	Hispanic (of any Race)	Mean	324.05	304.24	268.25	271.29	299.18	287.5	310.87	291.69
		N	25838	23100	28070	27882	25660	23094	21812	21657
	Non-Hispanic American Indian	Mean	328.92	310	280.36	276.25	303.71	295.84	317.87	300.01
		N	104	85	107	107	103	85	82	81
	Non-Hispanic Asian	Mean	336.14	321.21	287.89	290.06	314.13	305.56	326.43	309.11
		N	1139	1032	1243	1233	1130	1032	966	958
	Non-Hispanic Black	Mean	323.97	301.54	262.59	284.32	305.66	283.26	308.94	290.88
		N	2323	1993	2587	2566	2305	1992	1854	1843
	Non-Hispanic Multiracial	Mean	334.7	315.74	281.5	304.48	319.83	298.44	321.37	304.23
		N	186	160	197	196	185	160	153	152
	Non-Hispanic Pacific Islander	Mean	316.15	292.33	255.33	254.54	288.89	277.88	301.81	283.71
		N	39	33	42	41	38	33	32	31
	Non-Hispanic White	Mean	329.33	311.38	274.81	284.44	308.23	294.69	317.66	299.6
		N	2430	2149	2624	2609	2414	2148	2030	2015
	Unknown	Mean	307.37	295.21	250.66	247.73	280.92	275.37	299.44	277.33
		N	235	201	257	256	234	201	193	192
3	Hispanic (of any Race)	Mean	350.76	329.18	286.51	284.66	318.59	308.66	335.93	311.9
		N	26321	23357	28110	27938	26157	23356	22302	22169
	Non-Hispanic American Indian	Mean	354.72	334.49	288.24	289.66	322.22	312.82	341.35	316.03
		N	120	99	125	124	119	99	96	96
	Non-Hispanic Asian	Mean	359.97	339.53	299.9	300.78	331.11	320.47	345.62	323.46
		N	937	819	996	989	931	818	785	781
	Non-Hispanic Black	Mean	353.61	329.54	283.8	294.24	325.31	307.3	337.11	313.09
		N	2637	2325	2886	2864	2616	2324	2165	2149
	Non-Hispanic Multiracial	Mean	359.75	336.35	300.66	307.19	333.9	318.66	343.92	323.08
		N	166	138	182	181	165	138	132	131
	Non-Hispanic Pacific Islander	Mean	346.26	330.46	291.9	281.05	312.42	311.43	335.11	311.51
		N	38	37	40	40	38	37	35	35
	Non-Hispanic White	Mean	355.37	332.82	291.07	293.07	324.96	312.86	339.96	316.62
		N	2174	1919	2344	2329	2158	1919	1816	1802
	Unknown	Mean	342.44	323.46	272.64	259	301.79	299.08	329.79	301.54
		N	232	196	254	250	229	196	183	180

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
4	Hispanic (of any Race)	Mean	363.38	339.43	314.86	312.9	339.26	328.02	346.97	331.7
		N	18708	16938	19579	19448	18583	16934	16390	16289
	Non-Hispanic American Indi...	Mean	369.89	343.72	319.01	312.39	343.51	332.23	352.78	336.63
		N	91	81	96	96	91	81	79	79
	Non-Hispanic Asian	Mean	378.78	353.43	327.91	337.85	359.17	341.2	361.7	347.17
		N	730	656	771	767	726	656	632	629
	Non-Hispanic Black	Mean	368.87	342.21	315.81	333.91	352.37	329.9	350.39	336.58
		N	2067	1774	2174	2164	2057	1774	1712	1705
	Non-Hispanic Multiracial	Mean	379.39	350.71	329.23	349.86	364.72	339.8	358.91	347.33
		N	111	93	118	117	111	93	90	90
	Non-Hispanic Pacific Islander	Mean	358.4	336.62	315.62	324.16	343.02	324.78	341.39	328.49
		N	43	37	45	43	41	37	36	35
	Non-Hispanic White	Mean	369.58	344.63	319.35	325.13	349.14	332.9	352.64	338.22
		N	1599	1468	1699	1691	1591	1468	1409	1402
	Unknown	Mean	347	327.15	296.88	283.31	317.32	312.69	333.32	314.32
		N	222	193	231	231	222	193	188	188
5	Hispanic (of any Race)	Mean	368.71	344.02	322.78	313.24	342.23	334.22	351.94	337.09
		N	17383	16137	18067	17989	17308	16135	15686	15617
	Non-Hispanic American Indi...	Mean	379.44	346.26	326.71	326.35	354.32	337.94	355.71	342.24
		N	88	82	93	93	88	82	80	80
	Non-Hispanic Asian	Mean	385.58	361.79	338.56	339.04	363.1	351.01	369.56	355.12
		N	574	530	595	586	566	530	516	509
	Non-Hispanic Black	Mean	376.75	347.6	326.27	337.01	357.9	337.93	356.78	344.48
		N	1854	1642	1953	1940	1841	1642	1586	1575
	Non-Hispanic Multiracial	Mean	389.62	362.42	336.11	365.4	378.62	350.91	370.77	359.25
		N	91	77	93	92	91	77	77	77
	Non-Hispanic Pacific Islander	Mean	370.84	345.39	338.03	335.47	352.74	343.96	351.74	345.78
		N	31	23	32	32	31	23	23	23
	Non-Hispanic White	Mean	376.16	351.21	328.32	329.09	353.91	340.93	359.73	345.89
		N	1441	1351	1519	1512	1435	1351	1302	1297
	Unknown	Mean	362.08	338.6	317.62	309.7	338.01	330.48	347.32	334.17
		N	171	147	183	183	171	147	139	139

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
6	Hispanic (of any Race)	Mean	361.03	343.64	307.28	324.84	344.12	326.12	349.07	331.69
		N	14934	14034	15754	15634	14815	14026	13474	13363
	Non-Hispanic American Indian	Mean	370.72	348.19	311.78	339.26	355.4	332.18	355.65	339.46
		N	83	73	86	86	83	73	72	72
	Non-Hispanic Asian	Mean	383.66	359.44	324.72	356.21	370.66	342.98	367.36	351.6
		N	417	404	446	443	414	404	391	388
	Non-Hispanic Black	Mean	371.86	345.47	308.51	346.23	359.87	328.25	353.78	338.13
		N	1711	1579	1831	1814	1699	1579	1503	1491
	Non-Hispanic Multiracial	Mean	381.12	355.6	320.86	356.71	368.89	340.34	363.84	349.92
		N	73	65	77	77	73	65	63	63
	Non-Hispanic Pacific Islander	Mean	366.52	345.75	320.87	332.97	354.04	335.89	353.16	342.84
		N	27	28	30	30	27	28	25	25
	Non-Hispanic White	Mean	370.15	351	312.99	341.32	356.58	332.79	357.23	340.27
		N	1219	1152	1285	1275	1210	1152	1112	1103
	Unknown	Mean	355.21	346.09	303.33	317.14	338.79	326.26	349.17	331.52
		N	210	204	227	221	204	204	195	189
7	Hispanic (of any Race)	Mean	362.67	348.24	310.13	322.45	343.9	329.86	352.97	334.43
		N	13237	12599	13974	13852	13118	12595	12119	12001
	Non-Hispanic American Indian	Mean	386.29	356.32	326.58	351.43	371.13	342.34	365.45	351.47
		N	83	79	90	90	83	79	77	77
	Non-Hispanic Asian	Mean	390.01	365.93	332.06	364.47	378.66	349.7	373.26	358.25
		N	378	358	396	394	376	358	344	342
	Non-Hispanic Black	Mean	376.61	351.54	314.27	348.62	363.22	333.93	359.43	342.96
		N	1339	1200	1423	1415	1331	1200	1156	1150
	Non-Hispanic Multiracial	Mean	382.02	359.71	331.49	381.16	380.72	345.85	366.79	356.23
		N	65	62	68	67	64	62	61	60
	Non-Hispanic Pacific Islander	Mean	380.08	345.54	314.22	343.52	367.88	329.25	357.82	342.77
		N	25	24	27	27	25	24	22	22
	Non-Hispanic White	Mean	374.28	357.79	317.09	347.32	362.07	338.16	363.33	345.94
		N	1110	1072	1179	1171	1102	1072	1027	1021
	Unknown	Mean	364.41	352.76	310.49	321.08	344.84	334.57	357.48	339.52
		N	170	160	186	186	170	160	155	155

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
8	Hispanic (of any Race)	Mean	367.76	354.96	315.55	329.17	349.77	336.05	359.39	340.64
		N	12523	12041	13249	13140	12411	12030	11566	11450
	Non-Hispanic American Indian	Mean	392.78	359.8	317.12	344.32	370.17	338.04	370.41	349.3
		N	77	74	83	82	76	74	70	69
	Non-Hispanic Asian	Mean	393.06	371.48	338.58	368.82	382	355.82	378.39	363.96
		N	372	355	393	392	371	355	341	340
	Non-Hispanic Black	Mean	383.51	358.9	320.74	356.21	370.89	341.15	367.19	350.84
		N	1286	1209	1394	1383	1275	1207	1134	1124
	Non-Hispanic Multiracial	Mean	391.89	371.23	333.63	352.27	372.34	353.32	378.28	359.91
		N	56	56	59	60	56	56	54	54
	Non-Hispanic Pacific Islander	Mean	361.67	344.53	307.76	332.75	353	327.27	350.79	334.29
		N	15	15	17	16	15	15	14	14
	Non-Hispanic White	Mean	376.68	362.23	321.8	348.01	363.23	342.78	367.58	349.84
		N	1001	967	1066	1059	993	966	925	917
	Unknown	Mean	362.71	355.65	312.18	318.44	342.31	335.28	358.35	338.8
		N	143	134	150	147	140	134	130	127
9	Hispanic (of any Race)	Mean	365.91	370.24	335.04	321.39	344.99	353.03	369.24	350.78
		N	12297	11696	13103	12987	12178	11690	11165	11052
	Non-Hispanic American Indian	Mean	361.14	369.91	325.08	311.58	337.42	348.43	366.51	344.53
		N	50	47	53	53	50	47	45	45
	Non-Hispanic Asian	Mean	392.95	385.29	355.74	358.65	377.86	371.3	388.12	374.04
		N	386	371	415	410	382	371	353	350
	Non-Hispanic Black	Mean	373.66	372.04	335.34	350.12	364.49	354.54	373.06	358.29
		N	1185	1058	1274	1260	1170	1058	1005	993
	Non-Hispanic Multiracial	Mean	382.25	373.35	341.6	355.73	370.25	357.49	375.92	361.02
		N	56	55	60	60	56	55	53	53
	Non-Hispanic Pacific Islander	Mean	377.14	366.94	338.75	318.52	352.7	353.39	373.19	353.56
		N	21	18	24	23	20	18	16	16
	Non-Hispanic White	Mean	379.07	379.2	345.54	343.76	362.83	362.68	379.7	363.12
		N	981	944	1060	1051	970	942	898	887
	Unknown	Mean	355.71	363.31	321.55	302.89	331.36	342.48	360.9	339.25
		N	330	316	372	367	327	316	290	287

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
10	Hispanic (of any Race)	Mean	369.17	374.59	339.72	325.46	348.28	357.64	373.23	354.95
		N	10768	10310	11405	11306	10673	10310	9898	9808
	Non-Hispanic American Indian	Mean	382.55	380	333.57	327.1	357.03	356.34	381.14	357.46
		N	29	29	30	30	29	29	28	28
	Non-Hispanic Asian	Mean	400.07	391.25	364.98	364.67	383.98	378.55	394.67	380.77
		N	314	308	331	329	312	308	293	291
	Non-Hispanic Black	Mean	377.83	375.13	340.9	350.06	364.49	358.91	376.23	360.63
		N	1032	918	1130	1116	1018	918	864	854
	Non-Hispanic Multiracial	Mean	405.33	393.24	358.03	372.66	388.9	375.8	397.05	380.15
		N	60	59	64	62	58	59	56	54
	Non-Hispanic Pacific Islander	Mean	380.2	378.09	346.18	348.9	366.74	362.41	378.8	364.53
		N	20	22	22	21	19	22	20	19
	Non-Hispanic White	Mean	383.59	384.67	348.87	348.74	366.99	367.38	384.47	367.46
		N	860	847	920	915	855	847	806	801
	Unknown	Mean	373.75	372.73	336.2	328.76	355.13	354.83	373.98	356.67
		N	112	110	128	127	112	110	100	100
11	Hispanic (of any Race)	Mean	377.22	381.36	348.53	339.62	359.42	365.61	380.48	364.02
		N	8823	8427	9352	9276	8746	8421	8073	7995
	Non-Hispanic American Indian	Mean	386.58	381.28	348.37	342.3	365.31	365.28	383.62	364.92
		N	26	25	27	27	26	25	24	24
	Non-Hispanic Asian	Mean	403.03	395.2	367.34	378.68	391.68	381.81	398.45	385.3
		N	306	291	319	317	304	291	281	279
	Non-Hispanic Black	Mean	383.48	381.12	348.14	358.5	372.65	365.4	382.29	367.91
		N	1108	1014	1205	1196	1099	1014	953	947
	Non-Hispanic Multiracial	Mean	403.2	392.88	368.18	384.58	394.18	379.64	395.9	382.93
		N	45	42	45	45	45	42	42	42
	Non-Hispanic Pacific Islander	Mean	386.55	378.4	369.17	393.64	387.2	373	379.9	375.22
		N	11	10	12	11	10	10	10	9
	Non-Hispanic White	Mean	392.44	391.26	359.94	364.29	379.67	375.91	391.94	377.36
		N	786	748	836	828	778	748	718	712
	Unknown	Mean	372.05	377.93	342.11	330.64	353.81	361.94	377.34	361.22
		N	108	98	115	115	108	98	94	94

Grade	Ethnicity		Listening	Reading	Writing	Speaking	Oral	Literacy	Compre-hension	Overall
12	Hispanic (of any Race)	Mean	382.25	383.93	352.32	352.51	368.04	368.59	383.82	368.6
		N	5870	5699	6177	6134	5824	5695	5482	5437
	Non-Hispanic American Indian	Mean	391.09	378.67	340.83	371.75	382.68	360.9	381.58	367.11
		N	22	21	24	24	22	21	19	19
	Non-Hispanic Asian	Mean	403.25	394.41	370.61	380.96	392.64	383.48	397.71	386.82
		N	248	241	256	254	245	240	235	231
	Non-Hispanic Black	Mean	381.43	379.86	348.45	363.88	373.17	364.8	380.34	367.1
		N	894	828	955	945	882	827	791	781
	Non-Hispanic Multiracial	Mean	396.72	390.74	366	372.28	385.87	378.53	392.61	380.46
		N	40	38	41	40	39	38	38	37
	Non-Hispanic Pacific Islander	Mean	399.14	388.92	362	378.23	389.46	375.46	391.62	378.75
		N	14	13	14	13	13	13	13	12
	Non-Hispanic White	Mean	392.17	390.29	358.92	367.3	380.69	375.5	391.52	377.76
		N	496	480	521	516	491	480	463	458
	Unknown	Mean	382.68	384.09	352.68	353.44	370.8	371.64	384.42	371.77
		N	71	66	75	75	71	66	66	66

1.2.3 Correlations

The tables in this section show Pearson correlations among the four domain scale scores by grade-level clusters across all tiers, as well as the number of students included in each correlation. Results are provided by grade-level cluster. In Kindergarten, the correlation between Listening and Speaking, and the correlation between Reading and Writing were pronounced. In Grades 1 to 12, the highest correlations were between Listening and Reading and between Reading and Writing in general. Speaking and Writing had a high correlation for grades 6-8 and 9-12 as well.

Table 1.2.3.1

Correlations Among Scale Scores: K, S601 Paper

Domains	Correlations and N counts	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.539	0.806	0.593
	N	247975	247963	247955	247959
Reading	Pearson Correlation		1	0.499	0.704
	N		247965	247945	247957
Speaking	Pearson Correlation			1	0.563
	N			247956	247945
Writing	Pearson Correlation				1
	N				247962

Table 1.2.3.2

Correlations Among Scale Scores: Grade 1, S601 Paper

Domains	Correlations and N counts	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.523	0.581	0.521
	N	33452	27121	33201	33439
Reading	Pearson Correlation		1	0.439	0.503
	N		30322	30086	30310
Speaking	Pearson Correlation			1	0.498
	N			39030	39009
Writing	Pearson Correlation				1
	N				39328

Table 1.2.3.3

Correlations Among Scale Scores: Grade 2, S601 Paper

Domains	Correlations and N counts	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.614	0.629	0.593
	N	32294	27122	32069	32283
Reading	Pearson Correlation		1	0.54	0.672
	N		28753	28556	28745
Speaking	Pearson Correlation			1	0.562
	N			34890	34880
Writing	Pearson Correlation				1
	N				35127

Table 1.2.3.4

Correlations Among Scale Scores: Grade 3, S601 Paper

Domains	Correlations and N counts	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.66	0.577	0.573
	N	32625	27514	32413	32619
Reading	Pearson Correlation		1	0.577	0.64
	N		28890	28713	28887
Speaking	Pearson Correlation			1	0.585
	N			34715	34709
Writing	Pearson Correlation				1
	N				34937

Table 1.2.3.5

Correlations Among Scale Scores: Grades 4–5, S601 Paper

Domains	Correlations and N counts	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.75	0.681	0.661
	N	45204	39945	44953	45197
Reading	Pearson Correlation		1	0.671	0.709
	N		41229	41014	41223
Speaking	Pearson Correlation			1	0.692
	N			46984	46977
Writing	Pearson Correlation				1
	N				47248

Table 1.2.3.6

Correlations Among Scale Scores: Grades 6–8, S601 Paper

Domains	Correlations and N counts	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.701	0.68	0.708
	N	50554	46030	50131	50529
Reading	Pearson Correlation		1	0.598	0.68
	N		47944	47528	47918
Speaking	Pearson Correlation			1	0.702
	N			53061	53031
Writing	Pearson Correlation				1
	N				53490

Table 1.2.3.7

Correlations Among Scale Scores: Grades 9–12, S601 Paper

Domains	Correlations and N counts	Listening	Reading	Speaking	Writing
Listening	Pearson Correlation	1	0.726	0.68	0.69
	N	47369	43192	46932	47348
Reading	Pearson Correlation		1	0.633	0.7
	N		45149	44731	45129
Speaking	Pearson Correlation			1	0.7
	N			49933	49899
Writing	Pearson Correlation				1
	N				50365

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, Speaking, Reading and Writing. Percentages of PL 5 and 6 across grades ranged 4%-31% (Listening), 0%-14% (Reading), 1%-23% (Speaking), and 0% - 0.7% (Writing). Grades with the highest percentages in PL5 and PL6 in each domain are grades 3-4 (Listening), grades 4-5 (Reading), grade K (Speaking), and grades 4-5 (Writing).

1.3.1 Domains

1.3.1.1 Listening

1.3.1.1.1   By Cluster by Tier

Table 1.3.1.1.1.1

Proficiency Level by Cluster (Count): Listening, S601 Paper

Cluster	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	89121	23174	19459	12544	31918	71759	247975
1	A	2403	3457	3441	2116	4893	3220	19530
	BC	234	489	1811	2204	2976	6208	13922
2	A	2249	2648	2040	1335	2580	0	10852
	BC	164	1180	3146	1962	5635	9355	21442
3	A	361	2671	2714	1410	1110	737	9003
	BC	10	415	3157	2208	8820	9012	23622
4–5	A	1306	4802	3523	1761	1150	844	13386
	BC	27	615	2981	4961	12447	10787	31818
6–8	A	6204	7540	2838	1345	1054	143	19124
	BC	61	1250	4984	9285	8513	7337	31430
9–12	A	8835	5073	2589	972	263	0	17732
	BC	486	2790	7621	9317	5776	3647	29637

Table 1.3.1.1.1.2

Proficiency Level by Cluster (Percent): Listening, S601 Paper

Cluster	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	35.9%	9.3%	7.8%	5.1%	12.9%	28.9%	100.0%
1	A	12.3%	17.7%	17.6%	10.8%	25.1%	16.5%	100.0%
	BC	1.7%	3.5%	13.0%	15.8%	21.4%	44.6%	100.0%
2	A	20.7%	24.4%	18.8%	12.3%	23.8%	0.0%	100.0%
	BC	0.8%	5.5%	14.7%	9.2%	26.3%	43.6%	100.0%
3	A	4.0%	29.7%	30.1%	15.7%	12.3%	8.2%	100.0%
	BC	0.0%	1.8%	13.4%	9.3%	37.3%	38.2%	100.0%
4–5	A	9.8%	35.9%	26.3%	13.2%	8.6%	6.3%	100.0%
	BC	0.1%	1.9%	9.4%	15.6%	39.1%	33.9%	100.0%
6–8	A	32.4%	39.4%	14.8%	7.0%	5.5%	0.7%	100.0%
	BC	0.2%	4.0%	15.9%	29.5%	27.1%	23.3%	100.0%
9–12	A	49.8%	28.6%	14.6%	5.5%	1.5%	0.0%	100.0%
	BC	1.6%	9.4%	25.7%	31.4%	19.5%	12.3%	100.0%

1.3.1.1.2   *By Grade by Tier*
Table 1.3.1.1.2.1

Proficiency Level by Grade (Count): Listening, S601 Paper

Grade	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	89121	23174	19459	12544	31918	71759	247975
1	A	2403	3457	3441	2116	4893	3220	19530
	BC	234	489	1811	2204	2976	6208	13922
2	A	2249	2648	2040	1335	2580	0	10852
	BC	164	1180	3146	1962	5635	9355	21442
3	A	361	2671	2714	1410	1110	737	9003
	BC	10	415	3157	2208	8820	9012	23622
4	A	526	2448	1748	851	743	350	6666
	BC	17	230	1490	2593	6607	5968	16905
5	A	780	2354	1775	910	407	494	6720
	BC	10	385	1491	2368	5840	4819	14913
6	A	1732	2870	917	563	353	59	6494
	BC	16	436	1732	3981	3161	2854	12180
7	A	2422	2106	1210	298	366	84	6486
	BC	19	462	1462	3305	2656	2017	9921
8	A	2050	2564	711	484	335	0	6144
	BC	26	352	1790	1999	2696	2466	9329
9	A	2962	2209	894	448	113	0	6626
	BC	56	493	1689	2920	1921	1601	8680
10	A	2857	1128	901	140	109	0	5135

	BC	73	872	1988	2758	1558	811	8060
11	A	1842	1250	481	239	23	0	3835
	BC	126	623	2431	1771	1520	907	7378
12	A	1174	486	313	145	18	0	2136
	BC	231	802	1513	1868	777	328	5519

Table 1.3.1.1.2.2

Proficiency Level by Grade (Percent): Listening, S601 Paper

Grade	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	35.90%	9.30%	7.80%	5.10%	12.90%	28.90%	100.00%
1	A	12.30%	17.70%	17.60%	10.80%	25.10%	16.50%	100.00%
	BC	1.70%	3.50%	13.00%	15.80%	21.40%	44.60%	100.00%
2	A	20.70%	24.40%	18.80%	12.30%	23.80%	0.00%	100.00%
	BC	0.80%	5.50%	14.70%	9.20%	26.30%	43.60%	100.00%
3	A	4.00%	29.70%	30.10%	15.70%	12.30%	8.20%	100.00%
	BC	0.00%	1.80%	13.40%	9.30%	37.30%	38.20%	100.00%
4	A	7.90%	36.70%	26.20%	12.80%	11.10%	5.30%	100.00%
	BC	0.10%	1.40%	8.80%	15.30%	39.10%	35.30%	100.00%
5	A	11.60%	35.00%	26.40%	13.50%	6.10%	7.40%	100.00%
	BC	0.10%	2.60%	10.00%	15.90%	39.20%	32.30%	100.00%
6	A	26.70%	44.20%	14.10%	8.70%	5.40%	0.90%	100.00%
	BC	0.10%	3.60%	14.20%	32.70%	26.00%	23.40%	100.00%
7	A	37.30%	32.50%	18.70%	4.60%	5.60%	1.30%	100.00%
	BC	0.20%	4.70%	14.70%	33.30%	26.80%	20.30%	100.00%
8	A	33.40%	41.70%	11.60%	7.90%	5.50%	0.00%	100.00%
	BC	0.30%	3.80%	19.20%	21.40%	28.90%	26.40%	100.00%
9	A	44.70%	33.30%	13.50%	6.80%	1.70%	0.00%	100.00%
	BC	0.60%	5.70%	19.50%	33.60%	22.10%	18.40%	100.00%
10	A	55.60%	22.00%	17.50%	2.70%	2.10%	0.00%	100.00%

	BC	0.90%	10.80%	24.70%	34.20%	19.30%	10.10%	100.00%
11	A	48.00%	32.60%	12.50%	6.20%	0.60%	0.00%	100.00%
	BC	1.70%	8.40%	32.90%	24.00%	20.60%	12.30%	100.00%
12	A	55.00%	22.80%	14.70%	6.80%	0.80%	0.00%	100.00%
	BC	4.20%	14.50%	27.40%	33.80%	14.10%	5.90%	100.00%

1.3.1.1.3   By Grade
Table 1.3.1.1.3.1

Proficiency Level by Grade (Count): Listening, S601 Paper

Grade	Listening Proficiency Range						Total
	1	2	3	4	5	6	
K	89121	23174	19459	12544	31918	71759	247975
1	2637	3946	5252	4320	7869	9428	33452
2	2413	3828	5186	3297	8215	9355	32294
3	371	3086	5871	3618	9930	9749	32625
4	543	2678	3238	3444	7350	6318	23571
5	790	2739	3266	3278	6247	5313	21633
6	1748	3306	2649	4544	3514	2913	18674
7	2441	2568	2672	3603	3022	2101	16407
8	2076	2916	2501	2483	3031	2466	15473
9	3018	2702	2583	3368	2034	1601	15306
10	2930	2000	2889	2898	1667	811	13195
11	1968	1873	2912	2010	1543	907	11213
12	1405	1288	1826	2013	795	328	7655

Table 1.3.1.1.3.2

Proficiency Level by Grade (Percent): Listening, S601 Paper

Grade	Listening Proficiency Range						Total
	1	2	3	4	5	6	
K	35.90%	9.30%	7.80%	5.10%	12.90%	28.90%	100.00%
1	7.90%	11.80%	15.70%	12.90%	23.50%	28.20%	100.00%
2	7.50%	11.90%	16.10%	10.20%	25.40%	29.00%	100.00%
3	1.10%	9.50%	18.00%	11.10%	30.40%	29.90%	100.00%
4	2.30%	11.40%	13.70%	14.60%	31.20%	26.80%	100.00%
5	3.70%	12.70%	15.10%	15.20%	28.90%	24.60%	100.00%
6	9.40%	17.70%	14.20%	24.30%	18.80%	15.60%	100.00%
7	14.90%	15.70%	16.30%	22.00%	18.40%	12.80%	100.00%
8	13.40%	18.80%	16.20%	16.00%	19.60%	15.90%	100.00%
9	19.70%	17.70%	16.90%	22.00%	13.30%	10.50%	100.00%
10	22.20%	15.20%	21.90%	22.00%	12.60%	6.10%	100.00%
11	17.60%	16.70%	26.00%	17.90%	13.80%	8.10%	100.00%
12	18.40%	16.80%	23.90%	26.30%	10.40%	4.30%	100.00%

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, S601 Paper

Cluster	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	194858	6164	16369	10986	19588	0	247965
1	A	8207	6180	2116	835	524	361	18223
	BC	143	2924	4933	1367	1503	1229	12099
2	A	5816	2507	928	248	407	99	10005
	BC	1290	5544	4995	1967	2743	2209	18748
3	A	4282	2500	909	197	276	95	8259
	BC	139	2217	9197	3740	3612	1726	20631
4–5	A	7342	3308	922	408	552	43	12575
	BC	349	5268	9653	4997	5383	3004	28654
6–8	A	9337	6563	1623	322	380	242	18467
	BC	1485	12616	8453	2873	2813	1237	29477
9–12	A	6687	6902	2227	732	749	335	17632
	BC	666	8605	7915	3392	4011	2928	27517

Table 1.3.1.2.1.2

Proficiency Level by Cluster (Percent): Reading, S601 Paper

Cluster	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	78.6%	2.5%	6.6%	4.4%	7.9%	0.0%	100.0%
1	A	45.0%	33.9%	11.6%	4.6%	2.9%	2.0%	100.0%
	BC	1.2%	24.2%	40.8%	11.3%	12.4%	10.2%	100.0%
2	A	58.1%	25.1%	9.3%	2.5%	4.1%	1.0%	100.0%
	BC	6.9%	29.6%	26.6%	10.5%	14.6%	11.8%	100.0%
3	A	51.8%	30.3%	11.0%	2.4%	3.3%	1.2%	100.0%
	BC	0.7%	10.7%	44.6%	18.1%	17.5%	8.4%	100.0%
4–5	A	58.4%	26.3%	7.3%	3.2%	4.4%	0.3%	100.0%
	BC	1.2%	18.4%	33.7%	17.4%	18.8%	10.5%	100.0%
6–8	A	50.6%	35.5%	8.8%	1.7%	2.1%	1.3%	100.0%
	BC	5.0%	42.8%	28.7%	9.7%	9.5%	4.2%	100.0%
9–12	A	37.9%	39.1%	12.6%	4.2%	4.2%	1.9%	100.0%
	BC	2.4%	31.3%	28.8%	12.3%	14.6%	10.6%	100.0%

1.3.1.2.2   *By Grade by Tier*
Table 1.3.1.2.2.1

Proficiency Level by Grade (Count): Reading, S601 Paper

Grade	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	194858	6164	16369	10986	19588	0	247965
1	A	8207	6180	2116	835	524	361	18223
	BC	143	2924	4933	1367	1503	1229	12099
2	A	5816	2507	928	248	407	99	10005
	BC	1290	5544	4995	1967	2743	2209	18748
3	A	4282	2500	909	197	276	95	8259
	BC	139	2217	9197	3740	3612	1726	20631
4	A	3452	1811	404	165	290	43	6165
	BC	133	2212	5042	3306	2779	1603	15075
5	A	3890	1497	518	243	262	0	6410
	BC	216	3056	4611	1691	2604	1401	13579
6	A	2959	2410	546	110	127	84	6236
	BC	418	5448	3192	979	972	294	11303
7	A	3213	2186	522	100	160	58	6239
	BC	474	3787	2959	869	807	419	9315
8	A	3165	1967	555	112	93	100	5992
	BC	593	3381	2302	1025	1034	524	8859
9	A	2686	2282	926	264	304	120	6582
	BC	103	2157	2123	1435	1174	931	7923
10	A	1857	2151	656	183	137	129	5113
	BC	111	2327	2329	837	1055	831	7490
11	A	1346	1595	391	193	238	55	3818
	BC	159	2130	2002	814	1041	691	6837
12	A	798	874	254	92	70	31	2119
	BC	293	1991	1461	306	741	475	5267

Table 1.3.1.2.2.2

Proficiency Level by Grade (Percent): Reading, S601 Paper

Grade	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	78.60%	2.50%	6.60%	4.40%	7.90%	0.00%	100.00%
1	A	45.00%	33.90%	11.60%	4.60%	2.90%	2.00%	100.00%
	BC	1.20%	24.20%	40.80%	11.30%	12.40%	10.20%	100.00%
2	A	58.10%	25.10%	9.30%	2.50%	4.10%	1.00%	100.00%
	BC	6.90%	29.60%	26.60%	10.50%	14.60%	11.80%	100.00%
3	A	51.80%	30.30%	11.00%	2.40%	3.30%	1.20%	100.00%
	BC	0.70%	10.70%	44.60%	18.10%	17.50%	8.40%	100.00%
4	A	56.00%	29.40%	6.60%	2.70%	4.70%	0.70%	100.00%
	BC	0.90%	14.70%	33.40%	21.90%	18.40%	10.60%	100.00%
5	A	60.70%	23.40%	8.10%	3.80%	4.10%	0.00%	100.00%
	BC	1.60%	22.50%	34.00%	12.50%	19.20%	10.30%	100.00%
6	A	47.50%	38.60%	8.80%	1.80%	2.00%	1.30%	100.00%
	BC	3.70%	48.20%	28.20%	8.70%	8.60%	2.60%	100.00%
7	A	51.50%	35.00%	8.40%	1.60%	2.60%	0.90%	100.00%
	BC	5.10%	40.70%	31.80%	9.30%	8.70%	4.50%	100.00%
8	A	52.80%	32.80%	9.30%	1.90%	1.60%	1.70%	100.00%
	BC	6.70%	38.20%	26.00%	11.60%	11.70%	5.90%	100.00%
9	A	40.80%	34.70%	14.10%	4.00%	4.60%	1.80%	100.00%
	BC	1.30%	27.20%	26.80%	18.10%	14.80%	11.80%	100.00%
10	A	36.30%	42.10%	12.80%	3.60%	2.70%	2.50%	100.00%
	BC	1.50%	31.10%	31.10%	11.20%	14.10%	11.10%	100.00%
11	A	35.30%	41.80%	10.20%	5.10%	6.20%	1.40%	100.00%
	BC	2.30%	31.20%	29.30%	11.90%	15.20%	10.10%	100.00%
12	A	37.70%	41.20%	12.00%	4.30%	3.30%	1.50%	100.00%
	BC	5.60%	37.80%	27.70%	5.80%	14.10%	9.00%	100.00%

1.3.1.2.3   By Grade
Table 1.3.1.2.3.1

Proficiency Level by Grade (Count): Reading, S601 Paper

Grade	Reading Proficiency Range						Total
	1	2	3	4	5	6	
K	194858	6164	16369	10986	19588	0	247965
1	8350	9104	7049	2202	2027	1590	30322
2	7106	8051	5923	2215	3150	2308	28753
3	4421	4717	10106	3937	3888	1821	28890
4	3585	4023	5446	3471	3069	1646	21240
5	4106	4553	5129	1934	2866	1401	19989
6	3377	7858	3738	1089	1099	378	17539
7	3687	5973	3481	969	967	477	15554
8	3758	5348	2857	1137	1127	624	14851
9	2789	4439	3049	1699	1478	1051	14505
10	1968	4478	2985	1020	1192	960	12603
11	1505	3725	2393	1007	1279	746	10655
12	1091	2865	1715	398	811	506	7386

Table 1.3.1.2.3.2

Proficiency Level by Grade (Percent): Reading, S601 Paper

Grade	Reading Proficiency Range						Total
	1	2	3	4	5	6	
K	78.60%	2.50%	6.60%	4.40%	7.90%	0.00%	100.00%
1	27.50%	30.00%	23.20%	7.30%	6.70%	5.20%	100.00%
2	24.70%	28.00%	20.60%	7.70%	11.00%	8.00%	100.00%
3	15.30%	16.30%	35.00%	13.60%	13.50%	6.30%	100.00%
4	16.90%	18.90%	25.60%	16.30%	14.40%	7.70%	100.00%
5	20.50%	22.80%	25.70%	9.70%	14.30%	7.00%	100.00%
6	19.30%	44.80%	21.30%	6.20%	6.30%	2.20%	100.00%
7	23.70%	38.40%	22.40%	6.20%	6.20%	3.10%	100.00%
8	25.30%	36.00%	19.20%	7.70%	7.60%	4.20%	100.00%
9	19.20%	30.60%	21.00%	11.70%	10.20%	7.20%	100.00%
10	15.60%	35.50%	23.70%	8.10%	9.50%	7.60%	100.00%
11	14.10%	35.00%	22.50%	9.50%	12.00%	7.00%	100.00%
12	14.80%	38.80%	23.20%	5.40%	11.00%	6.90%	100.00%

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, S601 Paper

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	177909	39391	24089	6573	0	0	247962
1	A	13058	9815	666	0	0	0	23539
	BC	3019	6172	6411	185	2	0	15789
2	A	6389	3943	1897	1	0	0	12230
	BC	2344	5952	13218	1373	10	0	22897
3	A	4735	3463	1632	9	0	0	9839
	BC	1470	3991	17455	2170	12	0	25098
4–5	A	5453	3870	5014	62	0	0	14399
	BC	495	1278	20296	10456	316	8	32849
6–8	A	10647	6501	3363	64	0	0	20575
	BC	1489	3198	22516	5692	20	0	32915
9–12	A	7756	6193	4794	346	0	0	19089
	BC	1793	2781	17516	9078	108	0	31276

Table 1.3.1.3.1.2

Proficiency Level by Cluster (Percent): Writing, S601 Paper

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	71.7%	15.9%	9.7%	2.7%	0.0%	0.0%	100.0%
1	A	55.5%	41.7%	2.8%	0.0%	0.0%	0.0%	100.0%
	BC	19.1%	39.1%	40.6%	1.2%	0.0%	0.0%	100.0%
2	A	52.2%	32.2%	15.5%	0.0%	0.0%	0.0%	100.0%
	BC	10.2%	26.0%	57.7%	6.0%	0.0%	0.0%	100.0%
3	A	48.1%	35.2%	16.6%	0.1%	0.0%	0.0%	100.0%
	BC	5.9%	15.9%	69.5%	8.6%	0.0%	0.0%	100.0%
4–5	A	37.9%	26.9%	34.8%	0.4%	0.0%	0.0%	100.0%
	BC	1.5%	3.9%	61.8%	31.8%	1.0%	0.0%	100.0%
6–8	A	51.7%	31.6%	16.3%	0.3%	0.0%	0.0%	100.0%
	BC	4.5%	9.7%	68.4%	17.3%	0.1%	0.0%	100.0%
9–12	A	40.6%	32.4%	25.1%	1.8%	0.0%	0.0%	100.0%
	BC	5.7%	8.9%	56.0%	29.0%	0.3%	0.0%	100.0%

1.3.1.3.2 By Grade by Tier

Table 1.3.1.3.2.1

Proficiency Level by Grade (Count): Writing, S601 Paper

Grade	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	177909	39391	24089	6573	0	0	247962
1	A	13058	9815	666	0	0	0	23539
	BC	3019	6172	6411	185	2	0	15789
2	A	6389	3943	1897	1	0	0	12230
	BC	2344	5952	13218	1373	10	0	22897
3	A	4735	3463	1632	9	0	0	9839
	BC	1470	3991	17455	2170	12	0	25098
4	A	2986	1903	2273	26	0	0	7188
	BC	239	685	11660	4786	152	3	17525
5	A	2467	1967	2741	36	0	0	7211
	BC	256	593	8636	5670	164	5	15324
6	A	3408	2283	1277	16	0	0	6984
	BC	490	1423	8636	2193	10	0	12752
7	A	3574	2415	951	25	0	0	6965
	BC	442	1093	7102	1736	5	0	10378
8	A	3665	1803	1135	23	0	0	6626
	BC	557	682	6778	1763	5	0	9785
9	A	2710	2315	1970	202	0	0	7197
	BC	281	581	5089	3147	66	0	9164
10	A	2145	1999	1282	76	0	0	5502
	BC	467	715	4687	2636	23	0	8528
11	A	1740	1431	925	41	0	0	4137
	BC	440	772	4195	2354	13	0	7774
12	A	1161	448	617	27	0	0	2253
	BC	605	713	3545	941	6	0	5810

Table 1.3.1.3.2.2

Proficiency Level by Grade (Percent): Writing, S601 Paper

Grade	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	71.70%	15.90%	9.70%	2.70%	0.00%	0.00%	100.00%
1	A	55.50%	41.70%	2.80%	0.00%	0.00%	0.00%	100.00%
	BC	19.10%	39.10%	40.60%	1.20%	0.00%	0.00%	100.00%
2	A	52.20%	32.20%	15.50%	0.00%	0.00%	0.00%	100.00%
	BC	10.20%	26.00%	57.70%	6.00%	0.00%	0.00%	100.00%
3	A	48.10%	35.20%	16.60%	0.10%	0.00%	0.00%	100.00%
	BC	5.90%	15.90%	69.50%	8.60%	0.00%	0.00%	100.00%
4	A	41.50%	26.50%	31.60%	0.40%	0.00%	0.00%	100.00%
	BC	1.40%	3.90%	66.50%	27.30%	0.90%	0.00%	100.00%
5	A	34.20%	27.30%	38.00%	0.50%	0.00%	0.00%	100.00%
	BC	1.70%	3.90%	56.40%	37.00%	1.10%	0.00%	100.00%
6	A	48.80%	32.70%	18.30%	0.20%	0.00%	0.00%	100.00%
	BC	3.80%	11.20%	67.70%	17.20%	0.10%	0.00%	100.00%
7	A	51.30%	34.70%	13.70%	0.40%	0.00%	0.00%	100.00%
	BC	4.30%	10.50%	68.40%	16.70%	0.00%	0.00%	100.00%
8	A	55.30%	27.20%	17.10%	0.30%	0.00%	0.00%	100.00%
	BC	5.70%	7.00%	69.30%	18.00%	0.10%	0.00%	100.00%
9	A	37.70%	32.20%	27.40%	2.80%	0.00%	0.00%	100.00%
	BC	3.10%	6.30%	55.50%	34.30%	0.70%	0.00%	100.00%
10	A	39.00%	36.30%	23.30%	1.40%	0.00%	0.00%	100.00%
	BC	5.50%	8.40%	55.00%	30.90%	0.30%	0.00%	100.00%
11	A	42.10%	34.60%	22.40%	1.00%	0.00%	0.00%	100.00%
	BC	5.70%	9.90%	54.00%	30.30%	0.20%	0.00%	100.00%
12	A	51.50%	19.90%	27.40%	1.20%	0.00%	0.00%	100.00%
	BC	10.40%	12.30%	61.00%	16.20%	0.10%	0.00%	100.00%

1.3.1.3.3 By Grade

Table 1.3.1.3.3.1

Proficiency Level by Grade (Count): Writing, S601 Paper

Grade	Writing Proficiency Range						Total
	1	2	3	4	5	6	
K	177909	39391	24089	6573	0	0	247962
1	16077	15987	7077	185	2	0	39328
2	8733	9895	15115	1374	10	0	35127
3	6205	7454	19087	2179	12	0	34937
4	3225	2588	13933	4812	152	3	24713
5	2723	2560	11377	5706	164	5	22535
6	3898	3706	9913	2209	10	0	19736
7	4016	3508	8053	1761	5	0	17343
8	4222	2485	7913	1786	5	0	16411
9	2991	2896	7059	3349	66	0	16361
10	2612	2714	5969	2712	23	0	14030
11	2180	2203	5120	2395	13	0	11911
12	1766	1161	4162	968	6	0	8063

Table 1.3.1.3.3.2

Proficiency Level by Grade (Percent): Writing, S601 Paper

Grade	Writing Proficiency Range						Total
	1	2	3	4	5	6	
K	71.70%	15.90%	9.70%	2.70%	0.00%	0.00%	100.00%
1	40.90%	40.70%	18.00%	0.50%	0.00%	0.00%	100.00%
2	24.90%	28.20%	43.00%	3.90%	0.00%	0.00%	100.00%
3	17.80%	21.30%	54.60%	6.20%	0.00%	0.00%	100.00%
4	13.00%	10.50%	56.40%	19.50%	0.60%	0.00%	100.00%
5	12.10%	11.40%	50.50%	25.30%	0.70%	0.00%	100.00%
6	19.80%	18.80%	50.20%	11.20%	0.10%	0.00%	100.00%
7	23.20%	20.20%	46.40%	10.20%	0.00%	0.00%	100.00%
8	25.70%	15.10%	48.20%	10.90%	0.00%	0.00%	100.00%
9	18.30%	17.70%	43.10%	20.50%	0.40%	0.00%	100.00%
10	18.60%	19.30%	42.50%	19.30%	0.20%	0.00%	100.00%
11	18.30%	18.50%	43.00%	20.10%	0.10%	0.00%	100.00%
12	21.90%	14.40%	51.60%	12.00%	0.10%	0.00%	100.00%

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, S601 Paper

Cluster	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	79987	51124	16861	18218	23857	57909	247956
1	A	8382	7277	4362	2510	836	0	23367
	BC	571	3350	5212	4734	1355	441	15663
2	A	5765	2578	2890	660	268	0	12161
	BC	1328	4611	9314	5140	1552	784	22729
3	A	5672	1962	1402	736	0	0	9772
	BC	1382	5099	10676	5475	1097	1214	24943
4–5	A	9179	2661	1354	896	207	0	14297
	BC	867	3056	8699	12657	4944	2464	32687
6–8	A	12664	2786	2895	1480	415	144	20384
	BC	1603	4637	8583	11247	3915	2692	32677
9–12	A	12891	2073	2858	971	103	0	18896
	BC	3421	4387	10644	7393	2163	3029	31037

Table 1.3.1.4.1.2

Proficiency Level by Cluster (Percent): Speaking, S601 Paper

Cluster	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	32.3%	20.6%	6.8%	7.3%	9.6%	23.4%	100.0%
1	A	35.9%	31.1%	18.7%	10.7%	3.6%	0.0%	100.0%
	BC	3.6%	21.4%	33.3%	30.2%	8.7%	2.8%	100.0%
2	A	47.4%	21.2%	23.8%	5.4%	2.2%	0.0%	100.0%
	BC	5.8%	20.3%	41.0%	22.6%	6.8%	3.4%	100.0%
3	A	58.0%	20.1%	14.3%	7.5%	0.0%	0.0%	100.0%
	BC	5.5%	20.4%	42.8%	22.0%	4.4%	4.9%	100.0%
4–5	A	64.2%	18.6%	9.5%	6.3%	1.4%	0.0%	100.0%
	BC	2.7%	9.3%	26.6%	38.7%	15.1%	7.5%	100.0%
6–8	A	62.1%	13.7%	14.2%	7.3%	2.0%	0.7%	100.0%
	BC	4.9%	14.2%	26.3%	34.4%	12.0%	8.2%	100.0%
9–12	A	68.2%	11.0%	15.1%	5.1%	0.5%	0.0%	100.0%
	BC	11.0%	14.1%	34.3%	23.8%	7.0%	9.8%	100.0%

1.3.1.4.2 By Grade by Tier
Table 1.3.1.4.2.1

Proficiency Level by Grade (Count): Speaking, S601 Paper

Grade	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	79987	51124	16861	18218	23857	57909	247956
1	A	8382	7277	4362	2510	836	0	23367
	BC	571	3350	5212	4734	1355	441	15663
2	A	5765	2578	2890	660	268	0	12161
	BC	1328	4611	9314	5140	1552	784	22729
3	A	5672	1962	1402	736	0	0	9772
	BC	1382	5099	10676	5475	1097	1214	24943
4	A	4367	1544	680	398	139	0	7128
	BC	372	1505	4533	6898	2619	1502	17429
5	A	4812	1117	674	498	68	0	7169
	BC	495	1551	4166	5759	2325	962	15258
6	A	4167	1180	856	475	179	72	6929
	BC	389	2091	3456	4228	1533	954	12651
7	A	4122	1113	872	630	78	72	6887
	BC	583	1296	2608	3882	959	987	10315
8	A	4375	493	1167	375	158	0	6568
	BC	631	1250	2519	3137	1423	751	9711
9	A	5238	606	936	228	103	0	7111
	BC	712	1428	2849	2181	1050	880	9100
10	A	3835	533	788	299	0	0	5455
	BC	1113	1143	2587	2505	397	706	8451
11	A	2587	458	759	288	0	0	4092
	BC	819	957	2999	1719	419	810	7723
12	A	1231	476	375	156	0	0	2238
	BC	777	859	2209	988	297	633	5763

Table 1.3.1.4.2.2

Proficiency Level by Grade (Percent): Speaking, S601 Paper

Grade	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	32.30%	20.60%	6.80%	7.30%	9.60%	23.40%	100.00%
1	A	35.90%	31.10%	18.70%	10.70%	3.60%	0.00%	100.00%
	BC	3.60%	21.40%	33.30%	30.20%	8.70%	2.80%	100.00%
2	A	47.40%	21.20%	23.80%	5.40%	2.20%	0.00%	100.00%
	BC	5.80%	20.30%	41.00%	22.60%	6.80%	3.40%	100.00%
3	A	58.00%	20.10%	14.30%	7.50%	0.00%	0.00%	100.00%
	BC	5.50%	20.40%	42.80%	22.00%	4.40%	4.90%	100.00%
4	A	61.30%	21.70%	9.50%	5.60%	2.00%	0.00%	100.00%
	BC	2.10%	8.60%	26.00%	39.60%	15.00%	8.60%	100.00%
5	A	67.10%	15.60%	9.40%	6.90%	0.90%	0.00%	100.00%
	BC	3.20%	10.20%	27.30%	37.70%	15.20%	6.30%	100.00%
6	A	60.10%	17.00%	12.40%	6.90%	2.60%	1.00%	100.00%
	BC	3.10%	16.50%	27.30%	33.40%	12.10%	7.50%	100.00%
7	A	59.90%	16.20%	12.70%	9.10%	1.10%	1.00%	100.00%
	BC	5.70%	12.60%	25.30%	37.60%	9.30%	9.60%	100.00%
8	A	66.60%	7.50%	17.80%	5.70%	2.40%	0.00%	100.00%
	BC	6.50%	12.90%	25.90%	32.30%	14.70%	7.70%	100.00%
9	A	73.70%	8.50%	13.20%	3.20%	1.40%	0.00%	100.00%
	BC	7.80%	15.70%	31.30%	24.00%	11.50%	9.70%	100.00%
10	A	70.30%	9.80%	14.40%	5.50%	0.00%	0.00%	100.00%
	BC	13.20%	13.50%	30.60%	29.60%	4.70%	8.40%	100.00%
11	A	63.20%	11.20%	18.50%	7.00%	0.00%	0.00%	100.00%
	BC	10.60%	12.40%	38.80%	22.30%	5.40%	10.50%	100.00%
12	A	55.00%	21.30%	16.80%	7.00%	0.00%	0.00%	100.00%
	BC	13.50%	14.90%	38.30%	17.10%	5.20%	11.00%	100.00%

1.3.1.4.3   **By Grade**
Table 1.3.1.4.3.1

Proficiency Level by Grade (Count): Speaking, S601 Paper

Grade	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
K	79987	51124	16861	18218	23857	57909	247956
1	8953	10627	9574	7244	2191	441	39030
2	7093	7189	12204	5800	1820	784	34890
3	7054	7061	12078	6211	1097	1214	34715
4	4739	3049	5213	7296	2758	1502	24557
5	5307	2668	4840	6257	2393	962	22427
6	4556	3271	4312	4703	1712	1026	19580
7	4705	2409	3480	4512	1037	1059	17202
8	5006	1743	3686	3512	1581	751	16279
9	5950	2034	3785	2409	1153	880	16211
10	4948	1676	3375	2804	397	706	13906
11	3406	1415	3758	2007	419	810	11815
12	2008	1335	2584	1144	297	633	8001

Table 1.3.1.4.3.2

Proficiency Level by Grade (Percent): Speaking, S601 Paper

Grade	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
K	32.30%	20.60%	6.80%	7.30%	9.60%	23.40%	100.00%
1	22.90%	27.20%	24.50%	18.60%	5.60%	1.10%	100.00%
2	20.30%	20.60%	35.00%	16.60%	5.20%	2.20%	100.00%
3	20.30%	20.30%	34.80%	17.90%	3.20%	3.50%	100.00%
4	19.30%	12.40%	21.20%	29.70%	11.20%	6.10%	100.00%
5	23.70%	11.90%	21.60%	27.90%	10.70%	4.30%	100.00%
6	23.30%	16.70%	22.00%	24.00%	8.70%	5.20%	100.00%
7	27.40%	14.00%	20.20%	26.20%	6.00%	6.20%	100.00%
8	30.80%	10.70%	22.60%	21.60%	9.70%	4.60%	100.00%
9	36.70%	12.50%	23.30%	14.90%	7.10%	5.40%	100.00%
10	35.60%	12.10%	24.30%	20.20%	2.90%	5.10%	100.00%
11	28.80%	12.00%	31.80%	17.00%	3.50%	6.90%	100.00%
12	25.10%	16.70%	32.30%	14.30%	3.70%	7.90%	100.00%

1.3.2 Composites

Performance of composites is observed in their percentage in PL5 and 6 in grades: Comprehension (2-24%), Oral (2-20%), Overall (0-6%), and Literacy (0-4%). 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, S601 Paper

Cluster	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	87630	35293	27104	20312	35826	41790	247955
1	A	4197	4318	5926	3480	1409	61	19391
	BC	201	1241	4169	4095	3225	879	13810
2	A	3960	2445	2855	1242	278	0	10780
	BC	224	1833	6708	7948	3495	1081	21289
3	A	3539	2353	1879	953	207	10	8941
	BC	84	1581	7869	9217	3821	900	23472
4–5	A	6408	3240	1973	1294	342	31	13288
	BC	147	1111	5905	13107	8337	3058	31665
6–8	A	9607	4915	2629	1408	334	35	18928
	BC	304	2041	8029	12316	6312	2201	31203
9–12	A	10872	3417	2397	813	49	0	17548
	BC	1020	3595	10024	10186	3494	1065	29384

Table 1.3.2.1.1.2

Proficiency Level by Cluster (Percent): Oral, S601 Paper

Cluster	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	35.3%	14.2%	10.9%	8.2%	14.4%	16.9%	100.0%
1	A	21.6%	22.3%	30.6%	17.9%	7.3%	0.3%	100.0%
	BC	1.5%	9.0%	30.2%	29.7%	23.4%	6.4%	100.0%
2	A	36.7%	22.7%	26.5%	11.5%	2.6%	0.0%	100.0%
	BC	1.1%	8.6%	31.5%	37.3%	16.4%	5.1%	100.0%
3	A	39.6%	26.3%	21.0%	10.7%	2.3%	0.1%	100.0%
	BC	0.4%	6.7%	33.5%	39.3%	16.3%	3.8%	100.0%
4–5	A	48.2%	24.4%	14.8%	9.7%	2.6%	0.2%	100.0%
	BC	0.5%	3.5%	18.6%	41.4%	26.3%	9.7%	100.0%
6–8	A	50.8%	26.0%	13.9%	7.4%	1.8%	0.2%	100.0%
	BC	1.0%	6.5%	25.7%	39.5%	20.2%	7.1%	100.0%
9–12	A	62.0%	19.5%	13.7%	4.6%	0.3%	0.0%	100.0%
	BC	3.5%	12.2%	34.1%	34.7%	11.9%	3.6%	100.0%

1.3.2.1.2 By Grade by Tier
Table 1.3.2.1.2.1

Proficiency Level by Grade (Count): Oral, S601 Paper

Grade	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	87630	35293	27104	20312	35826	41790	247955
1	A	4197	4318	5926	3480	1409	61	19391
	BC	201	1241	4169	4095	3225	879	13810
2	A	3960	2445	2855	1242	278	0	10780
	BC	224	1833	6708	7948	3495	1081	21289
3	A	3539	2353	1879	953	207	10	8941
	BC	84	1581	7869	9217	3821	900	23472
4	A	3050	1702	978	668	179	31	6608
	BC	61	473	3147	6777	4548	1808	16814
5	A	3358	1538	995	626	163	0	6680
	BC	86	638	2758	6330	3789	1250	14851
6	A	2939	1855	987	483	154	21	6439
	BC	86	657	3167	4784	2467	925	12086
7	A	3385	1608	841	468	95	14	6411
	BC	98	740	2478	3921	1943	678	9858
8	A	3283	1452	801	457	85	0	6078
	BC	120	644	2384	3611	1902	598	9259
9	A	4081	1307	816	316	22	0	6542
	BC	128	879	2609	3252	1363	380	8611
10	A	3296	921	648	203	23	0	5091
	BC	305	1086	2710	2675	902	307	7985
11	A	2289	768	566	168	4	0	3795
	BC	300	899	2593	2470	834	225	7321
12	A	1206	421	367	126	0	0	2120
	BC	287	731	2112	1789	395	153	5467

Table 1.3.2.1.2.2

Proficiency Level by Grade (Percent): Oral, S601 Paper

Grade	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	35.30%	14.20%	10.90%	8.20%	14.40%	16.90%	100.00%
1	A	21.60%	22.30%	30.60%	17.90%	7.30%	0.30%	100.00%
	BC	1.50%	9.00%	30.20%	29.70%	23.40%	6.40%	100.00%
2	A	36.70%	22.70%	26.50%	11.50%	2.60%	0.00%	100.00%
	BC	1.10%	8.60%	31.50%	37.30%	16.40%	5.10%	100.00%
3	A	39.60%	26.30%	21.00%	10.70%	2.30%	0.10%	100.00%
	BC	0.40%	6.70%	33.50%	39.30%	16.30%	3.80%	100.00%
4	A	46.20%	25.80%	14.80%	10.10%	2.70%	0.50%	100.00%
	BC	0.40%	2.80%	18.70%	40.30%	27.00%	10.80%	100.00%
5	A	50.30%	23.00%	14.90%	9.40%	2.40%	0.00%	100.00%
	BC	0.60%	4.30%	18.60%	42.60%	25.50%	8.40%	100.00%
6	A	45.60%	28.80%	15.30%	7.50%	2.40%	0.30%	100.00%
	BC	0.70%	5.40%	26.20%	39.60%	20.40%	7.70%	100.00%
7	A	52.80%	25.10%	13.10%	7.30%	1.50%	0.20%	100.00%
	BC	1.00%	7.50%	25.10%	39.80%	19.70%	6.90%	100.00%
8	A	54.00%	23.90%	13.20%	7.50%	1.40%	0.00%	100.00%
	BC	1.30%	7.00%	25.70%	39.00%	20.50%	6.50%	100.00%
9	A	62.40%	20.00%	12.50%	4.80%	0.30%	0.00%	100.00%
	BC	1.50%	10.20%	30.30%	37.80%	15.80%	4.40%	100.00%
10	A	64.70%	18.10%	12.70%	4.00%	0.50%	0.00%	100.00%
	BC	3.80%	13.60%	33.90%	33.50%	11.30%	3.80%	100.00%
11	A	60.30%	20.20%	14.90%	4.40%	0.10%	0.00%	100.00%
	BC	4.10%	12.30%	35.40%	33.70%	11.40%	3.10%	100.00%
12	A	56.90%	19.90%	17.30%	5.90%	0.00%	0.00%	100.00%
	BC	5.20%	13.40%	38.60%	32.70%	7.20%	2.80%	100.00%

1.3.2.1.3   By Grade
Table 1.3.2.1.3.1

Proficiency Level by Grade (Count): Oral, S601 Paper

Grade	Oral Language Proficiency Range						Total
	1	2	3	4	5	6	
K	87630	35293	27104	20312	35826	41790	247955
1	4398	5559	10095	7575	4634	940	33201
2	4184	4278	9563	9190	3773	1081	32069
3	3623	3934	9748	10170	4028	910	32413
4	3111	2175	4125	7445	4727	1839	23422
5	3444	2176	3753	6956	3952	1250	21531
6	3025	2512	4154	5267	2621	946	18525
7	3483	2348	3319	4389	2038	692	16269
8	3403	2096	3185	4068	1987	598	15337
9	4209	2186	3425	3568	1385	380	15153
10	3601	2007	3358	2878	925	307	13076
11	2589	1667	3159	2638	838	225	11116
12	1493	1152	2479	1915	395	153	7587

Table 1.3.2.1.3.2

Proficiency Level by Grade (Percent): Oral, S601 Paper

Grade	Oral Language Proficiency Range						Total
	1	2	3	4	5	6	
K	35.30%	14.20%	10.90%	8.20%	14.40%	16.90%	100.00%
1	13.20%	16.70%	30.40%	22.80%	14.00%	2.80%	100.00%
2	13.00%	13.30%	29.80%	28.70%	11.80%	3.40%	100.00%
3	11.20%	12.10%	30.10%	31.40%	12.40%	2.80%	100.00%
4	13.30%	9.30%	17.60%	31.80%	20.20%	7.90%	100.00%
5	16.00%	10.10%	17.40%	32.30%	18.40%	5.80%	100.00%
6	16.30%	13.60%	22.40%	28.40%	14.10%	5.10%	100.00%
7	21.40%	14.40%	20.40%	27.00%	12.50%	4.30%	100.00%
8	22.20%	13.70%	20.80%	26.50%	13.00%	3.90%	100.00%
9	27.80%	14.40%	22.60%	23.50%	9.10%	2.50%	100.00%
10	27.50%	15.30%	25.70%	22.00%	7.10%	2.30%	100.00%
11	23.30%	15.00%	28.40%	23.70%	7.50%	2.00%	100.00%
12	19.70%	15.20%	32.70%	25.20%	5.20%	2.00%	100.00%

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, S601 Paper

Cluster	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	195005	25122	19619	8211	0	0	247957
1	A	8926	7470	1811	8	0	0	18215
	BC	1215	4157	5821	830	69	3	12095
2	A	5626	2850	1475	49	0	0	10000
	BC	1424	5089	9255	2808	163	6	18745
3	A	4226	2646	1319	65	0	0	8256
	BC	328	2806	14082	3239	159	17	20631
4–5	A	6104	3651	2632	184	0	0	12571
	BC	229	1344	15728	9969	1278	104	28652
6–8	A	9526	6329	2407	189	0	0	18451
	BC	723	5334	18635	4513	257	5	29467
9–12	A	6679	6556	3906	477	9	0	17627
	BC	741	3999	14179	7578	997	8	27502

Table 1.3.2.2.1.2

Proficiency Level by Cluster (Percent): Literacy, S601 Paper

Cluster	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	78.6%	10.1%	7.9%	3.3%	0.0%	0.0%	100.0%
1	A	49.0%	41.0%	9.9%	0.0%	0.0%	0.0%	100.0%
	BC	10.0%	34.4%	48.1%	6.9%	0.6%	0.0%	100.0%
2	A	56.3%	28.5%	14.8%	0.5%	0.0%	0.0%	100.0%
	BC	7.6%	27.1%	49.4%	15.0%	0.9%	0.0%	100.0%
3	A	51.2%	32.0%	16.0%	0.8%	0.0%	0.0%	100.0%
	BC	1.6%	13.6%	68.3%	15.7%	0.8%	0.1%	100.0%
4–5	A	48.6%	29.0%	20.9%	1.5%	0.0%	0.0%	100.0%
	BC	0.8%	4.7%	54.9%	34.8%	4.5%	0.4%	100.0%
6–8	A	51.6%	34.3%	13.0%	1.0%	0.0%	0.0%	100.0%
	BC	2.5%	18.1%	63.2%	15.3%	0.9%	0.0%	100.0%
9–12	A	37.9%	37.2%	22.2%	2.7%	0.1%	0.0%	100.0%
	BC	2.7%	14.5%	51.6%	27.6%	3.6%	0.0%	100.0%

1.3.2.2.2   By Grade by Tier

Table 1.3.2.2.2.1

Proficiency Level by Grade (Count): Literacy, S601 Paper

Grade	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	195005	25122	19619	8211	0	0	247957
1	A	8926	7470	1811	8	0	0	18215
	BC	1215	4157	5821	830	69	3	12095
2	A	5626	2850	1475	49	0	0	10000
	BC	1424	5089	9255	2808	163	6	18745
3	A	4226	2646	1319	65	0	0	8256
	BC	328	2806	14082	3239	159	17	20631
4	A	3040	1748	1289	86	0	0	6163
	BC	100	630	8862	4965	477	39	15073
5	A	3064	1903	1343	98	0	0	6408
	BC	129	714	6866	5004	801	65	13579
6	A	3041	2210	922	59	0	0	6232
	BC	224	2099	7395	1495	82	4	11299
7	A	3172	2250	757	59	0	0	6238
	BC	224	1735	5866	1406	80	1	9312
8	A	3313	1869	728	71	0	0	5981
	BC	275	1500	5374	1612	95	0	8856
9	A	2361	2450	1549	212	6	0	6578
	BC	93	821	4083	2564	354	4	7919
10	A	2014	1920	1028	149	2	0	5113
	BC	163	997	3909	2105	313	3	7490
11	A	1448	1425	858	86	1	0	3818
	BC	182	1000	3492	1926	230	1	6831
12	A	856	761	471	30	0	0	2118
	BC	303	1181	2695	983	100	0	5262

Table 1.3.2.2.2.2

Proficiency Level by Grade (Percent): Literacy, S601 Paper

Grade	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	78.60%	10.10%	7.90%	3.30%	0.00%	0.00%	100.00%
1	A	49.00%	41.00%	9.90%	0.00%	0.00%	0.00%	100.00%
	BC	10.00%	34.40%	48.10%	6.90%	0.60%	0.00%	100.00%
2	A	56.30%	28.50%	14.80%	0.50%	0.00%	0.00%	100.00%
	BC	7.60%	27.10%	49.40%	15.00%	0.90%	0.00%	100.00%
3	A	51.20%	32.00%	16.00%	0.80%	0.00%	0.00%	100.00%
	BC	1.60%	13.60%	68.30%	15.70%	0.80%	0.10%	100.00%
4	A	49.30%	28.40%	20.90%	1.40%	0.00%	0.00%	100.00%
	BC	0.70%	4.20%	58.80%	32.90%	3.20%	0.30%	100.00%
5	A	47.80%	29.70%	21.00%	1.50%	0.00%	0.00%	100.00%
	BC	0.90%	5.30%	50.60%	36.90%	5.90%	0.50%	100.00%
6	A	48.80%	35.50%	14.80%	0.90%	0.00%	0.00%	100.00%
	BC	2.00%	18.60%	65.40%	13.20%	0.70%	0.00%	100.00%
7	A	50.80%	36.10%	12.10%	0.90%	0.00%	0.00%	100.00%
	BC	2.40%	18.60%	63.00%	15.10%	0.90%	0.00%	100.00%
8	A	55.40%	31.20%	12.20%	1.20%	0.00%	0.00%	100.00%
	BC	3.10%	16.90%	60.70%	18.20%	1.10%	0.00%	100.00%
9	A	35.90%	37.20%	23.50%	3.20%	0.10%	0.00%	100.00%
	BC	1.20%	10.40%	51.60%	32.40%	4.50%	0.10%	100.00%
10	A	39.40%	37.60%	20.10%	2.90%	0.00%	0.00%	100.00%
	BC	2.20%	13.30%	52.20%	28.10%	4.20%	0.00%	100.00%
11	A	37.90%	37.30%	22.50%	2.30%	0.00%	0.00%	100.00%
	BC	2.70%	14.60%	51.10%	28.20%	3.40%	0.00%	100.00%
12	A	40.40%	35.90%	22.20%	1.40%	0.00%	0.00%	100.00%
	BC	5.80%	22.40%	51.20%	18.70%	1.90%	0.00%	100.00%

1.3.2.2.3 *By Grade*
Table 1.3.2.2.3.1

Proficiency Level by Grade (Count): Literacy, S601 Paper

Grade	Literacy Proficiency Range						Total
	1	2	3	4	5	6	
K	195005	25122	19619	8211	0	0	247957
1	10141	11627	7632	838	69	3	30310
2	7050	7939	10730	2857	163	6	28745
3	4554	5452	15401	3304	159	17	28887
4	3140	2378	10151	5051	477	39	21236
5	3193	2617	8209	5102	801	65	19987
6	3265	4309	8317	1554	82	4	17531
7	3396	3985	6623	1465	80	1	15550
8	3588	3369	6102	1683	95	0	14837
9	2454	3271	5632	2776	360	4	14497
10	2177	2917	4937	2254	315	3	12603
11	1630	2425	4350	2012	231	1	10649
12	1159	1942	3166	1013	100	0	7380

Table 1.3.2.2.3.2

Proficiency Level by Grade (Percent): Literacy, S601 Paper

Grade	Literacy Proficiency Range						Total
	1	2	3	4	5	6	
K	78.60%	10.10%	7.90%	3.30%	0.00%	0.00%	100.00%
1	33.50%	38.40%	25.20%	2.80%	0.20%	0.00%	100.00%
2	24.50%	27.60%	37.30%	9.90%	0.60%	0.00%	100.00%
3	15.80%	18.90%	53.30%	11.40%	0.60%	0.10%	100.00%
4	14.80%	11.20%	47.80%	23.80%	2.20%	0.20%	100.00%
5	16.00%	13.10%	41.10%	25.50%	4.00%	0.30%	100.00%
6	18.60%	24.60%	47.40%	8.90%	0.50%	0.00%	100.00%
7	21.80%	25.60%	42.60%	9.40%	0.50%	0.00%	100.00%
8	24.20%	22.70%	41.10%	11.30%	0.60%	0.00%	100.00%
9	16.90%	22.60%	38.80%	19.10%	2.50%	0.00%	100.00%
10	17.30%	23.10%	39.20%	17.90%	2.50%	0.00%	100.00%
11	15.30%	22.80%	40.80%	18.90%	2.20%	0.00%	100.00%
12	15.70%	26.30%	42.90%	13.70%	1.40%	0.00%	100.00%

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, S601 Paper

Cluster	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	176845	16975	18379	9257	21289	5218	247963
1	A	3385	5482	4728	1215	896	331	16037
	BC	26	697	3908	2131	2759	1563	11084
2	A	3411	3267	1453	499	539	58	9227
	BC	161	2757	5293	3032	4061	2591	17895
3	A	1828	3781	1134	372	399	193	7707
	BC	3	381	5451	5617	6069	2286	19807
4–5	A	4932	4051	1554	622	667	98	11924
	BC	24	1490	7394	6451	8160	4502	28021
6–8	A	7909	6544	1909	589	463	100	17514
	BC	158	5562	10872	5740	4562	1622	28516
9–12	A	6592	6681	2180	743	452	54	16702
	BC	263	5040	8568	5291	4806	2522	26490

Table 1.3.2.3.1.2

Proficiency Level by Cluster (Percent): Comprehension, S601 Paper

Cluster	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	71.3%	6.8%	7.4%	3.7%	8.6%	2.1%	100.0%
1	A	21.1%	34.2%	29.5%	7.6%	5.6%	2.1%	100.0%
	BC	0.2%	6.3%	35.3%	19.2%	24.9%	14.1%	100.0%
2	A	37.0%	35.4%	15.7%	5.4%	5.8%	0.6%	100.0%
	BC	0.9%	15.4%	29.6%	16.9%	22.7%	14.5%	100.0%
3	A	23.7%	49.1%	14.7%	4.8%	5.2%	2.5%	100.0%
	BC	0.0%	1.9%	27.5%	28.4%	30.6%	11.5%	100.0%
4–5	A	41.4%	34.0%	13.0%	5.2%	5.6%	0.8%	100.0%
	BC	0.1%	5.3%	26.4%	23.0%	29.1%	16.1%	100.0%
6–8	A	45.2%	37.4%	10.9%	3.4%	2.6%	0.6%	100.0%
	BC	0.6%	19.5%	38.1%	20.1%	16.0%	5.7%	100.0%
9–12	A	39.5%	40.0%	13.1%	4.4%	2.7%	0.3%	100.0%
	BC	1.0%	19.0%	32.3%	20.0%	18.1%	9.5%	100.0%

1.3.2.3.2 By Grade by Tier
Table 1.3.2.3.2.1

Proficiency Level by Grade (Count): Comprehension, S601 Paper

Grade	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	176845	16975	18379	9257	21289	5218	247963
1	A	3385	5482	4728	1215	896	331	16037
	BC	26	697	3908	2131	2759	1563	11084
2	A	3411	3267	1453	499	539	58	9227
	BC	161	2757	5293	3032	4061	2591	17895
3	A	1828	3781	1134	372	399	193	7707
	BC	3	381	5451	5617	6069	2286	19807
4	A	2219	2128	799	275	334	80	5835
	BC	10	438	3859	3482	4585	2327	14701
5	A	2713	1923	755	347	333	18	6089
	BC	14	1052	3535	2969	3575	2175	13320
6	A	2278	2527	703	219	135	54	5916
	BC	26	2045	4603	2248	1500	497	10919
7	A	2819	2146	605	189	139	35	5933
	BC	36	1815	3440	1776	1412	549	9028
8	A	2812	1871	601	181	189	11	5665
	BC	96	1702	2829	1716	1650	576	8569
9	A	2319	2567	821	278	174	32	6191
	BC	23	997	2427	1729	1668	790	7634
10	A	1967	1995	557	198	132	15	4864
	BC	45	1259	2449	1508	1194	746	7201
11	A	1479	1361	492	166	105	7	3610
	BC	75	1337	2136	1128	1247	662	6585
12	A	827	758	310	101	41	0	2037
	BC	120	1447	1556	926	697	324	5070

Table 1.3.2.3.2.2

Proficiency Level by Grade (Percent): Comprehension, S601 Paper

Grade	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	71.30%	6.80%	7.40%	3.70%	8.60%	2.10%	100.00%
1	A	21.10%	34.20%	29.50%	7.60%	5.60%	2.10%	100.00%
	BC	0.20%	6.30%	35.30%	19.20%	24.90%	14.10%	100.00%
2	A	37.00%	35.40%	15.70%	5.40%	5.80%	0.60%	100.00%
	BC	0.90%	15.40%	29.60%	16.90%	22.70%	14.50%	100.00%
3	A	23.70%	49.10%	14.70%	4.80%	5.20%	2.50%	100.00%
	BC	0.00%	1.90%	27.50%	28.40%	30.60%	11.50%	100.00%
4	A	38.00%	36.50%	13.70%	4.70%	5.70%	1.40%	100.00%
	BC	0.10%	3.00%	26.20%	23.70%	31.20%	15.80%	100.00%
5	A	44.60%	31.60%	12.40%	5.70%	5.50%	0.30%	100.00%
	BC	0.10%	7.90%	26.50%	22.30%	26.80%	16.30%	100.00%
6	A	38.50%	42.70%	11.90%	3.70%	2.30%	0.90%	100.00%
	BC	0.20%	18.70%	42.20%	20.60%	13.70%	4.60%	100.00%
7	A	47.50%	36.20%	10.20%	3.20%	2.30%	0.60%	100.00%
	BC	0.40%	20.10%	38.10%	19.70%	15.60%	6.10%	100.00%
8	A	49.60%	33.00%	10.60%	3.20%	3.30%	0.20%	100.00%
	BC	1.10%	19.90%	33.00%	20.00%	19.30%	6.70%	100.00%
9	A	37.50%	41.50%	13.30%	4.50%	2.80%	0.50%	100.00%
	BC	0.30%	13.10%	31.80%	22.60%	21.80%	10.30%	100.00%
10	A	40.40%	41.00%	11.50%	4.10%	2.70%	0.30%	100.00%
	BC	0.60%	17.50%	34.00%	20.90%	16.60%	10.40%	100.00%
11	A	41.00%	37.70%	13.60%	4.60%	2.90%	0.20%	100.00%
	BC	1.10%	20.30%	32.40%	17.10%	18.90%	10.10%	100.00%
12	A	40.60%	37.20%	15.20%	5.00%	2.00%	0.00%	100.00%
	BC	2.40%	28.50%	30.70%	18.30%	13.70%	6.40%	100.00%

1.3.2.3.3 By Grade
Table 1.3.2.3.3.1

Proficiency Level by Grade (Count): Comprehension, S601 Paper

Grade	Comprehension Proficiency Range						Total
	1	2	3	4	5	6	
K	176845	16975	18379	9257	21289	5218	247963
1	3411	6179	8636	3346	3655	1894	27121
2	3572	6024	6746	3531	4600	2649	27122
3	1831	4162	6585	5989	6468	2479	27514
4	2229	2566	4658	3757	4919	2407	20536
5	2727	2975	4290	3316	3908	2193	19409
6	2304	4572	5306	2467	1635	551	16835
7	2855	3961	4045	1965	1551	584	14961
8	2908	3573	3430	1897	1839	587	14234
9	2342	3564	3248	2007	1842	822	13825
10	2012	3254	3006	1706	1326	761	12065
11	1554	2698	2628	1294	1352	669	10195
12	947	2205	1866	1027	738	324	7107

Table 1.3.2.3.3.2

Proficiency Level by Grade (Percent): Comprehension, S601 Paper

Grade	Comprehension Proficiency Range						Total
	1	2	3	4	5	6	
K	71.30%	6.80%	7.40%	3.70%	8.60%	2.10%	100.00%
1	12.60%	22.80%	31.80%	12.30%	13.50%	7.00%	100.00%
2	13.20%	22.20%	24.90%	13.00%	17.00%	9.80%	100.00%
3	6.70%	15.10%	23.90%	21.80%	23.50%	9.00%	100.00%
4	10.90%	12.50%	22.70%	18.30%	24.00%	11.70%	100.00%
5	14.10%	15.30%	22.10%	17.10%	20.10%	11.30%	100.00%
6	13.70%	27.20%	31.50%	14.70%	9.70%	3.30%	100.00%
7	19.10%	26.50%	27.00%	13.10%	10.40%	3.90%	100.00%
8	20.40%	25.10%	24.10%	13.30%	12.90%	4.10%	100.00%
9	16.90%	25.80%	23.50%	14.50%	13.30%	5.90%	100.00%
10	16.70%	27.00%	24.90%	14.10%	11.00%	6.30%	100.00%
11	15.20%	26.50%	25.80%	12.70%	13.30%	6.60%	100.00%
12	13.30%	31.00%	26.30%	14.50%	10.40%	4.60%	100.00%

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, S601 Paper

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	158734	38318	29540	18340	3008	0	247940
1	A	4961	6570	4284	110	1	0	15926
	BC	651	1825	6550	1736	222	5	10989
2	A	4008	3213	1818	123	0	0	9162
	BC	441	3168	9479	4277	390	12	17767
3	A	3327	2681	1464	178	2	0	7652
	BC	132	1504	12216	5481	335	23	19691
4–5	A	5643	3347	2453	392	2	0	11837
	BC	154	731	10867	13756	2250	139	27897
6–8	A	8925	5284	2698	402	8	0	17317
	BC	310	2742	15414	9223	598	13	28300
9–12	A	7898	5001	3183	446	3	0	16531
	BC	521	3004	12749	8891	1073	12	26250

Table 1.3.2.4.1.2

Proficiency Level by Grade-Level Cluster (Percent): Overall, S601 Paper

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	64.0%	15.5%	11.9%	7.4%	1.2%	0.0%	100.0%
1	A	31.2%	41.3%	26.9%	0.7%	0.0%	0.0%	100.0%
	BC	5.9%	16.6%	59.6%	15.8%	2.0%	0.0%	100.0%
2	A	43.7%	35.1%	19.8%	1.3%	0.0%	0.0%	100.0%
	BC	2.5%	17.8%	53.4%	24.1%	2.2%	0.1%	100.0%
3	A	43.5%	35.0%	19.1%	2.3%	0.0%	0.0%	100.0%
	BC	0.7%	7.6%	62.0%	27.8%	1.7%	0.1%	100.0%
4–5	A	47.7%	28.3%	20.7%	3.3%	0.0%	0.0%	100.0%
	BC	0.6%	2.6%	39.0%	49.3%	8.1%	0.5%	100.0%
6–8	A	51.5%	30.5%	15.6%	2.3%	0.0%	0.0%	100.0%
	BC	1.1%	9.7%	54.5%	32.6%	2.1%	0.0%	100.0%
9–12	A	47.8%	30.3%	19.3%	2.7%	0.0%	0.0%	100.0%
	BC	2.0%	11.4%	48.6%	33.9%	4.1%	0.0%	100.0%

1.3.2.4.2 *By Grade by Tier*
Table 1.3.2.4.2.1

Proficiency Level by Grade (Count): Overall, S601 Paper

Grade	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	158734	38318	29540	18340	3008	0	247940
1	A	4961	6570	4284	110	1	0	15926
	BC	651	1825	6550	1736	222	5	10989
2	A	4008	3213	1818	123	0	0	9162
	BC	441	3168	9479	4277	390	12	17767
3	A	3327	2681	1464	178	2	0	7652
	BC	132	1504	12216	5481	335	23	19691
4	A	2739	1610	1233	205	1	0	5788
	BC	70	310	5920	7252	1014	63	14629
5	A	2904	1737	1220	187	1	0	6049
	BC	84	421	4947	6504	1236	76	13268
6	A	2771	1964	984	136	6	0	5861
	BC	90	954	6339	3257	186	7	10833
7	A	3122	1735	870	133	1	0	5861
	BC	97	942	4787	2931	208	2	8967
8	A	3032	1585	844	133	1	0	5595
	BC	123	846	4288	3035	204	4	8500
9	A	2849	1867	1204	193	1	0	6114
	BC	66	586	3451	3041	416	9	7569
10	A	2438	1427	821	136	1	0	4823
	BC	135	814	3461	2406	315	1	7132
11	A	1673	1067	753	79	1	0	3573
	BC	134	728	3238	2190	237	2	6529
12	A	938	640	405	38	0	0	2021
	BC	186	876	2599	1254	105	0	5020

Table 1.3.2.4.2.2

Proficiency Level by Grade (Percent): Overall, S601 Paper

| Grade | Tier | Overall Proficiency Range | | | | | | Total |
		1	2	3	4	5	6	
K	-	64.00%	15.50%	11.90%	7.40%	1.20%	0.00%	100.00%
1	A	31.20%	41.30%	26.90%	0.70%	0.00%	0.00%	100.00%
	BC	5.90%	16.60%	59.60%	15.80%	2.00%	0.00%	100.00%
2	A	43.70%	35.10%	19.80%	1.30%	0.00%	0.00%	100.00%
	BC	2.50%	17.80%	53.40%	24.10%	2.20%	0.10%	100.00%
3	A	43.50%	35.00%	19.10%	2.30%	0.00%	0.00%	100.00%
	BC	0.70%	7.60%	62.00%	27.80%	1.70%	0.10%	100.00%
4	A	47.30%	27.80%	21.30%	3.50%	0.00%	0.00%	100.00%
	BC	0.50%	2.10%	40.50%	49.60%	6.90%	0.40%	100.00%
5	A	48.00%	28.70%	20.20%	3.10%	0.00%	0.00%	100.00%
	BC	0.60%	3.20%	37.30%	49.00%	9.30%	0.60%	100.00%
6	A	47.30%	33.50%	16.80%	2.30%	0.10%	0.00%	100.00%
	BC	0.80%	8.80%	58.50%	30.10%	1.70%	0.10%	100.00%
7	A	53.30%	29.60%	14.80%	2.30%	0.00%	0.00%	100.00%
	BC	1.10%	10.50%	53.40%	32.70%	2.30%	0.00%	100.00%
8	A	54.20%	28.30%	15.10%	2.40%	0.00%	0.00%	100.00%
	BC	1.40%	10.00%	50.40%	35.70%	2.40%	0.00%	100.00%
9	A	46.60%	30.50%	19.70%	3.20%	0.00%	0.00%	100.00%
	BC	0.90%	7.70%	45.60%	40.20%	5.50%	0.10%	100.00%
10	A	50.50%	29.60%	17.00%	2.80%	0.00%	0.00%	100.00%
	BC	1.90%	11.40%	48.50%	33.70%	4.40%	0.00%	100.00%
11	A	46.80%	29.90%	21.10%	2.20%	0.00%	0.00%	100.00%
	BC	2.10%	11.20%	49.60%	33.50%	3.60%	0.00%	100.00%
12	A	46.40%	31.70%	20.00%	1.90%	0.00%	0.00%	100.00%
	BC	3.70%	17.50%	51.80%	25.00%	2.10%	0.00%	100.00%

1.3.2.4.3   By Grade
Table 1.3.2.4.3.1

Proficiency Level by Grade (Count): Overall, S601 Paper

Grade	Overall Proficiency Range						Total
	1	2	3	4	5	6	
K	158734	38318	29540	18340	3008	0	247940
1	5612	8395	10834	1846	223	5	26915
2	4449	6381	11297	4400	390	12	26929
3	3459	4185	13680	5659	337	23	27343
4	2809	1920	7153	7457	1015	63	20417
5	2988	2158	6167	6691	1237	76	19317
6	2861	2918	7323	3393	192	7	16694
7	3219	2677	5657	3064	209	2	14828
8	3155	2431	5132	3168	205	4	14095
9	2915	2453	4655	3234	417	9	13683
10	2573	2241	4282	2542	316	1	11955
11	1807	1795	3991	2269	238	2	10102
12	1124	1516	3004	1292	105	0	7041

Table 1.3.2.4.3.2

Proficiency Level by Grade (Percent): Overall, S601 Paper

Grade	Overall Proficiency Range						Total
	1	2	3	4	5	6	
K	64.00%	15.50%	11.90%	7.40%	1.20%	0.00%	100.00%
1	20.90%	31.20%	40.30%	6.90%	0.80%	0.00%	100.00%
2	16.50%	23.70%	42.00%	16.30%	1.40%	0.00%	100.00%
3	12.70%	15.30%	50.00%	20.70%	1.20%	0.10%	100.00%
4	13.80%	9.40%	35.00%	36.50%	5.00%	0.30%	100.00%
5	15.50%	11.20%	31.90%	34.60%	6.40%	0.40%	100.00%
6	17.10%	17.50%	43.90%	20.30%	1.20%	0.00%	100.00%
7	21.70%	18.10%	38.20%	20.70%	1.40%	0.00%	100.00%
8	22.40%	17.20%	36.40%	22.50%	1.50%	0.00%	100.00%
9	21.30%	17.90%	34.00%	23.60%	3.00%	0.10%	100.00%
10	21.50%	18.70%	35.80%	21.30%	2.60%	0.00%	100.00%
11	17.90%	17.80%	39.50%	22.50%	2.40%	0.00%	100.00%
12	16.00%	21.50%	42.70%	18.30%	1.50%	0.00%	100.00%

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), Pni1/Pni0 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 3.2.2, ratings on the ACCESS Writing Scoring Scale range from 0, 1, 1+,…, 6, and the possible raw scores range from 0 to 9. All Writing tasks are scored using this scoring scale except for Grade 1 Tier A Tasks 1 and 2. The profiles of the responses to these two tasks do not fit the generic scoring scale well, so additional task-specific instructions are provided to raters. These instructions guide raters in applying a limited number of score points on the scoring scale to responses elicited by these two tasks. The possible ratings for Grade 1 Tier A Task 1 are 0 or 1, and the possible ratings for Grade 1 Tier A Task 2 are 0, 1, 1+, or 2. To simplify the year-to-year linking process, the Grade 1 Writing Tier A Task 1 is treated as a

dichotomously scored task. The Grade 1 Writing Tier A Task 2 is modeled using a rating scale with a possible raw score of 0 to 3. All other Writing tasks are modeled using a rating scale with possible raw scores of 0 to 9. Thus, a total of two rating scales are modeled for ACCESS Writing. One rating scale is associated with the Grade 1 Writing Tier A Task 2, and the other rating scale is associated with all Writing tasks that are scored using the rating scale with raw score values of 0–9. We conducted a study in the summer of 2016 to reconstruct the logit scales. Detailed information about the derivation of the Writing rating scales as well as the psychometric properties of Writing rating scales are available in the scaling report (see Center for Applied Linguistics, 2017).

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

Scale Scores and Proficiency Level Scores

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

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

- L: (Ability Measure in Logits * 37.571) + 316.637

- R: (Ability Measure in Logits * 26.000) + 323.272

- W: (Ability Measure in Logits * 26.851) + 303.332

- S: (Ability Measure in Logits * 29.248) + 265.076

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

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

- L: (Ability Measure in Logits * 37.571) + 316.637

- R: (Ability Measure in Logits * 26.000) + 323.272

- W: (Ability Measure in Logits * 31.097) + 317.068

- S: (Ability Measure in Logits * 20.084) + 322.686

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

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

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

PL score of 5.8 for a Grade 2 student, a 3.8 for a Grade 5 student, a 3.1 for a Grade 8 student, and a 2.3 for a Grade 12 student.

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

2.1   Complete Item or Task Analysis and Summary

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

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

All Rasch analyses were conducted using the Rasch measurement software program *Winsteps 4.8.2.0* (Linacre, 2006). When speaking of the measure of student ability, we use the term *ability measure* (rather than *theta* used commonly when discussing models based on item response theory). When speaking of the measure of how hard an item is, we use the term *item difficulty measure* (rather than *b parameter* used commonly when discussing models based on item response theory). *Step measures* refer to the calibration of the steps in the Rasch rating scale model previously presented. All three measures (ability, difficulty, and step) are expressed in terms of Rasch logits, which are then converted into scores on the ACCESS score scale for reporting purposes.

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

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

- Values greater than 2.0 "distort or degrade[1] the measurement system."

- Values between 1.5 and 2.0 are "unproductive for construction of measurement, but not degrading."

- Values between 0.5 and 1.5 should be considered "productive for measurement."

- Values below 0.5 are "less productive for measurement, but not degrading."

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

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

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

---

[1] We interpret "degrade" here in the sense of lowering the quality of the measurement system.

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

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

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

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

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

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

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

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

All Listening and Reading Grades 1-12 items have infit mean square statistics and outfit mean square statistics between 0.5 and 1.5. The percentage of Listening Kindergarten items that have infit mean square statistics between 0.5 and 1.5 is 97% and the percentage of outfit mean square statistics between 0.5 and 1.5 is 70%. The percentages of Kindergarten Reading items with infit mean square statistics between 0.5 and 1.5 is 100%, while the percentage of those items with outfit mean square statistics between 0.5 and 1.5 is 37%.

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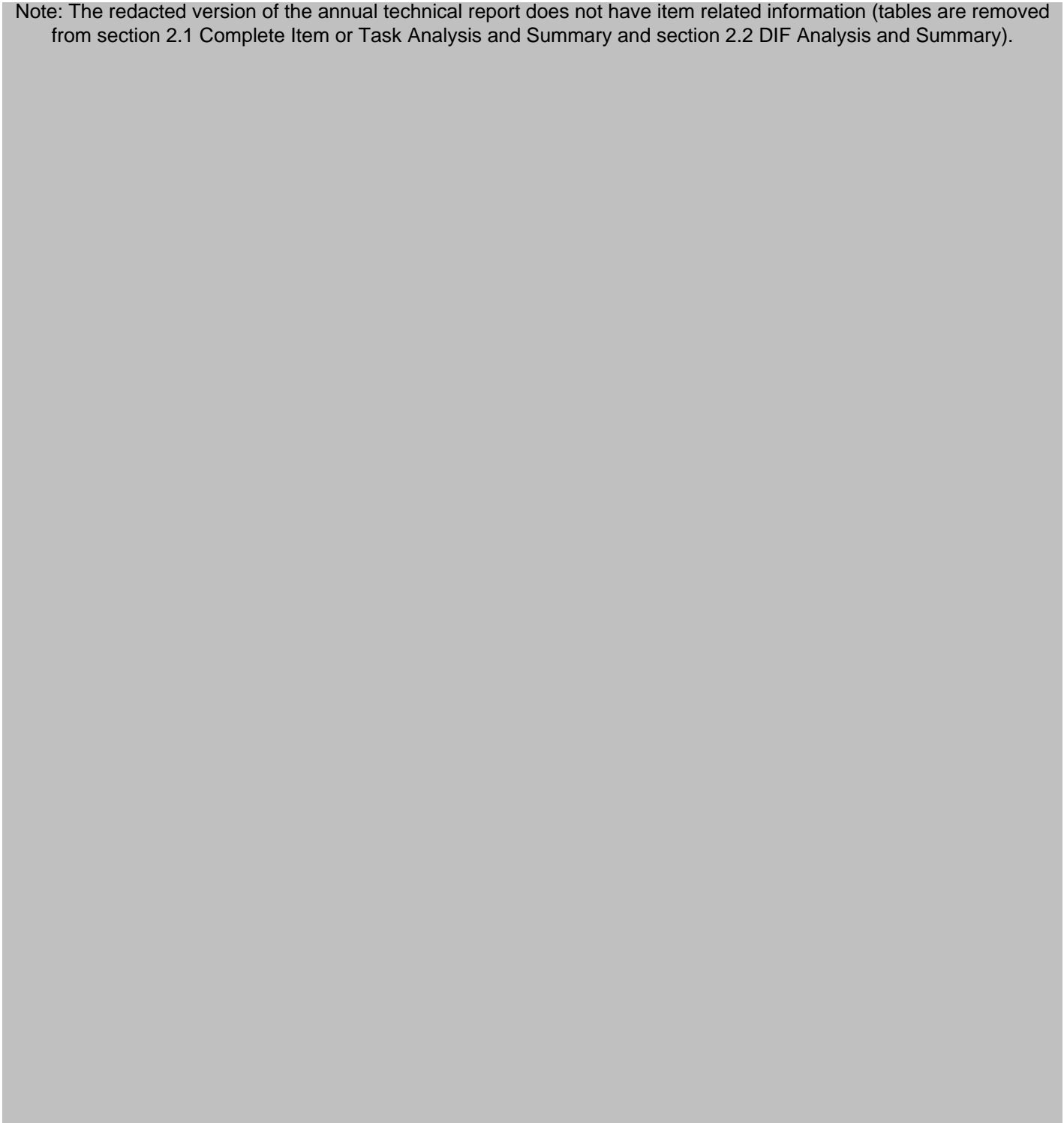
2.2 DIF Analysis and Summary

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

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

It should be noted that for ACCESS Paper Listening, Reading, Writing and Speaking, static forms are used. As such, the DIF analysis was conducted the first year these forms were used operationally, using ATR data. However, for Listening domain, Grades 9-12, Tier A there was C-level DIF found in the original 502 data so the three items that demonstrated C-level DIF were removed and replaced with 503 ATR data. Please see section 2.7, below, for further explanation and operational history of forms.

Dichotomous Items

---

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

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

- A (no DIF) when the absolute value of delta is <1.0

- B (weak DIF) when the absolute value of delta is 1.0 to 1.5

- C (strong DIF) when the absolute value of the delta is >1.5

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

Polytomous Items

For polytomous items (i.e., Writing and Speaking tasks), we take a similar approach. Our approach is based on the M-H chi-square statistic and the standardized mean difference following procedures that ETS developed (Allen, Carlson, & Zalanak, 1999; Zwick, Donoghue, & Grima, 1993). These DIF procedures for polytomous items were used to identify tasks that

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

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

- AA (no DIF), when the Cochran-Mantel-Haenszel chi-square statistic is not significant or when it is significant and standardized P-DIF* is <0.05

- BB (weak DIF), when the Cochran-Mantel-Haenszel chi-square statistic is significant and standardized P-DIF* is ≥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 (focal group) versus female (reference group) or Hispanic (reference group) versus non-Hispanic ethnicities (focal group). The top part of the table summarizes the number of items that exhibit DIF falling into each of the three categories (A, B, or C for

Listening and Reading, and AA, BB, or CC for Writing and Speaking). Any items that show B (or BB) or C (or CC)–level DIF are reported in the bottom part of the table.

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

The DIF panel manager, from CAL, draws panelists from CAL staff members. Members are chosen so that a diverse background is represented. Therefore, the panel manager considers gender, first/second language backgrounds, and ethnicity when empaneling judges. The manager also ensures that some members have expertise in English as a Second Language instruction and/or professional development for teachers of ESL students. Without being told which, if any, items have an initial DIF finding, the panel is asked to discuss all items in the affected folder and come to a consensus on whether they believe or do not believe that the item demonstrates bias against a particular group and is or is not appropriate to place on the operational test.

Series 601 Paper test showed no items with C-level DIF for Listening and Reading.

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

2.3   Raw Score Distribution

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

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

- The number of students in the analyses (the number of students who were not absent, invalid, refused, exempt, or in the wrong grade-level cluster)

- The minimum observed raw score

- The maximum observed raw score

- The mean (average) raw score

- The standard deviation (std. dev.) of the raw scores

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

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

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

As mentioned, students routed to the A form take three P1 tasks, scored 0 to 2. They also take three P3 tasks, scored 0 to 4, for a total raw score range of 0 to 18. Students routed to take the B/C form do not take the P1 tasks, as it is assumed that they would be able to get the full two

points on these very easy P1 tasks. These students take three P3 and three P5 tasks, each scored 0 to 4, and they are awarded two points on each of three P1 tasks. The total raw score range for Tier B/C form is 6 to 30.

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

2.3.1 Listening

2.3.1.0 Kindergarten

Table 2.3.1.0

Raw Score Descriptive Statistics: List K S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	256,847	0	30	19.94	8.56
Total	256,847	0	30	19.94	8.56

Figure 2.3.1.0
Raw Scores: List K S601 Paper
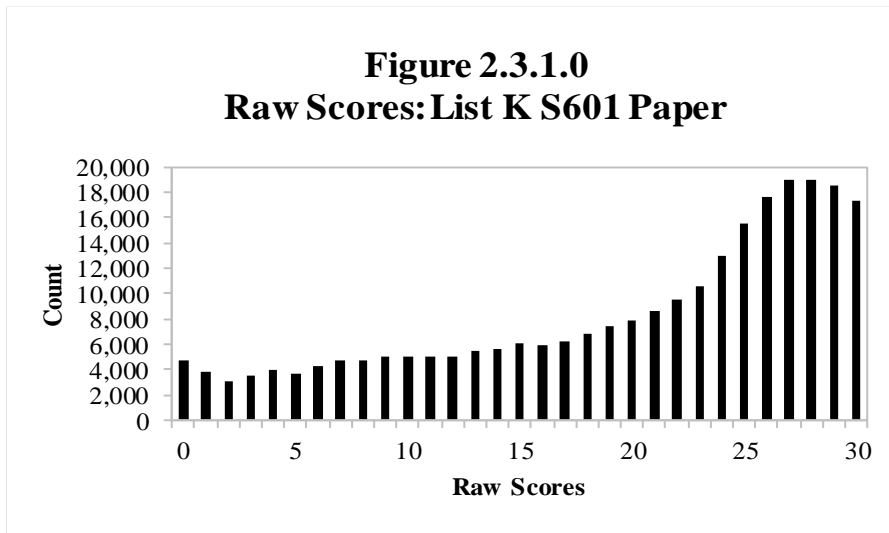
2.3.1.1 Grade 1

Table 2.3.1.1.1

Raw Score Descriptive Statistics: List 1 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	19,651	0	18	13.09	3.42
Total	19,651	0	18	13.09	3.42

Figure 2.3.1.1.1
Raw Scores: List 1A S601 Paper
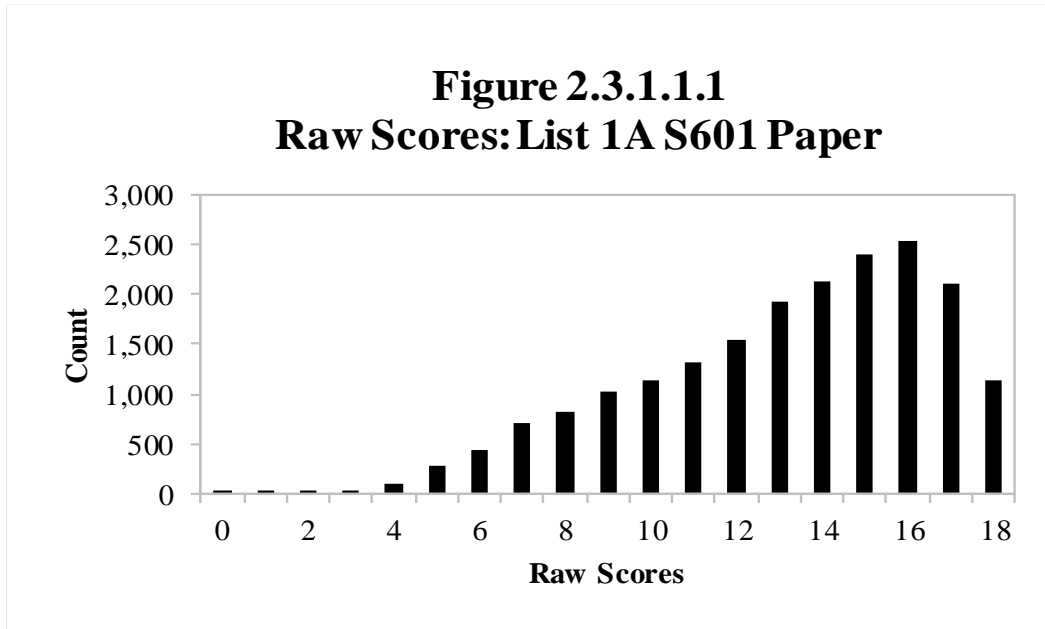
Table 2.3.1.1.2

Raw Score Descriptive Statistics: List 1 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	13,958	2	21	14.63	3.39
Total	13,958	2	21	14.63	3.39

Figure 2.3.1.1.2
Raw Scores: List 1B/C S601 Paper
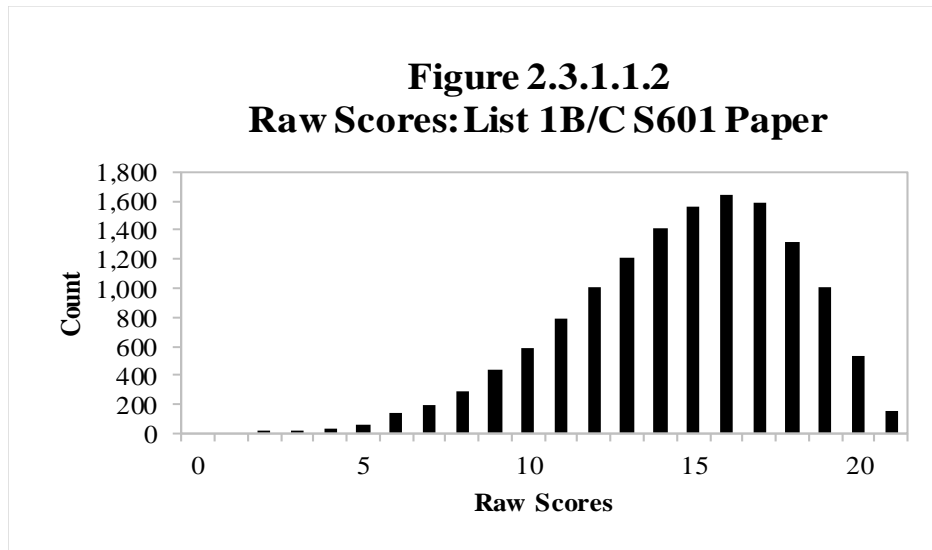
2.3.1.2 Grade 2

Table 2.3.1.2.1

Raw Score Descriptive Statistics: List 2 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	10,904	0	18	13.20	3.78
Total	10,904	0	18	13.20	3.78

Figure 2.3.1.2.1
Raw Scores: List 2A S601 Paper
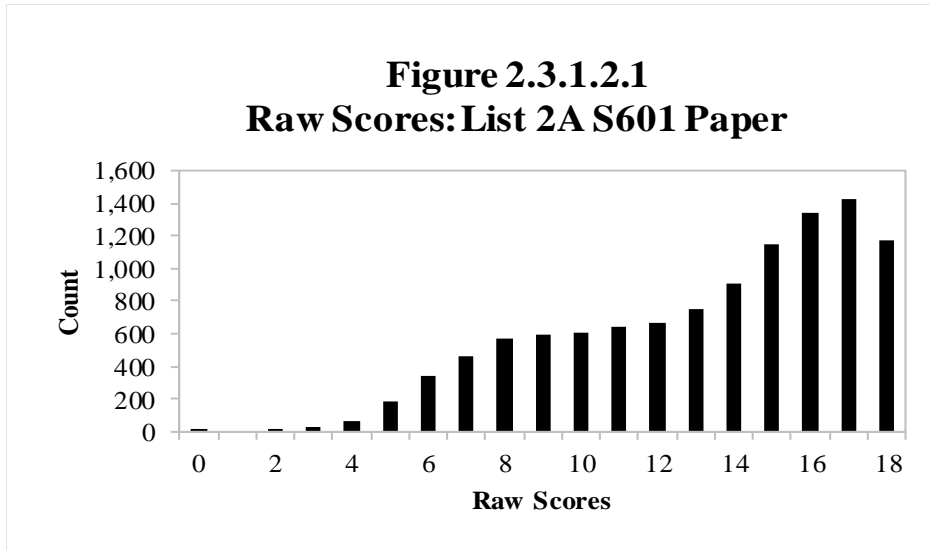
Table 2.3.1.2.2

Raw Score Descriptive Statistics: List 2 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	21,515	2	21	16.56	2.89
Total	21,515	2	21	16.56	2.89

Figure 2.3.1.2.2
Raw Scores: List 2B/C S601 Paper
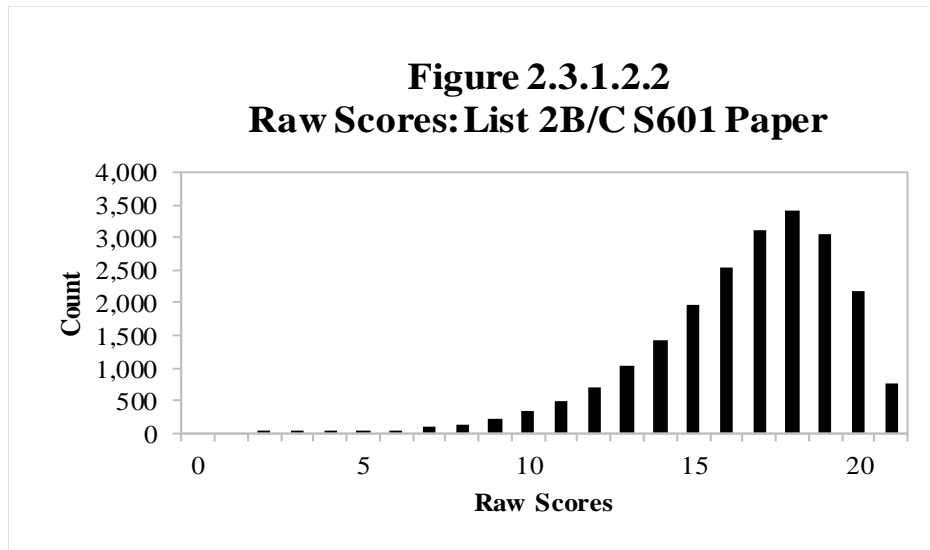
2.3.1.3 Grade 3

Table 2.3.1.3.1

Raw Score Descriptive Statistics: List 3 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	9,041	1	18	10.26	3.51
Total	9,041	1	18	10.26	3.51

Figure 2.3.1.3.1
Raw Scores: List 3A S601 Paper
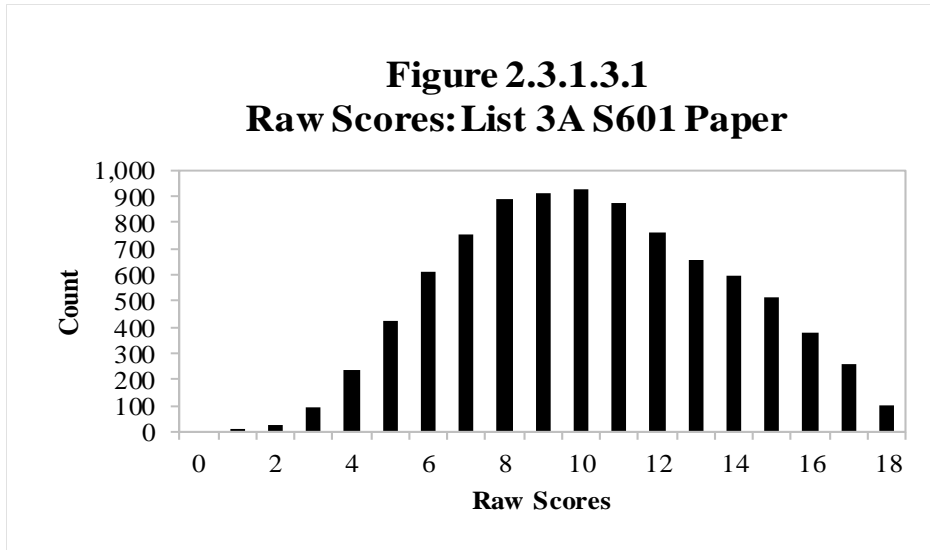
Table 2.3.1.3.2

Raw Score Descriptive Statistics: List 3 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	23,706	2	21	12.53	2.88
Total	23,706	2	21	12.53	2.88

Figure 2.3.1.3.2
Raw Scores: List 3B/C S601 Paper
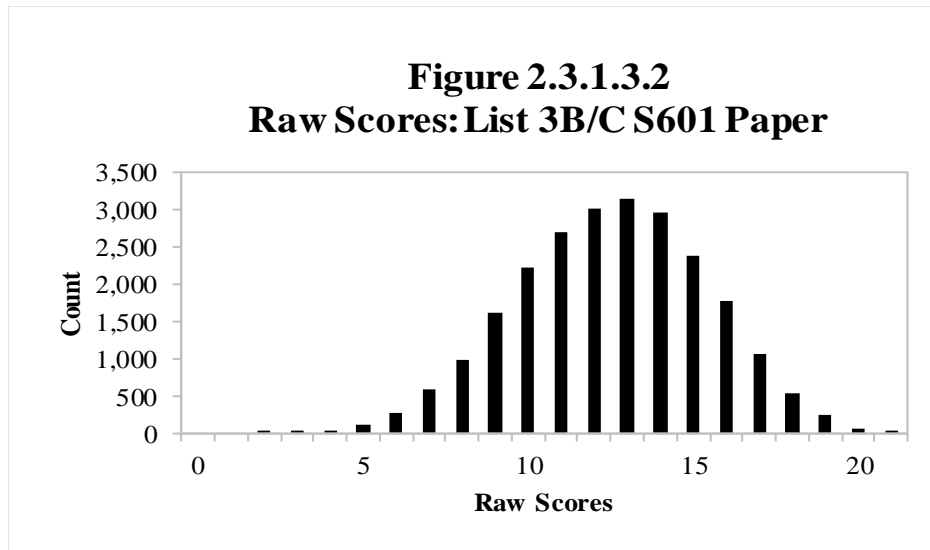
2.3.1.4 Grades 4-5

Table 2.3.1.4.1

Raw Score Descriptive Statistics: List 4-5 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	6,691	1	18	10.38	3.61
5	6,738	1	18	10.98	3.60
Total	13,429	1	18	10.68	3.62

Figure 2.3.1.4.1
Raw Scores: List 4-5A S601 Paper
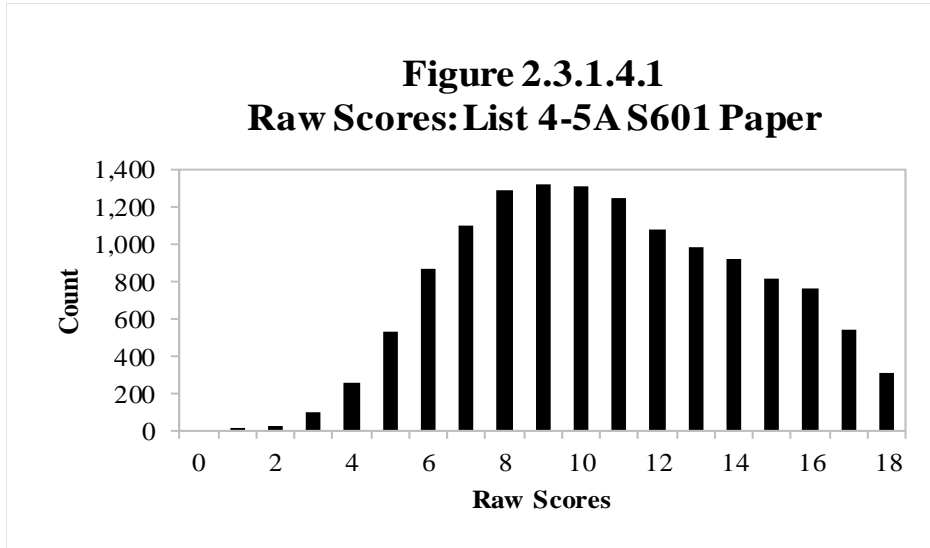
Table 2.3.1.4.2

Raw Score Descriptive Statistics: List 4-5 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	16,974	1	21	14.25	2.85
5	14,957	3	21	14.94	2.93
Total	31,931	1	21	14.57	2.91

Figure 2.3.1.4.2
Raw Scores: List 4-5B/C S601 Paper
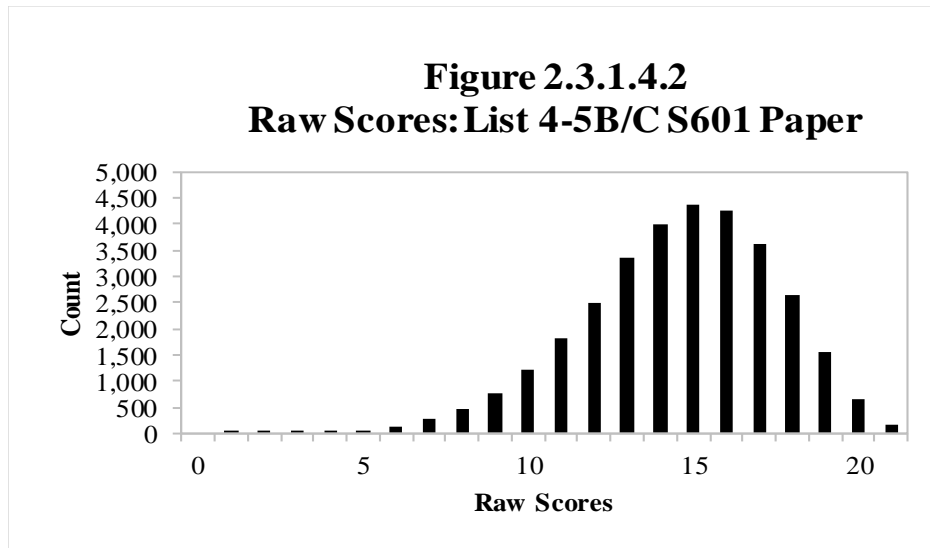
2.3.1.5 Grades 6-8

Table 2.3.1.5.1

Raw Score Descriptive Statistics: List 6-8 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	6,503	1	18	9.78	3.32
7	6,496	1	18	9.93	3.35
8	6,156	0	18	10.26	3.40
Total	19,155	0	18	9.98	3.37

Figure 2.3.1.5.1
Raw Scores: List 6-8A S601 Paper
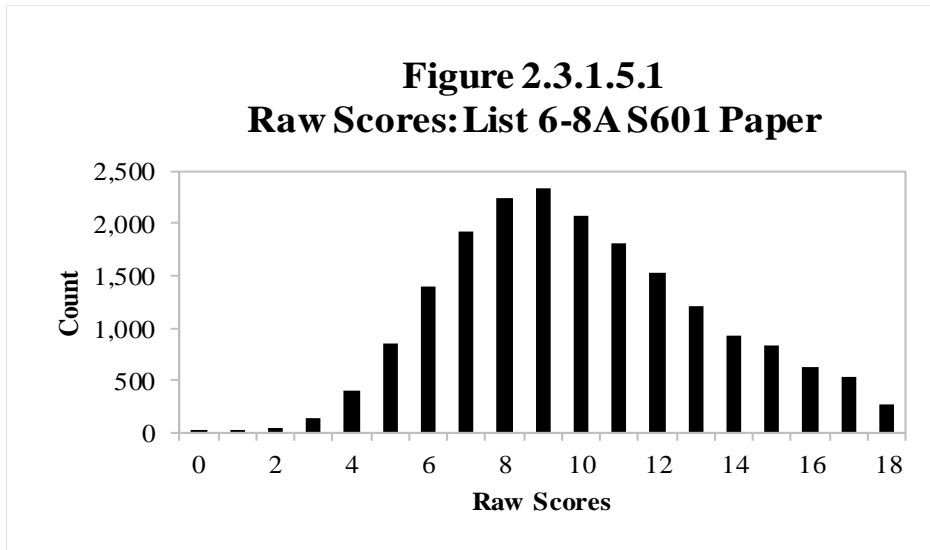
Table 2.3.1.5.2

Raw Score Descriptive Statistics: List 6-8 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	12,212	2	21	14.23	2.95
7	9,962	1	21	14.95	2.94
8	9,380	2	21	15.45	2.96
Total	31,554	1	21	14.82	2.99

Figure 2.3.1.5.2
Raw Scores: List 6-8B/C S601 Paper
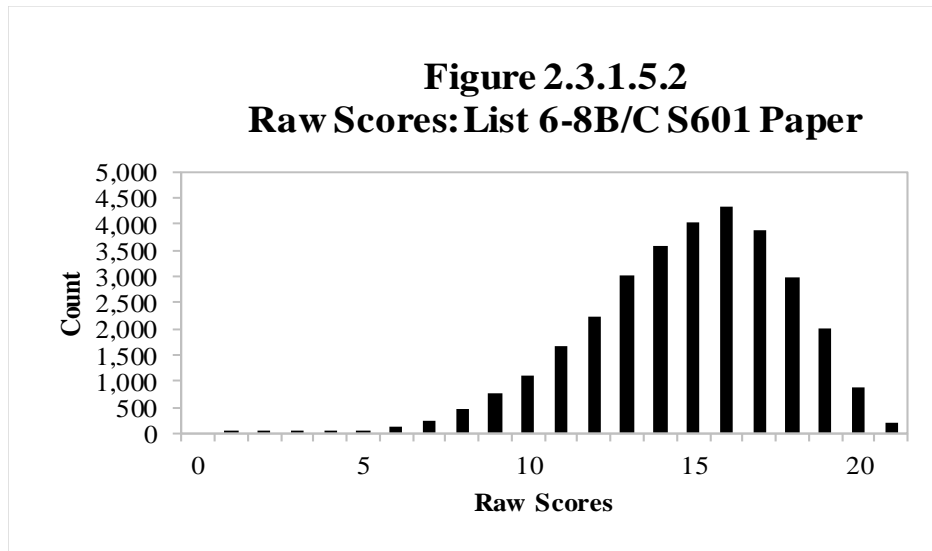
2.3.1.6 Grades 9-12

Table 2.3.1.6.1

Raw Score Descriptive Statistics: List 9-12 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	6,628	1	18	10.10	3.01
10	5,148	2	18	10.26	2.97
11	3,846	1	18	10.77	2.99
12	2,140	3	18	11.16	2.90
Total	17,762	1	18	10.42	3.01

Figure 2.3.1.6.1
Raw Scores: List 9-12 A S601 Paper
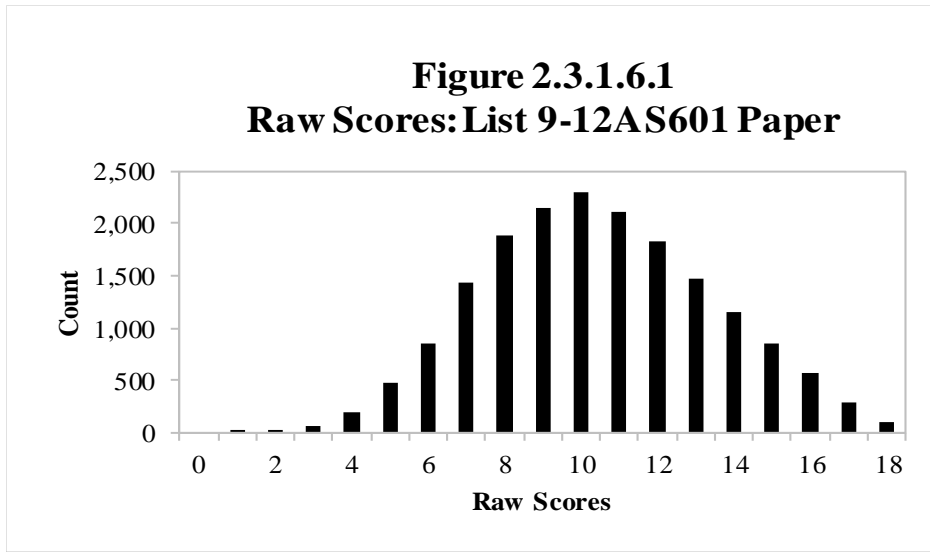
Table 2.3.1.6.2

Raw Score Descriptive Statistics: List 9-12 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	8,687	2	21	13.55	3.14
10	8,072	1	21	13.51	3.24
11	7,405	2	21	13.84	3.21
12	5,551	3	21	13.59	3.29
Total	29,715	1	21	13.62	3.22

Figure 2.3.1.6.2
**Raw Scores: List 9-12B/C S601 Paper**
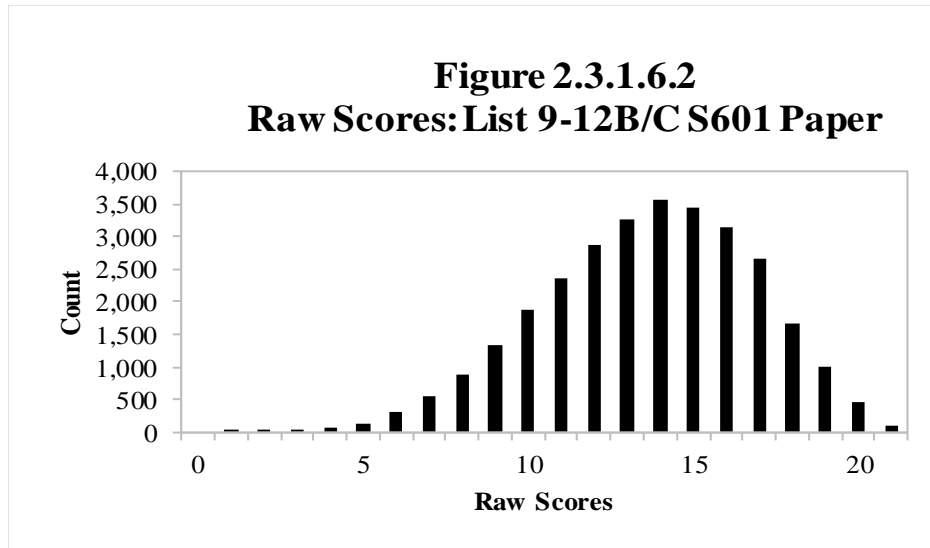
2.3.2 Reading

2.3.2.0 Kindergarten

Table 2.3.2.0

Raw Score Descriptive Statistics: Read K S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	256,837	0	30	16.81	7.91
Total	256,837	0	30	16.81	7.91

Figure 2.3.2.0
Raw Scores: Read K S601 Paper
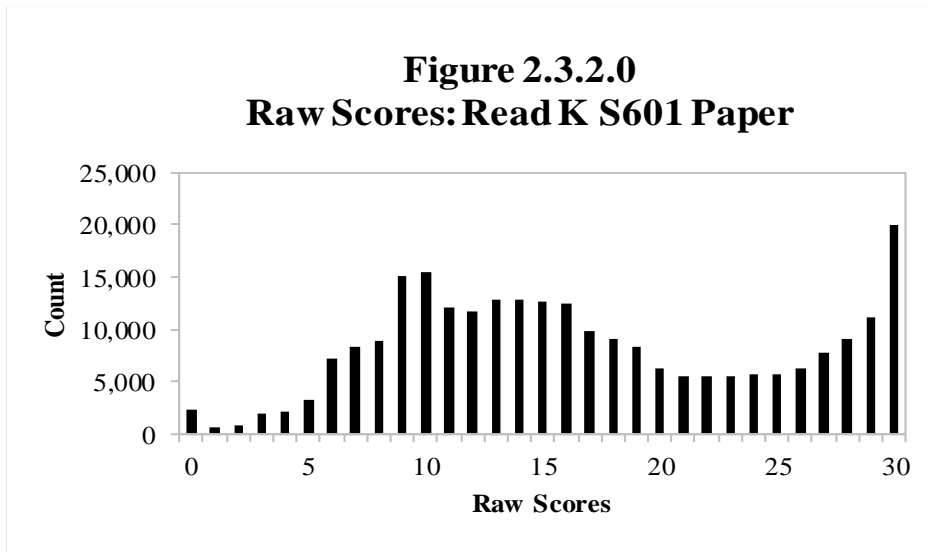
2.3.2.1 Grade 1

Table 2.3.2.1.1

Raw Score Descriptive Statistics: Read 1 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	18,345	0	24	10.61	3.97
Total	18,345	0	24	10.61	3.97

Figure 2.3.2.1.1
Raw Scores: Read 1A S601 Paper
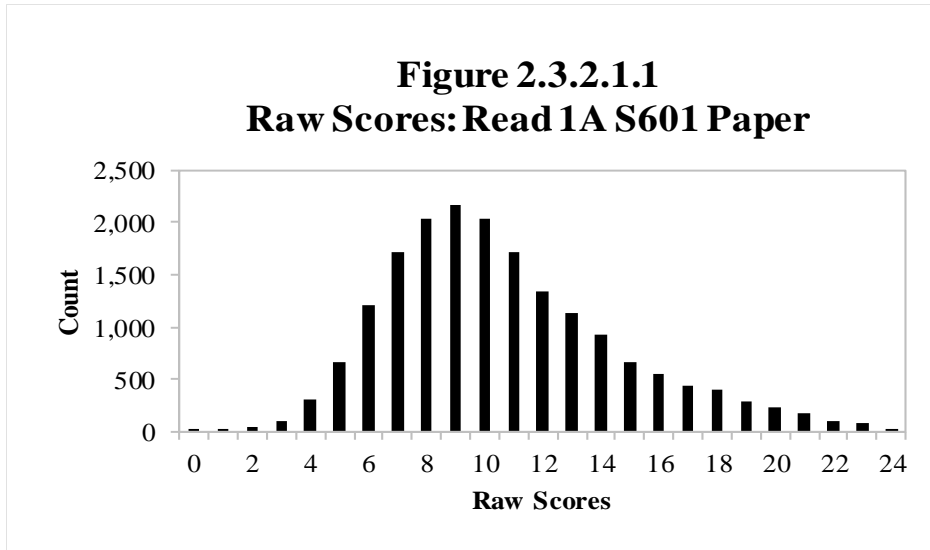
Table 2.3.2.1.2

Raw Score Descriptive Statistics: Read 1 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	12,132	2	27	11.69	4.54
Total	12,132	2	27	11.69	4.54

Figure 2.3.2.1.2
Raw Scores: Read 1B/C S601 Paper
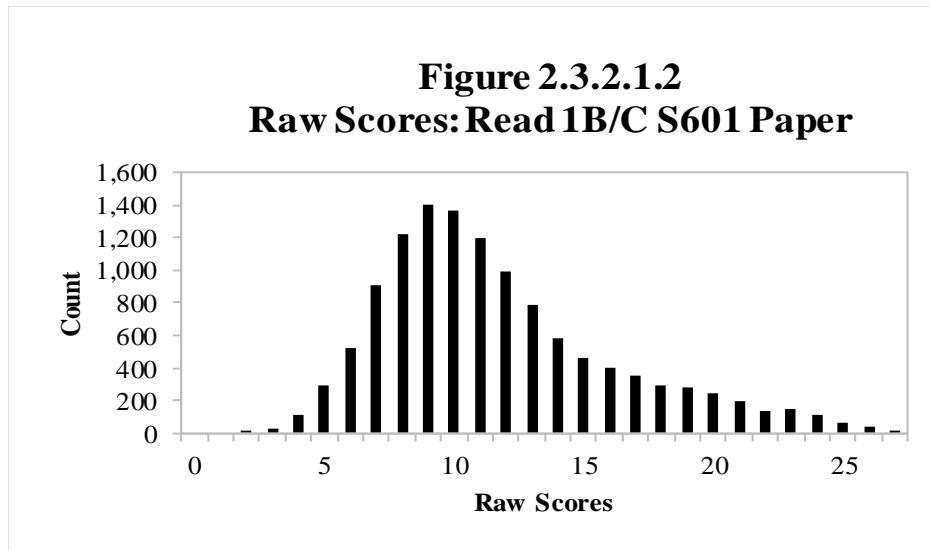
2.3.2.2 Grade 2

Table 2.3.2.2.1

Raw Score Descriptive Statistics: Read 2 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	10,057	1	24	12.32	4.75
Total	10,057	1	24	12.32	4.75

Figure 2.3.2.2.1
Raw Scores: Read 2A S601 Paper
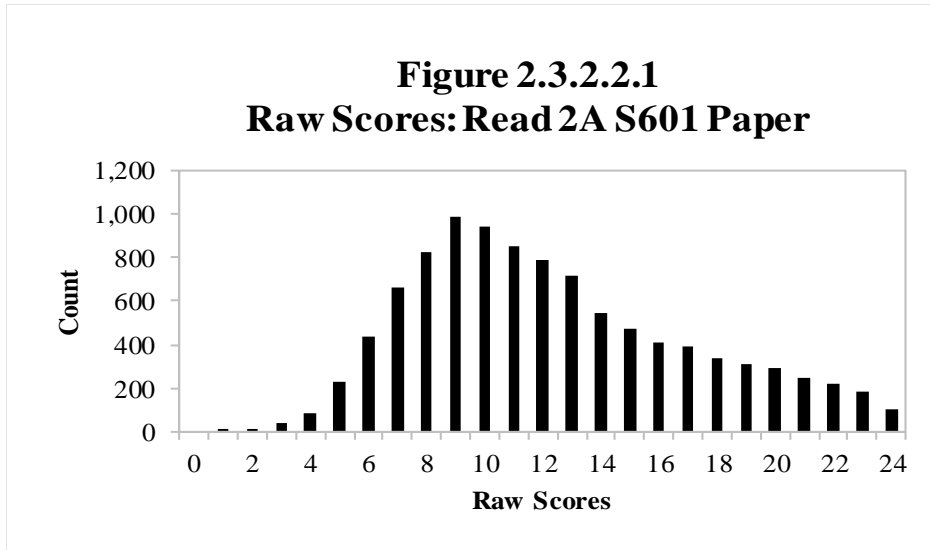
Table 2.3.2.2.2

Raw Score Descriptive Statistics: Read 2 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	18,817	1	27	15.27	5.51
Total	18,817	1	27	15.27	5.51

Figure 2.3.2.2.2
Raw Scores: Read 2B/C S601 Paper
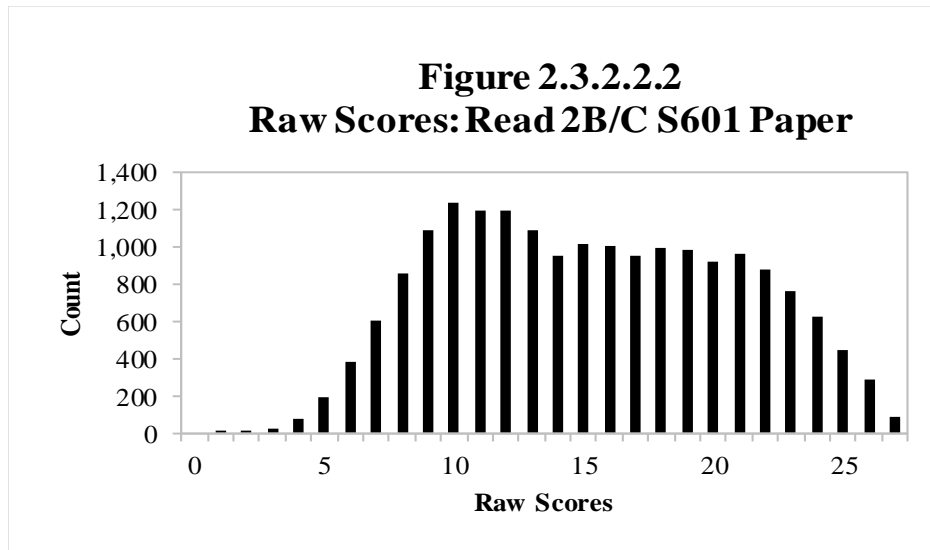
2.3.2.3 Grade 3

Table 2.3.2.3.1

Raw Score Descriptive Statistics: Read 3 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	8,294	1	24	11.22	4.48
Total	8,294	1	24	11.22	4.48

Figure 2.3.2.3.1
Raw Scores: Read 3A S601 Paper
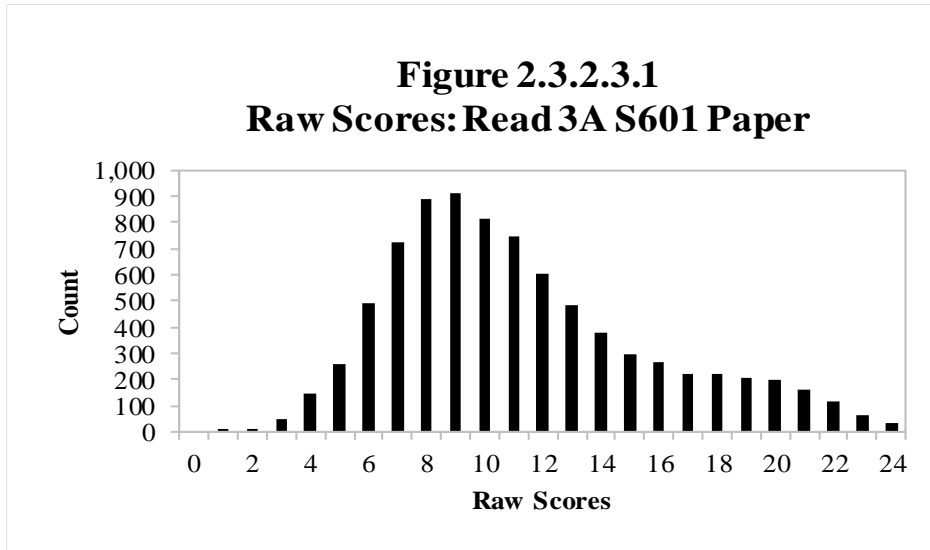
Table 2.3.2.3.2

Raw Score Descriptive Statistics: Read 3 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	20,701	0	27	9.60	3.87
Total	20,701	0	27	9.60	3.87

Figure 2.3.2.3.2
Raw Scores: Read 3B/C S601 Paper
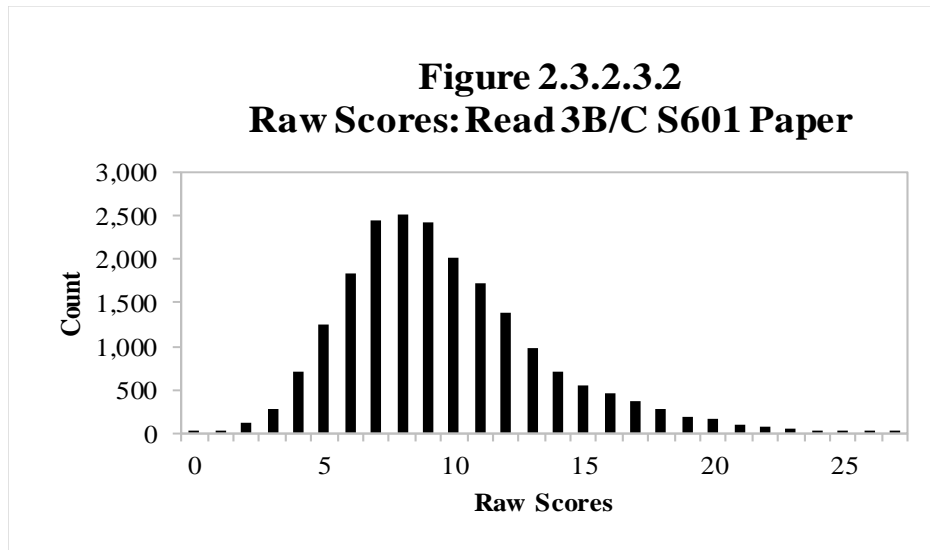
2.3.2.4 Grades 4-5

Table 2.3.2.4.1

Raw Score Descriptive Statistics: Read 4-5 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	6,187	1	24	12.56	4.82
5	6,428	0	24	13.50	4.99
Total	12,615	0	24	13.04	4.93

Figure 2.3.2.4.1
Raw Scores: Read 4-5 A S601 Paper
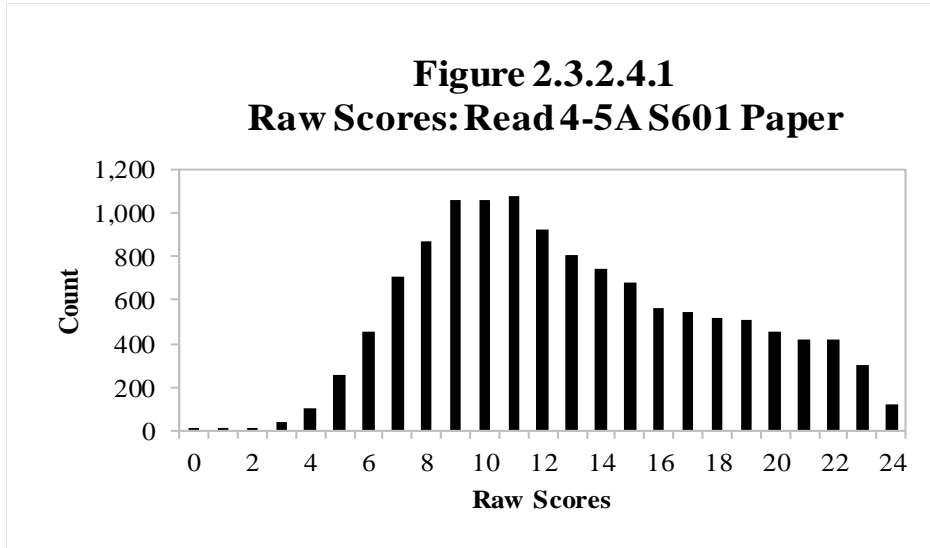
Table 2.3.2.4.2

Raw Score Descriptive Statistics: Read 4-5 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	15,138	0	27	12.15	4.60
5	13,620	1	27	13.58	5.08
Total	28,758	0	27	12.83	4.89

Figure 2.3.2.4.2
Raw Scores: Read 4-5B/C S601 Paper
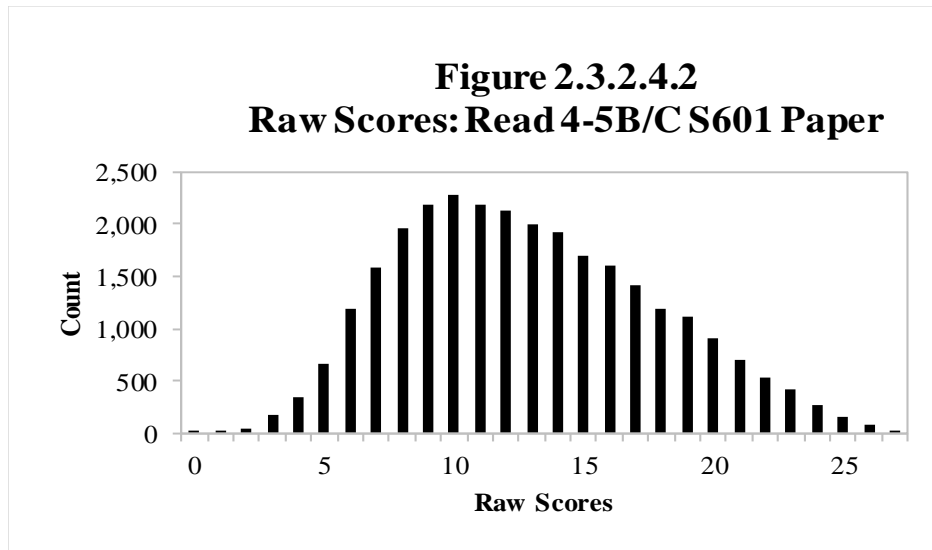
2.3.2.5 Grades 6-8

Table 2.3.2.5.1

Raw Score Descriptive Statistics: Read 6-8 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	6,244	0	24	10.47	4.25
7	6,249	1	24	11.06	4.43
8	6,005	0	24	11.85	4.68
Total	18,498	0	24	11.12	4.49

Figure 2.3.2.5.1
Raw Scores: Read 6-8A S601 Paper
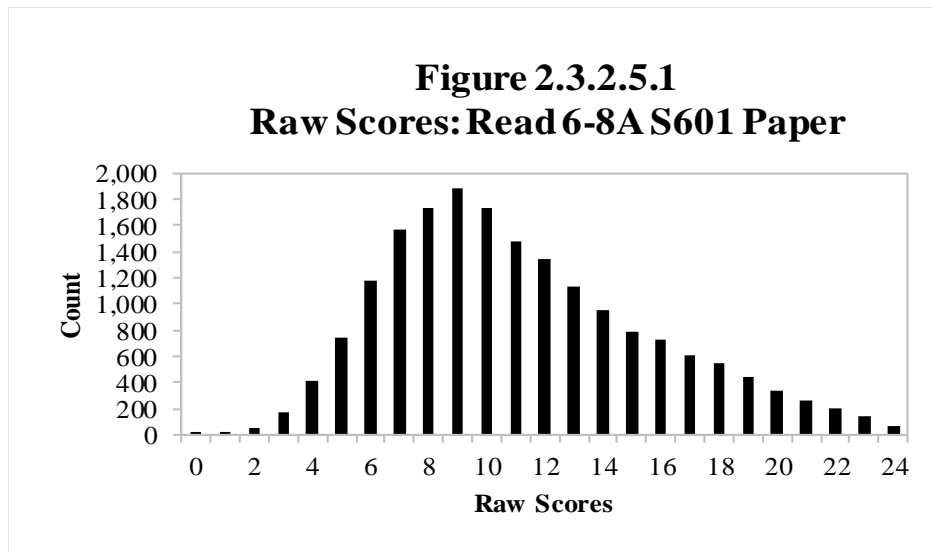
Table 2.3.2.5.2

Raw Score Descriptive Statistics: Read 6-8 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	11,336	0	27	10.88	4.28
7	9,349	1	27	12.56	4.87
8	8,908	1	27	14.20	5.23
Total	29,593	0	27	12.41	4.96

Figure 2.3.2.5.2
Raw Scores: Read 6-8 B/C S601 Paper
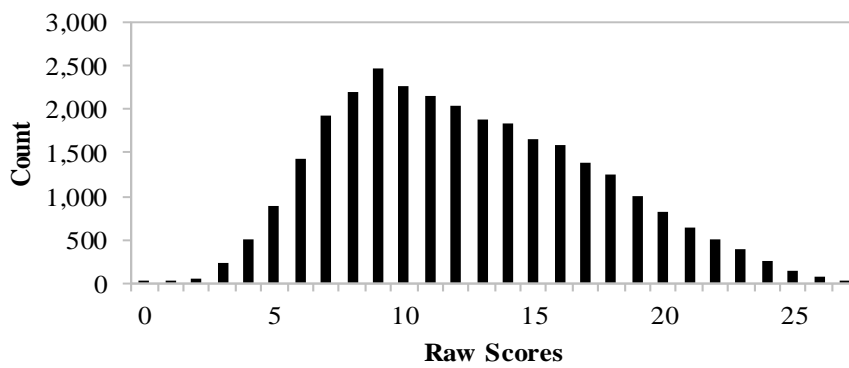
2.3.2.6 Grades 9-12

Table 2.3.2.6.1

Raw Score Descriptive Statistics: Read 9-12 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	6,583	0	24	13.05	4.57
10	5,126	1	24	13.57	4.53
11	3,831	3	24	14.59	4.63
12	2,123	3	24	15.17	4.56
Total	17,663	0	24	13.79	4.63

Figure 2.3.2.6.1
Raw Scores: Read 9-12A S601 Paper
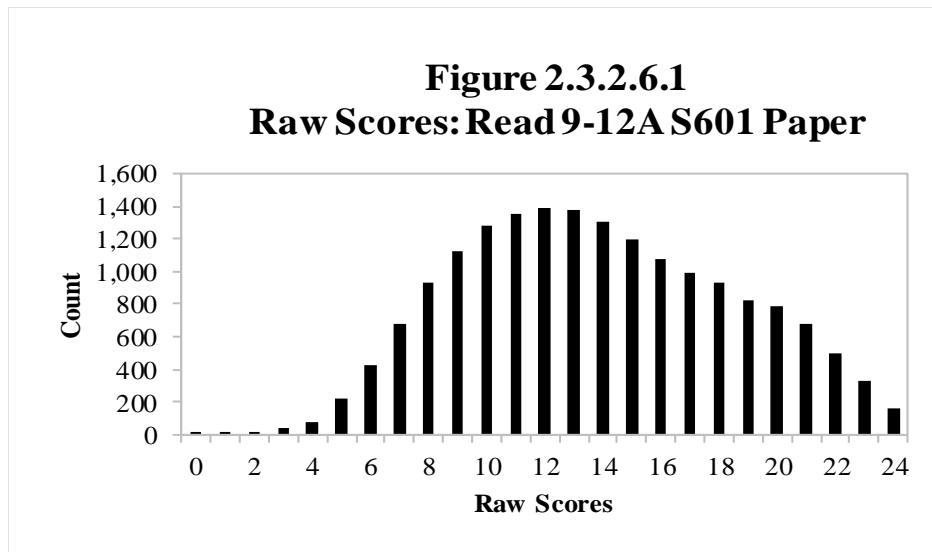
Table 2.3.2.6.2

Raw Score Descriptive Statistics: Read 9-12 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	7,929	1	27	13.95	5.13
10	7,502	1	27	14.46	5.28
11	6,862	1	27	15.31	5.38
12	5,292	2	27	14.83	5.44
Total	27,585	1	27	14.60	5.32

Figure 2.3.2.6.2
Raw Scores: Read 9-12B/C S601 Paper
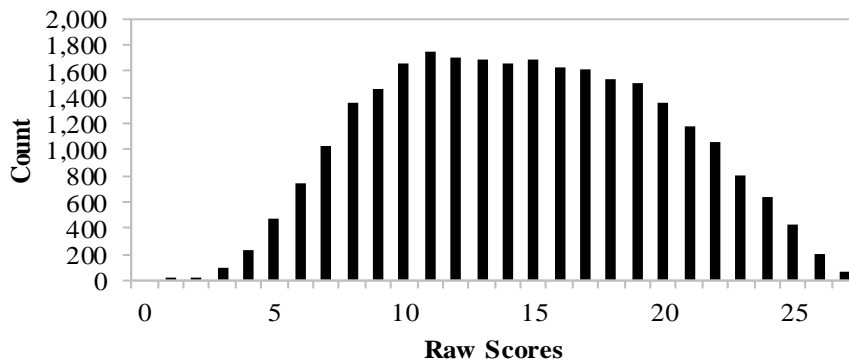
2.3.3 Writing

2.3.3.0 Kindergarten

Table 2.3.3.0

Raw Score Descriptive Statistics: Writ K S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	256,834	0	17	6.96	4.17
Total	256,834	0	17	6.96	4.17

Figure 2.3.3.0
Raw Scores: Writ K S601 Paper
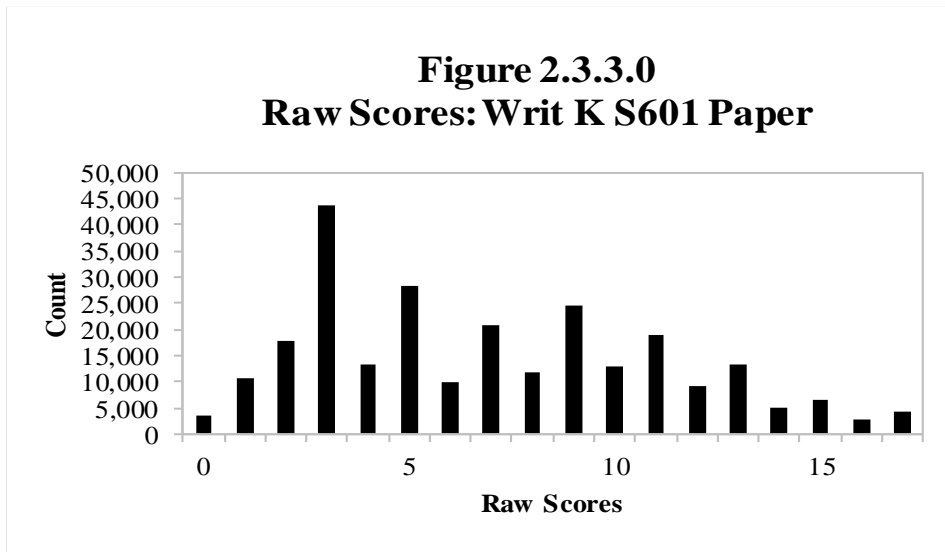
2.3.3.1 Grade 1

Table 2.3.3.1.1

Raw Score Descriptive Statistics: Writ 1 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	23,675	0	25	10.19	5.86
Total	23,675	0	25	10.19	5.86

Figure 2.3.3.1.1
Raw Scores: Writ 1A S601 Paper
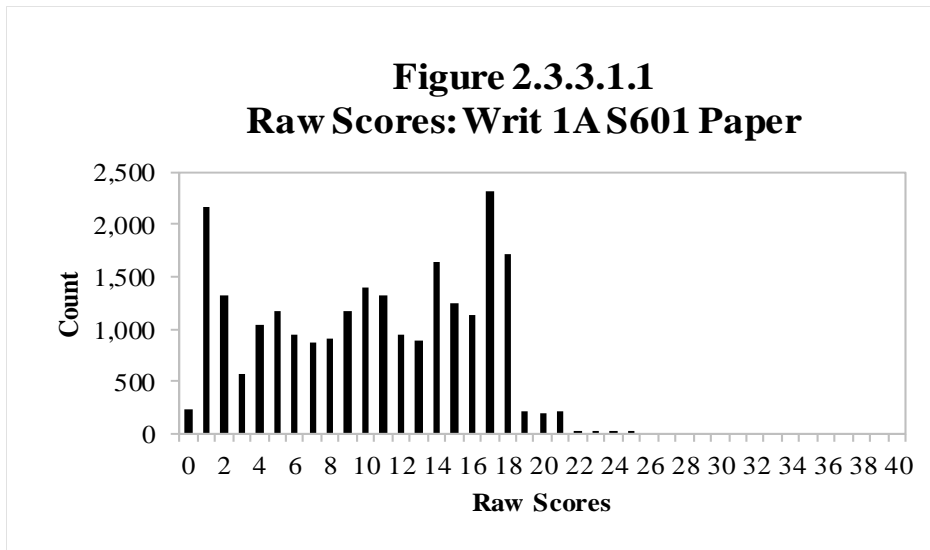
Table 2.3.3.1.2

Raw Score Descriptive Statistics: Writ 1 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,827	0	44	18.49	9.10
Total	15,827	0	44	18.49	9.10

Figure 2.3.3.1.2
Raw Scores: Writ 1B/C S601 Paper
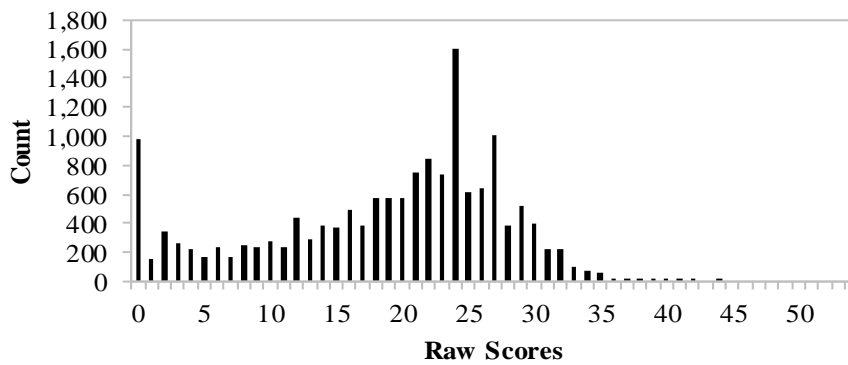
2.3.3.2 Grade 2

Table 2.3.3.2.1

Raw Score Descriptive Statistics: Writ 2 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	12,283	0	18	6.34	4.43
Total	12,283	0	18	6.34	4.43

Figure 2.3.3.2.1
Raw Scores: Writ 2A S601 Paper
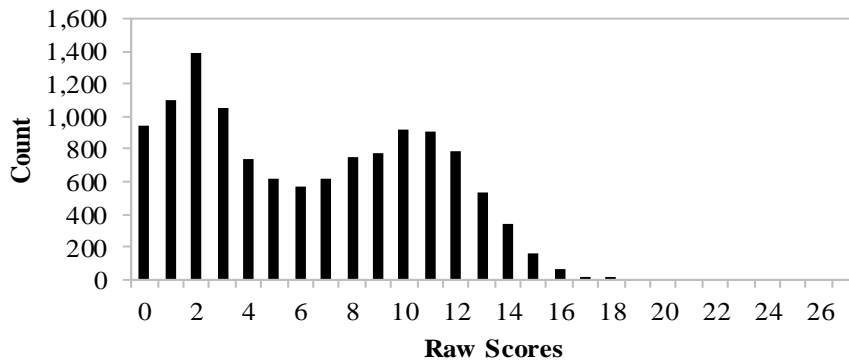
Table 2.3.3.2.2

Raw Score Descriptive Statistics: Writ 2 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	22,974	0	45	23.45	8.28
Total	22,974	0	45	23.45	8.28

**Figure 2.3.3.2.2
Raw Scores: Writ 2B/C S601 Paper**
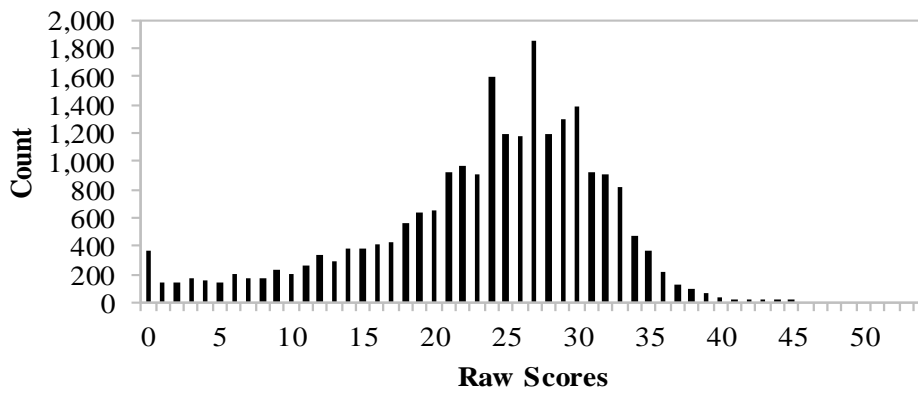
2.3.3.3 Grade 3

Table 2.3.3.3.1

Raw Score Descriptive Statistics: Writ 3 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	9,880	0	20	7.52	4.61
Total	9,880	0	20	7.52	4.61

Figure 2.3.3.3.1
Raw Scores: Writ 3A S601 Paper
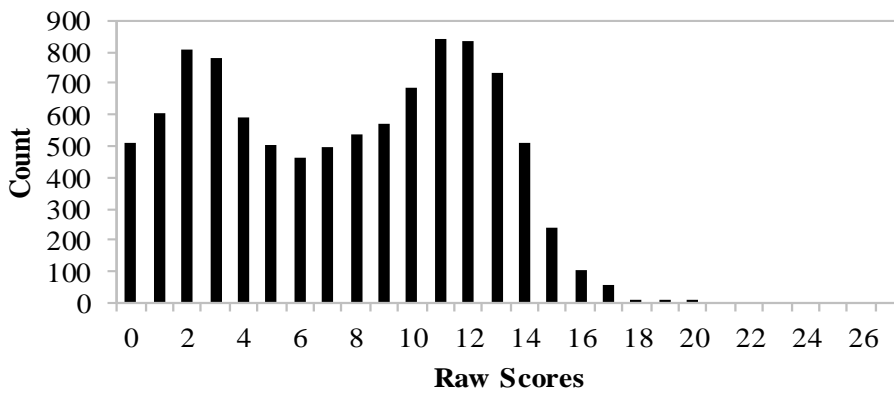
Table 2.3.3.3.2

Raw Score Descriptive Statistics: Writ 3 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	25,187	0	47	26.87	7.15
Total	25,187	0	47	26.87	7.15

Figure 2.3.3.3.2
Raw Scores: Writ 3B/C S601 Paper
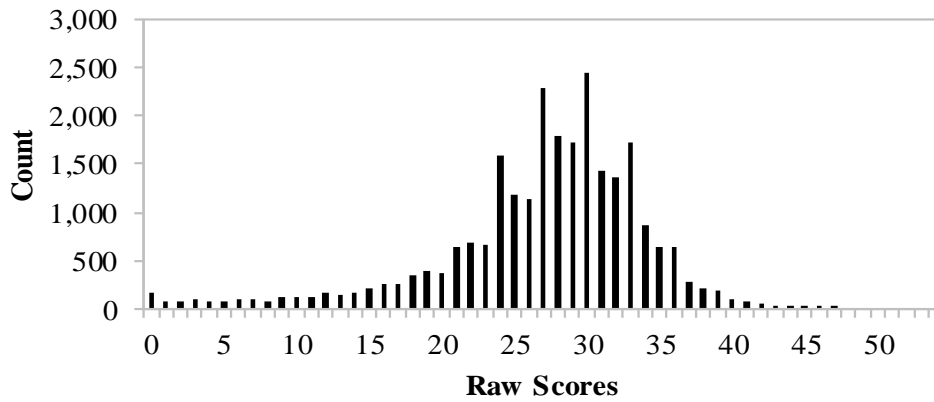
2.3.3.4 Grades 4-5

Table 2.3.3.4.1

Raw Score Descriptive Statistics: Writ 4-5 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	7,213	0	19	6.63	4.32
5	7,228	0	19	7.50	4.34
Total	14,441	0	19	7.07	4.35

Figure 2.3.3.4.1
Raw Scores: Writ 4-5A S601 Paper
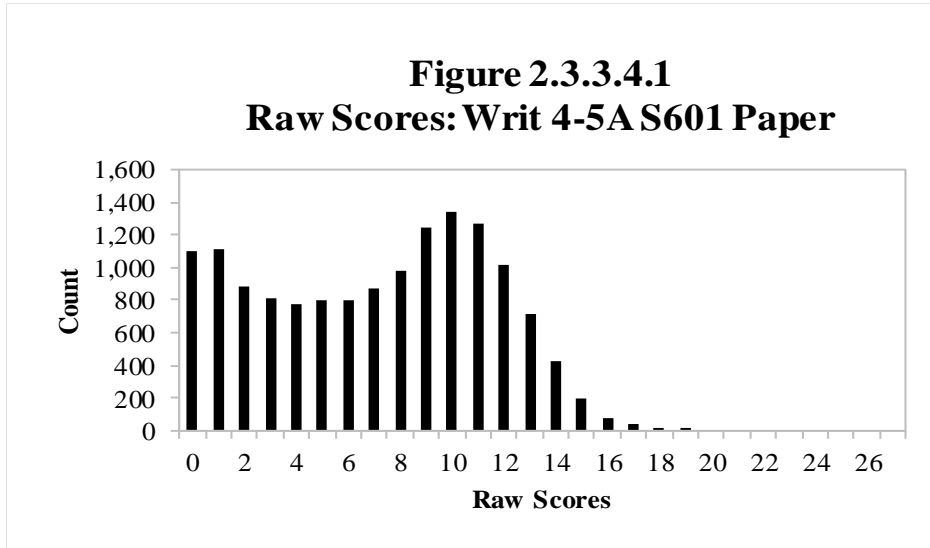
Table 2.3.3.4.2

Raw Score Descriptive Statistics: Writ 4-5 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	17,595	0	44	23.77	6.47
5	15,367	0	44	26.10	6.73
Total	32,962	0	44	24.86	6.70

Figure 2.3.3.4.2
Raw Scores: Writ 4-5B/C S601 Paper
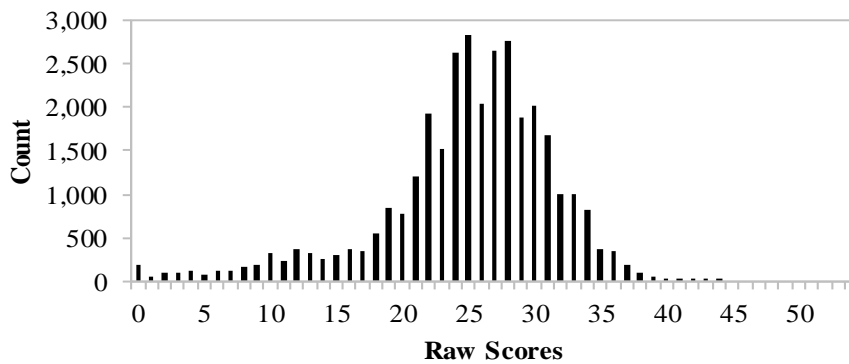
2.3.3.5 Grades 6-8

Table 2.3.3.5.1

Raw Score Descriptive Statistics: Writ 6-8 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	6,993	0	19	7.42	4.17
7	6,977	0	20	7.99	4.13
8	6,639	0	21	8.52	4.16
Total	20,609	0	21	7.97	4.18

Figure 2.3.3.5.1
Raw Scores: Writ 6-8A S601 Paper
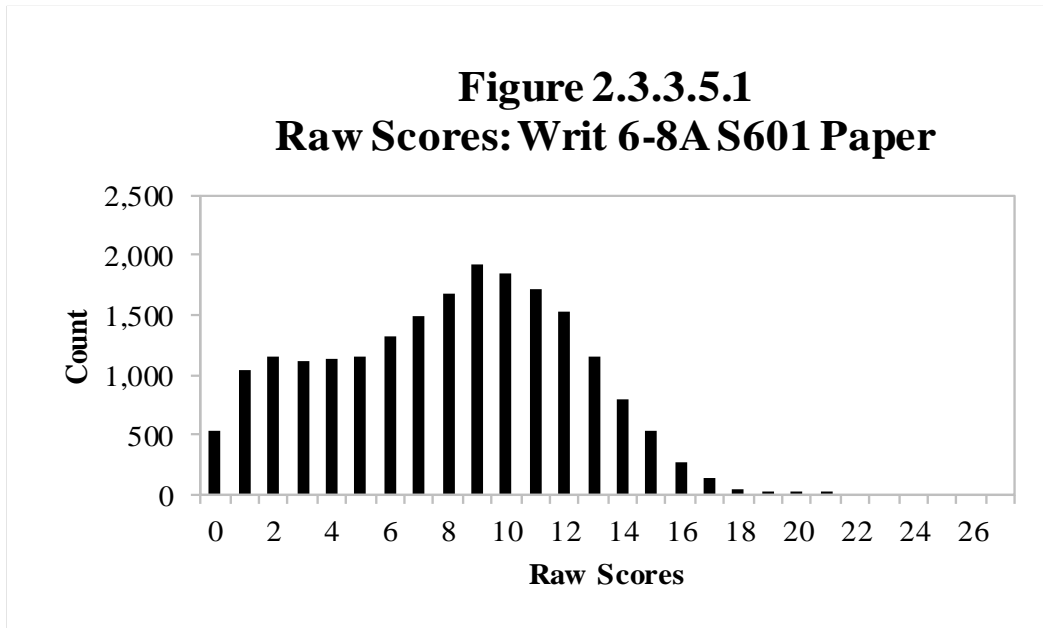
Table 2.3.3.5.2

Raw Score Descriptive Statistics: Writ 6-8 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	12,787	0	45	26.51	7.05
7	10,421	0	47	28.03	6.84
8	9,841	0	47	29.23	6.73
Total	33,049	0	47	27.80	6.98

Figure 2.3.3.5.2
Raw Scores: Writ 6-8B/C S601 Paper
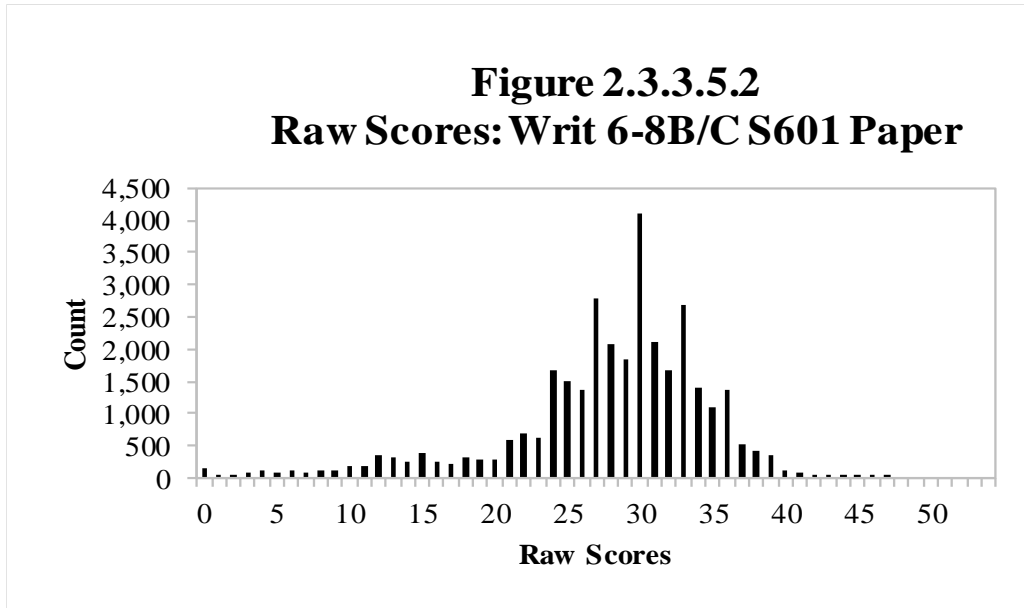
2.3.3.6 Grades 9-12

Table 2.3.3.6.1

Raw Score Descriptive Statistics: Writ 9-12 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	7,199	0	20	7.52	4.57
10	5,517	0	20	7.97	4.41
11	4,151	0	21	8.97	4.36
12	2,257	0	20	9.81	4.16
Total	19,124	0	21	8.23	4.50

Figure 2.3.3.6.1
Raw Scores: Writ 9-12 A S601 Paper
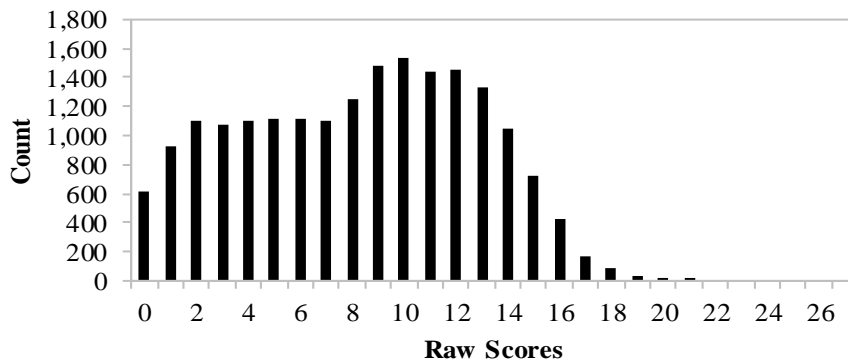
Table 2.3.3.6.2

Raw Score Descriptive Statistics: Writ 9-12 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	9,171	0	47	29.60	6.80
10	8,540	0	50	29.80	7.40
11	7,803	0	51	30.86	7.02
12	5,843	0	48	30.19	7.61
Total	31,357	0	51	30.08	7.19

Figure 2.3.3.6.2
Raw Scores: Writ 9-12B/C S601 Paper
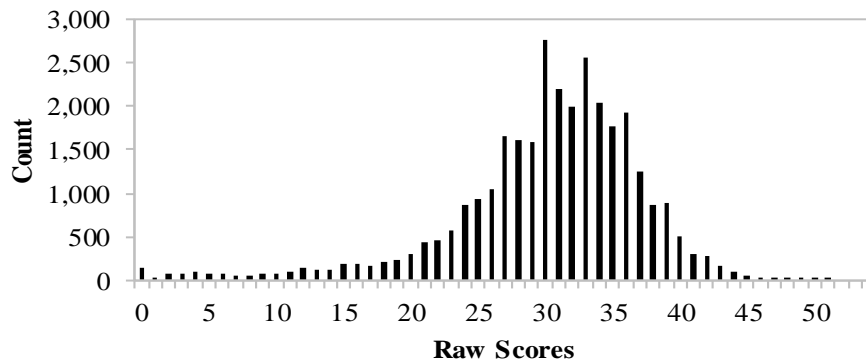
2.3.4 Speaking

2.3.4.0 Kindergarten

Table 2.3.4.0

Raw Score Descriptive Statistics: Spek K S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	256,828	0	10	5.73	3.54
Total	256,828	0	10	5.73	3.54

Figure 2.3.4.0
Raw Scores: Spek K S601 Paper
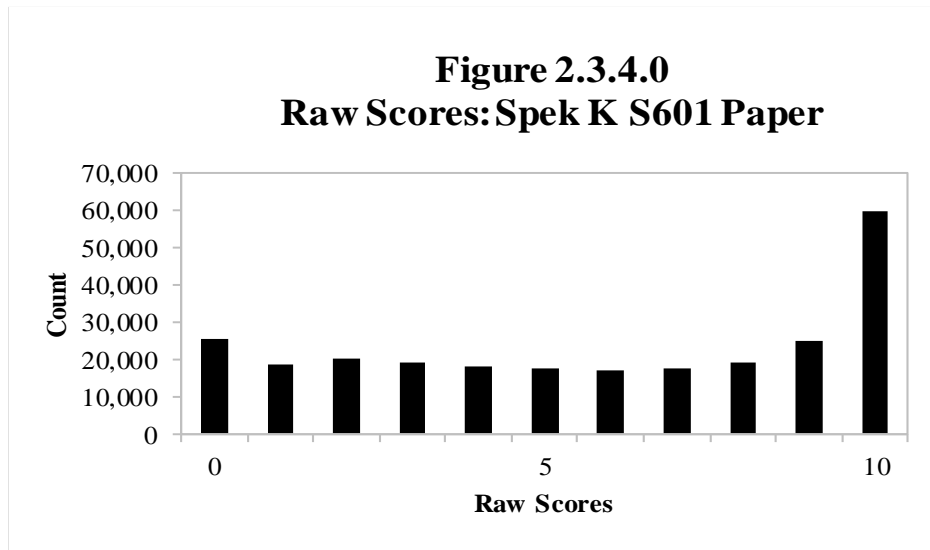
2.3.4.1 Grade 1

Table 2.3.4.1.1

Raw Score Descriptive Statistics: Spek 1 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	23,502	0	18	10.12	4.52
Total	23,502	0	18	10.12	4.52

Figure 2.3.4.1.1
Raw Scores: Spek 1A S601 Paper
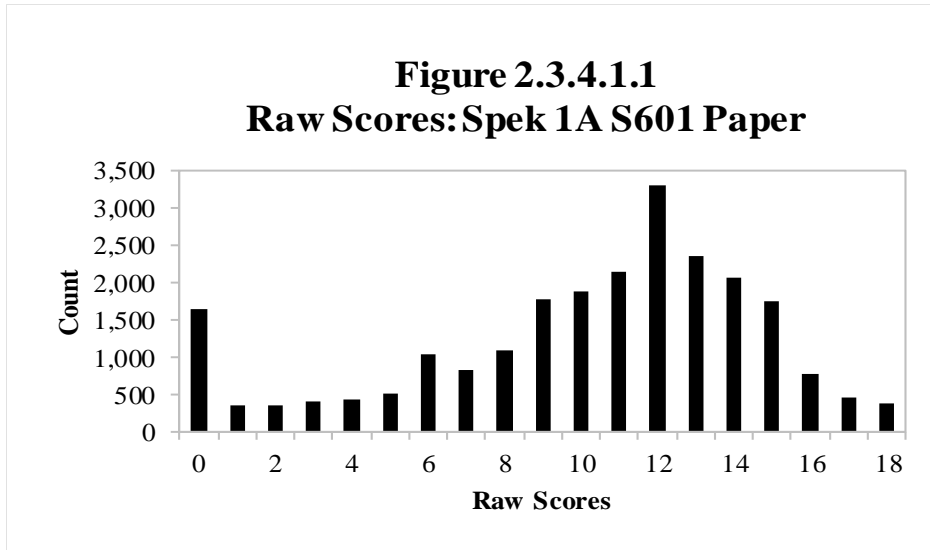
Table 2.3.4.1.2

Raw Score Descriptive Statistics: Spek 1 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,701	6	30	20.40	4.87
Total	15,701	6	30	20.40	4.87

Figure 2.3.4.1.2
Raw Scores: Spek 1B/C S601 Paper
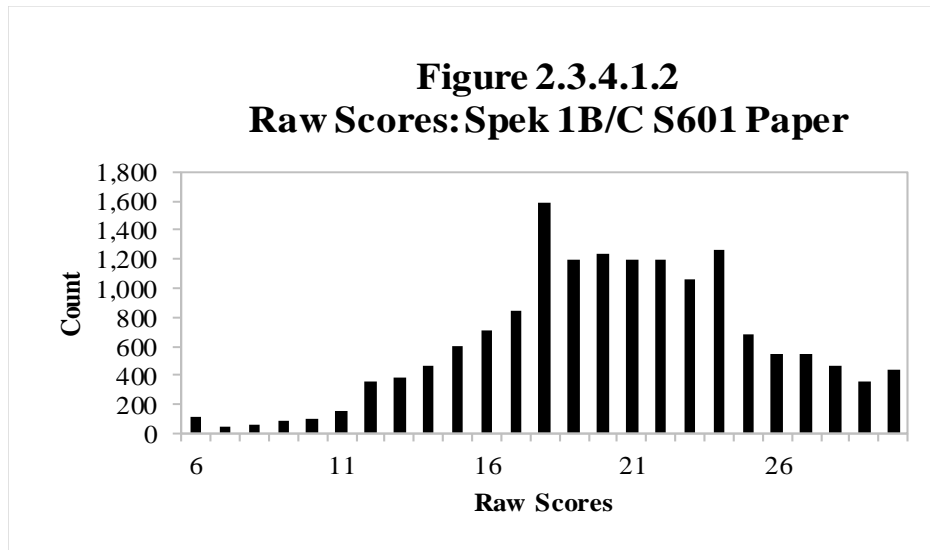
2.3.4.2 Grade 2

Table 2.3.4.2.1

Raw Score Descriptive Statistics: Spek 2 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	12,214	0	18	9.56	5.02
Total	12,214	0	18	9.56	5.02

Figure 2.3.4.2.1
Raw Scores: Spek 2A S601 Paper
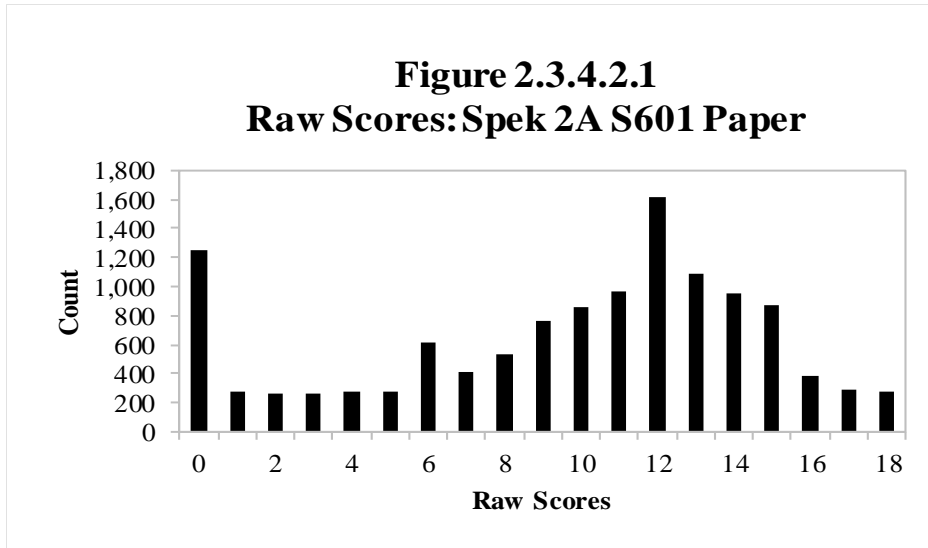
Table 2.3.4.2.2

Raw Score Descriptive Statistics: Spek 2 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	22,806	6	30	21.46	4.47
Total	22,806	6	30	21.46	4.47

Figure 2.3.4.2.2
Raw Scores: Spek 2B/C S601 Paper
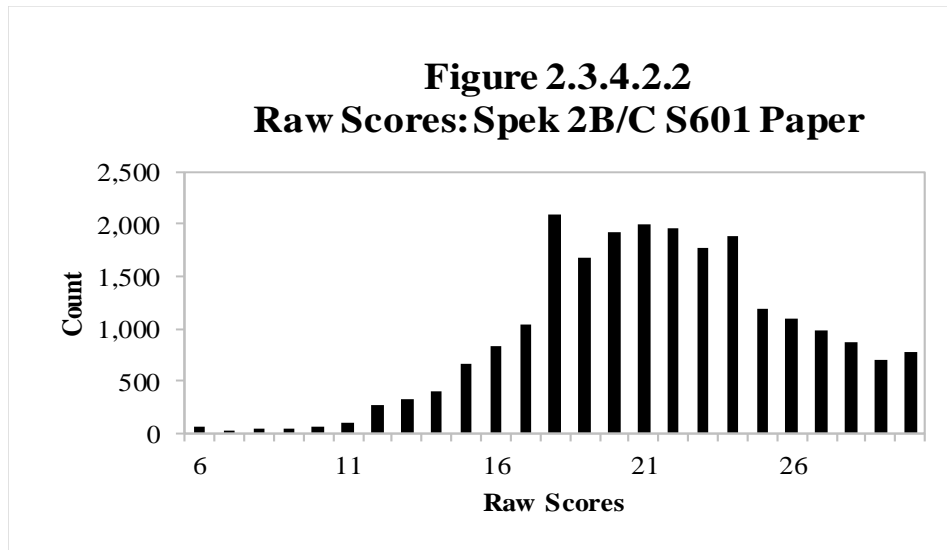
2.3.4.3 Grade 3

Table 2.3.4.3.1

Raw Score Descriptive Statistics: Spek 3 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	9,813	0	18	9.37	5.01
Total	9,813	0	18	9.37	5.01

Figure 2.3.4.3.1
Raw Scores: Spek 3A S601 Paper
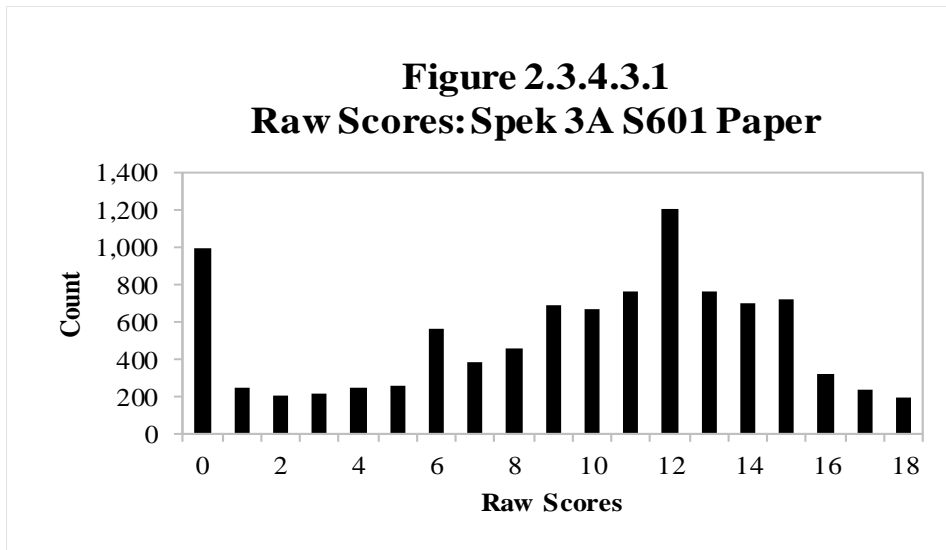
Table 2.3.4.3.2

Raw Score Descriptive Statistics: Spek 3 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	25,031	6	30	22.36	4.34
Total	25,031	6	30	22.36	4.34

Figure 2.3.4.3.2
Raw Scores: Spek 3B/C S601 Paper
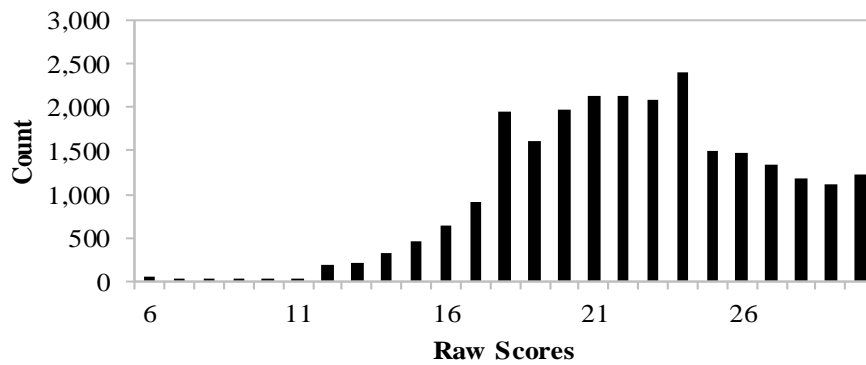
2.3.4.4 Grades 4-5

Table 2.3.4.4.1

Raw Score Descriptive Statistics: Spek 4-5 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	7,153	0	18	7.79	4.77
5	7,187	0	18	8.08	4.67
Total	14,340	0	18	7.93	4.73

Figure 2.3.4.4.1
Raw Scores: Spek 4-5A S601 Paper
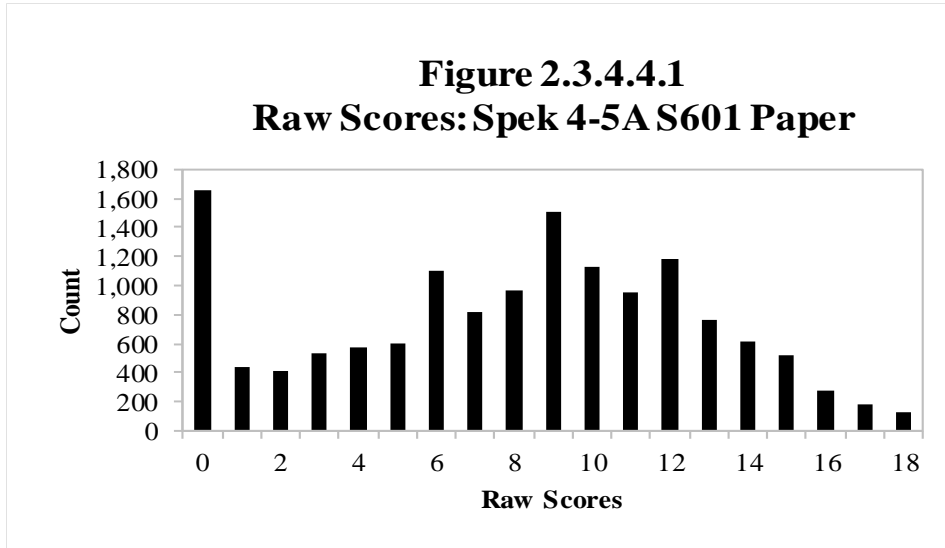
Table 2.3.4.4.2

Raw Score Descriptive Statistics: Spek 4-5 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	17,499	6	30	22.01	4.52
5	15,301	6	30	22.46	4.62
Total	32,800	6	30	22.22	4.57

**Figure 2.3.4.4.2
Raw Scores: Spek 4-5B/C S601 Paper**
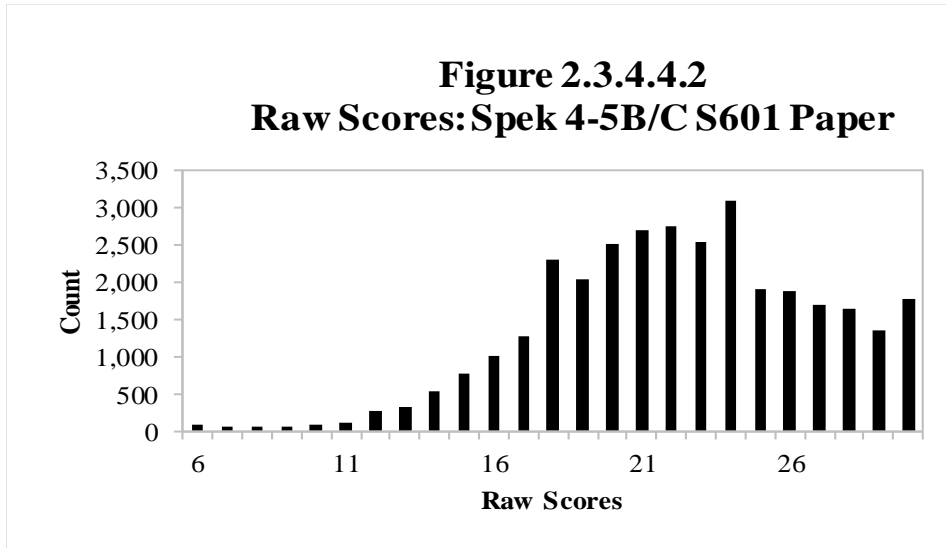
2.3.4.5 Grades 6-8

Table 2.3.4.5.1

Raw Score Descriptive Statistics: Spek 6-8 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	6,938	0	18	8.12	4.48
7	6,897	0	18	8.26	4.42
8	6,581	0	18	8.59	4.34
Total	20,416	0	18	8.32	4.42

**Figure 2.3.4.5.1
Raw Scores: Spek 6-8A S601 Paper**
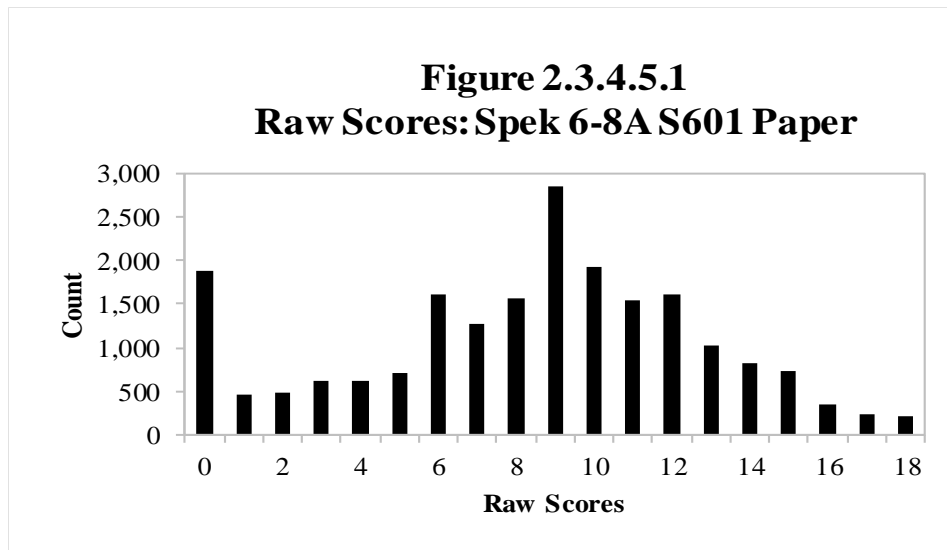
Table 2.3.4.5.2

Raw Score Descriptive Statistics: Spek 6-8 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	12,685	6	30	20.86	5.08
7	10,358	6	30	21.30	5.25
8	9,767	6	30	21.92	5.21
Total	32,810	6	30	21.32	5.19

**Figure 2.3.4.5.2
Raw Scores: Spek 6-8B/C S601 Paper**
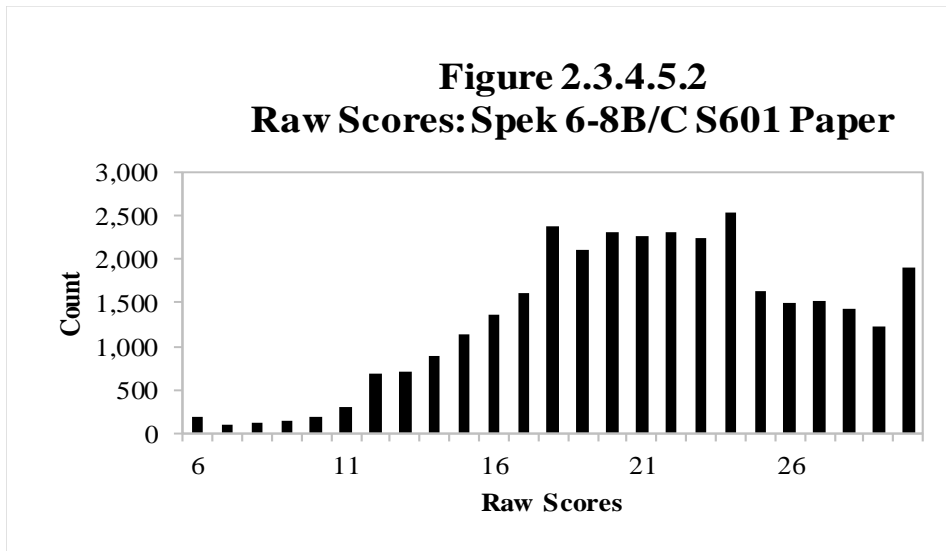
2.3.4.6 Grades 9-12

Table 2.3.4.6.1

Raw Score Descriptive Statistics: Spek 9-12 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	7,113	0	18	8.20	4.66
10	5,470	0	18	8.75	4.50
11	4,106	0	18	9.62	4.37
12	2,242	0	18	10.57	4.13
Total	18,931	0	18	8.95	4.56

**Figure 2.3.4.6.1
Raw Scores: Spek 9-12A S601 Paper**
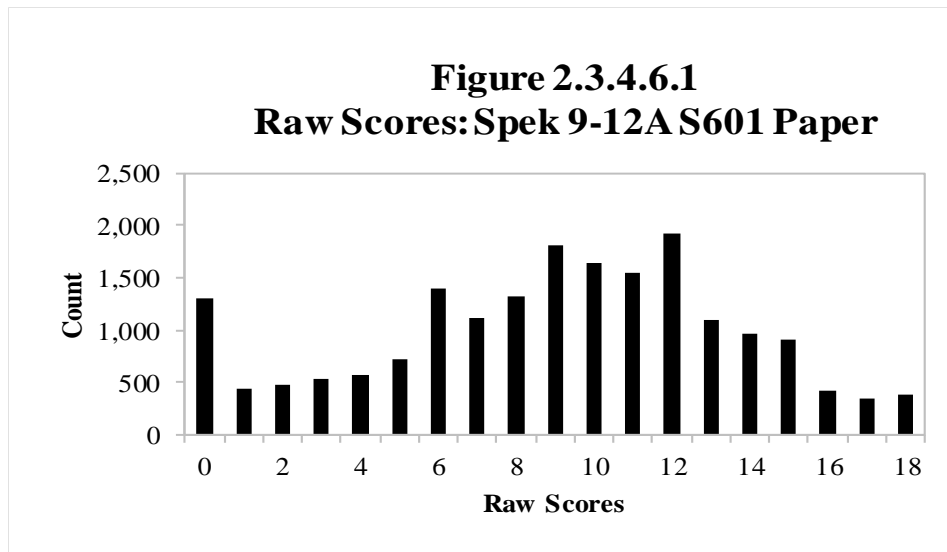
Table 2.3.4.6.2

Raw Score Descriptive Statistics: Spek 9-12 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	9,107	6	30	22.46	5.25
10	8,463	6	30	21.91	5.45
11	7,750	6	30	22.52	5.31
12	5,796	6	30	22.60	5.31
Total	31,116	6	30	22.35	5.34

Figure 2.3.4.6.2
Raw Scores: Spek 9-12B/C S601 Paper
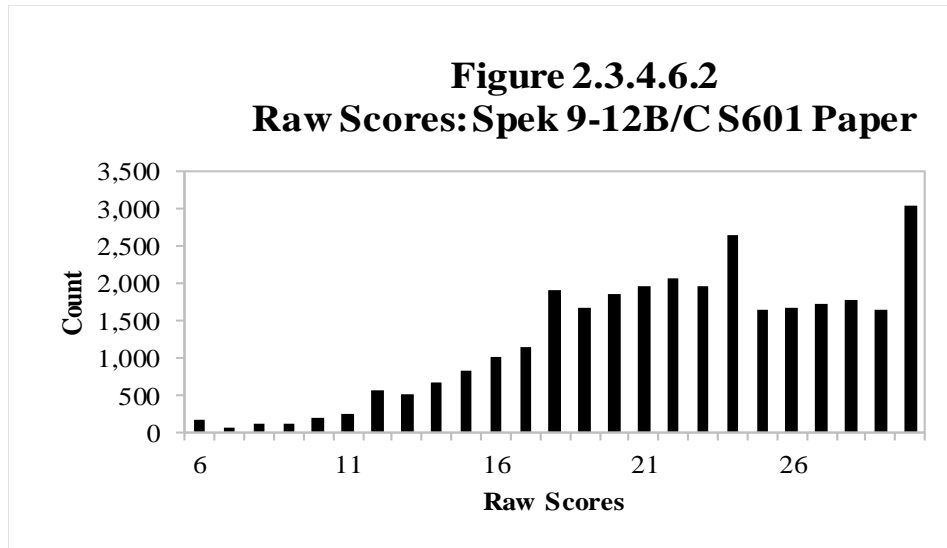
2.4   Scale Score Distribution

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 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	256,847	100	363	249.14	82.55
Total	256,847	100	363	249.14	82.55

Figure 2.4.1.0
Scale Scores: List K S601 Paper
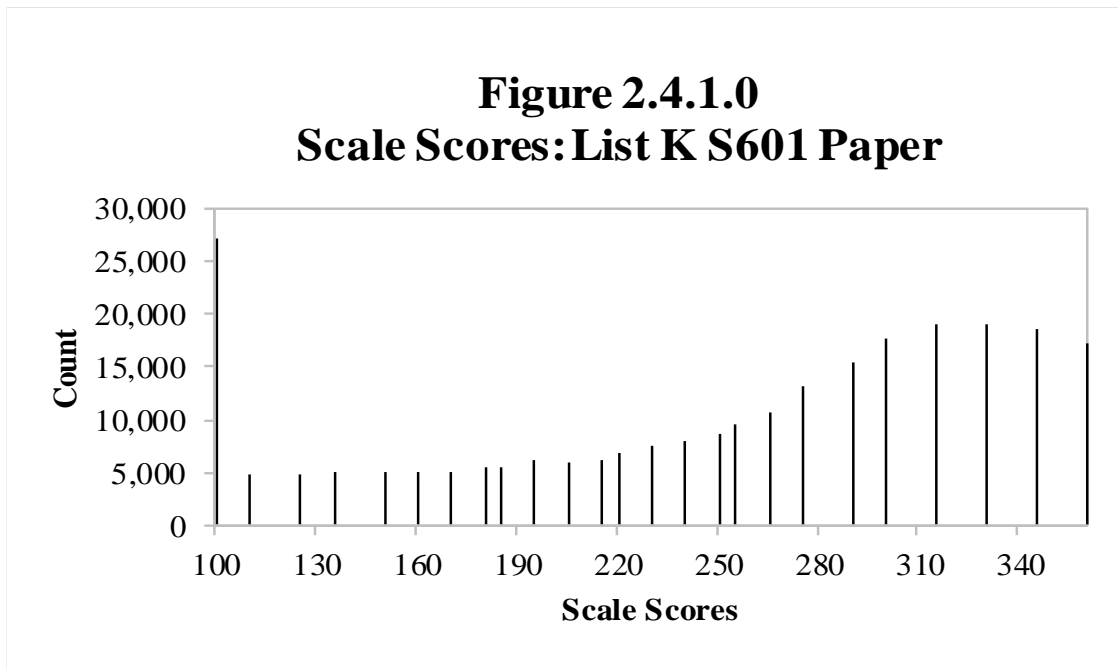
2.4.1.1 Grade 1

Table 2.4.1.1.1

Scale Score Descriptive Statistics: List 1 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	19,530	104	352	285.50	41.33
Total	19,530	104	352	285.50	41.33

Figure 2.4.1.1.1
Scale Scores: List 1A S601 Paper
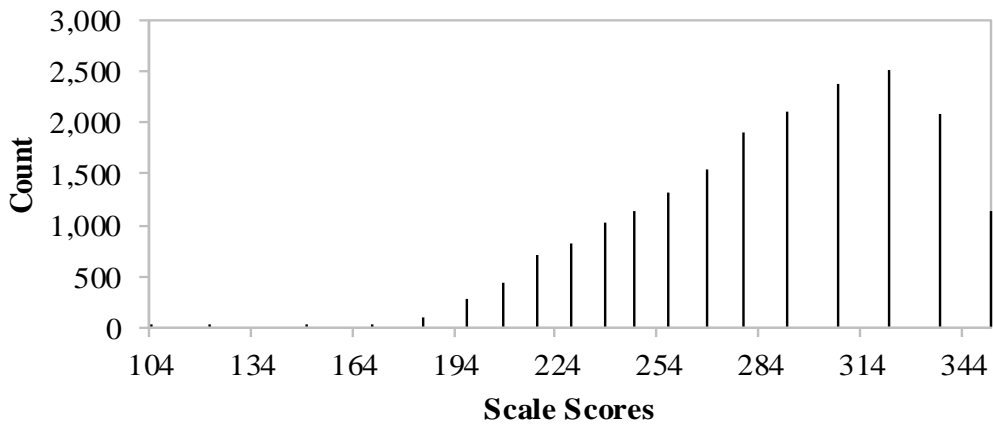
Table 2.4.1.1.2

Scale Score Descriptive Statistics: List 1 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	13,922	175	405	320.79	37.84
Total	13,922	175	405	320.79	37.84

Figure 2.4.1.1.2
Scale Scores: List 1B/C S601 Paper
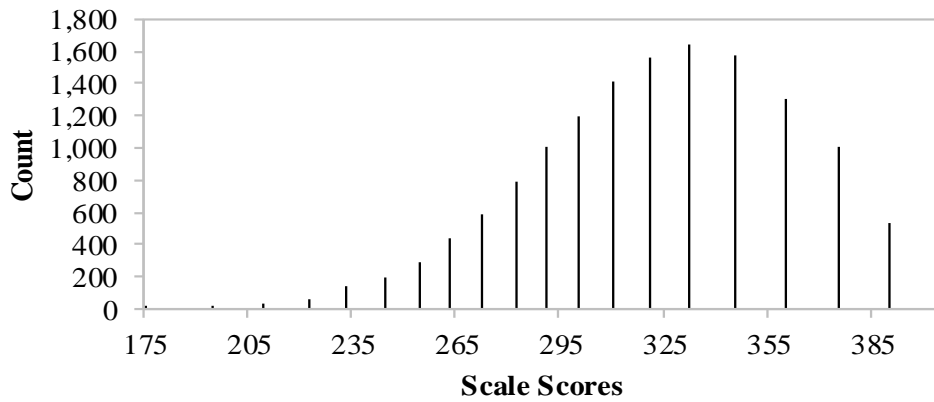
Table 2.4.1.1.3

Scale Score Descriptive Statistics: List 1 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	33,452	104	405	300.19	43.54
Total	33,452	104	405	300.19	43.54

Figure 2.4.1.1.3
Scale Scores: List 1 S601 Paper
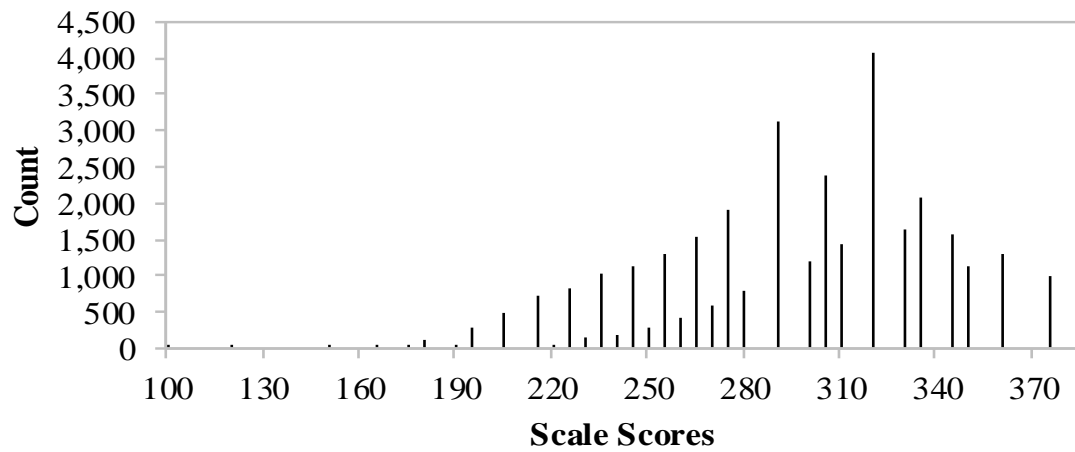
2.4.1.2 Grade 2

Table 2.4.1.2.1

Scale Score Descriptive Statistics: List 2 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	10,852	112	352	287.92	45.82
Total	10,852	112	352	287.92	45.82

**Figure 2.4.1.2.1
Scale Scores: List 2A S601 Paper**
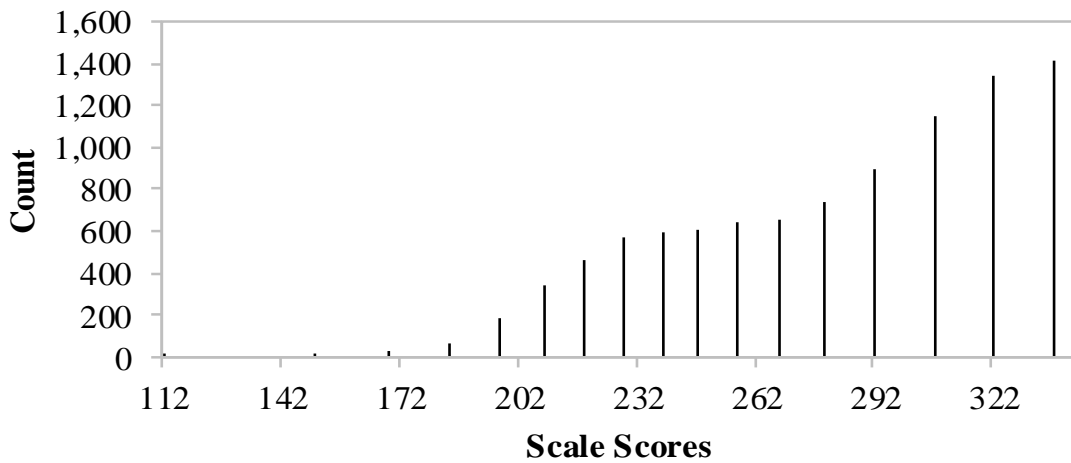
Table 2.4.1.2.2

Scale Score Descriptive Statistics: List 2 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	21,442	175	405	343.48	35.03
Total	21,442	175	405	343.48	35.03

Figure 2.4.1.2.2
Scale Scores: List 2B/C S601 Paper
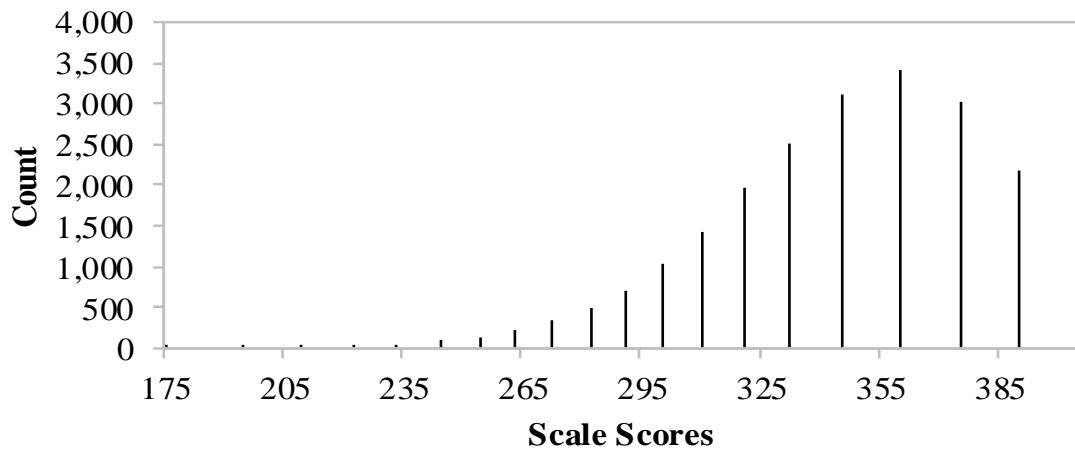
Table 2.4.1.2.3

Scale Score Descriptive Statistics: List 2 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	32,294	112	405	324.81	47.00
Total	32,294	112	405	324.81	47.00

Figure 2.4.1.2.3
Scale Scores: List 2 S601 Paper
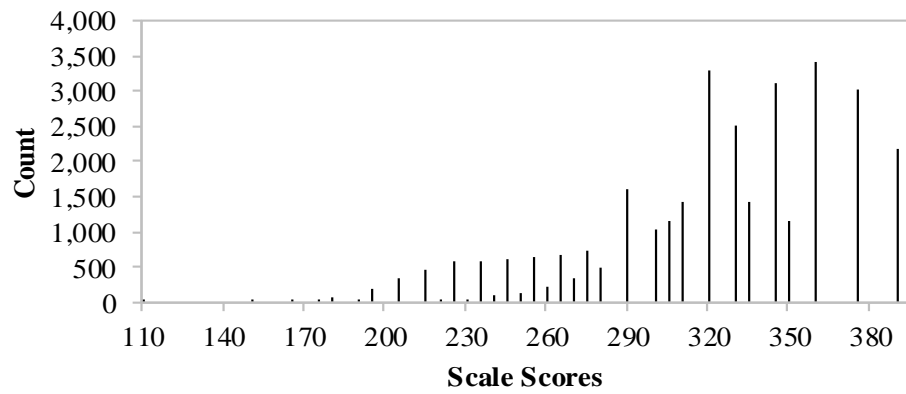
2.4.1.3 Grade 3

Table 2.4.1.3.1

Scale Score Descriptive Statistics: List 3 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	9,003	184	416	318.57	38.46
Total	9,003	184	416	318.57	38.46

Figure 2.4.1.3.1
Scale Scores: List 3A S601 Paper
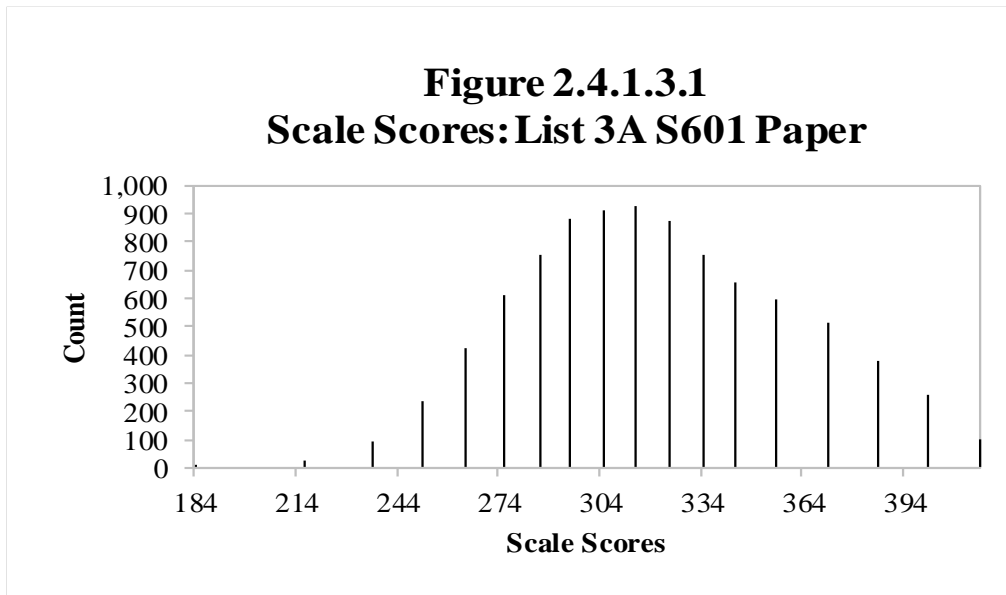
Table 2.4.1.3.2

Scale Score Descriptive Statistics: List 3 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	23,622	234	471	364.13	29.30
Total	23,622	234	471	364.13	29.30

Figure 2.4.1.3.2
Scale Scores: List 3B/C S601 Paper
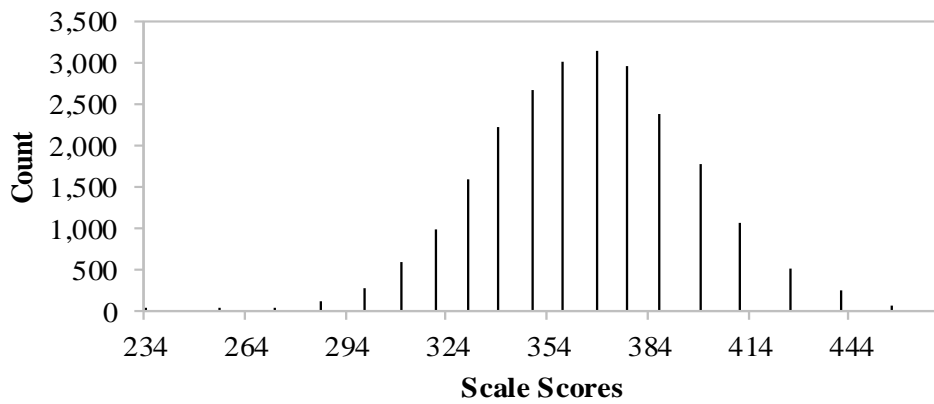
Table 2.4.1.3.3

Scale Score Descriptive Statistics: List 3 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	32,625	184	471	351.56	38.01
Total	32,625	184	471	351.56	38.01

Figure 2.4.1.3.3
Scale Scores: List 3 S601 Paper
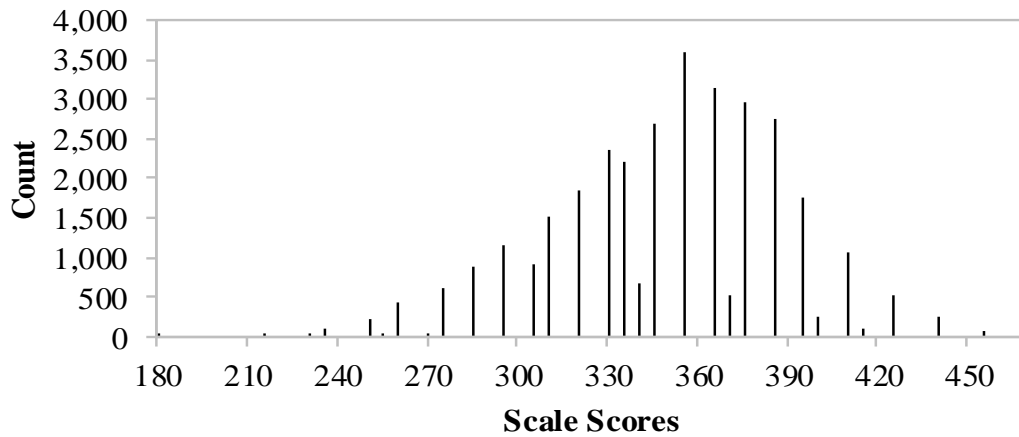
2.4.1.4 Grades 4-5

Table 2.4.1.4.1

Scale Score Descriptive Statistics: List 4-5 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	6,666	184	416	320.26	39.89
5	6,720	184	416	326.70	40.18
Total	13,386	184	416	323.49	40.16

Figure 2.4.1.4.1
Scale Scores: List 4-5A S601 Paper
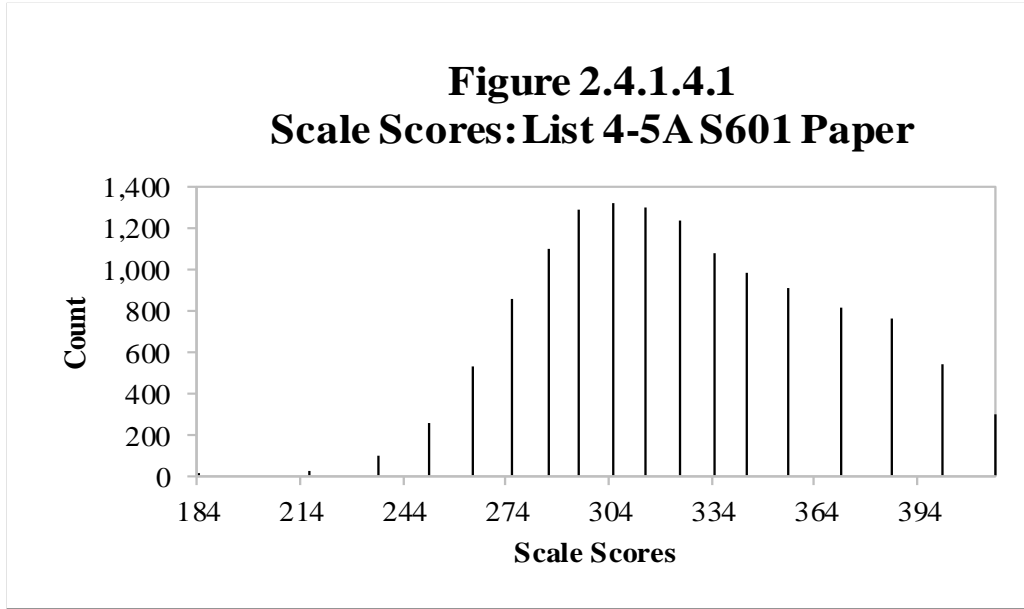
Table 2.4.1.4.2

Scale Score Descriptive Statistics: List 4-5 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	16,905	201	471	382.22	30.88
5	14,913	256	471	390.13	32.71
Total	31,818	201	471	385.93	31.99

Figure 2.4.1.4.2
Scale Scores: List 4-5B/C S601 Paper
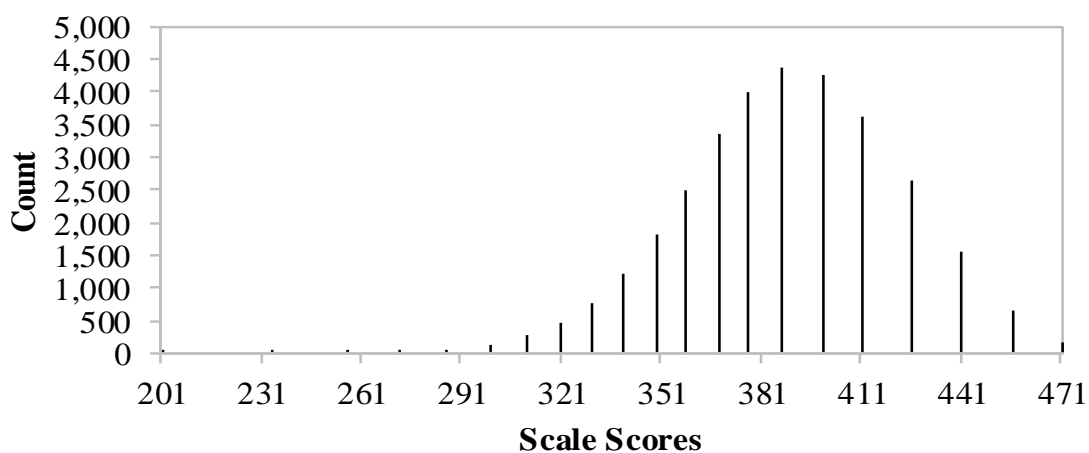
Table 2.4.1.4.3

Scale Score Descriptive Statistics: List 4-5 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	23,571	184	471	364.70	43.73
5	21,633	184	471	370.42	45.83
Total	45,204	184	471	367.44	44.84

Figure 2.4.1.4.3
Scale Scores: List 4-5 S601 Paper
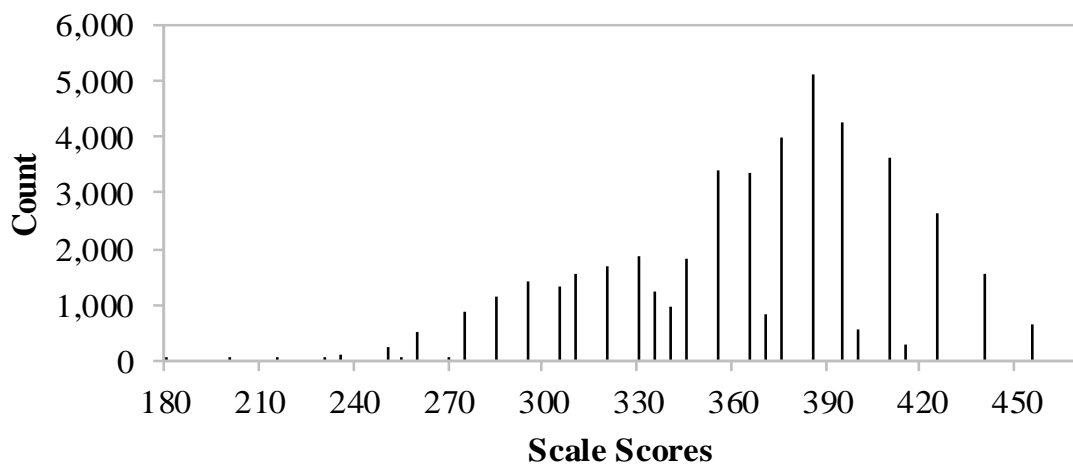
2.4.1.5 Grades 6-8

Table 2.4.1.5.1

Scale Score Descriptive Statistics: List 6-8 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	6,494	181	424	317.45	38.87
7	6,486	181	424	319.33	39.28
8	6,144	132	424	323.26	40.19
Total	19,124	132	424	319.96	39.51

Figure 2.4.1.5.1
Scale Scores: List 6-8A S601 Paper
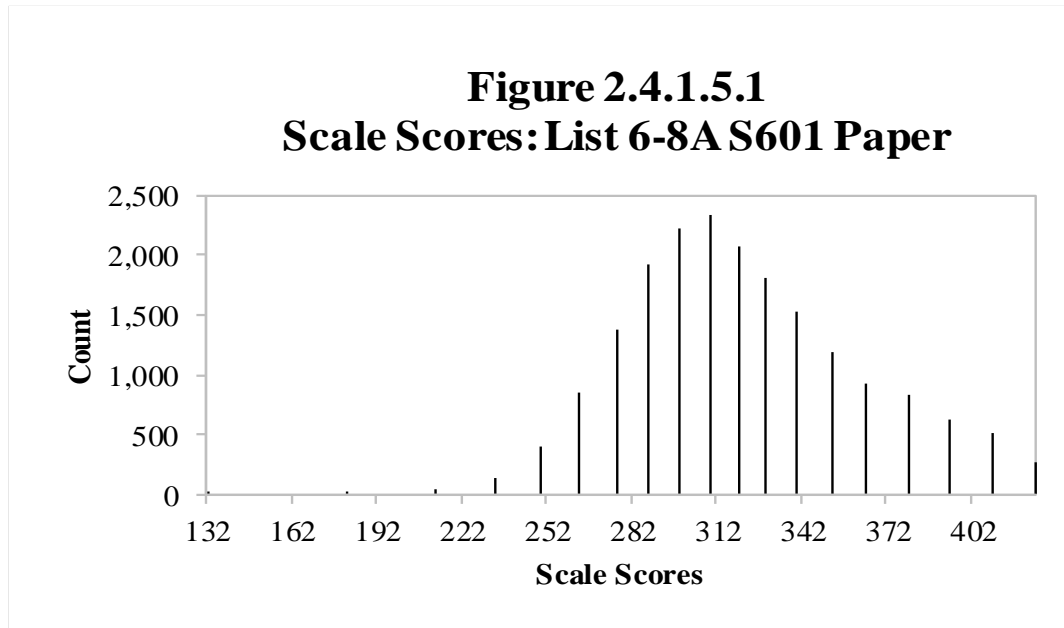
Table 2.4.1.5.2

Scale Score Descriptive Statistics: List 6-8 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	12,180	255	476	387.57	31.02
7	9,921	226	476	395.63	32.13
8	9,329	255	476	401.47	33.10
Total	31,430	226	476	394.24	32.51

Figure 2.4.1.5.2
Scale Scores: List 6-8B/C S601 Paper
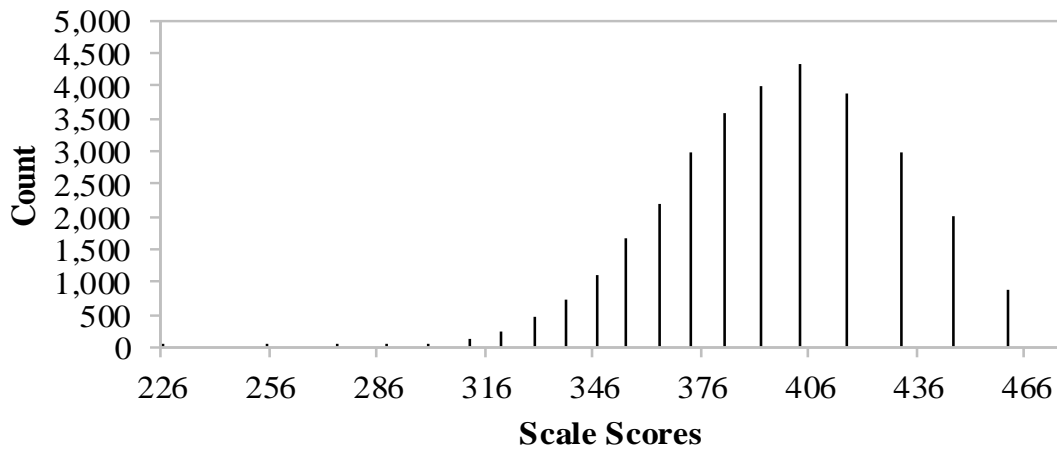
Table 2.4.1.5.3

Scale Score Descriptive Statistics: List 6-8 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	18,674	181	476	363.19	47.63
7	16,407	181	476	365.46	51.24
8	15,473	132	476	370.41	52.59
Total	50,554	132	476	366.14	50.45

Figure 2.4.1.5.3
Scale Scores: List 6-8 S601 Paper
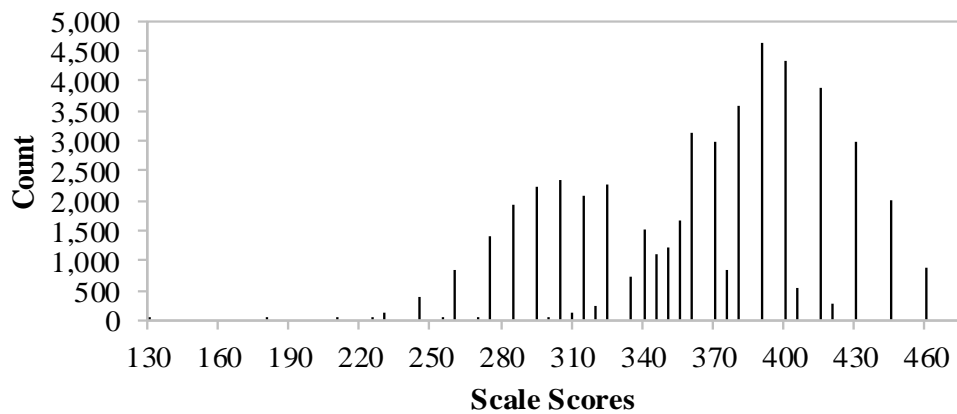
2.4.1.6 Grades 9-12

Table 2.4.1.6.1

Scale Score Descriptive Statistics: List 9-12 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	6,626	172	432	324.88	37.26
10	5,135	205	432	327.01	36.88
11	3,835	172	432	333.22	37.25
12	2,136	228	432	338.03	36.09
Total	17,732	172	432	328.89	37.29

Figure 2.4.1.6.1
Scale Scores: List 9-12A S601 Paper
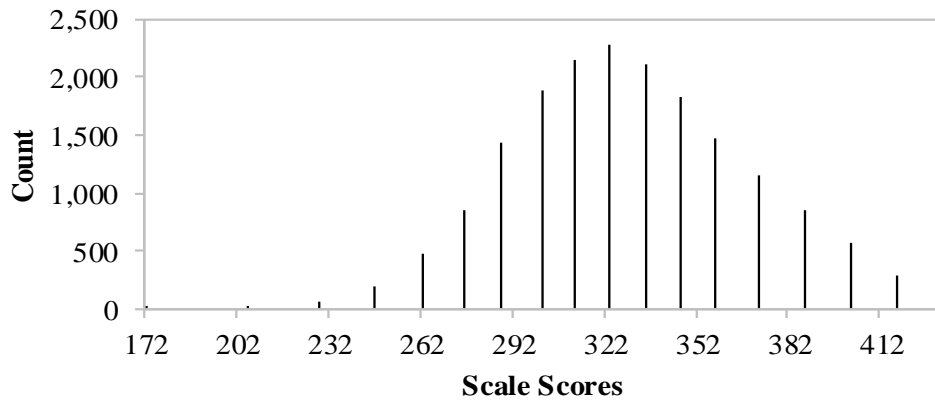
Table 2.4.1.6.2

Scale Score Descriptive Statistics: List 9-12 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	8,680	263	493	400.69	33.00
10	8,060	232	493	400.29	34.05
11	7,378	263	493	403.85	34.02
12	5,519	283	493	401.26	34.79
Total	29,637	232	493	401.47	33.91

Figure 2.4.1.6.2
Scale Scores: List 9-12B/C S601 Paper
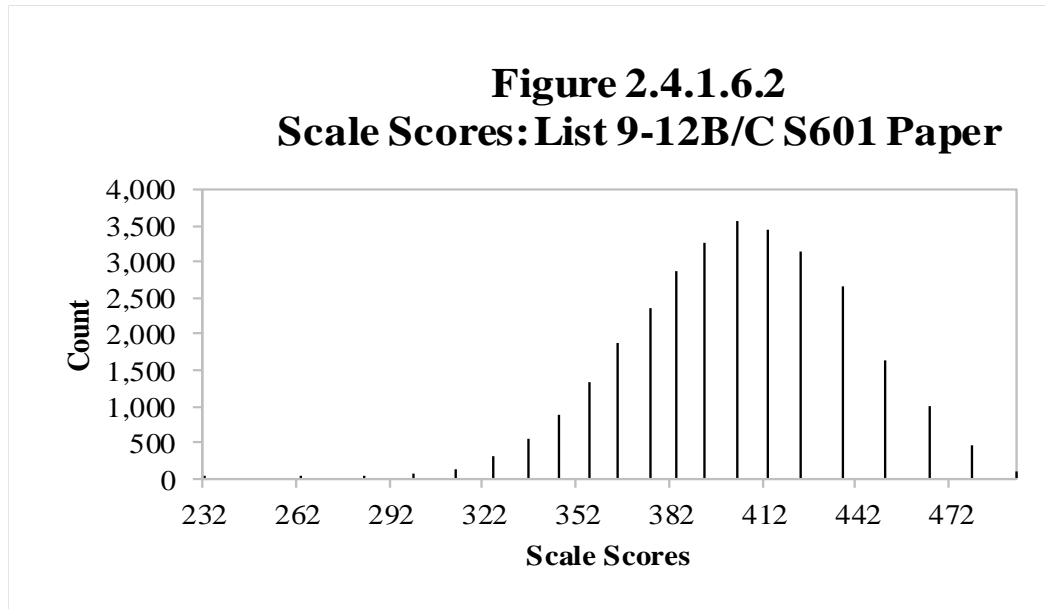
Table 2.4.1.6.3

Scale Score Descriptive Statistics: List 9-12 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	15,306	172	493	367.87	51.28
10	13,195	205	493	371.77	50.14
11	11,213	172	493	379.69	48.57
12	7,655	228	493	383.62	45.17
Total	47,369	172	493	374.30	49.74

Figure 2.4.1.6.3
Scale Scores: List 9-12 S601 Paper
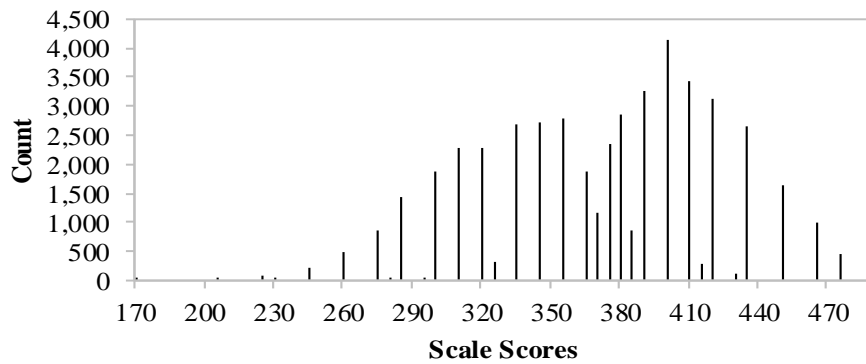
2.4.2 Reading

2.4.2.0 Kindergarten

Table 2.4.2.0

Scale Score Descriptive Statistics: Read K S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	256,837	100	290	171.80	66.56
Total	256,837	100	290	171.80	66.56

Figure 2.4.2.0
Scale Scores: Read K S601 Paper
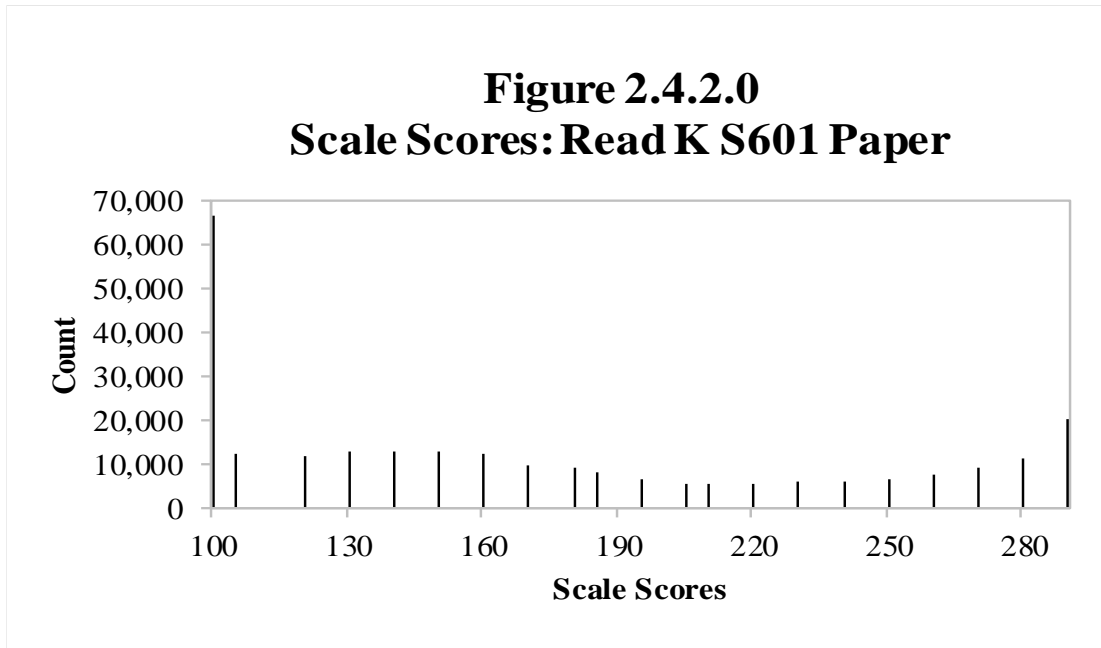
2.4.2.1 Grade 1

Table 2.4.2.1.1

Scale Score Descriptive Statistics: Read 1 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	18,223	141	361	270.46	22.56
Total	18,223	141	361	270.46	22.56

Figure 2.4.2.1.1
Scale Scores: Read 1A S601 Paper
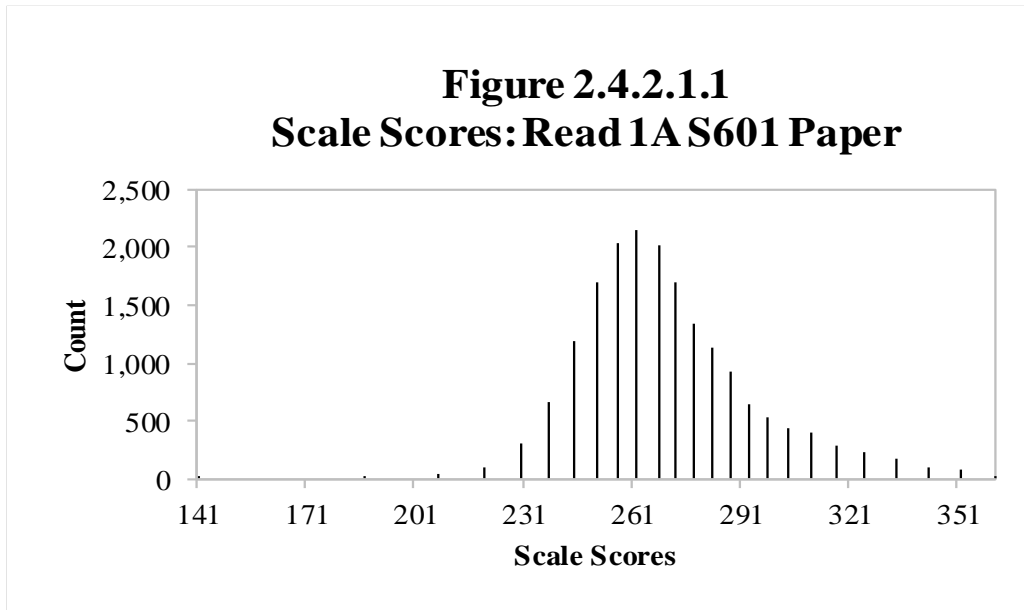
Table 2.4.2.1.2

Scale Score Descriptive Statistics: Read 1 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	12,099	236	394	300.70	22.34
Total	12,099	236	394	300.70	22.34

**Figure 2.4.2.1.2
Scale Scores: Read 1B/C S601 Paper**
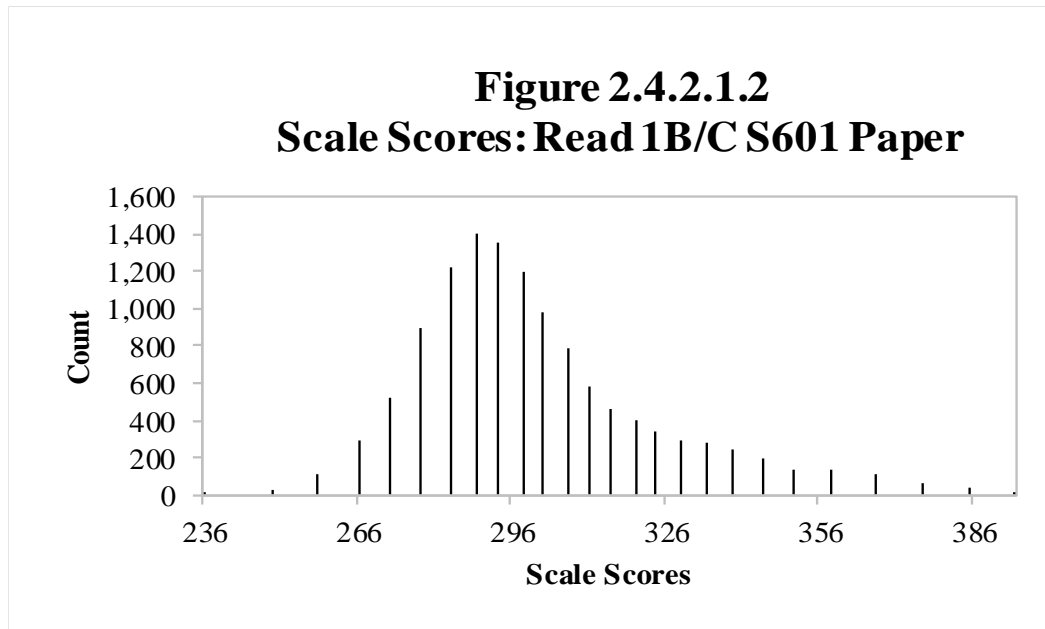
Table 2.4.2.1.3

Scale Score Descriptive Statistics: Read 1 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	30,322	141	394	282.52	26.91
Total	30,322	141	394	282.52	26.91

Figure 2.4.2.1.3
Scale Scores: Read 1 S601 Paper
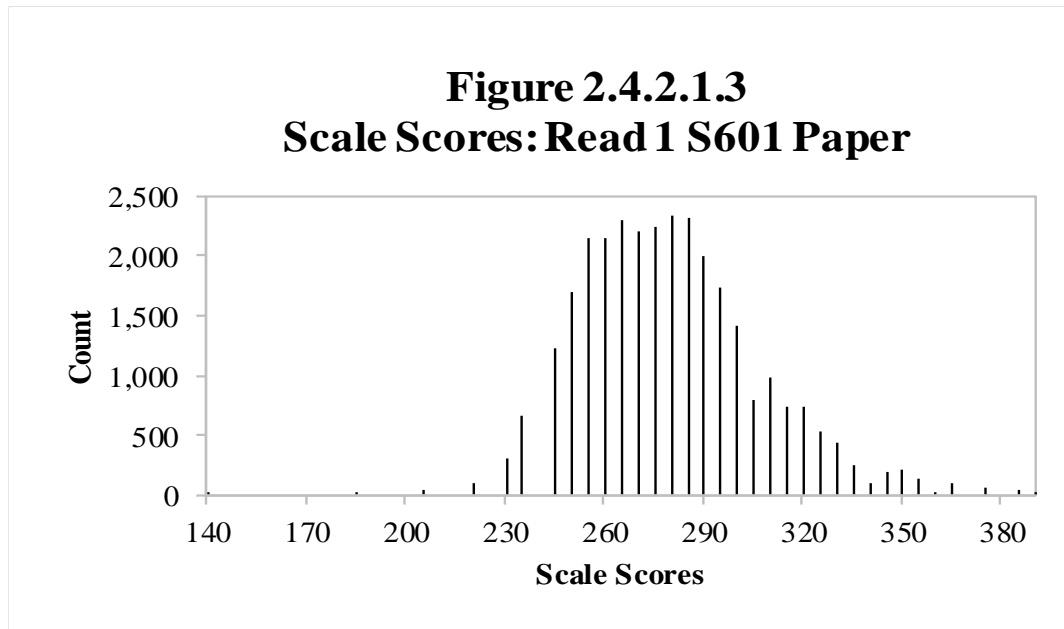
2.4.2.2 Grade 2

Table 2.4.2.2.1

Scale Score Descriptive Statistics: Read 2 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	10,005	187	361	280.55	27.97
Total	10,005	187	361	280.55	27.97

**Figure 2.4.2.2.1
Scale Scores: Read 2A S601 Paper**
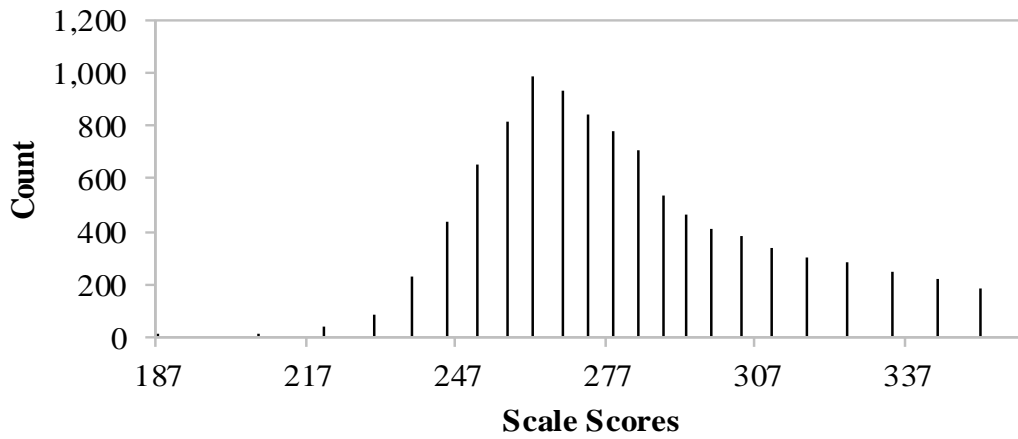
Table 2.4.2.2.2

Scale Score Descriptive Statistics: Read 2 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	18,748	216	394	318.36	28.15
Total	18,748	216	394	318.36	28.15

Figure 2.4.2.2.2
Scale Scores: Read 2B/C S601 Paper
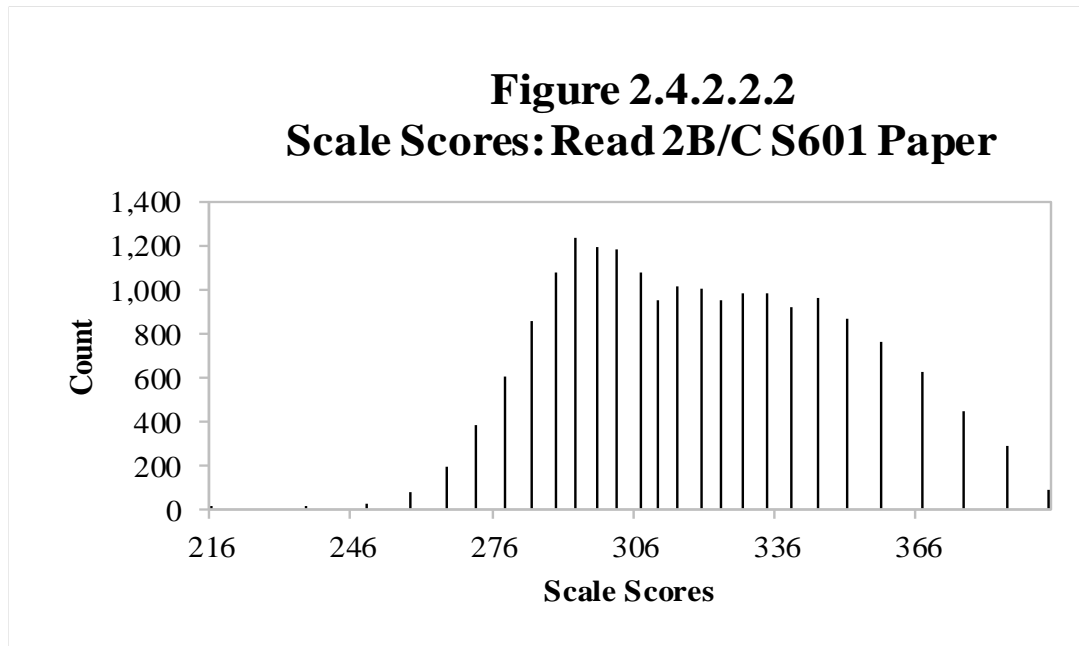
Table 2.4.2.2.3

Scale Score Descriptive Statistics: Read 2 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	28,753	187	394	305.20	33.36
Total	28,753	187	394	305.20	33.36

Figure 2.4.2.2.3
Scale Scores: Read 2 S601 Paper
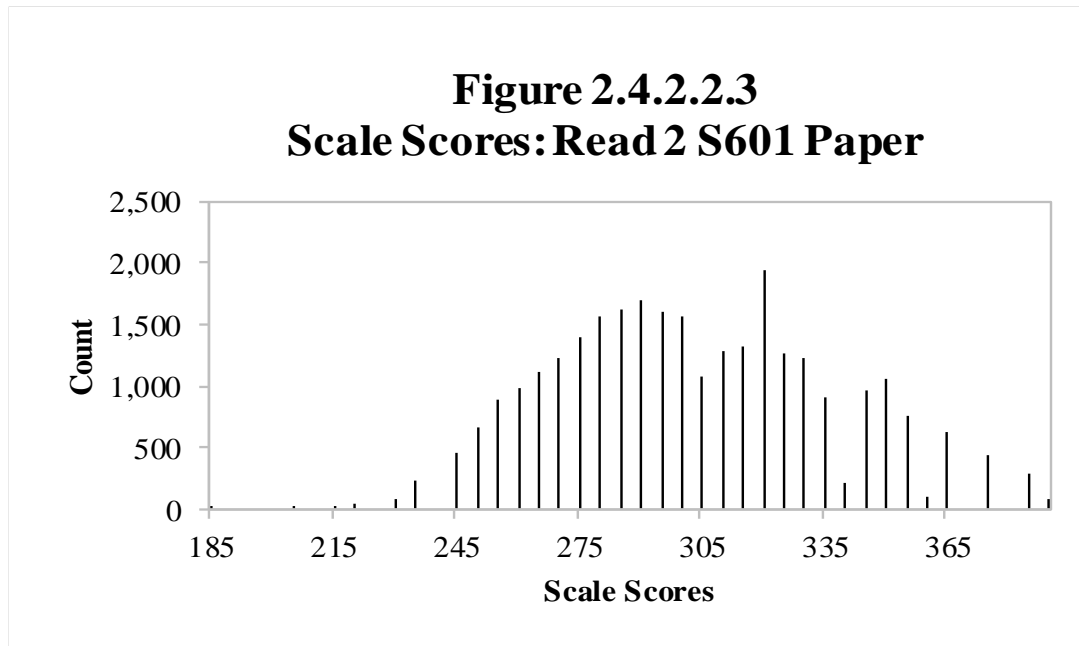
2.4.2.3 Grade 3

Table 2.4.2.3.1

Scale Score Descriptive Statistics: Read 3 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	8,259	213	385	298.59	25.59
Total	8,259	213	385	298.59	25.59

Figure 2.4.2.3.1
Scale Scores: Read 3A S601 Paper
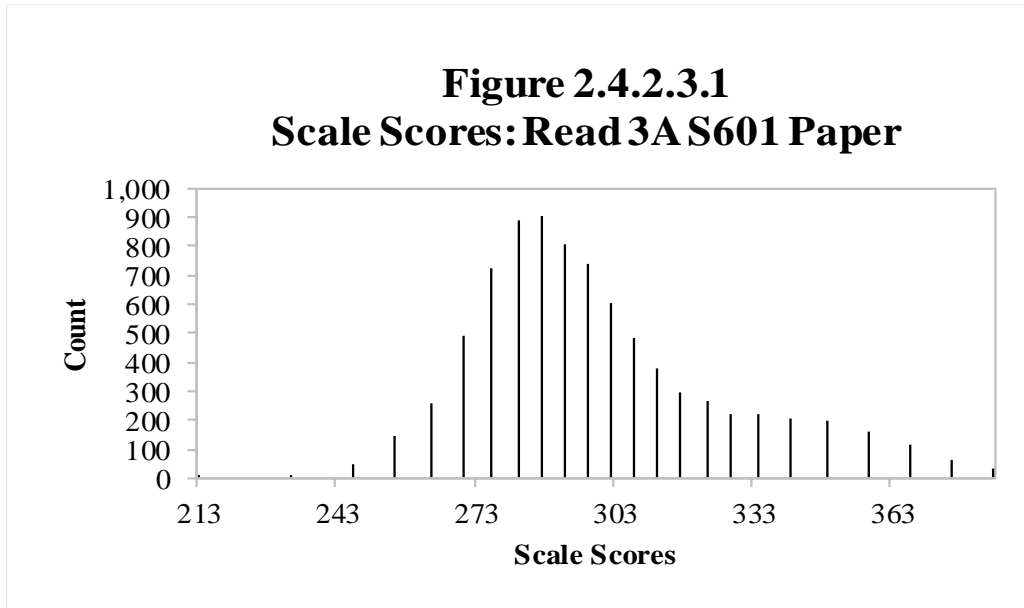
Table 2.4.2.3.2

Scale Score Descriptive Statistics: Read 3 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	20,631	158	445	342.24	19.11
Total	20,631	158	445	342.24	19.11

Figure 2.4.2.3.2
Scale Scores: Read 3B/C S601 Paper
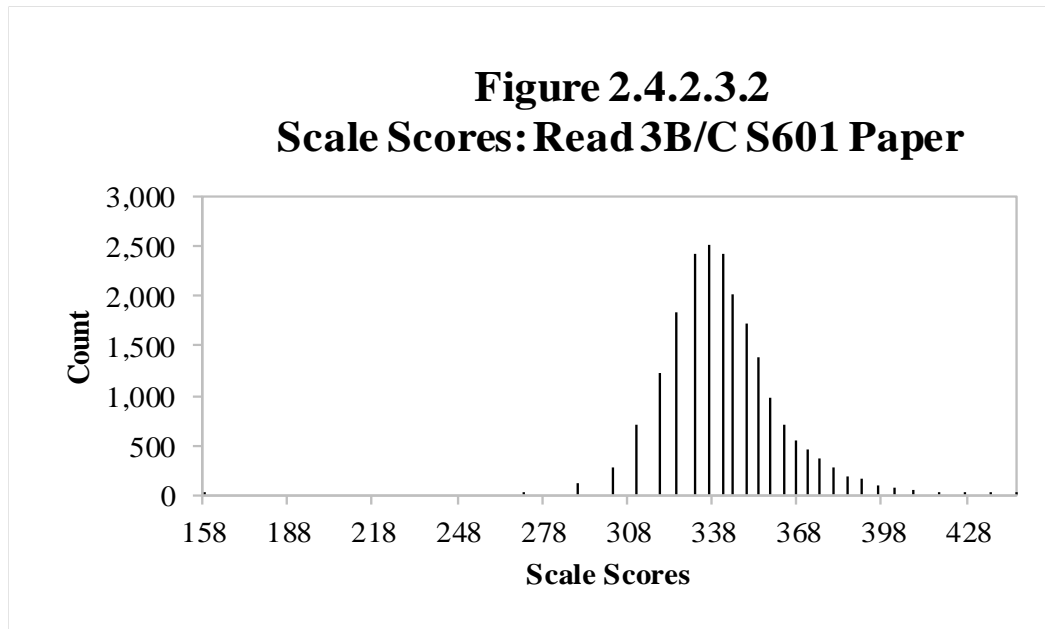
Table 2.4.2.3.3

Scale Score Descriptive Statistics: Read 3 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	28,890	158	445	329.76	28.93
Total	28,890	158	445	329.76	28.93

Figure 2.4.2.3.3
Scale Scores: Read 3 S601 Paper
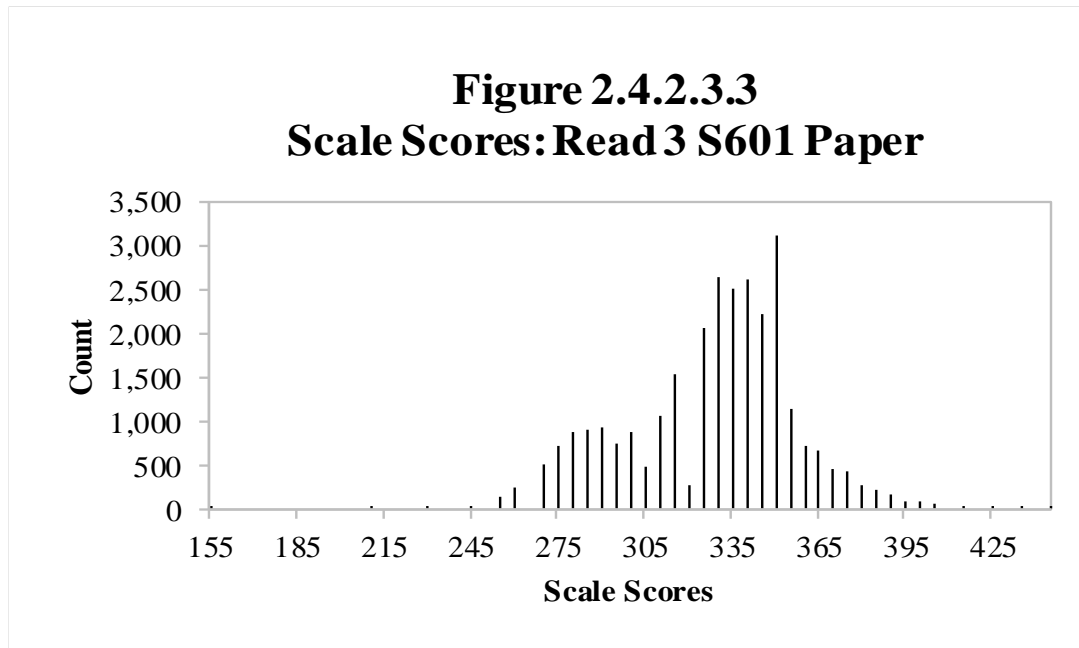
2.4.2.4 Grades 4-5

Table 2.4.2.4.1

Scale Score Descriptive Statistics: Read 4-5 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	6,165	213	385	306.22	27.97
5	6,410	175	385	311.68	29.59
Total	12,575	175	385	309.00	28.94

Figure 2.4.2.4.1
Scale Scores: Read 4-5 A S601 Paper
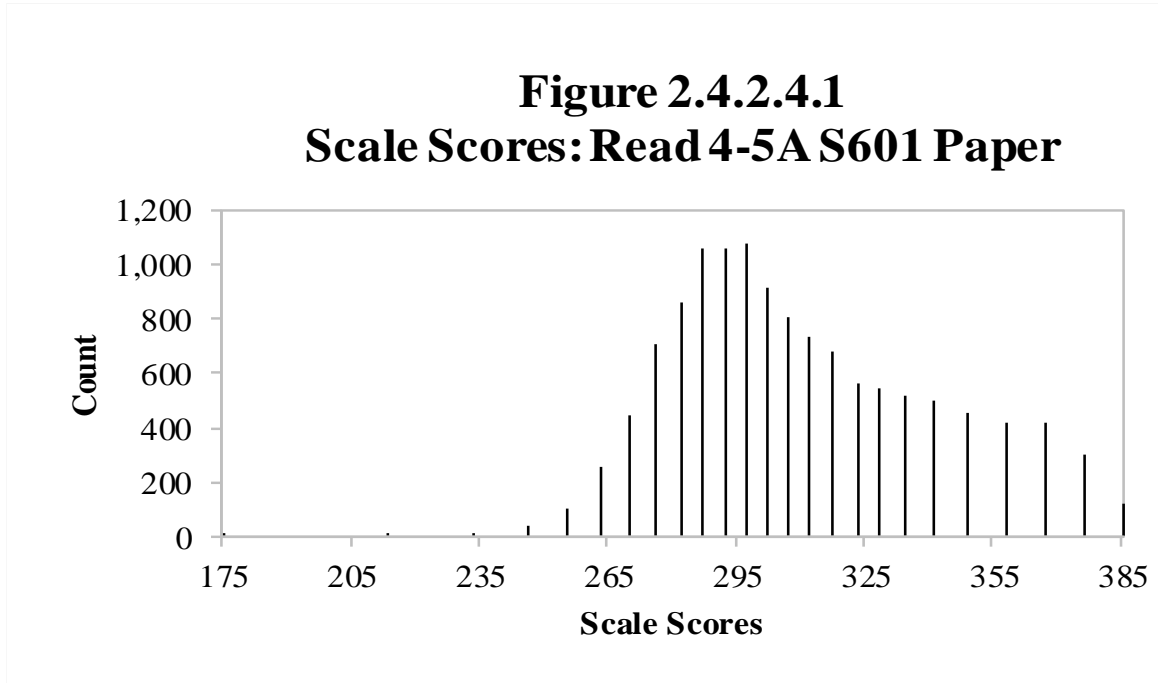
Table 2.4.2.4.2

Scale Score Descriptive Statistics: Read 4-5 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	15,075	175	445	354.38	22.07
5	13,579	271	445	361.19	24.59
Total	28,654	175	445	357.61	23.55

Figure 2.4.2.4.2
Scale Scores: Read 4-5B/C S601 Paper
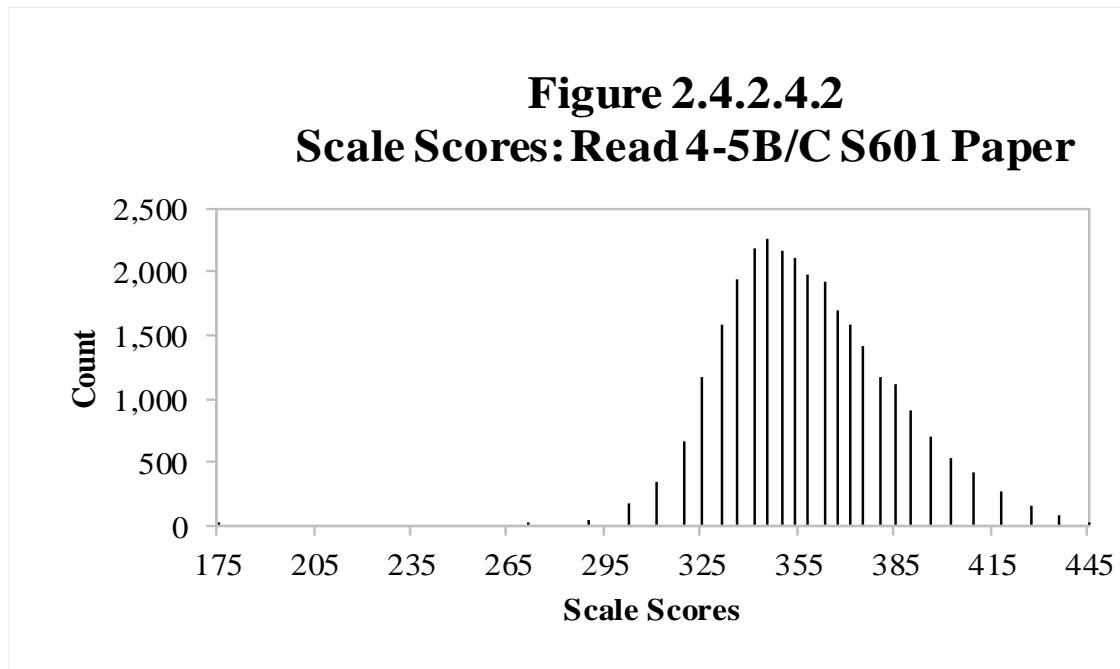
Table 2.4.2.4.3

Scale Score Descriptive Statistics: Read 4-5 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	21,240	175	445	340.40	32.41
5	19,989	175	445	345.31	35.01
Total	41,229	175	445	342.78	33.78

Figure 2.4.2.4.3
Scale Scores: Read 4-5 S601 Paper
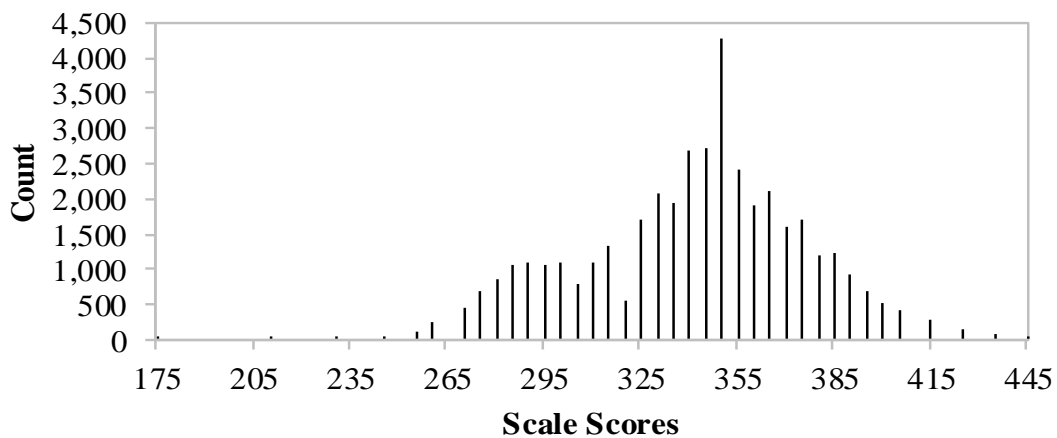
2.4.2.5 Grades 6-8

Table 2.4.2.5.1

Scale Score Descriptive Statistics: Read 6-8 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	6,236	200	422	326.75	24.67
7	6,239	245	422	330.19	25.86
8	5,992	200	422	334.82	27.57
Total	18,467	200	422	330.53	26.25

Figure 2.4.2.5.1
Scale Scores: Read 6-8A S601 Paper
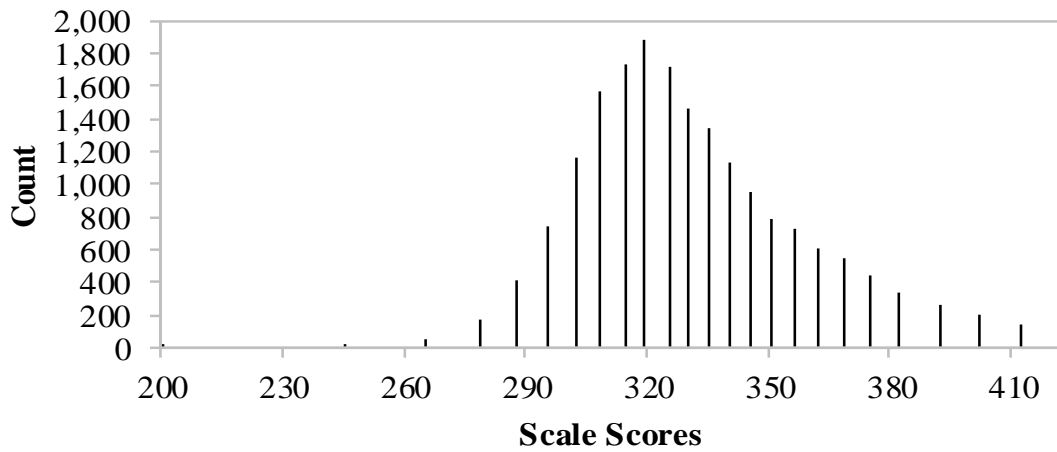
Table 2.4.2.5.2

Scale Score Descriptive Statistics: Read 6-8 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	11,303	200	452	354.67	20.93
7	9,315	277	452	362.75	23.66
8	8,859	277	452	370.71	25.86
Total	29,477	200	452	362.05	24.28

Figure 2.4.2.5.2
Scale Scores: Read 6-8B/C S601 Paper
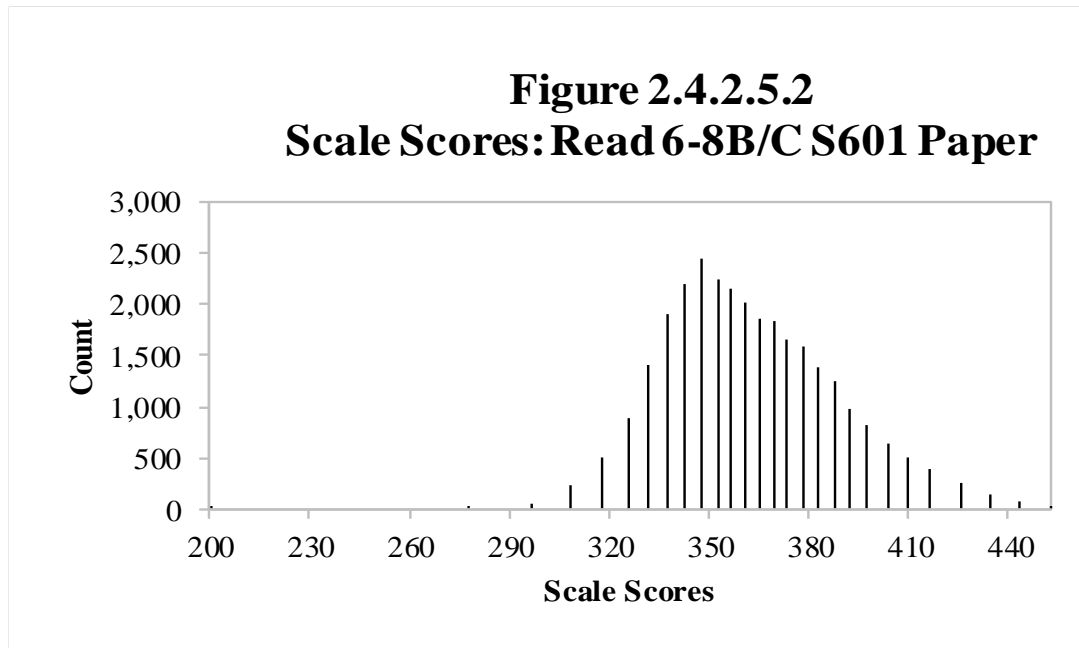
Table 2.4.2.5.3

Scale Score Descriptive Statistics: Read 6-8 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	17,539	200	452	344.74	26.03
7	15,554	245	452	349.69	29.29
8	14,851	200	452	356.23	31.87
Total	47,944	200	452	349.91	29.38

Figure 2.4.2.5.3
Scale Scores: Read 6-8 S601 Paper
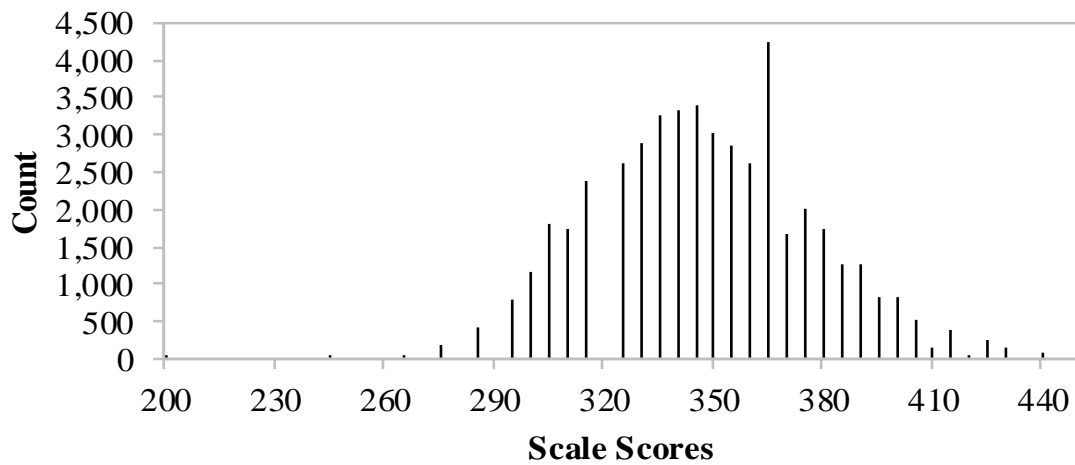
2.4.2.6 Grades 9-12

Table 2.4.2.6.1

Scale Score Descriptive Statistics: Read 9-12 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	6,582	233	433	351.23	27.68
10	5,113	252	433	354.51	27.83
11	3,818	285	433	360.75	29.06
12	2,119	285	433	364.29	29.00
Total	17,632	233	433	355.81	28.58

**Figure 2.4.2.6.1
Scale Scores: Read 9-12A S601 Paper**
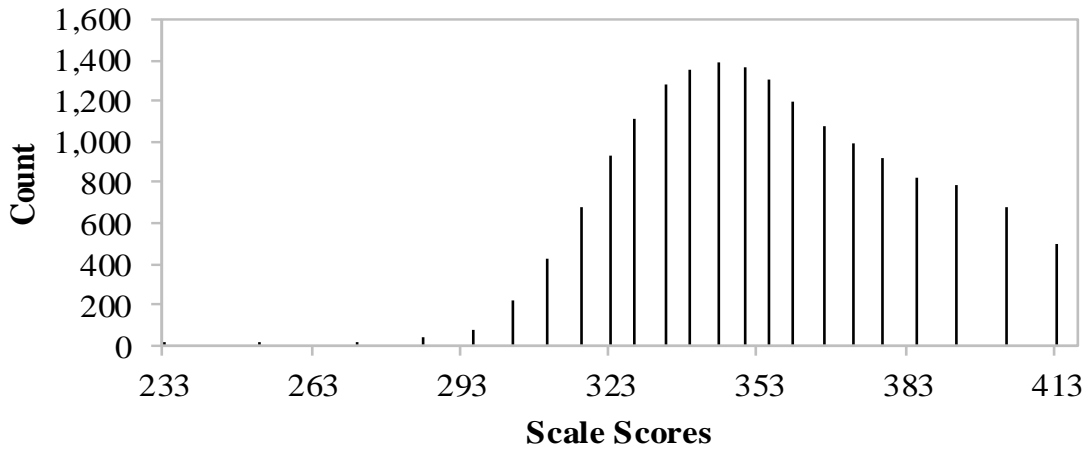
Table 2.4.2.6.2

Scale Score Descriptive Statistics: Read 9-12 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	7,923	297	470	387.79	24.83
10	7,490	297	470	390.34	25.70
11	6,837	297	470	394.52	26.59
12	5,267	316	470	392.29	26.79
Total	27,517	297	470	391.02	26.02

Figure 2.4.2.6.2
Scale Scores: Read 9-12B/C S601 Paper
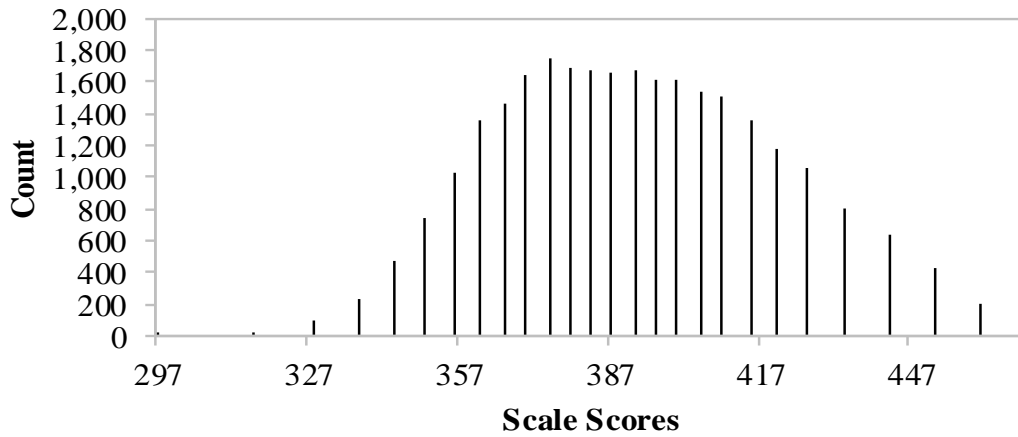
Table 2.4.2.6.3

Scale Score Descriptive Statistics: Read 9-12 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	14,505	233	470	371.20	31.87
10	12,603	252	470	375.80	31.88
11	10,655	285	470	382.42	31.91
12	7,386	285	470	384.26	30.22
Total	45,149	233	470	377.27	32.04

Figure 2.4.2.6.3
Scale Scores: Read 9-12 S601 Paper
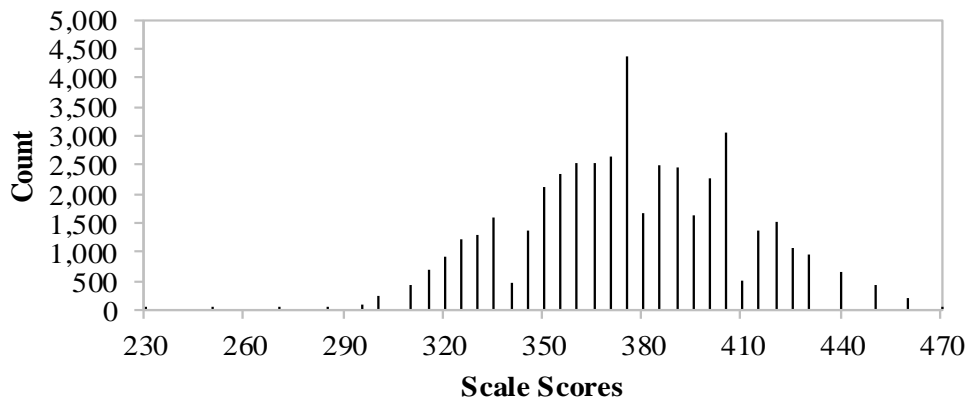
2.4.3 Writing

2.4.3.0 Kindergarten

Table 2.4.3.0

Scale Score Descriptive Statistics: Writ K S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	256,834	100	339	186.74	66.97
Total	256,834	100	339	186.74	66.97

Figure 2.4.3.0
Scale Scores: Writ K S601 Paper
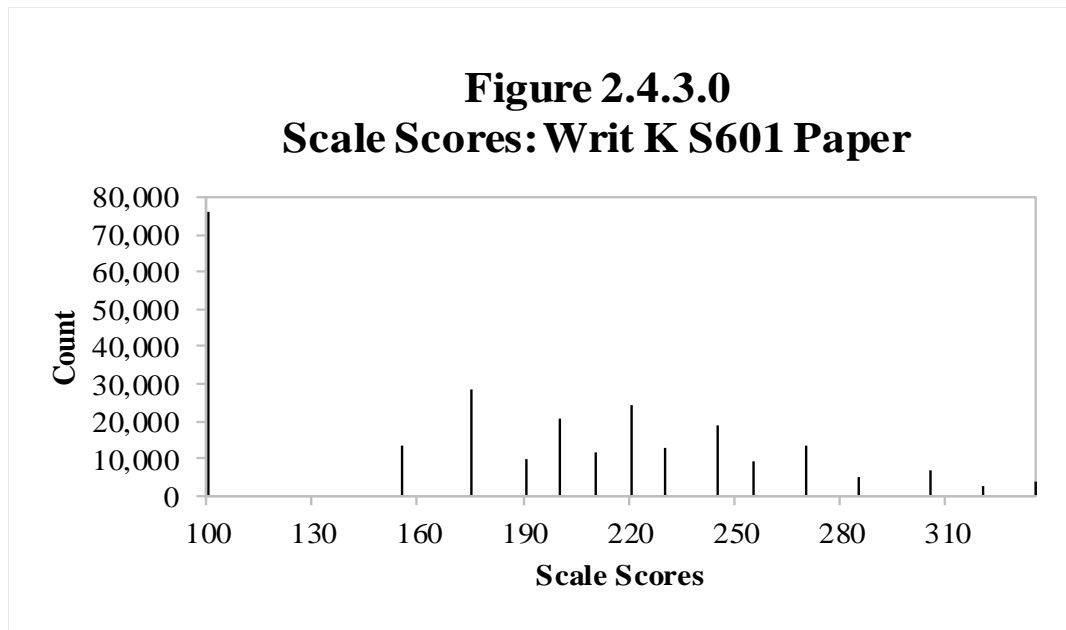
2.4.3.1 Grade 1

Table 2.4.3.1.1

Scale Score Descriptive Statistics: Writ 1 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	23,675	111	321	226.17	37.54
Total	23,675	111	321	226.17	37.54

**Figure 2.4.3.1.1
Scale Scores: Writ 1A S601 Paper**
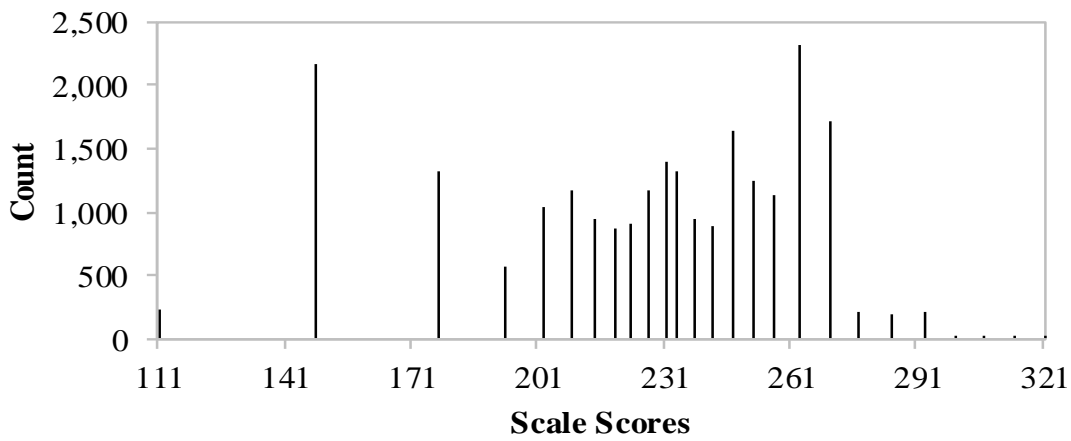
Table 2.4.3.1.2

Scale Score Descriptive Statistics: Writ 1 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,827	111	389	260.39	48.79
Total	15,827	111	389	260.39	48.79

Figure 2.4.3.1.2
Scale Scores: Writ 1B/C S601 Paper
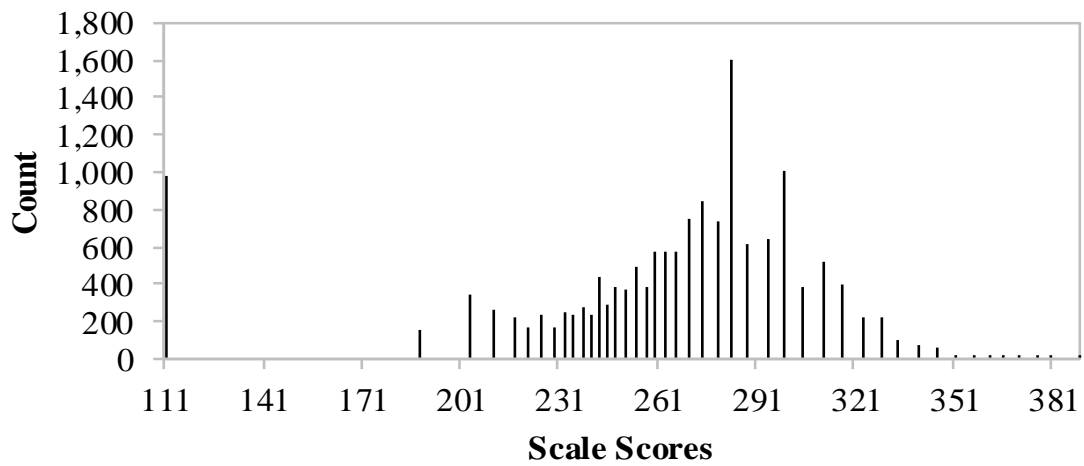
Table 2.4.3.1.3

Scale Score Descriptive Statistics: Writ 1 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	39,502	111	389	239.88	45.60
Total	39,502	111	389	239.88	45.60

**Figure 2.4.3.1.3
Scale Scores: Writ 1 S601 Paper**
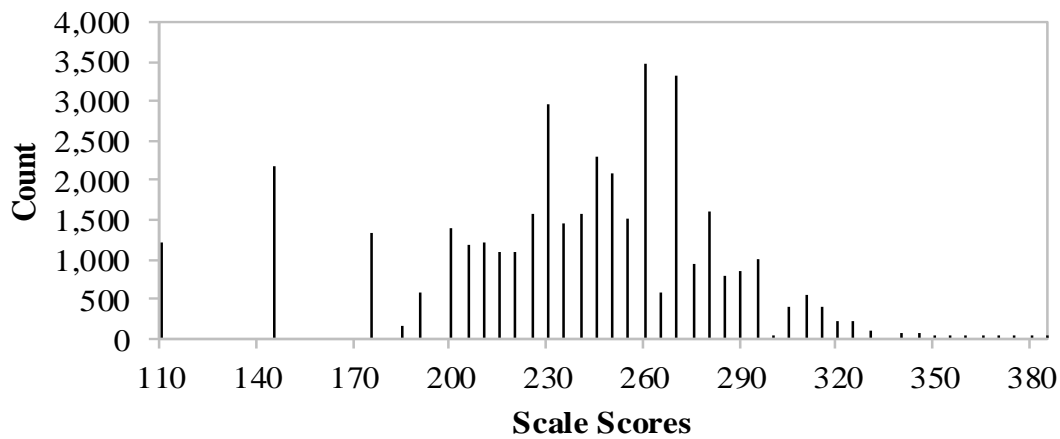
2.4.3.2 Grade 2

Table 2.4.3.2.1

Scale Score Descriptive Statistics: Writ 2 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	12,283	133	347	235.89	41.08
Total	12,283	133	347	235.89	41.08

Figure 2.4.3.2.1
Scale Scores: Writ 2A S601 Paper
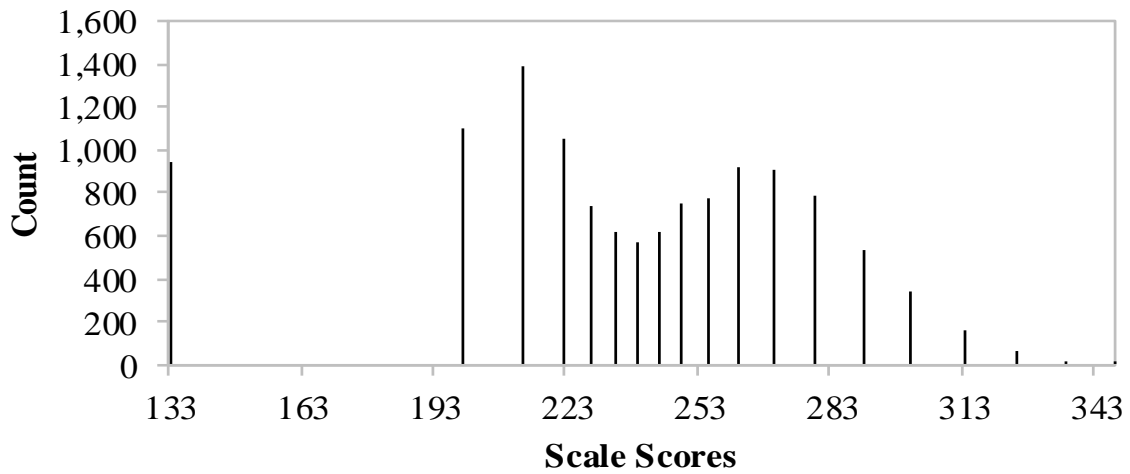
Table 2.4.3.2.2

Scale Score Descriptive Statistics: Writ 2 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	22,974	133	394	286.72	38.85
Total	22,974	133	394	286.72	38.85

Figure 2.4.3.2.2
Scale Scores: Writ B/C S601 Paper
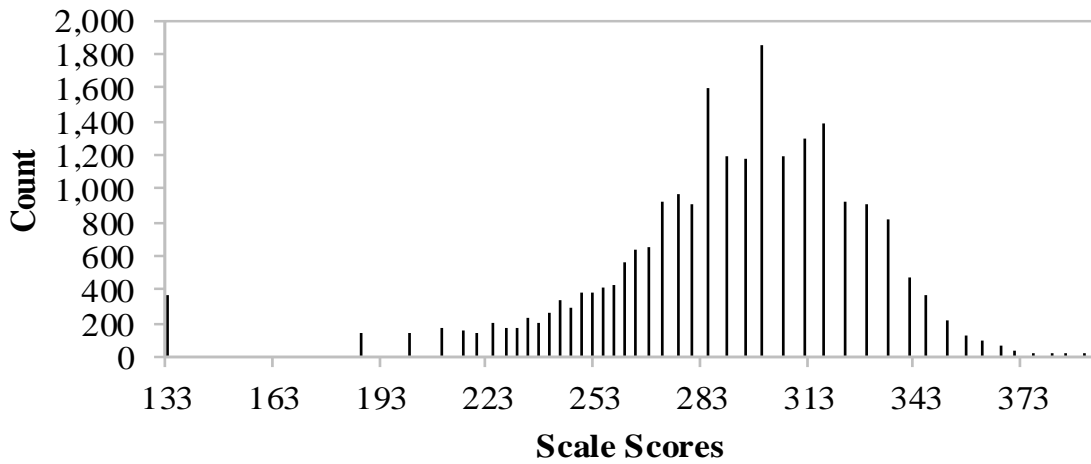
Table 2.4.3.2.3

Scale Score Descriptive Statistics: Writ 2 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	35,257	133	394	269.01	46.45
Total	35,257	133	394	269.01	46.45

Figure 2.4.3.2.3
Scale Scores: Writ 2 S601 Paper
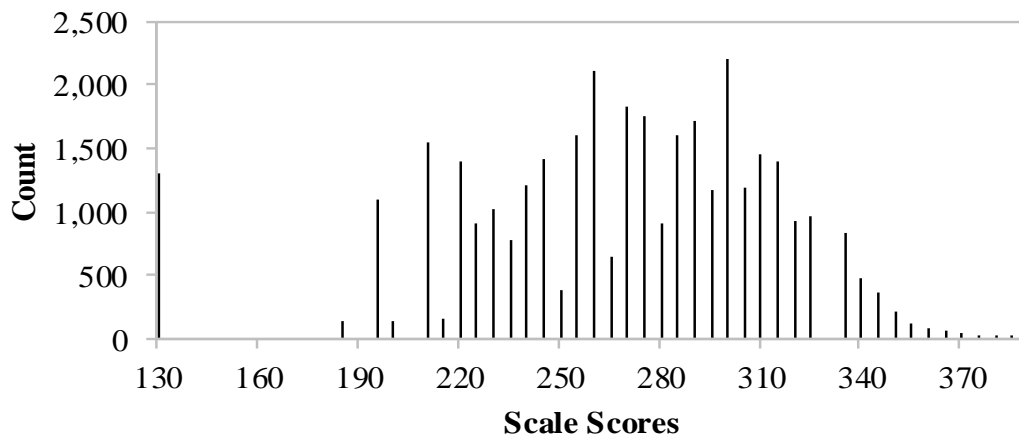
2.4.3.3 Grade 3

Table 2.4.3.3.1

Scale Score Descriptive Statistics: Writ 3 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	9,880	133	367	246.35	40.67
Total	9,880	133	367	246.35	40.67

Figure 2.4.3.3.1
Scale Scores: Writ 3A S601 Paper
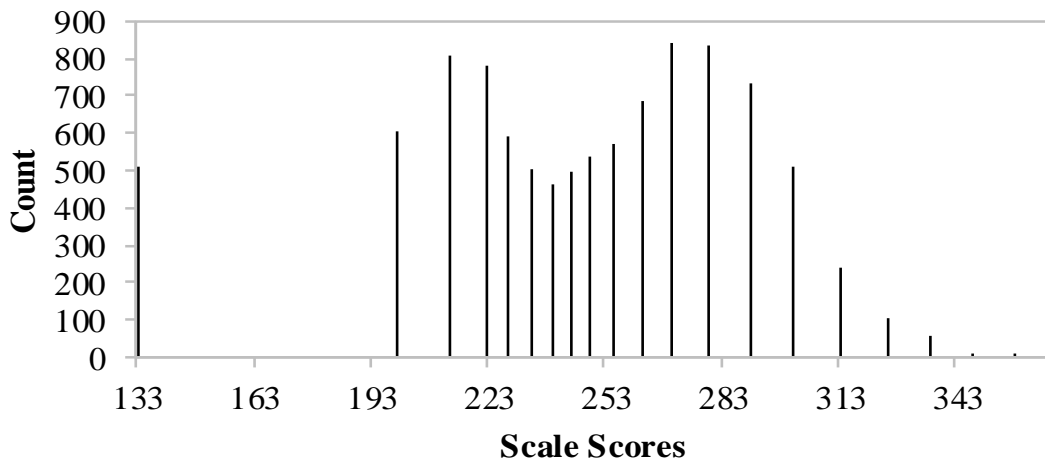
Table 2.4.3.3.2

Scale Score Descriptive Statistics: Writ 3 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	25,187	133	404	302.93	34.45
Total	25,187	133	404	302.93	34.45

Figure 2.4.3.3.2
Scale Scores: Writ 3B/C S601 Paper
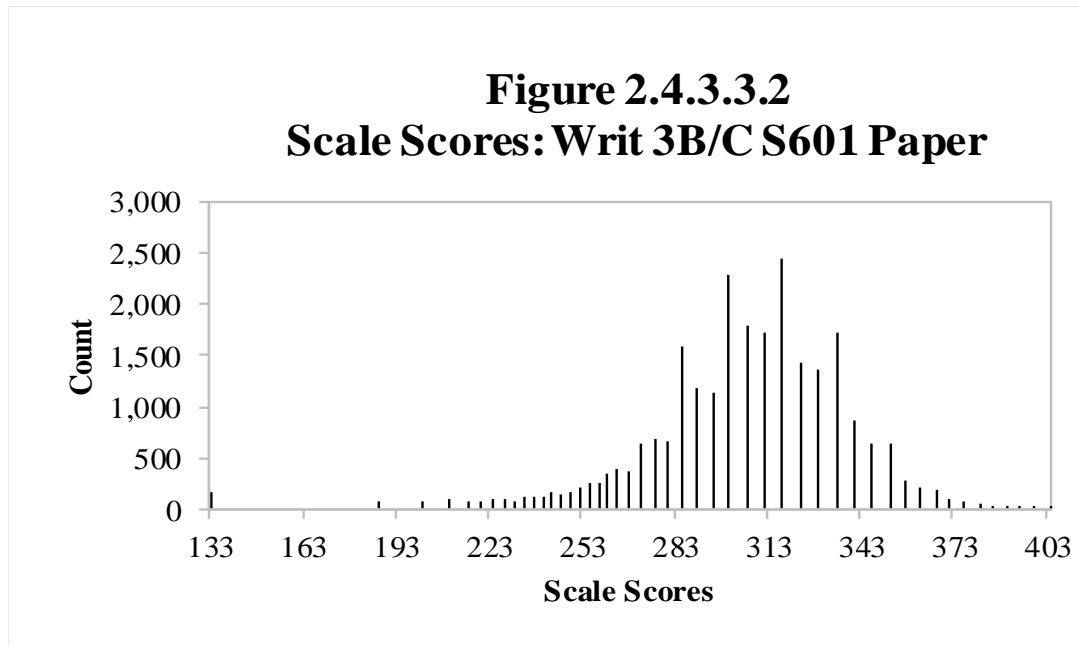
Table 2.4.3.3.3

Scale Score Descriptive Statistics: Writ 3 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	35,067	133	404	286.99	44.34
Total	35,067	133	404	286.99	44.34

Figure 2.4.3.3.3
Scale Scores: Writ 3 S601 Paper
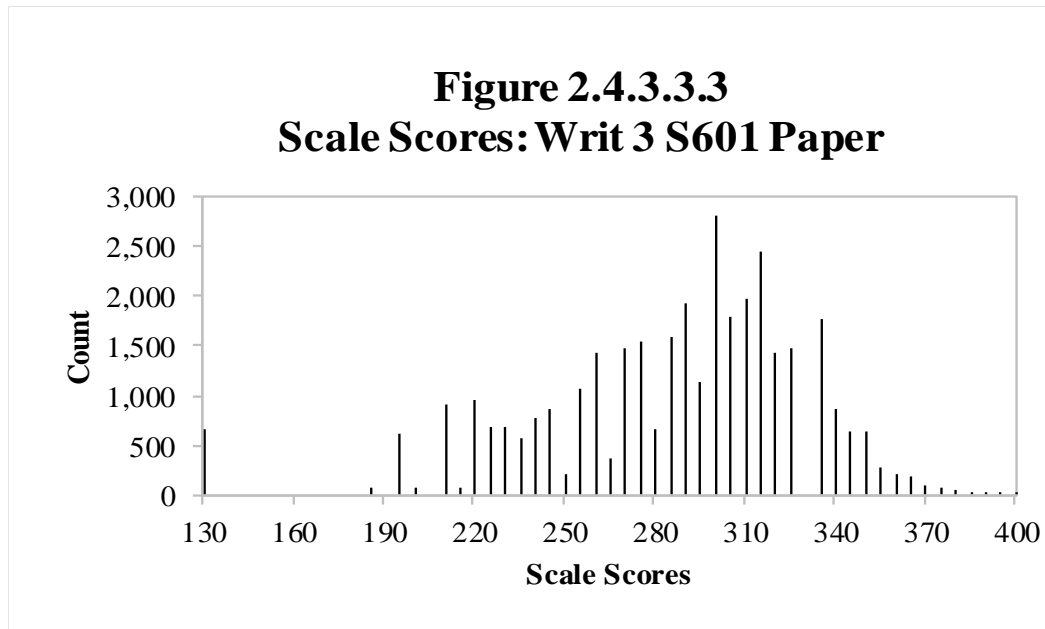
2.4.3.4 Grades 4-5

Table 2.4.3.4.1

Scale Score Descriptive Statistics: Writ 4-5 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	7,188	155	388	266.96	43.62
5	7,211	155	388	274.98	41.72
Total	14,399	155	388	270.97	42.87

Figure 2.4.3.4.1
Scale Scores: Writ 4-5A S601 Paper
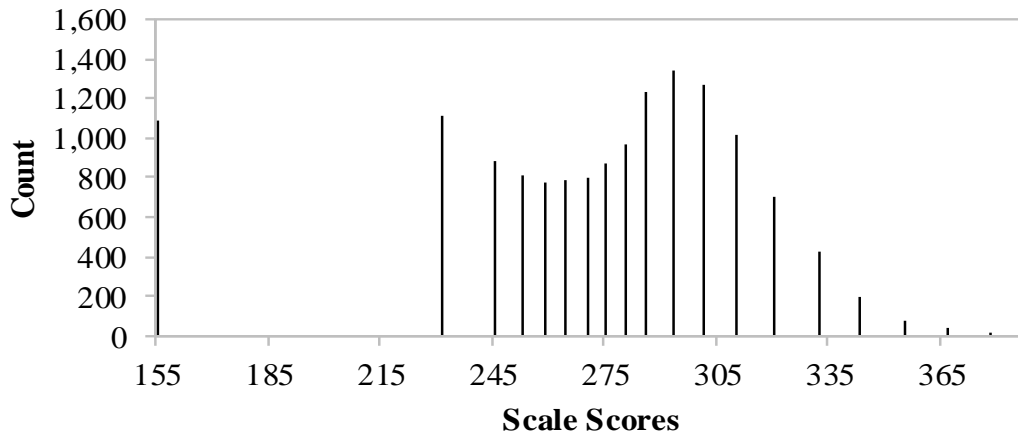
Table 2.4.3.4.2

Scale Score Descriptive Statistics: Writ 4-5 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	17,525	155	438	335.52	29.74
5	15,324	155	438	346.96	33.35
Total	32,849	155	438	340.85	31.99

Figure 2.4.3.4.2
Scale Scores: Writ 4-5B/C S601 Paper
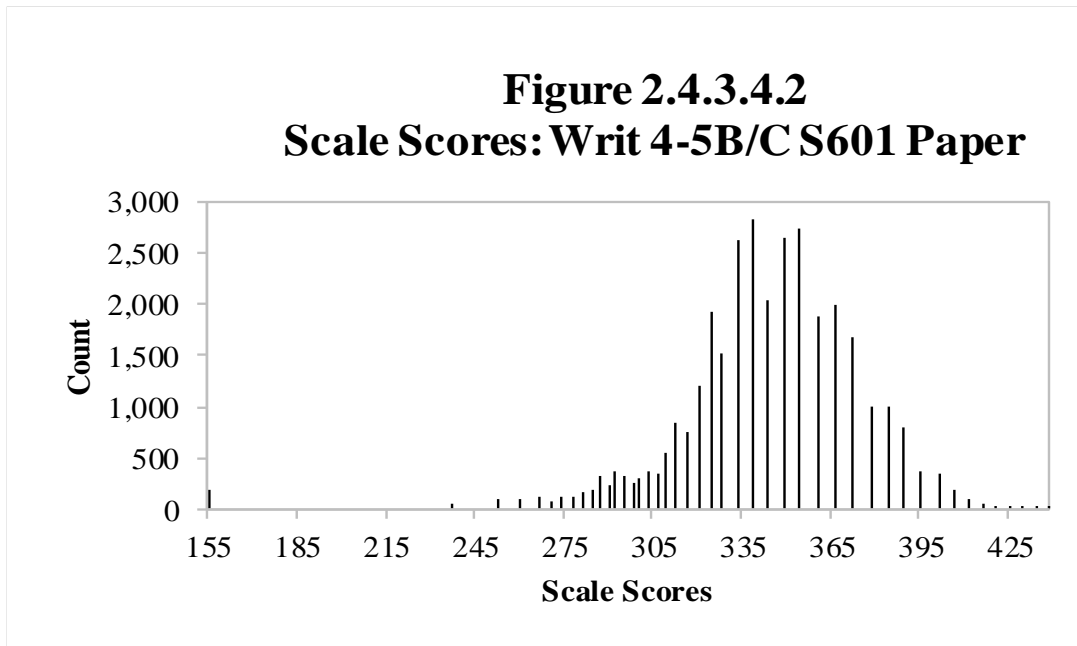
Table 2.4.3.4.3

Scale Score Descriptive Statistics: Writ 4-5 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	24,713	155	438	315.58	46.37
5	22,535	155	438	323.92	49.41
Total	47,248	155	438	319.56	48.02

Figure 2.4.3.4.3
Scale Scores: Writ 4-5 S601 Paper
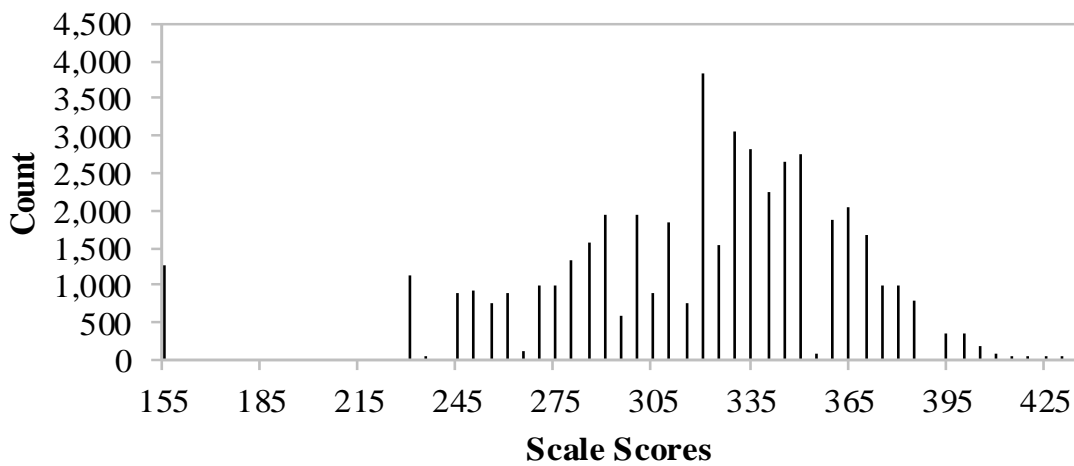
2.4.3.5 Grades 6-8

Table 2.4.3.5.1

Scale Score Descriptive Statistics: Writ 6-8 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	6,984	188	379	268.97	32.30
7	6,965	188	389	273.25	32.24
8	6,626	188	398	277.49	32.71
Total	20,575	188	398	273.16	32.59

Figure 2.4.3.5.1
Scale Scores: Writ 6-8A S601 Paper
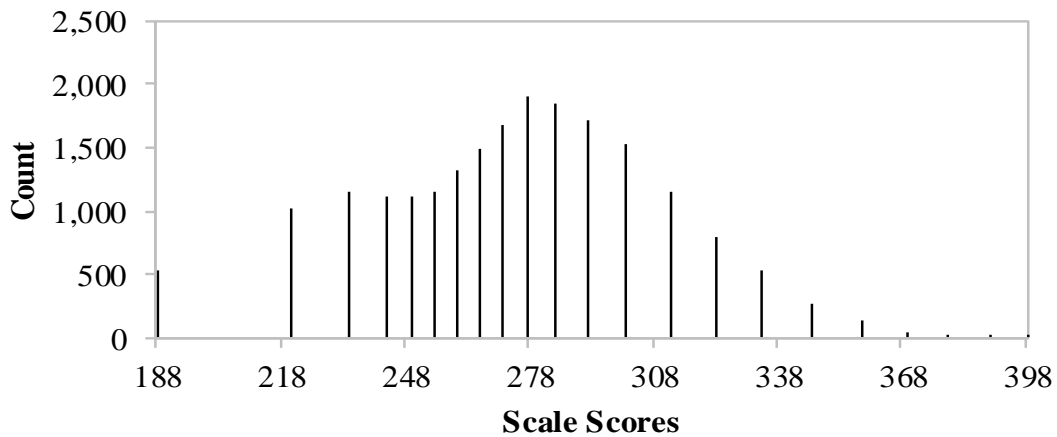
Table 2.4.3.5.2

Scale Score Descriptive Statistics: Writ 6-8 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	12,752	188	422	329.69	32.90
7	10,378	188	431	337.37	33.00
8	9,785	188	431	343.72	32.78
Total	32,915	188	431	336.28	33.40

Figure 2.4.3.5.2
Scale Scores: Writ 6-8B/C S601 Paper
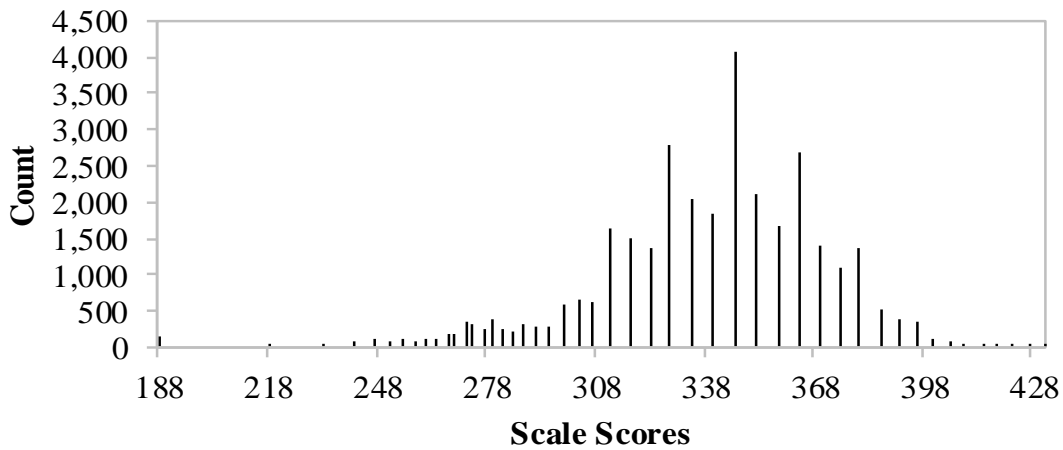
Table 2.4.3.5.3

Scale Score Descriptive Statistics: Writ 6-8 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	19,736	188	422	308.21	43.72
7	17,343	188	431	311.62	45.36
8	16,411	188	431	316.98	46.14
Total	53,490	188	431	312.00	45.15

Figure 2.4.3.5.3
Scale Scores: Writ 6-8 S601 Paper
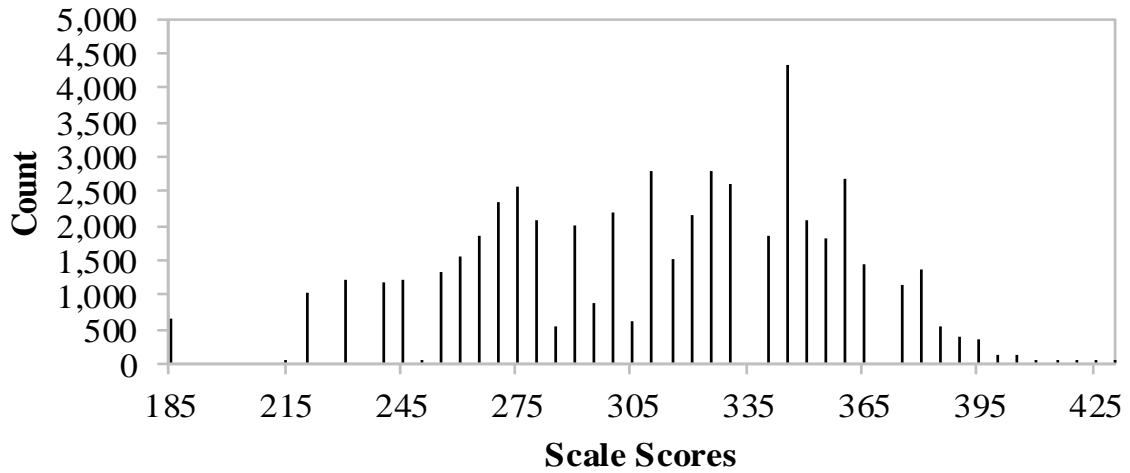
2.4.3.6 Grades 9-12

Table 2.4.3.6.1

Scale Score Descriptive Statistics: Writ 9-12 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	7,197	232	421	303.16	35.57
10	5,502	232	421	306.64	34.13
11	4,137	232	430	314.37	34.35
12	2,253	232	421	320.67	33.36
Total	19,089	232	430	308.66	35.16

**Figure 2.4.3.6.1
Scale Scores: Writ 9-12 A S601 Paper**
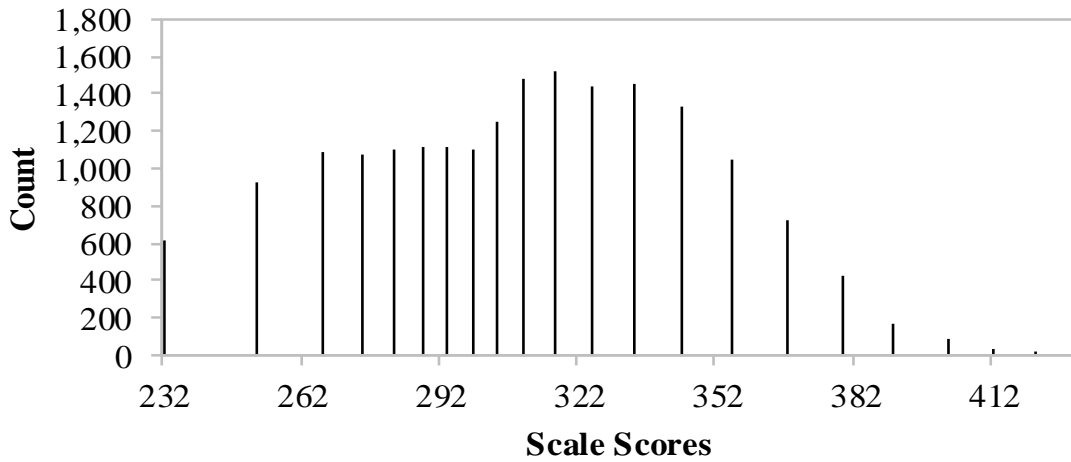
Table 2.4.3.6.2

Scale Score Descriptive Statistics: Writ 9-12 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	9,164	232	448	361.72	33.30
10	8,528	232	464	363.27	35.73
11	7,774	232	472	368.70	34.61
12	5,810	232	453	365.43	36.88
Total	31,276	232	472	364.57	35.08

Figure 2.4.3.6.2
Scale Scores: Writ 9-12B/C S601 Paper
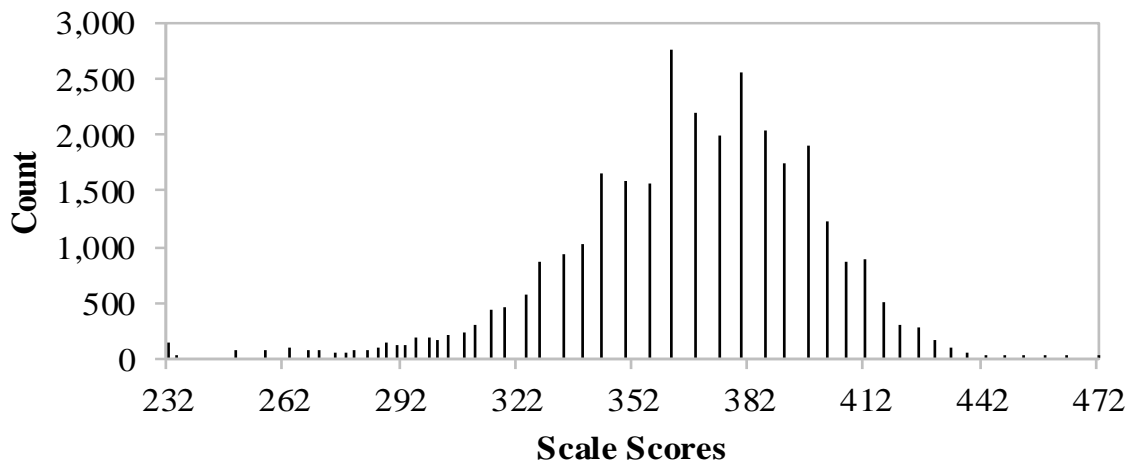
Table 2.4.3.6.3

Scale Score Descriptive Statistics: Writ 9-12 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	16,361	232	448	335.96	44.97
10	14,030	232	464	341.06	44.69
11	11,911	232	472	349.83	43.13
12	8,063	232	453	352.92	41.16
Total	50,365	232	472	343.38	44.37

Figure 2.4.3.6.3
Scale Scores: Writ 9-12 S601 Paper
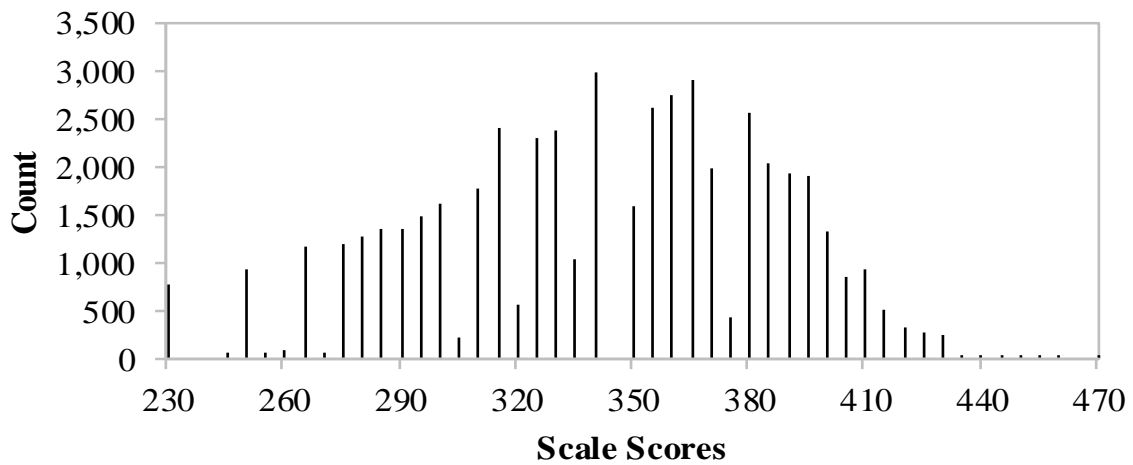
2.4.4 Speaking

2.4.4.0 Kindergarten

Table 2.4.4.0

Scale Score Descriptive Statistics: Spek K S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	256,828	100	392	249.99	104.44
Total	256,828	100	392	249.99	104.44

Figure 2.4.4.0
Scale Scores: Spek K S601 Paper
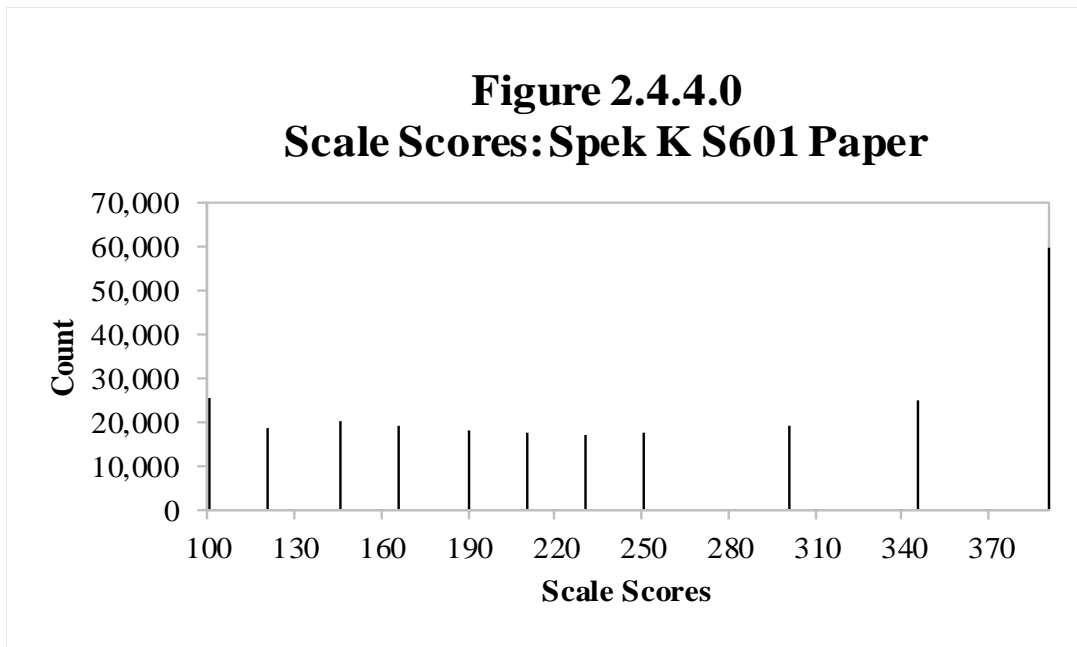
2.4.4.1 Grade 1

Table 2.4.4.1.1

Scale Score Descriptive Statistics: Spek 1 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	23,367	106	391	235.88	73.10
Total	23,367	106	391	235.88	73.10

Figure 2.4.4.1.1
Scale Scores: Spek 1A S601 Paper
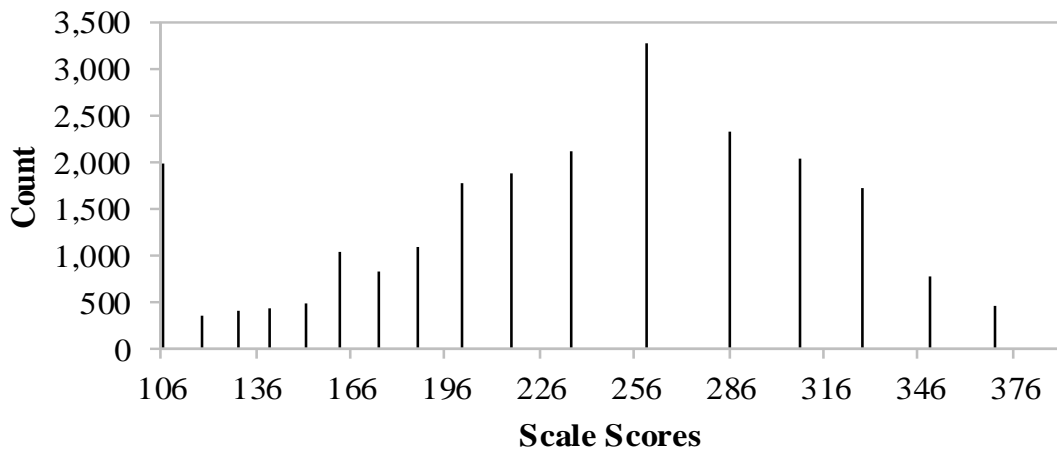
Table 2.4.4.1.2

Scale Score Descriptive Statistics: Spek 1 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	15,663	106	407	294.31	55.26
Total	15,663	106	407	294.31	55.26

**Figure 2.4.4.1.2
Scale Scores: Spek 1B/C S601 Paper**
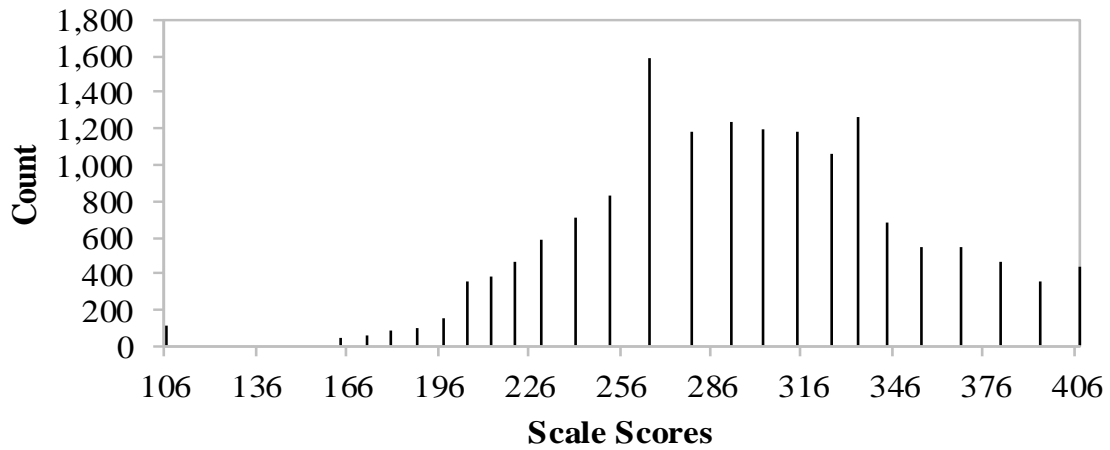
Table 2.4.4.1.3

Scale Score Descriptive Statistics: Spek 1 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	39,030	106	407	259.33	72.42
Total	39,030	106	407	259.33	72.42

Figure 2.4.4.1.3
Scale Scores: Spek 1 S601 Paper
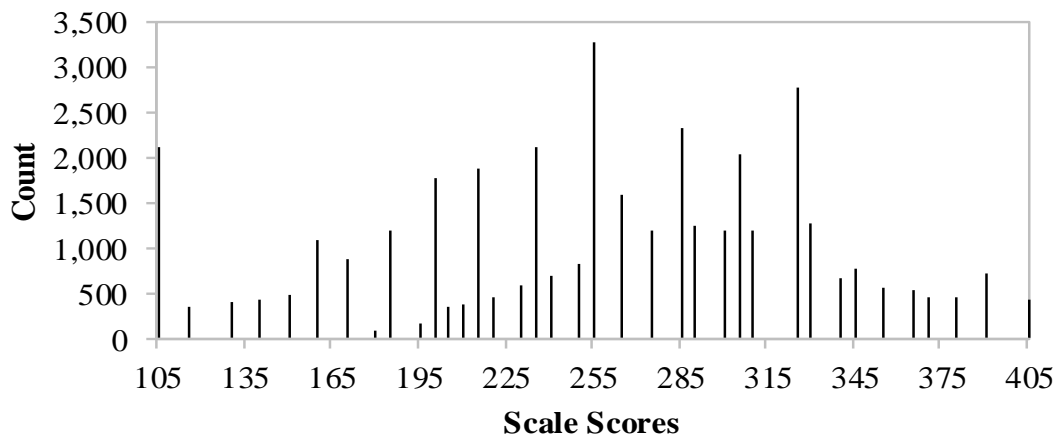
2.4.4.2 Grade 2

Table 2.4.4.2.1

Scale Score Descriptive Statistics: Spek 2 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	12,161	118	383	218.06	77.37
Total	12,161	118	383	218.06	77.37

Figure 2.4.4.2.1
Scale Scores: Spek 2A S601 Paper
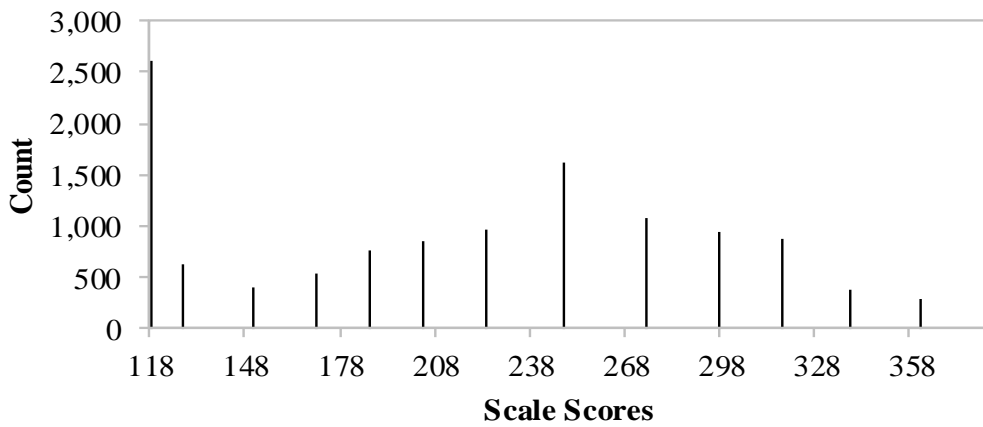
Table 2.4.4.2.2

Scale Score Descriptive Statistics: Spek 2 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	22,729	118	425	303.78	53.19
Total	22,729	118	425	303.78	53.19

**Figure 2.4.4.2.2
Scale Scores: Spek 2B/C S601 Paper**
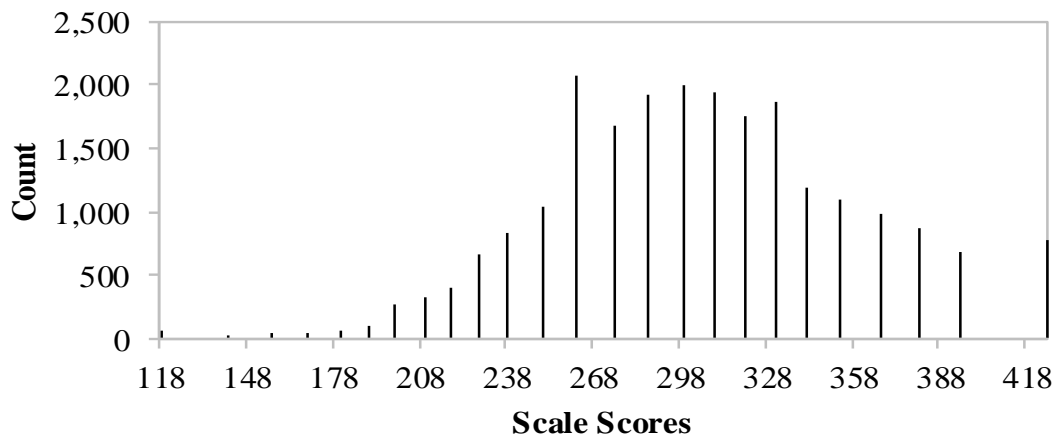
Table 2.4.4.2.3

Scale Score Descriptive Statistics: Spek 2 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	34,890	118	425	273.90	74.82
Total	34,890	118	425	273.90	74.82

Figure 2.4.4.2.3
Scale Scores: Spek 2 S601 Paper
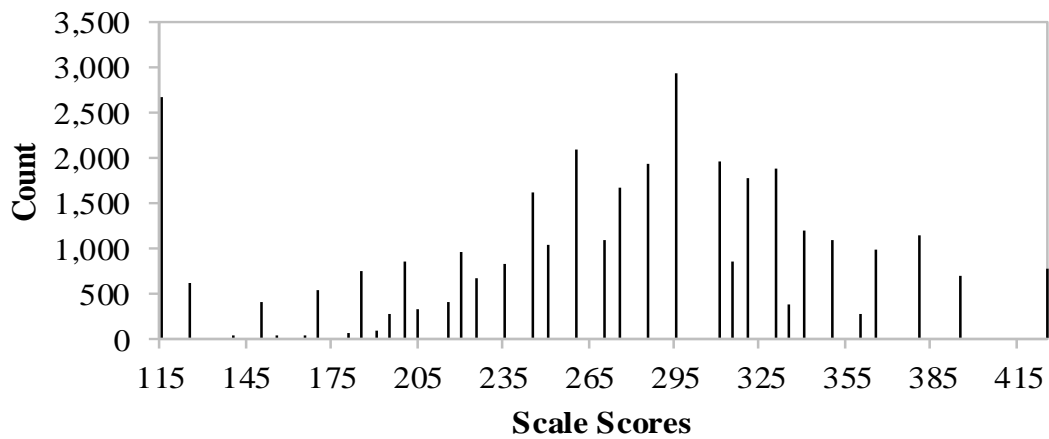
2.4.4.3 Grade 3

Table 2.4.4.3.1

Scale Score Descriptive Statistics: Spek 3 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	9,772	118	383	214.46	77.72
Total	9,772	118	383	214.46	77.72

Figure 2.4.4.3.1
Scale Scores: Spek 3A S601 Paper
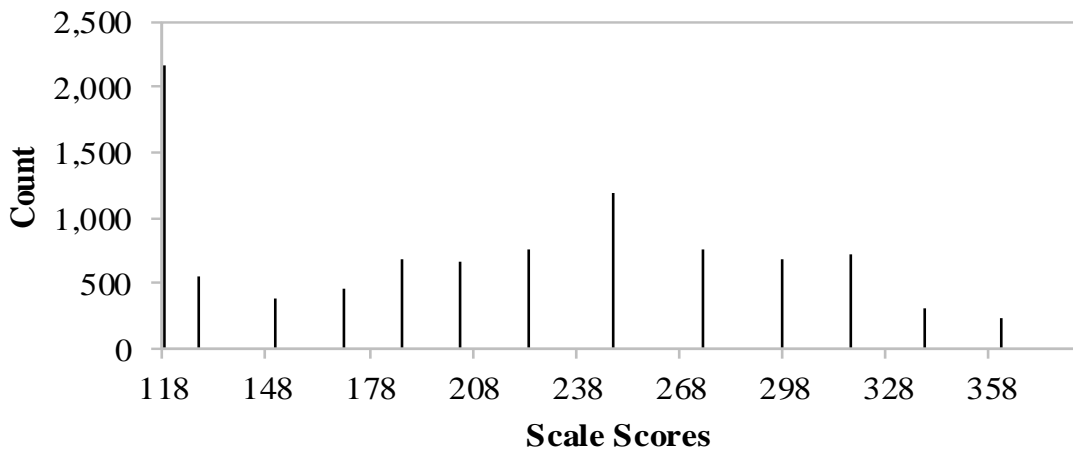
Table 2.4.4.3.2

Scale Score Descriptive Statistics: Spek 3 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	24,943	118	425	314.61	52.52
Total	24,943	118	425	314.61	52.52

Figure 2.4.4.3.2
Scale Scores: Spek 3B/C S601 Paper
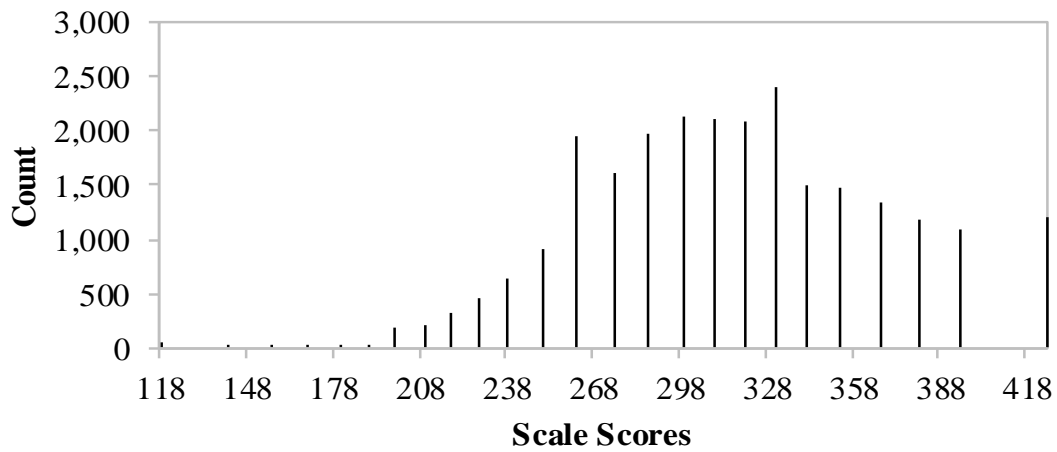
Table 2.4.4.3.3

Scale Score Descriptive Statistics: Spek 3 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	34,715	118	425	286.42	75.57
Total	34,715	118	425	286.42	75.57

Figure 2.4.4.3.3
Scale Scores: Spek 3 S601 Paper
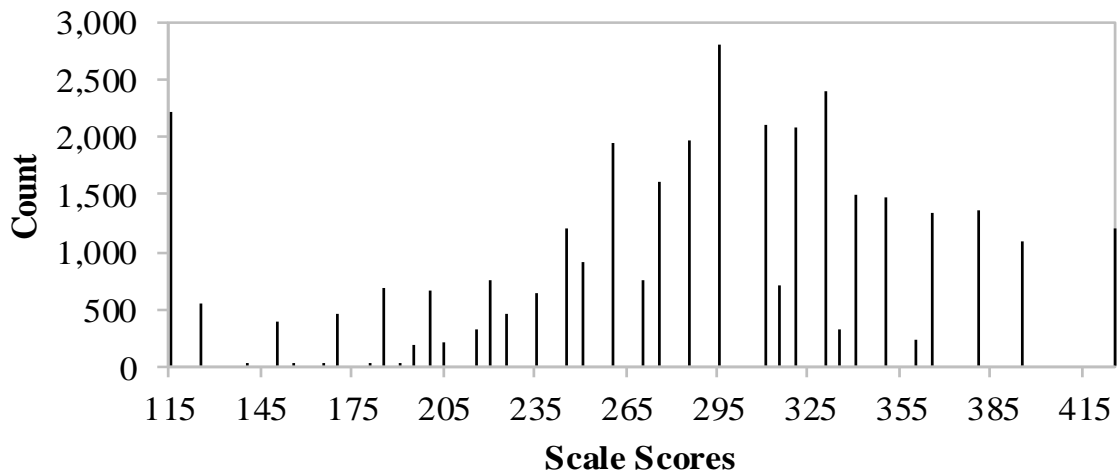
2.4.4.4 Grades 4-5

Table 2.4.4.4.1

Scale Score Descriptive Statistics: Spek 4-5 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	7,128	130	423	224.98	75.85
5	7,169	130	423	228.88	75.22
Total	14,297	130	423	226.93	75.56

Figure 2.4.4.4.1
Scale Scores: Spek 4-5 A S601 Paper
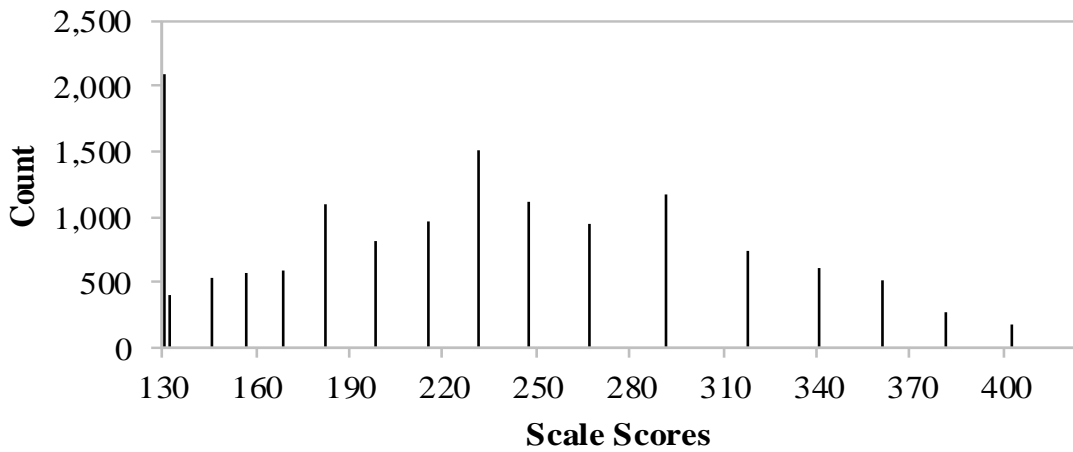
Table 2.4.4.4.2

Scale Score Descriptive Statistics: Spek 4–5 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	17,429	130	450	353.64	52.71
5	15,258	130	450	358.85	53.91
Total	32,687	130	450	356.07	53.33

Figure 2.4.4.4.2
Scale Scores: Spek 4-5B/C S601 Paper
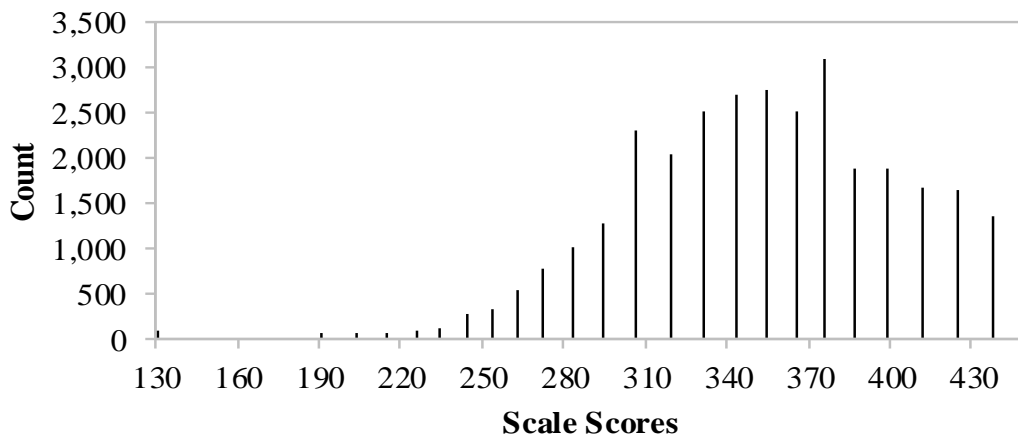
Table 2.4.4.4.3

Scale Score Descriptive Statistics: Spek 4-5 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	24,557	130	450	316.29	83.98
5	22,427	130	450	317.31	86.37
Total	46,984	130	450	316.78	85.13

Figure 2.4.4.4.3
Scale Scores: Spek 4-5 S601 Paper
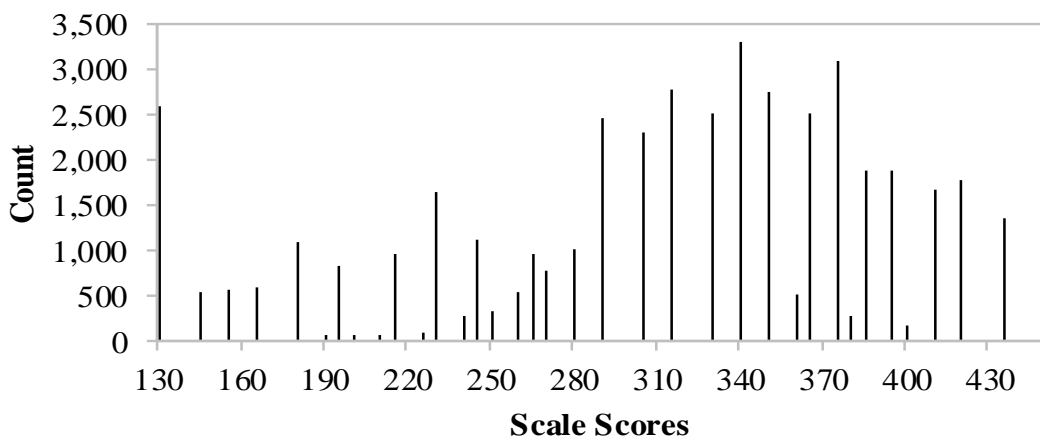
2.4.4.5 Grades 6-8

Table 2.4.4.5.1

Scale Score Descriptive Statistics: Spek 6-8 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	6,929	148	459	259.70	76.85
7	6,887	148	459	262.07	76.57
8	6,568	148	459	267.43	76.14
Total	20,384	148	459	262.99	76.59

Figure 2.4.4.5.1
Scale Scores: Spek 6-8A S601 Paper
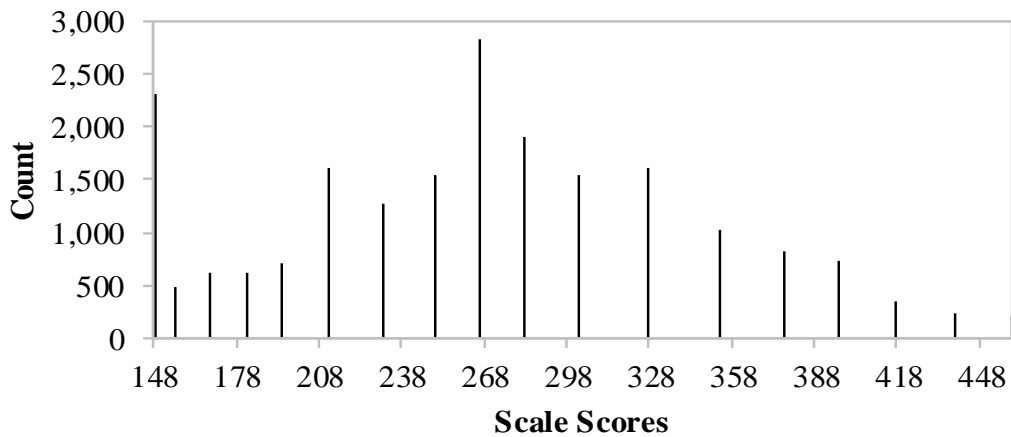
Table 2.4.4.5.2

Scale Score Descriptive Statistics: Spek 6-8 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	12,651	148	471	366.52	58.30
7	10,315	148	471	371.45	60.16
8	9,711	148	471	378.55	59.75
Total	32,677	148	471	371.65	59.53

Figure 2.4.4.5.2
Scale Scores: Spek 6-8B/C S601 Paper
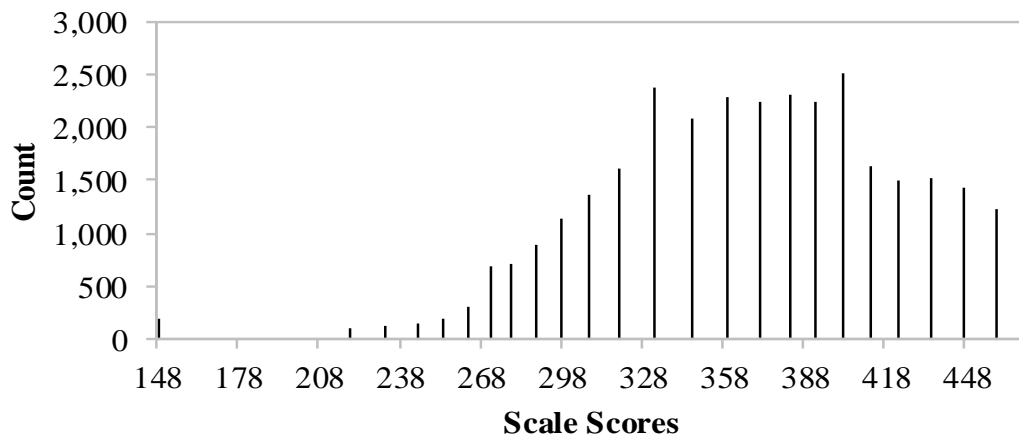
Table 2.4.4.5.3

Scale Score Descriptive Statistics: Spek 6-8 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	19,580	148	471	328.72	83.04
7	17,202	148	471	327.66	85.96
8	16,279	148	471	333.72	86.26
Total	53,061	148	471	329.91	85.02

Figure 2.4.4.5.3
Scale Scores: Spek 6-8 S601 Paper
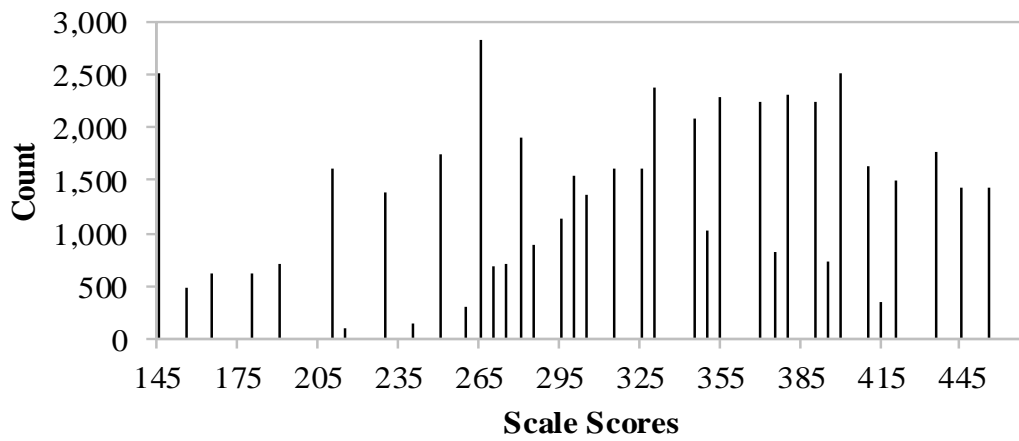
2.4.4.6 Grades 9-12

Table 2.4.4.6.1

Scale Score Descriptive Statistics: Spek 9-12 A S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	7,111	172	445	258.54	70.25
10	5,455	172	445	265.80	70.43
11	4,092	172	445	279.48	70.98
12	2,238	172	445	294.71	72.02
Total	18,896	172	445	269.45	71.69

Figure 2.4.4.6.1
Scale Scores: Spek 9-12 A S601 Paper
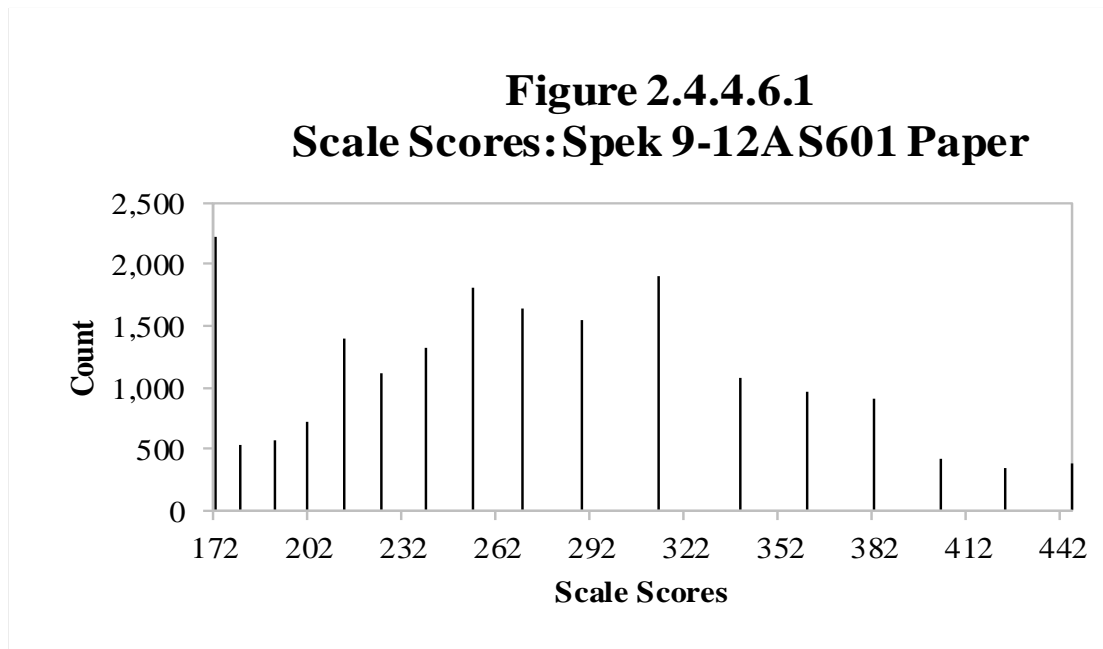
Table 2.4.4.6.2

Scale Score Descriptive Statistics: Spek 9-12 B/C S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	9,100	172	476	378.16	61.68
10	8,451	172	476	371.72	63.65
11	7,723	172	476	378.87	62.49
12	5,763	172	476	379.69	62.68
Total	31,037	172	476	376.87	62.69

Figure 2.4.4.6.2
Scale Scores: Spek 9-12B/C S601 Paper
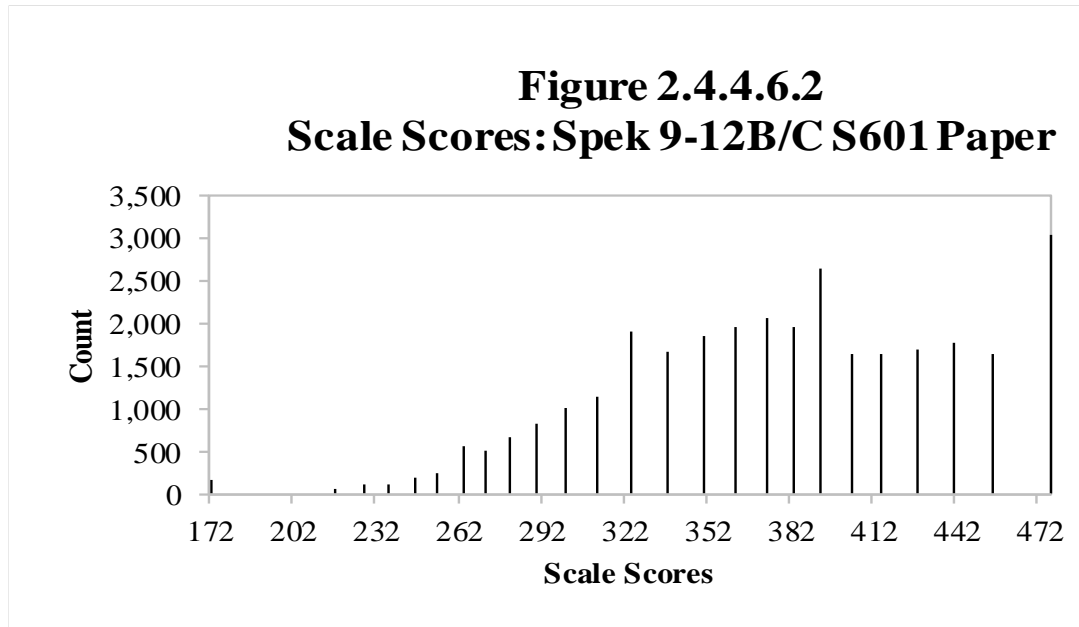
Table 2.4.4.6.3

Scale Score Descriptive Statistics: Spek 9-12 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	16,211	172	476	325.69	88.46
10	13,906	172	476	330.17	84.16
11	11,815	172	476	344.45	80.83
12	8,001	172	476	355.92	75.73
Total	49,933	172	476	336.22	84.27

Figure 2.4.4.6.3
Scale Scores: Spek 9-12 S601 Paper
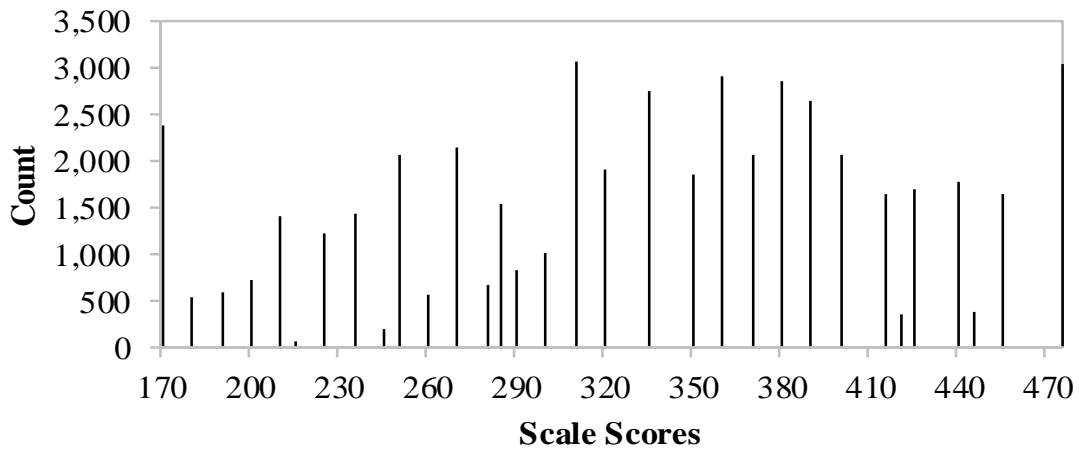
2.5 Proficiency Level Distribution

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 S601 Paper

Level	Grade K		Total	
	Count	**Percent**	**Count**	**Percent**
1	92,574	36.04%	92,574	36.04%
2	24,082	9.38%	24,082	9.38%
3	20,163	7.85%	20,163	7.85%
4	13,055	5.08%	13,055	5.08%
5	33,140	12.90%	33,140	12.90%
6	73,833	28.75%	73,833	28.75%
Total	256,847	100.00%	256,847	100.00%

**Figure 2.5.1.0
Proficiency Level: List K S601 Paper**
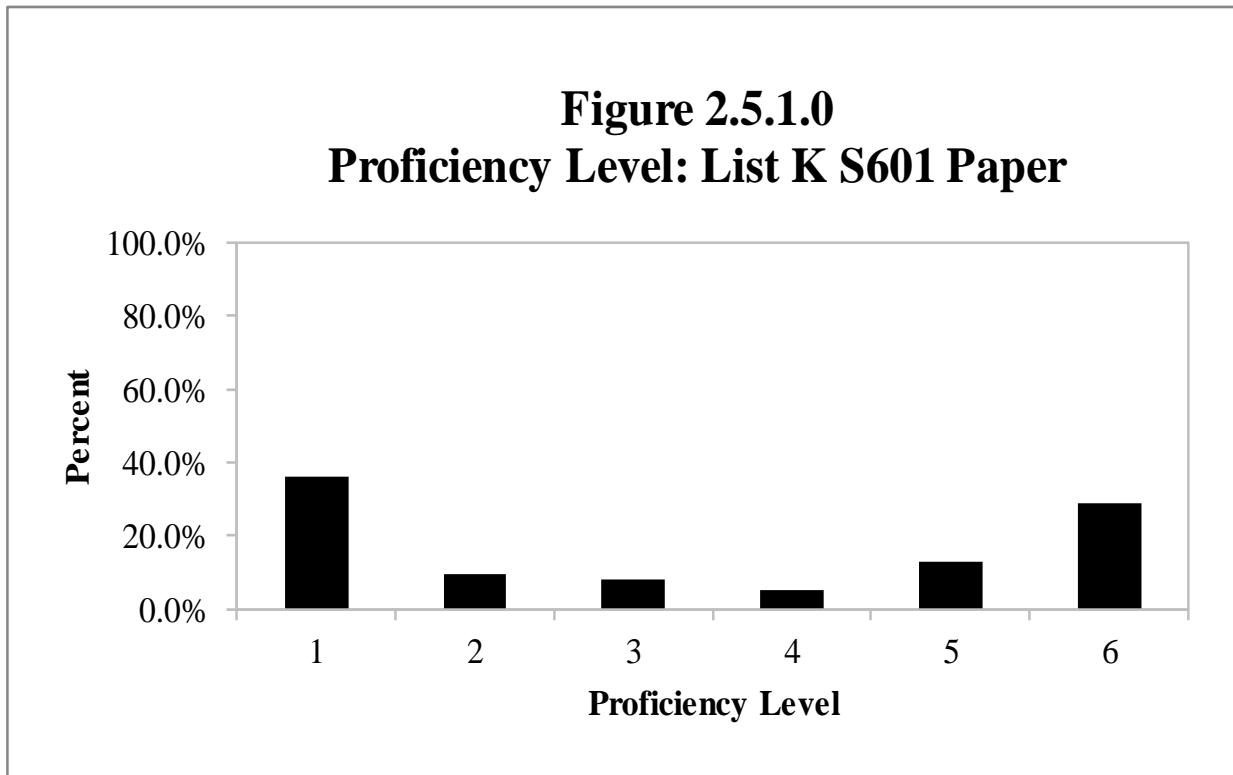
2.5.1.1 Grade 1

Table 2.5.1.1.1

Proficiency Level Distribution: List 1 A S601 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	2,403	12.30%	2,403	12.30%
2	3,457	17.70%	3,457	17.70%
3	3,441	17.62%	3,441	17.62%
4	2,116	10.83%	2,116	10.83%
5	4,893	25.05%	4,893	25.05%
6	3,220	16.49%	3,220	16.49%
Total	19,530	100.00%	19,530	100.00%

Figure 2.5.1.1.1
Proficiency Level: List 1A S601 Paper
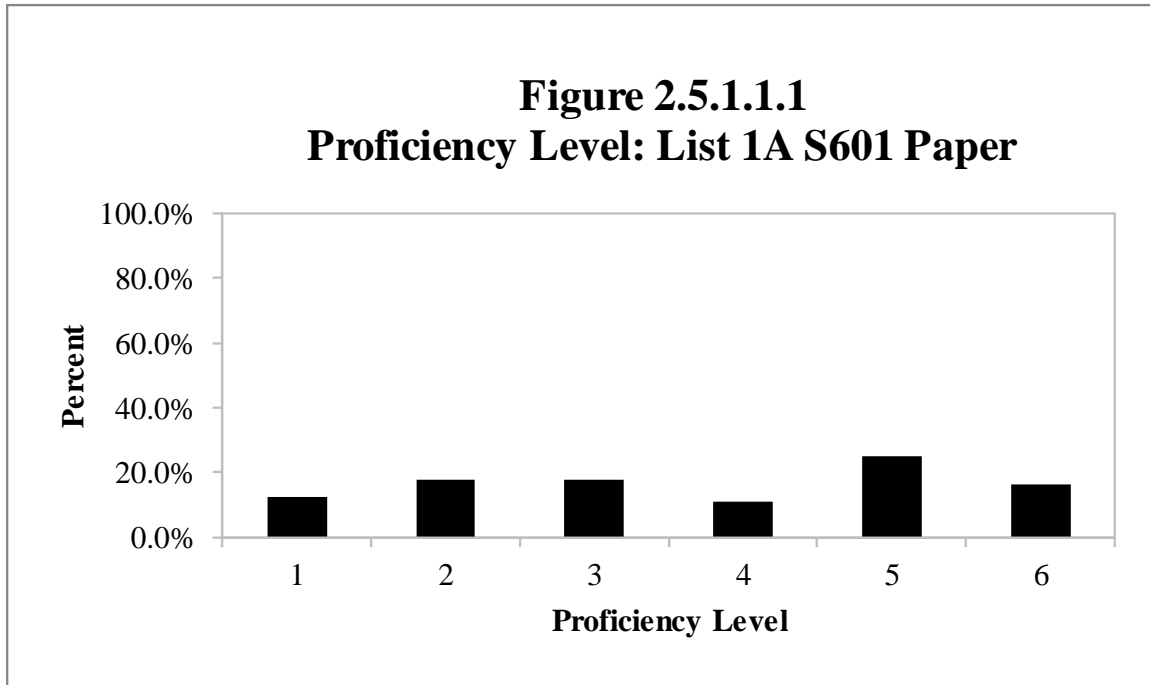
Table 2.5.1.1.2

Proficiency Level Distribution: List 1 B/C S601 Paper

Level	Grade 1		Total	
	Count	**Percent**	**Count**	**Percent**
1	234	1.68%	234	1.68%
2	489	3.51%	489	3.51%
3	1,811	13.01%	1,811	13.01%
4	2,204	15.83%	2,204	15.83%
5	2,976	21.38%	2,976	21.38%
6	6,208	44.59%	6,208	44.59%
Total	13,922	100.00%	13,922	100.00%

**Figure 2.5.1.1.2
Proficiency Level: List 1B/C S601 Paper**
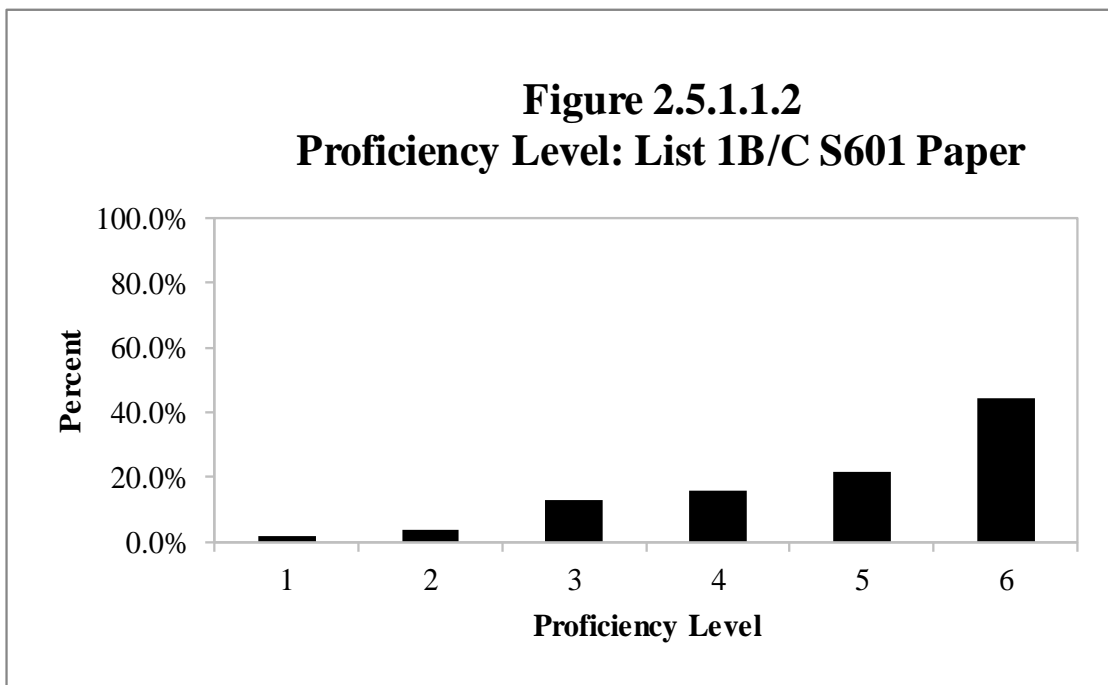
Table 2.5.1.1.3

Proficiency Level Distribution: List 1 S601 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	2,637	7.88%	2,637	7.88%
2	3,946	11.80%	3,946	11.80%
3	5,252	15.70%	5,252	15.70%
4	4,320	12.91%	4,320	12.91%
5	7,869	23.52%	7,869	23.52%
6	9,428	28.18%	9,428	28.18%
Total	33,452	100.00%	33,452	100.00%

Figure 2.5.1.1.3
Proficiency Level: List 1 S601 Paper
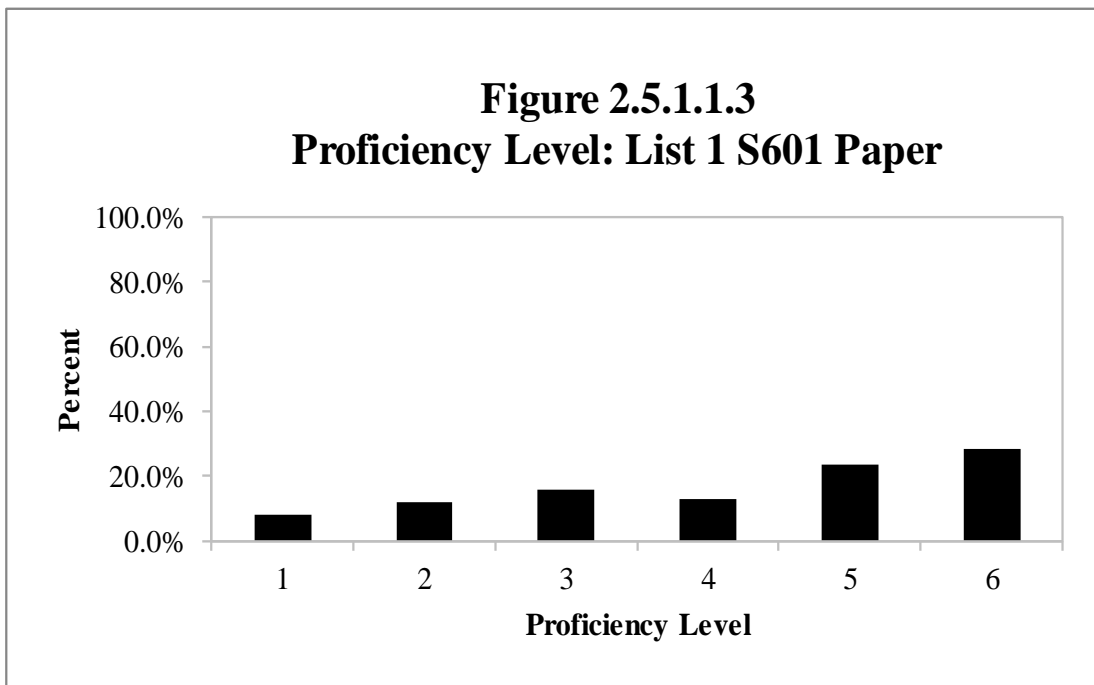
2.5.1.2 Grade 2

Table 2.5.1.2.1

Proficiency Level Distribution: List 2 A S601 Paper

Level	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	2,249	20.72%	2,249	20.72%
2	2,648	24.40%	2,648	24.40%
3	2,040	18.80%	2,040	18.80%
4	1,335	12.30%	1,335	12.30%
5	2,580	23.77%	2,580	23.77%
6	0	0.00%	0	0.00%
Total	10,852	100.00%	10,852	100.00%

**Figure 2.5.1.2.1
Proficiency Level: List 2A S601 Paper**
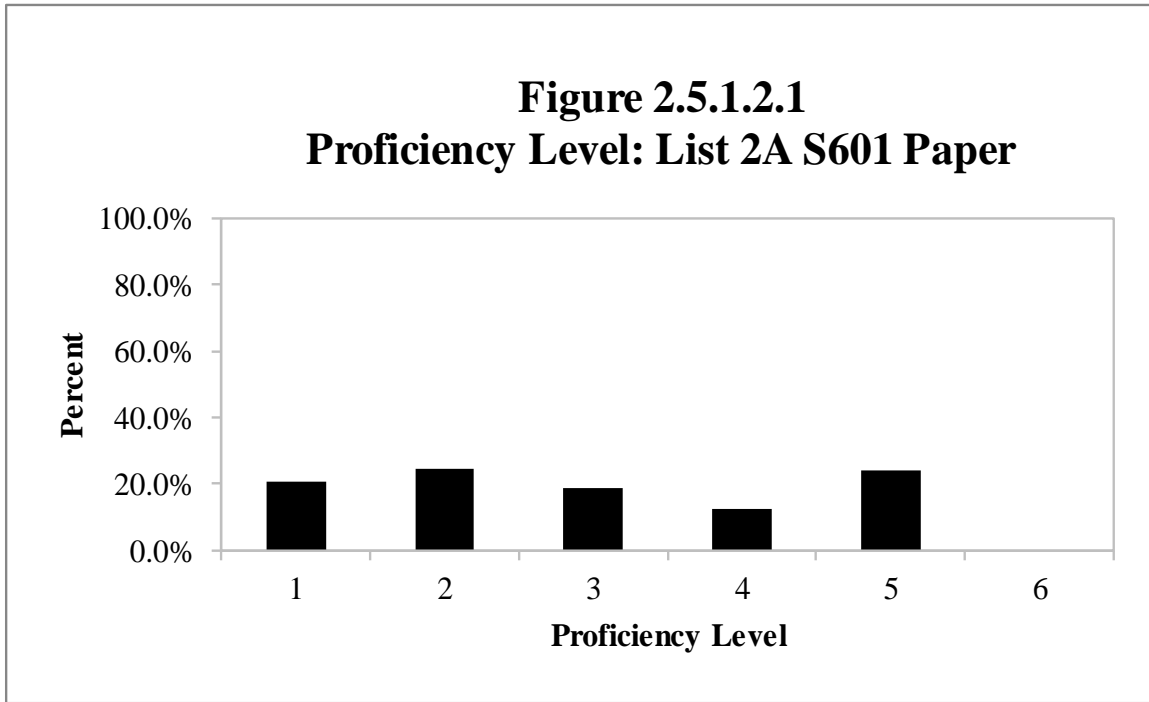
Table 2.5.1.2.2

Proficiency Level Distribution: List 2 B/C S601 Paper

Level	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	164	0.76%	164	0.76%
2	1,180	5.50%	1,180	5.50%
3	3,146	14.67%	3,146	14.67%
4	1,962	9.15%	1,962	9.15%
5	5,635	26.28%	5,635	26.28%
6	9,355	43.63%	9,355	43.63%
Total	21,442	100.00%	21,442	100.00%

Figure 2.5.1.2.2
Proficiency Level: List 2B/C S601 Paper
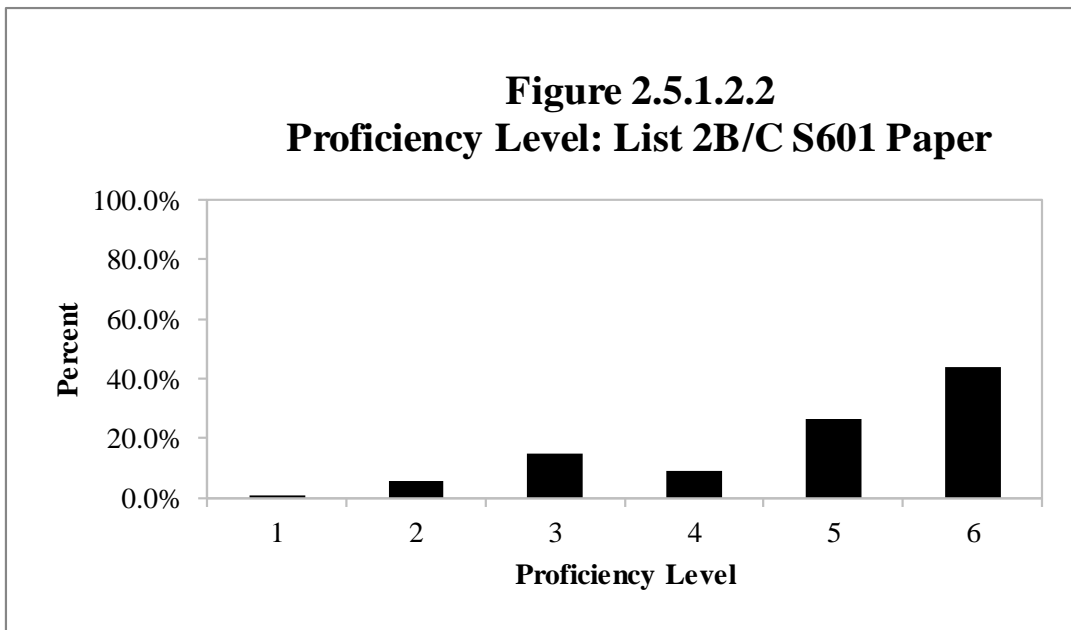
Table 2.5.1.2.3

Proficiency Level Distribution: List 2 S601 Paper

Level	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	2,413	7.47%	2,413	7.47%
2	3,828	11.85%	3,828	11.85%
3	5,186	16.06%	5,186	16.06%
4	3,297	10.21%	3,297	10.21%
5	8,215	25.44%	8,215	25.44%
6	9,355	28.97%	9,355	28.97%
Total	32,294	100.00%	32,294	100.00%

Figure 2.5.1.2.3
Proficiency Level: List 2 S601 Paper
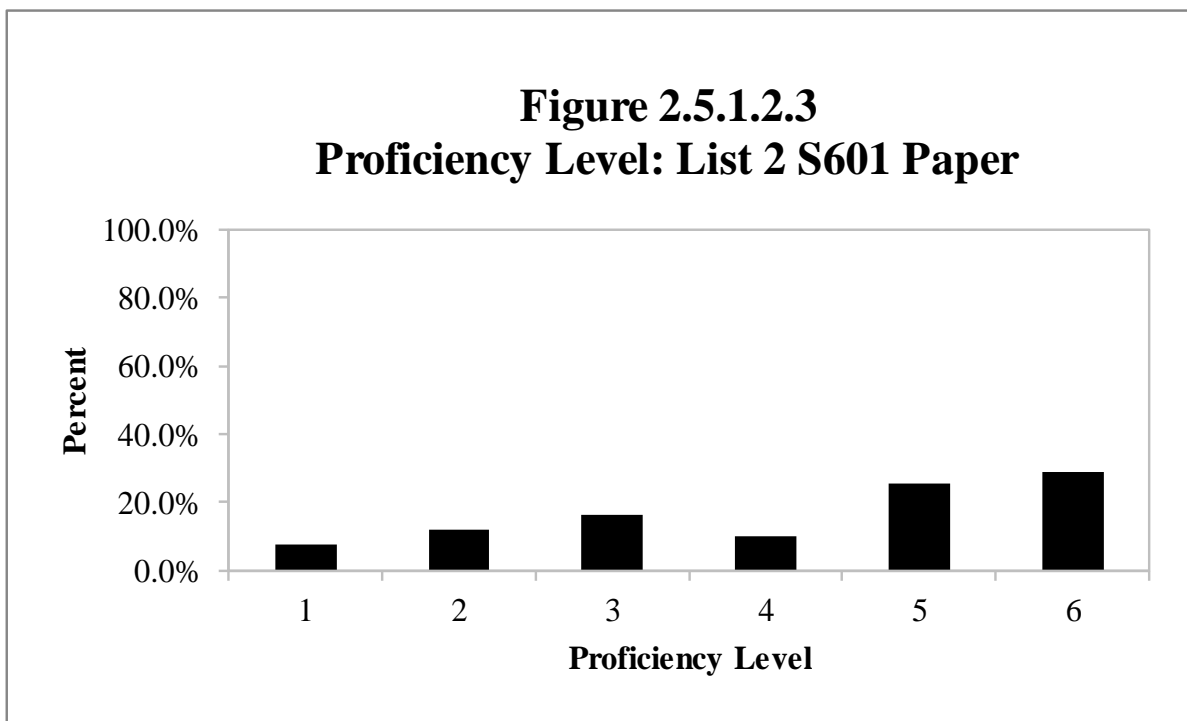
2.5.1.3 Grade 3

Table 2.5.1.3.1

Proficiency Level Distribution: List 3 A S601 Paper

Level	Grade 3		Total	
	Count	**Percent**	**Count**	**Percent**
1	361	4.01%	361	4.01%
2	2,671	29.67%	2,671	29.67%
3	2,714	30.15%	2,714	30.15%
4	1,410	15.66%	1,410	15.66%
5	1,110	12.33%	1,110	12.33%
6	737	8.19%	737	8.19%
Total	9,003	100.00%	9,003	100.00%

**Figure 2.5.1.3.1
Proficiency Level: List 3A S601 Paper**
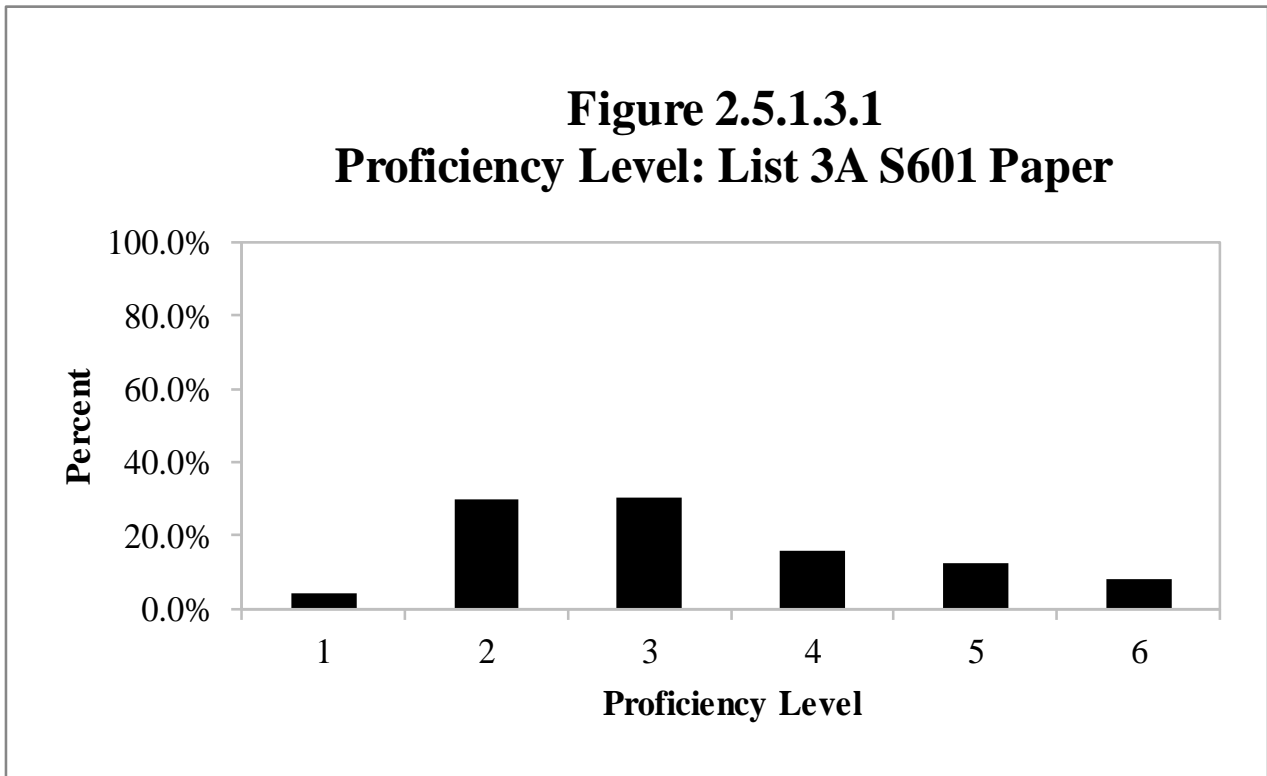
Table 2.5.1.3.2

Proficiency Level Distribution: List 3 B/C S601 Paper

Level	Grade 3		Total	
	Count	**Percent**	**Count**	**Percent**
1	10	0.04%	10	0.04%
2	415	1.76%	415	1.76%
3	3,157	13.36%	3,157	13.36%
4	2,208	9.35%	2,208	9.35%
5	8,820	37.34%	8,820	37.34%
6	9,012	38.15%	9,012	38.15%
Total	23,622	100.00%	23,622	100.00%

Figure 2.5.1.3.2
Proficiency Level: List 3B/C S601 Paper
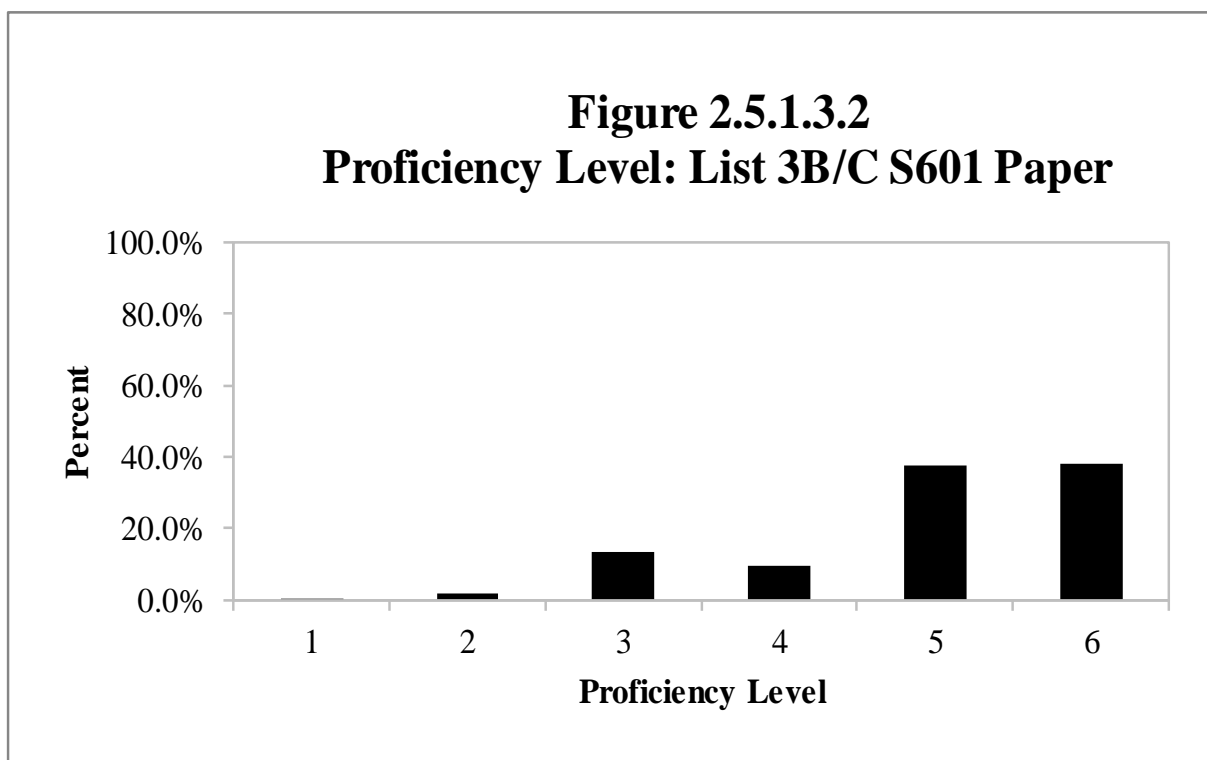
Table 2.5.1.3.3

Proficiency Level Distribution: List 3 S601 Paper

Level	Grade 3		Total	
	Count	**Percent**	**Count**	**Percent**
1	371	1.14%	371	1.14%
2	3,086	9.46%	3,086	9.46%
3	5,871	18.00%	5,871	18.00%
4	3,618	11.09%	3,618	11.09%
5	9,930	30.44%	9,930	30.44%
6	9,749	29.88%	9,749	29.88%
Total	32,625	100.00%	32,625	100.00%

Figure 2.5.1.3.3
Proficiency Level: List 3 S601 Paper
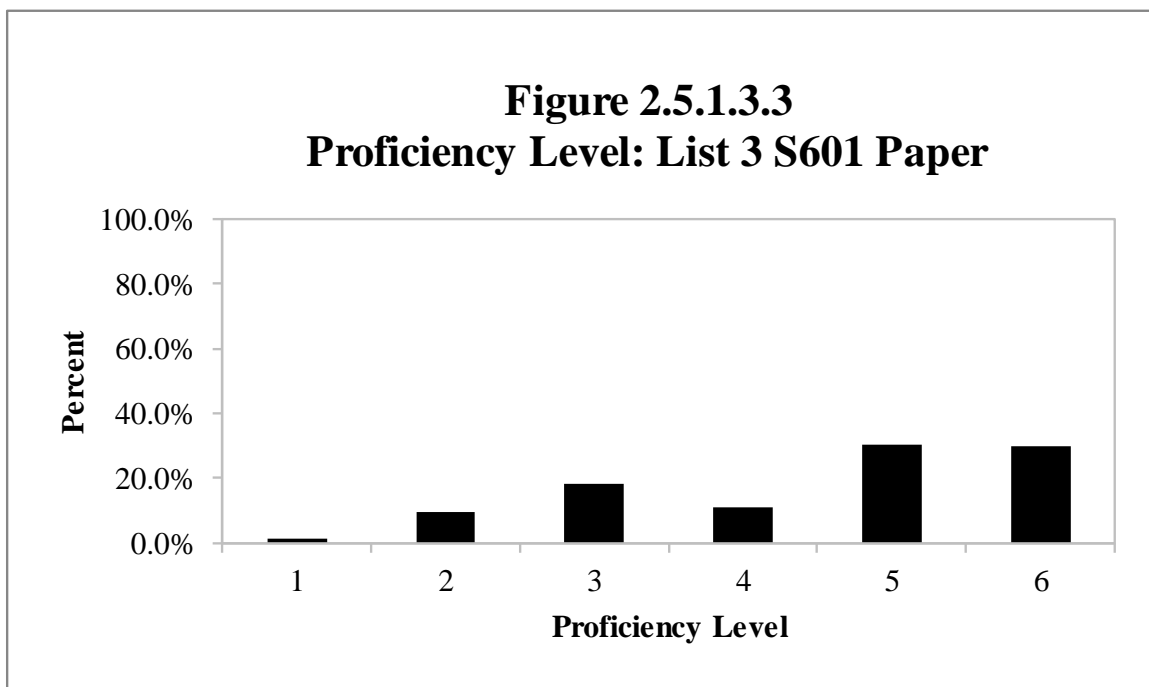
2.5.1.4 Grades 4-5

Table 2.5.1.4.1

Proficiency Level Distribution: List 4-5 A S601 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	526	7.89%	780	11.61%	1,306	9.76%
2	2,448	36.72%	2,354	35.03%	4,802	35.87%
3	1,748	26.22%	1,775	26.41%	3,523	26.32%
4	851	12.77%	910	13.54%	1,761	13.16%
5	743	11.15%	407	6.06%	1,150	8.59%
6	350	5.25%	494	7.35%	844	6.31%
Total	6,666	100.00%	6,720	100.00%	13,386	100.00%

Figure 2.5.1.4.1
Proficiency Level: List 4-5A S601 Paper
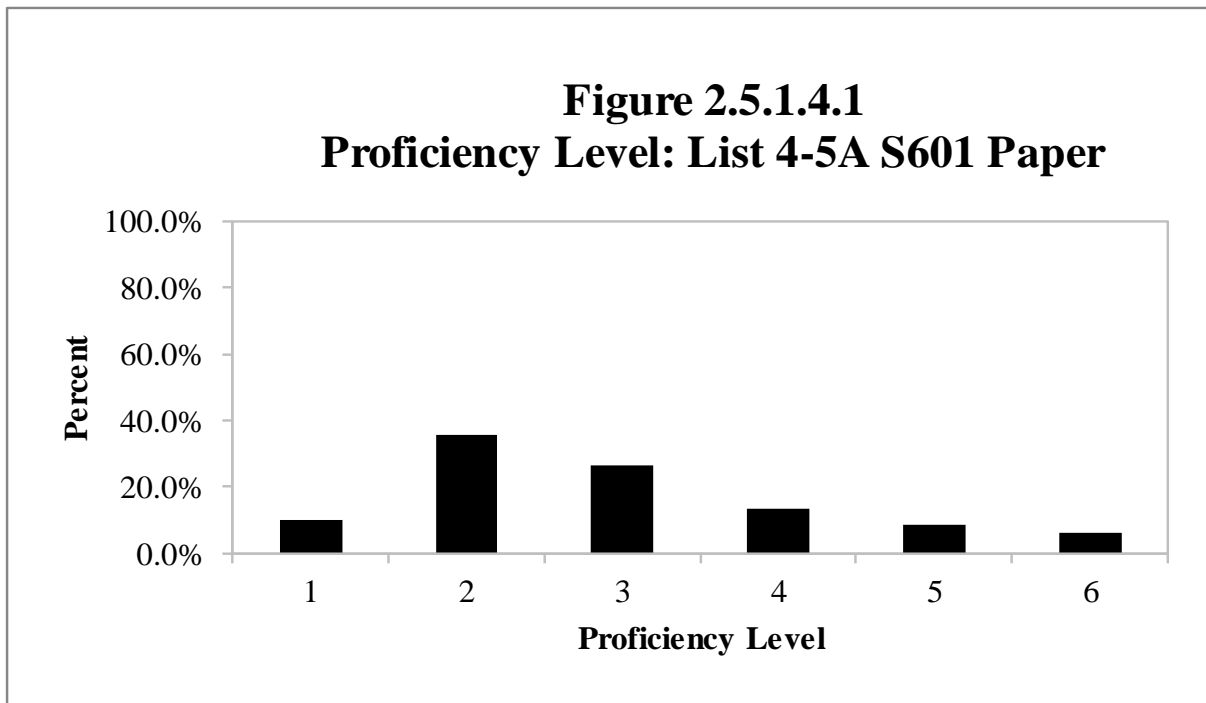
Table 2.5.1.4.2

Proficiency Level Distribution: List 4-5 B/C S601 Paper

Level	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	17	0.10%	10	0.07%	27	0.08%
2	230	1.36%	385	2.58%	615	1.93%
3	1,490	8.81%	1,491	10.00%	2,981	9.37%
4	2,593	15.34%	2,368	15.88%	4,961	15.59%
5	6,607	39.08%	5,840	39.16%	12,447	39.12%
6	5,968	35.30%	4,819	32.31%	10,787	33.90%
Total	16,905	100.00%	14,913	100.00%	31,818	100.00%

Figure 2.5.1.4.2
Proficiency Level: List 4-5B/C S601 Paper
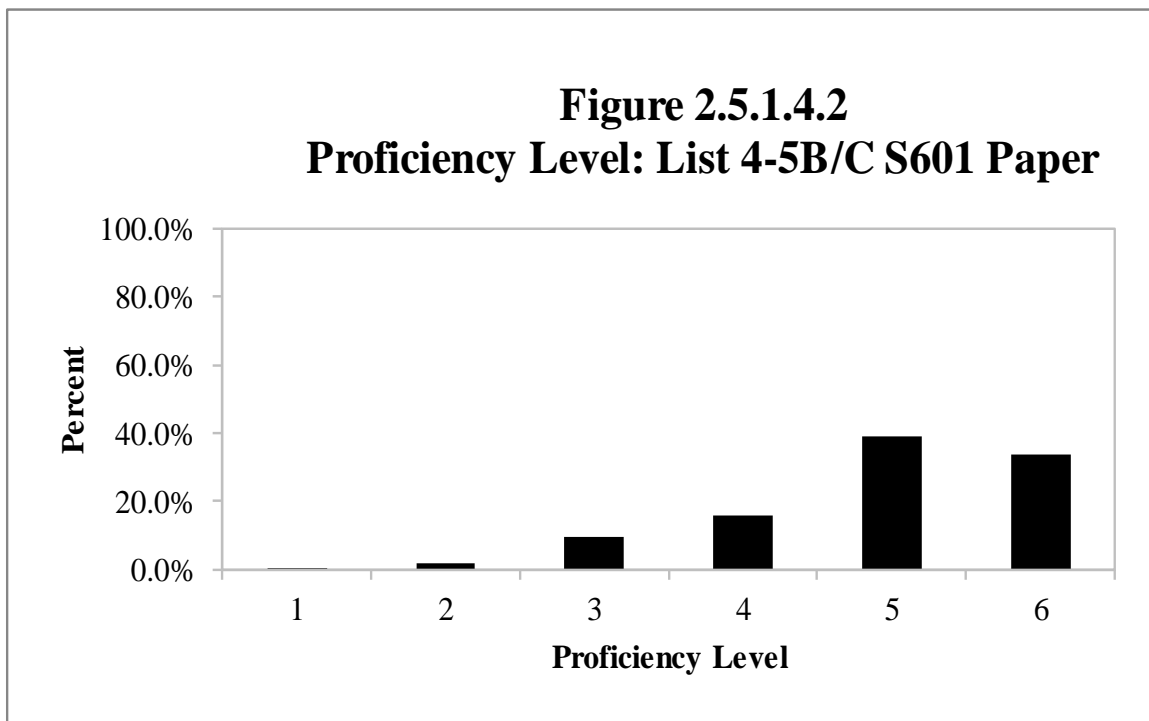
Table 2.5.1.4.3

Proficiency Level Distribution: List 4-5 S601 Paper

Level	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	543	2.30%	790	3.65%	1,333	2.95%
2	2,678	11.36%	2,739	12.66%	5,417	11.98%
3	3,238	13.74%	3,266	15.10%	6,504	14.39%
4	3,444	14.61%	3,278	15.15%	6,722	14.87%
5	7,350	31.18%	6,247	28.88%	13,597	30.08%
6	6,318	26.80%	5,313	24.56%	11,631	25.73%
Total	23,571	100.00%	21,633	100.00%	45,204	100.00%

Figure 2.5.1.4.3
Proficiency Level: List 4-5 S601 Paper
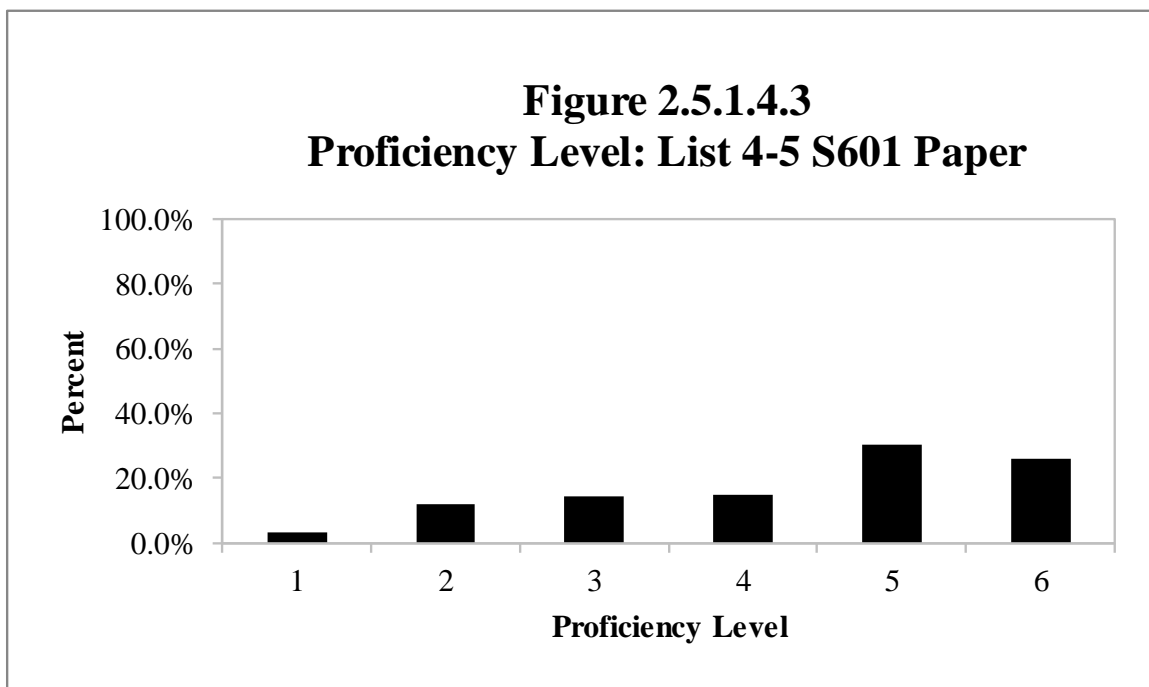
2.5.1.5 Grades 6-8

Table 2.5.1.5.1

Proficiency Level Distribution: List 6-8 A S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,732	26.67%	2,422	37.34%	2,050	33.37%	6,204	32.44%
2	2,870	44.19%	2,106	32.47%	2,564	41.73%	7,540	39.43%
3	917	14.12%	1,210	18.66%	711	11.57%	2,838	14.84%
4	563	8.67%	298	4.59%	484	7.88%	1,345	7.03%
5	353	5.44%	366	5.64%	335	5.45%	1,054	5.51%
6	59	0.91%	84	1.30%	0	0.00%	143	0.75%
Total	6,494	100.00%	6,486	100.00%	6,144	100.00%	19,124	100.00%

Figure 2.5.1.5.1
Proficiency Level: List 6-8A S601 Paper
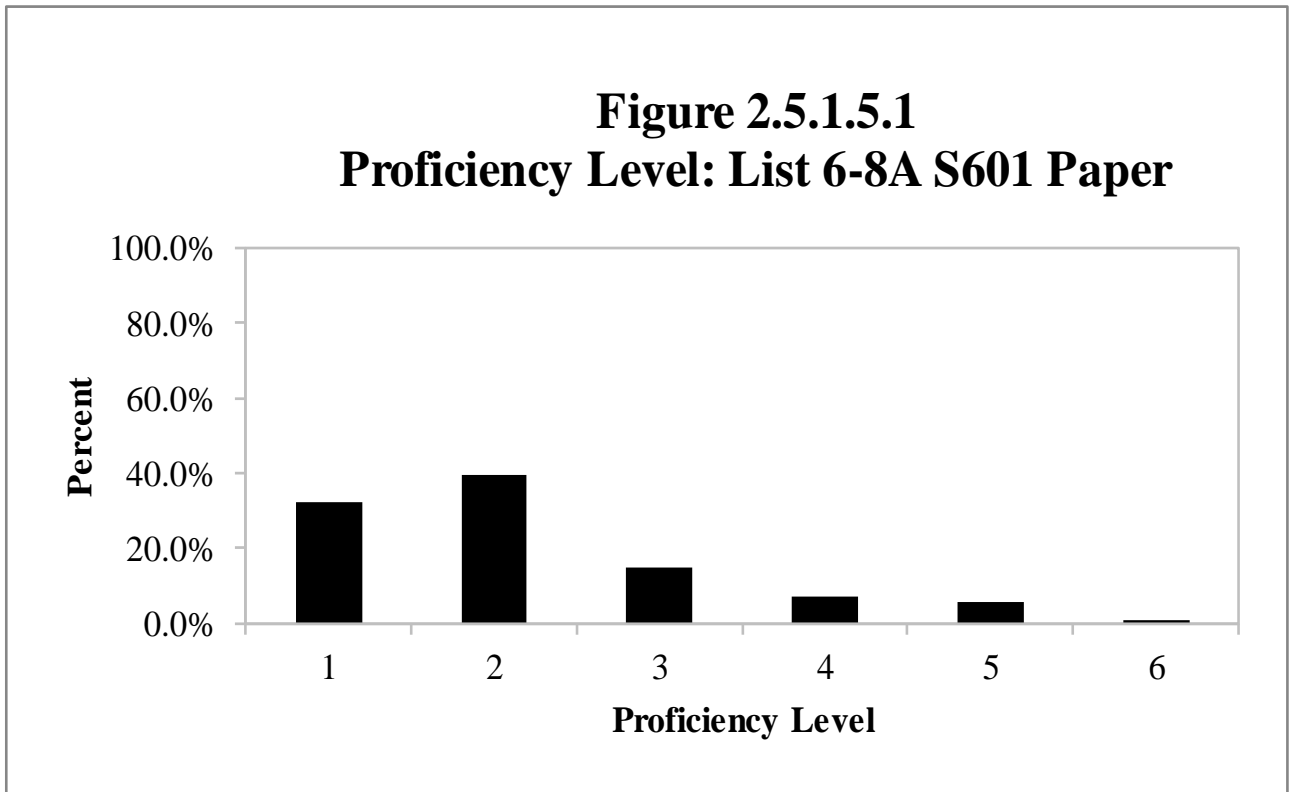
Table 2.5.1.5.2

Proficiency Level Distribution: List 6-8 B/C S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	16	0.13%	19	0.19%	26	0.28%	61	0.19%
2	436	3.58%	462	4.66%	352	3.77%	1,250	3.98%
3	1,732	14.22%	1,462	14.74%	1,790	19.19%	4,984	15.86%
4	3,981	32.68%	3,305	33.31%	1,999	21.43%	9,285	29.54%
5	3,161	25.95%	2,656	26.77%	2,696	28.90%	8,513	27.09%
6	2,854	23.43%	2,017	20.33%	2,466	26.43%	7,337	23.34%
Total	12,180	100.00%	9,921	100.00%	9,329	100.00%	31,430	100.00%

**Figure 2.5.1.5.2
Proficiency Level: List 6-8B/C S601 Paper**
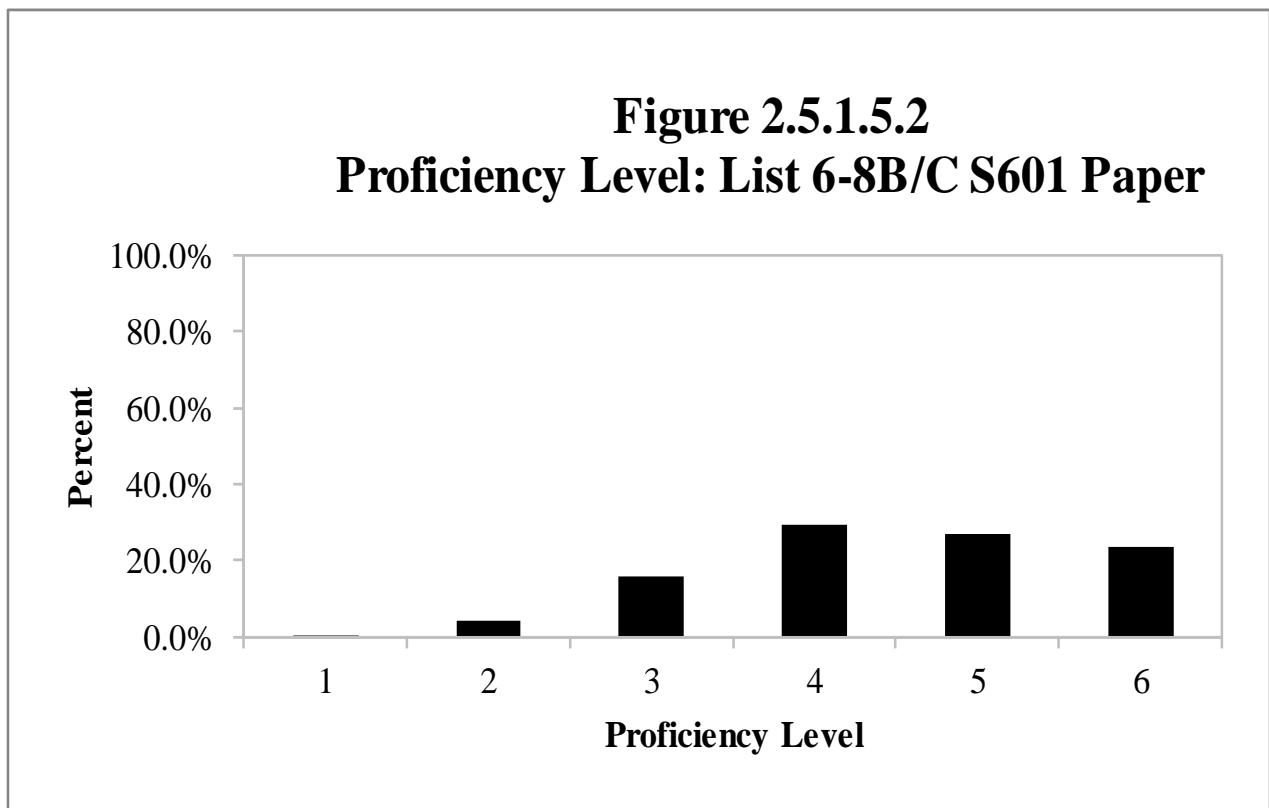
Table 2.5.1.5.3

Proficiency Level Distribution: List 6-8 S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,748	9.36%	2,441	14.88%	2,076	13.42%	6,265	12.39%
2	3,306	17.70%	2,568	15.65%	2,916	18.85%	8,790	17.39%
3	2,649	14.19%	2,672	16.29%	2,501	16.16%	7,822	15.47%
4	4,544	24.33%	3,603	21.96%	2,483	16.05%	10,630	21.03%
5	3,514	18.82%	3,022	18.42%	3,031	19.59%	9,567	18.92%
6	2,913	15.60%	2,101	12.81%	2,466	15.94%	7,480	14.80%
Total	18,674	100.00%	16,407	100.00%	15,473	100.00%	50,554	100.00%

Figure 2.5.1.5.3
Proficiency Level: List 6-8 S601 Paper
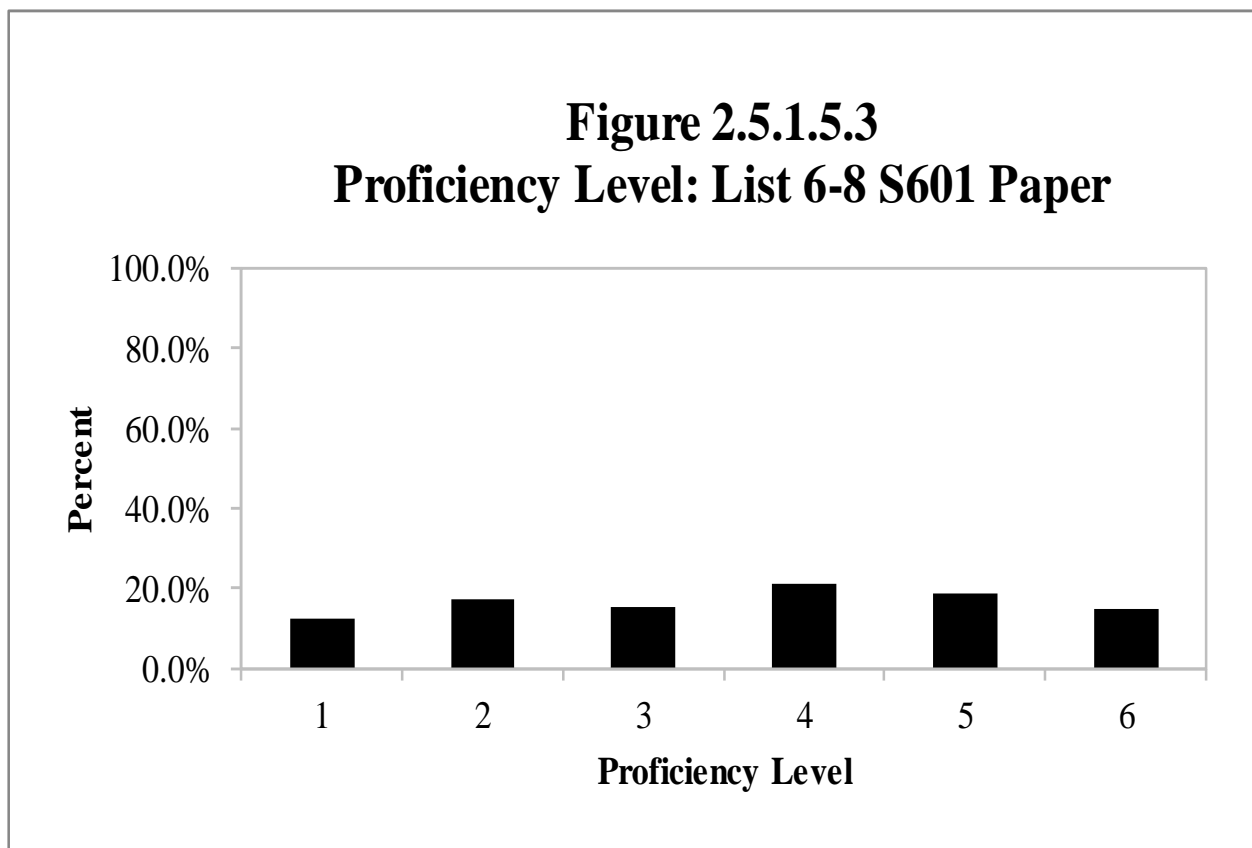
2.5.1.6 Grades 9-12

Table 2.5.1.6.1

Proficiency Level Distribution: List 9-12 A S601 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,962	44.70%	2,857	55.64%	1,842	48.03%	1,174	54.96%	8,835	49.83%
2	2,209	33.34%	1,128	21.97%	1,250	32.59%	486	22.75%	5,073	28.61%
3	894	13.49%	901	17.55%	481	12.54%	313	14.65%	2,589	14.60%
4	448	6.76%	140	2.73%	239	6.23%	145	6.79%	972	5.48%
5	113	1.71%	109	2.12%	23	0.60%	18	0.84%	263	1.48%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	6,626	100.00%	5,135	100.00%	3,835	100.00%	2,136	100.00%	17,732	100.00%

Figure 2.5.1.6.1
Proficiency Level: List 9-12A S601 Paper
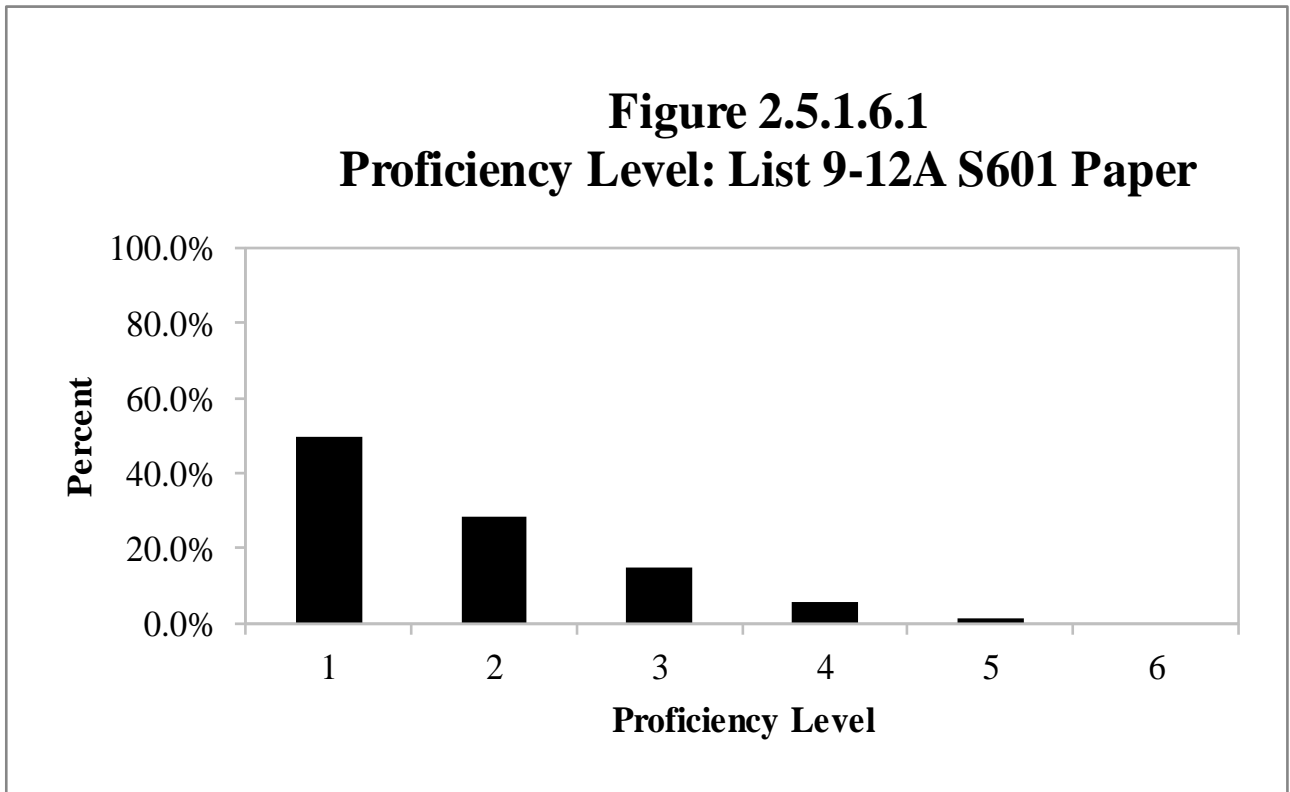
Table 2.5.1.6.2

Proficiency Level Distribution: List 9-12 B/C S601 Paper

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	56	0.65%	73	0.91%	126	1.71%	231	4.19%	486	1.64%
2	493	5.68%	872	10.82%	623	8.44%	802	14.53%	2,790	9.41%
3	1,689	19.46%	1,988	24.67%	2,431	32.95%	1,513	27.41%	7,621	25.71%
4	2,920	33.64%	2,758	34.22%	1,771	24.00%	1,868	33.85%	9,317	31.44%
5	1,921	22.13%	1,558	19.33%	1,520	20.60%	777	14.08%	5,776	19.49%
6	1,601	18.44%	811	10.06%	907	12.29%	328	5.94%	3,647	12.31%
Total	8,680	100.00%	8,060	100.00%	7,378	100.00%	5,519	100.00%	29,637	100.00%

Figure 2.5.1.6.2
Proficiency Level: List 9-12B/C S601 Paper
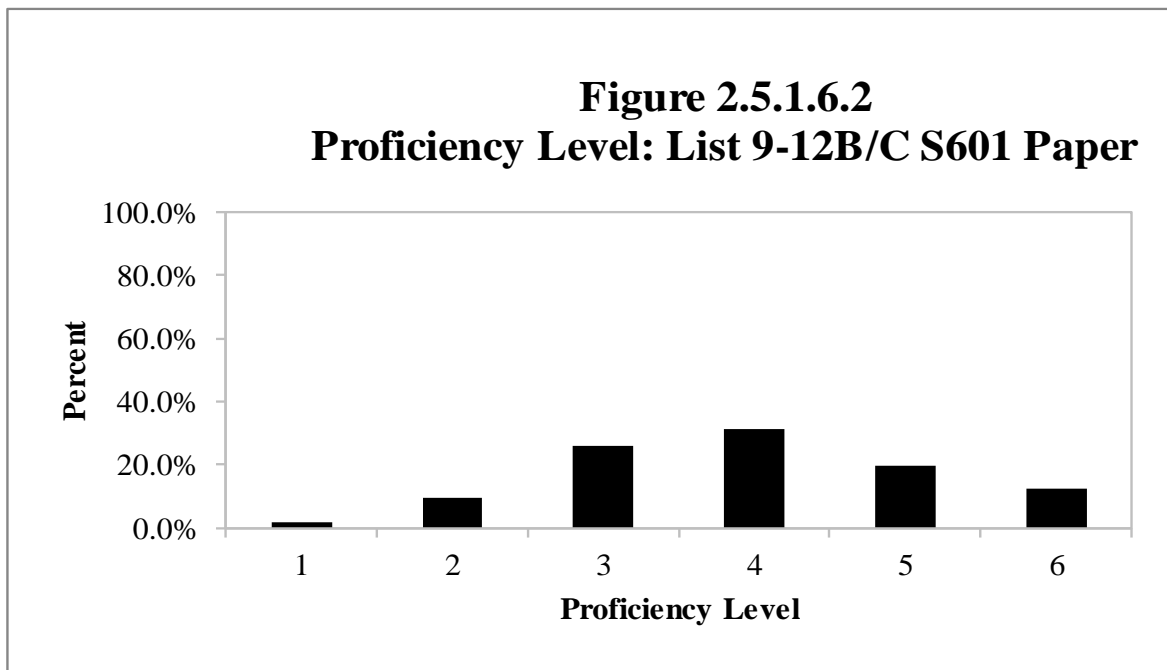
Table 2.5.1.6.3

Proficiency Level Distribution: List 9-12 S601 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	3,018	19.72%	2,930	22.21%	1,968	17.55%	1,405	18.35%	9,321	19.68%
2	2,702	17.65%	2,000	15.16%	1,873	16.70%	1,288	16.83%	7,863	16.60%
3	2,583	16.88%	2,889	21.89%	2,912	25.97%	1,826	23.85%	10,210	21.55%
4	3,368	22.00%	2,898	21.96%	2,010	17.93%	2,013	26.30%	10,289	21.72%
5	2,034	13.29%	1,667	12.63%	1,543	13.76%	795	10.39%	6,039	12.75%
6	1,601	10.46%	811	6.15%	907	8.09%	328	4.28%	3,647	7.70%
Total	15,306	100.00%	13,195	100.00%	11,213	100.00%	7,655	100.00%	47,369	100.00%

Figure 2.5.1.6.3
Proficiency Level: List 9-12 S601 Paper
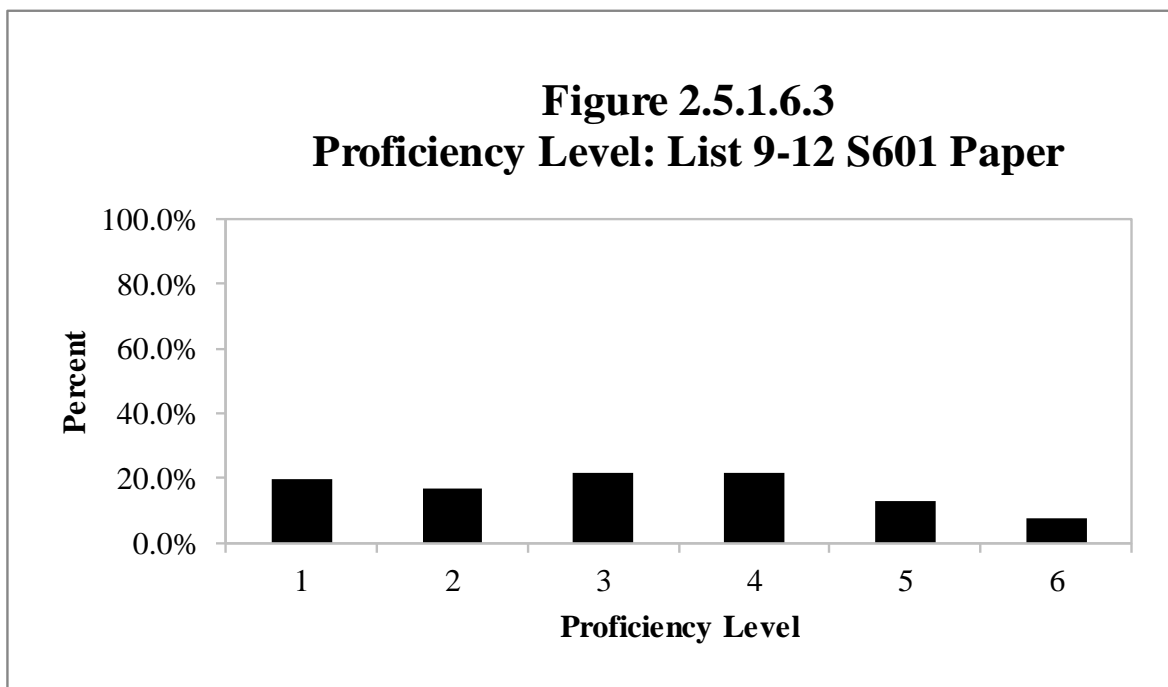
2.5.2 Reading

2.5.2.0 Kindergarten

Table 2.5.2.0

Proficiency Level Distribution: Read K S601 Paper

Level	Grade K		Total	
	Count	**Percent**	**Count**	**Percent**
1	202,530	78.86%	202,530	78.86%
2	6,348	2.47%	6,348	2.47%
3	16,787	6.54%	16,787	6.54%
4	11,214	4.37%	11,214	4.37%
5	19,958	7.77%	19,958	7.77%
6	0	0.00%	0	0.00%
Total	256,837	100.00%	256,837	100.00%

Figure 2.5.2.0
Proficiency Level: Read K S601 Paper
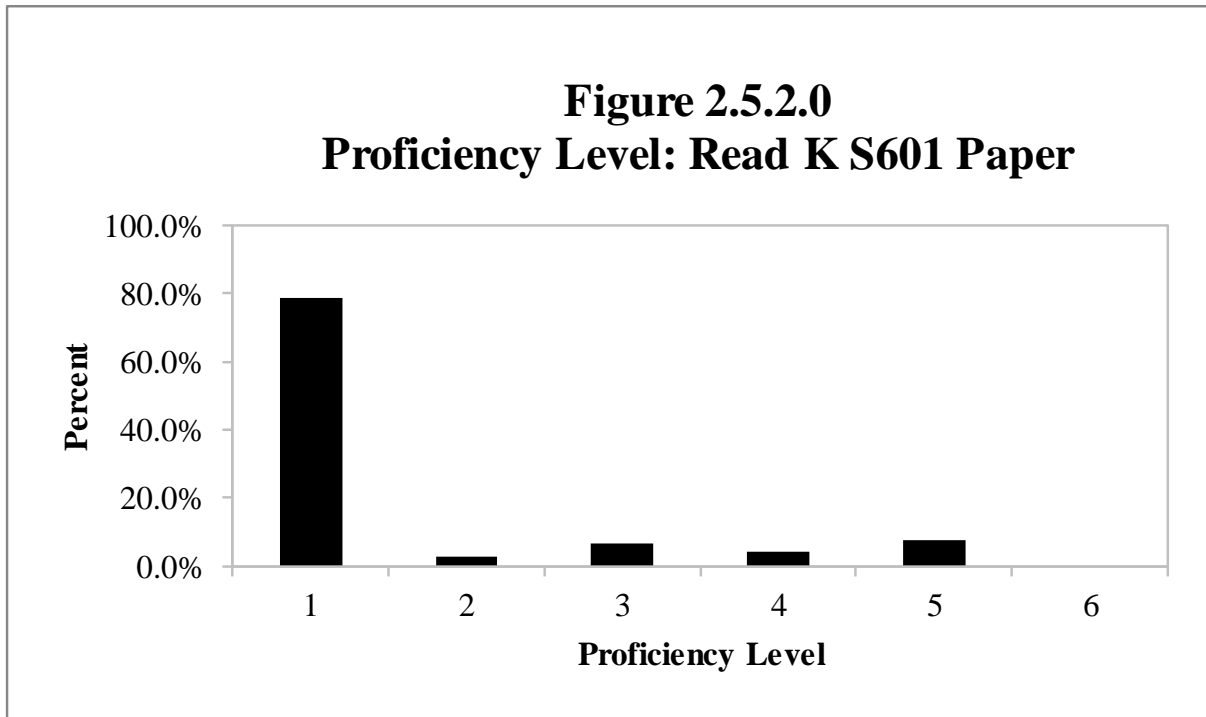
2.5.2.1 Grade 1

Table 2.5.2.1.1

Proficiency Level Distribution: Read 1 A S601 Paper

Level	Grade 1		Total	
	Count	**Percent**	**Count**	**Percent**
1	8,207	45.04%	8,207	45.04%
2	6,180	33.91%	6,180	33.91%
3	2,116	11.61%	2,116	11.61%
4	835	4.58%	835	4.58%
5	524	2.88%	524	2.88%
6	361	1.98%	361	1.98%
Total	18,223	100.00%	18,223	100.00%

**Figure 2.5.2.1.1
Proficiency Level: Read 1A S601 Paper**
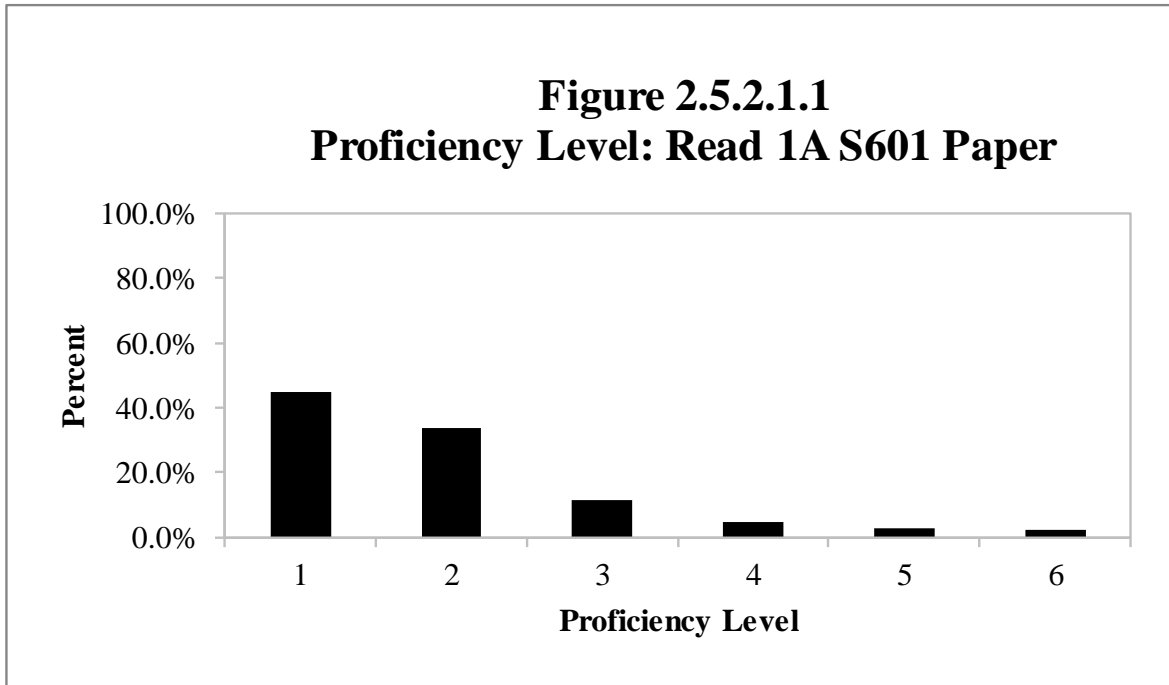
Table 2.5.2.1.2

Proficiency Level Distribution: Read 1 B/C S601 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	143	1.18%	143	1.18%
2	2,924	24.17%	2,924	24.17%
3	4,933	40.77%	4,933	40.77%
4	1,367	11.30%	1,367	11.30%
5	1,503	12.42%	1,503	12.42%
6	1,229	10.16%	1,229	10.16%
Total	12,099	100.00%	12,099	100.00%

Figure 2.5.2.1.2
Proficiency Level: Read 1B/C S601 Paper
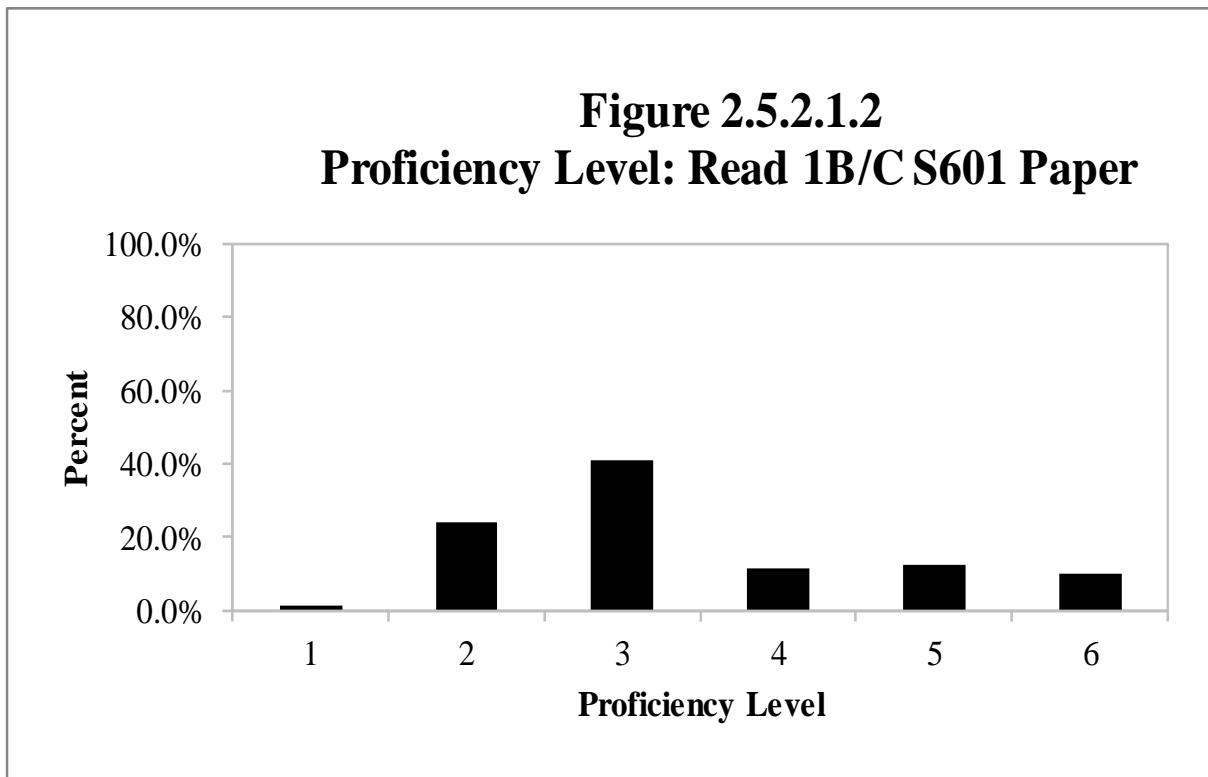
Table 2.5.2.1.3

Proficiency Level Distribution: Read 1 S601 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	8,350	27.54%	8,350	27.54%
2	9,104	30.02%	9,104	30.02%
3	7,049	23.25%	7,049	23.25%
4	2,202	7.26%	2,202	7.26%
5	2,027	6.68%	2,027	6.68%
6	1,590	5.24%	1,590	5.24%
Total	30,322	100.00%	30,322	100.00%

Figure 2.5.2.1.3
Proficiency Level: Read 1 S601 Paper
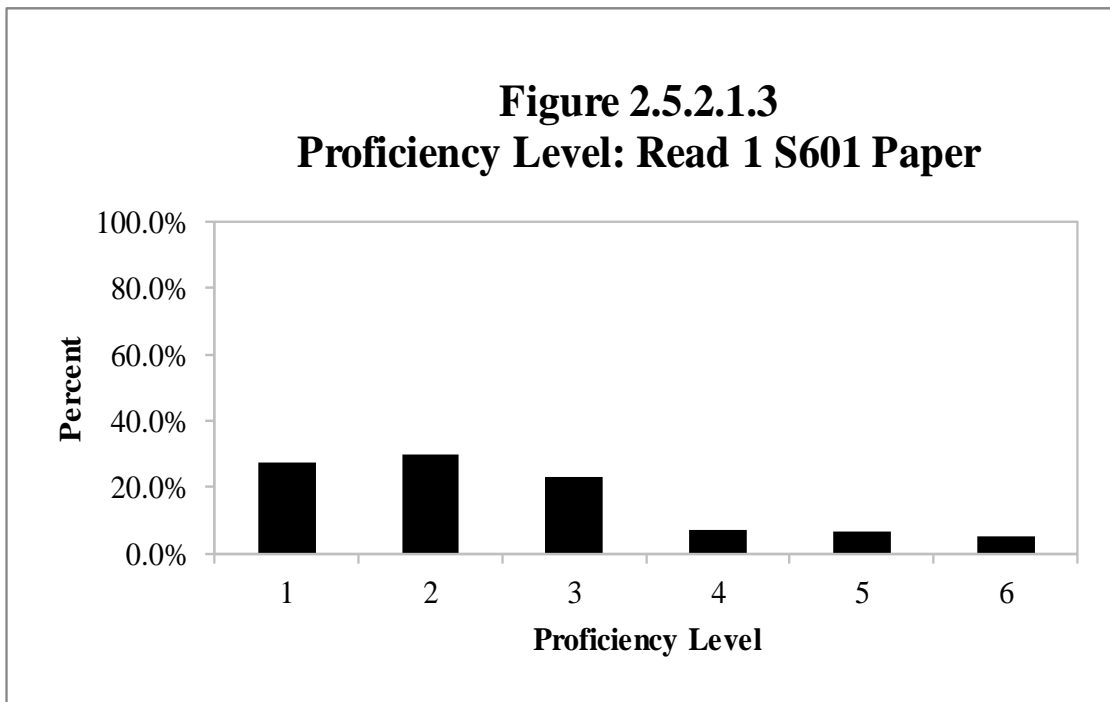
2.5.2.2 Grade 2

Table 2.5.2.2.1

Proficiency Level Distribution: Read 2 A S601 Paper

Level	Grade 2		Total	
	Count	**Percent**	**Count**	**Percent**
1	5,816	58.13%	5,816	58.13%
2	2,507	25.06%	2,507	25.06%
3	928	9.28%	928	9.28%
4	248	2.48%	248	2.48%
5	407	4.07%	407	4.07%
6	99	0.99%	99	0.99%
Total	10,005	100.00%	10,005	100.00%

Figure 2.5.2.2.1
Proficiency Level: Read 2A S601 Paper
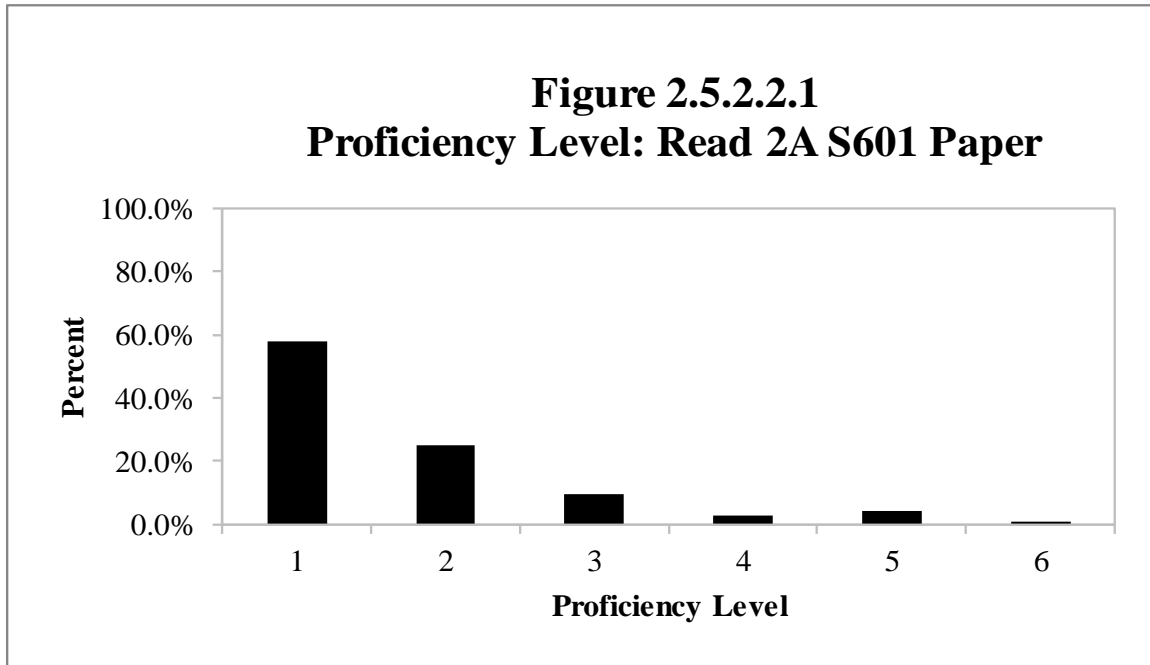
Table 2.5.2.2.2

Proficiency Level Distribution: Read 2 B/C S601 Paper

Level	Grade 2		Total	
	Count	Percent	Count	Percent
1	1,290	6.88%	1,290	6.88%
2	5,544	29.57%	5,544	29.57%
3	4,995	26.64%	4,995	26.64%
4	1,967	10.49%	1,967	10.49%
5	2,743	14.63%	2,743	14.63%
6	2,209	11.78%	2,209	11.78%
Total	18,748	100.00%	18,748	100.00%

Figure 2.5.2.2.2
Proficiency Level: Read 2B/C S601 Paper
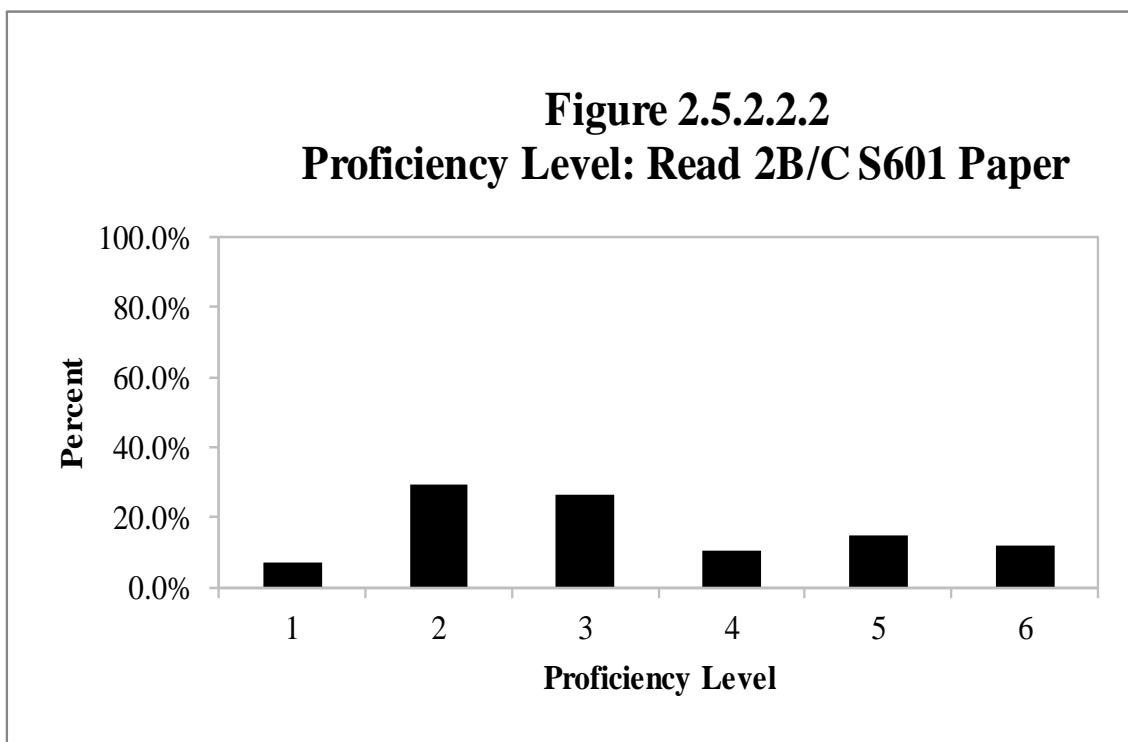
Table 2.5.2.2.3

Proficiency Level Distribution: Read 2 S601 Paper

Level	Grade 2		Total	
	Count	**Percent**	**Count**	**Percent**
1	7,106	24.71%	7,106	24.71%
2	8,051	28.00%	8,051	28.00%
3	5,923	20.60%	5,923	20.60%
4	2,215	7.70%	2,215	7.70%
5	3,150	10.96%	3,150	10.96%
6	2,308	8.03%	2,308	8.03%
Total	28,753	100.00%	28,753	100.00%

Figure 2.5.2.2.3
Proficiency Level: Read 2 S601 Paper
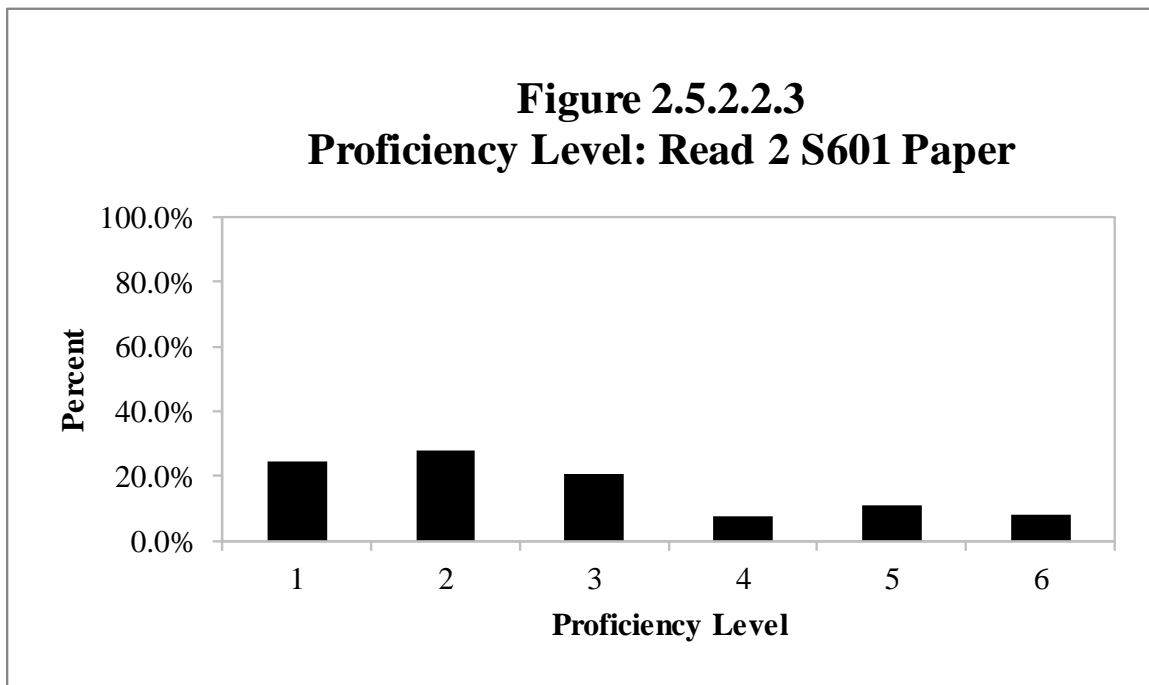
2.5.2.3 Grade 3

Table 2.5.2.3.1

Proficiency Level Distribution: Read 3 A S601 Paper

Level	Grade 3		Total	
	Count	**Percent**	**Count**	**Percent**
1	4,282	51.85%	4,282	51.85%
2	2,500	30.27%	2,500	30.27%
3	909	11.01%	909	11.01%
4	197	2.39%	197	2.39%
5	276	3.34%	276	3.34%
6	95	1.15%	95	1.15%
Total	8,259	100.00%	8,259	100.00%

Figure 2.5.2.3.1
Proficiency Level: Read 3A S601 Paper
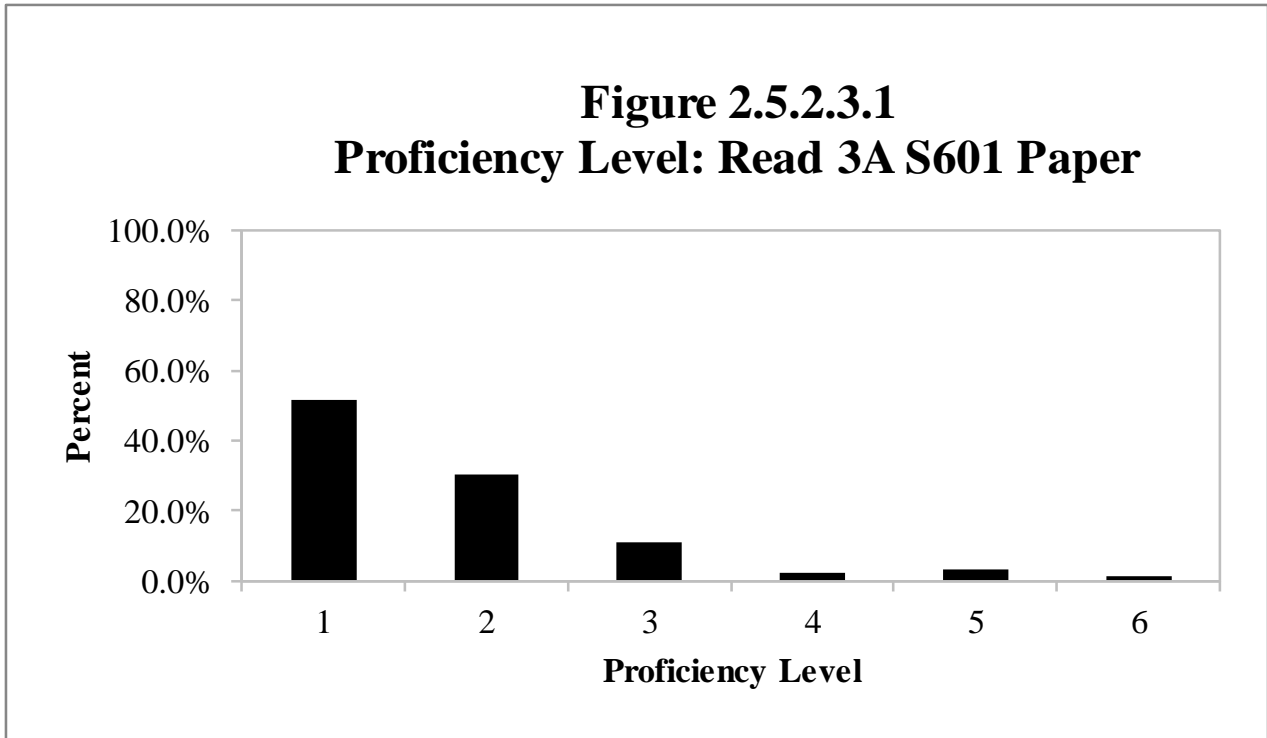
Table 2.5.2.3.2

Proficiency Level Distribution: Read 3 B/C S601 Paper

Level	Grade 3 Count	Grade 3 Percent	Total Count	Total Percent
1	139	0.67%	139	0.67%
2	2,217	10.75%	2,217	10.75%
3	9,197	44.58%	9,197	44.58%
4	3,740	18.13%	3,740	18.13%
5	3,612	17.51%	3,612	17.51%
6	1,726	8.37%	1,726	8.37%
Total	20,631	100.00%	20,631	100.00%

**Figure 2.5.2.3.2
Proficiency Level: Read 3B/C S601 Paper**
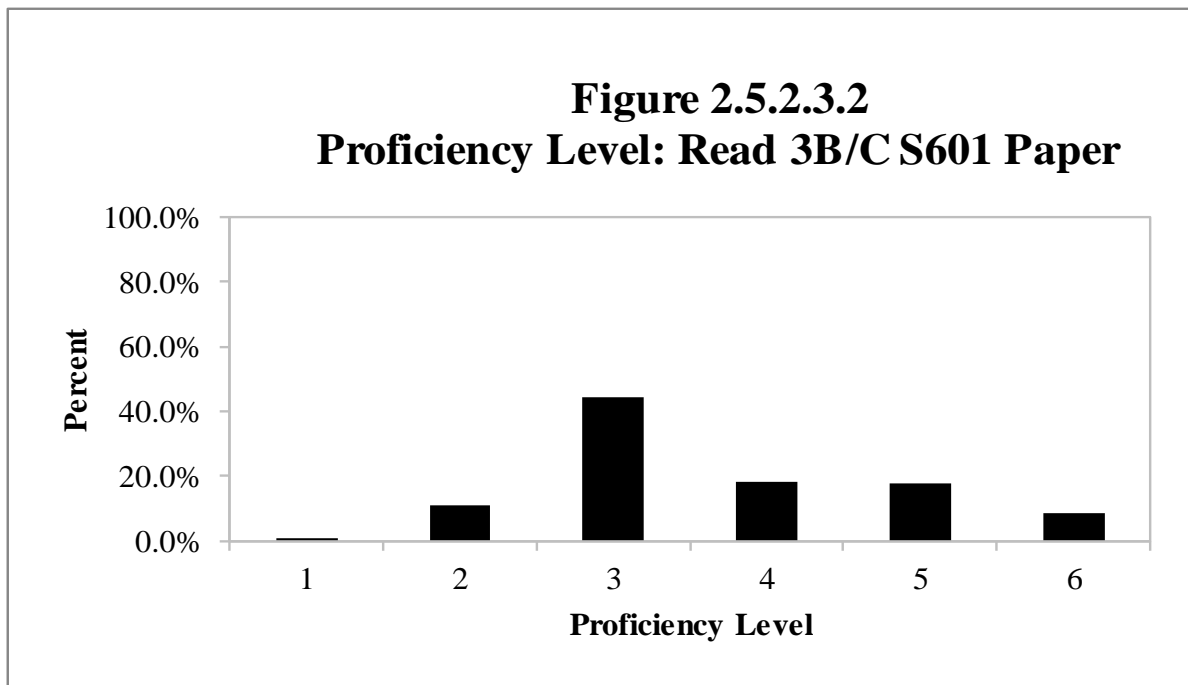
Table 2.5.2.3.3

Proficiency Level Distribution: Read 3 S601 Paper

Level	Grade 3 Count	Grade 3 Percent	Total Count	Total Percent
1	4,421	15.30%	4,421	15.30%
2	4,717	16.33%	4,717	16.33%
3	10,106	34.98%	10,106	34.98%
4	3,937	13.63%	3,937	13.63%
5	3,888	13.46%	3,888	13.46%
6	1,821	6.30%	1,821	6.30%
Total	28,890	100.00%	28,890	100.00%

Figure 2.5.2.3.3
Proficiency Level: Read 3 S601 Paper
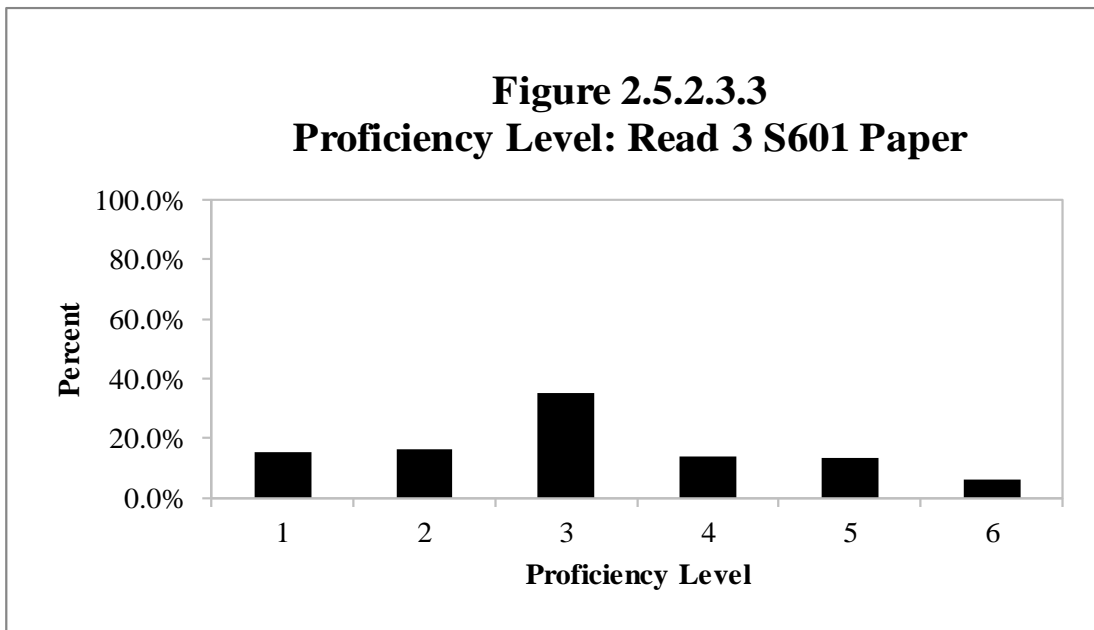
2.5.2.4 Grades 4-5

Table 2.5.2.4.1

Proficiency Level Distribution: Read 4-5 A S601 Paper

Level	Grade 4		Grade 5		Total	
	Count	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,452	55.99%	3,890	60.69%	7,342	58.39%
2	1,811	29.38%	1,497	23.35%	3,308	26.31%
3	404	6.55%	518	8.08%	922	7.33%
4	165	2.68%	243	3.79%	408	3.24%
5	290	4.70%	262	4.09%	552	4.39%
6	43	0.70%	0	0.00%	43	0.34%
Total	6,165	100.00%	6,410	100.00%	12,575	100.00%

Figure 2.5.2.4.1
Proficiency Level: Read 4-5A S601 Paper
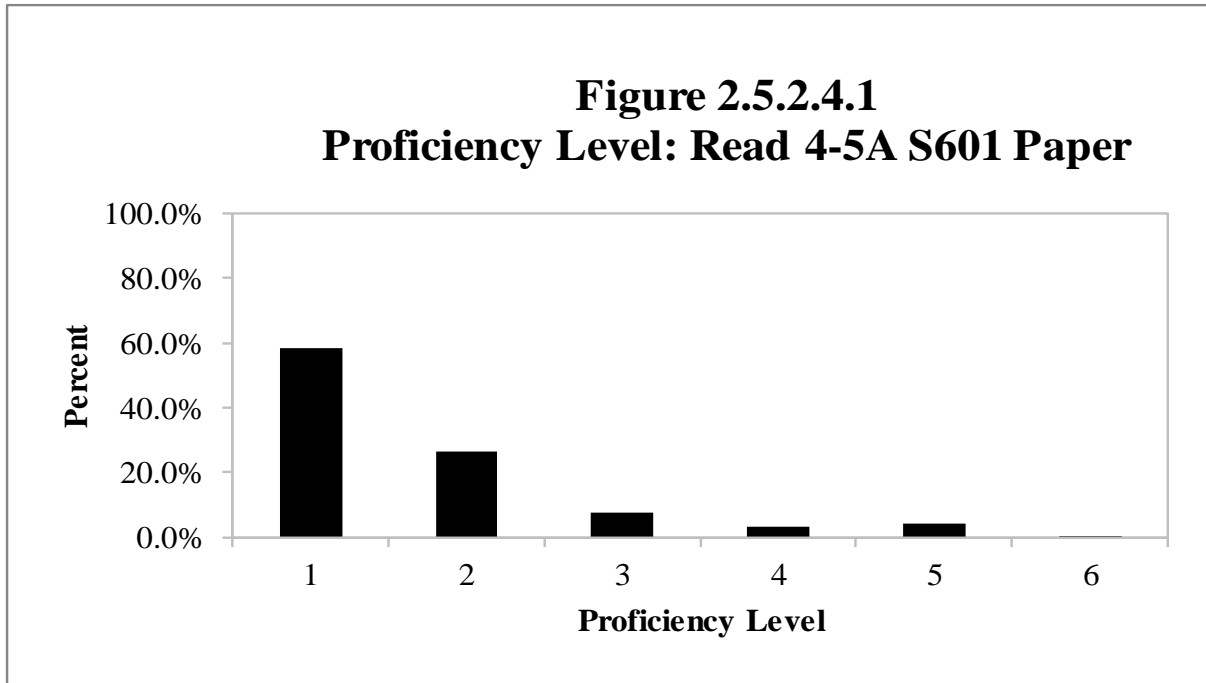
Table 2.5.2.4.2

Proficiency Level Distribution: Read 4-5 B/C S601 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	133	0.88%	216	1.59%	349	1.22%
2	2,212	14.67%	3,056	22.51%	5,268	18.38%
3	5,042	33.45%	4,611	33.96%	9,653	33.69%
4	3,306	21.93%	1,691	12.45%	4,997	17.44%
5	2,779	18.43%	2,604	19.18%	5,383	18.79%
6	1,603	10.63%	1,401	10.32%	3,004	10.48%
Total	15,075	100.00%	13,579	100.00%	28,654	100.00%

Figure 2.5.2.4.2
Proficiency Level: Read 4-5B/C S601 Paper
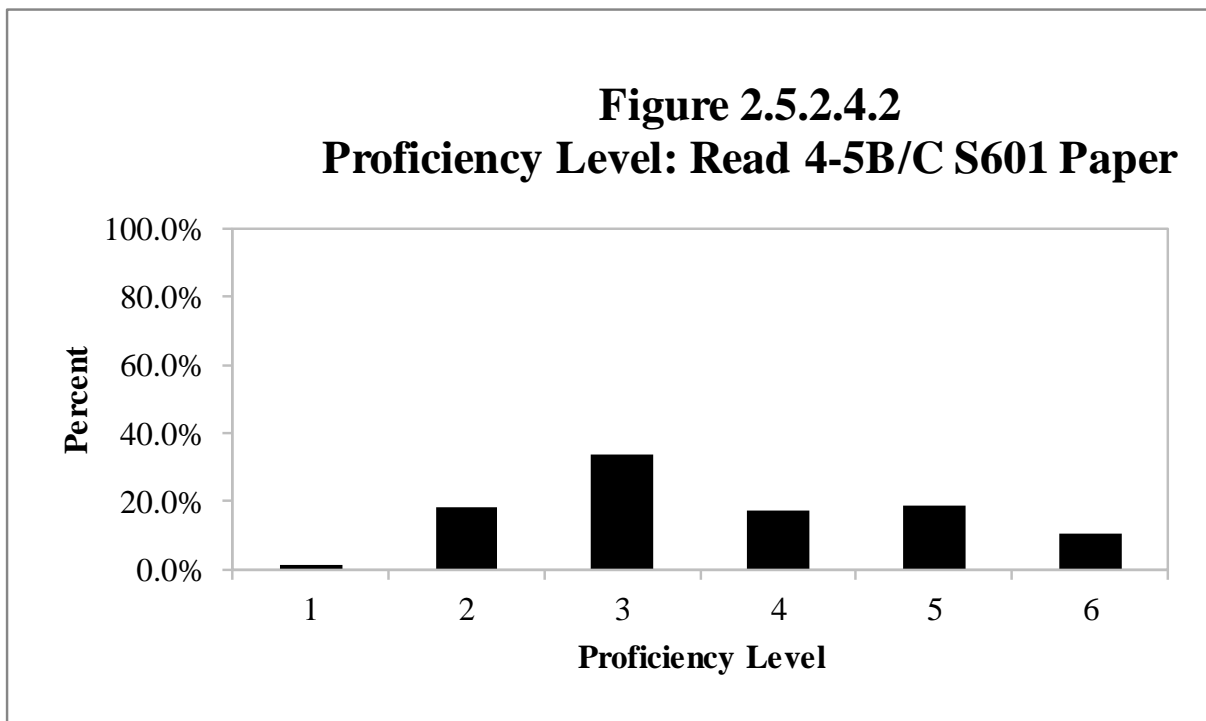
Table 2.5.2.4.3

Proficiency Level Distribution: Read 4-5 S601 Paper

Level	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	3,585	16.88%	4,106	20.54%	7,691	18.65%
2	4,023	18.94%	4,553	22.78%	8,576	20.80%
3	5,446	25.64%	5,129	25.66%	10,575	25.65%
4	3,471	16.34%	1,934	9.68%	5,405	13.11%
5	3,069	14.45%	2,866	14.34%	5,935	14.40%
6	1,646	7.75%	1,401	7.01%	3,047	7.39%
Total	21,240	100.00%	19,989	100.00%	41,229	100.00%

Figure 2.5.2.4.3
Proficiency Level: Read 4-5 S601 Paper
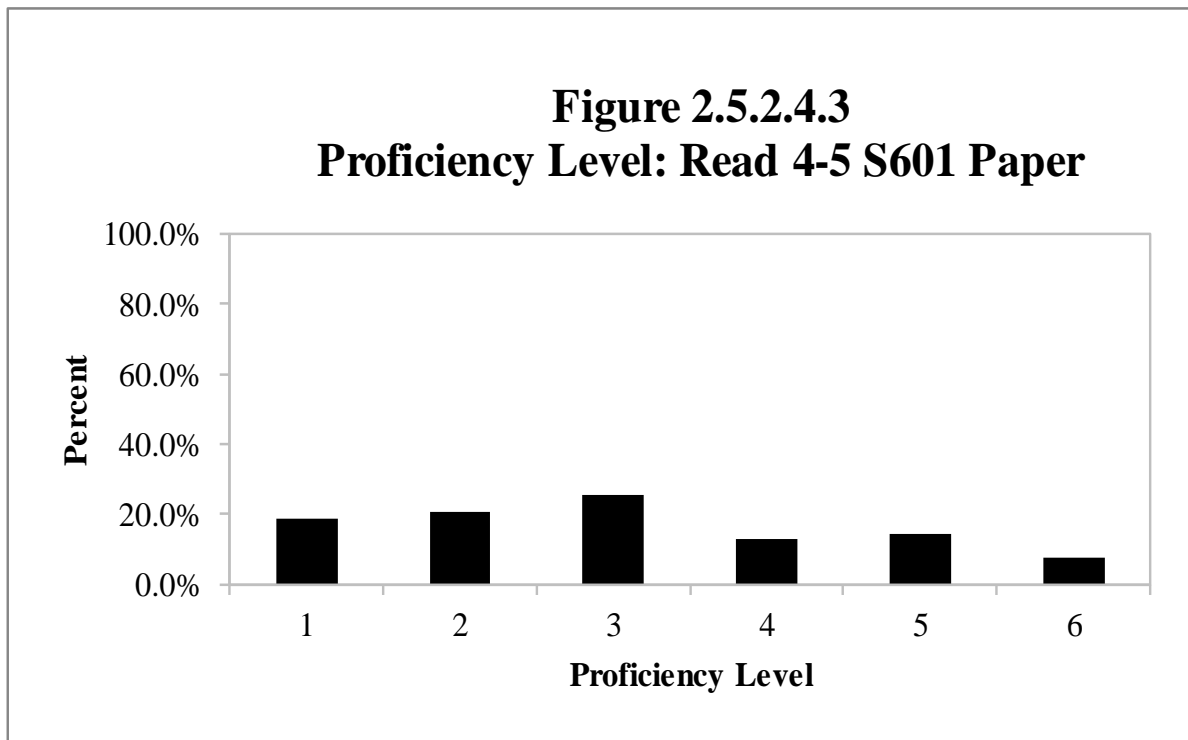
2.5.2.5 Grades 6-8

Table 2.5.2.5.1

Proficiency Level Distribution: Read 6-8 A S601 Paper

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	2,959	47.45%	3,213	51.50%	3,165	52.82%	9,337	50.56%
2	2,410	38.65%	2,186	35.04%	1,967	32.83%	6,563	35.54%
3	546	8.76%	522	8.37%	555	9.26%	1,623	8.79%
4	110	1.76%	100	1.60%	112	1.87%	322	1.74%
5	127	2.04%	160	2.56%	93	1.55%	380	2.06%
6	84	1.35%	58	0.93%	100	1.67%	242	1.31%
Total	6,236	100.00%	6,239	100.00%	5,992	100.00%	18,467	100.00%

Figure 2.5.2.5.1
Proficiency Level: Read 6-8A S601 Paper
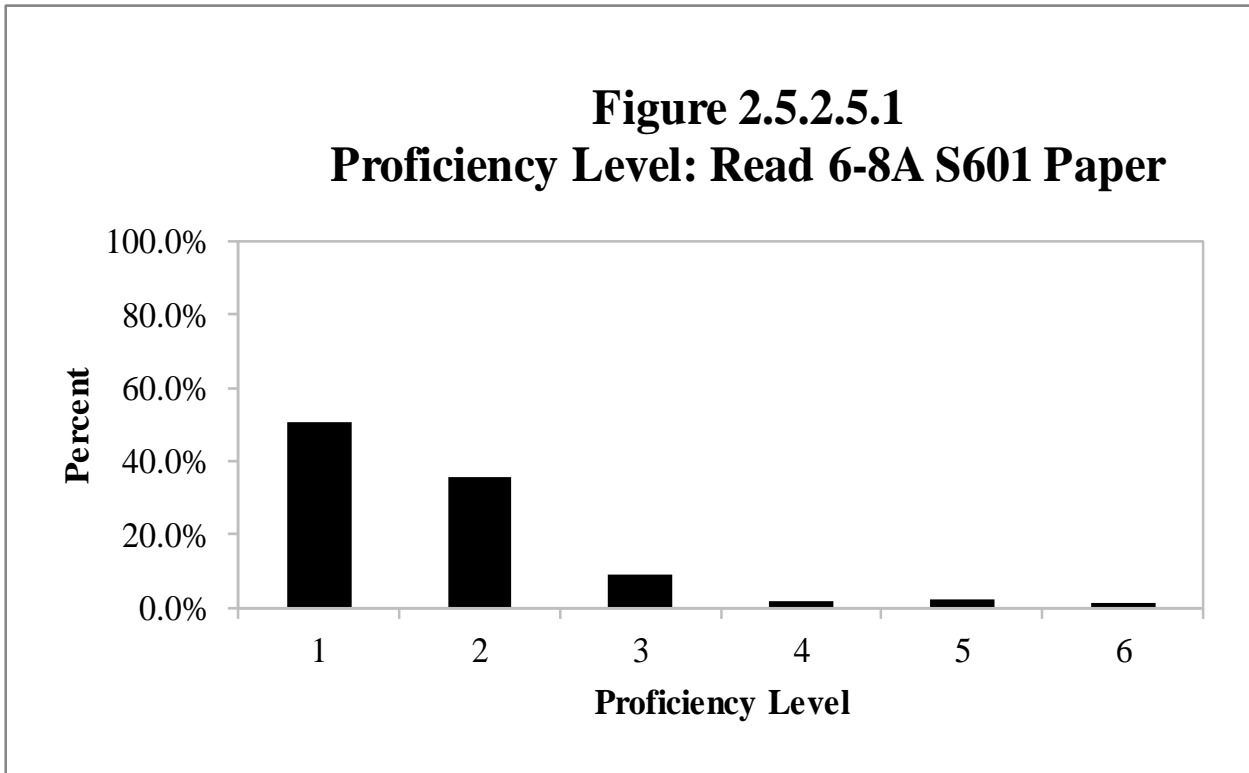
Table 2.5.2.5.2

Proficiency Level Distribution: Read 6-8 B/C S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	418	3.70%	474	5.09%	593	6.69%	1,485	5.04%
2	5,448	48.20%	3,787	40.65%	3,381	38.16%	12,616	42.80%
3	3,192	28.24%	2,959	31.77%	2,302	25.98%	8,453	28.68%
4	979	8.66%	869	9.33%	1,025	11.57%	2,873	9.75%
5	972	8.60%	807	8.66%	1,034	11.67%	2,813	9.54%
6	294	2.60%	419	4.50%	524	5.91%	1,237	4.20%
Total	11,303	100.00%	9,315	100.00%	8,859	100.00%	29,477	100.00%

Figure 2.5.2.5.2
Proficiency Level: Read 6-8B/C S601 Paper
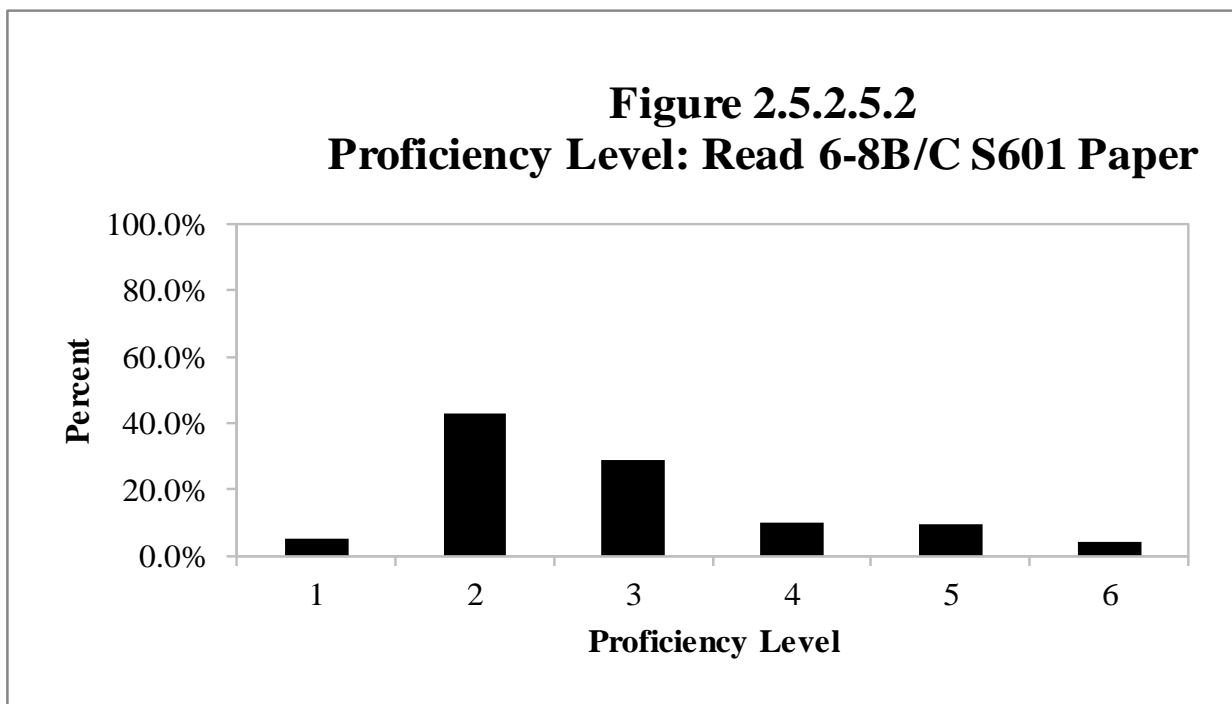
Table 2.5.2.5.3

Proficiency Level Distribution: Read 6-8 S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	3,377	19.25%	3,687	23.70%	3,758	25.30%	10,822	22.57%
2	7,858	44.80%	5,973	38.40%	5,348	36.01%	19,179	40.00%
3	3,738	21.31%	3,481	22.38%	2,857	19.24%	10,076	21.02%
4	1,089	6.21%	969	6.23%	1,137	7.66%	3,195	6.66%
5	1,099	6.27%	967	6.22%	1,127	7.59%	3,193	6.66%
6	378	2.16%	477	3.07%	624	4.20%	1,479	3.08%
Total	17,539	100.00%	15,554	100.00%	14,851	100.00%	47,944	100.00%

Figure 2.5.2.5.3
Proficiency Level: Read 6-8 S601 Paper
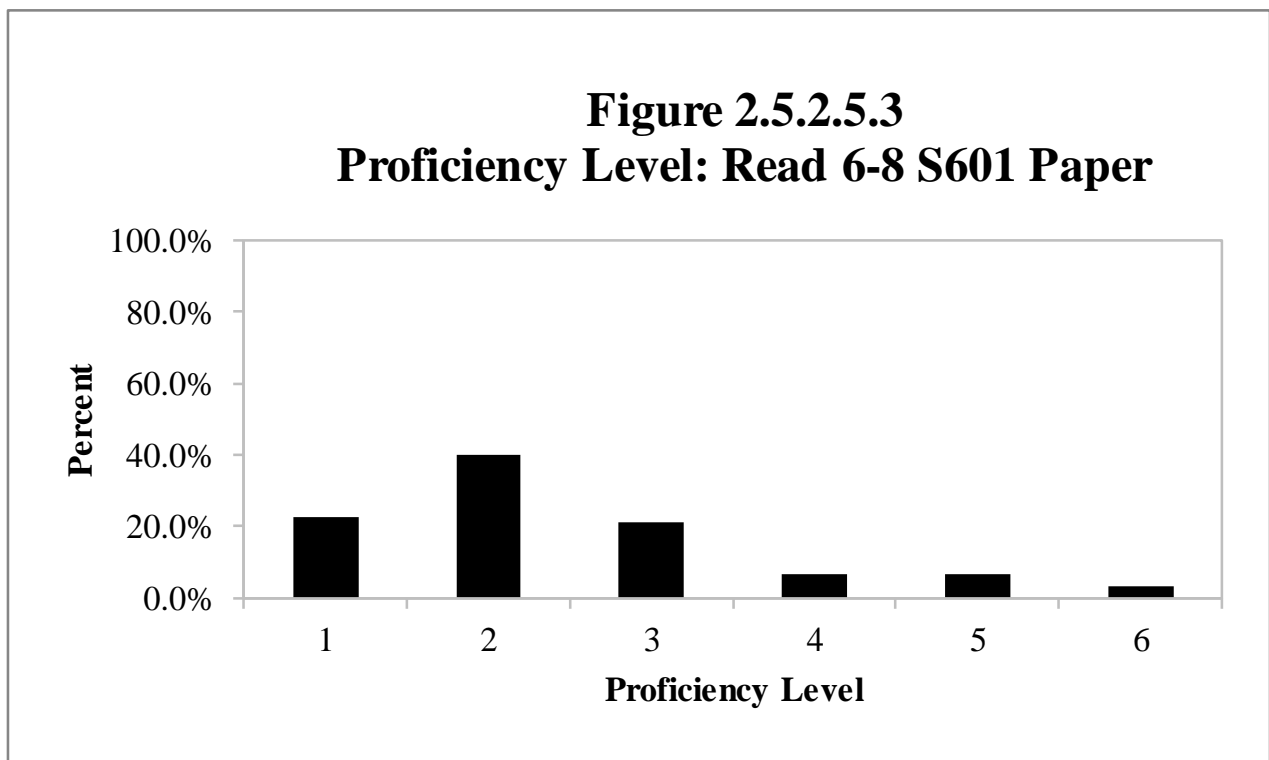
2.5.2.6 Grades 9-12

Table 2.5.2.6.1

Proficiency Level Distribution: Read 9-12 A S601 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,686	40.81%	1,857	36.32%	1,346	35.25%	798	37.66%	6,687	37.93%
2	2,282	34.67%	2,151	42.07%	1,595	41.78%	874	41.25%	6,902	39.14%
3	926	14.07%	656	12.83%	391	10.24%	254	11.99%	2,227	12.63%
4	264	4.01%	183	3.58%	193	5.06%	92	4.34%	732	4.15%
5	304	4.62%	137	2.68%	238	6.23%	70	3.30%	749	4.25%
6	120	1.82%	129	2.52%	55	1.44%	31	1.46%	335	1.90%
Total	6,582	100.00%	5,113	100.00%	3,818	100.00%	2,119	100.00%	17,632	100.00%

Figure 2.5.2.6.1
Proficiency Level: Read 9-12A S601 Paper
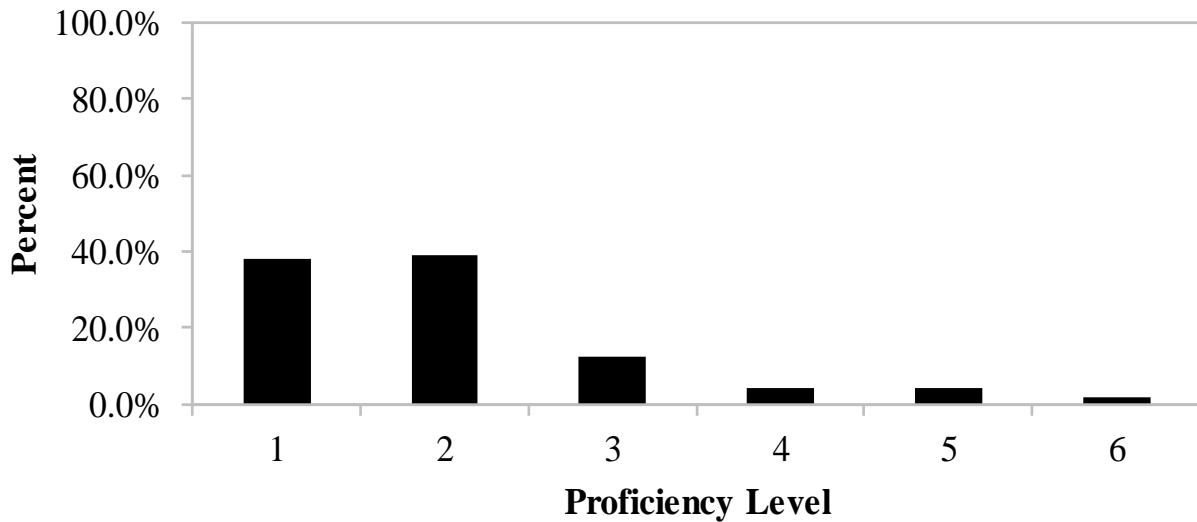
Table 2.5.2.6.2

Proficiency Level Distribution: Read 9-12 B/C S601 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	103	1.30%	111	1.48%	159	2.33%	293	5.56%	666	2.42%
2	2,157	27.22%	2,327	31.07%	2,130	31.15%	1,991	37.80%	8,605	31.27%
3	2,123	26.80%	2,329	31.09%	2,002	29.28%	1,461	27.74%	7,915	28.76%
4	1,435	18.11%	837	11.17%	814	11.91%	306	5.81%	3,392	12.33%
5	1,174	14.82%	1,055	14.09%	1,041	15.23%	741	14.07%	4,011	14.58%
6	931	11.75%	831	11.09%	691	10.11%	475	9.02%	2,928	10.64%
Total	7,923	100.00%	7,490	100.00%	6,837	100.00%	5,267	100.00%	27,517	100.00%

Figure 2.5.2.6.2
Proficiency Level: Read 9-12B/C S601 Paper
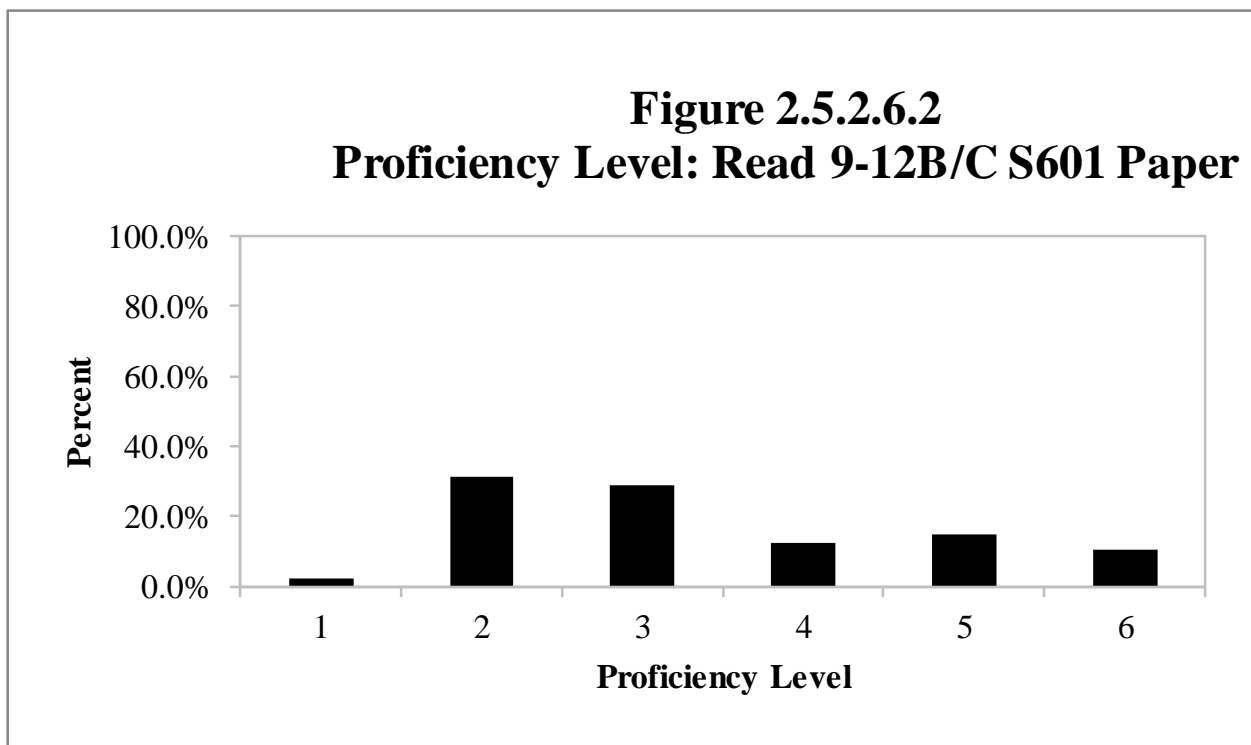
Table 2.5.2.6.3

Proficiency Level Distribution: Read 9-12 S601 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,789	19.23%	1,968	15.62%	1,505	14.12%	1,091	14.77%	7,353	16.29%
2	4,439	30.60%	4,478	35.53%	3,725	34.96%	2,865	38.79%	15,507	34.35%
3	3,049	21.02%	2,985	23.68%	2,393	22.46%	1,715	23.22%	10,142	22.46%
4	1,699	11.71%	1,020	8.09%	1,007	9.45%	398	5.39%	4,124	9.13%
5	1,478	10.19%	1,192	9.46%	1,279	12.00%	811	10.98%	4,760	10.54%
6	1,051	7.25%	960	7.62%	746	7.00%	506	6.85%	3,263	7.23%
Total	14,505	100.00%	12,603	100.00%	10,655	100.00%	7,386	100.00%	45,149	100.00%

**Figure 2.5.2.6.3
Proficiency Level: Read 9-12 S601 Paper**
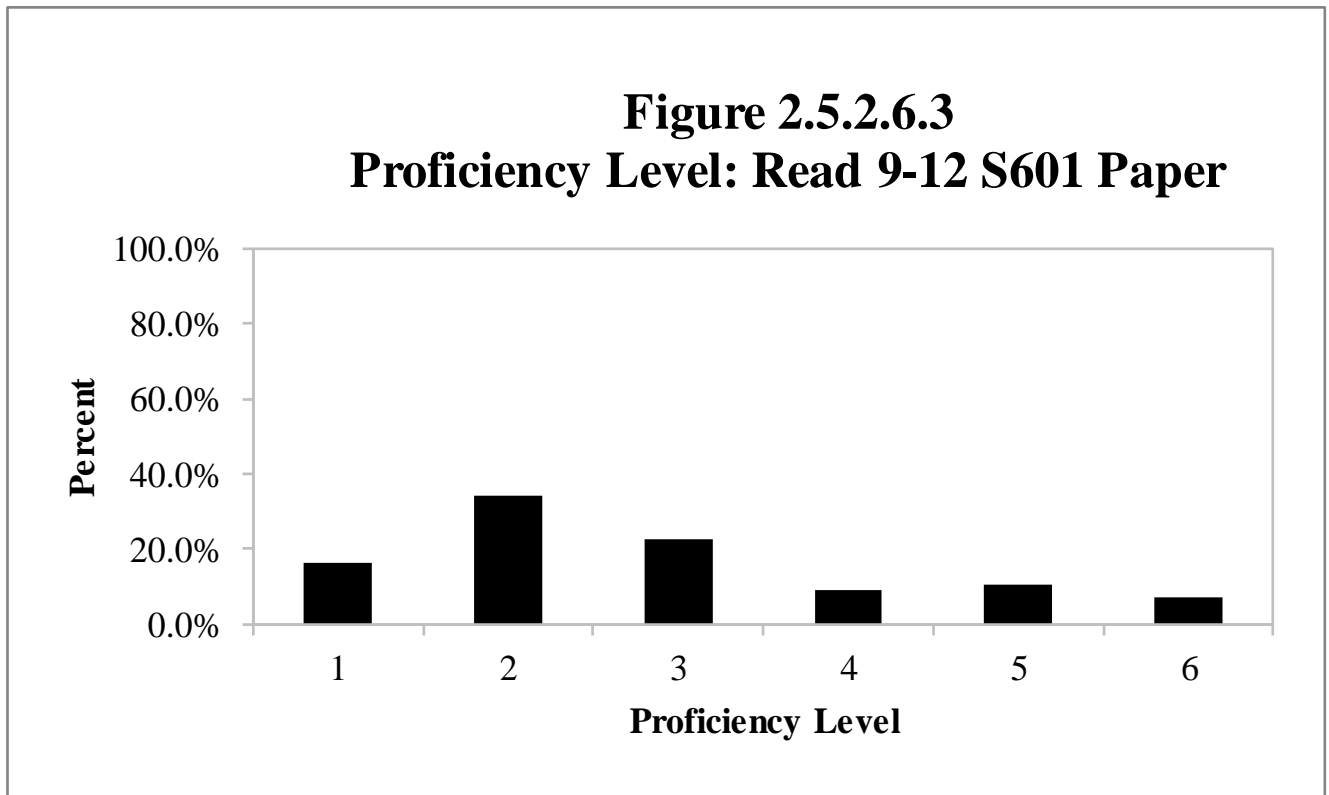
2.5.3 Writing

2.5.3.0 Kindergarten

Table 2.5.3.0

Proficiency Level Distribution: Writ K S601 Paper

Level	Grade K		Total	
	Count	Percent	Count	Percent
1	184,749	71.93%	184,749	71.93%
2	40,661	15.83%	40,661	15.83%
3	24,735	9.63%	24,735	9.63%
4	6,689	2.60%	6,689	2.60%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	256,834	100.00%	256,834	100.00%

Figure 2.5.3.0
Proficiency Level: Writ K S601 Paper
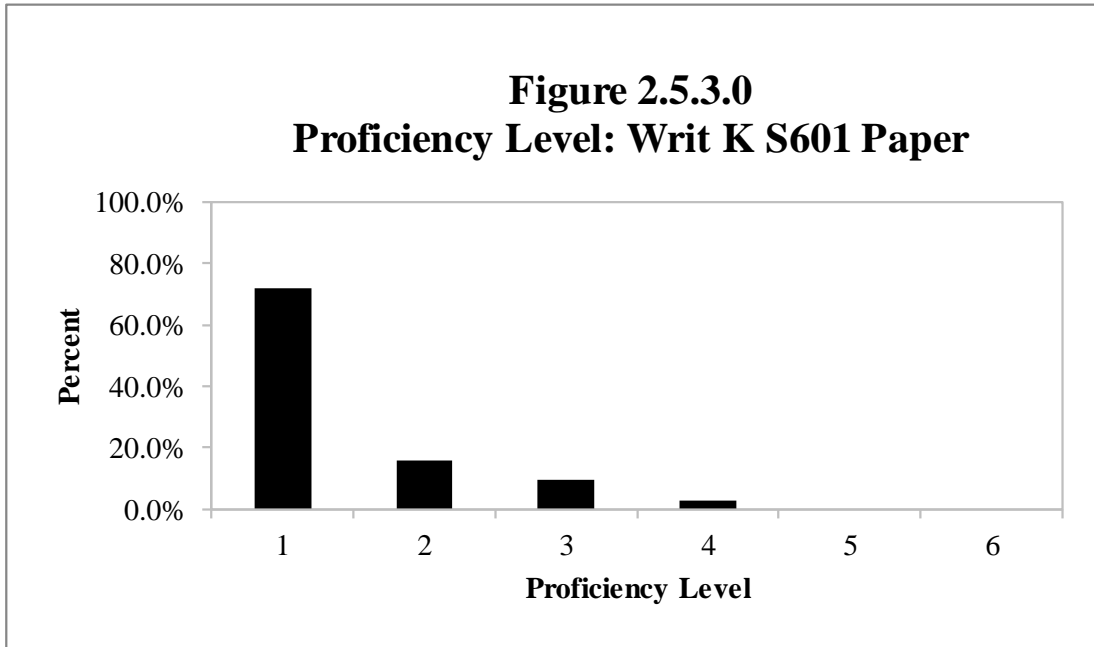
2.5.3.1 Grade 1

Table 2.5.3.1.1

Proficiency Level Distribution: Writ 1 A S601 Paper

Level	Grade 1		Total	
	Count	**Percent**	**Count**	**Percent**
1	13,125	55.44%	13,125	55.44%
2	9,874	41.71%	9,874	41.71%
3	676	2.86%	676	2.86%
4	0	0.00%	0	0.00%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	23,675	100.00%	23,675	100.00%

Figure 2.5.3.1.1
Proficiency Level: Writ 1A S601 Paper
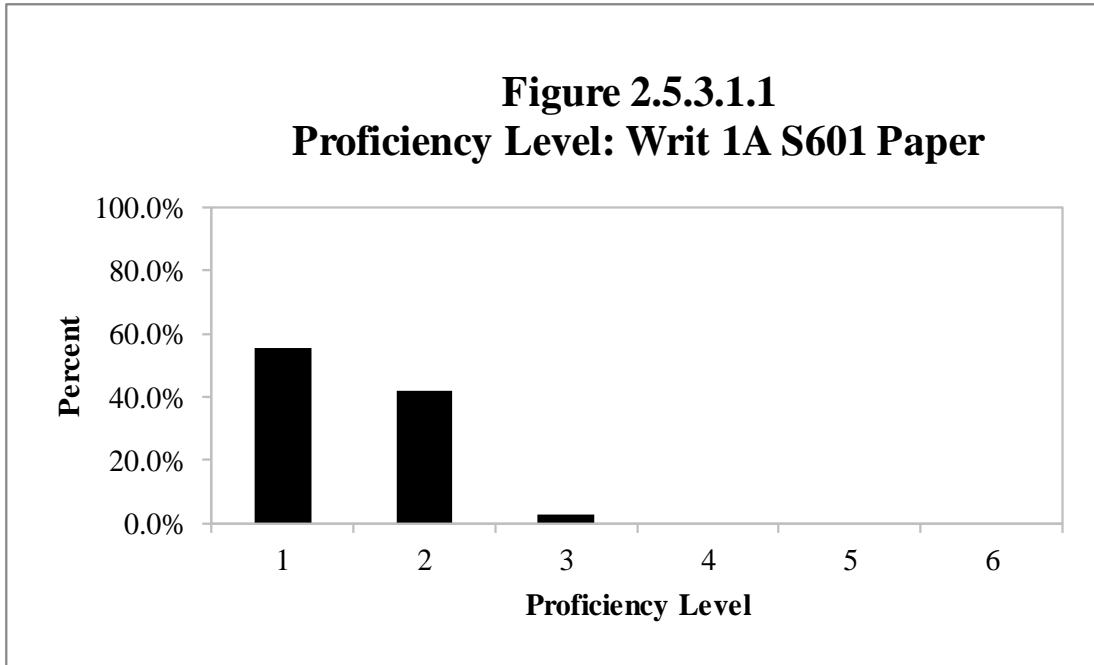
Table 2.5.3.1.2

Proficiency Level Distribution: Writ 1 B/C S601 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	3,024	19.11%	3,024	19.11%
2	6,181	39.05%	6,181	39.05%
3	6,431	40.63%	6,431	40.63%
4	189	1.19%	189	1.19%
5	2	0.01%	2	0.01%
6	0	0.00%	0	0.00%
Total	15,827	100.00%	15,827	100.00%

Figure 2.5.3.1.2
Proficiency Level: Writ 1B/C S601 Paper
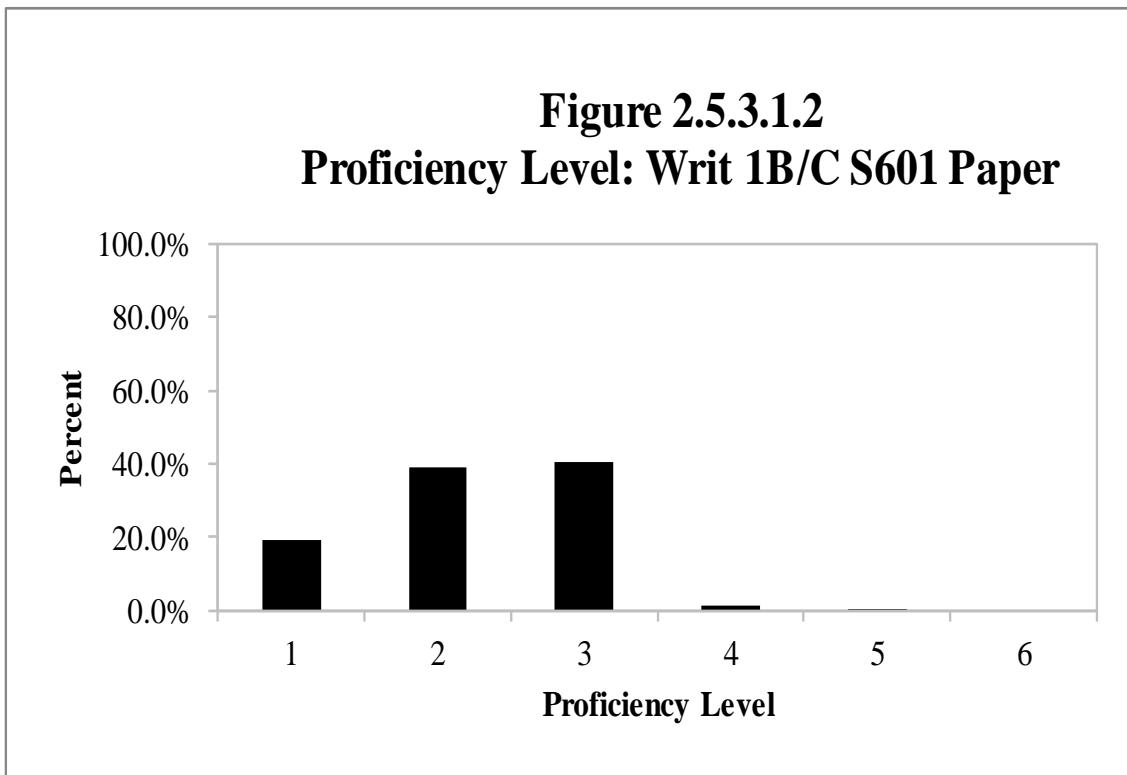
Table 2.5.3.1.3

Proficiency Level Distribution: Writ 1 S601 Paper

Level	Grade 1 Count	Grade 1 Percent	Total Count	Total Percent
1	16,149	40.88%	16,149	40.88%
2	16,055	40.64%	16,055	40.64%
3	7,107	17.99%	7,107	17.99%
4	189	0.48%	189	0.48%
5	2	0.01%	2	0.01%
6	0	0.00%	0	0.00%
Total	39,502	100.00%	39,502	100.00%

Figure 2.5.3.1.3
Proficiency Level: Writ 1 S601 Paper
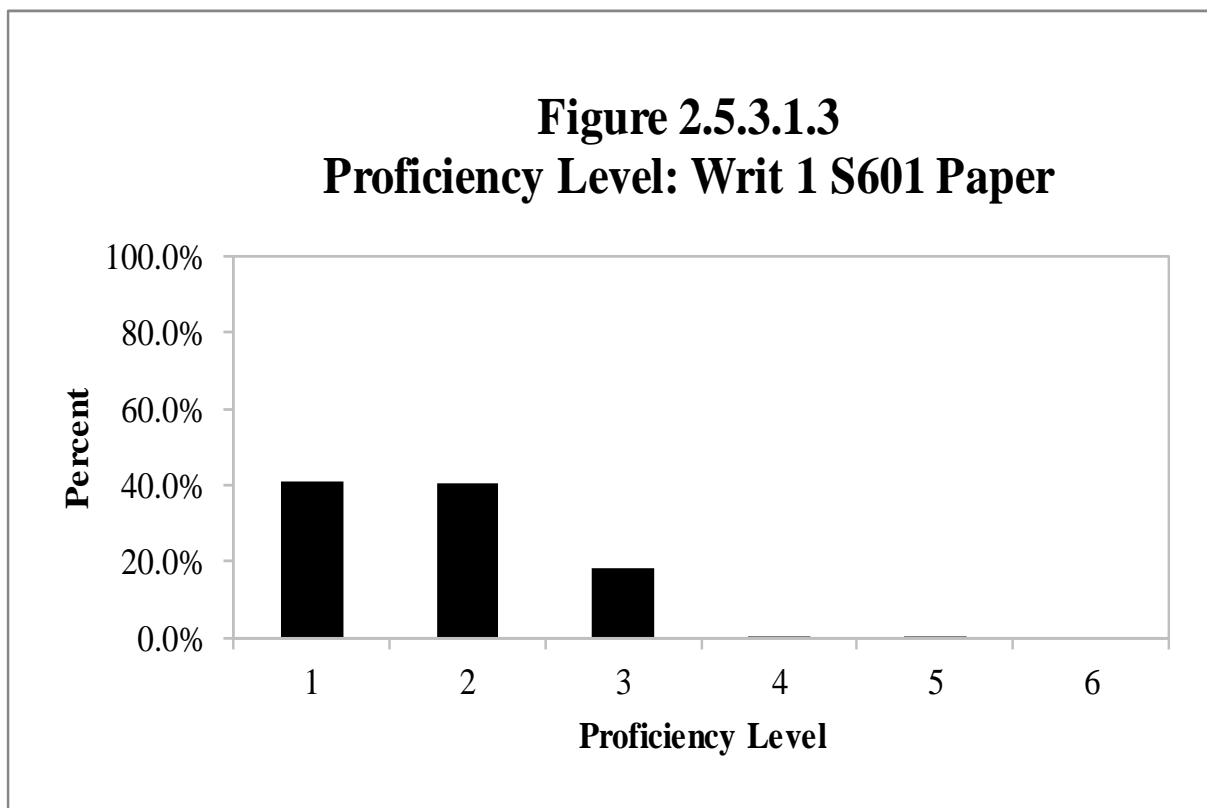
2.5.3.2 Grade 2

Table 2.5.3.2.1

Proficiency Level Distribution: Writ 2 A S601 Paper

Level	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	6,418	52.25%	6,418	52.25%
2	3,958	32.22%	3,958	32.22%
3	1,906	15.52%	1,906	15.52%
4	1	0.01%	1	0.01%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	12,283	100.00%	12,283	100.00%

Figure 2.5.3.2.1
Proficiency Level: Writ 2A S601 Paper
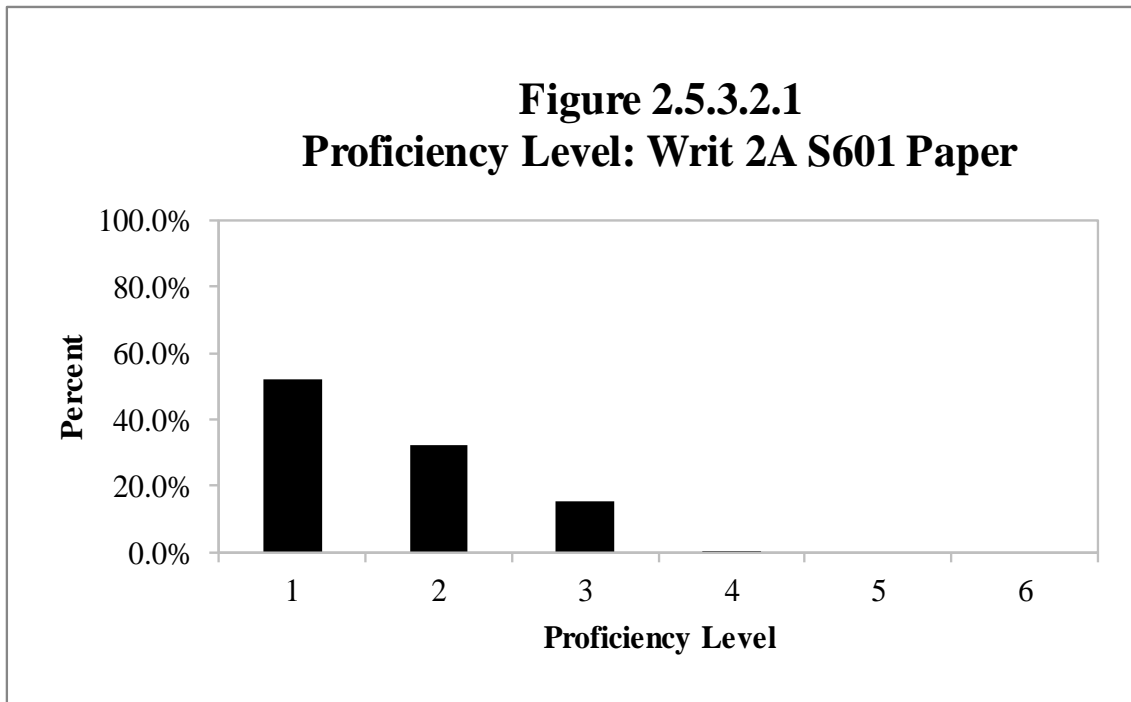
Table 2.5.3.2.2

Proficiency Level Distribution: Writ 2 B/C S601 Paper

Level	Grade 2		Total	
	Count	**Percent**	**Count**	**Percent**
1	2,345	10.21%	2,345	10.21%
2	5,971	25.99%	5,971	25.99%
3	13,264	57.73%	13,264	57.73%
4	1,384	6.02%	1,384	6.02%
5	10	0.04%	10	0.04%
6	0	0.00%	0	0.00%
Total	22,974	100.00%	22,974	100.00%

Figure 2.5.3.2.2
Proficiency Level: Writ 2B/C S601 Paper
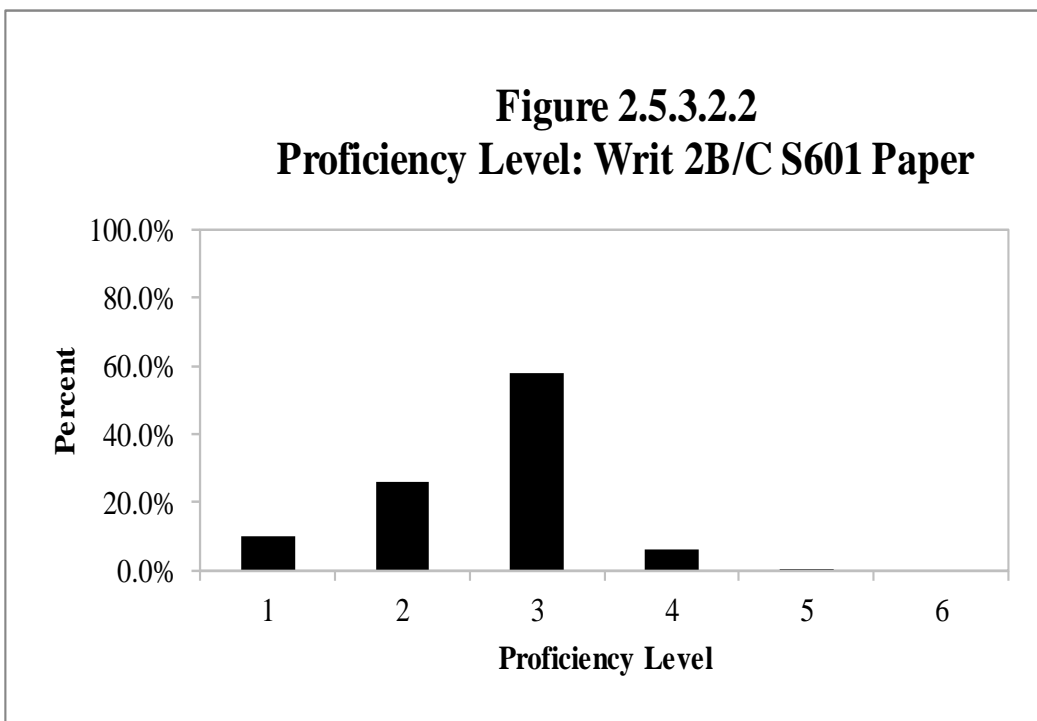
Table 2.5.3.2.3

Proficiency Level Distribution: Writ 2 S601 Paper

Level	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	8,763	24.85%	8,763	24.85%
2	9,929	28.16%	9,929	28.16%
3	15,170	43.03%	15,170	43.03%
4	1,385	3.93%	1,385	3.93%
5	10	0.03%	10	0.03%
6	0	0.00%	0	0.00%
Total	35,257	100.00%	35,257	100.00%

Figure 2.5.3.2.3
Proficiency Level: Writ 2 S601 Paper
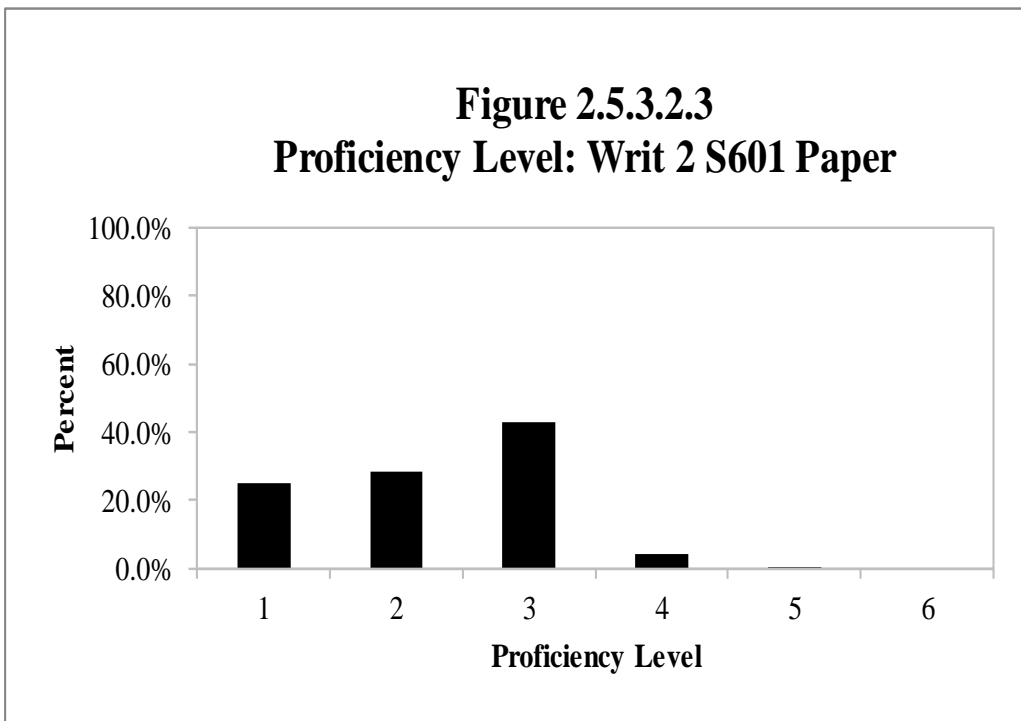
2.5.3.3 Grade 3

Table 2.5.3.3.1

Proficiency Level Distribution: Writ 3 A S601 Paper

Level	Grade 3		Total	
	Count	**Percent**	**Count**	**Percent**
1	4,759	48.17%	4,759	48.17%
2	3,472	35.14%	3,472	35.14%
3	1,640	16.60%	1,640	16.60%
4	9	0.09%	9	0.09%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	9,880	100.00%	9,880	100.00%

Figure 2.5.3.3.1
Proficiency Level: Writ 3A S601 Paper
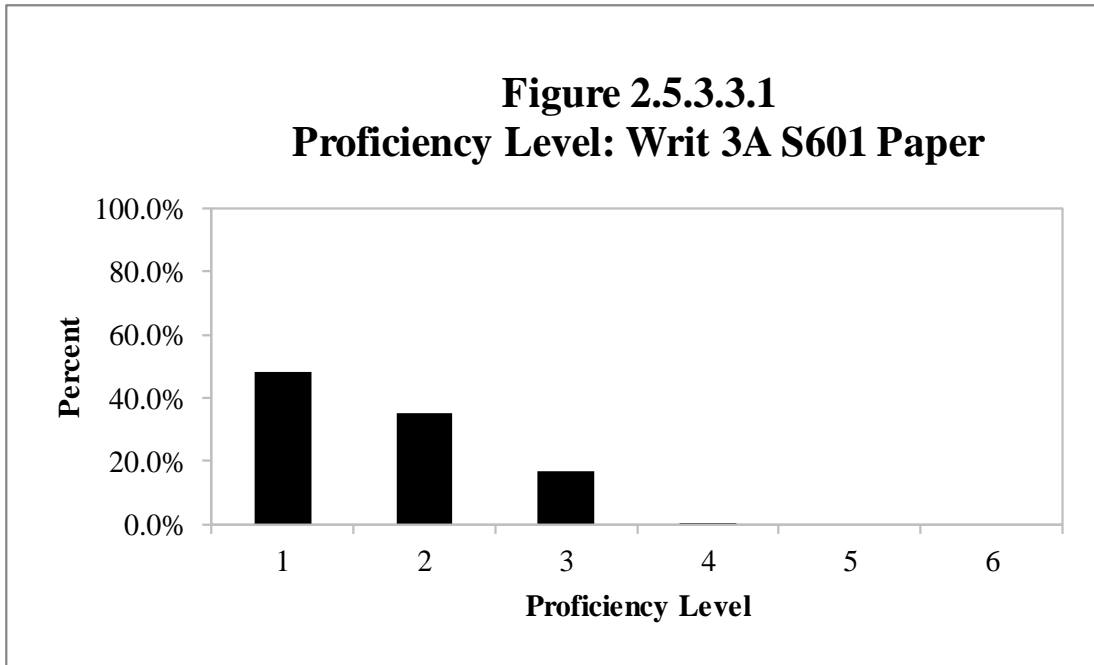
Table 2.5.3.3.2

Proficiency Level Distribution: Writ 3 B/C S601 Paper

Level	Grade 3		Total	
	Count	Percent	Count	Percent
1	1,473	5.85%	1,473	5.85%
2	3,995	15.86%	3,995	15.86%
3	17,510	69.52%	17,510	69.52%
4	2,195	8.71%	2,195	8.71%
5	14	0.06%	14	0.06%
6	0	0.00%	0	0.00%
Total	25,187	100.00%	25,187	100.00%

Figure 2.5.3.3.2
Proficiency Level: Writ 3B/C S601 Paper
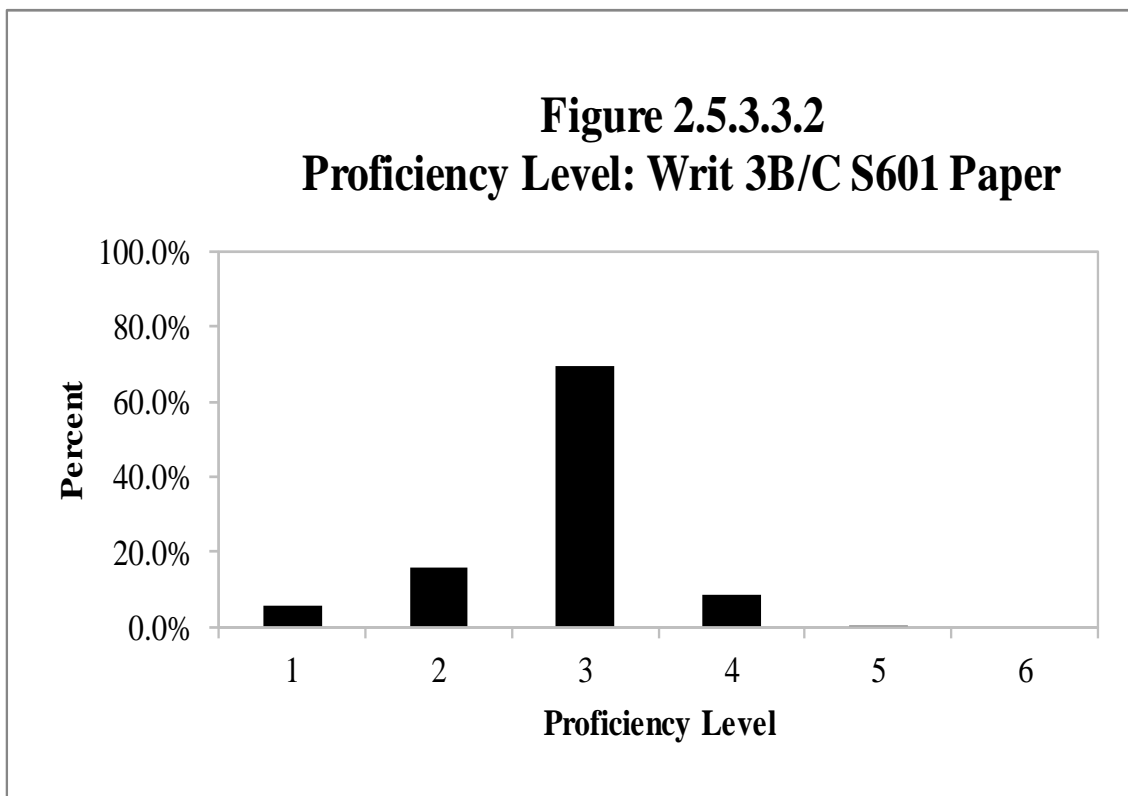
Table 2.5.3.3.3

Proficiency Level Distribution: Writ 3 S601 Paper

Level	Grade 3		Total	
	Count	**Percent**	**Count**	**Percent**
1	6,232	17.77%	6,232	17.77%
2	7,467	21.29%	7,467	21.29%
3	19,150	54.61%	19,150	54.61%
4	2,204	6.29%	2,204	6.29%
5	14	0.04%	14	0.04%
6	0	0.00%	0	0.00%
Total	35,067	100.00%	35,067	100.00%

Figure 2.5.3.3.3
Proficiency Level: Writ 3 S601 Paper
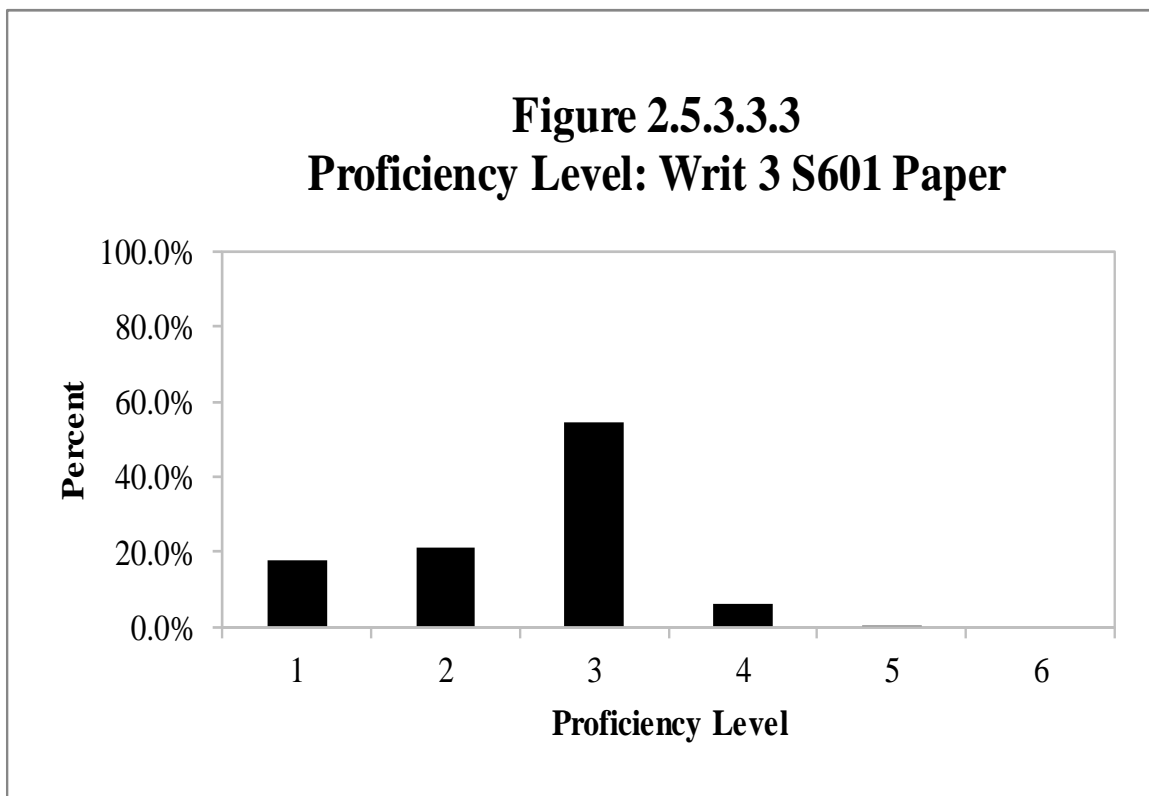
2.5.3.4 Grades 4-5

Table 2.5.3.4.1

Proficiency Level Distribution: Writ 4-5 A S601 Paper

Level	Grade 4		Grade 5		Total	
	Count	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,986	41.54%	2,467	34.21%	5,453	37.87%
2	1,903	26.47%	1,967	27.28%	3,870	26.88%
3	2,273	31.62%	2,741	38.01%	5,014	34.82%
4	26	0.36%	36	0.50%	62	0.43%
5	0	0.00%	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%	0	0.00%
Total	7,188	100.00%	7,211	100.00%	14,399	100.00%

Figure 2.5.3.4.1
Proficiency Level: Writ 4-5A S601 Paper
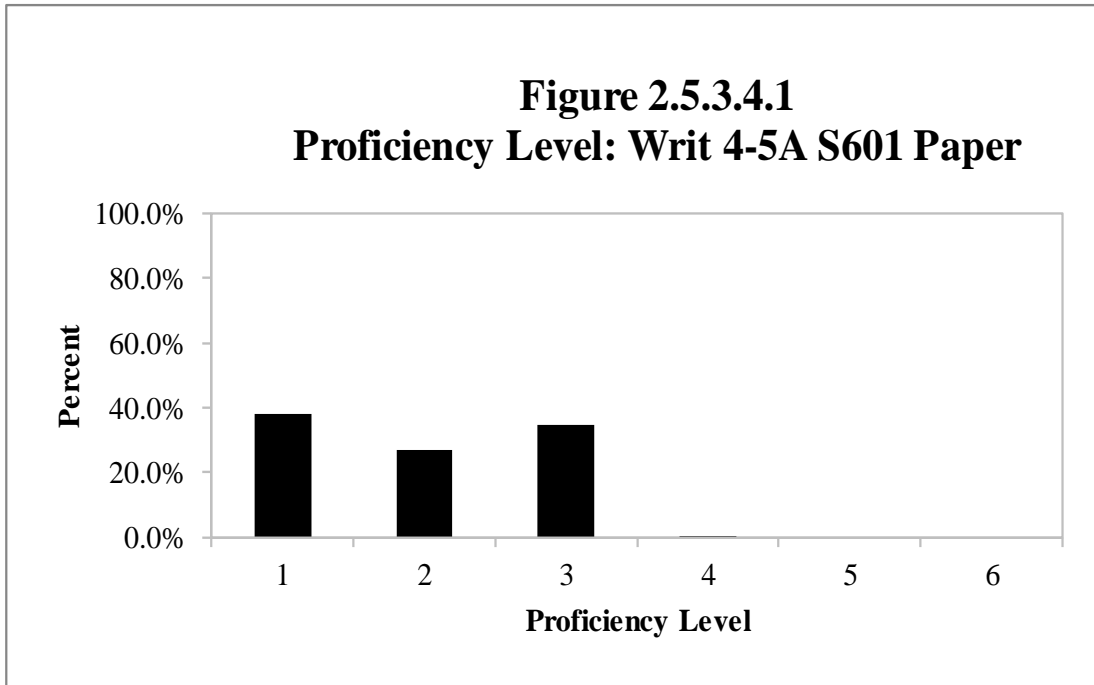
Table 2.5.3.4.2

Proficiency Level Distribution: Writ 4-5 B/C S601 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	239	1.36%	256	1.67%	495	1.51%
2	685	3.91%	593	3.87%	1,278	3.89%
3	11,660	66.53%	8,636	56.36%	20,296	61.79%
4	4,786	27.31%	5,670	37.00%	10,456	31.83%
5	152	0.87%	164	1.07%	316	0.96%
6	3	0.02%	5	0.03%	8	0.02%
Total	17,525	100.00%	15,324	100.00%	32,849	100.00%

Figure 2.5.3.4.2
Proficiency Level: Writ 4-5B/C S601 Paper
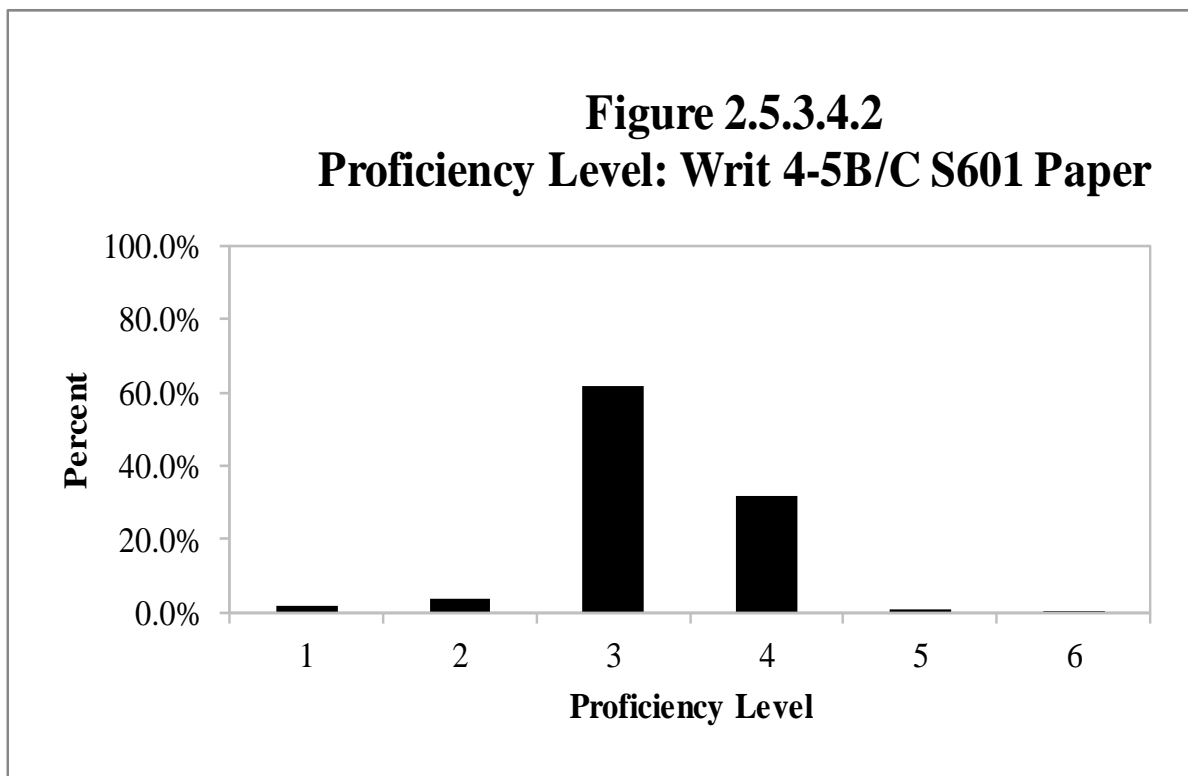
Table 2.5.3.4.3

Proficiency Level Distribution: Writ 4-5 S601 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	3,225	13.05%	2,723	12.08%	5,948	12.59%
2	2,588	10.47%	2,560	11.36%	5,148	10.90%
3	13,933	56.38%	11,377	50.49%	25,310	53.57%
4	4,812	19.47%	5,706	25.32%	10,518	22.26%
5	152	0.62%	164	0.73%	316	0.67%
6	3	0.01%	5	0.02%	8	0.02%
Total	24,713	100.00%	22,535	100.00%	47,248	100.00%

Figure 2.5.3.4.3
Proficiency Level: Writ 4-5 S601 Paper
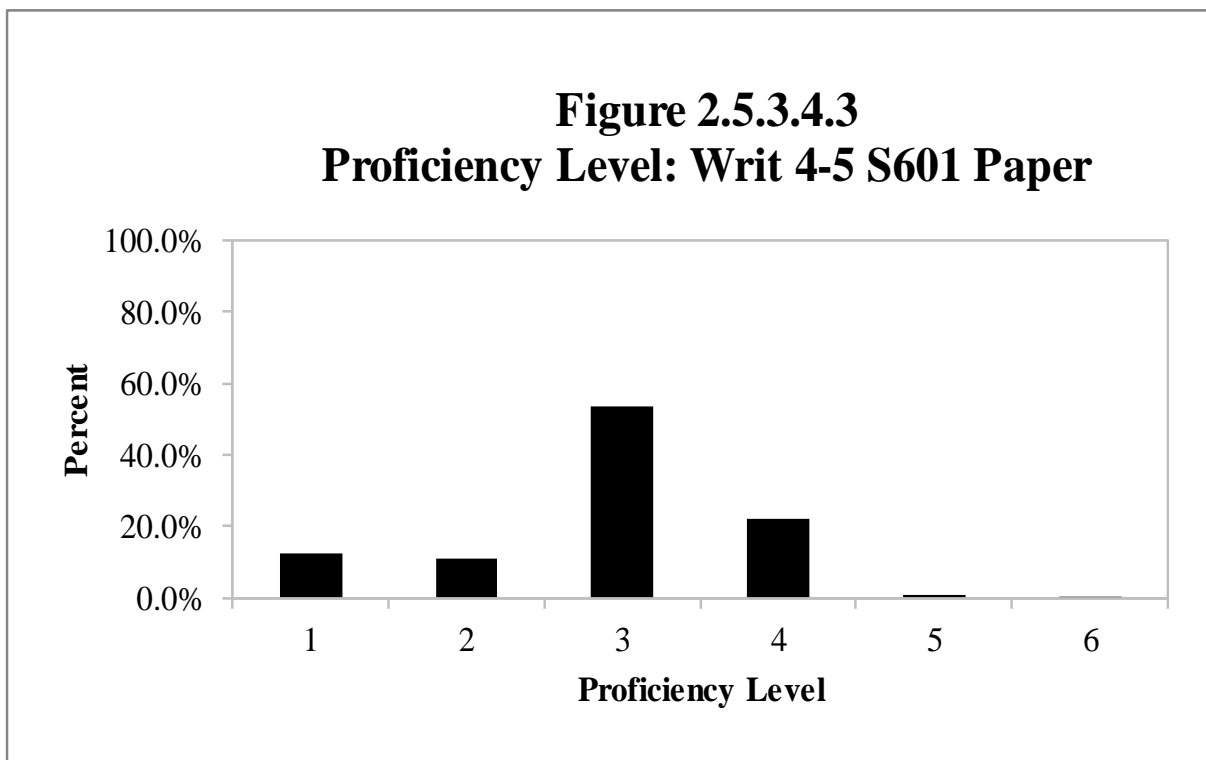
2.5.3.5 Grades 6-8

Table 2.5.3.5.1

Proficiency Level Distribution: Writ 6-8 A S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	3,408	48.80%	3,574	51.31%	3,665	55.31%	10,647	51.75%
2	2,283	32.69%	2,415	34.67%	1,803	27.21%	6,501	31.60%
3	1,277	18.28%	951	13.65%	1,135	17.13%	3,363	16.35%
4	16	0.23%	25	0.36%	23	0.35%	64	0.31%
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	6,984	100.00%	6,965	100.00%	6,626	100.00%	20,575	100.00%

Figure 2.5.3.5.1
Proficiency Level: Writ 6-8A S601 Paper
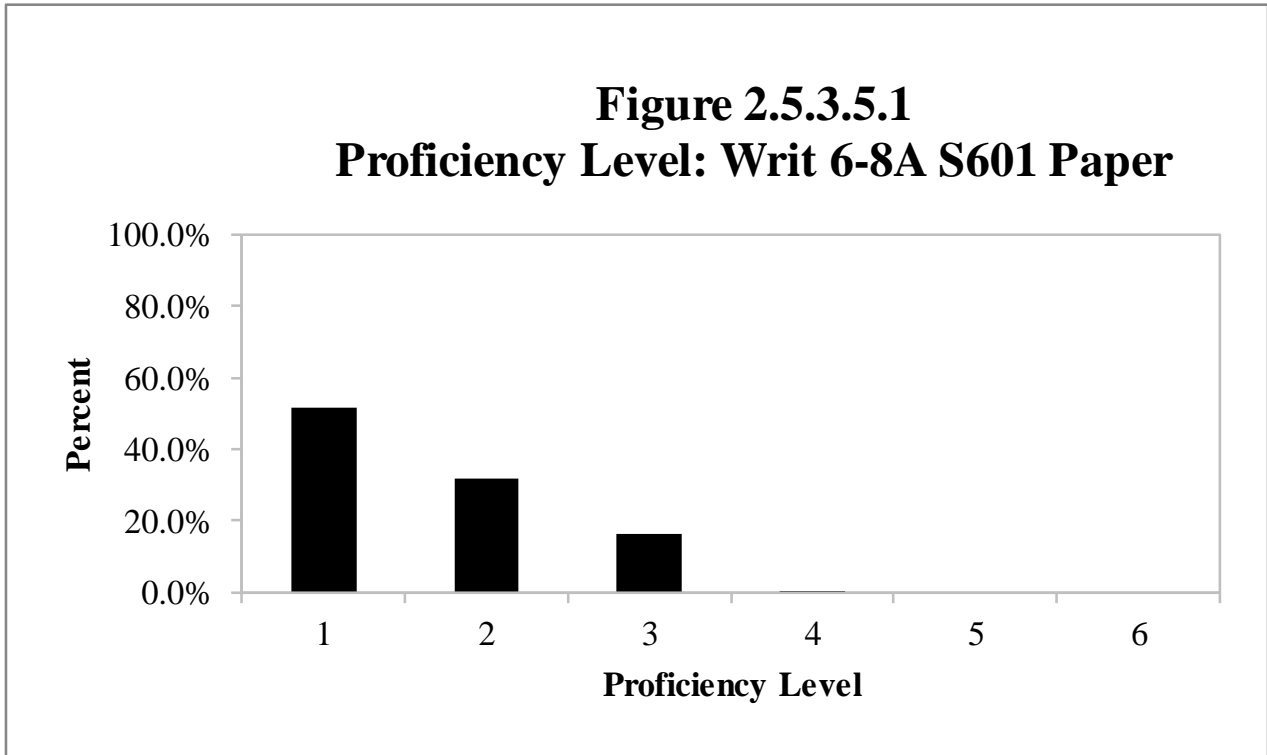
Table 2.5.3.5.2

Proficiency Level Distribution: Writ 6-8 B/C S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	490	3.84%	442	4.26%	557	5.69%	1,489	4.52%
2	1,423	11.16%	1,093	10.53%	682	6.97%	3,198	9.72%
3	8,636	67.72%	7,102	68.43%	6,778	69.27%	22,516	68.41%
4	2,193	17.20%	1,736	16.73%	1,763	18.02%	5,692	17.29%
5	10	0.08%	5	0.05%	5	0.05%	20	0.06%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	12,752	100.00%	10,378	100.00%	9,785	100.00%	32,915	100.00%

Figure 2.5.3.5.2
Proficiency Level: Writ 6-8B/C S601 Paper
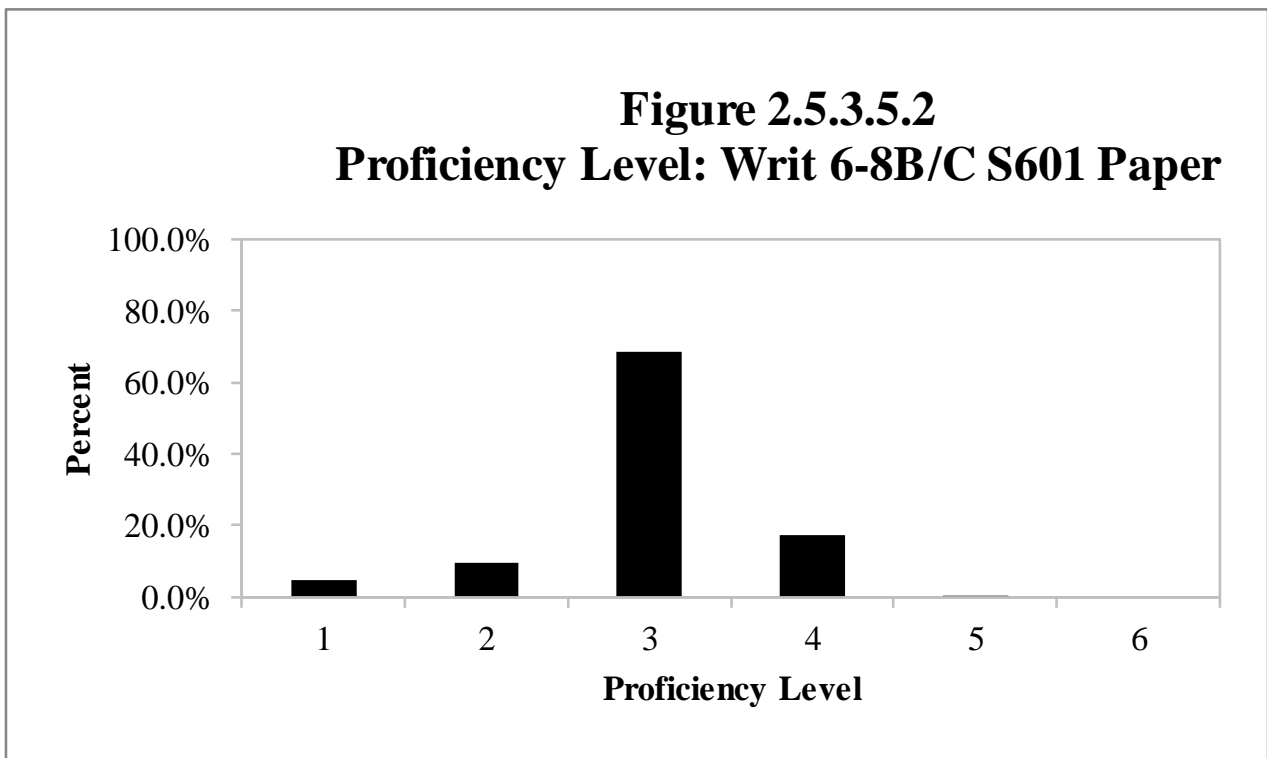
Table 2.5.3.5.3

Proficiency Level Distribution: Writ 6-8 S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	3,898	19.75%	4,016	23.16%	4,222	25.73%	12,136	22.69%
2	3,706	18.78%	3,508	20.23%	2,485	15.14%	9,699	18.13%
3	9,913	50.23%	8,053	46.43%	7,913	48.22%	25,879	48.38%
4	2,209	11.19%	1,761	10.15%	1,786	10.88%	5,756	10.76%
5	10	0.05%	5	0.03%	5	0.03%	20	0.04%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	19,736	100.00%	17,343	100.00%	16,411	100.00%	53,490	100.00%

Figure 2.5.3.5.3
Proficiency Level: Writ 6-8 S601 Paper
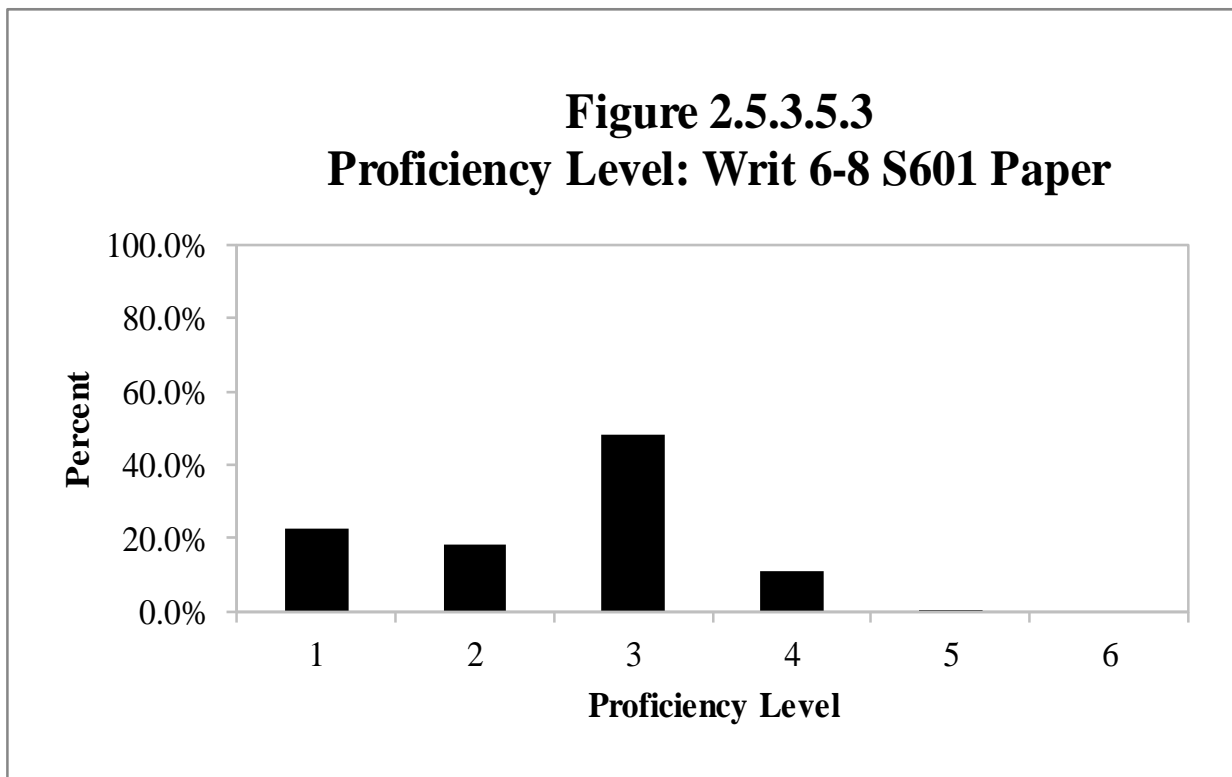
2.5.3.6 Grades 9-12

Table 2.5.3.6.1

Proficiency Level Distribution: Writ 9-12 A S601 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,710	37.65%	2,145	38.99%	1,740	42.06%	1,161	51.53%	7,756	40.63%
2	2,315	32.17%	1,999	36.33%	1,431	34.59%	448	19.88%	6,193	32.44%
3	1,970	27.37%	1,282	23.30%	925	22.36%	617	27.39%	4,794	25.11%
4	202	2.81%	76	1.38%	41	0.99%	27	1.20%	346	1.81%
5	0	0.00%	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%	0	0.00%
Total	7,197	100.00%	5,502	100.00%	4,137	100.00%	2,253	100.00%	19,089	100.00%

Figure 2.5.3.6.1
Proficiency Level: Writ 9-12A S601 Paper
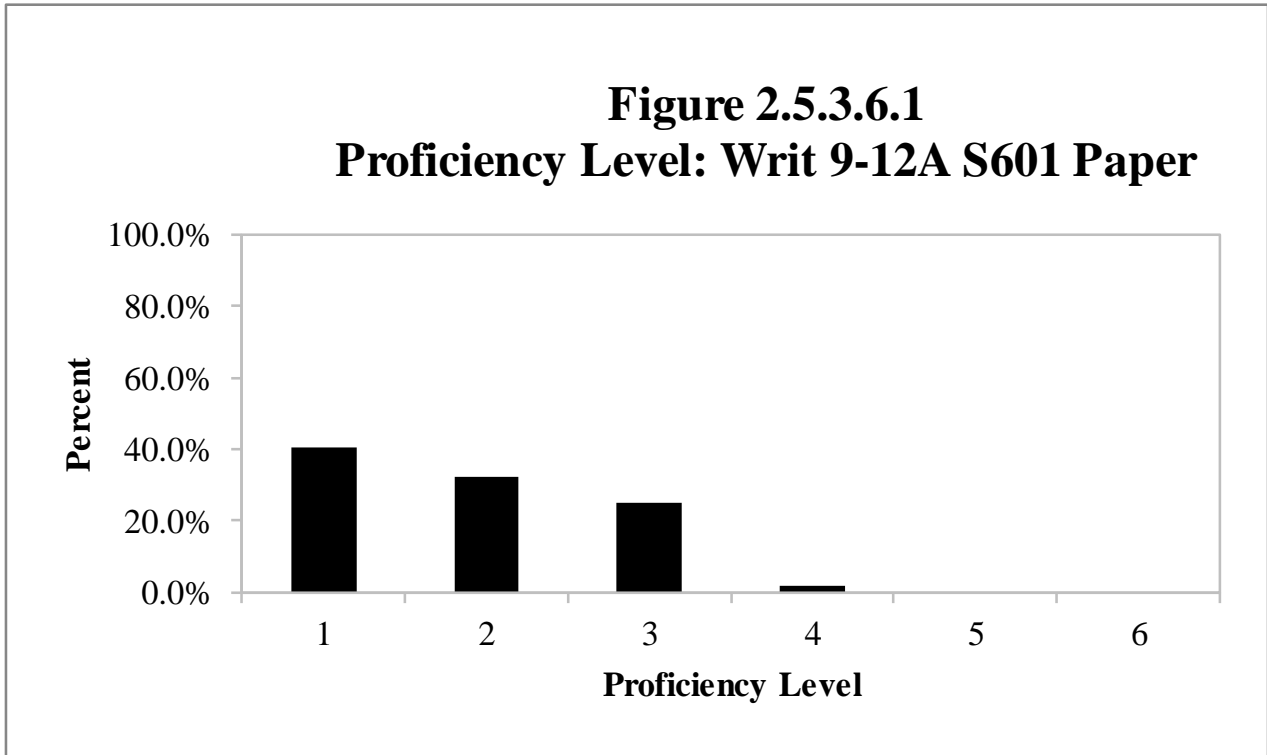
Table 2.5.3.6.2

Proficiency Level Distribution: Writ 9-12 B/C S601 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	281	3.07%	467	5.48%	440	5.66%	605	10.41%	1,793	5.73%
2	581	6.34%	715	8.38%	772	9.93%	713	12.27%	2,781	8.89%
3	5,089	55.53%	4,687	54.96%	4,195	53.96%	3,545	61.02%	17,516	56.00%
4	3,147	34.34%	2,636	30.91%	2,354	30.28%	941	16.20%	9,078	29.03%
5	66	0.72%	23	0.27%	13	0.17%	6	0.10%	108	0.35%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	9,164	100.00%	8,528	100.00%	7,774	100.00%	5,810	100.00%	31,276	100.00%

Figure 2.5.3.6.2
Proficiency Level: Writ 9-12B/C S601 Paper
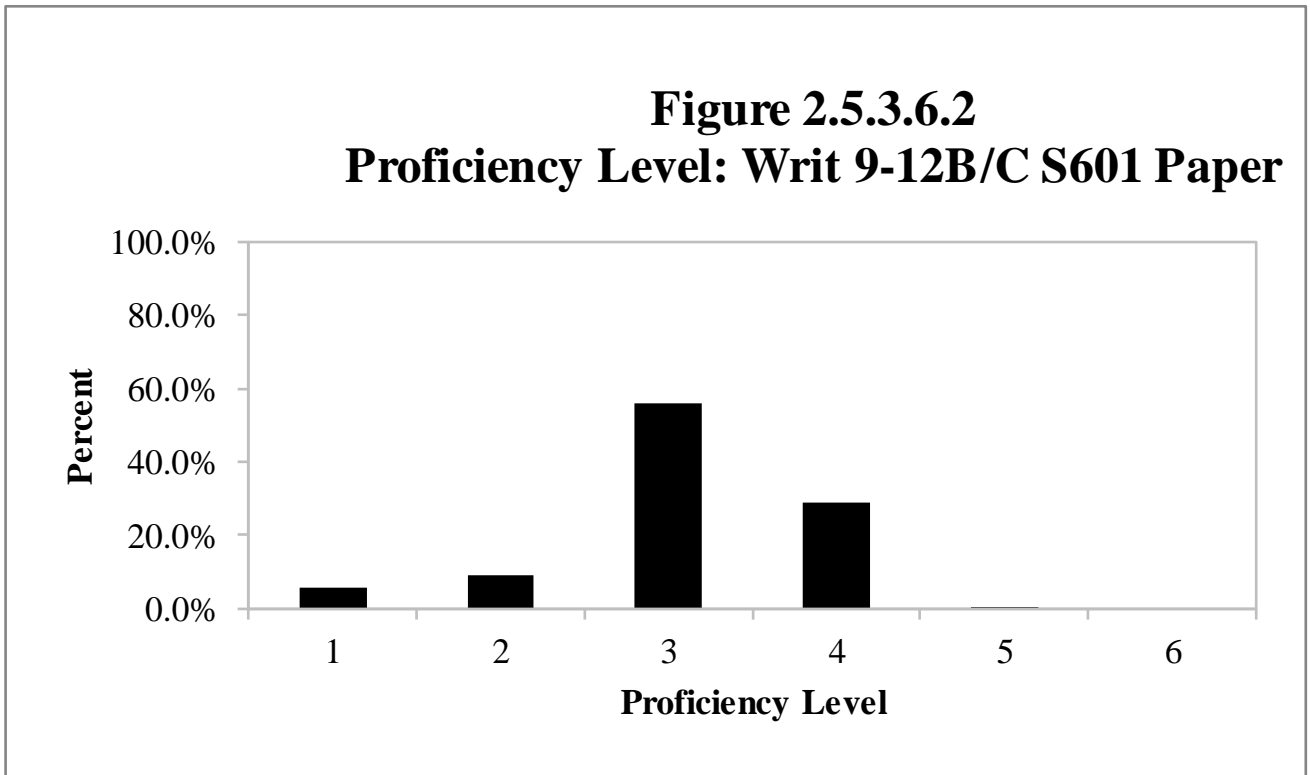
Table 2.5.3.6.3

Proficiency Level Distribution: Writ 9-12 S601 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,991	18.28%	2,612	18.62%	2,180	18.30%	1,766	21.90%	9,549	18.96%
2	2,896	17.70%	2,714	19.34%	2,203	18.50%	1,161	14.40%	8,974	17.82%
3	7,059	43.15%	5,969	42.54%	5,120	42.99%	4,162	51.62%	22,310	44.30%
4	3,349	20.47%	2,712	19.33%	2,395	20.11%	968	12.01%	9,424	18.71%
5	66	0.40%	23	0.16%	13	0.11%	6	0.07%	108	0.21%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	16,361	100.00%	14,030	100.00%	11,911	100.00%	8,063	100.00%	50,365	100.00%

Figure 2.5.3.6.3
Proficiency Level: Writ 9-12 S601 Paper
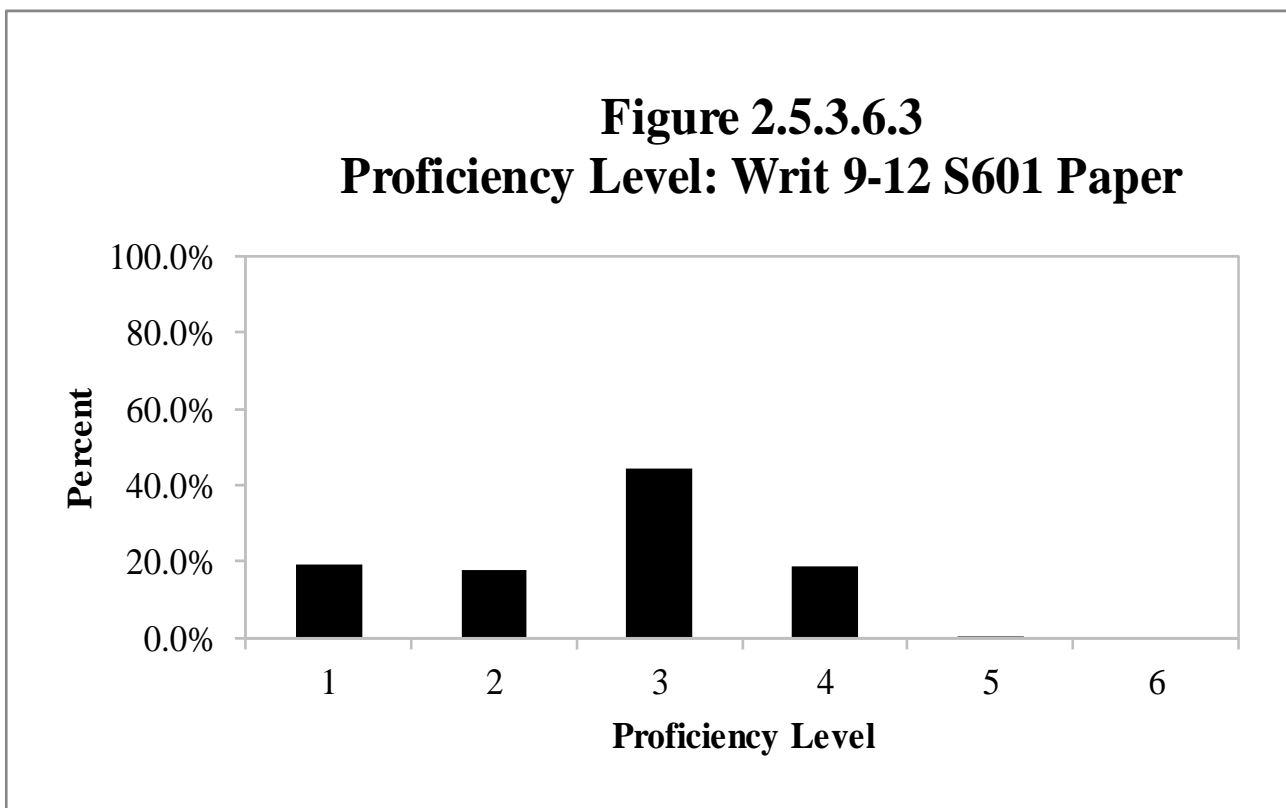
2.5.4 Speaking

2.5.4.0 Kindergarten

Table 2.5.4.0

Proficiency Level Distribution: Spek K S601 Paper

Level	Grade K		Total	
	Count	**Percent**	**Count**	**Percent**
1	83,152	32.38%	83,152	32.38%
2	52,953	20.62%	52,953	20.62%
3	17,437	6.79%	17,437	6.79%
4	18,874	7.35%	18,874	7.35%
5	24,726	9.63%	24,726	9.63%
6	59,686	23.24%	59,686	23.24%
Total	256,828	100.00%	256,828	100.00%

**Figure 2.5.4.0
Proficiency Level: Spek K S601 Paper**
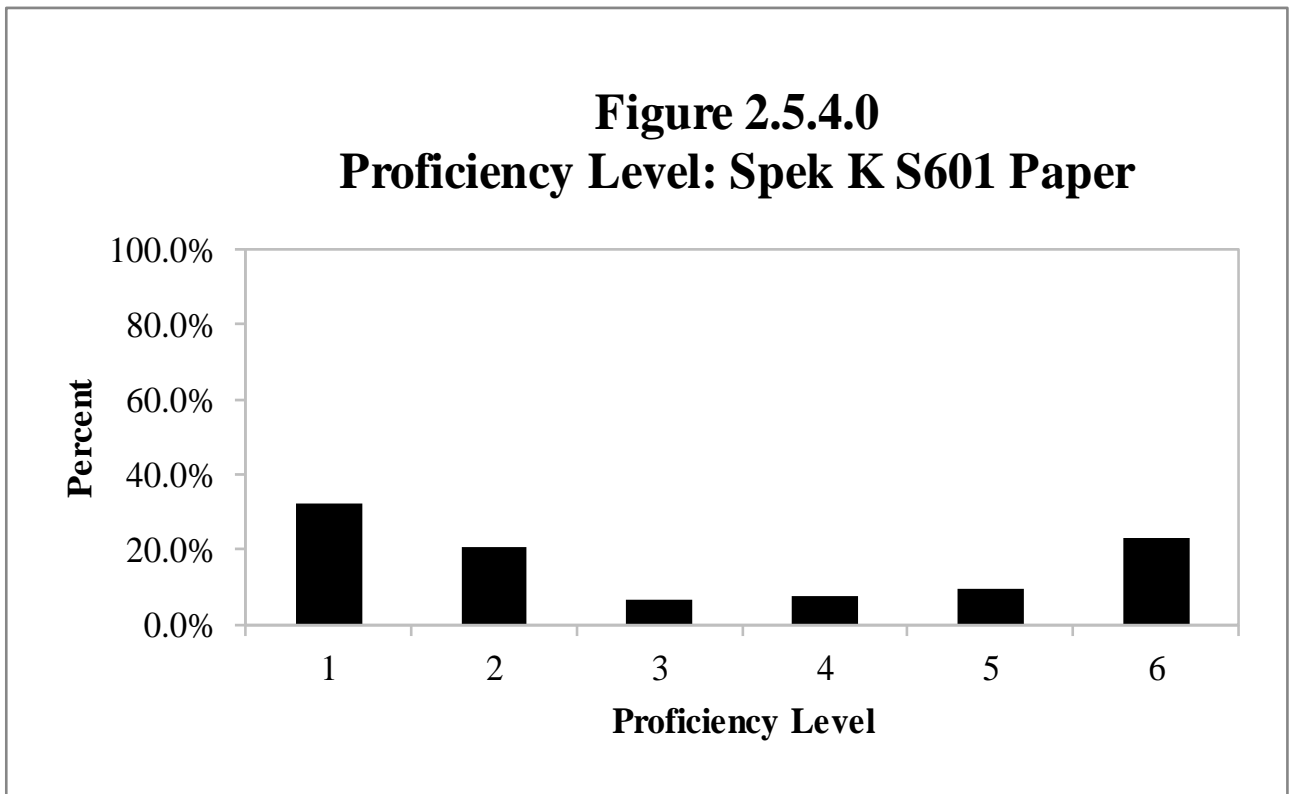
2.5.4.1 Grade 1

Table 2.5.4.1.1

Proficiency Level Distribution: Spek 1 A S601 Paper

Level	Grade 1 Count	Grade 1 Percent	Total Count	Total Percent
1	8,382	35.87%	8,382	35.87%
2	7,277	31.14%	7,277	31.14%
3	4,362	18.67%	4,362	18.67%
4	2,510	10.74%	2,510	10.74%
5	836	3.58%	836	3.58%
6	0	0.00%	0	0.00%
Total	23,367	100.00%	23,367	100.00%

Figure 2.5.4.1.1
Proficiency Level: Spek 1A S601 Paper
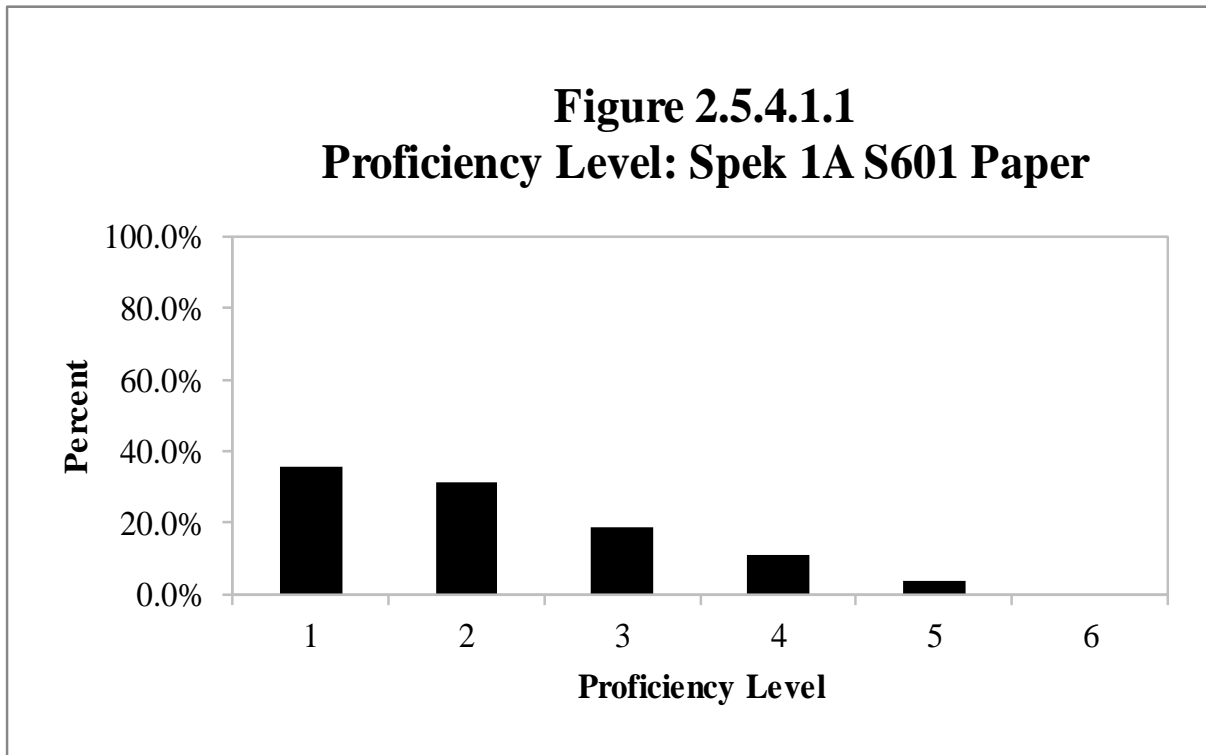
Table 2.5.4.1.2

Proficiency Level Distribution: Spek 1 B/C S601 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	571	3.65%	571	3.65%
2	3,350	21.39%	3,350	21.39%
3	5,212	33.28%	5,212	33.28%
4	4,734	30.22%	4,734	30.22%
5	1,355	8.65%	1,355	8.65%
6	441	2.82%	441	2.82%
Total	15,663	100.00%	15,663	100.00%

**Figure 2.5.4.1.2
Proficiency Level: Spek 1B/C S601 Paper**
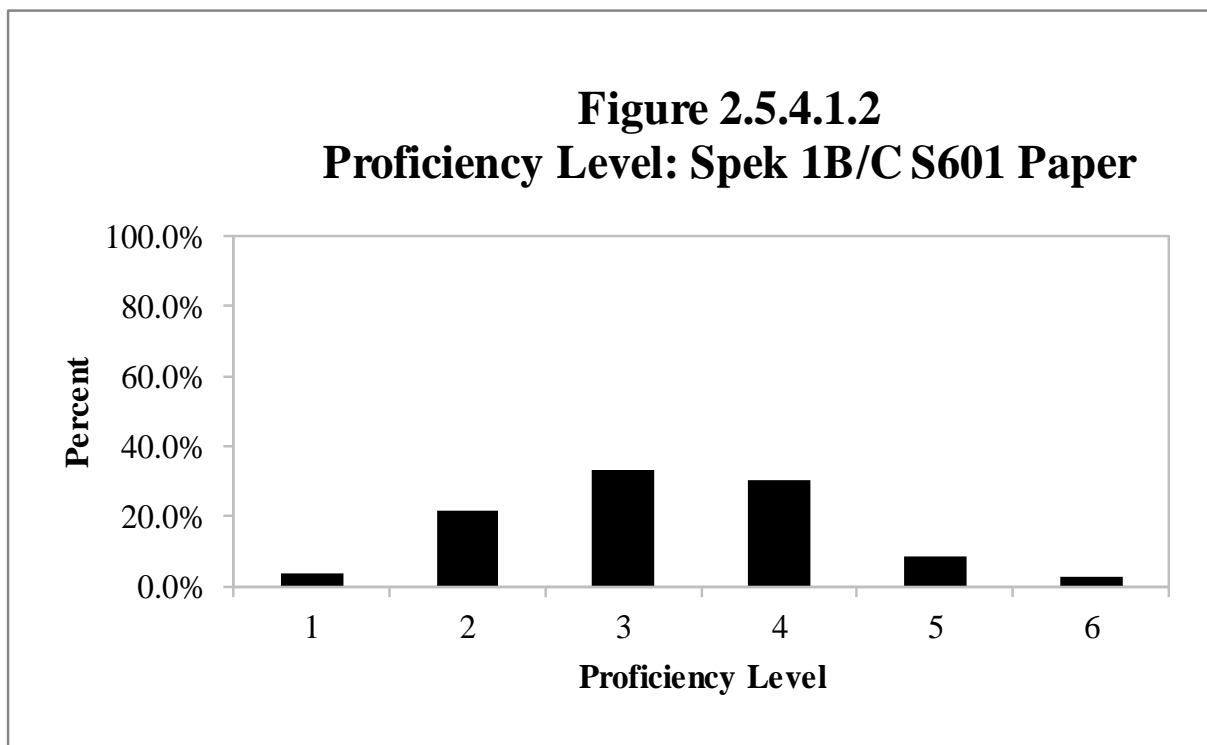
Table 2.5.4.1.3

Proficiency Level Distribution: Spek 1 S601 Paper

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	8,953	22.94%	8,953	22.94%
2	10,627	27.23%	10,627	27.23%
3	9,574	24.53%	9,574	24.53%
4	7,244	18.56%	7,244	18.56%
5	2,191	5.61%	2,191	5.61%
6	441	1.13%	441	1.13%
Total	39,030	100.00%	39,030	100.00%

Figure 2.5.4.1.3
Proficiency Level: Spek 1 S601 Paper
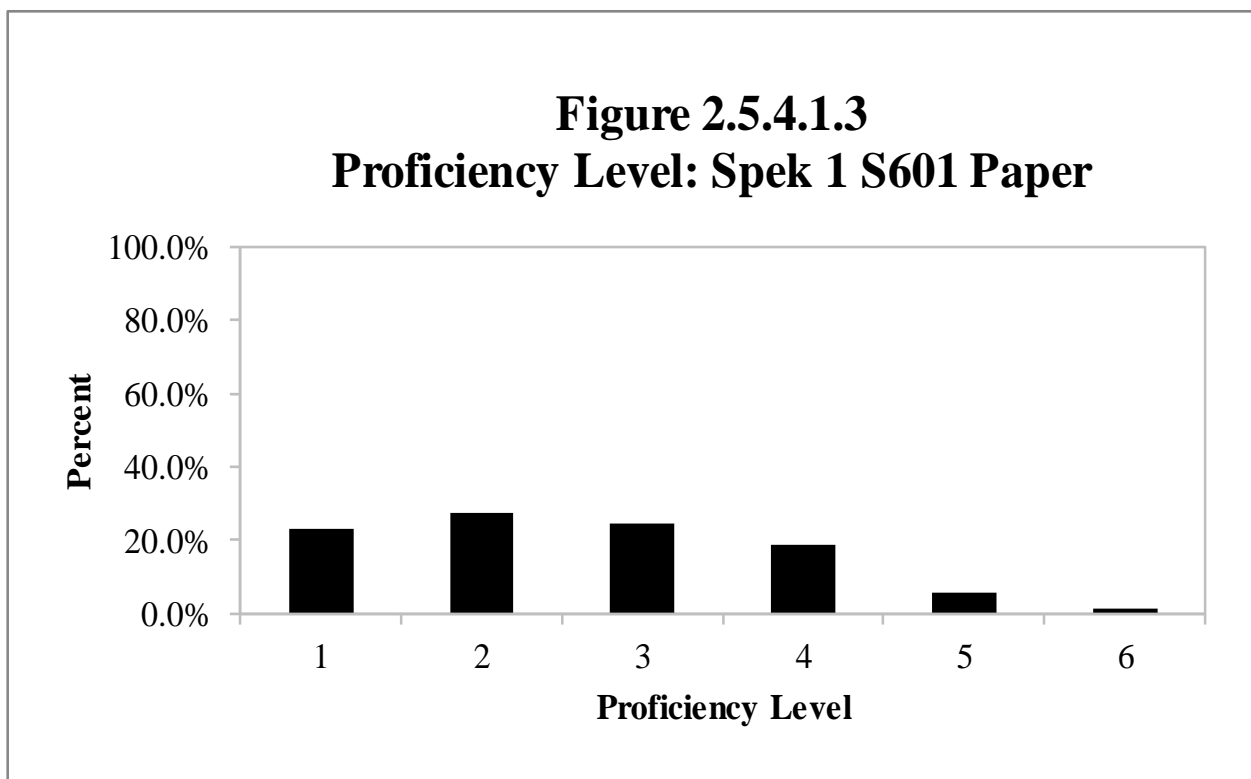
2.5.4.2 Grade 2

Table 2.5.4.2.1

Proficiency Level Distribution: Spek 2 A S601 Paper

Level	Grade 2		Total	
	Count	**Percent**	**Count**	**Percent**
1	5,765	47.41%	5,765	47.41%
2	2,578	21.20%	2,578	21.20%
3	2,890	23.76%	2,890	23.76%
4	660	5.43%	660	5.43%
5	268	2.20%	268	2.20%
6	0	0.00%	0	0.00%
Total	12,161	100.00%	12,161	100.00%

Figure 2.5.4.2.1
Proficiency Level: Spek 2A S601 Paper
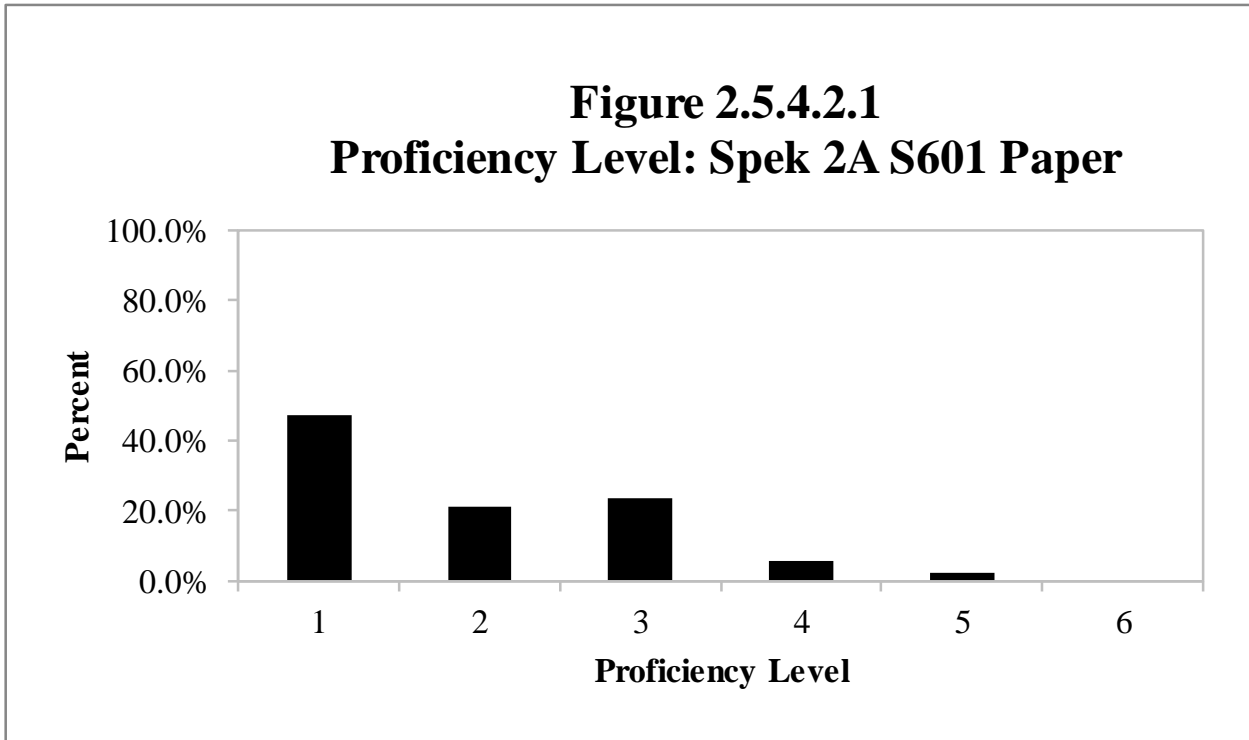
Table 2.5.4.2.2

Proficiency Level Distribution: Spek 2 B/C S601 Paper

Level	Grade 2		Total	
	Count	**Percent**	**Count**	**Percent**
1	1,328	5.84%	1,328	5.84%
2	4,611	20.29%	4,611	20.29%
3	9,314	40.98%	9,314	40.98%
4	5,140	22.61%	5,140	22.61%
5	1,552	6.83%	1,552	6.83%
6	784	3.45%	784	3.45%
Total	22,729	100.00%	22,729	100.00%

**Figure 2.5.4.2.2
Proficiency Level: Spek 2B/C S601 Paper**
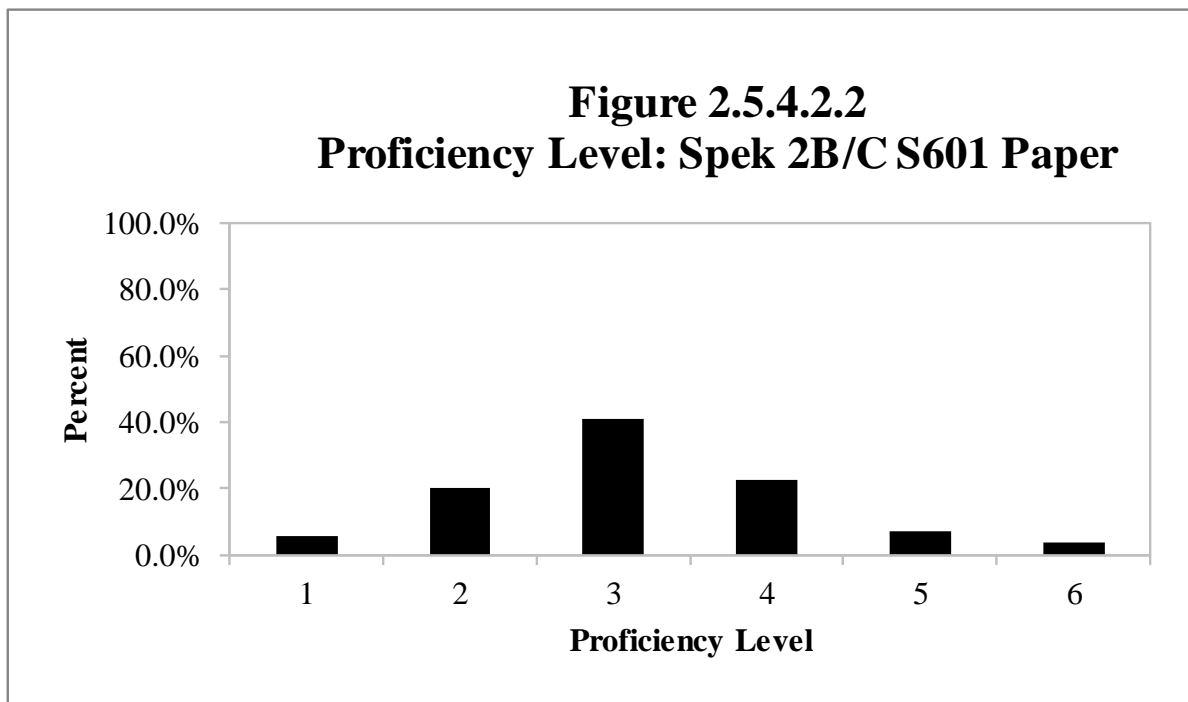
Table 2.5.4.2.3

Proficiency Level Distribution: Spek 2 S601 Paper

Level	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	7,093	20.33%	7,093	20.33%
2	7,189	20.60%	7,189	20.60%
3	12,204	34.98%	12,204	34.98%
4	5,800	16.62%	5,800	16.62%
5	1,820	5.22%	1,820	5.22%
6	784	2.25%	784	2.25%
Total	34,890	100.00%	34,890	100.00%

Figure 2.5.4.2.3
Proficiency Level: Spek 2 S601 Paper
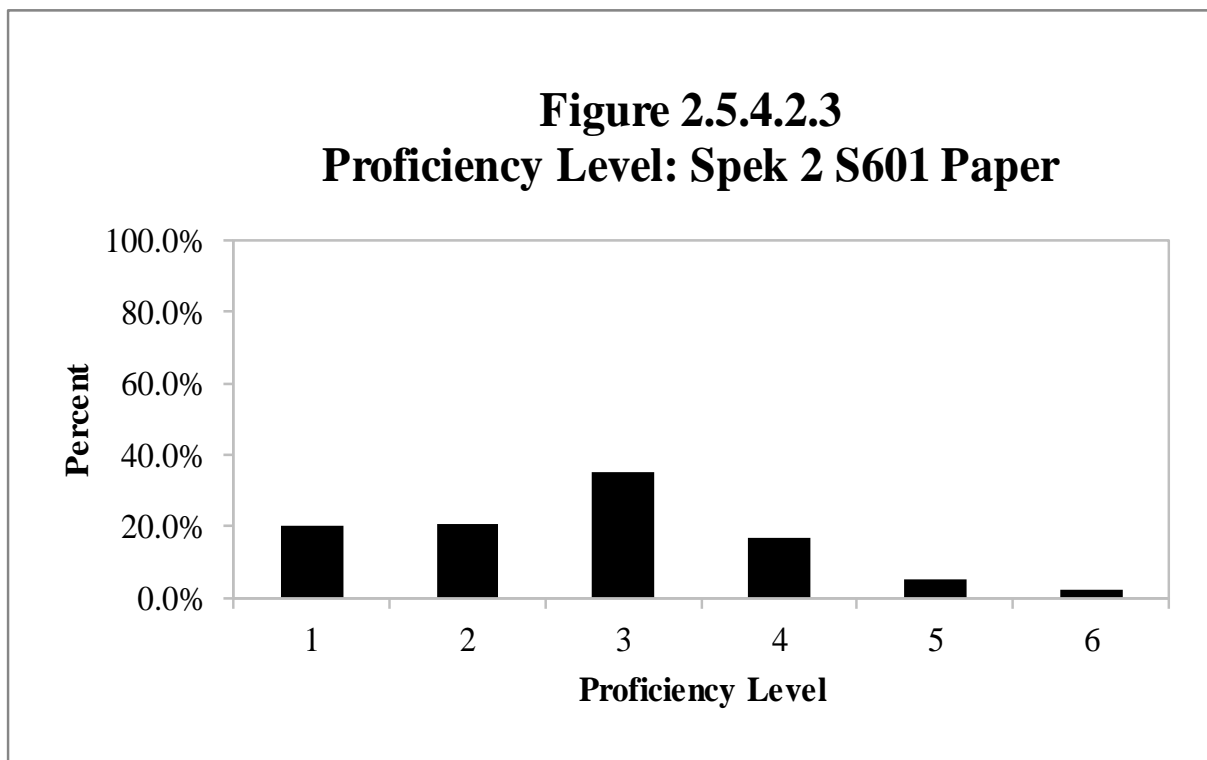
2.5.4.3 Grade 3

Table 2.5.4.3.1

Proficiency Level Distribution: Spek 3 A S601 Paper

Level	Grade 3		Total	
	Count	**Percent**	**Count**	**Percent**
1	5,672	58.04%	5,672	58.04%
2	1,962	20.08%	1,962	20.08%
3	1,402	14.35%	1,402	14.35%
4	736	7.53%	736	7.53%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	9,772	100.00%	9,772	100.00%

Figure 2.5.4.3.1
Proficiency Level: Spek 3A S601 Paper
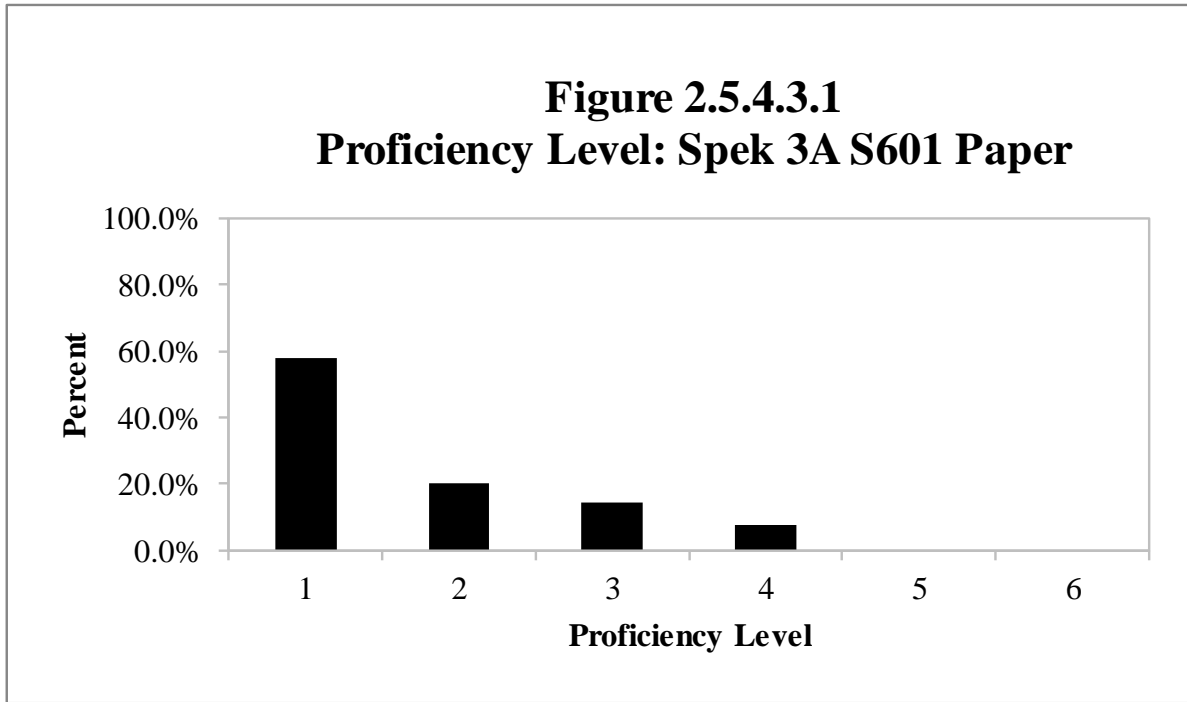
Table 2.5.4.3.2

Proficiency Level Distribution: Spek 3 B/C S601 Paper

Level	Grade 3		Total	
	Count	**Percent**	**Count**	**Percent**
1	1,382	5.54%	1,382	5.54%
2	5,099	20.44%	5,099	20.44%
3	10,676	42.80%	10,676	42.80%
4	5,475	21.95%	5,475	21.95%
5	1,097	4.40%	1,097	4.40%
6	1,214	4.87%	1,214	4.87%
Total	24,943	100.00%	24,943	100.00%

Figure 2.5.4.3.2
Proficiency Level: Spek 3B/C S601 Paper
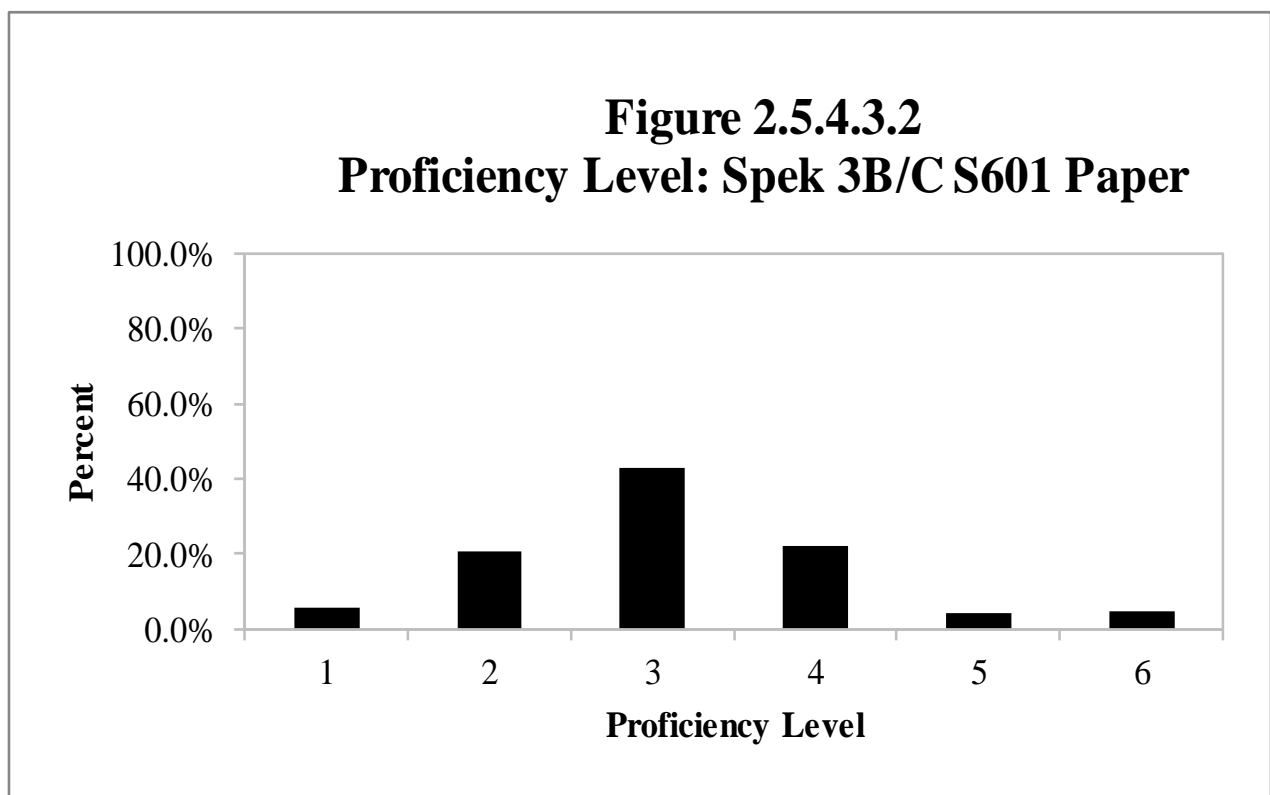
Table 2.5.4.3.3

Proficiency Level Distribution: Spek 3 S601 Paper

Level	Grade 3		Total	
	Count	**Percent**	**Count**	**Percent**
1	7,054	20.32%	7,054	20.32%
2	7,061	20.34%	7,061	20.34%
3	12,078	34.79%	12,078	34.79%
4	6,211	17.89%	6,211	17.89%
5	1,097	3.16%	1,097	3.16%
6	1,214	3.50%	1,214	3.50%
Total	34,715	100.00%	34,715	100.00%

Figure 2.5.4.3.3
Proficiency Level: Spek 3 S601 Paper
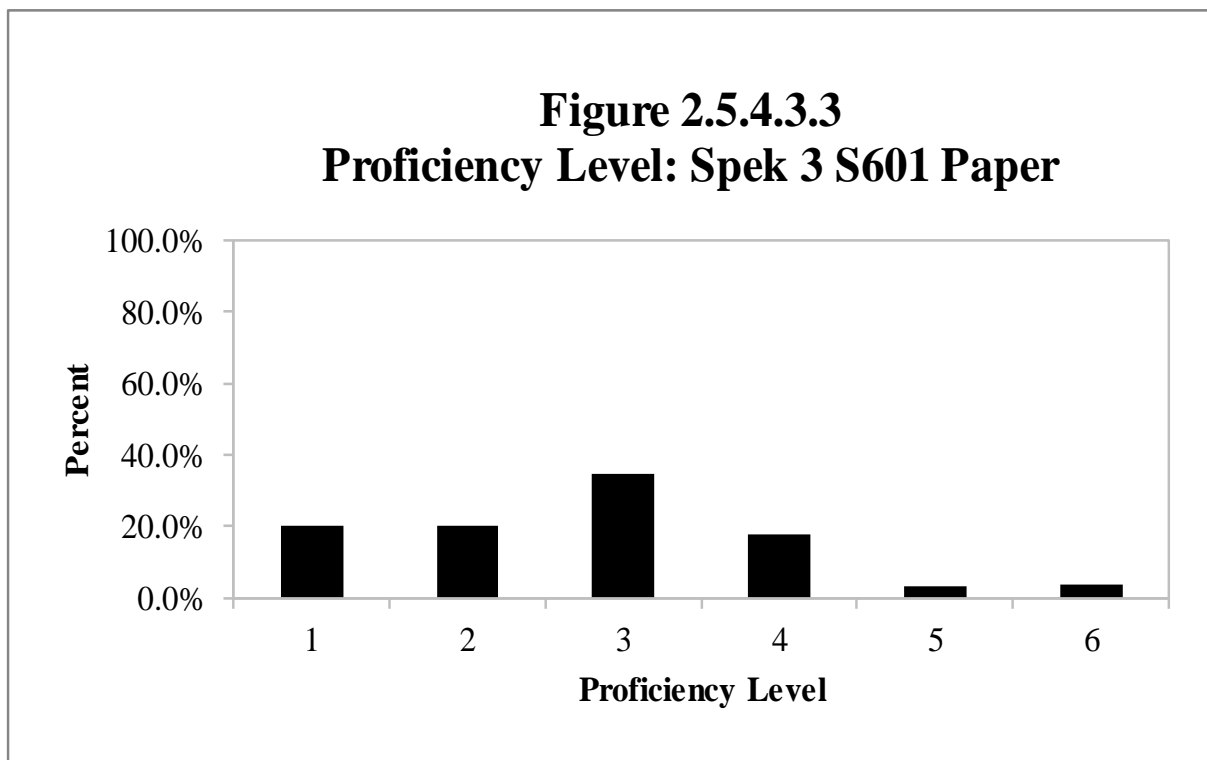
2.5.4.4 Grades 4-5

Table 2.5.4.4.1

Proficiency Level Distribution: Spek 4-5 A S601 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	4,367	61.27%	4,812	67.12%	9,179	64.20%
2	1,544	21.66%	1,117	15.58%	2,661	18.61%
3	680	9.54%	674	9.40%	1,354	9.47%
4	398	5.58%	498	6.95%	896	6.27%
5	139	1.95%	68	0.95%	207	1.45%
6	0	0.00%	0	0.00%	0	0.00%
Total	7,128	100.00%	7,169	100.00%	14,297	100.00%

Figure 2.5.4.4.1
Proficiency Level: Spek 4-5A S601 Paper
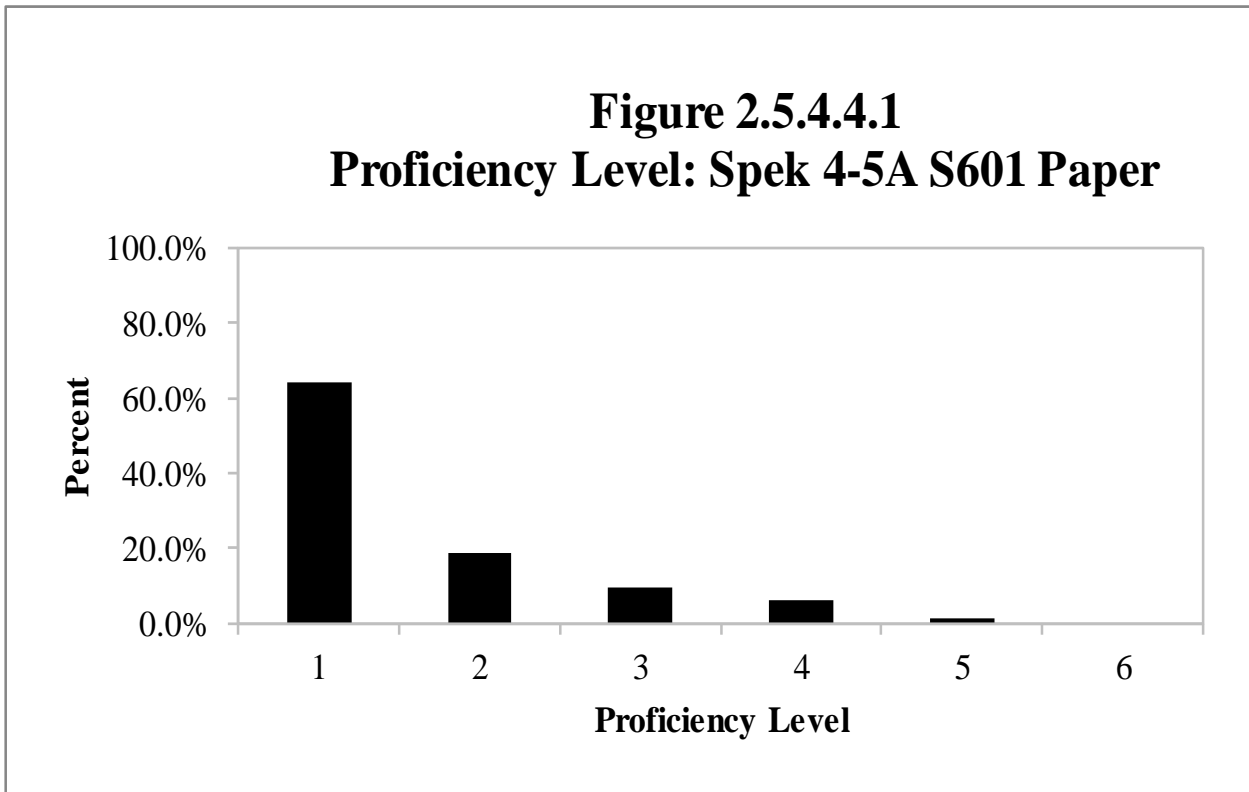
Table 2.5.4.4.2

Proficiency Level Distribution: Spek 4-5 B/C S601 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	372	2.13%	495	3.24%	867	2.65%
2	1,505	8.64%	1,551	10.17%	3,056	9.35%
3	4,533	26.01%	4,166	27.30%	8,699	26.61%
4	6,898	39.58%	5,759	37.74%	12,657	38.72%
5	2,619	15.03%	2,325	15.24%	4,944	15.13%
6	1,502	8.62%	962	6.30%	2,464	7.54%
Total	17,429	100.00%	15,258	100.00%	32,687	100.00%

Figure 2.5.4.4.2
Proficiency Level: Spek 4-5B/C S601 Paper
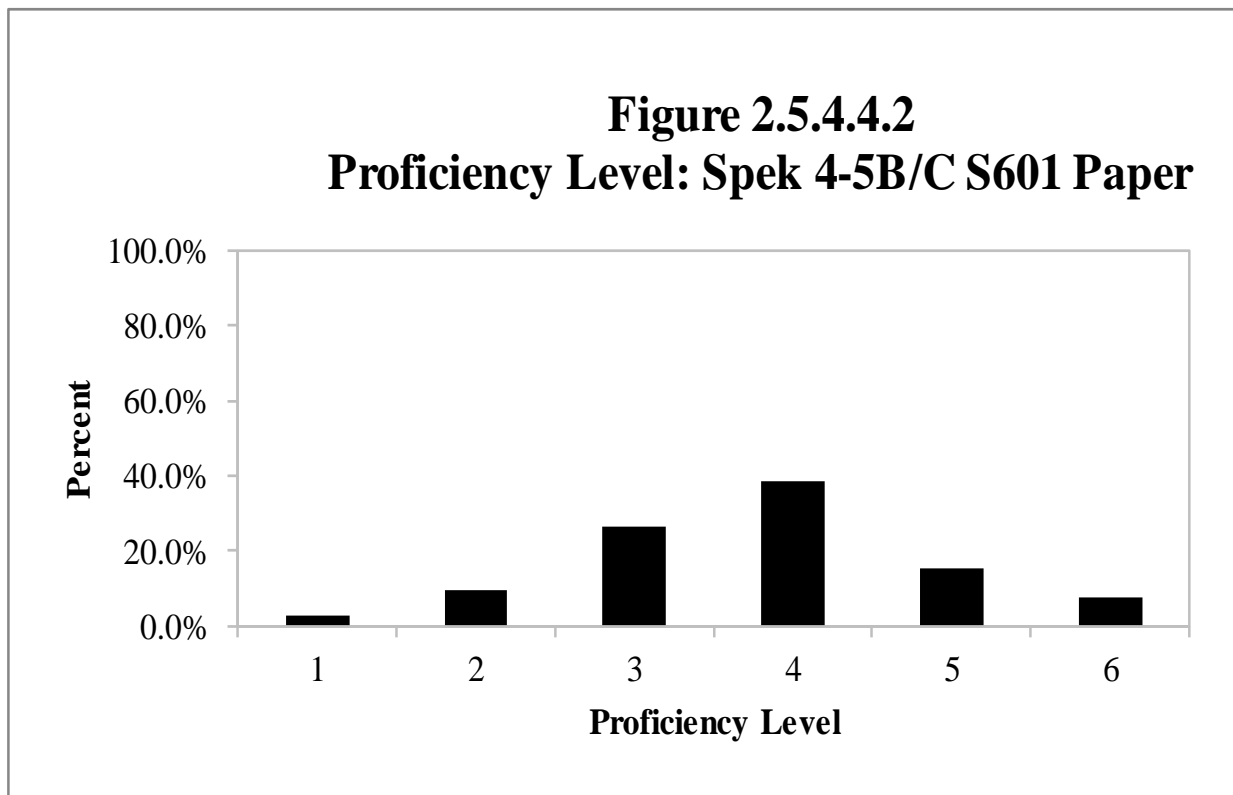
Table 2.5.4.4.3

Proficiency Level Distribution: Spek 4-5 S601 Paper

Level	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	4,739	19.30%	5,307	23.66%	10,046	21.38%
2	3,049	12.42%	2,668	11.90%	5,717	12.17%
3	5,213	21.23%	4,840	21.58%	10,053	21.40%
4	7,296	29.71%	6,257	27.90%	13,553	28.85%
5	2,758	11.23%	2,393	10.67%	5,151	10.96%
6	1,502	6.12%	962	4.29%	2,464	5.24%
Total	24,557	100.00%	22,427	100.00%	46,984	100.00%

Figure 2.5.4.4.3
Proficiency Level: Spek 4-5 S601 Paper
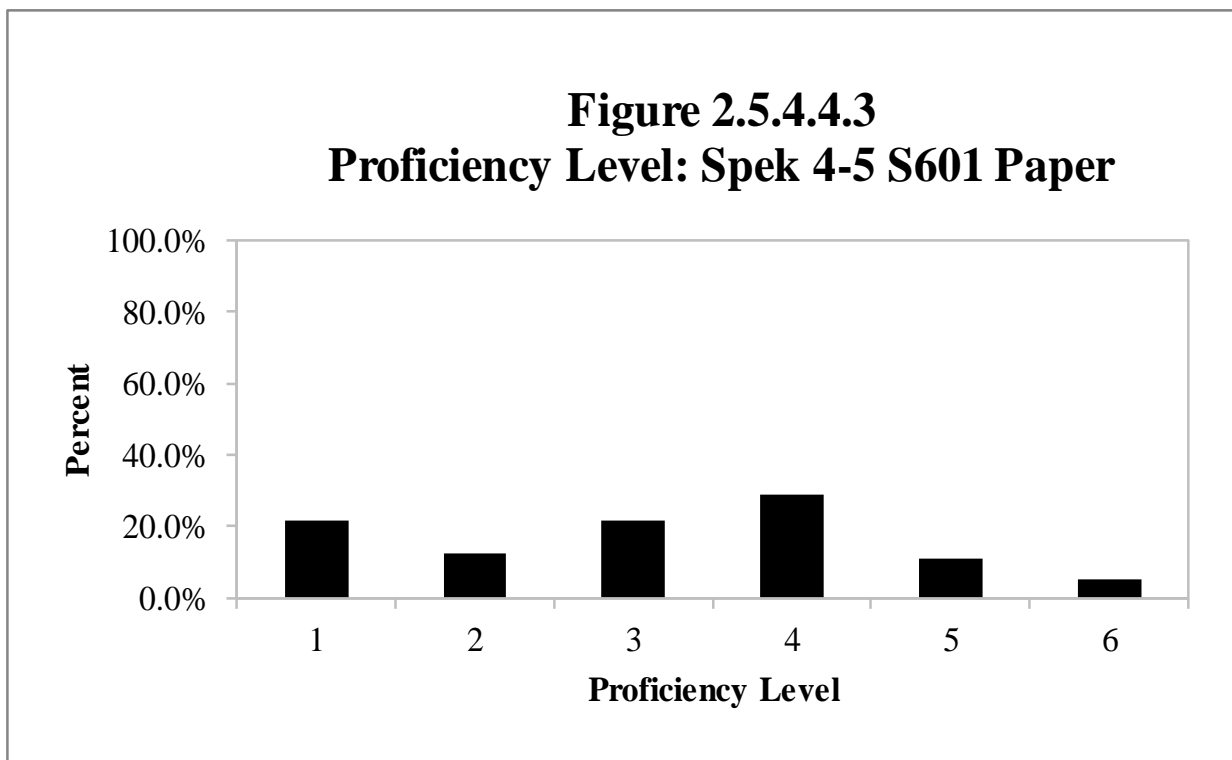
2.5.4.5 Grades 6-8

Table 2.5.4.5.1

Proficiency Level Distribution: Spek 6-8 A S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	4,167	60.14%	4,122	59.85%	4,375	66.61%	12,664	62.13%
2	1,180	17.03%	1,113	16.16%	493	7.51%	2,786	13.67%
3	856	12.35%	872	12.66%	1,167	17.77%	2,895	14.20%
4	475	6.86%	630	9.15%	375	5.71%	1,480	7.26%
5	179	2.58%	78	1.13%	158	2.41%	415	2.04%
6	72	1.04%	72	1.05%	0	0.00%	144	0.71%
Total	6,929	100.00%	6,887	100.00%	6,568	100.00%	20,384	100.00%

Figure 2.5.4.5.1
Proficiency Level: Spek 6-8A S601 Paper
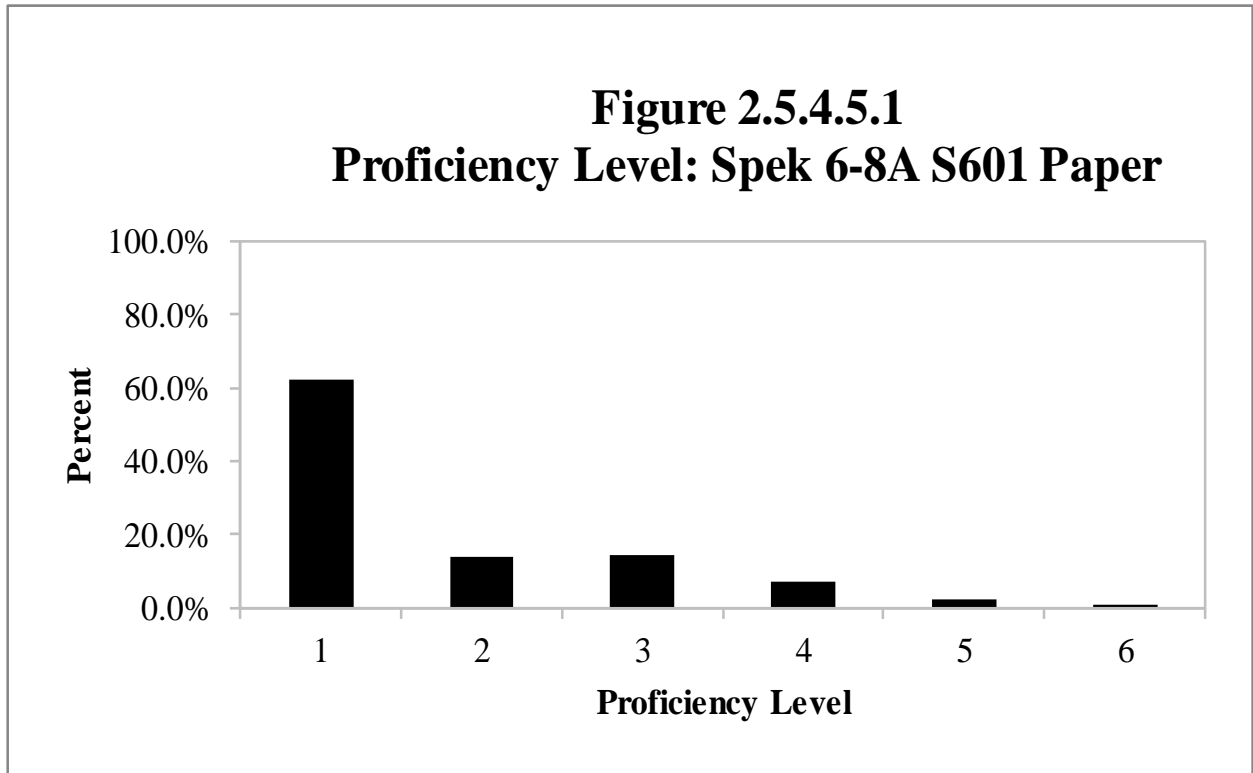
Table 2.5.4.5.2

Proficiency Level Distribution: Spek 6-8 B/C S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	389	3.07%	583	5.65%	631	6.50%	1,603	4.91%
2	2,091	16.53%	1,296	12.56%	1,250	12.87%	4,637	14.19%
3	3,456	27.32%	2,608	25.28%	2,519	25.94%	8,583	26.27%
4	4,228	33.42%	3,882	37.63%	3,137	32.30%	11,247	34.42%
5	1,533	12.12%	959	9.30%	1,423	14.65%	3,915	11.98%
6	954	7.54%	987	9.57%	751	7.73%	2,692	8.24%
Total	12,651	100.00%	10,315	100.00%	9,711	100.00%	32,677	100.00%

Figure 2.5.4.5.2
Proficiency Level: Spek 6-8B/C S601 Paper
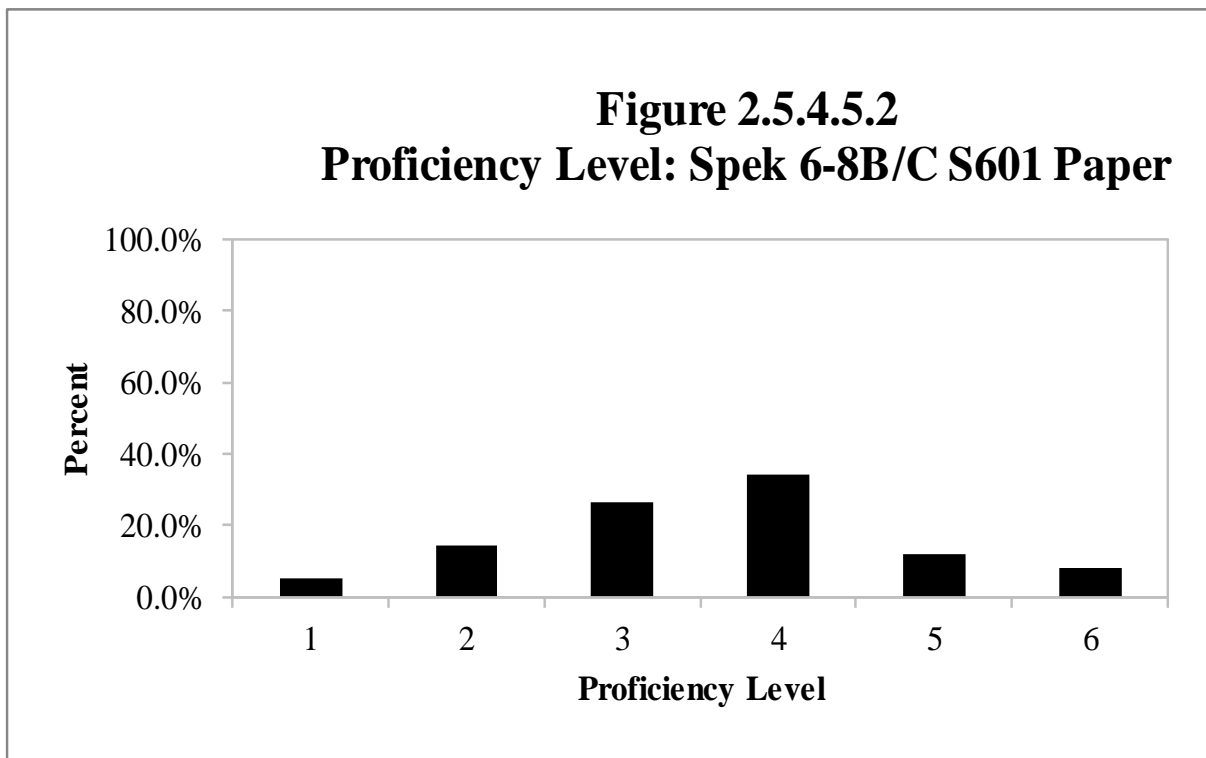
Table 2.5.4.5.3

Proficiency Level Distribution: Spek 6-8 S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	4,556	23.27%	4,705	27.35%	5,006	30.75%	14,267	26.89%
2	3,271	16.71%	2,409	14.00%	1,743	10.71%	7,423	13.99%
3	4,312	22.02%	3,480	20.23%	3,686	22.64%	11,478	21.63%
4	4,703	24.02%	4,512	26.23%	3,512	21.57%	12,727	23.99%
5	1,712	8.74%	1,037	6.03%	1,581	9.71%	4,330	8.16%
6	1,026	5.24%	1,059	6.16%	751	4.61%	2,836	5.34%
Total	19,580	100.00%	17,202	100.00%	16,279	100.00%	53,061	100.00%

Figure 2.5.4.5.3
Proficiency Level: Spek 6-8 S601 Paper
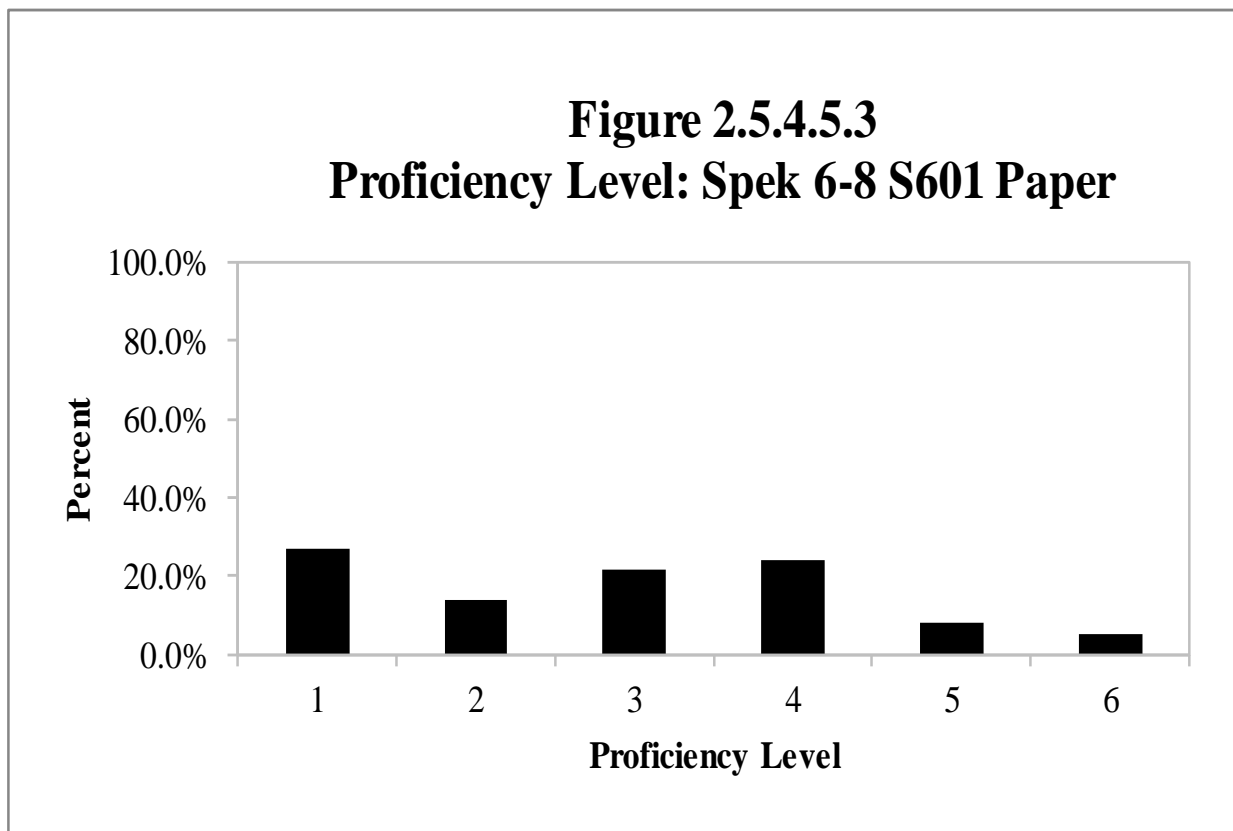
2.5.4.6 Grades 9-12

Table 2.5.4.6.1

Proficiency Level Distribution: Spek 9-12 A S601 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	5,238	73.66%	3,835	70.30%	2,587	63.22%	1,231	55.00%	12,891	68.22%
2	606	8.52%	533	9.77%	458	11.19%	476	21.27%	2,073	10.97%
3	936	13.16%	788	14.45%	759	18.55%	375	16.76%	2,858	15.12%
4	228	3.21%	299	5.48%	288	7.04%	156	6.97%	971	5.14%
5	103	1.45%	0	0.00%	0	0.00%	0	0.00%	103	0.55%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	7,111	100.00%	5,455	100.00%	4,092	100.00%	2,238	100.00%	18,896	100.00%

Figure 2.5.4.6.1
Proficiency Level: Spek 9-12A S601 Paper
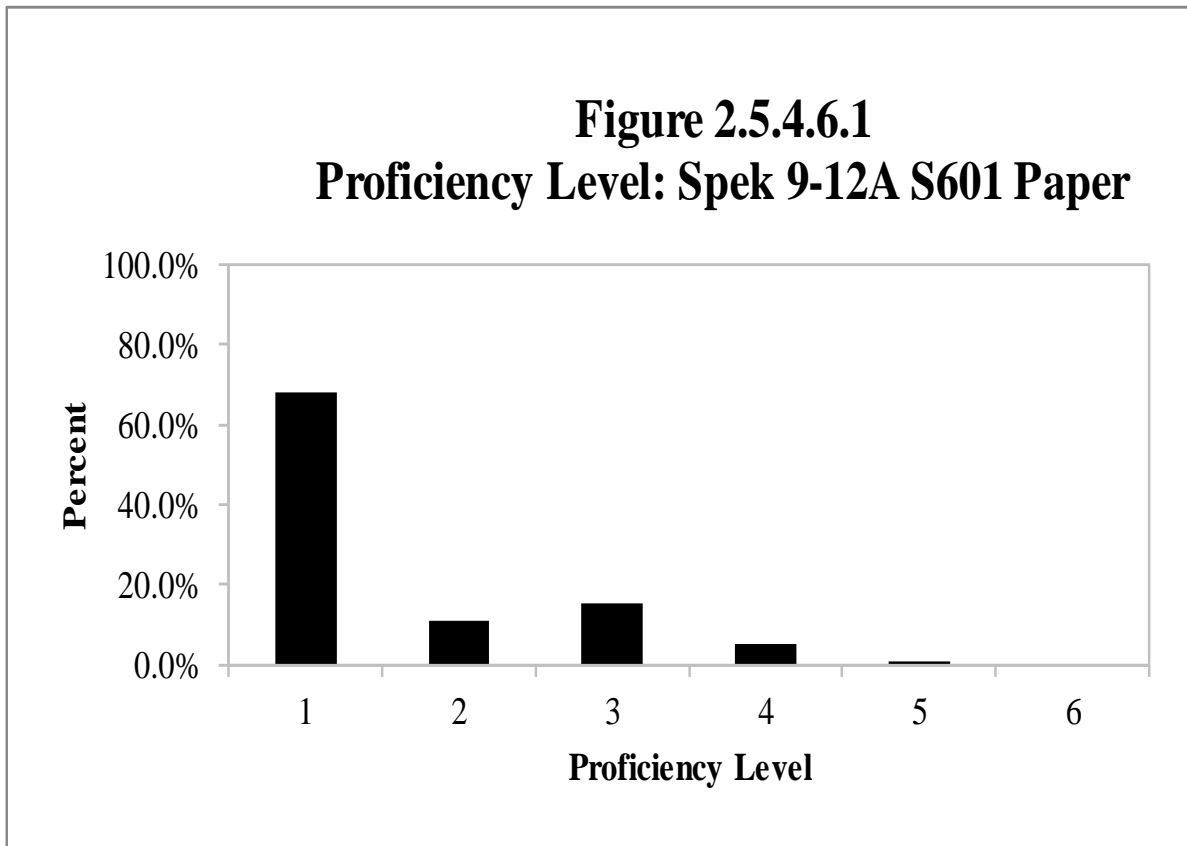
Table 2.5.4.6.2

Proficiency Level Distribution: Spek 9-12 B/C S601 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	712	7.82%	1,113	13.17%	819	10.60%	777	13.48%	3,421	11.02%
2	1,428	15.69%	1,143	13.53%	957	12.39%	859	14.91%	4,387	14.13%
3	2,849	31.31%	2,587	30.61%	2,999	38.83%	2,209	38.33%	10,644	34.29%
4	2,181	23.97%	2,505	29.64%	1,719	22.26%	988	17.14%	7,393	23.82%
5	1,050	11.54%	397	4.70%	419	5.43%	297	5.15%	2,163	6.97%
6	880	9.67%	706	8.35%	810	10.49%	633	10.98%	3,029	9.76%
Total	9,100	100.00%	8,451	100.00%	7,723	100.00%	5,763	100.00%	31,037	100.00%

**Figure 2.5.4.6.2
Proficiency Level: Spek 9-12B/C S601 Paper**
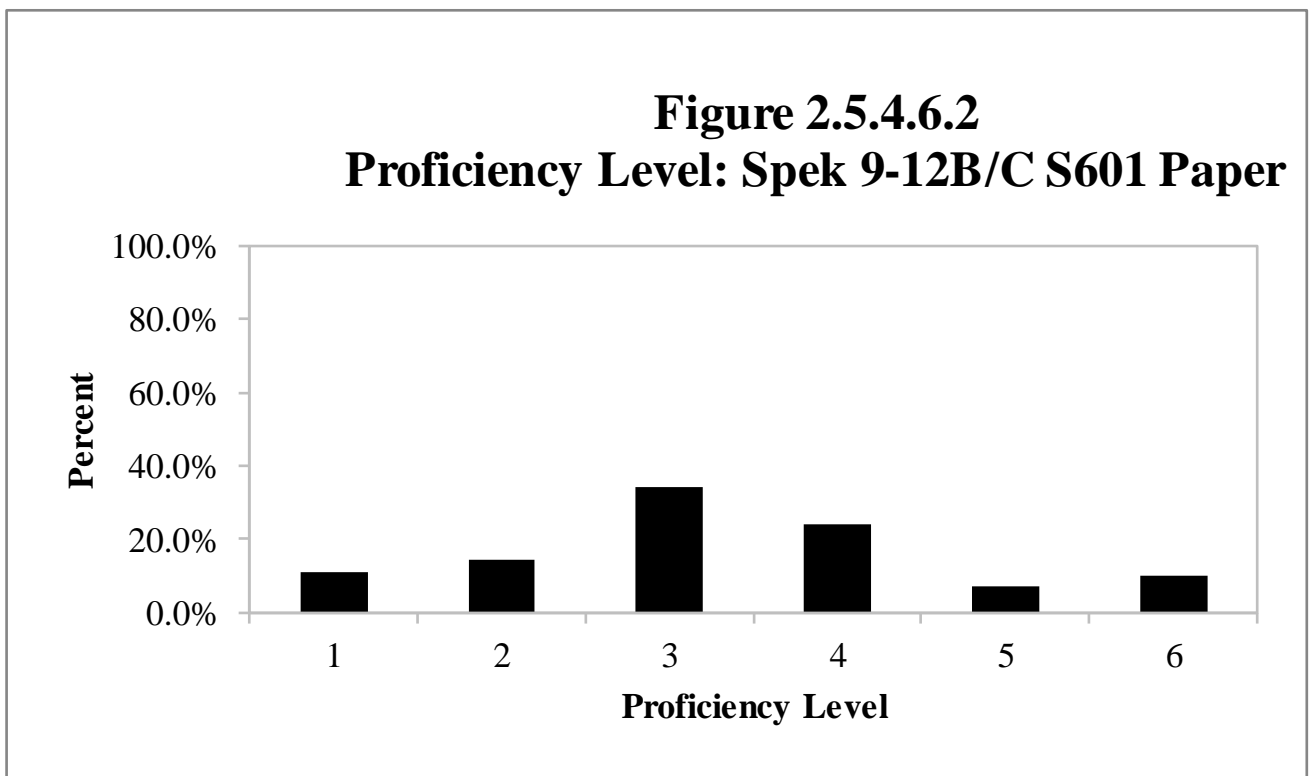
Table 2.5.4.6.3

Proficiency Level Distribution: Spek 9-12 S601 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	5,950	36.70%	4,948	35.58%	3,406	28.83%	2,008	25.10%	16,312	32.67%
2	2,034	12.55%	1,676	12.05%	1,415	11.98%	1,335	16.69%	6,460	12.94%
3	3,785	23.35%	3,375	24.27%	3,758	31.81%	2,584	32.30%	13,502	27.04%
4	2,409	14.86%	2,804	20.16%	2,007	16.99%	1,144	14.30%	8,364	16.75%
5	1,153	7.11%	397	2.85%	419	3.55%	297	3.71%	2,266	4.54%
6	880	5.43%	706	5.08%	810	6.86%	633	7.91%	3,029	6.07%
Total	16,211	100.00%	13,906	100.00%	11,815	100.00%	8,001	100.00%	49,933	100.00%

Figure 2.5.4.6.3
Proficiency Level: Spek 9-12 S601 Paper
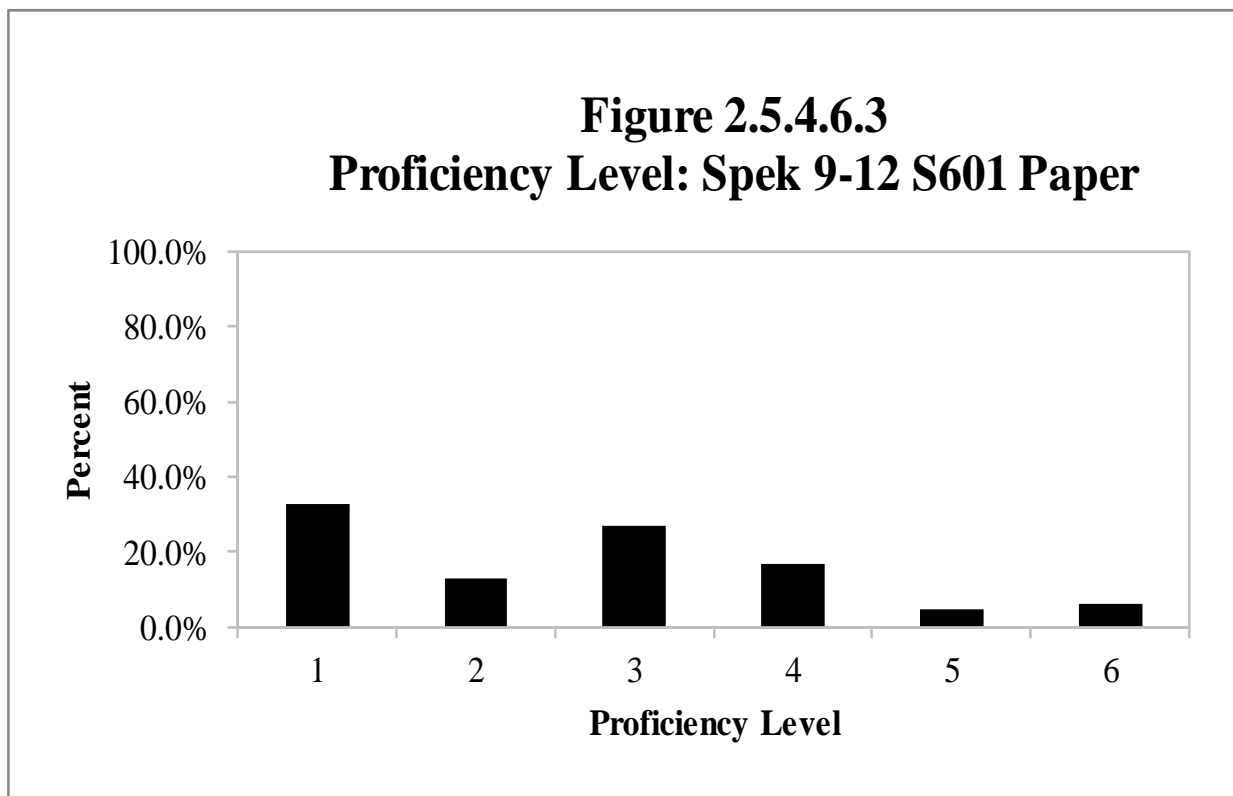
2.6 Raw Score to Scale Score to Proficiency Level Conversion

This section presents raw score to scale score conversions and associated proficiency levels for each test form.

The first column in the tables shows all possible raw scores. The following column shows the corresponding scale score. The next column shows the conditional standard error of measurement (CSEM) in the metric of the scale score, multiplied by 1.96. The resulting number (CSEM x 1.96) is used to construct the confidence band as reported on students' score reports. For example, if a student receives a scale score of 199 and if the CSEM multiplied by 1.96 is 45, then there is a 95% chance that the student's true scale score will be found somewhere between 154-244. For additional detail on conditional standard error of measurement, see Section 5, Reliability. Following the CSEM, columns provide the proficiency level interpretation for each grade in the grade-level cluster.

Performances that gain very few score points, and performances from students who gain all or almost all the score points, will have high CSEM values. The model does not precisely estimate these students' abilities; they may be well below or well above the range that is measured by the test, and therefore the error of measurement is large. We provide further detail on the CSEM because of its importance in interpretating 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 Paper

Raw Score	Scale Score	CSEM x 1.96	PL for G9	PL for G10	PL for G11	PL for G12
0	148	103	1.0	1.0	1.0	1.0
1	172	80	1.2	1.1	1.1	1.1
2	205	61	1.3	1.3	1.3	1.2
3	228	53	1.5	1.4	1.4	1.4
4	246	48	1.6	1.5	1.5	1.5
5	262	45	1.7	1.6	1.6	1.5
6	276	43	1.7	1.7	1.6	1.6
7	288	42	1.8	1.8	1.7	1.7
8	301	41	1.9	1.8	1.8	1.7
9	312	40	1.9	1.9	1.8	1.8
10	323	40	2.2	1.9	1.9	1.9
11	335	40	2.5	2.3	2.0	1.9
12	346	41	2.8	2.6	2.3	2.1
13	358	42	3.1	3.0	2.7	2.6
14	372	45	3.6	3.4	3.2	3.1
15	387	49	4.1	3.9	3.7	3.6
16	402	55	4.7	4.5	4.3	4.1
17	417	63	5.3	5.0	4.8	4.6
18	432	73	5.9	5.6	5.4	5.2

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

Table 2.6.1.6.2

Raw Score to Scale Score to Proficiency Level Conversion: List 9-12 B/C S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 Pap

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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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

Note: Score reports provided to students include the CSEM value multiplied by 1.96.

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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 S601 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 601 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 601 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 502, previously used in 403, and 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 601.

Table 2.7.1

Sources of Series 601 Paper Listening and Reading Forms

	Listening		**Reading**	
Tier A	Years previously used:		Years previously used:	
	Series 503 Paper	2020-21	Series 502 Paper	2020-21
	Series 502 Paper	2018-19	Series 403 Paper	2018-19
	Series 501 Paper	2017-18	Series 401 Paper	2016-17
	Series 403 Paper	2016-17	Series 303	2014-15
	Series 402 Paper	2015-16		
	Series 401 Paper	2014-15		
	Series 400 Paper			
	Series 303			
G9-12 Tier A	Series 503			
Tier B/C	Series 502		Series 502	
	Series 403		Series 403	

The S601 Tier A and BC forms are the same as S502, except for grades 9-12, Listening Tier A, which had one folder removed and replaced with an item from S503, due to having a C-level DIF. The S502 Tier B/C forms are the same as 403, which 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. We constructed the Series 502 Paper Listening and Reading Grades 1–12 Tier B/C forms at this meeting.

Writing and Speaking

Writing and Speaking are also static forms. Table 2.7.2 summarizes prior uses of these forms. Please see the Annual Technical Report for ACCESS for ELLs Paper Series 401 (CAL, 2018) for equating summaries for Writing and Speaking.

Table 2.7.2

Sources of Series 601 Paper Writing and Speaking Forms

	Writing		**Speaking**	
Tier A	Years previously used:		Years previously used:	
	Series 502 Paper	2020-21	Series 502 Paper	2020-21
	Series 403 Paper	2018-19	Series 403 Paper	2018-19
	Series 401 Paper	2016-17	Series 401 Paper	2016-17
Tier B/C	Years previously used:		Years previously used:	
	Series 502 Paper	2020-21	Series 502 Paper	2020-21
	Series 403 Paper	2018-19	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 in higher proficiency levels. The relative width of each section between the cut score lines, however, gives an indication of how many items on that form must be answered correctly (for Listening or Reading) or how many points must be earned (for Writing or Speaking) to be placed into a WIDA proficiency level.

2.8.1 Listening

2.8.1.0 Kindergarten

Figure 2.8.1.0
Test Characteristic Curve: List K S601 Paper

2.8.1.1 Grade 1

Figure 2.8.1.1.1
Test Characteristic Curve: List 1 A S601 Paper
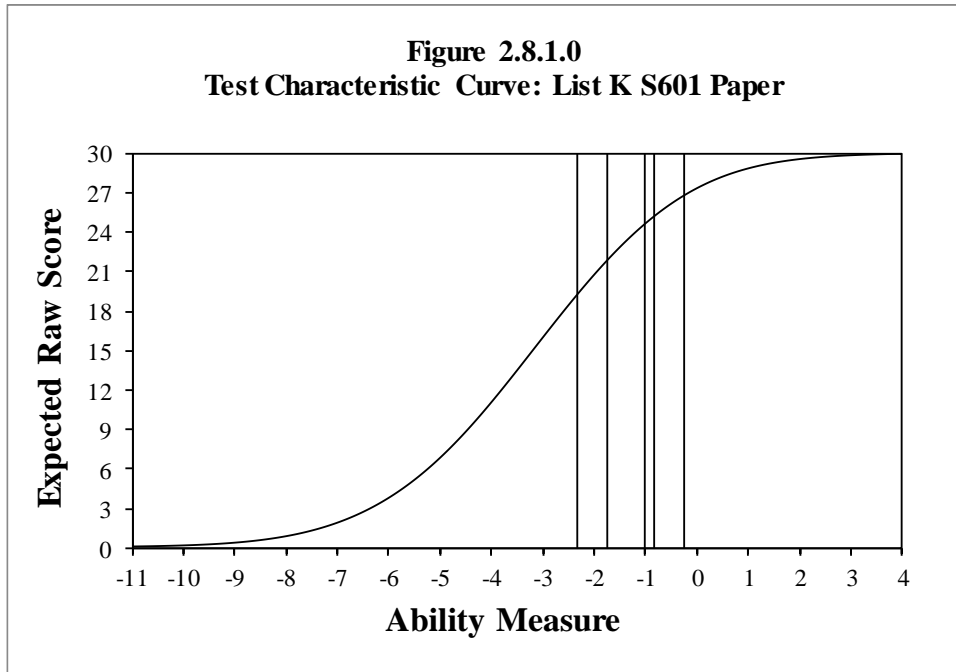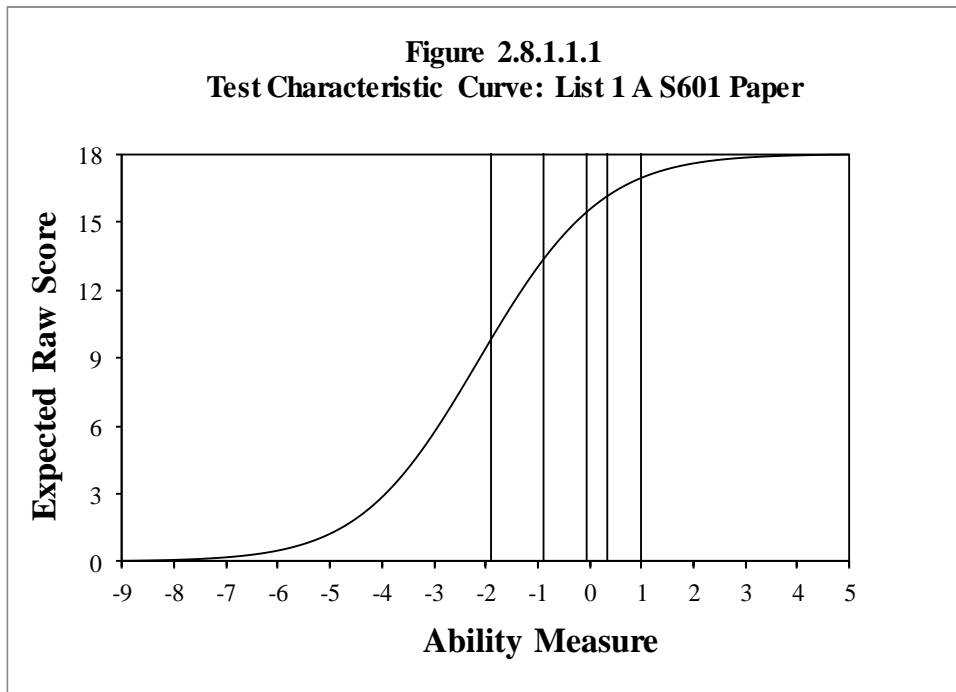
Note: The test form is shared between 1A and 2A.

Figure 2.8.1.1.2
Test Characteristic Curve: List 1 B/C S601 Paper
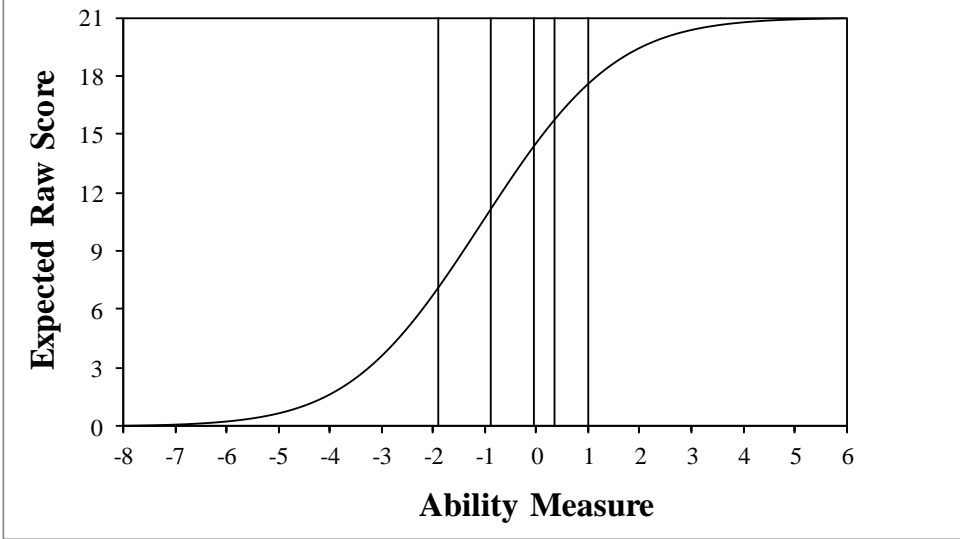
Note: The test form is shared between 1B/C and 2B/C.

2.8.1.2 Grade 2

Figure 2.8.1.2.1
Test Characteristic Curve: List 2 A S601 Paper
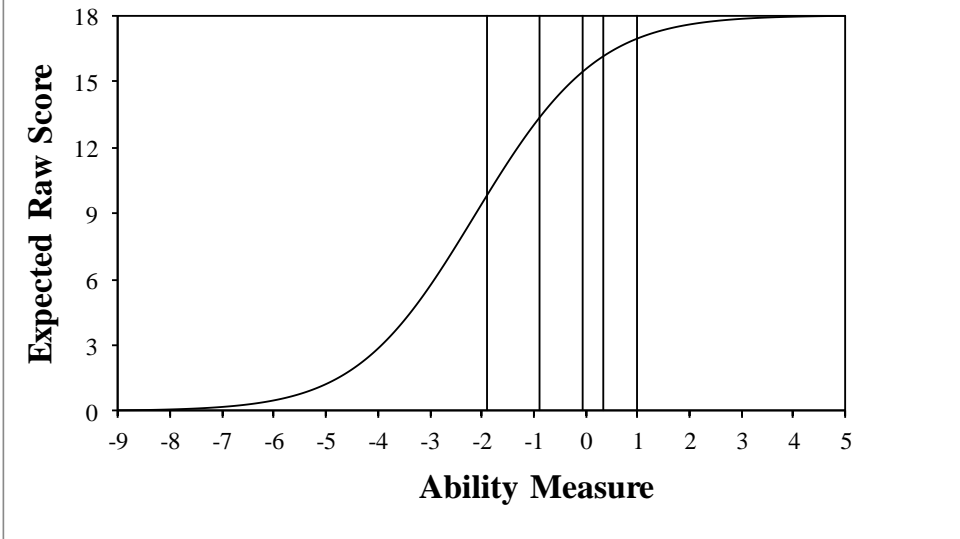
Note: The test form is shared between 1A and 2A.

Figure 2.8.1.2.2
Test Characteristic Curve: List 2 B/C S601 Paper
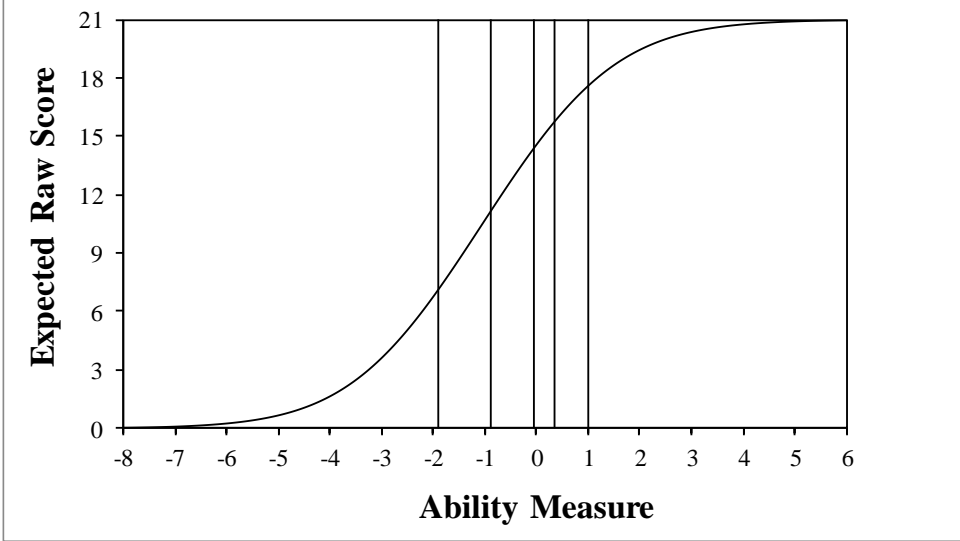
Note: The test form is shared between 1B/C and 2B/C.

2.8.1.3 Grade 3

Figure 2.8.1.3.1
Test Characteristic Curve: List 3 A S601 Paper
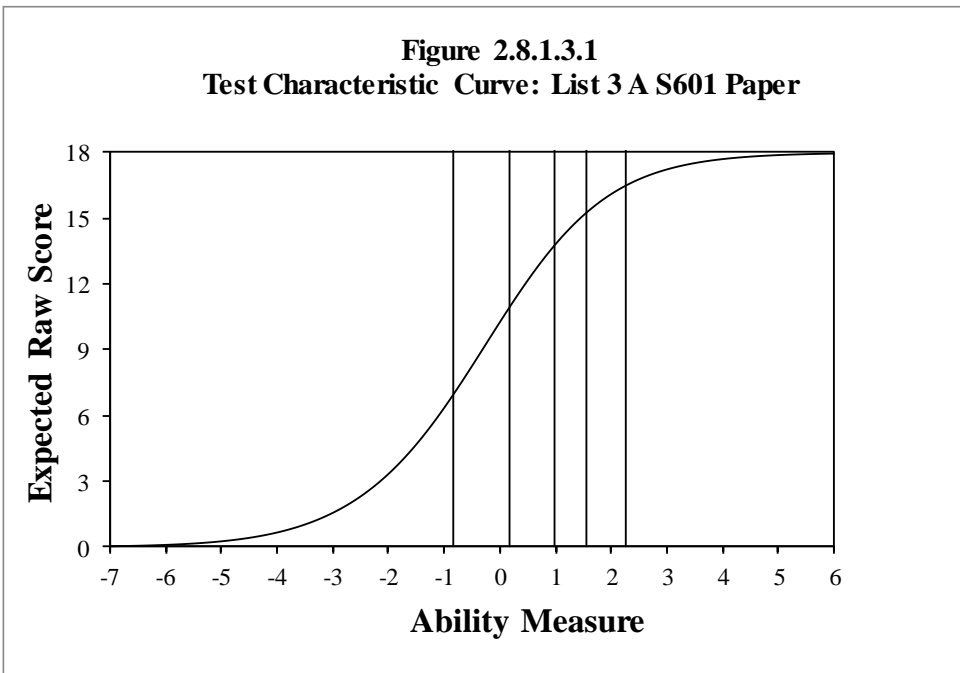
Note: The test form is shared between 3A and 4-5A.

Figure 2.8.1.3.2
Test Characteristic Curve: List 3 B/C S601 Paper
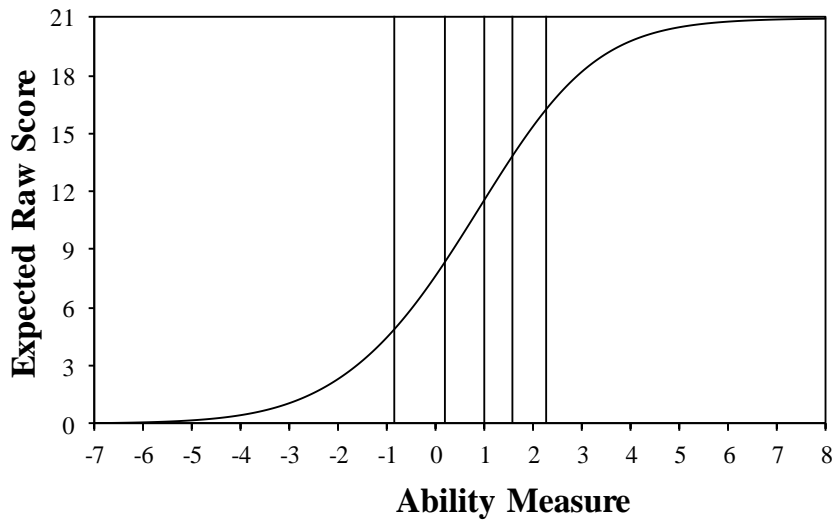
Note: The test form is shared between 3B/C and 4-5B/C.

2.8.1.4 Grades 4-5

Figure 2.8.1.4.1
Test Characteristic Curve: List 4-5 A S601 Paper
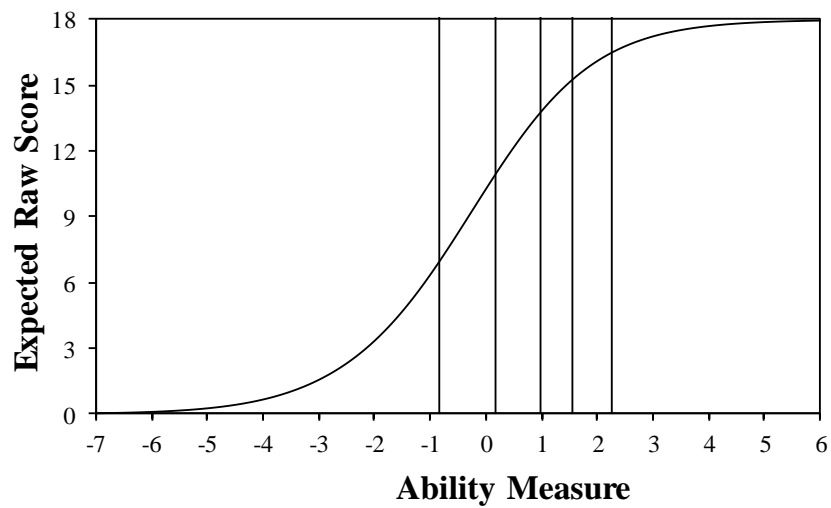
Note: The test form is shared between 3A and 4-5A.

Figure 2.8.1.4.2
Test Characteristic Curve: List 4-5 B/C S601 Paper
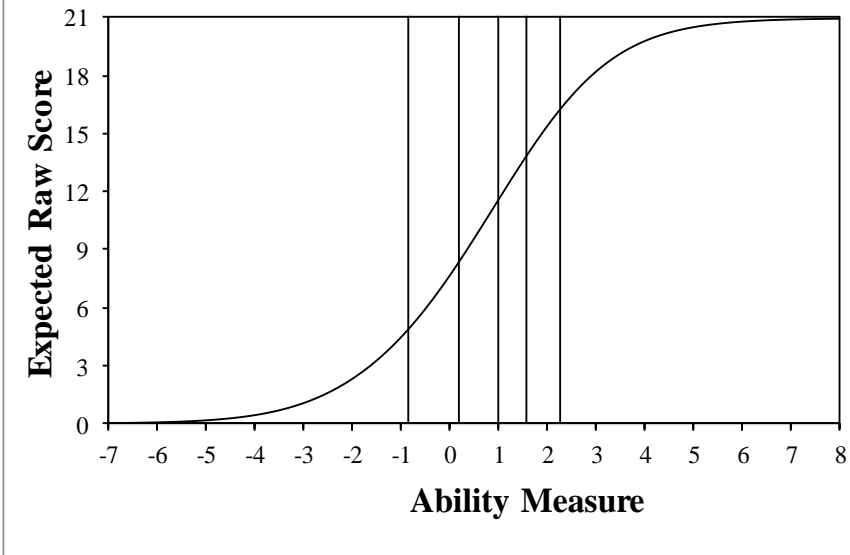
Note: The test form is shared between 3B/C and 4-5B/C.

2.8.1.5 Grades 6-8

Figure 2.8.1.5.1
Test Characteristic Curve: List 6-8 A S601 Paper
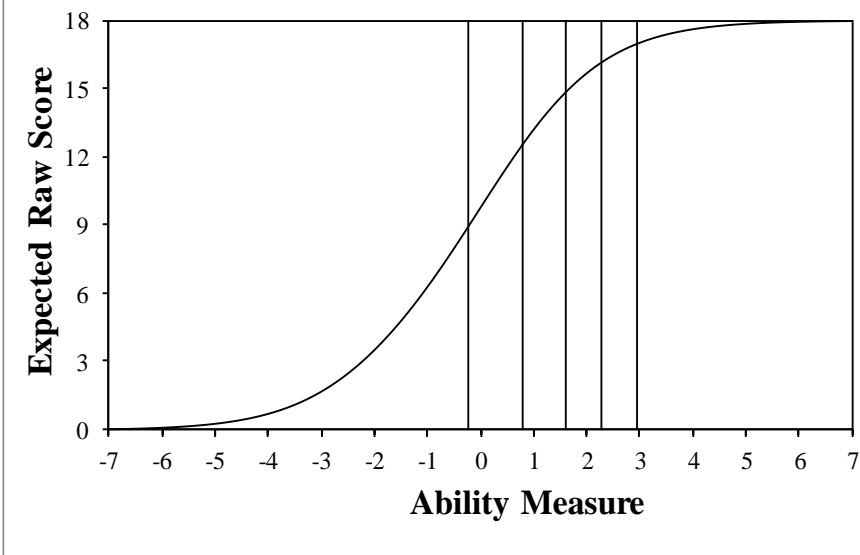
Figure 2.8.1.5.2
Test Characteristic Curve: List 6-8 B/C S601 Paper
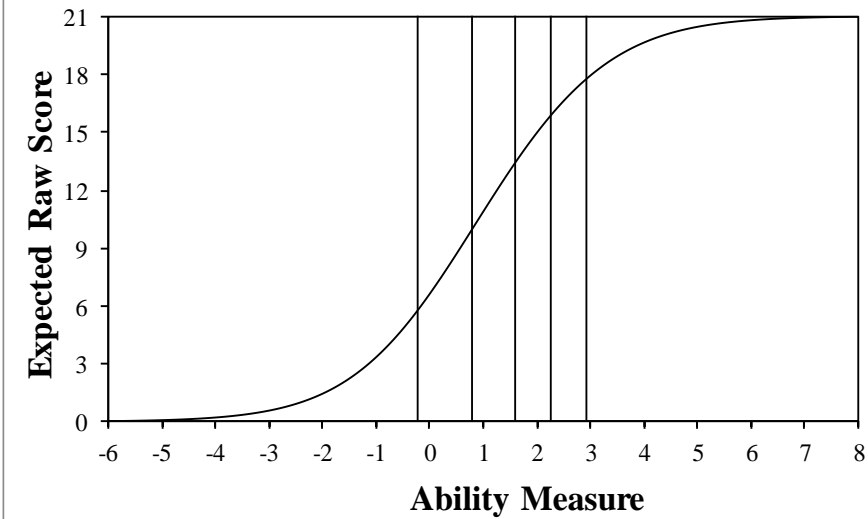
2.8.1.6 Grades 9-12

Figure 2.8.1.6.1
Test Characteristic Curve: List 9-12 A S601 Paper
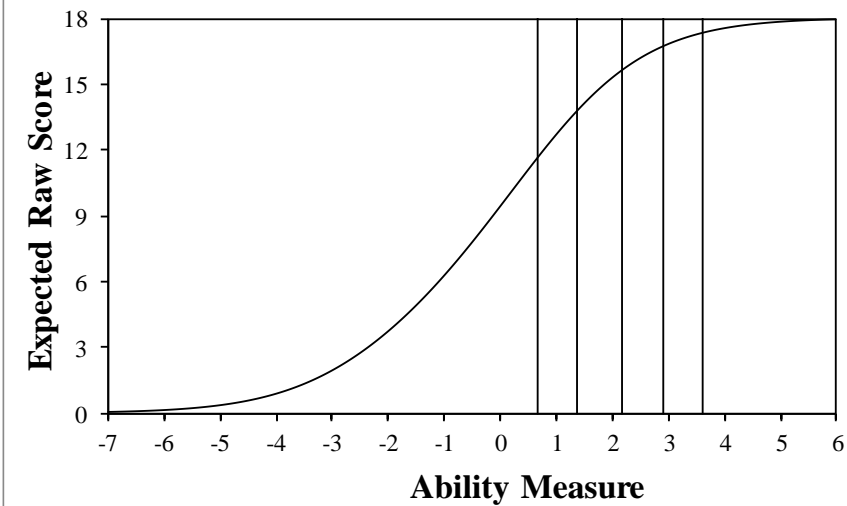
Figure 2.8.1.6.2
Test Characteristic Curve: List 9-12 B/C S601 Paper
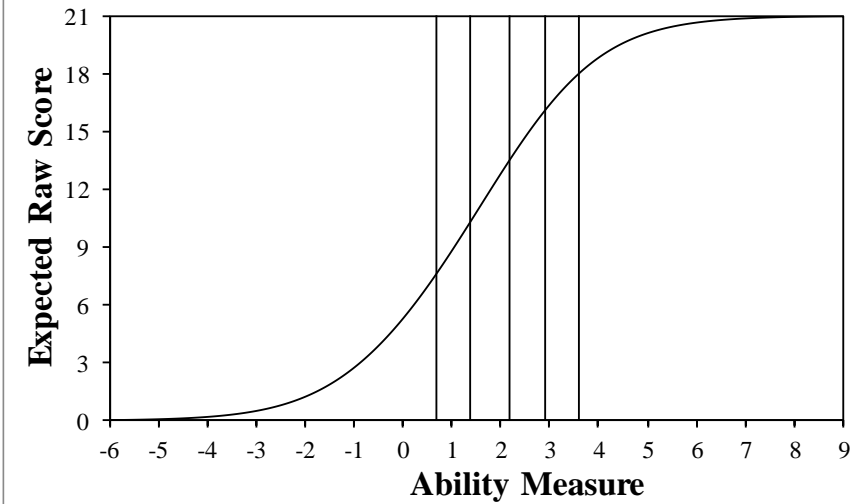
2.8.2 Reading

2.8.2.0 Kindergarten

Figure 2.8.2.0
Test Characteristic Curve: Read K S601 Paper
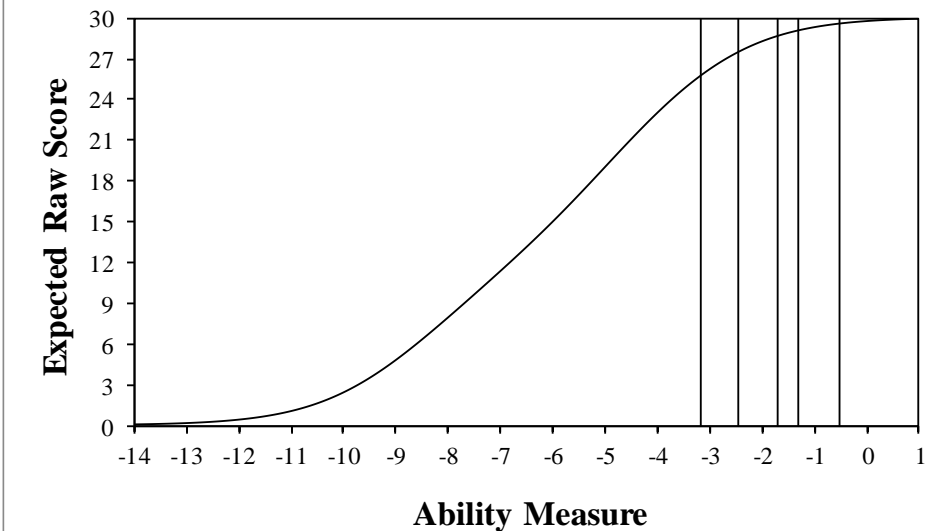
2.8.2.1 Grade 1
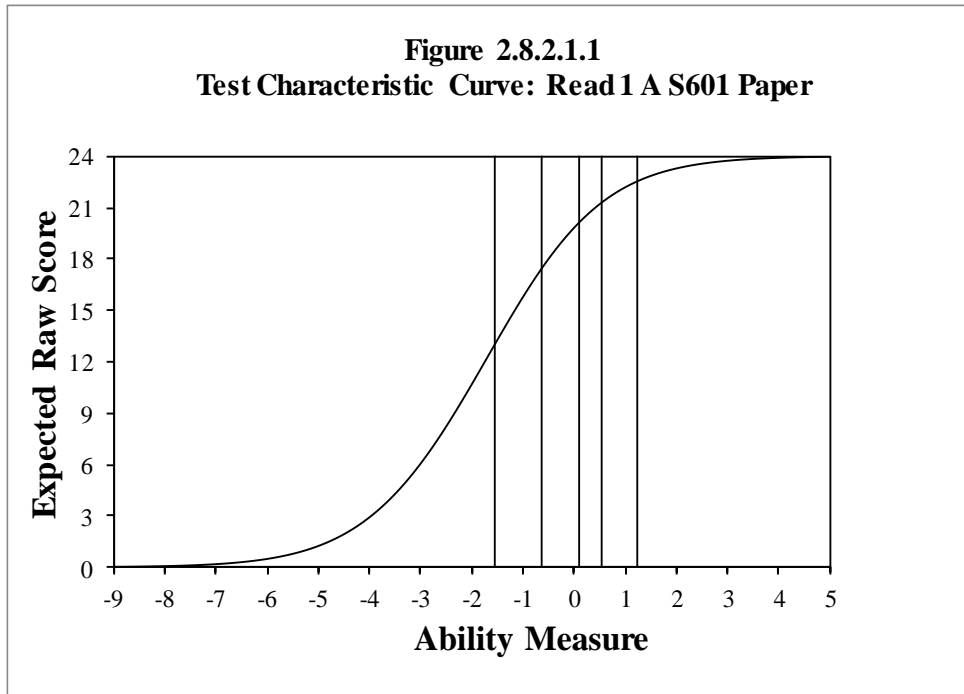
Figure 2.8.2.1.1
Test Characteristic Curve: Read 1 A S601 Paper

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

Figure 2.8.2.1.2
Test Characteristic Curve: Read 1 B/C S601 Paper
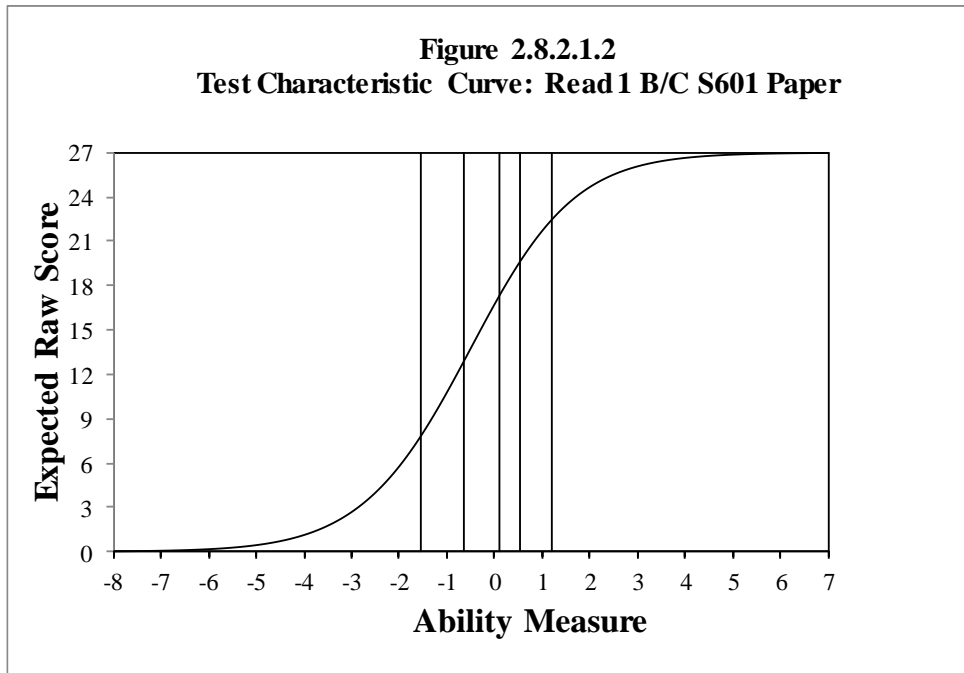
Note: The test form is shared between 1B/C and 2B/C.

2.8.2.2 Grade 2
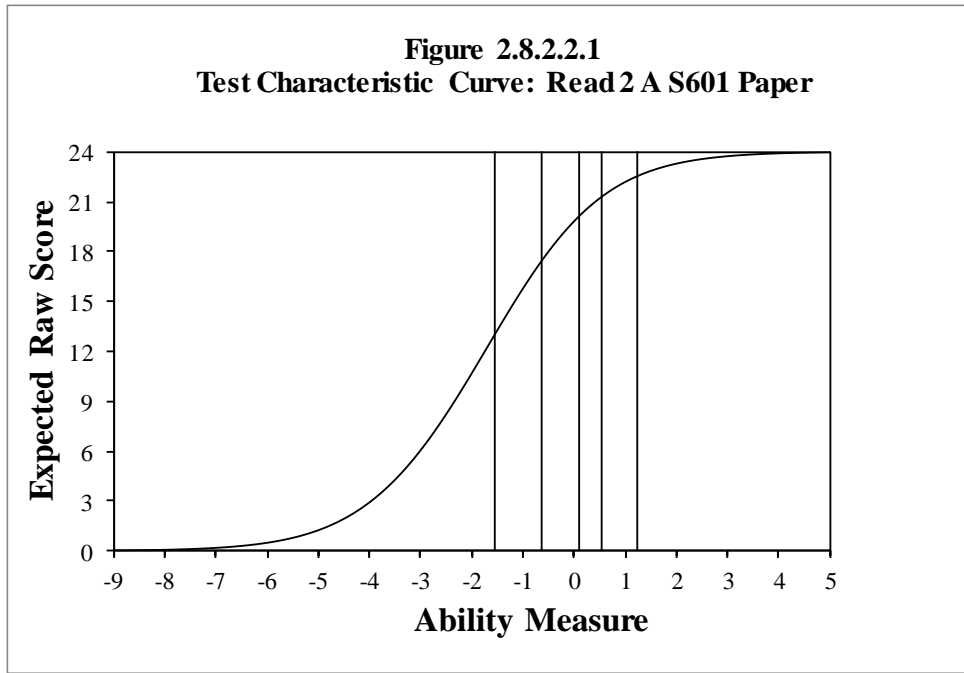
Figure 2.8.2.2.1
Test Characteristic Curve: Read 2 A S601 Paper

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

Figure 2.8.2.2.2
Test Characteristic Curve: Read 2 B/C S601 Paper
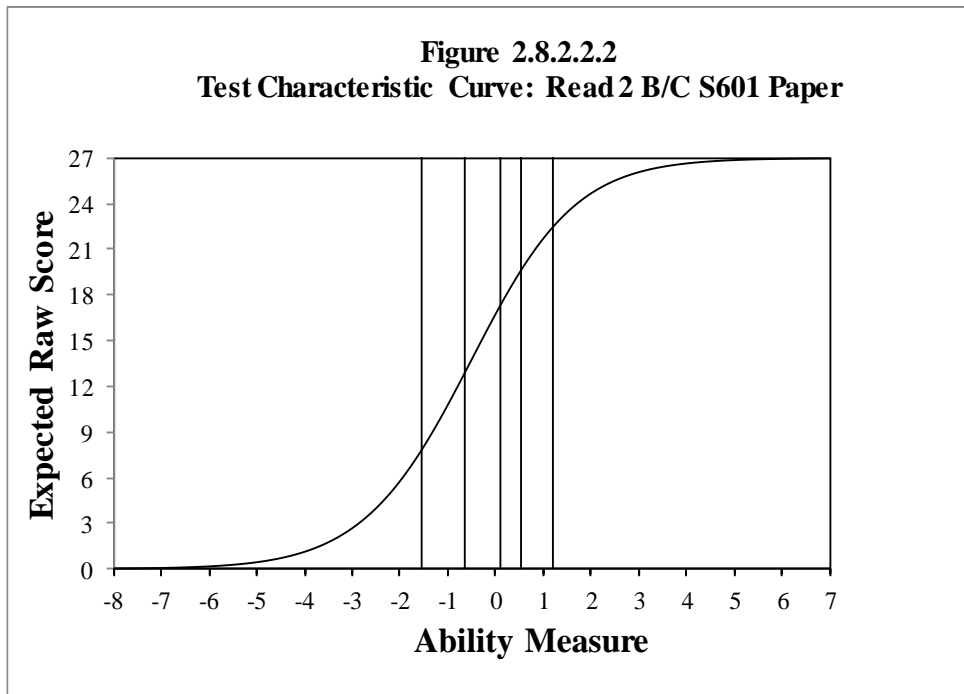
Note: The test form is shared between 1B/C and 2B/C.

2.8.2.3 Grade 3
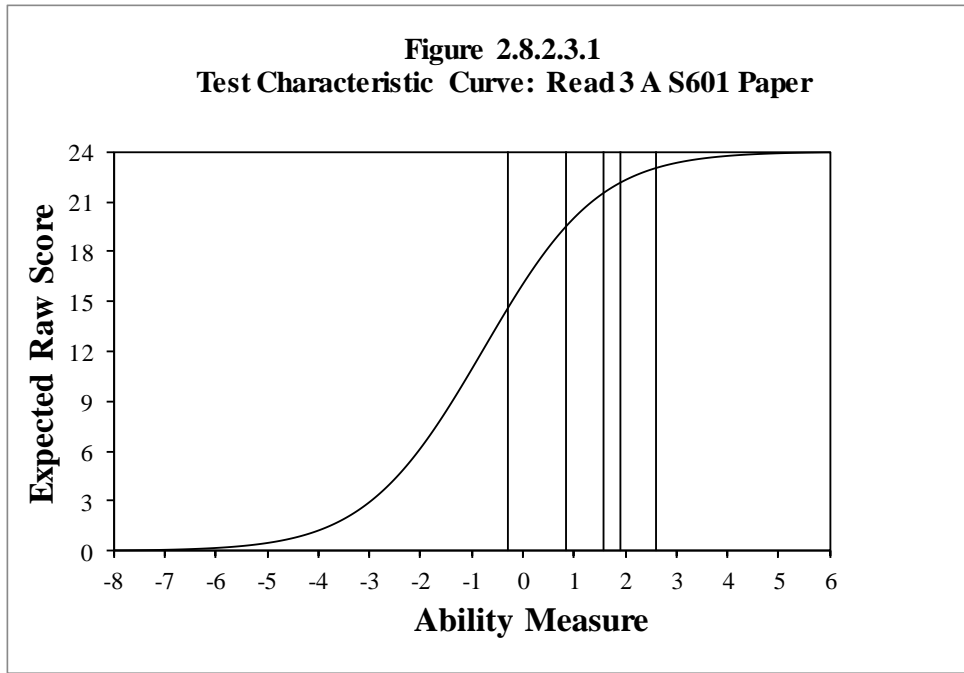
Figure 2.8.2.3.1
Test Characteristic Curve: Read 3 A S601 Paper

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

Figure 2.8.2.3.2
Test Characteristic Curve: Read 3 B/C S601 Paper
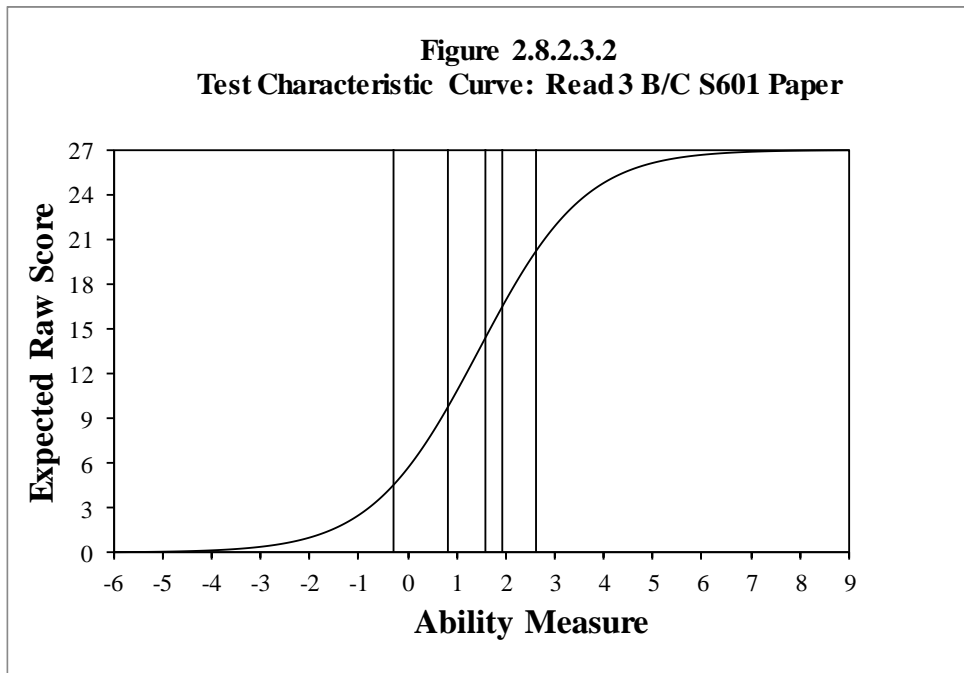
Note: The test form is shared between 3B/C and 4-5B/C.

2.8.2.4 Grades 4-5
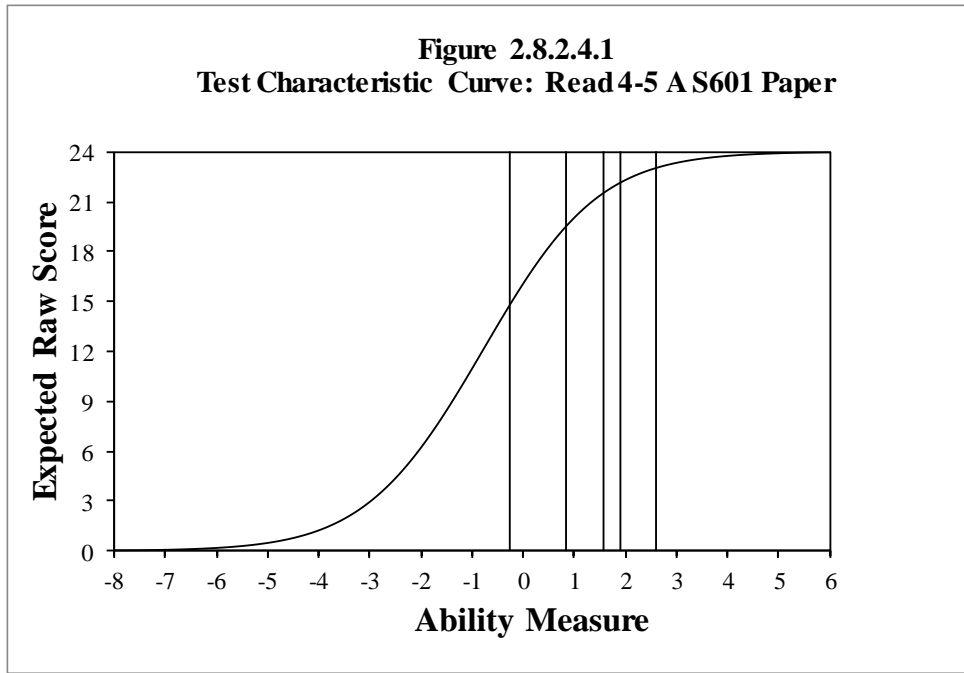
Figure 2.8.2.4.1
Test Characteristic Curve: Read 4-5 A S601 Paper

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

Figure 2.8.2.4.2
Test Characteristic Curve: Read 4-5 B/C S601 Paper
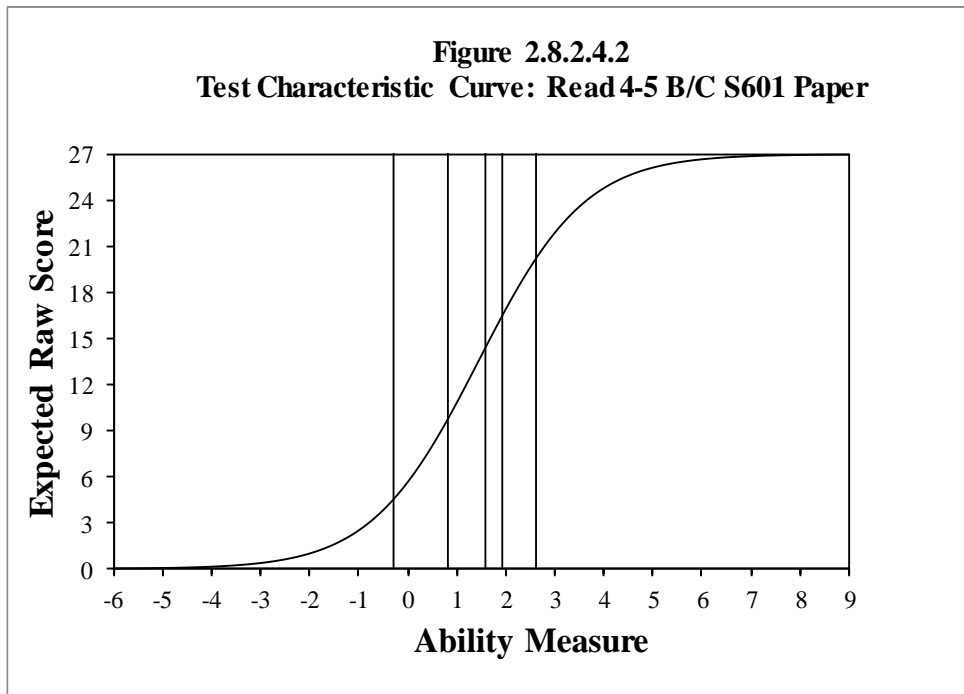
Note: The test form is shared between 3B/C and 4-5B/C.

2.8.2.5 Grades 6-8

Figure 2.8.2.5.1
Test Characteristic Curve: Read 6-8 A S601 Paper
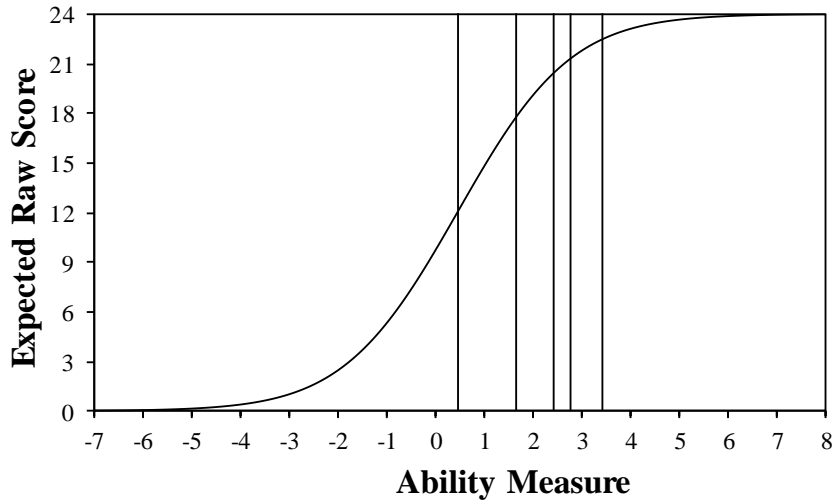
Figure 2.8.2.5.2
Test Characteristic Curve: Read 6-8 B/C S601 Paper
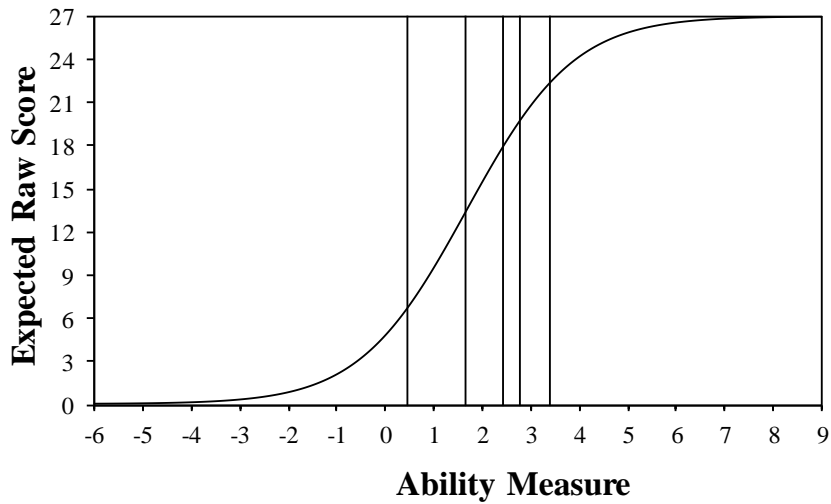
2.8.2.6 Grades 9-12

Figure 2.8.2.6.1
Test Characteristic Curve: Read 9-12 A S601 Paper
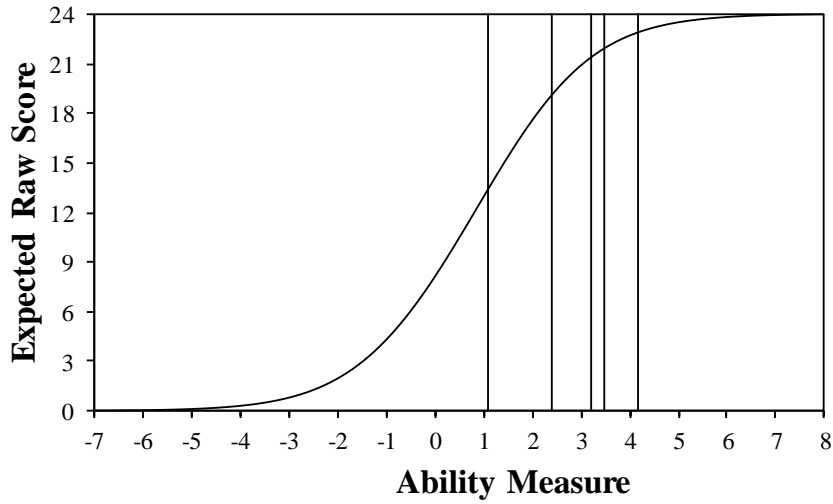
Figure 2.8.2.6.2
Test Characteristic Curve: Read 9-12 B/C S601 Paper
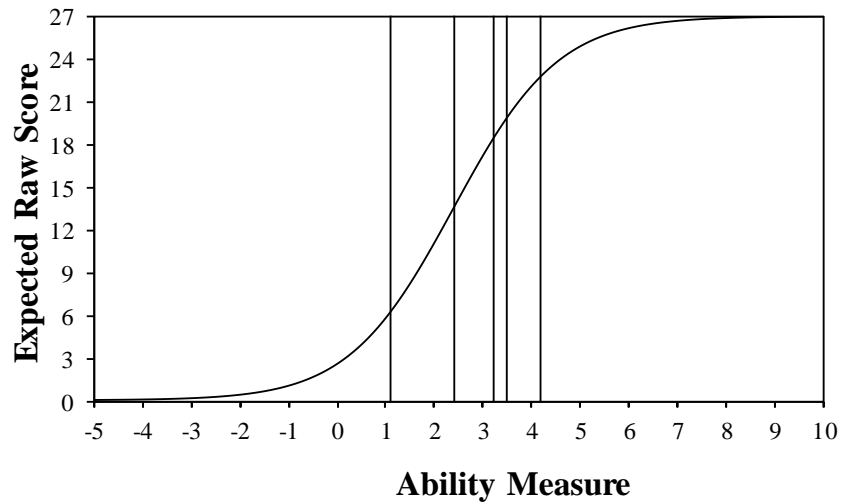
2.8.3 Writing

2.8.3.0 Kindergarten
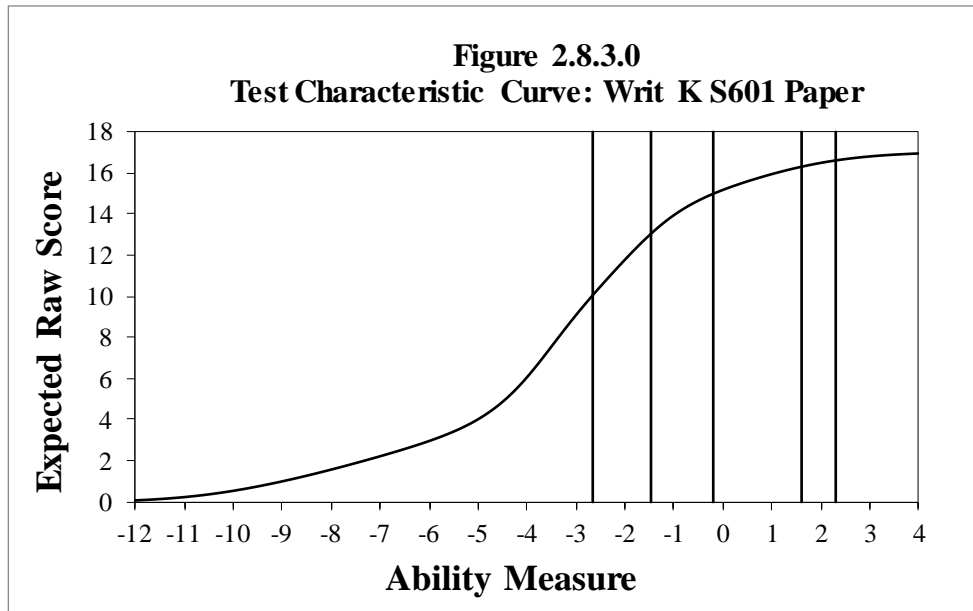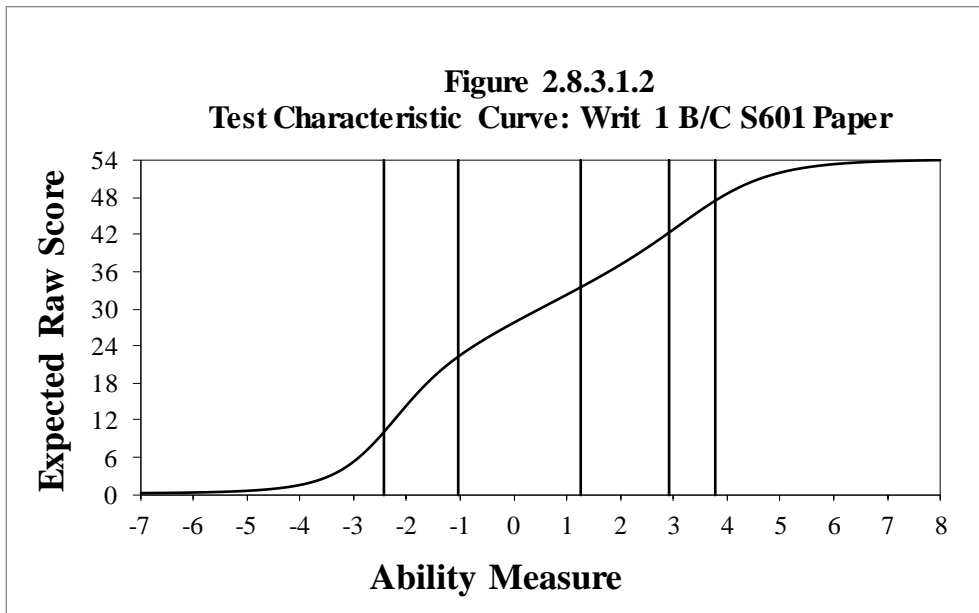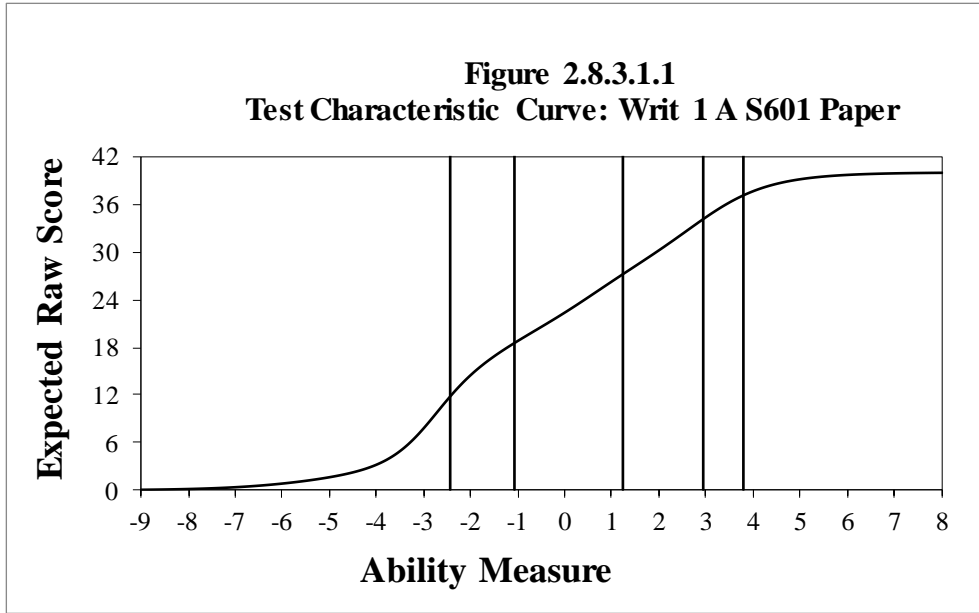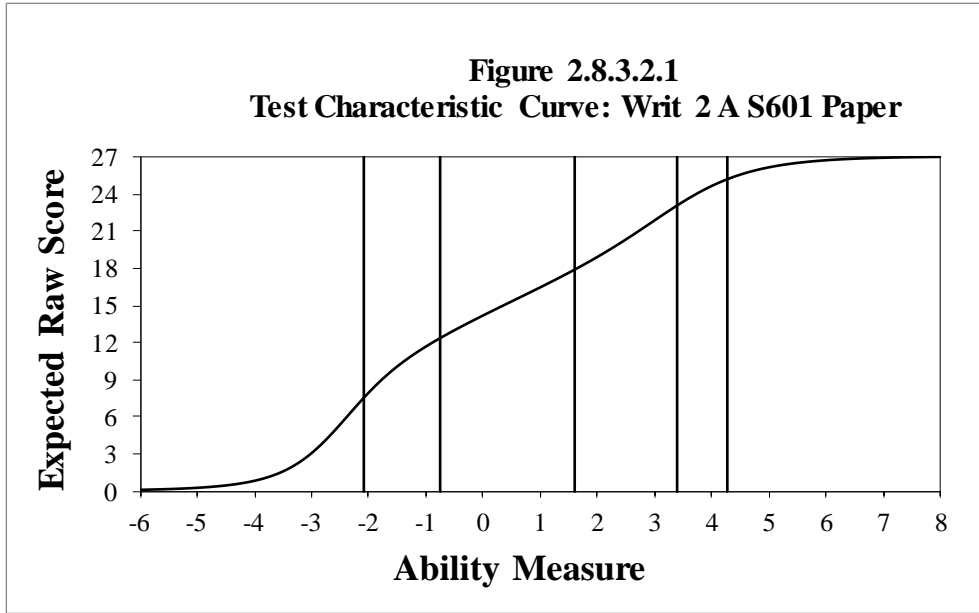
Figure 2.8.3.0
Test Characteristic Curve: Writ K S601 Paper

2.8.3.1 Grade 1

Figure 2.8.3.1.1
Test Characteristic Curve: Writ 1 A S601 Paper

Figure 2.8.3.1.2
Test Characteristic Curve: Writ 1 B/C S601 Paper

2.8.3.2 Grade 2

Figure 2.8.3.2.1
Test Characteristic Curve: Writ 2 A S601 Paper

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

Figure 2.8.3.2.2
Test Characteristic Curve: Writ 2 B/C S601 Paper
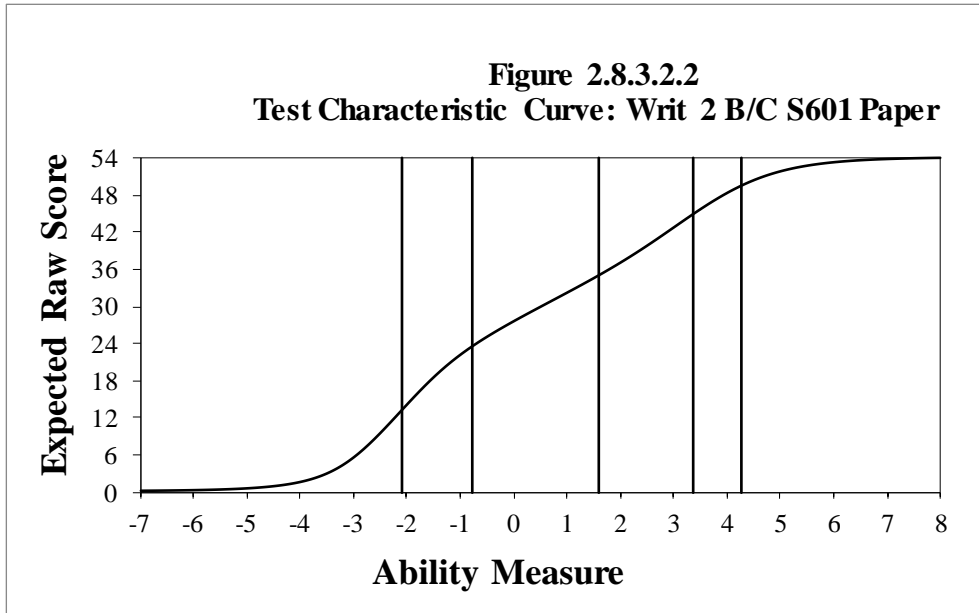
Note: The test form is shared between 2B/C and 3B/C.

2.8.3.3 Grade 3
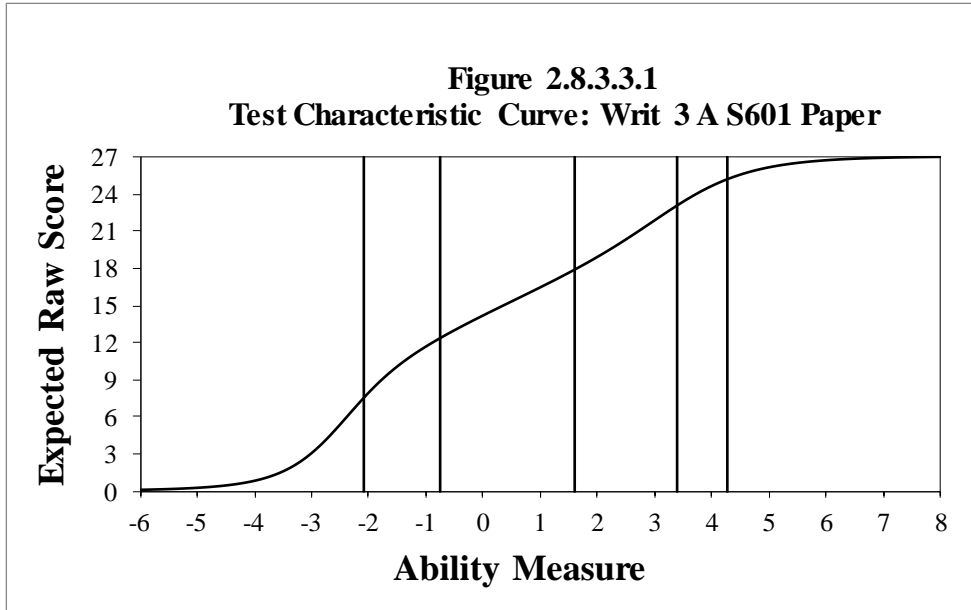
Figure 2.8.3.3.1
Test Characteristic Curve: Writ 3 A S601 Paper

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

Figure 2.8.3.3.2
Test Characteristic Curve: Writ 3 B/C S601 Paper
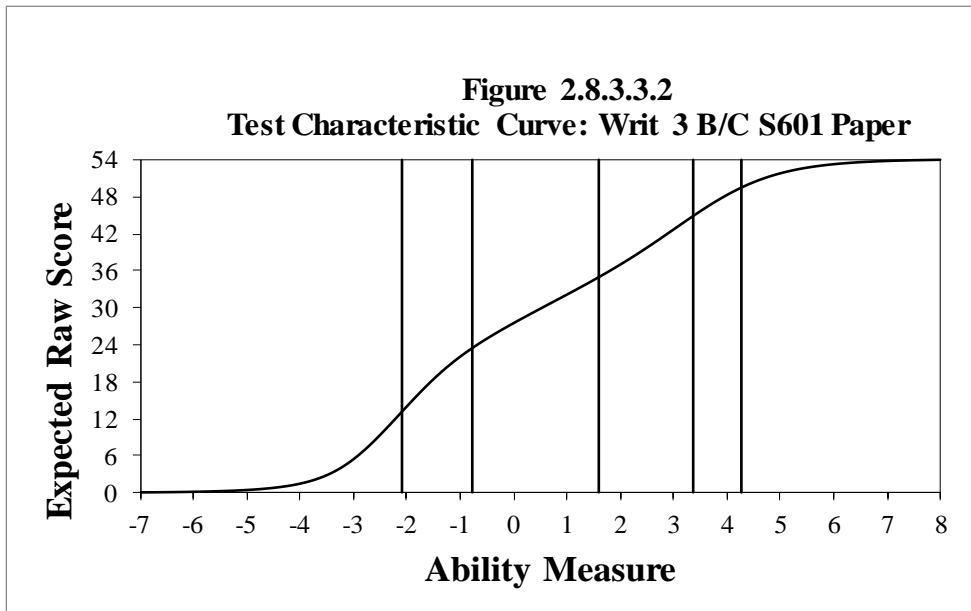
Note: The test form is shared between 2B/C and 3B/C.

2.8.3.4 Grades 4-5

Figure 2.8.3.4.1
Test Characteristic Curve: Writ 4-5 A S601 Paper

Figure 2.8.3.4.2
Test Characteristic Curve: Writ 4-5 B/C S601 Paper

2.8.3.5 Grades 6-8

Figure 2.8.3.5.1
Test Characteristic Curve: Writ 6-8 A S601 Paper

Figure 2.8.3.5.2
Test Characteristic Curve: Writ 6-8 B/C S601 Paper

2.8.3.6 rades 9-12

Figure 2.8.3.6.1
Test Characteristic Curve: Writ 9-12 A S601 Paper

Figure 2.8.3.6.2
Test Characteristic Curve: Writ 9-12 B/C S601 Paper
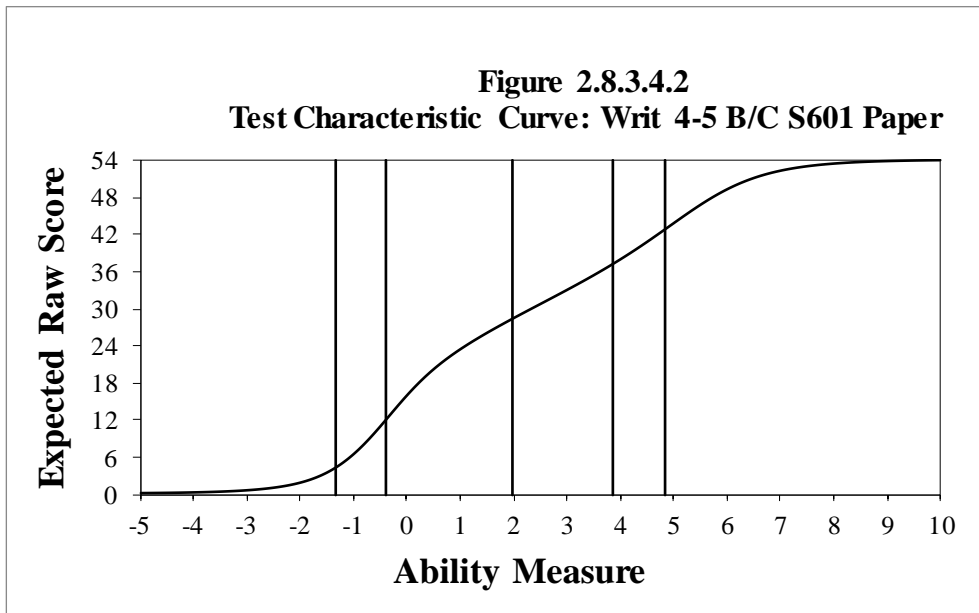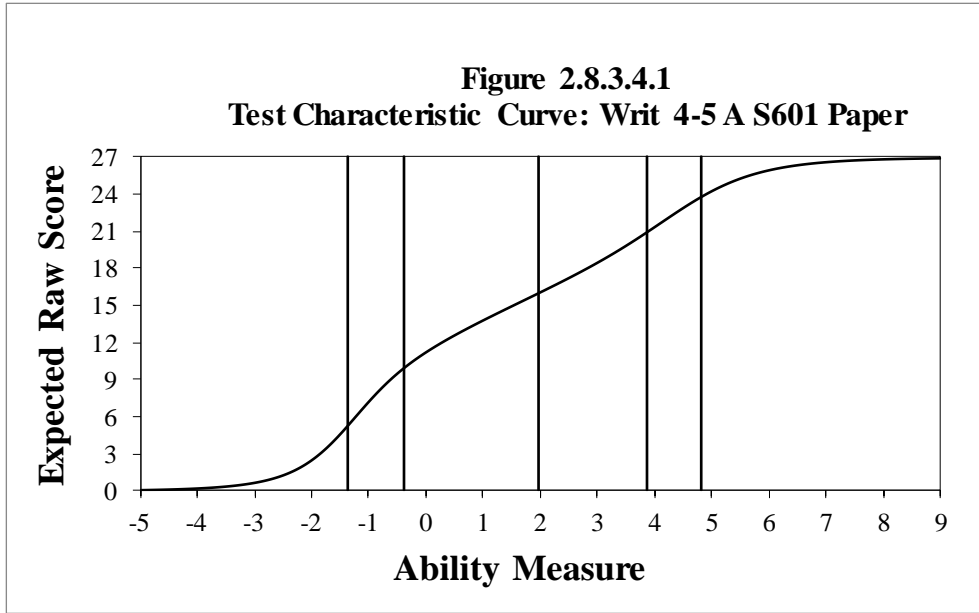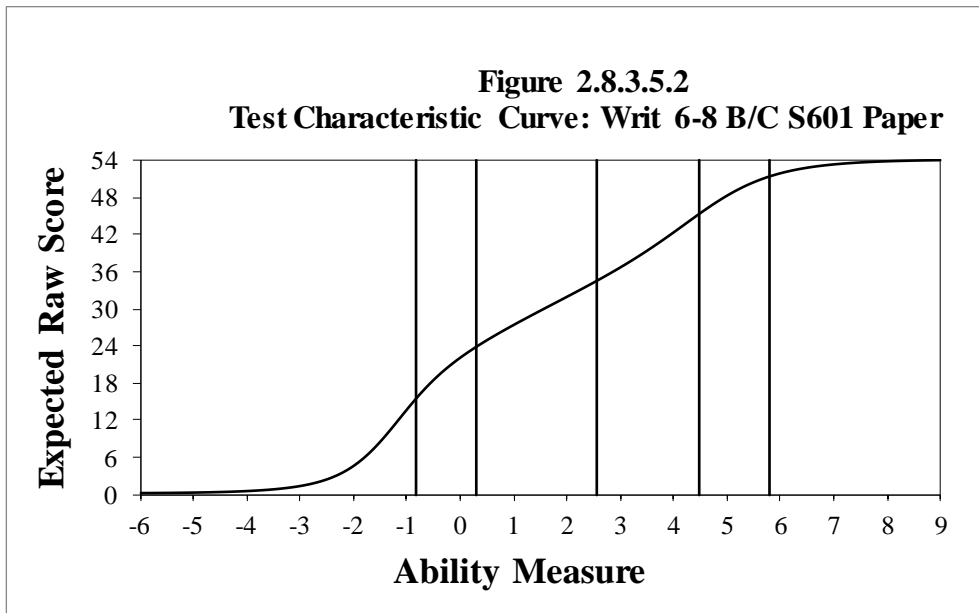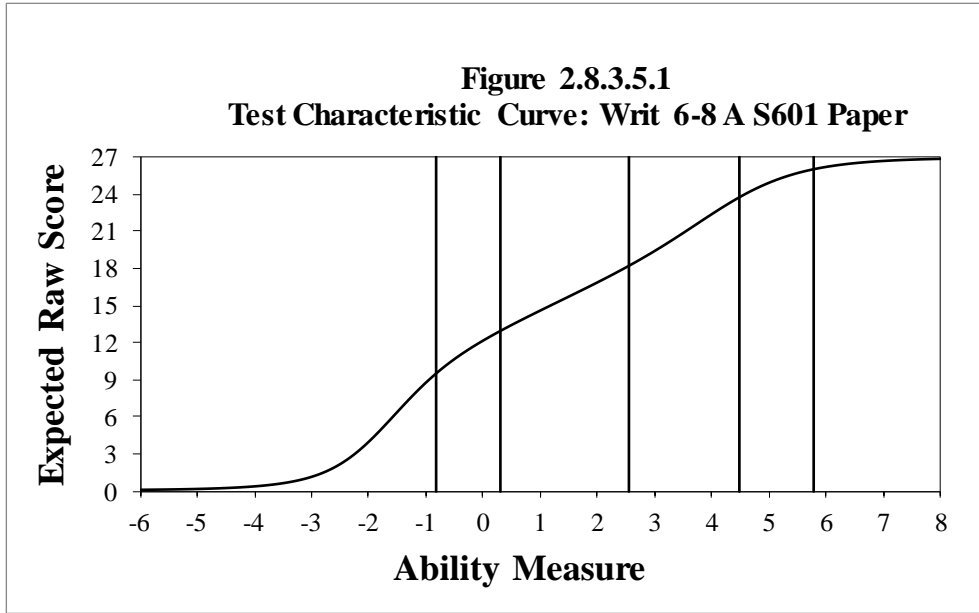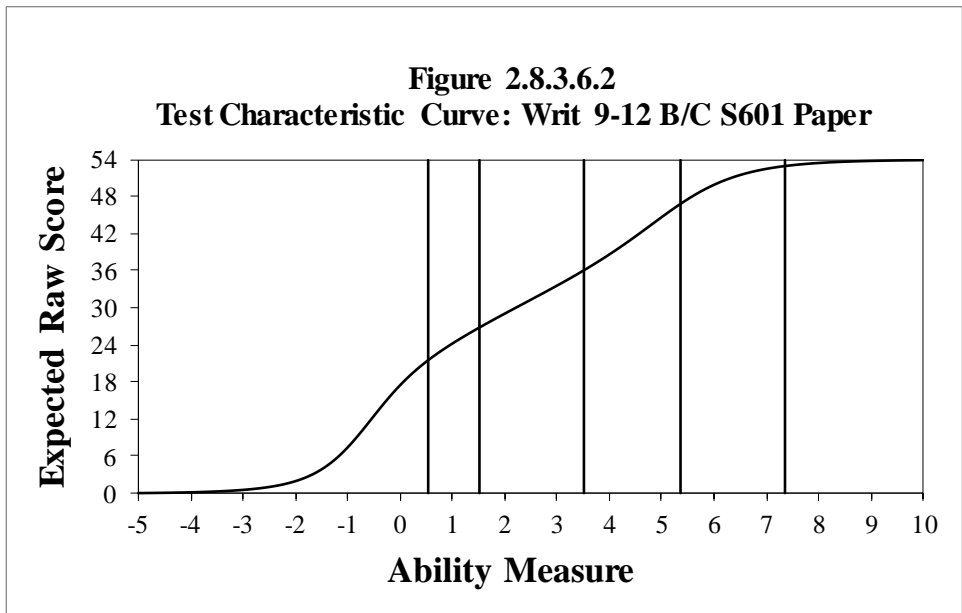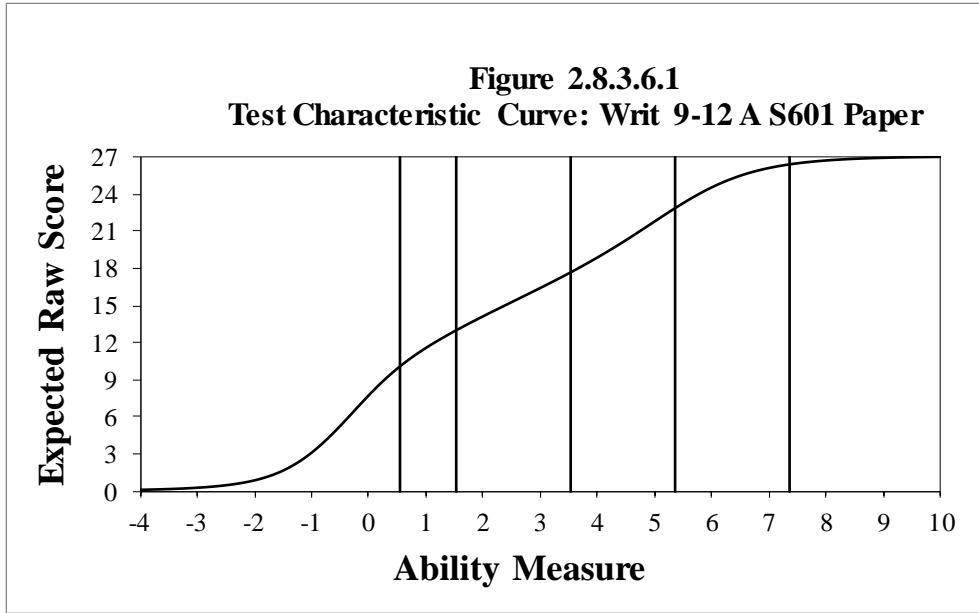
2.8.4 Speaking

2.8.4.0 Kindergarten
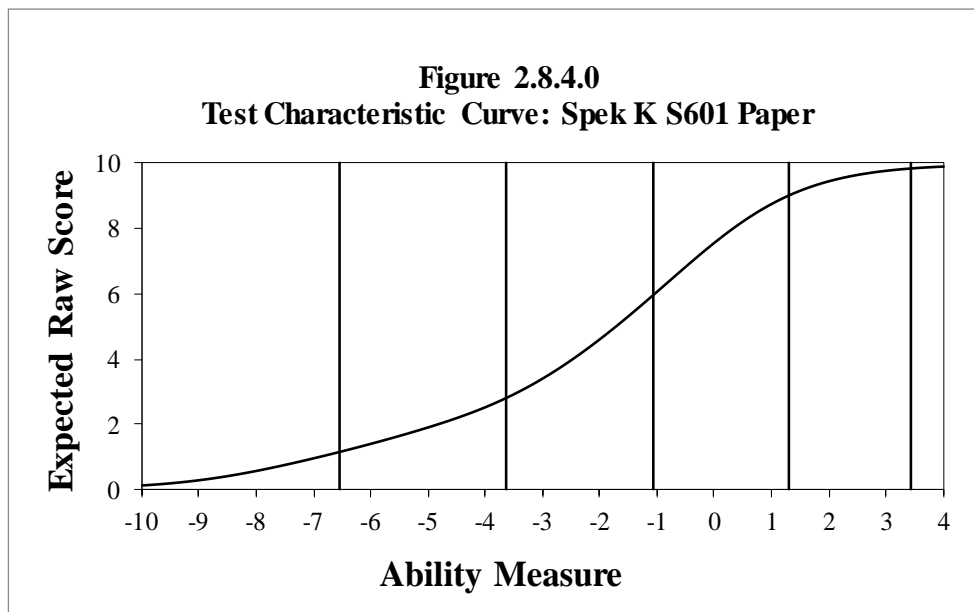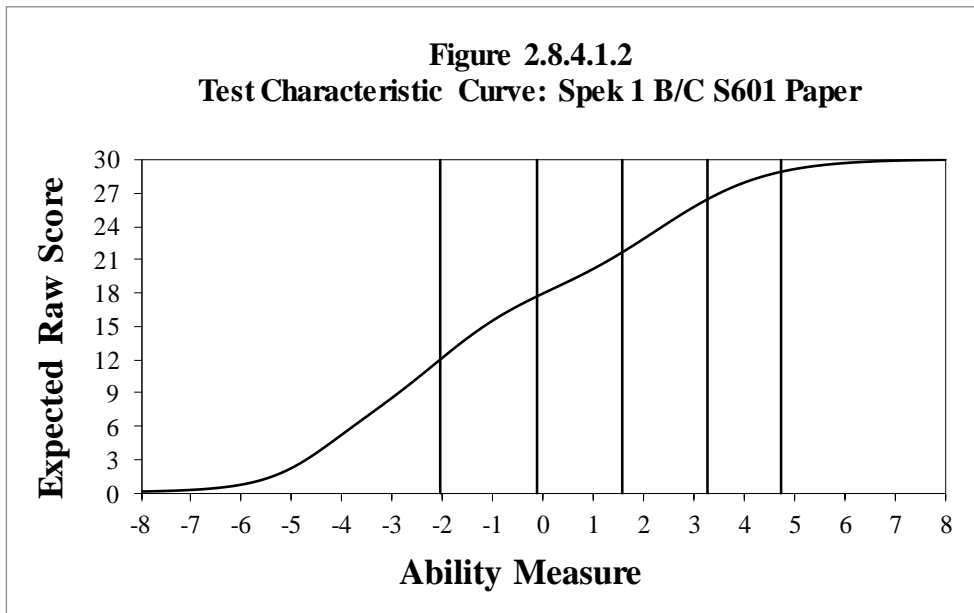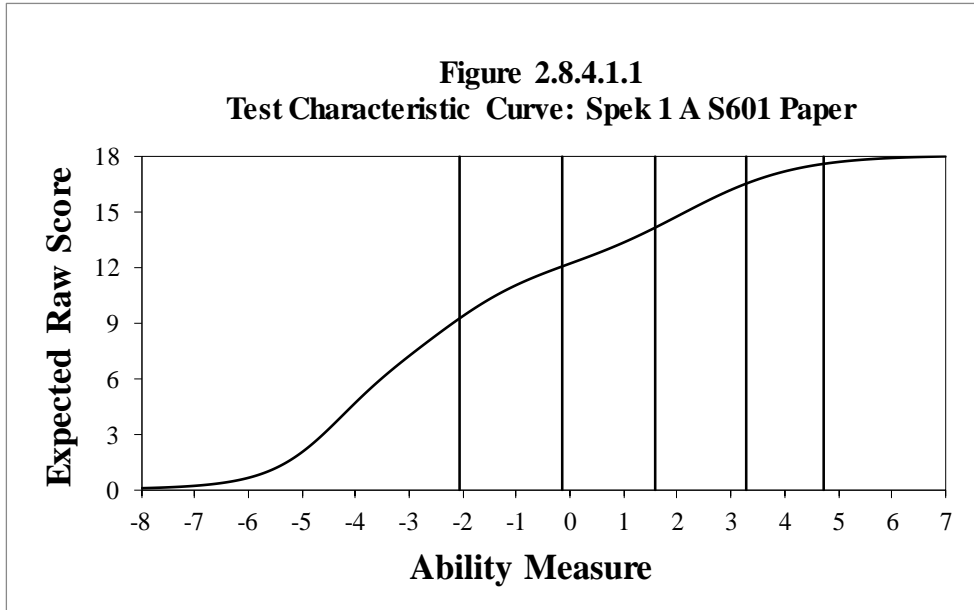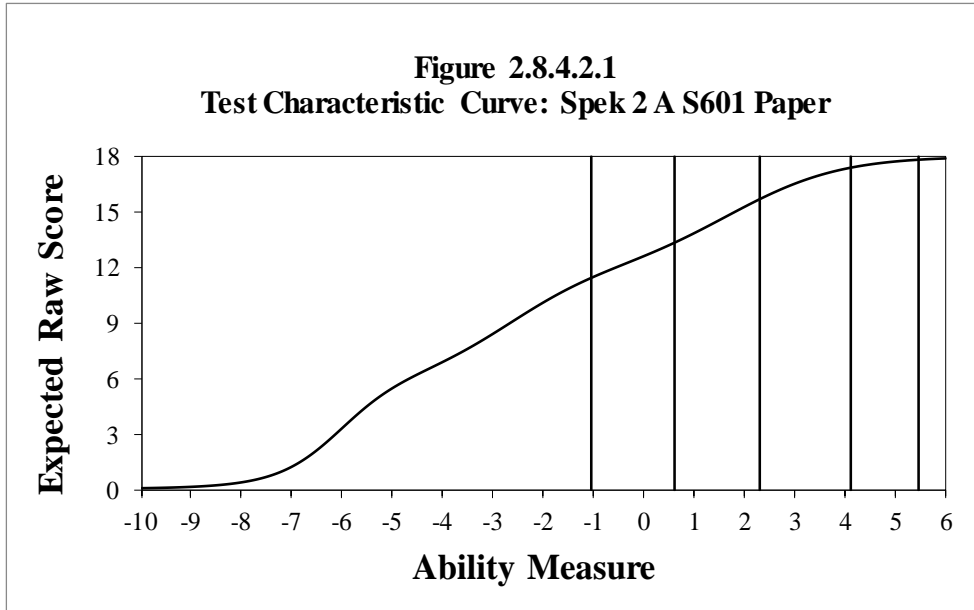
Figure 2.8.4.0
Test Characteristic Curve: Spek K S601 Paper

2.8.4.1 Grade 1

Figure 2.8.4.1.1
Test Characteristic Curve: Spek 1 A S601 Paper

Figure 2.8.4.1.2
Test Characteristic Curve: Spek 1 B/C S601 Paper

2.8.4.2 Grade 2

Figure 2.8.4.2.1
Test Characteristic Curve: Spek 2 A S601 Paper

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

Figure 2.8.4.2.2
Test Characteristic Curve: Spek 2 B/C S601 Paper
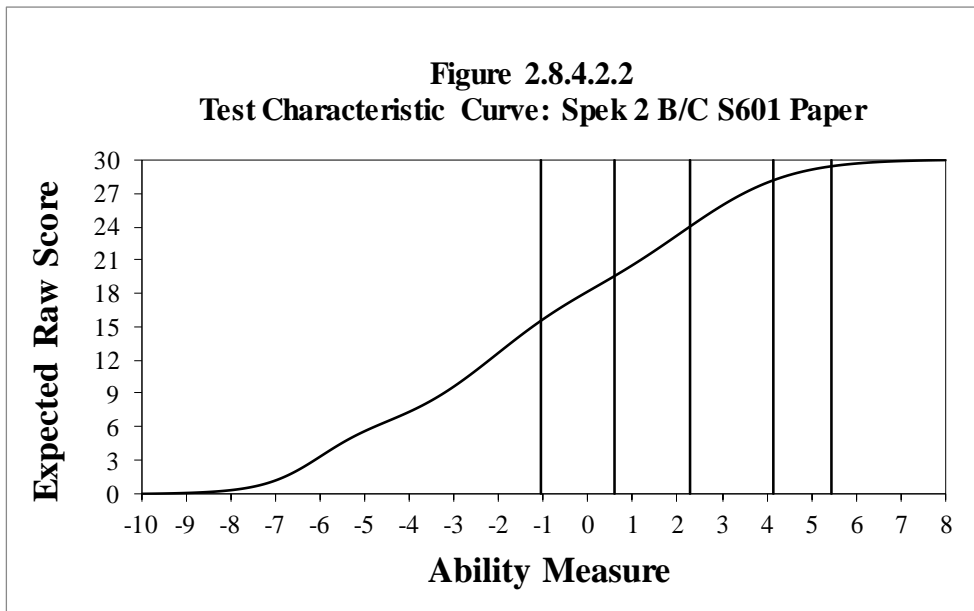
Note: The test form is shared between 2B/C and 3B/C.

2.8.4.3 Grade 3
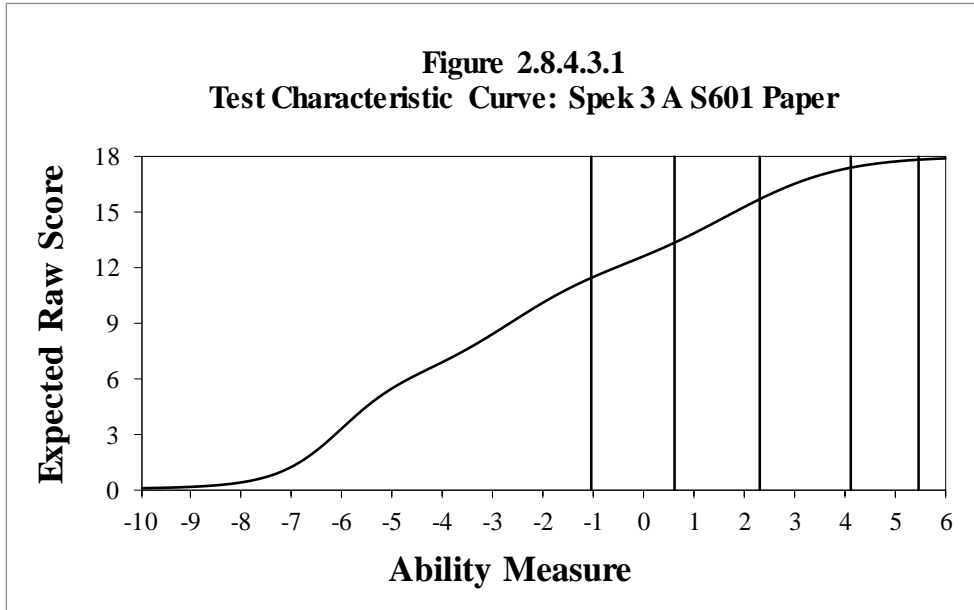
Figure 2.8.4.3.1
Test Characteristic Curve: Spek 3 A S601 Paper

Expected Raw Score vs *Ability Measure*

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

Figure 2.8.4.3.2
Test Characteristic Curve: Spek 3 B/C S601 Paper

Expected Raw Score vs *Ability Measure*
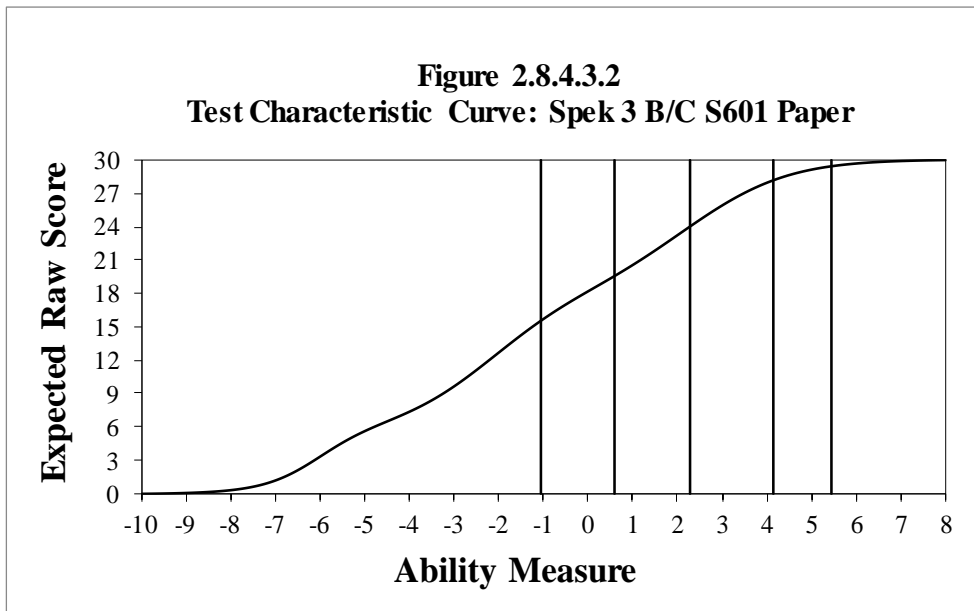
Note: The test form is shared between 2B/C and 3B/C.

2.8.4.4 Grades 4-5

Figure 2.8.4.4.1
Test Characteristic Curve: Spek 4-5 A S601 Paper

Figure 2.8.4.4.2
Test Characteristic Curve: Spek 4-5 B/C S601 Paper

2.8.4.5 Grades 6-8

Figure 2.8.4.5.1
Test Characteristic Curve: Spek 6-8 A S601 Paper

Figure 2.8.4.5.2
Test Characteristic Curve: Spek 6-8 B/C S601 Paper

2.8.4.6 Grades 9-12

Figure 2.8.4.6.1
Test Characteristic Curve: Spek 9-12 A S601 Paper

Figure 2.8.4.6.2
Test Characteristic Curve: Spek 9-12 B/C S601 Paper
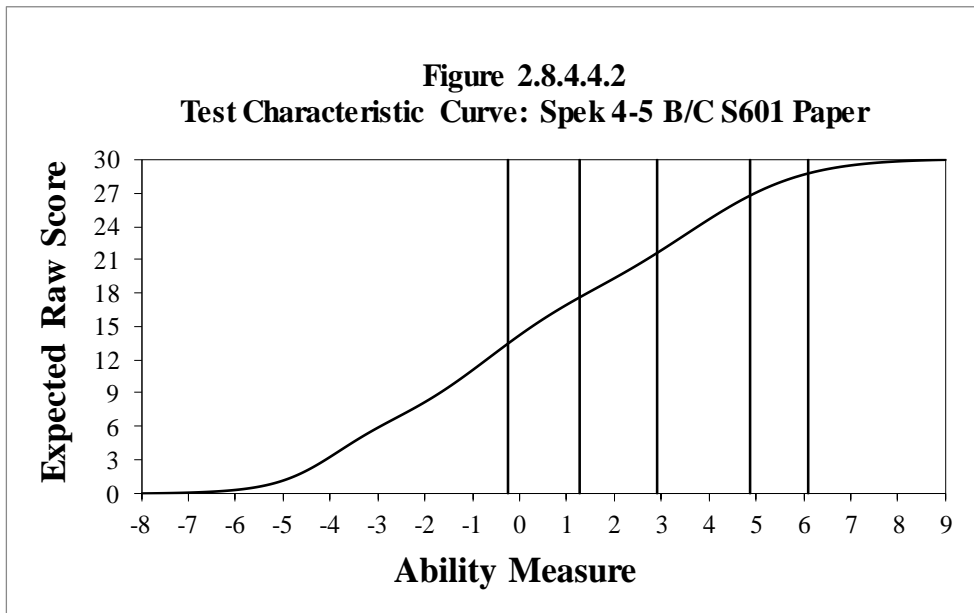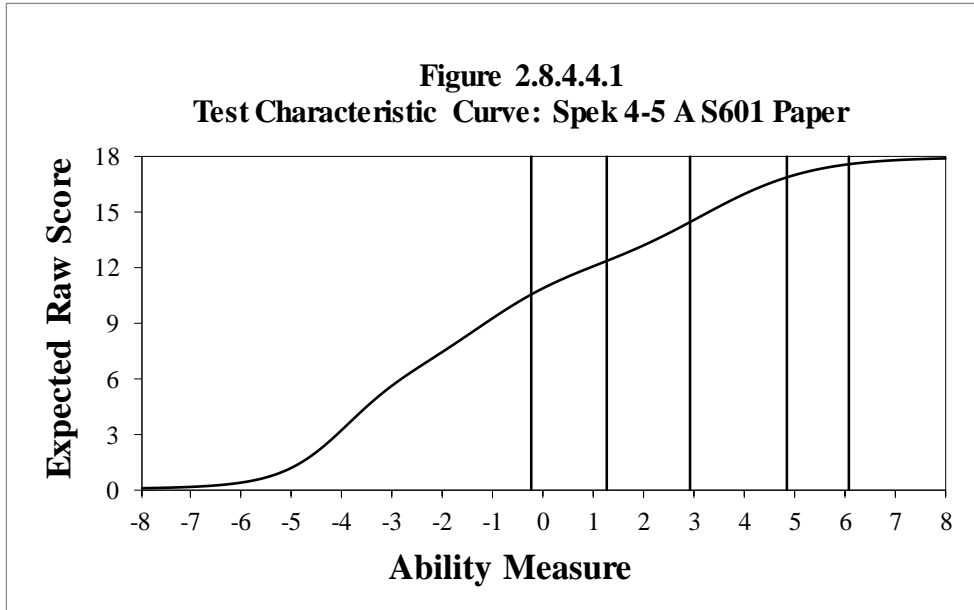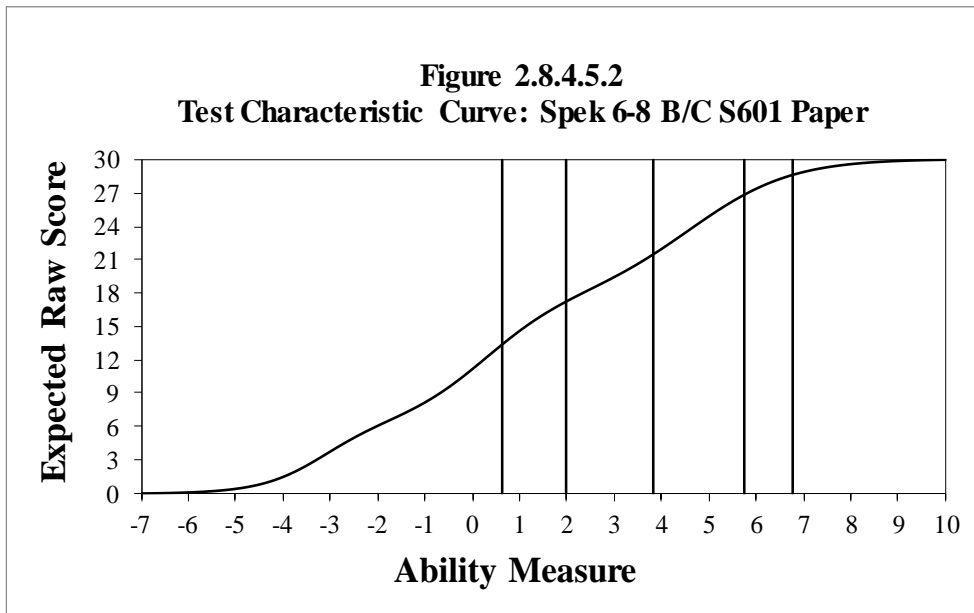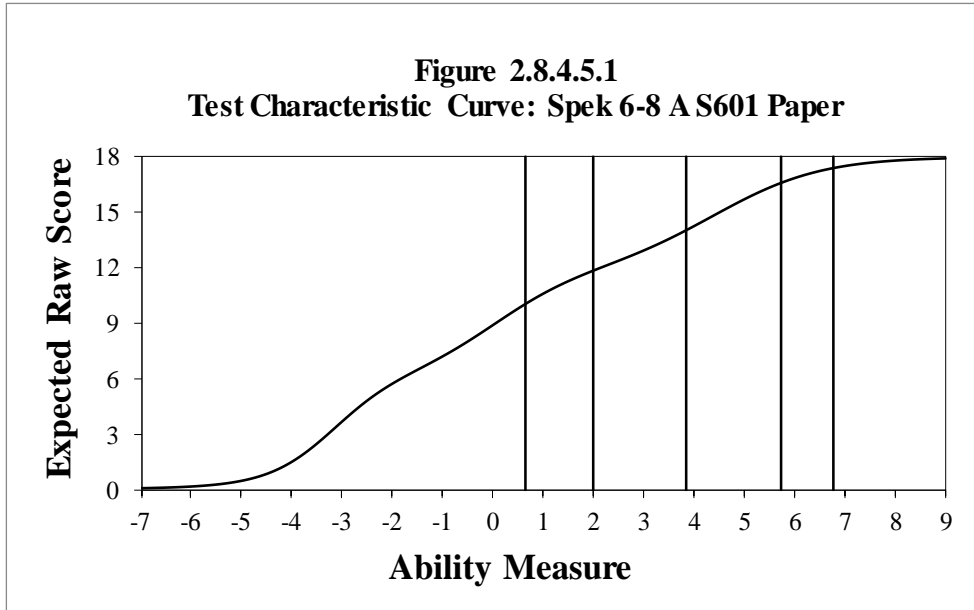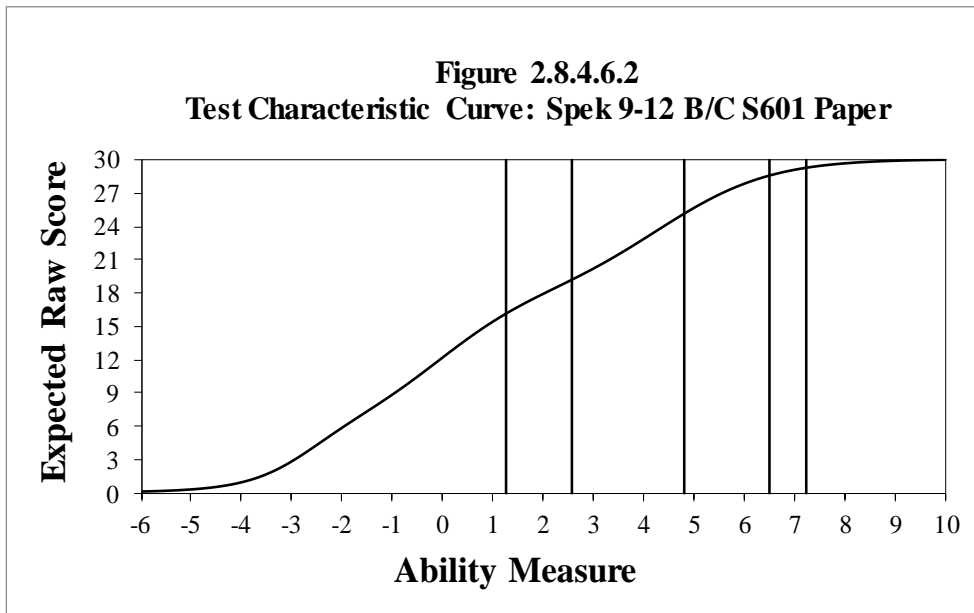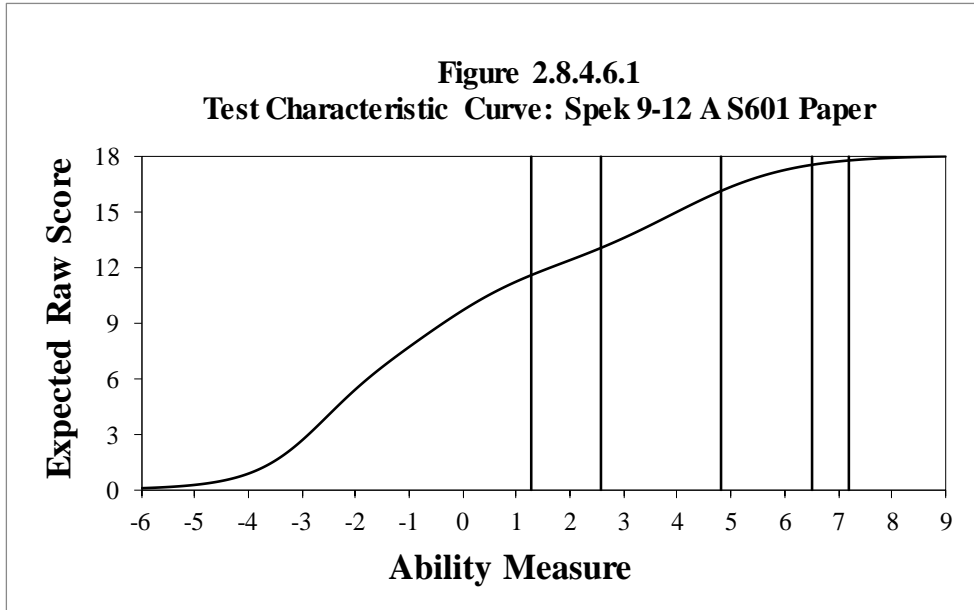
2.9   Test Information Function

With the Rasch measurement model, as with any measurement model following item response theory, one can use the item information function (Lord, 1980) to model the relationship between the ability measure (in logits) and the accuracy of the ability measure by item. The item information function indicates the amount of information we have about the ability estimate provided by the item, as a function of the ability level. The more information we have about the ability estimate, the more certain or confident we are about the ability estimate. If the amount of information is large, we can estimate the student whose true ability is at that level with a higher degree of certainty, and all the estimates will be close to the true values. Conversely, if the amount of information is small, we can estimate the student whose true ability is at that level with a lower degree of certainty, and estimates will be further away from the true values.

The **item/task information function** indicates the amount of information student responses to that item (or task) provides to help reduce our uncertainty regarding a student's true ability measure. The more information we have about the ability measure, the more certain or confident we can be in that estimate of the student's ability. If the amount of information is large, that means that we have estimated with a higher degree of certainty a student whose true ability is at that level. Therefore, the ability measures for students whose scores lie within that region of the ability continuum will be close to their true values. Conversely, if the amount of information is small, that means that we have estimated with a lower degree of certainty the student whose true ability is at that level. Consequently, the ability measures for students whose scores lie within that region of the ability continuum will be further away from their true values.

Mathematically, the amount of item information at a given ability level is the reciprocal of the variance of the ability estimate at the level for the item. In other words, item information value is the inverse squared of the standard errors of measurement of a given ability measure for the item. Therefore, for that item (or task), the information value also provides information about the precision of the ability measure along the ability continuum.

The **test information function** (TIF) aggregates the item information functions across all the items on the test form or item pool. Since the item information value is the inverse squared of the standard errors of measurement of a given ability measure for the item, the test information value reflects the standard errors of measurement of a given ability level for the test. When the TIF is presented graphically as the test information curve, it shows how well the test is measuring

across the continuum of student ability in terms of the amount of information, certainty, or the amount of measurement precision the test provides at each ability level. The higher the curve, the more information the test provides at the ability level.

Since the TIF is the sum of all item/task information functions on the test form (Lord, 1980), the TIF depends on the information functions (Lord, 1980) of the individual items/tasks included on the test form or in the item pool. The shape of the test information curve depends on several factors, including the number and characteristics of items/tasks, the item response theory used, and the values of the item/task parameters. With some exceptions, there is a general pattern to the shape of test information curves. Test information curves peak in the region of the student ability continuum where the test provides higher discrimination and more precise measurement as compared to other regions where the curve is less peaked, normally at the lower and upper ends of the ability continuum. When the test form consists of multiple-choice items such as on the Listening and Reading domains, the test information curve is usually unimodal.

The parameter values for the individual categories on the scoring tools that raters use to evaluate students' responses to the tasks, in addition to the factors mentioned earlier, affect the shape of the test information curves for the Writing and Speaking tests. Accordingly, some refer to these test information curves as "category information functions" (Engelhard & Wind, 2018). The scoring scales that the raters use have more score categories than the scoring schemes used for evaluating students' responses to multiple-choice items, which typically have just two categories— "right" or "wrong." Additionally, we designed the scoring scales to measure a wide range of student performance on a task. Consequently, the resulting adjacent score category boundaries may not be equidistant, and, indeed, in some cases, they may even be far apart if raters assign few scores in certain categories. In this situation, a test information curve will have one (or more) dips in the region(s) between the adjacent score category boundaries, indicating the loss of information in the corresponding ability range(s) and a decrease in the amount of information that certain score categories provide (Engelhard & Wind, 2018). Therefore, the shape of a test information curve for an ACCESS Writing or Speaking test may not be unimodal and instead may have two (or more) peaks. For example, suppose that a test information curve reveals a dip in the region of the student writing ability continuum where raters would have assigned a score of 3. That suggests that students who received a score of 3 may have displayed potentially substantively meaningful differences in writing ability that the raters were not able to adequately distinguish when they used the 9-point Writing Scoring Scale to assign scores or,

alternatively, that the score categories did not describe salient characteristics of students' writing that would make it possible for the raters to distinguish reliability among the students' responses in that region of the student ability continuum (Engelhard & Wind, 2018, pp. 316-319). The ACCESS Writing and Speaking tests are not the only assessments that have test information curves with these unusual shapes. The test information curves for other tests composed of open-ended tasks, such as the National Assessment of Educational Progress Writing assessment, also show a similar "dipping" pattern (Muraki, 1993).

The figures in this section plot the TIFs and show graphically the amount of information that the test provided across the continuum of student ability. For each test form, the five vertical lines in the figure indicate the ACCESS cut scores for the highest grade in each grade-level cluster, dividing the figure into six sections denoting the WIDA proficiency levels (PL 1– PL 6) for the domain. The test information curve and the corresponding ACCESS cut-score lines are both expressed on the ACCESS logit scale.

Inclusion of the ACCESS cut-score lines in these figures is meant only to facilitate the visual interpretation of the test information curves relative to the ACCESS cut scores by domains. These lines provide a benchmark for WIDA and CAL assessment experts to examine the ability range for which each test seems to be more (or less) accurate in estimating student ability. Readers should note that most states that use ACCESS for ELLs do not make reclassification decisions based solely on students' domain scale scores. Rather, the majority of these states set their reclassification (or exit) criterion based on a student's Overall composite scale score, which is a weighted sum of a student's four domain scale scores. Only a few states use those four domain scale scores in addition to the student's Overall composite scale score when making a reclassification decision. Therefore, from the WIDA policy perspective, it is more important to ensure that we minimize the measurement error near the cut score that most states use to set their reclassification criterion on the Overall composite scale score. We report the conditional standard errors of measurement (CSEMs) for the Overall composite scale scores in Section 5.6.

In addition to the TIF graphs by tier, we provide plots of the TIFs across tiers, by grade-level cluster, in the same graph. Test users may find it useful to compare the ability ranges across tiers, noting for each tier where the curve displays its highest peaks (i.e., where the most measurement information is provided). For example, as shown in Figure 2.9.3.1.3, the test information curve across tiers for Writing Grade 1 reveals that the Writing Grade 1 Tier A form provided more

information about student ability measures that were either just below the PL 2 cut score or just below the PL 5 cut score. By contrast, the Writing Grade 1 Tier B/C form provided more information about the student ability measures that were either just above the PL 2 cut score or just above the PL 5 cut score. The plot also shows that the Writing Grade 1 Tier A form provided more information for those student ability measures in the lowest range (i.e., ability measures of -0.5 logits or lower), while the Writing Grade 1 Tier B/C form provided more information than the Grade 1 Tier A form for the rest of the student ability measures, especially those in the higher ability range. Lastly, consistent with the purposes of the test design, there is also considerable overlap between the ranges of writing ability that the two forms cover.

2.9.1 Listening

2.9.1.0 Kindergarten
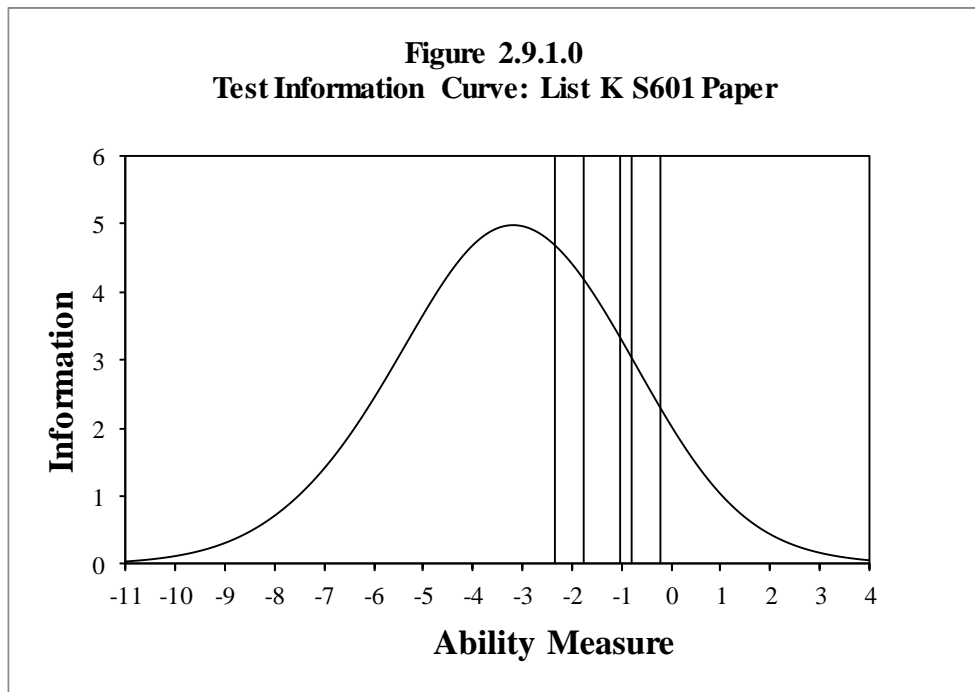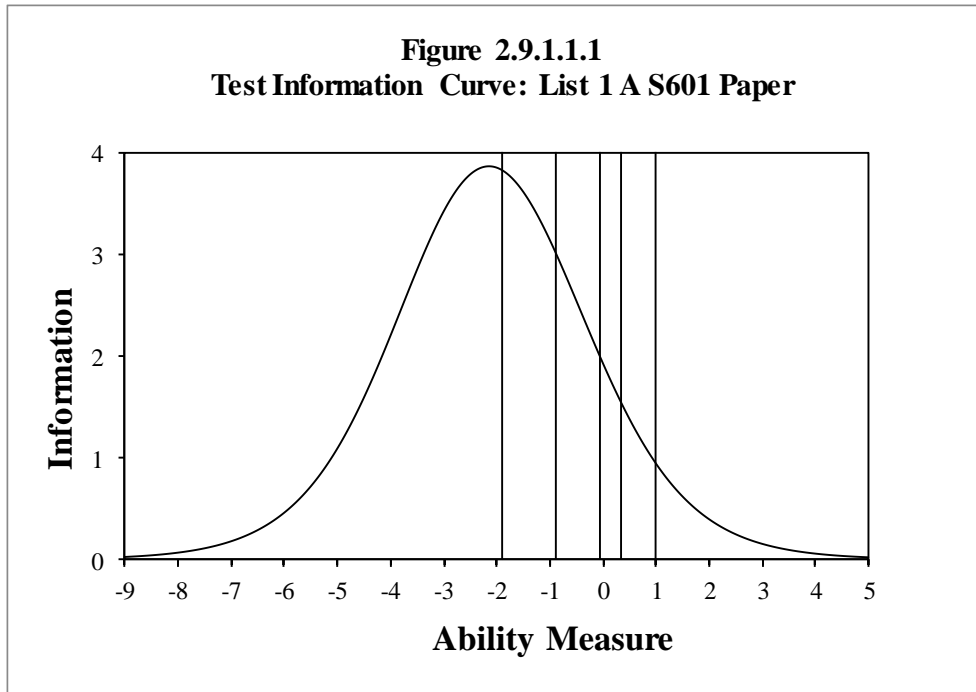
Figure 2.9.1.0
Test Information Curve: List K S601 Paper

2.9.1.1 Grade 1

Figure 2.9.1.1.1
Test Information Curve: List 1 A S601 Paper

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

Figure 2.9.1.1.2
Test Information Curve: List 1 B/C S601 Paper
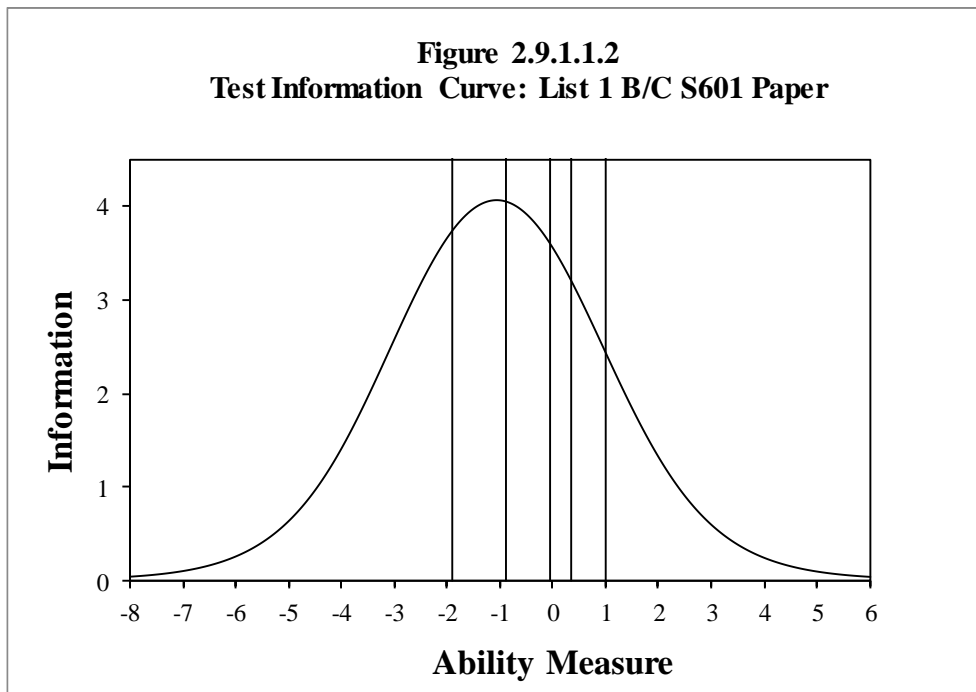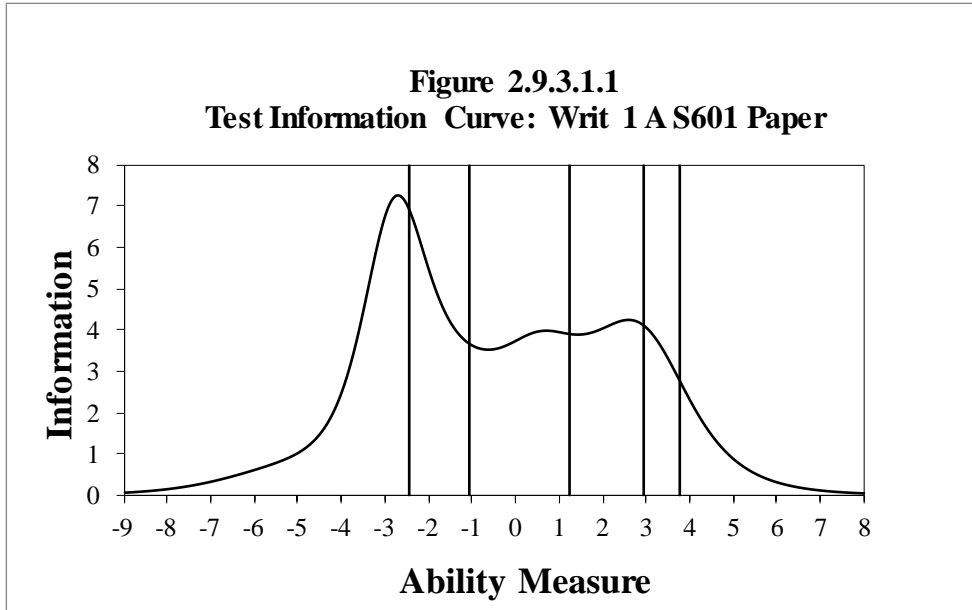

Note: The test form is shared between 1B/C and 2B/C.

Figure 2.9.1.1.3
Test Information Curve: List 1 S601 Paper
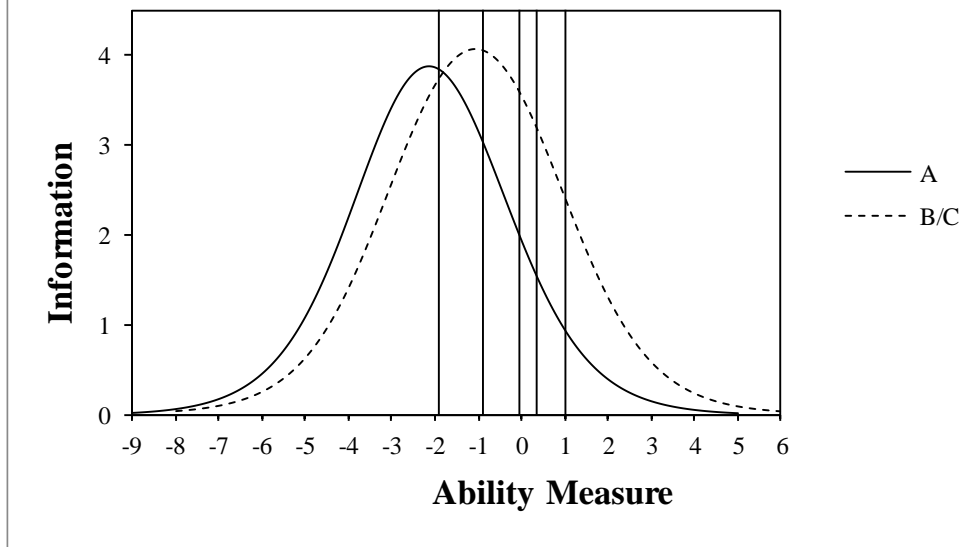
Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

2.9.1.2 Grade 2
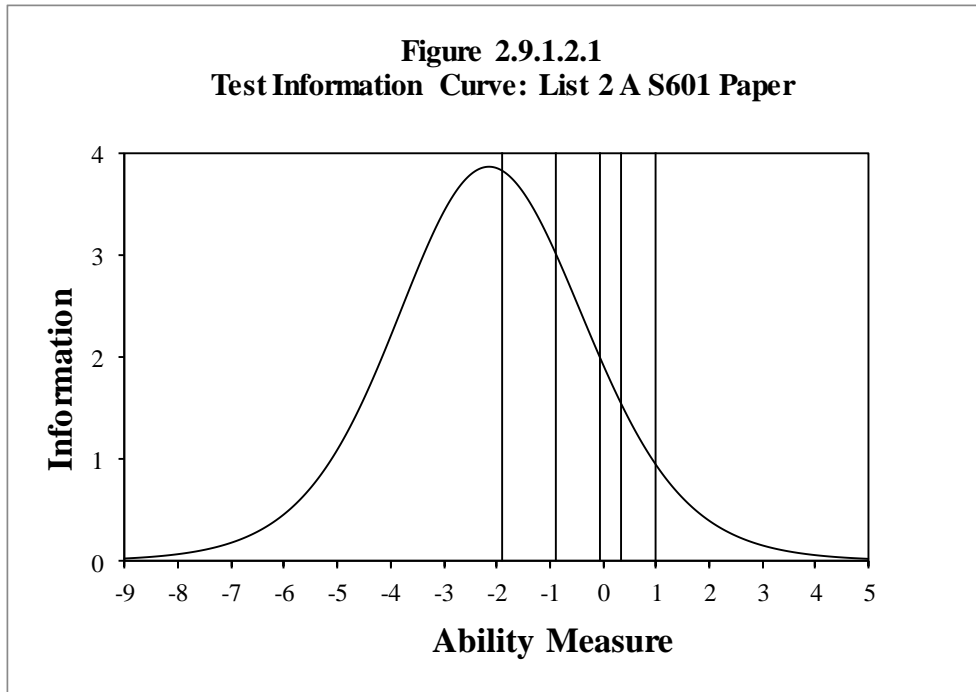
Figure 2.9.1.2.1
Test Information Curve: List 2 A S601 Paper

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

Figure 2.9.1.2.2
Test Information Curve: List 2 B/C S601 Paper
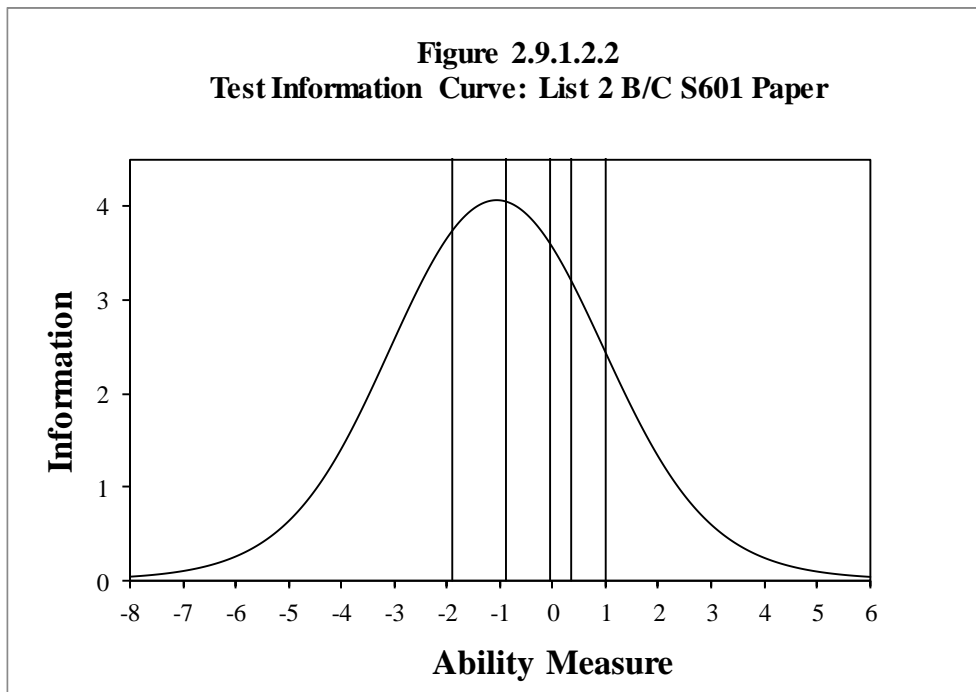
Note: The test form is shared between 1B/C and 2B/C.

Figure 2.9.1.2.3
Test Information Curve: List 2 S601 Paper
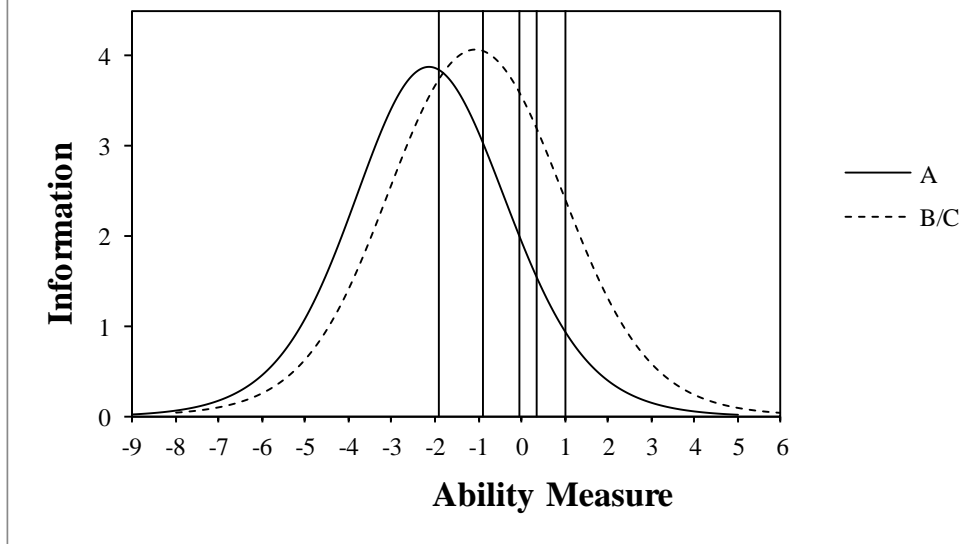
Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

2.9.1.3 Grade 3
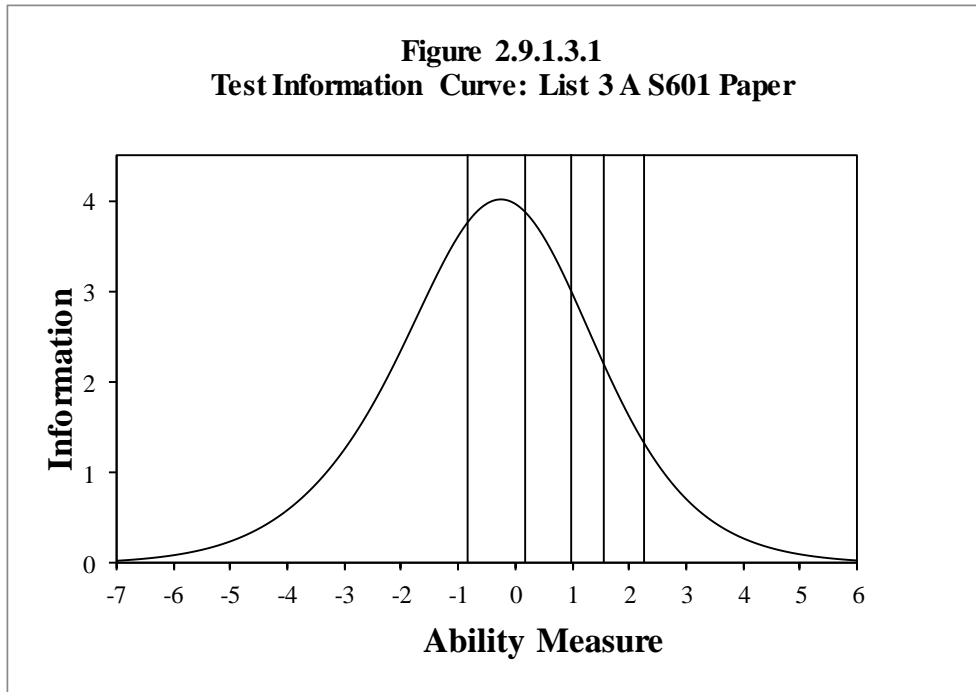
Figure 2.9.1.3.1
Test Information Curve: List 3 A S601 Paper

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

Figure 2.9.1.3.2
Test Information Curve: List 3 B/C S601 Paper
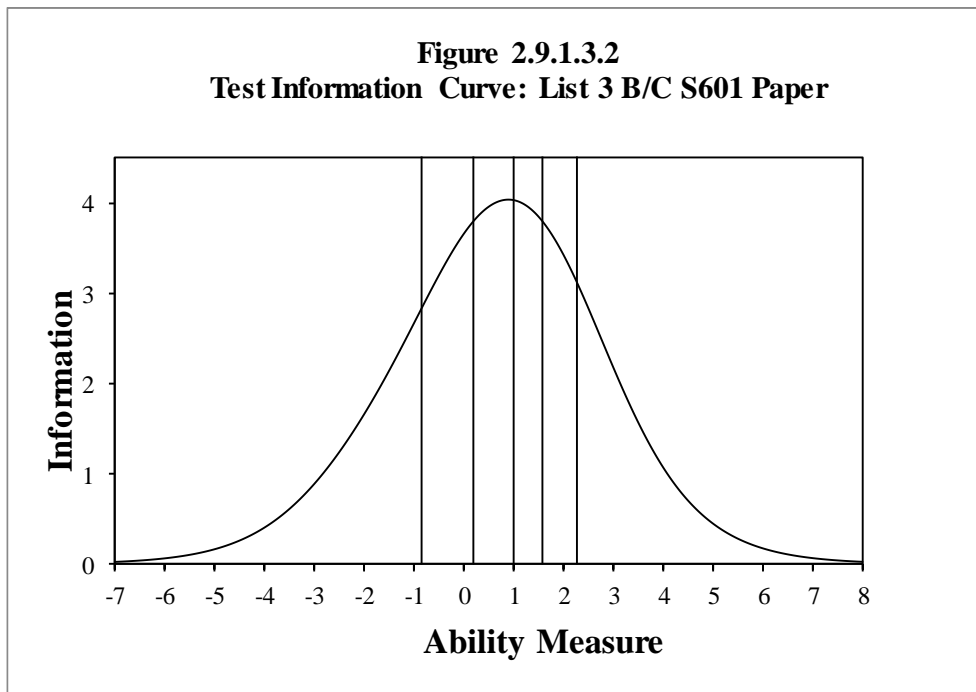
Note: The test form is shared between 3B/C and 4-5B/C.

Figure 2.9.1.3.3
Test Information Curve: List 3 S601 Paper
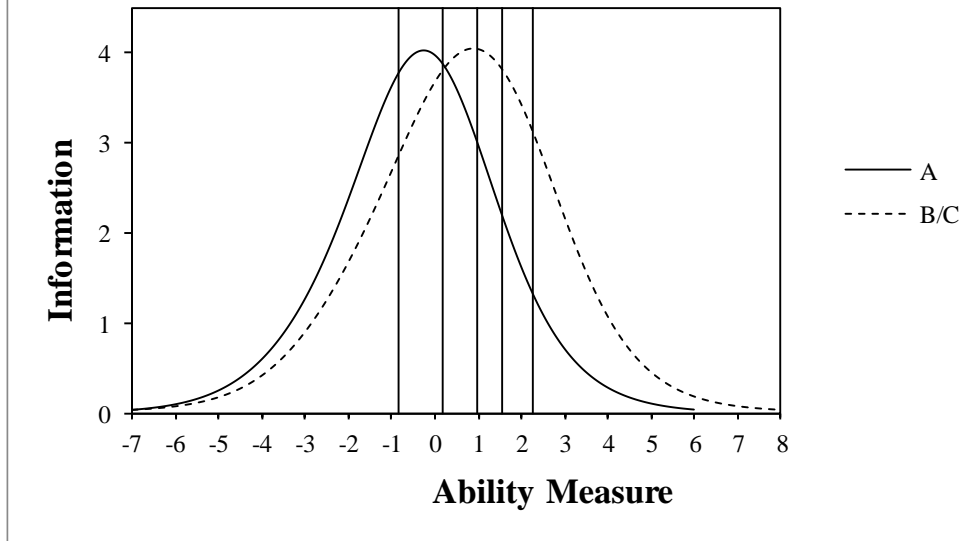
Note: The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

2.9.1.4 Grades 4-5
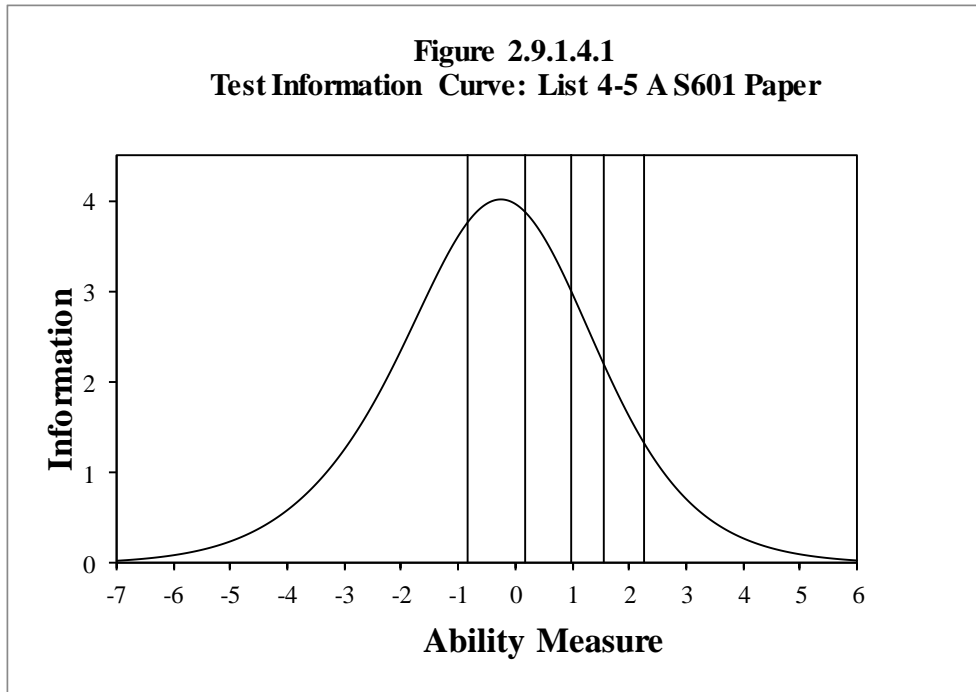
Figure 2.9.1.4.1
Test Information Curve: List 4-5 A S601 Paper

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

Figure 2.9.1.4.2
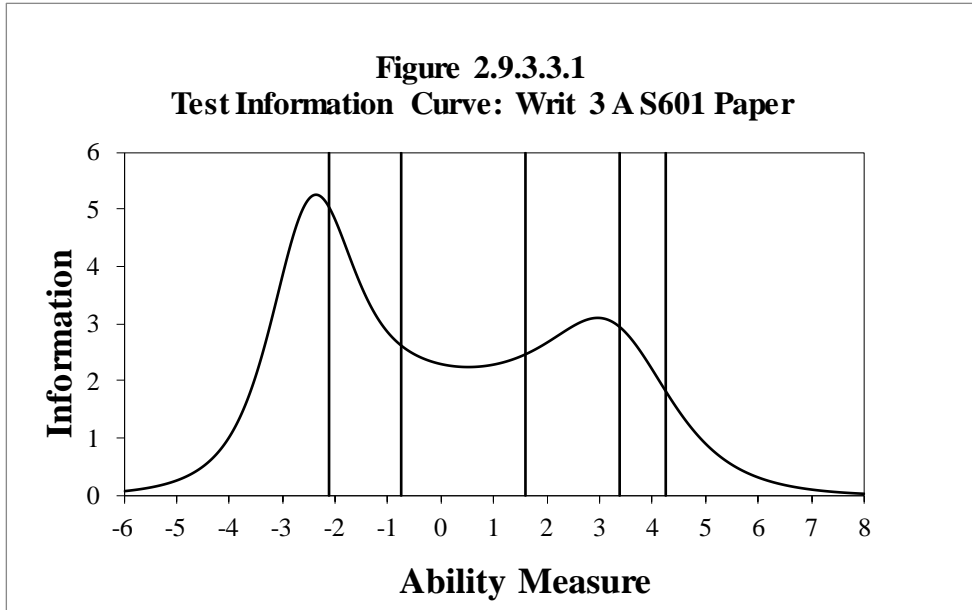**Test Information Curve: List 4-5 B/C S601 Paper**
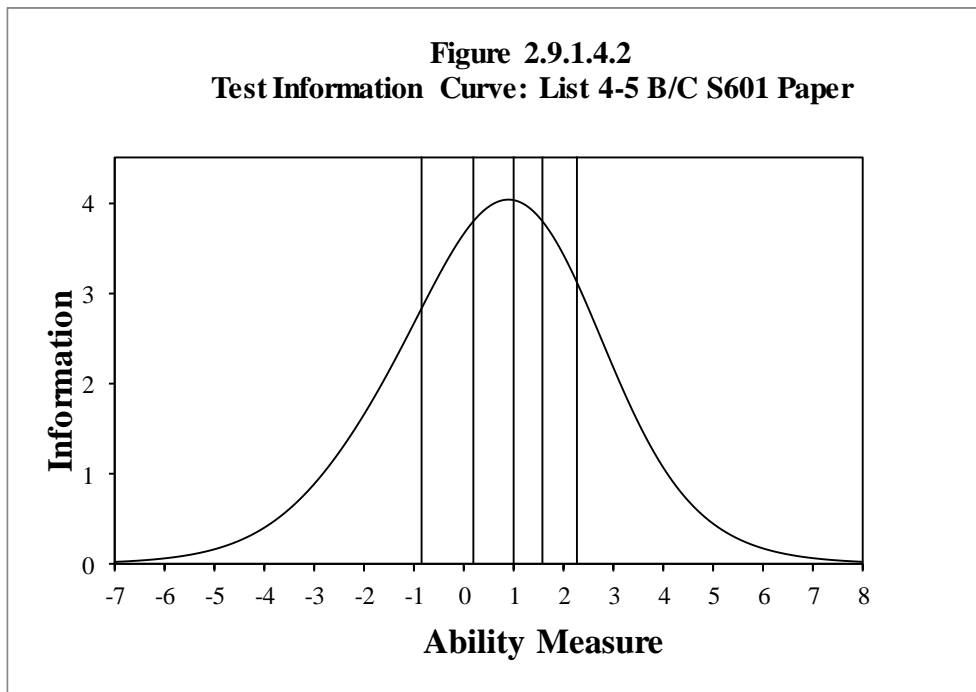
Note: The test form is shared between 3B/C and 4-5B/C.

Figure 2.9.1.4.3
Test Information Curve: List 4-5 S601 Paper
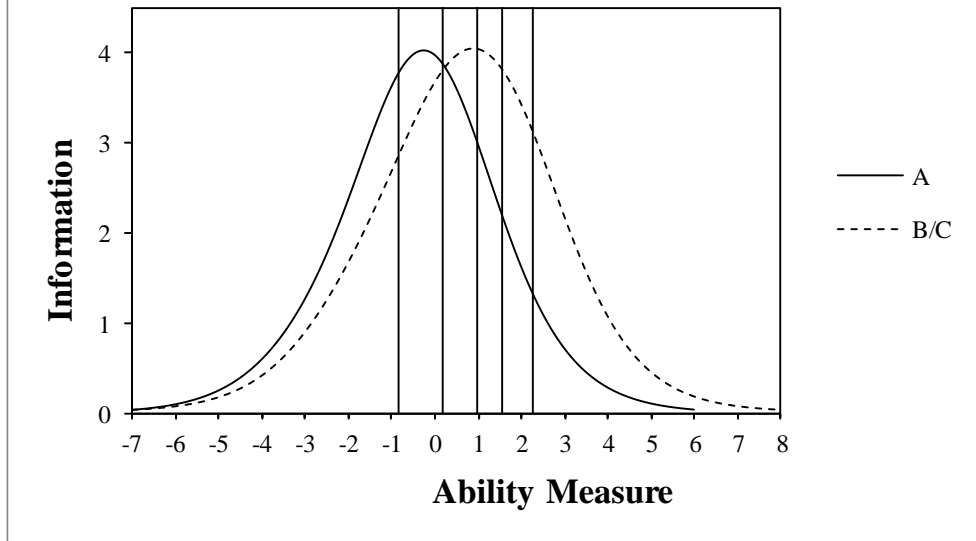
Note: The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

2.9.1.5 Grades 6-8

Figure 2.9.1.5.1
Test Information Curve: List 6-8 A S601 Paper
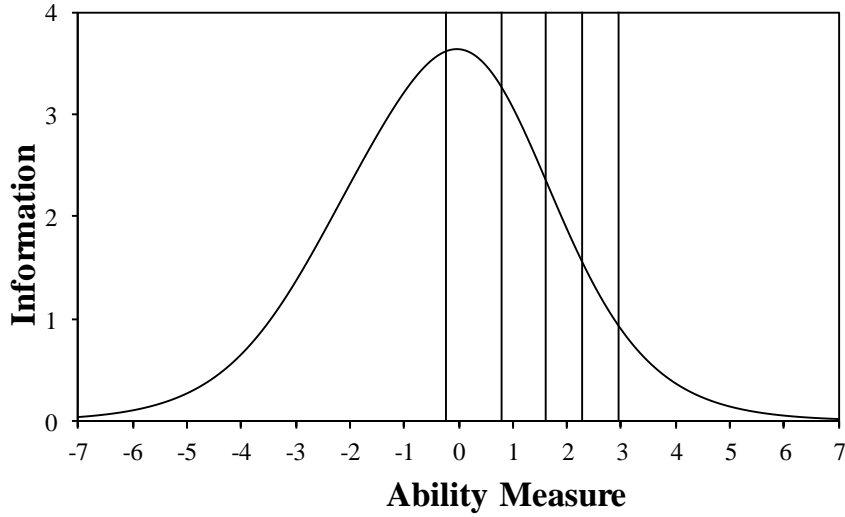
Figure 2.9.1.5.2
Test Information Curve: List 6-8 B/C S601 Paper
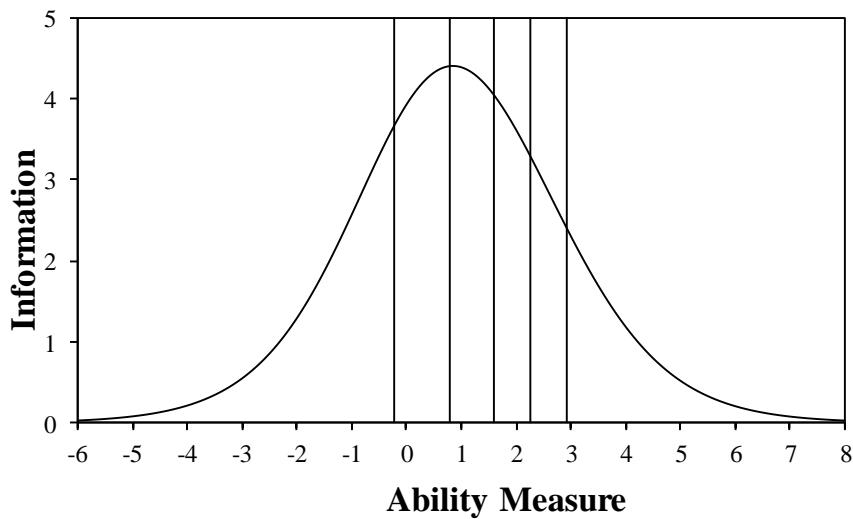
Figure 2.9.1.5.3
Test Information Curve: List 6-8 S601 Paper
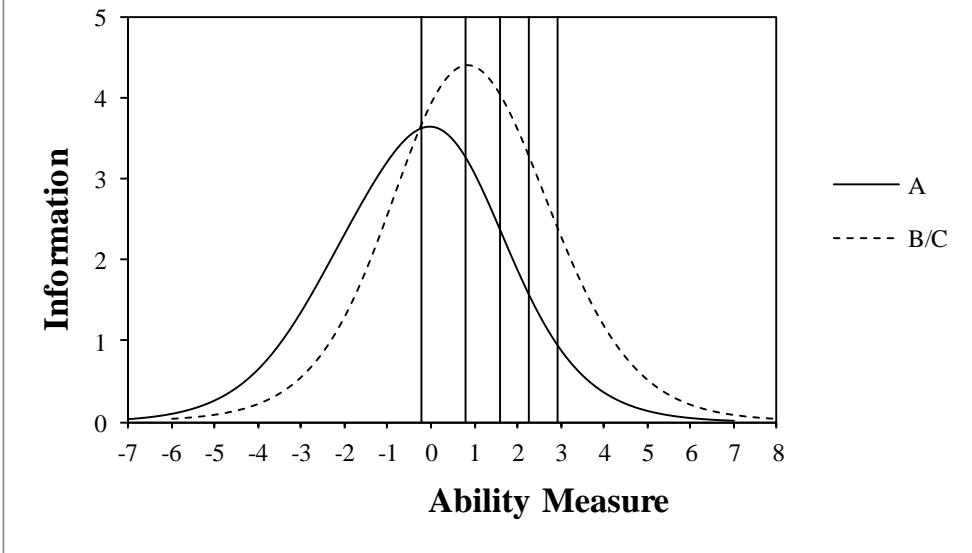
2.9.1.6 Grades 9-12

Figure 2.9.1.6.1
Test Information Curve: List 9-12 A S601 Paper
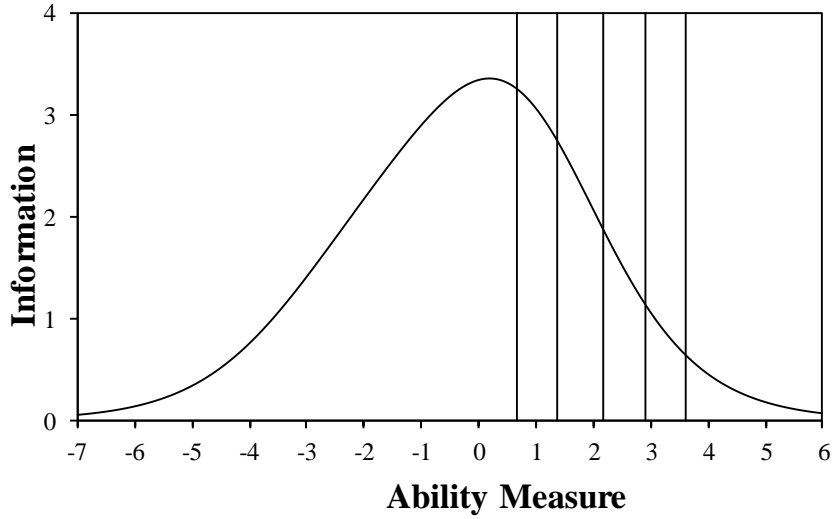
Figure 2.9.1.6.2
Test Information Curve: List 9-12 B/C S601 Paper
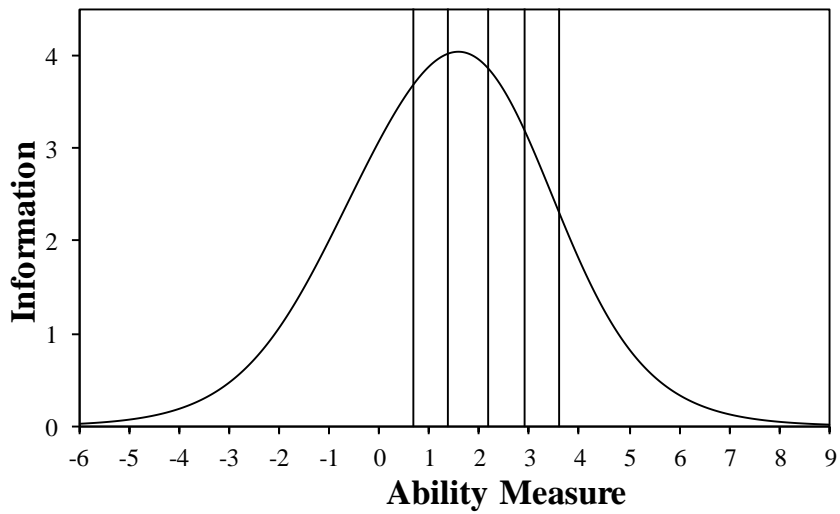
Figure 2.9.1.6.3
Test Information Curve: List 9-12 S601 Paper
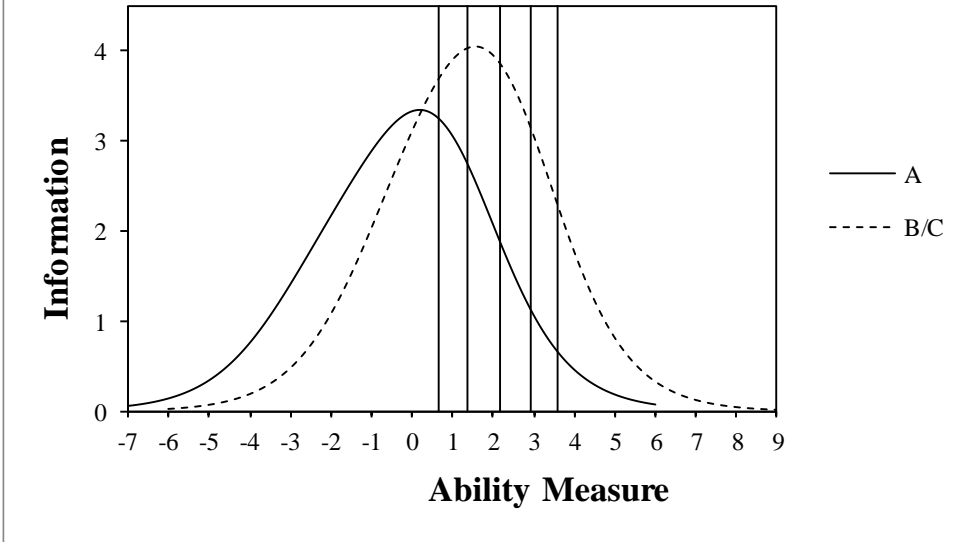
2.9.2 Reading

2.9.2.0 Kindergarten
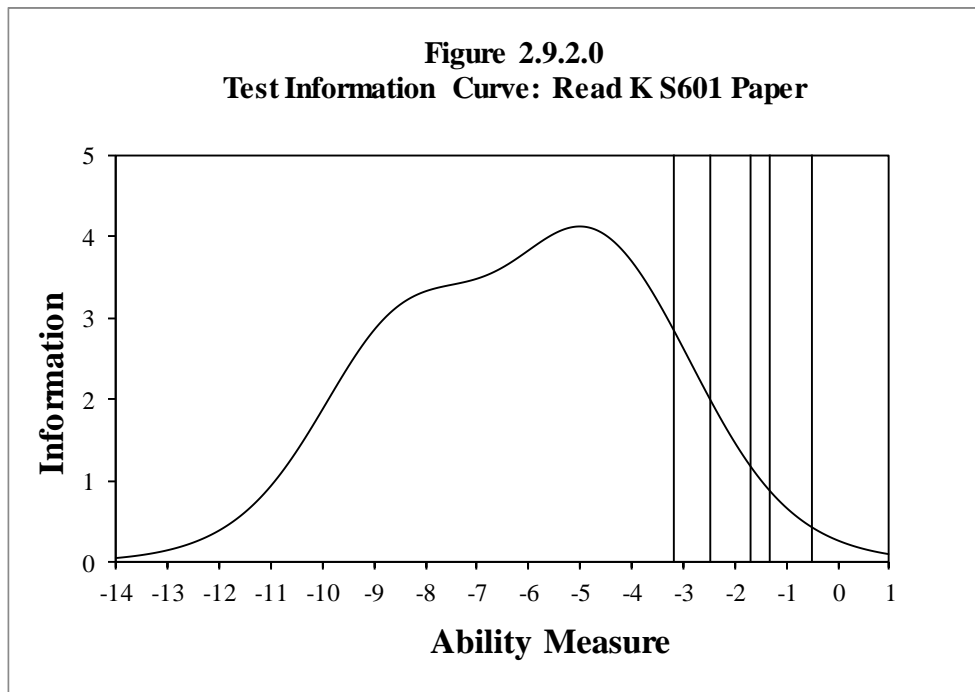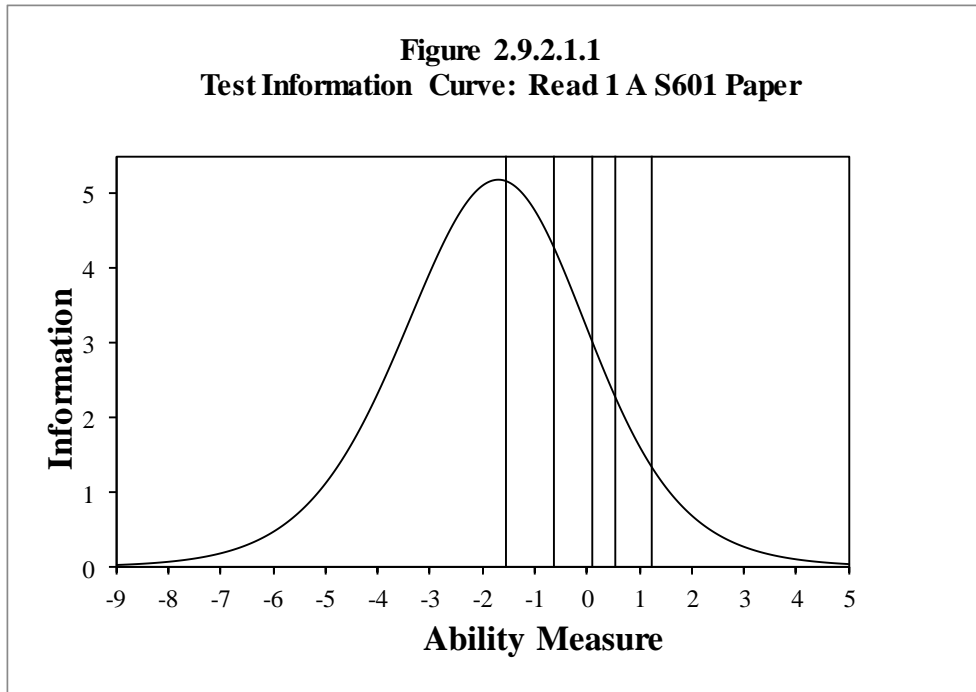
Figure 2.9.2.0
Test Information Curve: Read K S601 Paper

2.9.2.1 Grade 1

Figure 2.9.2.1.1
Test Information Curve: Read 1 A S601 Paper

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

Figure 2.9.2.1.2
Test Information Curve: Read 1 B/C S601 Paper
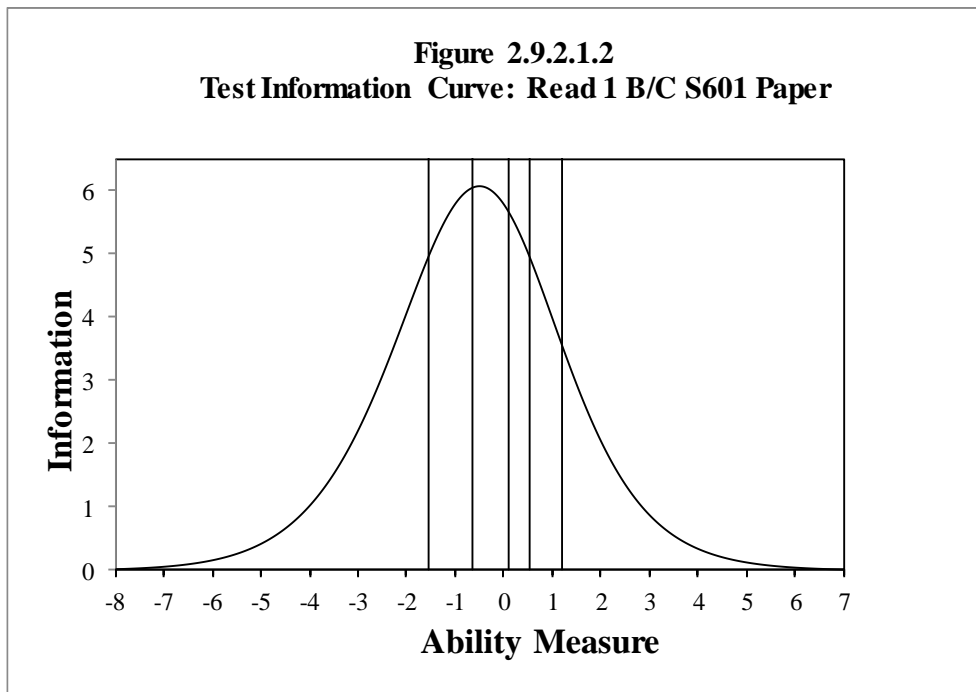
Note: The test form is shared between 1B/C and 2B/C.

Figure 2.9.2.1.3
Test Information Curve: Read 1 S601 Paper
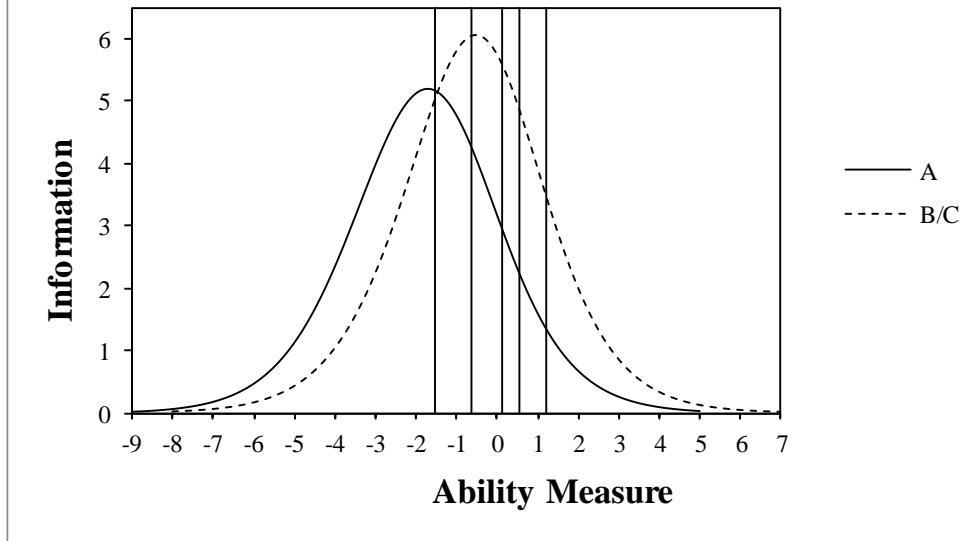
Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.

2.9.2.2 Grade 2
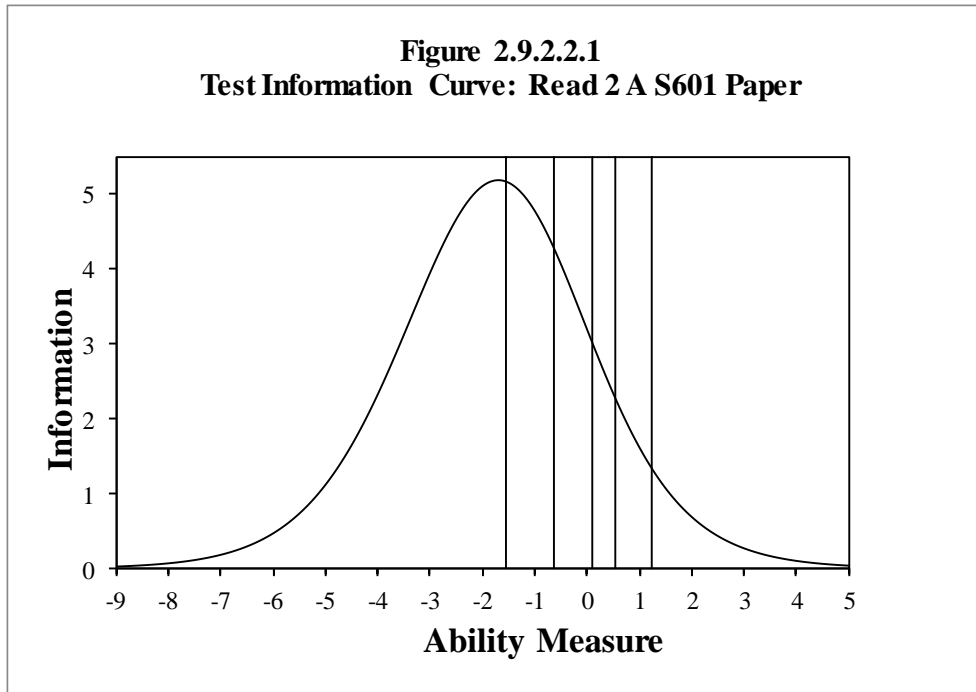
Figure 2.9.2.2.1
Test Information Curve: Read 2 A S601 Paper

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

Figure 2.9.2.2.2
Test Information Curve: Read 2 B/C S601 Paper
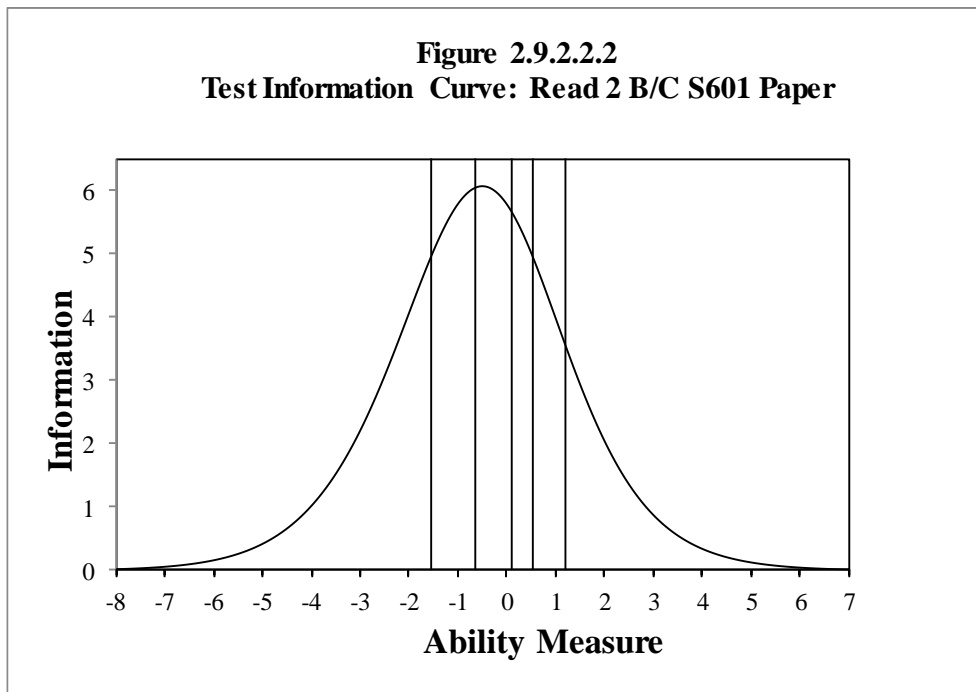
Note: The test form is shared between 1B/C and 2B/C.

Figure 2.9.2.2.3
Test Information Curve: Read 2 S601 Paper
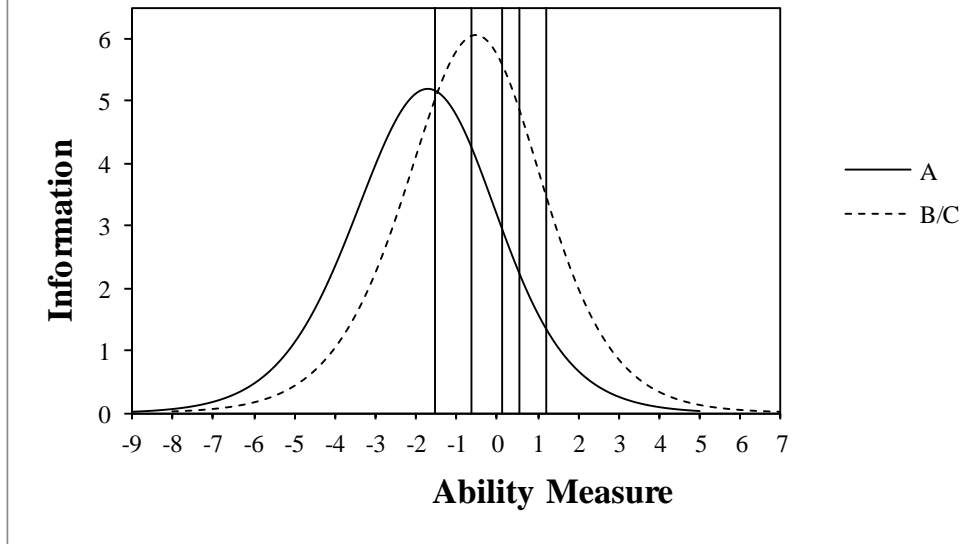
Note: The test form is shared between 1A and 2A, 1B/C and 2B/C.
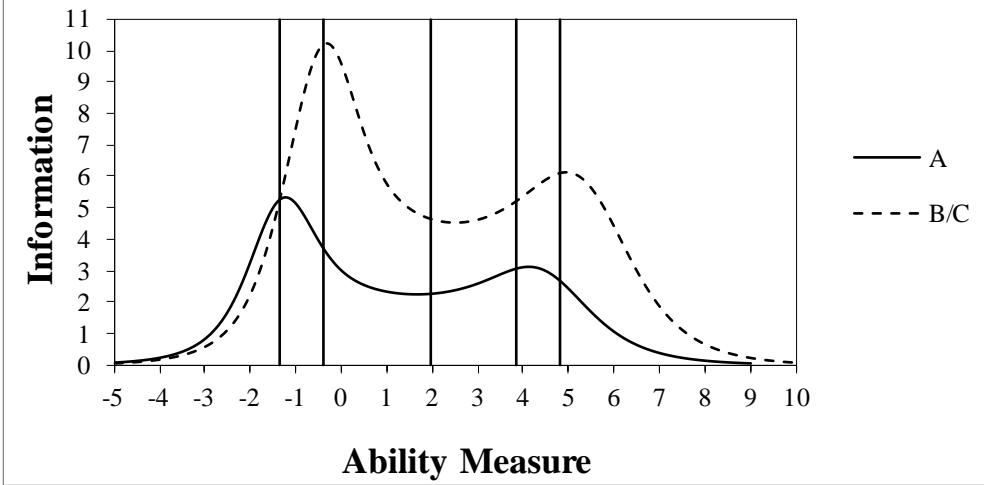
2.9.2.3 Grade 3
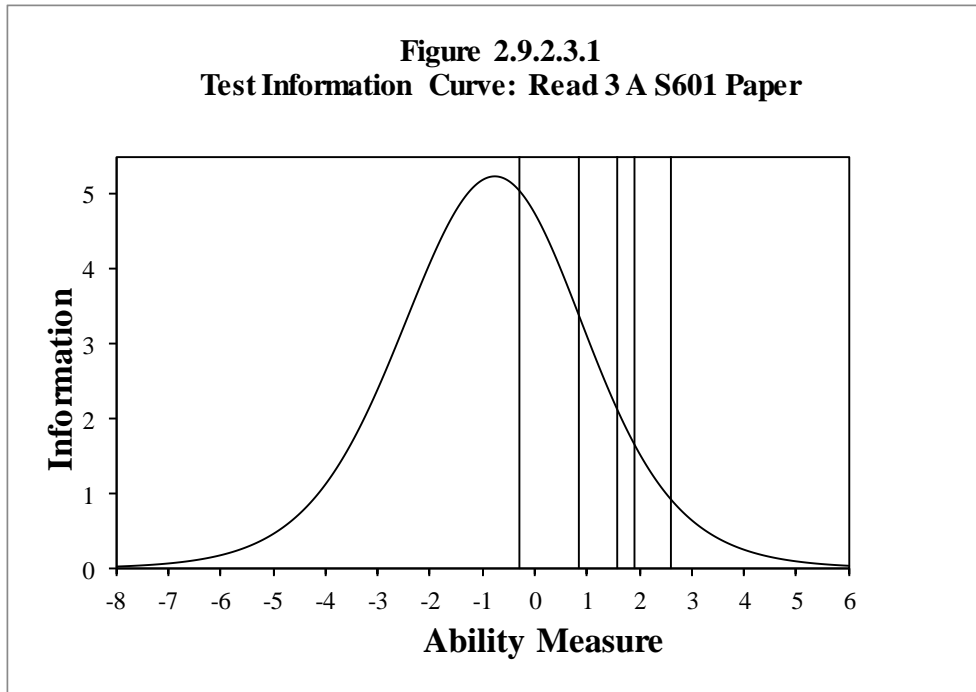
Figure 2.9.2.3.1
Test Information Curve: Read 3 A S601 Paper

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

Figure 2.9.2.3.2
Test Information Curve: Read 3 B/C S601 Paper
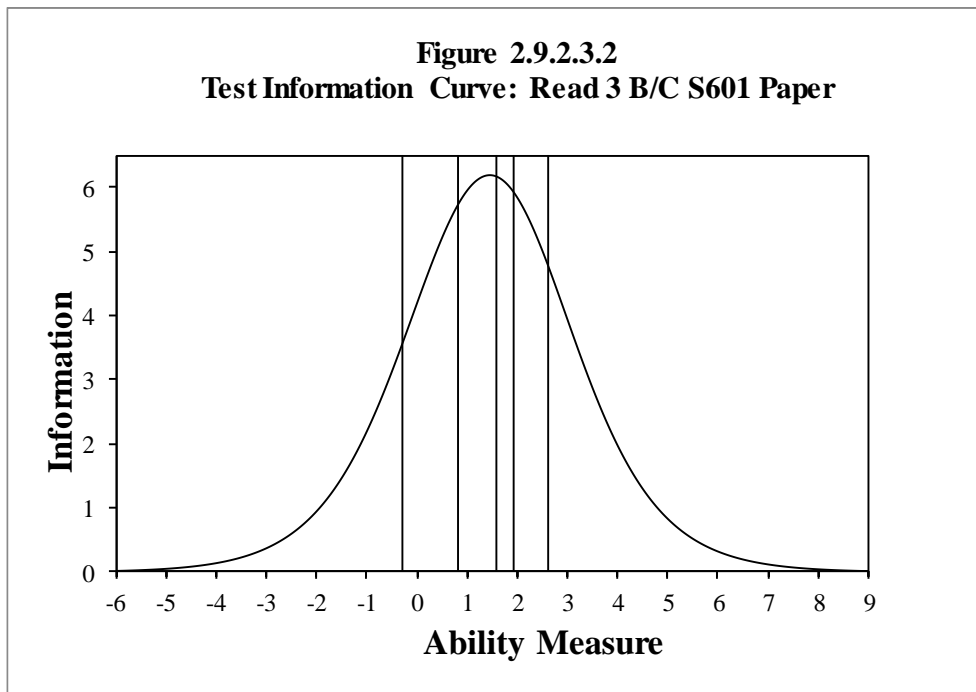
Note: The test form is shared between 3B/C and 4-5B/C.

Figure 2.9.2.3.3
Test Information Curve: Read 3 S601 Paper
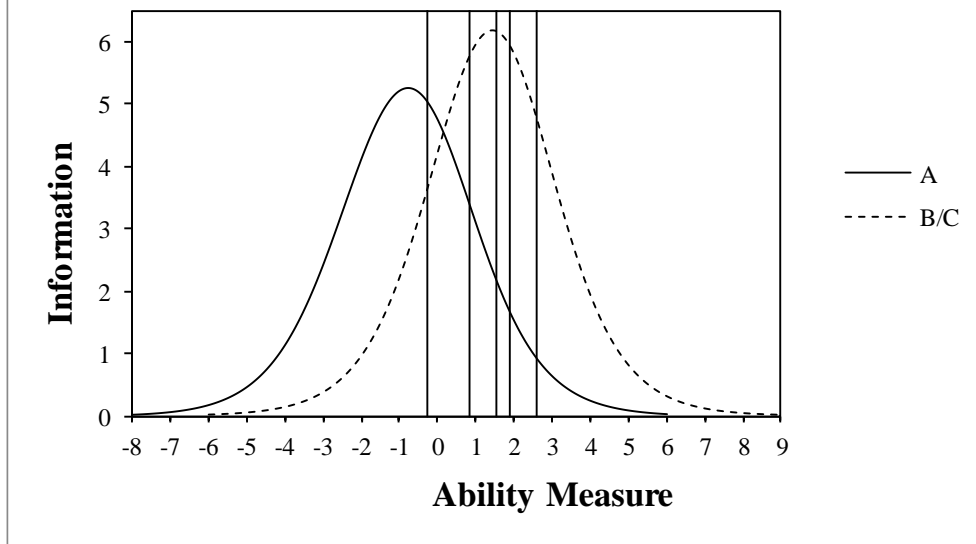
Note: The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

2.9.2.4 Grades 4-5
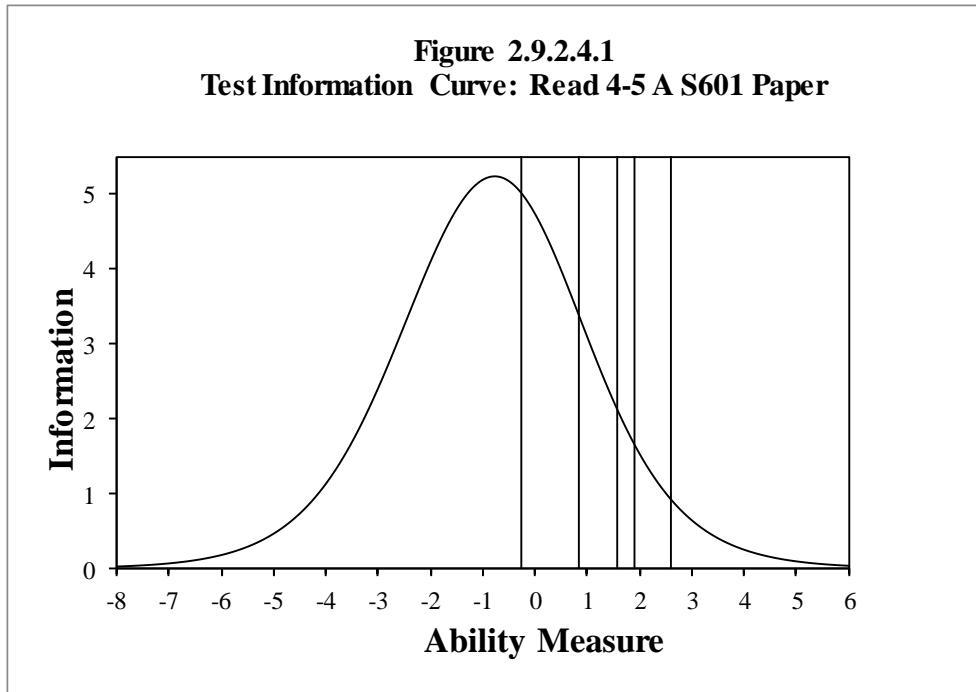
Figure 2.9.2.4.1
Test Information Curve: Read 4-5 A S601 Paper

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

Figure 2.9.2.4.2
Test Information Curve: Read 4-5 B/C S601 Paper
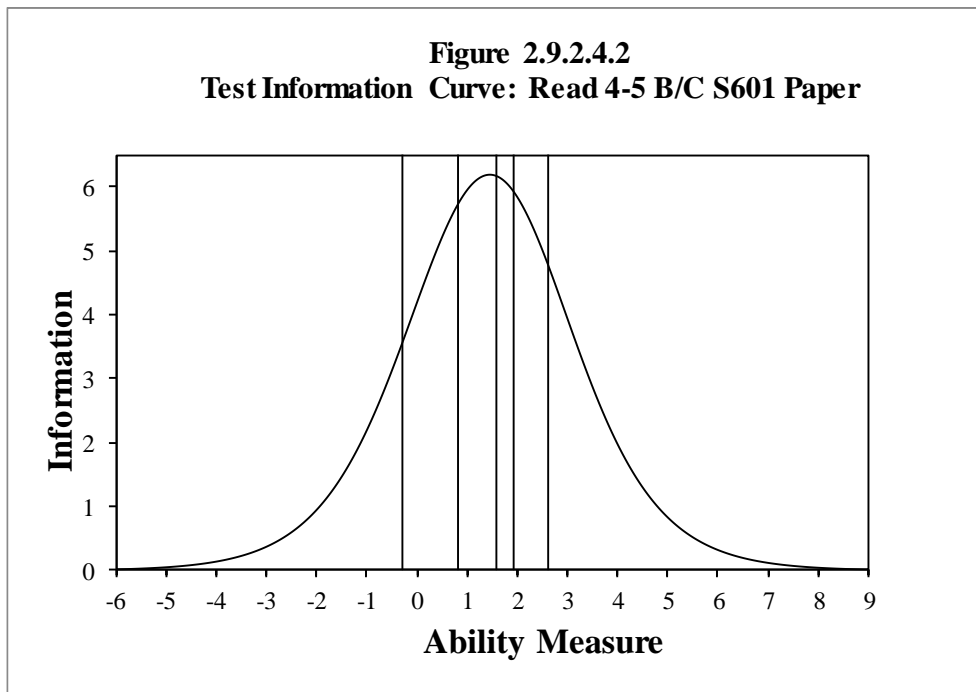
Note: The test form is shared between 3B/C and 4-5B/C.

Figure 2.9.2.4.3
Test Information Curve: Read 4-5 S601 Paper
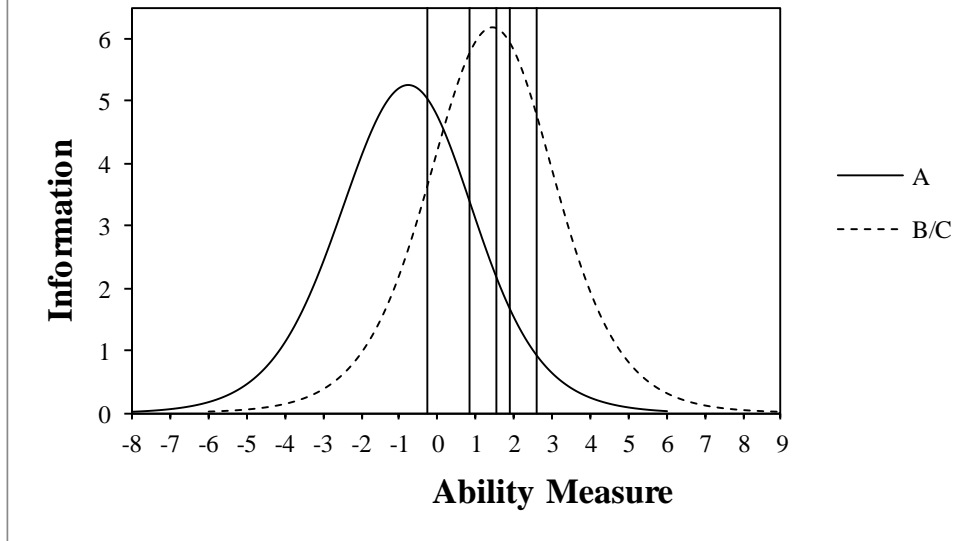
Note: The test form is shared between 3A and 4-5A, 3B/C and 4-5B/C.

2.9.2.5 Grades 6-8

Figure 2.9.2.5.1
Test Information Curve: Read 6-8 A S601 Paper
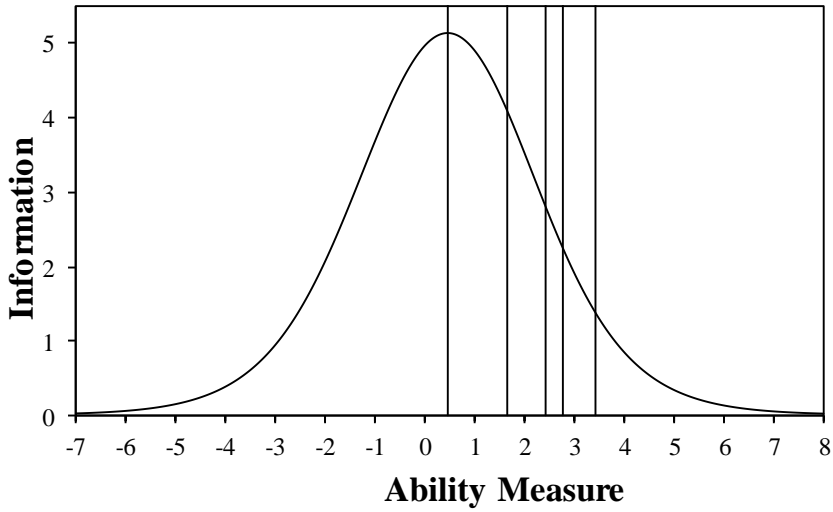
Figure 2.9.2.5.2
Test Information Curve: Read 6-8 B/C S601 Paper
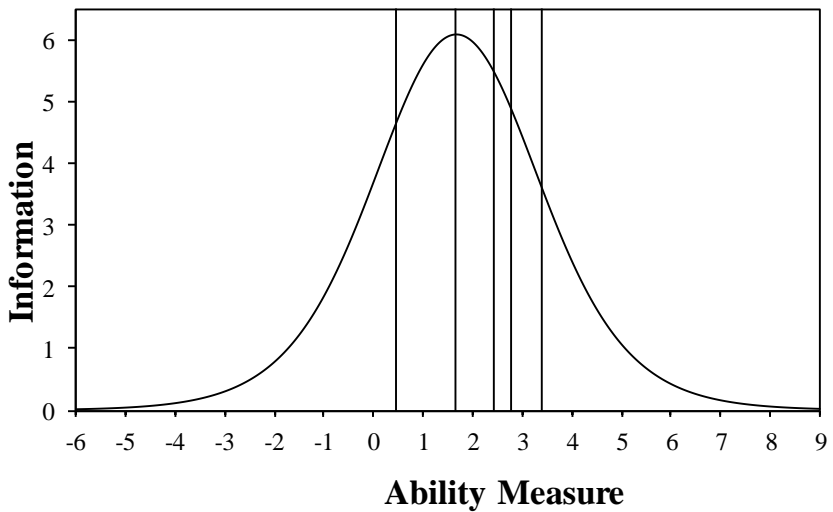
Figure 2.9.2.5.3
Test Information Curve: Read 6-8 S601 Paper
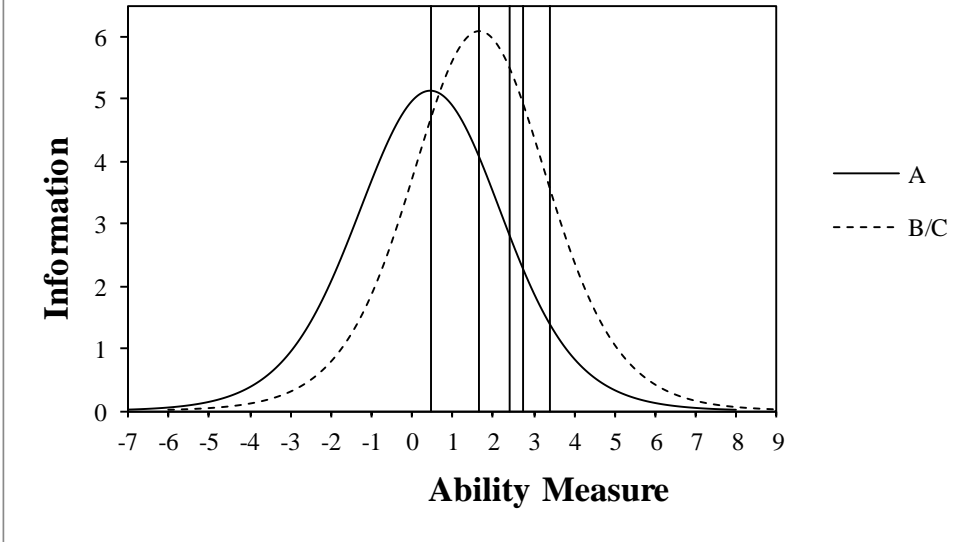
2.9.2.6 Grades 9-12

Figure 2.9.2.6.1
Test Information Curve: Read 9-12 A S601 Paper
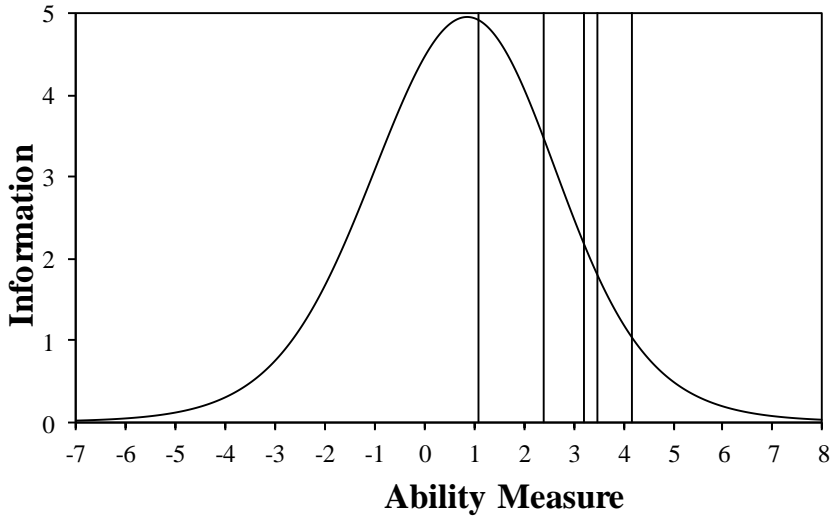
Figure 2.9.2.6.2
Test Information Curve: Read 9-12 B/C S601 Paper
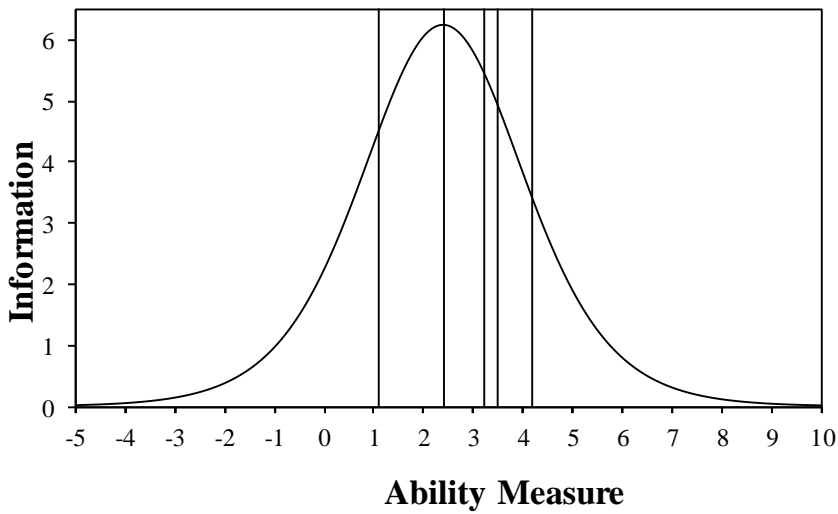
Figure 2.9.2.6.3
Test Information Curve: Read 9-12 S601 Paper
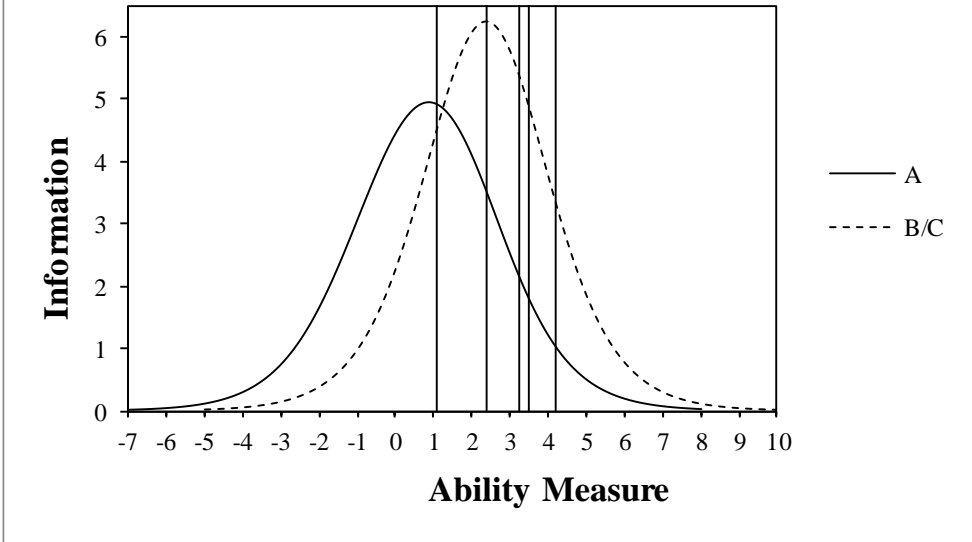
2.9.3 Writing

2.9.3.0 Kindergarten

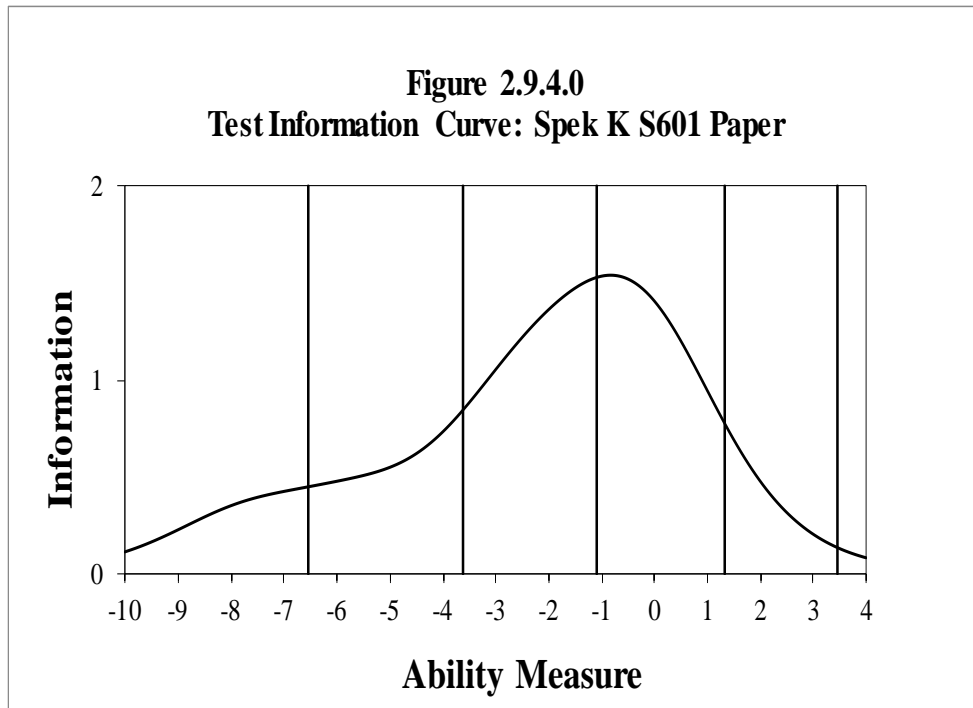**Figure 2.9.3.0**
Test Information Curve: Writ K S601 Paper

2.9.3.1 Grade 1

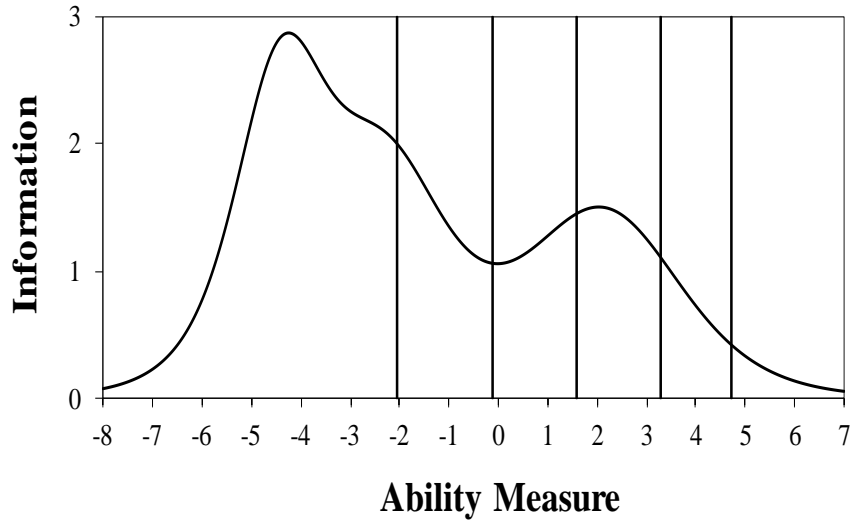**Figure 2.9.3.1.1**
Test Information Curve: Writ 1 A S601 Paper

Figure 2.9.3.1.2
Test Information Curve: Writ 1 B/C S601 Paper
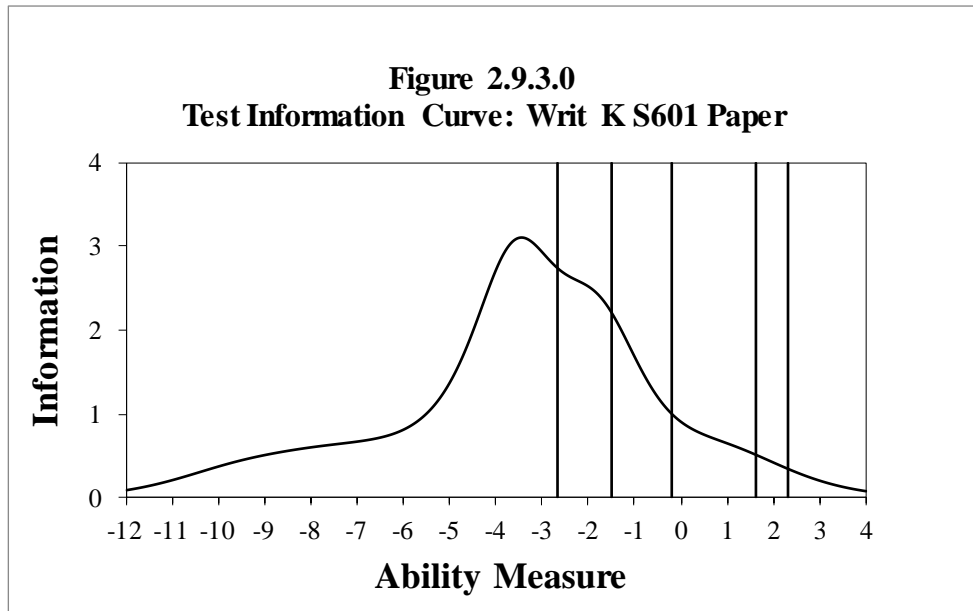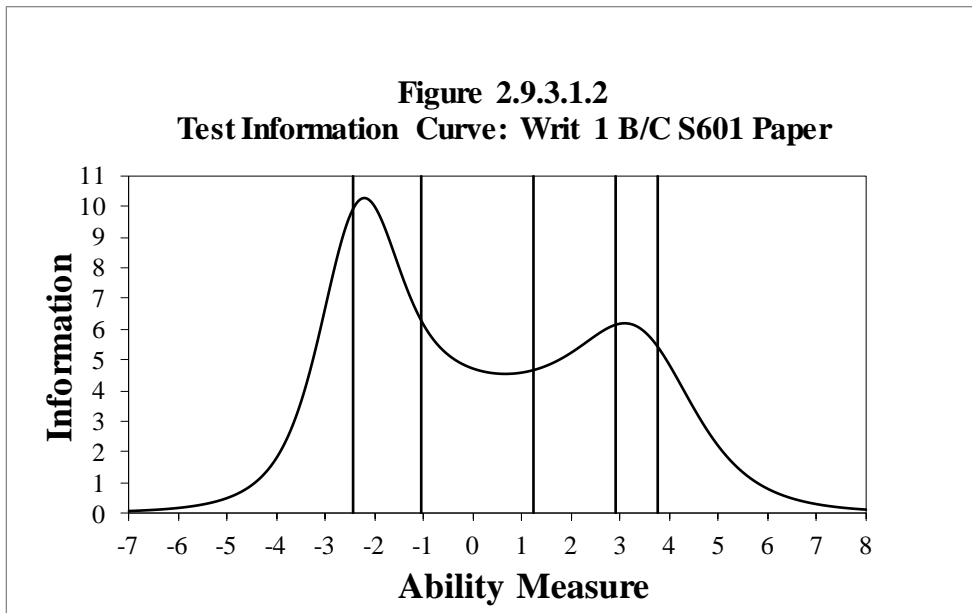
Figure 2.9.3.1.3
Test Information Curve: Writ 1 S601 Paper
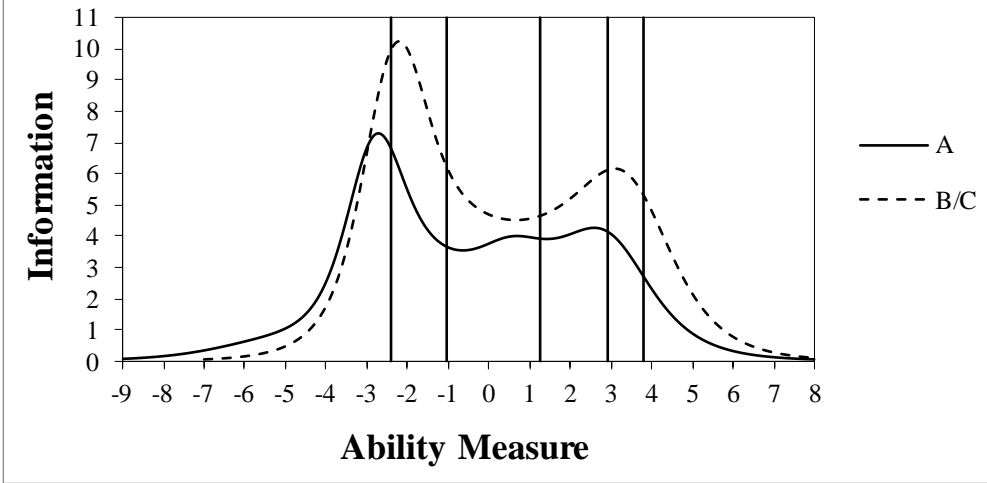
2.9.3.2 Grade 2
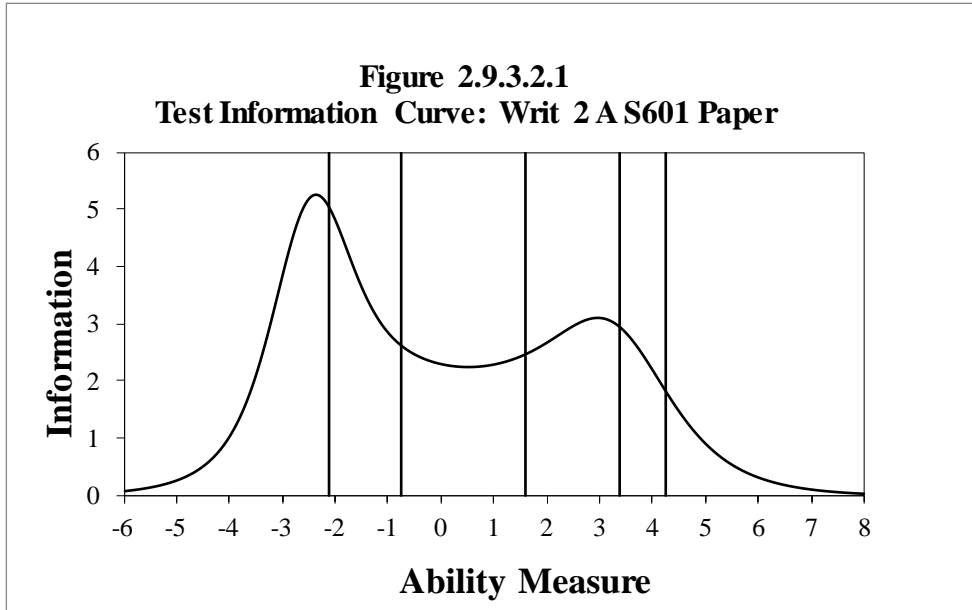
Figure 2.9.3.2.1
Test Information Curve: Writ 2 A S601 Paper

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

Figure 2.9.3.2.2
Test Information Curve: Writ 2 B/C S601 Paper
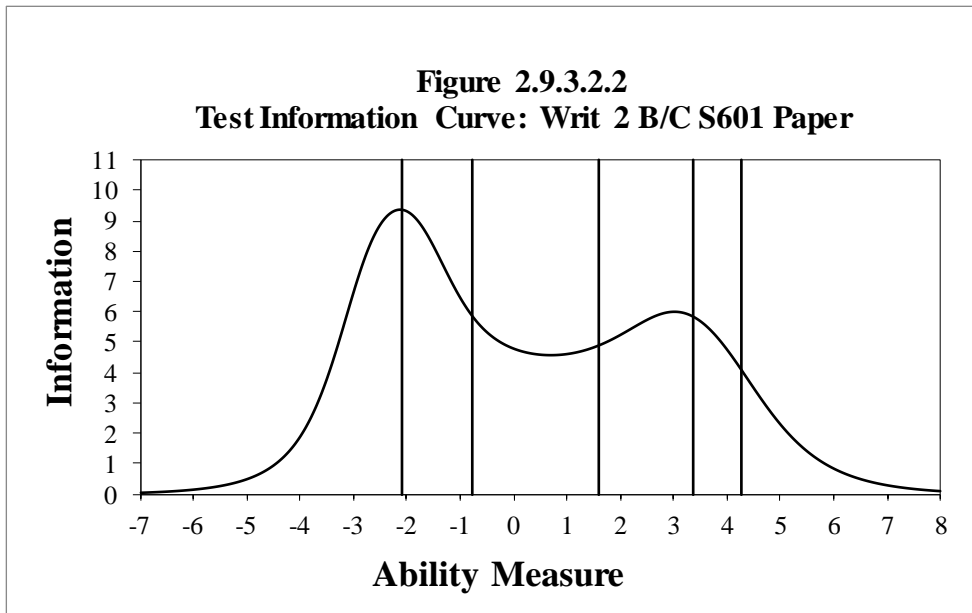
Note: The test form is shared between 2B/C and 3B/C.

Figure 2.9.3.2.3
Test Information Curve: Writ 2 S601 Paper
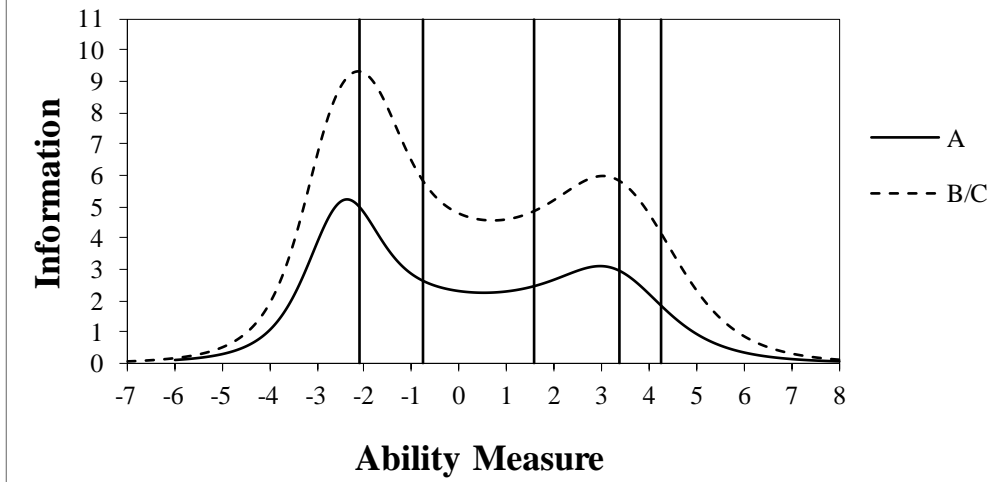
Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

2.9.3.3 Grade 3

Figure 2.9.3.3.1
Test Information Curve: Writ 3 A S601 Paper

Information vs *Ability Measure*

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

Figure 2.9.3.3.2
Test Information Curve: Writ 3 B/C S601 Paper

Information vs *Ability Measure*
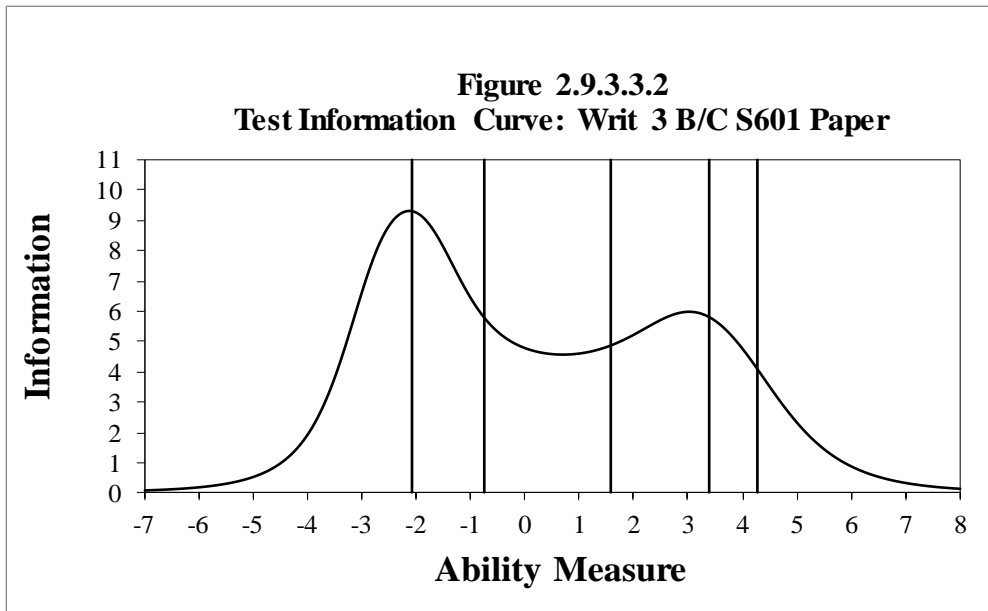
Note: The test form is shared between 2B/C and 3B/C.

Figure 2.9.3.3.3
Test Information Curve: Writ 3 S601 Paper
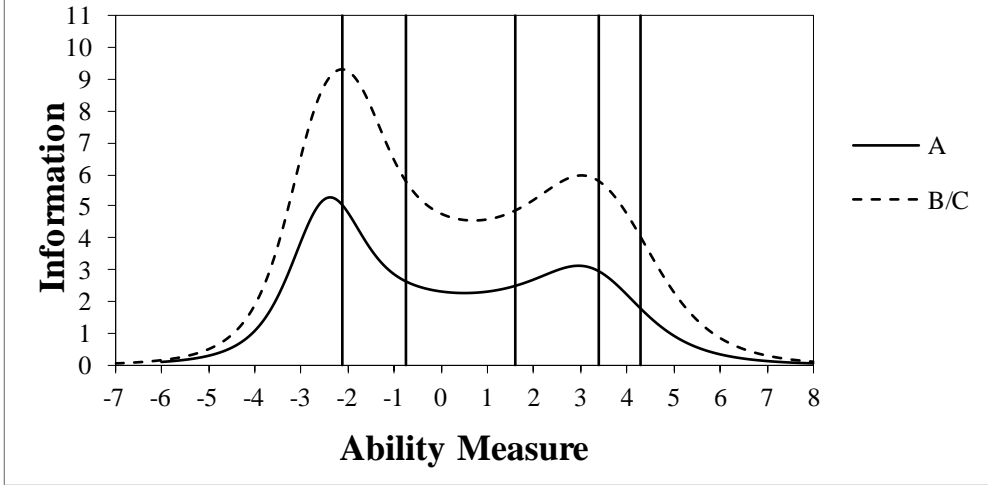
Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

2.9.3.4 Grades 4-5

Figure 2.9.3.4.1
Test Information Curve: Writ 4-5 A S601 Paper

Figure 2.9.3.4.2
Test Information Curve: Writ 4-5 B/C S601 Paper
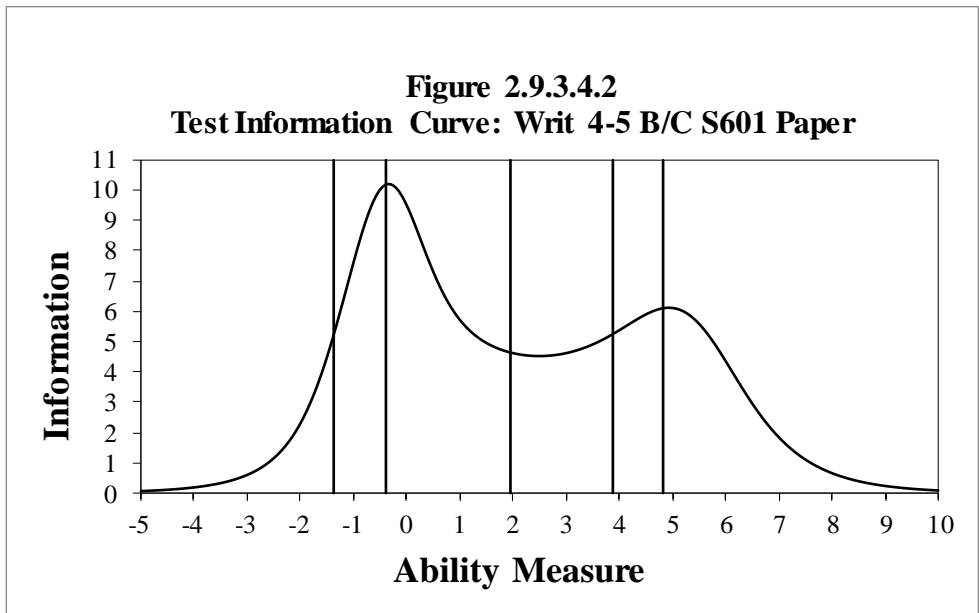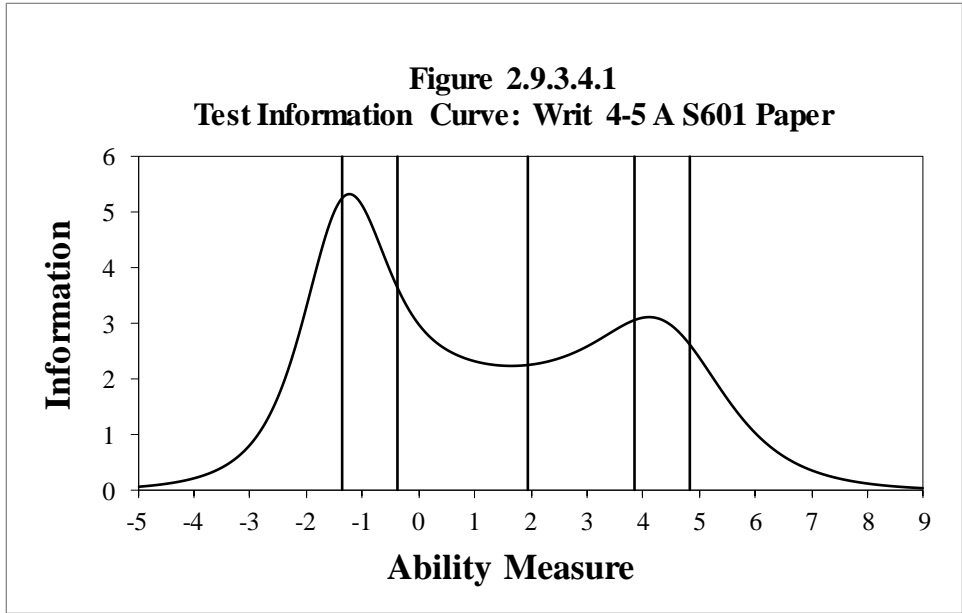
Figure 2.9.3.4.3
Test Information Curve: Writ 4-5 S601 Paper

2.9.3.5 Grades 6-8

Figure 2.9.3.5.1
Test Information Curve: Writ 6-8 A S601 Paper

Figure 2.9.3.5.2
Test Information Curve: Writ 6-8 B/C S601 Paper
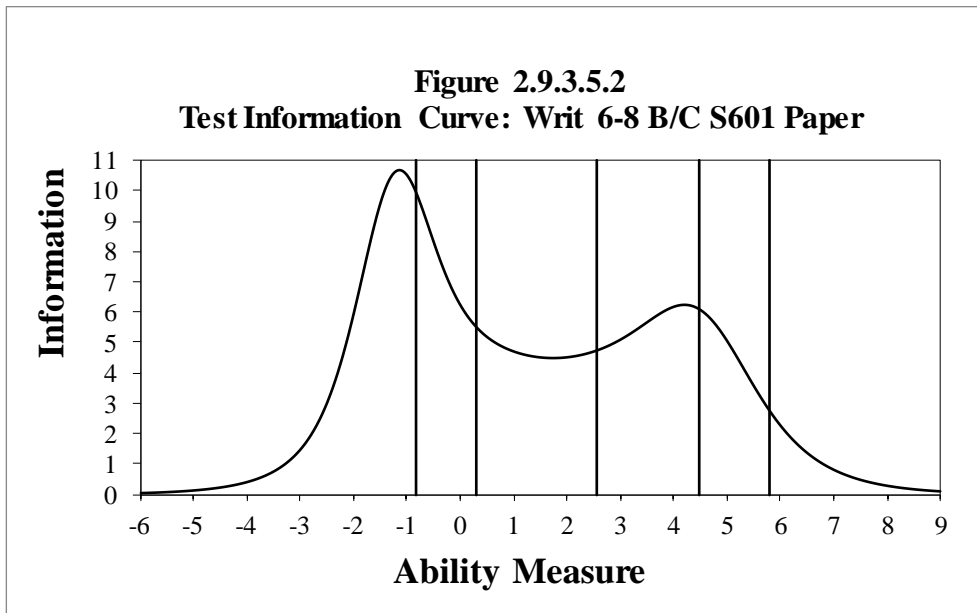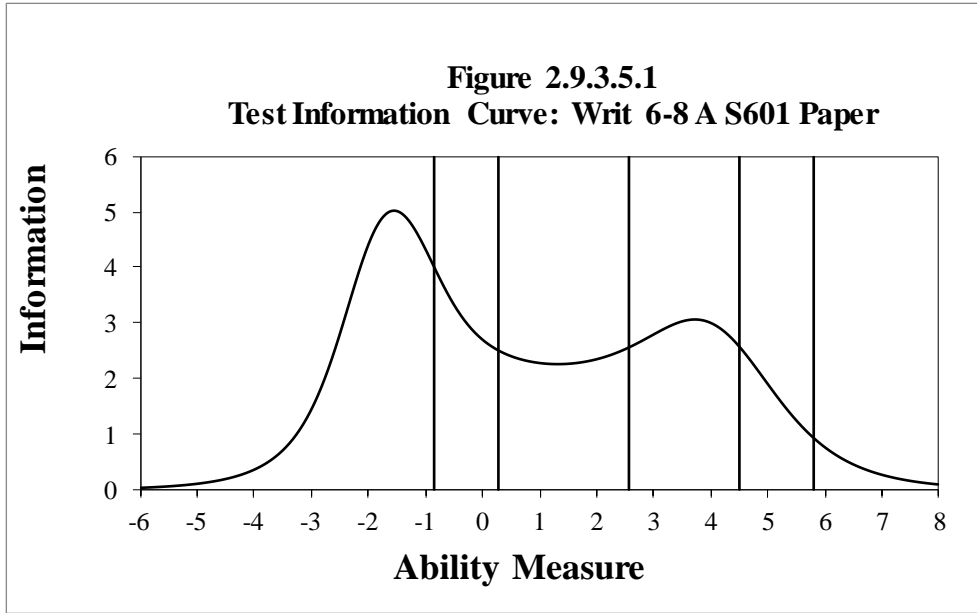
Figure 2.9.3.5.3
Test Information Curve: Writ 6-8 S601 Paper

2.9.3.6 Grades 9-12

Figure 2.9.3.6.1
Test Information Curve: Writ 9-12 A S601 Paper

Figure 2.9.3.6.2
Test Information Curve: Writ 9-12 B/C S601 Paper
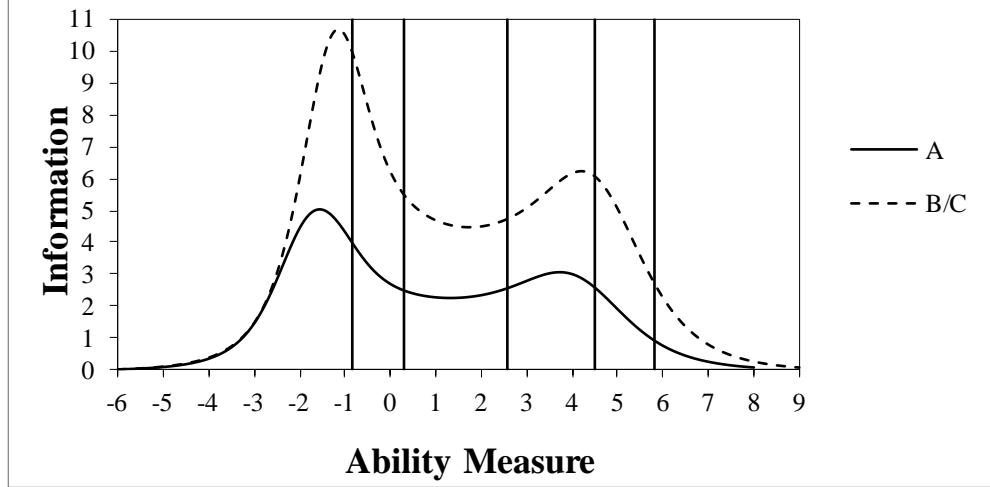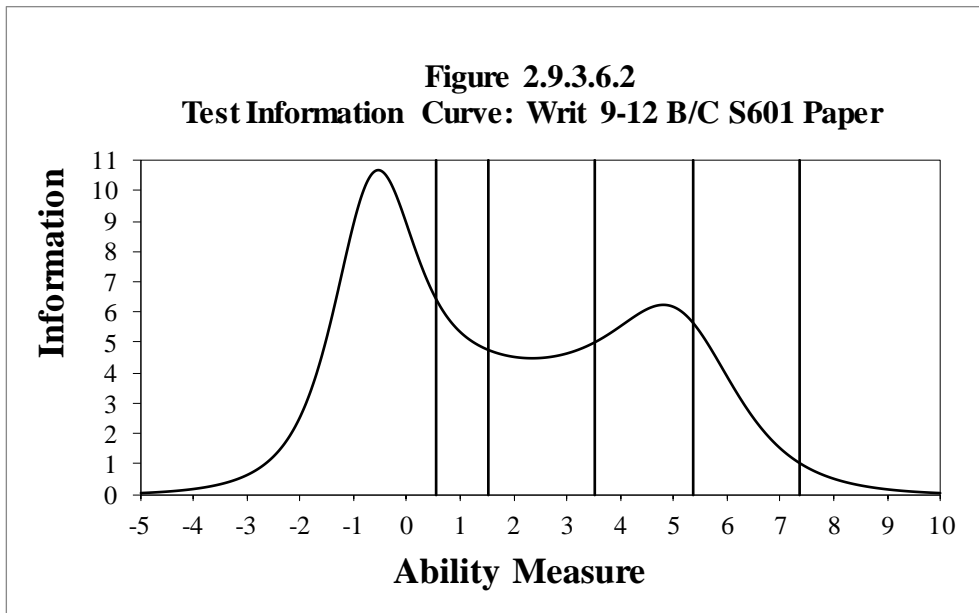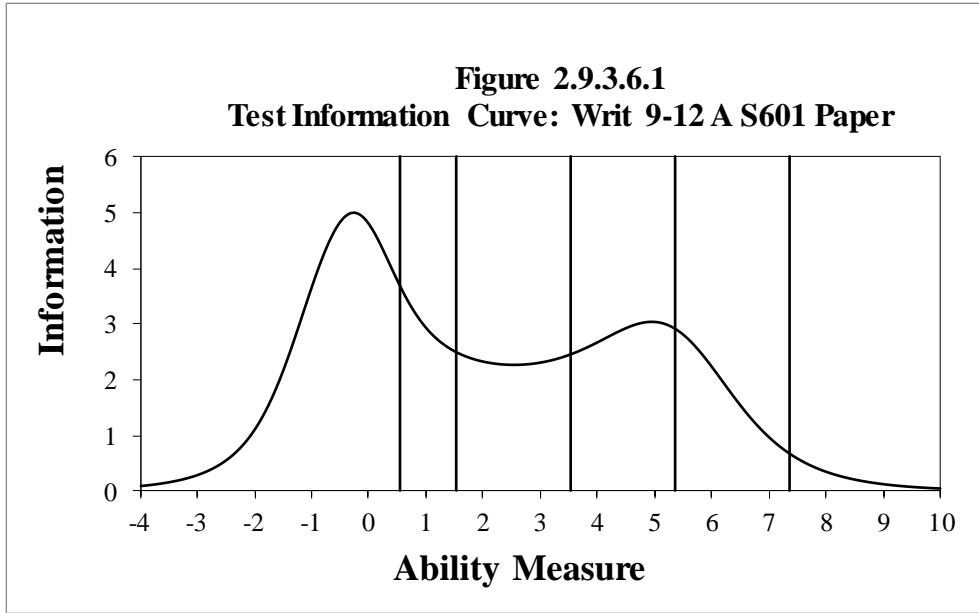
Figure 2.9.3.6.3
Test Information Curve: Writ 9-12 S601 Paper
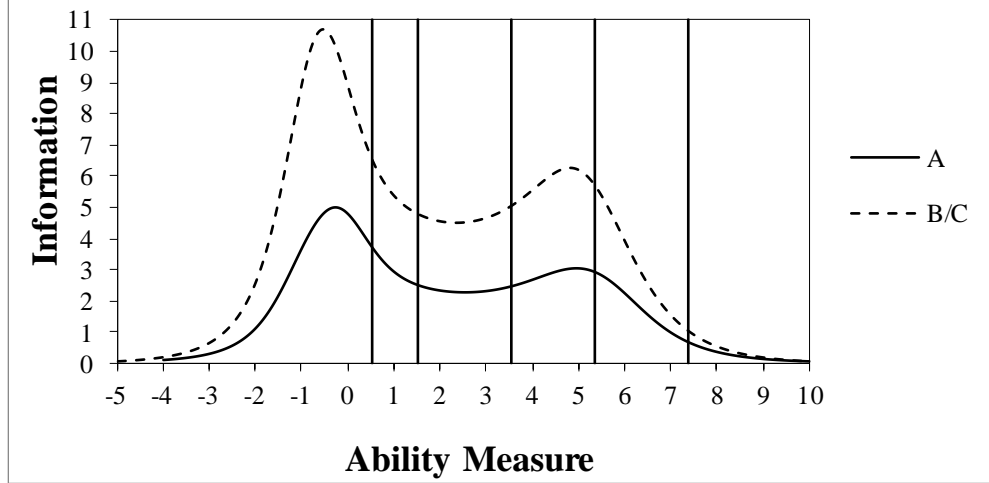
2.9.4 Speaking

2.9.4.0 Kindergarten

Figure 2.9.4.0
Test Information Curve: Spek K S601 Paper

2.9.4.1 Grade 1

Figure 2.9.4.1.1
Test Information Curve: Spek 1 A S601 Paper

Figure 2.9.4.1.2
Test Information Curve: Spek 1 B/C S601 Paper
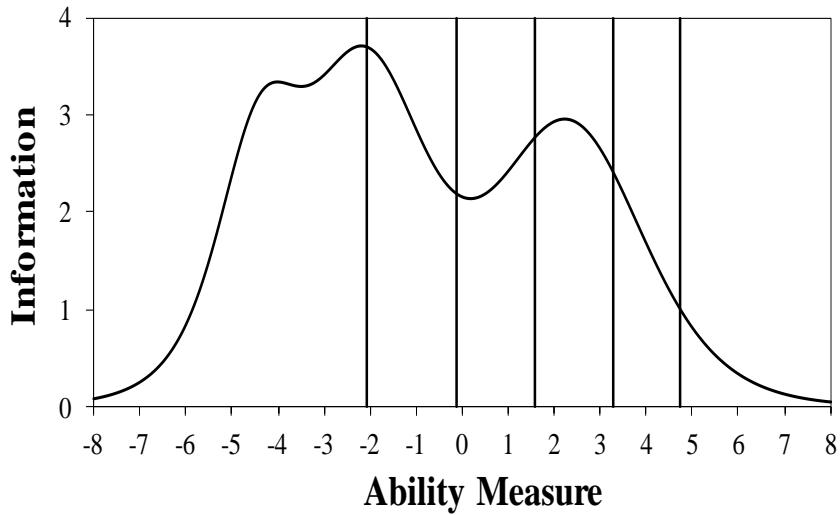
Figure 2.9.4.1.3
Test Information Curve: Spek 1 S601 Paper
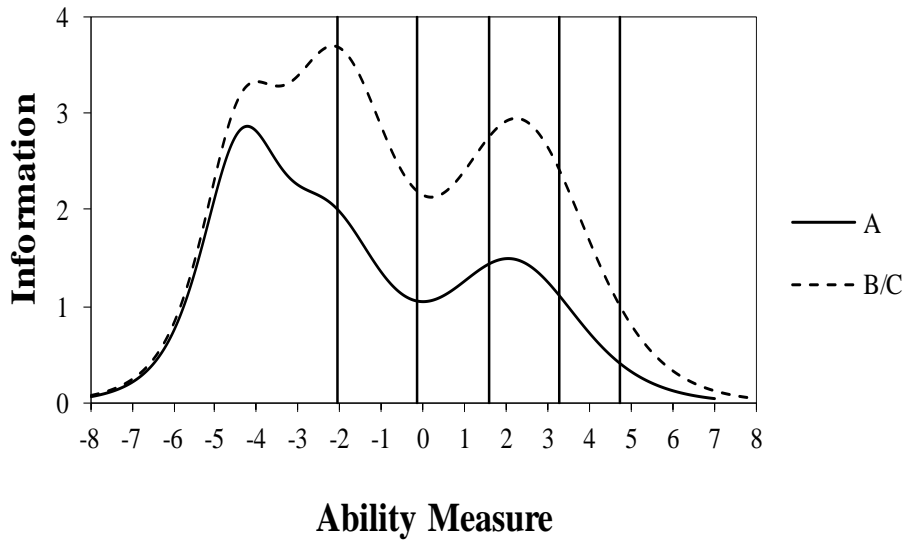
2.9.4.2 Grade 2
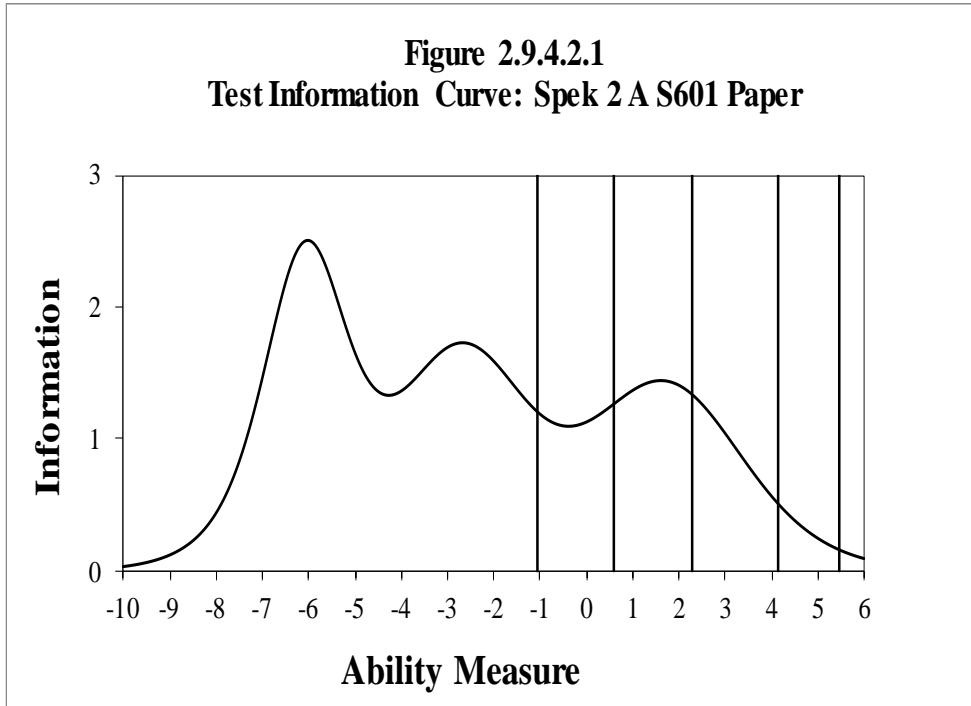
Figure 2.9.4.2.1
Test Information Curve: Spek 2 A S601 Paper

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

Figure 2.9.4.2.2
Test Information Curve: Spek 2 B/C S601 Paper
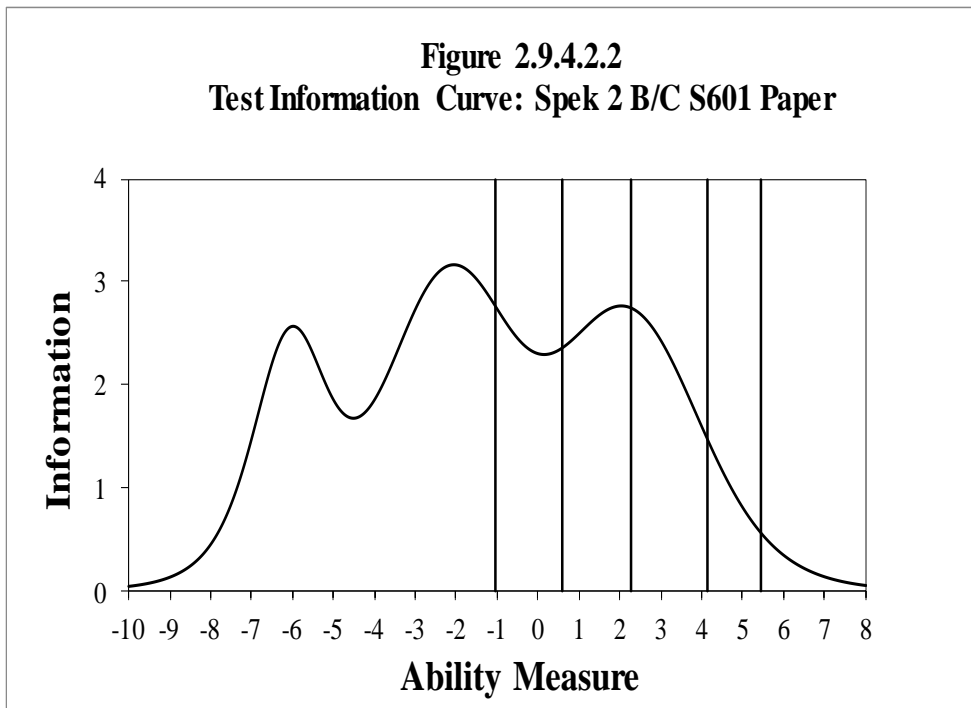
Note: The test form is shared between 2B/C and 3B/C.

Figure 2.9.4.2.3
Test Information Curve: Spek 2 S601 Paper
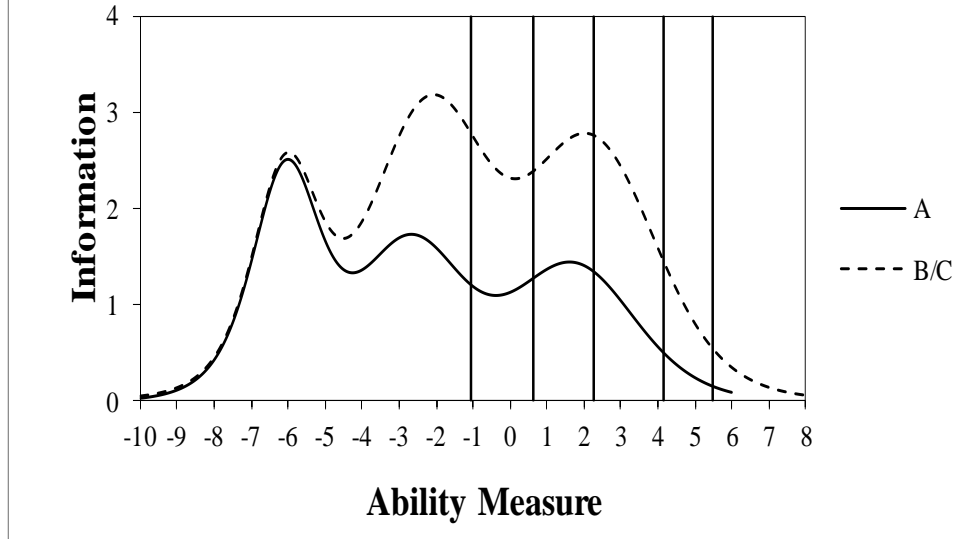
Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.
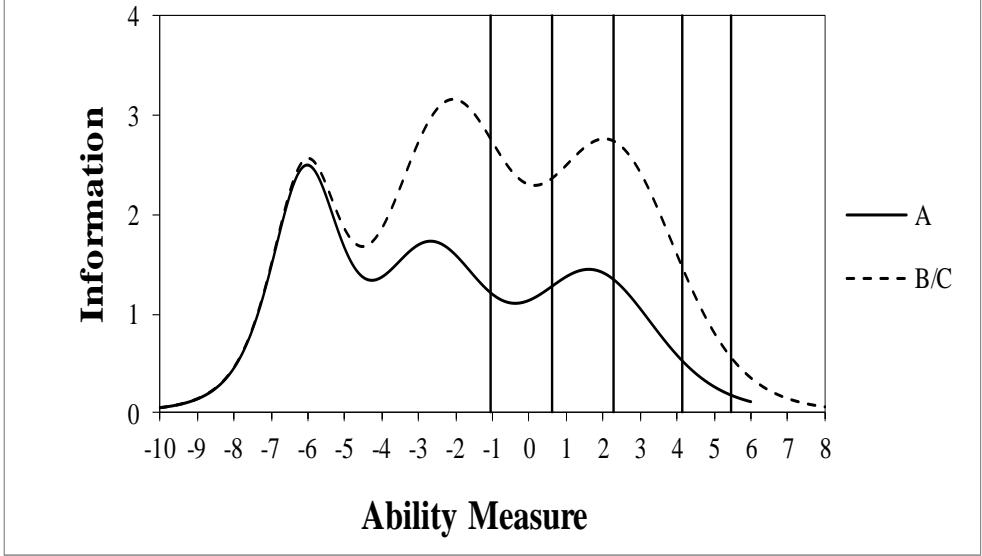
2.9.4.3 Grade 3
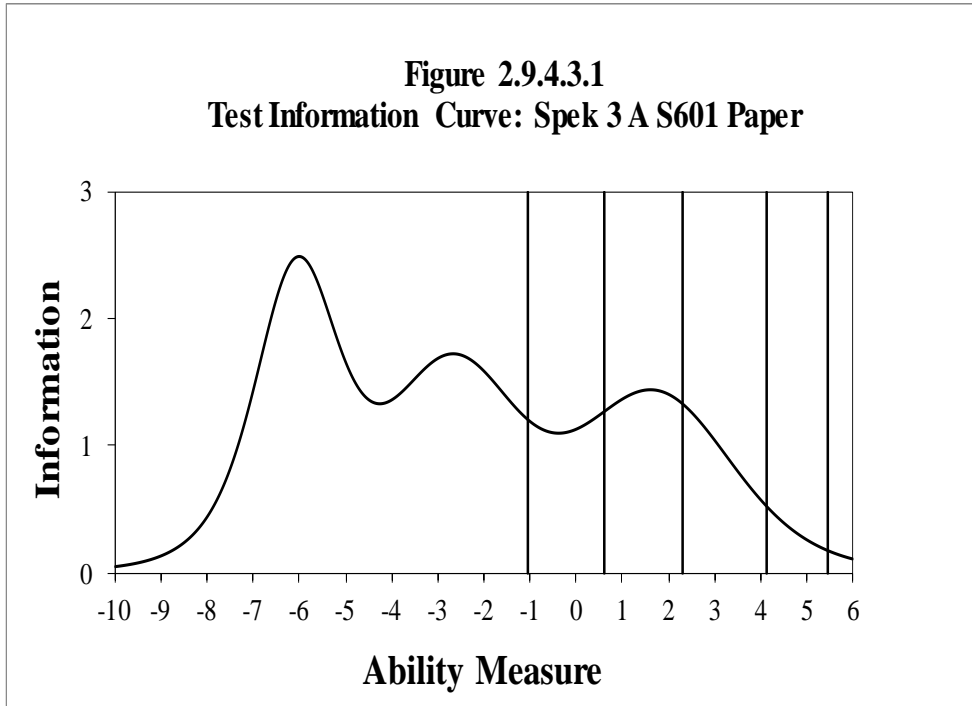
Figure 2.9.4.3.1
Test Information Curve: Spek 3 A S601 Paper

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

Figure 2.9.4.3.2
Test Information Curve: Spek 3 B/C S601 Paper
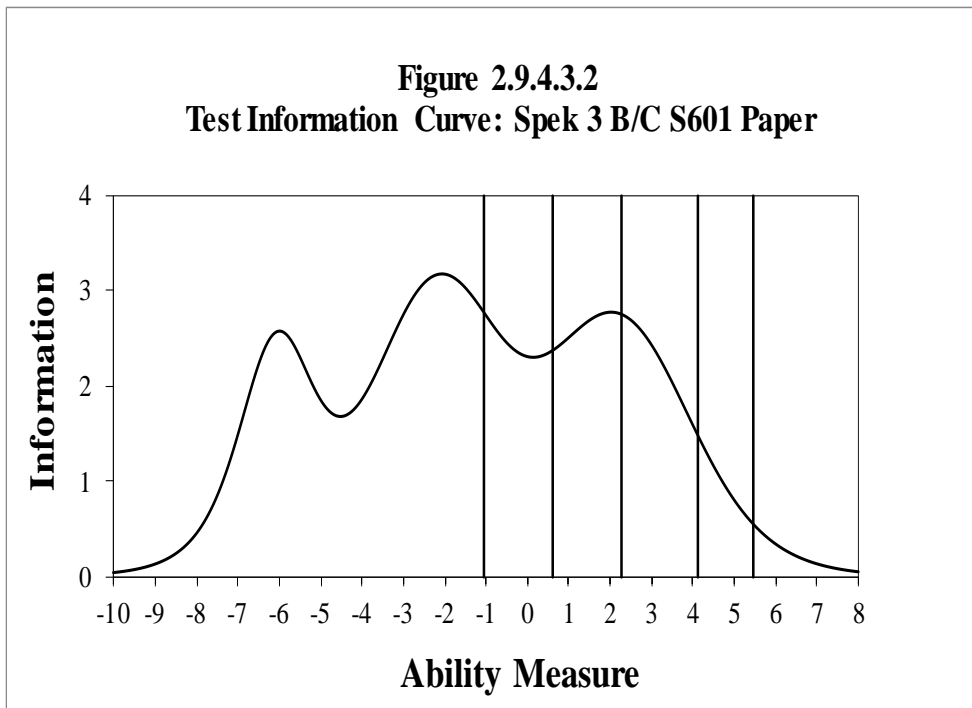
Note: The test form is shared between 2B/C and 3B/C.

Figure 2.9.4.3.3
Test Information Curve: Spek 3 S601 Paper

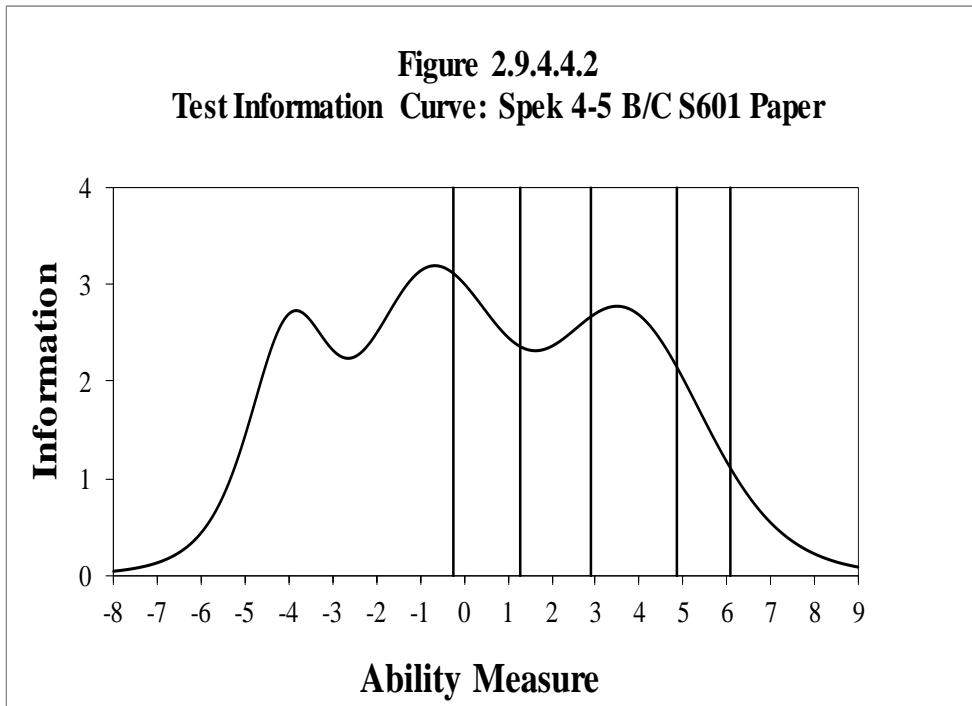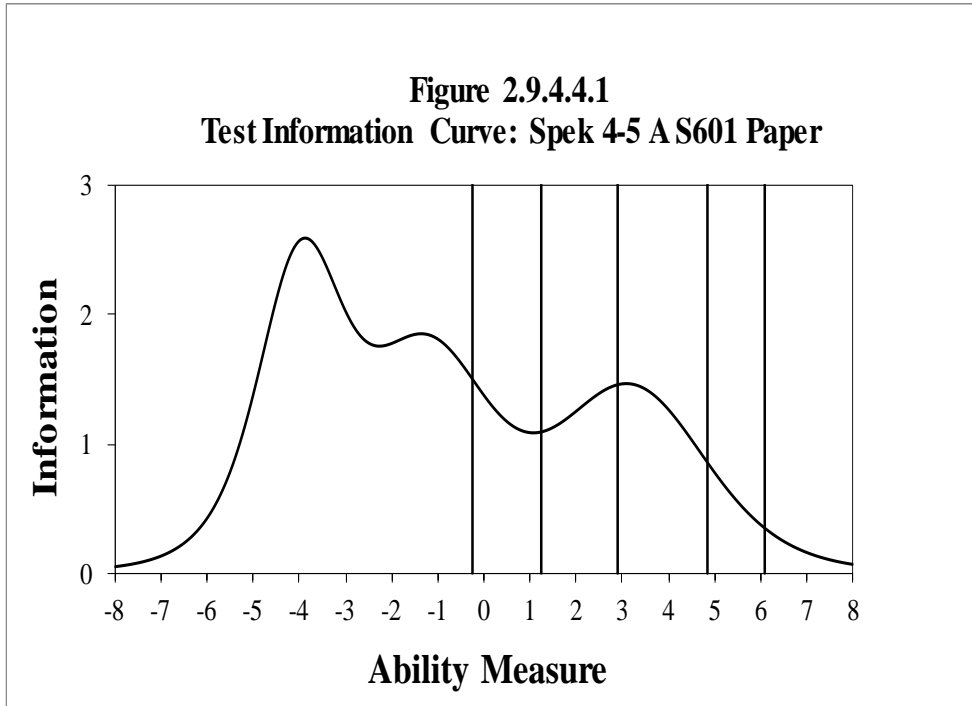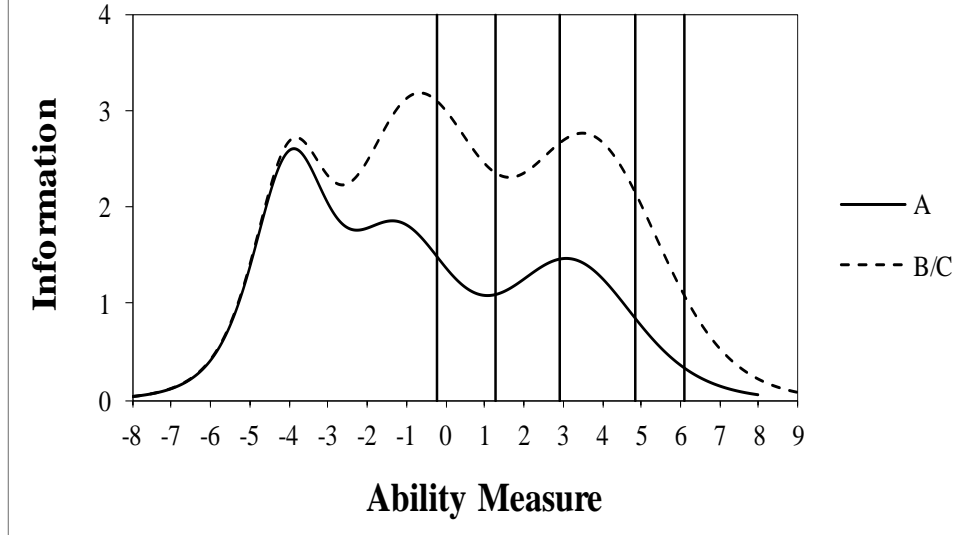Note: The test form is shared between 2A and 3A, 2B/C and 3B/C.

2.9.4.4 Grades 4-5

Figure 2.9.4.4.1
Test Information Curve: Spek 4-5 A S601 Paper

Figure 2.9.4.4.2
Test Information Curve: Spek 4-5 B/C S601 Paper

Figure 2.9.4.4.3
Test Information Curve: Spek 4-5 S601 Paper

2.9.4.5 Grades 6-8

Figure 2.9.4.5.1
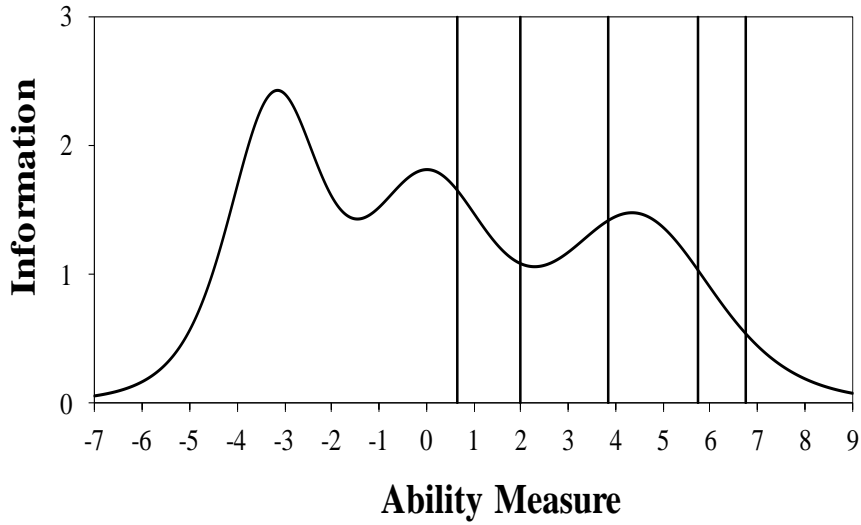**Test Information Curve: Spek 6-8 A S601 Paper**

Figure 2.9.4.5.2
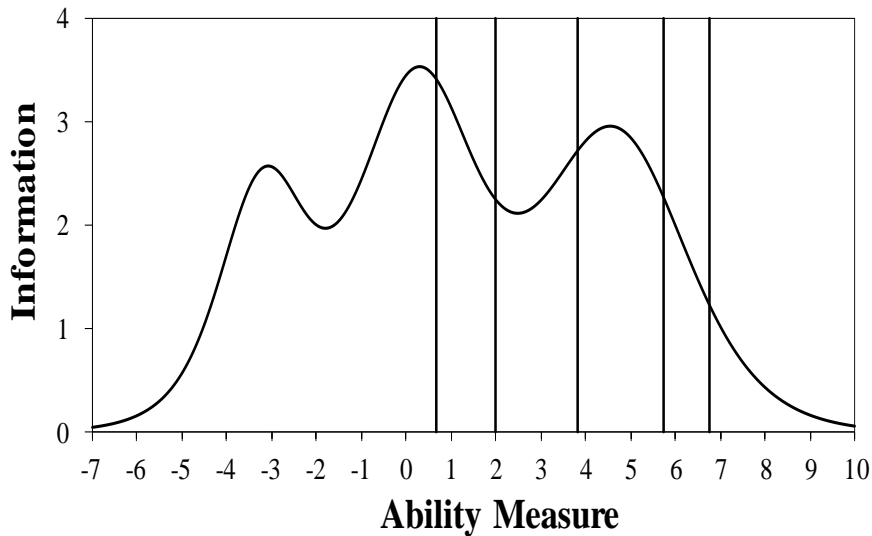**Test Information Curve: Spek 6-8 B/C S601 Paper**
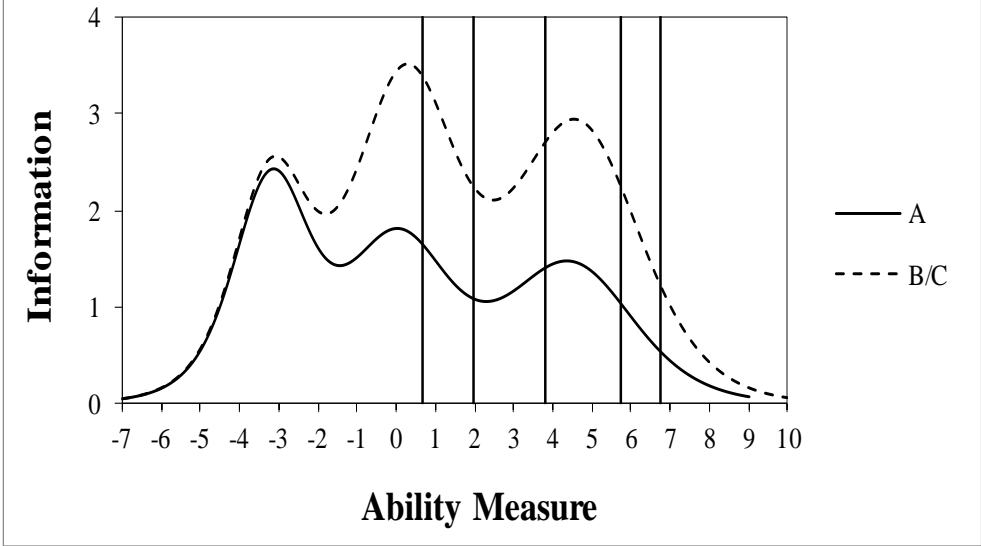
Figure 2.9.4.5.3
Test Information Curve: Spek 6-8 S601 Paper

2.9.4.6 Grades 9-12

Figure 2.9.4.6.1
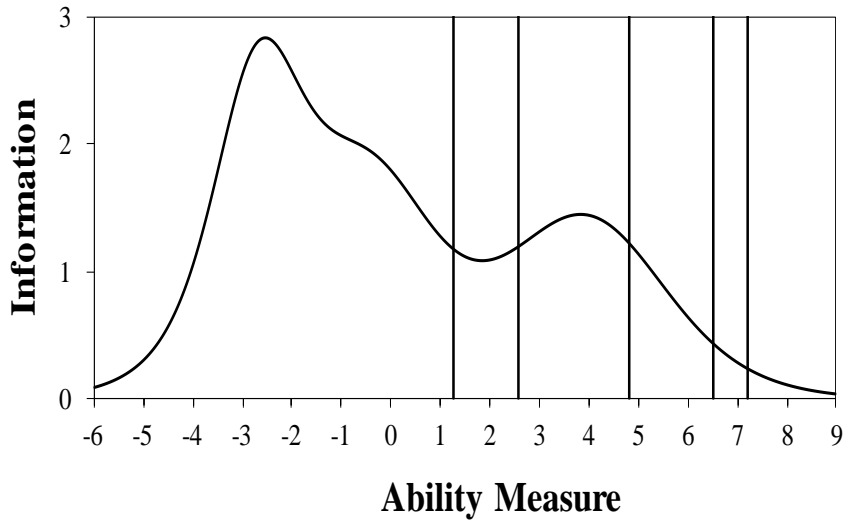**Test Information Curve: Spek 9-12 A S601 Paper**

Figure 2.9.4.6.2
Test Information Curve: Spek 9-12 B/C S601 Paper
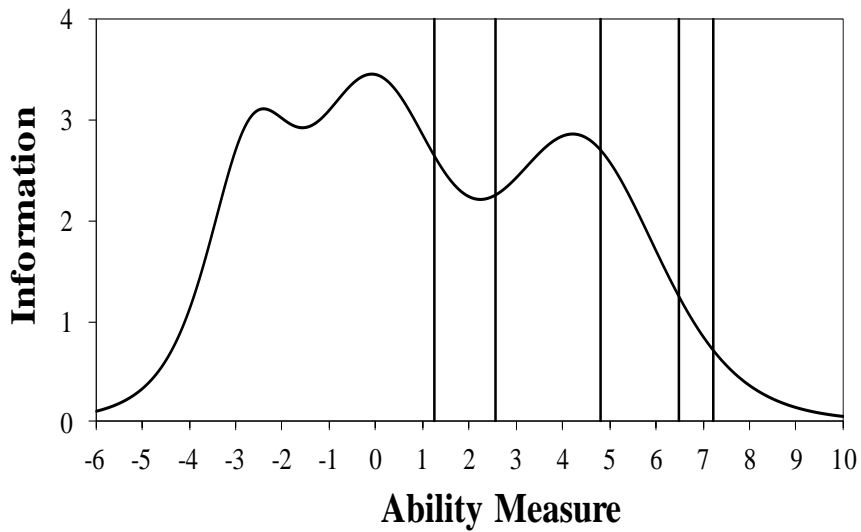
Figure 2.9.4.6.3
Test Information Curve: Spek 9-12 S601 Paper
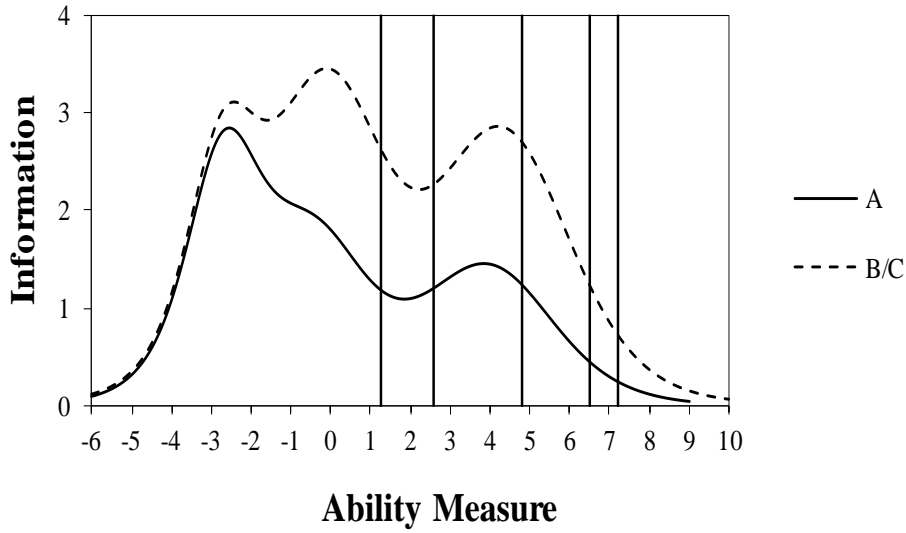
3.   Analyses of Composite Scores

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 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	256,827	100	378	249.82	88.92

Figure 3.1.1.0
Scale Scores: Oral K S601 Paper
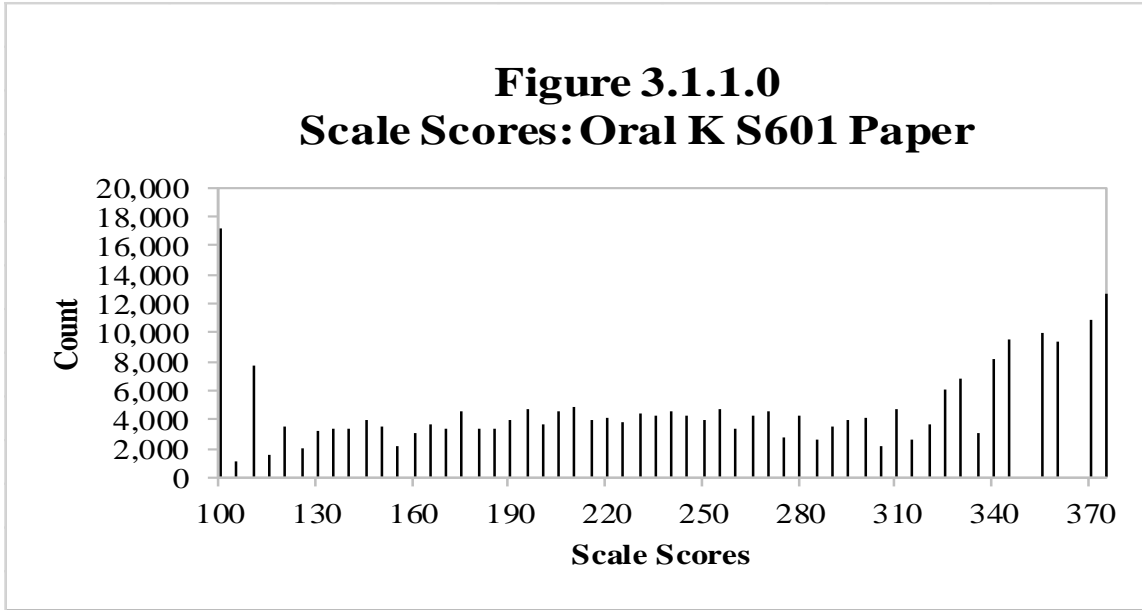
3.1.1.1 Grade 1

Table 3.1.1.1

Scale Score Descriptive Statistics: Oral 1 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	33,201	105	406	281.99	51.32

Figure 3.1.1.1
Scale Scores: Oral 1 S601 Paper
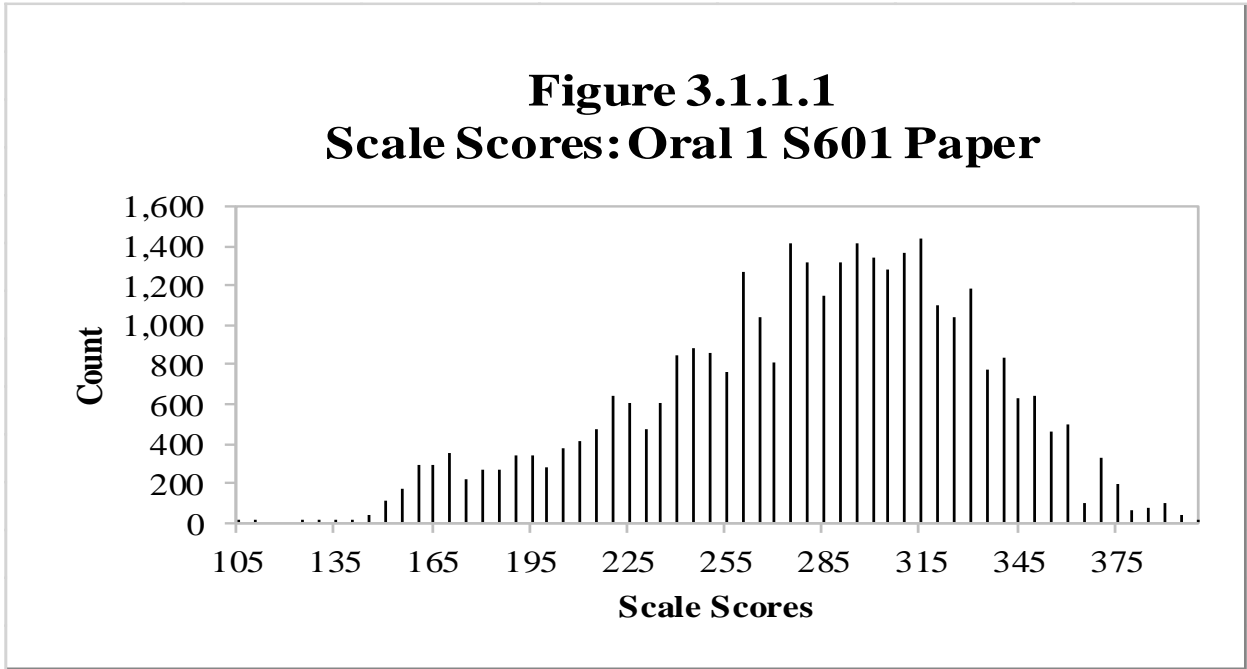
3.1.1.2 Grade 2

Table 3.1.1.2

Scale Score Descriptive Statistics: Oral 2 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	32,069	115	415	300.84	54.81

Figure 3.1.1.2
Scale Scores: Oral 2 S601 Paper
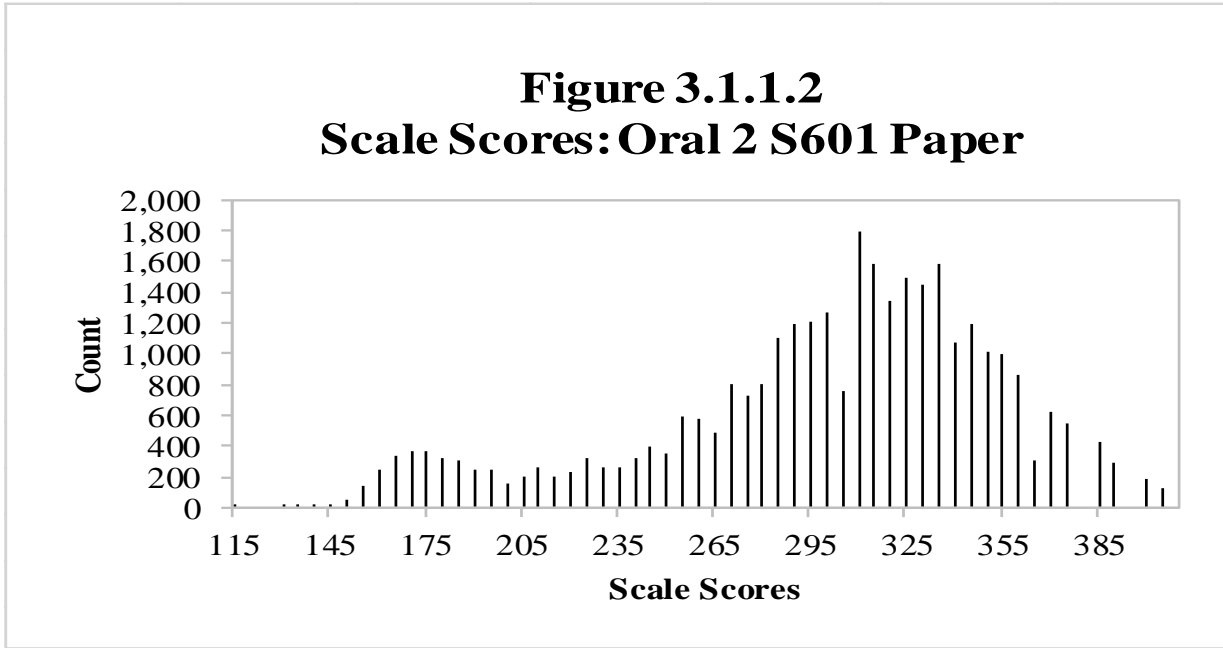
3.1.1.3 Grade 3

Table 3.1.1.3

Scale Score Descriptive Statistics: Oral 3 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	32,413	151	448	319.88	50.76

Figure 3.1.1.3
Scale Scores: Oral 3 S601 Paper
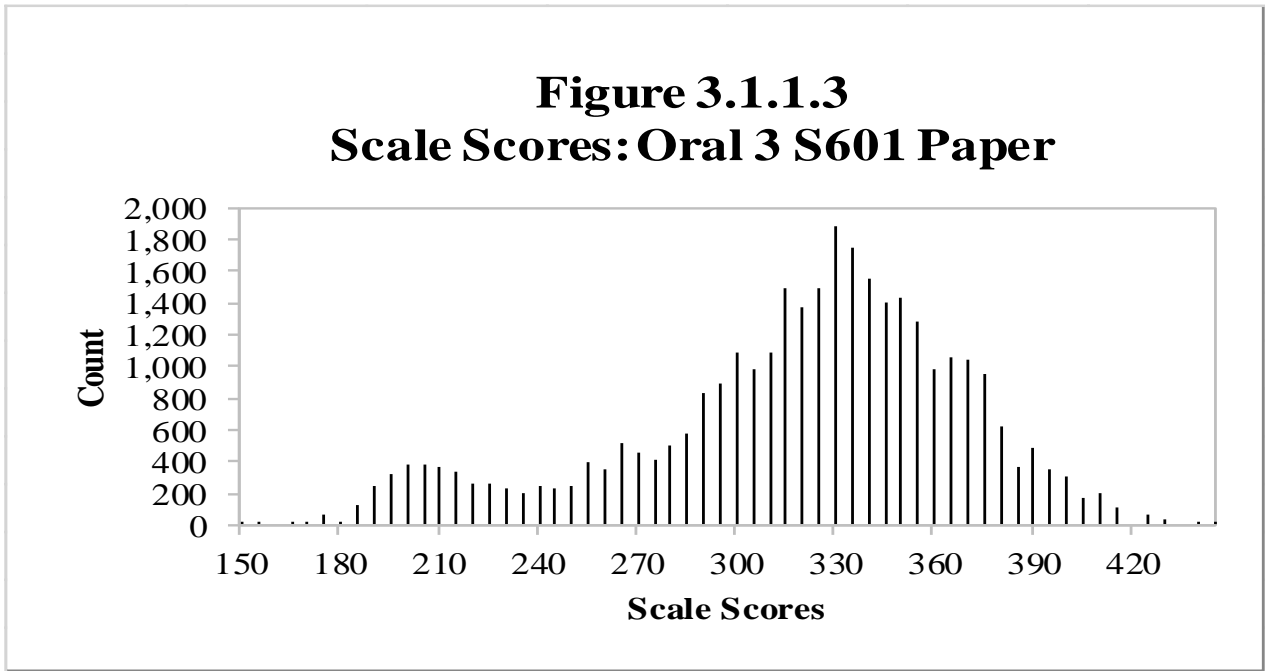
3.1.1.4 Grades 4-5

Table 3.1.1.4

Scale Score Descriptive Statistics: Oral 4-5 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	23,422	173	461	341.64	58.50
5	21,531	157	461	345.08	60.79
Total	44,953	157	461	343.29	59.63

Figure 3.1.1.4
Scale Scores: Oral 4-5 S601 Paper
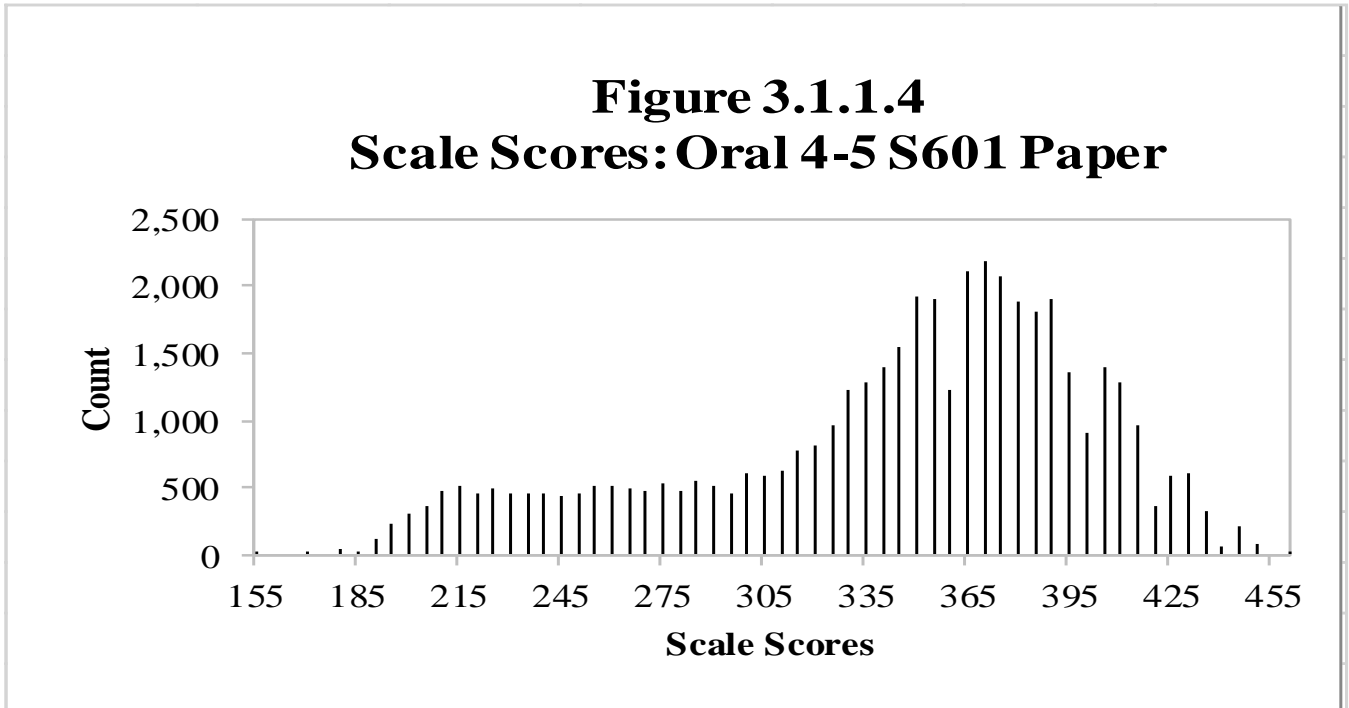
3.1.1.5 Grades 6-8

Table 3.1.1.5

Scale Score Descriptive Statistics: Oral 6-8 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	18,525	165	474	347.07	59.54
7	16,269	168	474	347.85	63.13
8	15,337	180	474	353.29	63.95
Total	50,131	165	474	349.23	62.14

**Figure 3.1.1.5
Scale Scores: Oral 6-8 S601 Paper**
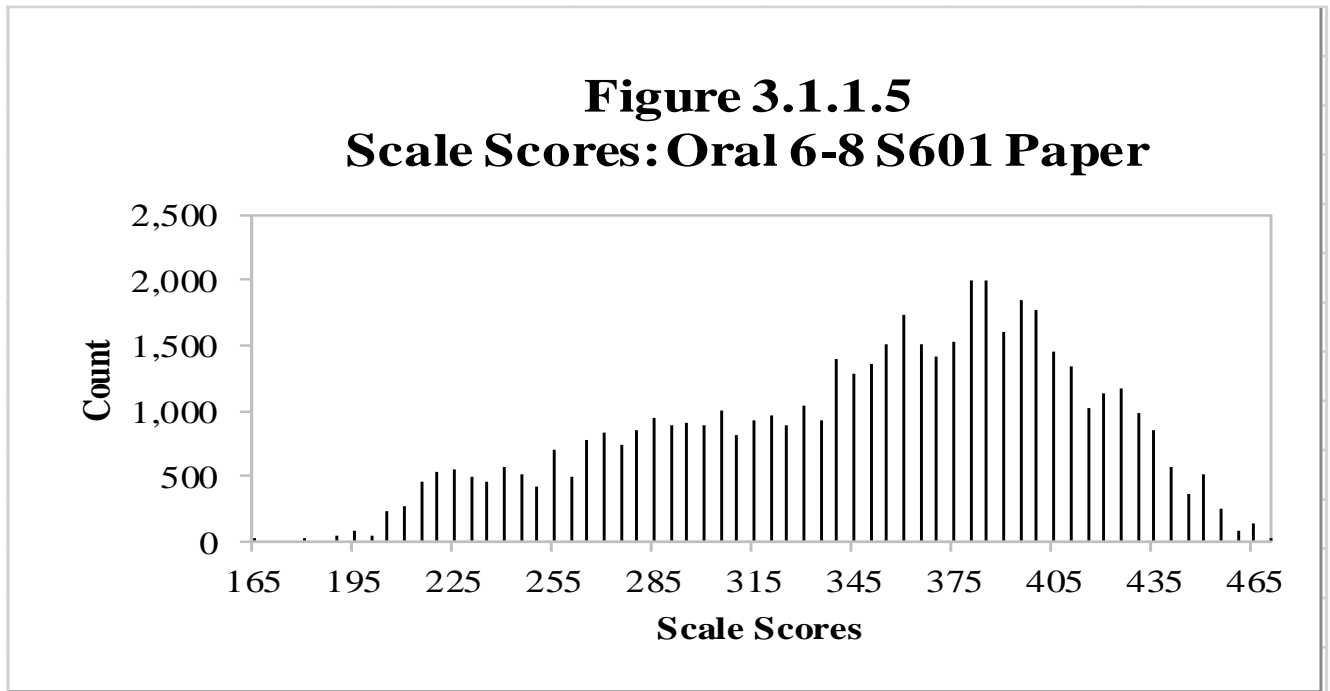
3.1.1.6 Grades 9-12

Table 3.1.1.6

Scale Score Descriptive Statistics: Oral 9-12 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	15,153	172	485	348.25	64.56
10	13,076	189	485	351.91	61.93
11	11,116	200	485	363.15	59.20
12	7,587	200	485	370.45	53.86
Total	46,932	172	485	356.39	61.50

Figure 3.1.1.6
Scale Scores: Oral 9-12 S601 Paper
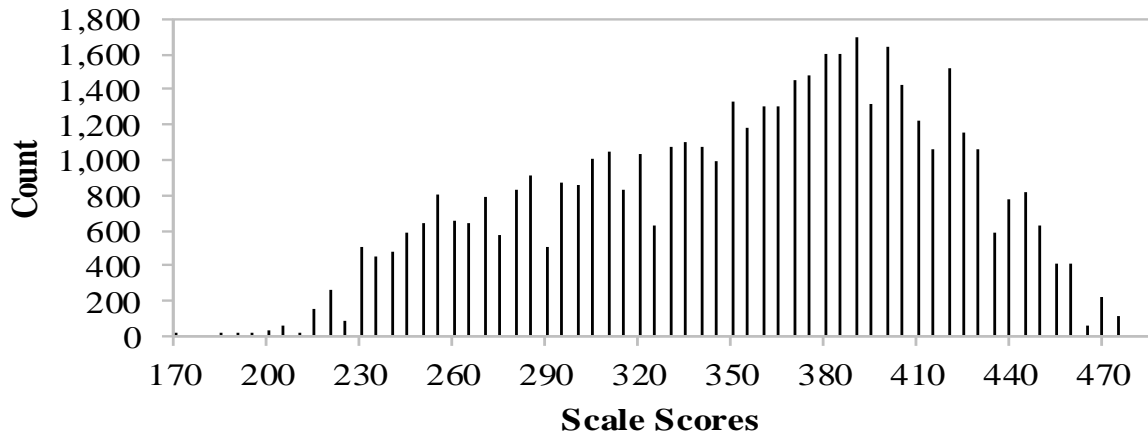
3.1.2 Literacy

3.1.2.0 Kindergarten

Table 3.1.2.0

Scale Score Descriptive Statistics: Litr K S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	256,829	100	315	179.49	61.69

Figure 3.1.2.0
Scale Scores: Litr K S601 Paper
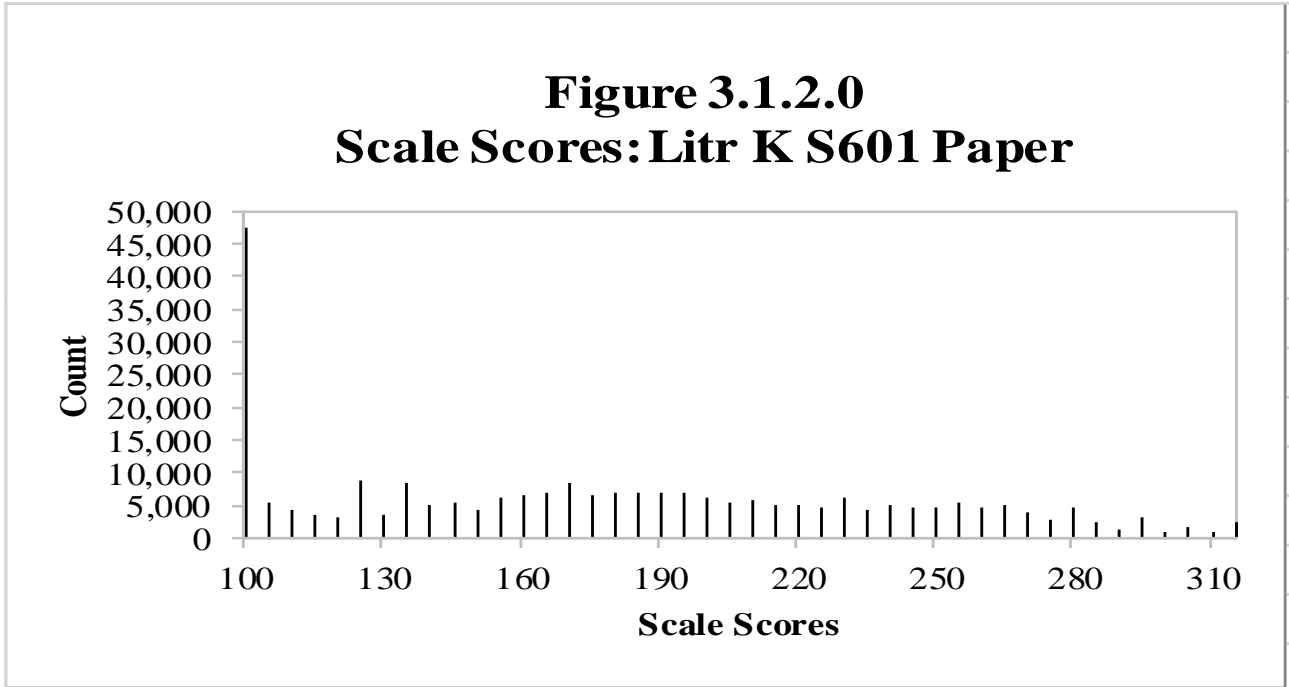
3.1.2.1 Grade 1

Table 3.1.2.1

Scale Score Descriptive Statistics: Litr 1 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	30,310	145	372	262.51	31.24

Figure 3.1.2.1
Scale Scores: Litr 1 S601 Paper
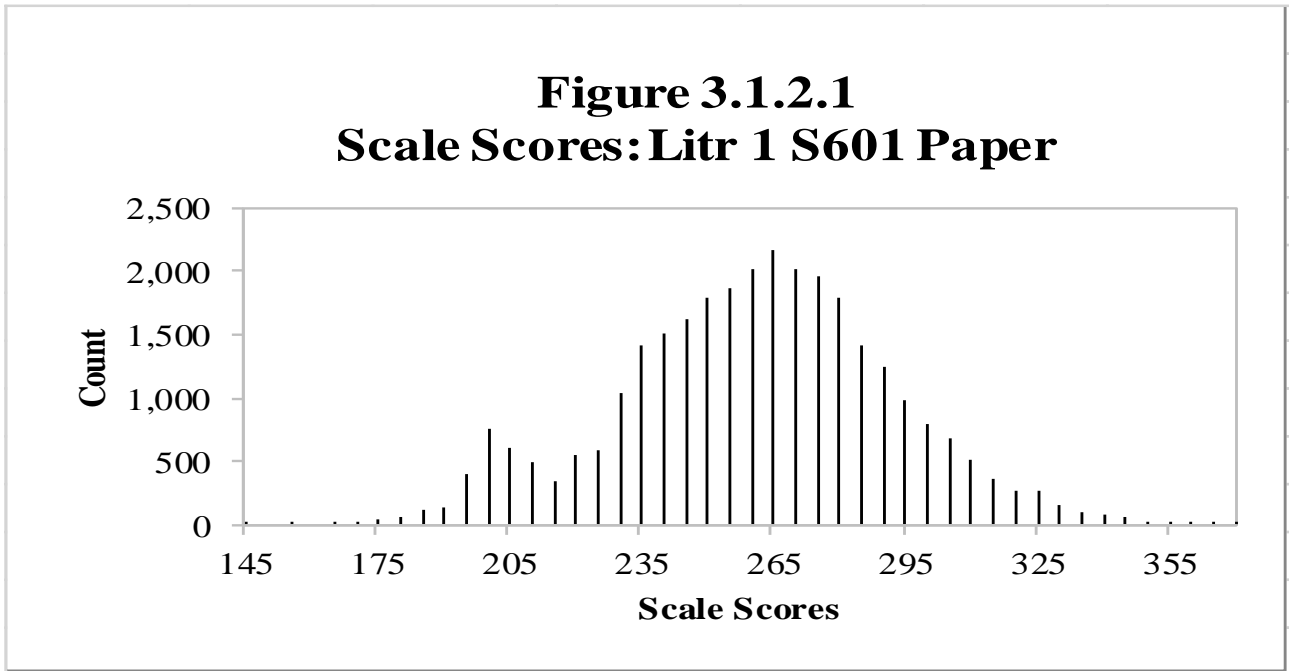
3.1.2.2 Grade 2

Table 3.1.2.2

Scale Score Descriptive Statistics: Litr 2 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	28,745	177	388	288.38	36.28

Figure 3.1.2.2
Scale Scores: Litr 2 S601 Paper
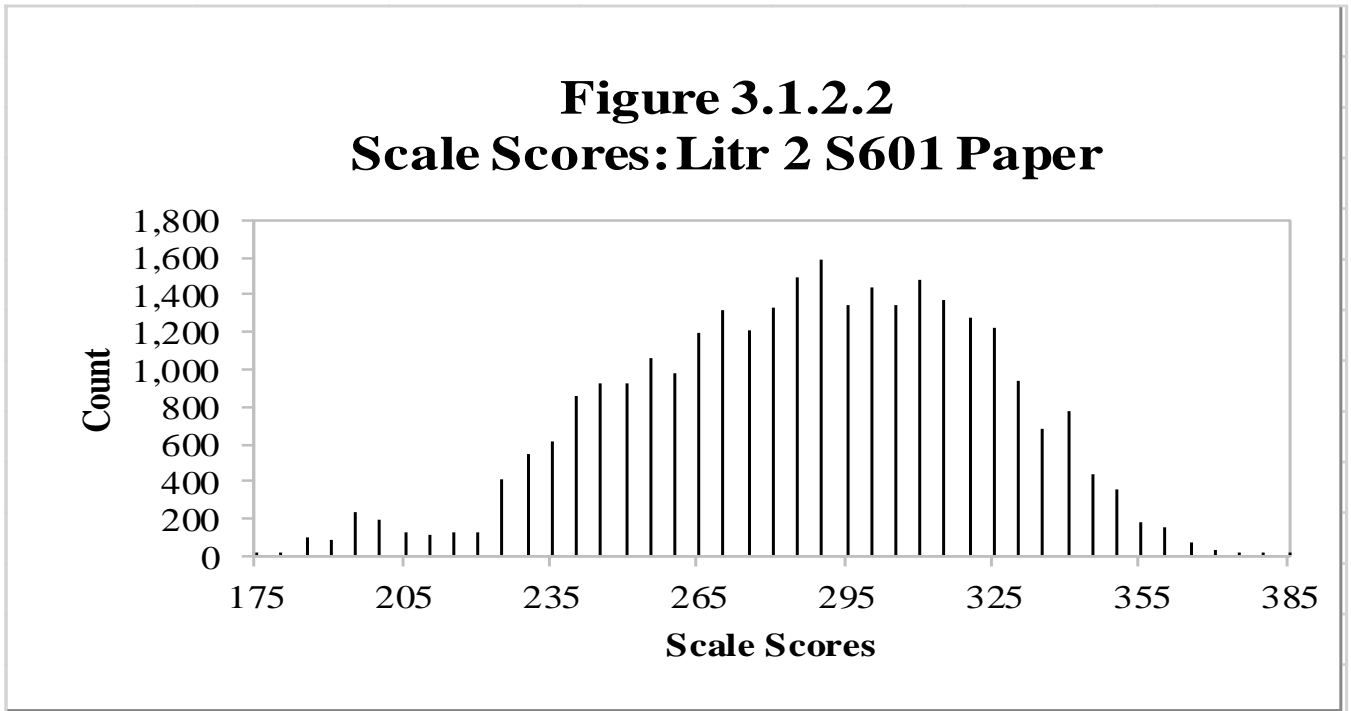
3.1.2.3 Grade 3

Table 3.1.2.3

Scale Score Descriptive Statistics: Litr 3 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	28,887	190	418	309.16	33.12

**Figure 3.1.2.3
Scale Scores: Litr 3 S601 Paper**
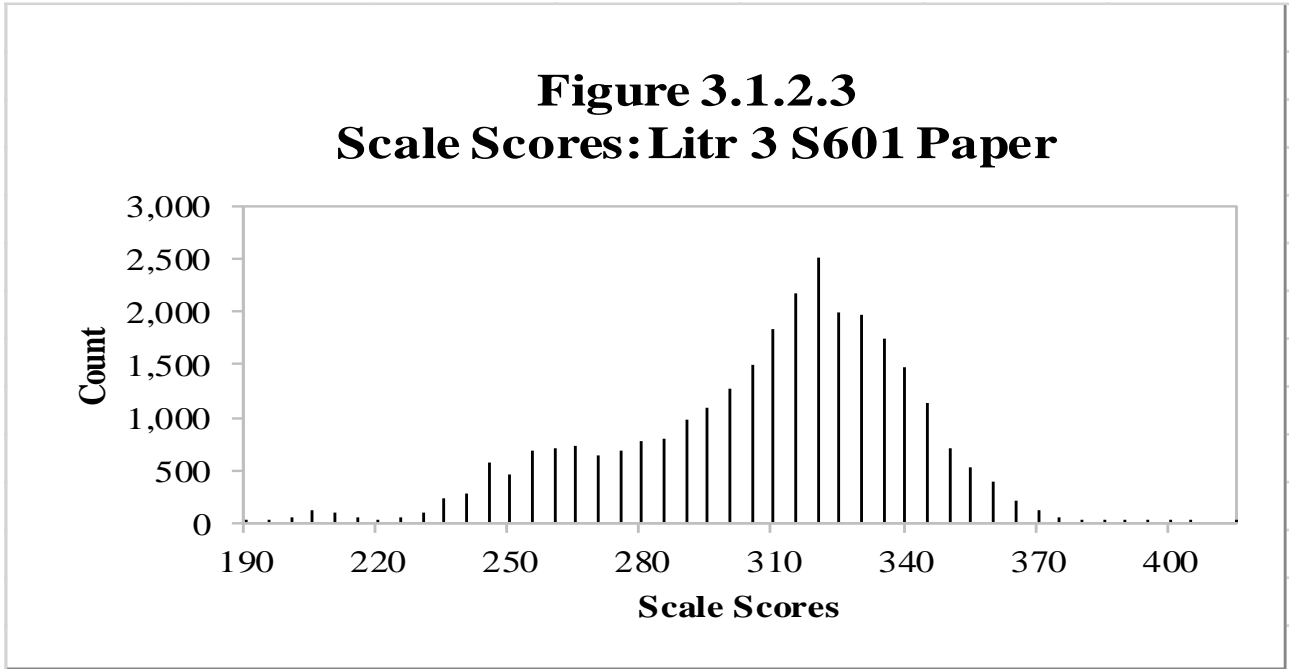
3.1.2.4 Grades 4-5

Table 3.1.2.4

Scale Score Descriptive Statistics: Litr 4-5 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	21,236	184	428	328.85	36.10
5	19,987	194	433	335.49	38.89
Total	41,223	184	433	332.07	37.62

Figure 3.1.2.4
Scale Scores: Litr 4-5 S601 Paper
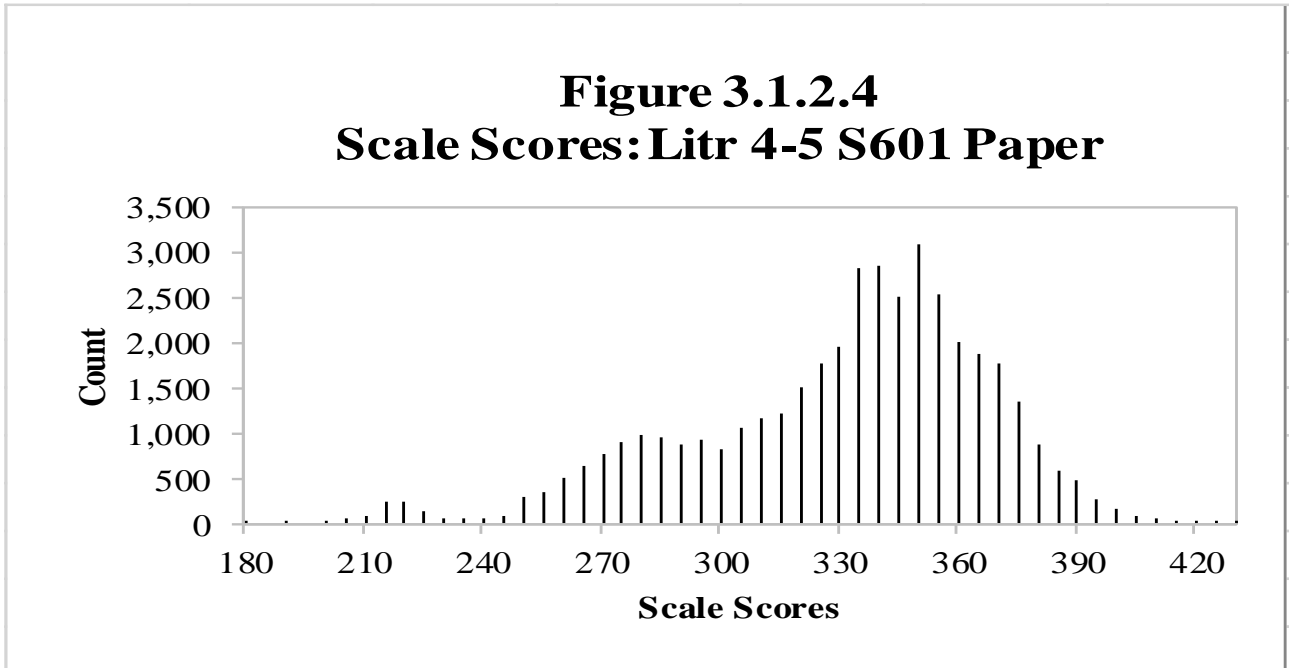
3.1.2.5 Grades 6-8

Table 3.1.2.5

Scale Score Descriptive Statistics: Litr 6-8 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	17,531	194	431	327.23	31.68
7	15,550	217	429	331.38	34.30
8	14,837	210	433	337.44	36.05
Total	47,918	194	433	331.74	34.19

Figure 3.1.2.5
Scale Scores: Litr 6-8 S601 Paper
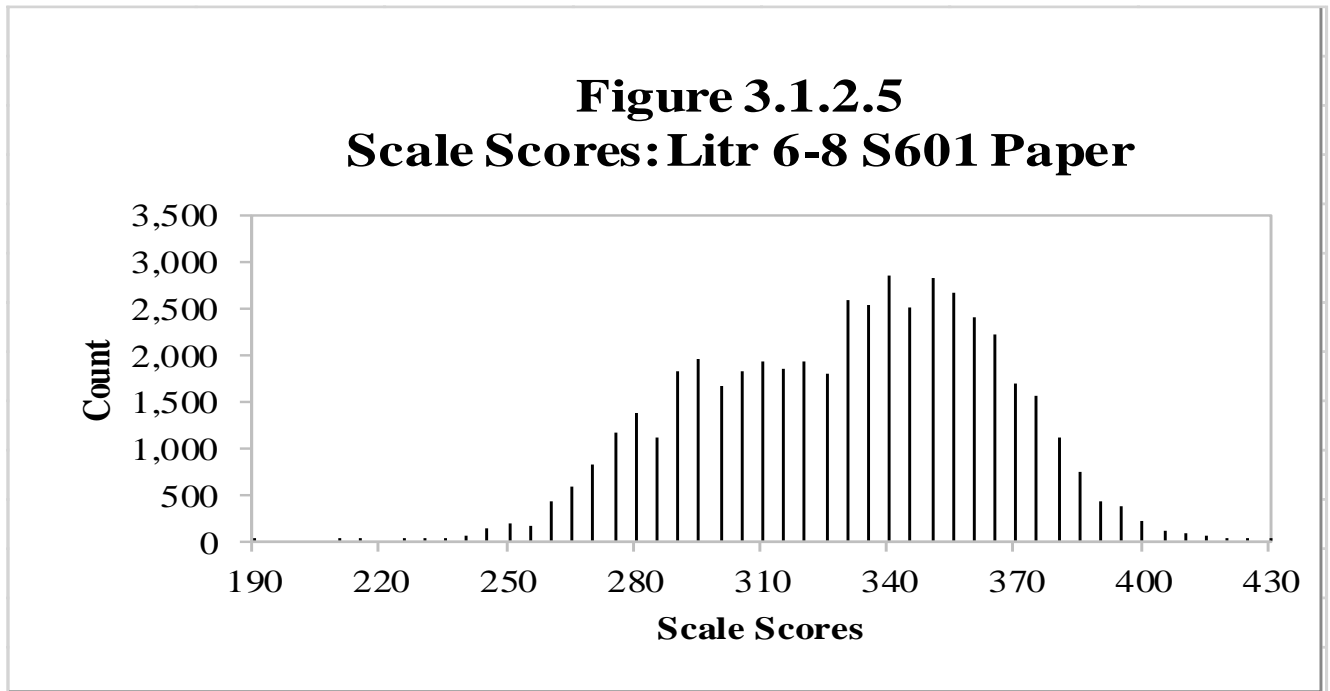
3.1.2.6 Grades 9-12

Table 3.1.2.6

Scale Score Descriptive Statistics: Litr 9-12 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	14,497	233	448	354.00	35.66
10	12,603	252	463	358.96	35.42
11	10,649	268	467	366.78	34.34
12	7,380	268	459	369.17	32.10
Total	45,129	233	467	360.88	35.23

Figure 3.1.2.6
Scale Scores: Litr 9-12 S601 Paper

3.1.3 Comprehension

3.1.3.0 Kindergarten

Table 3.1.3.0

Scale Score Descriptive Statistics: Cphn K S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	256,835	100	312	194.99	63.44

Figure 3.1.3.0
Scale Scores: Cphn K S601 Paper
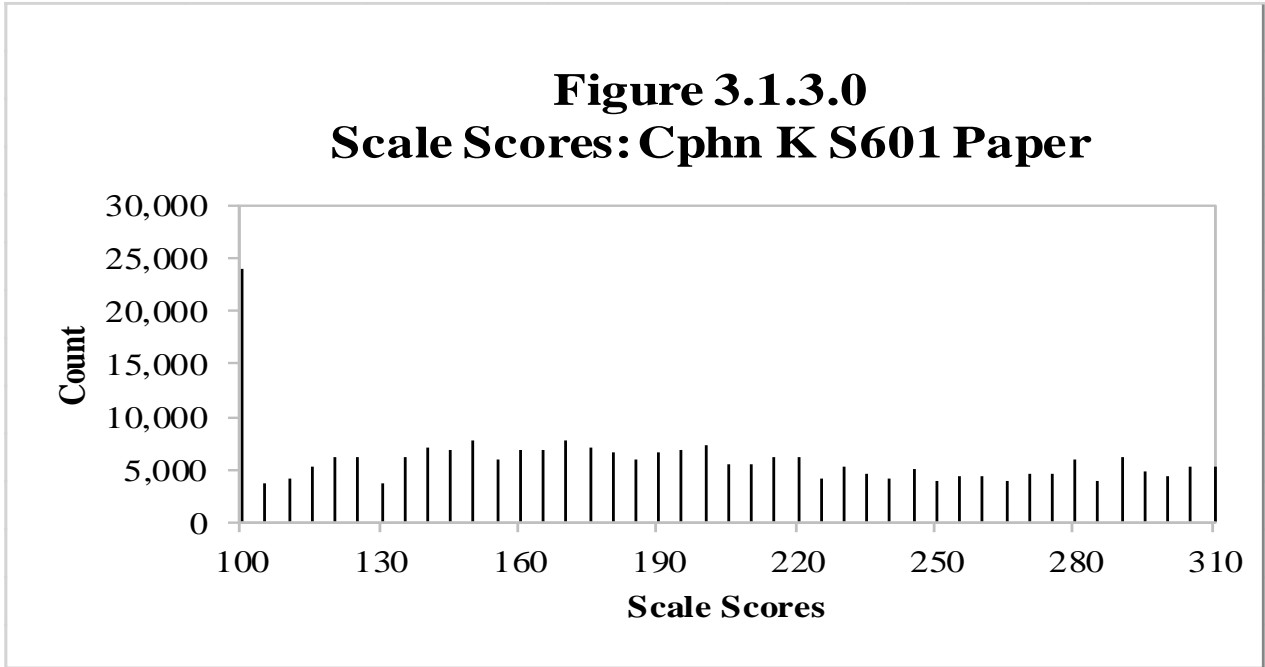
3.1.3.1 Grade 1

Table 3.1.3.1

Scale Score Descriptive Statistics: Cphn 1 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	27,121	130	397	288.49	28.01

Figure 3.1.3.1
Scale Scores: Cphn 1 S601 Paper
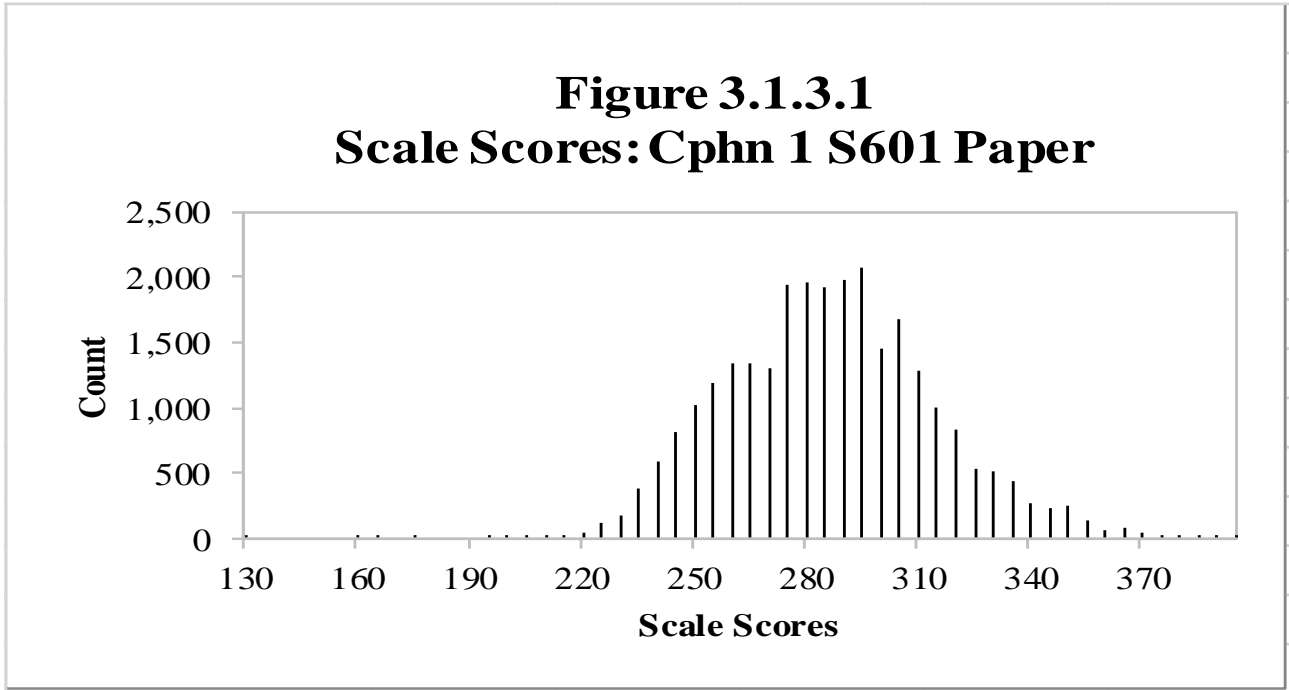
3.1.3.2 Grade 2

Table 3.1.3.2

Scale Score Descriptive Statistics: Cphn 2 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	27,122	206	397	311.79	33.78

Figure 3.1.3.2
Scale Scores: Cphn 2 S601 Paper
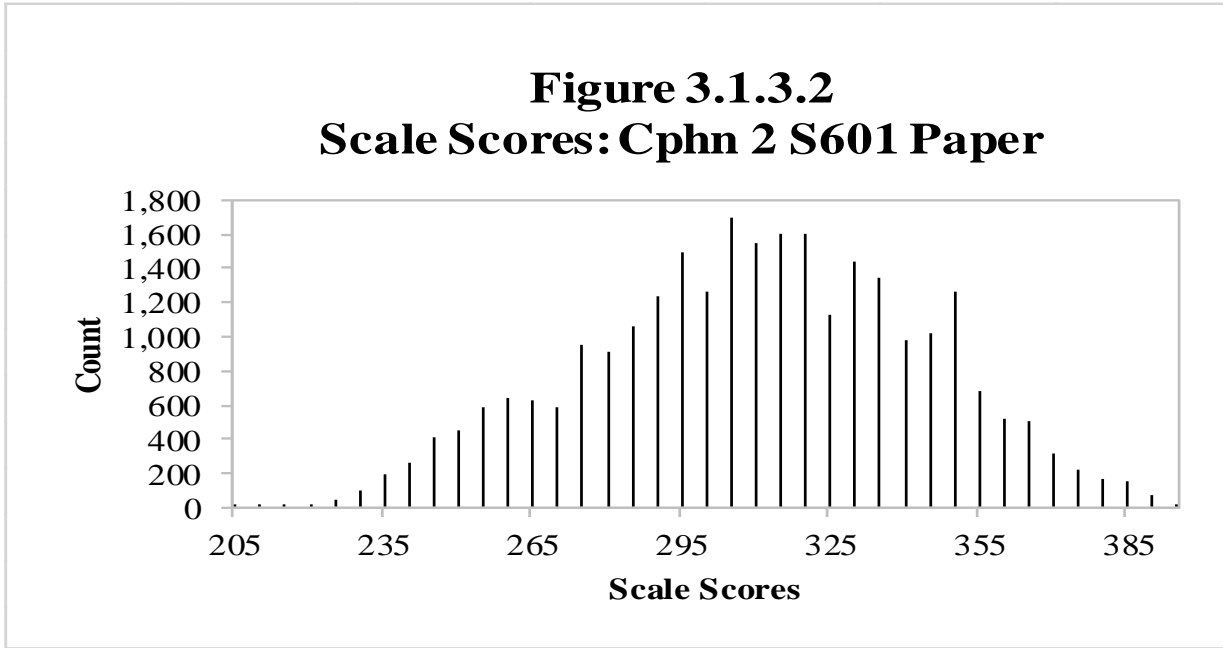
3.1.3.3 Grade 3

Table 3.1.3.3

Scale Score Descriptive Statistics: Cphn 3 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	27,514	212	448	336.58	28.93

Figure 3.1.3.3
Scale Scores: Cphn 3 S601 Paper
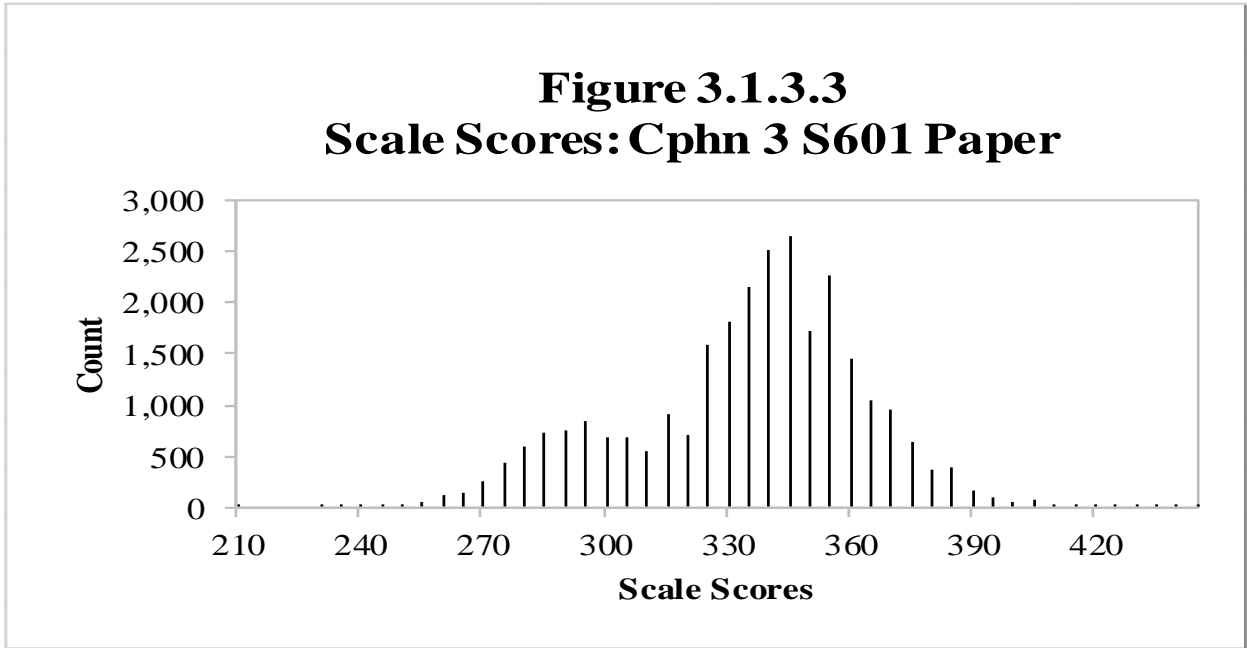
3.1.3.4 Grades 4-5

Table 3.1.3.4

Scale Score Descriptive Statistics: Cphn 4-5 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	20,536	183	448	348.04	33.38
5	19,409	228	453	353.39	35.90
Total	39,945	183	453	350.64	34.73

Figure 3.1.3.4
Scale Scores: Cphn 4-5 S601 Paper
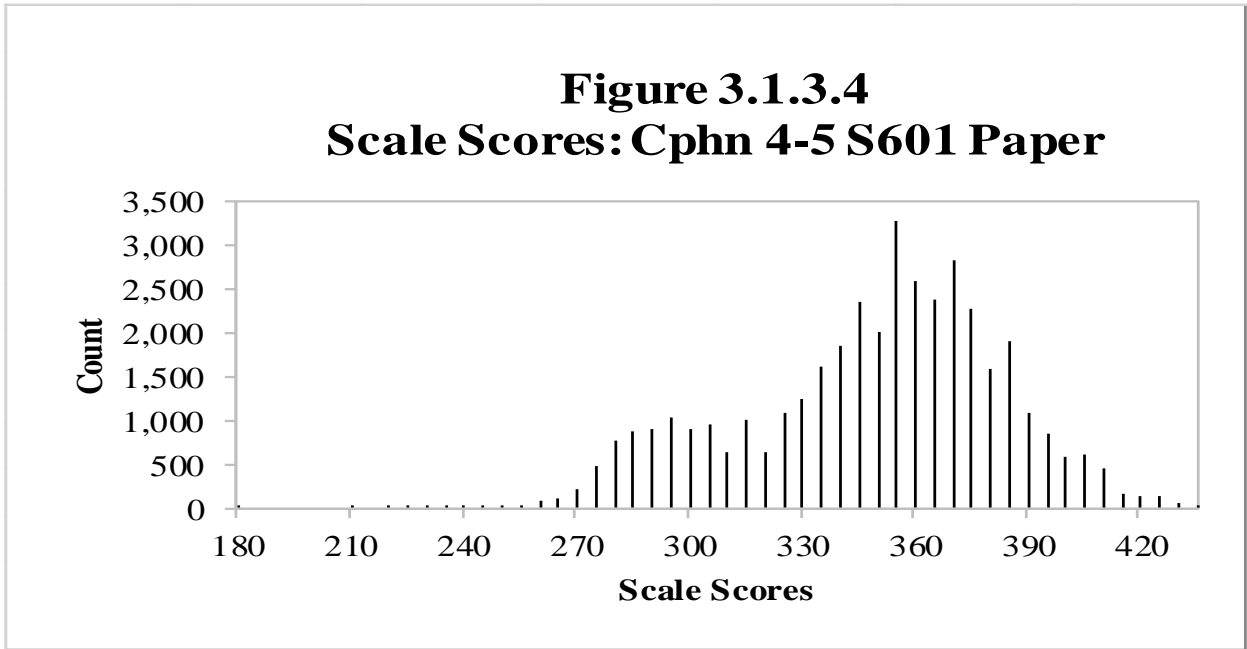
3.1.3.5 Grades 6-8

Table 3.1.3.5

Scale Score Descriptive Statistics: Cphn 6-8 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	16,835	194	453	350.55	29.68
7	14,961	250	459	354.82	33.20
8	14,234	254	459	361.10	35.43
Total	46,030	194	459	355.20	32.97

Figure 3.1.3.5
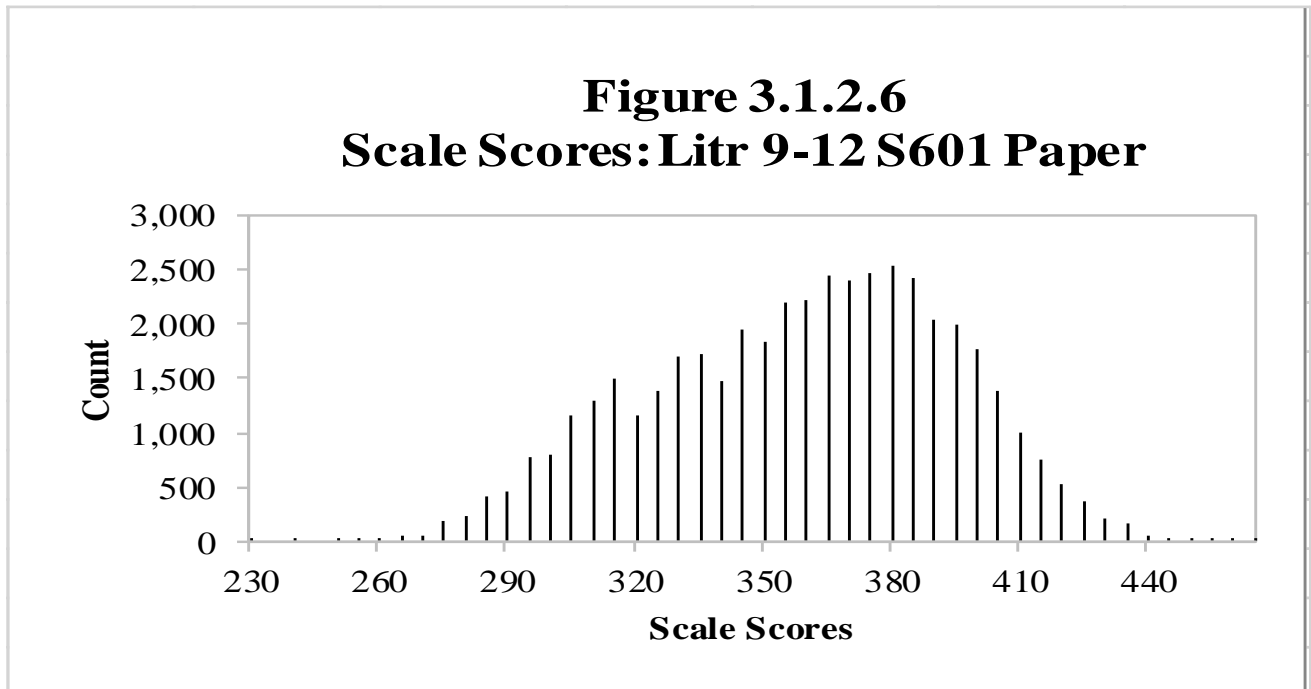## Scale Scores: Cphn 6-8 S601 Paper

3.1.3.6 Grades 9-12

Table 3.1.3.6

Scale Score Descriptive Statistics: Cphn 9-12 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	13,825	250	477	370.53	35.20
10	12,065	250	471	374.86	34.91
11	10,195	274	477	381.99	34.30
12	7,107	282	473	384.46	31.82
Total	43,192	250	477	376.73	34.80

Figure 3.1.3.6
Scale Scores: Cphn 9-12 S601 Paper

3.1.4 Overall

3.1.4.0 Kindergarten

Table 3.1.4.0

Scale Score Descriptive Statistics: Over K S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	256,812	100	333	200.38	63.29

Figure 3.1.4.0
Scale Scores: Over K S601 Paper
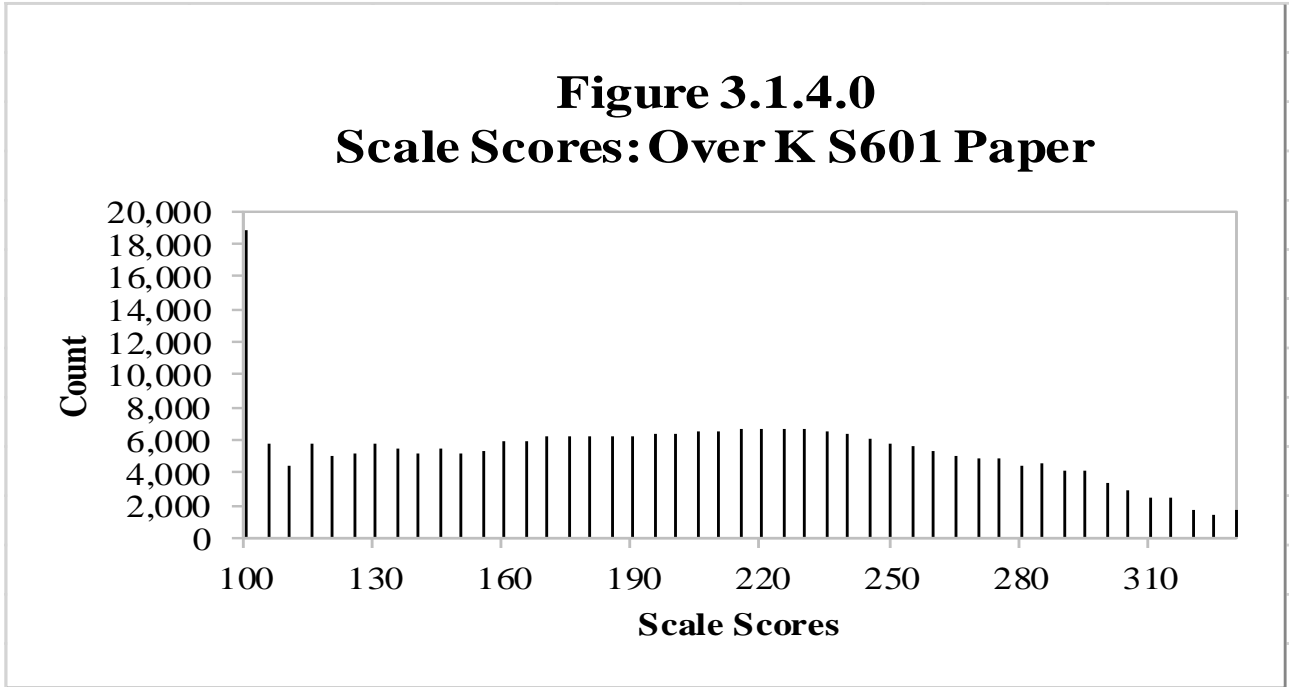
3.1.4.1 Grade 1

Table 3.1.4.1

Scale Score Descriptive Statistics: Over 1 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	26,915	133	374	269.08	33.49

Figure 3.1.4.1
Scale Scores: Over 1 S601 Paper
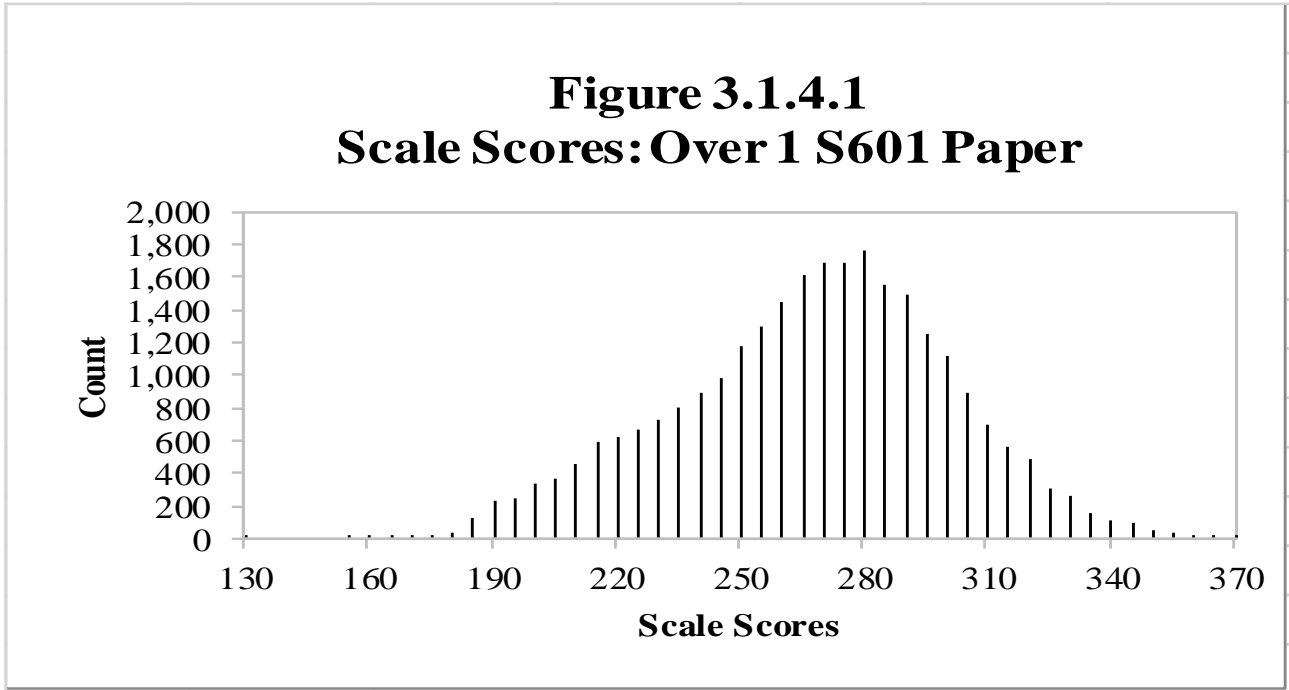
3.1.4.2 Grade 2

Table 3.1.4.2

Scale Score Descriptive Statistics: Over 2 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	26,929	175	394	292.83	38.25

Figure 3.1.4.2
Scale Scores: Over 2 S601 Paper
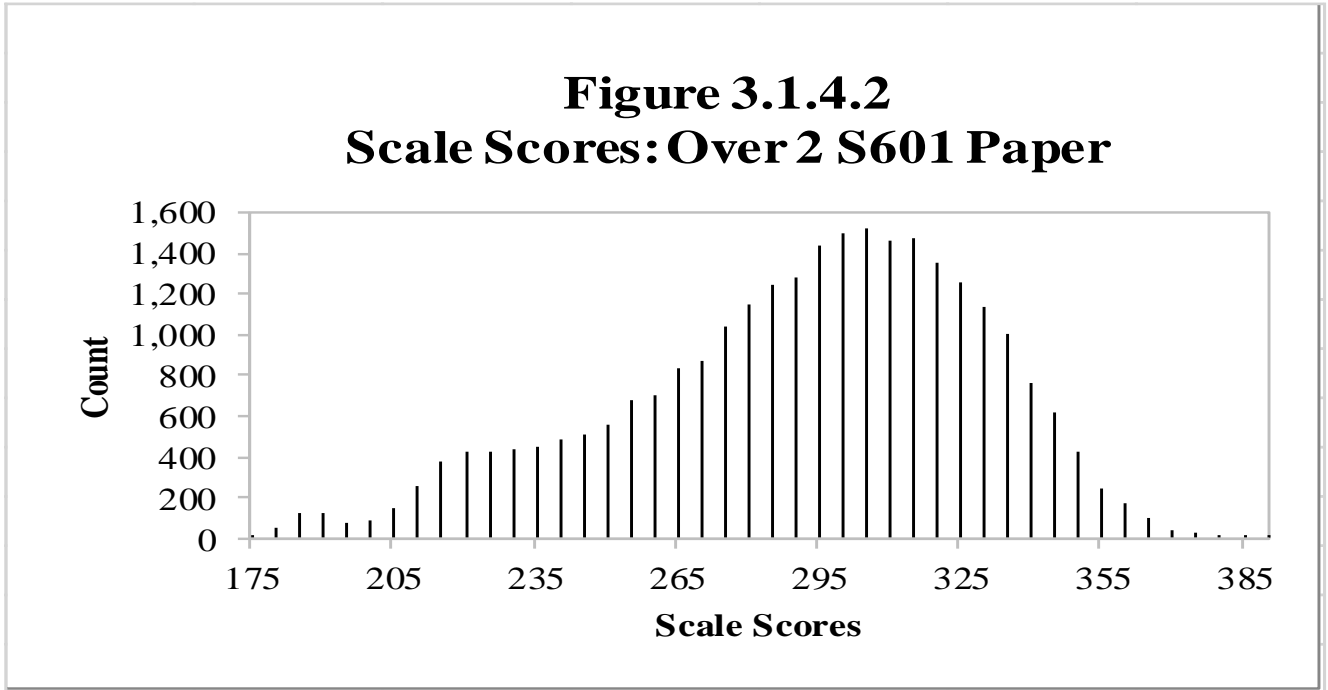
3.1.4.3 Grade 3

Table 3.1.4.3

Scale Score Descriptive Statistics: Over 3 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	27,343	189	418	312.64	35.48

**Figure 3.1.4.3**
Scale Scores: Over 3 S601 Paper
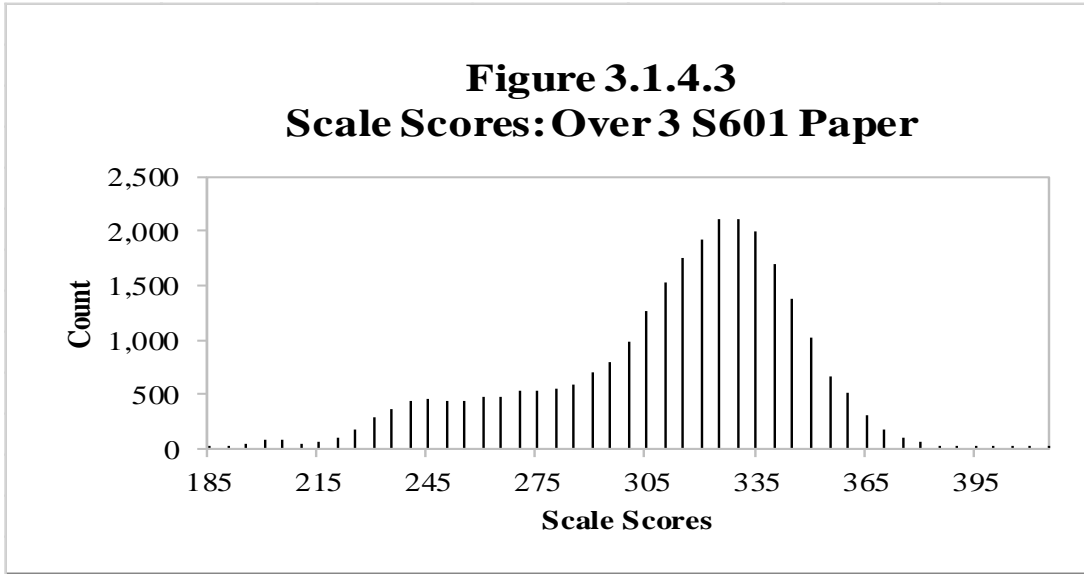
3.1.4.4 Grades 4-5

Table 3.1.4.4

Scale Score Descriptive Statistics: Over 4-5 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	20,417	186	429	332.96	40.29
5	19,317	193	435	338.86	43.00
Total	39,734	186	435	335.82	41.73

Figure 3.1.4.4
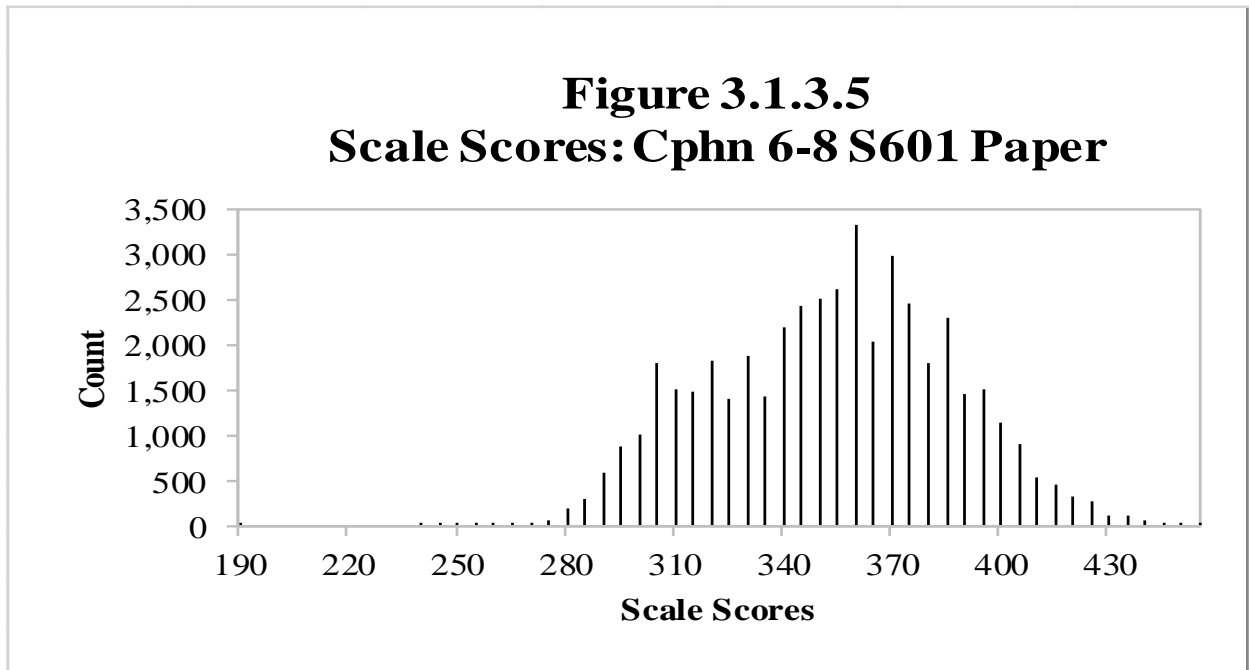**Scale Scores: Over 4-5 S601 Paper**
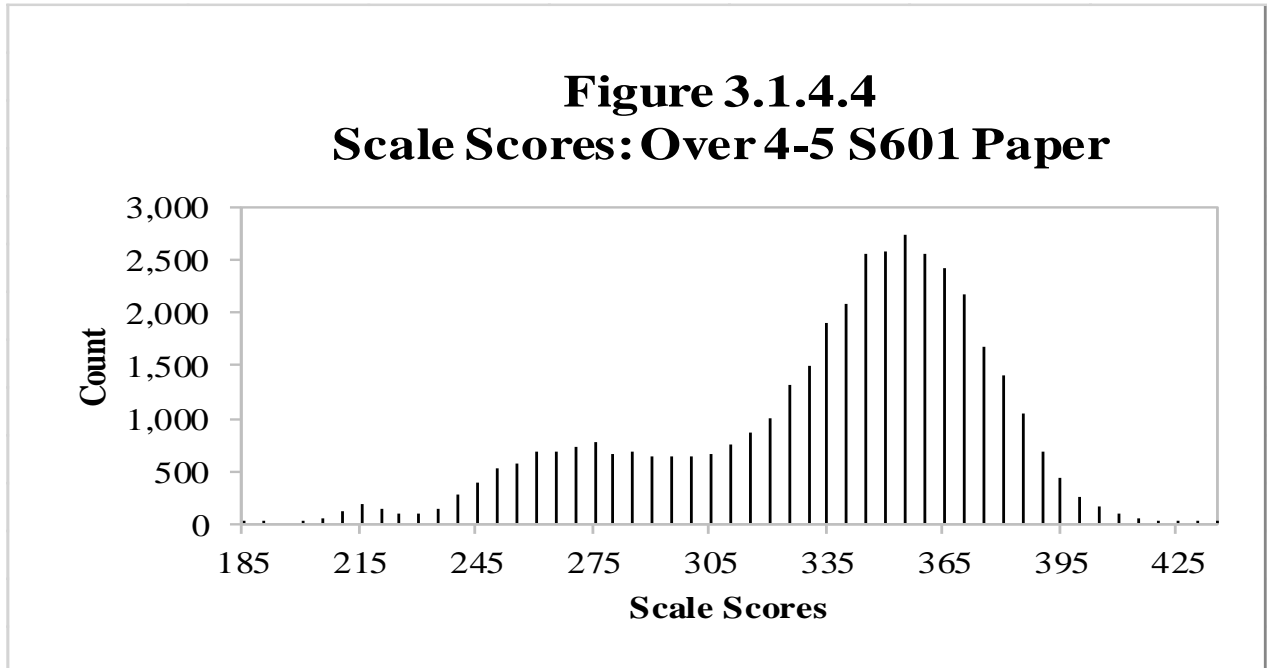
3.1.4.5 Grades 6-8

Table 3.1.4.5

Scale Score Descriptive Statistics: Over 6-8 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	16,694	185	429	333.41	37.50
7	14,828	218	437	336.67	40.69
8	14,095	224	440	342.71	42.29
Total	45,617	185	440	337.35	40.25

Figure 3.1.4.5
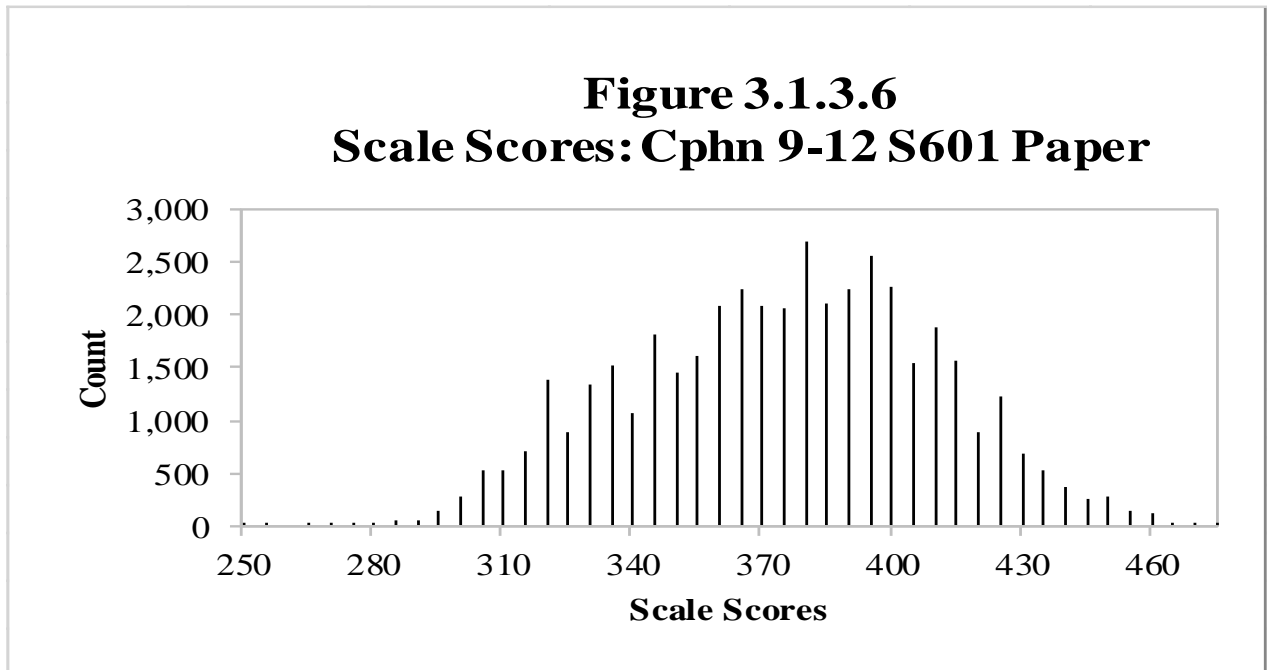**Scale Scores: Over 6-8 S601 Paper**
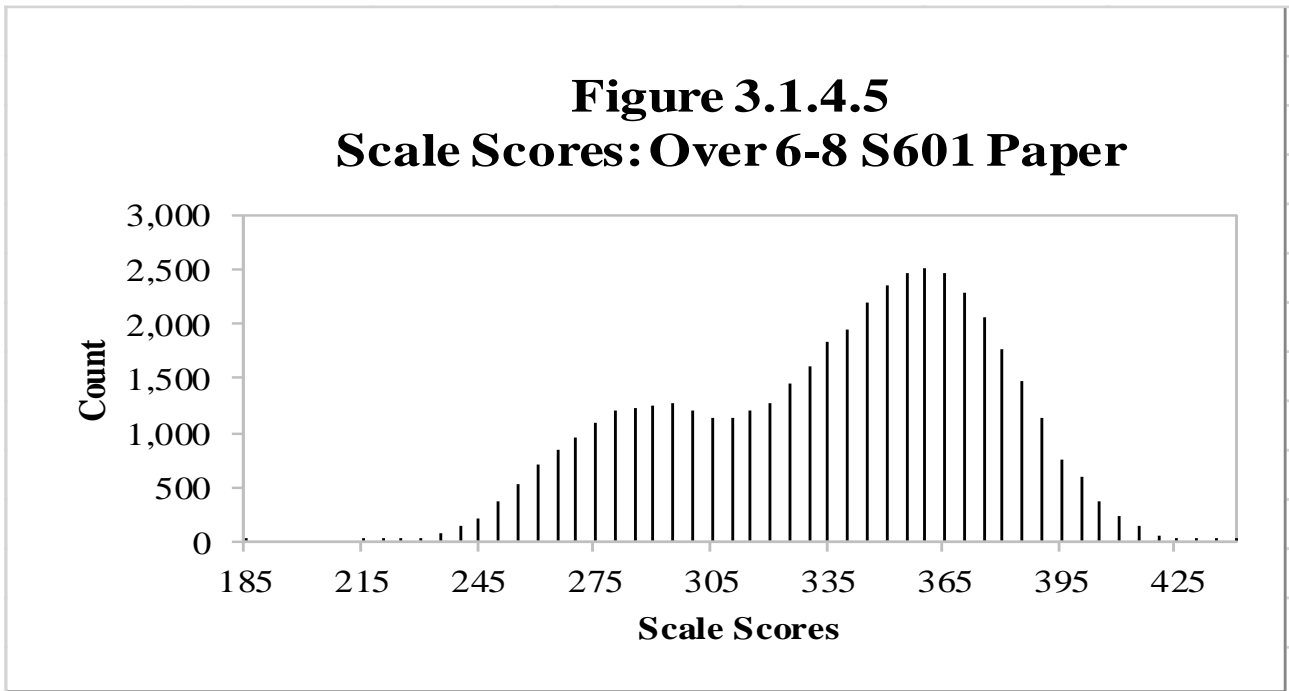
3.1.4.6 Grades 9-12

Table 3.1.4.6

Scale Score Descriptive Statistics: Over 9-12 S601 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	13,683	238	455	352.50	42.21
10	11,955	239	465	356.97	41.41
11	10,102	250	466	365.98	39.56
12	7,041	255	462	369.73	35.82
Total	42,781	238	466	359.77	40.92

Figure 3.1.4.6
Scale Scores: Over 9-12 S601 Paper

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 S601 Paper

Level	Count	Percent
1	91,025	35.44%
2	36,608	14.25%
3	28,057	10.92%
4	21,095	8.21%
5	37,112	14.45%
6	42,930	16.72%
Total	256,827	100.00%

Figure 3.2.1.0
Proficiency Level: Oral K S601 Paper
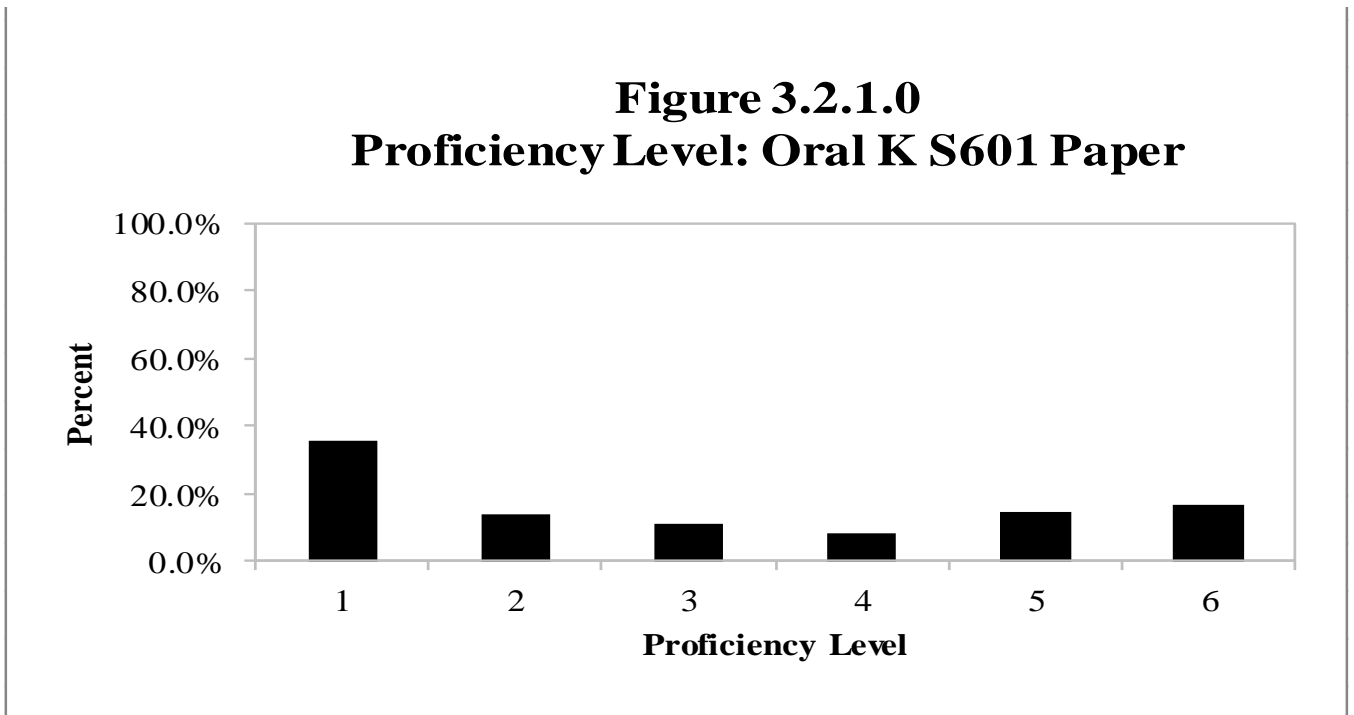
3.2.1.1 Grade 1

Table 3.2.1.1

Proficiency Level Distribution: Oral 1 S601 Paper

Level	Count	Percent
1	4,398	13.25%
2	5,559	16.74%
3	10,095	30.41%
4	7,575	22.82%
5	4,634	13.96%
6	940	2.83%
Total	33,201	100.00%

Figure 3.2.1.1
Proficiency Level: Oral 1 S601 Paper
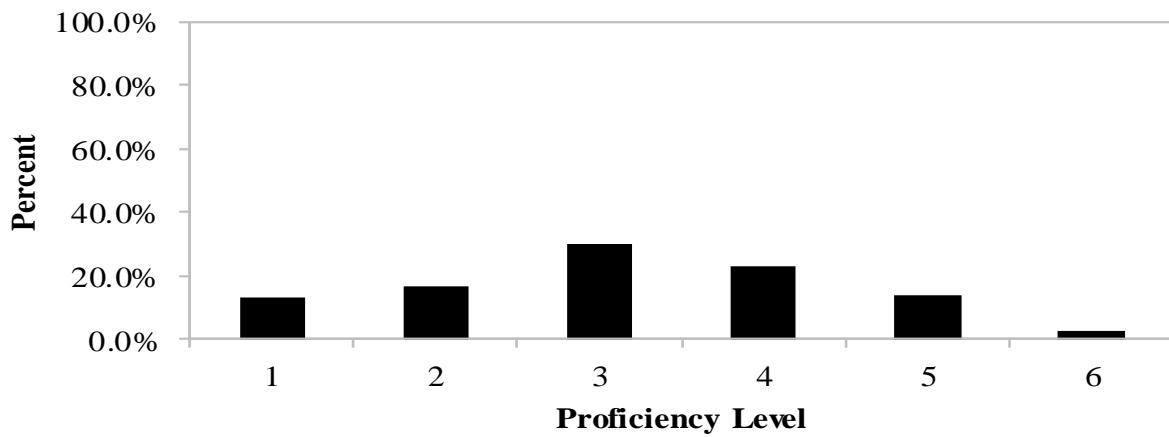
3.2.1.2 Grade 2

Table 3.2.1.2

Proficiency Level Distribution: Oral 2 S601 Paper

Level	Count	Percent
1	4,184	13.05%
2	4,278	13.34%
3	9,563	29.82%
4	9,190	28.66%
5	3,773	11.77%
6	1,081	3.37%
Total	32,069	100.00%

Figure 3.2.1.2
Proficiency Level: Oral 2 S601 Paper
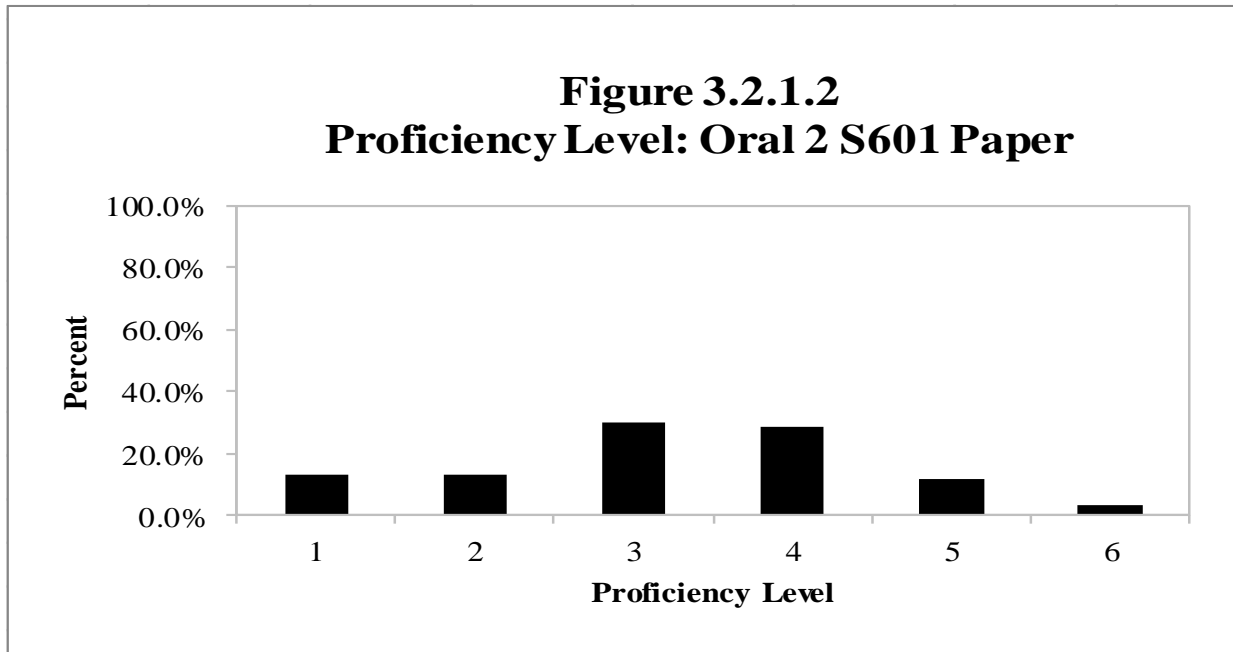
3.2.1.3 Grade 3

Table 3.2.1.3

Proficiency Level Distribution: Oral 3 S601 Paper

Level	Count	Percent
1	3,623	11.18%
2	3,934	12.14%
3	9,748	30.07%
4	10,170	31.38%
5	4,028	12.43%
6	910	2.81%
Total	32,413	100.00%

Figure 3.2.1.3
Proficiency Level: Oral 3 S601 Paper
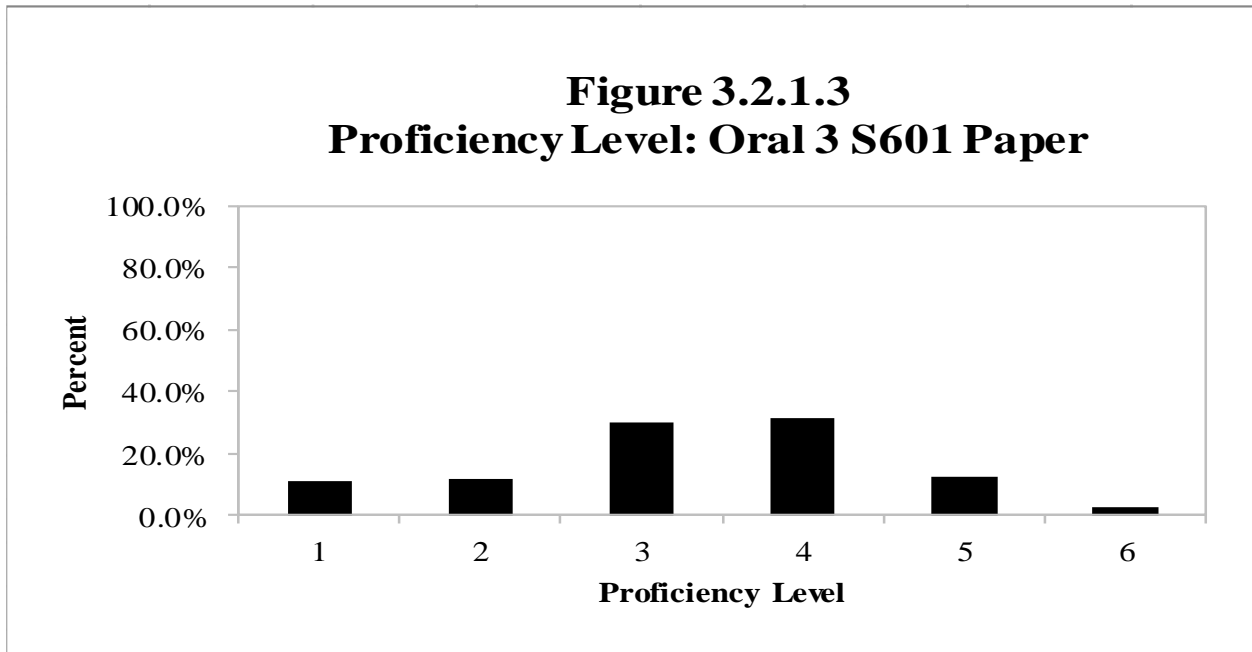
3.2.1.4 Grades 4-5

Table 3.2.1.4

Proficiency Level Distribution: Oral 4-5 S601 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	3,111	13.28%	3,444	16.00%	6,555	14.58%
2	2,175	9.29%	2,176	10.11%	4,351	9.68%
3	4,125	17.61%	3,753	17.43%	7,878	17.52%
4	7,445	31.79%	6,956	32.31%	14,401	32.04%
5	4,727	20.18%	3,952	18.35%	8,679	19.31%
6	1,839	7.85%	1,250	5.81%	3,089	6.87%
Total	23,422	100.00%	21,531	100.00%	44,953	100.00%

Figure 3.2.1.4
Proficiency Level: Oral 4-5 S601 Paper
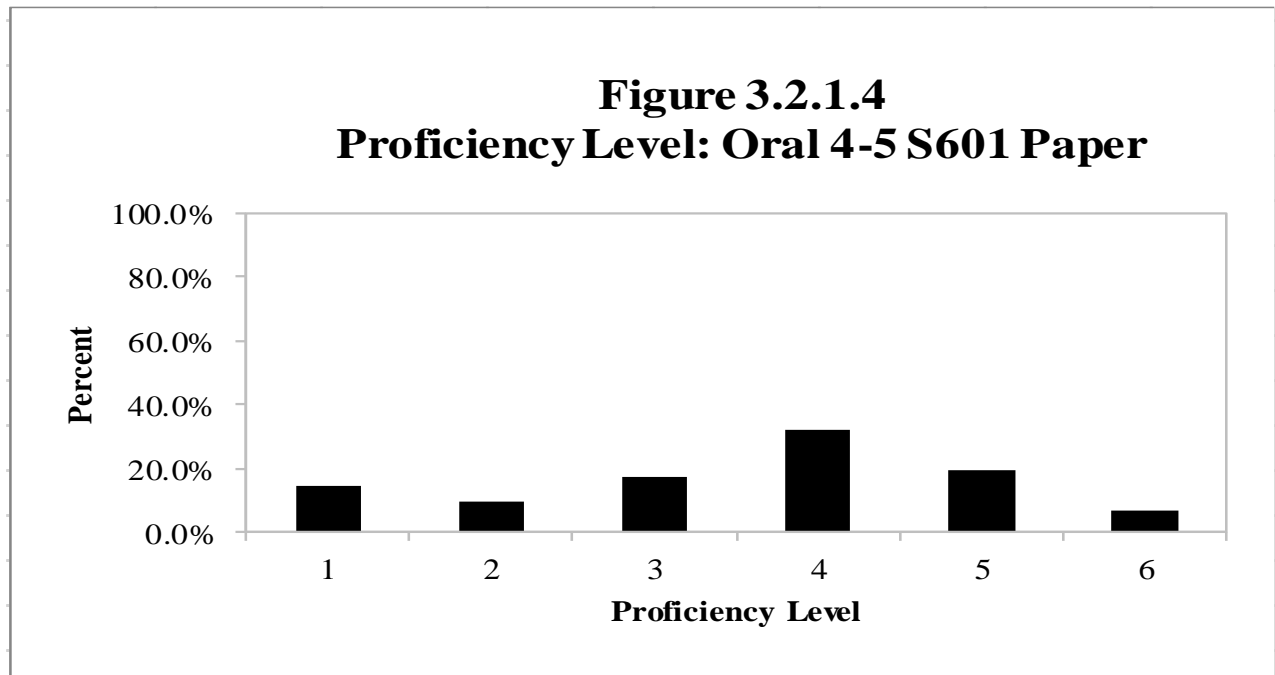
3.2.1.5 Grades 6-8

Table 3.2.1.5

Proficiency Level Distribution: Oral 6-8 S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	3,025	16.33%	3,483	21.41%	3,403	22.19%	9,911	19.77%
2	2,512	13.56%	2,348	14.43%	2,096	13.67%	6,956	13.88%
3	4,154	22.42%	3,319	20.40%	3,185	20.77%	10,658	21.26%
4	5,267	28.43%	4,389	26.98%	4,068	26.52%	13,724	27.38%
5	2,621	14.15%	2,038	12.53%	1,987	12.96%	6,646	13.26%
6	946	5.11%	692	4.25%	598	3.90%	2,236	4.46%
Total	18,525	100.00%	16,269	100.00%	15,337	100.00%	50,131	100.00%

Figure 3.2.1.5
Proficiency Level: Oral 6-8 S601 Paper
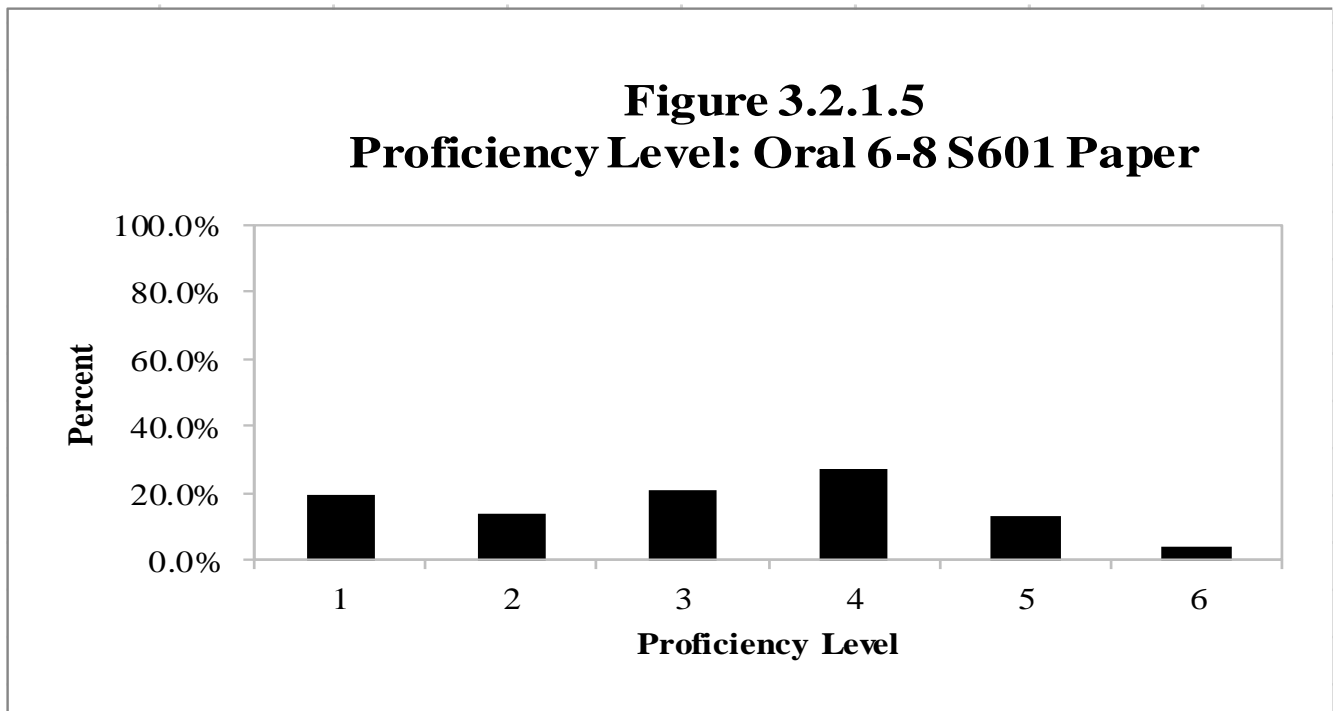
3.2.1.6 Grades 9-12

Table 3.2.1.6

Proficiency Level Distribution: Oral 9-12 S601 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	4,209	27.78%	3,601	27.54%	2,589	23.29%	1,493	19.68%	11,892	25.34%
2	2,186	14.43%	2,007	15.35%	1,667	15.00%	1,152	15.18%	7,012	14.94%
3	3,425	22.60%	3,358	25.68%	3,159	28.42%	2,479	32.67%	12,421	26.47%
4	3,568	23.55%	2,878	22.01%	2,638	23.73%	1,915	25.24%	10,999	23.44%
5	1,385	9.14%	925	7.07%	838	7.54%	395	5.21%	3,543	7.55%
6	380	2.51%	307	2.35%	225	2.02%	153	2.02%	1,065	2.27%
Total	15,153	100.00%	13,076	100.00%	11,116	100.00%	7,587	100.00%	46,932	100.00%

Figure 3.2.1.6
Proficiency Level: Oral 9-12 S601 Paper
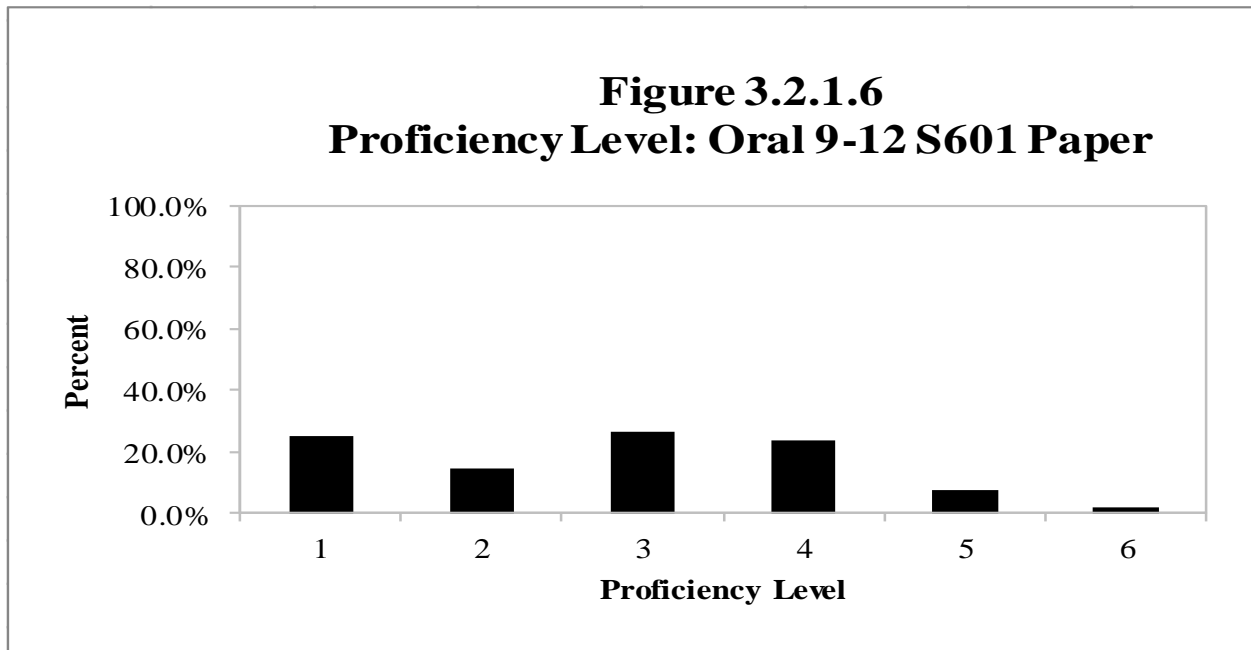
3.2.2 Literacy

3.2.2.0 Kindergarten

Table 3.2.2.0

Proficiency Level Distribution: Litr K S601 Paper

Level	Count	Percent
1	202,587	78.88%
2	25,813	10.05%
3	20,080	7.82%
4	8,349	3.25%
5	0	0.00%
6	0	0.00%
Total	256,829	100.00%

Figure 3.2.2.0
Proficiency Level: Litr K S601 Paper
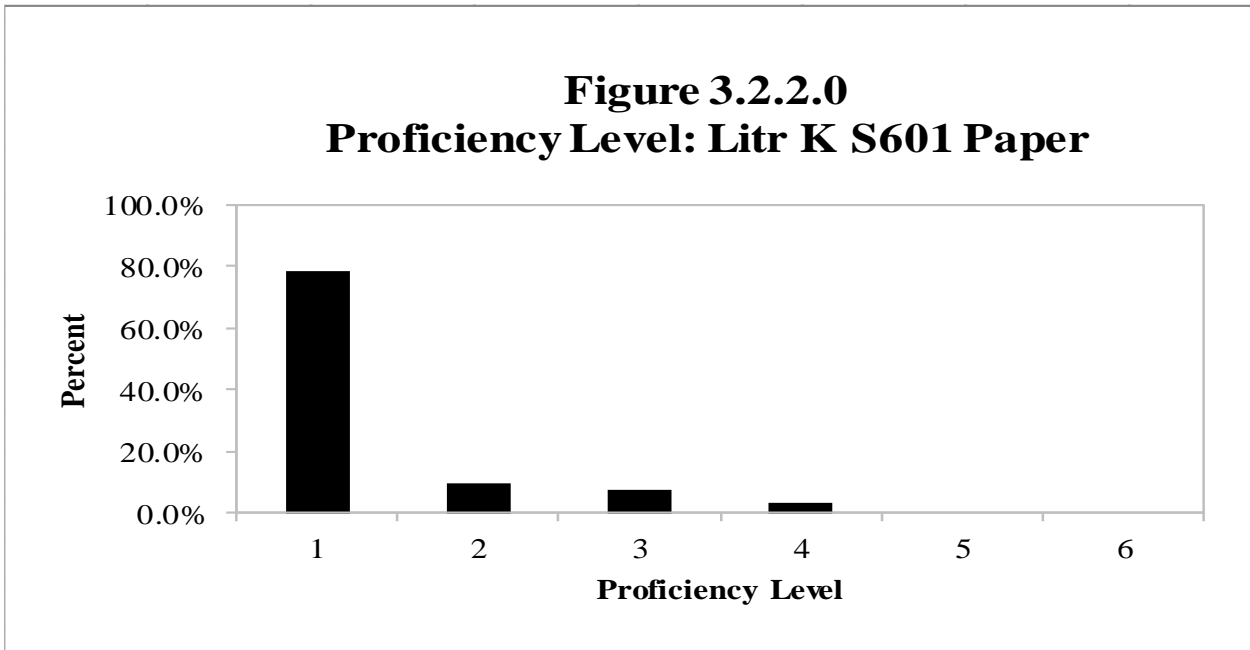
3.2.2.1 Grade 1

Table 3.2.2.1

Proficiency Level Distribution: Litr 1 S601 Paper

Level	Count	Percent
1	10,141	33.46%
2	11,627	38.36%
3	7,632	25.18%
4	838	2.76%
5	69	0.23%
6	3	0.01%
Total	30,310	100.00%

Figure 3.2.2.1
Proficiency Level: Litr 1 S601 Paper
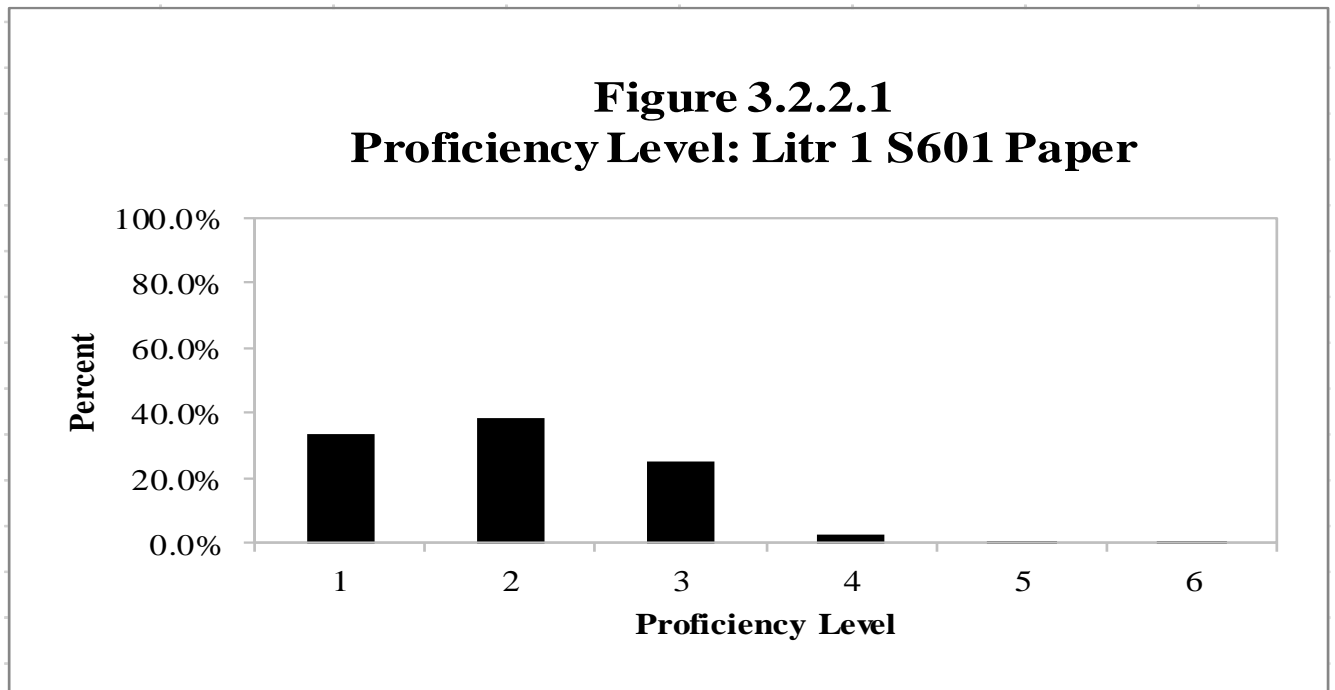
3.2.2.2 Grade 2

Table 3.2.2.2

Proficiency Level Distribution: Litr 2 S601 Paper

Level	Count	Percent
1	7,050	24.53%
2	7,939	27.62%
3	10,730	37.33%
4	2,857	9.94%
5	163	0.57%
6	6	0.02%
Total	28,745	100.00%

Figure 3.2.2.2
Proficiency Level: Litr 2 S601 Paper
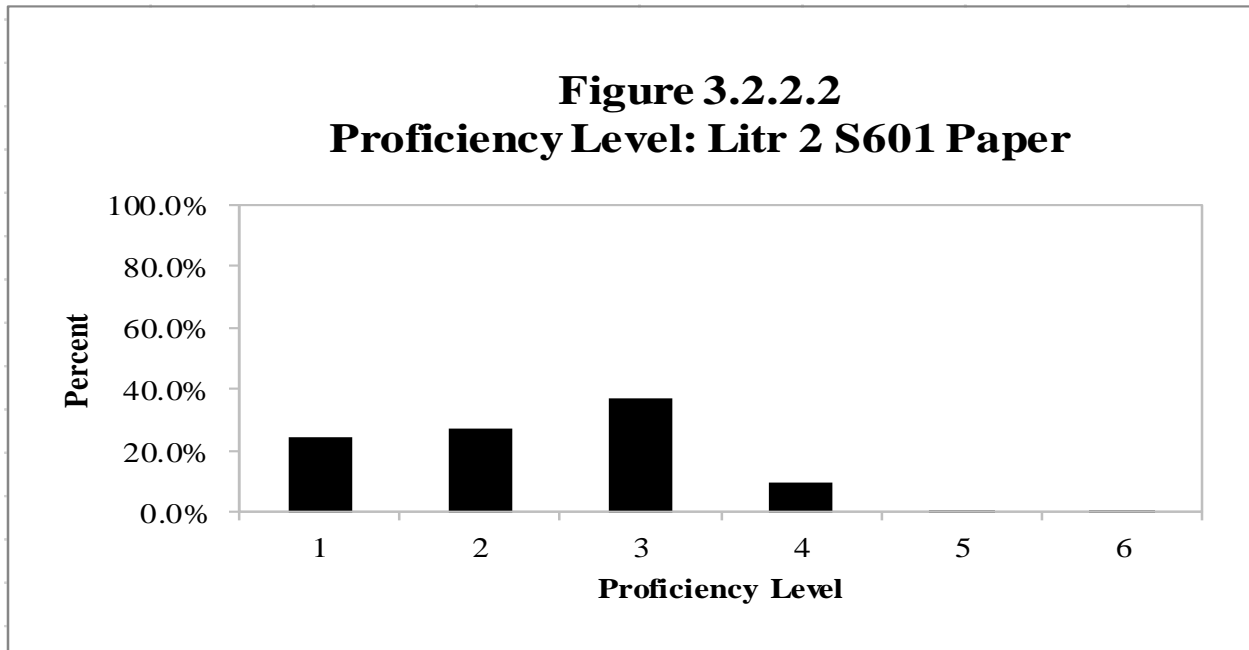
3.2.2.3 Grade 3

Table 3.2.2.3

Proficiency Level Distribution: Litr 3 S601 Paper

Level	Count	Percent
1	4,554	15.76%
2	5,452	18.87%
3	15,401	53.31%
4	3,304	11.44%
5	159	0.55%
6	17	0.06%
Total	28,887	100.00%

Figure 3.2.2.3
Proficiency Level: Litr 3 S601 Paper
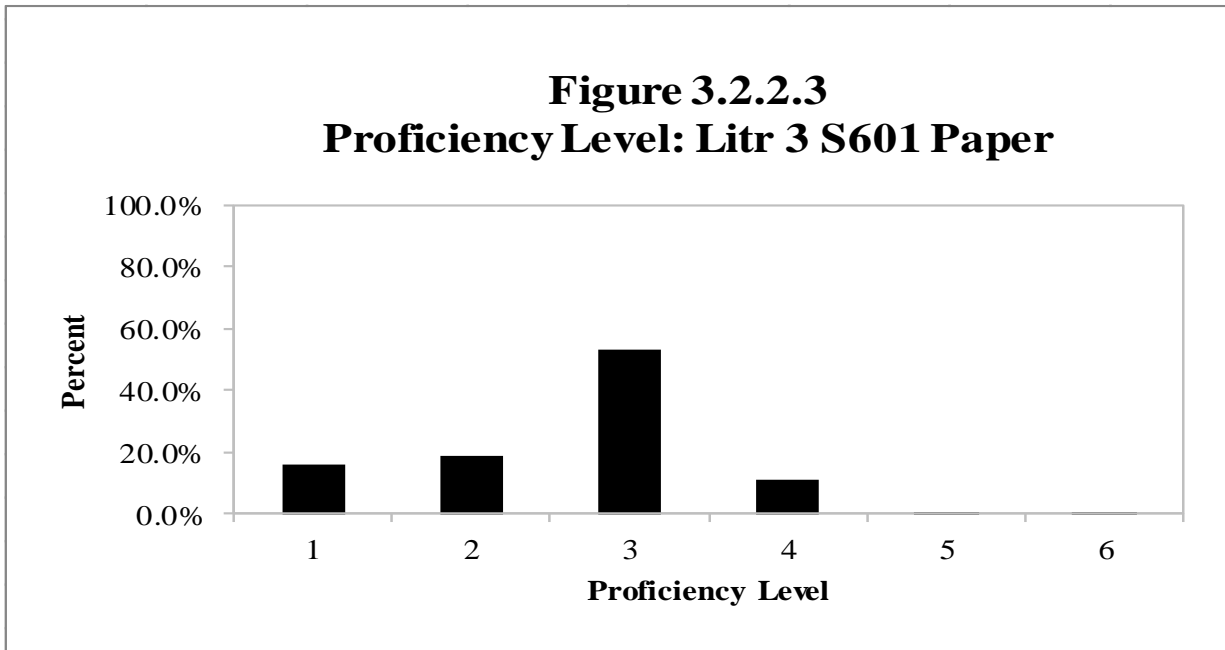
3.2.2.4 Grades 4-5

Table 3.2.2.4

Proficiency Level Distribution: Litr 4-5 S601 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	3,140	14.79%	3,193	15.98%	6,333	15.36%
2	2,378	11.20%	2,617	13.09%	4,995	12.12%
3	10,151	47.80%	8,209	41.07%	18,360	44.54%
4	5,051	23.79%	5,102	25.53%	10,153	24.63%
5	477	2.25%	801	4.01%	1,278	3.10%
6	39	0.18%	65	0.33%	104	0.25%
Total	21,236	100.00%	19,987	100.00%	41,223	100.00%

Figure 3.2.2.4
Proficiency Level: Litr 4-5 S601 Paper
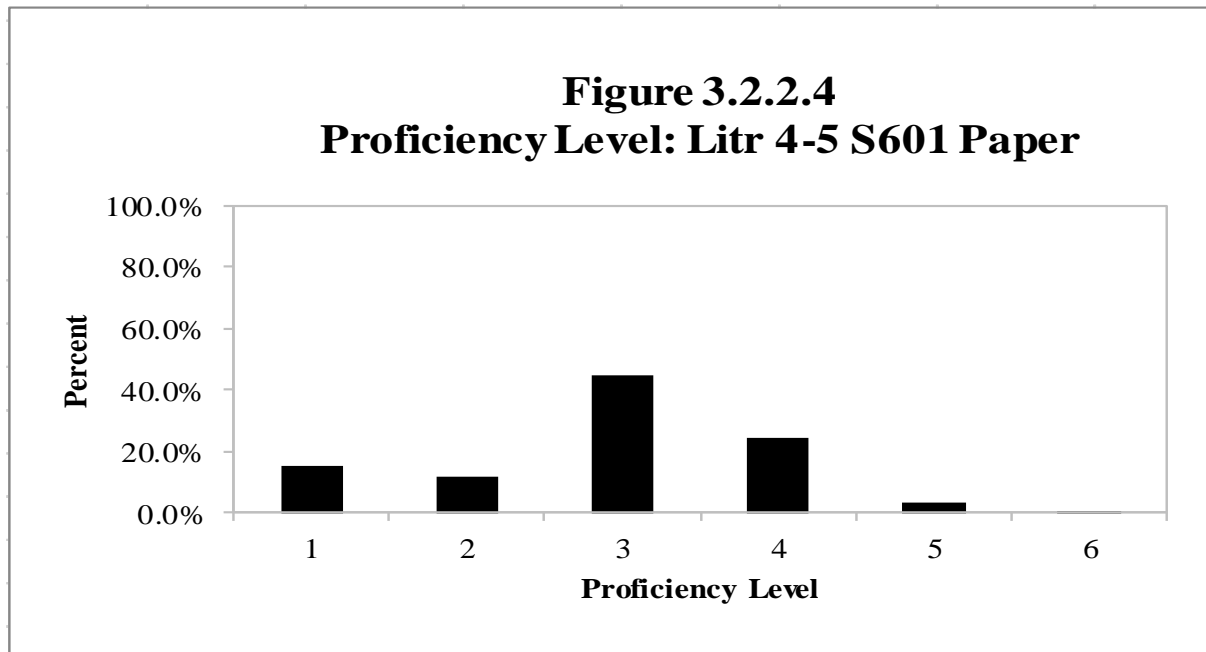
3.2.2.5 Grades 6-8

Table 3.2.2.5

Proficiency Level Distribution: Litr 6-8 S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	3,265	18.62%	3,396	21.84%	3,588	24.18%	10,249	21.39%
2	4,309	24.58%	3,985	25.63%	3,369	22.71%	11,663	24.34%
3	8,317	47.44%	6,623	42.59%	6,102	41.13%	21,042	43.91%
4	1,554	8.86%	1,465	9.42%	1,683	11.34%	4,702	9.81%
5	82	0.47%	80	0.51%	95	0.64%	257	0.54%
6	4	0.02%	1	0.01%	0	0.00%	5	0.01%
Total	17,531	100.00%	15,550	100.00%	14,837	100.00%	47,918	100.00%

Figure 3.2.2.5
Proficiency Level: Litr 6-8 S601 Paper
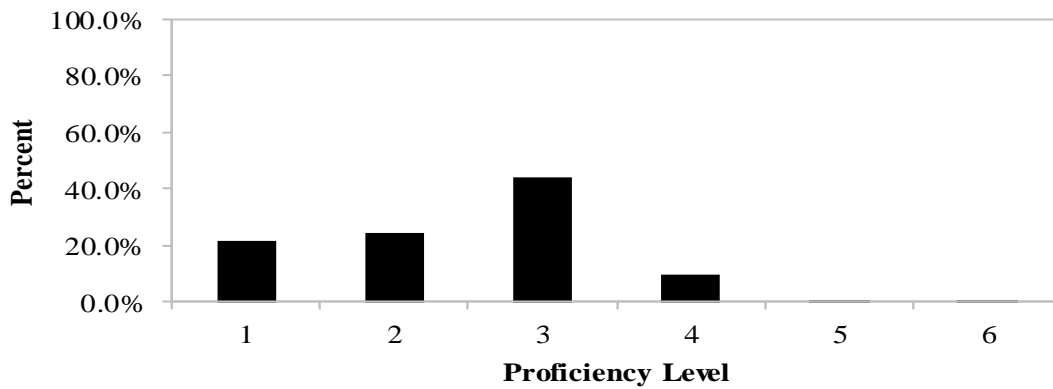
3.2.2.6 Grades 9-12

Table 3.2.2.6

Proficiency Level Distribution: Litr 9-12 S601 Paper

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	2,454	16.93%	2,177	17.27%	1,630	15.31%	1,159	15.70%	7,420	16.44%
2	3,271	22.56%	2,917	23.15%	2,425	22.77%	1,942	26.31%	10,555	23.39%
3	5,632	38.85%	4,937	39.17%	4,350	40.85%	3,166	42.90%	18,085	40.07%
4	2,776	19.15%	2,254	17.88%	2,012	18.89%	1,013	13.73%	8,055	17.85%
5	360	2.48%	315	2.50%	231	2.17%	100	1.36%	1,006	2.23%
6	4	0.03%	3	0.02%	1	0.01%	0	0.00%	8	0.02%
Total	14,497	100.00%	12,603	100.00%	10,649	100.00%	7,380	100.00%	45,129	100.00%

Figure 3.2.2.6
Proficiency Level: Litr 9-12 S601 Paper
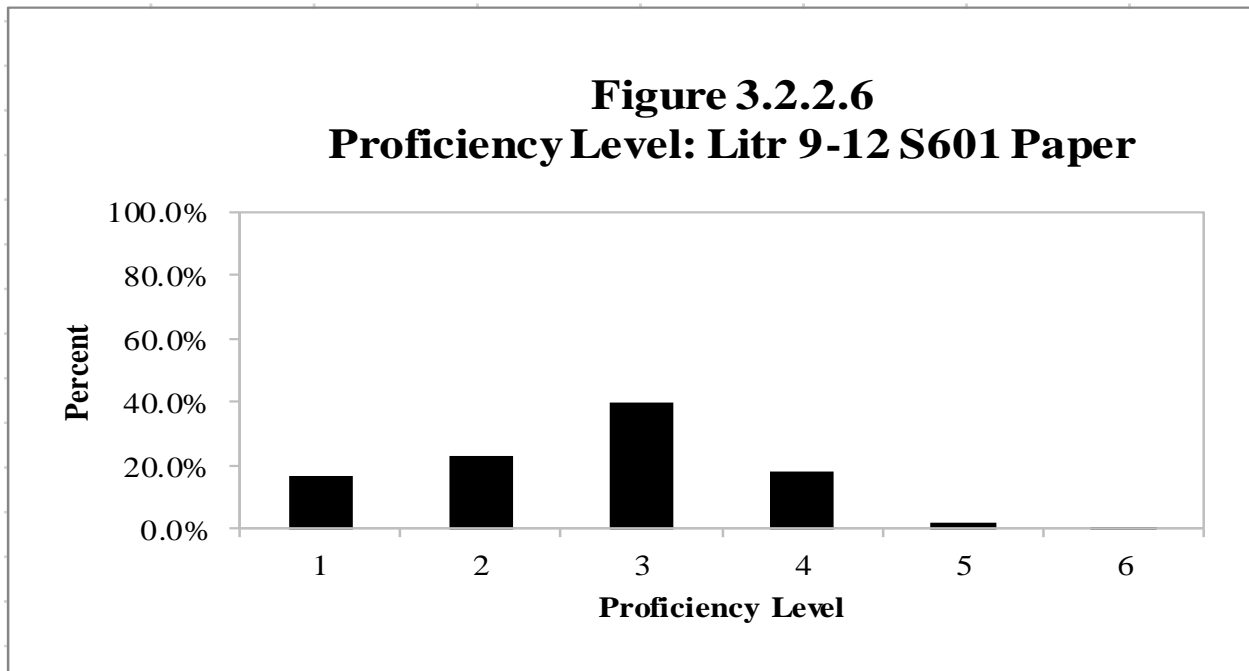
3.2.3 Comprehension

3.2.3.0 Kindergarten

Table 3.2.3.0

Proficiency Level Distribution: Cphn K S601 Paper

Level	Count	Percent
1	183,949	71.62%
2	17,497	6.81%
3	18,902	7.36%
4	9,470	3.69%
5	21,728	8.46%
6	5,289	2.06%
Total	256,835	100.00%

Figure 3.2.3.0
Proficiency Level: Cphn K S601 Paper
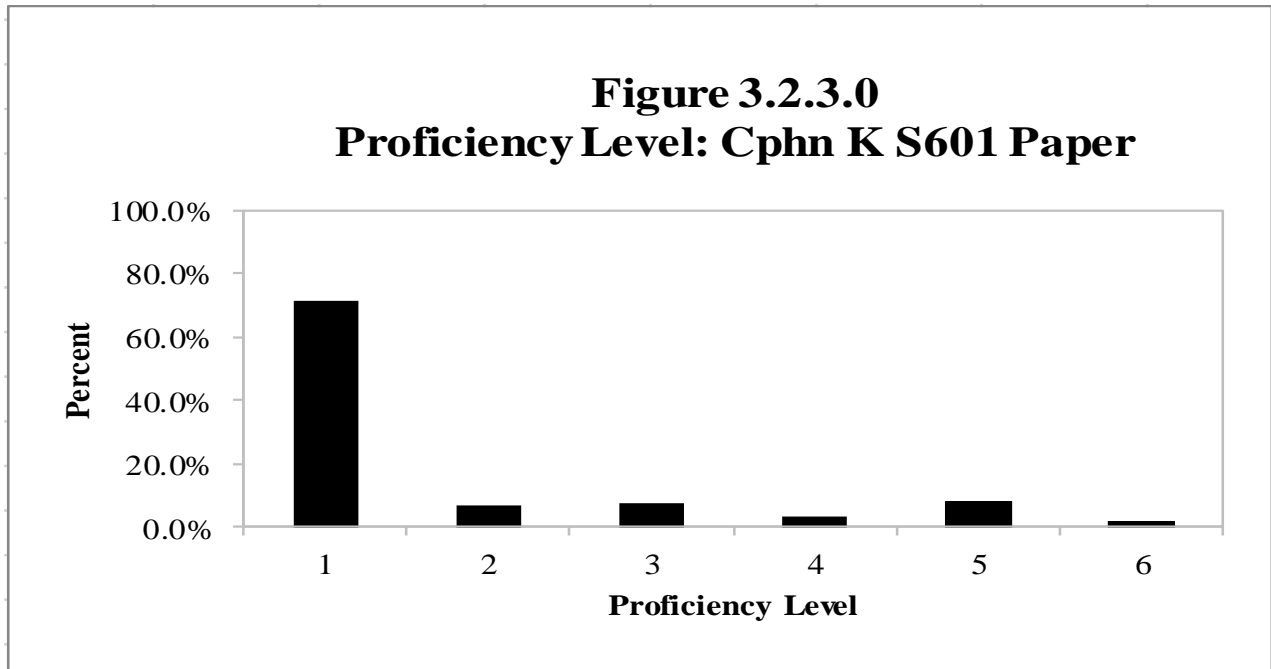
3.2.3.1 Grade 1

Table 3.2.3.1

Proficiency Level Distribution: Cphn 1 S601 Paper

Level	Count	Percent
1	3,411	12.58%
2	6,179	22.78%
3	8,636	31.84%
4	3,346	12.34%
5	3,655	13.48%
6	1,894	6.98%
Total	27,121	100.00%

Figure 3.2.3.1
Proficiency Level: Cphn 1 S601 Paper
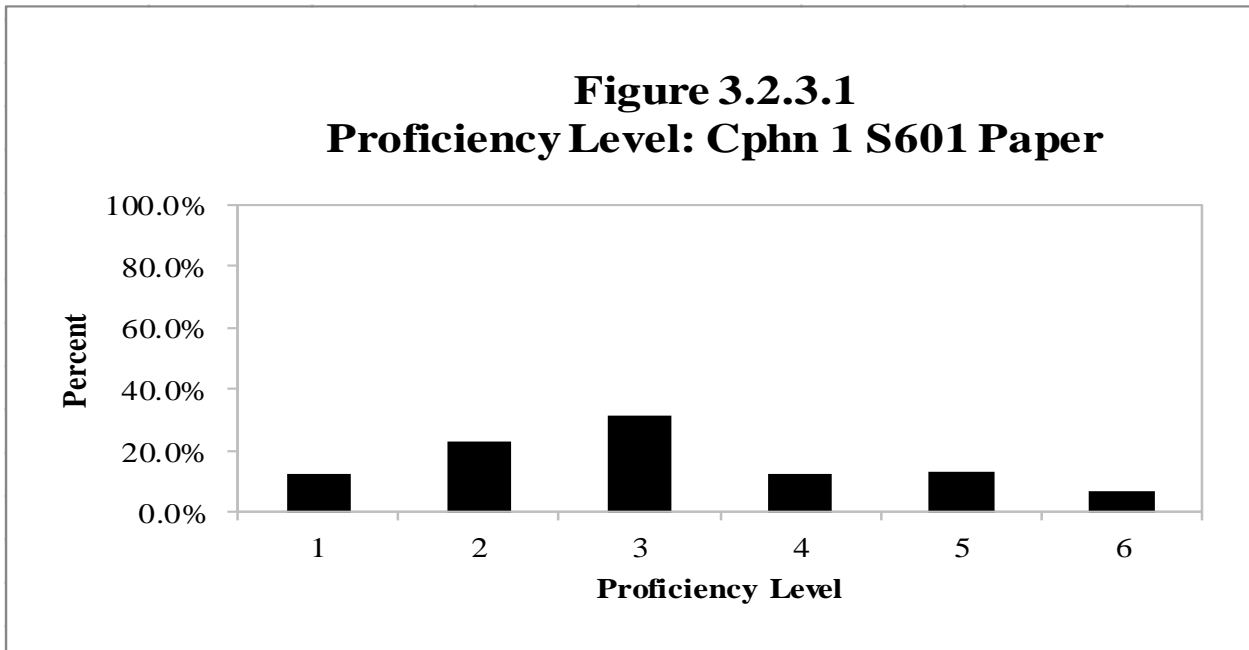
3.2.3.2 Grade 2

Table 3.2.3.2

Proficiency Level Distribution: Cphn 2 S601 Paper

Level	Count	Percent
1	3,572	13.17%
2	6,024	22.21%
3	6,746	24.87%
4	3,531	13.02%
5	4,600	16.96%
6	2,649	9.77%
Total	27,122	100.00%

**Figure 3.2.3.2
Proficiency Level: Cphn 2 S601 Paper**
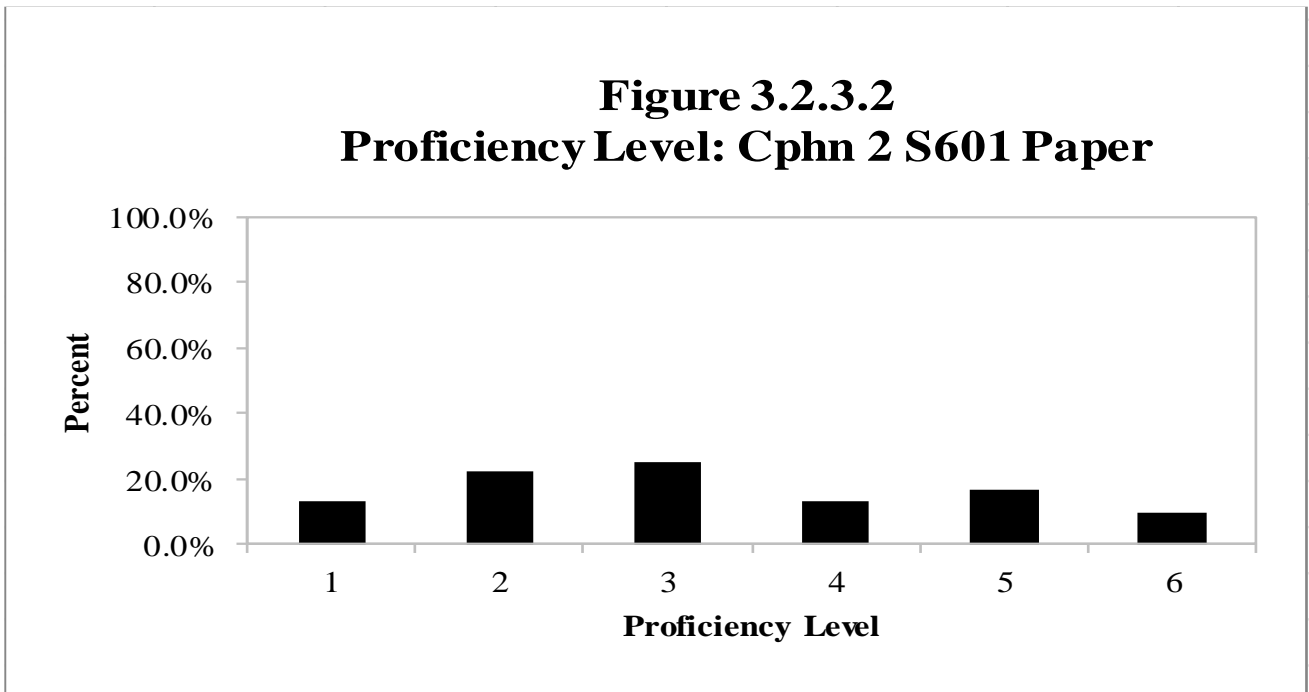
3.2.3.3 Grade 3

Table 3.2.3.3

Proficiency Level Distribution: Cphn 3 S601 Paper

Level	Count	Percent
1	1,831	6.65%
2	4,162	15.13%
3	6,585	23.93%
4	5,989	21.77%
5	6,468	23.51%
6	2,479	9.01%
Total	27,514	100.00%

Figure 3.2.3.3
Proficiency Level: Cphn 3 S601 Paper
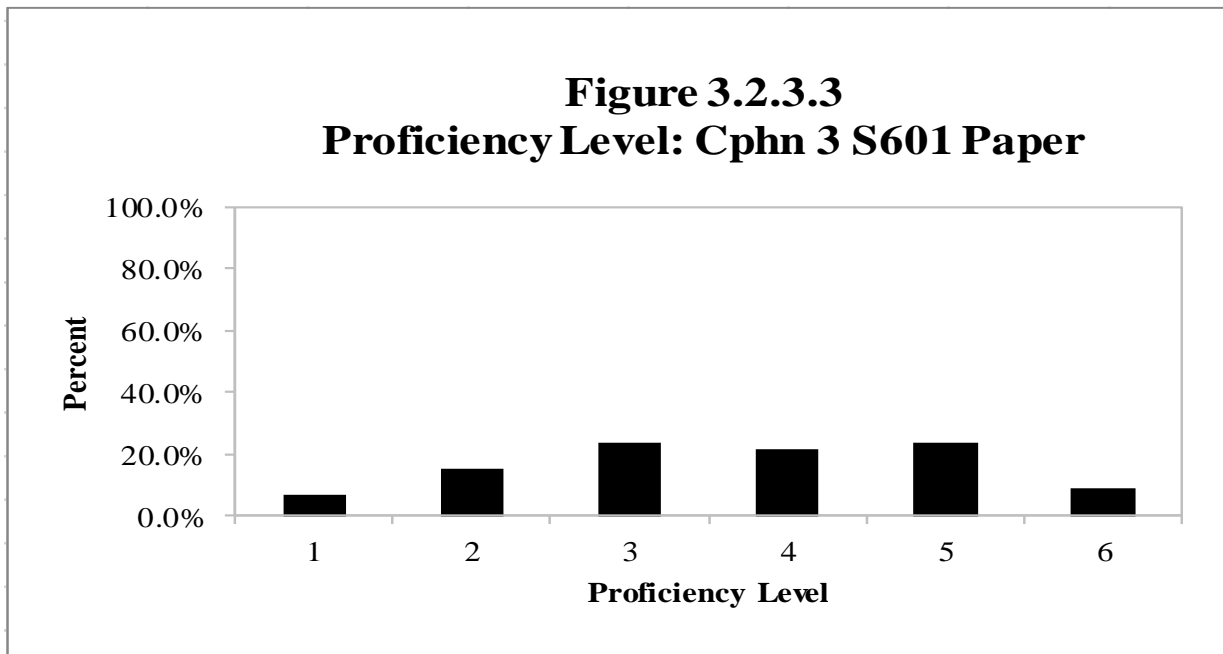
3.2.3.4 Grades 4-5

Table 3.2.3.4

Proficiency Level Distribution: Cphn 4-5 S601 Paper

Level	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	2,229	10.85%	2,727	14.05%	4,956	12.41%
2	2,566	12.50%	2,975	15.33%	5,541	13.87%
3	4,658	22.68%	4,290	22.10%	8,948	22.40%
4	3,757	18.29%	3,316	17.08%	7,073	17.71%
5	4,919	23.95%	3,908	20.13%	8,827	22.10%
6	2,407	11.72%	2,193	11.30%	4,600	11.52%
Total	20,536	100.00%	19,409	100.00%	39,945	100.00%

Figure 3.2.3.4
Proficiency Level: Cphn 4-5 S601 Paper
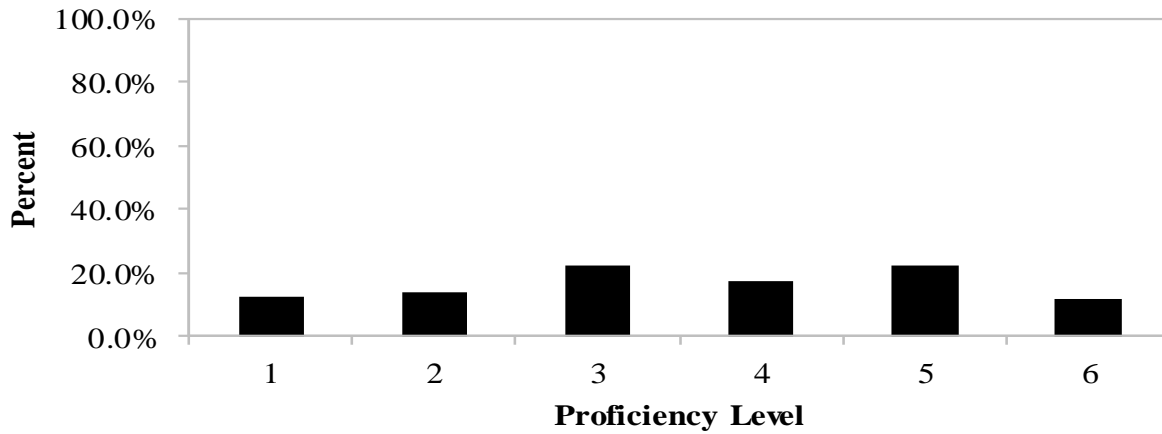
3.2.3.5 Grades 6-8

Table 3.2.3.5

Proficiency Level Distribution: Cphn 6-8 S601 Paper

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	2,304	13.69%	2,855	19.08%	2,908	20.43%	8,067	17.53%
2	4,572	27.16%	3,961	26.48%	3,573	25.10%	12,106	26.30%
3	5,306	31.52%	4,045	27.04%	3,430	24.10%	12,781	27.77%
4	2,467	14.65%	1,965	13.13%	1,897	13.33%	6,329	13.75%
5	1,635	9.71%	1,551	10.37%	1,839	12.92%	5,025	10.92%
6	551	3.27%	584	3.90%	587	4.12%	1,722	3.74%
Total	16,835	100.00%	14,961	100.00%	14,234	100.00%	46,030	100.00%

Figure 3.2.3.5
Proficiency Level: Cphn 6-8 S601 Paper
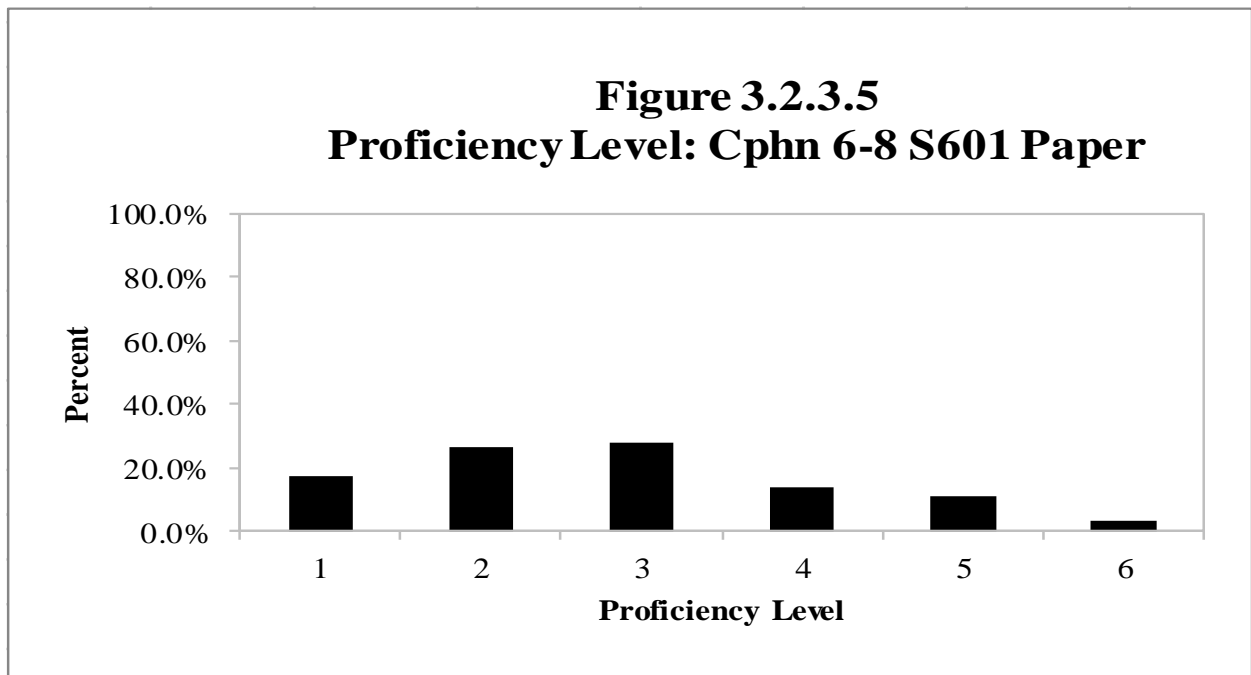
3.2.3.6 Grades 9-12

Table 3.2.3.6

Proficiency Level Distribution: Cphn 9-12 S601 Paper

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,342	16.94%	2,012	16.68%	1,554	15.24%	947	13.32%	6,855	15.87%
2	3,564	25.78%	3,254	26.97%	2,698	26.46%	2,205	31.03%	11,721	27.14%
3	3,248	23.49%	3,006	24.92%	2,628	25.78%	1,866	26.26%	10,748	24.88%
4	2,007	14.52%	1,706	14.14%	1,294	12.69%	1,027	14.45%	6,034	13.97%
5	1,842	13.32%	1,326	10.99%	1,352	13.26%	738	10.38%	5,258	12.17%
6	822	5.95%	761	6.31%	669	6.56%	324	4.56%	2,576	5.96%
Total	13,825	100.00%	12,065	100.00%	10,195	100.00%	7,107	100.00%	43,192	100.00%

Figure 3.2.3.6
Proficiency Level: Cphn 9-12 S601 Paper
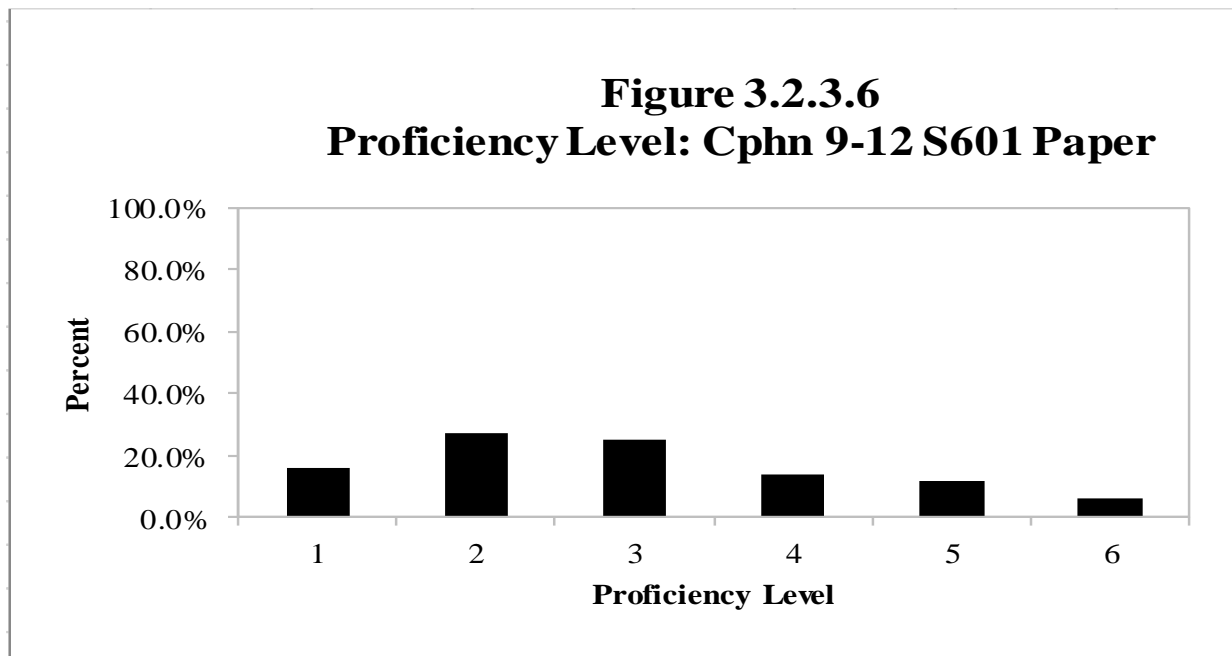
3.2.4 Overall

3.2.4.0 Kindergarten

Table 3.2.4.0

Proficiency Level Distribution: Over K S601 Paper

Level	Count	Percent
1	165,094	64.29%
2	39,565	15.41%
3	30,375	11.83%
4	18,737	7.30%
5	3,041	1.18%
6	0	0.00%
Total	256,812	100.00%

Figure 3.2.4.0
Proficiency Level: Over K S601 Paper
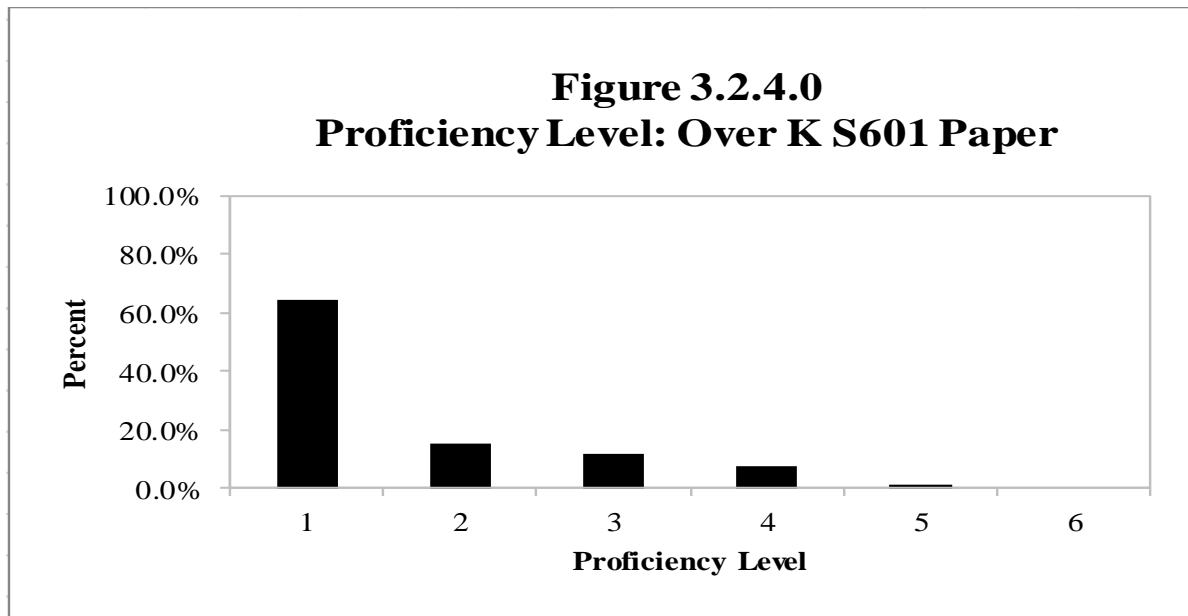
3.2.4.1 Grade 1

Table 3.2.4.1

Proficiency Level Distribution: Over 1 S601 Paper

Level	Count	Percent
1	5,612	20.85%
2	8,395	31.19%
3	10,834	40.25%
4	1,846	6.86%
5	223	0.83%
6	5	0.02%
Total	26,915	100.00%

Figure 3.2.4.1
Proficiency Level: Over 1 S601 Paper
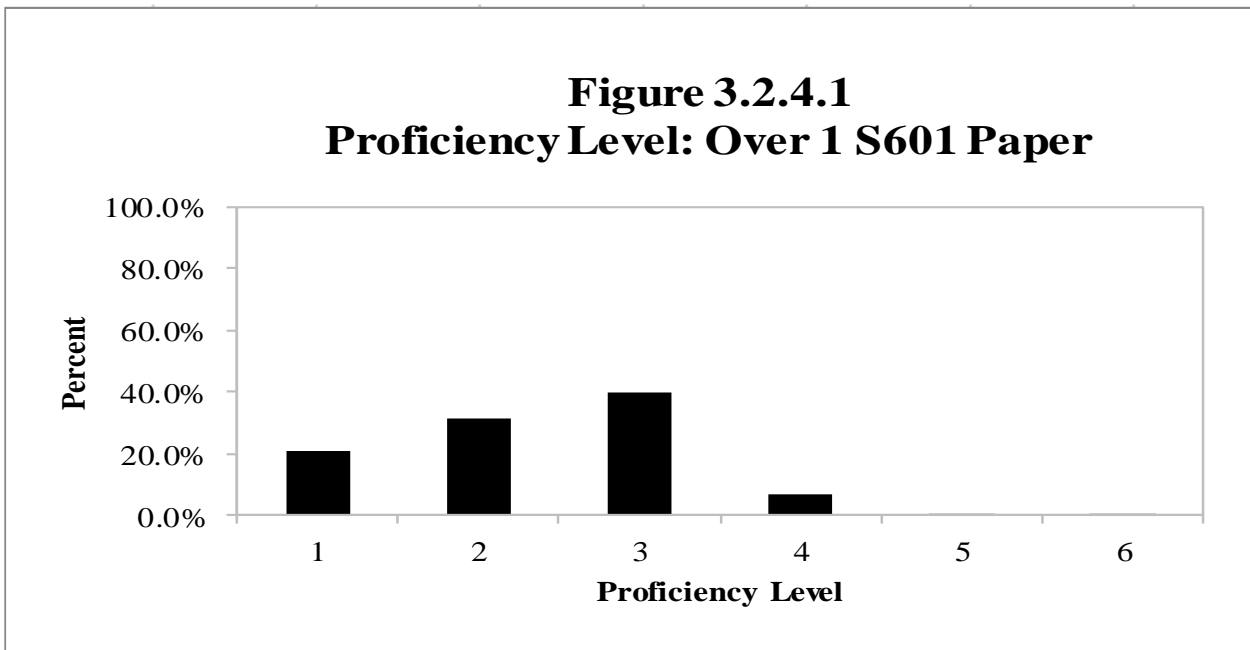
3.2.4.2 Grade 2

Table 3.2.4.2

Proficiency Level Distribution: Over 2 S601 Paper

Level	Count	Percent
1	4,449	16.52%
2	6,381	23.70%
3	11,297	41.95%
4	4,400	16.34%
5	390	1.45%
6	12	0.04%
Total	26,929	100.00%

Figure 3.2.4.2
Proficiency Level: Over 2 S601 Paper
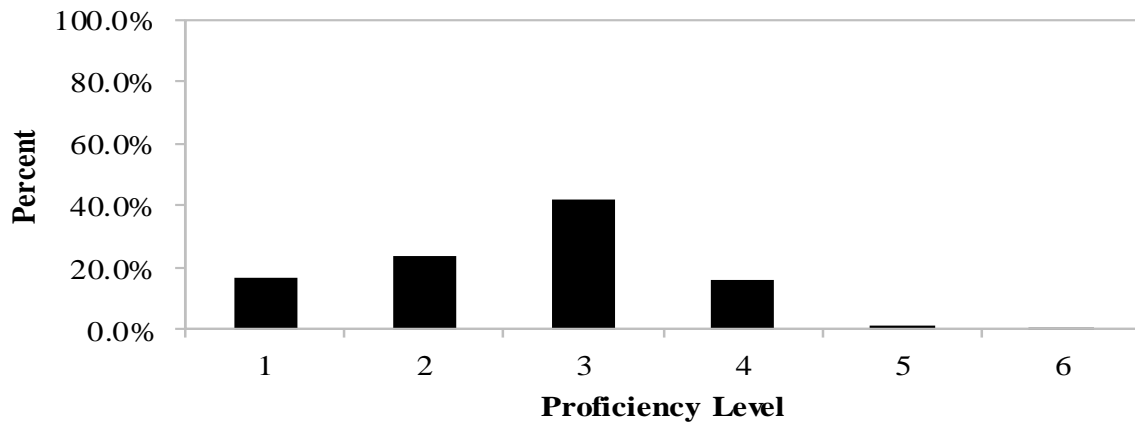
3.2.4.3 Grade 3

Table 3.2.4.3

Proficiency Level Distribution: Over 3 S601 Paper

Level	Count	Percent
1	3,459	12.65%
2	4,185	15.31%
3	13,680	50.03%
4	5,659	20.70%
5	337	1.23%
6	23	0.08%
Total	27,343	100.00%

Figure 3.2.4.3
Proficiency Level: Over 3 S601 Paper
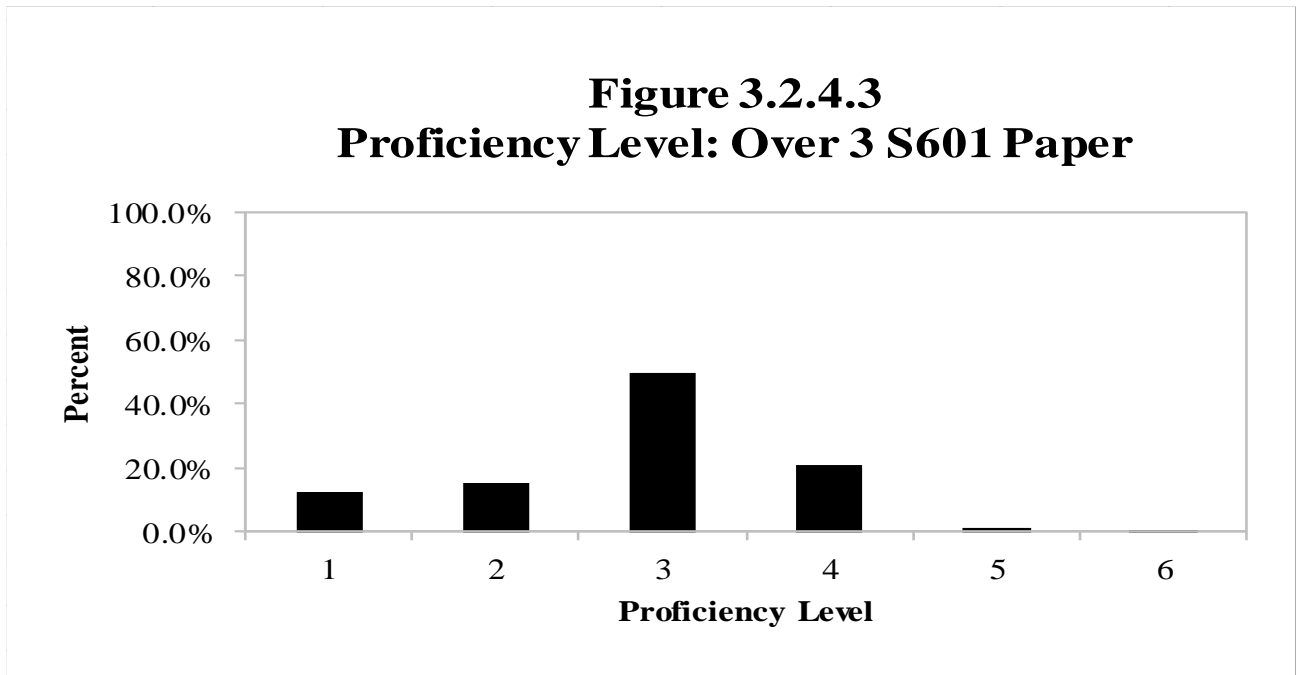
3.2.4.4 Grades 4-5

Table 3.2.4.4

Proficiency Level Distribution: Over 4-5 S601 Paper

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	2,809	13.76%	2,988	15.47%	5,797	14.59%
2	1,920	9.40%	2,158	11.17%	4,078	10.26%
3	7,153	35.03%	6,167	31.93%	13,320	33.52%
4	7,457	36.52%	6,691	34.64%	14,148	35.61%
5	1,015	4.97%	1,237	6.40%	2,252	5.67%
6	63	0.31%	76	0.39%	139	0.35%
Total	20,417	100.00%	19,317	100.00%	39,734	100.00%

**Figure 3.2.4.4
Proficiency Level: Over 4-5 S601 Paper**
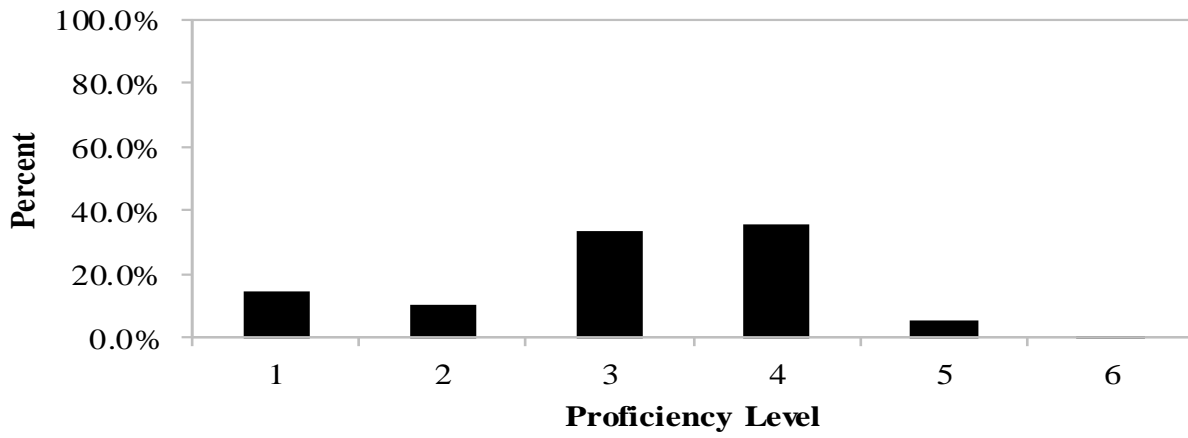
3.2.4.5 Grades 6-8

Table 3.2.4.5

Proficiency Level Distribution: Over 6-8 S601 Paper

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,861	17.14%	3,219	21.71%	3,155	22.38%	9,235	20.24%
2	2,918	17.48%	2,677	18.05%	2,431	17.25%	8,026	17.59%
3	7,323	43.87%	5,657	38.15%	5,132	36.41%	18,112	39.70%
4	3,393	20.32%	3,064	20.66%	3,168	22.48%	9,625	21.10%
5	192	1.15%	209	1.41%	205	1.45%	606	1.33%
6	7	0.04%	2	0.01%	4	0.03%	13	0.03%
Total	16,694	100.00%	14,828	100.00%	14,095	100.00%	45,617	100.00%

Figure 3.2.4.5
Proficiency Level: Over 6-8 S601 Paper
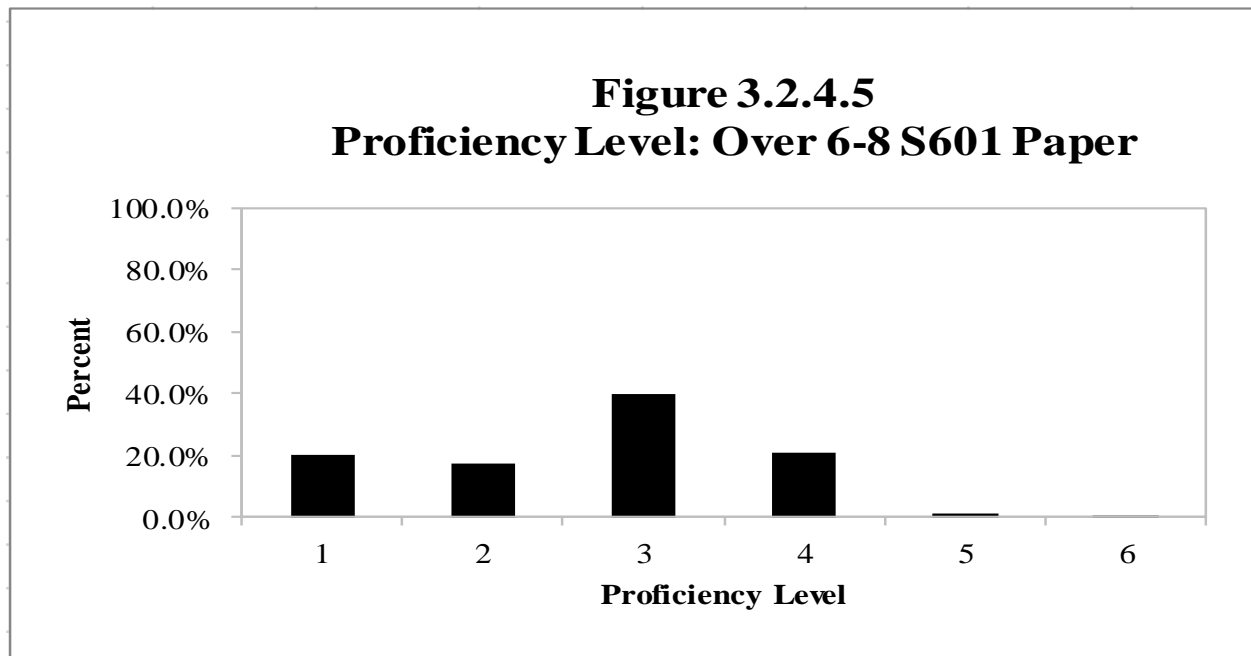
3.2.4.6 Grades 9-12

Table 3.2.4.6

Proficiency Level Distribution: Over 9-12 S601 Paper

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	2,915	21.30%	2,573	21.52%	1,807	17.89%	1,124	15.96%	8,419	19.68%
2	2,453	17.93%	2,241	18.75%	1,795	17.77%	1,516	21.53%	8,005	18.71%
3	4,655	34.02%	4,282	35.82%	3,991	39.51%	3,004	42.66%	15,932	37.24%
4	3,234	23.64%	2,542	21.26%	2,269	22.46%	1,292	18.35%	9,337	21.83%
5	417	3.05%	316	2.64%	238	2.36%	105	1.49%	1,076	2.52%
6	9	0.07%	1	0.01%	2	0.02%	0	0.00%	12	0.03%
Total	13,683	100.00%	11,955	100.00%	10,102	100.00%	7,041	100.00%	42,781	100.00%

Figure 3.2.4.6
Proficiency Level: Over 9-12 S601 Paper
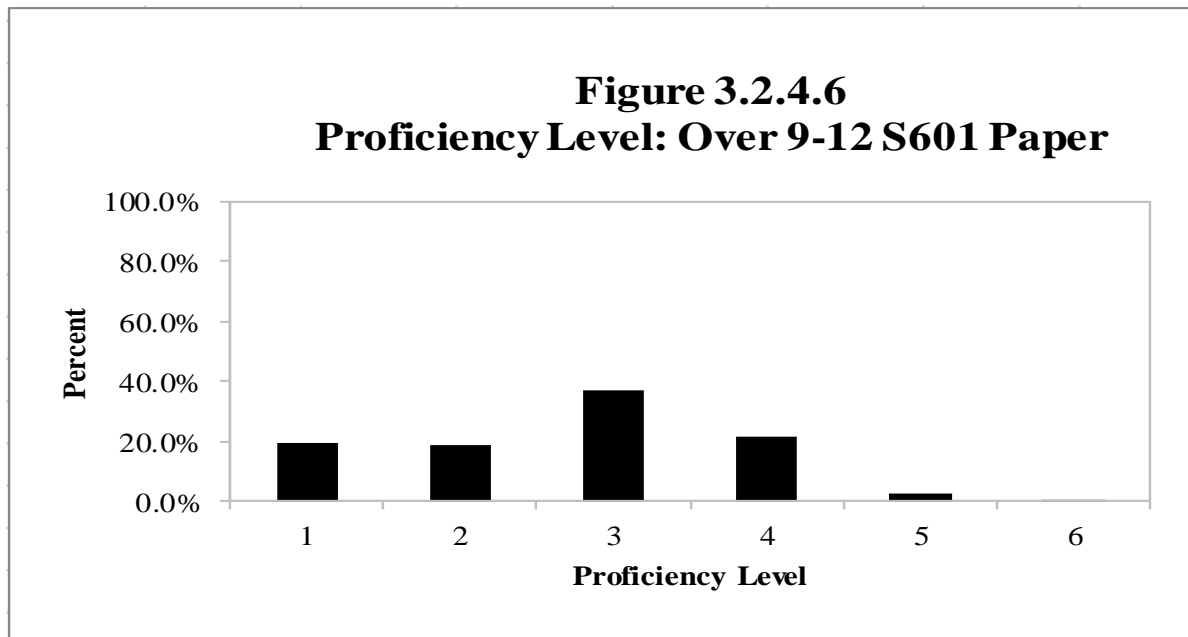
4. Annual Updates of Validity Evidence

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. Particular interpretations for specified uses begin by specifying the construct the test is intended to measure. Rather than referring to distinct types of validity, the Standards refer to types of validity evidence. According to the Standards, the evidence can be based on (1) test content, (2) response processes, (3) internal structure, and (4) relation to other variables.

The validity evidence of the Standards is also observed in "A State's guidance to the U.S. Department of Education's Assessment Peer Review Process" document (Department of Education, 2018 https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf) to support states' use of ELP assessments for reviewing of validity evidence, as well as is linked to Assessment User Argument (AUA) to support the claims of validity of Online ACCESS assessment. WIDA structures its validity arguments using AUA model in lieu of the model highlighted in the *Standards for Educational and Psychological Testing.* AUA has similar topics; however, they are organized differently. Below is a short summary of each AUA claim. For the full AUA validity claims, please refer to WIDA Assessment User Argument document.

Claim 1 (Consequences): With the use of ACCESS, the intended decisions will have beneficial consequences for stakeholders, in terms of using ACCESS and the decisions made based on ACCESS.

Claim 2 (Decisions): Decisions based on ACCESS test results are made by individuals, in a timely manner, and affect a variety of stakeholders. Two types of decisions that are made based on ACCESS results are classification and programming decisions. The decisions take into consideration educational and societal values, and relevant laws, rules, and regulations, and they are equitable for the intended stakeholders.

Claim 3 (Interpretations): The interpretations of students' academic English language proficiency in four domains are *relevant* to the classification, placement and programming decisions; *sufficient*, in conjunction with additional information as outlined in state and local policies, to make such decisions; *meaningful* with respect to the WIDA English Language

Development (ELD) Standards; *generalizable* to the academic English language used in K–12 instructional settings, and *impartial* to all students.

Claim 4 (Assessment records: Scores): ACCESS scores are consistent across different aspects of test administration, different test tasks, and different groups of students. Test forms and metrics accurately represent the construct being measured and result in expected test taker performances.

4.1 Standards

4.1.1 Test Content

The relationship between the content of a test and the construct to measure is called content validity. Test content includes the themes, wording, and format of the items, tasks, or questions on a test. Administration and scoring may also be part of the content. Empirical or logical evidence can show how appropriately the content reflects the domain as we interpret test scores.

4.1.2 Response Processes

Empirical analysis of how test takers process tests provides evidence of the nature between performance and the construct. Examples of this validity include analyzing individual item responses, different response processes in answering questions by subgroups or evaluating test-takers' performance.

4.1.3 Internal Structure

Validity related to internal structure indicates how test items/components agree with the construct we base for the score interpretation. The internal structure of the construct can be unidimensional or contain multidimensional components.

4.1.4 Relation to Other Variables

The interpretation of the test scores with an external indicator provides valuable validity evidence. We often ask how accurately the test score predicts the criterion variable. The test criterion validity has two different validities: concurrent and predictive validity. Predictive validity is how accurately test scores predict the future performance of criterion scores. Concurrent validity indicates how test scores relate to criterion scores at the same time.

4.2   Annual Validity Studies

4.2.1 English Learner Identification Project: Focus on Kindergarten Children

The WIDA screener for Kindergarten (hereafter K Screener) became available in March 2021. Due to the novelty of K Screener, WIDA states have decided on new criteria, including K Screener cut scores, that will be used to identify ELs. As is usually the case, there is variability in the EL identification criteria used across states; it is also possible for districts within the same state to have slightly different criteria and/or decision-making process for EL identification.

In addition to understanding the factors and the decision-making process involved with EL identification, it is necessary to examine the appropriateness of the decisions, as they often have significant consequences on student outcomes. Also, it is important to understand the extent of misidentification and how it is addressed. Research questions that are explored in this study are listed below:

1. Who makes decisions on EL identification?
2. What are the criteria used for EL identification? (Including proficiency level used for determining ELs vs. non-ELs.)
3. What are the procedures involved in EL identification (e.g., home language survey is completed first and then screener is administered)?
4. To what extent are the EL identification decisions appropriate?

To address the research questions, we conducted a two-phase study and collected data from one northeastern state in Fall 2022. Phase 1 involved distribution of an online survey across all districts. In Phase 2, we collected and analyzed home language survey and K screener data from three districts, varying in EL enrollment size (small, medium, large). In addition, educators completed a review of the appropriateness of the EL identification decisions in their schools; follow-up interviews were conducted with select educators to further discuss the appropriateness of EL identification.

The major findings In Phase I were as follows:

- There is variation in the EL identification process at the kindergarten level across districts, with multiple types of educators involved in the process, including EL/bilingual teachers, district EL coordinators, and EL/bilingual program directors.
- Districts used multiple types of data and tools to identify potential ELs, including but not limited to, the home language survey (HLS), English language proficiency (ELP) screener, parent/guardian input, teacher input, classroom observation of students, home visits, and students' prior educational history.
- For districts that used the K Screener, most reported using similar criteria as the state's. However, many reported that they believed the cut score should be lower and some expressed concerns about the K Screener.
- Educators believed that misidentifying non-ELs as ELs was more likely than misidentifying ELs as non-ELs.

In Phase II, all three districts reported that they adhered to the state's guidelines where they used both the HLS and ELP assessment to screen and identify EL students. Some variations in the kindergarten EL identification process included the ways they collected HLS data (online vs. paper format), the data educators used to make decisions (ELP assessment only vs. multiple types of data), and the ELP assessment used by the district in the 2022-2023 school year (transition to K Screener from WIDA MODEL for Kindergarten and other assessments). Educators from all three districts also shared some similar concerns: high cut scores and additional factors that may affect students' performance on the ELP assessment (e.g., familiarity with the test administrator, students' personalities, and special needs). In terms of suggestions for improving the EL identification process in their respective districts, educators shared several ideas: collecting more student data from different sources and administering the assessment in a familiar setting or with a familiar test administrator.

A report on this study is forthcoming.

5. Reliability

In accordance with the *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014), when interpreting test scores, it is important to evaluate their reliability, as the interpretation of test scores depends on the assumption that students exhibit some degree of consistency in their scores across independent administrations of the same testing procedure. We expect that students mastering the domain will consistently perform well, and those who have not mastered the domain will consistently perform less well, regardless of the sample of items and tasks used to assess students. Furthermore, because we assume that all items and tasks on such a test measure some aspect of the domain of interest, we expect that students will perform consistently across different items and tasks measuring the same ability within the test. Therefore, it is important to evaluate the degree to which students' test scores are consistent across replications of the same testing condition.

However, different samples of performances from the same student are rarely identical. A student's responses to sets of test items or tasks vary from one sample of test items or tasks targeting the domain to another, and from one occasion to another, even under strictly controlled conditions. In addition, different raters may award different scores to the same student performance on a test task. These sources of variation are reflected in the students' scores. Therefore, it is important to evaluate the extent to which differences in students' test scores reflect true differences in the knowledge, skills, or ability being tested, rather than fluctuations due to chance.

The reliability of the test scores depends on how much the scores vary across replications of the testing procedure, and analyses of reliability depend on the types of variability likely to be of concern in the testing procedure. There are several ways to collect reliability data and to estimate reliability, some of which depend on the exact nature of the measurement, the intended use of the test scores, the assessment design, and the potential sources of measurement error that might contribute to inconsistency in students' scores across different test administrations.

The reliability information presented in this section is organized to be in compliance with Critical Element 4.1 of the Every Student Succeeds Act Peer Review requirements (U.S. Department of Education, 2018) and follows the guidelines of the *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014). We present

information regarding the reliability of the domain scale scores first, followed by information about the reliability of the composite scale scores.

Policymakers in states and districts use ACCESS Listening, Reading, Writing, and Speaking tests to determine the English language proficiency of students based on their scores in each of the four domains. Therefore, the main concern in interpreting these scores is how consistent the scores would be over replications of the same testing procedure. We use **internal consistency reliability statistics** to address this question (Section 5.1).

Additionally, for the Writing and Speaking domains, because having different raters evaluate the same students' responses to tasks may result in inconsistent scoring, a potential source of variation of those scores is the rater. In Section 5.2, we report the **interrater agreement** rates that the raters achieved when evaluating students' responses to the Writing and Speaking tasks. We can use these statistics to determine how consistent the students' scores would have been if different raters had evaluated their responses. Since we use an item response theory (IRT)–based method to estimate students' **latent scores** (i.e., test scores based on variables that we cannot see or directly measure but which we can infer mathematically through advanced statistical techniques by using students' scores on variables that we can observe), we also examine the amount of **measurement error** in students' scores using the **conditional standard error of measurement** (CSEM) (Section 5.3). Lastly, in Section 5.4, we evaluate the reliability of the classifications of students into WIDA proficiency levels based on their domain scores (the most important interpretation of the test scores) in terms of the **accuracy and consistency** of the classification decisions made. In each subsection, we present detailed descriptions of the methods, data sources, and procedures.

Policymakers in states and districts use ACCESS **composite scale scores** to describe the English language proficiency of students in the respective composites. Therefore, the most important concern in interpreting these scores is how consistent the scores would be over replications of the same testing procedure. We use internal consistency reliability statistics to address this question, and in Section 5.5 we provide the results. In addition, in Section 5.6, we examine the CSEM of these scores. Lastly, in Section 5.7, we evaluate the reliability of the classifications in terms of the accuracy and consistency of the decisions made about students' levels of English language proficiency based on their composite scale scores. In each subsection, we present detailed descriptions of the methods, data sources, and procedures.

Internal Consistency Reliability Statistics

One way to evaluate the consistency of students' test scores across test administrations is to examine how the students would have performed on alternate forms of the same test (i.e., **parallel test form reliability**). Given our assumption that the ability the test measures is constant for each student over two administrations of alternate forms, the more variation found across the two administrations, the more evidence for lower reliability. The **measurement error** represents the sources of inconsistency across the two administrations, taken together. We consider measurement error to be random and to occur by chance. For example, there may be some construct-irrelevant knowledge and/or skills that some items or tasks measure that affect students' scores but are not part of the ability that the test intends to measure.

Unless students take two alternate versions of the same test, we cannot calculate test score reliability directly. Thus, we usually estimate it from student responses to a single form of the test. Methods employed to estimate reliability using test scores from a single test administration are based on classical test theory and are referred to as estimates of **internal consistency**. An internal consistency reliability statistic is a useful estimate of alternate-forms reliability, providing an estimate of the consistency of students' performances across items and tasks within a test. The most common index of internal consistency reliability is **Cronbach's coefficient alpha** (Cronbach, 1951), which is a lower-bound estimate of test reliability. Conceptually, we think of Cronbach's coefficient alpha as the correlation obtained between performances on two halves of the same test if every possible way of dividing the test items and tasks in two were attempted. Because Cronbach's coefficient alpha is a correlation of students' performances on all possible pairs of test items and tasks, it may be low if some items or tasks are measuring something other than what most of the other items and tasks are measuring (and thus leading to inconsistent student performances). In this way, Cronbach's coefficient alpha expresses how well the items and tasks on a test appear to measure the same ability. The Cronbach's coefficient alpha of internal consistency ranges from 0 to 1. If students achieve their scores by a completely random process (i.e., their scores are not correlated or share no covariance), then the reliability estimate is very close to 0. On the other hand, if students' scores are perfectly consistent (i.e., their scores have high covariances), then the internal consistency coefficient will approach 1.

While there is no one set of criteria that the testing community uses when interpreting Cronbach's coefficient alpha values, from time to time, researchers have proposed various

arbitrary criteria that one could apply. Initially, Cronbach (1951) argued that it was "desirable" to have a high alpha value for an instrument that test developers were using to report individual scores, since the scores on that instrument needed to be interpretable, and that would require a high alpha value. Later, Nunnally (1978) suggested that researchers should consider a value of 0.70 as an acceptable lower limit if they were engaged in the early stages of research (e.g., when developing a scale). Today, it has become common practice to cite Nunnally's suggested 0.70 criterion as a minimum acceptable lower limit for this value for all types of research. However, in so doing, researchers ignore Nunnally's more nuanced guidance: If researchers were engaged in basic research, Nunnally advised that they should use a higher cut-off value (i.e., 0.80 or higher), and those engaged in applied research should use a much higher cut-off value (0.90 or higher) (Lance et al., 2006). Since Nunnally's time, some researchers have suggested even more nuanced interpretations of various alpha values. For example, George and Mallery (2003) proposed the following interpretations: "≥ 0.90 – Excellent, ≥ 0.80 – Good, ≥ 0.70 – Acceptable, ≥ 0.60 – Questionable, ≥ 0.50 – Poor, and ≤ 0.50 – Unacceptable" (p. 231). Clearly, there is little consensus among the experts in their views of what the acceptable lower limit of the Cronbach's coefficient alpha value should be, or for that matter, how one should interpret various values. This lack of consensus led the authors of the *Standards for Educational and Psychological Measurement* (2014) to conclude, "The choice of [reliability/precision] estimation and the minimum acceptable level for any index remain a matter of professional judgment" (p. 41). For the purposes of this report then, WIDA has made the decision that within the domains of Listening, Reading, and Speaking, an alpha value of ≥ 0.80 is acceptable, while an alpha value of ≥ 0.65 is acceptable for the Writing domain.

Reliability statistics such as the Cronbach's coefficient alpha of internal consistency are affected by two factors: (1) the number of test items or tasks, and (2) the total number of score points students achieve. That is, all things being equal, the greater the number of items or tasks measuring the same ability there are on the test, the higher the internal consistency reliability statistics. Additionally, because reliability statistics refer to the consistency of scores *for a group of students*, the distribution of that specific group's ability measures affects these statistics. If the students in the group are nearly equal in the ability that the test measures (i.e., their scores are concentrated in the center of the ability distribution), small changes in their scores can easily change their relative positions in the group. Consequently, the internal consistency reliability statistics will be low. In this case, the statistic may be telling us more about the group of students

tested than about the test itself. On the other hand, if the students in the group differ widely in the ability that the test measures (i.e., their scores are distributed across the ability continuum), small changes in their scores will not affect their relative positions in the group as much, and the internal consistency reliability statistics will be higher. Therefore, reliability can be as much a function of the performance of test items and tasks as of the performance of the sample of students tested. That is, the exact same test can produce widely disparate reliability indices based on the ability distribution of the group of students. This means, in turn, that when interpreting estimates of internal consistency, it is wise to keep in mind the specific set of test items and tasks and the distribution of ability measures in the group of students used in the estimation.

Interrater Agreement

The behavior of raters is a potential source of variance in students' scores for the productive domains of ACCESS (i.e., Writing and Speaking). ACCESS scoring procedures and rater training and quality control monitoring processes are described elsewhere in this report (see Part 1, Section 3.2.2). In Section 5.2, we report the **interrater agreement rates** for the scoring of students' responses to the Writing and Speaking tasks. These values reflect how consistent the students' scores would be if different groups of raters scored their responses. Additionally, in this section of the report we present a detailed description of the methods, data sources, and procedures we used when calculating interrater agreement rates.

Measurement Error

In addition to evaluating test score reliability in terms of estimates of internal consistency, we can calculate the amount of measurement error in students' test scores in two different ways. One way is to hypothesize that there is an error-free measure of each student's true ability, referred to as the **true score** in classical test theory. The true score is a theoretical value, so it is not a known quantity. Rather, we view it as the hypothetical average score over repeated replications of the same testing condition (Livingston, 2018, p. 9). Under the assumptions of classical test theory, the **error of measurement** over a replication of a testing condition provides an estimate of the amount of variability from students' true scores that we would expect. In practical testing contexts, it is generally not possible to replicate a testing condition (i.e., have students take the same test form multiple times), so it is not possible to estimate the standard error of each student's score using a repeated measures design. Instead, we calculate the average error of measurement over the population of students who take the test, and then we use that as an

indication of the amount of variation in any individual student's score that we would expect. Classical test theory refers to this average as the **standard error of measurement** (SEM), which provides an indication of how much students' scores differ from their true scores, on average, on the raw score metric. Because it is a standard deviation of the distribution of errors of measurement, we can construct a **confidence interval** to indicate how the errors of measurement are affecting the scores. Test scores with large SEMs pose a challenge to the interpretation of the reliability of any single test score.

A second way to address the impact of measurement errors on students' test scores is to estimate the SEM for specific scores using IRT. IRT addresses reliability using the **test information function**, which indicates the precision with which we can use student performances on items and tasks to estimate the **latent** (i.e., true) **ability** of each student (i.e., **latent scores**). The square root of the inverse of the information function at any point on the latent ability distribution is the **conditional standard error of measurement** (**CSEM**). The CSEM provides information about the amount of error we would expect in any student's score at that point on the underlying latent ability scale, which IRT refers to in terms of the **latent score metric** (i.e., the IRT metric for expressing student ability, as opposed to the raw score metric). In addition, by using IRT, we can estimate indices analogous to traditional reliability coefficients such as Cronbach's coefficient alpha from the test information function and the distribution of the latent scores in the same student population.

Classification Accuracy and Consistency

One of the main purposes of the WIDA ACCESS program is to identify the English language proficiency levels of students with respect to the WIDA ELD Standards. Because of the emphasis on the classification of student performance into six WIDA proficiency levels, it is important to know how consistently ACCESS scores do indeed classify students into those proficiency levels (American Educational Research Association et al., 2014). The questions that we want to answer are different from the questions that the reliability coefficient answers. Instead of looking at the reliability of a specific student score, we want to know the consistency of the decisions we make when we use students' test scores to classify them into a smaller number of proficiency levels. One way to approach this question is to estimate the degree to which the classification decisions we are making based on the students' **observed test scores** agree with the classification decisions we would make based on students' **theoretical true**

scores. This estimate is known as **decision accuracy**. A second way to approach this question is to estimate the degree to which the classification decisions we are making based on the students' test scores agree with the classification decisions we would make based on students' scores on an alternate form of the test. This estimate is known as **decision consistency**.

5.1   Reliability of the Domain Scores

Cronbach's coefficient alpha is widely used as an estimate of reliability, particularly for the internal consistency of test items and/or tasks, and this statistic is appropriate for calculating the reliabilities of students' scores from the administration of the fixed forms of the Writing and Speaking tests. Conceptually, we can think of it as the correlation obtained between students' performances on two halves of the Writing or Speaking test if every possible way of dividing the test tasks in two was attempted. Thus, Cronbach's coefficient alpha may be low if some tasks are measuring something other than what the majority of the tasks are measuring. In this way, Cronbach's coefficient alpha expresses how well the tasks on a test appear to measure the same ability.

The formula for calculating Cronbach's coefficient alpha for the fixed forms of the Writing and Speaking tests is

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n}\sigma_i^2}{\sigma_t^2}\right]$$

where

n = number of items i

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

$$\text{SEM} = SD\sqrt{1 - reliability}$$

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-level 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 0.95 to 0.96. For the Kindergarten Reading test, the reliability for all students was 0.95, and reliability values across subgroups ranged from 0.94 to 0.96. For the Kindergarten Writing test, the reliability for all students was 0.92, and reliability values across subgroups ranged from 0.91 to 0.93. For the Kindergarten Speaking test, the reliability for all students was 0.92, and reliability values across subgroups ranged from 0.90 to 0.92.

Listening Tier A: The Listening Tier A Cronbach's coefficient alphas computed for all students ranged from 0.64 to 0.78. The Listening Tier A Cronbach's alpha ranged from 0.64 to 0.79 for male students; 0.63 to 0.78 for female students; 0.62 to 0.78 for Hispanic students; 0.68 to 0.78 for non-Hispanic students; and 0.64 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 0.60 to 0.69. The Listening Tier B/C Cronbach's coefficient alphas ranged from 0.61 to 0.69 for male students; 0.58 to 0.68 for female students; 0.60 to 0.68 for Hispanic students; 0.59 to 0.71 for non-Hispanic students; and 0.58 to 0.73 for students with an IEP.

Reading Tier A: The Reading Tier A Cronbach's coefficient alphas computed for all students ranged from 0.74 to 0.80. The Reading Tier A Cronbach's coefficient alphas ranged from 0.75 to 0.80 for male students; 0.73 to 0.79 for female students; 0.73 to 0.79 for Hispanic students; 0.78 to 0.82 for non-Hispanic students; and 0.66 to 0.76 for students with an IEP.

Reading Tier B/C: The Reading Tier B/C Cronbach's coefficient alphas computed for all students ranged from 0.75 to 0.82. The Reading Tier B/C Cronbach's coefficient alphas ranged from 0.76 to 0.82 for male students; 0.73 to 0.81 for female students; 0.74 to 0.81 for Hispanic students; 0.77 to 0.83 for non-Hispanic students; and 0.69 to 0.77 for students with an IEP.

Writing Tier A: The Writing Tier A Cronbach's coefficient alphas computed for all students ranged from 0.87 to 0.90. The Writing Tier A Cronbach's coefficient alphas ranged from 0.87 to

0.90 for male students; 0.86 to 0.90 for female students; 0.87 to 0.90 for Hispanic students; 0.87 to 0.90 for non-Hispanic students; and 0.85 to 0.89 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 0.95. The Writing Tier B/C Cronbach's coefficient alphas ranged from 0.91 to 0.95 for male students; 0.90 to 0.95 for female students; 0.91 to 0.95 for Hispanic students; 0.91 to 0.95 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 0.90 to 0.93. Cronbach's coefficient alphas ranged from 0.90 to 0.93 for male students; 0.89 to 0.93 for female students; 0.90 to 0.93 for Hispanic students; 0.88 to 0.90 for non-Hispanic students; and 0.86 to 0.90 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 0.93 for students with an IEP.

5.1.1 Listening

Table 5.1.1.1

Reliabilities of Domain Scores: List S601 Paper

Cluster	Tier	No. of Students	No. of Items	Cronbach's Alpha	SEM
K	-	256,847	30	0.95	1.85
1	A	30,555	18	0.78	1.66
	B/C	35,473	21	0.69	1.80
2	A	30,555	18	0.78	1.66
	B/C	35,473	21	0.69	1.80
3	A	22,470	18	0.72	1.89
	B/C	55,637	21	0.61	1.92
4-5	A	22,470	18	0.72	1.89
	B/C	55,637	21	0.61	1.92
6-8	A	19,155	18	0.69	1.87
	B/C	31,554	21	0.60	1.90
9-12	A	17,762	18	0.64	1.81
	B/C	29,715	21	0.63	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 S601 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	106,546	0.95	1.83	118,129	0.96	1.86
1	A	18	14,387	0.78	1.64	16,052	0.79	1.67
	B/C	21	17,278	0.68	1.78	18,145	0.69	1.82
2	A	18	14,387	0.78	1.64	16,052	0.79	1.67
	B/C	21	17,278	0.68	1.78	18,145	0.69	1.82
3	A	18	10,548	0.71	1.89	11,816	0.73	1.88
	B/C	21	25,832	0.60	1.92	29,736	0.61	1.92
4-5	A	18	10,548	0.71	1.89	11,816	0.73	1.88
	B/C	21	25,832	0.60	1.92	29,736	0.61	1.92
6-8	A	18	9,056	0.69	1.86	10,016	0.70	1.88
	B/C	21	14,466	0.58	1.90	17,034	0.61	1.90
9-12	A	18	8,397	0.63	1.81	9,267	0.64	1.82
	B/C	21	13,783	0.62	1.95	15,814	0.64	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 S601 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	172,325	0.95	1.88	76,515	0.95	1.79
1	A	18	24,731	0.78	1.66	5,592	0.78	1.61
	B/C	21	27,892	0.68	1.80	7,434	0.71	1.80
2	A	18	24,731	0.78	1.66	5,592	0.78	1.61
	B/C	21	27,892	0.68	1.80	7,434	0.71	1.80
3	A	18	18,567	0.72	1.89	3,702	0.74	1.84
	B/C	21	44,170	0.60	1.93	11,198	0.63	1.91
4-5	A	18	18,567	0.72	1.89	3,702	0.74	1.84
	B/C	21	44,170	0.60	1.93	11,198	0.63	1.91
6-8	A	18	16,142	0.68	1.88	2,820	0.73	1.82
	B/C	21	24,739	0.60	1.90	6,598	0.59	1.87
9-12	A	18	14,820	0.62	1.82	2,674	0.68	1.79
	B/C	21	23,077	0.63	1.95	6,401	0.64	1.94

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 S601 Paper by IEP Status

Cluster	Tier	No. of Students	No. of Items	Cronbach's Alpha	SEM
K	-	22,257	30	0.96	1.88
1	A	3,332	18	0.77	1.70
	B/C	4,022	21	0.73	1.90
2	A	3,332	18	0.77	1.70
	B/C	4,022	21	0.73	1.90
3	A	1,779	18	0.70	1.87
	B/C	10,567	21	0.61	1.98
4-5	A	1,779	18	0.70	1.87
	B/C	10,567	21	0.61	1.98
6-8	A	565	18	0.67	1.87
	B/C	4,987	21	0.58	1.98
9-12	A	473	18	0.64	1.83
	B/C	2,291	21	0.59	1.99

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 S601 Paper

Cluster	Tier	No. of Students	No. of Items	Cronbach's Alpha	SEM
K	-	256,837	30	0.95	1.75
1	A	28,402	24	0.74	2.22
	B/C	30,949	27	0.82	2.33
2	A	28,402	24	0.74	2.22
	B/C	30,949	27	0.82	2.33
3	A	20,909	24	0.80	2.19
	B/C	49,459	27	0.75	2.39
4-5	A	20,909	24	0.80	2.19
	B/C	49,459	27	0.75	2.39
6-8	A	18,498	24	0.76	2.19
	B/C	29,593	27	0.77	2.36
9-12	A	17,663	24	0.79	2.10
	B/C	27,585	27	0.80	2.37

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 S601 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	106,543	0.95	1.75	118,123	0.95	1.74
1	A	24	13,290	0.73	2.22	15,000	0.75	2.21
	B/C	27	15,054	0.81	2.33	15,851	0.82	2.32
2	A	24	13,290	0.73	2.22	15,000	0.75	2.21
	B/C	27	15,054	0.81	2.33	15,851	0.82	2.32
3	A	24	9,925	0.79	2.18	10,892	0.80	2.19
	B/C	27	23,071	0.73	2.39	26,331	0.76	2.38
4-5	A	24	9,925	0.79	2.18	10,892	0.80	2.19
	B/C	27	23,071	0.73	2.39	26,331	0.76	2.38
6-8	A	24	8,792	0.77	2.18	9,623	0.76	2.19
	B/C	27	13,743	0.77	2.35	15,799	0.78	2.37
9-12	A	24	8,418	0.79	2.09	9,146	0.80	2.12
	B/C	27	12,884	0.80	2.36	14,592	0.81	2.37

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 S601 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	172,319	0.94	1.74	76,511	0.96	1.73
1	A	24	23,030	0.73	2.22	5,164	0.78	2.19
	B/C	27	24,350	0.81	2.33	6,474	0.83	2.31
2	A	24	23,030	0.73	2.22	5,164	0.78	2.19
	B/C	27	24,350	0.81	2.33	6,474	0.83	2.31
3	A	24	17,304	0.79	2.19	3,433	0.82	2.16
	B/C	27	39,425	0.74	2.39	9,814	0.77	2.38
4-5	A	24	17,304	0.79	2.19	3,433	0.82	2.16
	B/C	27	39,425	0.74	2.39	9,814	0.77	2.38
6-8	A	24	15,646	0.75	2.19	2,677	0.81	2.15
	B/C	27	23,207	0.77	2.37	6,175	0.78	2.35
9-12	A	24	14,742	0.79	2.11	2,647	0.81	2.07
	B/C	27	21,519	0.80	2.37	5,859	0.80	2.37

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 S601 Paper by IEP Status

Cluster	Tier	No. of Students	No. of Items	Cronbach's Alpha	SEM
K	-	22,254	30	0.96	1.73
1	A	3,102	24	0.66	2.25
	B/C	3,578	27	0.77	2.37
2	A	3,102	24	0.66	2.25
	B/C	3,578	27	0.77	2.37
3	A	1,621	24	0.73	2.24
	B/C	9,489	27	0.69	2.38
4-5	A	1,621	24	0.73	2.24
	B/C	9,489	27	0.69	2.38
6-8	A	537	24	0.68	2.22
	B/C	4,616	27	0.69	2.38
9-12	A	473	24	0.76	2.15
	B/C	2,114	27	0.74	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 S601 Paper

Cluster	Tier	No. of Students	No. of Tasks	Total Possible Raw Score Points	Cronbach's Alpha*	SEM
K	-	256,834	6	0-17	0.92	1.15
1	A	23,675	4	0-40	0.88	2.02
	B/C	15,827	3	0-54	0.95	2.01
2	A	22,163	3	0-27	0.90	1.41
	B/C	48,161	3	0-54	0.95	1.85
3	A	22,163	3	0-27	0.90	1.41
	B/C	48,161	3	0-54	0.95	1.85
4-5	A	14,441	3	0-27	0.88	1.48
	B/C	32,962	3	0-54	0.91	2.01
6-8	A	20,609	3	0-27	0.87	1.52
	B/C	33,049	3	0-54	0.92	2.02
9-12	A	19,124	3	0-27	0.87	1.63
	B/C	31,357	3	0-54	0.92	2.02

*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 S601 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	106,542	0.92	1.15	118,121	0.93	1.14
1	A	4	0-40	11,019	0.88	2.01	12,577	0.88	2.02
	B/C	3	0-54	7,569	0.95	2.01	8,222	0.95	2.02
2	A	3	0-27	10,169	0.90	1.42	11,887	0.90	1.40
	B/C	3	0-54	22,808	0.94	1.83	25,301	0.95	1.87
3	A	3	0-27	10,169	0.90	1.42	11,887	0.90	1.40
	B/C	3	0-54	22,808	0.94	1.83	25,301	0.95	1.87
4-5	A	3	0-27	6,759	0.88	1.51	7,620	0.88	1.46
	B/C	3	0-54	15,152	0.90	1.99	17,768	0.91	2.03
6-8	A	3	0-27	9,723	0.86	1.54	10,797	0.87	1.51
	B/C	3	0-54	15,121	0.91	1.95	17,871	0.92	2.06
9-12	A	3	0-27	9,042	0.86	1.65	9,970	0.87	1.62
	B/C	3	0-54	14,456	0.92	1.99	16,778	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 5.1.3.3

Reliabilities of Domain Scores: Writ S601 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	172,318	0.91	1.13	76,509	0.93	1.16
1	A	4	0-40	19,074	0.88	2.01	4,441	0.88	2.03
	B/C	3	0-54	12,187	0.95	2.01	3,560	0.95	2.04
2	A	3	0-27	18,069	0.90	1.40	3,878	0.90	1.43
	B/C	3	0-54	38,384	0.95	1.84	9,588	0.94	1.87
3	A	3	0-27	18,069	0.90	1.40	3,878	0.90	1.43
	B/C	3	0-54	38,384	0.95	1.84	9,588	0.94	1.87
4-5	A	3	0-27	11,958	0.88	1.48	2,346	0.88	1.48
	B/C	3	0-54	25,874	0.91	2.01	6,913	0.91	2.04
6-8	A	3	0-27	17,354	0.87	1.51	3,039	0.88	1.54
	B/C	3	0-54	25,826	0.92	2.00	6,996	0.92	2.08
9-12	A	3	0-27	15,885	0.87	1.61	2,925	0.87	1.72
	B/C	3	0-54	24,299	0.92	2.00	6,805	0.92	2.07

*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 S601 Paper by IEP Status

Cluster	Tier	No. of Students	No. of Tasks	Total Possible Raw Score Points	Cronbach's Alpha*	SEM
K	-	22,249	6	0-17	0.92	1.10
1	A	2,540	4	0-40	0.89	1.92
	B/C	1,723	3	0-54	0.96	1.96
2	A	2,534	3	0-27	0.89	1.38
	B/C	7,175	3	0-54	0.96	1.90
3	A	2,534	3	0-27	0.89	1.38
	B/C	7,175	3	0-54	0.96	1.90
4-5	A	803	3	0-27	0.89	1.37
	B/C	6,676	3	0-54	0.92	2.06
6-8	A	607	3	0-27	0.87	1.50
	B/C	5,224	3	0-54	0.92	2.10
9-12	A	501	3	0-27	0.85	1.70
	B/C	2,424	3	0-54	0.92	2.01

*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 S601 Paper

Cluster	Tier	No. of Students	No. of Tasks	Total Possible Raw Score Points	Cronbach's Alpha	SEM
K	-	256,828	10	0-10	0.92	1.02
1	A	23,502	6	0-18	0.91	1.36
1	B/C	15,701	6	0-24	0.92	1.37
2	A	22,027	6	0-18	0.93	1.32
2	B/C	47,837	6	0-24	0.91	1.32
3	A	22,027	6	0-18	0.93	1.32
3	B/C	47,837	6	0-24	0.91	1.32
4-5	A	14,340	6	0-18	0.91	1.38
4-5	B/C	32,800	6	0-24	0.91	1.38
6-8	A	20,416	6	0-18	0.90	1.40
6-8	B/C	32,810	6	0-24	0.92	1.48
9-12	A	18,931	6	0-18	0.90	1.45
9-12	B/C	31,116	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 S601 Paper by Gender

Cluster	Tier	No. of Tasks	Total Possible Raw Score Points	Female No. of Students	Female Cronbach's Alpha	Female SEM	Male No. of Students	Male Cronbach's Alpha	Male SEM
K	-	10	0-10	106,542	0.92	1.01	118,119	0.91	1.03
1	A	6	0-18	10,950	0.91	1.37	12,473	0.91	1.34
1	B/C	6	0-24	7,513	0.92	1.37	8,153	0.92	1.37
2	A	6	0-18	10,115	0.93	1.32	11,806	0.93	1.32
2	B/C	6	0-24	22,649	0.91	1.31	25,136	0.91	1.32
3	A	6	0-18	10,115	0.93	1.32	11,806	0.93	1.32
3	B/C	6	0-24	22,649	0.91	1.31	25,136	0.91	1.32
4-5	A	6	0-18	6,713	0.91	1.40	7,565	0.92	1.36
4-5	B/C	6	0-24	15,063	0.91	1.38	17,695	0.91	1.37
6-8	A	6	0-18	9,647	0.89	1.42	10,683	0.90	1.38
6-8	B/C	6	0-24	15,020	0.92	1.49	17,733	0.92	1.47
9-12	A	6	0-18	8,957	0.89	1.46	9,863	0.90	1.43
9-12	B/C	6	0-24	14,343	0.93	1.41	16,650	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 S601 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	172,312	0.92	1.03	76,510	0.90	1.02
1	A	6	0-18	18,930	0.91	1.35	4,413	0.89	1.39
	B/C	6	0-24	12,090	0.92	1.37	3,531	0.92	1.38
2	A	6	0-18	17,956	0.93	1.30	3,859	0.90	1.37
	B/C	6	0-24	38,137	0.91	1.31	9,512	0.91	1.34
3	A	6	0-18	17,956	0.93	1.30	3,859	0.90	1.37
	B/C	6	0-24	38,137	0.91	1.31	9,512	0.91	1.34
4-5	A	6	0-18	11,879	0.92	1.36	2,324	0.89	1.46
	B/C	6	0-24	25,744	0.91	1.37	6,881	0.91	1.39
6-8	A	6	0-18	17,196	0.90	1.38	3,009	0.88	1.48
	B/C	6	0-24	25,630	0.92	1.48	6,956	0.92	1.47
9-12	A	6	0-18	15,728	0.90	1.43	2,894	0.88	1.49
	B/C	6	0-24	24,119	0.93	1.39	6,744	0.92	1.41

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 S601 Paper by IEP Status

Cluster	Tier	No. of Students	No. of Tasks	Total Possible Raw Score Points	Cronbach's Alpha	SEM
K	-	22,241	10	0-10	0.90	1.04
1	A	2,524	6	0-18	0.89	1.31
	B/C	1,706	6	0-24	0.93	1.32
2	A	2,507	6	0-18	0.87	1.33
	B/C	7,125	6	0-24	0.91	1.33
3	A	2,507	6	0-18	0.87	1.33
	B/C	7,125	6	0-24	0.91	1.33
4-5	A	801	6	0-18	0.86	1.39
	B/C	6,650	6	0-24	0.91	1.39
6-8	A	600	6	0-18	0.87	1.40
	B/C	5,181	6	0-24	0.91	1.50
9-12	A	497	6	0-18	0.90	1.42
	B/C	2,397	6	0-24	0.91	1.42

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

5.2   Interrater Agreement Rates

DRC raters score students' responses to the tasks included on the ACCESS Writing tests (except Kindergarten, which is scored by the Test Administrator). We describe the scoring of students' responses to these performance tasks in Section 3.2.2. DRC selects a sample of 20% of all responses scored, chosen at random during the operational scoring process, for double scoring. The tables below provide information on interrater agreement for a sample of 20% of task raters. These tables show, for each of the tasks, the percentage of agreement between two raters. The first column shows the task, and the second column shows the number of responses that were double scored. DRC selects a sample of 20% of all responses scored, chosen at random during the operational scoring process. The next columns show the rates of agreement.

For Writing, the scoring rubric that the raters used defines six levels of performance ranging from 0 to 6, with the possibility of awarding a "plus" score between levels (e.g., 3, 3+, or 4 are all valid scores). We considered scores that matched or were contiguous as signifying **agreement** (%AG)—for example, if Rater 1 assigned a score of 3+ while Rater 2 assigned a score of 3, 3+, or 4. We considered scores that were one whole score point apart as **adjacent scores** (%AD)— for example, if Rater 1 assigned a score of 3+ while Rater 2 assigned a score of 2+ or 4+. Finally, if two raters assigned scores that were more than one whole score point apart, we considered those scores to be **nonadjacent scores** (%NA).

As the Speaking test is scored locally, it is not possible to provide interrater agreement data for Speaking. Section 3.2.3 in Part 1 of this report describes training procedures that local raters must complete before being certified to administer and score the Speaking test.

WIDA stipulates a minimum interrater agreement rate of 70%. DRC defines this "**agreement**" as being scored as adjacent agreement (AG) for Writing. See Section 3.2.2 for more detail about how WIDA and DRC used the agreement rates to ensure that DRC maintains sufficient quality control throughout the course of scoring.

For Writing, the lowest value for interrater agreement was 95%.

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 S601 Paper

Interrater Agreement	n/a

5.2.3.1 Grade 1

Table 5.2.3.1.1

Interrater Agreement: Writ 1 A S601 Paper

Interrater Agreement	Task	No. in Sample	%AG	%AD	%NA
	1	9,948	100	0	0
	2	10,912	100	0	0
	3	18,844	98	2	0
	4	16,168	98	2	0

Table 5.2.3.1.2

Interrater Agreement: Writ 1 B/C S601 Paper

Interrater Agreement	Task	No. in Sample	%AG	%AD	%NA
	1	8,488	98	1	0
	2	10,622	98	2	0
	3	9,610	97	3	0

5.2.3.2 Grade 2

Table 5.2.3.2.1

Interrater Agreement: Writ 2 A S601 Paper

Interrater Agreement	Task	No. in Sample	%AG	%AD	%NA
	1	7,468	98	2	0
	2	9,142	98	2	0
	3	11,356	98	2	0

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

Table 5.2.3.2.2

Interrater Agreement: Writ 2 B/C S601 Paper

Interrater Agreement	Task	No. in Sample	%AG	%AD	%NA
	1	11,112	95	5	0
	2	11,474	97	3	0
	3	11,196	96	4	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 S601 Paper

Interrater Agreement	Task	No. in Sample	%AG	%AD	%NA
	1	5,800	97	3	0
	2	6,092	96	4	0
	3	7,162	98	2	0

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

Table 5.2.3.3.2

Interrater Agreement: Writ 3 B/C S601 Paper

Interrater Agreement	Task	No. in Sample	%AG	%AD	%NA
	1	11,244	95	5	0
	2	11,420	97	3	0
	3	11,256	95	5	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 S601 Paper

Interrater Agreement	Task	No. in Sample	%AG	%AD	%NA
	1	10,842	96	4	0
	2	8,782	97	3	0
	3	10,696	98	2	0

Table 5.2.3.4.2

Interrater Agreement: Writ 4-5 B/C S601 Paper

Interrater Agreement	Task	No. in Sample	%AG	%AD	%NA
	1	14,504	95	5	0
	2	14,560	96	4	0
	3	15,520	95	4	0

5.2.3.5 Grades 6–8

Table 5.2.3.5.1

Interrater Agreement: Writ 6-8 A S601 Paper

Interrater Agreement	Task	No. in Sample	%AG	%AD	%NA
	1	11,122	97	3	0
	2	12,898	97	3	0
	3	10,512	95	5	0

Table 5.2.3.5.2

Interrater Agreement: Writ 6-8 B/C S601 Paper

Interrater Agreement	Task	No. in Sample	%AG	%AD	%NA
	1	14,324	97	3	0
	2	14,432	97	3	0
	3	17,014	96	4	0

5.2.3.6 Grades 9–12

Table 5.2.3.6.1

Interrater Agreement: Writ 9-12 A S601 Paper

Interrater Agreement	Task	No. in Sample	%AG	%AD	%NA
	1	11,926	97	3	0
	2	9,104	97	3	0
	3	14,764	97	3	0

Table 5.2.3.6.2

Interrater Agreement: Writ 9-12 B/C S601 Paper

Interrater Agreement	Task	No. in Sample	%AG	%AD	%NA
	1	14,308	97	3	0
	2	14,434	97	3	0
	3	14,720	96	4	0

5.3 Conditional Standard Errors of Measurement at Cut Score

The tables in this section present information about the conditional standard errors of measurement (CSEM) values of scale scores at the most important points at which policy makers make decisions such as reclassification about students based on performance on ACCESS—the cut points between language proficiency levels. The CSEM provides information about the amount of measurement error we would expect in any student's scale score at that point on the underlying latent ability scale. We first computed CSEM values on the theta metric, which is the square root of the inverse of the Test Information Function. Next, we used the multiplicative constant of the linear equation for the domain to linearly transform those logit-based CSEM values so that we could report them on the ACCESS score scale (See Section 2.2).

When calculated using an IRT approach, CSEM values can vary across the scale scores. For example, in the Listening and Reading domains, if a student answers correctly either a very few or a very large number of items (i.e., scores at the extremes of the scale score distribution), the CSEM value will be larger than it would be if the student correctly answers a moderate number of items. Scale scores near the middle of the score distribution typically have lower CSEM values compared to scale scores near the extremes because many tests are comprised of a large proportion of moderately difficult items, which are well suited to measuring students of moderate proficiency.

We use the CSEM to construct an error band, quantifying the amount of uncertainty in a student's scale score. One CSEM below a student's scale score and one CSEM above that scale score indicates an approximate 68% confidence interval. To interpret this confidence interval, consider a student who takes the test 100 times. Assuming measurement error is normally distributed, the student's true proficiency would fall within the confidence interval 68% of the time (or 68 times out of 100).

As a rule, lower CSEM values around scale scores at important decision points are desirable. Generally speaking, the most important decision points for the ACCESS scores are at the PL 3/4 and PL 4/5 cut points, although the approaches that WIDA states use to make decisions about ACCESS scores differ. As discussed in Section 5, all WIDA states use composite scale scores when making reclassification decisions, and no WIDA state uses a single domain scale score when making those decisions. Because each grade has its own set of cut points, we provide information for each grade within a grade-level cluster.

Since we scale ACCESS test scores using an IRT approach, CSEM values for the scale scores at the highest cut points are typically large. Use of this approach tends to produce larger CSEM values at the lower and the higher ends of the score scale. In addition, because students exit the EL program when they demonstrate that they are English language proficient, there are typically fewer students at the highest cut points than at those other cut points. Therefore, the CSEM values associated with the scale scores at the highest cut points tend to be larger than those of the scale scores at the lower cut points since there are fewer students available for estimating the scores and the CSEM values for these scores.

For each domain, we present the values by tier. From these tables, it is possible to identify how well the different tiers are targeted for making decisions about students at the various proficiency level cuts.

In the tables below, the leftmost column shows the proficiency level cut (e.g., 1/2, which is the cut between PL 1 and PL 2). The second column shows the grade level. The third column shows the cut score in the scale score metric (e.g., 305). In the last column(s), the corresponding CSEM value for the scale score at each cut point are shown.

5.3.1 Listening

5.3.1.0 Kindergarten

Table 5.3.1.0

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List K S601 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 S601 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 S601 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 S601 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 Grade 4-5

Table 5.3.1.4

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 4-5 S601 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 Grade 6-8

Table 5.3.1.5

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 6-8 S601 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 Grade 9-12

Table 5.3.1.6

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: List 9-12 S601 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	20.66	19.91
	12	342	21.04	19.54
2/3	9	353	21.42	19.16
	10	358	21.80	18.79
	11	364	22.17	18.79
	12	368	22.54	18.79
3/4	9	383	24.80	18.79
	10	389	25.55	18.79
	11	394	26.68	19.16
	12	398	27.43	19.16
4/5	9	409	30.06	19.54
	10	415	31.94	20.29
	11	420	33.44	20.66
	12	426	35.32	21.04
5/6	9	434	38.32	22.17
	10	441	41.33	22.92
	11	447	44.33	24.05
	12	452	46.96	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 S601 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 S601 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 S601 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 S601 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 Grade 4-5

Table 5.3.2.4

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 4-5 S601 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 Grade 6-8

Table 5.3.2.5

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 6-8 S601 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 Grade 9-12

Table 5.3.2.6

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Read 9-12 S601 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 S601 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 S601 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 S601 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 S601 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 Grade 4-5

Table 5.3.3.4

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 4-5 S601 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 Grade 6-8

Table 5.3.3.5

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 6-8 S601 Paper

Proficiency Level Cut Point	Grade	Cut Score	CSEM	
			Tier A	Tier B/C
1/2	6	268	12.35	8.32
	7	273	12.62	8.32
	8	281	13.69	8.59
2/3	6	298	15.84	10.20
	7	305	16.65	11.01
	8	311	16.92	11.55
3/4	6	361	17.45	12.62
	7	367	17.18	12.35
	8	372	16.92	12.35
4/5	6	413	15.57	10.74
	7	419	16.11	10.74
	8	424	16.92	11.01
5/6	6	441	20.68	12.35
	7	450	23.90	14.23
	8	459	27.93	16.38

5.3.3.6 Grade 9-12

Table 5.3.3.6

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Writ 9-12 S601 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 S601 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 S601 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 S601 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 S601 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 Grade 4-5

Table 5.3.4.4

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 4-5 S601 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 Grade 6-8

Table 5.3.4.5

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 6-8 S601 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 Grade 9-12

Table 5.3.4.6

Conditional Standard Errors of Measurement of Scale Scores at the Cut Points: Spek 9-12 S601 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

One of the main purposes of the WIDA ACCESS program is to identify the English language proficiency level of students with respect to the WIDA ELD Standards. Because of the emphasis on the classification of student performance, a question of interest is how accurately and consistently ACCESS domain scale scores can classify students into the WIDA proficiency levels determined by the 2016 ACCESS standard-setting process (Cook & MacGregor, 2017). Test users can examine indices that report on the accuracy and consistency of these classifications and can use that information to judge the utility of WIDA's proficiency level categorization, while policy makers can use these indices to assist them when making decisions about ACCESS test design and score reporting (American Educational Research Association et al., 2014). The analyses we conduct to examine the accuracy and consistency of classifications utilize the methods that Livingston and Lewis (1995) and Young and Yoon (1998) outlined, as implemented in the software program BB-CLASS (Brennan, 2004; cf. also Lee, Hanson, & Brennan, 2002).

Classification accuracy is defined conceptually as the extent to which the proficiency classifications of students based on their observed raw scores or scale scores would agree with those made based on their true scores (Livingston, 2018; Livingston & Lewis, 1995). A student's true score is the average of the scores that the student would have received, averaging over some set of prespecified factors or conditions (e.g., different versions of the test, different times of test administration). Therefore, the calculation of the true scores depends upon the particular factors over which one chooses to average (Livingston, 2018). We assume that true scores measure perfectly, but those scores are unknown. Therefore, to provide the best estimation of classification accuracy for WIDA, we use test data from one ACCESS administration to estimate students' true scale scores based on their domain scale scores and the parameters of the model used in estimating those true scale scores. We can then use the results from our analysis to estimate the percentages of the students who were accurately classified into each proficiency level.

Classification consistency is defined conceptually as the extent to which the proficiency classifications of students agree, given two independent administrations of the same or two parallel test forms. It is impractical to obtain repeated administrations of the same or parallel test forms because of cost, testing burden, and the effects of student memory and practice. However,

it is possible to estimate the percentages of the students who would be consistently classified with the assumption that the same test is independently administered twice to the same group of students.

The approach that Livingston and Lewis (1995) took, which we implemented here, uses information about the reliability of the students' domain scale scores, the cut points, and the observed distribution of scores. Then, using a four-parameter beta distribution, we model the distribution of the true scale scores and of the domain scale scores on a parallel form. The Livingston and Lewis procedure requires that the reliability estimate of the students' scores on a test form be provided when calculating the classification consistency and accuracy indices. For Listening and Reading, we used the Rasch student separation reliability estimates by grade-level clusters in the procedure. Since the Writing and Speaking tests were tiered, we needed to produce a single reliability estimate across tiers to implement the Livingston and Lewis procedure. This is a weighted reliability estimate across tiers (see Section 5.1).

Overall Classification Accuracy and Consistency

Overall classification accuracy indicates the percentage of all students whom we would classify into the same language proficiency level by both their domain scale scores and their true scale scores (i.e., the percentage of students whom we accurately classified). For example, an overall classification accuracy index of 0.774 means that we would classify 77% of the students into the same proficiency level according to their domain scale scores and their true scale scores. **Overall classification consistency** indicates the percentage of all students whom we would classify into the same language proficiency levels by their performances on both the administered test and on a parallel test. For example, an overall classification consistency index of 0.664 means that we would classify 66% of the students into the same proficiency level if they took two parallel forms of the test. A classification consistency index is always lower than its corresponding classification accuracy index, because in classification consistency, a classification based on a student's performance on the administered test and a classification based on that student's performance on a parallel test are both subject to measurement error. In contrast, in classification accuracy, only the classification based on a student's performance on the administered test contains error while we assume that the classification based on that student's true scale score is free of measurement error.

Marginal Classification Accuracy and Consistency

Overall classification accuracy and consistency indices indicate the degree to which we accurately and consistently classify students into the same WIDA proficiency levels, but not the degree to which we accurately or consistently classify students into the proficiency levels below or above the specific cut point (e.g., at the PL 4/PL 5 cut point). The indices that can address this question are **marginal classification accuracy and consistency indices based on domain scale scores at the cut points**. From an accountability perspective, the most important indices for test users and policy makers to examine are the marginal classification accuracy and consistency indices.

The **marginal classification accuracy indices based on domain scale scores at the cut points** report the percentage of students whom we accurately placed into proficiency levels above and below each cut point based on their domain scale scores. For example, a classification accuracy index of 0.774 at the PL 4/PL 5 cut point means that we would classify 77% of the students in the same way using their domain scale scores or their true scale scores, either into the proficiency levels below the cut point (i.e., PL 1 to PL 4) or into the proficiency levels above the cut point (i.e., PL 5 to PL 6). The **marginal classification consistency indices based on domain scale scores at the cut points** report the percentage of students whom we would classify consistently above and below each cut point based on their domain scale scores. For example, a classification consistency index of 0.664 at the PL 4/PL 5 cut point means that we would classify 66% of the students in the same way if they took two parallel forms, either into the proficiency levels below the cut point (i.e., PL 1 to PL 4) or into the proficiency levels above the cut point (i.e., PL 5 to PL 6). Note that the marginal accuracy and consistency indices are generally higher for students' domain scale scores at the cut points than are the overall classification accuracy and consistency indices (Livingston, 2018). This is because the marginal accuracy and consistency indices report the classification decisions at one cut point at a time while the overall accuracy and consistency indices report the classification decisions at all five cut points at the same time.

The interactions of a number of factors affect the calculation of classification accuracy and consistency: (1) the number of proficiency level cut points, (2) the magnitude of the test score reliability coefficient, (3) measurement accuracy for scale scores at the cut points, (4) the distances between adjacent cut points, (5) the locations of the cut points on the ability scale, and (6) the proportion of students' scale scores around a cut point (Ercikan & Julian, 2002; Lee et al., 2002). These factors are functions of the test design and, most importantly, the standard-setting decisions. The indices are lower when there is a greater number of proficiency levels, a lower

test score reliability coefficient, and higher measurement accuracy of the scale scores at the cut points, as well as when the two adjacent cut points are closer, and when more students' domain scale scores are around a cut point. Furthermore, the numbers and types of items on a test affect the calculation of the test score reliability coefficient. The lower the test score reliability, the lower the classification accuracy and consistency indices would be. For example, the test score reliability coefficient for the ACCESS Online Writing domain raw scores would be lower than the test score reliability coefficients for similar tests that include more items or tasks since we estimate the test score reliability coefficient for ACCESS Online Writing domain raw scores based on students' performance on only two tasks. Therefore, the classification accuracy and consistency indices for the Writing domain might be lower than those for other domains.

For each test domain, we present three tables. The first reports indices that describe the overall accuracy and overall consistency of the proficiency level classifications for each grade level. The second reports the marginal classification accuracy indices based on domain scale scores at the cut points for each grade level. The third reports the marginal classification consistency indices based on domain scale scores at the cut points for each grade level. If we could not estimate the overall and marginal classification accuracy and consistency indices because we classified fewer than 200 students into a given proficiency level, we combined the affected proficiency level and the proficiency level below it and placed 'N/A' in the table for the affected proficiency level.

Assessment experts have issued little guidance to aid in making judgments about the ideal or expected levels of decision consistency and accuracy needed for educational assessments since many different factors affect the calculation of these indices, as discussed earlier. To help test users and policy makers interpret the results from our classification analyses, for each of the ACCESS test domains, we report the range of the overall classification accuracy and consistency indices across grades. Additionally, we highlight the grade with the lowest classification accuracy and consistency indices. Since the overall accuracy and consistency indices are summaries of the degree of classification accuracy and consistency across all proficiency level cut points, we also report the marginal classification accuracy and consistency indices for these grades to identify the specific source(s) of low classification accuracy and consistency.

For Listening, as shown in Table 5.4.1.1, the overall classification accuracy indices ranged from 0.334 to 0.717 and the overall classification consistency indices ranged from 0.286 to 0.658. Grade 7 had the lowest overall classification accuracy and consistency indices for Listening.

For Reading, as shown in Table 5.4.2.1, the overall classification accuracy indices ranged from 0.436 to 0.854 and the overall classification consistency indices ranged from 0.342 to 0.833. Grade 3 had the lowest overall classification accuracy and consistency indices for Reading.

For Writing, as shown in Table 5.4.3.1, the overall classification accuracy indices ranged from 0.723 to 0.825, and the overall classification consistency indices ranged from 0.640 to 0.769. Grade 4 had the lowest overall classification accuracy and consistency indices for Writing.

For Speaking, as shown in Table 5.4.4.1, the overall classification accuracy indices ranged from 0.527 to 0.692 and the overall classification consistency indices ranged from 0.510 to 0.604. The lowest overall classification accuracy index for Speaking was Kindergarten and Grade 4 had the lowest overall classification consistency indices for Speaking.

From an accountability perspective, the most important indices for test users and policy makers to examine are the marginal classification accuracy and consistency indices. To help them interpret our results, we report for each domain the range of the marginal classification accuracy and consistency indices across grades and then highlight the grades (and the cut points within those grades) that had the lowest marginal classification accuracy and the lowest classification consistency.

For Listening, the marginal classification accuracy indices based on scale scores at the cut points ranged from 0.779 to 0.989 (Table 5.4.1.2) and the marginal classification consistency indices ranged from 0.708 to 0.980 (Table 5.4.1.3). Grade 7, at the PL 4/PL 5 cut point, had the lowest classification accuracy index for Listening and Grade 3, at the PL 4/PL 5 cut point, had the lowest classification consistency index for Listening. Note that Grade 7 was also identified as having the lowest overall classification accuracy and consistency indices in the Listening domain. The low marginal classification accuracy and consistency at the PL 4/PL 5 cut score appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal classification accuracy and consistency for Grade 7 and Grade 3 Listening were still in the .70's and .90's.

For Reading, the marginal classification accuracy indices based on scale scores at the cut points ranged from 0.786 to 0.984 (Table 5.4.2.2) and the marginal classification consistency indices scores ranged from 0.719 to 0.976 (Table 5.4.2.3). Grade 3, at PL 3/PL 4 cut point, had the lowest classification accuracy and consistency indices. Note that Grade 3 was also identified as having the lowest overall classification accuracy and consistency indices in the Reading domain.

The low marginal classification consistency at the PL 3/PL 4 cut appeared to have contributed to its low overall classification consistency. However, it should be noted that the marginal classification accuracy and consistency for Grade 3 Reading were still in the .70's and .90's.

For Writing, the marginal classification accuracy indices based on scale scores at the cut points ranged from 0.814 to 0.974 (Table 5.4.3.2) and the marginal classification consistency indices ranged from 0.759 to 0.972 (Table 5.4.3.3). Grade 4, at the PL 3/PL 4 cut point, had the lowest classification accuracy and consistency indices for Writing. Note that Grade 4 was also identified as having the lowest overall classification accuracy and consistency indices in the Writing domain. The low marginal classification accuracy and consistency at the PL 3/PL 4 cut appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal classification accuracy and consistency for Grade 4 Writing were still in the .70's and .90's.

For Speaking, the marginal classification accuracy indices based on scale scores at the cut points ranged from 0.768 to 0.989 (Table 5.4.4.2) and the marginal classification consistency indices ranged from 0.815 to 0.989 (Table 5.4.4.3). Kindergarten, at the PL 5/PL 6 cut point, had the lowest classification accuracy value for Speaking and Grade 4, at the PL 4/PL 5 cut point, had the lowest classification consistency index for Speaking. Note that Kindergarten was also identified as having the lowest overall classification accuracy index and Grade 4 was also identified as having the lowest overall classification consistency index in the Speaking domain. The low marginal classification accuracy and consistency at the PL 4/PL 5 and PL 5/PL 6 cut appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal classification accuracy and consistency for Kindergarten Speaking were still in the .70's and .90's.

When we compared the overall and marginal classification accuracy and consistency indices based on the domain scale scores for a particular grade, we saw that in many instances they told the same story (i.e., for a given grade, when the overall classification accuracy and consistency indices were low, then the marginal classification accuracy and consistency indices also tended to be low).

We observed that in the domains of Reading and Writing, the marginal classification accuracy and consistency indices for PL cut points in the middle of the proficiency level range (e.g., the PL 2/3 and PL 3/4 cut points) tended, on average, to be lower than the marginal classification

accuracy and consistency indices for cut points at the lower and upper ends of the range, a finding that is consistent with findings from previous researchers (Ercikan & Julian, 2002; Lee et al., 2002). One possible reason might be that the cut points for the proficiency levels in the middle of the proficiency level range tend to be closer together than the cut points for the proficiency levels at the ends of that range. (Cut points tend to be closer to each other when there are a large number of proficiency levels.) We would expect marginal classification accuracy and consistency to vary for different ability levels due to variation in measurement accuracy. That is, the further away the students' domain scale scores are from the cut points, the smaller the classification errors would be, or the more accurate the classification decisions would be. With many proficiency levels, there are more student domain scale scores near the cut points than there would be if there were fewer proficiency levels. Therefore, the higher the number of proficiency levels, the higher the probability that we would misclassify students (Ercikan & Julian, 2002). Additionally, the intervals between cut points that are in the middle of the ACCESS proficiency level range are smaller than the intervals between cut points that are at the upper and lower ends of the proficiency level range. Consequently, the marginal classification accuracy and consistency indices based on the domain scale scores for the PL 2/3 and PL 3/4 cut points tend to be lower than for other cut points, as we might expect.

Although assessment experts have issued little guidance to aid in making judgments about the ideal or expected levels of decision consistency and accuracy needed for educational assessments since many different factors affect the calculation of these indices, as discussed earlier, the ranges of the classification accuracy and consistency indices for the ACCESS domains are very similar to those reported for similar testing programs such as ELPA21 (American Institutes of Research, 2018), with the exception of the Writing domain. Since the ACCESS Online Writing test consists of only two tasks, the test score reliability estimate may be lower than similar writing tests that include more tasks. The classification accuracy and consistency indices derived using the Livingston and Lewis (1995) procedure are affected by the magnitude of the test score reliability, which is lower when a test has fewer tasks. Also note that we would not expect the indices estimated for ACCESS domains to be exactly the same as those computed in other programs, because testing programs differ in their student populations, the numbers of proficiency levels, their test designs, their score distributions, and the methods used to compute classification accuracy and consistency indices. For example, compared to similar testing programs, students taking ACCESS represent a much larger and more diverse population.

Additionally, the ACCESS testing program defines more proficiency levels than other similar testing programs, and the ACCESS test design is more complex. Therefore, it is difficult to compare the classification accuracy and consistency indices for ACCESS domains to those for other testing programs.

5.4.1 Listening

Table 5.4.1.1

Overall Accuracy and Consistency of Classification Indices: List S601 Paper

Grade	Accuracy	Consistency
K	0.717	0.658
1	0.478	0.395
2	0.472	0.384
3	0.450	0.360
4	0.454	0.359
5	0.437	0.344
6	0.382	0.303
7	0.334	0.286
8	0.363	0.292
9	0.390	0.316
10	0.414	0.327
11	0.408	0.322
12	0.428	0.332

Table 5.4.1.2

Marginal Classification Accuracy Indices Based on the Domain Scale Scores at the Cut Points: List S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.944	0.936	0.929	0.922	0.911
1	0.940	0.888	0.840	0.820	0.850
2	0.946	0.897	0.851	0.821	0.817
3	0.989	0.919	0.837	0.793	0.781
4	0.978	0.907	0.849	0.792	0.797
5	0.965	0.891	0.838	0.793	0.813
6	0.923	0.861	0.816	0.787	0.850
7	0.851	0.860	0.836	0.779	0.872
8	0.866	0.854	0.870	0.809	0.841
9	0.860	0.849	0.822	0.829	0.895
10	0.863	0.834	0.800	0.842	0.939
11	0.877	0.819	0.804	0.849	0.921
12	0.876	0.813	0.788	0.875	0.958

Table 5.4.1.3

Marginal Classification Consistency Indices Based on the Domain Scale Scores at the Cut Points: List S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.922	0.911	0.901	0.894	0.875
1	0.913	0.846	0.775	0.758	0.791
2	0.924	0.852	0.789	0.755	0.752
3	0.980	0.887	0.767	0.708	0.711
4	0.963	0.869	0.782	0.720	0.725
5	0.945	0.846	0.768	0.721	0.745
6	0.892	0.803	0.741	0.717	0.795
7	0.830	0.794	0.760	0.720	0.799
8	0.819	0.790	0.795	0.738	0.782
9	0.817	0.779	0.754	0.771	0.855
10	0.813	0.761	0.732	0.783	0.893
11	0.824	0.753	0.732	0.787	0.890
12	0.822	0.741	0.718	0.812	0.927

5.4.2 Reading

Table 5.4.2.1

Overall Accuracy and Consistency of Classification Indices: Read S601 Paper

Grade	Accuracy	Consistency
K	0.854	0.833
1	0.503	0.402
2	0.567	0.462
3	0.436	0.342
4	0.492	0.393
5	0.509	0.410
6	0.625	0.511
7	0.600	0.489
8	0.580	0.471
9	0.557	0.454
10	0.574	0.464
11	0.561	0.454
12	0.596	0.486

Table 5.4.2.2

Marginal Classification Accuracy Indices Based on the Domain Scale Scores at the Cut Points: Read S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.966	0.968	0.960	0.937	N/A
1	0.795	0.826	0.906	0.942	0.971
2	0.871	0.861	0.901	0.921	0.957
3	0.910	0.830	0.786	0.843	0.939
4	0.918	0.861	0.824	0.860	0.939
5	0.899	0.842	0.845	0.875	0.946
6	0.884	0.856	0.921	0.948	0.984
7	0.873	0.851	0.911	0.942	0.978
8	0.862	0.859	0.902	0.932	0.972
9	0.895	0.865	0.874	0.907	0.956
10	0.902	0.852	0.888	0.917	0.956
11	0.916	0.852	0.870	0.896	0.953
12	0.906	0.856	0.900	0.916	0.957

Table 5.4.2.3

Marginal Classification Consistency Indices Based on the Domain Scale Scores at the Cut Points: Read S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.948	0.953	0.946	0.933	N/A
1	0.733	0.761	0.863	0.911	0.960
2	0.820	0.810	0.857	0.887	0.938
3	0.874	0.756	0.719	0.775	0.904
4	0.886	0.801	0.770	0.806	0.904
5	0.856	0.782	0.791	0.825	0.914
6	0.836	0.798	0.887	0.926	0.976
7	0.821	0.796	0.873	0.914	0.967
8	0.808	0.803	0.864	0.902	0.957
9	0.856	0.809	0.832	0.869	0.934
10	0.861	0.796	0.843	0.879	0.936
11	0.879	0.795	0.823	0.854	0.927
12	0.867	0.801	0.856	0.880	0.940

5.4.3 Writing

Table 5.4.3.1

Overall Accuracy and Consistency of Classification Indices: Writ S601 Paper

Grade	Accuracy	Consistency
K	0.799	0.766
1	0.767	0.694
2	0.824	0.767
3	0.825	0.769
4	0.723	0.640
5	0.739	0.652
6	0.777	0.696
7	0.776	0.693
8	0.782	0.701
9	0.772	0.686
10	0.767	0.682
11	0.765	0.679
12	0.783	0.700

Table 5.4.3.2

Marginal Classification Accuracy Indices Based on the Domain Scale Scores at the Cut Points: Writ S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.920	0.878	0.974	N/A	N/A
1	0.919	0.847	N/A	N/A	N/A
2	0.947	0.917	0.960	N/A	N/A
3	0.959	0.930	0.937	N/A	N/A
4	0.964	0.943	0.814	N/A	N/A
5	0.962	0.936	0.838	N/A	N/A
6	0.940	0.913	0.923	N/A	N/A
7	0.933	0.910	0.933	N/A	N/A
8	0.934	0.916	0.931	N/A	N/A
9	0.944	0.920	0.907	N/A	N/A
10	0.940	0.914	0.912	N/A	N/A
11	0.943	0.917	0.904	N/A	N/A
12	0.941	0.917	0.923	N/A	N/A

Table 5.4.3.3

Marginal Classification Consistency Indices Based on the Domain Scale Scores at the Cut Points: Writ S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.885	0.867	0.972	N/A	N/A
1	0.885	0.804	N/A	N/A	N/A
2	0.924	0.884	0.958	N/A	N/A
3	0.941	0.902	0.925	N/A	N/A
4	0.948	0.923	0.759	N/A	N/A
5	0.945	0.912	0.784	N/A	N/A
6	0.915	0.878	0.895	N/A	N/A
7	0.906	0.873	0.906	N/A	N/A
8	0.907	0.879	0.902	N/A	N/A
9	0.920	0.888	0.870	N/A	N/A
10	0.915	0.881	0.877	N/A	N/A
11	0.919	0.885	0.867	N/A	N/A
12	0.917	0.881	0.890	N/A	N/A

5.4.4 Speaking

Table 5.4.4.1

Overall Accuracy and Consistency of Classification Indices: Spek S601 Paper

Grade	Accuracy	Consistency
K	0.527	0.527
1	0.692	0.597
2	0.659	0.566
3	0.658	0.563
4	0.601	0.510
5	0.621	0.532
6	0.626	0.530
7	0.632	0.542
8	0.650	0.560
9	0.666	0.584
10	0.690	0.604
11	0.675	0.589
12	0.666	0.585

Table 5.4.4.2

Marginal Classification Accuracy Indices Based on the Domain Scale Scores at the Cut Points: Spek S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.892	0.931	0.954	0.957	0.768
1	0.935	0.906	0.917	0.933	0.989
2	0.949	0.903	0.880	0.925	0.978
3	0.946	0.901	0.876	0.933	0.965
4	0.952	0.924	0.897	0.851	0.939
5	0.944	0.919	0.898	0.831	0.957
6	0.933	0.906	0.906	0.913	0.948
7	0.931	0.912	0.905	0.921	0.938
8	0.931	0.917	0.910	0.914	0.954
9	0.924	0.911	0.925	0.935	0.946
10	0.928	0.918	0.919	0.953	0.949
11	0.933	0.913	0.920	0.956	0.931
12	0.927	0.906	0.931	0.965	0.921

Table 5.4.4.3

Marginal Classification Consistency Indices Based on the Domain Scale Scores at the Cut Points: Spek S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.856	0.905	0.930	0.933	0.826
1	0.906	0.870	0.880	0.924	0.989
2	0.925	0.866	0.836	0.915	0.977
3	0.920	0.863	0.831	0.923	0.965
4	0.930	0.894	0.857	0.815	0.934
5	0.918	0.888	0.858	0.838	0.954
6	0.904	0.872	0.868	0.876	0.941
7	0.902	0.878	0.867	0.882	0.934
8	0.901	0.883	0.873	0.882	0.947
9	0.892	0.877	0.892	0.908	0.942
10	0.897	0.885	0.886	0.926	0.949
11	0.904	0.878	0.885	0.931	0.934
12	0.897	0.870	0.900	0.948	0.929

5.5　Reliabilities of Students' Composite Scale Scores

The reliabilities of the ACCESS composite scale scores indicate the consistency of those scores over replications of the testing procedure. Because the domains that make up the composites consist of different test items, and because items from different domains may measure different abilities (even though items within the domain are assumed to measure a single ability), a traditional internal consistency index such as Cronbach's coefficient alpha is not appropriate, since statisticians who devised such indices assumed that items in a test measure similar ability. It is more appropriate to report a stratified Cronbach's coefficient alpha (Feldt & Brennan, 1989), which measures consistency in students' composite scale scores when those scores are based on students' responses to sets of items that measure different abilities. A stratified alpha is a weighted average of Cronbach's coefficient alphas for item sets that differ in the maximum score points or "strata." Stratified alpha is a reliability estimate computed by dividing the test into components (strata), computing a Cronbach's coefficient alpha separately for the scale scores for each component, and then using the results to estimate a reliability coefficient for the composite scale scores.

In computing the stratified Cronbach's coefficient alphas for ACCESS composite scale scores, we treated each domain that makes up a composite as a separate component (or stratum). For example, when computing the stratified Cronbach's coefficient alphas for students' Literacy scale scores, we entered the variances of the students' scale scores for two components (i.e., Reading and Writing) and the weights of those two components. The stratified Cronbach's coefficient alpha is interpreted like other traditional internal consistency statistics such as Cronbach's coefficient alpha. Like Cronbach's coefficient alpha, a stratified Cronbach's coefficient alpha is an estimate of the proportion of the total variance in the students' composite scale scores that the variance in their true composite scale scores can explain.

Because of the differential weights applied to the ACCESS domains that contribute to the students' composite scale scores, the stratified Cronbach's coefficient alpha is weighted by the contribution that each domain makes to the students' composite scale scores (Kamata, Turhan, & Darandari, 2003; Kane & Case, 2004; Rudner, 2001). Specifically, the formula is

$$\alpha_c = 1 - \frac{\sum_{j=1}^{k} w_j^2 \sigma_j^2 (1-\rho_j)}{\sigma_c^2}$$

where

k = the number of components (domains) j that contribute to the composite

w_j = the weight of component (domain) j

σ_j^2 = the variance of the students' scale scores for component (domain) j

σ_c^2 = the variance of the students' composite scale scores

ρ_j = the reliability coefficient for students' scale scores for component (domain) j.

As is true for the Cronbach's coefficient alpha (see the explanation in Section 5), there is no one set of criteria that the testing community uses when interpreting stratified Cronbach's coefficient alpha values. There is little consensus among the experts in their views of what the acceptable lower limit of the stratified Cronbach's coefficient alpha value should be, or for that matter, how one should interpret various values. This lack of consensus led the authors of the *Standards for Educational and Psychological Measurement* (2014) to conclude, "The choice of [reliability/precision] estimation and the minimum acceptable level for any index remain a matter of professional judgment" (p. 41).

The tables report the stratified Cronbach's coefficient alphas for the students' scale scores for each of the four composites (Oral, Literacy, Comprehension, Overall). The first table for each composite provides stratified Cronbach's coefficient alphas for all students' composite scale scores. The second table for each composite provides the same information for the population of female students and for the population of male students. The third table provides information by ethnicity, for Hispanic and for non-Hispanic students, and the fourth table provides information for the population of students who have an IEP.

The first column of each table shows the grade-level clusters. The tables report the input values that we used to compute the stratified Cronbach's coefficient alphas (i.e., the number of components for each composite, each component's weight, and the variance of the students' scale scores for each component). See Chapter 3 for an explanation of the procedures we used to compute the composite scale scores.

For each grade-level cluster excluding Kindergarten, we derive a reliability coefficient across tiers for each domain. (The Kindergarten test is not tiered and so this step is not necessary.) To produce this coefficient, values for Cronbach's alpha for each of the tiers in the grade-level cluster (provided in Section 5.1) are weighted by the number of students who were administered the tier form, and a weighted average is expressed in the tables.

For each relevant domain component, we report the variance of the students' domain scale scores. We also report the variance of the students' composite scale scores. When we computed the variances of the students' domain scale scores and the variances of the students' composite scale scores, we included the students who had valid scores for all four domains.

Finally, the tables present the computed stratified Cronbach's coefficient alphas for students' scale scores for each composite, by grade-level cluster.

Additionally, we used the stratified Cronbach's coefficient alphas, presented in the tables in this section, to produce the **Accuracy and Consistency** classification tables for the composites (Section 5.7).

The stratified Cronbach's alpha of the Oral composite computed for all students ranged from 0.89 to 0.96. The stratified Cronbach's alpha of the Oral composite ranged from 0.89 to 0.96 for male students; from 0.90 to 0.96 for female students; from 0.90 to 0.96 for Hispanic students; from 0.89 to 0.95 for non-Hispanic students; and from 0.86 to 0.96 for students with an IEP.

The stratified Cronbach's alpha of the Literacy composite computed for all students ranged from 0.90 to 0.96. The stratified Cronbach's alpha of the Literacy composite ranged from 0.90 to 0.97 for male students; from 0.89 to 0.96 for female students; from 0.89 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 0.77 to 0.97. The stratified Cronbach's alpha of the Comprehension composite ranged from 0.78 to 0.97 for male students; from 0.75 to 0.96 for female students; from 0.76 to 0.96 for Hispanic students; from 0.79 to 0.97 for non-Hispanic students; and from 0.71 to 0.97 for students with an IEP.

The stratified Cronbach's alpha of the Overall composite computed for all students ranged from 0.94 to 0.98. The stratified Cronbach's alpha of the Overall composite ranged from 0.94 to 0.98 for male students; from 0.94 to 0.98 for female students; from 0.94 to 0.97 for Hispanic students; from 0.94 to 0.98 for non-Hispanic students; and from 0.92 to 0.98 for students with an IEP.

5.5.1 Oral

Table 5.5.1.1

Reliabilities of Composite Scale Scores: Oral S601 Paper

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.50	6813.57	0.95
	Speaking	0.50	10907.52	0.92
	Oral		7906.12	0.96
1	Listening	0.50	1894.47	0.73
	Speaking	0.50	5034.72	0.91
	Oral		2634.76	0.91
2	Listening	0.50	2197.67	0.70
	Speaking	0.50	5443.56	0.92
	Oral		3001.06	0.91
3	Listening	0.50	1441.93	0.57
	Speaking	0.50	5561.84	0.92
	Oral		2566.44	0.89
4-5	Listening	0.50	1996.78	0.63
	Speaking	0.50	7039.72	0.91
	Oral		3533.57	0.90
6-8	Listening	0.50	2538.38	0.63
	Speaking	0.50	7077.62	0.91
	Oral		3840.49	0.90
9-12	Listening	0.50	2495.26	0.63
	Speaking	0.50	7040.02	0.92
	Oral		3813.24	0.90

Table 5.5.1.2

Reliabilities of Composite Scale Scores: Oral S601 Paper by Gender

Cluster	Component	Weight	Female		Male	
			Variance	Reliability	Variance	Reliability
K	Listening	0.50	6561.45	0.95	7070.66	0.96
	Speaking	0.50	11150.35	0.92	10588.09	0.91
	Oral		7875.26	0.96	7910.57	0.96
1	Listening	0.50	1841.21	0.72	1934.55	0.73
	Speaking	0.50	5107.70	0.91	4941.96	0.91
	Oral		2629.07	0.91	2622.76	0.91
2	Listening	0.50	2173.52	0.69	2204.03	0.71
	Speaking	0.50	5492.31	0.92	5386.49	0.92
	Oral		3012.48	0.91	2977.57	0.91
3	Listening	0.50	1386.40	0.56	1484.61	0.58
	Speaking	0.50	5743.11	0.92	5381.80	0.92
	Oral		2590.62	0.90	2532.02	0.89
4-5	Listening	0.50	1944.58	0.62	2036.58	0.64
	Speaking	0.50	7375.29	0.91	6730.68	0.91
	Oral		3633.57	0.90	3435.91	0.90
6-8	Listening	0.50	2527.00	0.62	2546.83	0.64
	Speaking	0.50	7129.85	0.91	7024.07	0.91
	Oral		3868.99	0.90	3811.06	0.90
9-12	Listening	0.50	2391.82	0.62	2590.99	0.64
	Speaking	0.50	6974.71	0.92	7099.92	0.92
	Oral		3730.96	0.90	3889.40	0.90

Table 5.5.1.3

Reliabilities of Composite Scale Scores: Oral S601 Paper by Ethnicity

Cluster	Component	Weight	Hispanic		Other	
			Variance	Reliability	Variance	Reliability
K	Listening	0.50	6842.60	0.95	5974.65	0.95
	Speaking	0.50	10851.28	0.92	10050.34	0.90
	Oral		7891.34	0.96	7056.27	0.95
1	Listening	0.50	1891.90	0.72	1860.19	0.74
	Speaking	0.50	5119.02	0.92	4524.46	0.90
	Oral		2664.17	0.91	2417.82	0.90
2	Listening	0.50	2236.55	0.70	1969.13	0.71
	Speaking	0.50	5579.75	0.92	4613.85	0.90
	Oral		3085.09	0.91	2504.93	0.90
3	Listening	0.50	1437.28	0.56	1428.86	0.60
	Speaking	0.50	5702.52	0.92	4784.98	0.91
	Oral		2620.74	0.90	2245.55	0.89
4-5	Listening	0.50	2029.84	0.63	1776.15	0.65
	Speaking	0.50	7342.32	0.91	5360.20	0.90
	Oral		3672.95	0.90	2732.28	0.89
6-8	Listening	0.50	2575.19	0.63	2198.07	0.63
	Speaking	0.50	7306.35	0.91	5553.03	0.90
	Oral		3953.97	0.90	3014.01	0.89
9-12	Listening	0.50	2511.38	0.63	2269.71	0.65
	Speaking	0.50	7165.76	0.92	5883.61	0.91
	Oral		3873.04	0.90	3207.75	0.90

Table 5.5.1.4

Reliabilities of Composite Scale Scores: Oral S601 Paper by IEP Status

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.50	7135.62	0.96
	Speaking	0.50	8954.07	0.90
	Oral		7218.70	0.96
1	Listening	0.50	1921.36	0.74
	Speaking	0.50	4175.41	0.91
	Oral		2277.80	0.90
2	Listening	0.50	1925.51	0.73
	Speaking	0.50	3826.00	0.90
	Oral		2120.85	0.89
3	Listening	0.50	1164.38	0.56
	Speaking	0.50	3486.41	0.90
	Oral		1576.16	0.86
4-5	Listening	0.50	1241.09	0.61
	Speaking	0.50	3497.40	0.90
	Oral		1636.94	0.87
6-8	Listening	0.50	1403.30	0.59
	Speaking	0.50	3924.74	0.91
	Oral		1870.32	0.87
9-12	Listening	0.50	1669.15	0.60
	Speaking	0.50	5035.82	0.91
	Oral		2467.27	0.89

5.5.2 Literacy

Table 5.5.2.1

Reliabilities of Composite Scale Scores: Litr S601 Paper

Cluster	Component	Weight	Variance	Reliability
K	Reading	0.50	4429.83	0.95
	Writing	0.50	4484.68	0.92
	Literacy		3805.48	0.96
1	Reading	0.50	727.09	0.70
	Writing	0.50	1916.53	0.91
	Literacy		957.76	0.90
2	Reading	0.50	1105.17	0.81
	Writing	0.50	2025.71	0.93
	Literacy		1283.86	0.93
3	Reading	0.50	824.78	0.66
	Writing	0.50	1881.70	0.93
	Literacy		1073.14	0.90
4-5	Reading	0.50	1126.68	0.78
	Writing	0.50	2199.86	0.90
	Literacy		1387.45	0.92
6-8	Reading	0.50	862.17	0.77
	Writing	0.50	1990.92	0.90
	Literacy		1158.18	0.91
9-12	Reading	0.50	1019.36	0.80
	Writing	0.50	1924.87	0.90
	Literacy		1224.67	0.92

Table 5.5.2.2

Reliabilities of Composite Scale Scores: Litr S601 Paper by Gender

Cluster	Component	Weight	Female		Male	
			Variance	Reliability	Variance	Reliability
K	Reading	0.50	4251.74	0.95	4617.62	0.95
	Writing	0.50	4448.69	0.92	4545.46	0.93
	Literacy		3716.30	0.96	3915.95	0.97
1	Reading	0.50	690.89	0.69	760.10	0.71
	Writing	0.50	1773.19	0.91	2017.60	0.91
	Literacy		898.66	0.89	1003.79	0.90
2	Reading	0.50	1081.13	0.81	1127.60	0.82
	Writing	0.50	1989.56	0.93	2013.28	0.93
	Literacy		1266.96	0.93	1285.35	0.93
3	Reading	0.50	765.72	0.63	874.83	0.68
	Writing	0.50	1832.03	0.92	1867.10	0.93
	Literacy		1032.67	0.90	1094.09	0.90
4-5	Reading	0.50	1075.45	0.77	1167.63	0.79
	Writing	0.50	2283.27	0.90	2090.65	0.90
	Literacy		1403.47	0.91	1364.03	0.92
6-8	Reading	0.50	863.12	0.77	857.74	0.77
	Writing	0.50	2030.88	0.89	1908.76	0.90
	Literacy		1180.54	0.91	1119.13	0.91
9-12	Reading	0.50	1000.24	0.79	1031.02	0.80
	Writing	0.50	1971.21	0.90	1854.76	0.90
	Literacy		1239.65	0.92	1196.36	0.92

Table 5.5.2.3

Reliabilities of Composite Scale Scores: Litr S601 Paper by Ethnicity

Cluster	Component	Weight	Hispanic		Other	
			Variance	Reliability	Variance	Reliability
K	Reading	0.50	3794.23	0.94	5030.12	0.96
	Writing	0.50	4060.46	0.91	4695.06	0.93
	Literacy		3285.65	0.96	4183.54	0.97
1	Reading	0.50	701.43	0.68	806.43	0.74
	Writing	0.50	1892.64	0.91	1938.19	0.91
	Literacy		931.21	0.89	1023.44	0.91
2	Reading	0.50	1090.45	0.81	1133.14	0.83
	Writing	0.50	2032.57	0.93	1948.75	0.93
	Literacy		1280.12	0.93	1261.05	0.94
3	Reading	0.50	818.89	0.65	832.69	0.70
	Writing	0.50	1906.89	0.93	1739.59	0.93
	Literacy		1082.83	0.90	1008.09	0.91
4-5	Reading	0.50	1131.16	0.77	1065.51	0.79
	Writing	0.50	2277.95	0.90	1791.40	0.90
	Literacy		1422.95	0.91	1182.95	0.92
6-8	Reading	0.50	852.14	0.76	861.75	0.79
	Writing	0.50	2025.89	0.90	1781.74	0.90
	Literacy		1167.60	0.91	1067.59	0.92
9-12	Reading	0.50	1026.24	0.80	953.05	0.80
	Writing	0.50	1949.40	0.90	1741.32	0.90
	Literacy		1240.17	0.92	1101.78	0.92

Table 5.5.2.4

Reliabilities of Composite Scale Scores: Litr S601 Paper by IEP Status

Cluster	Component	Weight	Variance	Reliability
K	Reading	0.50	4656.00	0.96
	Writing	0.50	4090.10	0.92
	Literacy		3623.94	0.97
1	Reading	0.50	616.23	0.62
	Writing	0.50	2293.10	0.92
	Literacy		947.72	0.89
2	Reading	0.50	888.72	0.77
	Writing	0.50	2101.10	0.93
	Literacy		1143.37	0.92
3	Reading	0.50	679.03	0.57
	Writing	0.50	1747.73	0.94
	Literacy		884.41	0.89
4-5	Reading	0.50	699.36	0.72
	Writing	0.50	1414.15	0.92
	Literacy		792.93	0.90
6-8	Reading	0.50	524.06	0.69
	Writing	0.50	1377.82	0.91
	Literacy		693.35	0.90
9-12	Reading	0.50	697.58	0.74
	Writing	0.50	1441.70	0.91
	Literacy		816.01	0.90

5.5.3 Comprehension

Table 5.5.3.1

Reliabilities of Composite Scale Scores: Cphn S601 Paper

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.30	6813.57	0.95
	Reading	0.70	4429.83	0.95
	Comprehension		4024.04	0.97
1	Listening	0.30	1894.47	0.73
	Reading	0.70	727.09	0.70
	Comprehension		783.84	0.80
2	Listening	0.30	2197.67	0.70
	Reading	0.70	1105.17	0.81
	Comprehension		1140.40	0.86
3	Listening	0.30	1441.93	0.57
	Reading	0.70	824.78	0.66
	Comprehension		836.51	0.77
4-5	Listening	0.30	1996.78	0.63
	Reading	0.70	1126.68	0.78
	Comprehension		1204.12	0.84
6-8	Listening	0.30	2538.38	0.63
	Reading	0.70	862.17	0.77
	Comprehension		1086.80	0.83
9-12	Listening	0.30	2495.26	0.63
	Reading	0.70	1019.36	0.80
	Comprehension		1209.54	0.85

Table 5.5.3.2

Reliabilities of Composite Scale Scores: Cphn S601 Paper by Gender

Cluster	Component	Weight	Female		Male	
			Variance	Reliability	Variance	Reliability
K	Listening	0.30	6561.45	0.95	7070.66	0.96
	Reading	0.70	4251.74	0.95	4617.62	0.95
	Comprehension		3879.63	0.96	4185.12	0.97
1	Listening	0.30	1841.21	0.72	1934.55	0.73
	Reading	0.70	690.89	0.69	760.10	0.71
	Comprehension		754.53	0.80	809.47	0.81
2	Listening	0.30	2173.52	0.69	2204.03	0.71
	Reading	0.70	1081.13	0.81	1127.60	0.82
	Comprehension		1119.82	0.86	1156.73	0.86
3	Listening	0.30	1386.40	0.56	1484.61	0.58
	Reading	0.70	765.72	0.63	874.83	0.68
	Comprehension		781.17	0.75	882.12	0.78
4-5	Listening	0.30	1944.58	0.62	2036.58	0.64
	Reading	0.70	1075.45	0.77	1167.63	0.79
	Comprehension		1157.22	0.84	1240.79	0.85
6-8	Listening	0.30	2527.00	0.62	2546.83	0.64
	Reading	0.70	863.12	0.77	857.74	0.77
	Comprehension		1092.95	0.83	1079.98	0.83
9-12	Listening	0.30	2391.82	0.62	2590.99	0.64
	Reading	0.70	1000.24	0.79	1031.02	0.80
	Comprehension		1180.66	0.85	1234.59	0.85

Table 5.5.3.3

Reliabilities of Composite Scale Scores: Cphn S601 Paper by Ethnicity

Cluster	Component	Weight	Hispanic		Other	
			Variance	Reliability	Variance	Reliability
K	Listening	0.30	6842.60	0.95	5974.65	0.95
	Reading	0.70	3794.23	0.94	5030.12	0.96
	Comprehension		3578.95	0.96	4229.52	0.97
1	Listening	0.30	1891.90	0.72	1860.19	0.74
	Reading	0.70	701.43	0.68	806.43	0.74
	Comprehension		763.70	0.80	838.18	0.83
2	Listening	0.30	2236.55	0.70	1969.13	0.71
	Reading	0.70	1090.45	0.81	1133.14	0.83
	Comprehension		1138.57	0.86	1107.65	0.87
3	Listening	0.30	1437.28	0.56	1428.86	0.60
	Reading	0.70	818.89	0.65	832.69	0.70
	Comprehension		832.94	0.76	831.14	0.79
4-5	Listening	0.30	2029.84	0.63	1776.15	0.65
	Reading	0.70	1131.16	0.77	1065.51	0.79
	Comprehension		1217.15	0.84	1096.45	0.85
6-8	Listening	0.30	2575.19	0.63	2198.07	0.63
	Reading	0.70	852.14	0.76	861.75	0.79
	Comprehension		1086.37	0.83	1014.57	0.84
9-12	Listening	0.30	2511.38	0.63	2269.71	0.65
	Reading	0.70	1026.24	0.80	953.05	0.80
	Comprehension		1218.73	0.85	1105.83	0.85

Table 5.5.3.4

Reliabilities of Composite Scale Scores: Cphn S601 Paper by IEP Status

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.30	7135.62	0.96
	Reading	0.70	4656.00	0.96
	Comprehension		4019.81	0.97
1	Listening	0.30	1921.36	0.74
	Reading	0.70	616.23	0.62
	Comprehension		667.61	0.76
2	Listening	0.30	1925.51	0.73
	Reading	0.70	888.72	0.77
	Comprehension		896.72	0.83
3	Listening	0.30	1164.38	0.56
	Reading	0.70	679.03	0.57
	Comprehension		642.66	0.71
4-5	Listening	0.30	1241.09	0.61
	Reading	0.70	699.36	0.72
	Comprehension		675.90	0.80
6-8	Listening	0.30	1403.30	0.59
	Reading	0.70	524.06	0.69
	Comprehension		582.24	0.77
9-12	Listening	0.30	1669.15	0.60
	Reading	0.70	697.58	0.74
	Comprehension		777.81	0.81

5.5.4 Overall

Table 5.5.4.1

Reliabilities of Composite Scale Scores: Over S601 Paper

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.15	6813.57	0.95
	Reading	0.35	4429.83	0.95
	Writing	0.35	4484.68	0.92
	Speaking	0.15	10907.52	0.92
	Overall Composite		4005.10	0.98
1	Listening	0.15	1894.47	0.73
	Reading	0.35	727.09	0.70
	Writing	0.35	1916.53	0.91
	Speaking	0.15	5034.72	0.91
	Overall Composite		1121.41	0.94
2	Listening	0.15	2197.67	0.70
	Reading	0.35	1105.17	0.81
	Writing	0.35	2025.71	0.93
	Speaking	0.15	5443.56	0.92
	Overall Composite		1463.42	0.95
3	Listening	0.15	1441.93	0.57
	Reading	0.35	824.78	0.66
	Writing	0.35	1881.70	0.93
	Speaking	0.15	5561.84	0.92
	Overall Composite		1258.96	0.94
4-5	Listening	0.15	1996.78	0.63
	Reading	0.35	1126.68	0.78
	Writing	0.35	2199.86	0.90
	Speaking	0.15	7039.72	0.91
	Overall Composite		1741.66	0.95
6-8	Listening	0.15	2538.38	0.63
	Reading	0.35	862.17	0.77
	Writing	0.35	1990.92	0.90
	Speaking	0.15	7077.62	0.91
	Overall Composite		1620.06	0.95
9-12	Listening	0.15	2495.26	0.63
	Reading	0.35	1019.36	0.80
	Writing	0.35	1924.87	0.90
	Speaking	0.15	7040.02	0.92
	Overall Composite		1674.07	0.95

Table 5.5.4.2

Reliabilities of Composite Scale Scores: Over S601 Paper by Gender

Cluster	Component	Weight	Female		Male	
			Variance	Reliability	Variance	Reliability
K	Listening	0.15	6561.45	0.95	7070.66	0.96
	Reading	0.35	4251.74	0.95	4617.62	0.95
	Writing	0.35	4448.69	0.92	4545.46	0.93
	Speaking	0.15	11150.35	0.92	10588.09	0.91
	Overall Composite		3938.58	0.98	4087.36	0.98
1	Listening	0.15	1841.21	0.72	1934.55	0.73
	Reading	0.35	690.89	0.69	760.10	0.71
	Writing	0.35	1773.19	0.91	2017.60	0.91
	Speaking	0.15	5107.70	0.91	4941.96	0.91
	Overall Composite		1079.89	0.94	1149.10	0.94
2	Listening	0.15	2173.52	0.69	2204.03	0.71
	Reading	0.35	1081.13	0.81	1127.60	0.82
	Writing	0.35	1989.56	0.93	2013.28	0.93
	Speaking	0.15	5492.31	0.92	5386.49	0.92
	Overall Composite		1462.97	0.95	1450.90	0.95
3	Listening	0.15	1386.40	0.56	1484.61	0.58
	Reading	0.35	765.72	0.63	874.83	0.68
	Writing	0.35	1832.03	0.92	1867.10	0.93
	Speaking	0.15	5743.11	0.92	5381.80	0.92
	Overall Composite		1239.33	0.94	1265.34	0.94
4-5	Listening	0.15	1944.58	0.62	2036.58	0.64
	Reading	0.35	1075.45	0.77	1167.63	0.79
	Writing	0.35	2283.27	0.90	2090.65	0.90
	Speaking	0.15	7375.29	0.91	6730.68	0.91
	Overall Composite		1784.94	0.95	1697.33	0.95
6-8	Listening	0.15	2527.00	0.62	2546.83	0.64
	Reading	0.35	863.12	0.77	857.74	0.77
	Writing	0.35	2030.88	0.89	1908.76	0.90
	Speaking	0.15	7129.85	0.91	7024.07	0.91
	Overall Composite		1658.72	0.95	1577.05	0.95
9-12	Listening	0.15	2391.82	0.62	2590.99	0.64
	Reading	0.35	1000.24	0.79	1031.02	0.80
	Writing	0.35	1971.21	0.90	1854.76	0.90
	Speaking	0.15	6974.71	0.92	7099.92	0.92
	Overall Composite		1682.32	0.95	1662.17	0.95

Table 5.5.4.3

Reliabilities of Composite Scale Scores: Over S601 Paper by Ethnicity

Cluster	Component	Weight	Hispanic		Other	
			Variance	Reliability	Variance	Reliability
K	Listening	0.15	6842.60	0.95	5974.65	0.95
	Reading	0.35	3794.23	0.94	5030.12	0.96
	Writing	0.35	4060.46	0.91	4695.06	0.93
	Speaking	0.15	10851.28	0.92	10050.34	0.90
	Overall Composite		3606.08	0.97	4096.73	0.98
1	Listening	0.15	1891.90	0.72	1860.19	0.74
	Reading	0.35	701.43	0.68	806.43	0.74
	Writing	0.35	1892.64	0.91	1938.19	0.91
	Speaking	0.15	5119.02	0.92	4524.46	0.90
	Overall Composite		1103.27	0.94	1139.60	0.94
2	Listening	0.15	2236.55	0.70	1969.13	0.71
	Reading	0.35	1090.45	0.81	1133.14	0.83
	Writing	0.35	2032.57	0.93	1948.75	0.93
	Speaking	0.15	5579.75	0.92	4613.85	0.90
	Overall Composite		1479.67	0.95	1333.56	0.95
3	Listening	0.15	1437.28	0.56	1428.86	0.60
	Reading	0.35	818.89	0.65	832.69	0.70
	Writing	0.35	1906.89	0.93	1739.59	0.93
	Speaking	0.15	5702.52	0.92	4784.98	0.91
	Overall Composite		1278.12	0.94	1138.43	0.94
4-5	Listening	0.15	2029.84	0.63	1776.15	0.65
	Reading	0.35	1131.16	0.77	1065.51	0.79
	Writing	0.35	2277.95	0.90	1791.40	0.90
	Speaking	0.15	7342.32	0.91	5360.20	0.90
	Overall Composite		1802.54	0.95	1394.69	0.95
6-8	Listening	0.15	2575.19	0.63	2198.07	0.63
	Reading	0.35	852.14	0.76	861.75	0.79
	Writing	0.35	2025.89	0.90	1781.74	0.90
	Speaking	0.15	7306.35	0.91	5553.03	0.90
	Overall Composite		1654.58	0.95	1364.05	0.95
9-12	Listening	0.15	2511.38	0.63	2269.71	0.65
	Reading	0.35	1026.24	0.80	953.05	0.80
	Writing	0.35	1949.40	0.90	1741.32	0.90
	Speaking	0.15	7165.76	0.92	5883.61	0.91
	Overall Composite		1703.74	0.95	1429.86	0.95

Table 5.5.4.4

Reliabilities of Composite Scale Scores: Over S601 Paper by IEP Status

Cluster	Component	Weight	Variance	Reliability
K	Listening	0.15	7135.62	0.96
	Reading	0.35	4656.00	0.96
	Writing	0.35	4090.10	0.92
	Speaking	0.15	8954.07	0.90
	Overall Composite		3670.58	0.98
1	Listening	0.15	1921.36	0.74
	Reading	0.35	616.23	0.62
	Writing	0.35	2293.10	0.92
	Speaking	0.15	4175.41	0.91
	Overall Composite		1008.68	0.93
2	Listening	0.15	1925.51	0.73
	Reading	0.35	888.72	0.77
	Writing	0.35	2101.10	0.93
	Speaking	0.15	3826.00	0.90
	Overall Composite		1145.62	0.95
3	Listening	0.15	1164.38	0.56
	Reading	0.35	679.03	0.57
	Writing	0.35	1747.73	0.94
	Speaking	0.15	3486.41	0.90
	Overall Composite		866.87	0.92
4-5	Listening	0.15	1241.09	0.61
	Reading	0.35	699.36	0.72
	Writing	0.35	1414.15	0.92
	Speaking	0.15	3497.40	0.90
	Overall Composite		817.17	0.93
6-8	Listening	0.15	1403.30	0.59
	Reading	0.35	524.06	0.69
	Writing	0.35	1377.82	0.91
	Speaking	0.15	3924.74	0.91
	Overall Composite		798.57	0.93
9-12	Listening	0.15	1669.15	0.60
	Reading	0.35	697.58	0.74
	Writing	0.35	1441.70	0.91
	Speaking	0.15	5035.82	0.91
	Overall Composite		1013.53	0.94

5.6 Conditional Standard Error of Measurement for Composites

CSEMs for the four ACCESS composite scale scores provide test users with a benchmark indicating how free a student's composite scale score is from measurement errors at different WIDA proficiency levels. Due to the differential weights applied to different ACCESS domains (see the introduction to Section 3 for weighting conventions), WIDA estimates the CSEMs using a procedure that is based on IRT (Lord, 1980) and developed by Price, Lurie, Raju, Wilkins, and Zhu (2006). Price et al. (2006) extended the work by Lord (1980) and Kolen, Hanson, and Brennan (1992) in estimating the CSEMs of students' composite scale scores consisting of components. The basic premise of this procedure is that one can estimate empirically the CSEM for a student's weighted composite scale score using the IRT-based CSEMs for each student's component scale scores and the weights associated with the components. We used this method to estimate the CSEMs for ACCESS composite scale scores by treating the ACCESS domains as components.

We used a three-step process to derive the CSEM for each ACCESS composite scale score. We calculated a unique CSEM for each composite scale score by grade. Since this procedure relies on empirical student data, which are subject to year-to-year fluctuations, we used all population student data from all previous three ACCESS 2.0 series in our calculations to obtain more stable estimates than using data from just a single series.

Step 1. Since we calibrated ACCESS domains separately, measurement errors associated with each of the ACCESS domains, as expressed in the CSEM, were independent of each other. Therefore, we estimated the CSEM for a student's composite scale score x, SEM_x, using the equation derived by Price et al. (2006):

$$SEM_x = \sqrt{W_1^2 SEM_1^2 + W_2^2 SEM_2^2 + W_3^2 SEM_3^2 + \cdots + W_k^2 SEM_k^2}$$

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,…,k.

Step 2. Due to the differential weights applied to different ACCESS domains, two students with the same sum of weighted domain score, or composite, may obtain different CSEMs; therefore, we took an additional step to obtain a unique value for each composite score. Specifically, we

estimated the expected value of the CSEM functions for a composite score using a regression approach, and we reported this expected value as the CSEM for that composite score.

Step 3. We applied a linear smoothing procedure to derive the CSEMs for composite scale scores that we did not observe in the data.

The figures in this section show graphically the CSEMs for various composite scale scores by grade level. The students' composite scale scores appear on the horizontal axis, and the corresponding CSEMs appear on the vertical axis. Each point in a figure represents a student in the dataset, showing the relationship between the CSEM and that student's composite scale score. We did not plot values for students who received the lowest possible scale scores for any ACCESS domains, as it is not possible to compute accurately the CSEM for these students' scale scores. For grade-level clusters with multiple grades, we use different colors in the figures to represent students in different grades.

The five vertical lines in the figure indicate the five ACCESS composite scale score cut points for the highest grade in the grade-level cluster for the test form, dividing the figure into six sections representing the six WIDA proficiency levels.

Smaller CSEM values indicate less measurement error (i.e., greater accuracy in measurement). In general, these figures show that the CSEMs are smaller and fairly constant in the middle of the composite scale score range but larger and more variable for extreme low and high composite scale scores. This is to be expected, since we used an IRT approach when scaling ACCESS, which typically produces larger CSEMs for scale scores that are at the lower and the higher ends of the scale score range. In addition, because students exit the EL program when they demonstrate that they are English language proficient, the number of students whose composite scale scores are at the extreme high end of the score range is typically small, as compared to the number of students whose composite scale scores are in the middle of the score range. Therefore, the measurement errors associated with the composite scale scores at the extreme high end of the score range tend to be larger since the calculation of these scale scores is based on the test performances of fewer students.

5.6.1 Oral

5.6.1.0 Kindergarten
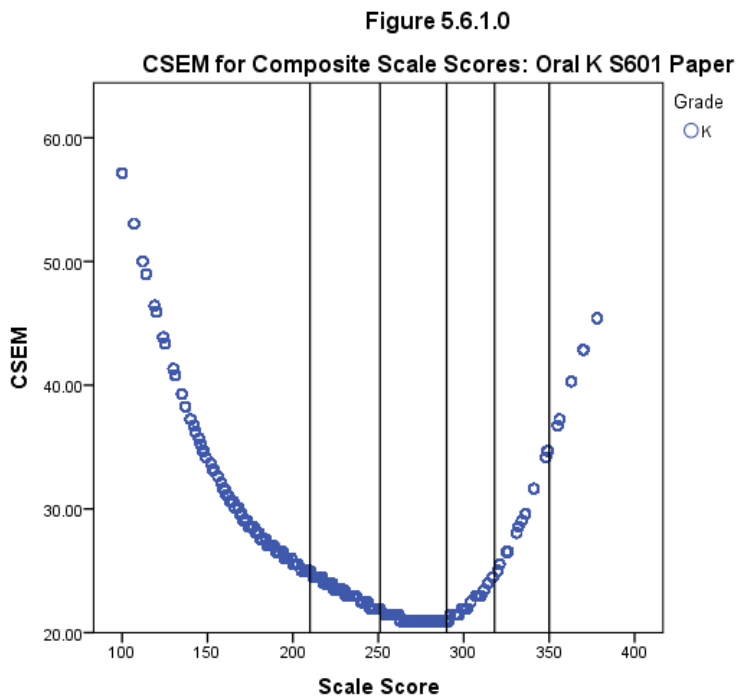
Figure 5.6.1.0
CSEM for Composite Scale Scores: Oral K S601 Paper

5.6.1.1 Grade 1

Figure 5.6.1.1

CSEM for Composite Scale Scores: Oral 1 S601 Paper

5.6.1.2 Grade 2

Figure 5.6.1.2

CSEM for Composite Scale Scores: Oral 2 S601 Paper

5.6.1.3 Grade 3

Figure 5.6.1.3

CSEM for Composite Scale Scores: Oral 3 S601 Paper

5.6.1.4 Grades 4-5

Figure 5.6.1.4

CSEM for Composite Scale Scores: Oral 4-5 S601 Paper

5.6.1.5 Grades 6-8

Figure 5.6.1.5

CSEM for Composite Scale Scores: Oral 6-8 S601 Paper

5.6.1.6 Grades 9-12

Figure 5.6.1.6

CSEM for Composite Scale Scores: Oral 9-12 S601 Paper
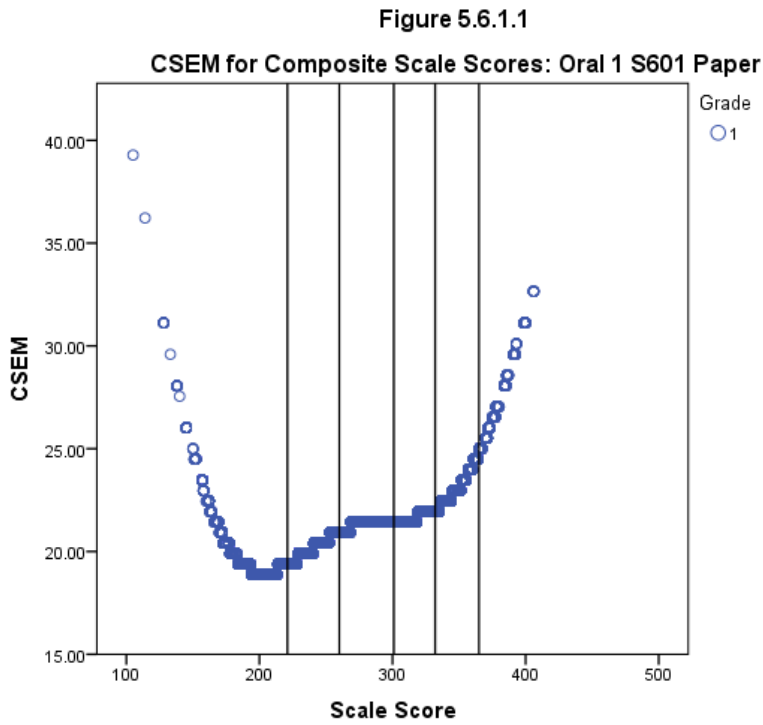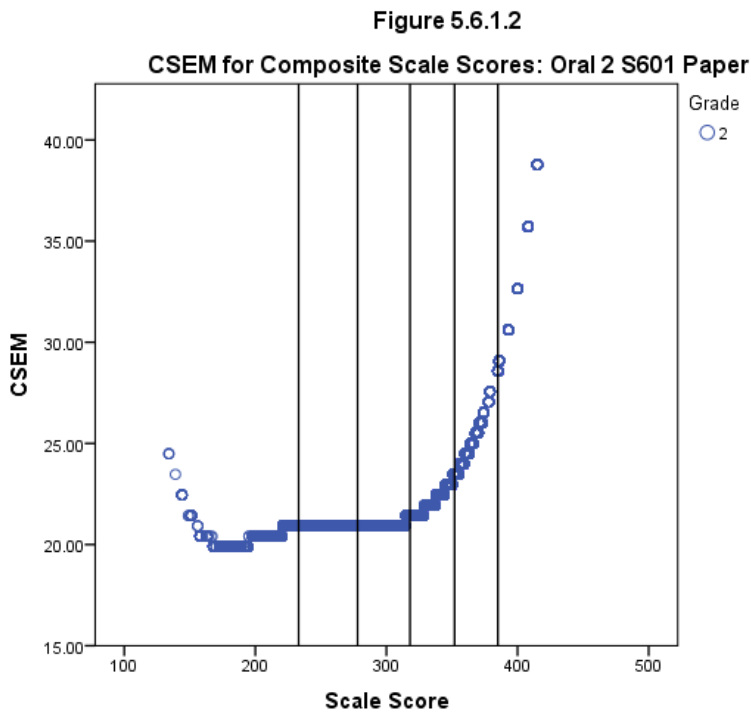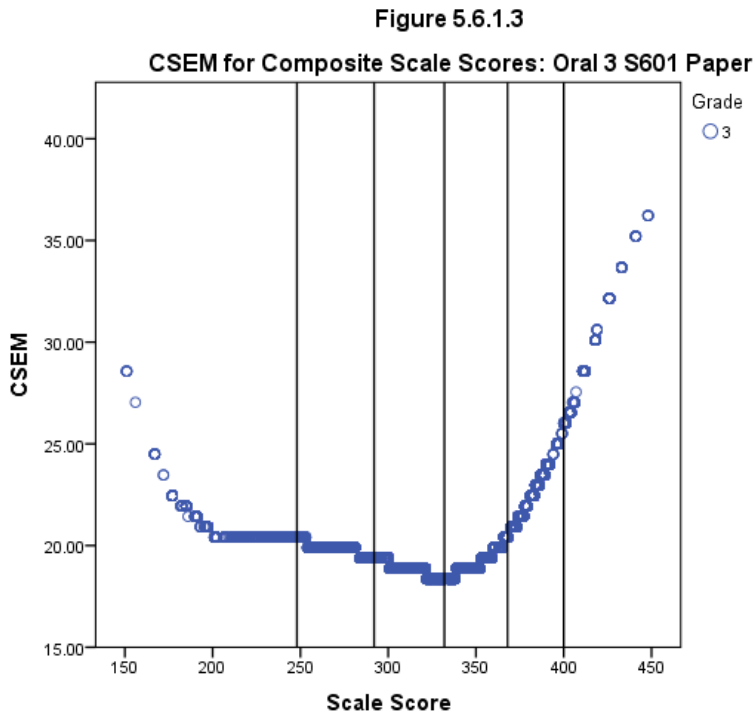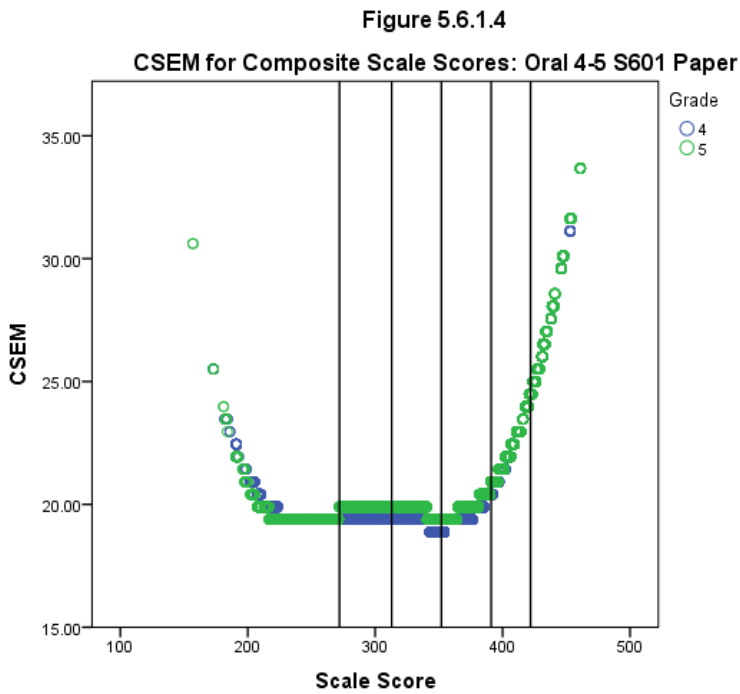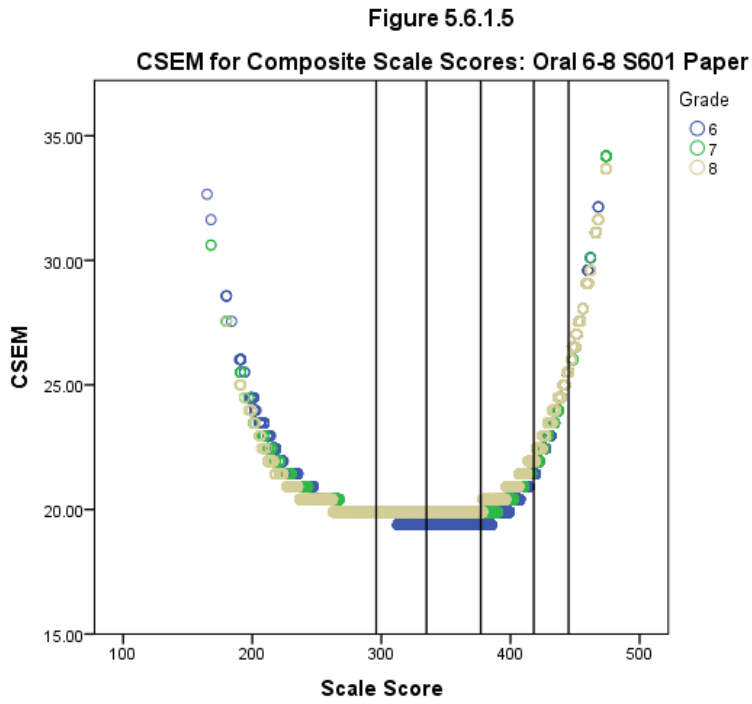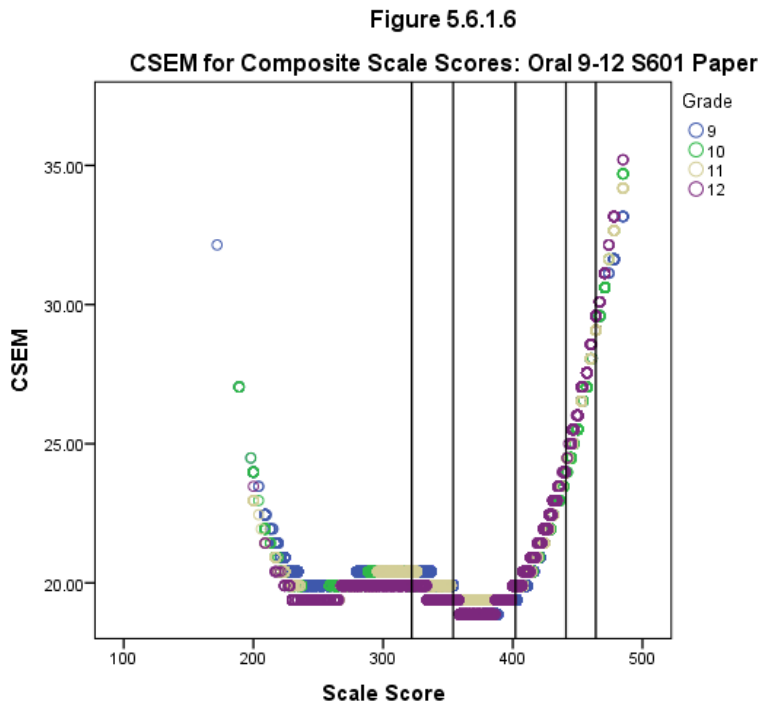
5.6.2 Literacy

5.6.2.0 Kindergarten
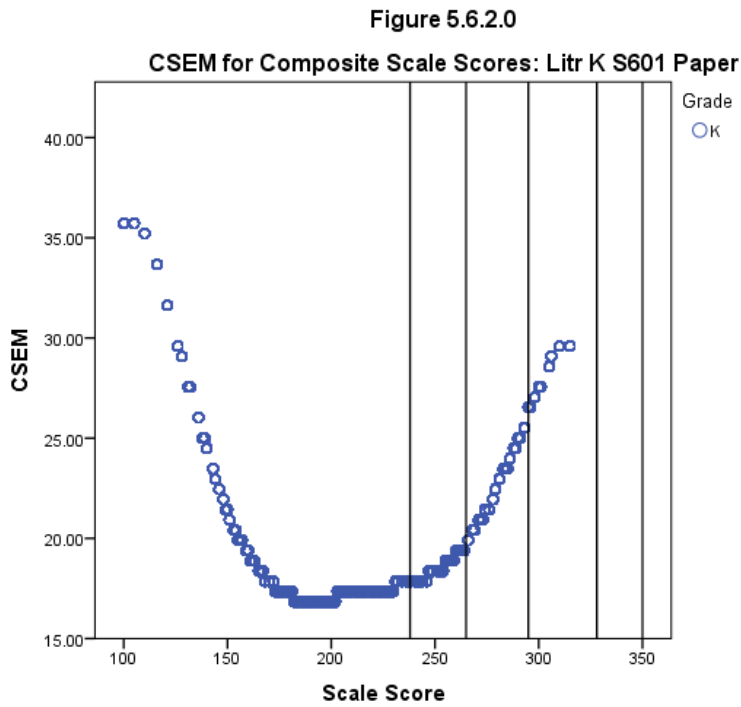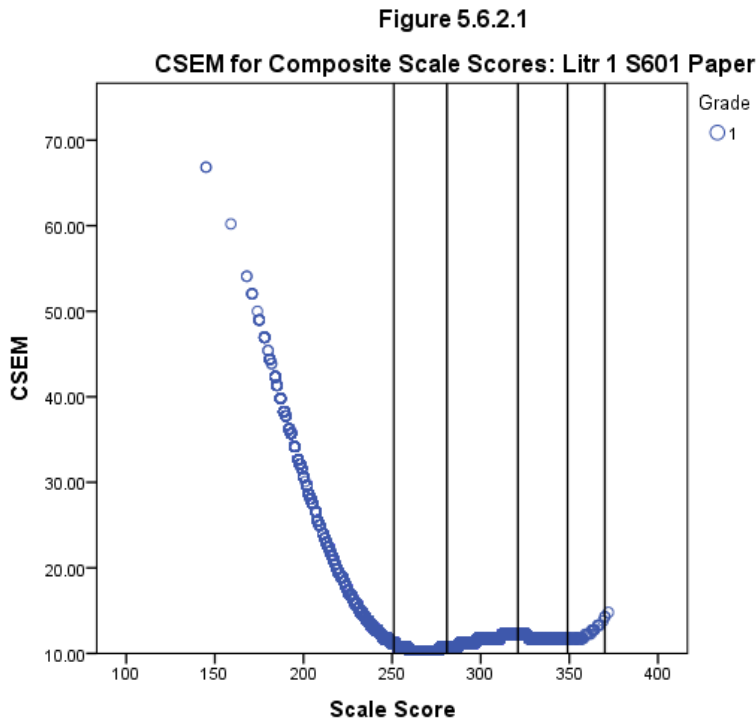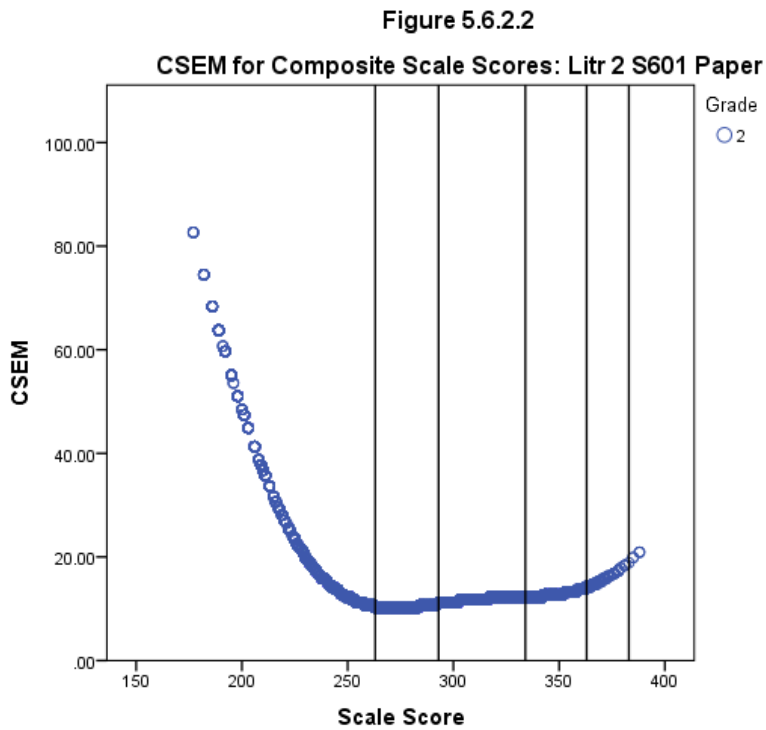
Figure 5.6.2.0
CSEM for Composite Scale Scores: Litr K S601 Paper

5.6.2.1 Grade 1

Figure 5.6.2.1
CSEM for Composite Scale Scores: Litr 1 S601 Paper

5.6.2.2 Grade 2

Figure 5.6.2.2
CSEM for Composite Scale Scores: Litr 2 S601 Paper

5.6.2.3 Grade 3

Figure 5.6.2.3

CSEM for Composite Scale Scores: Litr 3 S601 Paper

5.6.2.4 Grades 4-5

Figure 5.6.2.4

CSEM for Composite Scale Scores: Litr 4-5 S601 Paper

5.6.2.5 Grades 6-8

Figure 5.6.2.5

CSEM for Composite Scale Scores: Litr 6-8 S601 Paper

5.6.2.6 Grades 9-12

Figure 5.6.2.6

CSEM for Composite Scale Scores: Litr 9-12 S601 Paper
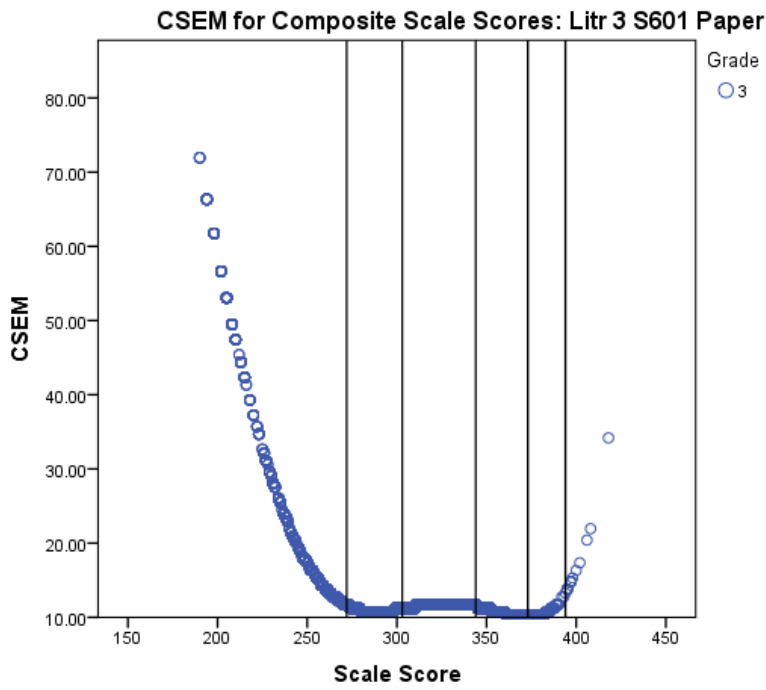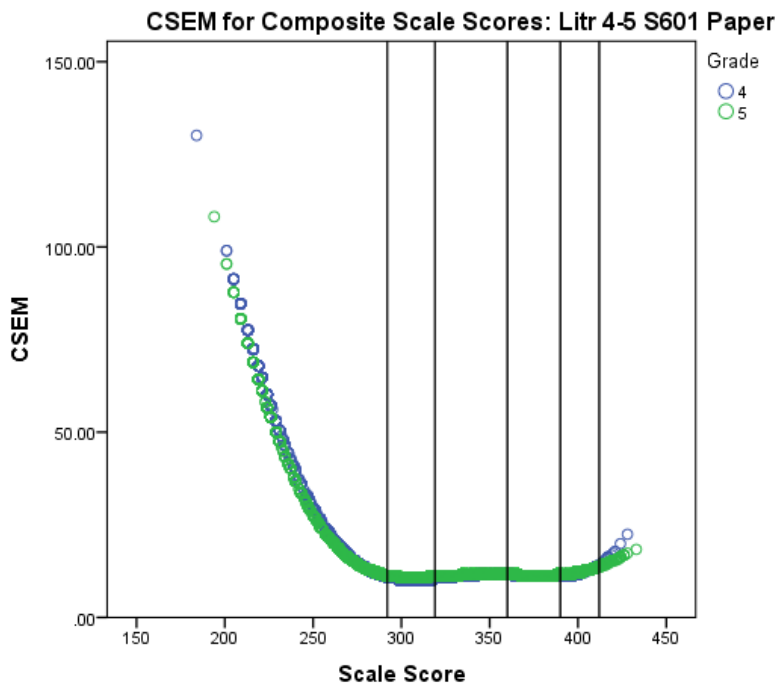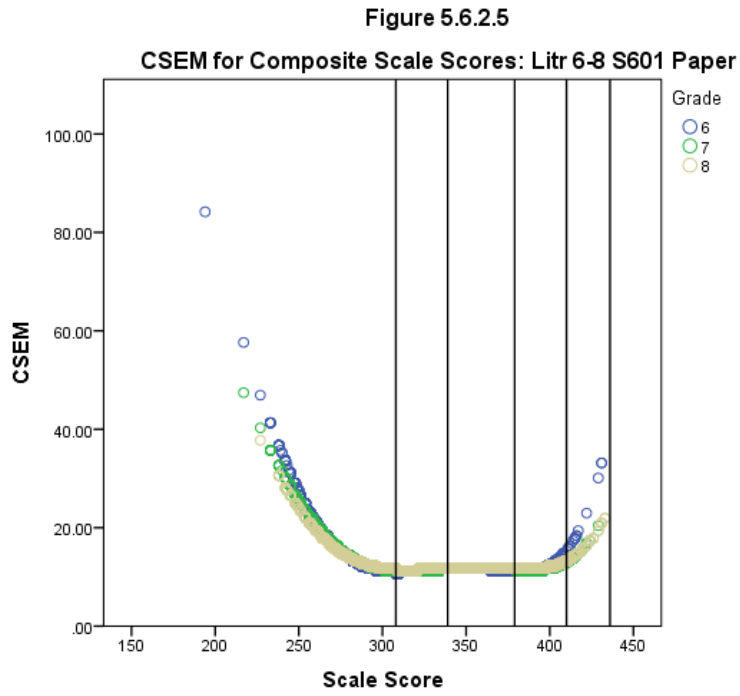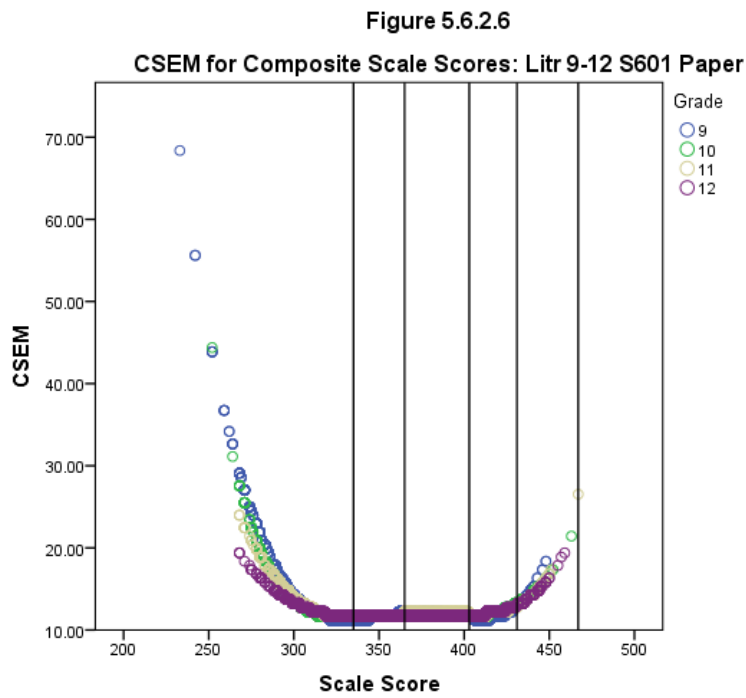
5.6.3 Comprehension

5.6.3.0 Kindergarten
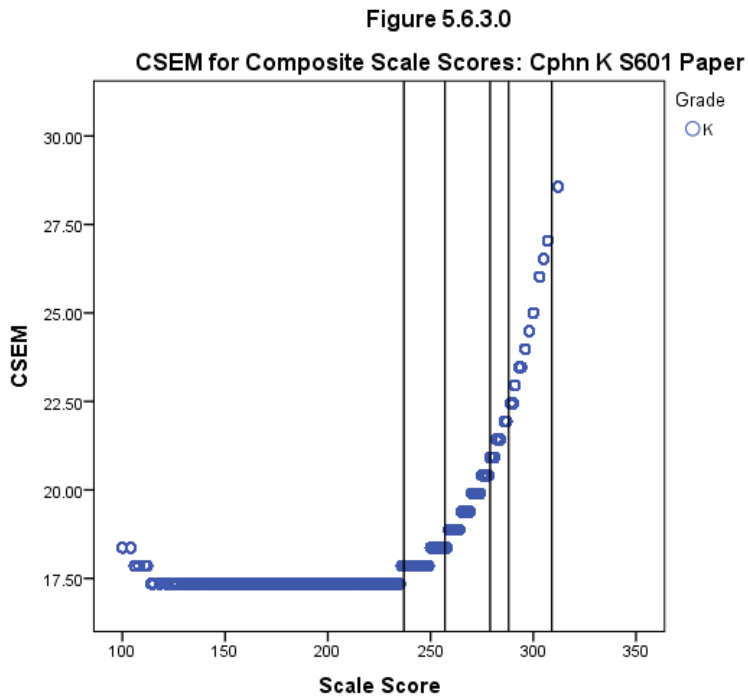
Figure 5.6.3.0

CSEM for Composite Scale Scores: Cphn K S601 Paper

5.6.3.1 Grade 1

Figure 5.6.3.1

CSEM for Composite Scale Scores: Cphn 1 S601 Paper

5.6.3.2 Grade 2

Figure 5.6.3.2

CSEM for Composite Scale Scores: Cphn 2 S601 Paper

5.6.3.3 Grade 3

Figure 5.6.3.3

CSEM for Composite Scale Scores: Cphn 3 S601 Paper

5.6.3.4 Grades 4-5

Figure 5.6.3.4

CSEM for Composite Scale Scores: Cphn 4-5 S601 Paper

5.6.3.5 Grades 6-8

Figure 5.6.3.5

CSEM for Composite Scale Scores: Cphn 6-8 S601 Paper

5.6.3.6 Grades 9-12

Figure 5.6.3.6

CSEM for Composite Scale Scores: Cphn 9-12 S601 Paper
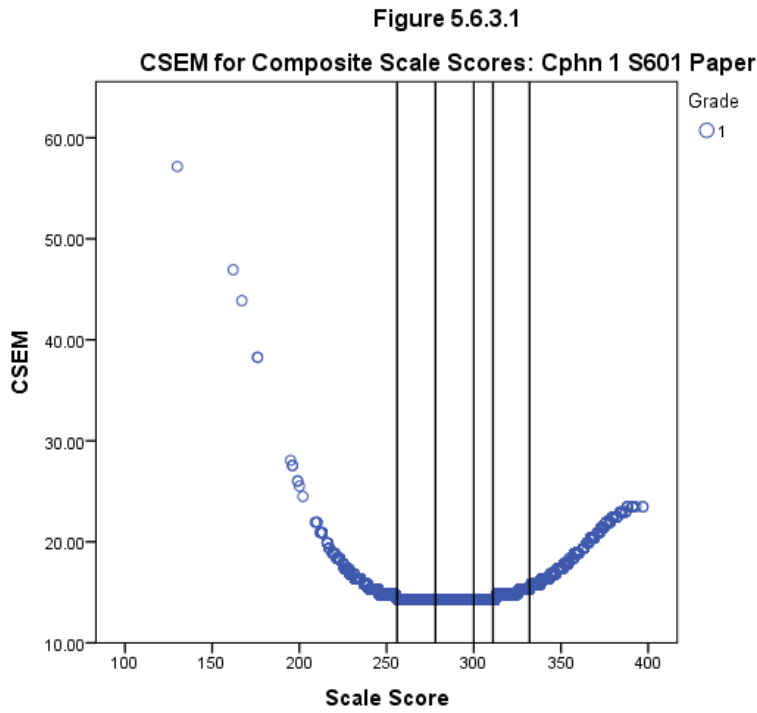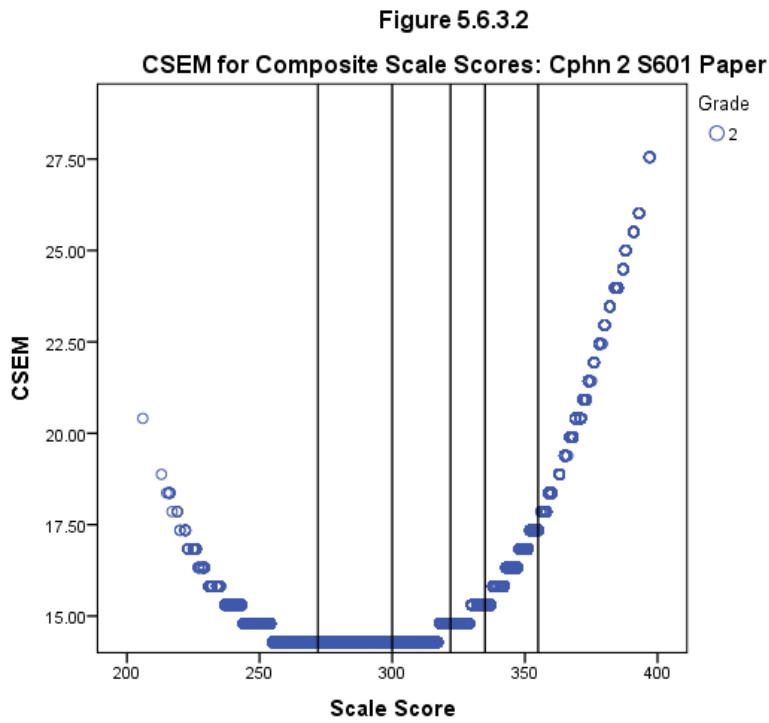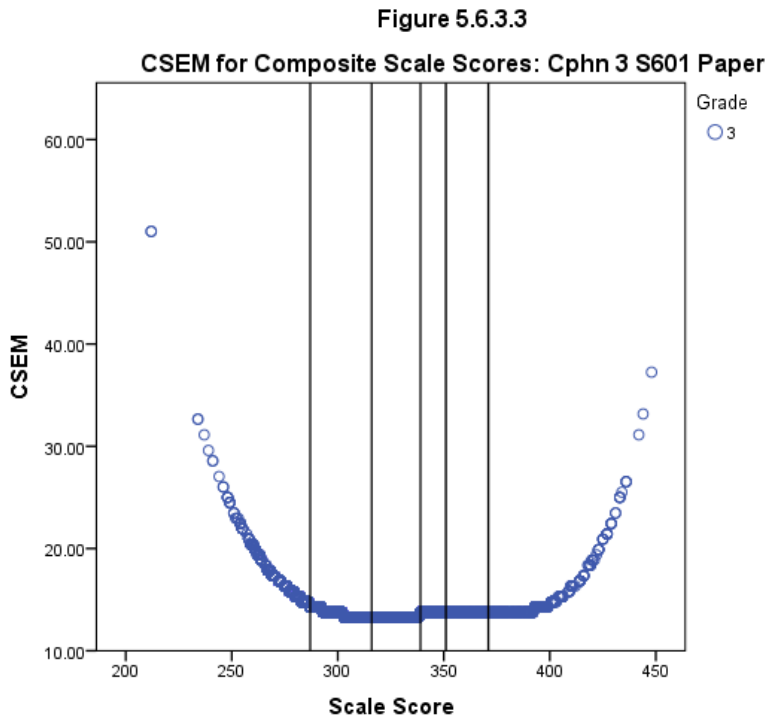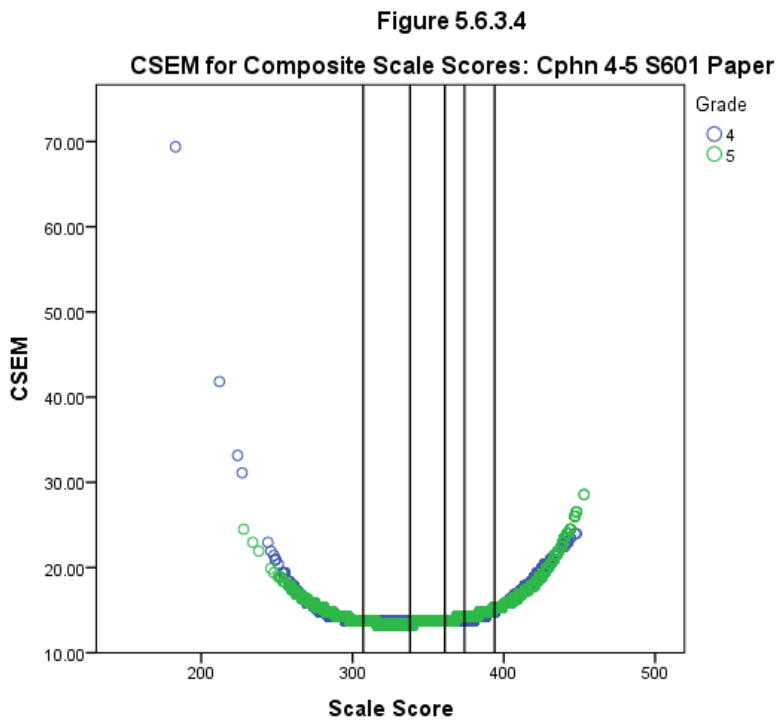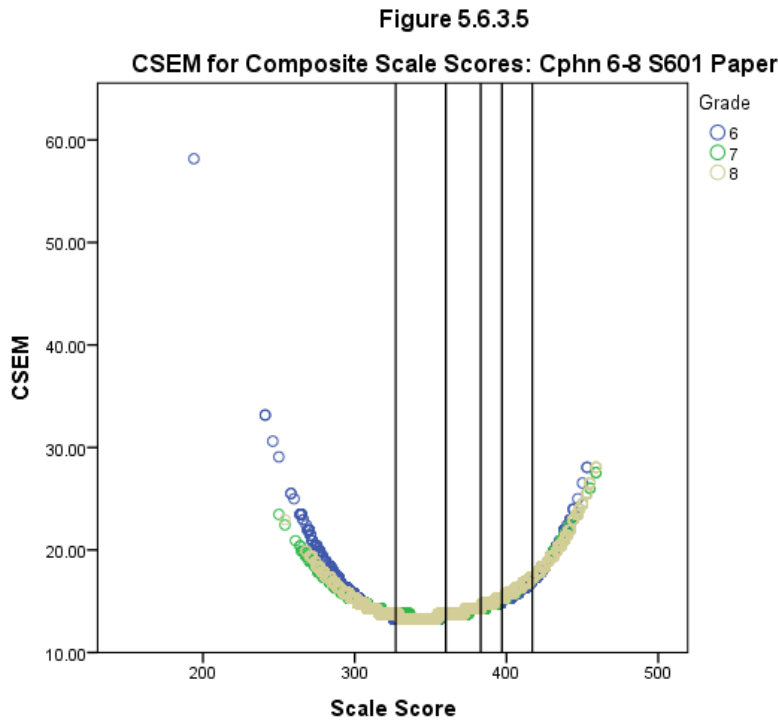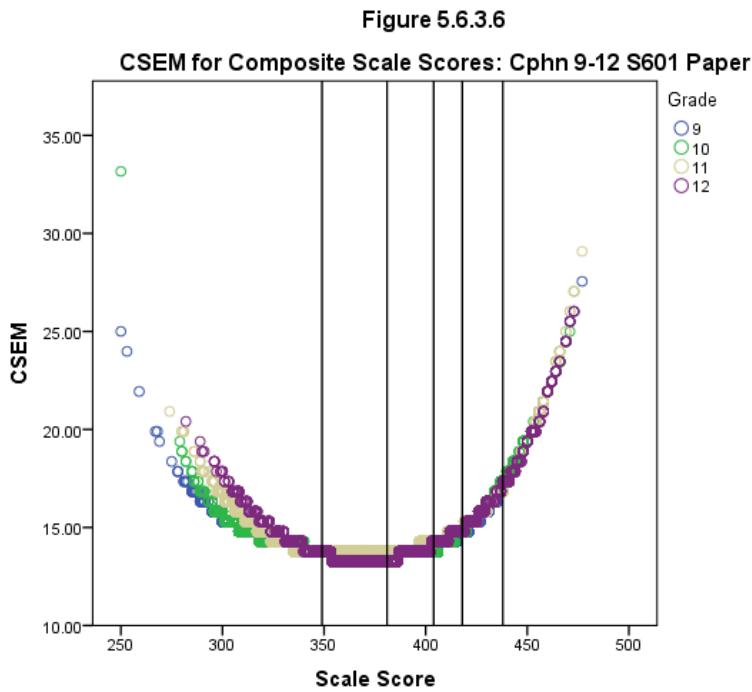
5.6.4 Overall

5.6.4.0 Kindergarten
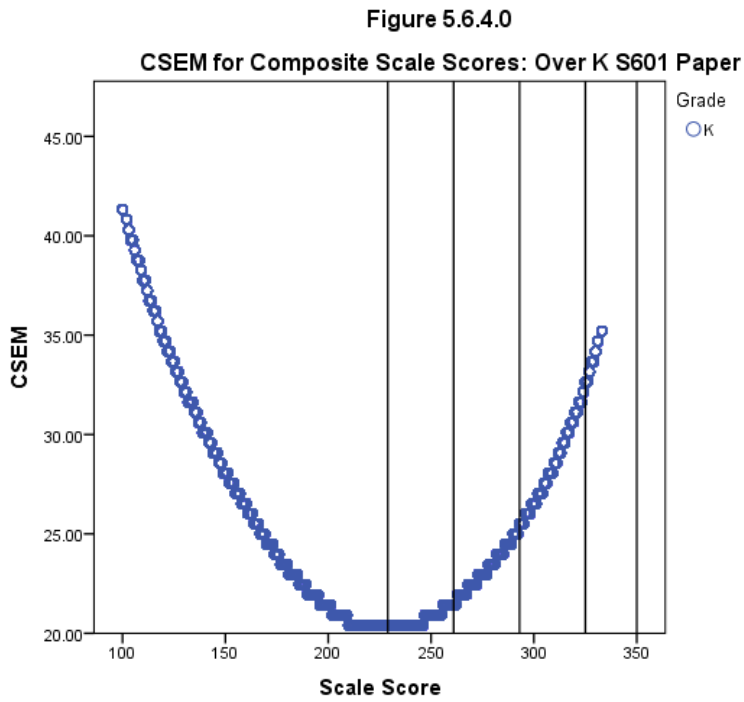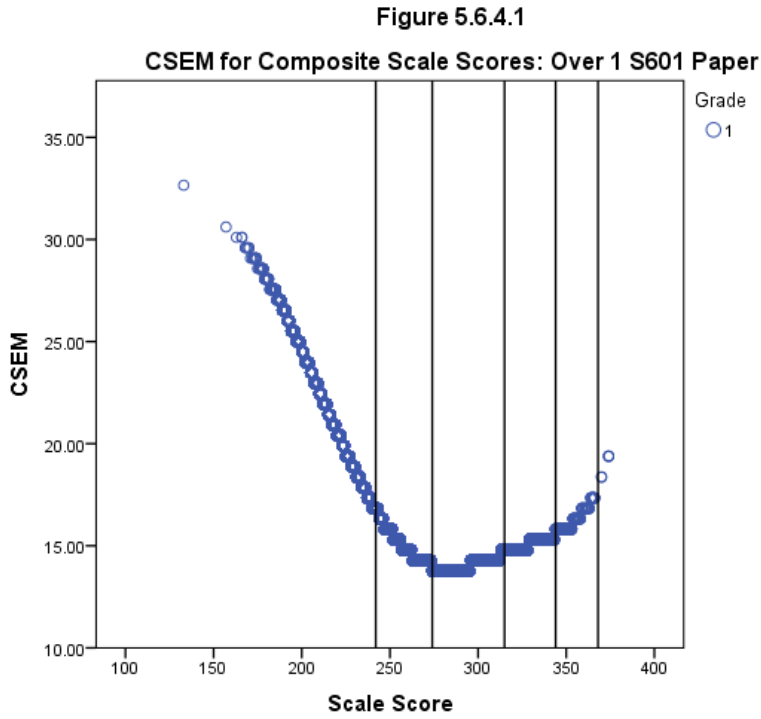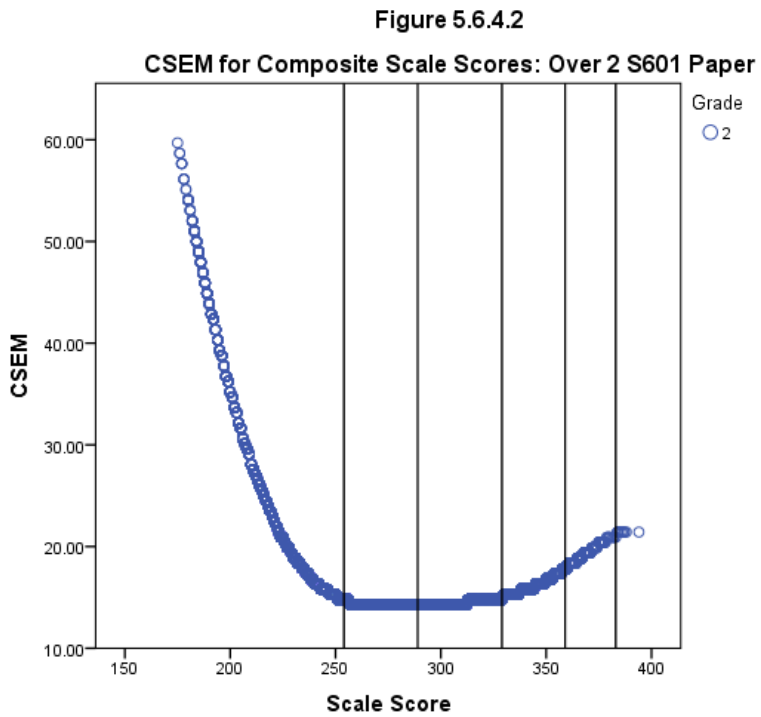
Figure 5.6.4.0

CSEM for Composite Scale Scores: Over K S601 Paper

5.6.4.1 Grade 1

Figure 5.6.4.1

CSEM for Composite Scale Scores: Over 1 S601 Paper

5.6.4.2 Grade 2

Figure 5.6.4.2

CSEM for Composite Scale Scores: Over 2 S601 Paper

5.6.4.3 Grade 3

Figure 5.6.4.3

CSEM for Composite Scale Scores: Over 3 S601 Paper

5.6.4.4 Grades 4-5

Figure 5.6.4.4

CSEM for Composite Scale Scores: Over 4-5 S601 Paper

5.6.4.5 Grades 6-8

Figure 5.6.4.5

CSEM for Composite Scale Scores: Over 6-8 S601 Paper

5.6.4.6 Grades 9-12

Figure 5.6.4.6

CSEM for Composite Scale Scores: Over 9-12 S601 Paper
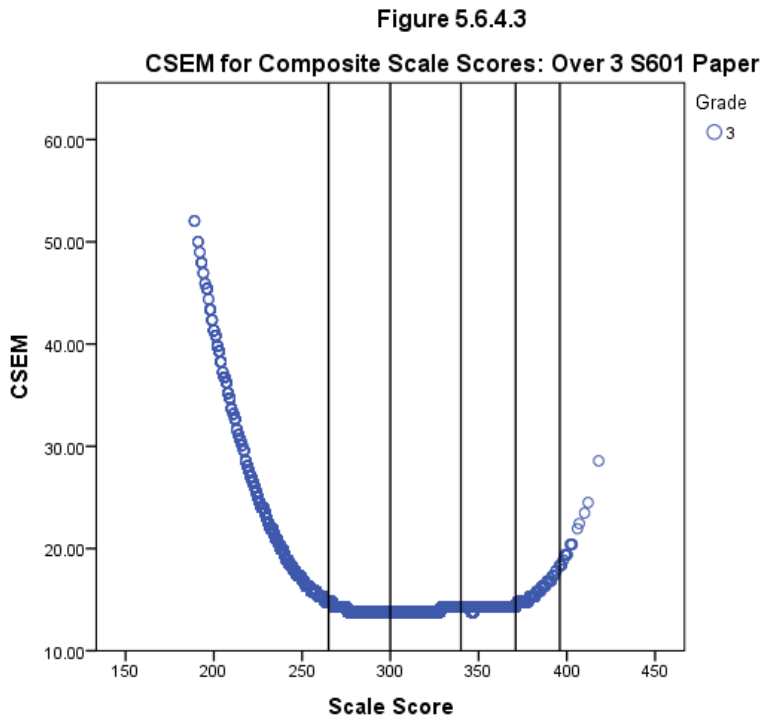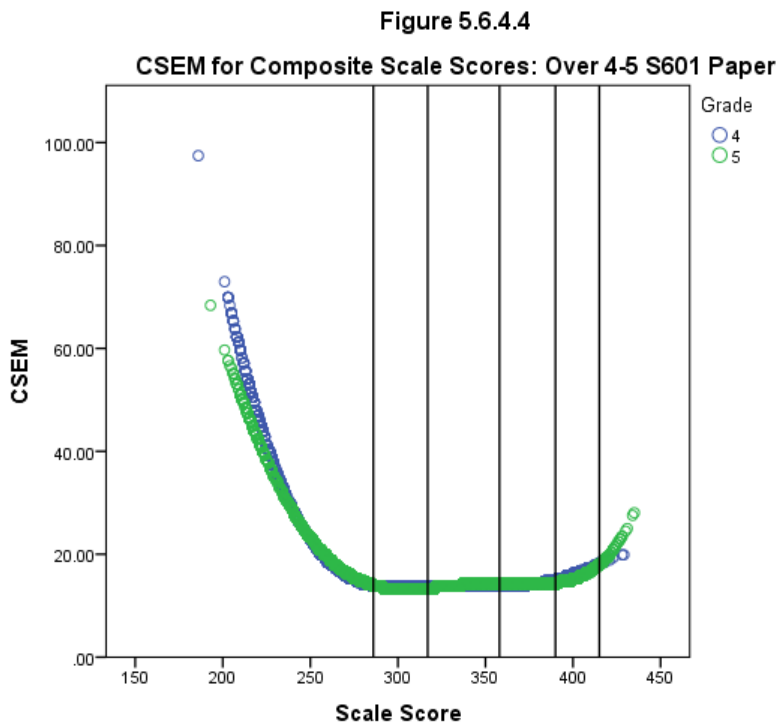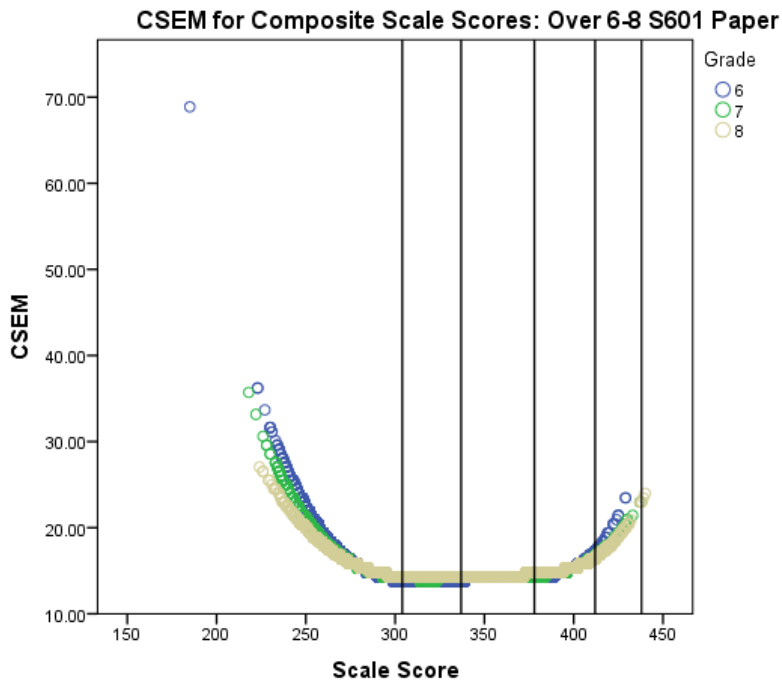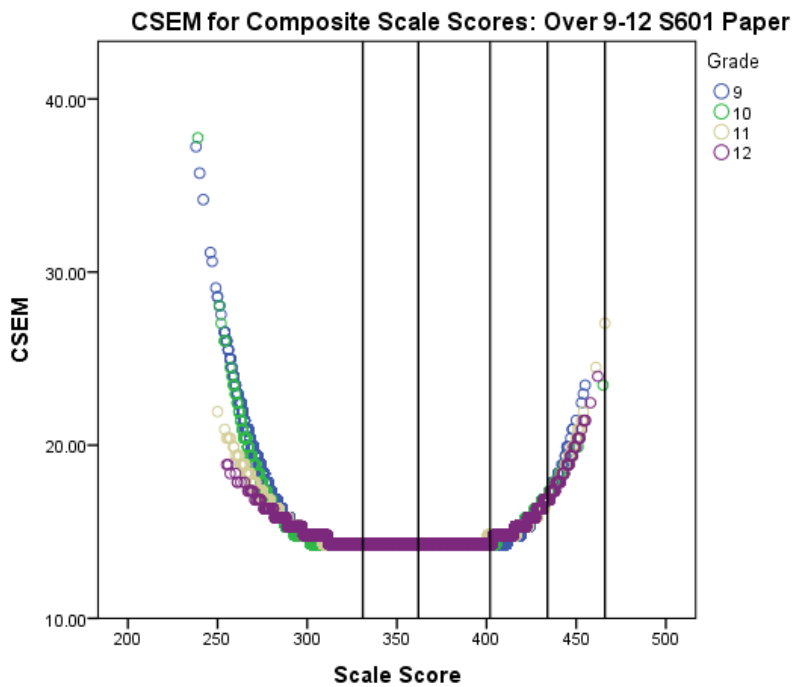
5.7   Accuracy and Consistency of Composites

One of the main purposes of the WIDA ACCESS program is to identify the English language proficiency level of students with respect to the WIDA ELD Standards. Because of the emphasis on the classification of student performance, a question of interest is how accurately and consistently the ACCESS composite scale scores can classify students into WIDA proficiency categories determined by the 2016 ACCESS standard-setting process (Cook & MacGregor, 2017). Although states in the WIDA Consortium take into consideration one or more of the domain and composite scale scores when making accountability decisions, all WIDA Consortium states use the Overall composite scale score as the primary score when making classification decisions about students. Therefore, it is especially important to examine the accuracy and consistency of the classifications based on the Overall composite scale scores to help test users and policy makers judge the utility of this information and to make decisions about score reporting (American Educational Research Association et al., 2014). The analyses utilize the methods that Livingston and Lewis (1995) and Young and Yoon (1998) outlined, as implemented in the software program BB-CLASS (Brennan, 2004; cf. also Lee et al., 2002).

The method and descriptions of the classification accuracy and consistency indices reported in this section appear in detail in Section 5.4. The only substantive methodological difference between the estimation of the classification accuracy and consistency of the domain scale scores versus the composite scale scores is that to estimate the classification accuracy and consistency of the composite scale scores, we first estimate the reliability of the composite scale scores using a stratified Cronbach's coefficient alpha, as described in Section 5.4.

For each composite, we present three tables. The first reports the overall accuracy and the overall consistency indices for each grade. The second reports the marginal classification accuracy indices based on the composite scale scores at the cut points for each grade. The third reports the marginal classification consistency indices based on the composite scale scores at the cut points for each grade.

If we could not estimate the overall and marginal classification accuracy and consistency indices because there were fewer than 200 students in the proficiency level, we collapsed the affected proficiency level with the level below it and placed 'N/A' in the table for the affected proficiency level.

As noted in Section 5.4, assessment experts have issued 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 for each composite the range of these indices, highlighting the grade with the lowest classification accuracy and consistency indices for that composite. Since overall accuracy and consistency indices are summaries of the degree of classification accuracy and consistency for the composite scale scores across all proficiency level cut points, we also examine the marginal classification accuracy and consistency indices for these grades to identify the specific source(s) of low classification accuracy and consistency.

For the Oral composite, as shown in Table 5.7.1.1, overall classification accuracy ranged from 0.617 to 0.740 and overall classification consistency ranged from 0.512 to 0.662 across grades. The lowest overall classification accuracy and consistency values were found for students in Grade 4.

For the Literacy composite, overall classification accuracy ranged from 0.750 to 0.887 and overall classification consistency ranged from 0.659 to 0.858 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 0.504 to 0.858 and overall classification consistency ranged from 0.395 to 0.817 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 0.784 to 0.880 and overall classification consistency ranged from 0.705 to 0.837 across grades. The lowest overall classification accuracy and consistency values were found for students in Grade 5.

The results reveal that Grade 5 had the lowest overall classification accuracy and consistency indices for two out of the four composites (Literacy and Overall), while Grade 4 had the lowest overall classification accuracy and consistency indices for the Oral composite and Grade 3 had the lowest overall classification accuracy and consistency indices for the Comprehension composite.

From an accountability perspective, the most important indices for test users and policy makers to examine are the marginal classification accuracy and consistency indices. We report for each composite the range of the marginal classification accuracy and consistency indices for the composite scale scores across grades and then highlight the grade (and the cut point within that grade) that had the lowest marginal classification accuracy and the lowest consistency indices.

For the Oral composite, classification accuracy indices at the cut ranged from 0.863 to 0.980 (Table 5.7.1.2) and classification consistency at the cut ranged from 0.809 to 0.972 (Table 5.7.1.3). The lowest marginal classification accuracy and consistency values were found for students in Grade 4 at the PL 4/PL 5 cut. Note that Grade 4 was also identified as having the lowest overall classification accuracy and consistency in the Overall Oral composite. However, it should be noted that the marginal classification accuracy and consistency for the Grade 4 Oral composite are still in the .80's and 90's.

For the Literacy composite, classification accuracy indices at the cut ranged from 0.885 to 0.981 (Table 5.7.2.2) and classification consistency at the cut ranged from 0.838 to 0.975 (Table 5.7.2.3). Grade 4, at the PL 3/4 cut point, had the lowest marginal classification accuracy and consistency indices. However, it should be noted that the marginal classification accuracy and consistency for the Grade 4 Literacy composite are still in the .80's and 90's.

For the Comprehension composite, classification accuracy indices at the cut ranged from 0.824 to 0.986 (Table 5.7.3.2) and classification consistency at the cut ranged from 0.765 to 0.981 (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 Overall Comprehension composite. The low marginal classification accuracy and consistency at the PL 4/PL 5 cut appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal classification accuracy and consistency for the Grade 3 Comprehension composite are still in the high .70's and 90's.

For the Overall composite, classification accuracy indices at the cut ranged from 0.896 to 0.994 (Table 5.7.4.2) and classification consistency at the cut ranged from 0.854 to 0.993 (Table 5.7.4.3). The lowest marginal classification accuracy and consistency values were found for students in Grade 3 at the PL 3/PL 4 cut. Note that Grade 3 was also identified as having the lowest marginal classification accuracy and consistency in the Comprehension composite. The

low marginal classification accuracy and consistency at the PL 3/PL 4 cut appeared to have contributed to its low overall classification accuracy and consistency. However, it should be noted that the marginal classification accuracy and consistency for the Grade 3 Overall composite are still in the .80's and 90's .

When we compared the overall and marginal classification accuracy and consistency indices for the composites for a particular grade, we saw that in many instances they told the same story (i.e., for a given grade, if the overall classification accuracy and consistency indices were low, then the marginal classification accuracy and consistency indices also tended to be low). This was especially true for Grade 5 for two of the four composites (Literacy and Overall). Grade 3 had the lowest overall and marginal classification accuracy and consistency indices for the Comprehension composites.

Grade 4 had the lowest marginal classification accuracy and consistency in two of the four composites (Oral and Literacy). Grade 3 had the lowest marginal classification accuracy and consistency in the Comprehension composite. Grade 3 also had the lowest marginal classification accuracy and consistency in the Overall composite.

In addition, the lowest marginal classification accuracy and consistency of the composites occurred at the PL 3/PL 4 and PL 4/PL 5 cut points. This finding is consistent with previous research (Lee et al., 2002), in that classification accuracy and consistency at cut points in the middle of the proficiency level range are lower than those at the lower and upper ends.

A higher number of proficiency levels typically results in cut scores that are closer to each other than if a smaller number of proficiency levels is used. Classification accuracy and consistency are expected to vary for different ability levels due to variation in measurement accuracy. The further away the scores are from the cut scores, the smaller the classification errors would be or the more accurate the classification decisions would be. When there is a large number of proficiency levels, more students are near the cut scores than there would be if there were fewer proficiency levels. Therefore, the higher the number of proficiency levels, the higher the probability that students are misclassified (Ercikan & Julian, 2002). Since ACCESS has six proficiency levels and PL 3 and PL 4 occupy relatively narrow ranges on the ability scale compared with other proficiency levels, the classification accuracy and consistency for the 3/4 and 4/5 cuts are lower than for other cuts.

Assessment experts have issued little guidance to aid in making judgments about the ideal or expected levels of decision consistency and accuracy needed for educational assessments that report composite scale scores. From an accountability perspective, the most important indices are the marginal classification accuracy and consistency indices. The marginal classification accuracy and consistency indices were at or above 0.800 for all four composites. Additionally, the marginal classification accuracy and consistency indices were at or above 0.838 for the Overall composite scale score, which is the primary score that WIDA Consortium states use when making accountability decisions.

5.7.1 Oral

Table 5.7.1.1

Overall Accuracy and Consistency of Classification Indices: Oral S601 Paper

Grade	Accuracy	Consistency
K	0.740	0.662
1	0.681	0.572
2	0.655	0.548
3	0.653	0.545
4	0.617	0.512
5	0.630	0.522
6	0.632	0.524
7	0.635	0.529
8	0.640	0.533
9	0.668	0.565
10	0.676	0.575
11	0.675	0.570
12	0.710	0.606

Table 5.7.1.2

Marginal Classification Accuracy Indices Based on the Composite Scale Scores at the Cut Points: Oral S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.945	0.944	0.950	0.946	0.945
1	0.956	0.922	0.900	0.927	0.973
2	0.964	0.930	0.888	0.902	0.966
3	0.967	0.935	0.884	0.888	0.972
4	0.966	0.951	0.908	0.863	0.921
5	0.958	0.944	0.905	0.874	0.942
6	0.950	0.927	0.896	0.902	0.949
7	0.938	0.922	0.900	0.907	0.957
8	0.939	0.923	0.902	0.904	0.961
9	0.932	0.921	0.908	0.921	0.975
10	0.930	0.915	0.908	0.936	0.977
11	0.934	0.913	0.901	0.939	0.980
12	0.939	0.915	0.903	0.949	N/A

Table 5.7.1.3

Marginal Classification Consistency Indices Based on the Composite Scale Scores at the Cut Points: Oral S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.923	0.921	0.928	0.924	0.923
1	0.938	0.889	0.861	0.895	0.966
2	0.949	0.899	0.845	0.858	0.963
3	0.955	0.906	0.839	0.844	0.967
4	0.953	0.928	0.872	0.809	0.897
5	0.941	0.917	0.868	0.824	0.919
6	0.930	0.895	0.855	0.861	0.935
7	0.913	0.889	0.861	0.868	0.942
8	0.914	0.891	0.864	0.865	0.943
9	0.904	0.888	0.872	0.890	0.963
10	0.900	0.880	0.872	0.909	0.969
11	0.907	0.877	0.862	0.912	0.972
12	0.914	0.880	0.864	0.927	N/A

5.7.2 Literacy

Table 5.7.2.1

Overall Accuracy and Consistency of Classification Indices: Litr S601 Paper

Grade	Accuracy	Consistency
K	0.887	0.858
1	0.796	0.716
2	0.811	0.735
3	0.779	0.695
4	0.766	0.680
5	0.750	0.659
6	0.808	0.730
7	0.798	0.717
8	0.790	0.707
9	0.775	0.688
10	0.775	0.687
11	0.776	0.689
12	0.802	0.722

Table 5.7.2.2

Marginal Classification Accuracy Indices Based on the Composite Scale Scores at the Cut Points: Litr S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.956	0.963	0.967	N/A	N/A
1	0.905	0.910	0.981	N/A	N/A
2	0.941	0.921	0.950	N/A	N/A
3	0.959	0.924	0.896	N/A	N/A
4	0.965	0.941	0.885	0.976	N/A
5	0.959	0.935	0.900	0.957	N/A
6	0.946	0.914	0.948	N/A	N/A
7	0.939	0.914	0.946	N/A	N/A
8	0.937	0.917	0.937	N/A	N/A
9	0.946	0.921	0.930	0.979	N/A
10	0.948	0.921	0.929	0.977	N/A
11	0.955	0.922	0.922	0.978	N/A
12	0.951	0.915	0.937	N/A	N/A

Table 5.7.2.3

Marginal Classification Consistency Indices Based on the Composite Scale Scores at the Cut Points: Litr S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.938	0.945	0.966	N/A	N/A
1	0.866	0.873	0.975	N/A	N/A
2	0.916	0.888	0.930	N/A	N/A
3	0.943	0.891	0.860	N/A	N/A
4	0.950	0.914	0.838	0.970	N/A
5	0.942	0.906	0.859	0.944	N/A
6	0.924	0.879	0.925	N/A	N/A
7	0.915	0.879	0.922	N/A	N/A
8	0.911	0.883	0.910	N/A	N/A
9	0.925	0.889	0.901	0.971	N/A
10	0.927	0.889	0.900	0.970	N/A
11	0.937	0.889	0.889	0.971	N/A
12	0.930	0.880	0.910	N/A	N/A

5.7.3 Comprehension

Table 5.7.3.1

Overall Accuracy and Consistency of Classification Indices: Cphn S601 Paper

Grade	Accuracy	Consistency
K	0.858	0.817
1	0.553	0.441
2	0.588	0.479
3	0.504	0.395
4	0.557	0.447
5	0.546	0.442
6	0.612	0.501
7	0.591	0.483
8	0.577	0.470
9	0.590	0.482
10	0.598	0.489
11	0.592	0.484
12	0.614	0.502

Table 5.7.3.2

Marginal Classification Accuracy Indices Based on the Composite Scale Scores at the Cut Points: Cphn S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.958	0.965	0.968	0.970	0.986
1	0.908	0.857	0.876	0.911	0.963
2	0.938	0.889	0.882	0.899	0.946
3	0.962	0.907	0.834	0.824	0.917
4	0.954	0.923	0.866	0.858	0.916
5	0.943	0.910	0.862	0.863	0.917
6	0.933	0.876	0.880	0.927	0.976
7	0.918	0.877	0.881	0.917	0.967
8	0.914	0.882	0.879	0.904	0.961
9	0.924	0.888	0.882	0.907	0.958
10	0.922	0.884	0.887	0.918	0.961
11	0.931	0.881	0.882	0.908	0.956
12	0.933	0.876	0.889	0.926	0.970

Table 5.7.3.3

Marginal Classification Consistency Indices Based on the Composite Scale Scores at the Cut Points: Cphn S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.940	0.949	0.955	0.958	0.981
1	0.874	0.800	0.828	0.872	0.946
2	0.913	0.844	0.835	0.859	0.922
3	0.949	0.863	0.772	0.765	0.875
4	0.939	0.885	0.816	0.808	0.877
5	0.924	0.867	0.813	0.814	0.880
6	0.909	0.825	0.836	0.893	0.966
7	0.886	0.827	0.837	0.882	0.950
8	0.880	0.834	0.834	0.865	0.940
9	0.895	0.841	0.839	0.870	0.937
10	0.891	0.836	0.846	0.883	0.943
11	0.904	0.833	0.836	0.870	0.937
12	0.904	0.826	0.846	0.893	0.958

5.7.4 Overall

Table 5.7.4.1

Overall Accuracy and Consistency of Classification Indices: Over S601 Paper

Grade	Accuracy	Consistency
K	0.880	0.837
1	0.828	0.759
2	0.826	0.759
3	0.804	0.731
4	0.804	0.729
5	0.784	0.705
6	0.839	0.772
7	0.819	0.748
8	0.816	0.744
9	0.808	0.734
10	0.812	0.740
11	0.815	0.744
12	0.841	0.776

Table 5.7.4.2

Marginal Classification Accuracy Indices Based on the Composite Scale Scores at the Cut Points: Over S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.957	0.963	0.972	0.988	N/A
1	0.946	0.921	0.968	0.994	N/A
2	0.966	0.940	0.936	0.985	N/A
3	0.975	0.948	0.896	0.987	N/A
4	0.976	0.962	0.920	0.947	N/A
5	0.971	0.957	0.924	0.933	N/A
6	0.964	0.944	0.932	N/A	N/A
7	0.958	0.942	0.933	0.986	N/A
8	0.958	0.943	0.930	0.985	N/A
9	0.958	0.945	0.936	0.969	N/A
10	0.958	0.943	0.939	0.973	N/A
11	0.963	0.943	0.933	0.976	N/A
12	0.964	0.939	0.939	N/A	N/A

Table 5.7.4.3

Marginal Classification Consistency Indices Based on the Composite Scale Scores at the Cut Points: Over S601 Paper

Grade	PL 1/2	PL 2/3	PL 3/4	PL 4/5	PL 5/6
K	0.939	0.948	0.960	0.988	N/A
1	0.924	0.888	0.954	0.993	N/A
2	0.952	0.915	0.909	0.984	N/A
3	0.965	0.925	0.854	0.986	N/A
4	0.967	0.944	0.888	0.930	N/A
5	0.960	0.938	0.893	0.913	N/A
6	0.949	0.920	0.903	N/A	N/A
7	0.941	0.918	0.905	0.983	N/A
8	0.941	0.919	0.901	0.982	N/A
9	0.942	0.922	0.910	0.960	N/A
10	0.941	0.919	0.913	0.967	N/A
11	0.949	0.919	0.906	0.970	N/A
12	0.949	0.913	0.914	N/A	N/A

6.   Quality Control

6.1   Content Development Quality Control

The Center for Applied Linguistics (CAL) utilizes educators and other consultants at a number of phases throughout the test-development cycle. These educators and consultants are recruited, vetted, and trained by CAL and/or WIDA and make crucial contributions to these phases of the test development cycle. The phases of development in which educators or consultants are involved, as well as the procedures and criteria for recruitment and training, are described below.

Theme Generation

During theme generation, CAL and WIDA recruit educators to generate raw ideas to be used in new item development. Educators with ESL or content-area expertise and two or more years of teaching experience in a WIDA state (in the grade-level cluster for which they will generate themes) are invited to participate. Recruitment also focuses on a geographical distribution of educators from across the consortium. Upon selection, educators participate in a short training that introduces the theme-generation process, along with how to understand the item specifications that they use to generate themes.

Item Writing

CAL recruits professional item writers to generate raw item/task content based on the ideas from theme generation. To recruit item writers, CAL has a standing announcement on its website asking prospective item writers to submit their resume and fill out a survey describing their past item-writing experience. CAL selects individuals with significant experience in writing items, both in large-scale assessment programs (ESL/EFL or ELA) and in other contexts (e.g., writing items for assessment-programs in university-based ESL programs).

Item writers undergo a 90-minute orientation prior to beginning item writing. This training focuses on the item specifications, the process and procedures, the item writing checklist, the acceptance criteria for the items, and the security protocols. Item writers also receive an item writing handbook, which formalizes the content of the orientation, along with assignment of themes to develop and the associated item specifications. After the orientation, CAL Language Testing Specialists and managers provide feedback to the item writers on the items, focusing on alignment with the item writing checklist and the item specifications. After completion of item

writing for a given development cycle, item writers are evaluated by CAL staff for their compliance with the requirements and the quality of their items.

Standards Expert Review

After items have been drafted by item writers, CAL Language Testing Specialists review all of the raw content internally. This review focuses on determining which sets of items will move on to further development and which will be discontinued, based on criteria from an item review checklist. The Language Testing Specialists then do minor editing and formatting to the items to make sure that they are complete, with no stray comments or other editorial notes from previous drafts, and they produce a short questionnaire for each set of items that becomes part of Standards Expert review. The purpose of Standards Expert review is to ensure that the items are appropriate for the grade-level and intended difficulty level in terms of both the content and the language, and the items have not drifted from their intended target between theme generation and item writing. The questionnaires produced by CAL's Language Testing Specialists guide the Standards Experts through the review process, asking questions specific to the purpose of this review.

Educators are recruited jointly by CAL and WIDA to serve as Standards Experts; educators with ESL or content-area expertise and two or more years of teaching experience in a WIDA state are invited to participate. Recruitment also focuses on a geographical distribution of educators from across the consortium. Standards Experts receive written instructions and a questionnaire to complete for each set of items they review.

Bias and Sensitivity and Content Review

After Standards Expert Review has been completed, all items undergo an additional phase of review and revision internal to CAL, leading up to Bias & Sensitivity and Content Review. These are technically two separate reviews, although a single recruitment effort is conducted by WIDA, and the reviews occur consecutively in a single week (generally 3 days for Content review followed by 2 days for Bias & Sensitivity review). As with other reviews, educators for Content review must have at least 2 years of ESL teaching experience (with a preference for content-area experience as well). Recruitment also focuses on selecting educators with a variety of cultural and linguistic backgrounds and obtaining a geographical distribution of educators from across the consortium. Recruitment for Bias & Sensitivity review focuses on selecting educators with culturally and linguistically diverse backgrounds who have experience interacting

with English learners from a range of cultural, regional, religious, linguistic, ethnic, and socioeconomic backgrounds.

At the beginning of both Bias & Sensitivity and Content review meetings, CAL and WIDA staff conduct an intensive training to orient the reviewers to the specific purpose of the review (Bias & Sensitivity or Content), how to use the review checklist and what to look for in the review, and the procedures and security protocols for the review. Then, the reviews are conducted in breakout groups by grade-level cluster (or combinations of grade-level clusters; for example, Bias & Sensitivity review of Grade 1 and Grades 2–3 is often combined). Although Bias & Sensitivity and Content reviews are generally held in -person, the reviews for the Writing domain occur virtually each year due to timeline constraints. For both the in-person and virtual contexts, CAL and WIDA facilitators are present in each breakout group to guide the educators in their reviews of the materials.

Writing Tryouts

For the Writing domain, all tasks in the Writing domain are subject to tryouts in the field. The Writing tryouts only occur once the tasks have been through a thorough Bias & Sensitivity and Content review and subsequent revision. CAL and WIDA recruit educators who are willing to administer the Writing tasks to their students; these educators are classroom ESL or content teachers who work with ELLs. All students who participate are required to have parent/guardian consent.

Once the students complete the Writing tasks, both the students and educators fill out questionnaires. Student questionnaires focus on whether the students understood the task, their engagement with the task, and their ability to complete the task; educator surveys ask the teachers to evaluate the effectiveness of the task input, the appropriateness of the task, the comparability of the task with other classroom-based writing tasks, and the ability of the students to complete the task.

CAL provides the teachers with a number of documents outlining the procedures for administering the tasks, recording student responses to the tasks, recording student and teacher responses to the questionnaires, and protecting the personally identifiable information of the students. CAL staff are also available throughout the tryouts process to answer any questions the teachers might have. Following the Writing tryouts, CAL specialists review the writing

responses both qualitatively and quantitatively, providing WIDA with a report on how the Writing tasks performed.

6.2   Test Administration Quality Control

This section describes how WIDA monitors test administration to ensure standardized test administration procedures are implemented with fidelity across districts and schools. To support standardized administrations, WIDA provides test administrators with a series of resources, such as a Test Administration Manual, a training course, and a Test Administration Script for each assessment.

Qualifications of Test Administrators

Before, during, and after a state's testing window, educators hold various roles to ensure all tasks are carried out for successful test administration. These roles include Test Coordinators at the district and school level and Test Administrators. The Test Administrator administers and monitors the test. They are also responsible for managing student data prior to, during, and after testing.

WIDA has worked directly with each state education agency to develop the ACCESS for ELLs Checklist for the school year. This list highlights all tasks that need to be completed before, during, and after testing within a school or district and outlines which tasks are assigned to Test Coordinators at the district and school level and to Test Administrators. It also provides additional guidance that a state expects test administrators to follow as they prepare for and administer the ACCESS for ELLs suite of assessments.

Test administrators are responsible for reviewing each state's checklist in detail prior to completing any training and for working with the district or school Test Coordinator to complete these tasks. The state's checklist can be found in the training course and on each state's WIDA webpage at www.wida.us/membership/states.

The training course within the WIDA Secure Portal (https://www.grow.wida.us/) is where educators can access both training to become certified to administer ACCESS for ELLs as well as additional materials and resources to assist administrators and coordinators before, during, and after each state's testing window. WIDA user accounts provide access to the training course and Facilitator Toolkit within the WIDA Secure Portal. Educators must pass an administration quiz at the end of the training with a score of 80% or higher. WIDA recommends taking the quiz

immediately after completing the training. There is no limit to the number of times educators can attempt the quiz. Once individuals pass an administration quiz, training certificates within the WIDA Secure Portal are updated to reflect their status as a certified test administrator for that component of the assessment suite.

Paper Testing (for Writing Grades 1–3)

Depending on state, district, and school policy, not all Test Administrators will be responsible for initially labeling and/or bubbling booklets. However, it is the responsibility of all Test Administrators and Test Coordinators to ensure that correct and complete information is either labeled or bubbled in each student booklet. Each state's ACCESS for ELLs checklist has more information on who is responsible for each task related to materials management in the state.

To ensure all booklets have the detailed and necessary information needed to score, all Test Administrators must adhere to the following:

- Prior to administration

  o Review labels and/or bubbled information to ensure all student information is accurate.

  o Complete labeling or bubbling if needed.

- During administration

  o Distribute the test booklets, as applicable, to the correct students.

  o Verify that students have been given their assigned booklet.

- Immediately following administration

  o Collect all material from all students.

  o Review student test booklets once more for any errors or discrepancies in student information.

  o Confirm all necessary fields are completed and all necessary labels are correctly adhered to student test booklets.

  o Ensure all booklets are in proper condition to be returned, with no loose or damaged pages.

  o 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:

- Overview and Test Structure

- Speaking Assessment Scoring Practice

- Speaking Assessment Recommended Practice

WIDA strongly recommends that all new raters complete all three of these modules. These modules provide a comprehensive introduction to the ACCESS for ELLs Speaking test and the

opportunity to learn how to score students' spoken English reliably using the ACCESS for ELLs Speaking Scoring Scale.

In addition to the modules described above, WIDA also releases supplemental training materials each year to refamiliarize experienced raters with the Speaking Scoring Scale and introduce new Speaking tasks and sample responses for the coming year. These materials, called Supplemental Training for the Speaking Assessment, reflect the Speaking tasks that will appear on the test in the current year. WIDA recommends that all raters (new and experienced) engage with these supplementary materials at the start of each scoring season. Reading and reviewing these materials will help raters maintain their reliability from year to year and contribute to the fairness of test scores awarded to all students.

Rater Certification

After completing the training modules described in the section above, new raters should take the relevant certification quiz. WIDA provides two quizzes: one for raters who will evaluate students in Grades 1–5 and another for raters who will evaluate students in Grades 6–12. Raters should take the appropriate quiz.

The purpose of the quiz is to ensure that raters have internalized the Speaking Scoring Scale and can apply it consistently. Only raters who pass the quiz(zes) should administer and score the ACCESS for ELLs Paper Speaking test.

Checklist for Rater Training, Monitoring, and Recertification

✓ New raters complete all Speaking assessment training

✓ New raters take and pass the appropriate certification quizzes

✓ All raters recertify at the start of each testing season (review new materials, retake quiz)

✓ Only certified raters administer and score the ACCESS for ELLs 2.0 Speaking test

✓ Raters do not evaluate their own students, if at all possible

✓ Rater reliability and/or score point distributions are monitored regularly

For more information on Writing rater QC, please refer to section 3.2.2.

6.4　Score Reporting Quality Control

WIDA conducts an annual score reporting quality control process to (1) verify the accuracy of paper-based test scores (i.e., ACCESS for ELLs Paper, Kindergarten ACCESS for ELLs, and Alternate ACCESS) and (2) verify the accuracy of all score reports (the Individual Student Report, the Student Roster Report, the School Frequency Report, the District Frequency Report, and the State Frequency Report) for both ACCESS (Online, Paper, and Kindergarten) and Alternate ACCESS.

The Score Reporting quality control is conducted at DRC's offices in Maple Grove, Minnesota. The team generally includes five state education agency representatives, one CAL employee, and four WIDA employees. This team examines data from three districts: a primary district, for quality control of all score reports; a secondary district, for quality control of State Frequency Reports only; and a tertiary district for quality control of paper-based tests only.

After an introductory presentation, which includes details of the quality control processes undertaken by DRC and WIDA and instructions on using the data entry tools, panelists begin by confirming the scoring of ACCESS Paper. Using the information in the State Student Response file, panelists enter the grade level, grade level cluster, tier, the Listening and Reading responses, and the Speaking and Writing scores into the data entry tool. The tool then calculates the student's raw scores and, using a series of look-ups, the student's scale score, proficiency level score, and confidence bands for all domains and composites. Panelists check student scores on the Individual Student Reports against those calculations. Any discrepancies are brought to the attention of the WIDA facilitator who investigates and, if there seems to be an issue with the report (rather than the data entry or data entry tool), discusses the issue further with DRC.

The panelists follow a similar process with the Kindergarten ACCESS tests, but with the raw scores for these tests copied directly from the response booklets.

After checking the paper-based tests, panelists turn their attention to the score reports. Panelists first check both the demographic information and the student scores in the Individual Student Reports against the information in the Student Roster Reports. Again, any discrepancies are brought to the attention of the facilitator, who investigates and discusses the issue with DRC if necessary. Panelists use the verified Individual Student Reports to check the Student Roster Report. Once the Student Roster Report is verified, panelists use it to check the State Frequency Report; they then use the verified State Frequency Reports to check the District Frequency

Report. Finally, panelists check the State Frequency Reports against verified District Frequency Reports from the primary district along with District Frequency Reports from the secondary district.

6.5   Data Forensic Quality Control

Caveon Data Forensic Analysis Results

WIDA hired Caveon to perform data forensic analysis during the 2022–2023 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 2022 through Spring 2023 online multistage adaptive test administrations, Listening and Reading domains

December 2022 through Spring 2023 paper fixed-form administrations, Listening and Reading domains

The analysis utilized several of Caveon's security statistics to detect evidence of whether the assessment instrument has been compromised through disclosure of the content. This analysis attempted to understand where and when disclosure of the test content may have occurred and what items and forms may have been affected. Results of this analysis enable WIDA to take specific actions to limit the impact of disclosed content. Such actions may include

- Republishing or reworking items or forms

- Rotating disclosed items to limit their exposure

- Designing a republication or rotation strategy for future items and forms

Caveon security statistics were computed for each individual test instance. These data were aggregated or summarized at the group level. The aggregated statistics were compared against the population model.

Analysis of Tests

Caveon aggregated the data according to individual test forms using the security statistics to determine whether rates of detections by the security statistics were higher for certain test forms. For fixed-form paper tests, two forms—A and B/C—were analyzed. For the multistage adaptive test, there is a finite number of ways a student could progress through the test. Caveon analyzed each pathway as a separate form. Higher rates of security detections for a specific form of the test suggest that compromise of the form may have occurred.

Analysis of Items

Item security: In this portion of the analysis, the security of the items was evaluated using aberrance statistics. Aberrance statistics detect test-taking behaviors such as answering difficult items correctly but answering easy items incorrectly, or unusual patterns in the time taken to answer test items. In the absence of security issues, aberrant test taking is expected to be the result of poor or uneven test preparation, illness or other physical malady, mental and emotional distractions, and so forth. These factors usually result in lower levels of test performance. When aberrance is associated with higher performance, however, test fraud may have occurred, such as pre-knowledge of test content. By applying aberrance measures and comparing the performance between aberrant and non-aberrant test instances on individual items, inferences can be made about item security.

Item performance changes: Analysis of item performance changes tracks individual item performance rates over time. The item performance shifts are measured within the context of the item response theory model and adjusted for varying test-taker performance levels. This means that detected performance shifts are invariant to fluctuations in the test-taker population. When performance shifts indicate the item has become significantly easier, the item may have been disclosed. Items with significant performance shifts become candidates for revision or replacement. Item performance shifts were detected with a granularity of 1 week, where Monday to Sunday represents 1 week.

Analysis of Groups

Analysis by week: This analysis aggregates the data according to the week in which the test was taken to identify whether security threats and pass rates appeared to be more prevalent at certain times during the testing window. Increases in scores or security detections during certain periods of time suggest the content may have been disclosed at some point prior to that time. This analysis also includes a form-date grouping to determine if increasing security threats are associated with a particular form of the test. This analysis is performed for online and paper tests, where relevant test date data are provided.

Analysis of WIDA jurisdictions: Caveon analyzed WIDA member jurisdictions (states and districts) to determine whether rates of detections by the security statistics were higher for certain jurisdictions. This analysis is intended to detect whether compromise at the state or member jurisdiction level potentially occurred. This analysis is performed for online and paper tests.

Analysis of administration mode: Caveon aggregates the data according to administration mode (i.e., online versus paper) to determine if security threats are associated with the mode of testing.

Other Analyses

Analysis of mean score over time was used to identify whether mean scores increased over time during the testing window. Increases in scores over time suggest the content may have been disclosed during the testing window.

Findings of Data Forensic Analyses

Generally, no major data forensic anomalies were observed across WIDA states. There were some general findings and a few minor localized anomalies. States where these anomalies occurred were notified.

References

ACCESS Test Practice and Sample Items | WIDA. (n.d.). Wida.wisc.edu; Wisconsin Center for Education Research at the University of Wisconsin-Madison. https://wida.wisc.edu/assess/access/preparing-students/practice

Allen, N. L., Carlson, J. E., & Zalanak, C. A. (1999). *The NAEP 1996 technical report.* 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*. American Psychological Association.

American Institutes of Research. (2018). *ELPA21 technical report, part I—Summative assessment*. 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]. Center for Advanced Studies in Measurement and Assessment.

Center for Applied Linguistics. (2016). *ACCESS for ELLs Series 400 listening and reading scale maintenance: Technical brief.* Author.

Center for Applied Linguistics. (2017). *ACCESS for ELLs 2.0 speaking and writing score scale reconstruction: Technical brief.* Author.

Center for Applied Linguistics. (2018). *Annual technical report for ACCESS for ELLs 2.0 Online English Language Proficiency Test, Series 401, 2016–2017 administration*. 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). Author.

Cook, H. G., & MacGregor, D. (2017). *The ACCESS for ELLs 2.0 2016 Standard-setting study* (Technical Report). 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.

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). 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.* WIDA Consortium.

Kamata, A., Turhan, A., & Darandari, E. (2003, April). *Estimating reliability for multidimensional composite scale scores.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Kane, M., & Case, S. M. (2004). The reliability and validity of weighted composite scores. *Applied Measurement in Education, 17,* 221–240.

Kenyon, D. M. (2006). *Development and field test of ACCESS for ELLs®* (WIDA Consortium Technical Report No. 1). 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). 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, Autumn). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878. 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). JAM Press.

Linacre, J. M. (2006). Winsteps Rasch analysis (Version 3.60.1) [Computer software]. http://www.winsteps.com

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*. Lawrence Erlbaum Associates.

MacGregor, D., Kenyon, D. M., Gibson, S., & Evans, E. (2009). *Development and field test of Kindergarten ACCESS for ELLs*. 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), 151–363.

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]. University of California – Davis.

WIDA Consortium. (2007). *English language proficiency standards and resource guide, 2007 edition, prekindergarten through grade 12*. Board of Regents of the University of Wisconsin System.

WIDA Consortium. (2012). *2012 amplification of the English language development standards kindergarten–grade 12*. Board of Regents of the University of Wisconsin System.

WIDA Consortium. (2022). *ACCESS for ELLs Interpretive Guide for Score Reports.* Board of Regents of the University of Wisconsin System.

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement.* 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). 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). 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.

From CAL: