

World-Class Instructional Design and Assessment



**Annual Technical Report for
ALTERNATE ACCESS for ELLs[®] English Language Proficiency
Test, Series 403, 2018-2019 Administration**

Annual Technical Report No. 7

Prepared by:

Psychometrics team

WIDA

January, 21, 2020

Table of Contents

1. Description of Alternate ACCESS for ELLs English Language Proficiency Test	1
1.1. Purpose of Alternate ACCESS for ELLs	1
1.2. Format of Alternate ACCESS	1
1.2.1 Integration with the Standards	1
1.2.1. Grade-level Clusters.....	2
1.2.2. Language Domains	2
1.2.3. Language Proficiency Levels.....	2
1.3. Test Development	3
1.3.1. Item Development.....	3
1.3.2. Field Test	4
1.3.3. Scaling.....	4
1.3.4. Standard Setting.....	5
1.4. Reporting of Results.....	8
1.4.1. Scale Scores	8
1.4.2. Language Proficiency Level Scores.....	8
1.5. Test Administration.....	9
1.5.1. Test Administrator Training	9
1.5.2. Test Security	9
1.5.3. Test Accommodations	9
1.6. Scoring.....	9
1.6.1. Listening and Reading	9
1.6.2. Writing	10
1.6.3. Speaking.....	11
2 An Assessment Use Argument for Alternate ACCESS for ELLs: Focus on Assessment Records	14
2.1 The Generic Validation Framework for Alternate ACCESS	15
2.2 Focus on Assessment Records	16
2.2.1 Breakdown of Claims for the Assessment Records Produced in the Alternate ACCESS for ELLs Assessment Program	18
2.3 Evidence for Assessment Records Claims of Alternate ACCESS for ELLs	19
2.4 Summary of Assessment Records Claims, Actions, and Evidence.....	25
2.5 Visual Guide to Tables and Figures.....	26
2.5.1 Chapter 4 Visual Guide to Tables and Figures	28
2.5.2 Chapter 6 Visual Guide to Tables and Figures	29
3. Descriptions of Student Results	30
3.1 Participation.....	30

3.2	Scale Score Results	31
3.2.1	Mean Scale Scores Across Domain and Composite Scores	31
3.2.2	Correlations	31
3.3	Proficiency Level Results	31
3.4	Participation by Disability	32
4.	Student Results	33
4.1	Students excluded from Analysis	33
4.1.1	Out-of-grade-level Test Administration	33
4.2	Participation by Grade-level Cluster	34
4.2.1	Participation by Grade-level Cluster by State	34
4.2.2	Participation by Grade-level Cluster by Gender	35
4.2.3	Participation by Grade-level Cluster by Ethnicity	35
4.3	Participation by Grade	36
4.3.1	Participation by Grade by State	36
4.3.2	Participation by Grade by Gender	37
4.3.3	Participation by Grade by Ethnicity	37
4.4	Participation by Domain	38
4.4.1	Participation by Grade-level Cluster by Domain	38
4.4.2	Participation by Grade by Domain	38
4.5	Scale Scores by Domain and Composite	39
4.5.1	Mean Scale Scores by Domain and Composite	39
4.6	Scale Scores by Grade-level Cluster	41
4.6.1	Mean Scale Scores by Gender	41
4.6.2	Mean Scale Scores by Ethnicity	43
4.7	Scale Scores by Grade	47
4.7.1	Mean Scale Scores by Gender	47
4.7.2	Mean Scale Scores by Ethnicity	53
4.8	Correlations among Scale Scores by Grade-level Cluster	65
4.8.1	Correlations among Scale Scores: Grade-level Cluster 1-2	65
4.8.2	Correlations among Scale Scores: Grade-level Cluster 3-5	65
4.8.3	Correlations among Scale Scores: Grade-level Cluster 6-8	66
4.8.4	Correlations among Scale Scores: Grade-level Cluster 9-12	66
4.9	Proficiency Levels	67
4.9.1	Proficiency Level by Grade-level Cluster	67
4.9.2	Proficiency Level by Grade	70
4.10	Participation by Disability	74
4.10.1	Participation by Disability	74
5.	Analyses of Test Forms: Overview	76
5.1	Background	76
5.1.1	Measurement Models Used	76
5.1.2	Sampling	77
5.1.3	Scaling	77

5.1.4	DIF Analyses	78
5.1.5	Reliability of Composites.....	79
5.1.6	Accuracy and Consistency of Classification.....	79
5.2	<i>Descriptions</i>	82
5.2.1	Raw Score Information (Figure A and Table A)	82
5.2.2	Scale Score Information (Figure B and Table B)	82
5.2.3	Proficiency Level Information (Figure C and Table C).....	83
5.2.4	Equating Summary Table (Table D).....	83
5.2.5	Reliability (Table E).....	83
5.2.6	Test Characteristic Curve (Figure D).....	84
5.2.7	Test Information Function (Figure E).....	85
5.2.8	Item Analysis Summary (Table F).....	85
5.2.9	Complete Item Analysis Table (Table G).....	85
5.2.10	Complete Raw Score to Scale Score Conversion Chart (Table H).....	86
5.2.11	Raw Score to Proficiency Level Score Conversion Table (Table I).....	86
5.2.12	Accuracy and Consistency of Classification Table (Table J)	87
5.2.13	Conditional Standard Error of Measurement for Composite Figure (Figure F)	87
6.	<i>Analyses of Test Forms: Results</i>	88
6.1	<i>Grades: 1-2</i>	88
6.1.1	Listening 1-2	88
6.1.2	Reading 1-2.....	94
6.1.3	Speaking 1-2	100
6.1.4	Writing 1-2.....	104
6.1.5	Oral Language Composite 1-2	110
6.1.6	Literacy Composite 1-2.....	113
6.1.7	Comprehension Composite 1-2.....	116
6.1.8	Overall Composite 1-2.....	119
6.2	<i>Grades: 3-5</i>	122
6.2.1	Listening 3-5	122
6.2.2	Reading 3-5.....	128
6.2.3	Speaking 3-5	134
6.2.4	Writing 3-5.....	139
6.2.5	Oral Language Composite 3-5	145
6.2.6	Literacy Composite 3-5.....	148
6.2.7	Comprehension Composite 3-5.....	151
6.2.8	Overall Composite 3-5.....	154
6.3	<i>Grades: 6-8</i>	157
6.3.1	Listening 6-8	157
6.3.2	Reading 6-8.....	163
6.3.3	Speaking 6-8	169
6.3.4	Writing 6-8.....	174

6.3.5	Oral Language Composite 6-8	180
6.3.6	Literacy Composite 6-8.....	183
6.3.7	Comprehension Composite 6-8.....	186
6.3.8	Overall Composite 6-8.....	189
6.4	Grades: 9-12	192
6.4.1	Listening 9-12	192
6.4.2	Reading 9-12.....	198
6.4.3	Speaking 9-12	204
6.4.4	Writing 9-12.....	209
6.4.5	Oral Language Composite 9-12	215
6.4.6	Literacy Composite 9-12.....	218
6.4.7	Comprehension Composite 9-12.....	221
6.4.8	Overall Composite 9-12	224
	References	227

1. Description of Alternate ACCESS for ELLs English Language Proficiency Test

1.1. Purpose of Alternate ACCESS for ELLs

The purpose of Alternate ACCESS for ELLs (hereafter, Alternate ACCESS) is to assess the developing English language proficiency (ELP) of English language learners (ELLs) with the most significant cognitive disabilities in Grades 1–12 in the states of the WIDA consortium. The assessment is rooted in the *Alternate English Language Development (ELD) Standards for English Language Learners with Significant Cognitive Disabilities* of the WIDA Consortium. Alternate ACCESS is a first of its kind attempt made by WIDA to assess ELP for ELLs with the most significant cognitive disabilities. As such, the assessment continues to be refined to clarify the construct and to develop a test design that better reflects the diversity of student language use within this population.

The WIDA ELD Standards are aligned to WIDA Consortium state academic content standards and form the core of the WIDA Consortium’s approach to instructing and testing academic English for ELLs with significant cognitive disabilities. Alternate ACCESS, which was developed based on the WIDA ELD Standards, may thus be described as a standards-based ELP test designed to measure proficiency for ELLs with significant cognitive disabilities. It assesses social and instructional English as well as the language associated with Language Arts, Mathematics, and Science within the school context across the four language domains of Listening, Reading, Writing, and Speaking.

Major purposes of Alternate ACCESS include¹:

- To meet federal accountability requirements for assessment practice for ELLs and students with disabilities as specified in The Every Student Succeeds Act (ESSA; 2015) and the Individuals with Disabilities Education Act (IDEA; 2004)
- To provide educators with a measure sensitive to ELP growth of ELLs with significant cognitive disabilities

1.2. Format of Alternate ACCESS

1.2.1 Integration with the Standards

The design of Alternate ACCESS is built upon the foundational WIDA ELD Standards. The four WIDA ELD Standards represented are:

Standard 1—Social and Instructional Language

ELLs communicate in English for **social and instructional** purposes in the school setting.

Standard 2— Language of Language Arts

ELLs communicate information, ideas, and concepts necessary for academic success in the

¹ From the WIDA Alternate ACCESS website, <https://wida.wisc.edu/assess/alt-access>

content area of **Language Arts**.

Standard 3—Language of Mathematics

ELLs communicate information, ideas, and concepts necessary for academic success in the content area of **Mathematics**.

Standard 4—Language of Science

ELLs communicate information, ideas, and concepts necessary for academic success in the content area of **Science**.

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

Social and Instructional language: SI

Language of English Language Arts: LA

Language of Mathematics: MA

Language of Science: SC

The selected response items and performance-based tasks on Alternate ACCESS target these four Standards.

1.2.1. Grade-level Clusters

The WIDA ELD Standards describe developing ELP for five grade-level clusters. These are PreK-K, 1-2, 3-5, 6-8, and 9-12. A Kindergarten version of Alternate ACCESS, however, is not currently available. Thus, Alternate ACCESS is organized into the following grade-level clusters: 1-2, 3-5, 6-8, and 9-12.²

1.2.2. Language Domains

The Alternate ACCESS test includes individual sections to assess each of four language domains: Listening, Reading, Speaking, and Writing.

1.2.3. Language Proficiency Levels

Alternate ACCESS assesses growth in ELP over six levels. These six levels include three newly developed language proficiency levels and three levels derived from the WIDA ELD Standards for the general population. The most basic proficiency level is A1: ‘Initiating,’ and the most advanced stage of language proficiency described is P3: ‘Developing’. The first three levels of the Alternate ELD proficiency levels, A1 – A3, are language proficiency antecedents to the existing WIDA ELD P1 that applies to the general student population. An important aspect of the Alternate ELD levels (A1 – A3) is that they represent small chunks of language growth within P1. A highlight of this structure is that progress in language acquisition for students with significant cognitive disabilities can be identified in smaller and narrower gradations. Figure 1.2.4A below presents a conceptualization of the proficiency levels assessed in Alternate ACCESS. In this figure, P1 has

² The organization of grade-level clusters is based on the 2007 WIDA ELP Standards (WIDA, 2007).

been stretched for illustrative purposes to display levels A1 – A3.

ACCESS. In this figure, PL1 has been stretched for illustrative purposes to display levels A1 – A3.

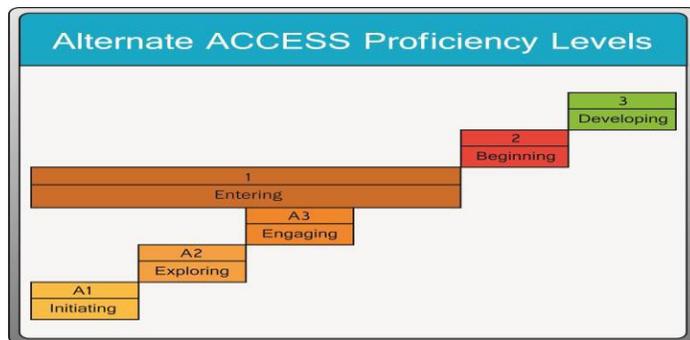


Figure 1.2.4A. Alternate ACCESS Proficiency Levels

These language proficiency levels are thoroughly embedded in the WIDA ELD Standards in a two-pronged fashion.

First, they appear in the **performance definitions**. According to the WIDA ELD Standards, the performance definitions provide a global overview of the stages of the language acquisition process. As such, they complement the **Alternate Model Performance Indicators (AMPIS)** for each language proficiency level (see the next paragraph for further description of the AMPIS). The performance definitions are based on three criteria. The first is students’ increasing comprehension and production of the technical language required for success in the academic content areas. The second criterion is students’ demonstration of oral interaction or writing of increasing linguistic complexity. The final criterion is the increasing development of phonological, syntactic, and semantic understanding in receptive skills or control in usage in productive language skills.

Second, the language proficiency levels of the WIDA ELD Standards are fully embedded in the accompanying AMPIS, which exemplify the Standards. The AMPIS describe the expectations for ELLs with significant cognitive disabilities for each of the four **Standards**, at the four different **grade-level clusters**, across four **language domains**, and at each of the **language proficiency levels**. The sequence of these five AMPIS together describes a logical progression and accumulation of skills on the path from the lowest level of ELP to full proficiency for academic success. This progression is called a “strand.”

Each selected-response item or performance-based task on Alternate ACCESS is carefully developed, reviewed, piloted, and field tested to ensure that it allows students to demonstrate accomplishment of the targeted AMPIS. (See the sample items at the WIDA website [<https://wida.wisc.edu/assess/alt-access>] for examples.)

1.3. Test Development

1.3.1. Item Development

Items developed for Alternate ACCESS were field tested on Form 100 and included on Form 101. The initial item writing for Alternate ACCESS was done during the grant phase of test

development at the University of Wisconsin. The subsequent pool of items was then refined by the CAL test development team. An internal review of the items was conducted, and items were chosen for further development based on how well they fit the Standards and AMPIs. The chosen items were refined by CAL staff before proceeding through further test development activities. Upon internal revision and development of test forms, CAL conducted the following test development activities, each followed by further internal review and revisions: Bias and Content Reviews, Pilot Testing, and WIDA/SEA's Forms Review. Details regarding this portion of the test development cycle can be found in the *Alternate ACCESS for ELLs Technical Report for Form 100*.

1.3.2. Field Test

Field testing of Alternate ACCESS Form 100 was conducted from March 12 to June 1, 2012. The purpose of the field test was to collect data on items and tasks, to judge the strength of individual items and tasks, to develop the Alternate ACCESS reporting scale, and to conduct the Standard Setting Study.

In total, 1,912 students in Grades 1-12 in 15 WIDA states participated in the field test. Participating SEAs encouraged educators in their states to sign up for the field test through the regular ACCESS for ELLs test ordering site provided by MetriTech, Inc. The administrations were labeled as an operational field test, meaning states had the option of designating participation in the testing as a field test activity or as the first operational testing opportunity of the Alternate ACCESS program. For more details about the field test please refer to the *Alternate ACCESS for ELLs Technical Report for Form 100*.

1.3.3. Scaling

Scaling is the process of developing a standard scale that maintains a consistent meaning across test administrations. Reporting scores on such a scale allows users to interpret test scores.

For Alternate ACCESS, a three-digit scale score (910 to 960) was selected to aid in score interpretation. The scale needed an interpretive center point across domains and composites, so the centering value of 935 was chosen to represent the midpoint of the cut score between proficiency levels A3 and P1 for the 3-5 grade-level cluster (see "Creating the Composite Scores" on the next page for more information about the composites). This is analogous to the ACCESS for ELLs scale, where the score of 350 is set as the center value and represents the cut score between proficiency levels P3 and P4 for Grade 5 (for more information see Kenyon, 2006).

Because the test blueprints across grade-level clusters by domain are the same and the Alternate PLs and AMPIs for the test tasks across grade-level clusters pose nearly identical linguistic challenges and differ only in the topics presented, it is desirable to have common cut scores across grade-level clusters by domain. In order to derive these common cut scores, however, test scores from all grade-level clusters need to be placed on a common scale. A common Rasch logit scale was developed to put the task parameters across grade-level clusters on the same scale, allowing test scores from all grade-level clusters to be placed on a common scale. Because the same scoring rules are used to convert students' original responses to raw scores by domain, a single rating scale was modeled across all grade-level clusters by domain. This was achieved by imposing the same threshold parameters across the four grade-level clusters by domain. Through this scaling process, task parameters as well as test scores across grade-level clusters are put on the same scale.

The procedure for developing the reporting scale for Alternate ACCESS was complex, but involved a number of basic steps. These were carried out separately for each domain until the last stage, when the separate domain scales were combined to form the composite scores. These steps, as conducted following the field test administration, are briefly summarized here. For more details about the field test please refer to the *Alternate ACCESS for ELLs Technical Report for Form 100*.

Scaling Design: The measurement model that formed the basis of the Alternate ACCESS scaling analyses was the Rasch Rating Scale Model (Andrich, 1978), as this model is appropriate for polytomously scored test tasks. For the initial Rasch calibration, the Rasch analyses were conducted separately by grade-level cluster and domain; therefore, the parameters for each grade-level cluster and domain were expressed on a unique logit scale. In the later stages of the psychometric analysis, the step or threshold parameters were constrained to be equal across grade-level clusters by domain through an anchoring process in order to put the task parameters across grade-level clusters by domain on the same logit scale. The Grade 3-5 step or threshold parameters were then used as the common step values, primarily because more Grade 3-5 students participated in the field test, therefore producing more stable parameters than other grade-level clusters. For each domain, the Grades 1-2, 6-8, and 9-12 rating scale threshold parameters were anchored to the Grade 3-5 domain values using Winsteps. The difficulty parameters for Grades 1-2, 6-8, and 9-12 were unanchored and thus were calibrated in the runs. All task parameters including the difficulty and threshold parameters were placed on the same logit scale across grade-level clusters by domain through this process. The logit scales were then transformed to the common reporting scale.

Developing the Logit Scale: A calibration of the ability of the students and items using Rasch procedures was applied to the scored student responses, putting the difficulty of the items or tasks and the ability of the students onto one common interval linear scale. The units of this scale are called logits, and by default the scale is usually centered at 0 (representing the average item difficulty for the ACCESS for ELLs items being calibrated). Theoretically, the logit scale runs from minus infinity to plus infinity, although in practice most tests run from about -4 logits to +4 logits.

Transforming the Logit Scale to the Reporting Scale: The logit scale has both negative numbers and decimals, which makes it confusing for many users. Therefore, scores on the logit scale were then transformed onto a reporting scale by means of a linear transformation of the Alternate ACCESS score scale. There is a separate scale for each of the four domains: Listening, Reading, Writing, and Speaking.

Creating the Composite Scores: The scores on the four reporting scales were then combined, in predetermined proportions, to create four composite scores: an Oral Language score (based on performances in Listening and Speaking), a Literacy score (based on performances in Reading and Writing), a Comprehension score (based on performances in Listening and Reading), and an Overall score (based on performances in all four domains).

1.3.4. Standard Setting

The goal of the Standard Setting Study was to interpret performances on the Alternate ACCESS operational field test form in terms of the WIDA ELD Standards, AMPIs, and the WIDA Alternate ELP levels. As discussed in 1.3.3., because the test blueprints across grade-level clusters by domain are the same, and the Alternate ELP levels and AMPIs for the test tasks across grade-level clusters pose nearly identical linguistic challenges and differ only in the topics presented, common cut

scores were set across grade-level clusters by domain. The study was held in Arlington, VA, on October 9-10, 2012.

The *Angoff Yes/No* methodology was used for all four domains because this method is thought to simplify the cognitive tasks that panelists are asked to perform (Cizek & Bunch, 2007). Having a straightforward cognitive task was important in this study as panelists had to examine many tasks to set four cut scores (A1/A2, A2/A3, A3/P1, and P1/P2) across the four domains (Listening, Speaking, Reading, and Writing).

The *Angoff Yes/No* method was designed for multiple choice and dichotomously scored tasks. This method asks the panelists to consider a student currently functioning at the borderline between two adjacent levels and then to review each question on the test, judging each task as either: a) *Yes, the borderline student is more likely than not to meet expectations for this task*; or b) *No, the borderline student is not more likely than not to meet expectations for this task*. Under this method, the average of the panelists' *Yes* decisions represents an estimated proportion of the target borderline group who would correctly answer the task.

Some modifications were made to the typical *Angoff Yes/No* methodology. First, for the two tasks in Writing Part C, which are scored using a rubric, panelists were shown various writing samples from all score points and asked to make the decision whether *Yes, the borderline student is more likely than not to have produced this sample*, or *No, the borderline student is not more likely than not to have produced this sample*. This approach to addressing the two rubric-scored tasks meant that the same judging procedures that the panelists used on all other tasks could also be used for these two tasks. The second modification was that the *Yes/No* judgment data collected from the panelists was analyzed using a logistic regression procedure to determine cuts. Logistic regression is a statistical technique for relating a continuous variable (i.e., the difficulty of the assessment tasks) to a dichotomous outcome (i.e., the *Yes/No* decisions made by the panelists). This approach was used to avoid limitations in the traditional summation approach of calculating final cut scores with the *Angoff Yes/No* method, which systematically makes lower cuts easier and higher cuts more difficult as compared to the typical *Angoff* method.

Standards were set on Writing Parts A and B and Speaking using the following procedure. Starting with a student at the lowest borderline within the WIDA Alternate ELP levels (i.e., between A1 and A2), panelists independently indicated whether that borderline student would be more likely than not to meet the expectation for the task. If their decision was *No*, panelists then went on to consider a borderline student at the next higher borderline on that same task (i.e., between A2 and A3). This process was continued, considering students at progressively higher levels of proficiency until they reached the highest borderline OR until they indicated *Yes*, that the borderline student would be more likely than not able to meet expectations for that task. Once a decision of *Yes* was made, then all higher borderlines would also necessarily be *Yes* and did not need to be individually considered. This aspect of the procedure greatly simplified the panelists' task.

After panelists considered the borderlines for one task, they then examined the next task and began again by considering a student at the lowest borderline. This process continued until panelists had considered all the borderlines on all the tasks. The test tasks were considered in the same order as they are presented in the Alternate ACCESS test booklets. Each panelist completed these evaluations independently. After the first round of evaluations, results for each task were tallied, allowing the panelists to see the 'average' borderline student (e.g., A2/A3) at which the group had determined the task to be more likely than not be answered correctly.

Writing Part C consisted of two writing tasks that were scored using a five-point rubric ('No Response,' 'Approaches,' 'Meets 1,' 'Meets 2,' and 'Meets 3') and therefore required a slightly different approach. Sample student responses to the two writing tasks were presented to panelists. Panelists were asked to determine whether a student at each borderline would be more likely than not able to have produced each writing sample.

For Listening and Reading, the prompts for the assessment tasks are repeated to students with increasing levels of support, allowing students multiple opportunities to respond. The repeated prompts are labeled as: CUE A: Initial Prompt; CUE B: Simplified Prompt; CUE C: Simplified Prompt & Answer. A response meeting expectations at CUE A (i.e., with minimal support) is interpreted as demonstrating a higher level of proficiency than a response meeting expectations at CUE B, and a response meeting expectations at CUE B exhibits higher proficiency than one at CUE C. For Listening and Reading, the panelists' task was the same as for Writing Parts A and B and Speaking, except that before moving on to the next task they first considered all borderlines on the first task at CUE A, then all borderlines on that task at CUE B, and, finally, all borderlines on that task at CUE C.

For all tasks across all four domains, panelists provided *Yes/No* decisions in a two-round process. In Round 1, panelists independently made their decisions. Staff members then typed the decisions into a specially prepared Excel spreadsheet which tallied the results by the total number of *Yes* and *No* responses. The tallied *Yes/No* decisions across panelists in the group were then revealed to all panelists on a screen with an LCD projector, at which point the panelists had the opportunity to comment on the tallies. Following this discussion, empirical data on student performances on the tasks were presented to the panelists. Using the results from the first round and this new information, the panelists then made a second round of independent *Yes/No* decisions. The Round 2 decisions were again entered and shared with the entire group. A brief opportunity was given to anyone who wanted to comment on the group results before moving on to the next language domain. At the conclusion of the study, researchers used the percentage of *Yes* decisions across panelists from Round 2 to derive the cut scores.

To derive the final cut scores by domain, a series of logistic regression analyses were conducted. A logistic regression analysis was conducted for each cut for each domain (e.g., the A3/P1 cut for Listening) using the panelists' *Yes/No* decisions across test tasks and grade clusters in that domain. The logistic function was used to find the location along the underlying ability continuum at which 50% of the panelists thought that the borderline student is more likely than not to meet the task expectations. This point became the cut point between the two adjacent proficiency levels being analyzed.

For more details regarding the Standard Setting Study, please refer to the *Alternate ACCESS for ELLs Standard Setting Study: Technical Brief* (CAL, 2012a).

1.4. Reporting of Results

1.4.1. Scale Scores

Alternate ACCESS scores are reported as both scale scores and proficiency level scores. Scores are given for all four language domains. In addition, four composite scores are given: Oral Language (based on performances in Listening and Speaking), Literacy (based on performances in Reading and Writing), Comprehension (based on performances in Listening and Reading), and Overall (based on performances in all four domains).

Raw scores are converted to scale scores through processes called scaling (see section 1.3.3 for details). These processes allow scores to be reported on a standard scale that is familiar to test users and that remains constant across test forms and grade-level clusters. Scale scores range from 910 to 960.

In determining the Oral Language and Literacy composite scores, equal weight is given to each domain. However, in determining the Comprehension and Overall composite scores, more weight is given to literacy skills than to oral skills. The scores are weighted as follows:

Comprehension = 70% Reading + 30% Listening

Overall = 35% Reading + 35% Writing + 15% Listening + 15% Speaking

1.4.2. Language Proficiency Level Scores

In addition to the scale scores, users of Alternate ACCESS also receive proficiency level scores. These scores are *interpretive*; that is, they interpret a student's scale score in terms of the results of the Standard Setting Study. The cut scores between proficiency levels are presented in Table 1.4.2.

Table 1.4.2A

Cut Scores by Domain and Composite

Domain	A1/A2	A2/A3	A3/P1	P1/P2
Listening	925	932	937	942
Reading	924	932	937	942
Speaking	925	930	939	945
Writing	923	931	938	947
Oral Composite	925	931	938	944
Literacy Composite	924	932	938	945
Comprehension Composite	924	932	937	942
Overall Composite	924	931	938	944

1.5. Test Administration

1.5.1. Test Administrator Training

Test administrators for Alternate ACCESS are required to take the appropriate steps to prepare themselves for test administration. The training steps included reading through the Alternate ACCESS Test Administration Manual (TAM) (WIDA, 2012a) and the Alternate ACCESS Test Administration Tutorial (available on the WIDA website). Test administrators are instructed to internalize the Writing and Speaking rubrics which are essential to consistent scoring across test administrations. For the Writing section, in addition to these materials, the Writing Scoring Guidance document provides sample student papers that help calibrate scoring for the Writing Section.

1.5.2. Test Security

Every effort is made to keep the test secure at all levels of development and administration. CAL and MetriTech follow policies and procedures regarding the security of the test, and every individual involved in the administration of the test from the district to the classroom level is trained in issues of test security.

1.5.3. Test Accommodations

Alternate ACCESS was designed for a population of students with a wide range of physical and cognitive disabilities. As such, the test design and layout reflect built-in features that aim to provide accessibility and are included as available accommodations on standardized tests for the general population. However, there are many situations where test administrators would need to modify the test administration in order to accommodate student-specific needs. In such cases, the criteria for implementation of any accommodation is determined primarily by the following: guidance in a student's Individual Education Plan (IEP), state accommodation policies, and the WIDA guidelines for appropriate test accommodations specified in the Alternate ACCESS TAM.

1.6. Scoring

All domains (Listening, Reading, Writing and Speaking) are scored locally by test administrators in individual Student Response Booklets. Test administrators must prepare for the scoring of each of the sections by following guidance provided in the TAM. Additional materials for ensuring that test administrators understand the correct scoring guidelines include the Alternate ACCESS Test Administration Video Tutorial and Writing Scoring Guidance document available through the WIDA website at <http://www.wida.us/assessment/alternateaccess.aspx>. Once a school has finished testing, all test booklets are returned to MetriTech, where they are electronically scanned and recorded in an electronic database in preparation for data analysis.

1.6.1. Listening and Reading

As with all sections of the Alternate ACCESS test, the Listening and Reading sections are scored by the test administrator. The Listening and Reading tests are identical in administration

procedures and consist of selected-response items that provide students with multiple opportunities to demonstrate their knowledge. It is helpful to understand the administration guidelines for the Listening and Reading tasks in order to understand the scoring procedures. The following steps are used to administer each task in the Listening and the Reading sections:

1. Administer CUE A (initial prompt and question for the task).
2. If the student does not respond, the test administrator must repeat CUE A again, as indicated in the test administrator's script.
3. If the student answers incorrectly or does not respond to CUE A, the test administrator will read CUE B. CUE B simplifies the initial prompt and asks the question again.
4. If the student responds incorrectly, or does not respond at all after the test administrator reads CUE B, the test administrator will administer CUE C. This cue provides the answer to the question, restates the prompt, and asks the question again.

Based on these administration guidelines for Listening and Reading, a student has a maximum of four opportunities to respond to each task (CUE A – 2, CUE B – 1, CUE C – 1). If a student responds correctly to the task at CUE A (including if the teacher repeated CUE A) the test administrator will score the task as **Correct at CUE A**. If after the two possible attempts at CUE A the test administrator moves on to CUE B and the student answers correctly, they will be scored as **Correct at CUE B**. Likewise, if the student has reached CUE C and answers correctly, they will be scored as **Correct at CUE C**. Finally, if after the four possible chances to answer the task the student has not selected the correct answer, the teacher will mark the task as **Incorrect**. If the student did not respond to any of the four opportunities, the task will be marked as **'No Response.'** Test administrators record all student responses in a Student Response Booklet.

1.6.2. Writing

As mentioned earlier, the Writing section is also scored by locally by the test administrator. It is important to understand the design and administration procedures of the Writing test in order to understand the scoring procedures.

The Writing section has three thematic folders, Parts A, B, and C.

- Part A of the Writing section has tasks at levels A1-P1.
- Part B of the Writing section has tasks at levels A1–P1.
- Part C provides the student with tasks at Levels P1 – P3; a student is only administered Part C if s/he scores 'Meets' on seven of the eight tasks in Parts A and B.

In Parts A and B of the Writing section, the script is designed for the test administrator to model each task for the student. This provides students the opportunity to observe the test administrator perform the task before trying it. For example, in the first task of the Writing section, the test administrator's script will instruct the test administrator to draw a circle around an image before asking the student to do the same. Similar to the Speaking section, each task in the Writing section provides the student with multiple opportunities for the student to produce a response. If the student produces a response that is appropriate for the task, a score of **'Meets'** is assigned, and if

the student does not produce a response that meets task expectations, a score of ‘**Approaches**’ is assigned. If the student does not respond during the task administration, ‘**No Response**’ is assigned to the task. The TAM instructs teachers to score the Writing section using scoring guidance provided in a column of the Writing score sheet termed the ‘Expect’ box. For each task in Parts A and B, the ‘Expect’ box provides the test administrator with a description of a response that would meet the task expectations (e.g., copy or write a word related to the task). The scoring guidelines in the ‘Expect’ boxes parallel the Writing rubric available in the TAM and the Student Response Booklet. Part C is scored based on the Writing rubric. Student performances can receive a score of ‘Meets 1,’ ‘Meets 2,’ ‘Meets 3,’ ‘Approaches,’ or ‘No Response.’ A score of ‘Meets’ 1, 2 or 3 corresponds to performances described in the Writing rubric for PL 1, 2, or 3. Test administrators are trained to follow the WIDA Consortium’s Writing Rubric for Alternate ACCESS and have access to Writing training materials through the WIDA website (www.wida.us). Table 1.6.2A presents the Writing Rubric.

Table 1.6.2A

Writing Rubric for Alternate ACCESS

Level	Text Features
3-Developing	One or more simple and expanded sentences. Words in the sentence(s) may be original or adapted from model or source text. Generally comprehensible. Comprehensibility may be impeded from time to time by errors when text becomes more complex. Text is related to the task.
2-Emerging	One or more simple phrases. Text is original or adapted from the model or source text. Comprehensible when text is adapted from model or source text. Comprehensibility may be impeded by errors in original text. Text is related to the task.
1-Entering	One or more general content words. Text is original or adapted from the model or source text. Generally comprehensible when text is adapted from model or source text. Comprehensibility may be significantly impeded in original text. Text is related to the task.
A3-Engaging	Single words and numbers. All or part of text is copied. If original text is present, it is not related to the task. Comprehensibility of the text may be significantly impeded by imprecise letter, symbol, or number formation. Text may or may not be related to the task.
A2-Exploring	Common single-digit numbers, letters, symbols, or syllables. All or part of text is copied. Comprehensibility of the text may be significantly impeded by imprecise letter, symbol, or number formation. Text may or may not be related to the task.
A1-Initiating	Pictorial representations and imprecise, but intentional markings, such as drawings and scribbles. Representations may or may not be related to the task.

1.6.3. Speaking

The Speaking section is also scored by the test administrator. As with other sections of the test, it is helpful to understand the design and administration guidelines for the Speaking section in order to understand the scoring criteria for the Speaking section.

The Speaking section has two thematic folders, Parts A and B. Thematic folders are a set of tasks based on a common setting or story (e.g., students in the library). The graphic(s) and character(s) often remain the same for all the tasks in a thematic folder.

- Part A of the Speaking section has tasks at levels A1 - A3.

- Part B of the Speaking section has tasks at levels A1 - P2.
- The script for all tasks includes three questions (Question 1, 2, and 3), which offers multiple opportunities for the student to provide a response at a given task level.

In the Speaking section, the student is given up to six opportunities to respond. This provides students with multiple opportunities to respond appropriately to the task in English. For each task, the test administrator reads Question 1 and prompts the student to respond. If the student does not score 'Meets,' the test administrator must repeat the task again. If the student still does not score 'Meets' after the repetition, the test administrator must ask Question 2, which simplifies the prompt and, in some tasks, models the expected response. If the student again does not score 'Meets,' Question 2 must be repeated. If the student does not score 'Meets' after that repetition, the test administrator must administer Question 3. Again, if the student does not score 'Meets,' this question is repeated once. The possibility of repetition for all three questions provides the student with six opportunities to produce a response in each Speaking task. If the student produces an appropriate response to the task at any point within the six provided opportunities, the task is scored as 'Meets.' If the student is not able at any point to produce a response that meets task expectations, a score of 'Approaches' is assigned. If the student does not make any attempt to respond to the task, a score of 'No Response' is assigned. The TAM instructs teachers to score the Speaking section using scoring guidance provided in a column of the Speaking score sheet termed the 'Expect' box. For each task, the 'Expect' box provides the test administrator with a description of a response that would meet the task expectations (e.g., repeat a word or produce a phrase related to the task). The scoring guidelines in the 'Expect' boxes parallel the Speaking rubric shown in Table 1.6.3A.

Table 1.6.3A

Alternate ACCESS Speaking Rubric

Level	Speech Features
2—Emerging	<p>Phrases or short sentences.</p> <p>General language related to the task; groping for vocabulary when going beyond the highly familiar is evident.</p> <p>When using simple discourse, is generally comprehensible and fluent; communication may be impeded by groping for language structures or by phonological, syntactic, or semantic errors when going beyond phrases and short, simple sentences.</p>
1—Entering	<p>Single words or chunks of memorized oral language.</p> <p>General vocabulary from school setting and related to task.</p> <p>When using memorized language, is generally comprehensible; communication may be significantly impeded when going beyond the highly familiar.</p>
A3—Engaging	<p>Single words or chunks of mimicked oral language.</p> <p>Mimicked high frequency vocabulary words related to the task.</p> <p>When using mimicked language, is generally comprehensible; communication may be significantly impeded when going beyond mimicked language.</p>
A2—Exploring	<p>Single syllables or syllables of single words; speech is mimicked.</p> <p>Mimicked sounds and syllables of high frequency vocabulary words related to the task.</p> <p>Language is minimal.</p>
A1—Initiating	<p>Communicative vocalizations, which may be imitated (e.g., grunts).</p> <p>Indiscriminant sounds and syllables.</p>

2 An Assessment Use Argument for Alternate ACCESS for ELLs: Focus on Assessment Records

Validity is “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 2014, p. 11). Evaluations of test validity assess whether there is evidence that supports the appropriateness and adequacy of the interpretations and decisions made about test takers on the basis of their performance on a test. This chapter contextualizes the information presented in this Annual Technical Report within an argument-based approach to addressing validity (Bachman & Palmer, 2010; Chapelle, Enright, & Jamieson, 2008; Kane, 2002, 2013; Mislevy, Almond, & Lukas, 2004) for Alternate ACCESS for ELLs.

A fully developed validation framework, including an Assessment Use Argument (AUA) (Bachman & Palmer, 2010), consists of several steps (described in Section 2.1 below) that connect test design and administration to intended and actual score interpretation and consequences. This chapter begins the process of developing a complete validation framework for Alternate ACCESS for ELLs. This argument-based structure organizes the information in this Annual Technical Report to support claims about Assessment Records (i.e., test scores and proficiency level descriptions collected via Alternate ACCESS for ELLs). Specifically, tables and figures from this report are explicitly linked to questions related assessment data. Chapelle, Enright, & Jamieson (2010) support using such a structure to present information to assessment users because “based on an analysis of four points of comparison—framing the intended score interpretation, outlining the essential research, structuring research results into a validity argument, and challenging the validity argument—we conclude that an argument-based approach to validity introduces some new and useful concepts and practices” (p.3). A larger, though yet undocumented (as of 2014), validity argument for the complete assessment from its inception to its consequences is currently under development by WIDA.

The complete validity argument that will be employed to support the use of Alternate ACCESS for ELLs will show the path from test design to test taker performance to the uses and interpretations of test scores and the subsequent consequences of test use. This framework is structured around assertions, or claims, about the assessment. The claims are presented as a series of statements that connect some aspect of the assessment process to the intended purposes of the assessment. Evidence for each claim is then organized by the action that is used to ensure each claim, and it includes results from analyses of test data, outside documentation, and other resources. In the complete validation argument, this process of identifying evidence to support claims will encompass the entire testing process, from the commencement of the test design to the consequences of test use (Bachman & Palmer, 2010; Llosa, 2008); Figure 2A shows the process by which evidence supports validation actions, which are used to establish larger claims about Alternate ACCESS for ELLs.

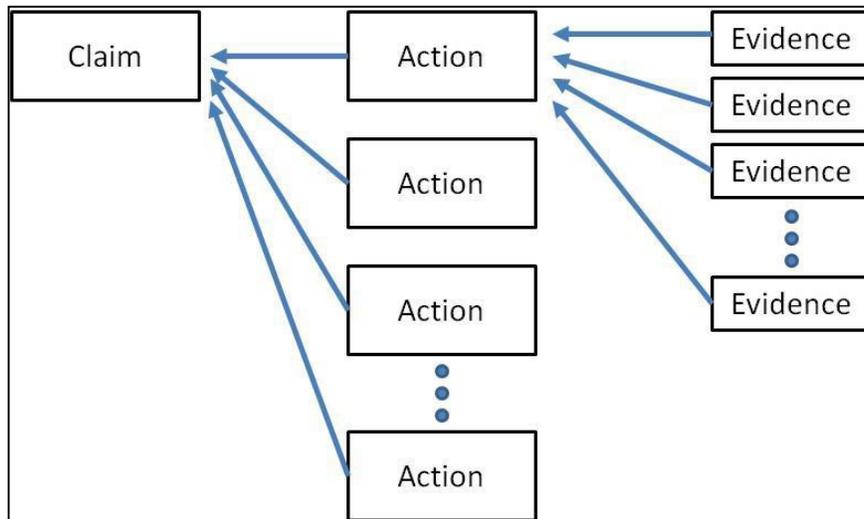


Figure 2A: General Argument Structure for Assessment Validation

2.1 The Generic Validation Framework for Alternate ACCESS

The generic validation framework that will be applied to the entire Alternate ACCESS for ELLs testing process was developed at the Center for Applied Linguistics (CAL) and is hereafter referred to as CAL's Validation Framework. CAL's Validation Framework, shown in Figure 2.1A, combines models for both test development (i.e., Evidence-Centered Design [Mislevy, Almond, & Lukas, 2004]) and assessment validation (i.e., Bachman and Palmer's (2010) AUA) to cover the assessment development and implementation process from initial conceptualization to the score interpretations and consequences of using the assessment. This framework constantly looks both forward and backward; for example, during the initial *Plan* step (Step 7), test developers state the anticipated decisions and consequences of implementing the assessment program, which are investigated in the *Decisions* step (Step 2) and *Consequences* step (Step 1). Because each subsequent step depends upon the strength of the step below it, the steps are numbered from 7 to 1, with *Consequences* being the culmination of the previous steps. This structure highlights the fact that any weakness in a lower step affects the steps above it.

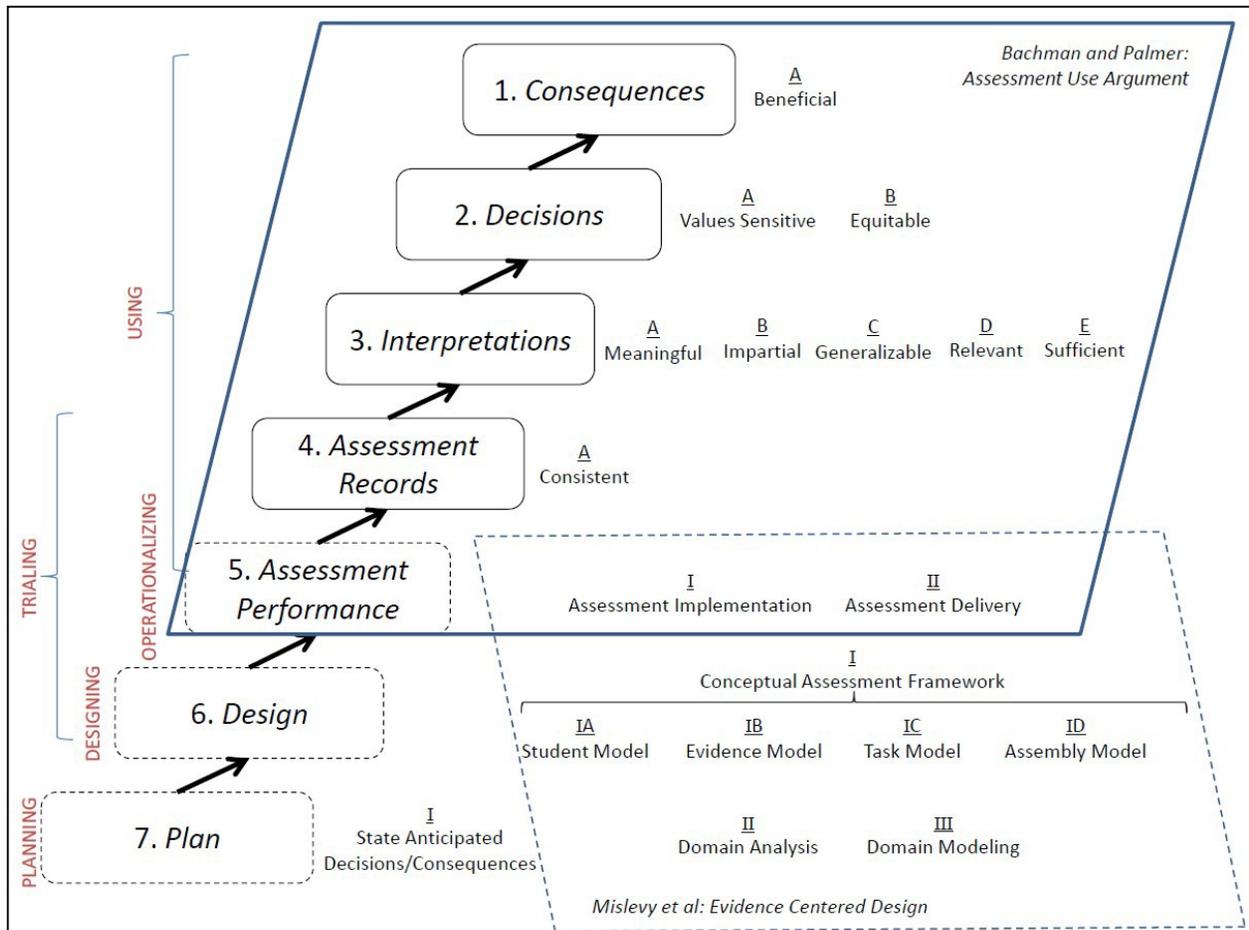


Figure 2.1A: CAL's Validation Framework (based on Bachman & Palmer, 2010; Mislevy, Almond, & Lukas, 2004)

In CAL's Validation Framework, the *Plan* step involves an examination of possible decisions states might make and consequences that might result from the assessment. This leads to the consideration of several models during the *Design* step, where specifications that answer such critical questions as "What are we measuring?" and "How do we measure it?" are developed (Mislevy, Almond, & Lukas, 2004). The subsequent steps of the validation framework highlight the trialing, implementation, and use of the assessment results, beginning with test takers' performance on the assessment (*Assessment Performance*) and continuing through the collection of test scores (*Assessment Records*), interpretations of those test scores (*Interpretations*), decisions made based on the test scores (*Decisions*), and the consequences of test use (*Consequences*).

The WIDA Consortium is using CAL's Validation Framework to present a complete validity argument, which will be updated as needed, for Alternate ACCESS for ELLs. To date, information related to Step 4, Assessment Records, has been explored and is found in this chapter.

2.2 Focus on Assessment Records

Although the complete validation framework for Alternate ACCESS for ELLs contains seven steps (see Figure 2.1A), the data presented in this document cover the Assessment Records step, which is part of Bachman and Palmer's (2010) AUA. By focusing on Assessment Records (i.e.,

test scores and proficiency level descriptions), the information in the Annual Technical Report will be used to support claims related to the quality and consistency of the assessment data gathered and analyzed using Alternate ACCESS for ELLs. The claims in this step of the AUA all pertain to the general question “How do we know that the reported language domain scores and composite scores on Alternate ACCESS for ELLs are consistent and dependable?” Other questions about the development, administration, and outcomes of Alternate ACCESS for ELLs will be evaluated in a forthcoming document, currently in development by WIDA.

The diagram in Figure 2.2A shows a visual representation of an argument-based approach for supporting claims related to Assessment Records. The figure shows how the Assessment Records step, Step 4 of the complete validation framework, will fit in the generic validation framework and be expanded into a series of claims and corresponding actions in this chapter of the Annual Technical Report. Evidence in the form of data from this report or other sources will be presented to support these claims as they relate to ACCESS for ELLs.

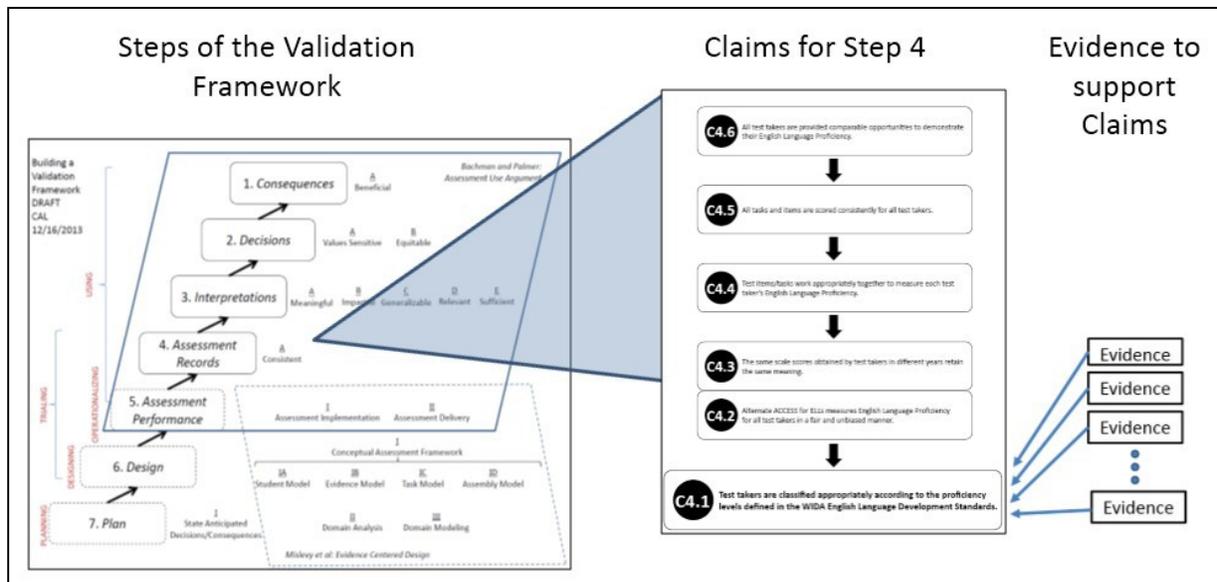


Figure 2.2A: Structure of the Argument-Based Approach Supporting Step 4 Contained in this Chapter

2.2.1 Breakdown of Claims for the Assessment Records Produced in the Alternate ACCESS for ELLs Assessment Program

The general *Assessment Records* step, Step 4 of the full Alternate ACCESS for ELLs validation framework, is broken down into the following six claims:

- C4.6. All test takers are provided comparable opportunities to demonstrate their English Language Proficiency.
- C4.5. All tasks and items are scored consistently for all test takers.
- C4.4. Test items/tasks work appropriately together to measure each test taker's English Language Proficiency.
- C4.3. The same scale scores obtained by test takers in different years retain the same meaning.
- C4.2. Alternate ACCESS for ELLs measures English Language Proficiency for all test takers in a fair and unbiased manner.
- C4.1. Test takers are classified appropriately according to the Alternate English Proficiency Levels defined in the WIDA English Language Development Standards.

As shown in Figure 2.2.1A, these claims depend upon each other, again moving from (4.6) up to (4.1). Within this organizational structure, each successive claim builds upon the previous one(s) (e.g., ratings are only useful to test developers and stakeholders if all test takers are provided comparable opportunities to demonstrate their proficiency). In the next section, these claims are broken down even further into actions that are taken to ensure the consistency and reliability of the assessment records.

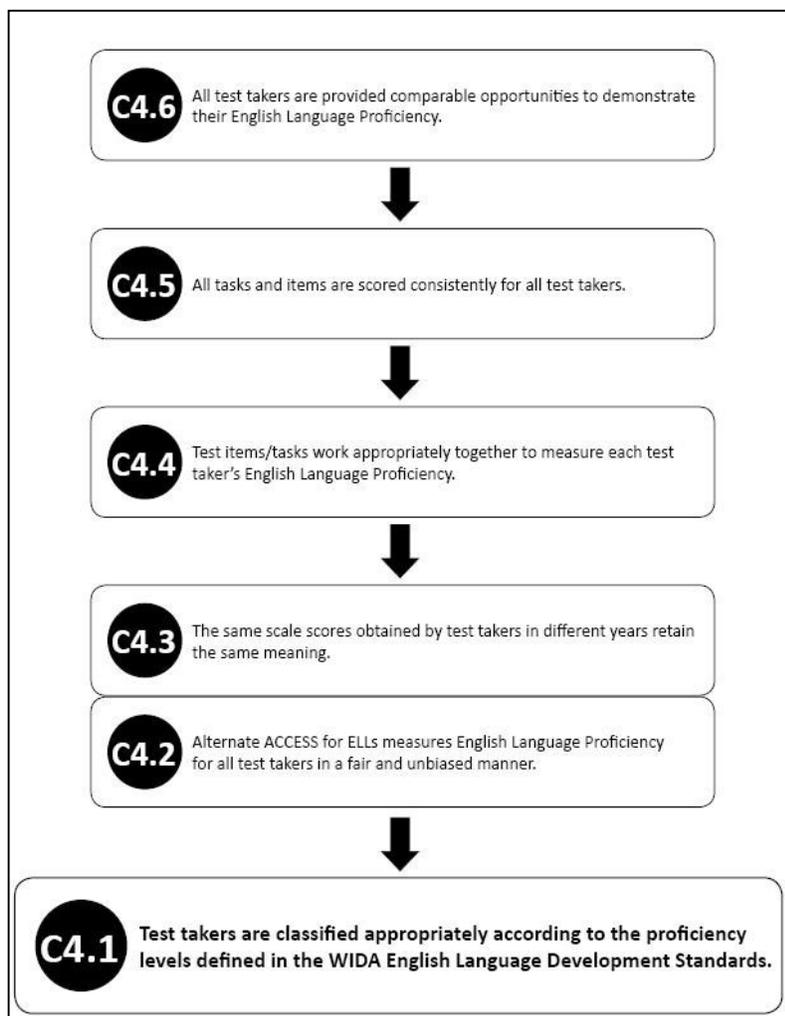


Figure 2.2.1A: Progression of Claims for Step 4: Assessment Records

2.3 Evidence for Assessment Records Claims of Alternate ACCESS for ELLs

In this section, evidence in the form of data or other sources (e.g., Test Administration Manuals, the technical brief of the Alternate ACCESS for ELLs standard setting study, the technical brief of the Alternate ACCESS for ELLs Series 100 development and operational field Test, and other information within this report, etc.) is connected to each of the *Assessment Records* claims via the actions taken to ensure those claims. This section denotes the tables, figures, and external sources that provide evidence related to each action. A summary table of the information presented in this section, including hyperlinks to the detailed description of each table or figure in Chapter 5 of this Annual Technical Report, is contained in Section 2.4. Information on how to navigate the tables and figures throughout this report is presented in Section 2.5.

Because these claims relate to Step 4 of the overall validation framework, their numbering begins with 4. The second number (after the decimal) denotes the level of the claim within Step 4. This numbering system is used in anticipation of the development of more complete documentation of a validity argument for Alternate ACCESS for ELLs, which will be completed by WIDA. Individual actions to ensure each claim are denoted by the final letter (a, b, c, and so on).

Claim 4.6 - All test takers are provided comparable opportunities to demonstrate their English Language Proficiency.

Action 4.6.a: The students that take Alternate ACCESS for ELLs have been identified as English language learners and participate in an alternate curriculum that aligns with the test.

Evidence: Exclusionary criteria and participation guidelines are closely followed by local test administrators (see Table 4.10.1 Participation by Disability, S403).

Action 4.6b: All test takers are given equal opportunities to demonstrate their English language proficiency.

Evidence: The Test Administration Manual provides clear guidance on the use of supporting features of Alternate ACCESS for ELLs, including repetition of questions, availability of cues, etc. (WIDA, 2013). If necessary, further accommodations for test takers are taken following the principles in the test administration manual.

Action 4.6c: Well-specified procedures were developed for test administrators so that they are able to administer the test consistently.

Evidence: Procedures for administering the test, stopping the test, and producing reported scores are documented in the Alternate ACCESS for ELLs Test Administration Manual (WIDA, 2013).

Action 4.6d: Test administrators document and report any irregularities that may occur so that appropriate action may be taken.

Evidence: Alternate ACCESS student response booklets contain a section for reporting irregular cases, such as invalid administration, absent student, or declined assessment. Test administration procedures are documented in the Alternate ACCESS for ELLs Test Administration Manual (WIDA, 2013).

Claim 4.5 – All items and tasks are scored consistently for all test takers.

Action 4.5a: A clear scoring design facilitates the task rating process for Test Administrators.

Evidence: The scoring procedures are clearly stated in the test administrator’s script and the Student Response Booklet is designed to match the scoring procedures and to avoid any scoring ambiguity.

Action 4.5b: Test Administrators undergo training so that they know how to score appropriately.

Evidence: Section 1.6 of this report specifies the scoring procedure for Alternate ACCESS for ELLs. Since all sections of Alternate ACCESS are scored locally, Test Administrators are provided with adequate training materials through an online program on the WIDA website to make sure they follow the test administration script and scoring rubrics for the Speaking and

Writing sections. The scoring rubrics for Speaking and Writing are in the Test Administration Manual (WIDA, 2013).

Claim 4.4 - Test items/tasks work appropriately together to measure each test taker’s English Language Proficiency.

Action 4.4a: For each *test form* (e.g., Reading 6–8), item and task analyses are performed and psychometric properties of the items and tasks are evaluated to confirm that scores are internally

consistent.

Evidence: Reliability information based on Classical Test Theory is calculated for each test form. This information includes Cronbach's alpha, which is a measure of internal consistency. Cronbach's coefficient alpha is widely used as an estimate of reliability and expresses how well the items on a test appear to work together to measure the same construct (see Table 6E).

Action 4.4b: For each *domain and composite score*, item and task analyses are performed and psychometric properties of the items and tasks are evaluated to confirm that scores are internally consistent.

Evidence: A single reliability estimate, a stratified Cronbach's alpha (Cronbach, Schonemann, & McKie, 1965), is calculated by grade-level cluster for each domain and composite score. Cronbach's alpha indicates the extent to which test items are consistent with each other. The stratified Cronbach's alpha is an average reliability, and it is used when test takers are administered several related subtests but are then evaluated based on a composite of those subtest scores. Table 6E presents the data used to calculate an estimate of the reliability of the composite scores using a stratified Cronbach's alpha.

Action 4.4c: Analyses of Rasch model fit statistics are conducted to show that individual tasks perform appropriately.

Evidence: The Complete Items Analysis table includes information on the Rasch fit statistics for each test item (see Table 6G). These statistics, called outfit mean square and infit mean square statistics, measure how well an item is measuring the same construct as other items on the test. Infit and outfit statistics indicate any consistently unusual performance in relation to the item's difficulty measure by measuring the degree to which examinees' responses to items deviate from expected responses. Both statistics have an expected value of 1.0. Items with infit and outfit mean square statistics between 0.5 and 1.5 are considered "productive for measurement" (Linacre, 2002). Values between 1.5 and 2.0 are "unproductive for construction of measurement, but not degrading." Values greater than 2.0 might "distort or degrade the measurement system." Values below 0.5 are "less productive for measurement, but not degrading." Infit helps ensure that test takers within range of the targeted proficiency level perform as expected. It is not as sensitive to outliers as Outfit. Outfit can be skewed if test takers with extreme (i.e., high-level or low-level) proficiency do not perform as expected. High infit is a bigger threat to validity, but is more difficult to explain than high outfit (Linacre, 2002). The infit and outfit mean square statistics are part of the evaluation criteria used to select the items and tasks that appear on the final operational forms. Alternate ACCESS for ELLs test items with infit or outfit values between 1.2 and 1.3 are reviewed and items with values greater than 1.3 are not used on operational forms of the test.

Claim 4.3 - The same scale scores obtained by test takers in different years retain the same meaning.

Action 4.3a: All test items and tasks have been field tested and anchored using items from the operational field test (Series 100) to maintain a consistent scale from year to year.

Evidence: These retained "anchor items" ensure that performances on the newer form may be interpreted in the same frame of reference as the previous year. Table 6G displays information on the anchor items for each test form.

Action 4.3b: The same scaling equation is applied from year to year to ensure that scale scores are obtained consistently over time.

Evidence: The scaling equation table is used to convert a test taker’s ability measure, which is calculated based on test performance using Rasch modeling, into an Alternate ACCESS for ELLs scale score (see Table 6H). The same equation is used across grade-level clusters within each domain.

Claim 4.2 – Alternate ACCESS for ELLs measures English Language Proficiency for all test takers in a fair and unbiased manner.

Action 4.2a: Differential Item Functioning (DIF) analyses are conducted to determine whether any items or tasks may be biased against certain subgroups in terms of gender and ethnicity.

Evidence: The Item Analysis Summary provides a summary of the findings of the differential item functioning (DIF) analyses, which look for measurement bias in test items (see Table 6F). Analyses search for bias in contrasting groups based on gender (male versus female) and ethnicity (Hispanic versus non-Hispanic). This table shows the number of items that favored one group or the other at all levels of DIF.

The Complete Items Analysis table includes more detailed information on the DIF analyses, showing the degree of measurement bias for each item and which group is favored (Table 6G). Each item is categorized into three levels of DIF: A, B, or C (Zieky, 1993). An item exhibiting A level DIF shows little or no evidence of bias toward a particular group, an item exhibiting B level DIF displays a moderate amount of bias, and an item exhibiting C level DIF is considered to display considerable evidence for potential bias and should be closely examined by test developers to identify any construct irrelevant factors that may contribute to DIF.

Action 4.2b: Items that show evidence of DIF are carefully reviewed so that any that indicate bias are not used for scoring and are removed from future test forms.

Evidence: As described in Chapter 5.1.4 (DIF Items), ethnicity and gender DIF analyses are conducted using all test taker data.

Claim 4.1 - Test takers are classified appropriately according to the Alternate proficiency levels defined in the WIDA English Language Development Standards.

Action 4.1a: Distributions of scale scores and proficiency levels for each domain are analyzed to confirm that Alternate ACCESS for ELLs effectively measures the performance of test takers across the range of Alternate English Language Proficiency levels as defined by the WIDA English Language Development (ELD) Standards.

Evidence: The distribution of test takers’ raw scores on Alternate ACCESS for ELLs, organized by individual test form (e.g., Reading 3–5), shows the extent to which Alternate ACCESS for ELLs effectively measures the performance of test takers across the range of ELD abilities that each form was designed to assess (see Table 6A; see Figure 6A).

The distribution of test takers’ scale scores on Alternate ACCESS for ELLs, organized by test form (e.g., Reading 3–5), shows that Alternate ACCESS for ELLs effectively measures the performance of test takers across the range of ELD abilities that each form was designed to assess (see Table 6B; see Figure 6B).

The proficiency level distribution of test takers' scores on Alternate ACCESS for ELLs, organized by individual test form (e.g., Reading 3–5), shows that Alternate ACCESS for ELLs effectively measures the performance of test takers across the range of proficiency levels that each form was designed to assess (see Table 6C; see Figure 6C).

The Raw Score to Proficiency Level Score table shows the interpretive proficiency level score associated with each raw score (see Table 6I). This distribution of scores shows that Alternate ACCESS for ELLs effectively measures the performance of test takers across the range of proficiency levels that each form was designed to assess.

The Test Characteristic Curve for each test form graphically shows the relationship between test takers' ability measure (which is calculated based on test performance using Rasch modeling) on the horizontal axis and the expected raw scores on the vertical axis (see Figure 6D). Four vertical lines indicate the four cut scores for the highest grade in the cluster, dividing the figure into five sections for each of the five WIDA proficiency levels. The curve shows that higher expected raw scores are required to be placed into higher language proficiency levels.

Action 4.1b: Distributions of scale scores and proficiency levels, organized by grade-level cluster, are analyzed to confirm that Alternate ACCESS for ELLs effectively measures the performance of test takers across the range of Alternate English Language Proficiency levels as defined by the WIDA ELD Standards.

Evidence: The distribution of test takers' scale scores on Alternate ACCESS for ELLs, organized by grade-level cluster, shows that Alternate ACCESS for ELLs effectively measures the performance of test takers across the range of abilities as described by the WIDA ELD Standards (see Table 6B; see Figure 6B).

The proficiency level distribution of test takers' scores on Alternate ACCESS for ELLs, organized by grade-level cluster, shows that Alternate ACCESS for ELLs effectively measures the performance of test takers across the range of Alternate proficiency levels as defined by the WIDA ELD Standards (see Table 6C; see Figure 6C).

The Test Characteristic Curve reflects test takers' mean raw scores by domain on Alternate ACCESS for ELLs across the entire test for each grade-level cluster (except for the Kindergarten level) (see Figure 6D).

Action 4.1c: For each test form, analyses are run to confirm that English Language Proficiency is measured with high precision at the cut points.

Evidence: The Test Information Function graphically shows how well the test is measuring across the ability measure spectrum, which is calculated based on test performance using Rasch modeling (see Figure 6E). High values indicate more accuracy in measurement.

In the Raw Score to Proficiency Level Conversion Chart, the proficiency level associated with each raw score shows the distribution of proficiency level scores associated with each raw score for each grade in the cluster, along with the percentage of test takers in that grade who scored at that raw score/proficiency level score (see Table 6I). The Raw Score to Scale Score Conversion Chart (Table 6H) presents the conditional standard error for each scale score, along with the upper and lower bound of the scale scores within this standard error of measurement. This value indicates how accurately or precisely the test is measuring test takers at a particular ability level by estimating the error measurement at each score point. Because there is usually more information about test takers with scores in the middle of the score distribution on each form, the conditional

standard error values are usually smallest and scores are more reliable in that region of the score distribution.

Action 4.1d: Classification and accuracy analyses are conducted by grade level to confirm that proficiency level classifications are reliable for all domain and composite scores.

Evidence: Information related to the accuracy of test takers' proficiency-level classifications is presented in multiple ways (see Table 6J). A separate table is provided for each grade level in a cluster. The table provides overall indices related to the accuracy and consistency of classification. These indices indicate the percent of all test takers who would be classified into the same language proficiency level by both the administered test and either the true score distribution (accuracy) or a parallel test (consistency). Cohen's kappa, which is a statistical measure of interrater agreement between two raters that takes chance agreement between raters into account, is also presented. A kappa value of 1 indicates complete agreement between the two raters, while a kappa value of 0 indicates no agreement other than what would be expected by chance. Table 6J also shows accuracy and consistency information conditional on level and provides indices of classification accuracy and consistency at the cut points.

2.4 Summary of Assessment Records Claims, Actions, and Evidence

Table 2.4A

Summary of Assessment Records Claims, Actions, and Evidence

Claim	Actions	Evidence
6. All test takers are provided comparable opportunities to demonstrate their English Language Proficiency	a. The students that take Alternate ACCESS have been identified as English language learners and participate in an alternate curriculum that aligns with the test.	a. Test Administration Manual Table 4.10.1 (<i>Participation by Disability</i>)
	b. All test takers are given supported opportunities to demonstrate their English language proficiency.	b. Test Administration Manual
	c. Well-specified procedures were developed for test administrators so that they are able to administer the test consistently.	c. Test Administration Manual
	d. Test administrators document and report any irregularities that may occur so that appropriate action may be taken	d. Test Administration Manual
5. All items and tasks are scored consistently for all test takers.	a. A clear scoring design facilitates the task rating process for Test Administrators.	a. Test Administration Manual; Student Response Booklets
	b. Raters of performance-based tasks undergo thorough training so that they know how to score appropriately.	b. Chapter 1.6
4. Test items/tasks work appropriately together to measure each test taker's English Language Proficiency.	a. For each test form (e.g., Reading 6-8), item and task analyses are performed and psychometric properties of the items and tasks are evaluated to confirm that scores are internally consistent.	a. Table 6E (<i>Reliability</i>)
	b. For each domain and composite score, item and task analyses are performed and psychometric properties of the items and tasks are evaluated to confirm that scores are internally consistent.	b. Table 6E (<i>Reliability</i>)
	c. Analyses of Rasch model fit statistics are conducted to show that individual tasks perform appropriately.	c. Table 6G (<i>Complete Item Analysis</i>)
3. The same scale scores obtained by test takers in	a. All the items and tasks have been field tested and are used as anchor items from the operational field test (Series 100) to maintain a consistent scale from year to year.	a. Table 6D (<i>Equating Summary</i>)

different years retain the same meaning.	b. The same scaling equation is applied from year to year to ensure that scale scores are obtained consistently over time.	b. Table 6H (Raw Score to Scale Score Conversation Chart)
2. Alternate ACCESS for ELLs measures English Language Proficiency for all test takers in a fair and unbiased manner.	<p>a. Differential Item Functioning (DIF) analyses are conducted to determine whether any items or tasks are biased against certain subgroups in terms of gender and ethnicity.</p> <p>b. Items that show evidence of DIF are carefully reviewed so that any that indicate bias are not used for scoring and are removed from future test forms.</p>	<p>a. Table 6F (<i>Item Analysis Summary</i>); Table 6G (<i>Complete Item Analysis</i>)</p> <p>b. Chapter 5.1.4 (<i>DIF Items</i>)</p>
1. Test takers are classified appropriately according to the Alternate proficiency levels defined in the WIDA English Language Development (ELD) Standards.	<p>a. Distributions of scale scores and proficiency levels for each domain are analyzed to confirm that Alternate ACCESS for ELLs effectively measures the performance of test takers across the range of Alternate English Language Proficiency levels as defined by the WIDA ELD Standards.</p> <p>b. Distributions of scale scores and proficiency levels, organized by grade-level cluster, are analyzed to confirm that Alternate ACCESS for ELLs effectively measures the performance of test takers across the range of Alternate English Language Proficiency levels as defined by the WIDA ELD Standards</p> <p>c. For each test form, analyses are run to confirm that English Language Proficiency is measured with high precision at the pertinent cutpoints.</p> <p>d. Classification and accuracy analyses are conducted by grade-level to confirm that proficiency level classifications are reliable for all domain and composite scores.</p>	<p>a. Figure 6A (<i>Raw Scores</i>) & Table 6A (<i>Raw Score Descriptive Statistics</i>); Figure 6B (<i>Scale Scores</i>) & Table 6B (<i>Scale Score Descriptive Statistics</i>); Figure 6C (<i>Proficiency Level</i>) & Table 6C (<i>Proficiency Level Distribution</i>); Table 6I (<i>Raw Score to Proficiency Level Score Conversion Chart</i>); Figure 6D (<i>Test Characteristic Curve</i>)</p> <p>b. Figure 6B (<i>Scale Scores</i>) & Table 6B (<i>Scale Score Descriptive Statistics</i>); Figure 6C (<i>Proficiency Level</i>) & Table 6C (<i>Proficiency Level Distribution</i>); Figure 6D (<i>Test Characteristic Curve</i>)</p> <p>c. Figure 6E (<i>Test Information Function</i>); Table 6H (<i>Raw Score to Scale Score Conversion Chart</i>)</p> <p>d. Table 6J (<i>Accuracy and Consistency of Classification Indices</i>)</p>

2.5 Visual Guide to Tables and Figures

This section provides navigational support for the tables and figures contained in the Alternate ACCESS for ELLs Annual Technical Report. The Visual Guide to Tables and Figures, shown in Figures 2.5.1 and 2.5.2, serves as a resource to quickly identify which table and/or figure to look for when seeking specific information based on grade, grade-level cluster, and demographic characteristics, such as state, gender, disability type, and ethnicity and race, as well as domains and domain composites.

To use the Visual Guide to Tables and Figures as a navigational tool, click on the links in Figures 2.5.1 through 2.5.3 to navigate to the selected tables and figures in the Annual Technical Report. A

link is provided at the end of each section in Chapters 4 and 6. Detailed descriptions of the information in each of the tables and figures is included in the preceding chapters (e.g., Chapter 5 contains information on tables and figures in Chapter 6). These descriptions may be accessed through links in Table 2.4A *Summary of Assessment Records Claims, Actions, and Evidence*.

Figure 2.5.1 displays the tables in Chapter 4 that provide information on participation, scale score, and proficiency level results, as well as results by standard. The key in the upper left corner of the figure describes the tables contained in each section of the chapter. For example, tables in Section 4.1 contain information about participation. To find specific information in Chapter 4, select the Grade or Grade Cluster tab, and then the Domain tab, and then choose from three categories: Demographic Characteristics, Domain Composites, or Domains. Within each of these categories, several additional options organize information so that individual tables can be accessed. For example, to find a table that displays information on the number of female Grade 2 students who completed the Speaking section, refer to Figure 2.5.1 and complete the following steps: one, select Grade; two, select Domains; three, select Demographic Characteristics; four, select Gender. The information is found in Table 4.2.2.2. Click on 4.2.2.2 to go to the appropriate table in Chapter 4.

Figure 2.5.2 displays the sections in Chapter 6 that contains analyses for each Alternate ACCESS for ELLs test form by grade-level cluster and domain. The key above the figure describes specific information in each table and figure. For example, to find the Reliability table for Grade-level Cluster 9–12 in the Reading domain, refer to Figure 2.5.2 and complete the following steps: one, select Grade Cluster 9–12; two, select ; three, select Reading under Domains. Information for 9–12 Reading is shown in section 6.5.2.3. Finally, look at the key that explains that reliability information is located in table F. The result is Table 6.5.2.3F. Click on 6.5.2.3 to go to the appropriate section, and then locate Table F.

2.5.1 Chapter 4 Visual Guide to Tables and Figures

		Test Form Characteristics			
		Grade		Grade-Level Cluster	
		Domain		Domain	
Demographic Characteristics	State		4.3.1		4.2.1
	Gender	4.7.1	4.3.2	4.6.1	4.2.2
	Ethnicity and Race	4.7.2	4.3.3	4.6.2	4.2.3
Domain Composites	Overall		4.9.2 H		4.9.1 H
	Oral Language		4.9.2 E		4.9.1 E
	Literacy		4.9.2 F		4.9.1 F
	Comprehension		4.9.2 G		4.9.1 G
Domains	Across All Domains		4.4.2 4.5.1		4.4.1 4.8 4.5.1
	Listening		4.9.2 A		4.9.1 A
	Reading		4.9.2 B		4.9.1 B
	Writing		4.9.2 D		4.9.1 D
	Speaking		4.9.2 C		4.9.1 C

Figure 2.5.1 Chapter 4 Visual Guide to Tables and Figures

2.5.2 Chapter 6 Visual Guide to Tables and Figures

Table A and Figure A	<i>Raw Score Descriptive Statistics</i>
Table B and Figure B	<i>Scale Score Descriptive Statistics</i>
Table C and Figure C	<i>Proficiency Level Distribution</i>
Table D	<i>Equating Summary</i>
Figure D	<i>Test Characteristic Curve</i>
Table E	<i>Reliability</i>
Figure E	<i>Test Information Function</i>
Table F	<i>Item Analysis Summary</i>
Table G	<i>Complete Item Analysis</i>
Table H	<i>Raw Score to Scale Score Conversion</i>
Table I	<i>Raw Score to Proficiency Level Conversion</i>
Table J	<i>Accuracy and Consistency of Classification Indices</i>

		Grade-Level Cluster			
		1-2	3-5	6-8	9-12
Domain Composites	Overall	6.1.8	6.2.8	6.3.8	6.4.8
	Oral Language	6.1.5	6.2.5	6.3.5	6.4.5
	Literacy	6.1.6	6.2.6	6.3.6	6.4.6
	Comprehension	6.1.7	6.2.7	6.3.7	6.4.7
Domains	Listening	6.1.1	6.2.1	6.3.1	6.4.1
	Reading	6.1.2	6.2.2	6.3.2	6.4.2
	Writing	6.1.4	6.2.4	6.3.4	6.4.4
	Speaking	6.1.3	6.2.3	6.3.3	6.4.3

Figure 2.5.2 Chapter 6 Visual Guide to Tables and Figures

3. Descriptions of Student Results

Chapter 3 provides a description of the Chapter 4 tables summarizing students' participation, scale scores, and proficiency levels; results are further subdivided by grade, grade-level cluster, state, domain, domain and composite scores, gender, ethnicity/race, and disability. Of the 39 WIDA Consortium states, 38 participated in the 2018-2019 Alternate ACCESS operational administration.

3.1 Participation

Table 4.1.1—Students Excluded from Analysis

In some circumstances there was a mismatch between a student's reported grade and the grade-level cluster (i.e., 1-2, 3-5, 6-8, or 9-12) actually administered (e.g., a student reported to be in Grade 1 who was administered a test intended for students in the 3-5 grade-level cluster). In all, 32 students were administered a test form not intended for their grade-level cluster. See Table 4.1.1 for a breakdown of the incorrect test forms assigned, by grade. The data from these 32 students were eliminated from all subsequent analyses in this report.

Section 4.2—Grade-Level Cluster, Gender, Ethnicity

Section 4.2 provides a breakdown of participation by *grade-level cluster* as a function of state (Table 4.2.1), gender (Table 4.2.2) and ethnicity (Table 4.2.3). For each of the 38 WIDA states who participated in the 2018-2019 operational testing program, Table 4.2.1 provides the number of test takers by grade-level cluster as well as total counts by state (final column) and grade-level cluster across all states (final row). For each grade-level cluster, Table 4.2.2 provides the distribution of test takers by gender (Female, Male, or Missing). Table 4.2.3 provides a similar breakdown of grade-level cluster by ethnicity (Hispanic or Non-Hispanic).

Section 4.3—Grade, Gender, Ethnicity

Section 4.3 duplicates the information provided by Section 4.2, but further breaks down the distribution of test takers by *grade* (Grades 1 to 12), instead of grade-level cluster. For each state, Table 4.3.1 provides the distribution of test takers by grade; for each grade, Table 4.3.2 provides the distribution of test takers by gender; for each grade, Table 4.3.3 provides the distribution of test takers by ethnicity.

Section 4.4—Domain, Grade-Level Cluster, Grade

Section 4.4 provides a breakdown of test taker counts by *domain* (Listening, Reading, Speaking, and Writing), with Table 4.4.1 summarizing the distribution by grade-level cluster and Table 4.4.2 summarizing the distribution by grade.

3.2 Scale Score Results

3.2.1 Mean Scale Scores Across Domain and Composite Scores

Overview of Sections 4.5 – 4.7

Sections 4.5 through 4.7 display the mean scale scores (Mean), standard deviation (Std. Dev.) and counts (N) by *grade and/or grade-level cluster* across the eight scores awarded on Alternate ACCESS for ELLs, first for each of the four domains (Listening, Reading, Speaking, and Writing) and then for each of the four composites (Oral Language, Literacy, Comprehension, and Overall). Sections 4.6 and 4.7 include gender and ethnicity information.

Section 4.5–Grade and Grade-Level Cluster

For each of the four grade-level clusters, Tables 4.5.1A through 4.5.1D display the mean scale scores for each domain and composite — first separately by grades within each cluster and then by the grade-level cluster overall (as the final column).

Section 4.6–Grade-Level Cluster, Gender, Ethnicity and Race

For each of the four grade-level clusters, Tables 4.6.1A through 4.6.1D display the mean scale scores for each domain and composite by gender. Correspondingly, Tables 4.6.2A through 4.6.2.D provide the mean scale score information by ethnicity and race. (Note that for the 4.6.1 Table series Domain is the row variable, and for the 4.6.2 table series Domain is the column variable.)

Section 4.7–Grade, Gender, Ethnicity and Race

For each of the 12 grades, Tables 4.7.1A through 4.7.1L display the mean scale scores for each domain and composite. Correspondingly, Tables 4.7.2.A through 4.7.2L display the mean scale scores by ethnicity and race.

3.2.2 Correlations

For each of the four grade-level clusters, Tables 4.8.1 through 4.8.4 display the Pearson correlations between scale scores on the four domains.

3.3 Proficiency Level Results

Section 3.3, Proficiency Level Results, displays the distribution of students' language proficiency level³ by grade-level cluster (Tables 4.9.1A-H) and grade (Tables 4.9.2A-H), with each sub-table presenting results by domain/composite:

A – Listening

B – Reading

³ The WIDA Alternate ELD Standards has six levels (A1-A3; P1; P2; P3). P3 was not part of the current analysis.

C – Speaking

D – Writing

E – Oral Language Composite

F – Literacy Composite

G – Comprehension Composite

H – Overall Composite

3.4 Participation by Disability

Table 4.10.1 displays the distribution of test takers as function of primary and secondary disability, each with 15 categories:

- No Primary Disability recorded (NPD)
- No Secondary Disability recorded (SPD)
- Autism Spectrum Disorder (AS)
- Deaf-blindness (DB)
- Developmental Delay (DD)
- Hearing Impairment, including Deafness (HI)
- Infant/Toddler with a Disability (ITD)
- Intellectual Disability (ID)
- Multiple Disability (MD)
- Orthopedic Impairment (OI)
- Other Health Impairment (OHI)
- Serious Emotional Disability (SED)
- Specific Learning Disability (SLD)
- Speech or Language Impairment (SLI)
- Traumatic Brain Injury (TBI)
- Visual Impairment, including Blindness (VI)

The accompanying *Acronyms for Table 4.10.1* table matches each disability category with its acronym to aid in interpretation.

4. Student Results

4.1 Students excluded from Analysis

4.1.1 Out-of-grade-level Test Administration

Table 4.1.1

Out-of-grade-level Test Administrations

Grade					Total
	1-2	3-5	6-8	9-12	
1		0	0	0	0
2		2	0	0	2
3	7		0	0	7
4	2		0	0	2
5	3		1	1	5
6	1	4		0	5
7	1	0		1	2
8	1	0		4	5
9	1	0	0		1
10	0	1	0		1
11	0	0	1		1
12	0	0	1		1
Total	16	7	3	6	32

4.2 Participation by Grade-level Cluster

4.2.1 Participation by Grade-level Cluster by State

Table 4.2.1

Participation by Cluster by State

State	Cluster				Total
	1-2	3-5	6-8	9-12	
AK	9	18	19	42	88
AL	70	114	83	76	343
BI	3	9	11	4	27
CO	154	324	301	370	1,149
DC	20	21	20	9	70
DE	12	15	12	8	47
FL	385	424	234	207	1,250
GA	243	435	368	242	1,288
HI	46	68	74	68	256
ID	37	78	59	47	221
IL	808	1,109	955	1,121	3,993
IN	149	271	286	329	1,035
KY	41	82	89	83	295
MA	301	472	326	385	1,484
MD	117	148	169	130	564
ME	13	17	14	14	58
MI	168	266	258	303	995
MN	191	307	227	225	950
MO	47	76	53	47	223
MT	3	8	5	3	19
NC	268	454	326	369	1,417
ND	6	7	12	10	35
NH	5	9	9	6	29
NJ	132	150	80	58	420
NM	86	147	136	134	503
NV	122	242	300	317	981
OK	134	263	164	158	719
PA	238	415	292	350	1,295
RI	22	55	34	42	153
SC	78	115	90	79	362
SD	10	11	15	25	61
TN	58	147	122	92	419
UT	68	126	114	154	462
VA	421	591	501	534	2,047
VI	1	1	1	0	3
VT	8	1	4	8	21
WI	86	139	147	216	588
WY	12	8	18	11	49
Total	4,572	7,143	5,928	6,276	23,919

4.2.2 Participation by Grade-level Cluster by Gender

Table 4.2.2

Participation by Cluster by Gender

Cluster	Gender						Total
	Female		Male		Missing		
	Count	% within Cluster	Count	% within Cluster	Count	% within Cluster	
1-2	1,367	29.90	3,155	69.01	50	1.09	4,572
3-5	2,343	32.80	4,762	66.67	38	0.49	7,143
6-8	2,108	35.56	3,771	63.61	49	0.83	5,928
9-12	2,376	37.86	3,843	61.23	57	0.91	6,276
Total	8,194	34.26	15,531	64.93	194	0.81	23,919

4.2.3 Participation by Grade-level Cluster by Ethnicity

Table 4.2.3

Participation by Cluster by Ethnicity

Cluster	Hispanic/Non-Hispanic						Total
	Hispanic		Non-Hispanic		Missing		
	Count	% within Cluster	Count	% within Cluster	Count	% within Cluster	
1-2	2,723	59.56	1,540	33.68	309	6.76	4,572
3-5	4,538	63.53	2,111	29.55	494	6.92	7,143
6-8	3,887	65.57	1,671	28.19	370	6.24	5,928
9-12	3,952	62.97	1,871	29.81	453	7.22	6,276
Total	15,100	63.13	7,193	30.07	1,626	6.80	23,919

4.3 Participation by Grade

4.3.1 Participation by Grade by State

Table 4.3.1

Participation by Grade by State

State	Grade												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
AK	2	7	6	2	10	6	7	6	6	10	14	12	88
AL	37	33	31	35	48	38	27	18	20	17	16	23	343
BI	0	3	1	4	4	2	6	3	0	3	1	0	27
CO	83	71	104	113	107	96	111	94	90	85	77	118	1,149
DC	14	6	7	8	6	7	3	10	4	2	3	0	70
DE	8	4	3	8	4	1	7	4	5	2	0	1	47
FL	194	191	167	141	116	99	61	74	49	39	44	75	1,250
GA	113	130	145	147	143	143	124	101	89	63	41	49	1,288
HI	16	30	28	20	20	35	20	19	12	19	12	25	256
ID	13	24	28	20	30	26	16	17	21	12	8	6	221
IL	423	385	381	378	350	352	306	297	225	236	226	434	3,993
IN	80	69	79	95	97	87	116	83	82	69	77	101	1,035
KY	18	23	31	24	27	27	38	24	24	26	18	15	295
MA	143	158	163	149	160	115	109	102	91	100	86	108	1,484
MD	63	54	51	46	51	63	57	49	32	31	38	29	564
ME	7	6	5	3	9	4	6	4	5	3	3	3	58
MI	86	82	91	89	86	90	80	88	88	81	66	68	995
MN	96	95	114	108	85	86	60	81	58	59	44	64	950
MO	27	20	34	21	21	25	16	12	15	11	5	16	223
MT	1	2	3	3	2	3	1	1	1	0	2	0	19
NC	110	158	159	157	138	104	111	111	93	77	75	124	1,417
ND	3	3	1	2	4	6	3	3	2	3	4	1	35
NH	3	2	3	1	5	3	5	1	1	3	0	2	29
NJ	70	62	59	51	40	24	26	30	16	23	14	5	420
NM	43	43	45	42	60	58	34	44	30	40	29	35	503
NV	60	62	80	63	99	94	104	102	91	70	73	83	981
OK	58	76	90	86	87	49	58	57	36	42	36	44	719
PA	117	121	159	138	118	88	95	109	94	90	73	93	1,295
RI	8	14	20	11	24	15	11	8	9	12	9	12	153
SC	45	33	37	31	47	37	29	24	17	20	22	20	362
SD	3	7	2	5	4	7	2	6	8	8	1	8	61
TN	27	31	45	43	59	41	43	38	30	22	18	22	419
UT	34	34	40	45	41	31	46	37	39	34	46	35	462
VA	208	213	187	201	203	176	157	168	123	118	98	195	2,047
VI	1	0	0	0	1	1	0	0	0	0	0	0	3
VT	7	1	0	0	1	2	2	0	2	1	3	2	21
WI	38	48	40	50	49	42	60	45	45	58	35	78	588
WY	7	5	1	5	2	4	5	9	4	4	0	3	49
Total	2,266	2,306	2,440	2,345	2,358	2,087	1,962	1,879	1,557	1,493	1,317	1,909	23,919

4.3.2 Participation by Grade by Gender

Table 4.3.2

Participation by Grade by Gender

Cluster	Gender						Total
	Female		Male		Missing		
	Count	% within Cluster	Count	% within Cluster	Count	% within Cluster	
1	655	28.91	1,581	69.77	30	1.32	2,266
2	712	30.88	1,574	68.26	20	0.87	2,306
3	754	30.90	1,669	68.40	17	0.70	2,440
4	766	32.67	1,567	66.82	12	0.51	2,345
5	823	34.90	1,526	64.72	9	0.38	2,358
6	726	34.79	1,330	63.73	31	1.49	2,087
7	715	36.44	1,237	63.05	10	0.51	1,962
8	667	35.50	1,204	64.08	8	0.43	1,879
9	555	35.65	973	62.49	29	1.86	1,557
10	573	38.38	910	60.95	10	0.67	1,493
11	483	36.67	825	62.64	9	0.68	1,317
12	765	40.07	1,135	59.46	9	0.47	1,909
Total	8,194	34.26	15,531	64.93	194	0.81	23,919

4.3.3 Participation by Grade by Ethnicity

Table 4.3.3

Participation by Grade by Ethnicity

Cluster	Hispanic/Non-Hispanic						Total
	Hispanic		Non-Hispanic		Missing		
	Count	% within Cluster	Count	% within Cluster	Count	% within Cluster	
1	1,318	58.16	785	34.64	163	7.19	2,266
2	1,405	60.93	755	32.74	146	6.33	2,306
3	1,460	59.84	789	32.34	191	7.83	2,440
4	1,504	64.14	684	29.17	157	6.70	2,345
5	1,574	66.75	638	27.06	146	6.19	2,358
6	1,386	66.41	569	27.26	132	6.32	2,087
7	1,281	65.29	570	29.05	111	5.66	1,962
8	1,220	64.93	532	28.31	127	6.76	1,879
9	955	61.34	474	30.44	128	8.22	1,557
10	930	62.29	450	30.14	113	7.57	1,493
11	857	65.07	364	27.64	96	7.29	1,317
12	1,210	63.38	583	30.54	116	6.08	1,909
Total	15,100	63.13	7,193	30.07	1,626	6.80	23,919

4.4 Participation by Domain

4.4.1 Participation by Grade-level Cluster by Domain

Table 4.4.1
Participation by Cluster by Domain

Cluster	Domain			
	Listening	Reading	Speaking	Writing
1-2	4,532	4,535	4,532	4,565
3-5	7,070	7,079	7,080	7,101
6-8	5,846	5,851	5,851	5,859
9-12	6,204	6,207	6,205	6,206
Total	23,652	23,672	23,668	23,731

4.4.2 Participation by Grade by Domain

Table 4.4.2
Participation by Grade by Domain

Grade	Domain			
	Listening	Reading	Speaking	Writing
1	2,243	2,243	2,244	2,262
2	2,289	2,292	2,288	2,303
3	2,405	2,408	2,411	2,429
4	2,325	2,330	2,330	2,331
5	2,340	2,341	2,339	2,341
6	2,061	2,062	2,063	2,065
7	1,933	1,934	1,932	1,935
8	1,852	1,855	1,856	1,859
9	1,531	1,532	1,530	1,532
10	1,472	1,474	1,476	1,475
11	1,301	1,301	1,301	1,301
12	1,900	1,900	1,898	1,898
Total	23,652	23,672	23,668	23,731

4.5 Scale Scores by Domain and Composite

4.5.1 Mean Scale Scores by Domain and Composite

Table 4.5.1 A
Mean Scale Scores: 1-2

	Grade 1			Grade 1			Cluster 1-2		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	931.56	11.00	2,243	934.18	10.33	2,289	932.88	10.74	4,532
Reading	931.50	12.62	2,243	935.17	12.86	2,292	933.35	12.87	4,535
Speaking	932.46	14.58	2,244	935.01	14.17	2,288	933.75	14.43	4,532
Writing	926.99	11.05	2,262	929.61	11.51	2,303	928.31	11.36	4,565
Oral	932.26	12.13	2,242	934.87	11.60	2,285	933.58	11.94	4,527
Literacy	929.52	11.02	2,243	932.70	11.45	2,291	931.13	11.35	4,534
Comprehension	931.58	11.78	2,243	934.96	11.77	2,289	933.29	11.90	4,532
Overall	930.10	10.93	2,242	933.10	11.08	2,284	931.61	11.11	4,526

Table 4.5.1 B
Mean Scale Scores: 3-5

	Grade 3			Grade 4			Grade 5			Cluster 3-5		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	936.30	10.76	2,405	937.79	10.36	2,325	938.15	10.37	2,340	937.40	10.53	7,070
Reading	934.78	10.85	2,408	936.75	10.74	2,330	937.17	10.74	2,341	936.22	10.83	7,079
Speaking	935.55	13.43	2,411	936.89	13.13	2,330	937.01	13.28	2,339	936.47	13.30	7,080
Writing	930.75	11.45	2,429	932.91	11.91	2,331	933.58	12.26	2,341	932.39	11.93	7,101
Oral	936.03	11.34	2,404	937.44	11.06	2,323	937.68	11.11	2,337	937.04	11.20	7,064
Literacy	933.06	10.50	2,408	935.11	10.73	2,330	935.65	10.96	2,340	934.59	10.79	7,078
Comprehension	935.26	10.51	2,403	937.07	10.36	2,325	937.48	10.35	2,340	936.59	10.45	7,068
Overall	933.72	10.43	2,402	935.58	10.52	2,323	936.04	10.70	2,337	935.10	10.60	7,062

Table 4.5.1 C
Mean Scale Scores: 6-8

	Grade 6			Grade 7			Grade 8			Cluster 6-8		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	937.56	10.50	2,061	938.29	10.30	1,933	938.99	9.76	1,852	938.25	10.22	5,846
Reading	937.65	11.76	2,062	938.62	11.52	1,934	939.69	11.33	1,855	938.62	11.57	5,851
Speaking	936.83	12.82	2,063	937.13	13.01	1,932	938.07	12.46	1,856	937.32	12.78	5,851
Writing	932.52	10.65	2,065	933.28	10.96	1,935	934.64	10.82	1,859	933.44	10.84	5,859
Oral	937.57	11.19	2,059	938.09	11.13	1,930	938.91	10.63	1,850	938.16	11.01	5,839
Literacy	935.33	10.56	2,062	936.18	10.57	1,934	937.40	10.40	1,855	936.26	10.54	5,851
Comprehension	937.63	11.11	2,059	938.55	10.89	1,933	939.50	10.57	1,852	938.53	10.89	5,844
Overall	935.75	10.42	2,058	936.53	10.39	1,930	937.61	10.11	1,850	936.60	10.34	5,838

Table 4.5.1 D
Mean Scale Scores: 9-12

	Grade 9			Grade 10			Grade 11			Grade 12			Cluster 9-12		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	938.37	10.43	1,531	939.05	9.90	1,472	939.32	9.92	1,301	938.58	10.10	1,900	938.80	10.10	6,204
Reading	938.30	11.03	1,532	938.85	10.45	1,474	938.91	10.68	1,301	938.44	10.75	1,900	938.60	10.74	6,207
Speaking	936.43	12.36	1,530	937.05	11.91	1,476	937.14	11.81	1,301	936.03	12.30	1,898	936.60	12.13	6,205
Writing	934.09	11.01	1,532	934.68	10.93	1,475	935.02	11.00	1,301	934.10	11.26	1,898	934.43	11.07	6,206
Oral	937.55	10.71	1,529	938.20	10.27	1,472	938.37	10.22	1,301	937.48	10.53	1,898	937.86	10.45	6,200
Literacy	936.40	10.38	1,532	936.98	10.09	1,474	937.18	10.23	1,300	936.48	10.44	1,898	936.72	10.30	6,204
Comprehension	938.40	10.63	1,531	938.99	10.11	1,472	939.12	10.27	1,301	938.57	10.35	1,900	938.74	10.35	6,204
Overall	936.59	10.20	1,529	937.16	9.89	1,472	937.38	9.97	1,300	936.61	10.20	1,897	936.90	10.08	6,198

4.6 Scale Scores by Grade-level Cluster

4.6.1 Mean Scale Scores by Gender

Table 4.6.1 A
Mean Scale Scores by Gender: 1-2

	Female			Male			Missing		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	932.85	10.49	1,355	932.85	10.87	3,127	935.44	9.55	50
Reading	933.02	12.09	1,356	933.45	13.20	3,129	936.42	11.48	50
Speaking	933.47	14.33	1,355	933.82	14.48	3,127	936.72	13.46	50
Writing	927.46	10.94	1,364	928.63	11.53	3,151	931.38	10.62	50
Oral	933.41	11.75	1,353	933.61	12.04	3,124	936.38	10.57	50
Literacy	930.53	10.75	1,356	931.34	11.60	3,128	934.20	10.51	50
Comprehension	933.06	11.26	1,355	933.35	12.18	3,127	936.20	10.46	50
Overall	931.13	10.63	1,353	931.78	11.31	3,123	934.50	10.15	50

Table 4.6.1 B
Mean Scale Scores by Gender: 3-5

	Female			Male			Missing		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	937.31	10.34	2,319	937.43	10.63	4,714	938.92	9.46	37
Reading	935.75	10.53	2,320	936.46	10.97	4,722	935.57	10.27	37
Speaking	935.99	13.38	2,320	936.73	13.24	4,723	934.03	14.14	37
Writing	931.42	11.91	2,325	932.89	11.91	4,739	929.43	12.70	37
Oral	936.76	11.12	2,318	937.18	11.23	4,709	936.57	10.61	37
Literacy	933.87	10.66	2,319	934.96	10.83	4,722	932.78	10.65	37
Comprehension	936.24	10.20	2,319	936.76	10.57	4,712	936.65	9.74	37
Overall	934.51	10.48	2,318	935.40	10.64	4,707	933.68	10.20	37

Table 4.6.1 C
Mean Scale Scores by Gender: 6-8

	Female			Male			Missing		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	938.41	10.03	2,073	938.14	10.35	3,725	940.35	7.30	48
Reading	938.54	11.28	2,076	938.64	11.75	3,727	940.19	10.07	48
Speaking	937.12	12.77	2,076	937.40	12.81	3,727	939.79	10.13	48
Writing	933.24	11.07	2,078	933.53	10.74	3,733	934.96	8.81	48
Oral	938.14	10.84	2,071	938.15	11.13	3,720	940.48	7.79	48
Literacy	936.13	10.51	2,076	936.32	10.58	3,727	937.79	8.56	48
Comprehension	938.52	10.63	2,072	938.51	11.05	3,724	940.23	8.97	48
Overall	936.50	10.25	2,071	936.63	10.41	3,719	938.40	7.92	48

Table 4.6.1 D
Mean Scale Scores by Gender: 9-12

	Female			Male			Missing		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	938.87	9.91	2,345	938.83	10.16	3,802	933.60	12.41	57
Reading	938.61	10.60	2,346	938.69	10.75	3,804	931.95	13.32	57
Speaking	936.56	11.99	2,346	936.71	12.15	3,802	930.96	14.40	57
Writing	934.32	11.23	2,348	934.56	10.93	3,801	929.84	12.68	57
Oral	937.87	10.32	2,344	937.93	10.47	3,799	932.44	12.97	57
Literacy	936.68	10.35	2,346	936.84	10.21	3,801	931.11	12.56	57
Comprehension	938.78	10.20	2,345	938.81	10.37	3,802	932.58	12.85	57
Overall	936.86	10.09	2,344	937.00	10.01	3,797	931.30	12.49	57

4.6.2 Mean Scale Scores by Ethnicity

Table 4.6.2 A

Mean Scale Scores by Ethnicity: 1-2

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	932.00	933.49	933.76	929.69	933.15	931.91	933.10	932.03
	Std. Dev.	10.55	13.69	13.88	11.90	11.49	11.99	12.40	11.38
	N	595	595	597	599	595	595	595	595
Non-Hispanic Pacific Islander	Mean	929.38	927.81	930.19	922.46	929.88	925.42	928.38	926.58
	Std. Dev.	11.34	12.45	13.73	8.83	11.68	9.35	12.01	9.67
	N	26	26	26	26	26	26	26	26
Non-Hispanic Black	Mean	932.23	932.74	934.32	927.37	933.53	930.35	932.65	931.06
	Std. Dev.	10.52	13.28	14.67	11.76	12.02	11.84	12.11	11.47
	N	348	348	348	349	348	348	348	348
Hispanic (Of Any Race)	Mean	933.30	933.57	933.73	928.05	933.79	931.12	933.57	931.67
	Std. Dev.	10.73	12.72	14.48	10.98	11.97	11.09	11.78	10.94
	N	2,698	2,700	2,698	2,721	2,696	2,699	2,698	2,695
Non-Hispanic American Indian	Mean	934.21	935.03	934.82	930.73	934.79	933.31	934.87	933.41
	Std. Dev.	9.84	11.04	14.84	10.66	11.63	10.19	10.30	10.26
	N	39	39	39	40	39	39	39	39
Non-Hispanic Multi-racial	Mean	926.81	929.90	927.57	924.81	927.48	927.57	929.10	927.33
	Std. Dev.	12.73	14.56	15.49	11.04	13.31	12.39	13.87	12.49
	N	21	21	21	21	21	21	21	21
Non-Hispanic White	Mean	931.42	931.95	931.96	927.13	931.96	929.78	931.84	930.22
	Std. Dev.	11.26	12.91	15.07	12.26	12.50	11.83	12.16	11.67
	N	500	500	499	501	499	500	500	499
Missing	Mean	934.57	934.67	936.77	931.32	935.93	933.22	934.78	933.76
	Std. Dev.	9.88	11.75	12.89	11.01	10.73	10.59	10.80	10.20
	N	305	306	304	308	303	306	305	303

Table 4.6.2 B
Mean Scale Scores by Ethnicity: 3-5

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	936.04	935.68	935.69	932.59	936.00	934.42	935.81	934.66
	Std. Dev.	10.92	11.54	13.77	12.63	11.58	11.49	11.06	11.22
	N	868	868	868	869	868	868	868	868
Non-Hispanic Pacific Islander	Mean	936.15	934.25	935.11	931.58	935.74	933.15	934.89	933.74
	Std. Dev.	12.10	10.86	13.33	11.83	12.15	10.96	10.97	11.07
	N	53	53	53	53	53	53	53	53
Non-Hispanic Black	Mean	936.84	935.80	936.27	931.71	936.67	934.04	936.13	934.61
	Std. Dev.	10.19	10.89	13.43	12.12	11.09	10.92	10.39	10.66
	N	525	525	525	526	524	525	525	524
Hispanic (Of Any Race)	Mean	937.55	936.22	936.42	932.23	937.09	934.51	936.63	935.06
	Std. Dev.	10.52	10.77	13.31	11.76	11.20	10.66	10.41	10.51
	N	4,480	4,488	4,489	4,505	4,475	4,487	4,478	4,473
Non-Hispanic American Indian	Mean	939.35	940.21	939.95	934.74	939.72	937.74	939.93	938.16
	Std. Dev.	9.66	9.34	11.25	11.43	9.54	9.55	8.84	9.24
	N	43	43	43	43	43	43	43	43
Non-Hispanic Multi-racial	Mean	935.62	933.63	935.08	930.79	935.43	932.50	934.11	933.00
	Std. Dev.	9.46	10.71	12.56	11.44	10.31	10.43	9.93	9.86
	N	37	38	38	38	37	38	37	37
Non-Hispanic White	Mean	936.81	935.59	935.71	931.55	936.37	933.84	935.99	934.37
	Std. Dev.	10.69	10.86	13.49	12.29	11.39	11.01	10.55	10.77
	N	574	574	574	577	574	574	574	574
Missing	Mean	939.77	938.49	939.37	935.22	939.64	937.14	938.89	937.67
	Std. Dev.	9.44	9.64	11.61	11.24	9.88	9.96	9.31	9.58
	N	490	490	490	490	490	490	490	490

Table 4.6.2 C
Mean Scale Scores by Ethnicity: 6-8

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	936.96	937.76	935.97	933.84	936.82	936.03	937.54	936.03
	Std. Dev.	10.56	12.14	12.85	10.83	11.07	10.80	11.40	10.56
	N	672	672	672	672	671	672	671	671
Non-Hispanic Pacific Islander	Mean	936.28	936.25	934.26	930.47	935.60	933.53	936.23	933.92
	Std. Dev.	10.99	12.90	14.25	10.12	12.04	10.67	11.96	10.54
	N	53	53	53	53	53	53	53	53
Non-Hispanic Black	Mean	937.05	937.17	936.81	932.56	937.29	935.11	937.15	935.54
	Std. Dev.	10.97	12.11	13.04	10.97	11.50	10.89	11.52	10.70
	N	414	414	414	414	414	414	414	414
Hispanic (Of Any Race)	Mean	938.43	938.77	937.33	933.29	938.26	936.27	938.69	936.63
	Std. Dev.	10.18	11.47	12.82	10.81	11.02	10.48	10.80	10.30
	N	3,825	3,829	3,828	3,835	3,820	3,829	3,824	3,819
Non-Hispanic American Indian	Mean	941.42	942.11	939.25	936.77	940.74	939.72	941.87	939.77
	Std. Dev.	8.61	9.62	11.76	9.46	9.46	9.01	8.98	8.55
	N	53	53	53	53	53	53	53	53
Non-Hispanic Multi-racial	Mean	936.57	937.79	935.71	933.79	936.46	936.04	937.43	935.96
	Std. Dev.	9.52	11.60	13.55	10.60	10.68	9.60	10.79	9.62
	N	28	28	28	28	28	28	28	28
Non-Hispanic White	Mean	937.50	937.76	936.93	932.97	937.57	935.55	937.71	935.90
	Std. Dev.	10.63	12.25	13.34	11.86	11.61	11.49	11.50	11.22
	N	438	438	440	440	438	438	438	438
Missing	Mean	940.96	941.18	941.07	935.86	941.43	938.76	941.12	939.34
	Std. Dev.	7.72	9.42	10.07	9.51	8.37	8.73	8.60	8.25
	N	363	364	363	364	362	364	363	362

Table 4.6.2 D
Mean Scale Scores by Ethnicity: 9-12

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	937.94	938.08	935.95	934.35	937.13	936.44	938.11	936.45
	Std. Dev.	9.97	10.82	12.37	11.25	10.48	10.40	10.35	10.16
	N	810	810	811	811	810	810	810	810
Non-Hispanic Pacific Islander	Mean	939.97	940.57	938.26	936.11	939.34	938.57	940.43	938.57
	Std. Dev.	8.67	9.71	10.61	10.48	9.24	9.78	9.22	9.43
	N	35	35	35	35	35	35	35	35
Non-Hispanic Black	Mean	937.58	937.07	936.05	933.59	936.96	935.55	937.28	935.81
	Std. Dev.	10.55	11.32	12.12	11.35	10.67	10.86	10.82	10.55
	N	440	440	440	440	440	440	440	440
Hispanic (Of Any Race)	Mean	938.98	938.83	936.50	934.44	937.90	936.84	938.96	936.99
	Std. Dev.	10.08	10.67	12.23	10.99	10.48	10.21	10.30	10.02
	N	3,906	3,909	3,907	3,909	3,903	3,908	3,906	3,902
Non-Hispanic American Indian	Mean	940.07	940.29	937.55	935.79	938.98	938.24	940.31	938.29
	Std. Dev.	9.01	10.53	10.35	11.20	9.06	9.88	9.86	9.41
	N	42	42	42	42	42	42	42	42
Non-Hispanic Multi-racial	Mean	938.38	938.69	937.81	932.38	938.27	935.77	938.62	936.27
	Std. Dev.	9.71	10.52	10.26	8.85	9.47	8.93	10.11	8.96
	N	26	26	26	26	26	26	26	26
Non-Hispanic White	Mean	938.93	938.45	937.36	934.34	938.29	936.62	938.69	936.97
	Std. Dev.	10.33	10.98	12.08	11.24	10.61	10.60	10.63	10.34
	N	495	495	494	494	494	494	495	494
Missing	Mean	939.56	938.92	938.08	935.21	938.96	937.31	939.19	937.67
	Std. Dev.	9.83	10.20	11.03	11.11	9.91	10.02	9.87	9.70
	N	450	450	450	449	450	449	450	449

4.7 Scale Scores by Grade

4.7.1 Mean Scale Scores by Gender

Table 4.7.1 A
Mean Scale Scores by Gender: Grade 1

	Female			Male			Missing			Total		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	931.51	10.72	647	931.53	11.12	1,566	934.10	9.72	30	931.56	11.00	2,243
Reading	931.13	11.86	647	931.59	12.94	1,566	934.37	10.90	30	931.50	12.62	2,243
Speaking	931.74	14.40	647	932.73	14.65	1,567	934.27	14.25	30	932.46	14.58	2,244
Writing	925.99	10.57	653	927.35	11.21	1,579	929.63	11.02	30	926.99	11.05	2,262
Oral	931.86	11.86	646	932.39	12.26	1,566	934.53	10.82	30	932.26	12.13	2,242
Literacy	928.82	10.41	647	929.76	11.26	1,566	932.27	10.45	30	929.52	11.02	2,243
Comprehension	931.32	11.17	647	931.64	12.05	1,566	934.37	9.93	30	931.58	11.78	2,243
Overall	929.48	10.43	646	930.31	11.13	1,566	932.50	10.16	30	930.10	10.93	2,242

Table 4.7.1 B
Mean Scale Scores by Gender: Grade 2

	Female			Male			Missing			Total		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	934.07	10.12	708	934.18	10.43	1,561	937.45	8.92	20	934.18	10.33	2,289
Reading	934.74	12.05	709	935.31	13.21	1,563	939.50	11.64	20	935.17	12.86	2,292
Speaking	935.06	14.09	708	934.91	14.23	1,560	940.40	11.22	20	935.01	14.17	2,288
Writing	928.81	11.09	711	929.91	11.70	1,572	934.00	9.40	20	929.61	11.51	2,303
Oral	934.83	11.46	707	934.84	11.68	1,558	939.15	9.55	20	934.87	11.60	2,285
Literacy	932.08	10.82	709	932.92	11.72	1,562	937.10	9.93	20	932.70	11.45	2,291
Comprehension	934.65	11.10	708	935.06	12.06	1,561	938.95	10.62	20	934.96	11.77	2,289
Overall	932.64	10.58	707	933.25	11.30	1,557	937.50	9.37	20	933.10	11.08	2,284

Table 4.7.1 C
Mean Scale Scores by Gender: Grade 3

	Female			Male			Missing			Total		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	936.18	10.71	742	936.34	10.77	1,646	937.71	11.33	17	936.30	10.75	2,405
Reading	934.08	10.46	743	935.10	11.00	1,648	934.53	11.13	17	934.78	10.85	2,408
Speaking	934.92	13.66	743	935.85	13.31	1,651	934.41	13.21	17	935.55	13.43	2,411
Writing	929.74	11.21	749	931.23	11.54	1,663	927.71	9.39	17	930.75	11.45	2,429
Oral	935.66	11.39	742	936.20	11.31	1,645	936.12	11.20	17	936.03	11.34	2,404
Literacy	932.21	10.28	743	933.46	10.59	1,648	931.35	9.39	17	933.06	10.50	2,408
Comprehension	934.73	10.27	742	935.49	10.60	1,644	935.65	10.95	17	935.26	10.51	2,403
Overall	933.01	10.29	742	934.05	10.48	1,643	932.53	9.59	17	933.72	10.43	2,402

Table 4.7.1 D
Mean Scale Scores by Gender: Grade 4

	Female			Male			Missing			Total		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	937.89	9.94	760	937.72	10.57	1,554	939.45	8.92	11	937.79	10.36	2,325
Reading	936.47	10.33	760	936.89	10.95	1,559	936.09	9.09	11	936.75	10.74	2,330
Speaking	936.94	12.82	760	936.90	13.25	1,559	931.45	15.59	11	936.89	13.13	2,330
Writing	932.17	11.86	760	933.31	11.88	1,560	927.91	15.17	11	932.91	11.91	2,331
Oral	937.52	10.72	760	937.41	11.22	1,552	935.64	11.23	11	937.44	11.06	2,323
Literacy	934.59	10.51	760	935.38	10.82	1,559	932.36	11.40	11	935.11	10.73	2,330
Comprehension	936.92	9.95	760	937.15	10.56	1,554	937.09	8.91	11	937.07	10.36	2,325
Overall	935.25	10.27	760	935.76	10.62	1,552	933.09	10.97	11	935.58	10.51	2,323

Table 4.7.1 E
Mean Scale Scores by Gender: Grade 5

	Female			Male			Missing			Total		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	937.80	10.28	817	938.32	10.43	1,514	940.56	4.72	9	938.15	10.37	2,340
Reading	936.59	10.59	817	937.49	10.80	1,515	936.89	9.72	9	937.17	10.73	2,341
Speaking	936.08	13.57	817	937.51	13.09	1,513	936.44	13.47	9	937.01	13.28	2,339
Writing	932.26	12.39	816	934.28	12.12	1,516	934.56	13.43	9	933.58	12.26	2,341
Oral	937.05	11.17	816	938.01	11.08	1,512	938.56	8.14	9	937.68	11.11	2,337
Literacy	934.70	10.96	816	936.16	10.93	1,515	936.00	11.26	9	935.65	10.96	2,340
Comprehension	936.98	10.22	817	937.75	10.42	1,514	938.00	7.92	9	937.48	10.35	2,340
Overall	935.19	10.70	816	936.49	10.68	1,512	936.56	9.77	9	936.04	10.70	2,337

Table 4.7.1 F
Mean Scale Scores by Gender: Grade 6

	Female			Male			Missing			Total		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	937.37	10.19	711	937.59	10.74	1,320	940.37	5.83	30	937.56	10.50	2,061
Reading	937.21	11.20	712	937.83	12.08	1,320	940.23	9.41	30	937.65	11.76	2,062
Speaking	936.38	12.72	713	937.03	12.91	1,320	938.57	10.50	30	936.83	12.82	2,063
Writing	932.00	10.73	714	932.77	10.65	1,321	933.83	8.02	30	932.52	10.65	2,065
Oral	937.25	10.86	710	937.69	11.43	1,319	939.87	6.78	30	937.57	11.19	2,059
Literacy	934.84	10.32	712	935.54	10.73	1,320	937.27	7.44	30	935.33	10.55	2,062
Comprehension	937.27	10.62	710	937.77	11.41	1,319	940.23	8.01	30	937.63	11.10	2,059
Overall	935.32	10.15	710	935.94	10.63	1,318	937.80	6.70	30	935.75	10.42	2,058

Table 4.7.1 G
Mean Scale Scores by Gender: Grade 7

	Female			Male			Missing			Total		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	938.57	10.34	706	938.10	10.30	1,217	941.50	4.22	10	938.29	10.30	1,933
Reading	938.72	11.40	707	938.53	11.62	1,217	941.90	4.97	10	938.62	11.52	1,934
Speaking	937.07	13.22	707	937.12	12.92	1,215	943.30	4.82	10	937.13	13.01	1,932
Writing	933.28	11.21	707	933.24	10.83	1,218	938.90	6.74	10	933.28	10.96	1,935
Oral	938.19	11.20	706	937.99	11.11	1,214	942.80	4.17	10	938.09	11.12	1,930
Literacy	936.23	10.58	707	936.11	10.59	1,217	940.60	5.52	10	936.18	10.57	1,934
Comprehension	938.70	10.85	706	938.43	10.94	1,217	941.80	4.12	10	938.55	10.88	1,933
Overall	936.60	10.41	706	936.45	10.39	1,214	941.10	4.72	10	936.53	10.39	1,930

Table 4.7.1 H
Mean Scale Scores by Gender: Grade 8

	Female			Male			Missing			Total		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	939.35	9.38	656	938.79	9.93	1,188	938.88	12.90	8	938.99	9.76	1,852
Reading	939.80	11.09	657	939.64	11.43	1,190	937.88	15.38	8	939.69	11.33	1,855
Speaking	937.99	12.25	656	938.10	12.57	1,192	940.00	12.34	8	938.07	12.46	1,856
Writing	934.56	11.12	657	934.68	10.64	1,194	934.25	11.95	8	934.64	10.82	1,859
Oral	939.06	10.35	655	938.82	10.76	1,187	939.88	12.70	8	938.91	10.63	1,850
Literacy	937.42	10.49	657	937.39	10.32	1,190	936.25	13.45	8	937.40	10.39	1,855
Comprehension	939.67	10.27	656	939.41	10.69	1,188	938.25	14.62	8	939.50	10.57	1,852
Overall	937.68	10.05	655	937.57	10.11	1,187	937.25	12.98	8	937.61	10.10	1,850

Table 4.7.1 I
Mean Scale Scores by Gender: Grade 9

	Female			Male			Missing			Total		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	938.60	10.07	542	938.43	10.44	960	932.03	14.11	29	938.37	10.43	1,531
Reading	938.68	10.69	542	938.28	11.06	961	931.93	13.67	29	938.30	11.02	1,532
Speaking	936.76	12.00	541	936.44	12.42	960	929.69	14.75	29	936.43	12.36	1,530
Writing	934.32	11.43	542	934.10	10.69	961	929.52	12.48	29	934.09	11.01	1,532
Oral	937.82	10.41	541	937.59	10.69	959	931.00	14.05	29	937.55	10.71	1,529
Literacy	936.73	10.43	542	936.39	10.21	961	930.86	12.80	29	936.40	10.37	1,532
Comprehension	938.73	10.30	542	938.40	10.65	960	932.07	13.69	29	938.40	10.63	1,531
Overall	936.88	10.18	541	936.60	10.05	959	930.83	13.05	29	936.59	10.19	1,529

Table 4.7.1 J
Mean Scale Scores by Gender: Grade 10

	Female			Male			Missing			Total		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	939.02	9.95	567	939.11	9.87	895	935.20	8.54	10	939.05	9.90	1,472
Reading	938.96	10.47	568	938.85	10.40	896	931.80	10.20	10	938.85	10.44	1,474
Speaking	937.15	11.83	569	937.04	11.93	897	932.50	12.61	10	937.05	11.91	1,476
Writing	934.69	11.10	569	934.75	10.80	896	927.90	10.00	10	934.68	10.93	1,475
Oral	938.23	10.30	567	938.23	10.24	895	934.10	9.74	10	938.20	10.27	1,472
Literacy	937.03	10.26	568	937.02	9.96	896	930.20	9.56	10	936.98	10.09	1,474
Comprehension	939.05	10.14	567	939.02	10.06	895	932.90	9.51	10	938.99	10.10	1,472
Overall	937.23	10.02	567	937.19	9.79	895	931.00	9.27	10	937.16	9.89	1,472

Table 4.7.1 K
Mean Scale Scores by Gender: Grade 11

	Female			Male			Missing			Total		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	939.52	9.87	477	939.26	9.89	815	934.56	13.08	9	939.32	9.92	1,301
Reading	939.08	10.69	477	938.89	10.58	815	931.67	15.67	9	938.91	10.68	1,301
Speaking	937.34	11.48	477	937.06	11.93	815	932.67	16.13	9	937.14	11.81	1,301
Writing	935.23	11.14	478	934.94	10.84	814	931.44	16.19	9	935.02	11.00	1,301
Oral	938.58	10.09	477	938.30	10.22	815	933.67	14.40	9	938.37	10.21	1,301
Literacy	937.36	10.38	477	937.13	10.04	814	931.78	15.73	9	937.18	10.23	1,300
Comprehension	939.30	10.26	477	939.08	10.19	815	932.67	14.75	9	939.12	10.27	1,301
Overall	937.54	10.08	477	937.35	9.81	814	932.11	15.10	9	937.38	9.96	1,300

Table 4.7.1 L
Mean Scale Scores by Gender: Grade 12

	Female			Male			Missing			Total		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Listening	938.54	9.78	759	938.63	10.31	1,132	935.89	8.01	9	938.58	10.09	1,900
Reading	938.02	10.54	759	938.77	10.85	1,132	932.44	12.66	9	938.44	10.75	1,900
Speaking	935.49	12.34	759	936.43	12.25	1,130	931.67	12.79	9	936.03	12.30	1,898
Writing	933.48	11.17	759	934.54	11.29	1,130	931.44	11.56	9	934.10	11.26	1,898
Oral	937.20	10.38	759	937.70	10.61	1,130	934.00	10.14	9	937.48	10.52	1,898
Literacy	935.97	10.29	759	936.86	10.51	1,130	932.22	10.96	9	936.48	10.44	1,898
Comprehension	938.27	10.11	759	938.82	10.49	1,132	933.78	11.02	9	938.57	10.35	1,900
Overall	936.15	10.03	759	936.95	10.28	1,129	932.33	10.62	9	936.61	10.20	1,897

4.7.2 Mean Scale Scores by Ethnicity

Table 4.7.2 A
Mean Scale Scores by Ethnicity: Grade 1

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	930.64	931.80	933.15	928.91	932.16	930.66	931.53	930.88
	Std. Dev.	10.64	13.66	13.94	11.57	11.60	11.69	12.39	11.18
	N	294	294	296	297	294	294	294	294
Non-Hispanic Pacific Islander	Mean	924.91	920.91	926.27	918.27	925.64	919.82	922.18	921.45
	Std. Dev.	11.33	10.23	13.80	7.02	11.53	7.54	10.61	8.61
	N	11	11	11	11	11	11	11	11
Non-Hispanic Black	Mean	930.78	931.09	932.76	925.14	932.03	928.37	931.05	929.21
	Std. Dev.	10.73	13.32	15.00	10.91	12.28	11.46	12.16	11.22
	N	175	175	175	175	175	175	175	175
Hispanic (Of Any Race)	Mean	931.90	931.58	932.25	926.55	932.33	929.37	931.75	930.00
	Std. Dev.	11.04	12.38	14.58	10.59	12.21	10.72	11.64	10.75
	N	1,306	1,306	1,306	1,318	1,306	1,306	1,306	1,306
Non-Hispanic American Indian	Mean	933.94	934.31	936.38	930.65	935.44	933.06	934.25	933.50
	Std. Dev.	10.50	10.17	13.05	10.99	10.79	9.73	9.92	9.71
	N	16	16	16	17	16	16	16	16
Non-Hispanic Multi-racial	Mean	927.00	929.15	927.15	924.46	927.38	927.00	928.62	926.85
	Std. Dev.	12.43	12.04	15.48	11.28	12.75	11.16	11.98	11.36
	N	13	13	13	13	13	13	13	13
Non-Hispanic White	Mean	930.49	930.71	931.16	926.71	931.07	928.95	930.69	929.38
	Std. Dev.	11.66	13.01	15.33	12.43	12.82	11.97	12.33	11.83
	N	268	268	268	269	268	268	268	268
Missing	Mean	933.65	932.69	935.28	929.89	934.74	931.42	933.04	932.12
	Std. Dev.	9.51	10.88	13.45	10.34	10.47	9.54	10.05	9.46
	N	160	160	159	162	159	160	160	159

Table 4.7.2 B
Mean Scale Scores by Ethnicity: Grade 2

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	933.33	935.14	934.36	930.45	934.11	933.13	934.64	933.15
	Std. Dev.	10.28	13.52	13.79	12.16	11.30	12.14	12.22	11.46
	N	301	301	301	302	301	301	301	301
Non-Hispanic Pacific Islander	Mean	932.67	932.87	933.07	925.53	933.00	929.53	932.93	930.33
	Std. Dev.	10.16	11.47	12.94	8.76	10.78	8.36	10.89	8.63
	N	15	15	15	15	15	15	15	15
Non-Hispanic Black	Mean	933.69	934.42	935.90	929.62	935.05	932.36	934.27	932.93
	Std. Dev.	10.10	13.02	14.16	12.14	11.55	11.87	11.83	11.41
	N	173	173	173	174	173	173	173	173
Hispanic (Of Any Race)	Mean	934.61	935.44	935.11	929.46	935.15	932.75	935.28	933.23
	Std. Dev.	10.25	12.74	14.25	11.15	11.58	11.18	11.65	10.88
	N	1,392	1,394	1,392	1,403	1,390	1,393	1,392	1,389
Non-Hispanic American Indian	Mean	934.39	935.52	933.74	930.78	934.35	933.48	935.30	933.35
	Std. Dev.	9.35	11.59	15.87	10.40	12.17	10.50	10.53	10.63
	N	23	23	23	23	23	23	23	23
Non-Hispanic Multi-racial	Mean	926.50	931.13	928.25	925.38	927.63	928.50	929.88	928.13
	Std. Dev.	13.18	17.85	15.47	10.61	14.17	14.12	16.45	14.10
	N	8	8	8	8	8	8	8	8
Non-Hispanic White	Mean	932.49	933.38	932.89	927.62	932.99	930.75	933.17	931.19
	Std. Dev.	10.68	12.65	14.72	12.05	12.04	11.59	11.81	11.40
	N	232	232	231	232	231	232	232	231
Missing	Mean	935.59	936.84	938.41	932.91	937.25	935.18	936.70	935.56
	Std. Dev.	10.17	12.27	12.04	11.49	10.87	11.31	11.25	10.68
	N	145	146	145	146	144	146	145	144

Table 4.7.2 C
Mean Scale Scores by Ethnicity: Grade 3

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	935.15	934.69	934.77	931.13	935.09	933.22	934.86	933.53
	Std. Dev.	11.37	11.92	14.13	12.25	12.02	11.40	11.46	11.28
	N	331	331	331	332	331	331	331	331
Non-Hispanic Pacific Islander	Mean	935.64	932.64	933.89	930.04	934.86	931.57	933.61	932.39
	Std. Dev.	11.69	9.90	13.00	10.47	11.71	9.81	10.14	10.06
	N	28	28	28	28	28	28	28	28
Non-Hispanic Black	Mean	935.57	933.88	934.53	930.00	935.19	932.26	934.43	932.89
	Std. Dev.	10.26	10.76	13.52	11.67	11.02	10.54	10.25	10.31
	N	195	195	195	196	194	195	195	194
Hispanic (Of Any Race)	Mean	936.51	934.88	935.76	930.61	936.23	933.04	935.37	933.76
	Std. Dev.	10.60	10.66	13.25	11.25	11.18	10.27	10.33	10.24
	N	1,432	1,435	1,438	1,451	1,432	1,435	1,430	1,430
Non-Hispanic American Indian	Mean	933.80	932.60	934.30	928.40	934.20	930.80	933.00	931.50
	Std. Dev.	12.47	12.85	14.11	10.97	13.04	11.64	12.58	11.98
	N	10	10	10	10	10	10	10	10
Non-Hispanic Multi-racial	Mean	935.47	932.35	934.47	928.94	935.18	930.82	933.29	932.00
	Std. Dev.	8.59	10.87	13.12	11.53	10.00	10.05	9.65	9.30
	N	17	17	17	17	17	17	17	17
Non-Hispanic White	Mean	935.83	934.17	934.64	930.73	935.37	932.71	934.72	933.28
	Std. Dev.	10.86	10.64	13.97	11.75	11.62	10.74	10.40	10.69
	N	201	201	201	204	201	201	201	201
Missing	Mean	938.25	936.47	937.79	932.30	938.10	934.66	937.03	935.48
	Std. Dev.	10.72	10.38	12.56	10.88	10.93	10.13	10.23	9.98
	N	191	191	191	191	191	191	191	191

Table 4.7.2 D
Mean Scale Scores by Ethnicity: Grade 4

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	936.89	936.40	936.73	933.05	936.94	935.00	936.58	935.35
	Std. Dev.	10.36	11.18	13.07	12.64	10.90	11.35	10.62	10.92
	N	293	293	293	293	293	293	293	293
Non-Hispanic Pacific Islander	Mean	934.31	934.00	935.15	931.92	934.85	933.23	934.15	933.46
	Std. Dev.	13.37	13.03	15.38	14.95	14.23	13.58	12.95	13.59
	N	13	13	13	13	13	13	13	13
Non-Hispanic Black	Mean	937.50	936.94	936.84	932.50	937.27	934.99	937.10	935.47
	Std. Dev.	10.43	11.16	13.51	12.61	11.44	11.23	10.71	11.02
	N	157	157	157	157	157	157	157	157
Hispanic (Of Any Race)	Mean	937.82	936.59	936.71	932.77	937.37	934.96	936.97	935.46
	Std. Dev.	10.56	10.80	13.22	11.67	11.22	10.65	10.47	10.51
	N	1,490	1,494	1,494	1,495	1,488	1,494	1,490	1,488
Non-Hispanic American Indian	Mean	940.67	942.13	940.13	933.07	940.47	937.87	941.60	938.53
	Std. Dev.	6.30	5.08	11.34	9.52	8.20	6.82	5.02	6.88
	N	15	15	15	15	15	15	15	15
Non-Hispanic Multi-racial	Mean	938.50	938.64	939.27	935.91	939.00	937.73	938.40	937.40
	Std. Dev.	6.22	6.50	8.09	8.46	5.73	7.12	5.73	6.10
	N	10	11	11	11	10	11	10	10
Non-Hispanic White	Mean	936.78	935.71	935.52	931.32	936.25	933.79	936.06	934.29
	Std. Dev.	10.28	10.64	13.54	12.30	11.17	10.74	10.29	10.48
	N	194	194	194	194	194	194	194	194
Missing	Mean	940.71	939.59	940.37	936.37	940.62	938.28	939.95	938.74
	Std. Dev.	7.64	8.58	10.76	11.03	8.51	9.31	7.96	8.72
	N	153	153	153	153	153	153	153	153

Table 4.7.2 E
Mean Scale Scores by Ethnicity: Grade 5

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	936.22	936.15	935.70	934.03	936.11	935.36	936.19	935.34
	Std. Dev.	10.87	11.33	14.02	12.94	11.67	11.65	10.94	11.37
	N	244	244	244	244	244	244	244	244
Non-Hispanic Pacific Islander	Mean	939.33	938.25	937.92	934.83	938.75	936.75	938.67	937.17
	Std. Dev.	10.94	9.30	11.09	10.19	10.03	9.34	9.53	9.44
	N	12	12	12	12	12	12	12	12
Non-Hispanic Black	Mean	937.67	936.92	937.72	932.92	937.80	935.20	937.17	935.76
	Std. Dev.	9.75	10.46	13.04	11.95	10.64	10.78	9.98	10.46
	N	173	173	173	173	173	173	173	173
Hispanic (Of Any Race)	Mean	938.27	937.08	936.76	933.23	937.61	935.44	937.45	935.87
	Std. Dev.	10.33	10.71	13.43	12.15	11.16	10.87	10.31	10.65
	N	1,558	1,559	1,557	1,559	1,555	1,558	1,558	1,555
Non-Hispanic American Indian	Mean	941.33	942.83	942.94	939.67	942.17	941.50	942.39	941.56
	Std. Dev.	9.01	7.29	7.62	10.99	6.58	7.93	6.52	6.98
	N	18	18	18	18	18	18	18	18
Non-Hispanic Multi-racial	Mean	933.00	930.30	931.50	928.30	932.30	929.60	931.20	930.30
	Std. Dev.	12.33	12.12	14.21	12.32	12.95	11.95	12.06	12.12
	N	10	10	10	10	10	10	10	10
Non-Hispanic White	Mean	937.93	937.04	937.13	932.73	937.63	935.17	937.34	935.68
	Std. Dev.	10.82	11.14	12.75	12.77	11.24	11.44	10.81	11.04
	N	179	179	179	179	179	179	179	179
Missing	Mean	940.79	939.97	940.37	937.82	940.64	939.19	940.21	939.40
	Std. Dev.	9.07	9.22	10.93	11.05	9.50	9.70	8.96	9.36
	N	146	146	146	146	146	146	146	146

Table 4.7.2 F
Mean Scale Scores by Ethnicity: Grade 6

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	934.63	935.41	933.94	931.64	934.60	933.74	935.18	933.78
	Std. Dev.	11.84	13.22	13.80	10.92	12.27	11.31	12.56	11.32
	N	216	216	216	216	215	216	215	215
Non-Hispanic Pacific Islander	Mean	936.04	936.91	935.91	930.61	936.30	933.96	936.61	934.39
	Std. Dev.	10.03	12.58	11.83	8.32	10.61	9.68	11.45	9.37
	N	23	23	23	23	23	23	23	23
Non-Hispanic Black	Mean	937.50	937.39	937.69	932.05	937.97	934.96	937.44	935.62
	Std. Dev.	10.12	11.37	12.20	10.28	10.56	10.10	10.70	9.86
	N	154	154	154	154	154	154	154	154
Hispanic (Of Any Race)	Mean	937.73	937.71	936.77	932.43	937.63	935.31	937.73	935.77
	Std. Dev.	10.48	11.70	12.92	10.72	11.25	10.60	11.07	10.48
	N	1,366	1,367	1,368	1,370	1,365	1,367	1,365	1,364
Non-Hispanic American Indian	Mean	941.52	942.87	940.35	936.91	941.39	940.17	942.39	940.26
	Std. Dev.	8.74	8.91	11.31	9.09	9.09	8.66	8.69	8.28
	N	23	23	23	23	23	23	23	23
Non-Hispanic Multi-racial	Mean	934.00	933.90	935.20	933.40	934.90	933.90	934.00	934.00
	Std. Dev.	9.58	11.84	12.97	8.62	11.02	9.08	10.95	9.37
	N	10	10	10	10	10	10	10	10
Non-Hispanic White	Mean	937.34	937.38	936.96	932.19	937.51	935.02	937.39	935.49
	Std. Dev.	10.51	11.71	13.21	11.32	11.50	10.91	11.08	10.82
	N	140	140	140	140	140	140	140	140
Missing	Mean	940.75	940.87	940.68	935.40	941.12	938.38	940.81	938.95
	Std. Dev.	7.34	9.16	8.94	9.06	7.49	8.34	8.20	7.64
	N	129	129	129	129	129	129	129	129

Table 4.7.2 G
Mean Scale Scores by Ethnicity: Grade 7

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	938.50	939.16	937.33	934.62	938.28	937.12	938.98	937.23
	Std. Dev.	9.49	10.92	11.92	10.50	9.94	10.02	10.24	9.68
	N	231	231	231	231	231	231	231	231
Non-Hispanic Pacific Islander	Mean	936.00	936.05	932.38	931.52	934.52	933.90	936.05	933.95
	Std. Dev.	11.93	12.38	16.44	11.86	13.42	11.36	12.01	11.52
	N	21	21	21	21	21	21	21	21
Non-Hispanic Black	Mean	935.35	935.36	934.86	931.70	935.45	933.77	935.39	934.07
	Std. Dev.	12.30	13.05	14.08	11.47	12.73	11.77	12.59	11.70
	N	139	139	139	139	139	139	139	139
Hispanic (Of Any Race)	Mean	938.67	939.05	937.35	933.27	938.40	936.39	938.97	936.77
	Std. Dev.	10.08	11.26	12.89	10.76	10.97	10.30	10.63	10.14
	N	1,261	1,261	1,259	1,262	1,258	1,261	1,261	1,258
Non-Hispanic American Indian	Mean	943.07	941.40	939.87	936.00	941.87	938.87	941.93	939.67
	Std. Dev.	5.78	8.68	10.65	9.03	7.92	8.12	7.47	7.43
	N	15	15	15	15	15	15	15	15
Non-Hispanic Multi-racial	Mean	937.17	939.50	934.67	935.17	936.17	937.67	938.83	937.00
	Std. Dev.	7.49	8.75	14.88	12.46	10.85	10.21	8.41	10.25
	N	6	6	6	6	6	6	6	6
Non-Hispanic White	Mean	935.78	935.85	935.36	931.54	935.94	933.91	935.87	934.30
	Std. Dev.	11.68	13.45	14.31	12.95	12.65	12.74	12.65	12.43
	N	152	152	152	152	152	152	152	152
Missing	Mean	940.46	940.57	940.20	934.83	940.72	937.92	940.58	938.56
	Std. Dev.	8.09	9.32	11.58	9.78	9.18	8.78	8.70	8.53
	N	108	109	109	109	108	109	108	108

Table 4.7.2 H
Mean Scale Scores by Ethnicity: Grade 8

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	937.60	938.58	936.51	935.15	937.44	937.11	938.32	936.95
	Std. Dev.	9.91	11.92	12.59	10.74	10.63	10.72	11.02	10.34
	N	225	225	225	225	225	225	225	225
Non-Hispanic Pacific Islander	Mean	937.56	935.00	934.44	927.67	936.33	931.56	935.67	932.67
	Std. Dev.	10.95	14.65	13.89	9.33	11.89	11.22	13.04	10.88
	N	9	9	9	9	9	9	9	9
Non-Hispanic Black	Mean	938.41	938.96	937.94	934.19	938.55	936.84	938.83	937.12
	Std. Dev.	10.10	11.60	12.57	11.07	10.86	10.57	10.95	10.27
	N	121	121	121	121	121	121	121	121
Hispanic (Of Any Race)	Mean	938.98	939.68	937.94	934.28	938.84	937.22	939.49	937.47
	Std. Dev.	9.88	11.31	12.59	10.88	10.76	10.43	10.58	10.17
	N	1,198	1,201	1,201	1,203	1,197	1,201	1,198	1,197
Non-Hispanic American Indian	Mean	939.60	941.67	936.93	937.33	938.60	939.87	941.00	939.13
	Std. Dev.	10.26	11.32	13.10	10.35	10.95	10.26	10.60	9.87
	N	15	15	15	15	15	15	15	15
Non-Hispanic Multi-racial	Mean	938.42	940.17	936.67	933.42	937.92	937.00	939.58	937.08
	Std. Dev.	9.89	11.79	13.26	11.02	10.08	9.40	10.98	9.23
	N	12	12	12	12	12	12	12	12
Non-Hispanic White	Mean	939.46	940.10	938.51	935.18	939.32	937.77	939.93	937.97
	Std. Dev.	9.18	10.99	12.21	10.83	10.25	10.24	10.18	9.88
	N	146	146	148	148	146	146	146	146
Missing	Mean	941.59	942.02	942.22	937.24	942.36	939.87	941.90	940.42
	Std. Dev.	7.74	9.70	9.63	9.56	8.43	8.96	8.84	8.49
	N	126	126	125	126	125	126	126	125

Table 4.7.2 I
Mean Scale Scores by Ethnicity: Grade 9

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	936.21	936.48	934.11	932.58	935.36	934.76	936.45	934.73
	Std. Dev.	10.85	11.93	13.16	11.58	11.19	10.94	11.31	10.69
	N	192	192	192	192	192	192	192	192
Non-Hispanic Pacific Islander	Mean	934.67	934.44	934.56	931.67	934.78	933.22	934.67	933.56
	Std. Dev.	13.67	13.91	13.61	12.58	13.64	13.05	13.86	13.08
	N	9	9	9	9	9	9	9	9
Non-Hispanic Black	Mean	937.78	937.37	936.46	934.59	937.27	936.21	937.53	936.35
	Std. Dev.	10.86	11.51	12.32	11.31	10.80	10.89	10.96	10.58
	N	120	120	120	120	120	120	120	120
Hispanic (Of Any Race)	Mean	938.83	938.75	936.71	934.19	937.91	936.68	938.86	936.90
	Std. Dev.	10.24	10.76	12.19	10.76	10.54	10.10	10.41	9.95
	N	938	939	937	939	936	939	938	936
Non-Hispanic American Indian	Mean	941.42	940.25	935.58	932.58	938.67	936.67	940.75	937.08
	Std. Dev.	6.05	7.89	11.15	8.07	7.56	7.05	7.32	6.91
	N	12	12	12	12	12	12	12	12
Non-Hispanic Multi-racial	Mean	918.00	920.50	918.00	917.00	918.00	919.00	919.50	918.50
	Std. Dev.	0.00	10.50	8.00	7.00	4.00	9.00	7.50	7.50
	N	2	2	2	2	2	2	2	2
Non-Hispanic White	Mean	938.54	938.56	936.97	934.62	937.89	936.81	938.62	936.98
	Std. Dev.	9.56	10.74	12.44	10.81	10.32	10.20	10.25	10.03
	N	132	132	132	132	132	132	132	132
Missing	Mean	938.92	938.70	937.73	935.17	938.45	937.10	938.85	937.40
	Std. Dev.	10.87	10.71	11.70	11.60	10.86	10.69	10.52	10.44
	N	126	126	126	126	126	126	126	126

Table 4.7.2 J
Mean Scale Scores by Ethnicity: Grade 10

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	939.32	939.52	937.57	936.03	938.61	938.02	939.59	937.97
	Std. Dev.	8.93	9.54	11.41	10.25	9.43	9.25	9.16	9.07
	N	188	188	189	189	188	188	188	188
Non-Hispanic Pacific Islander	Mean	941.40	942.60	939.10	935.30	940.50	939.20	942.20	939.40
	Std. Dev.	4.94	5.54	9.92	8.41	6.92	6.42	4.94	6.39
	N	10	10	10	10	10	10	10	10
Non-Hispanic Black	Mean	937.31	937.00	935.97	933.40	936.78	935.42	937.15	935.61
	Std. Dev.	10.99	11.29	12.25	11.73	11.24	11.03	11.06	10.77
	N	106	106	106	106	106	106	106	106
Hispanic (Of Any Race)	Mean	939.11	938.91	936.74	934.55	938.08	936.94	939.04	937.10
	Std. Dev.	9.93	10.51	12.12	10.79	10.36	10.06	10.15	9.89
	N	916	918	919	918	916	918	916	916
Non-Hispanic American Indian	Mean	936.56	937.22	933.22	935.44	935.00	936.56	937.11	935.89
	Std. Dev.	14.77	15.02	14.82	14.97	14.62	14.92	14.93	14.72
	N	9	9	9	9	9	9	9	9
Non-Hispanic Multi-racial	Mean	943.83	944.83	944.50	940.17	944.33	942.67	944.67	943.17
	Std. Dev.	4.63	5.37	1.12	3.62	2.21	3.20	5.28	2.61
	N	6	6	6	6	6	6	6	6
Non-Hispanic White	Mean	938.34	937.62	937.13	933.27	937.87	935.66	937.91	936.19
	Std. Dev.	11.29	11.68	12.46	11.90	11.45	11.30	11.42	11.12
	N	125	125	125	125	125	125	125	125
Missing	Mean	940.23	939.79	939.43	935.81	939.95	938.03	940.06	938.46
	Std. Dev.	7.92	8.49	9.39	10.83	7.94	8.85	8.11	8.28
	N	112	112	112	112	112	112	112	112

Table 4.7.2 K
Mean Scale Scores by Ethnicity: Grade 11

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	938.81	938.74	936.25	934.46	937.70	936.83	938.85	936.90
	Std. Dev.	9.48	10.37	12.26	10.84	10.18	10.09	9.95	9.89
	N	162	162	162	162	162	162	162	162
Non-Hispanic Pacific Islander	Mean	944.43	945.29	942.86	943.71	943.86	944.86	945.00	944.14
	Std. Dev.	3.50	4.30	2.10	6.65	2.36	5.28	3.89	3.98
	N	7	7	7	7	7	7	7	7
Non-Hispanic Black	Mean	939.46	938.28	937.93	934.54	938.82	936.61	938.72	937.18
	Std. Dev.	9.74	11.14	11.04	11.26	9.77	10.77	10.51	10.32
	N	71	71	71	71	71	71	71	71
Hispanic (Of Any Race)	Mean	939.29	938.91	936.76	935.09	938.17	937.19	939.12	937.33
	Std. Dev.	9.92	10.65	12.06	11.01	10.30	10.19	10.24	9.96
	N	846	846	846	847	846	846	846	846
Non-Hispanic American Indian	Mean	943.88	944.50	942.00	938.50	943.00	941.63	944.38	941.88
	Std. Dev.	4.51	4.44	4.50	9.97	3.28	6.74	3.46	5.64
	N	8	8	8	8	8	8	8	8
Non-Hispanic Multi-racial	Mean	933.60	934.80	933.80	930.40	934.00	933.00	934.40	932.80
	Std. Dev.	12.14	13.78	12.54	9.91	12.35	10.64	13.12	10.93
	N	5	5	5	5	5	5	5	5
Non-Hispanic White	Mean	938.76	938.36	938.25	935.37	938.65	937.07	938.57	937.39
	Std. Dev.	11.04	11.50	11.22	11.39	10.66	10.99	11.26	10.63
	N	106	106	106	106	106	106	106	106
Missing	Mean	940.48	939.56	939.45	934.72	940.10	937.53	939.89	938.22
	Std. Dev.	9.59	10.33	9.97	10.57	9.45	9.66	9.90	9.18
	N	96	96	96	95	96	95	96	95

Table 4.7.2 L
Mean Scale Scores by Ethnicity: Grade 12

Ethnicity		Listening	Reading	Speaking	Writing	Oral	Literacy	Comprehension	Overall
Non-Hispanic Asian	Mean	937.68	937.81	935.94	934.35	937.01	936.29	937.83	936.34
	Std. Dev.	10.09	10.93	12.34	11.73	10.64	10.77	10.47	10.45
	N	268	268	268	268	268	268	268	268
Non-Hispanic Pacific Islander	Mean	940.22	940.78	937.44	935.56	939.11	938.33	940.67	938.33
	Std. Dev.	4.76	8.11	10.39	9.50	7.08	8.62	6.82	8.16
	N	9	9	9	9	9	9	9	9
Non-Hispanic Black	Mean	936.66	936.27	934.83	932.43	935.92	934.59	936.47	934.81
	Std. Dev.	10.21	11.21	12.23	11.02	10.40	10.65	10.59	10.35
	N	143	143	143	143	143	143	143	143
Hispanic (Of Any Race)	Mean	938.79	938.77	935.98	934.11	937.55	936.64	938.87	936.74
	Std. Dev.	10.18	10.73	12.43	11.27	10.62	10.43	10.38	10.21
	N	1,206	1,206	1,205	1,205	1,205	1,205	1,206	1,204
Non-Hispanic American Indian	Mean	938.92	939.85	939.62	937.31	939.54	938.77	939.62	938.85
	Std. Dev.	6.68	10.70	5.84	10.61	6.06	8.72	9.21	7.67
	N	13	13	13	13	13	13	13	13
Non-Hispanic Multi-racial	Mean	940.85	940.15	939.31	931.92	940.23	936.23	940.38	937.15
	Std. Dev.	5.43	6.47	7.08	6.24	5.01	5.51	5.79	5.20
	N	13	13	13	13	13	13	13	13
Non-Hispanic White	Mean	940.01	939.22	937.24	934.25	938.80	936.98	939.58	937.37
	Std. Dev.	9.39	10.00	11.99	10.80	9.99	9.89	9.59	9.58
	N	132	132	131	131	131	131	132	131
Missing	Mean	938.84	937.79	936.03	935.09	937.62	936.67	938.15	936.75
	Std. Dev.	10.37	10.87	12.18	11.22	10.68	10.58	10.51	10.42
	N	116	116	116	116	116	116	116	116

4.8 Correlations among Scale Scores by Grade-level Cluster

4.8.1 Correlations among Scale Scores: Grade-level Cluster 1-2

Table 4.8.1
Correlations Among Scale Scores: 1-2

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	0.846**	0.767**	0.692**
	N	4,532	4,532	4,527	4,531
Reading	Pearson Correlation		1	0.748**	0.736**
	N		4,535	4,530	4,534
Writing	Pearson Correlation			1	0.726**
	N			4,532	4,531
Speaking	Pearson Correlation				1
	N				4,565

** . Correlation is significant at the 0.05 level (2-tailed).

4.8.2 Correlations among Scale Scores: Grade-level Cluster 3-5

Table 4.8.2
Correlations Among Scale Scores: 3-5

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	0.881**	0.786**	0.714**
	N	7,070	7,068	7,064	7,069
Reading	Pearson Correlation		1	0.791**	0.783**
	N		7,079	7,073	7,078
Writing	Pearson Correlation			1	0.743**
	N			7,080	7,078
Speaking	Pearson Correlation				1
	N				7,101

** . Correlation is significant at the 0.05 level (2-tailed).

4.8.3 Correlations among Scale Scores: Grade-level Cluster 6-8

Table 4.8.3

Correlations Among Scale Scores: 6-8

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	0.873**	0.781**	0.712**
	N	5,846	5,844	5,839	5,845
Reading	Pearson Correlation		1	0.780**	0.751**
	N		5,851	5,845	5,851
Writing	Pearson Correlation			1	0.729**
	N			5,851	5,850
Speaking	Pearson Correlation				1
	N				5,859

** . Correlation is significant at the 0.05 level (2-tailed).

4.8.4 Correlations among Scale Scores: Grade-level Cluster 9-12

Table 4.8.4

Correlations Among Scale Scores: 9-12

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	0.889**	0.775**	0.743**
	N	6,204	6,204	6,200	6,201
Reading	Pearson Correlation		1	0.785**	0.780**
	N		6,207	6,203	6,204
Writing	Pearson Correlation			1	0.734**
	N			6,205	6,202
Speaking	Pearson Correlation				1
	N				6,206

** . Correlation is significant at the 0.05 level (2-tailed).

4.9 Proficiency Levels

4.9.1 Proficiency Level by Grade-level Cluster

Table 4.9.1 A

Proficiency Level by Cluster: Listening

Cluster	Listening Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1-2	892	19.60	558	12.26	924	20.31	936	20.57	1,240	27.25	4,550
3-5	943	13.30	550	7.76	852	12.02	1,529	21.57	3,216	45.36	7,090
6-8	648	11.04	466	7.94	667	11.36	749	12.76	3,341	56.91	5,871
9-12	623	10.00	442	7.09	724	11.62	1,264	20.28	3,179	51.01	6,232
Total	3,106	13.08	2,016	8.49	3,167	13.34	4,478	18.86	10,976	46.23	23,743

Table 4.9.1 B

Proficiency Level by Cluster: Reading

Cluster	Reading Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1-2	1,006	22.10	667	14.65	800	17.57	980	21.52	1,100	24.16	4,553
3-5	1,011	14.24	792	11.16	948	13.35	1,689	23.79	2,659	37.46	7,099
6-8	727	12.37	415	7.06	580	9.87	1,000	17.02	3,154	53.68	5,876
9-12	677	10.86	527	8.45	597	9.57	1,095	17.56	3,339	53.55	6,235
Total	3,421	14.40	2,401	10.10	2,925	12.31	4,764	20.05	10,252	43.14	23,763

Table 4.9.1 C

Proficiency Level by Cluster: Speaking

Cluster	Speaking Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1-2	1,254	27.56	182	4.00	576	12.66	1,474	32.40	1,064	23.38	4,550
3-5	1,453	20.46	318	4.48	513	7.23	2,216	31.21	2,600	36.62	7,100
6-8	1,049	17.85	183	3.11	618	10.52	1,597	27.18	2,429	41.34	5,876
9-12	1,106	17.74	179	2.87	611	9.80	1,573	25.24	2,764	44.34	6,233
Total	4,862	20.46	862	3.63	2,318	9.76	6,860	28.87	8,857	37.28	23,759

Table 4.9.1 D

Proficiency Level by Cluster: Writing

Cluster	Writing Proficiency Range												Total
	A1		A2		A3		P1		P2		P3		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1-2	1,440	31.42	1,015	22.15	1,164	25.40	810	17.67	119	2.60	35	0.76	4,583
3-5	1,504	21.12	1,351	18.97	1,755	24.65	1,434	20.14	848	11.91	229	3.22	7,121
6-8	972	16.52	1,483	25.20	990	16.83	2,021	34.35	143	2.43	275	4.67	5,884
9-12	917	14.71	1,391	22.31	1,067	17.12	2,327	37.33	180	2.89	352	5.65	6,234
Total	4,833	20.29	5,240	22.00	4,976	20.89	6,592	27.67	1,290	5.42	891	3.74	23,822

Table 4.9.1 E

Proficiency Level by Cluster: Oral

Cluster	Oral Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1-2	1,115	24.53	368	8.10	659	14.50	1,268	27.90	1,135	24.97	4,545
3-5	1,255	17.72	374	5.28	767	10.83	1,861	26.27	2,827	39.91	7,084
6-8	888	15.14	333	5.68	679	11.58	1,122	19.13	2,842	48.47	5,864
9-12	895	14.37	324	5.20	721	11.58	1,604	25.75	2,684	43.10	6,228
Total	4,153	17.51	1,399	5.90	2,826	11.91	5,855	24.68	9,488	40.00	23,721

Table 4.9.1 F

Proficiency Level by Cluster: Literacy

Cluster	Literacy Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1-2	1,161	25.51	886	19.46	1,154	25.35	832	18.28	519	11.40	4,552
3-5	1,210	17.05	1,085	15.29	1,529	21.54	1,748	24.63	1,526	21.50	7,098
6-8	808	13.75	653	11.11	1,281	21.80	1,740	29.61	1,394	23.72	5,876
9-12	775	12.44	695	11.15	1,208	19.38	1,913	30.70	1,641	26.33	6,232
Total	3,954	16.64	3,319	13.97	5,172	21.77	6,233	26.24	5,080	21.38	23,758

Table 4.9.1 G

Proficiency Level by Cluster: Comprehension

Cluster	Comprehension Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1-2	952	20.92	638	14.02	805	17.69	1,150	25.27	1,005	22.09	4,550
3-5	965	13.61	730	10.30	883	12.46	1,528	21.56	2,982	42.07	7,088
6-8	698	11.89	436	7.43	536	9.13	1,141	19.44	3,058	52.10	5,869
9-12	647	10.38	472	7.57	614	9.85	1,190	19.09	3,309	53.10	6,232
Total	3,262	13.74	2,276	9.59	2,838	11.96	5,009	21.10	10,354	43.62	23,739

Table 4.9.1 H

Proficiency Level by Cluster: Overall

Cluster	Overall Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1-2	1,110	24.43	646	14.22	1,193	26.25	990	21.79	605	13.31	4,544
3-5	1,179	16.65	721	10.18	1,481	20.91	1,944	27.45	1,757	24.81	7,082
6-8	786	13.41	495	8.44	1,137	19.39	1,698	28.96	1,747	29.80	5,863
9-12	749	12.03	526	8.45	1,233	19.80	1,769	28.41	1,949	31.30	6,226
Total	3,824	16.12	2,388	10.07	5,044	21.27	6,401	26.99	6,058	25.55	23,715

4.9.2 Proficiency Level by Grade

Table 4.9.2 A

Proficiency Level by Grade: Listening

Grade	Listening Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1	519	23.02	329	14.59	469	20.80	451	20.00	487	21.60	2,255
2	373	16.25	229	9.98	455	19.83	485	21.13	753	32.81	2,295
3	363	15.07	207	8.59	338	14.03	555	23.04	946	39.27	2,409
4	292	12.51	182	7.80	254	10.88	507	21.72	1,099	47.09	2,334
5	288	12.27	161	6.86	260	11.08	467	19.90	1,171	49.89	2,347
6	262	12.63	177	8.53	255	12.29	270	13.01	1,111	53.54	2,075
7	212	10.94	149	7.69	221	11.40	243	12.54	1,113	57.43	1,938
8	174	9.36	140	7.53	191	10.28	236	12.70	1,117	60.12	1,858
9	166	10.74	117	7.57	192	12.43	303	19.61	767	49.64	1,545
10	139	9.41	95	6.43	158	10.70	322	21.80	763	51.66	1,477
11	123	9.40	84	6.42	133	10.17	273	20.87	695	53.13	1,308
12	195	10.25	146	7.68	241	12.67	366	19.24	954	50.16	1,902
Total	3,106	13.08	2,016	8.49	3,167	13.34	4,478	18.86	10,976	46.23	23,743

Table 4.9.2 B

Proficiency Level by Grade: Reading

Grade	Reading Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1	579	25.68	370	16.41	413	18.31	486	21.55	407	18.05	2,255
2	427	18.58	297	12.92	387	16.84	494	21.50	693	30.16	2,298
3	387	16.04	324	13.43	357	14.80	647	26.82	697	28.90	2,412
4	317	13.55	238	10.18	312	13.34	537	22.96	935	39.97	2,339
5	307	13.07	230	9.80	279	11.88	505	21.51	1,027	43.74	2,348
6	288	13.87	167	8.04	228	10.98	364	17.53	1,029	49.57	2,076
7	234	12.07	134	6.91	195	10.06	345	17.79	1,031	53.17	1,939
8	205	11.02	114	6.13	157	8.44	291	15.64	1,094	58.79	1,861
9	182	11.77	137	8.86	147	9.51	254	16.43	826	53.43	1,546
10	147	9.94	108	7.30	147	9.94	289	19.54	788	53.28	1,479
11	144	11.01	104	7.95	103	7.87	235	17.97	722	55.20	1,308
12	204	10.73	178	9.36	200	10.52	317	16.67	1,003	52.73	1,902
Total	3,421	14.40	2,401	10.10	2,925	12.31	4,764	20.05	10,252	43.14	23,763

Table 4.9.2 C

Proficiency Level by Grade: Speaking

Grade	Speaking Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1	690	30.59	105	4.65	293	12.99	725	32.14	443	19.64	2,256
2	564	24.59	77	3.36	283	12.34	749	32.65	621	27.07	2,294
3	530	21.95	118	4.89	210	8.70	794	32.88	763	31.59	2,415
4	468	20.01	99	4.23	153	6.54	726	31.04	893	38.18	2,339
5	455	19.39	101	4.31	150	6.39	696	29.67	944	40.24	2,346
6	386	18.58	65	3.13	254	12.23	579	27.88	793	38.18	2,077
7	360	18.59	58	2.99	185	9.55	544	28.08	790	40.78	1,937
8	303	16.27	60	3.22	179	9.61	474	25.46	846	45.44	1,862
9	286	18.52	39	2.53	145	9.39	387	25.06	687	44.49	1,544
10	247	16.68	38	2.57	135	9.12	386	26.06	675	45.58	1,481
11	208	15.90	40	3.06	130	9.94	314	24.01	616	47.09	1,308
12	365	19.21	62	3.26	201	10.58	486	25.58	786	41.37	1,900
Total	4,862	20.46	862	3.63	2,318	9.76	6,860	28.87	8,857	37.28	23,759

Table 4.9.2 D

Proficiency Level by Grade: Writing

Grade	Writing Proficiency Range												Total
	A1		A2		A3		P1		P2		P3		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1	790	34.74	559	24.58	549	24.14	326	14.34	39	1.72	11	0.48	2,274
2	650	28.15	456	19.75	615	26.63	484	20.96	80	3.46	24	1.04	2,309
3	576	23.67	527	21.66	654	26.88	420	17.26	209	8.59	47	1.93	2,433
4	455	19.44	445	19.02	573	24.49	477	20.38	312	13.33	78	3.33	2,340
5	473	20.14	379	16.14	528	22.49	537	22.87	327	13.93	104	4.43	2,348
6	371	17.85	570	27.42	368	17.70	651	31.31	46	2.21	73	3.51	2,079
7	332	17.11	480	24.74	320	16.49	675	34.79	45	2.32	88	4.54	1,940
8	269	14.42	433	23.22	302	16.19	695	37.27	52	2.79	114	6.11	1,865
9	240	15.52	362	23.42	252	16.30	573	37.06	42	2.72	77	4.98	1,546
10	211	14.26	308	20.81	259	17.50	590	39.86	33	2.23	79	5.34	1,480
11	175	13.38	281	21.48	229	17.51	494	37.77	48	3.67	81	6.19	1,308
12	291	15.32	440	23.16	327	17.21	670	35.26	57	3.00	115	6.05	1,900
Total	4,833	20.29	5,240	22.00	4,976	20.89	6,592	27.67	1,290	5.42	891	3.74	23,822

Table 4.9.2 E

Proficiency Level by Grade: Oral

Grade	Oral Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1	628	27.86	198	8.78	350	15.53	636	28.22	442	19.61	2,254
2	487	21.26	170	7.42	309	13.49	632	27.59	693	30.25	2,291
3	468	19.44	128	5.32	321	13.33	668	27.74	823	34.18	2,408
4	403	17.28	113	4.85	231	9.91	611	26.20	974	41.77	2,332
5	384	16.38	133	5.67	215	9.17	582	24.83	1,030	43.94	2,344
6	330	15.92	123	5.93	280	13.51	394	19.01	946	45.63	2,073
7	302	15.61	114	5.89	204	10.54	380	19.64	935	48.32	1,935
8	256	13.79	96	5.17	195	10.51	348	18.75	961	51.78	1,856
9	246	15.94	77	4.99	164	10.63	397	25.73	659	42.71	1,543
10	205	13.88	70	4.74	147	9.95	410	27.76	645	43.67	1,477
11	171	13.07	54	4.13	169	12.92	319	24.39	595	45.49	1,308
12	273	14.37	123	6.47	241	12.68	478	25.16	785	41.32	1,900
Total	4,153	17.51	1,399	5.90	2,826	11.91	5,855	24.68	9,488	40.00	23,721

Table 4.9.2 F

Proficiency Level by Grade: Literacy

Grade	Literacy Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1	646	28.65	509	22.57	569	25.23	364	16.14	167	7.41	2,255
2	515	22.42	377	16.41	585	25.47	468	20.37	352	15.32	2,297
3	475	19.69	408	16.92	607	25.17	571	23.67	351	14.55	2,412
4	372	15.90	338	14.45	508	21.72	567	24.24	554	23.69	2,339
5	363	15.47	339	14.44	414	17.64	610	25.99	621	26.46	2,347
6	317	15.27	249	11.99	479	23.07	635	30.59	396	19.08	2,076
7	271	13.98	210	10.83	424	21.87	587	30.27	447	23.05	1,939
8	220	11.82	194	10.42	378	20.31	518	27.83	551	29.61	1,861
9	209	13.52	178	11.51	290	18.76	484	31.31	385	24.90	1,546
10	169	11.43	151	10.21	309	20.89	442	29.89	408	27.59	1,479
11	162	12.39	125	9.56	241	18.44	418	31.98	361	27.62	1,307
12	235	12.37	241	12.68	368	19.37	569	29.95	487	25.63	1,900
Total	3,954	16.64	3,319	13.97	5,172	21.77	6,233	26.24	5,080	21.38	23,758

Table 4.9.2 G

Proficiency Level by Grade: Comprehension

Grade	Comprehension Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1	551	24.43	358	15.88	413	18.31	565	25.06	368	16.32	2,255
2	401	17.47	280	12.20	392	17.08	585	25.49	637	27.76	2,295
3	374	15.54	286	11.88	343	14.25	597	24.80	807	33.53	2,407
4	305	13.07	218	9.34	276	11.83	487	20.87	1,048	44.90	2,334
5	286	12.19	226	9.63	264	11.25	444	18.92	1,127	48.02	2,347
6	277	13.36	162	7.81	218	10.52	416	20.07	1,000	48.24	2,073
7	225	11.61	142	7.33	183	9.44	389	20.07	999	51.55	1,938
8	196	10.55	132	7.10	135	7.27	336	18.08	1,059	57.00	1,858
9	172	11.13	131	8.48	144	9.32	285	18.45	813	52.62	1,545
10	141	9.55	100	6.77	144	9.75	311	21.06	781	52.88	1,477
11	135	10.32	94	7.19	108	8.26	247	18.88	724	55.35	1,308
12	199	10.46	147	7.73	218	11.46	347	18.24	991	52.10	1,902
Total	3,262	13.74	2,276	9.59	2,838	11.96	5,009	21.10	10,354	43.62	23,739

Table 4.9.2 H

Proficiency Level by Grade: Overall

Grade	Overall Proficiency Range										Total
	A1		A2		A3		P1		P2		
	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	Count	% within PL	
1	632	28.04	346	15.35	637	28.26	440	19.52	199	8.83	2,254
2	478	20.87	300	13.10	556	24.28	550	24.02	406	17.73	2,290
3	448	18.62	276	11.47	592	24.61	669	27.81	421	17.50	2,406
4	370	15.87	226	9.69	474	20.33	633	27.14	629	26.97	2,332
5	361	15.40	219	9.34	415	17.70	642	27.39	707	30.16	2,344
6	306	14.77	189	9.12	423	20.42	646	31.18	508	24.52	2,072
7	256	13.23	168	8.68	377	19.48	551	28.48	583	30.13	1,935
8	224	12.07	138	7.44	337	18.16	501	26.99	656	35.34	1,856
9	203	13.16	141	9.14	286	18.54	439	28.45	474	30.72	1,543
10	163	11.04	115	7.79	299	20.24	426	28.84	474	32.09	1,477
11	151	11.55	97	7.42	249	19.05	370	28.31	440	33.66	1,307
12	232	12.22	173	9.11	399	21.01	534	28.12	561	29.54	1,899
Total	3,824	16.12	2,388	10.07	5,044	21.27	6,401	26.99	6,058	25.55	23,715

4.10 Participation by Disability

4.10.1 Participation by Disability

Table 4.10.1
Participation by Disability

		Secondary Disability														Total	
		NSD	AS	DB	DD	HI	ID	MD	OHI	OI	SED	SLD	SLI	TBI	ITD		VI
Primary Disability	NPD	2,395	7	0	0	4	11	8	6	7	0	3	57	0	0	5	2,503
	AS	4,348	7	1	29	12	543	42	77	6	14	44	1,343	5	0	11	6,482
	DB	4	1	0	0	0	3	0	1	0	0	0	2	0	0	1	12
	DD	723	25	0	2	21	59	7	43	22	4	24	262	0	0	16	1,208
	HI	36	1	0	2	0	11	1	7	3	0	3	14	0	0	2	80
	ID	6,514	225	10	77	127	18	121	527	238	42	111	2,258	12	0	125	10,405
	MD	375	26	6	10	11	41	51	42	21	6	16	262	2	0	10	879
	OHI	634	19	2	17	13	105	15	4	32	2	25	217	1	0	25	1,111
	OI	51	0	2	4	0	39	7	12	0	0	3	30	0	0	3	151
	SED	55	1	0	2	1	12	1	6	1	0	5	8	0	0	0	92
	SLD	441	2	0	2	3	5	5	17	5	2	0	142	0	0	4	628
	SLI	160	4	0	4	1	12	1	4	1	3	6	2	1	0	1	200
	TBI	69	1	1	2	0	11	5	7	1	0	0	21	1	0	8	127
	ITD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VI	12	4	0	3	1	9	1	3	1	2	0	5	0	0	0	41	
Total		15,811	323	22	154	194	879	265	756	338	75	240	4,623	22	0	211	23,919

Acronyms for Table 4.10.1

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

5. Analyses of Test Forms: Overview

This chapter contains two parts. The first part provides some background on the technical measurement and statistical tools used to analyze Alternate ACCESS for ELLs. The second part explains the results that are presented for each test form in Chapter 6.

5.1 Background

5.1.1 Measurement Models Used

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

For all domains, a Rasch Rating Scale model was used. Mathematically, this can be represented as

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

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

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

B_n = ability of person “n”

D_i = difficulty of task “i”

F_k = calibration of step “k” on the rating scale

All Rasch analyses were conducted using the Rasch measurement software program *Winsteps* (Linacre, 2006). When speaking of the measure of examinee ability, we use the term “ability measure” (rather than *theta*, which is used commonly when discussing models based on Item Response Theory [IRT]). When speaking of the measure of how hard an item was, we use the term “item difficulty measure” (rather than the term *b parameter*, which is used commonly when discussing models based on IRT). “Step measures” refer to the calibration of the steps in the Rasch Rating Scale model presented above. All three measures (ability, difficulty, and step) are expressed in terms of Rasch logits, which then are converted into scores on the Alternate ACCESS for ELLs score scale for reporting purposes (see WIDA Technical Report 1 for more details).

Rasch model standard errors also appear in the tables. These are an indication of the precision with which the measures have been estimated. Unlike the standard error of measurement (SEM) based on classical test theory, which posits the same SEM for all persons regardless of their position on the ability distribution, Rasch model standard errors are conditional on the individual’s ability

measure. All things being equal, if a person gets few items correct or few items incorrect, the standard error of that person's measure will be greater than if a person gets a moderate number of items correct. In addition, for ability measures, standard errors are a function of the number of items on a test form as well as the distribution and quality of the items (i.e., their fit to the Rasch model).

Fit statistics for the Rasch model are provided in Chapter 6. These statistics are calculated by comparing the observed empirical data with the data that would be expected to be produced by the Rasch model. Of the several statistics available, the mean square fit statistics were used to flag items in the development of Alternate ACCESS for ELLs that needed to be deleted or revised. Outfit mean square statistics are more sensitive to outliers. For example, a difficult item that some low ability examinees get correct will have a high outfit mean square statistic that indicates that the item may not be measuring the same thing as other items on the test. Infit mean square statistics are influenced by more aberrant response patterns and generally indicate a more serious measurement problem. The expectation for both of these statistics is 1.00 and values near 1.0 are not of great concern. Values less than 1.00 indicate that the observations are too predictable and thus redundant, but are not of great concern. High values are more of a concern.

According to Linacre (2002):

values greater than 2.0 “distort or degrade the measurement system”

values between 1.5 and 2.0 are “unproductive for construction of measurement, but not degrading”

values between 0.5 and 1.5 should be considered “productive for measurement” values below 0.5 are considered “less productive for measurement, but not degrading”

Because conservative guidelines were followed in the development of Alternate ACCESS for ELLs, the vast majority of items and tasks on the test forms have mean square fit statistics in the range of 0.75 and 1.25 and therefore fall within the range that is “productive for measurement” according to the guidelines above.

5.1.2 Sampling

The results presented in most of the tables in Chapter 6 are based on the full data set of all students who were administered operational Series 403 of Alternate ACCESS for ELLs in the academic year 2018-2019. The item analysis summary tables (Table F), the complete item analysis tables (Table G), and the raw score to scale score conversion tables (Table H) use item difficulties from this calibration.

5.1.3 Scaling

Complete information on the horizontal and vertical scaling of Alternate ACCESS for ELLs scores is provided in Technical Report 1, *Alternate Access for ELLs™ Series 100 Development and Operational Field Test: Technical Report*. In brief, this scaling was accomplished during the field test based on an elaborate common item design, across grade-level clusters, which spanned two series of complete test forms. Concurrent calibration was used to determine item difficulty measures. These item difficulty measures were used to create the Alternate ACCESS for ELLs scale scores used for reporting results on the test.

Table 5.1.3A provides the scaling equation for each domain. This equation is used to convert an examinee’s ability measure into the scale score. Since Alternate ACCESS for ELLs is vertically equated, though each domain has its own equation, the same equation is used across all grade-level clusters within each domain.

Table 5.1.3A

Scaling Equation for each Domain

Domain	Scale Score
Listening	$(\text{Ability Measure in Logits} * 7.913) + 925.056$
Reading	$(\text{Ability Measure in Logits} * 6.026) + 925.788$
Speaking	$(\text{Ability Measure in Logits} * 4.433) + 924.531$
Writing	$(\text{Ability Measure in Logits} * 2.4) + 926.408$

5.1.4 DIF Analyses

Differential item analyses (DIF) attempt to investigate whether performances on items or tasks were influenced by factors extraneous to English language proficiency (i.e., the construct being measured on the test). In other words, it attempts to find items or tasks that may be functioning differently for different groups based on criteria irrelevant to what is being tested. The performance of students on the Alternate ACCESS for ELLs tasks was compared by dividing students into two different groupings: first, males versus females; second, students of Hispanic ethnic background versus students of non-Hispanic ethnic background (For both analyses, students for whom test scores and gender or ethnicity was missing were excluded). The underlying assumption of DIF analysis is that students who performed similarly overall on the test should perform similarly on the individual tasks. To test this assumption, students are initially placed into groups based on their total raw scores by domain. Then, student performance on a task of interest within that domain, the studied item, is compared between groups.

The Mantel Chi-square statistic and the standardized P-DIF (i.e., the DIF procedure used for polytomous items) or the standardized mean difference (SMD) procedures developed by the Education Testing Service (ETS) (Zwick, Donoghue, & Grima, 1993; Allen, Carlson, & Zalanak, 1999) for polytomous items were used for identifying tasks that exhibit DIF. JMetrik (Meyer, 2014), an open source computer program for psychometric analysis, was used in conducting the analyses. The procedures first calculate the Mantel statistic and determine its probability of 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, the SMD between the performances of the two groups being compared is calculated. The SMD compares the means of the two groups, adjusting for differences in the distribution of the two groups being compared across the values of the total raw scores. To standardize the outcome, this difference is divided by the standard deviation (SD) of the task for the total group. The ratio of SMD over SD serves as an effect size measure for the Mantel Chi-

square statistic. Since this effect size measure can be positive or negative which may present some challenges when interpreting them, it is divided by the item score range in JMetrik (Meyer, 2014) such that the range of the rescaled effect size (called standardized P-DIF* on the JMetrik DIF output) is restricted to 0 and 1. The effect size flagging criterion for polytomous items, proposed by ETS (Allen, Carlson, & Zalanak, 1999) was also rescaled to the standardized P-DIF* metric (Meyer, 2014).

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

- AA (no DIF), when the Mantel Chi-square statistic is not significant or when it is significant and standardized P-DIF* is less than 0.05
- BB (weak DIF), when the Mantel Chi-square statistic is significant and standardized P-DIF* is greater than or equal to 0.05 but less than 0.10
- CC (strong DIF), when the Mantel Chi-square statistic is significant and standardized P-DIF* is greater than or equal to 0.10

5.1.5 Reliability of Composites

Four composite scores are reported for Alternate ACCESS: Oral Language Composite (oral), Literacy Composite (litr), Comprehension Composite (cphn), and Overall Composite (over). To estimate the reliability of these composite scores, a stratified Cronbach’s alpha coefficient (e.g., Kamata, Turhan, & Darandari, 2003; April, Kane, & Case, 2004; Rudner, 2001) is computed, weighted by the contribution of each domain score into the composite. Specifically, the formula is

$$\alpha_c = 1 - \frac{\sum_{j=1}^k w_j^2 \sigma_j^2 (1 - \rho_j)}{\sigma_c^2}$$

Where

k = number of components j

w_j = domain weight of component j

σ_j^2 = variance of component j

σ_c^2 = variance of composite

ρ_j = reliability coefficient of component j .

The data to compute the stratified Cronbach’s alpha is provided in the appropriate tables in Chapter 6.

5.1.6 Accuracy and Consistency of Classification

For each domain across grade-level clusters, as well as for the four composite scores, tables were produced that indicate estimates of the accuracy and consistency of classification of examinees into the Alternate ACCESS for ELLs language proficiency levels based on their performances on the test. It is important to know the reliability of any student’s test score and the degree of precision with which it has been measured (i.e., the estimate of the invariant standard error of

measure [SEM] of classical test theory and the estimate of the variable conditional standard error of the Rasch measurement model). However, because decisions about students are ultimately made on the basis of their classification into language proficiency levels on the basis of their performance on Alternate ACCESS for ELLs[®], it is important to know how well these classifications are made. The analyses that we employed make use of the methods outlined and implemented in Livingston and Lewis (1995) and Young and Yoon (1998) as implemented in the software program BB-CLASS (Brennan, 2004) (cf. also Lee, Hanson, & Brennan, 2002).

In the approach of Livingston and Lewis (1995), the accuracy of a decision is the extent to which decisions made on the basis of the administered test (i.e., the observed scores) would agree with the decisions that would be made if each student could somehow be tested with all possible parallel forms of the assessments; that is, decisions based on the examinees' "true score." On the other hand, the consistency of a decision is the extent to which decisions made on the basis of the administered test would agree with the decisions that would be made if the students had taken a different but parallel form of the test. Thus, in every analysis of classification, two parallel analyses are made: accuracy (that is, vis-à-vis "true scores") and consistency (that is, vis-à-vis a second form).

In terms of classifications around a single cut point, students can be misclassified in one of two ways. Students who were below the proficiency cut score (based on their "true score"), but were classified on the basis of the assessment as being above the cut score, are considered to be false positives. Students who were above the proficiency cut score (based on their "true score"), but were classified as being below a cut score, are considered to be false negatives. All other students are considered to be accurately placed either above or below the cut score.

Since a 'true score' is a theoretical construct, it is unknown for any given student. The approach taken by Livingston and Lewis (1995) and implemented here *to model true scores* uses information about the reliability of the test, the cut scores, and the observed distribution of scores. Then, using a four-parameter beta distribution, we modeled the distribution of the true scores and of scores on a parallel form. Overall accuracy and consistency indices are produced by comparing the percentage of students classified across all categories the same way by both the observed distribution and modeled distribution. These indices indicate the percent of all students who would be classified into the same language proficiency level by both the administered test and either the true score distribution (accuracy) or a parallel test (consistency). Our tables also provide an estimate of Cohen's kappa statistic, which is a very conservative estimate of the overall classification since it corrects for chance.

We also look at accuracy and consistency conditional on the language proficiency level. These indices examine the percent of students classified by both tests into a level divided by all students classified into that level according either to the true score distribution (accuracy) or based on a parallel test (consistency).

Finally, we look at what may be the most important set of indices, which are the indices at the cut points. That is, at every cut point, using the true score distribution (e.g., accuracy), we provide the percent of students who are consistently placed above and below the cut score, as well as those who are false positives and false negatives. For consistency, only the percent of students classified consistently above and below the cut score is calculated. Thus, for example, to evaluate the degree of confidence that one can have in a decision made based on the Overall Composite score as to whether students are being accurately classified into Alternate WIDA language proficiency level

P2 (“Beginning”) or not, one can look at the accuracy index provided in the table for the cut score P1/P2.

5.2 Descriptions

The following paragraphs describe the tables and figures that appear in Chapter 6. Each description applies to each test form in each domain. Information on raw and scale score descriptive statistics, proficiency level distribution, and the equating summary, are displayed in tables/figures A-D.

Reliability, item analysis summary, complete item analysis, raw score to scale score conversion, and raw score to proficiency level conversion tables are provided in tables E-I. These tables are organized by: grade, grade-level cluster, domain, domain and composite scores.

Note that because the composite scores do not have raw scores associated with them, any table or figure that draws on raw scores is not included for the composite scores. This includes Table A, Table D, Table F, Table G, Table H and Table I, and Figure A, Figure D and Figure E.

5.2.1 Raw Score Information (Figure A and Table A)

Figure A and Table A relate to the raw scores on each test form (the raw score to proficiency level conversion table for each test form is displayed in Table I in each section). All domains were scored polytomously. The highest possible score for Listening and Reading is 36 (4 points per item for 9 items). The highest possible score for Speaking is 16 (2 points per item for 8 items). The highest possible score for Writing is 24 (Writing parts A & B: 2 points per item for 8 items; Writing part C: 4 points per item for 2 items). For each test form, Figure A 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 were awarded each raw score.

Table A shows the following information, by each grade in the cluster and by total for the cluster:

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

5.2.2 Scale Score Information (Figure B and Table B)

Figure B and Table B relate to the *scale scores* on each test form. For each test form, raw scores were converted to vertically-equated scale scores. The raw score to scale score conversion table for each test form is displayed in Table H in each section. Thus, for each test form, Figure B shows the distribution of the scale scores. The horizontal axis shows the scale scores based on performances on the test form. The vertical axis shows the number of students (count). Each bar shows how many students were awarded each scale score.

Table B shows the following information, by each grade in the cluster and by total for the cluster:

- Number of students in the analyses

- The minimum observed scale score
- The maximum observed scale score
- The mean (average) scale score
- The standard deviation (std. dev.) of the scale scores

5.2.3 Proficiency Level Information (Figure C and Table C)

Figure C and Table C provide information on the proficiency level distribution of the students who took the test form based on their performance. Thus, for each test form, Figure C shows the information graphically for the cluster as a whole. The horizontal axis shows five out of six Alternate WIDA proficiency levels.⁴ The vertical axis shows the percent of students. Each bar shows the percent of students who were placed into each proficiency level in the domain being tested on this test form.

Table C shows the following information, by each grade in the cluster and by total for the cluster:

- The Alternate WIDA proficiency level designation (A1-A3;P1-P2)
- The number of students (count) whose performance on the test form placed them into that proficiency level in the domain being tested
- The percent of students, out of the total number of students taking the form (by grade or by total for the cluster), who were placed into that proficiency level in the domain being tested

5.2.4 Equating Summary Table (Table D)

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the Series 100 field test. Thus, the results from the original field test of Alternate ACCESS were used to determine raw-to-scale score conversions. Technical details of the analysis of this process can be found in the *Alternate ACCESS for ELLs™ Series 100 Development and Operational Field Test: Technical Report* (2013).

5.2.5 Reliability (Table E)

Table E presents reliability information based on Classical Test Theory and shows the following information:

- The number of students
- The number of items
- Cronbach’s coefficient alpha (as a measure of internal consistency)
- The classical standard error of measurement (SEM) in terms of *raw scores*

⁴In Series 403, only the Alternate WIDA proficiency levels A1, A2, A3, P1 and P2 were reported. In Series 102, the proficiency level P3 will be reported as well.

Cronbach's coefficient alpha is widely used as an estimate of reliability, particularly of the internal consistency of test items. It expresses how well the items on a test appear to measure the same construct. Conceptually, it may be thought of as the correlation obtained between performances on two halves of the test, if every possibility of dividing the test items in two were attempted. Thus, Cronbach's alpha may be low if some items are measuring something other than what the majority of the items are measuring. As with any reliability index, it is affected by the number of test items (or test score points that may be awarded). That is, all things being equal, the greater the number of items, the higher the reliability.

Cronbach's alpha is also affected by the distribution of ability within the group of students tested. All things being equal, the greater the heterogeneity of abilities within the group of students tested (i.e., the more widely the scores are distributed), the higher the reliability. In this sense, Cronbach's alpha is sample dependent. It is widely recognized that reliability can be as much a function of the test as of the sample of students tested. That is, the exact same test can produce widely disparate reliability indices based on ability distribution of the group of students tested.

The formula for Cronbach's alpha is

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_t^2} \right]$$

where

n = number of items i

σ_i^2 = variance of score on item i

σ_t^2 = variance of total score

Table E also presents the *standard error of measurement* (SEM) based on classical test theory. Unlike IRT, in this approach, SEM is seen as a constant across the spread of test scores (ability continuum). Thus, it is **not** conditional on ability being measured. It is, however, a function of two statistics: the reliability of the test and the (observed) standard deviation of the test scores. It is calculated as

$$SEM = SD \sqrt{1 - reliability}$$

Traditionally, SEM has been used to create a band around an examinee's observed score. The assertion in the view of classical test theory is that the examinee's true score (i.e., what the examinee's score would be if it could be measured without error) would lie with a certain degree of probability within this band. Therefore, the statistical expectation is that an examinee's true score has a 68% probability of lying within the band, extending from the observed score minus 1 SEM to the observed score plus 1 SEM.

5.2.6 Test Characteristic Curve (Figure D)

For each test form, Figure D graphically shows the relationship between the ability measure (in

logits) on the horizontal axis and the expected raw score on the vertical axis. Four vertical lines indicate the four cut scores, dividing the figure into five sections for each of the WIDA proficiency levels (A1-A3; P1-P2) for the domain being tested. As would be expected, higher raw scores are required to be placed into higher language proficiency levels. The relative width of each section between the cut score lines, however, gives an indication of how many points must be earned to be placed into a WIDA language proficiency level.

5.2.7 Test Information Function (Figure E)

With the Rasch measurement model, as with any measurement model following Item Response Theory (IRT), the relationship between the ability measure (in logits) and the accuracy of test scores can be modeled. It is recognized that tests measure most accurately when the abilities of the examinees and the difficulty of the items are most appropriate for each other. If a test is too difficult for an examinee (i.e., the examinee scores close to zero), or if the test is too easy for an examinee (i.e., the examinee “tops out”), accurate measurement of the examinee’s ability cannot be made. The test information function shows graphically how well the test is measuring across the ability measure spectrum. High values indicate more accuracy in measurement. Thus, for each test form, Figure E shows the relationship between the ability measure (in logits) on the horizontal axis and measurement accuracy, represented as the Fisher information value (which is the inverse squared of the standard error), on the vertical axis. The test information function, then, reflects the conditional standard error of measurement.

Again, as in Figure D, four vertical lines in Figure E indicate the four cut scores, dividing the figure into five sections for each of the WIDA language proficiency levels (A1-A3:P1-P2) for the domain being tested. It is important that each test form measure most accurately in the areas for which it is primarily used to make classification decisions. In other words, optimally the test information function should be high for the cuts between A1/A2, A2/A3, A3/P1, and P1/P2.

5.2.8 Item Analysis Summary (Table F)

Table F provides a summary of the analyses of the items. This table is divided into two parts: one, the item summary; two, the DIF summary. The upper half of the table displays the item summary. The first column in this part states the type of item (MOSR for multiple opportunities for selected response or CR for constructed response). The next columns show the number of items on the test form and average item or task difficulty value in logits, respectively. The following column displays the average percentage of maximum possible score points across items. The last two columns give information on the Rasch model fit statistics (see 5.1.1). The first is the average infit mean square statistic; the second is the average outfit mean square statistic. Optimally, these values should be close to 1.00.

The lower half of Table F provides a summary of the findings of the DIF analyses (see 5.1.4). The first column gives the DIF level: AA, BB, or CC. The next major columns show the contrasting groups in the DIF analyses: either male versus female (M/F) or Hispanic versus other ethnicities (H/O). Even though DIF may be negligible (category AA), this table shows the number of items that were favoring one group or the other at all levels of DIF. Optimally, even when items are all in category AA, there should be roughly an even number of items favoring each of the two groups to ensure that there is no systematic biasing test effect across items.

5.2.9 Complete Item Analysis Table (Table G)

Table G presents results of the analyses of all of the items or tasks on the test form. The first column provides a descriptive name of the item. The item names vary slightly across domains, consisting of characters that represent the domain (e.g., “R” for Reading), the language proficiency level targeted (e.g., “P2”), and the test series (e.g., 403).

The second column in Table G presents the item difficulty in logits, while the third column indicates whether that item served as a common item, anchoring the measurement scale to the results of the field test. The next column shows the percent of maximum possible score points (PMPS). This is obtained by dividing the average score by the maximum possible score point for that task, then multiplying by 100. It is basically a rescaling of the average score. The percentage of maximum possible score points is a common measure used to indicate the task difficulty for a polytomously scored task, with a higher value indicating an easier task. The next two columns show the Rasch fit statistics (see 5.1.1) for the item. The next column provides the point biserial correlation, a measure of the degree to which performance on an item corresponds with performance on the entire test form. In other words, it is a measure of how useful the item is at distinguishing between high-scoring and low-scoring test-takers. The following columns show the results of the two DIF analyses (see 5.2.8) for that item. These last columns are interpreted just as in Table F.

5.2.10 Complete Raw Score to Scale Score Conversion Chart (Table H)

Table H presents the raw score to scale score conversion for the test form. The first column shows all possible raw scores. The next column shows the corresponding scale score for the grade-level cluster.

The next column shows the *conditional* standard error (i.e., from the Rasch analysis) in the metric of the scale score. The last two columns show a lower bound (i.e., the scale score minus one standard error) and an upper bound (i.e., the scale score plus one standard error) around the scale score. In some cases the resulting lower bound or upper bound is below 910, which has been set as the lowest score on the scale.

All domains were adjusted for an end-of-scale effect by allowing the top scale scores to increase only at the same rate as the preceding scale scores. If they were not adjusted, their effect in the composite scores might be excessive.

Thus, if the scale scores towards the high end of the raw score scale were increasing with each raw score by 9 scale points before the group of adjusted scores, then each of the adjusted scores would increase by only 9 scale points each. Because the lower and upper bounds were calculated based on the original logit scores, these adjusted scores do not fall in the middle of the range; they fall toward the lower end of the range, but they always fall *within* the range. In other words, the adjusted scale score is a very possible observed score for that number of raw score points obtained.

In addition, at the lower end of the raw score scale, scale scores are truncated when necessary so that the lowest scale score given is the scale score corresponding to a proficiency level score of A1.

5.2.11 Raw Score to Proficiency Level Score Conversion Table (Table I)

Table I shows the interpretive proficiency level score associated with each raw score. The first column in Table I shows the raw score. The remaining columns show the proficiency level score associated with each raw score/scale score for each grade in the cluster, the percentage of students in that grade who scored at that raw score/scale score/proficiency level score, and the cumulative percentage of students in that grade who scored up to that raw score/scale score/proficiency level score.

There are two things to note about this table. First, unlike scale scores, which are determined psychometrically and have a one-to-one correspondence to raw scores regardless of the grade level of the student, proficiency level scores are interpretations of the scale score. Second, for Alternate ACCESS, cut scores between proficiency levels were determined by domain and do not change by grade level.

In students with severe cognitive disabilities, the cognitive abilities that support language proficiency development are not expected to increase dramatically from one grade level to the next. At this point in the understanding of the development of ELP in such students, it appears appropriate to use the same cut scores for all grade clusters (from grades 1 to 12) by domain. In this way, it becomes easier to detect growth in ELP from year to year for this population of English learners.

5.2.12 Accuracy and Consistency of Classification Table (Table J)

Table J presents three rows of information related to the accuracy and consistency of placement into proficiency categories based on Alternate ACCESS (see above). The first row provides overall indices related to the accuracy and consistency of classification, as well as Cohen's kappa. The second row of information shows accuracy and consistency information conditional on level. The third provides indices of classification accuracy and consistency at the cut points. These indices are perhaps the most important of all when using any of these as an absolute cut- point for placement decisions. Note that the consistency is generally higher at the cut points than over the levels. For practical purposes, the primary score used for such decisions are the Overall Composite scores. In general, the reliability and the accuracy and consistency of classification of the Overall Composite are very high for Alternate ACCESS for ELLs.

5.2.13 Conditional Standard Error of Measurement for Composite Figure (Figure F)

Figure F presents conditional standard error of measurement (CSEM) for composite score. CSEM is measurement errors computed by applying weights of individual domain scale scores in each composite score. The CSEM curves are presented by each proficiency levels in composite scores. This figure informs amount of error variability on scale score level. Higher CSEM informs more measurement error and lower CSEM indicates more reliability.

6. Analyses of Test Forms: Results

6.1 Grades: 1-2

6.1.1 Listening 1-2

Figure 6.1.1A

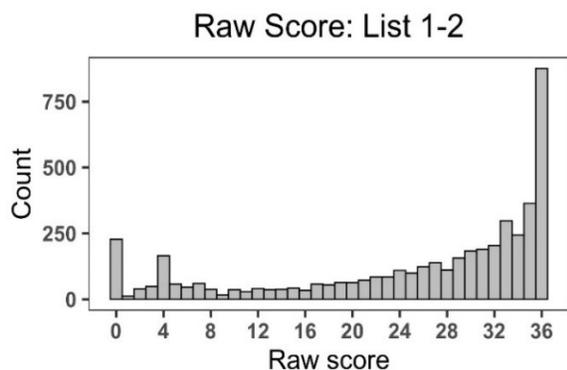


Table 6.1.1A

Raw Score Descriptive Statistics: List 1-2

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	2,255	0	36	23.98	11.73
2	2,295	0	36	26.79	11.01
Total	4,550	0	36	25.40	11.46

Figure 6.1.1B

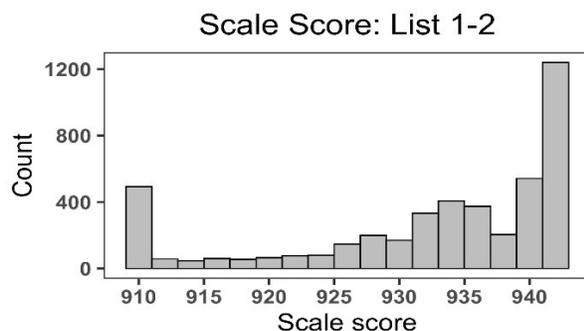


Table 6.1.1B

Scale Score Descriptive Statistics: List 1-2

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	2,255	910	943	931.53	11.02
2	2,295	910	943	934.16	10.34
Total	4,550	910	943	932.86	10.76

Figure 6.1.1C

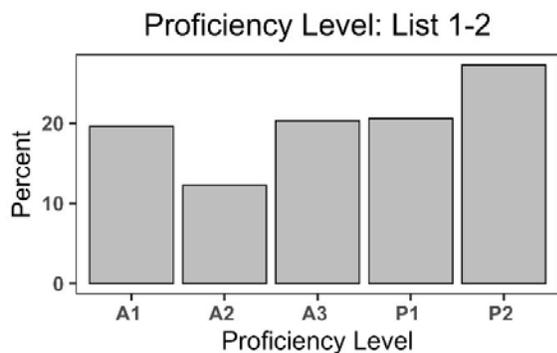


Table 6.1.1C

Proficiency Level Distribution: List 1-2

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
A1	519	23.02	373	16.25	892	19.60
A2	329	14.59	229	9.98	558	12.26
A3	469	20.80	455	19.83	924	20.31
P1	451	20.00	485	21.13	936	20.57
P2	487	21.60	753	32.81	1,240	27.25
Total	2,255	100.00	2,295	100.00	4,550	100.00

Table 6.1.1D

Equating Summary: List 1-2

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 102. Thus, the results from the S102 of the Alternate ACCESS were used to determine raw-to-scale score conversions.

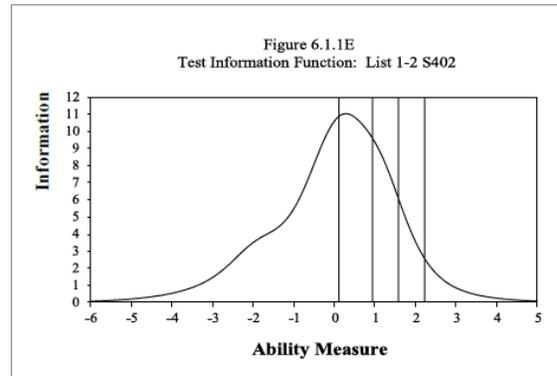
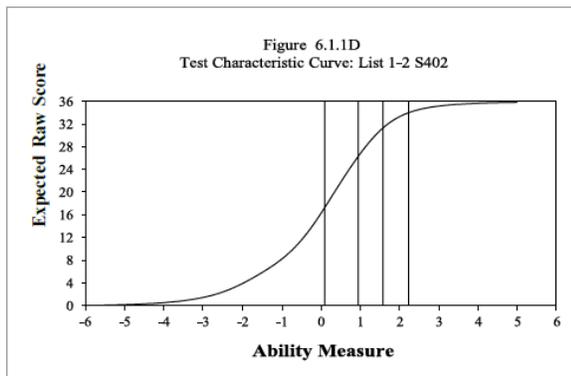


Table 6.1.1E

Reliability: List 1-2

No. of Students	No. of Items	Cronbach's Alpha	SEM
4,194	9	0.918	2.680

Table 6.1.1F

Item Analysis Summary: List 1-2

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	MOSR	9	0.00	71.17%	1.18	1.58
DIF Summary	DIF Level		Male/Female		Hispanic/Other	
			Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	3	6	6	3	
	BB	0	0	0	0	
CC	0	0	0	0		

Table 6.1.1G

Complete Item Analysis: List 1-2

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1. L1_A1_103	-1.59		86.00%	2.90	3.78	0.59	AA	F	AA	H
2. L2_A2_103	0.43	Yes	67.00%	1.33	1.48	0.78	AA	M	AA	O
3. L3_A2_103	0.17	Yes	73.25%	1.06	0.99	0.84	AA	M	AA	O
4. L4_A3_103	-0.56	Yes	75.75%	0.93	0.69	0.86	AA	F	AA	H
5. L5_A3_103	-0.12	Yes	73.50%	0.79	0.62	0.88	AA	F	AA	O
6. L6_P1_103	0.75	Yes	60.00%	1.10	1.08	0.79	AA	F	AA	H
7. L7_P1_103	0.98	Yes	58.50%	0.98	0.90	0.80	AA	M	AA	O
8. L8_P2_103	0.41	Yes	63.25%	0.85	0.65	0.85	AA	M	AA	O
9. L9_P2_103	0.96	Yes	55.00%	1.04	0.94	0.77	AA	F	AA	H

Table 6.1.1H

Raw Score to Scale Score Conversion: List 1-2

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	14.80	910.00^	910.00^
1	910^	8.23	910.00^	910.00^
2	910^	5.70	910.00^	910.00^
3	910^	4.67	910.00^	911.68
4	910^	4.19	910.00^	913.58
5	912	3.96	910.00^	915.48
6	914	3.88	910.00^	917.38
7	915	3.72	911.60	919.04
8	917	3.48	913.50	920.47
9	918	3.24	915.16	921.65
10	920	3.09	916.59	922.76
11	921	2.85	917.93	923.63
12	922	2.77	919.04	924.58
13	923	2.61	920.15	925.37
14	924	2.53	921.02	926.08
15	924	2.45	921.89	926.80
16	925	2.37	922.76	927.51
17	926	2.37	923.47	928.22
18	927	2.37	924.19	928.93
19	927	2.37	924.82	929.57
20	928	2.37	925.53	930.28
21	929	2.37	926.24	930.99
22	929	2.37	926.96	931.70
23	930	2.37	927.67	932.42
24	931	2.37	928.38	933.13
25	931	2.45	929.01	933.92
26	932	2.45	929.80	934.71
27	933	2.53	930.52	935.58
28	934	2.61	931.31	936.53
29	935	2.69	932.10	937.48
30	936	2.85	932.89	938.59
31	937	3.01	933.76	939.77
32	938	3.32	934.71	941.36
33	940	3.80	935.82	943.41
34	942*	4.67	937.16	946.50
35	944*	6.96	938.90	952.83
36	946*	13.85	940.17	967.87

^ Truncated

* Adjusted for end of scale effect

Table 6.1.11

Raw Score to Proficiency Level Conversion: List 1-2

Raw Score	Grade 1			Grade 2		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	5.68	5.68	A1	4.36	4.36
1	A1	0.40	6.08	A1	0.13	4.49
2	A1	0.75	6.83	A1	0.96	5.45
3	A1	1.37	8.20	A1	0.78	6.23
4	A1	4.39	12.59	A1	2.88	9.11
5	A1	1.60	14.19	A1	0.96	10.07
6	A1	1.11	15.30	A1	0.92	10.98
7	A1	1.46	16.76	A1	1.18	12.16
8	A1	1.06	17.83	A1	0.61	12.77
9	A1	0.31	18.14	A1	0.44	13.20
10	A1	1.02	19.16	A1	0.57	13.77
11	A1	0.75	19.91	A1	0.52	14.29
12	A1	1.02	20.93	A1	0.78	15.08
13	A1	1.02	21.95	A1	0.57	15.64
14	A1	1.06	23.02	A1	0.61	16.25
15	A2	1.15	24.17	A2	0.70	16.95
16	A2	0.75	24.92	A2	0.74	17.69
17	A2	1.55	26.47	A2	1.00	18.69
18	A2	1.15	27.63	A2	1.26	19.96
19	A2	1.73	29.36	A2	1.09	21.05
20	A2	1.86	31.22	A2	0.92	21.96
21	A2	2.13	33.35	A2	1.05	23.01
22	A2	2.44	35.79	A2	1.31	24.31
23	A2	1.82	37.61	A2	1.92	26.23
24	A3	2.13	39.73	A3	2.70	28.93
25	A3	2.17	41.91	A3	2.18	31.11
26	A3	2.84	44.75	A3	2.61	33.73
27	A3	3.33	48.07	A3	2.79	36.51
28	A3	2.39	50.47	A3	2.48	39.00
29	A3	3.77	54.24	A3	3.14	42.14
30	A3	4.17	58.40	A3	3.92	46.06
31	P1	4.12	62.53	P1	4.23	50.28
32	P1	4.66	67.18	P1	4.31	54.60
33	P1	6.30	73.48	P1	6.80	61.39
34	P1	4.92	78.40	P1	5.80	67.19
35	P2	6.92	85.32	P2	9.06	76.25
36	P2	14.68	100.00	P2	23.75	100.00

Table 6.1.1J

Accuracy and Consistency of Classification Indices: List 1-2

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.680	0.571		0.448	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.881		0.148	
	A2	0.620		0.274	
	A3	0.594		0.208	
	P1	0.333		0.228	
	P2	0.761		0.704	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.959	0.026	0.015	0.940
	A2/A3	0.931	0.033	0.036	0.909
	A3/P1	0.917	0.016	0.067	0.885
	P1/P2	0.852	0.057	0.091	0.778

6.1.2 Reading 1-2

Figure 6.1.2A

Raw Score: Read 1-2

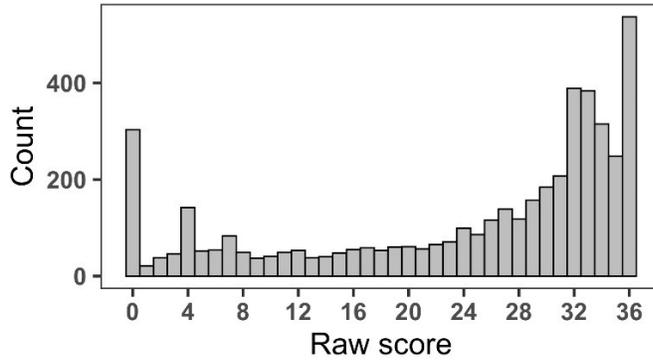


Table 6.1.2A

Raw Score Descriptive Statistics: Read 1-2

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	2,255	0	36	22.80	11.98
2	2,298	0	36	25.67	11.35
Total	4,553	0	36	24.25	11.76

Figure 6.1.2B

Scale Score: Read 1-2

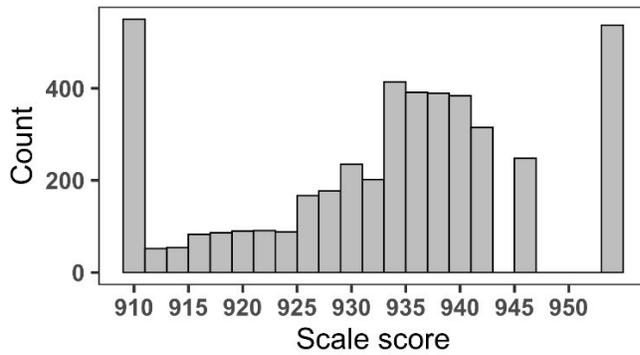


Table 6.1.2B

Scale Score Descriptive Statistics: Read 1-2

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	2,255	910	954	931.50	12.64
2	2,298	910	954	935.17	12.87
Total	4,553	910	954	933.35	12.89

Figure 6.1.2C

Proficiency Level: Read 1-2

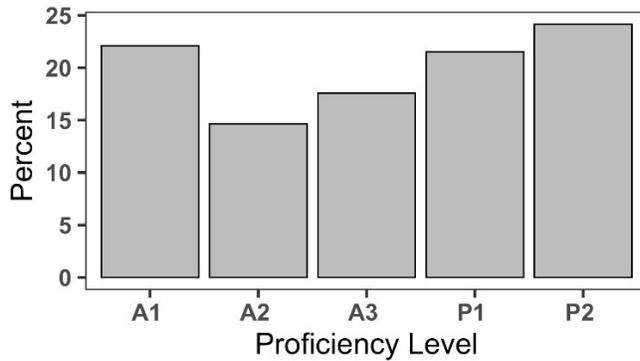


Table 6.1.2C

Proficiency Level Distribution: Read 1-2

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
A1	579	25.68	427	18.58	1,006	22.10
A2	370	16.41	297	12.92	667	14.65
A3	413	18.31	387	16.84	800	17.57
P1	486	21.55	494	21.50	980	21.52
P2	407	18.05	693	30.16	1,100	24.16
Total	2,255	100.00	2,298	100.00	4,553	100.00

Table 6.1.2D

Equating Summary: Read 1-2

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 100. Thus, the results from the original field test of the Alternate ACCESS were used to determine raw-to-scale score conversions. Technical details of the analysis of this process can be found in the Alternate ACCESS for ELLs™ Series 100 Development and Operational Field Test: Technical Report (2013).

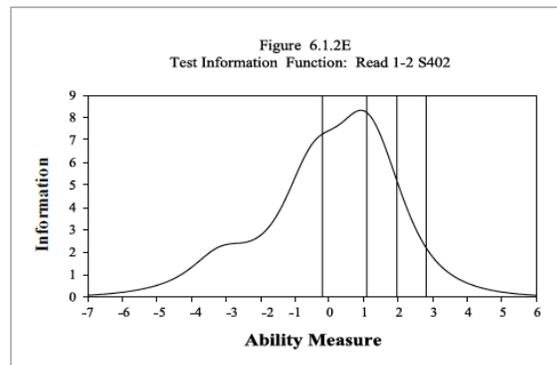
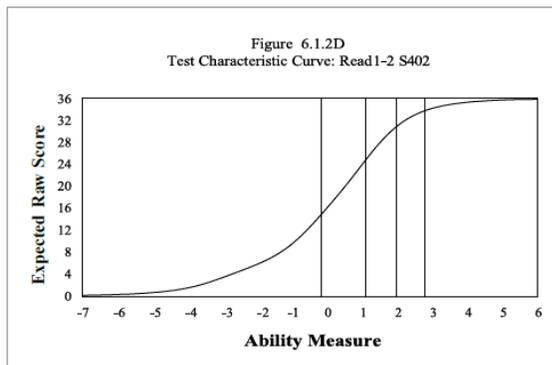


Table 6.1.2E

Reliability: Read 1-2

No. of Students	No. of Items	Cronbach's Alpha	SEM
4,116	9	0.927	3.070

Table 6.1.2F

Item Analysis Summary: Read 1-2

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	MOSR	9	0.00	66.42%	1.31	1.37

DIF Summary	DIF Level	Male/Female		Hispanic/Other	
		Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	4	5	4	5
	BB	0	0	0	0
CC	0	0	0	0	

Table 6.1.2G

Complete Item Analysis: Read 1-2

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1.R1_A1_103	-2.22		85.00%	3.02	3.61	0.60	AA	F	AA	H
2.R2_A2_103	-0.64	Yes	72.50%	1.39	1.11	0.83	AA	F	AA	O
3.R3_A2_103	-0.73	Yes	74.00%	1.26	1.04	0.83	AA	M	AA	H
4.R4_A3_103	-0.44	Yes	70.00%	1.18	0.98	0.86	AA	F	AA	O
5.R5_A3_103	0.73	Yes	59.00%	1.28	1.09	0.82	AA	M	AA	H
6.R6_P1_103	0.83	Yes	56.75%	0.94	0.83	0.85	AA	M	AA	O
7.R7_P1_103	1.01	Yes	54.25%	0.83	0.81	0.85	AA	F	AA	O
8.R8_P2_103	1.62	Yes	48.50%	0.82	0.87	0.81	AA	M	AA	H
9.R9_P2_103	0.94	Yes	56.75%	0.97	0.77	0.85	AA	M	AA	H

Table 6.1.2H

Raw Score to Scale Score Conversion: Read 1-2

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	11.27	910.00^	910.00^
1	910^	6.09	910.00^	910.00^
2	910^	4.40	910.00^	910.00^
3	910^	3.98	910.00^	911.57
4	910	3.92	910.00^	914.04
5	913	3.80	910.00^	916.39
6	915	3.50	911.33	918.32
7	917	3.13	913.49	919.76
8	918	2.83	915.24	920.91
9	919	2.65	916.69	921.99
10	920	2.47	917.95	922.90
11	921	2.35	919.04	923.74
12	922	2.29	920.00	924.58
13	923	2.23	920.91	925.37
14	924	2.23	921.69	926.15
15	925	2.17	922.59	926.93
16	926	2.17	923.38	927.72
17	926	2.17	924.16	928.50
18	927	2.11	924.94	929.16
19	928	2.11	925.67	929.89
20	929	2.11	926.45	930.67
21	929	2.05	927.17	931.27
22	930	2.05	927.90	931.99
23	931	2.05	928.56	932.66
24	931	2.05	929.28	933.38
25	932	2.05	929.95	934.04
26	933	2.11	930.61	934.83
27	934	2.17	931.33	935.67
28	934	2.23	932.06	936.51
29	935	2.29	932.84	937.42
30	936	2.47	933.62	938.56
31	937	2.65	934.53	939.83
32	938	2.95	935.49	941.40
33	940	3.37	936.70	943.44
34	942*	4.22	938.20	946.64
35	947*	6.03	940.55	952.60
36	954*	11.03	942.84	964.90

^ Truncated

* Adjusted for end of scale effect

Table 6.1.2I

Raw Score to Proficiency Level Conversion: Read 1-2

Raw Score	Grade 1			Grade 2		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	7.76	7.76	A1	5.57	5.57
1	A1	0.49	8.25	A1	0.44	6.01
2	A1	0.84	9.09	A1	0.83	6.83
3	A1	1.20	10.29	A1	0.83	7.66
4	A1	4.08	14.37	A1	2.18	9.83
5	A1	1.20	15.57	A1	1.09	10.92
6	A1	1.55	17.12	A1	0.83	11.75
7	A1	2.08	19.20	A1	1.57	13.32
8	A1	1.06	20.27	A1	1.09	14.40
9	A1	0.80	21.06	A1	0.83	15.23
10	A1	1.02	22.08	A1	0.78	16.01
11	A1	1.29	23.37	A1	0.87	16.88
12	A1	1.33	24.70	A1	1.00	17.89
13	A1	0.98	25.68	A1	0.70	18.58
14	A2	0.98	26.65	A2	0.78	19.36
15	A2	0.80	27.45	A2	1.31	20.67
16	A2	1.42	28.87	A2	1.00	21.67
17	A2	1.91	30.78	A2	0.70	22.37
18	A2	1.20	31.97	A2	1.13	23.50
19	A2	1.64	33.61	A2	1.00	24.50
20	A2	1.29	34.90	A2	1.39	25.89
21	A2	1.46	36.36	A2	1.00	26.89
22	A2	1.55	37.92	A2	1.31	28.20
23	A2	1.60	39.51	A2	1.52	29.72
24	A2	2.57	42.08	A2	1.78	31.51
25	A3	1.95	44.04	A3	1.83	33.33
26	A3	2.39	46.43	A3	2.70	36.03
27	A3	3.41	49.84	A3	2.70	38.73
28	A3	2.88	52.73	A3	2.31	41.04
29	A3	3.81	56.54	A3	3.09	44.13
30	A3	3.86	60.40	A3	4.22	48.35
31	P1	5.19	65.59	P1	3.92	52.26
32	P1	7.98	73.57	P1	9.09	61.36
33	P1	8.38	81.95	P1	8.49	69.84
34	P2	5.68	87.63	P2	8.14	77.98
35	P2	4.30	91.93	P2	6.57	84.55
36	P2	8.07	100.00	P2	15.45	100.00

Table 6.1.2J

Accuracy and Consistency of Classification Indices: Read 1-2

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.705	0.619		0.521	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.891		0.130	
	A2	0.649		0.218	
	A3	0.562		0.284	
	P1	0.585		0.271	
	P2	0.749		0.677	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.957	0.024	0.018	0.938
	A2/A3	0.924	0.043	0.032	0.894
	A3/P1	0.903	0.043	0.054	0.870
	P1/P2	0.908	0.024	0.067	0.873

6.1.3 Speaking 1-2

Figure 6.1.3A

Raw Score: Spek 1-2

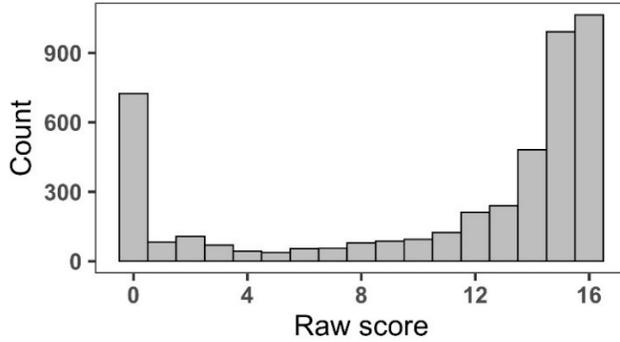


Table 6.1.3A

Raw Score Descriptive Statistics: Spek 1-2

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	2,256	0	16	10.42	6.12
2	2,294	0	16	11.38	5.85
Total	4,550	0	16	10.91	6.00

Figure 6.1.3B

Scale Score: Spek 1-2

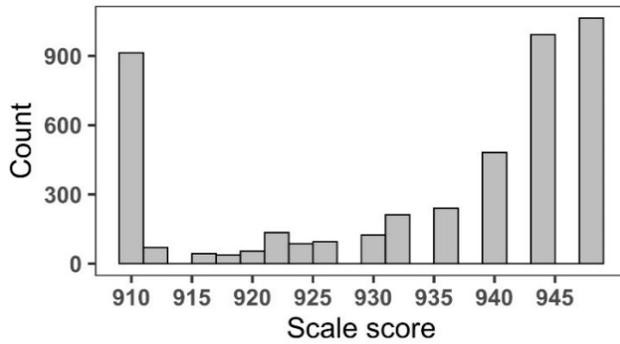


Table 6.1.3B

Scale Score Descriptive Statistics: Spek 1-2

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	2,256	910	948	932.44	14.59
2	2,294	910	948	934.98	14.18
Total	4,550	910	948	933.72	14.45

Figure 6.1.3C

Proficiency Level: Spek 1-2

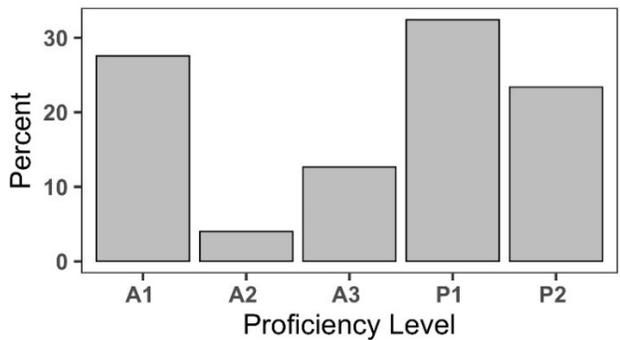


Table 6.1.3C

Proficiency Level Distribution: Spek 1-2

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
A1	690	30.59	564	24.59	1,254	27.56
A2	105	4.65	77	3.36	182	4.00
A3	293	12.99	283	12.34	576	12.66
P1	725	32.14	749	32.65	1,474	32.40
P2	443	19.64	621	27.07	1,064	23.38
Total	2,256	100.00	2,294	100.00	4,550	100.00

Table 6.1.3D

Equating Summary: Spek 1-2

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 100. Thus, the results from the original field test of the Alternate ACCESS were used to determine raw-to-scale score conversions. Technical details of the analysis of this process can be found in the Alternate ACCESS for ELLs™ Series 100 Development and Operational Field Test: Technical Report (2013).

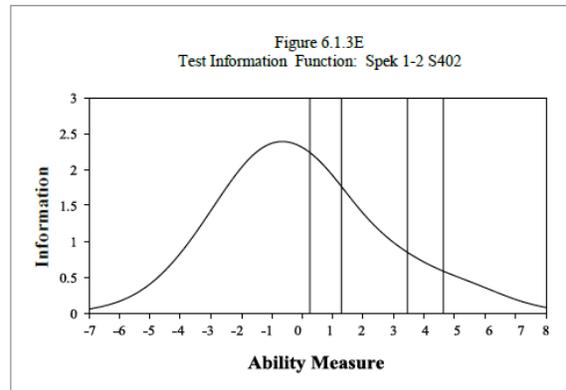
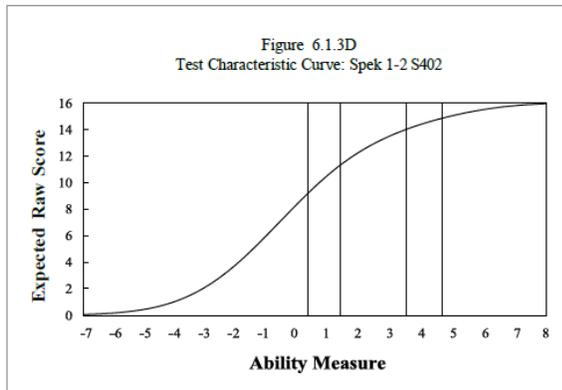


Table 6.1.3E

Reliability: Spek 1-2

No. of Students	No. of Items	Cronbach's Alpha	SEM
3,695	8	0.899	3.285

Table 6.1.3F

Item Analysis Summary: Spek 1-2

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	CR	8	0.00	67.13%	1.11	0.95
DIF Summary	DIF Level		Male/Female		Hispanic/Other	
			Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	5	3	5	3	
	BB	0	0	0	0	
CC	0	0	0	0		

Table 6.1.3G

Complete Item Analysis: Spek 1-2

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1.S1_A1_103	-2.09		75.50%	1.61	1.77	0.87	AA	F	AA	O
2.S2_A2_103	-1.04		72.00%	1.44	1.37	0.89	AA	M	AA	O
3.S3_A3_103	-0.41	Yes	70.00%	1.01	0.75	0.92	AA	M	AA	O
4.S4_A1_103	-1.20	Yes	72.00%	1.11	0.99	0.91	AA	F	AA	O
5.S5_A2_103	0.00	Yes	68.50%	1.08	0.92	0.91	AA	M	AA	H
6.S6_A3_103	-0.23	Yes	68.50%	0.81	0.58	0.93	AA	F	AA	O
7.S7_P1_103	1.51	Yes	61.00%	1.01	0.70	0.86	AA	F	AA	O
8.S8_P2_103	4.55	Yes	39.50%	0.67	1.08	0.67	AA	M	AA	H

Table 6.1.3H

Raw Score to Scale Score Conversion: Spek 1-2

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	8.42	910.00^	910.00^
1	910^	4.96	910.00^	911.23
2	910	3.86	910.00^	914.34
3	913	3.37	910.00^	916.73
4	916	3.10	912.61	918.81
5	918	2.97	914.78	920.72
6	920	2.88	916.82	922.58
7	922	2.88	918.68	924.44
8	923	2.88	920.54	926.30
9	925	2.97	922.40	928.34
10	927	3.06	924.35	930.47
11	930	3.28	926.39	932.95
12	932	3.59	928.70	935.88
13	936	4.08	931.49	939.65
14	940	4.83	935.17	944.83
15	944*	6.03	940.49	952.55
16	948*	8.95	945.50	963.41

^ Truncated

* Adjusted for end of scale effect

Table 6.1.31

Raw Score to Proficiency Level Conversion: Spek 1-2

Raw Score	Grade 1			Grade 2		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	17.60	17.60	A1	14.30	14.30
1	A1	2.04	19.64	A1	1.57	15.87
2	A1	2.75	22.38	A1	1.96	17.83
3	A1	1.55	23.94	A1	1.53	19.35
4	A1	0.98	24.91	A1	0.92	20.27
5	A1	0.93	25.84	A1	0.74	21.01
6	A1	1.29	27.13	A1	1.09	22.10
7	A1	1.51	28.63	A1	0.96	23.06
8	A1	1.95	30.59	A1	1.53	24.59
9	A2	2.26	32.85	A2	1.57	26.16
10	A2	2.39	35.24	A2	1.79	27.94
11	A3	2.62	37.85	A3	2.83	30.78
12	A3	4.83	42.69	A3	4.49	35.27
13	A3	5.54	48.23	A3	5.01	40.28
14	P1	11.57	59.80	P1	9.63	49.91
15	P1	20.57	80.36	P1	23.02	72.93
16	P2	19.64	100.00	P2	27.07	100.00

Table 6.1.3J

Accuracy and Consistency of Classification Indices: Spek 1-2

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.554	0.571		0.417	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.942		0.250	
	A2	0.505		0.147	
	A3	0.684		0.084	
	P1	0.416		0.405	
	P2	-		0.561	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.979	0.012	0.009	0.970
	A2/A3	0.974	0.012	0.014	0.965
	A3/P1	0.952	0.012	0.036	0.927
	P1/P2	0.646	0.354	0.000	0.681

6.1.4 Writing 1-2

Figure 6.1.4A

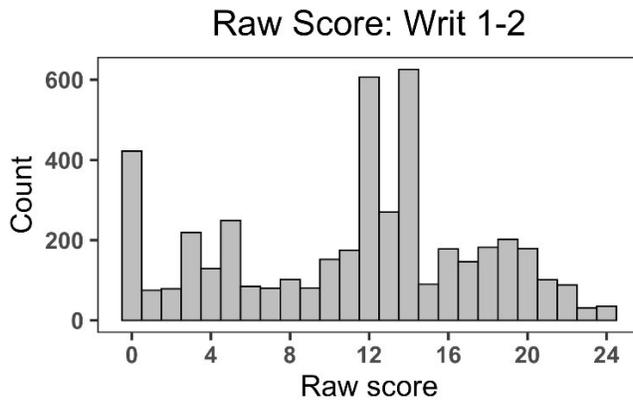


Table 6.1.4A

Raw Score Descriptive Statistics: Writ 1-2

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	2,274	0	24	10.46	6.20
2	2,309	0	24	11.89	6.42
Total	4,583	0	24	11.18	6.35

Figure 6.1.4B

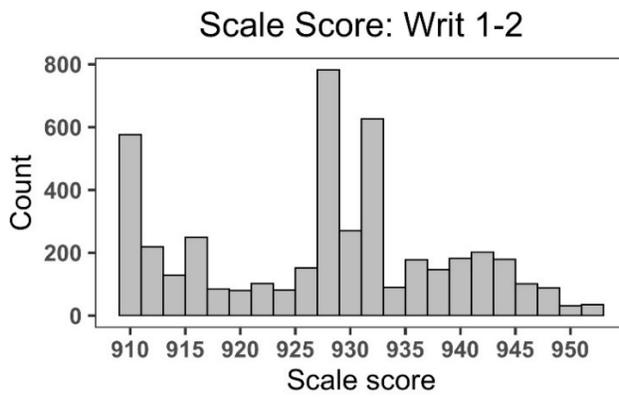


Table 6.1.4B

Scale Score Descriptive Statistics: Writ 1-2

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	2,274	910	953	926.95	11.05
2	2,309	910	953	929.60	11.51
Total	4,583	910	953	928.29	11.36

Figure 6.1.4C

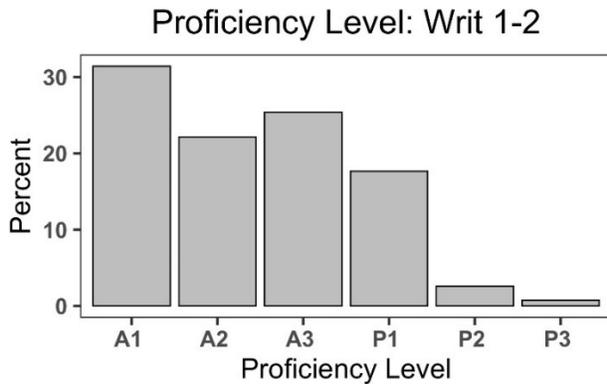


Table 6.1.4C

Proficiency Level Distribution: Writ 1-2

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
A1	790	34.74	650	28.15	1,440	31.42
A2	559	24.58	456	19.75	1,015	22.15
A3	549	24.14	615	26.63	1,164	25.40
P1	326	14.34	484	20.96	810	17.67
P2	39	1.72	80	3.46	119	2.60
P3	11	0.48	24	1.04	35	0.76
Total	2,274	100.00	2,309	100.00	4,583	100.00

Table 6.1.4D

Equating Summary: Writ 1-2

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 100. Thus, the results from the original field test of the Alternate ACCESS were used to determine raw-to-scale score conversions. Technical details of the analysis of this process can be found in the Alternate ACCESS for ELLs™ Series 100 Development and Operational Field Test: Technical Report (2013).

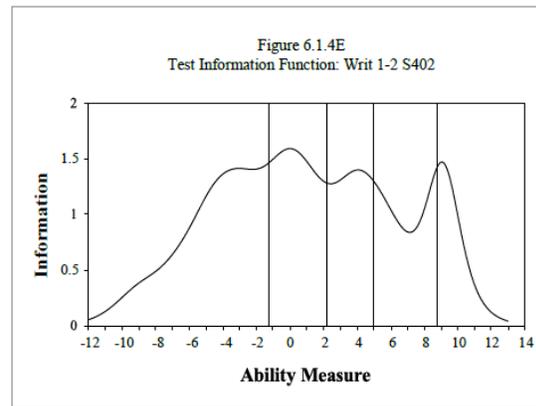
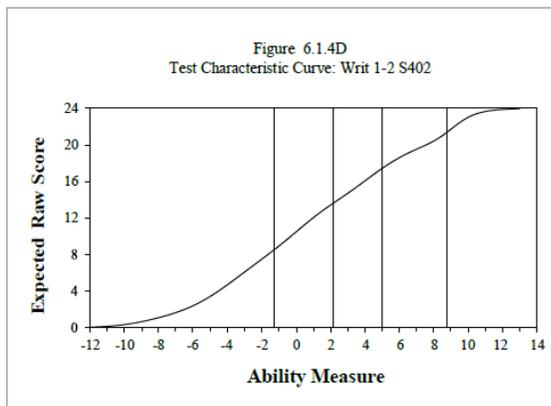


Table 6.1.4E

Reliability: Writ 1-2

No. of Students	No. of Items	Cronbach's Alpha	SEM
3,359	10	0.823	3.206

Table 6.1.4F

Item Analysis Summary: Writ 1-2

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	CR	10	0.00	52.65%	1.18	3.83
DIF Summary	DIF Level		Male/Female		Hispanic/Other	
			Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	6	4	5	5	
	BB	0	0	0	0	
	CC	0	0	0	0	

Table 6.1.4G

Complete Item Analysis: Writ 1-2

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1.W1_A1_103	-5.54		78.50%	2.00	9.90	0.69	AA	F	AA	H
2.W2_A2_103	-4.78	Yes	76.00%	1.42	9.90	0.76	AA	M	AA	O
3.W3_A3_103	-1.31	Yes	61.50%	1.41	7.92	0.81	AA	M	AA	O
4.W4_P1_103	1.73	Yes	41.00%	1.10	3.17	0.81	AA	F	AA	O
5.W5_A1_103	-2.63	Yes	67.50%	1.29	4.62	0.81	AA	F	AA	H
6.W6_A2_103	-2.18	Yes	66.00%	1.17	3.50	0.82	AA	M	AA	H
7.W7_A3_103	-0.34	Yes	54.00%	0.95	9.38	0.84	AA	M	AA	O
8.W8_P1_103	2.54	Yes	37.00%	0.73	0.96	0.83	AA	F	AA	O
9.W9_P3_103	6.84	Yes	12.75%	1.06	9.56	0.58	AA	M	AA	O
10.W10_P3_103	7.17	Yes	10.75%	1.00	9.90	0.53	AA	M	AA	H

Table 6.1.4H

Raw Score to Scale Score Conversion: Writ 1-2

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	4.99	910.00^	910.00^
1	910^	3.70	910.00^	910.00^
2	910	2.90	910.00^	912.78
3	913	2.40	910.38	915.18
4	915	2.18	912.75	917.12
5	917	2.11	914.72	918.94
6	919	2.11	916.54	920.77
7	921	2.14	918.39	922.66
8	922	2.14	920.31	924.58
9	924	2.04	922.23	926.31
10	926	1.97	923.98	927.92
11	928	1.97	925.59	929.53
12	929	2.04	927.20	931.28
13	931	2.18	928.90	933.27
14	933	2.23	930.92	935.38
15	935	2.14	933.03	937.30
16	937	2.06	934.93	939.06
17	939	2.06	936.68	940.81
18	941	2.14	938.43	942.70
19	943	2.23	940.30	944.77
20	945	2.18	942.42	946.78
21	946	2.02	944.41	948.44
22	948	2.02	946.06	950.10
23	950*	2.50	947.58	952.57
24	952*	4.34	948.63	957.32

^ Truncated

* Adjusted for end of scale effect

Table 6.1.4I

Raw Score to Proficiency Level Conversion: Writ 1-2

Raw Score	Grade 1			Grade 2		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	10.03	10.03	A1	8.40	8.40
1	A1	1.72	11.74	A1	1.56	9.96
2	A1	2.42	14.16	A1	1.04	11.00
3	A1	4.79	18.95	A1	4.76	15.76
4	A1	3.65	22.60	A1	1.99	17.76
5	A1	5.85	28.45	A1	5.02	22.78
6	A1	1.76	30.21	A1	1.95	24.73
7	A1	1.72	31.93	A1	1.78	26.50
8	A1	2.81	34.74	A1	1.65	28.15
9	A2	2.59	37.34	A2	0.95	29.10
10	A2	3.56	40.90	A2	3.07	32.18
11	A2	4.27	45.16	A2	3.38	35.56
12	A2	14.16	59.32	A2	12.34	47.90
13	A3	7.08	66.40	A3	4.72	52.62
14	A3	11.70	78.10	A3	15.59	68.21
15	A3	1.58	79.68	A3	2.34	70.55
16	A3	3.78	83.47	A3	3.98	74.53
17	P1	2.77	86.24	P1	3.59	78.13
18	P1	3.03	89.27	P1	4.89	83.02
19	P1	3.39	92.66	P1	5.41	88.44
20	P1	3.52	96.17	P1	4.29	92.72
21	P1	1.63	97.80	P1	2.77	95.50
22	P2	1.19	98.99	P2	2.64	98.14
23	P2	0.53	99.52	P2	0.82	98.96
24	P3	0.48	100.00	P3	1.04	100.00

Table 6.1.4J

Accuracy and Consistency of Classification Indices: Writ 1-2

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.733	0.650		0.535	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.899		0.147	
	A2	0.708		0.250	
	A3	0.652		0.312	
	P1	0.606		0.648	
	P2	-		0.180	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.943	0.033	0.024	0.919
	A2/A3	0.917	0.036	0.046	0.886
	A3/P1	0.903	0.026	0.070	0.865
	P1/P2	0.966	0.034	0.000	0.960

6.1.5 Oral Language Composite 1-2

Figure 6.1.5A

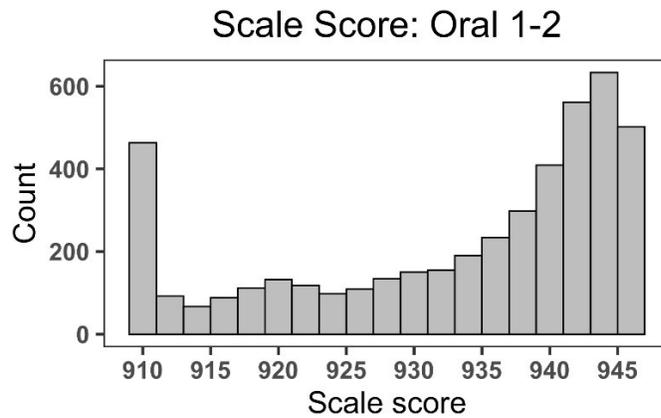


Table 6.1.5A

Scale Score Descriptive Statistics: Oral 1-2

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	2,254	910	946	932.24	12.15
2	2,291	910	946	934.85	11.62
Total	4,545	910	946	933.56	11.96

Figure 6.1.5B

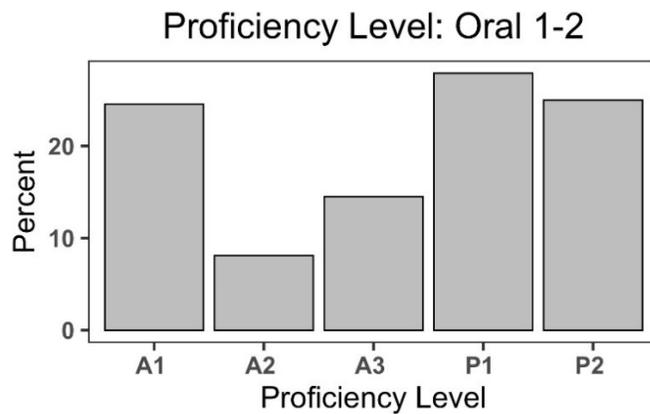


Table 6.1.5C

Proficiency Level Distribution: Oral 1-2

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
A1	628	27.86	487	21.26	1,115	24.53
A2	198	8.78	170	7.42	368	8.10
A3	350	15.53	309	13.49	659	14.50
P1	636	28.22	632	27.59	1,268	27.90
P2	442	19.61	693	30.25	1,135	24.97
Total	2,254	100.00	2,291	100.00	4,545	100.00

Table 6.1.5D

n/a

Figure 6.1.5D

n/a

Figure 6.1.5E

n/a

Table 6.1.5E

Reliability: Oral 1-2

Component	Weight	Variance	Reliability
Listening	0.5	115.348	.918
Speaking	0.5	208.225	.899
Oral		142.564	.947

*Variances from students who had results in all four domains

Table 6.1.5F

n/a

Table 6.1.5G

n/a

Table 6.1.5H

n/a

Table 6.1.5I

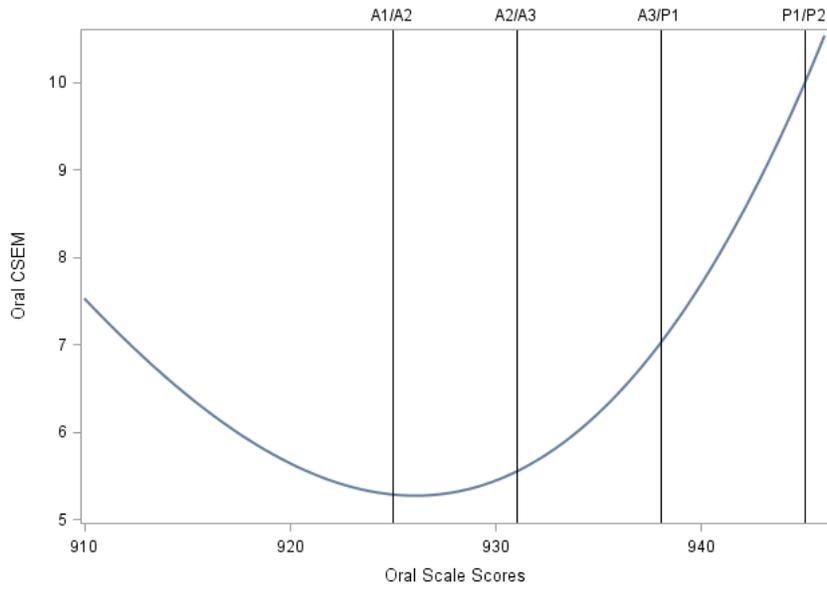
n/a

Table 6.1.5J

Accuracy and Consistency of Classification Indices: Oral 1-2

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.740	0.656		0.552	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.947		0.921	
	A2	0.639		0.517	
	A3	0.739		0.629	
	P1	0.632		0.507	
	P2	0.679		0.620	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.974	0.014	0.013	0.962
	A2/A3	0.967	0.017	0.016	0.953
	A3/P1	0.958	0.017	0.025	0.941
	P1/P2	0.841	0.067	0.092	0.794

Figure 6.1.5F CSEM for Oral Composite 1-2



6.1.6 Literacy Composite 1-2

Figure 6.1.6A

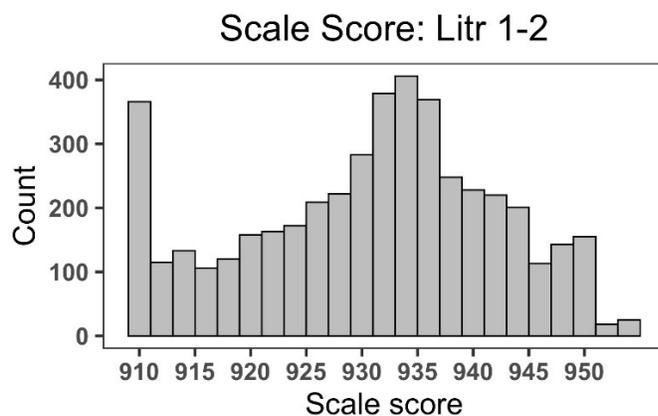


Table 6.1.6A

Scale Score Descriptive Statistics: Litr1-2

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	2,255	910	954	929.51	11.03
2	2,297	910	954	932.69	11.45
Total	4,552	910	954	931.11	11.36

Figure 6.1.6B

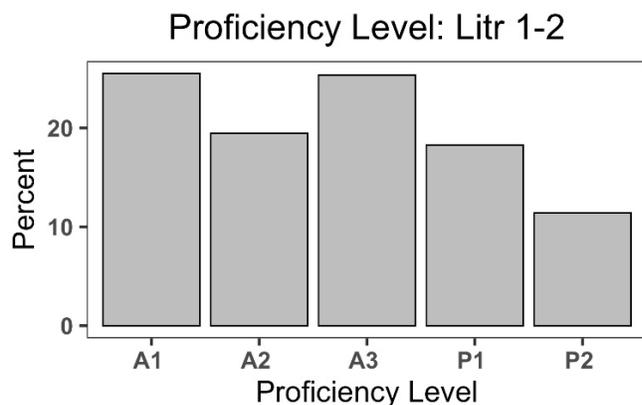


Table 6.1.6C

Proficiency Level Distribution: Litr 1-2

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
A1	646	28.65	515	22.42	1,161	25.51
A2	509	22.57	377	16.41	886	19.46
A3	569	25.23	585	25.47	1,154	25.35
P1	364	16.14	468	20.37	832	18.28
P2	167	7.41	352	15.32	519	11.40
Total	2,255	100.00	2,297	100.00	4,552	100.00

Table 6.1.6D

n/a

Figure 6.1.6D

n/a

Figure 6.1.6E

n/a

Table 6.1.6E

Reliability: Litr 1-2

Component	Weight	Variance	Reliability
Reading	0.5	165.637	.927
Writing	0.5	129.050	.823
Literacy		128.823	.932

*Variances from students who had results in all four domains

Table 6.1.6F

n/a

Table 6.1.6G

n/a

Table 6.1.6H

n/a

Table 6.1.6I

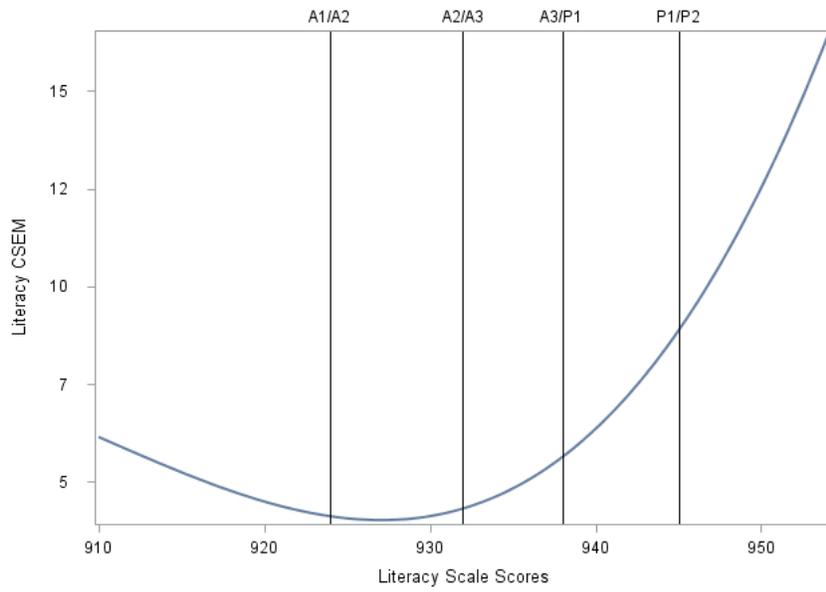
n/a

Table 6.1.6J

Accuracy and Consistency of Classification Indices: Litr 1-2

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.749	0.684		0.596	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.929		0.894	
	A2	0.736		0.635	
	A3	0.795		0.701	
	P1	0.564		0.521	
	P2	0.677		0.540	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.964	0.019	0.017	0.948
	A2/A3	0.939	0.035	0.026	0.914
	A3/P1	0.936	0.018	0.046	0.912
	P1/P2	0.910	0.081	0.009	0.906

Figure 6.1.6F CSEM for Literacy Composite 1-2



6.1.7 Comprehension Composite 1-2

Figure 6.1.7A

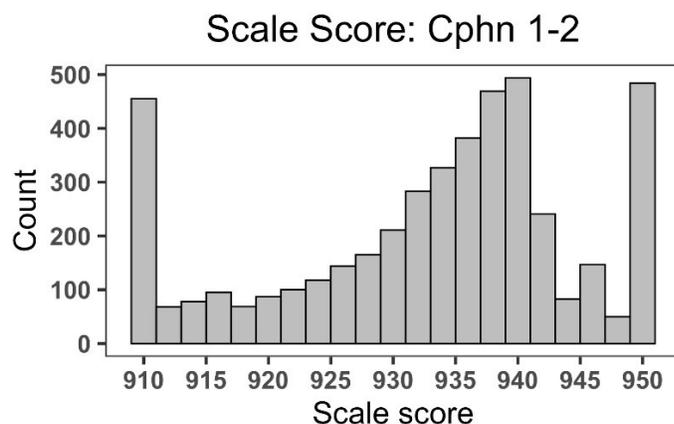


Table 6.1.7A

Scale Score Descriptive Statistics: Cphn 1-2

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	2,255	910	951	931.58	11.81
2	2,295	910	951	934.95	11.78
Total	4,550	910	951	933.28	11.91

Figure 6.1.7B

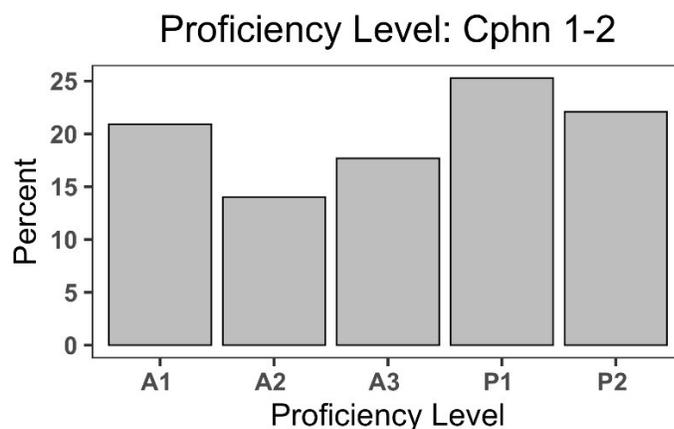


Table 6.1.7C

Proficiency Level Distribution: Cphn 1-2

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
A1	551	24.43	401	17.47	952	20.92
A2	358	15.88	280	12.20	638	14.02
A3	413	18.31	392	17.08	805	17.69
P1	565	25.06	585	25.49	1,150	25.27
P2	368	16.32	637	27.76	1,005	22.09
Total	2,255	100.00	2,295	100.00	4,550	100.00

Table 6.1.7D

n/a

Figure 6.1.7D

n/a

Figure 6.1.7E

n/a

Table 6.1.7E

Reliability: Cphn 1-2

Component	Weight	Variance	Reliability
Listening	0.3	115.348	.918
Reading	0.7	165.637	.927
Comprehension		141.610	.952

*Variances from students who had results in all four domains

Table 6.1.7F

n/a

Table 6.1.7G

n/a

Table 6.1.7H

n/a

Table 6.1.7I

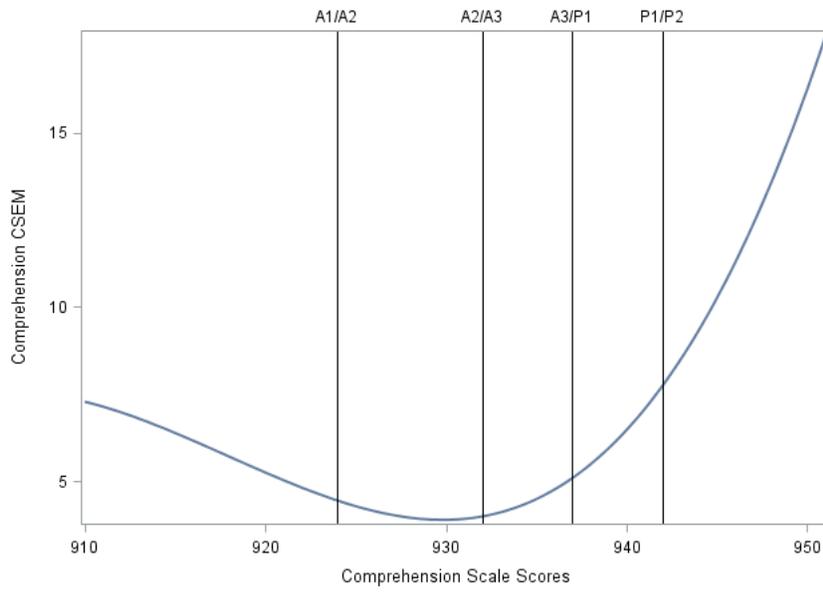
n/a

Table 6.1.7J

Accuracy and Consistency of Classification Indices: Cphn 1-2

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.750	0.666		0.580	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.922		0.883	
	A2	0.728		0.621	
	A3	0.663		0.545	
	P1	0.681		0.552	
	P2	0.730		0.674	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.971	0.016	0.012	0.958
	A2/A3	0.947	0.030	0.023	0.925
	A3/P1	0.927	0.031	0.042	0.900
	P1/P2	0.902	0.022	0.076	0.868

Figure 6.1.7F CSEM for Comprehension Composite 1-2



6.1.8 Overall Composite 1-2

Figure 6.1.8A

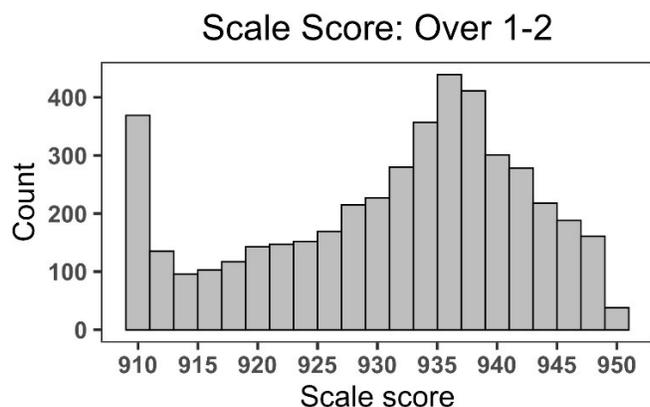


Table 6.1.8A
Scale Score Descriptive Statistics: Over 1-2

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	2,255	910	951	931.58	11.81
2	2,295	910	951	934.95	11.78
Total	4,550	910	951	933.28	11.91

Figure 6.1.8B

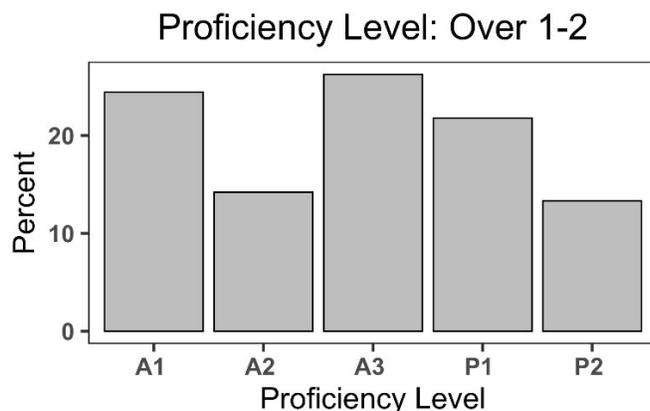


Table 6.1.8C

Proficiency Level Distribution: Over 1-2

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
A1	551	24.43	401	17.47	952	20.92
A2	358	15.88	280	12.20	638	14.02
A3	413	18.31	392	17.08	805	17.69
P1	565	25.06	585	25.49	1,150	25.27
P2	368	16.32	637	27.76	1,005	22.09
Total	2,255	100.00	2,295	100.00	4,550	100.00

Table
6.1.8D n/a

Figure 6.1.8D

n/a

Figure 6.1.8E

n/a

Table 6.1.8E

Reliability: Over 1-2

Component	Weight	Variance	Reliability
Listening	0.15	115.348	.918
Reading	0.35	165.637	.927
Speaking	0.15	208.225	.899
Writing	0.35	129.050	.823
Overall Composite		123.432	.960

*Variances from students who had results in all four domains

Table 6.1.8F

n/a

Table 6.1.8G

n/a

Table 6.1.8H

n/a

Table 6.1.8I

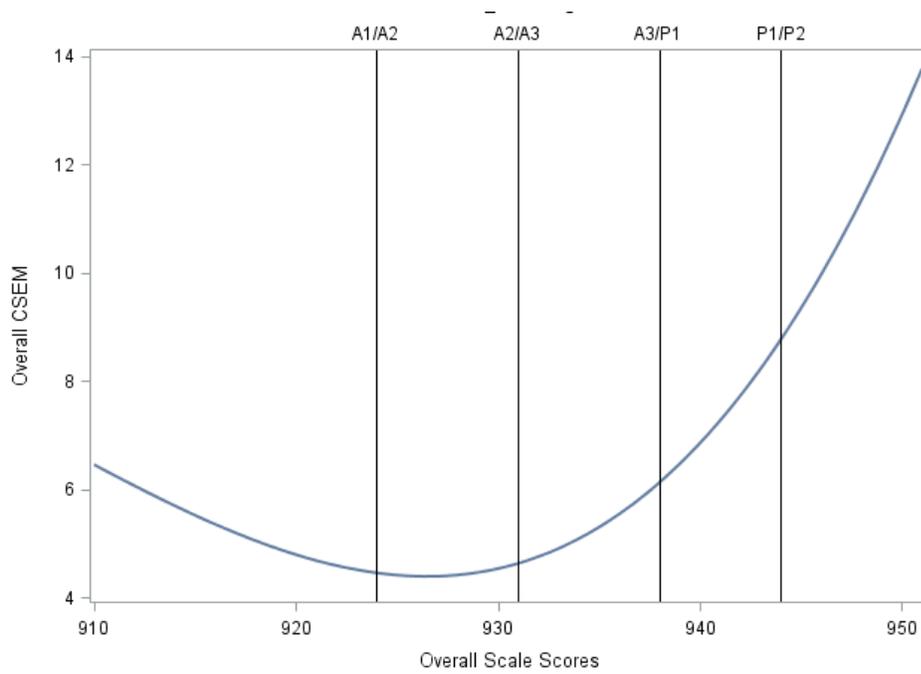
n/a

Table 6.1.8J

Accuracy and Consistency of Classification Indices: Over 1-2

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.761	0.722		0.644	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.950		0.926	
	A2	0.765		0.670	
	A3	0.875		0.813	
	P1	0.556		0.550	
	P2	-		0.539	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.976	0.013	0.011	0.966
	A2/A3	0.961	0.023	0.017	0.944
	A3/P1	0.952	0.013	0.035	0.933
	P1/P2	0.872	0.128	0.000	0.878

Figure 6.1.8F CSEM for Overall Composite 1-2



6.2 Grades: 3-5

6.2.1 Listening 3-5

Figure 6.2.1A

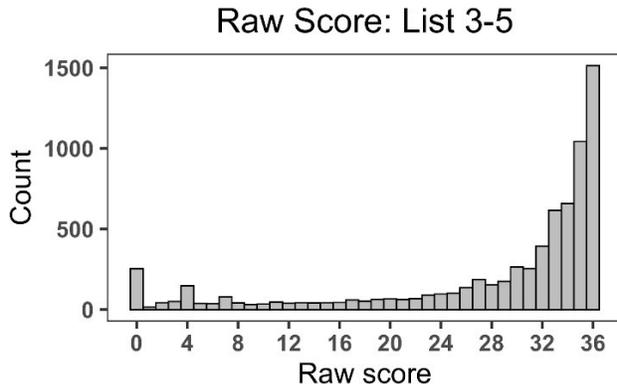


Table 6.2.1A

Raw Score Descriptive Statistics: List 3-5

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	2,409	0	36	27.39	10.62
4	2,334	0	36	28.63	10.14
5	2,347	0	36	28.91	10.07
Total	7,090	0	36	28.30	10.30

Figure 6.2.1B

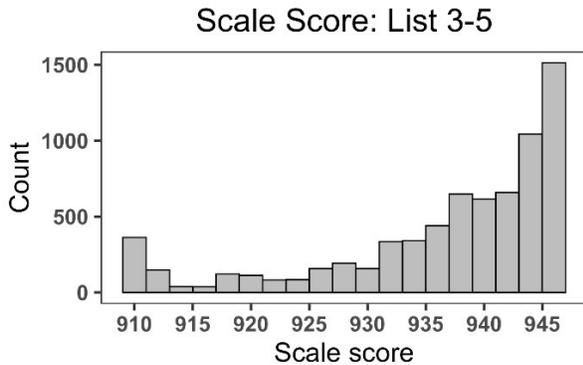


Table 6.2.1B

Scale Score Descriptive Statistics: List 3-5

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	2,409	910	947	936.31	10.75
4	2,334	910	947	937.78	10.36
5	2,347	910	947	938.12	10.39
Total	7,090	910	947	937.39	10.54

Figure 6.2.1C

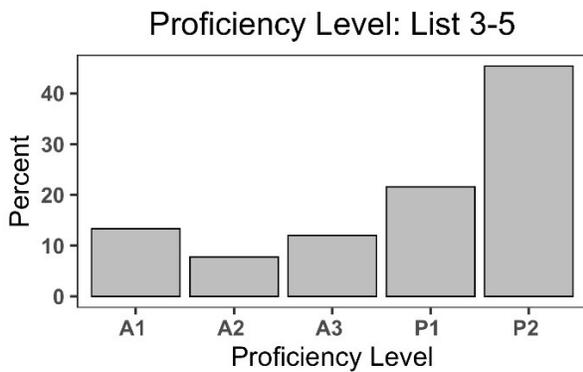


Table 6.2.1C

Proficiency Level Distribution: List 3-5

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	363	15.07	292	12.51	288	12.27	943	13.30
A2	207	8.59	182	7.80	161	6.86	550	7.76
A3	338	14.03	254	10.88	260	11.08	852	12.02
P1	555	23.04	507	21.72	467	19.90	1,529	21.57
P2	946	39.27	1,099	47.09	1,171	49.89	3,216	45.36
Total	2,409	100.00	2,334	100.00	2,347	100.00	7,090	100.00

Table 6.2.1D

Equating Summary: List 3-5

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 102. Thus, the results from the S102 of the Alternate ACCESS were used to determine raw-to-scale score conversions.

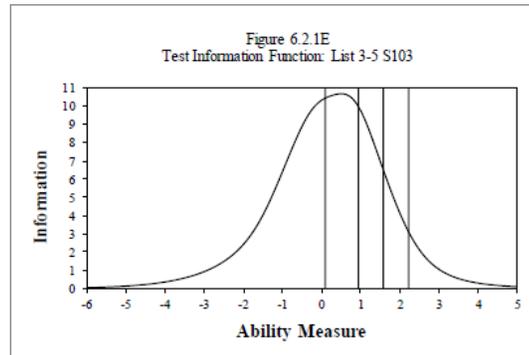
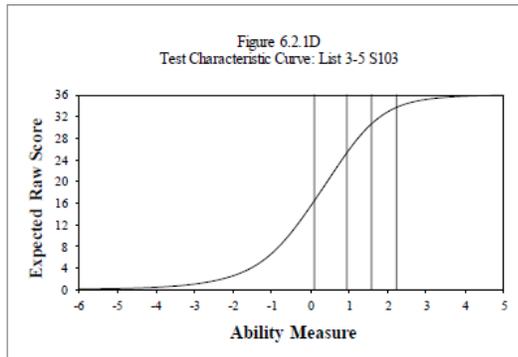


Table 6.2.1E

Reliability: List 3-5

No. of Students	No. of Items	Cronbach's Alpha	SEM
6,645	9	0.916	2.565

Table 6.2.1F

Item Analysis Summary: List 3-5

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	MOSR	9	0.00	79.97%	1.25	1.06
DIF Summary	DIF Level		Male/Female		Hispanic/Other	
			Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	4	5	6	3	
	BB	0	0	0	0	
CC	0	0	0	0		

Table 6.2.1G

Complete Item Analysis: List 3-5

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1. L1_A1_103	-1.12	Yes	90.50%	2.94	3.47	0.63	AA	F	AA	O
2. L2_A2_103	-0.31	Yes	84.25%	1.33	1.02	0.83	AA	M	AA	O
3. L3_A2_103	-0.45	Yes	85.75%	1.05	0.70	0.85	AA	F	AA	O
4. L4_A3_103	-0.43	Yes	84.25%	0.91	0.63	0.87	AA	F	AA	O
5. L5_A3_103	0.60	Yes	76.50%	1.14	0.99	0.82	AA	M	AA	O
6. L6_P1_103	0.64	Yes	73.25%	0.89	0.73	0.83	AA	F	AA	H
7. L7_P1_103	0.49	Yes	77.25%	0.85	0.66	0.87	AA	F	AA	H
8. L8_P2_103	0.76	Yes	72.50%	0.95	0.84	0.82	AA	M	AA	H
9. L9_P2_103	1.46	Yes	62.50%	1.00	1.00	0.74	AA	M	AA	H

Table 6.2.1H

Raw Score to Scale Score Conversion: List 3-5

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	14.64	910.00^	910.00^
1	910^	8.07	910.00^	910.00^
2	910^	5.70	910.00^	910.97
3	910^	4.59	910.00^	913.19
4	911	4.04	910.00^	915.01
5	913	3.64	910.00^	916.43
6	914	3.40	910.97	917.78
7	916	3.24	912.47	918.96
8	917	3.09	913.90	920.07
9	918	2.93	915.16	921.02
10	919	2.77	916.35	921.89
11	920	2.69	917.38	922.76
12	921	2.61	918.41	923.63
13	922	2.53	919.28	924.34
14	923	2.45	920.15	925.06
15	923	2.45	920.94	925.85
16	924	2.45	921.65	926.56
17	925	2.37	922.44	927.19
18	926	2.37	923.24	927.98
19	926	2.37	923.95	928.70
20	927	2.45	924.58	929.49
21	928	2.45	925.37	930.28
22	929	2.45	926.08	930.99
23	929	2.45	926.88	931.78
24	930	2.53	927.59	932.65
25	931	2.53	928.38	933.44
26	932	2.61	929.17	934.39
27	933	2.69	929.96	935.34
28	934	2.77	930.83	936.37
29	935	2.93	931.78	937.64
30	936	3.09	932.73	938.90
31	937	3.32	933.76	940.41
32	939	3.56	935.03	942.15
33	940	4.04	936.37	944.44
34	941*	4.91	937.95	947.77
35	942*	7.04	940.09	954.18
36	943*	13.85	941.44	969.13

^ Truncated

* Adjusted for end of scale effect

Table 6.2.11

Raw Score to Proficiency Level Conversion: List 3-5

Raw Score	Grade 3			Grade 4			Grade 5		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	3.74	3.74	A1	3.98	3.98	A1	3.03	3.03
1	A1	0.29	4.03	A1	0.17	4.16	A1	0.21	3.24
2	A1	0.79	4.82	A1	0.51	4.67	A1	0.51	3.75
3	A1	1.00	5.81	A1	0.43	5.10	A1	0.68	4.43
4	A1	2.41	8.22	A1	1.46	6.56	A1	2.39	6.82
5	A1	0.62	8.84	A1	0.51	7.07	A1	0.47	7.29
6	A1	0.75	9.59	A1	0.30	7.37	A1	0.51	7.80
7	A1	1.25	10.83	A1	1.11	8.48	A1	0.98	8.78
8	A1	0.58	11.42	A1	0.73	9.21	A1	0.47	9.25
9	A1	0.58	12.00	A1	0.43	9.64	A1	0.34	9.59
10	A1	0.46	12.45	A1	0.51	10.15	A1	0.47	10.06
11	A1	0.79	13.24	A1	0.47	10.63	A1	0.68	10.74
12	A1	0.37	13.62	A1	0.69	11.31	A1	0.60	11.33
13	A1	0.79	14.40	A1	0.43	11.74	A1	0.60	11.93
14	A1	0.66	15.07	A1	0.77	12.51	A1	0.34	12.27
15	A2	0.58	15.65	A2	0.56	13.07	A2	0.68	12.95
16	A2	0.58	16.23	A2	0.51	13.58	A2	0.77	13.72
17	A2	1.04	17.27	A2	0.99	14.57	A2	0.51	14.23
18	A2	0.91	18.18	A2	0.86	15.42	A2	0.47	14.70
19	A2	1.08	19.26	A2	1.03	16.45	A2	0.60	15.30
20	A2	1.08	20.34	A2	0.99	17.44	A2	0.72	16.02
21	A2	1.16	21.50	A2	0.77	18.21	A2	0.72	16.74
22	A2	0.91	22.42	A2	1.03	19.24	A2	0.89	17.64
23	A2	1.25	23.66	A2	1.07	20.31	A2	1.49	19.13
24	A3	1.45	25.11	A3	0.99	21.29	A3	1.66	20.79
25	A3	1.74	26.86	A3	1.16	22.45	A3	1.41	22.20
26	A3	2.24	29.10	A3	1.93	24.38	A3	1.58	23.78
27	A3	3.20	32.30	A3	2.31	26.69	A3	2.39	26.16
28	A3	2.66	34.95	A3	1.84	28.53	A3	2.00	28.16
29	A3	2.74	37.69	A3	2.66	31.19	A3	2.05	30.21
30	P1	3.78	41.47	P1	4.03	35.22	P1	3.37	33.57
31	P1	3.69	45.16	P1	3.81	39.03	P1	3.28	36.86
32	P1	6.19	51.35	P1	5.40	44.43	P1	5.07	41.93
33	P1	9.38	60.73	P1	8.48	52.91	P1	8.18	50.11
34	P2	9.30	70.03	P2	9.13	62.04	P2	9.42	59.52
35	P2	13.45	83.48	P2	15.47	77.51	P2	15.30	74.82
36	P2	16.52	100.00	P2	22.49	100.00	P2	25.18	100.00

Table 6.2.1J

Accuracy and Consistency of Classification Indices: List 3-5

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.483	0.506		0.338	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.899		0.158	
	A2	0.625		0.222	
	A3	0.633		0.100	
	P1	0.369		0.361	
	P2	-		0.598	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.976	0.014	0.010	0.966
	A2/A3	0.961	0.019	0.020	0.947
	A3/P1	0.937	0.011	0.052	0.910
	P1/P2	0.605	0.395	0.000	0.652

6.2.2 Reading 3-5

Figure 6.2.2A

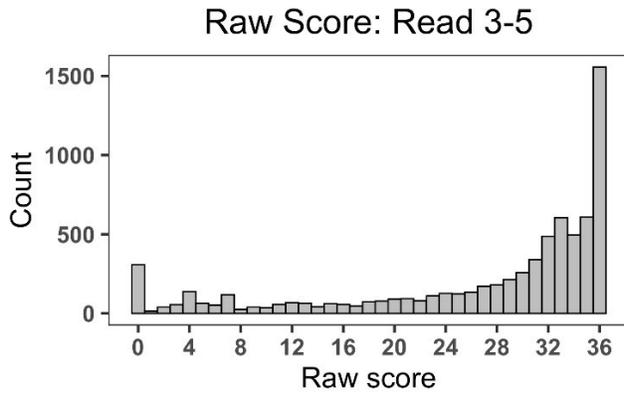


Table 6.2.2A

Raw Score Descriptive Statistics: Read 3-5

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	2,412	0	36	25.94	10.97
4	2,339	0	36	27.62	10.62
5	2,348	0	36	27.94	10.56
Total	7,099	0	36	27.16	10.76

Figure 6.2.2B

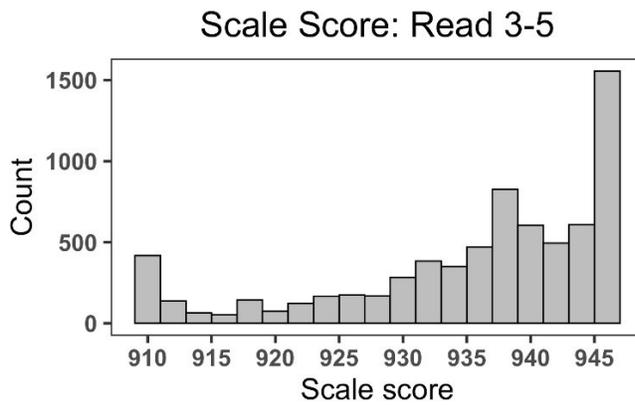


Table 6.2.2B

Scale Score Descriptive Statistics: Read 3-5

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	2,412	910	947	934.79	10.85
4	2,339	910	947	936.75	10.74
5	2,348	910	947	937.15	10.76
Total	7,099	910	947	936.21	10.83

Figure 6.2.2B

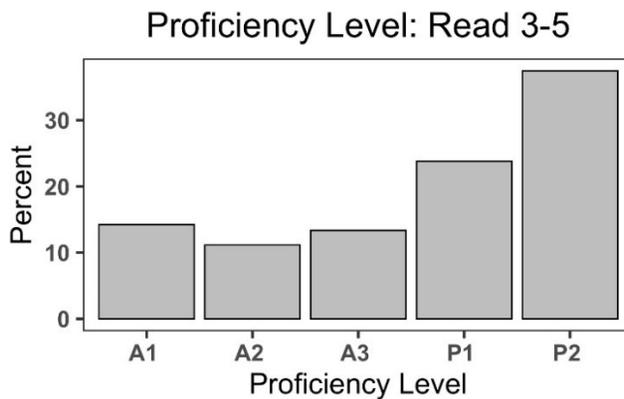


Table 6.2.2C

Proficiency Level Distribution: Read 3-5

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	387	16.04	317	13.55	307	13.07	1,011	14.24
A2	324	13.43	238	10.18	230	9.80	792	11.16
A3	357	14.80	312	13.34	279	11.88	948	13.35
P1	647	26.82	537	22.96	505	21.51	1,689	23.79
P2	697	28.90	935	39.97	1,027	43.74	2,659	37.46
Total	2,412	100.00	2,339	100.00	2,348	100.00	7,099	100.00

Table 6.2.2D

Equating Summary: Read 3-5

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 102. Thus, the results from the S102 of the Alternate ACCESS were used to determine raw-to-scale score conversions.

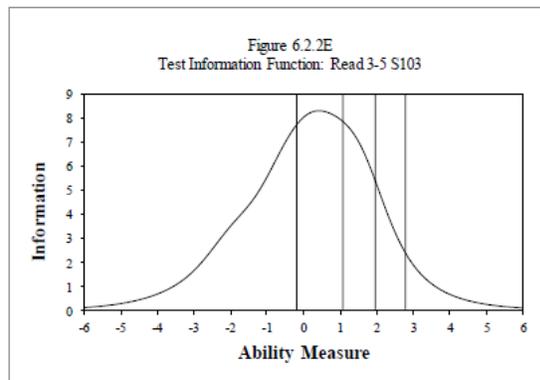
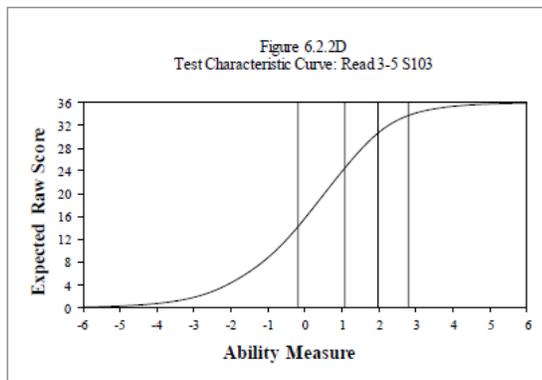


Table 6.2.2E

Reliability: Read 3-5

No. of Students	No. of Items	Cronbach's Alpha	SEM
6,638	9	0.931	2.403

Table 6.2.2F

Item Analysis Summary: Read 3-5

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	MOSR	9	0.00	75.94%	1.22	1.31
DIF Summary	DIF Level		Male/Female		Hispanic/Other	
			Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	5	4	4	5	
	BB	0	0	0	0	
	CC	0	0	0	0	

Table 6.2.2G

Complete Item Analysis: Read 3-5

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1. R1_A1_103	-2.13	Yes	90.00%	2.96	4.89	0.61	AA	F	AA	O
2. R2_A2_103	-0.44	Yes	81.25%	1.56	1.17	0.83	AA	M	AA	O
3. R3_A2_103	-0.82	Yes	82.75%	1.27	0.91	0.85	AA	F	AA	H
4. R4_A3_103	0.07	Yes	79.75%	1.22	1.00	0.86	AA	M	AA	O
5. R5_A3_103	0.11	Yes	77.50%	0.98	0.70	0.89	AA	F	AA	H
6. R6_P1_103	1.14	Yes	67.75%	0.86	0.91	0.87	AA	F	AA	O
7. R7_P1_103	0.64	Yes	72.75%	0.76	0.60	0.90	AA	M	AA	O
8. R8_P2_103	1.55	Yes	61.00%	0.72	0.69	0.83	AA	M	AA	O
9. R9_P2_103	1.45	Yes	62.75%	0.82	0.80	0.83	AA	M	AA	H

Table 6.2.2H

Raw Score to Scale Score Conversion: Read 3-5

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	11.45	910.00^	910.00^
1	910^	6.33	910.00^	910.00^
2	910^	4.52	910.00^	910.00^
3	910^	3.92	910.00^	911.57
4	910	3.74	910.00^	913.86
5	912	3.68	910.00^	916.09
6	915	3.50	911.02	918.01
7	916	3.19	913.19	919.58
8	918	2.95	915.00	920.91
9	919	2.71	916.57	921.99
10	920	2.53	917.89	922.96
11	921	2.41	919.04	923.86
12	922	2.35	920.06	924.76
13	923	2.23	921.03	925.49
14	924	2.17	921.93	926.27
15	925	2.17	922.71	927.05
16	926	2.11	923.50	927.72
17	926	2.11	924.22	928.44
18	927	2.05	925.00	929.10
19	928	2.05	925.73	929.83
20	928	2.05	926.39	930.49
21	929	2.05	927.11	931.21
22	930	2.05	927.78	931.87
23	931	2.05	928.50	932.60
24	931	2.11	929.16	933.38
25	932	2.11	929.89	934.10
26	933	2.17	930.55	934.89
27	934	2.17	931.33	935.67
28	934	2.29	932.06	936.63
29	935	2.35	932.90	937.60
30	936	2.47	933.74	938.68
31	937	2.65	934.65	939.95
32	939	2.95	935.67	941.58
33	940	3.37	936.88	943.62
34	941*	4.16	938.44	946.76
35	942*	6.03	940.67	952.72
36	943*	11.03	942.90	964.96

^ Truncated

* Adjusted for end of scale effect

Table 6.2.2I

Raw Score to Proficiency Level Conversion: Read 3-5

Raw Score	Grade 3			Grade 4			Grade 5		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	4.77	4.77	A1	4.15	4.15	A1	4.09	4.09
1	A1	0.41	5.18	A1	0.17	4.32	A1	0.04	4.13
2	A1	0.70	5.89	A1	0.38	4.70	A1	0.55	4.68
3	A1	0.87	6.76	A1	0.64	5.34	A1	0.81	5.49
4	A1	2.11	8.87	A1	1.97	7.31	A1	1.70	7.20
5	A1	1.04	9.91	A1	0.94	8.25	A1	0.72	7.92
6	A1	0.87	10.78	A1	0.56	8.81	A1	0.77	8.69
7	A1	1.82	12.60	A1	1.71	10.52	A1	1.45	10.14
8	A1	0.33	12.94	A1	0.47	10.99	A1	0.30	10.43
9	A1	0.54	13.47	A1	0.60	11.59	A1	0.51	10.95
10	A1	0.54	14.01	A1	0.34	11.93	A1	0.60	11.54
11	A1	1.04	15.05	A1	0.73	12.65	A1	0.60	12.14
12	A1	1.00	16.04	A1	0.90	13.55	A1	0.94	13.07
13	A2	1.12	17.16	A2	0.81	14.37	A2	0.77	13.84
14	A2	0.66	17.83	A2	0.60	14.96	A2	0.51	14.35
15	A2	1.08	18.91	A2	0.73	15.69	A2	0.77	15.12
16	A2	0.91	19.82	A2	0.77	16.46	A2	0.68	15.80
17	A2	0.91	20.73	A2	0.47	16.93	A2	0.55	16.35
18	A2	1.33	22.06	A2	0.94	17.87	A2	0.77	17.12
19	A2	1.24	23.30	A2	0.94	18.81	A2	1.06	18.19
20	A2	1.16	24.46	A2	1.50	20.31	A2	1.19	19.38
21	A2	1.70	26.16	A2	1.20	21.50	A2	1.02	20.40
22	A2	1.62	27.78	A2	0.86	22.36	A2	0.89	21.29
23	A2	1.70	29.48	A2	1.37	23.73	A2	1.58	22.87
24	A3	2.24	31.72	A3	1.50	25.22	A3	1.58	24.45
25	A3	1.74	33.46	A3	1.41	26.64	A3	2.09	26.53
26	A3	2.03	35.49	A3	1.92	28.56	A3	1.70	28.24
27	A3	2.49	37.98	A3	2.44	31.00	A3	2.30	30.54
28	A3	2.78	40.75	A3	2.78	33.78	A3	2.04	32.58
29	A3	3.52	44.28	A3	3.29	37.07	A3	2.17	34.75
30	P1	4.19	48.47	P1	3.25	40.32	P1	3.41	38.16
31	P1	5.47	53.94	P1	4.49	44.81	P1	4.39	42.55
32	P1	8.04	61.98	P1	6.41	51.22	P1	6.09	48.64
33	P1	9.12	71.10	P1	8.81	60.03	P1	7.62	56.26
34	P2	6.51	77.61	P2	6.33	66.35	P2	8.09	64.35
35	P2	7.17	84.78	P2	9.36	75.72	P2	9.20	73.55
36	P2	15.22	100.00	P2	24.28	100.00	P2	26.45	100.00

Table 6.2.2J

Accuracy and Consistency of Classification Indices: Read 3-5

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.527	0.528		0.389
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.896		0.125	
	A2	0.706		0.253	
	A3	0.653		0.110	
	P1	0.385		0.373	
	P2	-		0.591	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.973	0.016	0.011	0.961
	A2/A3	0.954	0.022	0.024	0.937
	A3/P1	0.939	0.012	0.049	0.912
	P1/P2	0.659	0.341	0.000	0.693

6.2.3 Speaking 3-5

Figure 6.2.3A

Raw Score: Spek 1-2

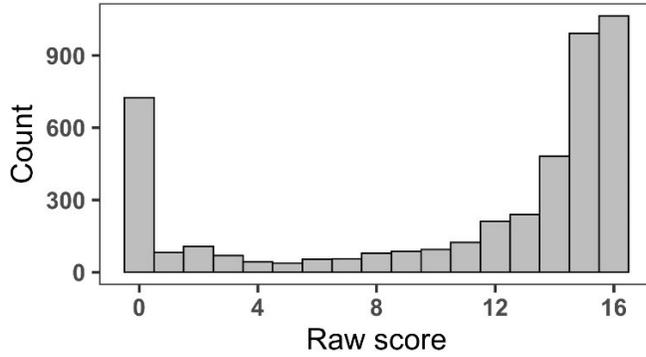


Table 6.2.3A

Raw Score Descriptive Statistics: Spek 3-5

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	2,415	0	16	11.83	5.72
4	2,339	0	16	12.30	5.53
5	2,346	0	16	12.31	5.63
Total	7,100	0	16	12.14	5.63

Figure 6.2.3B

Scale Score: Spek 1-2

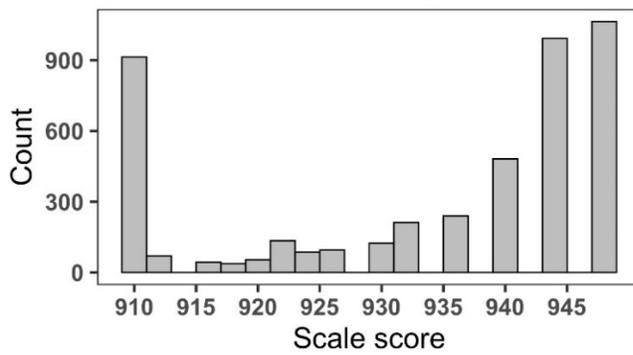


Table 6.2.3B

Scale Score Descriptive Statistics: Spek 3-5

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	2,415	910	947	935.56	13.43
4	2,339	910	947	936.87	13.14
5	2,346	910	947	936.98	13.31
Total	7,100	910	947	936.46	13.31

Figure 6.2.3C

Proficiency Level: Spek 1-2

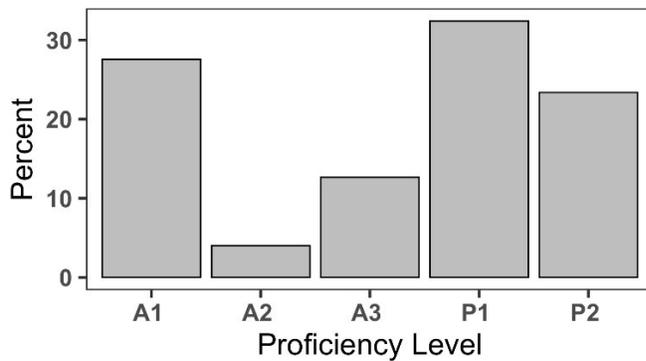


Table 6.2.3C

Proficiency Level Distribution: Spek 3-5

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	530	21.95	468	20.01	455	19.39	1,453	20.46
A2	118	4.89	99	4.23	101	4.31	318	4.48
A3	210	8.70	153	6.54	150	6.39	513	7.23
P1	794	32.88	726	31.04	696	29.67	2,216	31.21
P2	763	31.59	893	38.18	944	40.24	2,600	36.62
Total	2,415	100.00	2,339	100.00	2,346	100.00	7,100	100.00

Table 6.2.3D

Equating Summary: Spek 3-5

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 100. Thus, the results from the original field test of the Alternate ACCESS were used to determine raw-to-scale score conversions. Technical details of the analysis of this process can be found in the Alternate ACCESS for ELLs™ Series 100 Development and Operational Field Test: Technical Report (2013).

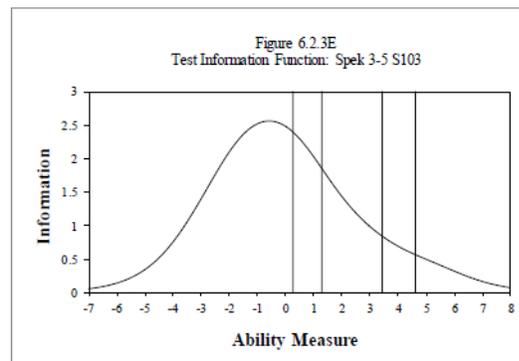
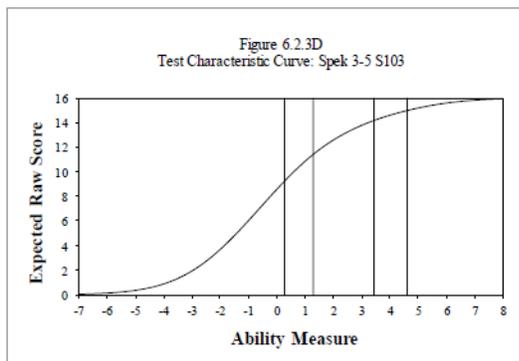


Table 6.2.3E

Reliability: Spek 3-5

No. of Students	No. of Items	Cronbach's Alpha	SEM
6,072	8	0.900	2.780

Table 6.2.3F

Item Analysis Summary: Spek 3-5

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	CR	8	0.00	75.88%	1.04	0.96
DIF Summary	DIF Level		Male/Female		Hispanic/Other	
			Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	5	3	3	5	
	BB	0	0	0	0	
CC	0	0	0	0		

Table 6.2.3G

Complete Item Analysis: Spek 3-5

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1.S1_A1_103	-2.15	Yes	82.50%	1.57	1.06	0.89	AA	F	AA	O
2.S2_A2_103	-1.31	Yes	80.50%	1.46	1.61	0.90	AA	F	AA	H
3.S3_A3_103	-0.63	Yes	79.00%	1.03	0.70	0.93	AA	M	AA	O
4.S4_A1_103	-0.84	Yes	80.00%	1.00	1.00	0.92	AA	F	AA	O
5.S5_A2_103	-0.08	Yes	78.50%	0.98	0.94	0.93	AA	M	AA	H
6.S6_A3_103	-0.38	Yes	77.50%	0.75	0.50	0.94	AA	F	AA	H
7.S7_P1_103	1.16	Yes	72.50%	0.92	0.67	0.89	AA	M	AA	O
8.S8_P2_103	4.23	Yes	54.00%	0.65	0.93	0.70	AA	M	AA	H

Table 6.2.3H

Raw Score to Scale Score Conversion: Spek 3-5

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	8.33	910.00^	910.00^
1	910^	4.88	910.00^	912.34
2	911	3.72	910.00^	915.13
3	914	3.24	910.88	917.35
4	916	3.01	913.27	919.30
5	918	2.88	915.35	921.12
6	920	2.79	917.22	922.80
7	922	2.75	919.03	924.53
8	924	2.79	920.72	926.30
9	925	2.84	922.45	928.12
10	927	2.97	924.22	930.16
11	929	3.15	926.13	932.42
12	932	3.46	928.25	935.17
13	935	3.95	930.83	938.72
14	939	4.70	934.19	943.59
15	943*	5.94	939.20	951.08
16	947*	8.95	943.99	961.90

^ Truncated

* Adjusted for end of scale effect

Table 6.2.3I

Raw Score to Proficiency Level Conversion: Spek 3-5

Raw Score	Grade 3			Grade 4			Grade 5		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	13.42	13.42	A1	11.63	11.63	A1	12.57	12.57
1	A1	1.12	14.53	A1	1.03	12.65	A1	1.41	13.98
2	A1	1.74	16.27	A1	1.07	13.72	A1	0.94	14.92
3	A1	1.37	17.64	A1	1.71	15.43	A1	1.07	15.98
4	A1	0.70	18.34	A1	0.81	16.25	A1	0.68	16.67
5	A1	0.66	19.01	A1	0.68	16.93	A1	0.60	17.26
6	A1	0.66	19.67	A1	0.90	17.83	A1	0.68	17.95
7	A1	0.95	20.62	A1	0.86	18.68	A1	0.51	18.46
8	A1	1.33	21.95	A1	1.33	20.01	A1	0.94	19.39
9	A2	1.04	22.98	A2	0.73	20.74	A2	1.41	20.80
10	A2	1.90	24.89	A2	1.97	22.70	A2	1.45	22.25
11	A2	1.95	26.83	A2	1.54	24.24	A2	1.45	23.70
12	A3	3.52	30.35	A3	2.57	26.81	A3	2.60	26.30
13	A3	5.18	35.53	A3	3.98	30.78	A3	3.79	30.09
14	P1	10.85	46.38	P1	8.89	39.68	P1	8.91	39.00
15	P1	22.03	68.41	P1	22.15	61.82	P1	20.76	59.76
16	P2	31.59	100.00	P2	38.18	100.00	P2	40.24	100.00

Table 6.2.3J

Accuracy and Consistency of Classification Indices: Spek 3-5

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.563	0.584		0.435
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.946		0.236	
	A2	0.522		0.148	
	A3	0.721		0.078	
	P1	0.421		0.406	
	P2	-		0.572	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.980	0.012	0.008	0.972
	A2/A3	0.975	0.012	0.013	0.966
	A3/P1	0.959	0.011	0.030	0.939
	P1/P2	0.646	0.354	0.000	0.687

6.2.4 Writing 3-5

Figure 6.2.4A

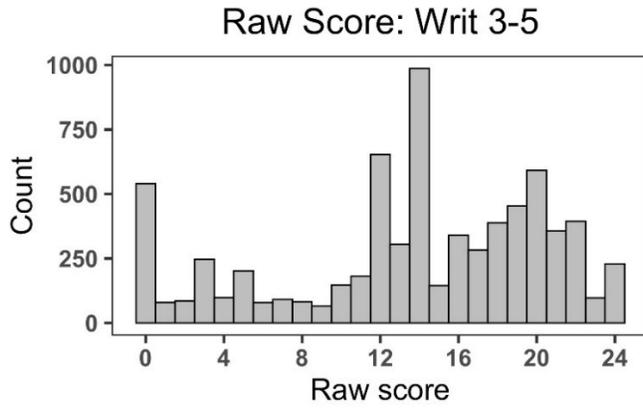


Table 6.2.4A

Raw Score Descriptive Statistics: Writ 3-5

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	2,433	0	24	12.84	6.55
4	2,340	0	24	14.01	6.71
5	2,348	0	24	14.29	6.90
Total	7,121	0	24	13.70	6.75

Figure 6.2.4B

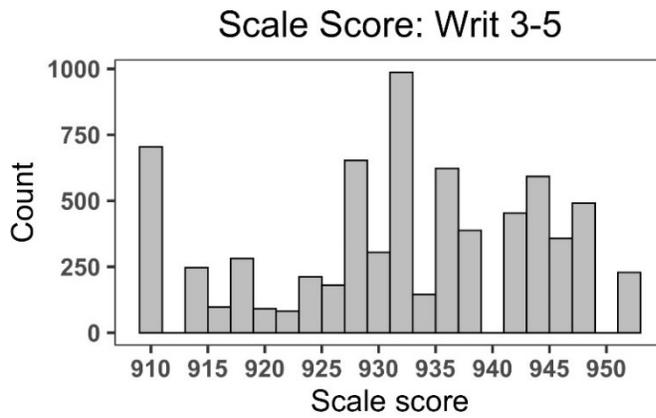


Table 6.2.4B

Scale Score Descriptive Statistics: Writ 3-5

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	2,433	910	953	930.75	11.44
4	2,340	910	953	932.92	11.92
5	2,348	910	953	933.55	12.26
Total	7,121	910	953	932.39	11.93

Figure 6.2.4C

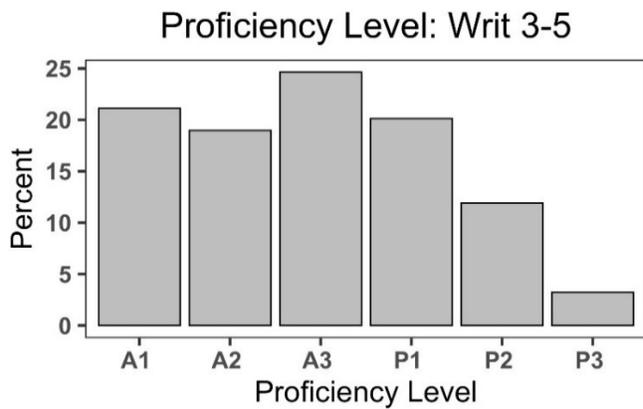


Table 6.2.4C

Proficiency Level Distribution: Writ 3-5

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	576	23.67	455	19.44	473	20.14	1,504	21.12
A2	527	21.66	445	19.02	379	16.14	1,351	18.97
A3	654	26.88	573	24.49	528	22.49	1,755	24.65
P1	420	17.26	477	20.38	537	22.87	1,434	20.14
P2	209	8.59	312	13.33	327	13.93	848	11.91
P3	47	1.93	78	3.33	104	4.43	229	3.22
Total	2,433	100.00	2,340	100.00	2,348	100.00	7,121	100.00

Table 6.2.4D

Equating Summary: Writ 3-5

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 100. Thus, the results from the original field test of the Alternate ACCESS were used to determine raw-to-scale score conversions. Technical details of the analysis of this process can be found in the Alternate ACCESS for ELLs™ Series 100 Development and Operational Field Test: Technical Report (2013).

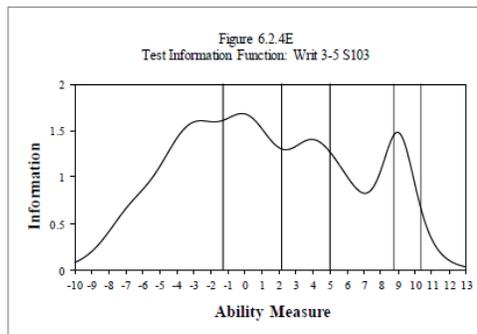
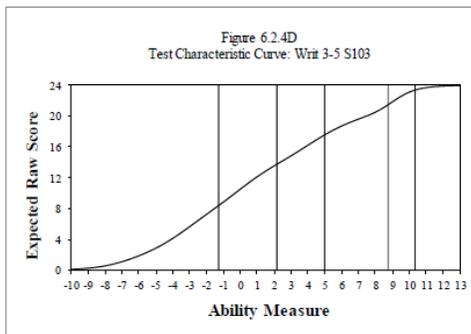


Table 6.2.4E

Reliability: Writ 3-5

No. of Students	No. of Items	Cronbach's Alpha	SEM
5,752	10	0.795	3.742

Table 6.2.4F

Item Analysis Summary: Writ 3-5

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	CR	10	0.00	63.93%	1.31	5.12
DIF Summary	DIF Level		Male/Female		Hispanic/Other	
			Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	4	6	4	6	
	BB	0	0	0	0	
CC	0	0	0	0		

Table 6.2.4G

Complete Item Analysis: Writ 3-5

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1.W1_A1_103	-5.15		85.00%	2.16	9.90	0.71	AA	M	AA	H
2.W2_A2_103	-4.34	Yes	83.50%	1.36	9.90	0.76	AA	F	AA	H
3.W3_A3_103	-1.34	Yes	74.00%	1.52	7.35	0.81	AA	M	AA	H
4.W4_P1_103	1.66	Yes	59.00%	1.12	2.45	0.85	AA	F	AA	O
5.W5_A1_103	-2.59	Yes	79.00%	1.31	8.91	0.79	AA	F	AA	H
6.W6_A2_103	-2.03		78.00%	1.23	5.78	0.80	AA	M	AA	H
7.W7_A3_103	-0.50	Yes	70.00%	1.28	5.37	0.82	AA	M	AA	H
8.W8_P1_103	2.33	Yes	55.50%	0.92	1.32	0.85	AA	F	AA	O
9.W9_P3_103	7.02	Yes	27.00%	0.99	9.90	0.68	AA	M	AA	H
10.W10_P3_103	6.84	Yes	27.50%	1.05	4.09	0.64	AA	F	AA	O

Table 6.2.4H

Raw Score to Scale Score Conversion: Writ 3-5

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	4.92	910.00^	910.00^
1	910^	3.43	910.00^	911.00
2	911	2.66	910.00^	914.05
3	914	2.26	911.60	916.11
4	916	2.06	913.71	917.84
5	918	2.02	915.49	919.52
6	919	1.99	917.17	921.15
7	921	1.99	918.82	922.81
8	922	1.97	920.46	924.39
9	924	1.92	922.06	925.90
10	925	1.87	923.60	927.34
11	927	1.90	925.06	928.86
12	928	1.97	926.53	930.46
13	930	2.06	928.11	932.24
14	932	2.14	929.89	934.16
15	934	2.09	931.78	935.96
16	936	2.04	933.58	937.66
17	937	2.06	935.29	939.42
18	939	2.21	937.04	941.46
19	942	2.50	939.01	944.00
20	944	2.57	941.77	946.90
21	947	2.11	944.48	948.70
22	948	1.99	946.28	950.26
23	949*	2.40	947.74	952.54
24	950*	4.20	948.63	957.03

^ Truncated

* Adjusted for end of scale effect

Table 6.2.4I

Raw Score to Proficiency Level Conversion: Writ 3-5

Raw Score	Grade 3			Grade 4			Grade 5		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	8.01	8.01	A1	6.79	6.79	A1	7.92	7.92
1	A1	0.90	8.92	A1	1.58	8.38	A1	0.89	8.82
2	A1	1.64	10.56	A1	1.07	9.44	A1	0.85	9.67
3	A1	3.62	14.18	A1	3.80	13.25	A1	2.98	12.65
4	A1	2.06	16.24	A1	1.11	14.36	A1	0.94	13.59
5	A1	3.12	19.36	A1	2.18	16.54	A1	3.19	16.78
6	A1	1.36	20.72	A1	0.90	17.44	A1	1.06	17.84
7	A1	1.48	22.19	A1	1.28	18.72	A1	1.06	18.91
8	A1	1.48	23.67	A1	0.73	19.44	A1	1.24	20.14
9	A2	1.23	24.91	A2	0.73	20.17	A2	0.77	20.91
10	A2	2.63	27.54	A2	1.41	21.58	A2	2.13	23.04
11	A2	2.92	30.46	A2	2.65	24.23	A2	2.04	25.09
12	A2	9.78	40.24	A2	10.21	34.44	A2	7.50	32.58
13	A2	5.10	45.33	A2	4.02	38.46	A2	3.71	36.29
14	A3	15.45	60.79	A3	13.89	52.35	A3	12.18	48.47
15	A3	2.51	63.30	A3	1.97	54.32	A3	1.62	50.09
16	A3	4.93	68.23	A3	4.53	58.85	A3	4.86	54.94
17	A3	3.99	72.22	A3	4.10	62.95	A3	3.83	58.77
18	P1	5.38	77.60	P1	5.68	68.63	P1	5.28	64.05
19	P1	5.80	83.39	P1	6.11	74.74	P1	7.24	71.29
20	P1	6.08	89.48	P1	8.59	83.33	P1	10.35	81.64
21	P2	3.62	93.09	P2	5.43	88.76	P2	6.05	87.69
22	P2	4.03	97.12	P2	6.32	95.09	P2	6.30	93.99
23	P2	0.95	98.07	P2	1.58	96.67	P2	1.58	95.57
24	P3	1.93	100.00	P3	3.33	100.00	P3	4.43	100.00

Table 6.2.4J

Accuracy and Consistency of Classification Indices: Writ 3-5

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.624	0.561		0.445	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.870		0.158	
	A2	0.639		0.245	
	A3	0.668		0.196	
	P1	0.474		0.454	
	P2	-		0.500	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.949	0.029	0.022	0.927
	A2/A3	0.921	0.039	0.040	0.892
	A3/P1	0.914	0.021	0.065	0.879
	P1/P2	0.835	0.165	0.000	0.835

6.2.5 Oral Language Composite 3-5

Figure 6.2.5A

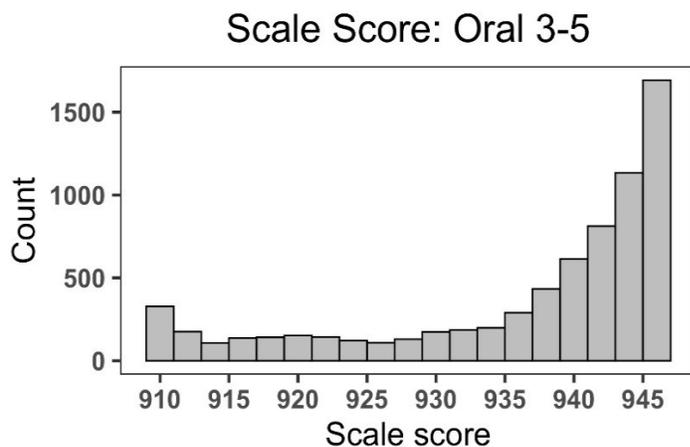


Table 6.2.5A

Scale Score Descriptive Statistics: Oral 3-5

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	2,408	910	947	936.04	11.34
4	2,332	910	947	937.43	11.07
5	2,344	910	947	937.65	11.14
Total	7,084	910	947	937.03	11.20

Figure 6.2.5B

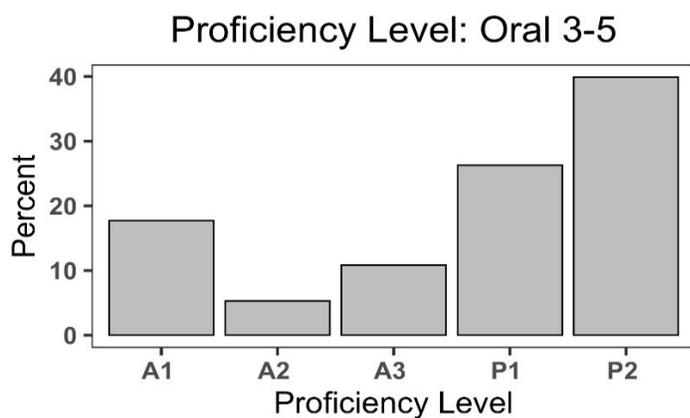


Table 6.2.5C

Proficiency Level Distribution: Oral 3-5

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	468	19.44	403	17.28	384	16.38	1,255	17.72
A2	128	5.32	113	4.85	133	5.67	374	5.28
A3	321	13.33	231	9.91	215	9.17	767	10.83
P1	668	27.74	611	26.20	582	24.83	1,861	26.27
P2	823	34.18	974	41.77	1,030	43.94	2,827	39.91
Total	2,408	100.00	2,332	100.00	2,344	100.00	7,084	100.00

Table 6.2.5D

n/a

Figure 6.2.5D
n/a

Figure 6.2.5E
n/a

Table 6.2.5E

Reliability: Oral 3-5

Component	Weight	Variance	Reliability
Listening	0.5	110.881	.916
Speaking	0.5	176.890	.900
Oral		125.440	.944

*Variances from students who had results in all four domains

Table 6.2.5F
n/a

Table 6.2.5G
n/a

Table 6.2.5H
n/a

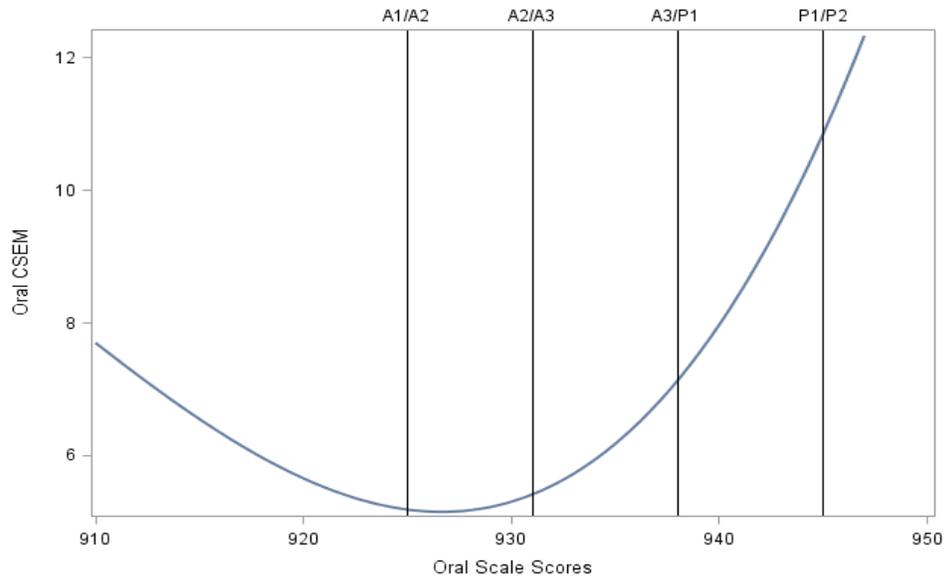
Table 6.2.5I
n/a

Table 6.2.5J

Accuracy and Consistency of Classification Indices: Oral 3-5

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.607	0.608		0.467	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.949		0.923	
	A2	0.632		0.509	
	A3	0.799		0.706	
	P1	0.487		0.500	
	P2	-		0.569	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.983	0.009	0.009	0.975
	A2/A3	0.976	0.013	0.011	0.966
	A3/P1	0.969	0.010	0.021	0.957
	P1/P2	0.680	0.320	0.000	0.708

Figure 6.2.5F CSEM for Oral Composite 3-5



6.2.6 Literacy Composite 3-5

Figure 6.2.6A

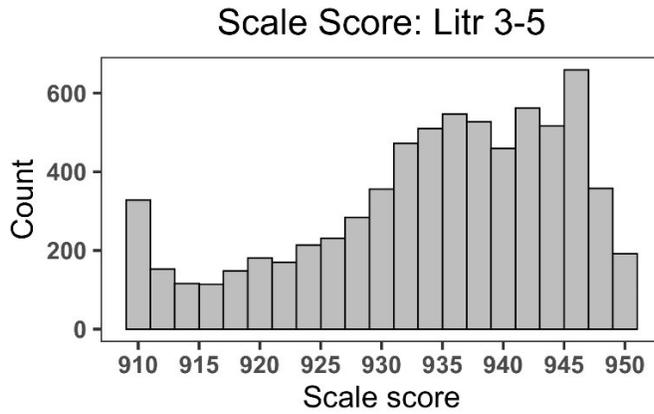


Table 6.2.6A

Scale Score Descriptive Statistics: Litr 3-5

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	2,412	910	950	933.07	10.49
4	2,339	910	950	935.11	10.74
5	2,347	910	950	935.63	10.97
Total	7,098	910	950	934.59	10.79

Figure 6.2.6B

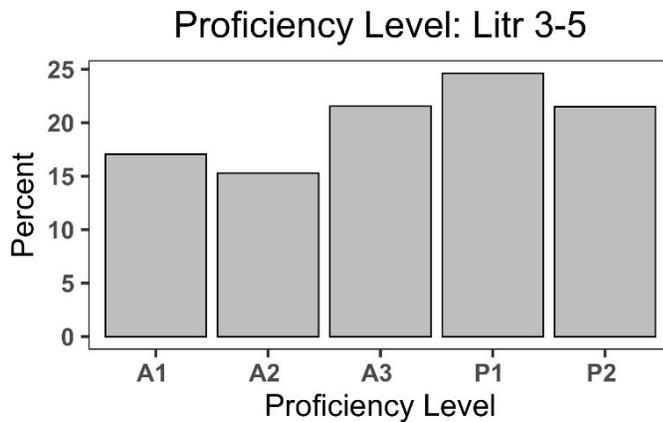


Table 6.2.6C

Proficiency Level Distribution: Litr 3-5

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	475	19.69	372	15.90	363	15.47	1,210	17.05
A2	408	16.92	338	14.45	339	14.44	1,085	15.29
A3	607	25.17	508	21.72	414	17.64	1,529	21.54
P1	571	23.67	567	24.24	610	25.99	1,748	24.63
P2	351	14.55	554	23.69	621	26.46	1,526	21.50
Total	2,412	100.00	2,339	100.00	2,347	100.00	7,098	100.00

Table 6.2.6D

n/a

Figure 6.2.6D
n/a

Figure 6.2.6E
n/a

Table 6.2.6E

Reliability: Litr 3-5

Component	Weight	Variance	Reliability
Reading	0.5	117.289	.931
Writing	0.5	142.325	.795
Literacy		116.424	.920

*Variances from students who had results in all four domains

Table 6.2.6F
n/a

Table 6.2.6G
n/a

Table 6.2.6H
n/a

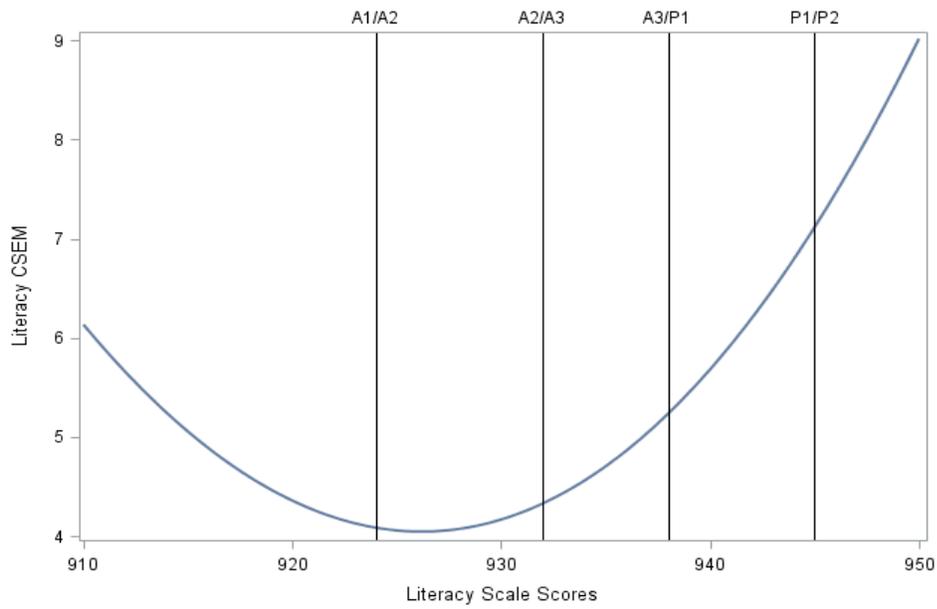
Table 6.2.6I
n/a

Table 6.2.6J

Accuracy and Consistency of Classification Indices: Litr 3-5

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.727	0.669		0.564	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.914		0.872	
	A2	0.753		0.653	
	A3	0.791		0.693	
	P1	0.625		0.602	
	P2	-		0.406	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.972	0.016	0.012	0.960
	A2/A3	0.951	0.026	0.023	0.931
	A3/P1	0.941	0.016	0.043	0.918
	P1/P2	0.863	0.137	0.000	0.857

Figure 6.2.6F CSEM for Literacy Composite 3-5



6.2.7 Comprehension Composite 3-5

Figure 6.2.7A

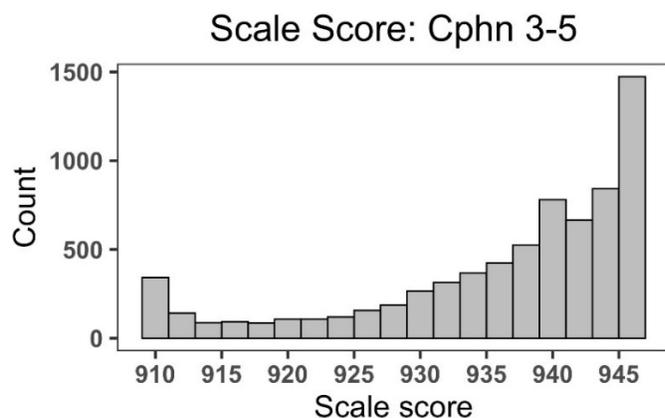


Table 6.2.7A

Scale Score Descriptive Statistics: Cphn 3-5

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	2,407	910	947	935.27	10.51
4	2,334	910	947	937.07	10.36
5	2,347	910	947	937.45	10.37
Total	7,088	910	947	936.58	10.46

Figure 6.2.7B

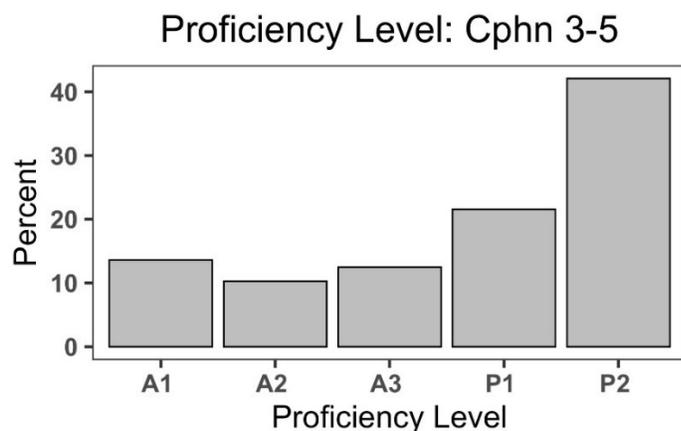


Table 6.2.7C

Proficiency Level Distribution: Cphn 3-5

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	374	15.54	305	13.07	286	12.19	965	13.61
A2	286	11.88	218	9.34	226	9.63	730	10.30
A3	343	14.25	276	11.83	264	11.25	883	12.46
P1	597	24.80	487	20.87	444	18.92	1,528	21.56
P2	807	33.53	1,048	44.90	1,127	48.02	2,982	42.07
Total	2,407	100.00	2,334	100.00	2,347	100.00	7,088	100.00

Table 6.2.7D

n/a

Figure 6.2.7D
n/a

Figure 6.2.7E
n/a

Table 6.2.7E

Reliability: Cphn 3-5

Component	Weight	Variance	Reliability
Listening	0.3	110.881	.916
Reading	0.7	117.289	.931
Comprehension		109.203	.959

*Variances from students who had results in all four domains

Table 6.2.7F
n/a

Table 6.2.7G
n/a

Table 6.2.7H
n/a

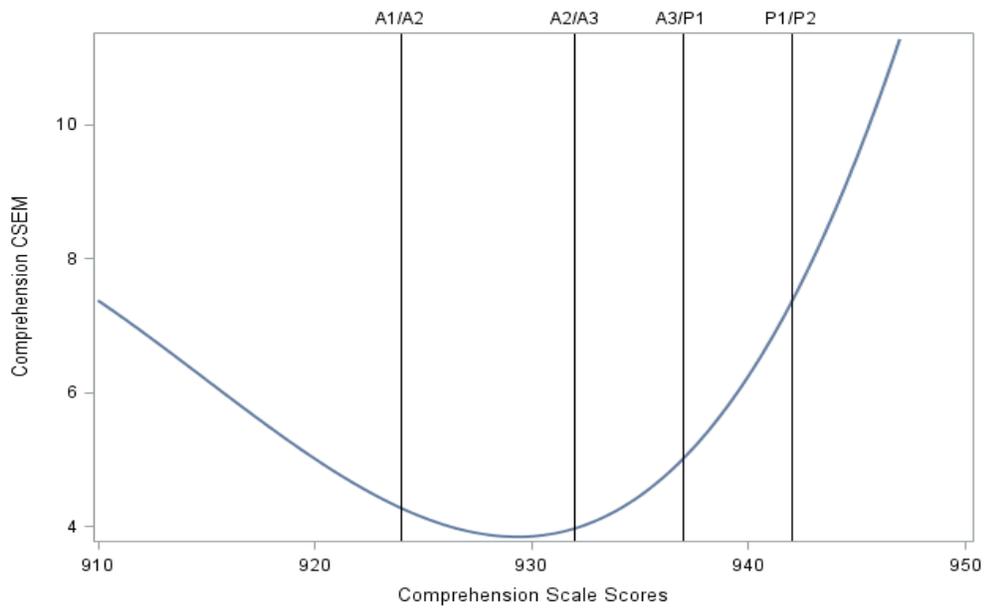
Table 6.2.7I
n/a

Table 6.2.7J

Accuracy and Consistency of Classification Indices: Cphn 3-5

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.562	0.560		0.423	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.918		0.880	
	A2	0.753		0.654	
	A3	0.748		0.630	
	P1	0.421		0.416	
	P2	-		0.595	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.980	0.012	0.008	0.972
	A2/A3	0.964	0.018	0.017	0.951
	A3/P1	0.957	0.009	0.034	0.940
	P1/P2	0.659	0.341	0.000	0.690

Figure 6.2.7F CSEM for Comprehension Composite 3-5



6.2.8 Overall Composite 3-5

Figure 6.2.8A

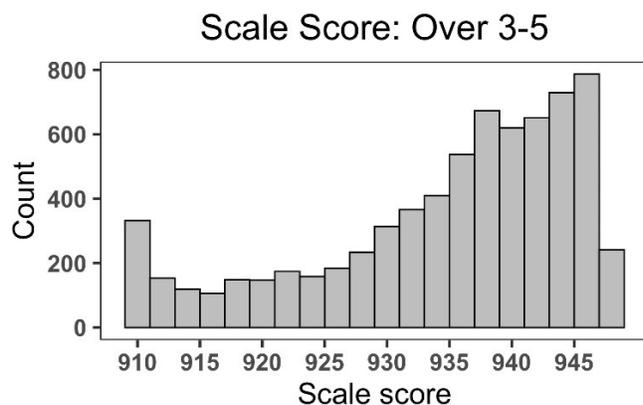


Table 6.2.8A
Scale Score Descriptive Statistics: Over 3-5

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	2,406	910	949	933.72	10.42
4	2,332	910	949	935.58	10.52
5	2,344	910	949	936.01	10.72
Total	7,082	910	949	935.09	10.60

Figure 6.2.8A

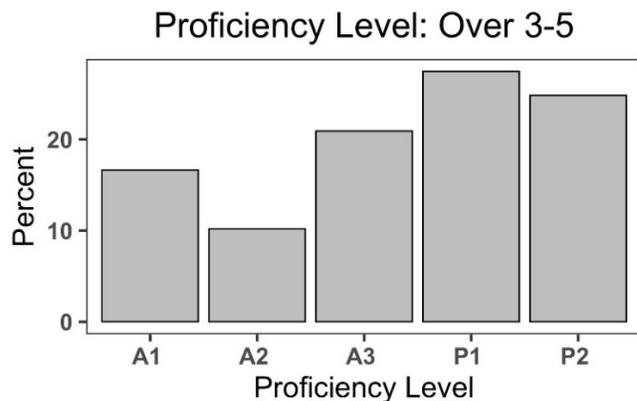


Table 6.2.8C

Proficiency Level Distribution: Over 3-5

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	448	18.62	370	15.87	361	15.40	1,179	16.65
A2	276	11.47	226	9.69	219	9.34	721	10.18
A3	592	24.61	474	20.33	415	17.70	1,481	20.91
P1	669	27.81	633	27.14	642	27.39	1,944	27.45
P2	421	17.50	629	26.97	707	30.16	1,757	24.81
Total	2,406	100.00	2,332	100.00	2,344	100.00	7,082	100.00

Table 6.2.8D

n/a

Figure 6.2.8D
n/a

Figure 6.2.8E
n/a

Table 6.2.8E

Reliability: Over 3-5

Component	Weight	Variance	Reliability
Listening	0.15	110.881	.916
Reading	0.35	117.289	.931
Speaking	0.15	176.890	.900
Writing	0.35	142.325	.795
Overall Composite		112.360	.954

*Variances from students who had results in all four domains

Table 6.2.8F
n/a

Table 6.2.8G
n/a

Table 6.2.8H
n/a

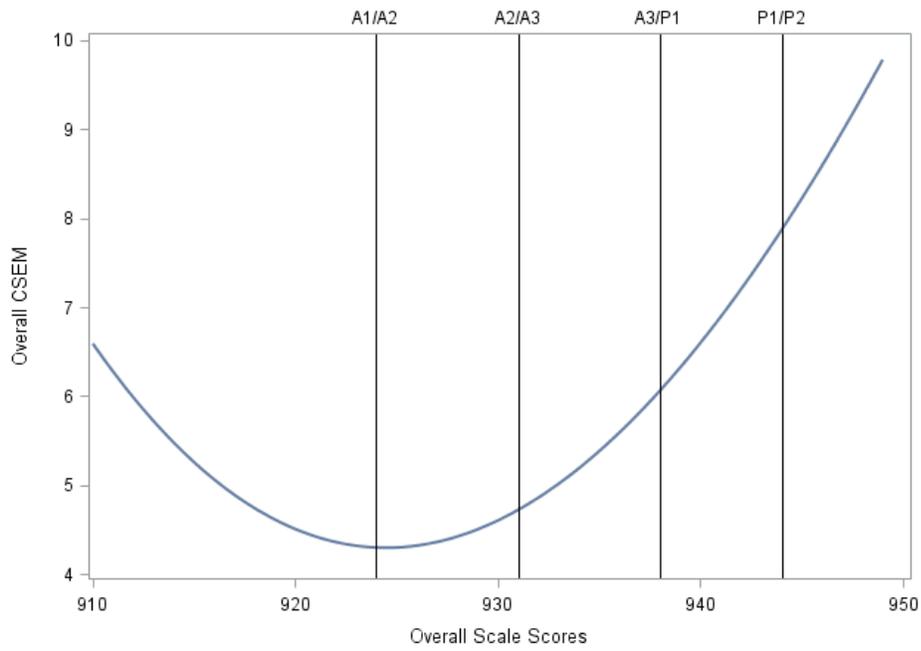
Table 6.2.8I
n/a

Table 6.2.8J

Accuracy and Consistency of Classification Indices: Over 3-5

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.708	0.671		0.567	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.941		0.913	
	A2	0.763		0.666	
	A3	0.876		0.814	
	P1	0.561		0.550	
	P2	-		0.499	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.982	0.010	0.008	0.974
	A2/A3	0.969	0.017	0.014	0.955
	A3/P1	0.958	0.010	0.031	0.943
	P1/P2	0.799	0.201	0.000	0.799

Figure 6.2.8F CSEM for Overall Composite 3-5



6.3 Grades: 6-8

6.3.1 Listening 6-8

Figure 6.3.1A

Raw Score: List 6-8

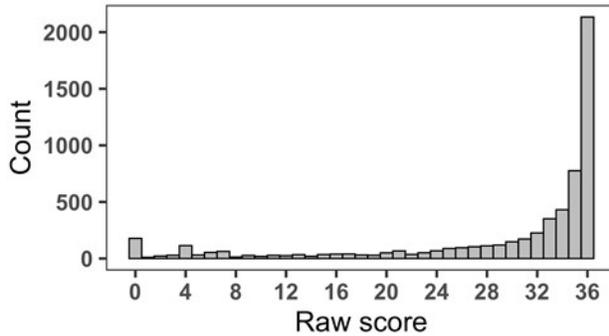


Table 6.3.1A

Raw Score Descriptive Statistics: List 6-8

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,075	0	36	28.91	10.38
7	1,938	0	36	29.63	10.02
8	1,858	0	36	30.26	9.48
Total	5,871	0	36	29.58	10.00

Figure 6.3.1B

Scale Score: List 6-8

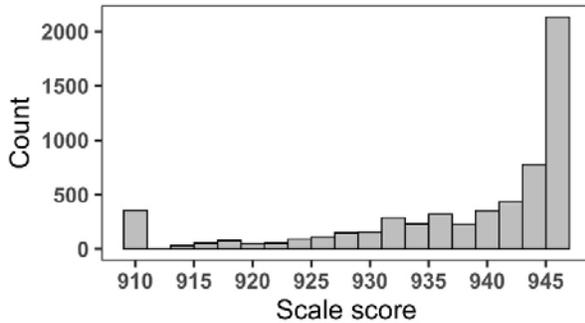


Table 6.3.1B

Scale Score Descriptive Statistics: List 6-8

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,075	910	946	937.51	10.54
7	1,938	910	946	938.27	10.31
8	1,858	910	946	938.98	9.75
Total	5,871	910	946	938.22	10.24

Figure 6.3.1C

Proficiency Level: List 6-8

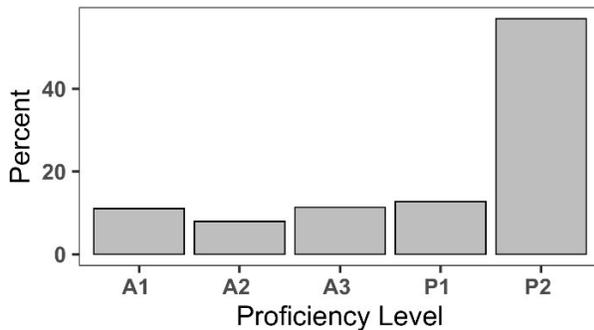


Table 6.3.1C

Proficiency Level Distribution: List 6-8

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	262	12.63	212	10.94	174	9.36	648	11.04
A2	177	8.53	149	7.69	140	7.53	466	7.94
A3	255	12.29	221	11.40	191	10.28	667	11.36
P1	270	13.01	243	12.54	236	12.70	749	12.76
P2	1,111	53.54	1,113	57.43	1,117	60.12	3,341	56.91
Total	2,075	100.00	1,938	100.00	1,858	100.00	5,871	100.00

Table 6.3.1D

Equating Summary: List 6-8

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 102. Thus, the results from the S102 of the Alternate ACCESS were used to determine raw-to-scale score conversion.

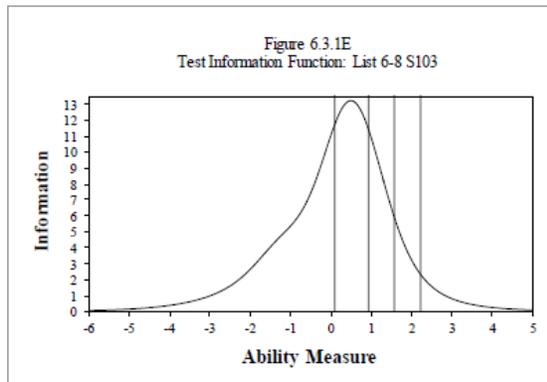
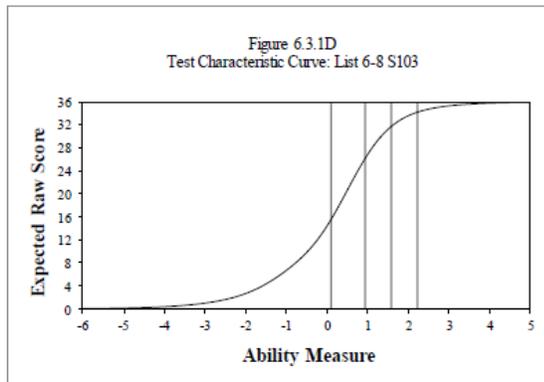


Table 6.3.1E

Reliability: List 6-8

No. of Students	No. of Items	Cronbach's Alpha	SEM
5,549	9	0.927	2.338

Table 6.3.1F

Item Analysis Summary: List 6-8

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	MOSR	9	0.00	81.89%	1.21	1.41

DIF Summary	DIF Level	Male/Female		Hispanic/Other	
		Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	4	5	5	4
	BB	0	0	0	0
CC	0	0	0	0	

Table 6.3.1G

Complete Item Analysis: List 6-8

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1. L1_A1_103	-1.53	Yes	91.00%	2.93	3.99	0.64	AA	M	AA	O
2. L2_A2_103	0.17	Yes	82.25%	1.16	0.97	0.85	AA	M	AA	H
3. L3_A2_103	0.05	Yes	83.00%	1.06	0.93	0.87	AA	F	AA	O
4. L4_A3_103	-0.04	Yes	82.50%	1.02	0.89	0.87	AA	M	AA	O
5. L5_A3_103	0.21	Yes	80.75%	0.90	0.84	0.88	AA	M	AA	O
6. L6_P1_103	0.95	Yes	74.50%	1.18	1.13	0.81	AA	F	AA	H
7. L7_P1_103	0.55	Yes	78.75%	0.78	0.70	0.89	AA	F	AA	H
8. L8_P2_103	0.37	Yes	78.00%	0.78	0.62	0.89	AA	F	AA	H
9. L9_P2_103	0.59	Yes	77.75%	0.87	0.70	0.87	AA	M	AA	O

Table 6.3.1H

Raw Score to Scale Score Conversion: List 6-8

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	14.72	910.00^	910.00^
1	910^	7.99	910.00^	910.00^
2	910^	5.54	910.00^	910.00^
3	910^	4.67	910.00^	911.13
4	910^	4.43	910.00^	913.42
5	911	4.35	910.00^	915.80
6	914	4.19	910.00^	917.93
7	916	3.96	911.84	919.75
8	918	3.64	913.98	921.26
9	919	3.32	915.80	922.44
10	920	3.09	917.38	923.55
11	922	2.85	918.73	924.42
12	923	2.69	919.83	925.21
13	923	2.53	920.78	925.85
14	924	2.45	921.65	926.56
15	925	2.37	922.44	927.19
16	926	2.29	923.24	927.83
17	926	2.22	923.95	928.38
18	927	2.22	924.58	929.01
19	927	2.22	925.21	929.65
20	928	2.14	925.85	930.12
21	929	2.14	926.48	930.75
22	929	2.22	926.96	931.39
23	930	2.22	927.59	932.02
24	930	2.22	928.22	932.65
25	931	2.29	928.78	933.36
26	932	2.37	929.41	934.16
27	932	2.45	930.04	934.95
28	933	2.53	930.67	935.74
29	934	2.69	931.39	936.77
30	935	2.85	932.18	937.88
31	936	3.09	933.05	939.22
32	937	3.40	934.08	940.88
33	939	3.96	935.18	943.10
34	941*	4.91	936.69	946.50
35	943*	7.36	938.75	953.46
36	945*	14.09	940.72	968.89

^ Truncated

* Adjusted for end of scale effect

WIDA ALTERNATE ACCE SS Annual Tech Rpt 7

Table 6.3.11

Raw Score to Proficiency Level Conversion: List 6-8

Raw Score	Grade 6			Grade 7			Grade 8		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	3.23	3.23	A1	3.20	3.20	A1	2.64	2.64
1	A1	0.24	3.47	A1	0.21	3.41	A1	0.11	2.74
2	A1	0.43	3.90	A1	0.36	3.77	A1	0.32	3.07
3	A1	0.63	4.53	A1	0.52	4.28	A1	0.32	3.39
4	A1	1.88	6.41	A1	2.27	6.55	A1	1.67	5.06
5	A1	0.77	7.18	A1	0.52	7.07	A1	0.27	5.33
6	A1	1.16	8.34	A1	0.52	7.59	A1	1.08	6.40
7	A1	1.16	9.49	A1	1.14	8.72	A1	0.91	7.32
8	A1	0.24	9.73	A1	0.15	8.88	A1	0.27	7.59
9	A1	0.63	10.36	A1	0.41	9.29	A1	0.32	7.91
10	A1	0.48	10.84	A1	0.31	9.60	A1	0.16	8.07
11	A1	0.58	11.42	A1	0.57	10.17	A1	0.27	8.34
12	A1	0.53	11.95	A1	0.31	10.47	A1	0.43	8.77
13	A1	0.67	12.63	A1	0.46	10.94	A1	0.59	9.36
14	A2	0.29	12.92	A2	0.31	11.25	A2	0.43	9.80
15	A2	0.82	13.73	A2	0.41	11.66	A2	0.54	10.33
16	A2	0.63	14.36	A2	0.67	12.33	A2	0.70	11.03
17	A2	0.82	15.18	A2	0.57	12.90	A2	0.65	11.68
18	A2	0.53	15.71	A2	0.46	13.36	A2	0.59	12.27
19	A2	0.58	16.29	A2	0.62	13.98	A2	0.27	12.54
20	A2	1.01	17.30	A2	0.77	14.76	A2	0.75	13.29
21	A2	1.16	18.46	A2	1.19	15.94	A2	1.08	14.37
22	A2	0.63	19.08	A2	0.72	16.67	A2	0.48	14.85
23	A2	0.92	20.00	A2	0.88	17.54	A2	0.81	15.66
24	A2	1.16	21.16	A2	1.08	18.63	A2	1.24	16.90
25	A3	1.73	22.89	A3	1.19	19.81	A3	1.61	18.51
26	A3	1.78	24.67	A3	1.75	21.57	A3	1.29	19.81
27	A3	1.83	26.51	A3	2.01	23.58	A3	1.45	21.26
28	A3	2.02	28.53	A3	2.12	25.70	A3	1.56	22.82
29	A3	2.27	30.80	A3	1.96	27.66	A3	1.83	24.65
30	A3	2.65	33.45	A3	2.37	30.03	A3	2.53	27.18
31	P1	3.33	36.77	P1	2.63	32.66	P1	2.80	29.98
32	P1	3.81	40.58	P1	4.23	36.89	P1	3.50	33.48
33	P1	5.88	46.46	P1	5.68	42.57	P1	6.40	39.88
34	P2	7.66	54.12	P2	7.43	50.00	P2	6.89	46.77
35	P2	12.43	66.55	P2	13.11	63.11	P2	14.21	60.98
36	P2	33.45	100.00	P2	36.89	100.00	P2	39.02	100.00

Table 6.3.1J

Accuracy and Consistency of Classification Indices: List 6-8

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.748	0.639		0.464	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.890		0.129	
	A2	0.606		0.229	
	A3	0.651		0.161	
	P1	0.448		0.202	
	P2	0.807		0.781	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.978	0.014	0.008	0.967
	A2/A3	0.959	0.023	0.018	0.945
	A3/P1	0.950	0.013	0.038	0.931
	P1/P2	0.855	0.029	0.116	0.770

6.3.2 Reading 6-8

Figure 6.3.2A

Raw Score: Read 6-8

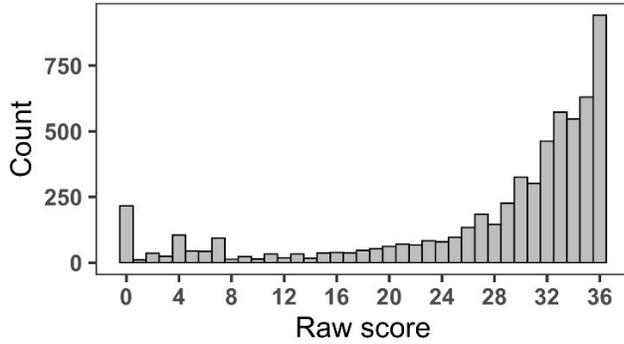


Table 6.3.2A

Raw Score Descriptive Statistics: Read 6-8

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,076	0	36	27.20	10.18
7	1,939	0	36	27.94	9.97
8	1,861	0	36	28.74	9.67
Total	5,876	0	36	27.93	9.97

Figure 6.3.2B

Scale Score: Read 6-8

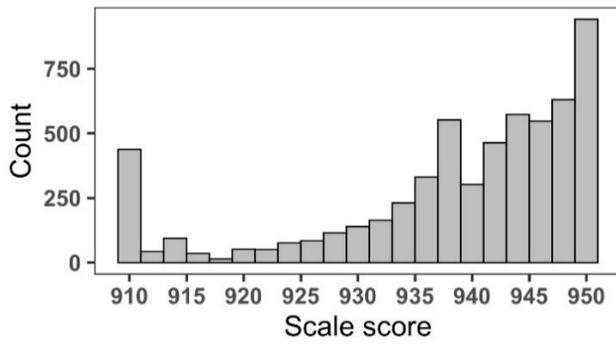


Table 6.3.2B

Scale Score Descriptive Statistics: Read 6-8

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,076	910	950	937.61	11.79
7	1,939	910	950	938.60	11.52
8	1,861	910	950	939.68	11.34
Total	5,876	910	950	938.59	11.59

Figure 6.3.2C

Proficiency Level: Read 6-8

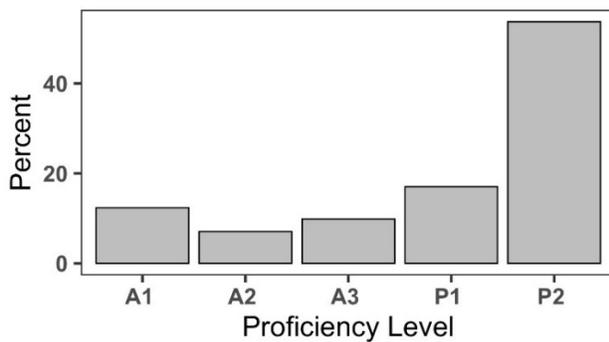


Table 6.3.2C

Proficiency Level Distribution: Read 6-8

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	288	13.87	234	12.07	205	11.02	727	12.37
A2	167	8.04	134	6.91	114	6.13	415	7.06
A3	228	10.98	195	10.06	157	8.44	580	9.87
P1	364	17.53	345	17.79	291	15.64	1,000	17.02
P2	1,029	49.57	1,031	53.17	1,094	58.79	3,154	53.68
Total	2,076	100.00	1,939	100.00	1,861	100.00	5,876	100.00

Table 6.3.2D

Equating Summary: Read 6-8

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 100. Thus, the results from the original field test of the Alternate ACCESS were used to determine raw-to-scale score conversions. Technical details of the analysis of this process can be found in the Alternate ACCESS for ELLs™ Series 100 Development and Operational Field Test: Technical Report (2013).

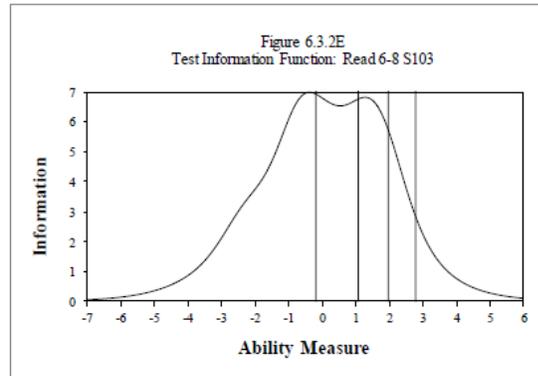
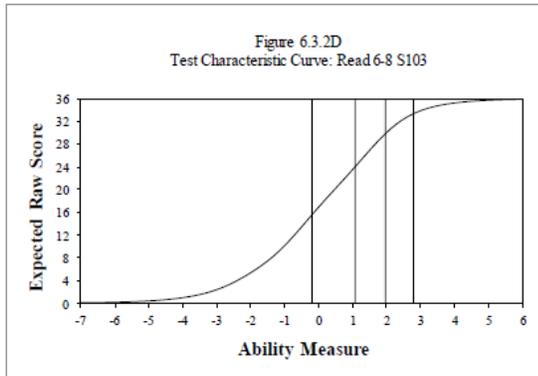


Table 6.3.2E

Reliability: Read 6-8

No. of Students	No. of Items	Cronbach's Alpha	SEM
5,540	9	0.915	2.904

Table 6.3.2F

Item Analysis Summary: Read 6-8

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	MOSR	9	0.00	76.22%	1.23	1.46
DIF Summary	DIF Level		Male/Female		Hispanic/Other	
			Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	4	5	4	5	
	BB	0	0	0	0	
	CC	0	0	0	0	

Table 6.3.2G

Complete Item Analysis: Read 6-8

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1.R1_A1_103	-2.50		91.00%	3.11	3.15	0.66	AA	F	AA	O
2.R2_A2_103	-0.76	Yes	84.50%	1.32	0.73	0.87	AA	M	AA	O
3.R3_A2_103	-0.93	Yes	84.50%	1.12	0.71	0.87	AA	F	AA	O
4.R4_A3_103	0.07	Yes	77.75%	1.18	0.92	0.87	AA	M	AA	O
5.R5_A3_103	-0.72	Yes	83.75%	0.90	0.52	0.89	AA	F	AA	O
6.R6_P1_103	1.17	Yes	69.50%	0.95	0.86	0.83	AA	F	AA	H
7.R7_P1_103	1.09	Yes	70.00%	0.96	0.91	0.83	AA	F	AA	H
8.R8_P2_103	1.66	Yes	63.00%	0.89	0.98	0.79	AA	F	AA	H
9.R9_P2_103	1.81	Yes	61.25%	0.92	0.95	0.76	AA	M	AA	H

Table 6.3.2H

Raw Score to Scale Score Conversion: Read 6-8

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	11.87	910.00^	910.00^
1	910^	6.99	910.00^	910.00^
2	910^	4.82	910.00^	910.00^
3	910^	4.10	910.00^	910.00^
4	910^	3.92	910.00^	910.60
5	910^	3.86	910.00^	913.07
6	912	3.80	910.00^	915.42
7	914	3.56	910.36	917.47
8	916	3.25	912.59	919.10
9	917	3.01	914.46	920.49
10	919	2.77	916.09	921.63
11	920	2.65	917.41	922.71
12	921	2.53	918.62	923.68
13	922	2.47	919.70	924.64
14	923	2.41	920.73	925.55
15	924	2.41	921.69	926.51
16	925	2.41	922.65	927.48
17	926	2.41	923.62	928.44
18	927	2.41	924.58	929.40
19	928	2.47	925.49	930.43
20	929	2.47	926.51	931.45
21	930	2.47	927.54	932.48
22	931	2.47	928.56	933.50
23	932	2.47	929.58	934.53
24	933	2.47	930.61	935.55
25	934	2.47	931.63	936.57
26	935	2.47	932.60	937.54
27	936	2.47	933.62	938.56
28	937	2.47	934.65	939.59
29	938	2.53	935.61	940.67
30	939	2.59	936.63	941.82
31	940	2.77	937.66	943.20
32	942	3.01	938.80	944.83
33	944	3.43	940.07	946.94
34	946*	4.16	941.70	950.01
35	948*	5.97	943.93	955.86
36	950*	11.03	946.10	968.15

^ Truncated

* Adjusted for end of scale effect

Table 6.3.2I

Raw Score to Proficiency Level Conversion: Read 6-8

Raw Score	Grade 6			Grade 7			Grade 8		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	3.52	3.52	A1	4.13	4.13	A1	3.39	3.39
1	A1	0.24	3.76	A1	0.15	4.28	A1	0.16	3.55
2	A1	0.87	4.62	A1	0.62	4.90	A1	0.32	3.87
3	A1	0.58	5.20	A1	0.46	5.36	A1	0.21	4.08
4	A1	1.88	7.08	A1	1.65	7.01	A1	1.83	5.91
5	A1	1.20	8.29	A1	0.62	7.63	A1	0.43	6.34
6	A1	0.87	9.15	A1	0.67	8.30	A1	0.64	6.99
7	A1	1.83	10.98	A1	1.44	9.75	A1	1.50	8.49
8	A1	0.29	11.27	A1	0.05	9.80	A1	0.32	8.81
9	A1	0.48	11.75	A1	0.31	10.11	A1	0.38	9.19
10	A1	0.05	11.80	A1	0.31	10.42	A1	0.38	9.56
11	A1	0.53	12.33	A1	0.67	11.09	A1	0.48	10.05
12	A1	0.39	12.72	A1	0.10	11.19	A1	0.48	10.53
13	A1	0.72	13.44	A1	0.52	11.71	A1	0.43	10.96
14	A1	0.43	13.87	A1	0.36	12.07	A1	0.05	11.02
15	A2	0.67	14.55	A2	0.57	12.64	A2	0.64	11.66
16	A2	0.63	15.17	A2	0.88	13.51	A2	0.48	12.14
17	A2	0.53	15.70	A2	0.83	14.34	A2	0.59	12.74
18	A2	0.63	16.33	A2	0.93	15.27	A2	0.86	13.59
19	A2	1.06	17.39	A2	0.77	16.04	A2	0.86	14.45
20	A2	1.45	18.83	A2	0.77	16.81	A2	0.91	15.37
21	A2	1.64	20.47	A2	1.08	17.90	A2	0.86	16.23
22	A2	1.45	21.92	A2	1.08	18.98	A2	0.91	17.14
23	A3	1.54	23.46	A3	1.19	20.17	A3	1.56	18.70
24	A3	1.69	25.14	A3	1.29	21.45	A3	1.07	19.77
25	A3	1.73	26.88	A3	1.96	23.41	A3	1.24	21.01
26	A3	2.36	29.24	A3	2.22	25.63	A3	2.26	23.27
27	A3	3.66	32.90	A3	3.40	29.04	A3	2.31	25.58
28	P1	2.41	35.31	P1	2.89	31.92	P1	2.15	27.73
29	P1	3.90	39.21	P1	4.02	35.95	P1	3.60	31.33
30	P1	5.64	44.85	P1	5.47	41.41	P1	5.53	36.86
31	P1	5.59	50.43	P1	5.42	46.83	P1	4.35	41.21
32	P2	8.57	59.01	P2	7.74	54.56	P2	7.25	48.47
33	P2	9.68	68.69	P2	9.75	64.31	P2	9.83	58.30
34	P2	8.77	77.46	P2	9.54	73.85	P2	9.67	67.97
35	P2	9.87	87.33	P2	10.42	84.27	P2	11.98	79.96
36	P2	12.67	100.00	P2	15.73	100.00	P2	20.04	100.00

Table 6.3.2J

Accuracy and Consistency of Classification Indices: Read 6-8

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.786	0.720		0.561	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.927		0.152	
	A2	0.564		0.184	
	A3	0.538		0.197	
	P1	0.561		0.157	
	P2	0.847		0.825	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.980	0.011	0.009	0.971
	A2/A3	0.963	0.022	0.015	0.948
	A3/P1	0.944	0.024	0.032	0.925
	P1/P2	0.888	0.019	0.093	0.845

6.3.3 Speaking 6-8

Figure 6.3.3A

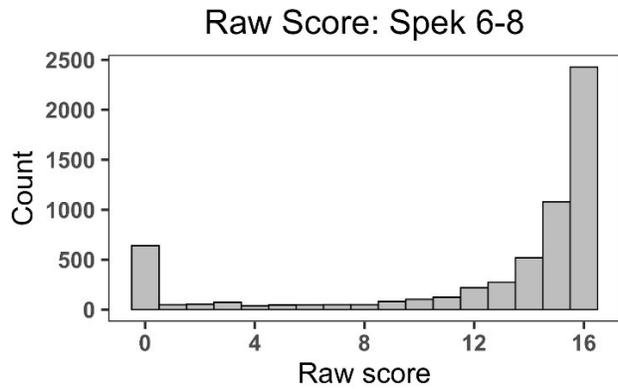


Table 6.3.3A
Raw Score Descriptive Statistics: Spek 6-8

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,077	0	16	12.33	5.40
7	1,937	0	16	12.39	5.48
8	1,862	0	16	12.79	5.17
Total	5,876	0	16	12.49	5.36

Figure 6.3.3B

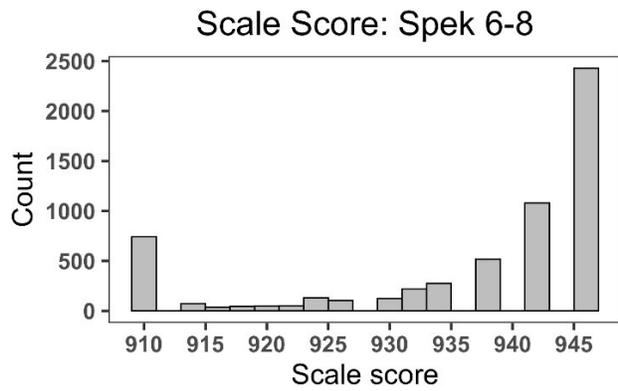


Table 6.3.3B
Scale Score Descriptive Statistics: Spek 6-8

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,077	910	947	936.79	12.85
7	1,937	910	947	937.10	13.03
8	1,862	910	947	938.05	12.47
Total	5,876	910	947	937.29	12.80

Figure 6.3.3C

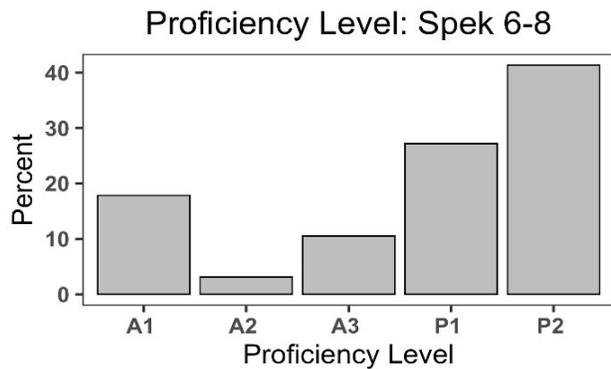


Table 6.3.3C

Proficiency Level Distribution: Spek 6-8

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	386	18.58	360	18.59	303	16.27	1,049	17.85
A2	65	3.13	58	2.99	60	3.22	183	3.11
A3	254	12.23	185	9.55	179	9.61	618	10.52
P1	579	27.88	544	28.08	474	25.46	1,597	27.18
P2	793	38.18	790	40.78	846	45.44	2,429	41.34
Total	2,077	100.00	1,937	100.00	1,862	100.00	5,876	100.00

Table 6.3.3D

Equating Summary: Spek 6-8

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 100. Thus, the results from the original field test of the Alternate ACCESS were used to determine raw-to-scale score conversions. Technical details of the analysis of this process can be found in the Alternate ACCESS for ELLs™ Series 100 Development and Operational Field Test: Technical Report (2013).

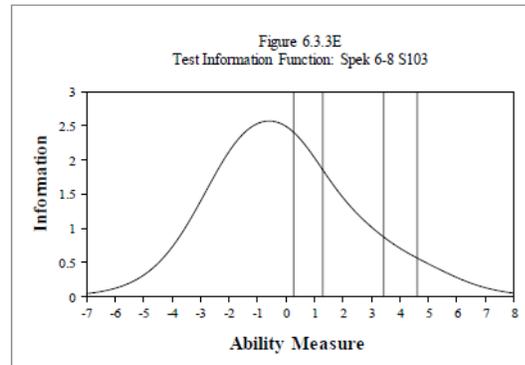
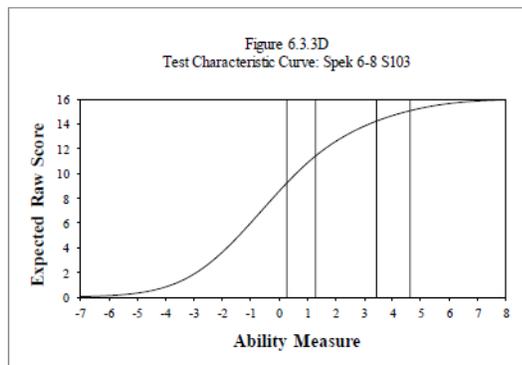


Table 6.3.3E

Reliability: Spek 6-8

No. of Students	No. of Items	Cronbach's Alpha	SEM
5,135	8	0.888	2.870

Table 6.3.3F

Item Analysis Summary: Spek 6-8

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	CR	8	0.00	77.19%	1.02	0.96

DIF Summary	DIF Level	Male/Female		Hispanic/Other	
		Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	5	3	3	5
	BB	0	0	0	0
CC	0	0	0	0	

Table 6.3.3G

Complete Item Analysis: Spek 6-8

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1.S1_A1_103	-1.91	Yes	84.00%	1.44	1.15	0.89	AA	F	AA	H
2.S2_A2_103	-1.23		82.50%	1.32	1.24	0.90	AA	M	AA	O
3.S3_A3_103	-0.52	Yes	80.50%	1.03	0.98	0.92	AA	F	AA	O
4.S4_A1_103	-0.62		81.00%	1.05	0.91	0.92	AA	F	AA	H
5.S5_A2_103	-0.71	Yes	81.50%	0.85	0.77	0.93	AA	F	AA	O
6.S6_A3_103	-0.47	Yes	79.50%	0.71	0.52	0.94	AA	F	AA	O
7.S7_P1_103	1.47	Yes	73.50%	0.91	0.76	0.87	AA	M	AA	O
8.S8_P2_103	3.95	Yes	58.50%	0.75	1.03	0.70	AA	M	AA	H

Table 6.3.3H

Raw Score to Scale Score Conversion: Spek 6-8

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	8.33	910.00^	910.00^
1	910^	4.83	910.00^	911.90
2	911	3.72	910.00^	914.73
3	914	3.28	910.43	916.99
4	916	3.06	912.92	919.03
5	918	2.93	915.04	920.90
6	920	2.88	916.99	922.76
7	922	2.88	918.86	924.62
8	924	2.88	920.72	926.48
9	925	2.93	922.54	928.39
10	927	3.01	924.40	930.43
11	930	3.19	926.39	932.78
12	932	3.50	928.57	935.57
13	935	3.95	931.18	939.07
14	939	4.61	934.59	943.81
15	943*	5.94	939.43	951.31
16	947*	8.95	944.21	962.12

^ Truncated

* Adjusted for end of scale effect

Table 6.3.3I

Raw Score to Proficiency Level Conversion: Spek 6-8

Raw Score	Grade 6			Grade 7			Grade 8		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	11.31	11.31	A1	11.87	11.87	A1	9.40	9.40
1	A1	0.91	12.23	A1	0.62	12.49	A1	1.02	10.42
2	A1	0.87	13.10	A1	1.08	13.58	A1	0.75	11.17
3	A1	1.25	14.35	A1	1.19	14.77	A1	1.29	12.46
4	A1	0.82	15.17	A1	0.57	15.33	A1	0.54	13.00
5	A1	0.58	15.74	A1	0.98	16.31	A1	0.81	13.80
6	A1	0.82	16.56	A1	0.98	17.29	A1	0.70	14.50
7	A1	1.01	17.57	A1	0.72	18.02	A1	0.81	15.31
8	A1	1.01	18.58	A1	0.57	18.59	A1	0.97	16.27
9	A2	1.20	19.79	A2	1.24	19.82	A2	1.66	17.94
10	A2	1.93	21.71	A2	1.76	21.58	A2	1.56	19.50
11	A3	2.36	24.07	A3	1.81	23.39	A3	2.15	21.64
12	A3	4.29	28.36	A3	3.67	27.05	A3	3.22	24.87
13	A3	5.58	33.94	A3	4.08	31.13	A3	4.24	29.11
14	P1	8.96	42.90	P1	9.03	40.17	P1	8.43	37.54
15	P1	18.92	61.82	P1	19.05	59.22	P1	17.02	54.56
16	P2	38.18	100.00	P2	40.78	100.00	P2	45.44	100.00

Table 6.3.3J

Accuracy and Consistency of Classification Indices: Spek 6-8

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.560	0.579		0.428
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.945		0.239	
	A2	0.513		0.149	
	A3	0.708		0.080	
	P1	0.419		0.406	
	P2	-		0.568	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.980	0.012	0.009	0.971
	A2/A3	0.975	0.012	0.014	0.965
	A3/P1	0.957	0.011	0.032	0.935
	P1/P2	0.646	0.354	0.000	0.684

6.3.4 Writing 6-8

Figure 6.3.4A

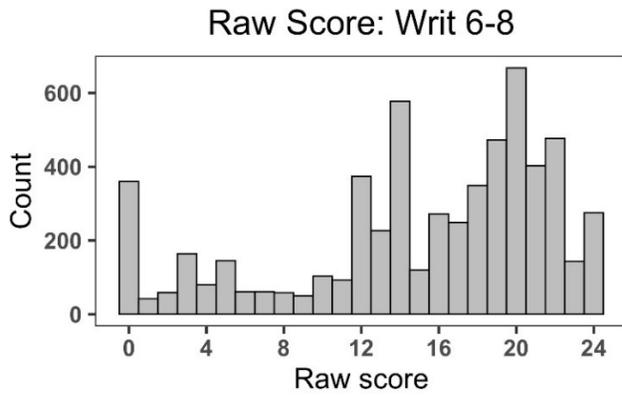


Table 6.3.4A

Raw Score Descriptive Statistics: Writ 6-8

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,079	0	24	14.59	6.74
7	1,940	0	24	15.01	6.83
8	1,865	0	24	15.76	6.59
Total	5,884	0	24	15.10	6.74

Figure 6.3.4B

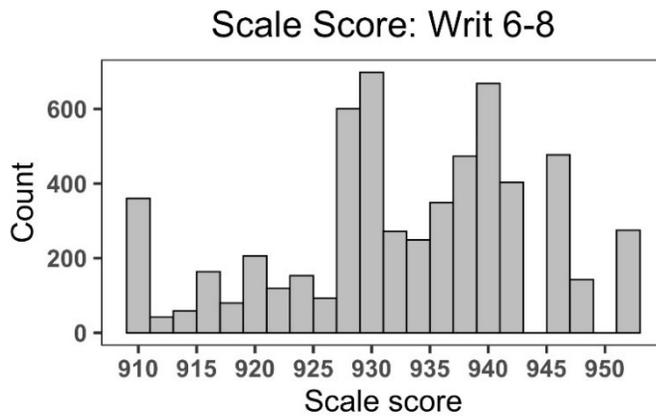


Table 6.3.4B

Scale Score Descriptive Statistics: Writ 6-8

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,079	910	953	932.48	10.68
7	1,940	910	953	933.27	10.95
8	1,865	910	953	934.60	10.83
Total	5,884	910	953	933.41	10.85

Figure 6.3.4C

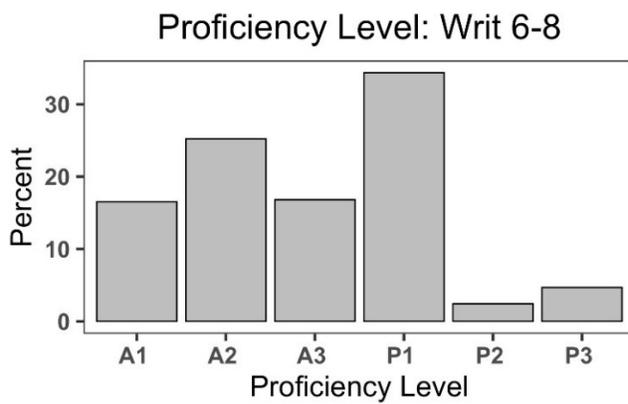


Table 6.3.4C

Proficiency Level Distribution: Writ 6-8

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	371	17.85	332	17.11	269	14.42	972	16.52
A2	570	27.42	480	24.74	433	23.22	1,483	25.20
A3	368	17.70	320	16.49	302	16.19	990	16.83
P1	651	31.31	675	34.79	695	37.27	2,021	34.35
P2	46	2.21	45	2.32	52	2.79	143	2.43
P3	73	3.51	88	4.54	114	6.11	275	4.67
Total	2,079	100.00	1,940	100.00	1,865	100.00	5,884	100.00

Table 6.3.4D

Equating Summary: Writ 6-8

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 100. Thus, the results from the original field test of the Alternate ACCESS were used to determine raw-to-scale score conversions. Technical details of the analysis of this process can be found in the Alternate ACCESS for ELLs™ Series 100 Development and Operational Field Test: Technical Report (2013).

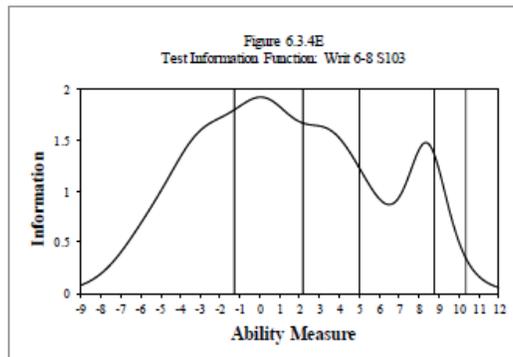
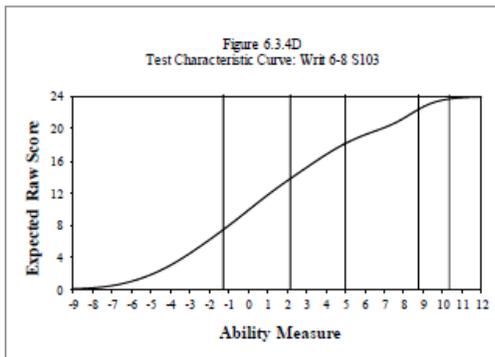


Table 6.3.4E

Reliability: Writ 6-8

No. of Students	No. of Items	Cronbach's Alpha	SEM
4,940	10	0.779	3.688

Table 6.3.4F

Item Analysis Summary: Writ 6-8

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	CR	10	0.00	67.23%	1.51	5.89

DIF Summary	DIF Level	Male/Female		Hispanic/Other	
		Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	6	4	6	4
	BB	0	0	0	0
CC	0	0	0	0	

Table 6.3.4G

Complete Item Analysis: Writ 6-8

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1.W1_A1_103	-4.12		85.00%	2.57	9.90	0.74	AA	F	AA	H
2.W2_A2_103	-2.91		83.00%	1.67	9.90	0.80	AA	F	AA	H
3.W3_A3_103	-1.32	Yes	78.50%	1.81	9.90	0.81	AA	F	AA	H
4.W4_P1_103	1.61	Yes	65.50%	1.34	2.05	0.84	AA	M	AA	O
5.W5_A1_103	-1.50		79.50%	1.44	7.30	0.82	AA	M	AA	H
6.W6_A2_103	-1.35		79.00%	1.33	3.08	0.83	AA	M	AA	H
7.W7_A3_103	-0.13		74.50%	1.43	6.94	0.84	AA	M	AA	O
8.W8_P1_103	1.83	Yes	64.50%	1.01	1.16	0.87	AA	M	AA	O
9.W9_P3_103	6.46	Yes	34.00%	1.04	3.94	0.71	AA	F	AA	O
10.W10_P3_103	6.16	Yes	34.75%	1.03	5.07	0.65	AA	F	AA	H

Table 6.3.4H

Raw Score to Scale Score Conversion: Writ 6-8

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	4.68	910.00^	913.09
1	912	2.90	910.00^	914.98
2	915	2.28	912.54	917.10
3	917	1.99	914.70	918.68
4	918	1.82	916.38	920.02
5	920	1.73	917.79	921.25
6	921	1.68	919.06	922.42
7	922	1.66	920.26	923.58
8	923	1.63	921.42	924.68
9	924	1.63	922.52	925.78
10	925	1.63	923.65	926.91
11	926	1.63	924.75	928.02
12	928	1.68	925.86	929.22
13	929	1.73	927.01	930.46
14	930	1.75	928.23	931.74
15	931	1.78	929.50	933.06
16	933	1.82	930.82	934.47
17	934	1.92	932.17	936.01
18	936	2.09	933.66	937.83
19	938	2.42	935.41	940.26
20	941	2.76	938.00	943.52
21	943	2.33	941.17	945.82
22	946	2.16	943.35	947.67
23	949*	2.57	945.18	950.31
24	952*	4.42	946.38	955.21

^ Truncated

* Adjusted for end of scale effect

Table 6.3.4I

Raw Score to Proficiency Level Conversion: Writ 6-8

Raw Score	Grade 6			Grade 7			Grade 8		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	6.59	6.59	A1	6.70	6.70	A1	4.99	4.99
1	A1	0.96	7.55	A1	0.52	7.22	A1	0.64	5.63
2	A1	1.06	8.61	A1	1.29	8.51	A1	0.64	6.27
3	A1	2.79	11.40	A1	3.04	11.55	A1	2.52	8.79
4	A1	1.49	12.89	A1	1.44	12.99	A1	1.13	9.92
5	A1	2.98	15.87	A1	1.96	14.95	A1	2.41	12.33
6	A1	0.96	16.84	A1	1.08	16.03	A1	1.07	13.40
7	A1	1.01	17.85	A1	1.08	17.11	A1	1.02	14.42
8	A2	1.15	19.00	A2	0.82	17.94	A2	0.97	15.39
9	A2	0.87	19.87	A2	0.88	18.81	A2	0.80	16.19
10	A2	1.97	21.84	A2	1.55	20.36	A2	1.72	17.91
11	A2	1.78	23.62	A2	1.49	21.86	A2	1.45	19.36
12	A2	6.01	29.63	A2	6.75	28.61	A2	6.33	25.68
13	A2	4.38	34.01	A2	3.76	32.37	A2	3.38	29.06
14	A2	11.26	45.26	A2	9.48	41.86	A2	8.58	37.64
15	A3	2.26	47.52	A3	1.86	43.71	A3	1.98	39.62
16	A3	4.62	52.14	A3	4.59	48.30	A3	4.66	44.29
17	A3	4.43	56.57	A3	4.28	52.58	A3	3.97	48.26
18	A3	6.40	62.96	A3	5.77	58.35	A3	5.58	53.83
19	P1	7.94	70.90	P1	8.51	66.86	P1	7.67	61.50
20	P1	10.29	81.19	P1	11.49	78.35	P1	12.39	73.89
21	P1	6.88	88.07	P1	6.75	85.10	P1	6.92	80.80
22	P1	6.20	94.28	P1	8.04	93.14	P1	10.29	91.10
23	P2	2.21	96.49	P2	2.32	95.46	P2	2.79	93.89
24	P3	3.51	100.00	P3	4.54	100.00	P3	6.11	100.00

Table 6.3.4J

Accuracy and Consistency of Classification Indices: Writ 6-8

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.748	0.652		0.531	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.858		0.129	
	A2	0.776		0.300	
	A3	0.524		0.121	
	P1	0.765		0.775	
	P2	-		0.190	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.953	0.030	0.018	0.932
	A2/A3	0.932	0.022	0.046	0.907
	A3/P1	0.925	0.031	0.044	0.888
	P1/P2	0.936	0.064	0.000	0.908

6.3.5 Oral Language Composite 6-8

Figure 6.3.5A

Scale Score: Oral 6-8

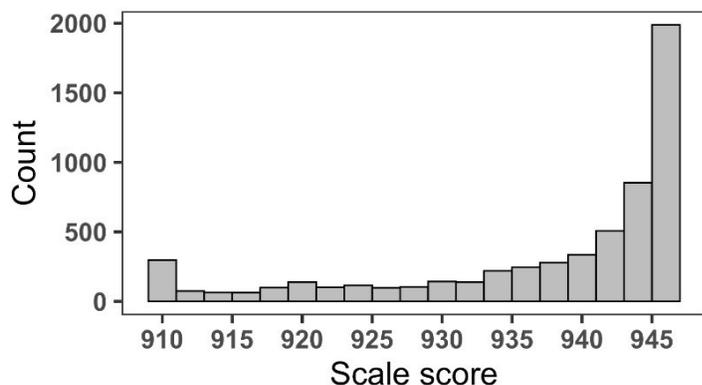


Table 6.3.5A

Scale Score Descriptive Statistics: Oral 6-8

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,073	910	947	937.52	11.22
7	1,935	910	947	938.06	11.14
8	1,856	910	947	938.89	10.63
Total	5,864	910	947	938.13	11.03

Figure 6.3.5B

Proficiency Level: Oral 6-8

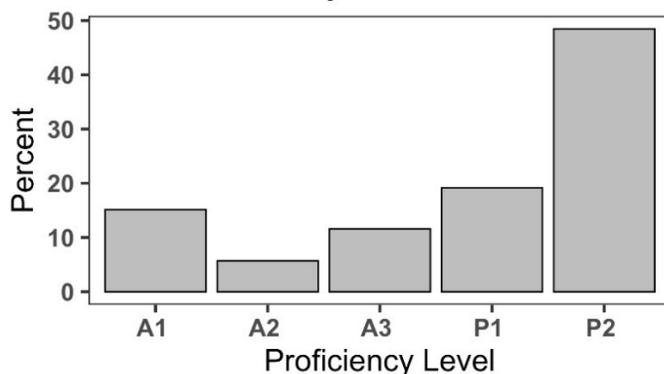


Table 6.3.5C

Proficiency Level Distribution: Oral 6-8

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	330	15.92	302	15.61	256	13.79	888	15.14
A2	123	5.93	114	5.89	96	5.17	333	5.68
A3	280	13.51	204	10.54	195	10.51	679	11.58
P1	394	19.01	380	19.64	348	18.75	1,122	19.13
P2	946	45.63	935	48.32	961	51.78	2,842	48.47
Total	2,073	100.00	1,935	100.00	1,856	100.00	5,864	100.00

Table 6.3.5D

n/a

Figure 6.3.5D

n/a

Figure 6.3.5E

n/a

Table 6.3.5E

Reliability: Oral 6-8

Component	Weight	Variance	Reliability
Listening	0.5	104.448	.927
Speaking	0.5	163.328	.899
Oral		121.220	.947

*Variances from students who had results in all four domains

Table 6.3.5F

n/a

Table 6.3.5G

n/a

Table 6.3.5H

n/a

Table 6.3.5I

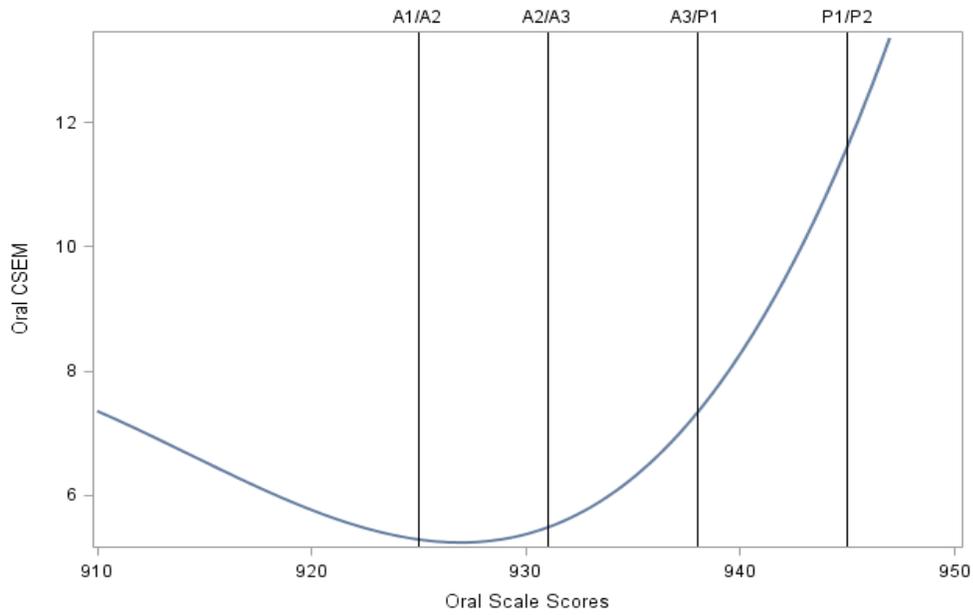
n/a

Table 6.3.5J

Accuracy and Consistency of Classification Indices: Oral 6-8

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.766	0.650		0.516	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.940		0.911	
	A2	0.648		0.526	
	A3	0.798		0.707	
	P1	0.603		0.393	
	P2	0.775		0.748	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.983	0.009	0.008	0.976
	A2/A3	0.975	0.014	0.012	0.964
	A3/P1	0.970	0.009	0.021	0.958
	P1/P2	0.838	0.048	0.114	0.749

Figure 6.3.5F CSEM for Oral Composite 6-8



6.3.6 Literacy Composite 6-8

Figure 6.3.6A

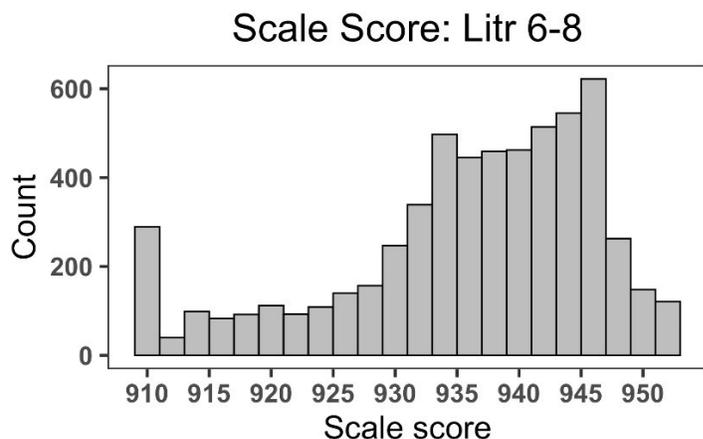


Table 6.3.6A

Scale Score Descriptive Statistics: Litr 6-8

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,076	910	952	935.29	10.59
7	1,939	910	952	936.17	10.57
8	1,861	910	952	937.37	10.40
Total	5,876	910	952	936.24	10.56

Figure 6.3.6B

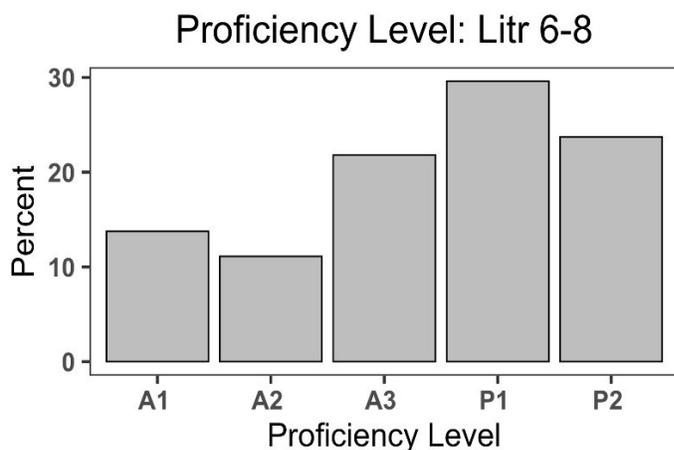


Table 6.3.6C

Proficiency Level Distribution: Litr 6-8

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	317	15.27	271	13.98	220	11.82	808	13.75
A2	249	11.99	210	10.83	194	10.42	653	11.11
A3	479	23.07	424	21.87	378	20.31	1,281	21.80
P1	635	30.59	587	30.27	518	27.83	1,740	29.61
P2	396	19.08	447	23.05	551	29.61	1,394	23.72
Total	2,076	100.00	1,939	100.00	1,861	100.00	5,876	100.00

Table 6.3.6D

n/a

Figure 6.3.6D

n/a

Figure 6.3.6E

n/a

Table 6.3.6E

Reliability: Litr 6-8

Component	Weight	Variance	Reliability
Reading	0.5	133.865	.915
Writing	0.5	117.506	.779
Literacy		111.092	.916

*Variances from students who had results in all four domains

Table 6.3.6F

n/a

Table 6.3.6G

n/a

Table 6.3.6H

n/a

Table 6.3.6I

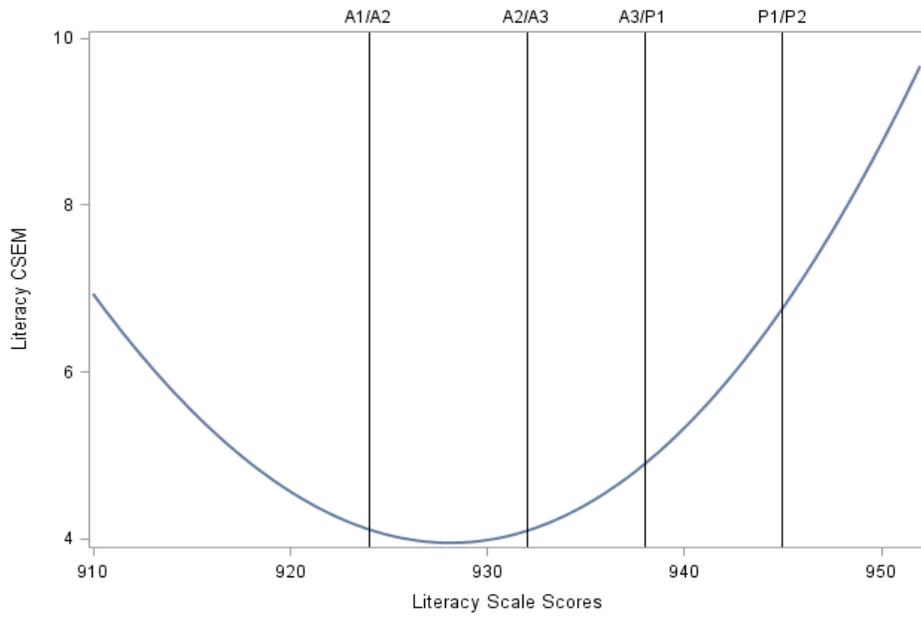
n/a

Table 6.3.6J

Accuracy and Consistency of Classification Indices: Litr 6-8

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.638	0.616		0.503	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.922		0.883	
	A2	0.740		0.636	
	A3	0.794		0.696	
	P1	0.499		0.503	
	P2	-		0.557	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.979	0.012	0.009	0.970
	A2/A3	0.959	0.023	0.018	0.943
	A3/P1	0.948	0.013	0.039	0.929
	P1/P2	0.751	0.249	0.000	0.771

Figure 6.3.6F CSEM for Literacy Composite 6-8



6.3.7 Comprehension Composite 6-8

Figure 6.3.7A

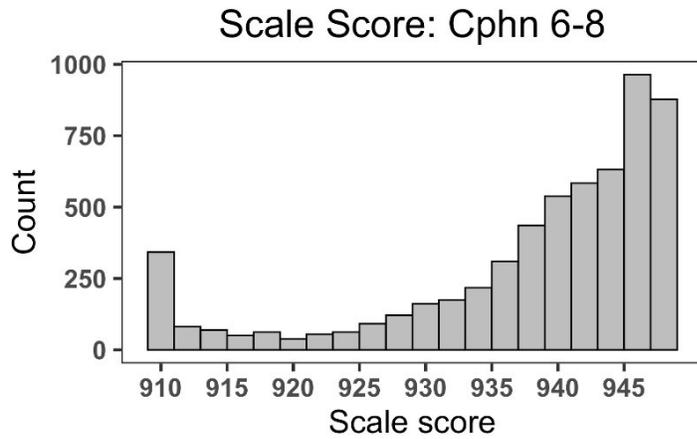


Table 6.3.7A
Scale Score Descriptive Statistics: Cphn 6-8

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,073	910	949	937.59	11.14
7	1,938	910	949	938.53	10.89
8	1,858	910	949	939.49	10.57
Total	5,869	910	949	938.50	10.91

Figure 6.3.7B

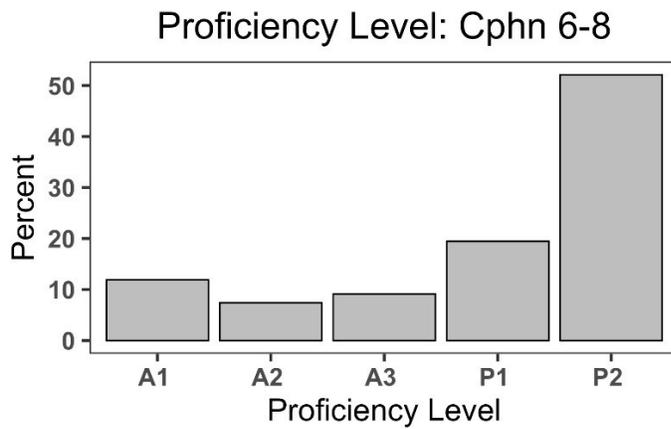


Table 6.3.7C

Proficiency Level Distribution: Cphn 6-8

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	277	13.36	225	11.61	196	10.55	698	11.89
A2	162	7.81	142	7.33	132	7.10	436	7.43
A3	218	10.52	183	9.44	135	7.27	536	9.13
P1	416	20.07	389	20.07	336	18.08	1,141	19.44
P2	1,000	48.24	999	51.55	1,059	57.00	3,058	52.10
Total	2,073	100.00	1,938	100.00	1,858	100.00	5,869	100.00

Table 6.3.7D
n/a

Figure 6.3.7D

n/a

Figure 6.3.7E

n/a

Table 6.3.7E

Reliability: Cphn 6-8

Component	Weight	Variance	Reliability
Listening	0.3	104.448	.927
Reading	0.7	133.865	.915
Comprehension		118.592	.947

*Variances from students who had results in all four domains

Table 6.3.7F

n/a

Table 6.3.7G

n/a

Table 6.3.7H

n/a

Table 6.3.7I

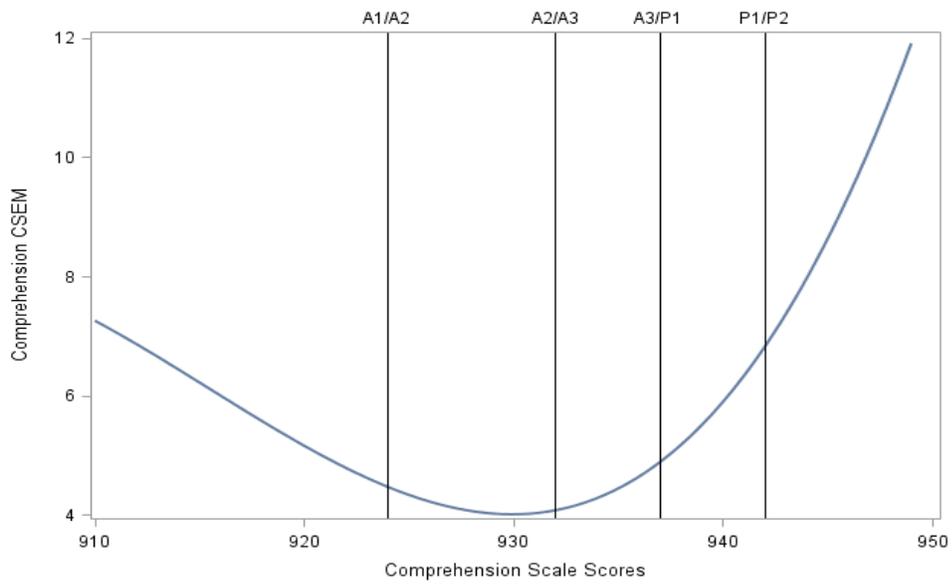
n/a

Table 6.3.7J

Accuracy and Consistency of Classification Indices: Cphn 6-8

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.823	0.763		0.633	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.948		0.920	
	A2	0.645		0.518	
	A3	0.603		0.478	
	P1	0.707		0.542	
	P2	0.870		0.849	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.985	0.007	0.008	0.979
	A2/A3	0.974	0.016	0.010	0.962
	A3/P1	0.955	0.022	0.023	0.938
	P1/P2	0.907	0.017	0.076	0.873

Figure 6.3.7F CSEM for Comprehension Composite 6-8



6.3.8 Overall Composite 6-8

Figure 6.3.8A

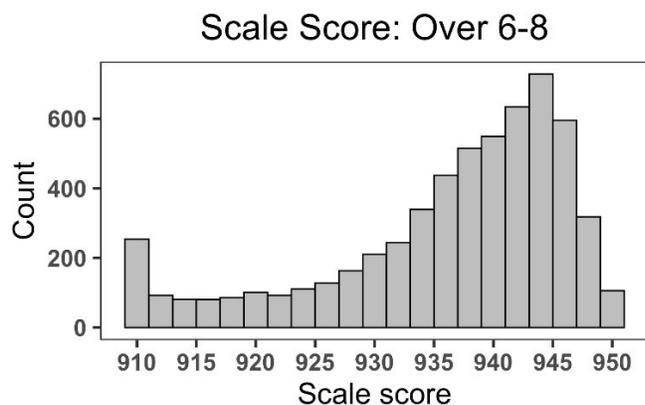


Table 6.3.8A
Scale Score Descriptive Statistics: Over 6-8

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,072	910	950	935.71	10.46
7	1,935	910	950	936.51	10.39
8	1,856	910	950	937.59	10.11
Total	5,863	910	950	936.57	10.35

Figure 6.3.8B

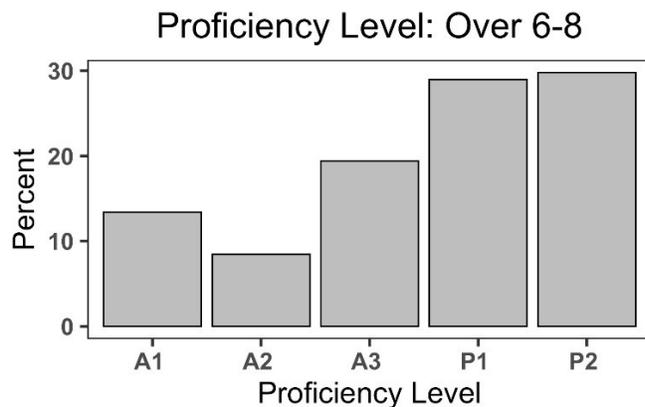


Table 6.3.8C

Proficiency Level Distribution: Over 6-8

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	306	14.77	256	13.23	224	12.07	786	13.41
A2	189	9.12	168	8.68	138	7.44	495	8.44
A3	423	20.42	377	19.48	337	18.16	1,137	19.39
P1	646	31.18	551	28.48	501	26.99	1,698	28.96
P2	508	24.52	583	30.13	656	35.34	1,747	29.80
Total	2,072	100.00	1,935	100.00	1,856	100.00	5,863	100.00

Table
6.3.8D n/a

Figure 6.3.8D

n/a

Figure 6.3.8E

n/a

Table 6.3.8E

Reliability: Over 6-8

Component	Weight	Variance	Reliability
Listening	0.15	104.448	.927
Reading	0.35	133.865	.915
Speaking	0.15	163.328	.888
Writing	0.35	117.506	.779
Overall Composite		106.916	.952

*Variances from students who had results in all four domains

Table 6.3.8F

n/a

Table 6.3.8G

n/a

Table 6.3.8H

n/a

Table 6.3.8I

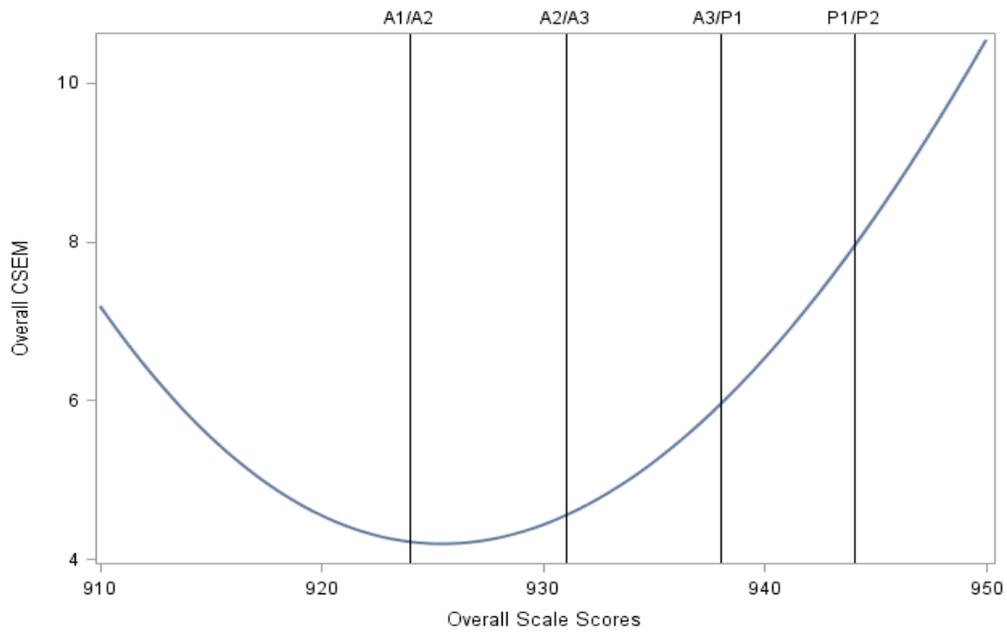
n/a

Table 6.3.8J

Accuracy and Consistency of Classification Indices: Over 6-8

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.738	0.657		0.552	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.954		0.932	
	A2	0.733		0.630	
	A3	0.887		0.829	
	P1	0.597		0.486	
	P2	0.684		0.637	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.987	0.007	0.006	0.981
	A2/A3	0.975	0.016	0.009	0.964
	A3/P1	0.963	0.010	0.027	0.950
	P1/P2	0.813	0.086	0.101	0.761

Figure 6.3.8F CSEM for Overall Composite 6-8



6.4 Grades: 9-12

6.4.1 Listening 9-12

Figure 6.4.1A

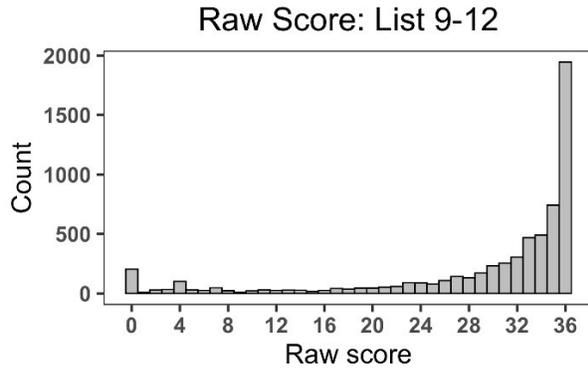


Table 6.4.1A

Raw Score Descriptive Statistics: List 9-12

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	1,545	0	36	29.12	9.92
10	1,477	0	36	29.76	9.45
11	1,308	0	36	29.91	9.51
12	1,902	0	36	29.31	9.65
Total	6,232	0	36	29.50	9.65

Figure 6.4.1B

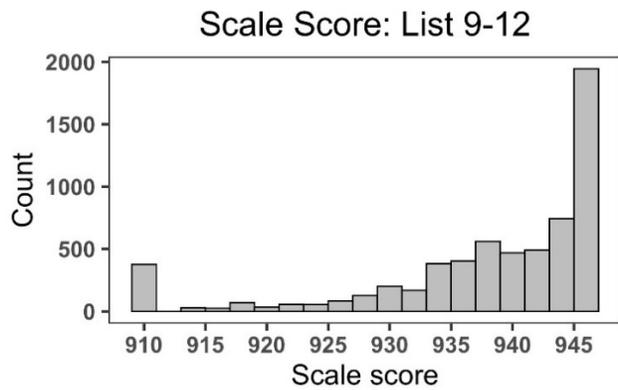


Table 6.4.1B

Scale Score Descriptive Statistics: List 9-12

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	1,545	910	947	938.37	10.42
10	1,477	910	947	939.03	9.92
11	1,308	910	947	939.29	9.96
12	1,902	910	947	938.58	10.09
Total	6,232	910	947	938.78	10.11

Figure 6.4.1B

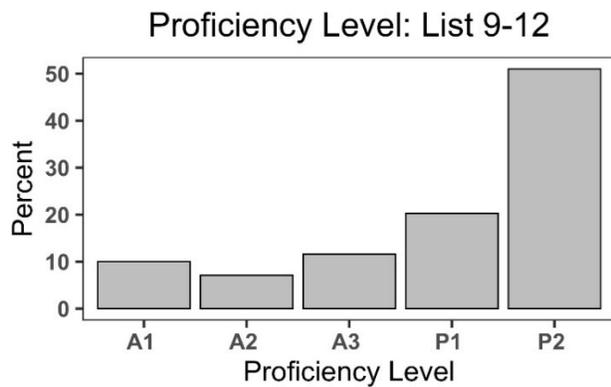


Table 6.4.1C

Proficiency Level Distribution: List 9-12

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	166	10.74	139	9.41	123	9.40	195	10.25	623	10.00
A2	117	7.57	95	6.43	84	6.42	146	7.68	442	7.09
A3	192	12.43	158	10.70	133	10.17	241	12.67	724	11.62
P1	303	19.61	322	21.80	273	20.87	366	19.24	1,264	20.28
P2	767	49.64	763	51.66	695	53.13	954	50.16	3,179	51.01
Total	1,545	100.00	1,477	100.00	1,308	100.00	1,902	100.00	6,232	100.00

Table 6.4.1D

Equating Summary: List 9-12

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 102. Thus, the results from the S102 of the Alternate ACCESS were used to determine raw-to-scale score conversion.

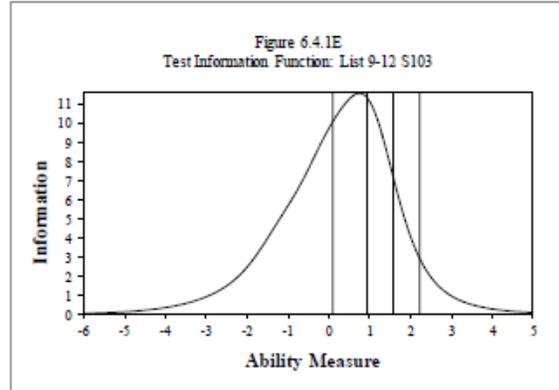
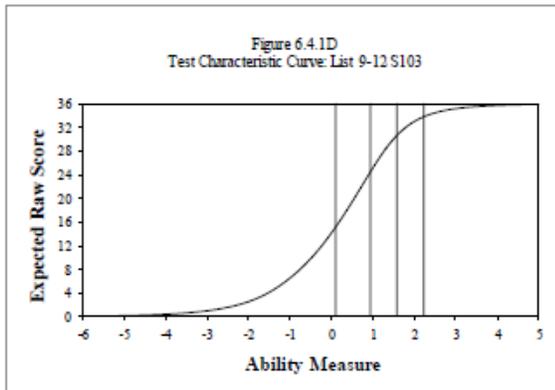


Table 6.4.1E

Reliability: List 9-12

No. of Students	No. of Items	Cronbach's Alpha	SEM
5,898	9	0.916	2.437

Table 6.4.1F

Item Analysis Summary: List 9-12

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	MOSR	9	0.00	82.83%	1.16	1.28
DIF Summary	DIF Level		Male/Female		Hispanic/Other	
			Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	6	3	3	6	
	BB	0	0	0	0	
CC	0	0	0	0		

Table 6.4.1G

Complete Item Analysis: List 9-12

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1. L1_A1_103	-1.38	Yes	88.00%	2.97	3.83	0.70	AA	M	AA	O
2. L2_A2_103	0.11	Yes	78.75%	1.24	1.00	0.86	AA	F	AA	O
3. L3_A2_103	-0.44	Yes	82.50%	1.00	0.73	0.89	AA	F	AA	O
4. L4_A3_103	0.71	Yes	73.25%	0.88	0.81	0.87	AA	F	AA	O
5. L5_A3_103	-0.23	Yes	80.50%	0.74	0.50	0.91	AA	M	AA	O
6. L6_P1_103	0.74	Yes	72.25%	0.71	0.69	0.89	AA	F	AA	H
7. L7_P1_103	0.77	Yes	72.25%	0.83	0.83	0.87	AA	F	AA	H
8. L8_P2_103	1.02	Yes	68.25%	0.91	0.86	0.83	AA	M	AA	H
9. L9_P2_103	0.77	Yes	70.25%	0.84	0.76	0.85	AA	M	AA	O

Table 6.4.1H

Raw Score to Scale Score Conversion: List 9-12

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	14.88	910.00^	910.00^
1	910^	8.23	910.00^	910.00^
2	910^	5.70	910.00^	910.00^
3	910^	4.75	910.00^	910.81
4	910^	4.27	910.00^	912.87
5	911	4.11	910.00^	914.93
6	913	4.04	910.00^	916.98
7	915	3.88	911.05	918.80
8	917	3.64	913.03	920.31
9	918	3.40	914.85	921.65
10	920	3.17	916.43	922.76
11	921	2.93	917.86	923.71
12	922	2.77	919.04	924.58
13	923	2.69	920.07	925.45
14	924	2.61	921.02	926.24
15	924	2.53	921.89	926.96
16	925	2.45	922.76	927.67
17	926	2.45	923.55	928.46
18	927	2.37	924.34	929.09
19	927	2.37	925.06	929.80
20	928	2.37	925.85	930.60
21	929	2.37	926.56	931.31
22	930	2.37	927.27	932.02
23	930	2.37	927.98	932.73
24	931	2.37	928.70	933.44
25	932	2.45	929.33	934.24
26	933	2.45	930.12	935.03
27	933	2.53	930.83	935.90
28	934	2.61	931.54	936.77
29	935	2.69	932.34	937.72
30	936	2.85	933.13	938.82
31	937	3.01	934.08	940.09
32	938	3.32	935.03	941.67
33	940	3.88	936.13	943.89
34	942*	4.83	937.48	947.13
35	944*	7.12	939.46	953.70
36	946*	13.93	940.96	968.81

^ Truncated

* Adjusted for end of scale effect

Table 6.4.11

Raw Score to Proficiency Level Conversion: List 9-12

Raw Score	Grade 9			Grade 10			Grade 11			Grade 12		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	2.85	2.85	A1	3.32	3.32	A1	3.75	3.75	A1	3.31	3.31
1	A1	0.19	3.04	A1	0.14	3.45	A1	0.08	3.82	A1	0.16	3.47
2	A1	0.52	3.56	A1	0.47	3.93	A1	0.38	4.20	A1	0.47	3.94
3	A1	0.78	4.34	A1	0.68	4.60	A1	0.46	4.66	A1	0.32	4.26
4	A1	2.27	6.60	A1	1.29	5.89	A1	1.22	5.89	A1	1.63	5.89
5	A1	0.58	7.18	A1	0.41	6.30	A1	0.38	6.27	A1	0.53	6.41
6	A1	0.71	7.90	A1	0.34	6.64	A1	0.00	6.27	A1	0.53	6.94
7	A1	1.04	8.93	A1	0.74	7.38	A1	0.76	7.03	A1	0.58	7.52
8	A1	0.13	9.06	A1	0.47	7.85	A1	0.46	7.49	A1	0.42	7.94
9	A1	0.19	9.26	A1	0.07	7.92	A1	0.08	7.57	A1	0.37	8.31
10	A1	0.39	9.64	A1	0.34	8.26	A1	0.46	8.03	A1	0.26	8.57
11	A1	0.58	10.23	A1	0.47	8.73	A1	0.46	8.49	A1	0.42	8.99
12	A1	0.19	10.42	A1	0.47	9.21	A1	0.15	8.64	A1	0.74	9.73
13	A1	0.32	10.74	A1	0.20	9.41	A1	0.76	9.40	A1	0.53	10.25
14	A2	0.58	11.33	A2	0.34	9.75	A2	0.46	9.86	A2	0.37	10.62
15	A2	0.65	11.97	A2	0.20	9.95	A2	0.08	9.94	A2	0.16	10.78
16	A2	0.58	12.56	A2	0.27	10.22	A2	0.54	10.47	A2	0.32	11.09
17	A2	0.78	13.33	A2	0.61	10.83	A2	0.61	11.09	A2	0.68	11.78
18	A2	0.45	13.79	A2	0.34	11.17	A2	0.84	11.93	A2	0.68	12.46
19	A2	0.71	14.50	A2	0.81	11.98	A2	0.54	12.46	A2	0.84	13.30
20	A2	0.65	15.15	A2	0.61	12.59	A2	0.76	13.23	A2	0.89	14.20
21	A2	1.04	16.18	A2	1.02	13.61	A2	0.46	13.69	A2	0.84	15.04
22	A2	0.84	17.02	A2	0.61	14.22	A2	0.84	14.53	A2	1.31	16.35
23	A2	1.29	18.32	A2	1.62	15.84	A2	1.30	15.83	A2	1.58	17.93
24	A3	1.36	19.68	A3	1.42	17.26	A3	1.45	17.28	A3	1.52	19.45
25	A3	1.23	20.91	A3	1.29	18.55	A3	0.69	17.97	A3	1.68	21.14
26	A3	1.88	22.78	A3	1.56	20.11	A3	1.38	19.34	A3	2.00	23.13
27	A3	2.39	25.18	A3	2.17	22.27	A3	2.45	21.79	A3	2.21	25.34
28	A3	2.20	27.38	A3	1.76	24.04	A3	2.45	24.24	A3	2.10	27.44
29	A3	3.37	30.74	A3	2.51	26.54	A3	1.76	25.99	A3	3.15	30.60
30	P1	3.43	34.17	P1	4.81	31.35	P1	3.36	29.36	P1	3.42	34.02
31	P1	3.56	37.73	P1	4.33	35.68	P1	4.28	33.64	P1	4.26	38.28
32	P1	4.47	42.20	P1	4.74	40.42	P1	5.12	38.76	P1	5.26	43.53
33	P1	8.16	50.36	P1	7.92	48.34	P1	8.10	46.87	P1	6.31	49.84
34	P2	7.90	58.25	P2	9.00	57.35	P2	6.73	53.59	P2	7.78	57.62
35	P2	11.97	70.23	P2	11.44	68.79	P2	12.08	65.67	P2	12.15	69.77
36	P2	29.77	100.00	P2	31.21	100.00	P2	34.33	100.00	P2	30.23	100.00

Table 6.4.1J

Accuracy and Consistency of Classification Indices: List 9-12

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.756	0.652		0.470
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.913		0.150	
	A2	0.456		0.146	
	A3	0.730		0.220	
	P1	0.399		0.166	
	P2	0.838		0.810	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.981	0.010	0.009	0.971
	A2/A3	0.958	0.029	0.013	0.942
	A3/P1	0.936	0.014	0.050	0.916
	P1/P2	0.872	0.030	0.098	0.791

6.4.2 Reading 9-12

Figure 6.4.2A

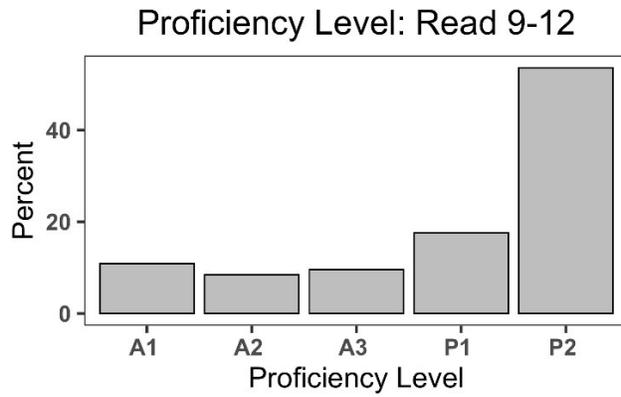


Figure 6.4.2B

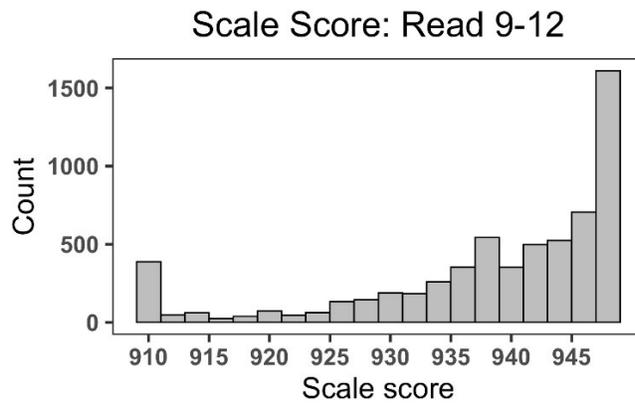


Figure 6.4.2C

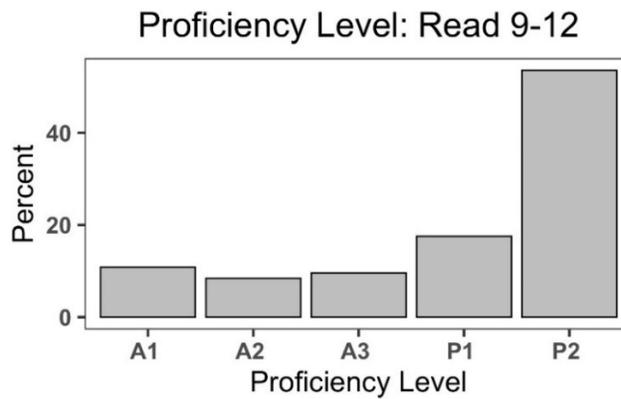


Table 6.4.2A

Raw Score Descriptive Statistics: Read 9-12

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	1,546	0	36	28.60	10.04
10	1,479	0	36	29.15	9.50
11	1,308	0	36	29.07	9.79
12	1,902	0	36	28.74	9.76
Total	6,235	0	36	28.87	9.78

Table 6.4.2B

Scale Score Descriptive Statistics: Read 9-12

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	1,546	910	948	938.31	11.01
10	1,479	910	948	938.83	10.46
11	1,308	910	948	938.86	10.72
12	1,902	910	948	938.44	10.75
Total	6,235	910	948	938.59	10.74

Table 6.4.2C

Proficiency Level Distribution: Read 9-12

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	182	11.77	147	9.94	144	11.01	204	10.73	677	10.86
A2	137	8.86	108	7.30	104	7.95	178	9.36	527	8.45
A3	147	9.51	147	9.94	103	7.87	200	10.52	597	9.57
P1	254	16.43	289	19.54	235	17.97	317	16.67	1,095	17.56
P2	826	53.43	788	53.28	722	55.20	1,003	52.73	3,339	53.55
Total	1,546	100.00	1,479	100.00	1,308	100.00	1,902	100.00	6,235	100.00

Table 6.4.2D

Equating Summary: Read 9-12

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 100. Thus, the results from the original field test of the Alternate ACCESS were used to determine raw-to-scale score conversions. Technical details of the analysis of this process can be found in the Alternate ACCESS for ELLs™ Series 100 Development and Operational Field Test: Technical Report (2013).

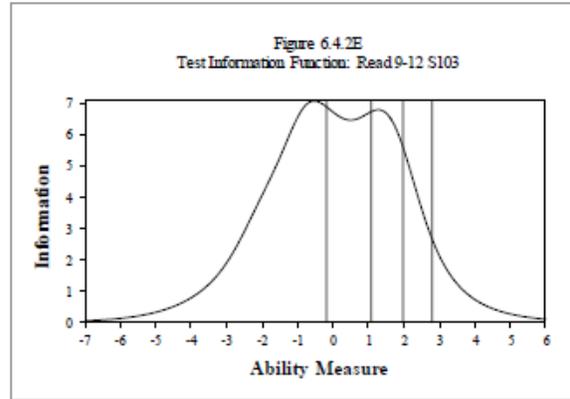
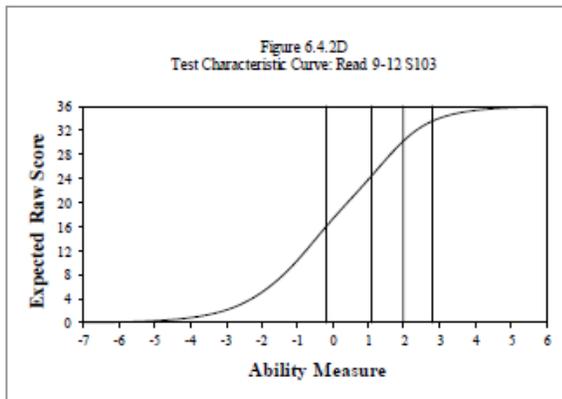


Table 6.4.2E

Reliability: Read 9-12

No. of Students	No. of Items	Cronbach's Alpha	SEM
5,866	9	0.908	2.677

Table 6.4.2F

Item Analysis Summary: Read 9-12

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	MOSR	9	0.00	79.19%	1.29	1.39

DIF Summary	DIF Level	Male/Female		Hispanic/Other	
		Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	7	2	5	4
	BB	0	0	0	0
CC	0	0	0	0	

Table 6.4.2G

Complete Item Analysis: Read 9-12

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1.R1_A1_103	-2.17		87.75%	3.45	3.63	0.72	AA	M	AA	O
2.R2_A2_103	-0.81	Yes	82.75%	1.32	0.93	0.88	AA	M	AA	O
3.R3_A2_103	-0.94	Yes	83.25%	1.08	0.70	0.88	AA	F	AA	H
4.R4_A3_103	0.01	Yes	78.00%	1.19	0.95	0.89	AA	F	AA	H
5.R5_A3_103	-0.93	Yes	82.00%	0.93	0.58	0.90	AA	M	AA	H
6.R6_P1_103	1.29	Yes	68.00%	0.91	0.85	0.85	AA	M	AA	O
7.R7_P1_103	0.95	Yes	69.00%	0.94	0.98	0.87	AA	M	AA	O
8.R8_P2_103	1.56	Yes	62.50%	0.85	0.96	0.81	AA	M	AA	O
9.R9_P2_103	1.74	Yes	60.50%	0.97	1.06	0.78	AA	F	AA	H

Table 6.4.2H

Raw Score to Scale Score Conversion: Read 9-12

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	11.51	910.00^	910.00^
1	910^	6.63	910.00^	910.00^
2	910^	4.70	910.00^	910.00^
3	910^	3.86	910.00^	911.45
4	910	3.50	910.00^	913.25
5	912	3.31	910.00^	915.00
6	913	3.25	910.24	916.75
7	915	3.13	912.05	918.32
8	917	2.95	913.74	919.64
9	918	2.77	915.30	920.85
10	919	2.59	916.63	921.81
11	920	2.47	917.83	922.78
12	921	2.35	918.92	923.62
13	922	2.29	919.88	924.46
14	923	2.29	920.73	925.31
15	924	2.29	921.63	926.21
16	925	2.29	922.47	927.05
17	926	2.29	923.32	927.90
18	927	2.35	924.16	928.86
19	927	2.35	925.06	929.77
20	928	2.35	926.03	930.73
21	929	2.41	926.87	931.69
22	930	2.41	927.84	932.66
23	931	2.41	928.80	933.62
24	932	2.35	929.83	934.53
25	933	2.35	930.73	935.43
26	934	2.35	931.63	936.33
27	935	2.35	932.54	937.24
28	936	2.35	933.44	938.14
29	937	2.41	934.34	939.17
30	938	2.47	935.25	940.19
31	939	2.65	936.15	941.46
32	940	2.83	937.24	942.90
33	942	3.25	938.32	944.83
34	944*	3.98	939.71	947.66
35	946*	5.72	941.64	953.09
36	948*	10.85	943.38	965.08

^ Truncated

* Adjusted for end of scale effect

WIDA ALTERNATE ACCE SS Annual Tech Rpt 7

201

Series 403 (2018-2019)

[Return to Chapter 4 Visual Guide to Tables and Figures](#)

Table 6.4.2I

Raw Score to Proficiency Level Conversion: Read 9-12

Raw Score	Grade 9			Grade 10			Grade 11			Grade 12		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	3.75	3.75	A1	3.85	3.85	A1	4.20	4.20	A1	3.94	3.94
1	A1	0.39	4.14	A1	0.07	3.92	A1	0.08	4.28	A1	0.16	4.10
2	A1	0.71	4.85	A1	0.34	4.26	A1	0.38	4.66	A1	0.16	4.26
3	A1	0.71	5.56	A1	0.47	4.73	A1	0.38	5.05	A1	0.47	4.73
4	A1	1.03	6.60	A1	1.28	6.02	A1	1.07	6.12	A1	1.42	6.15
5	A1	0.58	7.18	A1	0.54	6.56	A1	0.31	6.42	A1	0.37	6.52
6	A1	0.32	7.50	A1	0.47	7.03	A1	0.23	6.65	A1	0.26	6.78
7	A1	0.91	8.41	A1	0.61	7.64	A1	1.07	7.72	A1	1.26	8.04
8	A1	0.71	9.12	A1	0.07	7.71	A1	0.46	8.18	A1	0.37	8.41
9	A1	0.39	9.51	A1	0.27	7.98	A1	0.15	8.33	A1	0.11	8.52
10	A1	0.32	9.83	A1	0.41	8.38	A1	0.61	8.94	A1	0.26	8.78
11	A1	0.52	10.35	A1	0.81	9.20	A1	0.69	9.63	A1	0.37	9.15
12	A1	0.58	10.93	A1	0.27	9.47	A1	0.54	10.17	A1	0.84	9.99
13	A1	0.32	11.25	A1	0.27	9.74	A1	0.46	10.63	A1	0.53	10.52
14	A1	0.52	11.77	A1	0.20	9.94	A1	0.38	11.01	A1	0.21	10.73
15	A2	0.45	12.23	A2	0.20	10.14	A2	0.54	11.54	A2	0.63	11.36
16	A2	0.58	12.81	A2	0.54	10.68	A2	0.31	11.85	A2	0.63	11.99
17	A2	0.84	13.65	A2	0.54	11.22	A2	0.69	12.54	A2	0.68	12.67
18	A2	0.65	14.29	A2	0.54	11.76	A2	0.54	13.07	A2	0.58	13.25
19	A2	1.03	15.33	A2	0.81	12.58	A2	0.84	13.91	A2	0.74	13.99
20	A2	0.84	16.17	A2	1.01	13.59	A2	0.92	14.83	A2	1.10	15.09
21	A2	1.49	17.66	A2	1.42	15.01	A2	1.15	15.98	A2	1.31	16.40
22	A2	1.10	18.76	A2	0.68	15.69	A2	1.45	17.43	A2	1.52	17.93
23	A2	1.88	20.63	A2	1.56	17.24	A2	1.53	18.96	A2	2.16	20.08
24	A3	1.29	21.93	A3	1.28	18.53	A3	0.92	19.88	A3	1.79	21.87
25	A3	1.42	23.35	A3	1.76	20.28	A3	1.22	21.10	A3	1.84	23.71
26	A3	1.49	24.84	A3	2.16	22.45	A3	1.61	22.71	A3	1.79	25.50
27	A3	2.65	27.49	A3	2.37	24.81	A3	2.06	24.77	A3	2.42	27.92
28	A3	2.65	30.14	A3	2.37	27.18	A3	2.06	26.83	A3	2.68	30.60
29	P1	3.17	33.31	P1	3.38	30.56	P1	3.21	30.05	P1	3.10	33.70
30	P1	3.43	36.74	P1	5.48	36.04	P1	4.20	34.25	P1	4.52	38.22
31	P1	4.20	40.94	P1	4.73	40.77	P1	4.36	38.61	P1	4.00	42.22
32	P1	5.63	46.57	P1	5.95	46.72	P1	6.19	44.80	P1	5.05	47.27
33	P2	8.28	54.85	P2	8.45	55.17	P2	7.95	52.75	P2	7.41	54.68
34	P2	8.34	63.20	P2	8.25	63.42	P2	8.64	61.39	P2	8.46	63.14
35	P2	11.38	74.58	P2	11.43	74.85	P2	12.08	73.47	P2	10.67	73.82
36	P2	25.42	100.00	P2	25.15	100.00	P2	26.53	100.00	P2	26.18	100.00

Table 6.4.2J

Accuracy and Consistency of Classification Indices: Read 9-12

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.737	0.650		0.487	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.876		0.127	
	A2	0.605		0.207	
	A3	0.576		0.199	
	P1	0.530		0.196	
	P2	0.814		0.782	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.976	0.013	0.010	0.965
	A2/A3	0.951	0.027	0.021	0.932
	A3/P1	0.933	0.025	0.042	0.909
	P1/P2	0.867	0.030	0.103	0.809

6.4.3 Speaking 9-12

Figure 6.4.3A

Raw Score: Spek 9-12

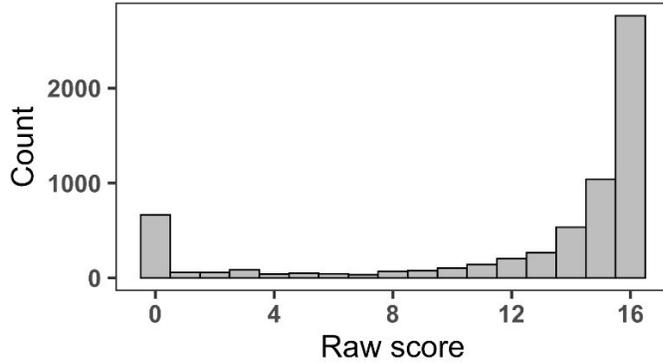


Table 6.4.3A

Raw Score Descriptive Statistics: Spek 9-12

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	1,544	0	16	12.49	5.46
10	1,481	0	16	12.76	5.26
11	1,308	0	16	12.78	5.23
12	1,900	0	16	12.35	5.44
Total	6,233	0	16	12.57	5.36

Figure 6.4.3B

Scale Score: Spek 9-12

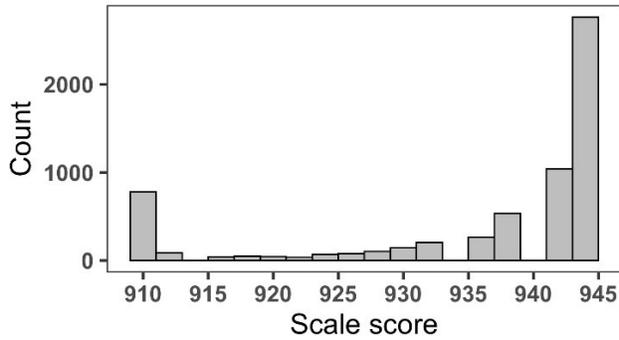


Table 6.4.3B

Scale Score Descriptive Statistics: Spek 9-12

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	1,544	910	945	936.42	12.37
10	1,481	910	945	937.04	11.91
11	1,308	910	945	937.08	11.83
12	1,900	910	945	936.02	12.31
Total	6,233	910	945	936.59	12.14

Figure 6.4.3C

Proficiency Level: Spek 9-12

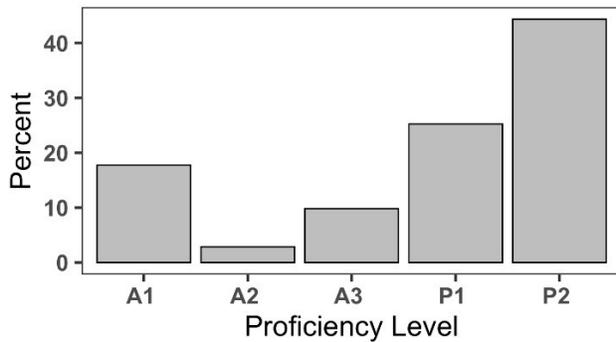


Table 6.4.3C

Proficiency Level Distribution: Spek 9-12

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	286	18.52	247	16.68	208	15.90	365	19.21	1,106	17.74
A2	39	2.53	38	2.57	40	3.06	62	3.26	179	2.87
A3	145	9.39	135	9.12	130	9.94	201	10.58	611	9.80
P1	387	25.06	386	26.06	314	24.01	486	25.58	1,573	25.24
P2	687	44.49	675	45.58	616	47.09	786	41.37	2,764	44.34
Total	1,544	100.00	1,481	100.00	1,308	100.00	1,900	100.00	6,233	100.00

Table 6.4.3D

Equating Summary: Spek 9-12

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 100. Thus, the results from the original field test of the Alternate ACCESS were used to determine raw-to-scale score conversions. Technical details of the analysis of this process can be found in the Alternate ACCESS for ELLs™ Series 100 Development and Operational Field Test: Technical Report (2013).

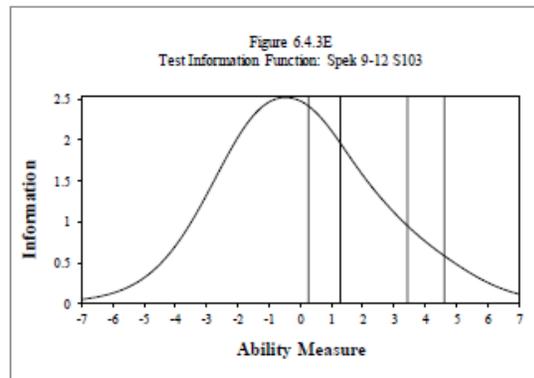
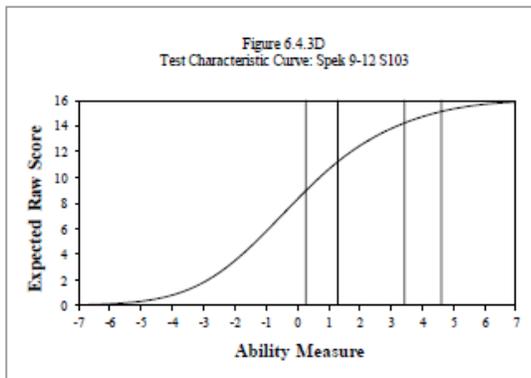


Table 6.4.3E

Reliability: Spek 9-12

No. of Students	No. of Items	Cronbach's Alpha	SEM
5,466	8	0.898	2.531

Table 6.4.3F

Item Analysis Summary: Spek 9-12

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	CR	8	0.00	76.94%	0.97	0.84
DIF Summary	DIF Level		Male/Female		Hispanic/Other	
			Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	4	4	5	3	
	BB	0	0	0	0	
CC	0	0	0	0		

Table 6.4.3G

Complete Item Analysis: Spek 9-12

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1.S1_A1_103	-2.09		80.50%	1.43	1.65	0.90	AA	F	AA	O
2.S2_A2_103	-1.16		79.00%	1.27	1.23	0.91	AA	M	AA	H
3.S3_A3_103	-0.43	Yes	77.00%	0.98	0.82	0.93	AA	F	AA	H
4.S4_A1_103	-0.69	Yes	77.00%	0.90	0.90	0.93	AA	M	AA	O
5.S5_A2_103	-0.47	Yes	77.00%	0.84	0.73	0.94	AA	F	AA	H
6.S6_A3_103	-0.16	Yes	76.00%	0.75	0.64	0.94	AA	M	AA	O
7.S7_P1_103	1.74	Yes	69.00%	0.88	0.76	0.87	AA	M	AA	H
8.S8_P2_103	3.64	Yes	56.00%	0.78	0.90	0.73	AA	M	AA	H

Table 6.4.3H

Raw Score to Scale Score Conversion: Spek 9-12

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	8.56	910.00^	910.00^
1	910^	5.19	910.00^	911.14
2	910	3.95	910.00^	914.42
3	913	3.41	910.08	916.91
4	916	3.15	912.78	919.08
5	918	3.01	915.04	921.07
6	920	2.97	917.13	923.07
7	922	2.93	919.12	924.97
8	924	2.93	921.07	926.92
9	926	2.97	922.98	928.92
10	928	3.01	924.97	931.00
11	930	3.19	926.97	933.35
12	933	3.41	929.19	936.01
13	936	3.81	931.71	939.34
14	939	4.43	934.90	943.77
15	942*	5.67	939.25	950.60
16	945*	8.82	943.46	961.10

^ Truncated

* Adjusted for end of scale effect

Table 6.4.3I

Raw Score to Proficiency Level Conversion: Spek 9-12

Raw Score	Grade 9			Grade 10			Grade 11			Grade 12		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	A1	11.14	11.14	A1	10.20	10.20	A1	10.17	10.17	A1	11.00	11.00
1	A1	0.84	11.98	A1	0.95	11.14	A1	0.99	11.16	A1	0.95	11.95
2	A1	1.36	13.34	A1	1.01	12.15	A1	0.61	11.77	A1	0.68	12.63
3	A1	1.42	14.77	A1	0.95	13.10	A1	1.15	12.92	A1	1.84	14.47
4	A1	0.78	15.54	A1	0.47	13.57	A1	0.61	13.53	A1	0.74	15.21
5	A1	0.39	15.93	A1	1.08	14.65	A1	0.69	14.22	A1	1.00	16.21
6	A1	1.04	16.97	A1	0.41	15.06	A1	0.23	14.45	A1	1.00	17.21
7	A1	0.52	17.49	A1	0.41	15.46	A1	0.46	14.91	A1	0.79	18.00
8	A1	1.04	18.52	A1	1.22	16.68	A1	0.99	15.90	A1	1.21	19.21
9	A2	1.04	19.56	A2	1.01	17.69	A2	1.30	17.20	A2	1.47	20.68
10	A2	1.49	21.05	A2	1.55	19.24	A2	1.76	18.96	A2	1.79	22.47
11	A3	2.33	23.38	A3	1.76	21.00	A3	2.37	21.33	A3	2.63	25.11
12	A3	3.17	26.55	A3	3.11	24.11	A3	3.82	25.15	A3	3.05	28.16
13	A3	3.89	30.44	A3	4.25	28.36	A3	3.75	28.90	A3	4.89	33.05
14	P1	8.35	38.80	P1	8.78	37.14	P1	9.25	38.15	P1	8.11	41.16
15	P1	16.71	55.51	P1	17.29	54.42	P1	14.76	52.91	P1	17.47	58.63
16	P2	44.49	100.00	P2	45.58	100.00	P2	47.09	100.00	P2	41.37	100.00

Table 6.4.3J

Accuracy and Consistency of Classification Indices: Spek 9-12

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.558	0.576		0.424
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.943		0.245	
	A2	0.502		0.150	
	A3	0.699		0.081	
	P1	0.419		0.406	
	P2	-		0.566	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.979	0.012	0.009	0.970
	A2/A3	0.974	0.012	0.014	0.964
	A3/P1	0.956	0.011	0.033	0.933
	P1/P2	0.646	0.354	0.000	0.683

6.4.4 Writing 9-12

Figure 6.4.4A

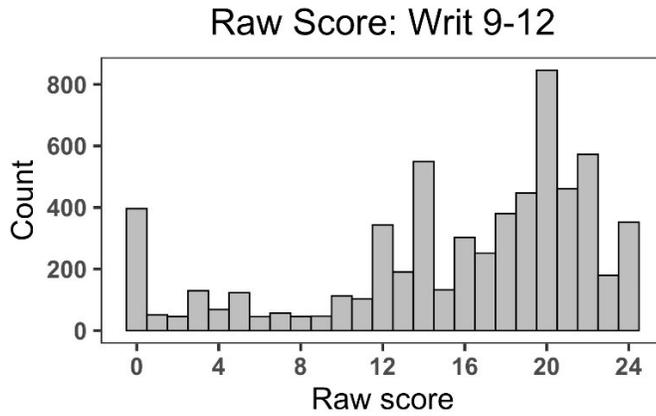


Table 6.4.4A

Raw Score Descriptive Statistics: Writ 9-12

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	1,546	0	24	15.45	6.73
10	1,480	0	24	15.84	6.64
11	1,308	0	24	15.99	6.63
12	1,900	0	24	15.47	6.82
Total	6,234	0	24	15.67	6.72

Figure 6.4.4B

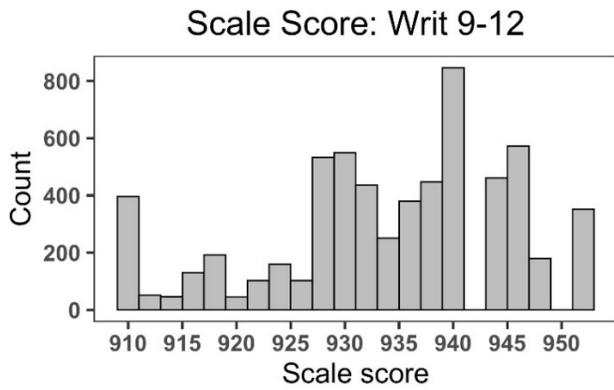


Table 6.4.4B

Scale Score Descriptive Statistics: Writ 9-12

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	1,546	910	953	934.06	11.01
10	1,480	910	953	934.67	10.94
11	1,308	910	953	934.96	11.03
12	1,900	910	953	934.10	11.26
Total	6,234	910	953	934.40	11.08

Figure 6.4.4C

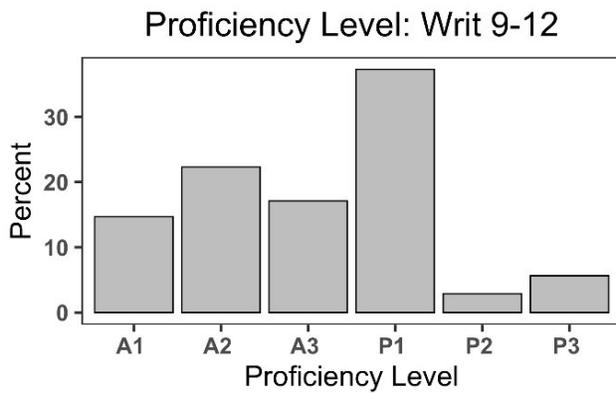


Table 6.4.4C

Proficiency Level Distribution: Writ 9-12

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	240	15.52	211	14.26	175	13.38	291	15.32	917	14.71
A2	362	23.42	308	20.81	281	21.48	440	23.16	1,391	22.31
A3	252	16.30	259	17.50	229	17.51	327	17.21	1,067	17.12
P1	573	37.06	590	39.86	494	37.77	670	35.26	2,327	37.33
P2	42	2.72	33	2.23	48	3.67	57	3.00	180	2.89
P3	77	4.98	79	5.34	81	6.19	115	6.05	352	5.65
Total	1,546	100.00	1,480	100.00	1,308	100.00	1,900	100.00	6,234	100.00

Table 6.4.4D

Equating Summary: Writ 9-12

No equating summary is presented because the Alternate ACCESS Series 403 was not equated. There is no change from the field test Series 100. Thus, the results from the original field test of the Alternate ACCESS were used to determine raw-to-scale score conversions. Technical details of the analysis of this process can be found in the Alternate ACCESS for ELLs™ Series 100 Development and Operational Field Test: Technical Report (2013).

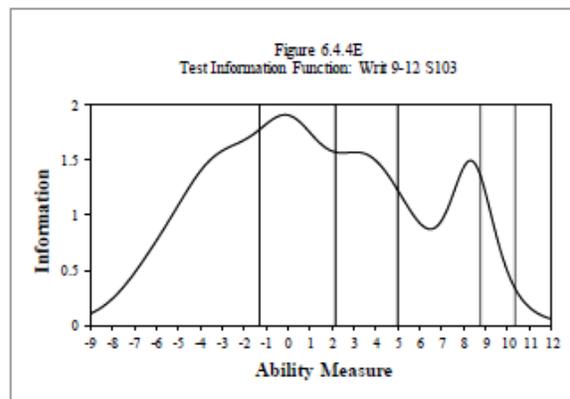
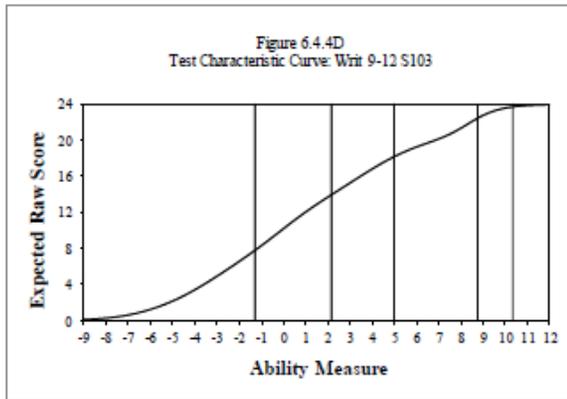


Table 6.4.4E

Reliability: Writ 9-12

No. of Students	No. of Items	Cronbach's Alpha	SEM
5,337	10	0.774	3.762

Table 6.4.4F

Item Analysis Summary: Writ 9-12

Item Summary	Item Type	No. of Items	Average Item Difficulty (in logits)	Average of % of Max. Possible Score Points	Average Infit Mean	Average Outfit Mean Square
	CR	10	0.00	68.75%	1.42	5.48

DIF Summary	DIF Level	Male/Female		Hispanic/Other	
		Favoring Male (M)	Favoring Female (F)	Favoring Hispanic (H)	Favoring Other (O)
	AA	4	6	4	6
	BB	0	0	0	0
CC	0	0	0	0	

Table 6.4.4G

Complete Item Analysis: Writ 9-12

Name	Item Difficulty (in logits)	Anchored?	% of Max. Possible Score Points	Fit Statistics		Point Biserial	DIF			
				Infit Mnsq	Outfit Mnsq		M/F		H/O	
							DIF Level	Favored Group	DIF Level	Favored Group
1.W1_A1_103	-4.41		83.50%	2.88	9.90	0.75	AA	M	AA	O
2.W2_A2_103	-3.09		81.00%	1.73	9.90	0.81	AA	F	AA	H
3.W3_A3_103	-1.35		77.00%	1.69	6.60	0.83	AA	M	AA	O
4.W4_P1_103	1.73	Yes	62.50%	1.43	2.18	0.84	AA	F	AA	O
5.W5_A1_103	-2.00	Yes	78.00%	1.40	9.90	0.83	AA	M	AA	H
6.W6_A2_103	-1.77	Yes	77.50%	1.25	8.39	0.84	AA	M	AA	H
7.W7_A3_103	-0.30		73.00%	1.50	6.32	0.84	AA	M	AA	O
8.W8_P1_103	1.83	Yes	62.50%	1.07	1.27	0.87	AA	F	AA	H
9.W9_P3_103	6.28	Yes	33.25%	1.03	3.13	0.71	AA	F	AA	O
10.W10_P3_103	6.27	Yes	33.75%	1.04	2.70	0.66	AA	M	AA	O

Table 6.4.4H

Raw Score to Scale Score Conversion: Writ 9-12

Raw Score	Scale Score	SE Scaled	Low Bound	High Bound
0	910^	4.68	910.00^	912.75
1	912	2.90	910.00^	914.67
2	915	2.30	912.20	916.81
3	916	1.99	914.43	918.42
4	918	1.82	916.11	919.76
5	919	1.75	917.50	921.01
6	921	1.70	918.80	922.21
7	922	1.70	920.00	923.41
8	923	1.68	921.22	924.58
9	924	1.68	922.40	925.76
10	925	1.66	923.58	926.89
11	926	1.68	924.70	928.06
12	928	1.70	925.88	929.29
13	929	1.78	927.08	930.63
14	930	1.82	928.38	932.02
15	932	1.82	929.74	933.39
16	933	1.82	931.14	934.78
17	934	1.87	932.50	936.25
18	936	2.04	933.92	938.00
19	938	2.42	935.58	940.42
20	941	2.88	938.14	943.90
21	944	2.38	941.58	946.33
22	946	2.18	943.86	948.22
23	948*	2.59	945.68	950.86
24	950*	4.44	946.93	955.81

^ Truncated

* Adjusted for end of scale effect

Table 6.4.4I

Raw Score to Proficiency Level Conversion: Writ 9-12

Raw Score	Grade 9			Grade 10			Grade 11			Grade 12		
	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students	Proficiency Level Score	% of Students	Cumulative % of Students
0	6.21	6.21	A1	6.01	6.01	A1	6.12	6.12	A1	6.89	6.89	6.21
1	0.45	6.66	A1	1.08	7.09	A1	0.69	6.80	A1	1.00	7.89	0.45
2	1.16	7.83	A1	0.68	7.77	A1	0.61	7.42	A1	0.53	8.42	1.16
3	2.13	9.96	A1	1.82	9.59	A1	2.06	9.48	A1	2.26	10.68	2.13
4	1.49	11.45	A1	1.22	10.81	A1	0.76	10.24	A1	0.95	11.63	1.49
5	2.01	13.45	A1	1.89	12.70	A1	1.91	12.16	A1	2.05	13.68	2.01
6	1.10	14.55	A1	0.68	13.38	A1	0.38	12.54	A1	0.68	14.37	1.10
7	0.97	15.52	A1	0.88	14.26	A1	0.84	13.38	A1	0.95	15.32	0.97
8	0.52	16.04	A2	0.61	14.86	A2	0.92	14.30	A2	0.89	16.21	0.52
9	1.10	17.14	A2	0.81	15.68	A2	0.69	14.98	A2	0.47	16.68	1.10
10	2.26	19.40	A2	1.35	17.03	A2	1.61	16.59	A2	1.95	18.63	2.26
11	1.42	20.83	A2	1.35	18.38	A2	1.53	18.12	A2	2.16	20.79	1.42
12	5.69	26.52	A2	4.86	23.24	A2	5.66	23.78	A2	5.74	26.53	5.69
13	3.30	29.82	A2	3.31	26.55	A2	2.22	25.99	A2	3.21	29.74	3.30
14	9.12	38.94	A2	8.51	35.07	A2	8.87	34.86	A2	8.74	38.47	9.12
15	1.94	40.88	A3	2.50	37.57	A3	1.99	36.85	A3	2.11	40.58	1.94
16	4.46	45.34	A3	5.07	42.64	A3	5.20	42.05	A3	4.79	45.37	4.46
17	4.72	50.06	A3	3.99	46.62	A3	3.44	45.49	A3	3.89	49.26	4.72
18	5.17	55.24	A3	5.95	52.57	A3	6.88	52.37	A3	6.42	55.68	5.17
19	6.53	61.77	P1	7.30	59.86	P1	6.88	59.25	P1	7.79	63.47	6.53
20	14.17	75.94	P1	14.66	74.53	P1	14.07	73.32	P1	11.89	75.37	14.17
21	7.83	83.76	P1	7.77	82.30	P1	7.49	80.81	P1	6.68	82.05	7.83
22	8.54	92.30	P1	10.14	92.43	P1	9.33	90.14	P1	8.89	90.95	8.54
23	2.72	95.02	P2	2.23	94.66	P2	3.67	93.81	P2	3.00	93.95	2.72
24	4.98	100.00	P3	5.34	100.00	P3	6.19	100.00	P3	6.05	100.00	4.98

Table 6.4.4J

Accuracy and Consistency of Classification Indices: Writ 9-12

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.732	0.639		0.514	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.835		0.134	
	A2	0.763		0.302	
	A3	0.535		0.125	
	P1	0.752		0.753	
	P2	-		0.221	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.949	0.031	0.020	0.926
	A2/A3	0.926	0.025	0.049	0.899
	A3/P1	0.923	0.033	0.044	0.887
	P1/P2	0.932	0.068	0.000	0.909

6.4.5 Oral Language Composite 9-12

Figure 6.4.5A

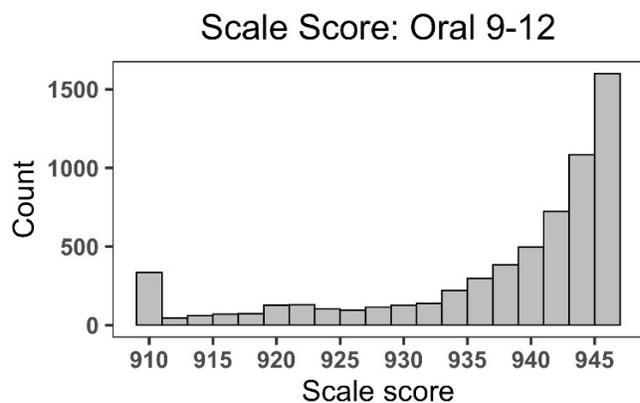


Table 6.4.5A

Scale Score Descriptive Statistics: Oral 9-12

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	1,543	910	946	937.54	10.71
10	1,477	910	946	938.19	10.28
11	1,308	910	946	938.33	10.25
12	1,900	910	946	937.48	10.53
Total	6,228	910	946	937.84	10.47

Figure 6.4.5B

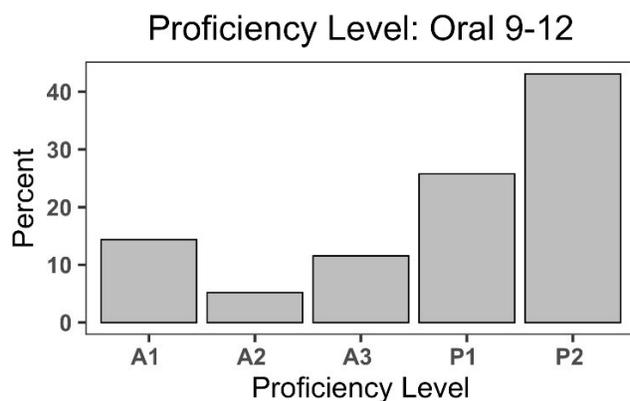


Table 6.4.5C

Proficiency Level Distribution: Oral 9-12

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	246	15.94	205	13.88	171	13.07	273	14.37	895	14.37
A2	77	4.99	70	4.74	54	4.13	123	6.47	324	5.20
A3	164	10.63	147	9.95	169	12.92	241	12.68	721	11.58
P1	397	25.73	410	27.76	319	24.39	478	25.16	1,604	25.75
P2	659	42.71	645	43.67	595	45.49	785	41.32	2,684	43.10
Total	1,543	100.00	1,477	100.00	1,308	100.00	1,900	100.00	6,228	100.00

Table 6.4.5D

n/a

Figure 6.4.5D

n/a

Figure 6.4.5E
n/a

Table 6.4.5E

Reliability: Oral 9-12

Component	Weight	Variance	Reliability
Listening	0.5	102.01	.916
Speaking	0.5	147.1369	.898
Oral		109.2025	.946

*Variances from students who had results in all four domains

Table 6.4.5F
n/a

Table 6.4.5G
n/a

Table 6.4.5H
n/a

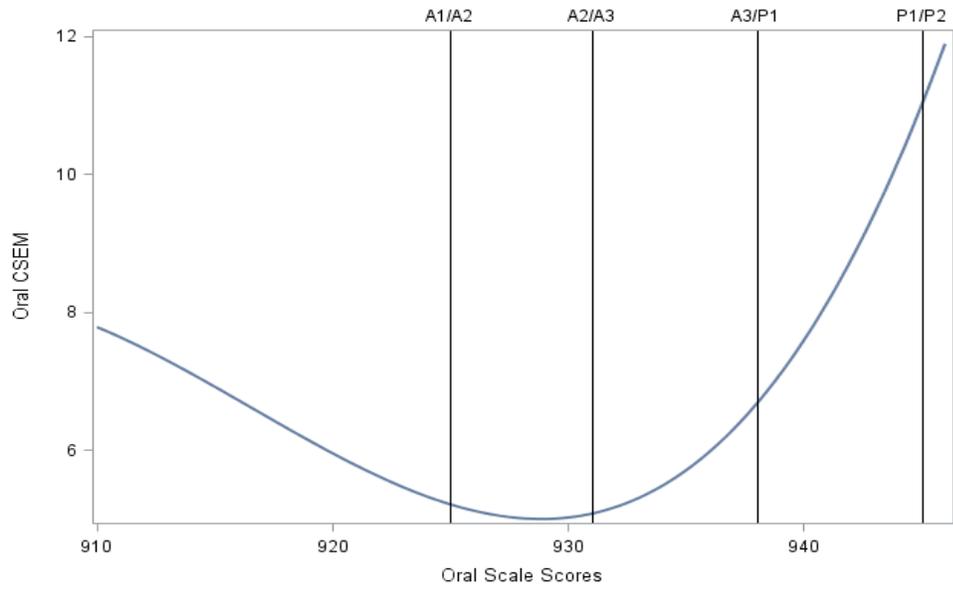
Table 6.4.5I
n/a

Table 6.4.5J

Accuracy and Consistency of Classification Indices: Oral 9-12

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.753	0.644		0.515	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.950		0.923	
	A2	0.589		0.463	
	A3	0.821		0.734	
	P1	0.626		0.441	
	P2	0.751		0.715	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.984	0.008	0.009	0.976
	A2/A3	0.975	0.015	0.010	0.965
	A3/P1	0.965	0.011	0.023	0.952
	P1/P2	0.829	0.057	0.114	0.749

Figure 6.4.5F CSEM for Oral Composite 9-12



6.4.6 Literacy Composite 9-12

Figure 6.4.6A

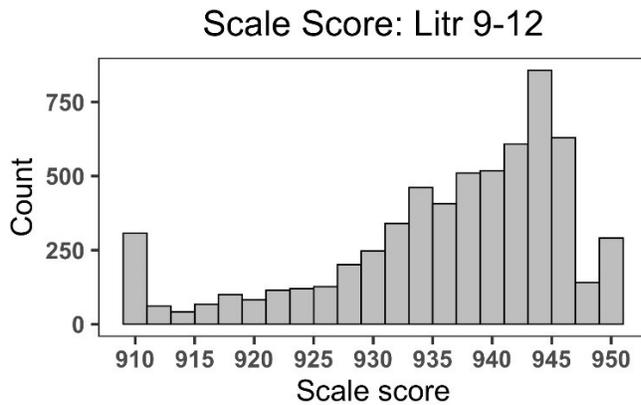


Table 6.4.6A

Scale Score Descriptive Statistics: Litr 9-12

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	1,546	910	951	936.39	10.37
10	1,479	910	951	936.96	10.11
11	1,307	910	951	937.12	10.27
12	1,900	910	951	936.48	10.44
Total	6,232	910	951	936.70	10.31

Figure 6.4.6B

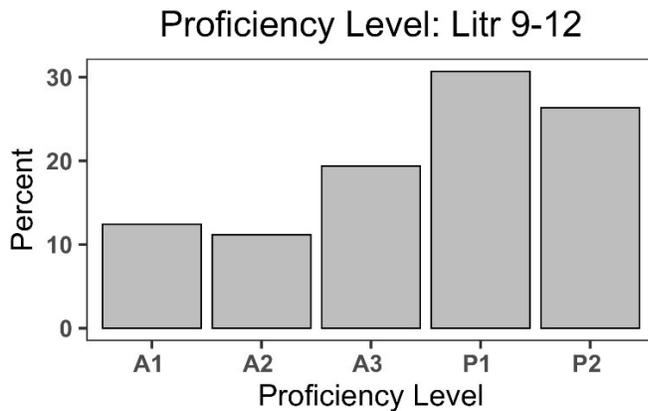


Table 6.4.6C

Proficiency Level Distribution: Litr 9-12

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	209	13.52	169	11.43	162	12.39	235	12.37	775	12.44
A2	178	11.51	151	10.21	125	9.56	241	12.68	695	11.15
A3	290	18.76	309	20.89	241	18.44	368	19.37	1,208	19.38
P1	484	31.31	442	29.89	418	31.98	569	29.95	1,913	30.70
P2	385	24.90	408	27.59	361	27.62	487	25.63	1,641	26.33
Total	1,546	100.00	1,479	100.00	1,307	100.00	1,900	100.00	6,232	100.00

Table 6.4.6D

n/a

Figure 6.4.6D

n/a

Figure 6.4.6E
n/a

Table 6.4.6E

Reliability: Litr 9-12

Component	Weight	Variance	Reliability
Reading	0.5	115.348	.908
Writing	0.5	122.545	.774
Literacy		106.090	.910

*Variances from students who had results in all four domains

Table 6.4.6F
n/a

Table 6.4.6G
n/a

Table 6.4.6H
n/a

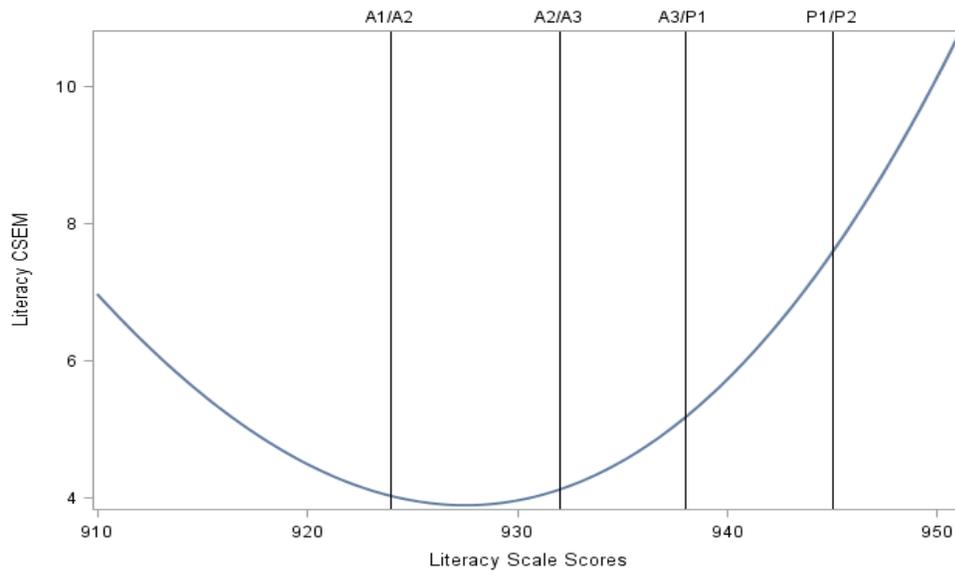
Table 6.4.6I
n/a

Table 6.4.6J

Accuracy and Consistency of Classification Indices: Litr 9-12

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.633	0.609		0.496	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.895		0.842	
	A2	0.747		0.644	
	A3	0.790		0.694	
	P1	0.492		0.495	
	P2	-		0.589	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.976	0.014	0.010	0.965
	A2/A3	0.952	0.026	0.022	0.933
	A3/P1	0.943	0.015	0.042	0.921
	P1/P2	0.762	0.238	0.000	0.786

Figure 6.4.6F CSEM for Literacy Composite 9-12



6.4.7 Comprehension Composite 9-12

Scale Score: Cphn 9-12

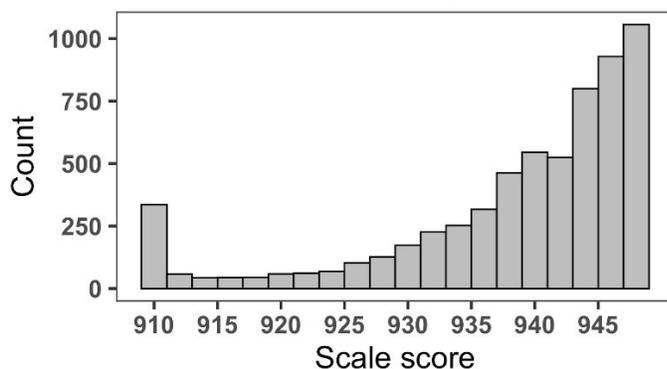


Table 6.4.7A

Scale Score Descriptive Statistics: Cphn 9-12

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	1,545	910	948	938.40	10.61
10	1,477	910	948	938.97	10.12
11	1,308	910	948	939.07	10.31
12	1,902	910	948	938.57	10.35
Total	6,232	910	948	938.73	10.36

Proficiency Level: Cphn 9-12

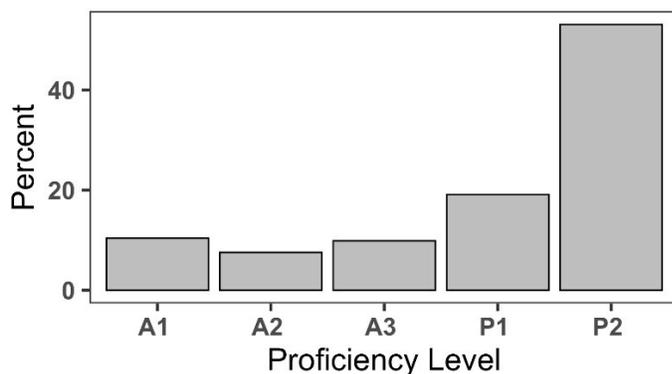


Table 6.4.7C

Proficiency Level Distribution: Cphn 9-12

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	172	11.13	141	9.55	135	10.32	199	10.46	647	10.38
A2	131	8.48	100	6.77	94	7.19	147	7.73	472	7.57
A3	144	9.32	144	9.75	108	8.26	218	11.46	614	9.85
P1	285	18.45	311	21.06	247	18.88	347	18.24	1,190	19.09
P2	813	52.62	781	52.88	724	55.35	991	52.10	3,309	53.10
Total	1,545	100.00	1,477	100.00	1,308	100.00	1,902	100.00	6,232	100.00

Table 6.4.7D

n/a

Figure 6.4.7D

n/a

Figure 6.4.7E
n/a

Table 6.4.7E

Reliability: Cphn 9-12

Component	Weight	Variance	Reliability
Listening	0.3	102.010	.916
Reading	0.7	115.348	.908
Comprehension		107.123	.944

*Variances from students who had results in all four domains

Table 6.4.7F
n/a

Table 6.4.7G
n/a

Table 6.4.7H
n/a

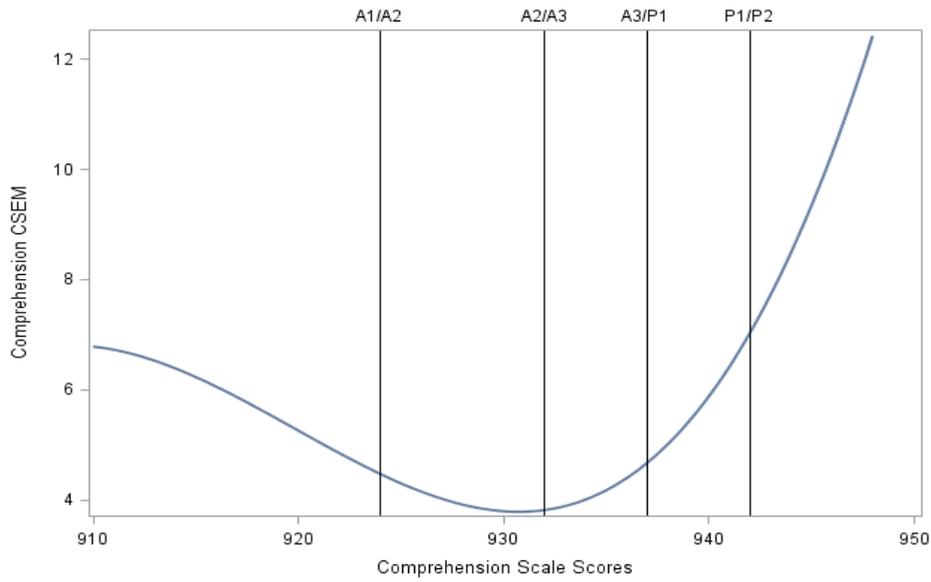
Table 6.4.7I
n/a

Table 6.4.7J

Accuracy and Consistency of Classification Indices: Cphn 9-12

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.778	0.699		0.560	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.918		0.874	
	A2	0.650		0.527	
	A3	0.661		0.538	
	P1	0.667		0.492	
	P2	0.826		0.798	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.983	0.009	0.008	0.976
	A2/A3	0.965	0.021	0.013	0.951
	A3/P1	0.945	0.023	0.032	0.925
	P1/P2	0.883	0.023	0.095	0.836

Figure 6.4.7F CSEM for Comprehension Composite 9-12



6.4.8 Overall Composite 9-12

Figure 6.4.8A

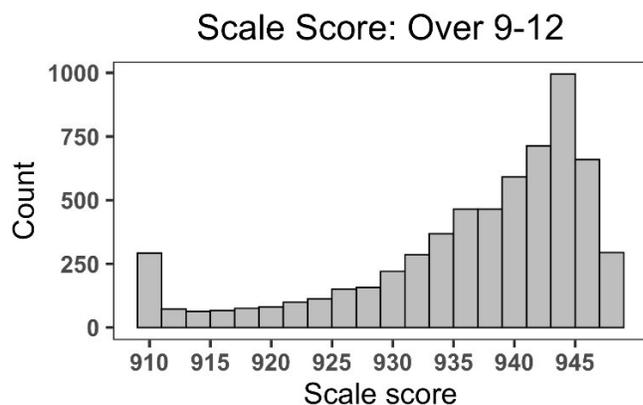


Table 6.4.8A

Scale Score Descriptive Statistics: Over 9-12

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	1,543	910	949	936.58	10.18
10	1,477	910	949	937.15	9.91
11	1,307	910	949	937.33	10.00
12	1,899	910	949	936.60	10.20
Total	6,226	910	949	936.88	10.09

Figure 6.4.8B

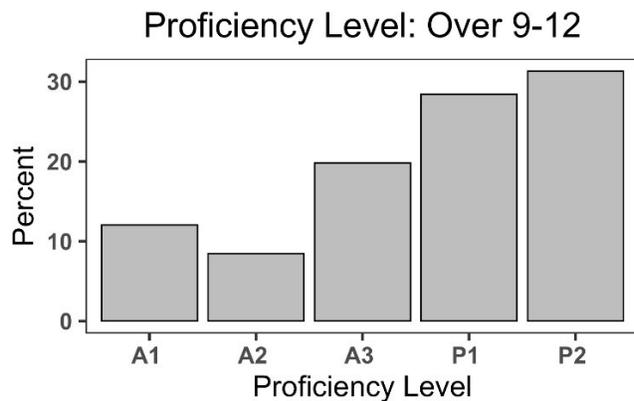


Table 6.4.8C

Proficiency Level Distribution: Over 9-12

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
A1	203	13.16	163	11.04	151	11.55	232	12.22	749	12.03
A2	141	9.14	115	7.79	97	7.42	173	9.11	526	8.45
A3	286	18.54	299	20.24	249	19.05	399	21.01	1,233	19.80
P1	439	28.45	426	28.84	370	28.31	534	28.12	1,769	28.41
P2	474	30.72	474	32.09	440	33.66	561	29.54	1,949	31.30
Total	1,543	100.00	1,477	100.00	1,307	100.00	1,899	100.00	6,226	100.00

Table 6.4.8D

n/a

Figure 6.4.8D

n/a

Figure 6.4.8E n/a

Table 6.4.8E

Reliability: Over 9-12

Component	Weight	Variance	Reliability
Listening	0.15	102.010	.916
Reading	0.35	115.348	.908
Speaking	0.15	147.137	.898
Writing	0.35	122.545	.774
Overall Composite		101.606	.949

*Variances from students who had results in all four domains

Table 6.4.8F n/a

Table 6.4.8G n/a

Table 6.4.8H n/a

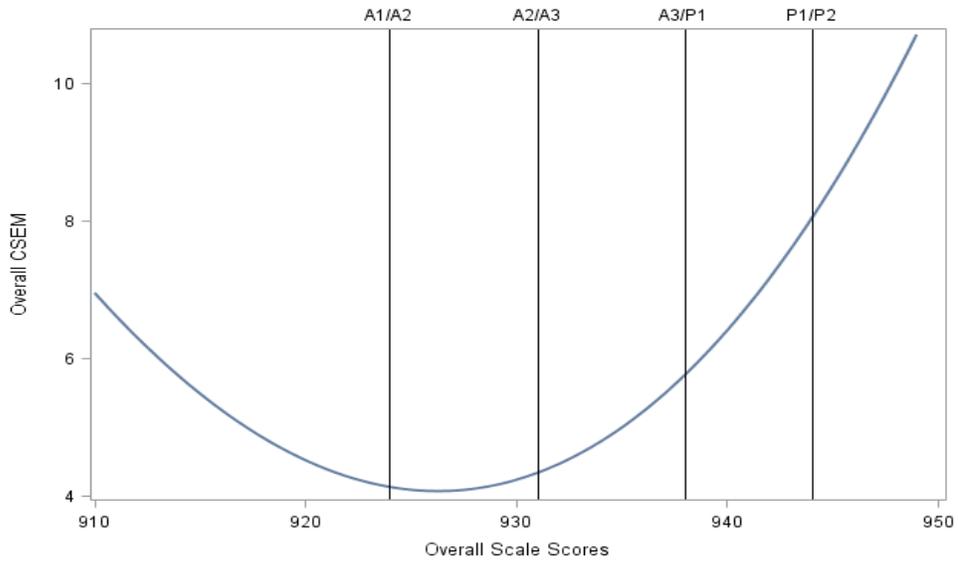
Table 6.4.8I n/a

Table 6.4.8J

Accuracy and Consistency of Classification Indices: Over 9-12

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.594	0.653		0.550	
Conditional on Level	Level	Accuracy		Consistency	
	A1	0.938		0.907	
	A2	0.725		0.619	
	A3	0.894		0.840	
	P1	0.424		0.487	
	P2	1.780		0.645	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	A1/A2	0.985	0.008	0.007	0.979
	A2/A3	0.969	0.019	0.011	0.956
	A3/P1	0.957	0.011	0.033	0.940
	P1/P2	0.683	0.305	0.012	0.778

Figure 6.4.8F CSEM for Overall Composite 9-12



References

- Allen, N.L., Carlson, J.E., & Zalanak, C.A. (1999). *The NAEP 1996 technical report*. Washington, DC: National Center for Education Statistics.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.
- Andrich D. (1978) A rating scale formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Bachman, L. F & A. B. Palmer (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. UK: Oxford University Press.
- Brennan, R.L. (2004). BB-CLASS: a computer program that uses the beta-binomial model for classification consistency and accuracy. [Computer Software]. Iowa City, IA: CASMA.
- Center for Applied Linguistics. (2012a). *Alternate ACCESS for ELLs™ Standard Setting Study: Technical Brief*. Available at www.wida.us.
- Center for Applied Linguistics (2013). *Alternate ACCESS for ELLs™ Series 100 Development and Operational Field Test: Technical Report*.
- Chapelle, C. A., Enright, M. & Jamieson, J. (Eds.) (2008). Building a validity argument for the Test of English as a Foreign Language. London: Routledge.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. (2010). Does an Argument-Based Approach to Validity Make a Difference?. *Educational Measurement: Issues and Practice*, 29(1), 3- 13.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.
- Cronbach, L.J., Schonemann, P., & McKie, D. (1965). Alpha coefficients for stratified-parallel tests. *Educational and Psychological Measurement*, 25, 291-312.
- Every Student Succeeds (ESSA) Act of 2015, Pub. L. No. 114-95, § 114 Stat. 1177 (2015-2016).
- Individuals With Disabilities Education Act, 20 U.S.C. § 1400 (2004).
- 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. (2002). Validating high-stakes testing programs. *Educational measurement: Issues and practice*, 21(1), 31-41.
- Kane, M., & Case, S.M. (2004). The reliability and validity of weighted composite scores. *Applied Measurement in Education*, 17, 221-240.

- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Kenyon, D.M. (2006). *Development and Field Test of ACCESS for ELLs®*. (WIDA Consortium Technical Report No. 1).
- 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, 878. Retrieved April 10, 2006, from www.rasch.org/rmt/rmt162f.htm
- Linacre, J.M. (2006). Winsteps (Version 3.60) [computer software]. Chicago, IL: 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.
- Llosa, L. (2008). Building and Supporting a Validity Argument for a Standards-Based Classroom Assessment of English Proficiency Based on Teacher Judgments. *Educational Measurement: Issues and Practice*, 27(3), 32-42.
- Meyer, J.P. (2014). jMetrik. [Computer Software]. <http://www.jmetrik.com>.
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2004). *A Brief Introduction to Evidence-Centered Design*. CSE Report 632. US Department of Education.
- Rudner, L. (2001, Spring). Informed test component weighting. *Educational Measurement: Issues and Practice*, 20:1, 16-19.
- WIDA Consortium. (2007). *English Language Proficiency Standards and Resource Guide, 2007 Edition, PreKindergarten through Grade 12*. Madison, Wisconsin: Board of Regents of the University of Wisconsin System.
- WIDA Consortium (2012a). *WIDA Alternate ACCESS for ELLs® Test Administration Manual*. Retrieved from <https://www.wida.us/assessment/alternateaccess.aspx>.
- WIDA Consortium (2013). *WIDA ACCESS for ELLs™ Test Administration Manual*. Retrieved from <https://www.wida.us/assessment/alternateaccess.aspx>.
- Wright, B.D., & Stone, M.H. (1979). *Best test design: Rasch measurement*. Chicago, IL: MESA Press.
- Young, M.J., & Yoon, B. (1998, April). *Estimating the consistency and accuracy of classifications*

in a standards-referenced assessment (CSE Tech. Rep. 475). Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education and Information Studies.

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. Holland & H. Wainer (Eds.), *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum.

Zwick, R., Donoghue, J.R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 30, 233-25.