World-Class Instructional Design and Assessment

WïDA™
C O N S O R T I U M

Annual Technical Report for
ACCESS for ELLs® English Language Proficiency Test,
Series 301, 2012-2013 Administration

Annual Technical Report No. 9
Volume 1 of 3: Description, Assessment Use Argument, and Student Results

Prepared by:

CAL/WIDA Partnership Activities
Psychometrics/Research Team

Center for Applied Linguistics

May 30, 2014

The WIDA ACCESS for ELLs Technical Advisory Committee

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

- Jamal Abedi, Ph.D., Professor at the Graduate School of Education at the University of California, Davis and a research partner at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
- Lyle Bachman, Ph.D., Professor Emeritus, Applied Linguistics, University of California, Los Angeles
- Akihito Kamata, Ph.D., Professor, Department of Education Policy and Leadership, Department of Psychology, Southern Methodist University.
- Timothy Kurtz, Hanover High School, Hanover, New Hampshire
- Carol Myford, Ph.D., Associate Professor, Educational Psychology at the University of Illinois at Chicago.

More information on the TAC members can be found at the WIDA website (www.wida.us/assessment/access/TAC/index.aspx).

Executive Summary

This is the ninth annual technical report on ACCESS for ELLs. This technical report is produced as a service to members and potential members of the WIDA Consortium. The technical information herein is intended for use by those who have technical knowledge of test construction and measurement procedures, as stated in *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999).

ACCESS for ELLs serves two purposes: 1) To assess reliably and validly the developing English language development (ELD) of English language learners in Grades K–12 according to WIDA *2012 Amplification of the English Language Development Standards Kindergarten–Grade 12* (2012); 2) To place students appropriately into proficiency levels described by the ELD Standards. Results on ACCESS for ELLs are used by WIDA Consortium states for monitoring the progress of students, for making decisions about exiting students from language support services, and for accountability.

This report provides detailed information from the analysis of the ninth series of the test, Series 301. Series 301 was administered during the academic year 2012–2013 in 31 WIDA Consortium states. Because the main focus of this report is on the technical quality of the test forms and not on the performance of students, analyses in this report are aggregated across all participating states.

As in the previous annual technical reports, this report provides background to the test (Chapter 1). The current report has been modified for Series 301 to introduce an argument-based validation framework to support the use of ACCESS for ELLs and to contextualize the data so that its interpretation and use are more transparent to stakeholders (Chapter 2). The rest of the report consists of paired chapters. The first chapter within each pair contains text that explains the data tables that follow in the second chapter. Information on the students who participated in the operational administration is presented (Chapters 3 and 4), followed by an explanation of the technical analyses conducted on each of the 44 test forms that constitute ACCESS for ELLs (Chapter 5) and the tables and figures of results (Chapter 6). The final chapters explain (Chapter 7) and present (Chapter 8) technical analyses based on the domain scores and composite scores by grade-level cluster. Note that Chapters 1–4 are in Volume 1, Chapters 5–6 are in Volume 2, and Chapters 7–8 are in Volume 3.

Summary Highlights

This report presents a wealth of data documenting the technical properties of the 44 test forms of ACCESS for ELLs Series 301, which is impossible to summarize here. In addition to information on validity, the report presents information on reliability of test scores and the accuracy and consistency of proficiency level classifications, including information on conditional standard errors of measurement for all scores and a separate table highlighting conditional standard errors around the cut scores. The report also provides details on scaling and the equating of the Series 301 test forms to those of Series 203. Item-level analyses include item difficulty levels, fit of the items to the Rasch measurement model, and differential item functioning (DIF) analyses for each item or assessment task. The annual analyses of the technical

properties of ACCESS for ELLs test forms are used in the continual refinement and improvement of ACCESS for ELLs.

Here we would like to highlight the following results of this report.

Argument-based validation framework for ACCESS for ELLs

Starting with Series 301, Chapter 2 of the ACCESS for ELLs Annual Technical Report will consist of an argument-based framework for supporting the validity of ACCESS for ELLs. This framework structures the information contained in this Annual Technical Report to support assertions about data collected via the assessment (i.e., *Assessment Records*). Specifically, tables and figures from this report are explicitly linked to claims related to *Assessment Records* through an Assessment Use Argument (AUA), which allows stakeholders to better interepret and use ACCESS for ELLs. A larger, forthcoming (as of 2014), validation framework for the complete assessment from its inception to its consequences is currently under development by WIDA.

Demographic data

The Series 301 data set for analyses included the results of 1,236,415 students. The largest grade was Kindergarten with 189,375 students, while the smallest was Grade 12 with 27,591 students. Of the participating WIDA states, the largest was Illinois with 179,712 students, while the smallest was Vermont with 1,519 students. Technical analyses in this report are based on the performance of all students who were administered Series 301 of ACCESS for ELLs.

Reliability and accuracy data

For most test users, the Overall Composite proficiency score, based on performances in Listening, Speaking, Reading, and Writing, is the major score used for making decisions about gains in student proficiency and exiting from language support services, and for Annual Measureable Achievement Objectives (AMAOs). As explained by Keng, Miller, O'Malley, and Turhan (2008), "the use of composite scores has become more widespread with federal testing requirements under Title III of No Child Left Behind now calling for states to assess students with limited English proficiency (LEP) annually from Kindergarten through 12th grade in the four language domains of listening, speaking, reading and writing. A composite of the student's performance on each of these domains is calculated to represent the student's overall English language proficiency." Results indicate that the reliability of the Overall Composite score for Series 301, presented in Chapter 8 Table D, is very high across all grade-level clusters. For Kindergarten it was .973; for Grades 1–2, .943; for Grades 3–5, .931; for Grades 6–8, .925; and for Grades 9–12, .943. Likewise, as Table 0.1 shows, the accuracy of decisions about student placement using the Overall Composite score around the proficiency level cut scores is very high across the grade and proficiency levels. Because many WIDA Consortium states use the proficiency level score of 5.0 as a criterion for exiting students from language support services, the column headed 4/5 Cut (the proficiency level score of 5.0) is of particular interest.

Table 0.1

Accuracy of Overall Score at Cut Points (Proficiency Level Score)

Grade	1/2 Cut (2.0)	2/3 Cut (3.0)	3/4 Cut (4.0)	4/5 Cut (5.0)	5/6 Cut (6.0)
K (instructional)	0.977	0.960	0.950	0.952	0.949
K (accountability)	0.953	0.951	0.953	0.951	0.992
1	0.981	0.936	0.923	0.966	0.991
2	0.990	0.961	0.915	0.942	0.978
3	0.997	0.984	0.939	0.905	0.923
4	0.995	0.981	0.934	0.894	0.914
5	0.993	0.978	0.928	0.884	0.932
6	0.986	0.959	0.904	1.020	0.985
7	0.984	0.955	0.896	0.914	0.985
8	0.983	0.949	0.898	0.895	0.984
9	0.982	0.961	0.929	0.902	0.902
10	0.985	0.955	0.922	0.911	0.941
11	0.985	0.955	0.920	0.892	0.943
12	0.984	0.953	0.908	0.809	0.954

Overview of the Annual Technical Report

The multistate WIDA Consortium's ACCESS for ELLs was first operationally administered in 2005 in three states: Alabama, Maine, and Vermont. Results of that administration were reported in Annual Technical Report 1 (Series 100, 2004–2005).

Each year, the Center for Applied Linguistics refreshes a third to a half of all ACCESS for ELLs test items. In academic year 2012–2013, Series 301 was administered in 31 WIDA Consortium states. The 44 test forms in Listening, Reading, Writing, and Speaking across the grades from Kindergarten to 12 constitute a test series. These test forms are grouped into five grade-level clusters: Kindergarten, Grades 1–2, Grades 3–5, Grades 6–8, and Grades 9–12. Within each grade-level cluster except Kindergarten, there are three overlapping tiers of test forms for Listening, Reading, and Writing: A, B, and C. This report presents the results of research into the technical properties of the 44 test forms (e.g., Grades 3–5, Reading, Tier C) that constitute Series 301. Data come from the 1,236,415 students who were administered the test operationally in the winter and spring of 2013.

Because of the size of the complete report, it is presented in three volumes.

Volume I contains an executive summary, this overview, an annotated bibliography, and Chapters 1 to 4. Chapter 1 provides background to the test. Readers unfamiliar with ACCESS for ELLs should pay particular attention to this chapter. Chapter 2 presents an argument-based approach for structuring the data contained in this report so that its interpretation and use are more transparent to stakeholders. Chapters 3 and 4 present information on the students who participated in the Series 301 (2012–2013) operational administration, including overall results.

Volume II contains Chapters 5 and 6. Chapter 5 presents background on the technical analyses conducted on each of the test forms and explains how to understand the tables and figures of results. Chapter 6 presents the results organized by

- Grade-level cluster (K, 1–2, 3–5, 6–8, 9–12)

- Domain (Listening, Reading, Writing, and Speaking, abbreviated List, Read, Writ, and Spek, respectively)

- Tier (A, B, C)

Thus, all of the results for Kindergarten are presented before the results for Grades 1–2, and all of the results for Listening (i.e., for tiers A, B, and C where applicable) are presented before results for Reading.

Volume III contains Chapters 7 and 8. These chapters focus on results across tiers within grade-level clusters, including the four composite scores (Oral Language, Literacy, Comprehension, and Overall). Chapter 7 presents background on the technical analyses and explains how to understand the tables and figures of results. Chapter 8 presents the results organized by

- Grade-level cluster (K, 1–2, 3–5, 6–8, 9–12)

- Score (Listening, Reading, Writing, Speaking, Oral Language Composite, Literacy Composite, Comprehension Composite, and Overall Composite, abbreviated List, Read, Writ, Spek, Oral, Litr, Cphn, and Over, respectively)

Annotated Bibliography: 2012-2013

Technical Reports

This is a list of reports that describe the development of ACCESS for ELLs.

Gottlieb, M., & Boals, T. (2005). Considerations in Reconfiguring Cohorts and Resetting Annual Measurable Achievement Objectives (AMAOs) based on ACCESS for ELLs Data (WIDA Consortium Technical Report No. 3).

> This report is intended to assist states with the transition to a standards-based test and determining their AMAOs using ACCESS for ELLs.

Gottlieb, M. & Kenyon, D.M. (2006). *The Bridge Study between Tests of English Language Proficiency and ACCESS for ELLs* (WIDA Consortium Technical Report No. 2).

> This report provides the background, procedures, and results of a study intended to establish estimates of comparability between ACCESS for ELLs and four other English language tests used by Consortium member states. Students in Illinois and Rhode Island were administered ACCESS for ELLs along with one of the other four tests, and results on the four tests were compared with results on ACCESS for ELLs. Results allow states, districts, and schools to understand and report ACCESS for ELLs scores and to establish continuity between previous tests and ACCESS for ELLs.

Kenyon, D. M. (2006). *Development and Field Test of ACCESS for ELLs* (WIDA Consortium Technical Report No. 1).

> This report provides detailed information on the conceptualization, development, and field testing of ACCESS for ELLs. It also provides technical data on equating and scaling procedures, standard setting and operational score reporting, analyses of reliability and errors of measurement, and two initial validity studies.

Kenyon, D. M., Ryu, J.R. (Willow), & MacGregor, D. (2013). *Setting Grade Level Cut Scores for ACCESS for ELLs.* (WIDA Consortium Technical Report No. 4).

> This report describes the technical procedures and outcomes of the process to move from grade-level-cluster cut scores to grade-level cut scores. Proposed cut scores were determined mathematically and then reviewed and revised in a standard setting process involving 75 teachers from 14 WIDA Consortium states.

MacGregor, D., Kenyon, D. M., Gibson, S., & Evans, E. (2009). *Development and Field Test of Kindergarten ACCESS for ELLs*. (WIDA Consortium).

> This report provides detailed information on the conceptualization, development, and field testing of Kindergarten ACCESS for ELLs. It also provides technical data on equating and scaling procedures, standard setting and operational score reporting, and analyses of reliability and errors of measurement.

Annual Technical Reports for ACCESS for ELLs

Below is a list of annual technical reports for ACCESS for ACCESS for ELLs, listed by year of publication. These reports provide extensive analysis of the results from the operational administration of ACCESS for ELLs. They provide detailed information on student results broken down by grade-level cluster, grade, and tier. They also provide detailed information on test and item characteristics.

Kenyon, D. M., MacGregor, D., Ryu, J.R. (Willow), Cho, B., and Louguit, M. (2006). *Annual Technical Report for ACCESS for ELLs[®] English Language Proficiency Test, Series 100, 2004-2005 Administration.* (WIDA Consortium Annual Technical Report No. 1).

Kenyon, D. M., MacGregor, D., Louguit, M. Cho, B., and Ryu, J.R. (Willow). (2007). *Annual Technical Report for ACCESS for ELLs[®] English Language Proficiency Test, Series 101, 2005-2006 Administration.* (WIDA Consortium Annual Technical Report No. 2).

MacGregor, D., Louguit, M., Ryu, J.R. (Willow), Kenyon, D.M., and Li, D. (2008). *Annual Technical Report for ACCESS for ELLs[®] English Language Proficiency Test, Series 102, 2006-2007 Administration* (WIDA Consortium Annual Technical Report No. 3).

MacGregor, D., Louguit, M., Huang, X., and Kenyon, D.M. (2009). Annual Technical Report for ACCESS for ELLs[®] English Language Proficiency Test, Series 103, 2007-2008 Administration (WIDA Consortium Annual Technical Report No. 4).

MacGregor, D., Louguit, M., Yanosky, T., Fidelman, C. G., Pan, M., Huang, X., and Kenyon, D.M. (2010). *Annual Technical Report for ACCESS for ELLs[®] English Language Proficiency Test, Series 200, 2008-2009 Administration* (WIDA Consortium Annual Technical Report No. 5).

Yanosky, T., Yen, S., Louguit, M., MacGregor, D., Zhang, Y., and Kenyon, D.M. (2011). *Annual Technical Report for ACCESS for ELLs[®] English Language Proficiency Test, Series 201, 2009-2010 Administration* (WIDA Consortium Annual Technical Report No. 6).

Yanosky, T., Chong, A., Louguit, M., Olson, E., Choi, Y., MacGregor, D., Yen, S., Cameron, C., and Kenyon, D.M. (2012). *Annual Technical Report for ACCESS for ELLs[®] English Language Proficiency Test, Series 202, 2010-2011 Administration* (WIDA Consortium Annual Technical Report No. 7).

Yanosky, T., Amos, M., Louguit, M., Olson, Cameron, C., Louguit, M., MacGregor, D., Yen, S., and Kenyon, D.M. (2013). *Annual Technical Report for ACCESS for ELLs[®] English Language Proficiency Test, Series 203, 2011-2012 Administration* (WIDA Consortium Annual Technical Report No. 8).

Other Documentation

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly, 2(1),* 1–34.

> This article describes how an argument for test use might be structured so as to provide a clear linkage from test performance to interpretations and from interpretations to uses.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice.* Oxford: Oxford University Press.

> This book presents the Assessment Use Argument, which provides a framework for justifying the intended uses of an assessment, as well as a guide for the design and development of the assessment itself.

Bauman, J., Boals, T., Cranley, E., Gottlieb, M., and Kenyon, D.M. (2007). The Newly Developed English Language Tests (World-Class Instructional Design and Assessment – WIDA). In Abedi, Jamal (Ed.), *English Language Proficiency Assessment in the Nation: Current Status and Future Practice.* Davis: University of California.

> In this book chapter, the authors describe the test development process, from the development of standards through the development of items, field testing, and operationalization. They also report on validation of the test, accommodations, the test administration and technical manuals, and score reporting.

Chapelle, C. A., Enright, M. E., & Jamieson, J. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice, 29(1),* 3–13.

> Drawing on experience between 2000 and 2007 in developing a validity argument for the high-stakes Test of English as a Foreign Language™, this paper evaluates the differences between the argument-based approach to validity as presented by Kane (2006) and that described in the 1999 AERA/APA/NCME Standards for Educational and Psychological Testing.

Chapelle, C. A., Enright, M. & Jamieson, J. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language.* London: Routledge.

> This book uses the Test of English as a Foreign Language™ as a case study for validating test design. It attempts to meet the standards of educational measurement while also drawing on theory related to English language proficiency.

Cook, H. G. (2007). Alignment Study Report: The WIDA Consortium's English Language Proficiency Standards for English Language Learners in Kindergarten through Grade 12 to ACCESS for ELLs® Assessment. Madison, WI: WIDA Consortium.

> In this report, the author describes a study to align the WIDA Standards to the ACCESS for ELLs test. The study was designed to address two questions: how well the test measures the proficiency levels described in the Standards, and how well the different domains of each standard are addressed by the domains of the test. The author concludes that overall ACCESS for ELLs is adequately aligned to the Standards.

Cook, H. G., Boals, T., Wilmes, C., and Santos, M. (2007). Issues in the Development of Annual Measurable Achievement Objectives (AMAOs) for WIDA Consortium States. Madison, WI: WIDA Consortium.

> In this paper, the authors offer guidance to states in formulating Annual Measurable Achievement Objectives for English language learners.

Fox, J. (2011). Test review: ACCESS for ELLs[®]. *Language Testing 28* (3): 425-431.

> The author provides a thorough review of ACCESS for ELLs, using the eight criteria enumerated in Fairbairn and Fox (2009).

Gottlieb, M. (2004). English Language Proficiency Standards for English Language Learners in Kindergarten through Grade 12: Framework for Large-Scale State and Classroom Assessment. Madison, WI: WIDA Consortium.

> These documents contain the WIDA Standards and describe the rationale behind and development of the frameworks for large-scale state and classroom assessments. These frameworks comprise English Language Development standards, language domains, grade-level clusters, language proficiency levels and the model performance indicators upon which ACCESS for ELLs is based. They are meant to guide curriculum development, instruction, and assessment of English language learners.

Kane, M. (2006). Validation. In R. Brennan, (Ed.), *Educational Measurement* (*4th Edition*), pp. 18-64. Westport, CT: Greenwood Publishing.

> This book chapter presents a conceptualization of test validity where evidence and logical argument are brought together to evaluate claims and propositions about the proposed uses and interpretations of test results.

Kenyon, D.M., MacGregor, D., Li, D., and Cook, H. G. (2011). Issues in vertical scaling of a K-12 English language proficiency test. *Language Testing 28* (3): 383-400.

> In this article, the authors describe the procedure used to place ACCESS for ELLs results on a vertical scale, and they discuss studies conducted to test the effectiveness of that scale.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2004). A Brief Introduction to Evidence-Centered Design (CSE Report 632). CA: Center for Research on Evaluation, Standards, and Student Testing.

> This paper provides an introduction to the basic ideas of Evidence Centered Design, an approach to constructing educational assessments in terms of evidentiary arguments. It includes some of the terminology and models that have been developed to implement the approach.

National Research Council. (2011). Allocating federal funds for state programs for English language learners. Panel to Review Alternative Data Sources for the Limited-English Proficiency Allocation Formula under Title III, Part A, Elementary and Secondary Education Act, Committee on National Statistics and Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

This report includes detailed descriptions of six English language proficiency tests, including ACCESS for ELLs, along with information about the reliability and validity of the tests.

Parker, C. E., Louie, J., and O'Dwyer, L. (2009). New measures of English language proficiency and their relationship to performance on large-scale content assessments (Issues & Answers Report, REL 2009–No. 066). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands. Retrieved from http://ies.ed.gov/ncee/edlabs, January 29, 2009.

This report describes a study investigating how well the domain tests on ACCESS for ELLs predict performance on a content test. Results indicate that the Reading and Writing tests are the strongest predictors.

Römhild, A., Kenyon, D. M., and MacGregor, D. (2011). Exploring domain-general and domain-specific linguistic knowledge in the assessment of academic English language proficiency. *Language Assessment Quarterly*, 8:213-228.

This article reports on a confirmatory factor analysis study conducted to model domain-specific and domain-general variance on ACCESS for ELLs. The authors found that, while domain-general linguistic knowledge represents the primary dimension across almost all test forms, domain-specific knowledge becomes increasingly salient as proficiency level increases.

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.

This document presents the second edition of the WIDA English Language Development Standards, which were released in 2007. The second edition included the addition of formative and summative frameworks for assessment and instruction, the separation of Kindergarten into its own grade-level cluster, and the addition of the sixth proficiency level, "Reaching".

WIDA Consortium. (2012a). *2012 Amplification of the English Language Development Standards Kindergarten–Grade 12*. Madison, Wisconsin: Board of Regents of the University of Wisconsin System.

This document describes the amplified Strands of Model Performance Indicators that represent the WIDA English Language Development Standards. The amplification reflects states' content standards and the fluid and ongoing process of language development.

WIDA Consortium (2012b). *WIDA ACCESS for ELLs® Test Administration Manual*. Retrieved from www.wida.us/assessment/ACCESS/#about.

This document details the test administration procedures for ACCESS for ELLs.

WIDA Consortium. (2013). *Interpretive Guide for Score Reports Spring 2013* (WIDA Consortium). Madison, WI: The Board of Regents of the University of Wisconsin System.

This report provides an overview on how ACCESS for ELLs is scored and how those scores are reported. Part 1 gives a description of scores for 2013. Part 2 gives suggestions on how states can use scores, as well as examples of score reports to various stakeholders. Part 3 provides guidance on interpreting the reports.

Wolf, M., Kao, J., Griffin, N., Herman, J., Bachman, P., Chang, S., and Farnsworth, T. (2008). Issues in assessing English language learners: English language proficiency measures and accommodation uses—Practice review. Retrieved from the University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing Web site: http://www.cse.ucla.edu/products/rsearch.asp.

This paper describes the English language proficiency tests in use in school year 2006, including ACCESS for ELLs, and provides a summary of validity evidence for the tests.

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

This book chapter describes prodedures for conducting Differential Item Functioning analysis.

Table of Contents

1. Description of ACCESS for ELLs English Language Proficiency Test

1.1   Purpose of ACCESS for ELLs

The overarching purpose of Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS for ELLs) is to assess the developing English language proficiency of English language learners in Grades K–12 in the United States following the English Language Development Standards (2012) of the multi-state WIDA Consortium. The WIDA English Language Proficiency (ELP) Standards (2004, 2007) were amplified in 2012 to become English Language Development (ELD) Standards, hereafter referred to as ELD Standards. The WIDA ELD Standards, aligned to state academic content standards, form the core of the WIDA Consortium's approach to instructing and testing English language learners and describe six levels of developing English language proficiency. ACCESS for ELLs may thus be described as a standards-based English language proficiency test designed to measure English language learners' social and academic language proficiency in English. It assesses social and instructional English as well as the language associated with language arts, mathematics, science, and social studies within the school context across the four language domains (Listening, Reading, Writing, and Speaking).

Other major purposes of ACCESS for ELLs include

Identifying the English language proficiency level of students with respect to the WIDA ELD Standards used in all member states of the WIDA Consortium,

Identifying students who have attained English language proficiency,

Assessing annual English language proficiency gains using a standards-based assessment instrument,

Providing districts with information that will help them to evaluate the effectiveness of their ESL/bilingual programs and determine staffing requirements,

Providing data for meeting federal and state statutory requirements with respect to student assessment, and

Providing information that enhances instruction and learning in programs for English language learners.

1.2   Format of ACCESS for ELLs

1.2.1 Integration with the Standards

The design of ACCESS for ELLs, from the structure of the assessment system to the content of each test booklet and item, is built upon the five foundational WIDA ELD Standards:

Standard 1 - English language learners communicate in English for **Social and Instructional** purposes within the school setting.

Standard 2 - English language learners communicate information, ideas, and concepts necessary for academic success in the content area of **Language Arts.**

Standard 3 - English language learners communicate information, ideas, and concepts necessary for academic success in the content area of **Mathematics.**

Standard 4 - English language learners communicate information, ideas, and concepts necessary for academic success in the content area of **Science.**

Standard 5 - English language learners communicate information, ideas, and concepts necessary for academic success in the content area of **Social Studies.**

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

Social and Instructional language: SI

Language of English Language Arts: LA

Language of Math: MA

Language of Science: SC

Language of Social Studies: SS

Every selected response item and every performance-based task on ACCESS for ELLs targets at least one of these five Standards.

1.2.2 Grade-level Clusters

The WIDA ELD Standards describe developing English language proficiency by five grade-level clusters. These are PreK-K, 1–2, 3–5, 6–8, and 9–12. Test booklets follow this grade-level clustering.

1.2.3 Language Domains

The WIDA ELD Standards describe developing English language proficiency for each of the four language domains: Listening, Speaking, Reading, and Writing. Thus, there is a section of the ACCESS for ELLs test assessing each of these four language domains.

1.2.4 Language Proficiency Levels

The WIDA ELD Standards describe the continuum of language development with five language proficiency levels that are fully delineated in the Standards document. These levels are "Entering," "Emerging," "Developing," "Expanding," and "Bridging." There is also a final exit stage known as Level 6, "Reaching," that describes students who have progressed across the entire WIDA English language proficiency continuum. These levels are shown graphically in Figure 1.2.4A.
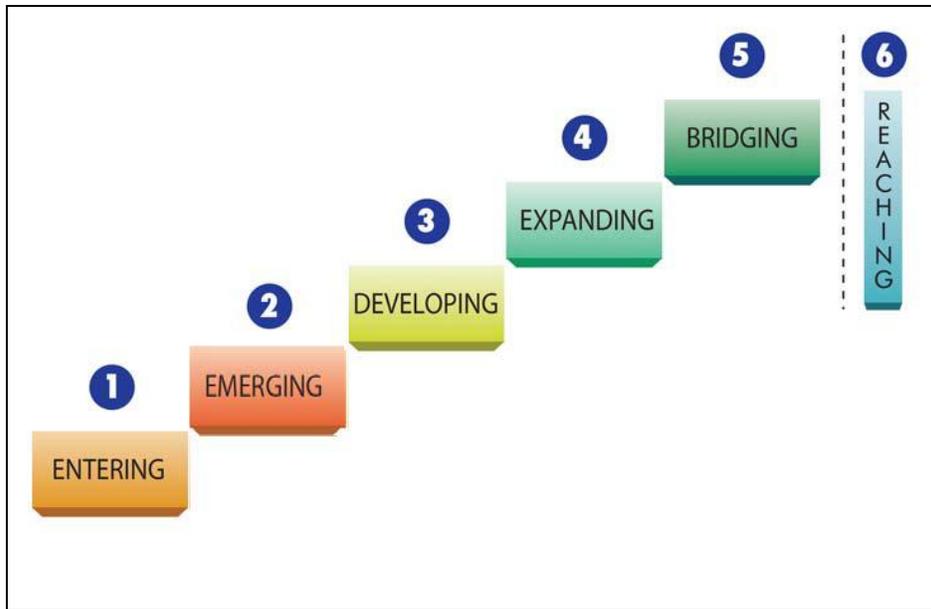
Figure 1.2.4A. The language proficiency levels of the *WIDA ELD Standards*

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 model **performance indicators** (PIs, see below) for each language proficiency level. Being general definitions applicable across the PIs, the performance definitions are not explicitly replicated within the PIs. 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 PIs, which exemplify the Standards. The PIs describe the expectations for ELL students for each of the five **Standards,** at five different **grade-level clusters**, across four **language domains**, and at the five **language proficiency levels**. That is, within each combination of standard, grade-level cluster, and language domain is a PI at each of the five language proficiency levels. Proficiency Level 6, Reaching, represents the end of the continuum rather than another level of language proficiency. The sequence of these five PIs together describes a logical progression and accumulation of skills on the path from the lowest level of English language proficiency to full English language proficiency for academic success. These groupings of five PIs in logical progression are called a "strand."

ACCESS for ELLs is based on the 80 strands, containing 400 individual PIs, within the WIDA ELD Standards. (The Standards and the accompanying model PIs are available at the WIDA web

Return to Summary of Assessment Records Claims, Actions, and Evidence

site, (www.wida.us.) Each selected-response item or performance-based task on ACCESS for ELLs is carefully developed, reviewed, piloted, and field tested to ensure that it allows students to demonstrate accomplishment of the targeted PI. (See the sample items at the WIDA web site for examples.)

1.2.5 Tiers

Obviously, test items and tasks suitable for allowing Entering (Level 1) or Emerging (Level 2) students to demonstrate accomplishment of the PIs at their level of language proficiency (i.e., that allow them to demonstrate what they can do) will not allow Expanding (Level 4) or Bridging (Level 5) students to demonstrate the full extent of their language proficiency. Likewise, items and tasks developed to allow Expanding (Level 4) and Bridging (Level 5) students to demonstrate accomplishment of the PIs at their level would be far too challenging for Entering (Level 1) or Emerging (Level 2) students. Items that are far too easy for test takers may be boring and lead to inattentiveness on the part of students. Likewise, items that are far too difficult for test takers may be frustrating, discouraging them from giving their best performance. But more importantly, a test is a measure, and items that are too easy or too hard for a student add very little to the accuracy or quality of the measurement of that student's proficiency. Tests need to be at the right difficulty level for individual test takers.

The solution to making ACCESS for ELLs appropriate to the proficiency level of individual students across the wide range of proficiencies described in the WIDA ELD Standards is to present the test items in three overlapping **tiers** for each grade-level cluster: A, B, and C. Figure 1.2.5A shows how the different tiers map to the language proficiency levels.
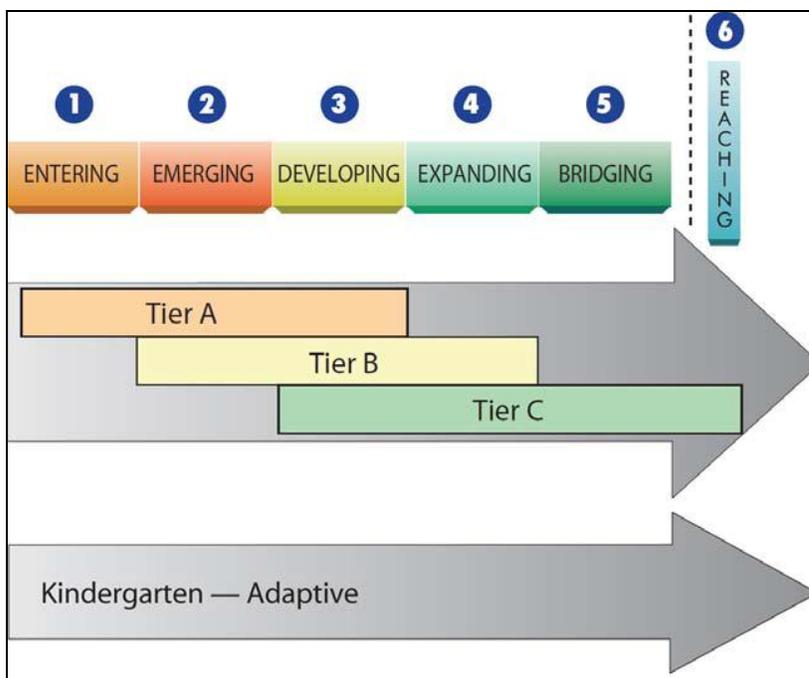
Figure 1.2.5A. Tier structure of ACCESS for ELLs

Thus, Tier A has items and tasks designed to allow students at the lowest language proficiency levels (Levels 1 and 2) to demonstrate meeting the WIDA ELD Standards at their language proficiency levels, and it includes some items targeted to Language Proficiency Level 3. Likewise, Tier C has items and tasks designed to allow students at the highest language proficiency levels (Levels 4 and 5) to demonstrate meeting the WIDA ELD Standards at their language proficiency levels, while also containing some items targeted to Language Proficiency Level 3. In this test design, the tiers overlap: while Tier A and Tier C would share little in common, Tier B is composed of tasks from both Tiers A (Level 2) and C (Level 4), as well as tasks from Level 3. This overlap of tiers ensures that all of the PIs from the WIDA ELD Standards appear on the assessment; however, each test booklet does not need to contain an unduly large number of test items. The overlap also ensures that the entire language proficiency range is covered. Finally, the overlap ensures that the assessment is *horizontally equated*; that is, common items and tasks across tiers ensure that each tier is measuring to a common language proficiency scale. Thus, a test booklet at any given tier is primarily composed of items and tasks that span three targeted language proficiency levels. (Note that in order to assure that students are accurately measured to Level 6, Tier C also includes some items that are slightly more difficult than Language Proficiency Level 5. The Tier structure only applies to the Grade 1-12 Listening, Reading, and Writing portions of ACCESS for ELLs. Kindergarten (PreK-K) does not have tiers because it is an adaptive assessment.

The individually administered Speaking portion of the assessment for each grade-level cluster is designed as an adaptive measure. In each of its three parts, the test administrator begins with a task that allows students to demonstrate meeting the performance level expectations of the PIs at Level 1 and then presents continually more challenging tasks (tasks at Level 2, then Level 3, and so on). Within each part, the administrator stops presenting additional tasks when the student can no longer demonstrate meeting the expectations of the tasks. Table 1.2.5A summarizes the main points in the above discussion and illustrates the number of unique components in ACCESS for ELLs.

Table 1.2.5A
Unique Components in ACCESS for ELLs

Grade-level Clusters	List/Read/Write			Speak (adaptive)
	Tier A	Tier B	Tier C	
9-12	x	x	x	x
6-8	x	x	x	x
3-5	x	x	x	x
1-2	x	x	x	x
K		x (adaptive)		

1.3   Test Development

1.3.1 Field Test

In 2004 the field test of ACCESS for ELLs was conducted. The purpose of the field test was to collect extensive data on items and forms in order to equate forms both horizontally (i.e., across tiers within the same grade-level clusters) and vertically (i.e., across grade-level clusters), as well

as to judge the strength of individual items. The item pool for the field test consisted of 376 Listening items, 355 Reading items, and 51 Writing tasks. Two forms were used for each tier in each grade-level cluster. For equating purposes, common items were used across tiers, as well as across forms, within grade-level clusters for the Listening, Reading, and Writing tests. In addition, common items were used across grade-level clusters for the Listening and Reading tests.

Table 1.3.1A shows the number of students who participated in the field test by grade-level cluster. 72.3% of the sample came from two states, Illinois and Wisconsin. Over half of the students (61.8%) had Spanish as their native language. The only other sizable language group was Hmong (13.8%). Indeed, of the 96 languages represented, only four languages (Spanish, Hmong, English, and French) had more than 100 students in the field test sample.

Table 1.3.1A
Field Test for Listening, Reading and Writing: Students per Grade-level Cluster

Grade-level Cluster	Students
1-2	1,647
3-5	1,850
6-8	1,449
9-12	1,716
Total	6,662

A separate, individually administered field test was conducted for Speaking. One form was developed for each grade-level cluster, using the adaptive design described in 1.2.5, for a total of 52 tasks. Field testing for Speaking was conducted in Wisconsin and the District of Columbia. Table 1.3.1.B shows the number of students who participated in the Speaking field test by grade-level cluster.

Table 1.3.1B
Speaking Field Test: Students per Grade-level Cluster

Grade-level Cluster	Students
1-2	103
3-5	159
6-8	136
9-12	125
Total	523

In addition, a separate field test was conducted in DC for the Kindergarten test. The final version of the adaptive Kindergarten assessment was produced by first choosing the Listening and Reading folders (i.e., sets of thematically related items) that contained items that were empirically the easiest for first graders based on the data collected from the field test. These folders were placed in the Kindergarten assessment in order from easiest to hardest. The Speaking portion of the Kindergarten assessment was the same as that for the 1–2 grade-level cluster, except it included only the SI and LA/SS folders, in order to reduce testing time. Special, very simple writing tasks were adapted from the 1–2 grade-level cluster Tier A SI writing folder.

The adaptive administration of the Kindergarten assessment is similar to that of the Speaking test. Thus, in any domain, if a student does not get at least two items in any folder (part) correct, the

administrator stops testing in that domain and moves on to the next domain. (The exception is Speaking, which operates exactly as the standard ACCESS Speaking assessment.)

A total of 154 students participated in the Kindergarten field test. Of those, 55% were boys (84 students) and 45% were girls (70 students). 90.2% (139) of the students were Spanish speakers; the only other language with more than one student was Vietnamese (3).

1.3.2 Equating and Scaling

If test results are to be meaningful, they need to be reported on a standard scale that is familiar to test users and that keeps the same meaning whenever it is used. Scaling is the process of developing such a scale. Equating, in the present context, is the process of putting all of the tests onto the same scale, such that results mean the same regardless of which test items the test taker takes.

Of particular challenge for ACCESS for ELLs and similar tests is the need to have a vertically equated scale (i.e., one that can measure progress across the grade levels from K to 12), in addition to the horizontal equating needed across the three tiers of ACCESS for ELLs within each grade-level cluster.

For ACCESS for ELLs, a three-digit scale score was chosen for reporting purposes. The reporting scale would have an interpretive center point across domains and composites. The centering value was chosen as 350, which would represent the cut score between language proficiency Levels 3 and 4 for the 3–5 grade-level cluster. As an additional defining characteristic, the scale would have a lower bound of 100 (i.e., 250 points lower than the center of 350) and an upper bound of 600 (i.e., 250 points higher than 350). In other words, conceptually, students from grades K–2 with the lowest language proficiency in any domain could go no lower than a scale score of 100, thus making 100 a lower bound. Conceptually, students from the 9–12 grade-level cluster with the highest language proficiency in any domain could go no higher than 600, thus making 600 a higher bound. Observed scores on all tests should fall between these extremes.

It should be kept in mind that a scale score is an interpretation of a latent ability measure and not a record of "points" earned on the test. In other words, 100 does not necessarily represent a score of 0 at all grade-level clusters, nor does 600 represent a perfect score. In fact, due to the technical nature of a vertical scale—and one of the criticisms of it—as one goes up the scale from grade level to grade level (or grade-level cluster to grade-level cluster in the case of ACCESS for ELLs), the scales adjust for the developmental growth. Thus, even if a student consistently gets a score of 0 while moving from grade-level cluster to grade-level cluster, the student's scale score on a vertical scale would show an increase, even if very slight.

Thus, to interpret appropriately what the scale scores mean, a standard-setting study was conducted (see Section 1.3.3). However, in this section, we focus on the creation of the ACCESS for ELLs scale score.

The procedure for developing the scale was complex but involved a number of basic steps. These steps were carried out separately for each of the four domains 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. They are explained more fully in ACCESS for ELLs Technical Report 1, *Development and Field Test of ACCESS for ELLs*, as well as in Kenyon et al. (2011).

Equating Design: As previously described in Section 1.3.1, within each grade-level cluster, the Listening, Reading, and Writing test booklets were presented in three tiers (A, B, and C) and two series (100 and 999), such that within each grade-level cluster, a different sample of test takers took different test booklets. However, the booklets had common folders of items (in the case of Listening and Reading) or common tasks on one or more test booklets for horizontal equating. In addition, there were common folders that went across grade-level clusters for vertical equating. Because of the adaptive design of the Speaking test, described in Section 1.3.1, there was only one form per grade and thus no need for horizontal equating.

For both Writing and Speaking, there were no common items in the equating design that linked the test booklets across different grade-level clusters. This was done intentionally as each task on these performance-based assessments was more complex, involved, and time-consuming. In addition, because these tasks targeted the WIDA Standards so closely, it would have been developmentally inappropriate to ask students to perform on tasks outside of their grade-level cluster. Thus, student performances on the Reading items were used as a scaling test for the Writing tasks, and performances on the Listening items were used as a scaling test for the Speaking tasks.

Creating the Data Matrix: The tests were scored, and the matrix of responses—every student's response to every Listening or Reading item or Writing or Speaking task—was the raw input into the scaling procedure.

Developing the Logit Scale: A calibration of the ability of the students and items using Rasch procedures was then applied to these data matrices, putting the difficulty of the items or tasks and the ability of the students onto one common interval linear scale. As described in ACCESS for ELLs Technical Report 1, *Development and Field Test of the ACCESS for ELLs®*, the steps of the common rating scale used to score the Writing items were also calibrated. 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 logit scores—in this case, the 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:

- Oral Language Composite (50% Listening + 50% Speaking)
- Literacy Composite (50% Reading + 50% Writing)
- Comprehension Composite (30% Listening + 70% Reading)
- Overall Composite (15% Listening + 15% Speaking + 35% Reading + 35% Writing).

The Comprehension Composite score (based on performances in Listening and Reading) and the Overall Composite score (based on performances in all four domains) were created with Series 100. Beginning with Series 101, the Oral Language Composite score (based on performances in

Listening and Speaking) and Literacy Composite score (based on performances in Reading and Writing) were added.

1.3.3 Standard Setting

In order to interpret appropriately what the scale scores mean, a standard-setting study was conducted. The standard-setting study was held in Madison, WI between April 20 and 27, 2005. The purpose of the study was not to set new standards on WIDA ACCESS for ELLs per se. Rather, the purpose was to use the WIDA ELD [1]Standards together with empirical information from the field test data to conduct a defensible and replicable approach to determine the relationship between student performances on the four domains of the ACCESS for ELLs and the language proficiency levels defined by the WIDA ELD Standards. The following is a brief summary of the Standard Setting study. For a fuller description, see ACCESS for ELLs Technical Report 1, *Development and Field Test of ACCESS for ELLs*®.

Four panels were convened, one for each major grade-level cluster: 1–2, 3–5, 6–8, and 9–12. On each panel were 20–22 teachers or administrators who were deemed qualified to participate in the study by the WIDA office, then located at the Wisconsin Department of Public Instruction. For Listening and Reading, a bookmarking procedure was used. Panelists were given books with all items within their grade-level cluster arranged by empirical difficulty, from least difficult to most difficult. After discussing the model performance indicators and the performance level descriptions from the WIDA ELD Standards, panelists were asked to read through the items and place a bookmark at the item that they determined a student at Language Proficiency Level 1 would have a 50% chance of answering a question correctly. They were then asked to repeat this procedure for all levels up to Level 5. During this procedure, panelists were encouraged to work independently.

After the initial round of bookmarking, the results were compiled and discussed with the panelists as a group. The panelists then were given the opportunity to reconsider and adjust their bookmarking, if they so chose. These results were compiled and presented to the WIDA management team, who used this data to help determine the final cut scores.

For Writing and Speaking, a modified body of work method was used. For Writing, the panelists were presented a book of portfolios from their grade-level cluster. Each portfolio consisted of all of the writings from the test of one student. The portfolios were chosen from students from each tier, and an attempt was made to choose students whose performances did not vary widely from one task to another. Within each grade-level cluster, portfolios were presented in ascending order; that is, the first portfolio represented a student's work that had received the lowest total raw score across the four pieces of writing, and the last portfolio presented was that of a student with a very high total raw score on the four pieces of writing.

After discussing the model performance indicators and the performance level descriptions as a group, the panelists were asked to read the portfolios and, working independently, make a judgment as to the probability that the work represents the work of a student at a given WIDA language proficiency level. For example, if they felt the portfolio represented the work of a

---

[1] Note: The 2005 ACCESS for ELLs field test and standard setting were based on the 2004 WIDA ELP standards. The WIDA English Language Proficiency (ELP) Standards (2004, 2007) were amplified in 2012 to become English Language Development (ELD) Standards (WIDA, 2012). In this section, the standards are referred to as ELD standards for consistency.

student at Language Proficiency Level 3, they would write 100% under the column "3" on their paper. If they felt that it was a borderline performance between Levels 2 and 3, they would write 50% under "2" and 50% under "3". They were allowed to indicate up to two language proficiency levels and a range in 10-point increments (i.e., 50/50, 60/40, 70/30, 80/20, or 90/10), or to indicate 100 under one language proficiency level. The results were compiled and discussed with the panelists as a group. The panelists then were given the opportunity to reconsider and adjust their bookmarking, if they so chose. The final results were analyzed by CAL using a logistic regression procedure to determine the points along the underlying writing proficiency continuum at which at least 50% of the panelists would be expected to agree that the writing represents the work of the next higher proficiency level than the current proficiency level. The results from this analysis were used to set the cut scores for the language proficiency levels.

The procedure for Speaking was similar, with the panelists listening to portfolios and recording their judgments.

1.4   Ongoing Item Development

To keep ACCESS for ELLs secure, as well as to incorporate a program of continual refinement to the assessment (e.g., using colored illustrations), new items are being developed and field tested every year. In fact, one third to one half of the items is replaced yearly. The intent of the ongoing item development is to replace completely all items or tasks in each test form over a three-year period.

The schedule for refreshing items in the ACCESS for ELLs is illustrated in Table 1.4A. This table applies to all grade-level clusters except K, which was redesigned for Series 200 and is not refreshed annually. As can be seen from this table, for the Listening and Reading tests, all of the LA and MA tasks are replaced in alternating years, while the SC, SS, and SIL tasks are replaced in a three-year cycle. For Speaking, the SIL task is replaced yearly, while the MA/SC and LA/SS tasks are replaced in alternating years. New items are field tested on separate forms during the operational administration of ACCESS for ELLs.

Table 1.4A also reflects a change in the Writing test that took effect starting with Series 201. In that series, the separate Math and Science folders were replaced with a combined Math/Science folder. Starting with that series, while the IT task will continue to be replaced yearly, the MA/SC and LA tasks will be replaced in a two-year cycle for Tier A, and the MA/SC and SIL tasks will be replaced in a two-year cycle for Tiers B and C.

From Table 1.4A, we see that between Series 101 and Series 100, the IT Writing task and the MA/SC Speaking task were replaced. In the Listening and Reading portion of the test, various item folders were replaced following analysis of the field test and operational Series 100. Because ACCESS for ELLs was so new, it was decided that it was most important to be able to improve and/or replace weaker items across all five Standards than to choose only two Standards to be replaced.

Table 1.4A
Schedule for Refreshing ACCESS for ELLs Items

Test Series	Year	Listening		Reading		Writing Tier A	Writing Tiers B &C		Speaking	
		LA/ MA	SC/SS/ SIL	LA/ MA	SC/SS/ SIL	SIL/LA/ MA/SC	SIL/MA/ SC	IT	SIL	LA/ SS/ MA/SC
100	04-05									
101	05-06	Various	Various	Various	Various	-		IT		MA/ SC
102	06-07	MA	SIL	LA	SC	-	MA	IT	SIL	LA/ SS
103	07-08	LA	SC	MA	SS	-	SC	IT	SIL	MA/ SC
200	08-09	Some MA	Some SS	Some LA	Some SIL	-	SIL	IT	SIL	LA/ SS
201	09-10	Various + Remaining MA	Various + Remaining SS	Various + Remaining LA	Various + Remaining SIL	-	-	IT	SIL	MA/ SC
202	10-11	LA	SIL	MA	SC	MA/SC	MA/SC	IT	SIL	LA/ SS
203	11-12	MA	SC	LA	SS	LA	SIL	IT	SIL	MA/ SC
301	12-13	LA	SS	MA	SIL	MA/SC	MA/SC	IT	SIL	LA/ SS

Social and Instructional language (SIL); Language of English Language Arts (LA); Language of Math (MA); Language of Science (SC); Language of Social Studies (SS); Integrated (SIL, LA, and SS)

The following paragraphs describe annual procedures currently in place that influence the development of future items.

1.4.1 Item Writing and Editing

The initial item writing is done by participants in an online item writing course conducted by CAL. An internal review of the items generated by that course is conducted, and items are chosen for further development based on how well they fit the Standards and PIs, and how different they are in terms of content from the previous year's items. Those items chosen are refined within CAL before undergoing an item bias and content review (see Section 1.4.2). Afterward, some items require further revision at CAL before being sent to MetriTech and WIDA central office for final review. Once returned to CAL, they are prepared for the field test.

1.4.2 Item Content and Bias Review

After items are internally refined, they are submitted to panels for content and bias review. The panels consist of educators from the WIDA Consortium states. Items are first submitted to the content review panel to assure that the content is accessible and relevant to students in the grade-level cluster, and that each item or assessment task matches the model performance indicators from the WIDA ELD Standards that it is intended to assess. After the items are revised based on the comments from the panel members, they are submitted to the bias review panel, which inspects the items for potential bias. For the bias review panel, panelists represent a wide variety of language backgrounds and ethnicities. Based on their recommendations, the items are revised as necessary.

1.4.3 Item Field Testing

All new items are field tested in conjunction with the current year's operational administration. Larger districts from across the WIDA Consortium states are invited to participate on a rotating schedule, and only districts that accept the invitation actually participate in the field test. Field testing occurs in WIDA states across the country immediately after the operational test is administered. Each participating student is administered items in only one domain. The field test is designed to take no more than 15 minutes on the part of any student participant.

For Listening and Reading, several forms of new items are prepared for each grade-level cluster, each containing two folders of new items and one folder of anchor items, in order to understand the difficulty of the new items in relation to the ACCESS for ELLs score scale. Thus, there are a total of three folders (9 items) per form. Within each form, an effort is made to alternate Standards. Thus, one form of the Listening field test might have two MA and one SI folders, while the other form has one MA and two SI folders. For Writing, four tasks are prepared per grade-level cluster: one task at each tier for the year's standard, and one IT task. Students are presented with just one task, when possible at the appropriate tier. For Speaking, two folders of tasks are prepared per grade-level cluster, and each student is presented with both folders.

1.4.4 Item Calibration and Analysis

After the items are field tested, the results are analyzed using a Rasch model to determine their difficulty measure on the ACCESS for ELLs score scale. Items are also analyzed as to all aspects of their functioning (e.g., fit statistics) to determine whether they may be included in the next year's operational form. Only folders of items meeting all technical requirements are placed into the operational form.

1.4.5 DIF Items

Starting with Series 203, two phases of analysis (Phase I and Phase II) for differential item functioning (DIF) are conducted on the operational form while operational testing is still ongoing, in addition to the DIF analysis conducted for the Annual Technical Report. 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 bias toward a particular group, and an item exhibiting C level DIF is considered to display bias and should be closely examined by test developers.

Phase I is conducted at the same time as equating (see Section 1.3.2) using two sources of data: one, all student data available a week before the equating sample is pulled, called Early Return; two, the equating sample, called Equating Sample. During Phase I analysis, only ethnicity DIF (Hispanic vs. Non-Hispanic) is investigated. In this phase, items that show high levels of DIF in both data sets are investigated by a team of content experts to determine if any construct-irrelevant factors can be identified that may contribute to DIF. Items which are identified as having construct-irrelevant sources of DIF will not be scored operationally. One item was identified as having a C-level ethnicity DIF favoring Hispanics in the Early Return data but a B-level DIF favoring Hispanics for the Equating Sample; therefore, no further action was required. For Series 301, no items were unscored because of DIF in Phase I.

Phase II is conducted using all student data available in early May. During Phase II analysis, ethnicity and gender DIF were investigated. As with Phase I, items that show high levels of DIF

are investigated by a team of content experts to determine if any construct-irrelevant factors can be identified that may contribute to DIF. Items which are identified as having construct-irrelevant sources of DIF will be removed from the test in the next operational year. For Series 301, one listening item was identified as having C-level ethnicity based DIF, favoring Hispanics. Since the Listening domain test format was revised to Media delivered for Series 302, all Listening are new for Series 302.

For the Annual Technical Report, an ethnicity and gender DIF analysis is conducted using all student data. For Series 301, two items showed DIF. Out of 270 Listening items, one (0.4%) showed C-level DIF based on ethnicity, favoring Hispanics. Out of 43 Writing items, one (2.3%) showed C-level DIF based on ethnicity, favoring Non-Hispanics. These items are thoroughly analyzed by the Psychometrics/Research team at CAL to determine the potential sources of DIF. In terms of DIF by ethnicity (Hispanics versus Non-Hispanics), special attention is paid to the presence of Spanish-English cognates or false cognates that may affect student performance. That information is provided to the test development team, which makes necessary revisions to continuing items and keeps a record of such cognates for future reference. The test development team uses this information to guide the item development and review process for future items.

For information on the procedures used to calculate DIF, see Section 5.1.4.

1.5   Reporting of Results

1.5.1 Scale Scores

ACCESS for ELLs 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 Composite (based on performances in Listening and Speaking), Literacy Composite (based on performances in Reading and Writing), Comprehension Composite (based on performances in Listening and Reading), and Overall Composite (based on performances in all four domains).

Raw scores are converted to scale scores through processes called equating and scaling (see section 1.3.2 for details). These processes allow us to report scores 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 100 to 600. Beginning with Series 102, the center point of the scale, 350, which formerly represented the cut score between Language Proficiency Levels 3 and 4 for the 3–5 Grade-level Cluster, represents the same cut score for Grade 5 only.

The scores for the four composite scores are calculated using the following weights:

- Oral Language Composite (50% Listening + 50% Speaking)
- Literacy Composite (50% Reading + 50% Writing)
- Comprehension Composite (30% Listening + 70% Reading)
- Overall Composite (15% Listening + 15% Speaking + 35% Reading + 35% Writing).

Figure 1.5.1A depicts the weighting for each of the composite scores. As shown, the Overall Composite is computed using scores from all four domains. Each of the other three composites is shown with the weighting of domains, in terms of the weighting used for the Overall Composite. As the diagram shows, more weighting is given to the literacy skills than to the oral skills for the Overall Composite
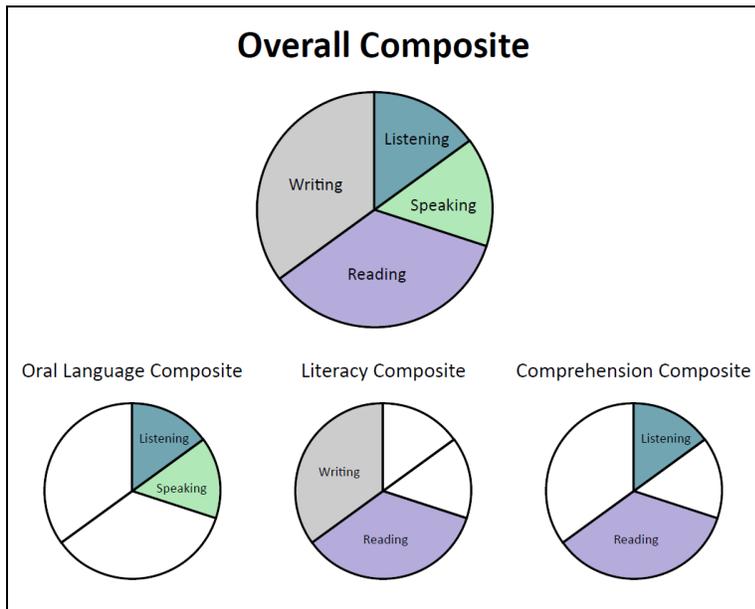
Figure 12.5.1A. Domain Composites

1.5.2 Language Proficiency Level Scores

In addition to the ACCESS scale scores, users of 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 Tables 1.5.2A–H and reflect the adoption of the grade-level cut scores for Series 102 and beyond, as well as the Instructional and Accountability cut scores adapted for Kindergarten for Series 200 and beyond.

Table 1.5.2A
Cut Scores (Listening)

Grades	Domain	Cut				
		1/2	2/3	3/4	4/5	5/6
K (Instructional)	List	175	204	240	279	322
K (Accountability)	List	229	251	278	286	308
1	List	238	267	295	305	330
2	List	247	281	311	324	350
3	List	255	295	325	340	367
4	List	264	307	338	355	383
5	List	274	318	350	368	397
6	List	283	328	359	380	409
7	List	293	337	368	390	418
8	List	302	345	375	399	426
9	List	312	352	381	406	432
10	List	322	358	386	412	436
11	List	332	363	389	416	438
12	List	343	366	391	418	439

Table 1.5.2B
Cut Scores (Reading)

Grades	Domain	Cut				
		1/2	2/3	3/4	4/5	5/6
K (Instructional)	Read	121	159	204	228	255
K (Accountability)	Read	238	251	261	274	295
1	Read	253	269	283	294	314
2	Read	267	286	303	312	331
3	Read	279	302	320	328	347
4	Read	291	316	336	343	360
5	Read	302	328	350	355	372
6	Read	312	340	360	366	382
7	Read	321	349	369	375	391
8	Read	329	358	376	382	398
9	Read	336	364	381	387	402
10	Read	341	370	383	390	406
11	Read	346	374	384	392	407
12	Read	350	376	385	393	408

Return to Summary of Assessment Records Claims, Actions, and Evidence

Table 1.5.2C
Cut Scores (Writing)

Grades	Domain	Cut				
		1/2	2/3	3/4	4/5	5/6
K (Instructional)	Writ	145	218	244	269	326
K (Accountability)	Writ	225	259	295	323	350
1	Writ	238	272	308	336	362
2	Writ	251	285	320	348	373
3	Writ	264	297	330	360	384
4	Writ	275	308	340	371	394
5	Writ	287	319	350	381	403
6	Writ	298	329	361	391	412
7	Writ	308	339	371	399	420
8	Writ	318	348	381	408	428
9	Writ	327	356	389	415	435
10	Writ	336	363	397	422	441
11	Writ	344	370	404	428	447
12	Writ	352	377	410	434	452

Table 1.5.2D
Cut Scores (Speaking)

Grades	Domain	Cut				
		1/2	2/3	3/4	4/5	5/6
K (Instructional)	Spek	256	285	308	342	365
K (Accountability)	Spek	269	314	343	366	383
1	Spek	278	318	344	367	385
2	Spek	286	322	345	368	386
3	Spek	293	326	346	369	389
4	Spek	299	329	348	371	391
5	Spek	305	333	350	374	394
6	Spek	310	337	353	377	397
7	Spek	314	340	358	380	400
8	Spek	317	344	361	384	404
9	Spek	319	347	366	388	407
10	Spek	321	351	371	393	412
11	Spek	322	354	377	399	416
12	Spek	322	357	384	405	421

Table 1.5.2E
Cut Scores (Oral Language Composite)

Grades	Domain	Cut				
		1/2	2/3	3/4	4/5	5/6
K (Instructional)	Oral	216	245	274	311	344
K (Accountability)	Oral	249	283	311	326	346
1	Oral	258	293	320	336	358
2	Oral	267	302	328	346	368
3	Oral	274	311	336	355	378
4	Oral	282	318	343	363	387
5	Oral	290	326	350	371	396
6	Oral	297	333	356	379	403
7	Oral	304	339	363	385	409
8	Oral	310	345	368	392	415
9	Oral	316	350	374	397	420
10	Oral	322	355	379	403	424
11	Oral	327	359	383	408	427
12	Oral	333	362	388	412	430

Table 1.5.2F
Cut Scores (Literacy Composite)

Grades	Domain	Cut				
		1/2	2/3	3/4	4/5	5/6
K (Instructional)	Litr	133	189	224	249	291
K (Accountability)	Litr	232	255	278	299	323
1	Litr	246	271	296	315	338
2	Litr	259	286	312	330	352
3	Litr	272	300	325	344	366
4	Litr	283	312	338	357	377
5	Litr	295	324	350	368	388
6	Litr	305	335	361	379	397
7	Litr	315	344	370	387	406
8	Litr	324	353	379	395	413
9	Litr	332	360	385	401	419
10	Litr	339	367	390	406	424
11	Litr	345	372	394	410	427
12	Litr	351	377	398	414	430

Table 1.5.2G
Cut Scores (Comprehension Composite)

Grades	Domain	Cut 1/2	2/3	3/4	4/5	5/6
K (Instructional)	Cphn	138	173	215	244	276
K (Accountability)	Cphn	235	251	266	278	299
1	Cphn	249	268	287	297	319
2	Cphn	261	285	305	316	337
3	Cphn	272	300	322	332	353
4	Cphn	283	313	337	347	367
5	Cphn	294	325	350	359	380
6	Cphn	303	336	360	370	390
7	Cphn	313	345	369	380	399
8	Cphn	321	354	376	387	406
9	Cphn	329	360	381	393	411
10	Cphn	335	366	384	397	415
11	Cphn	342	371	386	399	416
12	Cphn	348	373	387	401	417

Table 1.5.2H
Cut Scores (Overall Composite)

Grades	Domain	Cut 1/2	2/3	3/4	4/5	5/6
K (Instructional)	Over	158	206	239	268	307
K (Accountability)	Over	237	263	288	307	329
1	Over	249	277	303	321	344
2	Over	261	290	316	335	357
3	Over	272	303	328	347	369
4	Over	283	314	340	359	380
5	Over	293	324	350	369	390
6	Over	302	334	359	379	399
7	Over	311	342	368	386	407
8	Over	319	350	375	394	414
9	Over	327	357	382	400	419
10	Over	333	363	387	405	424
11	Over	340	368	391	409	427
12	Over	345	372	395	413	430

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

Unlike ACCESS scale scores, which form an interval scale and are continuous across the grades from K to 12, ACCESS proficiency level scores are, of course, dependent upon which grade a student was in when ACCESS for ELLs was taken. See, for example, the Listening cut scores in Table 1.5.2A. If a child is in Grade 2 and receives a 350 in Listening, that would be a proficiency level score of 6.0; if the child is in Grade 5 and receives a 350 in Listening, that would be 4.0; if the child is in Grade 8 and receives a 350 in Listening, that would be a 3.2; and if a child is in Grade 12 and receives a 350 in Listening, that would be a 2.3. (Note that grade-level-cluster cut scores were used to interpret performances on ACCESS for ELLs for Series 100 and 101. Beginning with Series 102, grade-level cut scores were used.)

Note that because the width between cut scores varies, proficiency level scores should not be considered as forming an interval scale. That is, it cannot be assumed to be the same distance between proficiency level scores of 1.5 and 2.5 as between 2.5 and 3.5. Only scale scores should be used as interval measures. Proficiency level scores are interval within a grade and level (e.g., it is the same distance in grade 3 between 3.1 and 3.2 as between 3.7 and 3.8), but they do not form an interval scale across language proficiency levels.

1.5.3 Results by English Language Development Standards

To provide a more complete picture of a student's performance, raw scores are reported by English Language Development Standards.

For Comprehension (combined Listening and Reading), the five ELD Standard scores (Social and Instructional language, Language Arts, Mathematics, Science, and Social Studies) are reported as number correct out of maximum possible (e.g., 3 of 8). It should be noted that the absolute number of items that a student sees in any given language proficiency area varies by tier.

For Speaking, standard scores are reported as raw numbers based on the number of tasks that the student met or exceeded in that standard. The maximum score for Social and Instructional language is 3; the maximum for Language Arts/Social Studies and for Mathematics/Science is 5.

For Writing tasks, three standard ratings are reported for each of the three or four tasks on the form. The three ratings are for Linguistic Complexity, Vocabulary Usage, and Language Control. Each of these scores can range from 0 to 6.

1.6 Test Administration

1.6.1 Test Administrator Training

To prepare individuals to serve as test administrators, test administrator training for Series 301 was conducted through an online course hosted at www.wida.us. Three certifications were offered to participants: a group test administration certification pertaining to the Listening, Reading, and Writing portions of ACCESS for ELLs; a certification for the Speaking test; and a certification for the Kindergarten test. In order to receive a certification, participants had to pass a quiz.

1.6.2 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.

1.6.3 Test Accommodations

As a test of developing English language proficiency designed for English language learners, there are no special test accommodations for this group of students. However, if a student also has an IEP, to the extent possible and practical, the recommendations in the student's IEP are to be followed. The extent to which this was successfully accomplished was a local decision during the administration of Series 301.

The WIDA Consortium for the first time ever during the 2011–2012 testing cycle has made available an alternate assessment for ACCESS for ELLs: Alternate ACCESS for ELLs (Alternate ACCESS). Alternate ACCESS is intended only for English language learners who have cognitive disabilities that are severe enough to prevent meaningful participation in the ACCESS for ELLs with accommodations. The results of the Alternate ACCESS operational administration will appear in a separate technical report (forthcoming).

The recommendations regarding physical disabilities, such as deafness or blindness, are available on the WIDA website (http://www.wida.us/get.aspx?id=289) but are being clarified for more standardization.

1.7 Scoring

Test booklets are returned to MetriTech, where they are electronically scanned in preparation for scoring. Listening, reading, and writing are scored by Metritech. Speaking is locally scored by the test administrator. Details of the scoring methods are described below.

1.7.1 Listening and Reading

In the case of the Listening and Reading tests, all items are selected-response and thus are dichotomously scored as correct or incorrect. Students have entered their answers directly into the test booklets, so each page is scanned into an electronic database.

1.7.2 Writing

Students' responses to the Writing tasks are centrally scored at MetriTech by raters who are trained to follow the WIDA Consortium's Writing Rubric (see 1.7.2.1). The rubric reflects the Performance Level Descriptions of the WIDA ELD Standards and is presented in Table 1.7.2A.

Table 1.7.2A
Performance Level Descriptions of the WIDA ELD Standards

At the given level of English language proficiency, English language learners will process, understand, produce or use:

Level	Descriptions
6 - Reaching	- specialized or technical language reflective of the content area at grade level - a variety of sentence lengths of varying linguistic complexity in extended oral or written discourse as required by the specified grade level - oral or written communication in English comparable to proficient English peers
5 - Bridging	- specialized or technical language of the content areas - a variety of sentence lengths of varying linguistic complexity in extended oral or written discourse, including stories, essays, or reports - oral or written language approaching comparability to that of English-proficient peers when presented with grade-level material
4 - Expanding	- specific and some technical language of the content areas - a variety of sentence lengths of varying linguistic complexity in oral discourse or multiple, related sentences or paragraphs - oral or written language with minimal phonological, syntactic, or semantic errors that do not impede the overall meaning of the communication when presented with oral or written connected discourse with sensory, graphic, or interactive support
3 - Developing	- general and some specific language of the content areas - expanded sentences in oral interaction or written paragraphs - oral or written language with phonological, syntactic, or semantic errors that may impede the communication, but retain much of its meaning, when presented with oral or written, narrative, or expository descriptions with sensory, graphic, or interactive support
2 - Emerging	- general language related to the content areas - phrases or short sentences - oral or written language with phonological, syntactic, or semantic errors that often impede the meaning of the communication when presented with one to multiple-step commands, directions, questions, or a series of statements with sensory, graphic, or interactive support
1 - Entering	- pictorial or graphic representation of the language of the content areas - words, phrases, or chunks of language when presented with one-step commands, directions, WH-, choice, or yes/no questions, or statements with sensory, graphic, or interactive support - oral language with phonological, syntactic, or semantic errors that often impede meaning when presented with basic oral commands, direct questions, or simple statement with sensory, graphic or interactive support

The Writing rubric contains expectations for three aspects of Writing that play an important role in determining proficiency level: Linguistic Complexity, Vocabulary Usage, and Language Control. Table 1.7.2B presents the WIDA Consortium's Writing Rubric.

Table 1.7.2B
WIDA Consortium's Writing Rubric for Grades 1-12

Level	Linguistic Complexity	Vocabulary Usage	Language Control
6 Reaching	A variety of sentence lengths of varying linguistic complexity in a single tightly organized paragraph or in well-organized extended text; tight cohesion and organization	Consistent use of just the right word in just the right place; precise Vocabulary Usage in general, specific, or technical language	Has reached comparability to that of English proficient peers functioning at the "proficient" level in state-wide assessments
5 Bridging	A variety of sentence lengths of varying linguistic complexity in a single organized paragraph or in extended text; cohesion and organization	Usage of technical language related to the content area; evident facility with needed vocabulary	Approaching comparability to that of English proficient peers; errors don't impede comprehensibility
4 Expanding	A variety of sentence lengths of varying linguistic complexity; emerging cohesion used to provide detail and clarity	Usage of specific and some technical language related to the content area; lack of needed vocabulary may be occasionally evident	Generally comprehensible at all times, errors don't impede the overall meaning; such errors may reflect first language interference
3 Developing	Simple and expanded sentences that show emerging complexity used to provide detail	Usage of general and some specific language related to the content area; lack of needed vocabulary may be evident	Generally comprehensible when writing in sentences; comprehensibility may from time to time be impeded by errors when attempting to produce more complex text
2 Emerging	Phrases and short sentences; varying amount of text may be copied or adapted; some attempt at organization may be evidenced	Usage of general language related to the content area; lack of vocabulary may be evident	Generally comprehensible when text is adapted from model or source text, or when original text is limited to simple text; comprehensibility may be often impeded by errors
1 Entering	Single words, set phrases, or chunks of simple language; varying amounts of text may be copied or adapted; adapted text contains original language	Usage of highest frequency vocabulary from school setting and content areas	Generally comprehensible when text is copied or adapted from model or source text; comprehensibility may be significantly impeded in original text

In addition to training in the generic rubric, training is provided to scorers as to expectations for each grade level and for each writing task. For example, exceptional vocabulary usage in the 1–2 grade-level cluster would not be so exceptional at the 9–12 grade-level cluster. The amount of writing and sophistication of thought at each performance level generally increases with moving up the grade-level clusters. Thus, a single generic rubric rooted in the WIDA ELD Standards lies at the core of the scoring of writing, but developmental differences between grade-level clusters are part of the additional training that each rater receives.

Scorers are provided anchor papers for each task. Training sets are also created, as well as calibration sets with which scorers are tested during the operational training session. Raters failing to meet standards on the calibration sets are removed from scoring.

In applying the rubric, the following method of scoring writing is used. First, the Metritech rater determines the language proficiency level that best characterizes the writing sample (e.g., Level 3). Then, the rater considers whether in any category the writing displayed any particular weakness (i.e., was lower in one of the three) or displayed any particular strength (i.e., was higher in one of the three categories). Finally, the rater awards three scores, one for each category: a 3-3-3 represents a solid Level 3 writing sample; a 3-3-2 is a low Level 3 writing sample that is a little weaker than expected in its language control; while a 3-4-3 is a high Level 3 writing sample that is a little stronger than expected in its vocabulary usage. The final score is the sum of the three scores; i.e., 9 for a solid Level 3 paper, 8 for a low Level 3 paper, and 10 for a high Level 3 paper.

In calculating an Overall Composite raw score for Writing, results from the different tasks are given different weights. These weights are intended to reflect the amount of writing that each task may be expected to produce. The weightings for the different tasks are as follows:

Kindergarten: 1-1-1-1-2-1

Grades 1–2 Tier A form: 1-1-1-3

Grades 1–12 Tier B and C forms: 1-2-3

Grades 3–12 Tier A forms: 1-1-1

For example, for all grades on Tier B and C tests the three tasks are given weights of 1, 2, and 3. Thus, a student who receives scores of 6, 5, and 4 on the three Writing tasks for that test would have an overall writing raw score of 28 ((6*1) +(5*2) + (4*3)).

1.7.2.1 Scoring Procedures for Writing

Scoring of ACCESS for ELLs is handled at the MetriTech scoring facilities in Illinois.

All constructed-response scoring for ACCESS for ELLs is performed utilizing a proprietary on-line scoring system (*MTscore*). As with all aspects of ACCESS for ELLs, MetriTech's top concern is security of student data and the items and forms eliciting student responses. Some of the strict security measures implemented as part of *MTscore* include:

- All students' identifying or biographical data (including name, ID number, gender, etc.) will be stripped from scorer images and will not be included in data transferred into *MTscore*

- Students' constructed responses will have an untraceable, non-identifying index number

- Item and student response images will be available only through *MTscore* and cannot be accessed by any outside network or saved on any media

- No image or portion of student response image can be printed, with the exception of Master Scorers only needing to print student responses in cases of alert papers.

- Score session access restrictions, requiring scorer login during predetermined times and dates only

With scoring centers located near several universities, MetriTech has a large pool of qualified scoring applicants from which to choose. Applicants must possess a minimum of a bachelor's degree and pass proprietary pre-employment tests found to predict performance. Many have backgrounds in education and are active or retired teachers.

Applicants are required to attend a pre-employment testing session where they review their already completed online application, answer additional questions specific to the project for which they are applying, and complete a series of proprietary pre-employment screening tests that reliably predict scorer performance. Hiring criteria include, but are not limited to, completion of at least a bachelor's degree from an accredited college or university; work experience, particularly teaching or education-related experience; and test scores.

Lead scoring staff members include master scorers and trainers (already on staff at the start of each project), as well as table leaders and senior scorers for each content area (usually assigned to specific projects based on their content qualifications and background). Before scorer training begins, group leaders and senior scorers are trained by master scoring staff so that they are familiar with the rubrics, annotated anchor papers (originally provided by CAL and augmented by senior MetriTech staff each year), training sets, calibration sets, and scoring procedures. MetriTech master scoring staff has been working with scoring protocols for various programs and states for an average of seven years. This core group works closely with the CAL development staff, augmenting originally supplied training materials for each year of ACCESS scoring.

Each potential reader has been selected to train on a particular grade-span. The training process starts with an on-line training session, where each reader will review the rubrics, the elements of analytic scoring, and anchor papers. Each score point on each rubric is defined, and approved examples of student work that meet the criteria for each score point are presented and analyzed. Following this presentation, the readers work through selected modules of sample papers. Each paper in a training module has already been reviewed and scored by the master reader. As the readers finish the training module, their recorded scores and rationales for their scoring are systematically scored. Discrepancies are noted and feedback and additional modules are presented to the reader to provide further training. Finally, each reader is given a post-test module containing sample student responses. The readers score these modules independently, and the final scores that they assign are compared with those assigned by the master reader. Readers need to reach the criterion of 70% exact agreement with the master reader's score to complete training and to be approved to score live test material. This process is repeated for each scorer selected for training. Training sessions utilizing one-on-one on-line interactive modular trainings supported by printed training manuals and master trainer Q&A provide readers by grade-cluster and typically include eight hours of material.

How scorers are supervised during the scoring process.

Group Leaders

- Prioritize work assignments for the scorers in their group for each shift

- Assign scorers work for each shift

- Review completed scoring for their group

- Track scorer attendance

- Monitor decorum within their group

Room Leaders

- Coordinate all Group Leaders on a shift

- Prioritize work assignments for the room for each shift

- Track scorer productivity on each shift

- Monitor decorum for the room

Master Scorers and Trainers

- Complete quality control/scoring checks on all employees on a daily basis, at pre-determined rates

- Provide written as well as verbal one-on-one feedback to scorers on a daily basis

- Provide retraining as needed

- Recommend scorer reassignment as needed

For the ACCESS for ELLs constructed-response scoring, papers from each scorer are randomly directed to the group leader for re-checking. If a group leader finds that a scorer's rates fall below the expected standard, the scorer is directed to retraining.

To monitor that the scoring rubric is being applied consistently across scoring sessions, specially prepared calibration sets are routed to each scorer daily. To the scorer, the calibration student images look like regular student responses. However, master scorers have already reviewed each response in these sets, and the master scorer has created a key of expected scores for each sample. Once the scorer completes the set, the scores that he or she assigned are immediately checked against the master key by the system. This approach allows for the immediate detection and correction of scorer drift. Exact agreement levels between the active and master scorer must exceed the standards established for the project (80% exact agreement) or the scorer is locked out of the system until they have successfully completed a retraining with the master scorer.

Twenty percent of all constructed-response items are blindly re-scored by another reader to provide overall inter-rater reliability. This information is kept for future analysis, reporting in the technical report, and reporting to the master scorer, allowing another avenue of feedback to the individual scorers.

1.7.3 Speaking

The Speaking test is administered individually to each test taker. Each task is immediately scored by the administrator while the test is being given. The administration and scoring procedure were designed together to be quite simple to implement. As described previously, the Speaking tasks are designed around the PIs to allow students to demonstrate mastery of the performance level for which the task is designed. After administering each task and listening to the student's responses, the administrator decides whether the student's performance exceeds, meets, or approaches task-level expectations. Specifically, the possible ratings are defined as follows:

Exceeds: The student's performance exceeds task-level expectations in quantity and/or quality.

Meets: The student's performance meets task-level expectations in quantity and quality.

Approaches: The student's performance approaches task-level expectations, but falls short in quantity and/or quality.

No Response: The student's performance is quite inadequate: there is no response, the response is incomprehensible or in a language other than English, or the student is unable to understand the task directions.

Operationally, a score of 1 is given for every task that either meets or exceeds expectations, and a 0 is given for any task that is rated as approaches or no response. The sum of those scores is the total Speaking raw score for that student.

Table 1.7.3A presents the WIDA Consortium's Speaking Rubric, which summarizes the expectations for each task level on the Speaking assessment. These expectations are drawn from the performance level descriptions of the WIDA ELD Standards and are divided into three components (Linguistic Complexity, Vocabulary Usage, and Language Control). The training for test administrators consists of familiarizing them with the tasks at each level and listening to responses to those tasks, determining whether they meet the task-level expectations or not.

1.7.3.1 Training Procedures for Scoring Speaking

The Speaking Test is the only portion of ACCESS for ELLs that the test administrator scores. Test administrators must complete the Speaking Test module of the online ACCESS for ELLs Test Administrator Training and the accompanying quiz. The training focuses on developing the test administrator's ability to conduct the test using standardized testing procedures and to score the test reliably. Test administrators are provided training on test administration procedures such as navigating the test, scores and ratings. To reliably score the test, test administrators are then trained on the Speaking Rubric of the WIDA Consortium (see Table 1.7.3A). Test administrators must study the rubric thoroughly to understand each of the requirements for speech, demonstrating proficiency at each of the different levels. Speaking Rubric training is accomplished by listening to online ACCESS for ELLs Test Administrator Training speech samples. Each sample presents a task targeted at a particular proficiency level to allow test administrators to evaluate the responses against the three criteria described in the rubric for the task. Scores and rationales that are provided for each sample demonstrate how and why a particular score is assigned. To be considered certified to administer the ACCESS for ELLs Grades 1–12 Speaking test, test administrators will then need to take the Speaking test quiz that accompanies the training test module.

Table 1.7.3A
WIDA Consortium's Speaking Rubric

Task Level	Linguistic Complexity	Vocabulary Usage	Language Control
1 **Entering**	Single words, set phrases, or chunks of memorized oral language	Highest frequency vocabulary from school setting and content areas	Generally comprehensible and fluent when using memorized language; communication may be significantly impeded when going beyond the highly familiar
2 **Emerging**	Phrases, short oral sentences	General language related to the content area; groping for vocabulary when going beyond the highly familiar is evident	Generally comprehensible and fluent when using simple discourse; communication may be impeded by groping for language structures or by phonological, syntactic, or semantic errors when going beyond phrases and short, simple sentences
3 **Developing**	Simple and expanded oral sentences; responses show emerging complexity used to add detail	General and some specific language related to the content area; may grope for needed vocabulary at times	Generally comprehensible and fluent when communicating in sentences; communication may from time to time be impeded by groping for language structures or by phonological, syntactic, or semantic errors, especially when attempting more complex oral discourse
4 **Expanding**	A variety of oral sentence lengths of varying linguistic complexity; responses show emerging cohesion used to provide detail and clarity	Specific and some technical language related to the content area; groping for needed vocabulary may be occasionally evident	Generally comprehensible and fluent at all times, though phonological, syntactic, or semantic errors that don't impede the overall meaning of the communication may appear at times; such errors may reflect first language interference
5 **Bridging**	A variety of sentence lengths of varying linguistic complexity in extended oral discourse; responses show cohesion and organization used to support main ideas	Technical language related to the content area; facility with needed vocabulary is evident	Approaching comparability to that of English proficient peers; errors don't impede communication and may be typical of those an English proficient peer may make

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

Validity is "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests." (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 1999, p. 9). 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 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 scores to intended consequences. This chapter begins the process of developing a complete validation framework for ACCESS for ELLs. This argument-based structure organizes the information in this Annual Technical Report to support assertions about *Assessment Records* (i.e., data collected via ACCESS for ELLs). Specifically, tables and figures from this report are explicitly linked to questions related assessment data. 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 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 testing 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; Wang, Choi, Schmidgall, & Bachman, 2012); Figure 2A shows the process by which evidence supports validation actions, which are used to advance larger claims about ACCESS for ELLs.
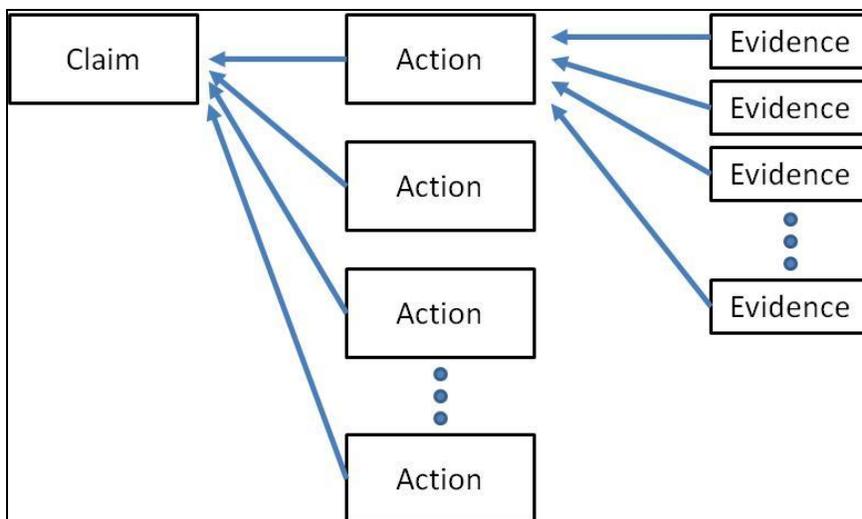
Figure 2A: General Argument Structure for Assessment Validation (simplified from Toulmin, 2003)

2.1   The Generic Validation Framework for ACCESS

The generic validation framework that will be applied to the entire ACCESS for ELLs testing process was developed at the Center for Applied Linguistics (CAL). This 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 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.

In this 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*).
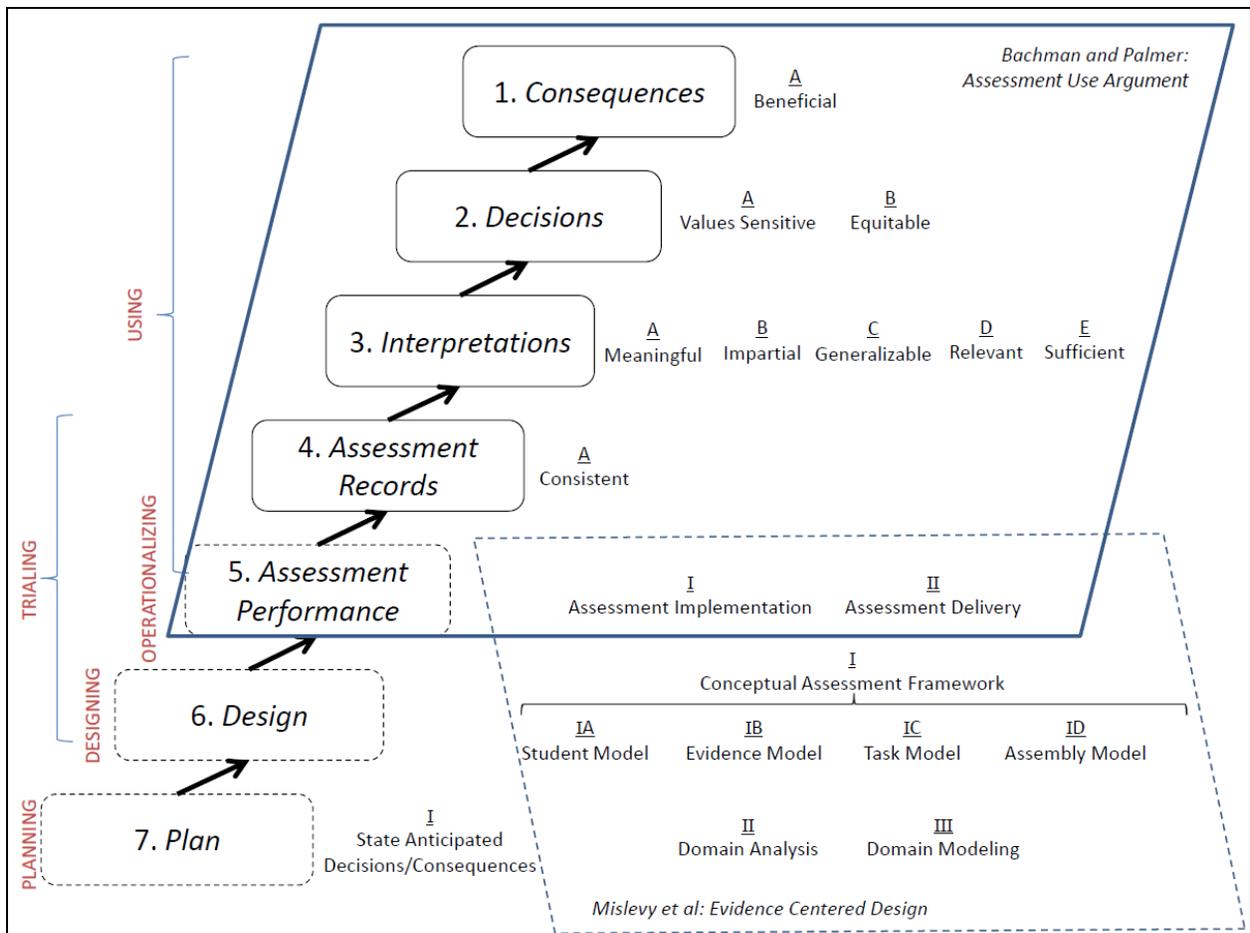
Figure 2.1A: CAL's validation framework (based on Bachman & Palmer, 2010; Mislevy, Almond, & Lukas, 2004)

The WIDA Consortium is using this generic validation framework to present a complete validity argument, which will be updated as needed, for 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 ACCESS for ELLs contains seven steps, 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*, 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 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 and composite scores on ACCESS for ELLs are consistent and dependable?" Other questions about the development, administration, and use of 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 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.
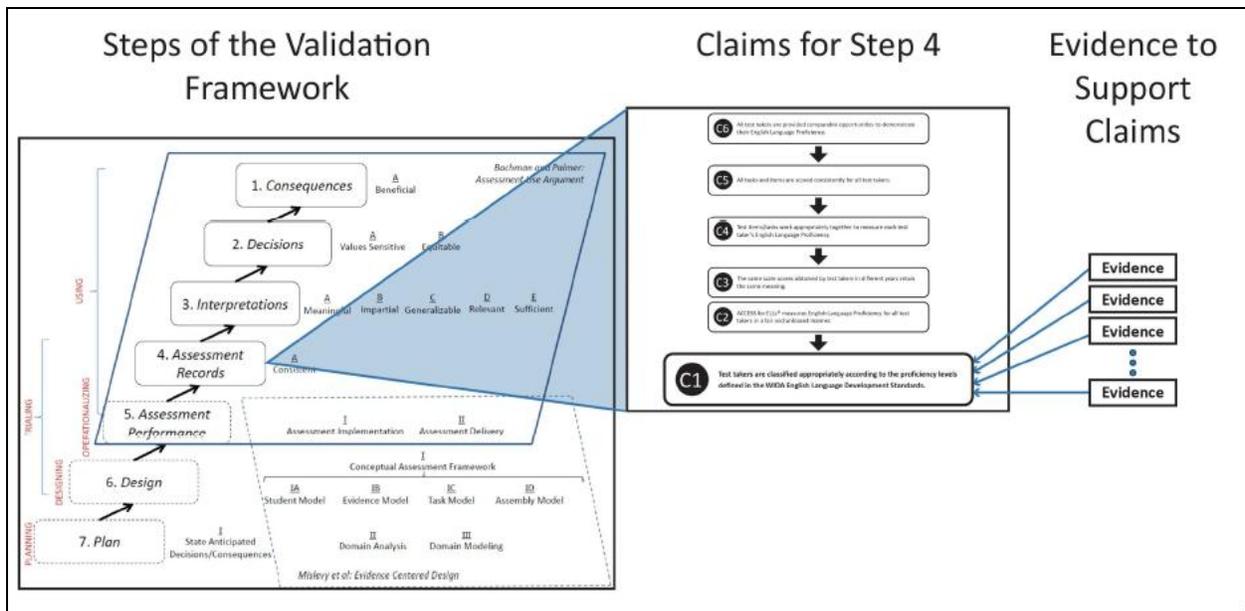
Figure 2.2A: Structure of the Argument-Based Approach Supporting Step 4 Contained in this Chapter

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

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

C6. All test takers are provided comparable opportunities to demonstrate their English Language Proficiency.
C5. All tasks and items are scored consistently for all test takers.
C4. Test items/tasks work appropriately together to measure each test taker's English Language Proficiency.
C3. The same scale scores obtained by test takers in different years retain the same meaning.
C2. ACCESS for ELLs measures English Language Proficiency for all test takers in a fair and unbiased manner.
C1. Test takers are classified appropriately according to the 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 lowest (6) to the highest (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.
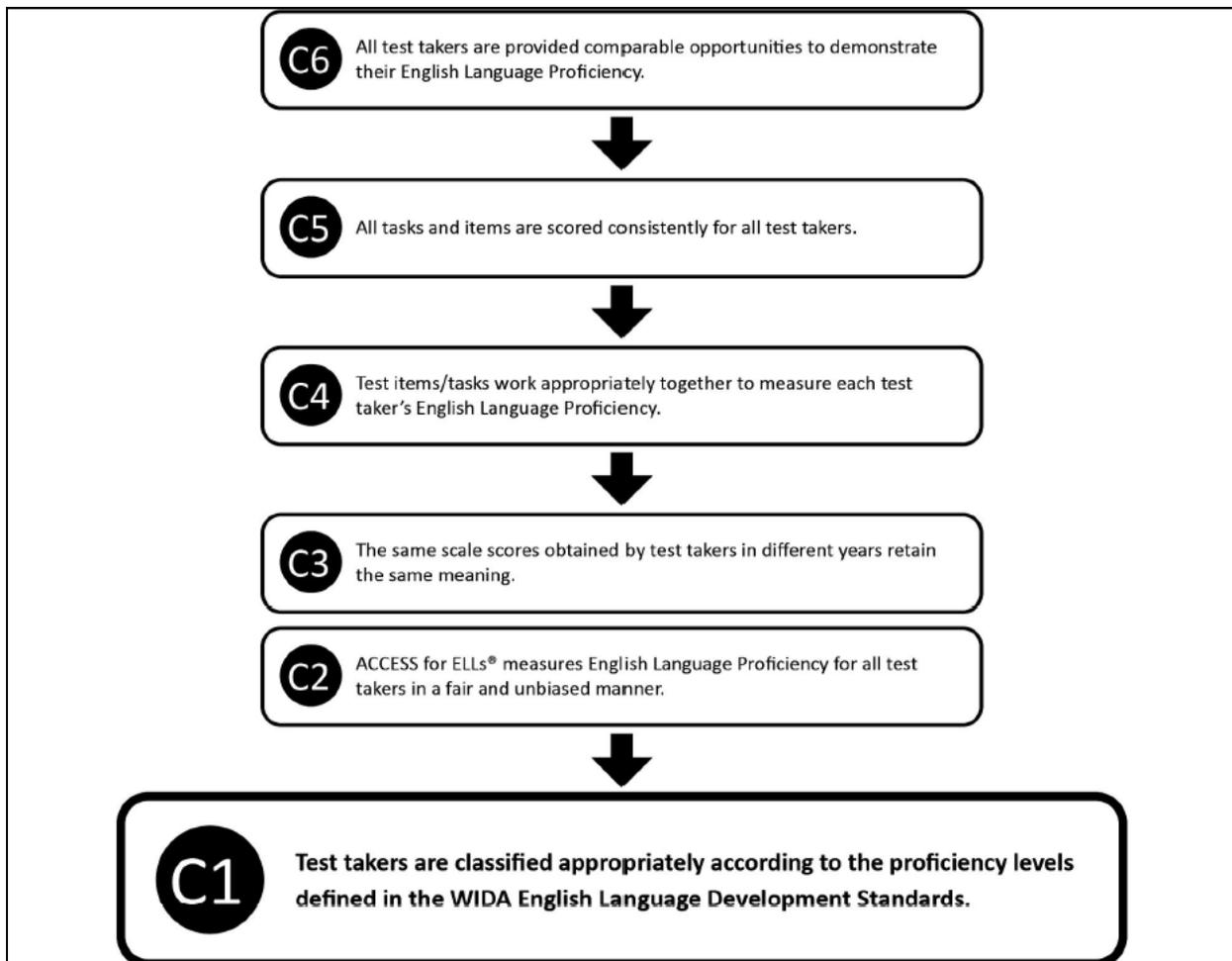
Figure 2.2.1A: Progression of Claims for Step 4: Assessment Records

2.3 Evidence for Assessment Records Claims of ACCESS for ELLs

In this section, evidence in the form of data or other sources (e.g., Test Administration Manuals, 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 Chapters 5 and 7 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 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 4.6a: Well-specified procedures were developed for test administrators so that they are able to administer the test consistently.

Evidence: Procedures for administering the test and producing reported scores are documented in the ACCESS for ELLs Test Administration Manual (WIDA, 2012a).

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

Evidence: Test administration procedures are documented in the ACCESS for ELLs Test Administration Manual (WIDA, 2012a).

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

Action 4.5a: Raters of performance-based tasks undergo thorough training so that they know how to score appropriately.

Evidence: Section 1.7 of this report specifies the scoring procedure for ACCESS for ELLs, with Section 1.7.2 providing information on the Writing domain and Section 1.7.3 explicating the procedure for Speaking tasks. Raters of Writing tasks are trained by MetriTech to follow the Writing rubric (see Table 1.7.2B). Since Speaking tasks are scored locally, raters are trained through an online program on the WIDA website to follow the Speaking rubric (see Table 1.7.3A).

Action 4.5b: Listening and Reading items are scored electronically using a carefully checked key.

Evidence: Section 1.7 of this report specifies the scoring procedure for ACCESS for ELLs. Listening and Reading items are dichotomous and are scored electronically by MetriTech (see Section 1.7.1).

Action 4.5c: Raters of performance-based tasks are certified, demonstrating that they can score appropriately.

Evidence: Section 1.7 of this report specifies the scoring procedure for ACCESS for ELLs. Writing tasks are centrally scored at MetriTech, and all raters are pre-screened and subsequently trained (see Section 1.7.2). Speaking is scored by the test administrator after the completion of training on test administration and on the Speaking rubric (see Section 1.7.3).

Action 4.5d: Raters of Writing tasks are monitored daily to ensure that they are scoring appropriately.

Evidence: MetriTech provides Raters of Writing tasks with specially prepared calibration sets each day to monitor that the scoring rubric is being applied consistently across scoring sessions (see Section 1.7.2.1).

Action 4.5e: Scoring data for Writing tasks are analyzed for rater agreement to understand how closely raters agree.

Evidence: Interrater reliability is calculated for each of the three or four Writing tasks. The percentage of agreement between two raters is calculated in terms of three features (i.e., Linguistic Complexity, Vocabulary Usage, and Language Control). When the two raters agree on a score, this is counted as exact agreement. If the two raters provide feature scores that differ by one point, this is counted as adjacent agreement (see Table 6F for percentages of exact and adjacent agreement).

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 6–8B), 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 and accuracy information based on Classical Test Theory is calculated for each test form (i.e., for each tier within each grade-level cluster). 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 6F).

Action 4.4b: For each domain and composite score across tiers, 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 across the three tiers for each domain. Cronbach's alpha indicates the extent to which items work together to measure the same construct. 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 8D 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 Properties table includes information on the Rasch fit statistics for each test item (see Table 6H). These statistics, called outfit mean square and infit mean square statistics, are calculated by comparing the observed empirical data with the values that the Rasch model expects test takers to produce. 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 can be skewed if test takers within range of the targeted proficiency level do not perform as expected. Outfit is not weighted and therefore is very sensitive to outliers. 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 diagnose 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.

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

Action 4.3a: A sufficient number of items and tasks are used as anchor items across adjacent years to maintain a consistent scale from year to year.

Evidence: Each year, while a certain percentage of items on each ACCESS for ELLs test form is refreshed, a number of items and tasks are retained from the previous year's assessment. These retained "anchor items" ensure that performances on the newer form may be interpreted in the same frame of reference as the previous year. For Listening and Reading, a majority of test items are anchor items, while one of three Writing tasks and one of three Speaking folders are retained annually as anchor tasks. Table 6E displays information on the anchor items for each test form.

Action 4.3b: New items and tasks are calibrated with anchor items to ensure that their difficulty measures are on the same consistent scale that is used from year to year.

Evidence: Both new and previously used items and tasks (i.e., anchor items) are included on each test form (see Table 6H for a list of new and anchored test items/tasks).

Action 4.3c: 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 ACCESS for ELLs scale score (see Table 6D). The same equation is used across all tiers and grade-level clusters within each domain.

Claim 4.2 - 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.

Evidence: The Item/Task 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 6G). 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 Properties table includes more detailed information on the DIF analyses, showing the degree of measurement bias for each item and which group is favored (ATR Table 6H). 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, and an item exhibiting C level DIF is considered to display evidence for potential bias and should be closely examined by test developers.

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 1.4.5 (*DIF Items*), ethnicity and gender DIF analyses are conducted using all test taker data. Information on DIF is gathered at different points in the testing cycle and is provided to the test development team. The test development team uses this information to guide the item development and review process for future items.

Claim 4.1 - Test takers are classified appropriately according to the 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 ACCESS for ELLs effectively measures the performance of test takers across the range of English Language Proficiency levels as defined by the WIDA English Language Development (ELD) Standards.

Evidence: The distribution of test takers' raw scores on ACCESS for ELLs, organized by individual test form (e.g., Reading 3–5B), shows the extent to which 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 ACCESS for ELLs, organized by test form (e.g., Reading 3–5B), shows that 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 ACCESS for ELLs, organized by individual test form (e.g., Reading 3–5B), shows that 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 6J). This distribution of scores shows that 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). Five vertical lines indicate the five cut scores for the highest grade in the cluster, dividing the figure into six

sections for each of the six 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 ACCESS for ELLs effectively measures the performance of test takers across the range of English Language Proficiency levels as defined by the WIDA English Language Development (ELD) Standards.

Evidence: The distribution of test takers' scale scores on ACCESS for ELLs, organized by grade-level cluster, shows that ACCESS for ELLs effectively measures the performance of test takers across the range of ELD abilities as described by the WIDA ELD Standards (see Table 8A; see Figure 8A).

The proficiency level distribution of test takers' scores on ACCESS for ELLs, organized by grade-level cluster, shows that ACCESS for ELLs effectively measures the performance of test takers across the range of proficiency levels as defined by the WIDA ELD Standards (see Table 8B; see Figure 8B).

The Test Characteristic Curve reflects test takers' mean raw scores by domain on ACCESS for ELLs across the entire test for Kindergarten and across the three tiers for the other grade-level clusters (see Figure 8C). It also graphically illustrates how the tiers differ in difficulty, showing that ACCESS for ELLs effectively captures a range of ELD ability levels. Tier A is represented by a dotted curve, Tier B by a light solid curve, and Tier C by a dark solid curve. As shown, Tier B is more difficult than Tier A, and Tier C is more difficult than Tier B.

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 pertinent to each tier.

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. Test forms for different tiers are designed to measure most accurately at certain proficiency levels (i.e., PL1 through PL3 for Tier A, PL2 through PL4 for Tier B, and PL3 and up for Tier C), and the expected peak of the distribution occurs within the desired range of the cut scores.

In the Raw Score to Scale Score Conversion Chart, the proficiency level associated with each raw score shows the distribution of proficiency level scores associated with each raw score/scale score for each grade in the cluster, along with the percentage of test takers in that grade who scored at that raw score/scale score/proficiency level score (see Table 6I). Additionally, this table 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: Across domains, analyses are run to confirm that English Language Proficiency is measured with high precision at the cut points pertinent to each tier.

Evidence: The conditional standard error of measurement provides information on how precisely test takers' performances on ACCESS for ELLs are measured at the cut points between language proficiency levels. These cut points are critical because they are the points at which decisions are made about test taker placements. Because the cut points depend on the grade level, information for each domain is provided for each grade level within the cluster. From Table 8C, it is possible to examine how well the different tiers measure the English Language Proficiency of test takers at the appropriate proficiency level cut scores (i.e., PL1 through PL3 for Tier A, PL2 through PL4 for Tier B, and PL3 and up for Tier C).

The Test Information Function reflects the precision of measurement by graphically presenting the standard error of measurement across tiers for grade-level clusters (see Figure 8D). Tier A is represented by a dotted curve, Tier B by a light solid curve, and Tier C by a dark solid curve. As shown, Tier B is more difficult than Tier A, and Tier C is more difficult than Tier B. As in Figure C, the cut scores at the highest grade in each cluster are indicated by vertical lines. These lines make it easy to see that the test forms for different tiers measure most accurately at the proficiency levels they are meant to capture.

Action 4.1e: 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 8E). 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 8E 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. Well-specified procedures were developed for test administrators so that they are able to administer the test consistently. b. Test administrators document and report any irregularities that may occur so that appropriate action may be taken.	a. Test Administration Manual b. Test Administration Manual
5. All items and tasks are scored consistently for all test takers.	a. Raters of performance-based tasks undergo thorough training so that they know how to score appropriately. b. Listening and Reading items are scored electronically onsite at MetriTech. c. Raters are of performance-based tasks are certified, demonstrating that they can score appropriately. d. Raters of Writing tasks are monitored daily to ensure that they are scoring appropriately. e. Scoring data for Writing tasks are analyzed for rater agreement to understand how closely raters agree.	a. Chapter 1.7.2 (*Scoring - Writing*); Chapter 1.7.3 (*Scoring - Speaking*) b. Chapter 1.7.1 (*Scoring - Listening and Reading*) c. Chapter 1.7.2 (*Scoring - Writing*); Chapter 1.7.3 (*Scoring - Speaking*) d. Chapter 1.7.2.1 (*Scoring Procedures for Writing*) e. Table 6F (*Reliability*)
4. Test items/tasks work appropriately together to measure each test taker's English Language Proficiency.	a. For each test form (e.g., Reading 6-8B), item and task analyses are performed and psychometric properties of the items and tasks are evaluated to confirm that scores are internally consistent. b. For each domain and composite score across tiers, item and task analyses are performed and psychometric properties of the items and tasks are evaluated to confirm that scores are internally consistent. c. Analyses of Rasch model fit statistics are conducted to show that individual tasks perform appropriately	a. Table 6F (*Reliability*) b. Table 8D (*Reliability*) c. Table 6H (*Complete Item Analysis*)

3. The same scale scores obtained by test takers in different years retain the same meaning.	a. A sufficient number of items and tasks are used as anchor items across adjacent years to maintain a consistent scale from year to year.	a. Table 6E (*Equating Summary*)
	b. New items and tasks are calibrated with anchor items to ensure that their difficulty measures are on the same consistent scale that is used from year to year.	b. Table 6D (*Scaling Equation*)
	c. The same scaling equation is applied from year to year to ensure that scale scores are obtained consistently over time.	c. Table 6H (*Complete Item Analysis*)
2. ACCESS for ELLs measures English Language Proficiency for all test takers in a fair and unbiased manner.	a. Differential Item Functioning (DIF) analyses are conducted to determine whether any items or tasks are biased against certain subgroups.	a. Table 6H (*Complete Item Analysis*); Table 6G (*Item/Task Analysis Summary*)
	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.	b. Chapter 1.4.5 (*DIF Items*)
1. Test takers are classified appropriately according to the proficiency levels defined in the WIDA English Language Development Standards.	a. Distributions of scale scores and proficiency levels for each domain are analyzed to confirm that ACCESS for ELLs effectively measures the performance of test takers across the range of English Language Proficiency levels as defined by the WIDA English Language Development Standards.	a. Figure 6A (*Raw Scores*) & Table 6A (*Raw Score Descriptive Statistics*); Figure 6B (*Scale Scores*) & Table 6B (*Scale Score Descriptive Statistics*); Figure 6C (*Proficiency Level*) & Table 6C (*Proficiency Level Distribution*); Table 6J (*Raw Score to Proficiency Level Score Conversion Chart*); Figure 6D (*Test Characteristic Curve*)
	b. Distributions of scale scores and proficiency levels, organized by grade-level cluster, are analyzed to confirm that ACCESS for ELLs effectively measures the performance of test takers across the range of English Language Proficiency levels as defined by the WIDA English Language Development Standards.	b. Figure 8A (*Scale Scores*) & Table 8A (*Scale Score Descriptive Statistics*); Figure 8B (*Proficiency Level*) & Table 8B (*Proficiency Level Distribution*); Figure 8C (*Test Characteristic Curve*)
	c. For each test form, analyses are run to confirm that English Language Proficiency is measured with high precision at the cut points pertinent to each tier.	c. Figure 6E (*Test Information Function*); Table 6I (*Raw Score to Scale Score Conversion Chart*)
	d. Across domains, analyses are run to confirm that English Language Proficiency is measured with high precision at the cut points pertinent to each tier.	d. Table 8C (*Conditional Standard Error of Measurement*) & Figure 8D (*Test Information Function*)
	e. Classification and accuracy analyses are conducted by grade-level to confirm that proficiency level classifications are reliable for all domain and composite scores.	e. Table 8E (*Accuracy and Consistency of Classification Indices*)

2.5 Visual Guide to Tables and Figures

This section provides navigational support for the tables and figures contained in the ACCESS for ELLs Annual Technical Report. The Visual Guide to Tables and Figures, shown in Figures 2.5.1 through 2.5.3, 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, tier, and demographic characteristics, such as state, gender, 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 4,6, and 8. Detailed descriptions of the information in each of the tables and figures is included in the preceding chapters (i.e., Chapter 5 contains information on tables and figures in Chapter 6, and Chapter 7 contains information on tables and figures in Chapter 8). 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, Domain or Tier 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 ACCESS for ELLs test form by grade-level cluster, tier, 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 9–12C in the Reading domain, refer to Figure 2.5.2 and complete the following steps: one, select Grade Cluster 9–12; two, select Tier C; three, select Reading under Domains. Information for 9–12C 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.

Figure 2.5.3 displays the sections in Chapter 8 that contain analyses across tiers, organized by grade-level cluster, domain composites, and domains. The key above the figure describes the specific information in each table and figure. For example, to find the Conditional Standard Error of Measurement table for Grade-level Cluster 6–8 in the Writing domain, refer to Figure 2.5.3 and complete the following steps: one, select Grade Cluster 6–8; two, select Domain; three, select Writing. Information for 6–8 Writing is shown in section 8.5.3. Finally, look at the key and find Conditional Standard of Error Measurement table. The result is 8.5.3C. Click on 8.5.3 to go to the appropriate section, and then locate Table C.

2.5.1. Chapter 4 Visual Guide to Tables and Figures

4.1	Participation
4.2	Scale Score Results
4.3	Proficiency Level Results
4.4	Results by Standard

Test Form Characteristics

		Grade		Grade-Level Cluster	
		Tier	Domain	Tier	Domain
Demographic Characteristics	State		4.1.2.1		4.1.1.1
	Gender		4.2.2.2 4.1.2.2	4.1.3.3	4.2.1.2 4.1.1.2
	Ethnicity and Race		4.2.2.3 4.1.2.3	4.1.3.4	4.2.1.3 4.1.1.3
Domain Composites	Overall	4.3.8.2	4.3.8.3	4.3.8.1	
	Oral Language	4.3.5.2	4.3.5.3	4.3.5.1	
	Literacy	4.3.6.2	4.3.6.3	4.3.6.1	
	Comprehension	4.3.7.2	4.3.7.3 4.4.2.1	4.3.7.1	4.4.1.1
Domains	Across All Domains	4.1.3.2	4.2.2.1	4.1.3.1	4.2.1.1 4.2.3
	Listening	4.3.1.2	4.3.1.3	4.3.1.1	
	Reading	4.3.2.2	4.3.2.3	4.3.2.1	
	Writing	4.3.3.2	4.3.3.3 4.4.2.2	4.3.3.1	4.4.2.1
	Speaking	4.3.4.2	4.3.4.3 4.4.3.2	4.3.4.1	4.4.3.1

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	*Raw Score Information*
Table B and Figure B	*Scale Score Information*
Table C and Figure C	*Proficiency and Distribution*
Table D	*Scaling Equation Table*
Table E	*Equating Summary*
Figure D	*Test Characteristic Curve*
Figure E	*Test Information Function*
Table F	*Reliability*
Table G	*Item/Task Analysis Summary*
Table H	*Complete Item Analysis Table*
Table I	*Complete Raw Score to Scale Score Conversion Table*
Table J	*Raw Score to Proficiency Level Score Conversion*

Domains

Grade-Level Cluster and Tier		Listening	Reading	Writing	Speaking
K		6.1.1	6.1.2	6.1.3	6.1.4
1-2	A	6.2.1.1	6.2.2.1	6.2.3.1	6.2.4
	B	6.2.1.2	6.2.2.2	6.2.3.2	
	C	6.2.1.3	6.2.2.3	6.2.3.3	
3-5	A	6.3.1.1	6.3.2.1	6.3.3.1	6.3.4
	B	6.3.1.2	6.3.2.2	6.3.3.2	
	C	6.3.1.3	6.3.2.3	6.3.3.3	
6-8	A	6.4.1.1	6.4.2.1	6.4.3.1	6.4.4
	B	6.4.1.2	6.4.2.2	6.4.3.2	
	C	6.4.1.3	6.4.2.3	6.4.3.3	
9-12	A	6.5.1.1	6.5.2.1	6.5.3.1	6.5.4
	B	6.5.1.2	6.5.2.2	6.5.3.2	
	C	6.5.1.3	6.5.2.3	6.5.3.3	

Figure 2.5.2 Chapter 6 Visual Guide to Tables and Figures

2.5.3. Chapter 8 Visual Guide to Tables and Figures

Table A and Figure A	Scale Score Information
Table B and Figure B	Proficiency Level Information
Table C and Figures C and D	Conditional Standard Error Measurement
Table D	Reliability Information
Table E	Accuracy and Consistency of Classification

Grade-Level Cluster

	K	1-2	3-5	6-8	9-12
Overall	8.1.8	8.2.8	8.3.8	8.4.8	8.5.8
Oral Language	8.1.5	8.2.5	8.3.5	8.4.5	8.5.5
Literacy	8.1.6	8.2.6	8.3.6	8.4.6	8.5.6
Comprehension	8.1.7	8.2.7	8.3.7	8.4.7	8.5.7
Listening	8.1.1	8.2.1	8.3.1	8.4.1	8.5.1
Reading	8.1.2	8.2.2	8.3.2	8.4.2	8.5.2
Writing	8.1.3	8.2.3	8.3.3	8.4.3	8.5.3
Speaking	8.1.4	8.2.4	8.3.4	8.4.4	8.5.4

Domain Composites: Overall, Oral Language, Literacy, Comprehension

Domains: Listening, Reading, Writing, Speaking

Figure 2.5.3 Chapter 8 Visual Guide to Tables and Figures

3.   Descriptions of Student Results

Chapter 3 provides a description of the tables that appear in Chapter 4.

3.1   Participation

Participation in ACCESS for ELLs is shown in three ways: grade-level cluster; grade, and tier.

3.1.1 Grade-Level Cluster

Section 4.1.1 gives information on participation by *grade-level cluster*.

Table 4.1.1.1 shows participation across the 31 WIDA states that participated in the operational testing program in 2012–2013. The first row shows the grade-level cluster, the next 31 rows show the number of students in that grade-level cluster who took the test, by state, and the final row shows the total across all 31 states.

Table 4.1.1.2 shows participation by cluster by gender across all 31 states combined, while Table 4.1.1.3 shows participation by cluster by ethnicity across all 31 states.

3.1.2 Grade

Section 4.1.2 gives similar data as in the previous section, but broken out by *grade* rather than by grade-level cluster.

3.1.3 Tier

Finally, Section 4.1.3 gives participation by *tier*.

Table 4.1.3.1 shows this information by cluster and domain. Because, for example, Listening in the 1–2 grade-level cluster for Tier A represents a specific test form, this table indicates how many students took each test form. Note that because Speaking is not administered by tiers, the total number shows how many took that cluster's Speaking test.

Table 4.1.3.2 shows the same information, but by grade rather than by grade-level cluster.

Table 4.1.3.3 shows the breakdown by cluster and tier for gender. When reviewing data on Differential Item Functioning (DIF) in Chapter 6, it may be useful to refer to these tables to understand the size of the comparison groups on each form.

Table 4.1.3.4 shows the same information for ethnicity (Hispanic vs. non-Hispanic). Consortium states use the Census Bureau categories for student ethnicity. Again, this data may be useful when reviewing analyses of DIF in Chapter 6.

Note that in some circumstances there was a mismatch between a student's reported grade and the reported cluster of the test the student took (for example, a student who was reported to be in Kindergarten but who was administered a test in the 1–2 grade-level cluster). In all, 349students were administered a test form from a cluster other than the grade in which they were reported to be. Table 3.1 below shows the number of students in each grade who were administered out-of-grade-level tests, and the test form that they were administered. The data for these students was eliminated from all analyses in this report.

Table 3.1
Students Excluded from Analysis due to Grade/Cluster Mismatch

Grade	Lower Test	N	Higher Test	N	Total
K			1-2A	52	85
			1-2B	14	
			1-2C	8	
			3-5A	3	
			3-5B	1	
			3-5C	2	
			6-8B	3	
			9-12A	1	
			9-12C	1	
1			3-5A	2	10
			3-5B	2	
			3-5C	3	
			6-8B	1	
			9-12B	2	
2			3-5A	15	49
			3-5B	18	
			3-5C	15	
			6-8B	1	
3	1-2A	20			77
	1-2B	39			
	1-2C	18			
			6-8A	1	2
			6-8C	1	
4	1-2A	5			8
	1-2B	1			
	1-2C	2			
5	1-2A	3			9
	1-2B	4			
	1-2C	2			
			6-8A	4	19
			6-8B	7	
			6-8C	8	
6	1-2A	1			36
	1-2C	1			
	3-5A	6			
	3-5B	15			
	3-5C	13			
			9-12A	1	1
7	1-2B	1			2
	3-5B	1			
			9-12A	1	6
			9-12B	3	
			9-12C	2	

8	1-2A	1			5
	3-5C	4			
			9-12A	3	23
			9-12B	12	
			9-12C	8	
9	6-8A	5			10
	6-8B	3			
	6-8C	2			
10	1-2A	3			5
	6-8A	1			
	6-8B	1			
11	1-2A	1			1
12	6-8A	1			1
Total	-	154	-	195	349

(Note that the apparent number of Kindergarten students reported as taking tests in the higher grade-level clusters is at least in part spurious. In some states, when a grade level has not been defined for a student before the identification labels for the operational test are sent out, the "Grade" field is filled in with a placeholder of 0, the same code that is used for Kindergarten. If that information is never updated, the grade for the operational data is recorded as Kindergarten. Thus, many of the students who are reported here as Kindergarten students taking tests from higher grade-level clusters may in fact be students for whom the grade level was never defined.)

3.2   Scale Score Results

3.2.1 Mean Scale Scores Across Domain and Composite Scores Section

4.2.1 shows mean (average) scale scores by *grade-level clusters* across the eight scores awarded on ACCESS, first for the four domains (Listening, Speaking, Reading, and Writing) and then for the four composites (Oral Language, Literacy, Comprehension, and Overall). In this section, under each average, the number of students in each group is also given.

Table 4.2.1.1 shows mean scale scores by grade-level cluster, while Table 4.2.1.2 shows the same information broken down by gender, and Table 4.2.1.3 shows the same information broken down by race and ethnicity. In 2010, the Census Bureau introduced a new approach to reporting race and ethnicity. Previously, race and ethnicity had been a single category with six values (Hispanic, Asian/Pacific Islander/Hawaiian, Black/African American, American Indian/Alaskan Native, White - Non Hispanic, and Multi-racial/Other). Under the new approach, ethnicity has become a binary category (Hispanic or non-Hispanic), with five categories for race (American Indian/Alaskan Native, Asian, Black/African American, Pacific Islander/Hawaiian, and White) that are not mutually exclusive. Thus, for example, Student A may be labeled as Hispanic for ethnicity and Asian for race, while Student B may be labeled as non-Hispanic for ethnicity and both American Indian/Alaskan Native and Black/African American for race. Starting with Series 202, students who were identified as non-Hispanic are included in the appropriate racial category (e.g., Asian) if they are identified in only one racial category; if they are identified in more than one racial category, they are labeled Multi-racial. Students who are labeled as Hispanic are included in the Hispanic category, regardless of how many racial categories that they are included in.

Section 4.2.2 shows the same scores broken down by *grade* rather than by grade-level cluster. Table 4.2.2.1 shows mean scale scores by grade, while Table 4.2.2.2 shows the same information broken down by gender, and Table 4.2.2.3 shows the same information broken down by ethnicity.

3.2.2 Correlations

Tables 4.2.3A through 4.2.3E show correlations among the four domain scale scores by grade-level clusters across all tiers, as well as the number of students included in each correlation. Table 4.2.3A shows the results for Kindergarten, Table 4.2.3B shows the results for the 1–2 grade-level cluster, Table 4.2.3C shows the results for the 3–5 grade-level cluster, Table 4.2.3D shows the results for the 6–8 grade-level cluster, and Table 4.2.3E shows the results for the 9–12 grade-level cluster. Beginning with Series 101, caps were placed on students taking Tier A and Tier B test forms in Listening and Reading. This capping of scores may raise the correlation between those two scores, while decreasing the correlation of those two scores with Speaking

and Writing. Note, all correlations in Tables 4.2.3A through 4.2.3E are significant at the 0.01 level (2-tailed).

3.3   Proficiency Level Results

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

4.3.1 – Listening

4.3.2 – Reading

4.3.3 – Writing

4.3.4 – Speaking

4.3.5 – Oral Language Composite

4.3.6 – Literacy Composite

4.3.7 – Comprehension Composite

4.3.8 – Overall Composite

Within each section, results are first presented by tier and *grade-level cluster* in Section 4.3.*.1 (note that the * indicates a subsection variable). Table 4.3.*.1A shows the number of students who were classified into each language proficiency level, while Table 4.3.*.1B shows the percent of students (within each row) classified into each language proficiency category. These tables clearly show the effect of the capping of scores on Tier A and Tier B for Listening and Reading.

Following the presentation by tier and cluster, results are presented by tier and *grade* in Section 4.3.*.2. Again, the first table in this section shows the number of students classified into each language proficiency level, while the second table shows the results in terms of percentages within each row.

Finally, in Section 4.3.*.3, results are presented by *grade alone*, that is, without the tiers. Again, the first table shows the number of students classified into each language proficiency level, while the second table shows the results in terms of percentages within each row.

3.4   Mean Raw Score Results by Standards

The tables in this section show information on mean raw score results by the five WIDA ELD Standards. These results are in terms of raw scores (i.e., the number of correct responses in Listening/Reading or Speaking or the points on the Writing rubric). Note that scores for Kindergarten students were not categorized by Standard; therefore, these tables include information only for grades 1–12.

3.4.1 Comprehension Composite

Section 4.4.1 shows the results for Comprehension (combined Listening and Reading items). The first section (4.4.1.1) shows results by *grade-level cluster,* while the second section (4.4.1.2) shows the results by *grade*. Within each table, the third column shows the Standard (Social and Instructional Language, Language of Language Arts, Language of Math, Language of Science,

and Language of Social Studies). The fourth column shows the maximum possible raw score by Standard, the fifth column shows the mean raw score, and the sixth column shows the mean raw score as a percentage of the maximum.

3.4.2 Writing

Section 4.4.2 shows the results for Writing. Again, the first section (4.4.2.1) shows results by *grade-level cluster,* while the second section (4.4.2.2) shows the results by *grade*. Within each table, the third column shows the Standard (Social and Instructional Language, Language of Language Arts/Social Studies, and Language of Mathematics/Science). The next three columns show the mean raw scores (out of a maximum of 6) of the three sub scores for the Writing test: Linguistic Complexity, Vocabulary Usage, and Language Control. The seventh column shows the total mean raw score for each Standard (out of a maximum of 18). The final column shows the mean raw score as a percentage of the maximum possible score.

3.4.3 Speaking

Finally, Section 4.4.3 presents the results for Speaking. As in the previous sections, the first section (4.4.3.1) shows results by *grade-level cluster*, while the second section (4.4.3.2) shows the results by *grade*. Note that the Speaking assessment itself is adaptive but not tiered. Student results are categorized here by tier according to the tier of the group-administered assessment that they took. Within each table, the third column shows the Standard (Social and Instructional Language, Language of Language Arts/Social Studies, and Language of Mathematics/Science). The fourth column shows the maximum possible score, the fifth columns shows the mean raw score, and the sixth column shows the mean raw score as a percentage of the maximum possible score.

4. Student Results

4.1 Participation

4.1.1 Participation by Grade-level Cluster

4.1.1.1 By State

Table 4.1.1.1

Participation by Cluster by State S301

| State | Cluster | | | | | |
	K	1-2	3-5	6-8	9-12	Total
AK	1,757	3,342	3,864	2,852	2,914	14,729
AL	3,827	5,763	3,840	2,396	1,890	17,716
CO	12,659	25,486	30,850	18,632	15,393	103,020
DC	1,079	1,651	1,197	963	987	5,877
DE	1,660	2,817	1,712	867	863	7,919
GA	18,259	31,222	21,560	10,031	7,834	88,906
HI	2,551	3,947	3,865	3,684	3,676	17,723
IL	31,343	59,986	46,571	23,915	17,897	179,712
KY	3,146	5,823	4,634	2,752	2,599	18,954
MA	9,413	17,452	18,923	12,490	13,123	71,401
MD	10,018	17,602	15,166	6,792	6,990	56,568
ME	520	971	1,196	1,191	1,298	5,176
MN	8,258	15,438	18,487	12,289	10,227	64,699
MO	4,478	7,338	6,771	4,274	3,555	26,416
MP	67	453	932	773	448	2,673
MS	1,291	2,080	2,234	1,226	913	7,744
MT	341	875	1,184	729	551	3,680
NC	14,034	26,403	26,079	17,307	13,484	97,307
ND	344	641	825	657	764	3,231
NH	439	1,116	1,046	776	962	4,339
NJ	11,690	17,762	12,121	7,717	10,311	59,601
NM	7,178	13,246	16,270	11,666	9,641	58,001
NV	9,826	19,390	21,722	10,404	7,030	68,372
OK	7,421	12,442	11,364	6,684	5,232	43,143
PA	4,760	10,669	11,853	10,121	11,983	49,386
RI	1,269	2,035	2,404	1,593	1,698	8,999
SD	672	1,048	1,242	938	931	4,831
VA	14,711	26,746	22,618	14,382	15,384	93,841
VT	186	371	362	254	346	1,519
WI	5,767	11,391	13,237	9,711	8,007	48,113
WY	411	737	788	498	385	2,819
Total	189,375	346,243	324,917	198,564	177,316	1,236,415

4.1.1.2 By Gender

Table 4.1.1.2
Participation by Cluster by Gender S301

Cluster		Gender			Total
		F	**M**	**Missing**	
K	Count	89,647	99,401	327	189,375
	% within Cluster	47.3%	52.5%	0.2%	100.0%
1-2	Count	164,213	181,608	422	346,243
	% within Cluster	47.4%	52.5%	0.1%	100.0%
3-5	Count	149,856	174,609	452	324,917
	% within Cluster	46.1%	53.7%	0.1%	100.0%
6-8	Count	88,602	109,585	377	198,564
	% within Cluster	44.6%	55.2%	0.2%	100.0%
9-12	Count	80,044	96,848	424	177,316
	% within Cluster	45.1%	54.6%	0.2%	100.0%
Total	Count	572,362	662,051	2,002	1,236,415
	% within Cluster	46.3%	53.5%	0.2%	100.0%

4.1.1.3 By Ethnicity

Table 4.1.1.3
Participation by Cluster by Ethnicity S301

Cluster		Hispanic/Non-Hispanic			Total
		Hispanic	**Other**	**Missing**	
K	Count	132,841	53,819	2,715	189,375
	% within Cluster	70.1%	28.4%	1.4%	100.0%
1-2	Count	244,853	97,602	3,788	346,243
	% within Cluster	70.7%	28.2%	1.1%	100.0%
3-5	Count	230,342	90,681	3,894	324,917
	% within Cluster	70.9%	27.9%	1.2%	100.0%
6-8	Count	134,209	61,476	2,879	198,564
	% within Cluster	67.6%	31.0%	1.4%	100.0%
9-12	Count	107,867	66,412	3,037	177,316
	% within Cluster	60.8%	37.5%	1.7%	100.0%
Total	Count	850,112	369,990	16,313	1,236,415
	% within Cluster	68.8%	29.9%	1.3%	100.0%

4.1.2 Participation by Grade

4.1.2.1 By State

Table 4.1.2.1

Participation by Grade by State S301

State	Grade													Total
	0	1	2	3	4	5	6	7	8	9	10	11	12	Total
AK	1,757	1,718	1,624	1,628	1,185	1,051	1,028	869	955	983	656	646	629	14,729
AL	3,827	3,118	2,645	2,025	1,032	783	735	852	809	829	466	349	246	17,716
CO	12,659	13,037	12,449	11,507	10,339	9,004	7,384	6,262	4,986	4,688	3,932	3,348	3,425	103,020
DC	1,079	936	715	549	361	287	319	332	312	529	187	137	134	5,877
DE	1,660	1,537	1,280	812	528	372	292	281	294	364	225	150	124	7,919
GA	18,259	16,642	14,580	11,017	5,993	4,550	3,641	3,381	3,009	3,661	2,047	1,267	859	88,906
HI	2,551	1,872	2,075	1,687	1,140	1,038	995	1,335	1,354	1,663	879	583	551	17,723
IL	31,343	30,431	29,555	24,037	12,972	9,562	8,299	8,256	7,360	8,156	4,363	3,252	2,126	179,712
KY	3,146	3,076	2,747	2,143	1,398	1,093	998	948	806	1,010	708	459	422	18,954
MA	9,413	9,257	8,195	7,146	6,406	5,371	4,610	4,174	3,706	4,496	3,460	2,960	2,207	71,401
MD	10,018	8,847	8,755	6,856	4,565	3,745	2,738	2,224	1,830	3,051	1,728	1,211	1,000	56,568
ME	520	477	494	464	375	357	416	405	370	406	295	281	316	5,176
MN	8,258	7,874	7,564	7,245	5,745	5,497	4,686	4,023	3,580	3,376	2,739	2,223	1,889	64,699
MO	4,478	3,903	3,435	2,719	2,227	1,825	1,540	1,497	1,237	1,387	896	683	589	26,416
MP	67	227	226	294	324	314	261	299	213	106	147	176	19	2,673
MS	1,291	1,114	966	1,000	677	557	475	403	348	378	263	181	91	7,744
MT	341	452	423	518	352	314	248	244	237	244	138	97	72	3,680
NC	14,034	13,418	12,985	11,385	7,620	7,074	5,809	5,742	5,756	6,611	3,166	2,064	1,643	97,307
ND	344	295	346	343	232	250	235	205	217	281	215	136	132	3,231
NH	439	524	592	485	293	268	227	269	280	424	235	176	127	4,339
NJ	11,690	9,833	7,929	5,773	3,607	2,741	2,444	2,633	2,640	3,508	2,950	2,233	1,620	59,601
NM	7,178	6,829	6,417	7,123	4,836	4,311	3,897	3,965	3,804	3,884	2,533	1,856	1,368	58,001
NV	9,826	9,959	9,431	8,782	7,388	5,552	4,379	3,429	2,596	2,072	1,869	1,735	1,354	68,372
OK	7,421	6,461	5,981	5,128	3,422	2,814	2,254	2,266	2,164	2,235	1,263	1,008	726	43,143
PA	4,760	5,344	5,325	4,589	3,803	3,461	3,379	3,324	3,418	3,785	3,168	2,672	2,358	49,386
RI	1,269	961	1,074	999	788	617	574	524	495	551	463	405	279	8,999
SD	672	518	530	532	366	344	329	324	285	356	228	168	179	4,831
VA	14,711	13,729	13,017	10,745	6,567	5,306	4,397	4,946	5,039	6,800	3,547	3,417	1,620	93,841
VT	186	175	196	172	112	78	84	84	86	130	81	69	66	1,519
WI	5,767	5,880	5,511	5,241	4,618	3,378	3,244	3,283	3,184	3,132	1,821	1,683	1,371	48,113
WY	411	343	394	396	210	182	174	174	150	181	87	68	49	2,819
Total	189,375	178,787	167,456	143,340	99,481	82,096	70,091	66,953	61,520	69,277	44,755	35,693	27,591	1,236,415

4.1.2.2 By Gender

Table 4.1.2.2

Participation by Grade by Gender S301

Grade		Gender			Total
		F	M	Missing	
K	Count	89,647	99,401	327	189,375
	% within Grade	47.3%	52.5%	0.2%	100.0%
1	Count	84,844	93,663	280	178,787
	% within Grade	47.5%	52.4%	0.2%	100.0%
2	Count	79,369	87,945	142	167,456
	% within Grade	47.4%	52.5%	0.1%	100.0%
3	Count	67,310	75,796	234	143,340
	% within Grade	47.0%	52.9%	0.2%	100.0%
4	Count	45,376	54,017	88	99,481
	% within Grade	45.6%	54.3%	0.1%	100.0%
5	Count	37,170	44,796	130	82,096
	% within Grade	45.3%	54.6%	0.2%	100.0%
6	Count	31,159	38,724	208	70,091
	% within Grade	44.5%	55.2%	0.3%	100.0%
7	Count	29,996	36,878	79	66,953
	% within Grade	44.8%	55.1%	0.1%	100.0%
8	Count	27,447	33,983	90	61,520
	% within Grade	44.6%	55.2%	0.1%	100.0%
9	Count	30,248	38,723	306	69,277
	% within Grade	43.7%	55.9%	0.4%	100.0%
10	Count	20,042	24,658	55	44,755
	% within Grade	44.8%	55.1%	0.1%	100.0%
11	Count	16,636	19,017	40	35,693
	% within Grade	46.6%	53.3%	0.1%	100.0%
12	Count	13,118	14,450	23	27,591
	% within Grade	47.5%	52.4%	0.1%	100.0%
Total	Count	572,362	662,051	2,002	1,236,415
	% within Grade	46.3%	53.5%	0.2%	100.0%

4.1.2.3 By Ethnicity

Table 4.1.2.3

Participation by Grade by Ethnicity S301

Grade		Hispanic/Non-Hispanic			Total
		Hispanic	Other	Missing	
K	Count	132,841	53,819	2,715	189,375
	% within Grade	70.1%	28.4%	1.4%	100.0%
1	Count	126,401	50,306	2,080	178,787
	% within Grade	70.7%	28.1%	1.2%	100.0%
2	Count	118,452	47,296	1,708	167,456
	% within Grade	70.7%	28.2%	1.0%	100.0%
3	Count	102,278	39,459	1,603	143,340
	% within Grade	71.4%	27.5%	1.1%	100.0%
4	Count	70,645	27,623	1,213	99,481
	% within Grade	71.0%	27.8%	1.2%	100.0%
5	Count	57,419	23,599	1,078	82,096
	% within Grade	69.9%	28.7%	1.3%	100.0%
6	Count	48,342	20,614	1,135	70,091
	% within Grade	69.0%	29.4%	1.6%	100.0%
7	Count	45,128	20,934	891	66,953
	% within Grade	67.4%	31.3%	1.3%	100.0%
8	Count	40,739	19,928	853	61,520
	% within Grade	66.2%	32.4%	1.4%	100.0%
9	Count	44,650	23,207	1,420	69,277
	% within Grade	64.5%	33.5%	2.0%	100.0%
10	Count	27,177	16,847	731	44,755
	% within Grade	60.7%	37.6%	1.6%	100.0%
11	Count	20,565	14,650	478	35,693
	% within Grade	57.6%	41.0%	1.3%	100.0%
12	Count	15,475	11,708	408	27,591
	% within Grade	56.1%	42.4%	1.5%	100.0%
Total	Count	850,112	369,990	16,313	1,236,415
	% within Grade	68.8%	29.9%	1.3%	100.0%

4.1.3 Participation by Tier

4.1.3.1 By Cluster by Domain (Test Form)

Table 4.1.3.1
Participation by Cluster by Tier by Domain S301

Cluster			Domain			
			Listening	Reading	Writing	Speaking
K	Tier	-	189,257	189,253	189,256	189,252
1-2	Tier	A	73,086	73,088	73,063	73,075
		B	188,331	188,330	188,218	188,225
		C	84,619	84,616	84,585	84,599
	Total		346,036	346,034	345,866	345,899
3-5	Tier	A	26,654	26,653	26,645	26,641
		B	141,544	141,536	141,481	141,490
		C	156,524	156,523	156,452	156,487
	Total		324,722	324,712	324,578	324,618
6-8	Tier	A	20,694	20,694	20,683	20,684
		B	77,582	77,579	77,468	77,533
		C	99,924	99,925	99,835	99,869
	Total		198,200	198,198	197,986	198,086
9-12	Tier	A	25,191	25,213	25,171	25,181
		B	69,112	69,163	69,123	69,074
		C	80,452	80,514	80,476	80,459
	Total		174,755	174,890	174,770	174,714

4.1.3.2 By Grade by Domain (Test Form)

Table 4.1.3.2
Participation by Grade by Tier by Domain S301

Grade			Domain			
			Listening	Reading	Writing	Speaking
K	Tier	-	189,257	189,253	189,256	189,252
1	Tier	A	56,649	56,651	56,636	56,641
		B	93,031	93,035	92,950	92,959
		C	28,989	28,988	28,976	28,984
	Total		178,669	178,674	178,562	178,584
2	Tier	A	16,437	16,437	16,427	16,434
		B	95,300	95,295	95,268	95,266
		C	55,630	55,628	55,609	55,615
	Total		167,367	167,360	167,304	167,315
3	Tier	A	11,694	11,694	11,685	11,687
		B	68,310	68,308	68,286	68,286
		C	63,261	63,262	63,240	63,250
	Total		143,265	143,264	143,211	143,223
4	Tier	A	8,052	8,050	8,052	8,048
		B	40,810	40,808	40,788	40,799
		C	50,556	50,552	50,529	50,540
	Total		99,418	99,410	99,369	99,387
5	Tier	A	6,908	6,909	6,908	6,906
		B	32,424	32,420	32,407	32,405
		C	42,707	42,709	42,683	42,697
	Total		82,039	82,038	81,998	82,008
6	Tier	A	7,221	7,222	7,217	7,218
		B	27,143	27,143	27,118	27,130
		C	35,639	35,634	35,623	35,618
	Total		70,003	69,999	69,958	69,966
7	Tier	A	6,837	6,836	6,835	6,833
		B	26,669	26,669	26,621	26,650
		C	33,321	33,318	33,287	33,301
	Total		66,827	66,823	66,743	66,784
8	Tier	A	6,636	6,636	6,631	6,633
		B	23,770	23,767	23,729	23,753
		C	30,964	30,973	30,925	30,950
	Total		61,370	61,376	61,285	61,336
9	Tier	A	12,903	12,908	12,883	12,895
		B	25,682	25,708	25,669	25,659
		C	29,875	29,891	29,841	29,879
	Total		68,460	68,507	68,393	68,433
10	Tier	A	6,363	6,368	6,361	6,357
		B	18,032	18,044	18,064	18,027
		C	19,809	19,828	19,829	19,809
	Total		44,204	44,240	44,254	44,193
11	Tier	A	3,849	3,855	3,854	3,855
		B	14,669	14,666	14,661	14,653
		C	16,643	16,666	16,660	16,652
	Total		35,161	35,187	35,175	35,160
12	Tier	A	2,076	2,082	2,073	2,074
		B	10,729	10,745	10,729	10,735
		C	14,125	14,129	14,146	14,119
	Total		26,930	26,956	26,948	26,928

4.1.3.3 By Cluster by Gender

Table 4.1.3.3

Participation by Cluster by Tier by Gender S301

Cluster	Tier		Gender			Total
			F	M	Missing	
K	-	Count	89,647	99,401	327	189,375
		% within Tier	47.3%	52.5%	0.2%	100.0%
1-2	A	Count	33,339	39,673	137	73,149
		% within Tier	45.6%	54.2%	0.2%	100.0%
	B	Count	88,663	99,590	202	188,455
		% within Tier	47.0%	52.8%	0.1%	100.0%
	C	Count	42,211	42,345	83	84,639
		% within Tier	49.9%	50.0%	0.1%	100.0%
3-5	A	Count	11,971	14,624	91	26,686
		% within Tier	44.9%	54.8%	0.3%	100.0%
	B	Count	63,269	78,176	199	141,644
		% within Tier	44.7%	55.2%	0.1%	100.0%
	C	Count	74,616	81,809	162	156,587
		% within Tier	47.7%	52.2%	0.1%	100.0%
6-8	A	Count	9,235	11,427	69	20,731
		% within Tier	44.5%	55.1%	0.3%	100.0%
	B	Count	33,321	44,276	165	77,762
		% within Tier	42.8%	56.9%	0.2%	100.0%
	C	Count	46,046	53,882	143	100,071
		% within Tier	46.0%	53.8%	0.1%	100.0%
9-12	A	Count	11,248	14,089	125	25,462
		% within Tier	44.2%	55.3%	0.5%	100.0%
	B	Count	31,448	38,506	180	70,134
		% within Tier	44.8%	54.9%	0.3%	100.0%
	C	Count	37,348	44,253	119	81,720
		% within Tier	45.7%	54.2%	0.1%	100.0%

4.1.3.4 By Cluster by Ethnicity

Table 4.1.3.4

Participation by Cluster by Tier by Ethnicity S301

| Cluster | Tier | | Hispanic/Non-Hispanic | | | Total |
			Hispanic	Other	Missing	
K	-	Count	132,841	53,819	2,715	189,375
		% within Tier	70.1%	28.4%	1.4%	100.0%
1-2	A	Count	53,285	18,807	1,057	73,149
		% within Tier	72.8%	25.7%	1.4%	100.0%
	B	Count	137,806	48,620	2,029	188,455
		% within Tier	73.1%	25.8%	1.1%	100.0%
	C	Count	53,762	30,175	702	84,639
		% within Tier	63.5%	35.7%	0.8%	100.0%
3-5	A	Count	16,061	9,732	893	26,686
		% within Tier	60.2%	36.5%	3.3%	100.0%
	B	Count	102,189	37,689	1,766	141,644
		% within Tier	72.1%	26.6%	1.2%	100.0%
	C	Count	112,092	43,260	1,235	156,587
		% within Tier	71.6%	27.6%	0.8%	100.0%
6-8	A	Count	12,247	7,741	743	20,731
		% within Tier	59.1%	37.3%	3.6%	100.0%
	B	Count	51,266	25,260	1,236	77,762
		% within Tier	65.9%	32.5%	1.6%	100.0%
	C	Count	70,696	28,475	900	100,071
		% within Tier	70.6%	28.5%	0.9%	100.0%
9-12	A	Count	15,320	9,184	958	25,462
		% within Tier	60.2%	36.1%	3.8%	100.0%
	B	Count	40,849	28,022	1,263	70,134
		% within Tier	58.2%	40.0%	1.8%	100.0%
	C	Count	51,698	29,206	816	81,720
		% within Tier	63.3%	35.7%	1.0%	100.0%

4.2 Scale Score Results

4.2.1 Mean Scale Scores by Grade-level Cluster Across Domain and Composite Scores

4.2.1.1 By Cluster

Table 4.2.1.1

Mean Scale Scores by Cluster S301

Cluster		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Mean	268.32	192.12	209.91	299.82	284.29	201.26	214.98	225.97
	N	188,955	188,946	188,946	188,946	188,946	188,940	188,938	188,926
1-2	Mean	312.45	296.29	277.11	348.60	330.79	286.95	301.21	299.90
	N	345,570	345,323	345,264	345,409	345,266	345,103	345,265	344,872
3-5	Mean	357.32	338.56	347.68	362.66	360.27	343.38	344.32	348.25
	N	324,096	323,496	323,491	323,993	323,801	323,250	323,428	322,971
6-8	Mean	379.11	357.82	357.53	372.29	375.97	357.95	364.31	363.18
	N	197,648	197,263	197,174	197,463	197,236	197,020	197,162	196,608
9-12	Mean	386.48	374.51	398.92	382.47	384.74	387.00	378.24	386.17
	N	173,540	173,452	173,150	173,608	172,527	172,840	173,052	171,739

4.2.1.2 By Cluster by Gender

Table 4.2.1.2

Mean Scale Scores by Cluster by Gender S301

Cluster	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	F	Mean	272.05	195.35	215.41	303.26	287.87	205.63	218.35	230.09
		N	89,450	89,450	89,449	89,448	89,448	89,447	89,447	89,443
	M	Mean	265.03	189.27	205.01	296.80	281.15	197.39	212.00	222.31
		N	99,184	99,175	99,176	99,177	99,177	99,172	99,170	99,162
	Missing	Mean	245.54	174.04	190.53	272.28	259.07	182.53	195.50	205.31
		N	321	321	321	321	321	321	321	321
1-2	F	Mean	313.36	297.53	279.72	350.41	332.15	288.88	302.35	301.65
		N	163,919	163,832	163,797	163,842	163,774	163,731	163,812	163,624
	M	Mean	311.66	295.19	274.77	346.99	329.59	285.24	300.20	298.34
		N	181,232	181,072	181,048	181,148	181,074	180,954	181,034	180,831
	Missing	Mean	299.48	283.66	265.03	335.25	317.62	274.61	288.47	287.26
		N	419	419	419	419	418	418	419	417
3-5	F	Mean	357.37	339.52	351.05	362.73	360.33	345.54	345.00	349.78
		N	149,518	149,303	149,311	149,476	149,402	149,213	149,277	149,098
	M	Mean	357.30	337.75	344.81	362.61	360.24	341.54	343.74	346.96
		N	174,132	173,747	173,735	174,071	173,954	173,592	173,705	173,429
	Missing	Mean	350.07	333.20	335.84	356.16	353.43	334.77	338.38	340.17
		N	446	446	445	446	445	445	446	444
6-8	F	Mean	379.16	359.15	361.59	370.47	375.09	360.64	365.25	364.79
		N	88,244	88,118	88,084	88,149	88,068	88,033	88,084	87,867
	M	Mean	379.12	356.77	354.29	373.77	376.72	355.80	363.58	361.91
		N	109,028	108,769	108,714	108,938	108,792	108,611	108,702	108,365
	Missing	Mean	363.63	351.68	345.85	367.17	365.64	349.02	355.36	353.76
		N	376	376	376	376	376	376	376	376
9-12	F	Mean	386.69	375.96	402.44	380.88	384.06	389.49	379.31	387.70
		N	78,457	78,450	78,352	78,478	78,024	78,220	78,266	77,714
	M	Mean	386.40	373.35	396.06	383.88	385.41	384.99	377.40	384.97
		N	94,676	94,595	94,391	94,720	94,102	94,213	94,379	93,624
	Missing	Mean	367.93	365.42	384.52	358.53	362.90	375.24	366.26	371.52
		N	407	407	407	410	401	407	407	401

4.2.1.3 By Cluster by Ethnicity

Table 4.2.1.3

Mean Scale Scores by Cluster by Ethnicity S301

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Non-Hispanic Asian	Mean	278.65	220.80	234.67	307.65	293.38	228.01	238.14	247.40
		N	24,291	24,289	24,290	24,290	24,290	24,288	24,288	24,287
	Non-Hispanic Pacific Islander	Mean	252.65	178.24	196.20	295.38	274.25	187.46	200.55	213.30
		N	1,671	1,671	1,671	1,671	1,671	1,671	1,671	1,671
	Non-Hispanic Black	Mean	279.57	209.44	225.43	319.49	299.75	217.69	230.47	242.09
		N	8,335	8,334	8,334	8,335	8,335	8,334	8,334	8,334
	Hispanic (Of Any Race)	Mean	264.24	185.00	203.55	295.80	280.25	194.52	208.77	220.04
		N	132,639	132,632	132,633	132,633	132,633	132,629	132,626	132,619
	Non-Hispanic American Indian	Mean	270.20	179.95	191.48	296.63	283.65	185.96	207.02	215.07
		N	3,482	3,482	3,480	3,482	3,482	3,480	3,482	3,480
	Non-Hispanic Multi-racial	Mean	301.96	220.59	235.34	330.79	316.60	228.24	244.98	254.52
		N	772	772	772	772	772	772	772	772
	Non-Hispanic White	Mean	283.47	203.49	223.74	313.30	298.61	213.87	227.47	239.09
		N	14,566	14,566	14,566	14,564	14,564	14,566	14,566	14,564
	Missing	Mean	258.73	186.38	203.19	292.41	275.77	195.02	208.08	219.06
		N	3,199	3,200	3,200	3,199	3,199	3,200	3,199	3,199
1-2	Non-Hispanic Asian	Mean	318.83	303.97	285.34	350.98	335.17	294.90	308.50	306.79
		N	43,744	43,715	43,718	43,725	43,707	43,701	43,704	43,665
	Non-Hispanic Pacific Islander	Mean	304.94	292.87	277.96	340.97	323.23	285.69	296.59	296.75
		N	2,899	2,899	2,897	2,898	2,897	2,896	2,896	2,893
	Non-Hispanic Black	Mean	313.22	297.95	278.56	354.32	334.04	288.51	302.60	301.97
		N	15,903	15,886	15,883	15,910	15,896	15,874	15,882	15,868
	Hispanic (Of Any Race)	Mean	310.94	294.60	275.32	347.12	329.30	285.22	299.58	298.23
		N	244,554	244,377	244,375	244,463	244,396	244,273	244,350	244,133
	Non-Hispanic American Indian	Mean	306.17	290.91	269.54	341.31	324.09	280.56	295.58	293.45
		N	6,567	6,561	6,500	6,516	6,493	6,488	6,556	6,467
	Non-Hispanic Multi-racial	Mean	325.68	305.54	283.97	365.40	345.76	295.02	311.64	310.05
		N	1,310	1,309	1,309	1,309	1,309	1,308	1,309	1,307
	Non-Hispanic White	Mean	318.40	300.39	281.31	358.14	338.53	291.10	305.86	305.14
		N	25,697	25,683	25,693	25,691	25,680	25,677	25,676	25,658
	Missing	Mean	306.22	291.53	273.55	342.25	324.50	282.78	296.00	295.08
		N	4,896	4,893	4,889	4,897	4,888	4,886	4,892	4,881

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
3-5	Non-Hispanic Asian	Mean	361.86	344.18	352.32	358.80	360.60	348.51	349.61	351.93
		N	39,889	39,869	39,854	39,881	39,857	39,842	39,861	39,815
	Non-Hispanic Pacific Islander	Mean	352.61	336.15	349.31	356.27	354.69	342.97	341.20	346.31
		N	3,052	3,051	3,051	3,048	3,047	3,048	3,048	3,043
	Non-Hispanic Black	Mean	356.78	337.62	346.33	362.78	360.07	342.23	343.49	347.40
		N	17,656	17,601	17,610	17,658	17,646	17,592	17,596	17,579
	Hispanic (Of Any Race)	Mean	356.83	337.72	347.16	363.42	360.40	342.70	343.58	347.81
		N	229,956	229,472	229,505	229,893	229,771	229,329	229,438	229,155
	Non-Hispanic American Indian	Mean	348.70	331.70	338.95	353.49	351.40	335.62	336.98	340.24
		N	8,105	8,087	8,066	8,087	8,067	8,048	8,079	8,016
	Non-Hispanic Multi-racial	Mean	365.18	345.51	352.03	371.96	368.80	348.94	351.54	354.62
		N	1,049	1,049	1,044	1,046	1,046	1,044	1,049	1,041
	Non-Hispanic White	Mean	360.32	342.01	350.33	368.23	364.55	346.44	347.64	351.67
		N	19,707	19,692	19,689	19,696	19,692	19,679	19,684	19,660
	Missing	Mean	348.51	332.65	340.42	352.20	350.67	336.81	337.55	340.80
		N	4,682	4,675	4,672	4,684	4,675	4,668	4,673	4,662
6-8	Non-Hispanic Asian	Mean	381.86	364.14	362.74	366.65	374.50	363.71	369.56	366.75
		N	26,785	26,783	26,778	26,770	26,753	26,767	26,766	26,726
	Non-Hispanic Pacific Islander	Mean	371.10	354.70	356.57	367.25	369.46	355.96	359.75	359.86
		N	2,646	2,641	2,637	2,638	2,635	2,634	2,639	2,627
	Non-Hispanic Black	Mean	376.22	356.99	356.26	371.59	374.18	356.89	362.90	361.93
		N	13,405	13,355	13,354	13,401	13,386	13,347	13,346	13,322
	Hispanic (Of Any Race)	Mean	379.20	356.63	356.62	373.65	376.69	356.89	363.50	362.66
		N	133,695	133,384	133,322	133,571	133,433	133,212	133,334	132,965
	Non-Hispanic American Indian	Mean	371.24	352.57	353.12	366.13	369.03	353.15	358.30	357.79
		N	6,262	6,260	6,247	6,237	6,202	6,231	6,246	6,173
	Non-Hispanic Multi-racial	Mean	388.09	365.09	363.25	385.14	386.90	364.33	372.05	370.92
		N	526	526	522	525	525	522	526	521
	Non-Hispanic White	Mean	384.38	362.62	361.59	378.18	381.55	362.37	369.26	367.96
		N	11,209	11,195	11,196	11,200	11,190	11,189	11,189	11,166
	Missing	Mean	366.02	352.88	351.68	358.67	362.74	352.53	356.93	355.40
		N	3,120	3,119	3,118	3,121	3,112	3,118	3,116	3,108

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
9-12	Non-Hispanic Asian	Mean	389.99	379.08	401.77	376.71	383.62	390.69	382.48	388.41
		N	30,124	30,136	30,112	30,136	30,024	30,087	30,085	29,957
	Non-Hispanic Pacific Islander	Mean	384.97	370.71	399.63	388.76	387.08	385.48	375.15	385.88
		N	2,073	2,080	2,067	2,077	2,061	2,063	2,070	2,046
	Non-Hispanic Black	Mean	384.67	372.59	397.24	380.80	382.98	385.18	376.35	384.35
		N	16,315	16,302	16,292	16,294	16,211	16,262	16,267	16,138
	Hispanic (Of Any Race)	Mean	385.16	373.30	398.17	383.74	384.73	386.03	376.99	385.49
		N	105,511	105,419	105,204	105,622	104,886	104,997	105,160	104,333
	Non-Hispanic American Indian	Mean	391.45	371.38	399.33	386.90	389.45	385.65	377.55	386.63
		N	5,407	5,396	5,373	5,368	5,319	5,352	5,386	5,281
	Non-Hispanic Multi-racial	Mean	399.47	384.27	406.79	394.40	397.16	395.80	388.88	395.95
		N	426	427	427	427	426	427	426	426
	Non-Hispanic White	Mean	395.25	380.81	403.45	390.56	393.21	392.41	385.27	392.49
		N	10,505	10,515	10,507	10,513	10,456	10,494	10,487	10,430
	Missing	Mean	368.16	367.04	387.98	363.16	365.73	377.84	367.46	374.03
		N	3,179	3,177	3,168	3,171	3,144	3,158	3,171	3,128

4.2.2 Mean Scale Scores by Grade Across Domain and Composite Scores

4.2.2.1 By Grade

Table 4.2.2.1

Mean Scale Scores by Grade S301

Grade		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Mean	268.32	192.12	209.91	299.82	284.29	201.26	214.98	225.97
	N	188,955	188,946	188,946	188,946	188,946	188,940	188,938	188,926
1	Mean	299.11	283.24	268.01	339.34	319.44	275.90	288.02	288.75
	N	178,422	178,300	178,273	178,328	178,240	178,166	178,263	178,034
2	Mean	326.70	310.21	286.82	358.48	342.90	298.75	315.29	311.79
	N	167,148	167,023	166,991	167,081	167,026	166,937	167,002	166,838
3	Mean	347.56	330.21	341.84	360.05	354.10	336.29	335.56	341.42
	N	143,026	142,788	142,788	142,994	142,919	142,686	142,762	142,581
4	Mean	360.57	341.10	349.44	362.98	362.03	345.52	347.07	350.30
	N	99,216	99,019	99,012	99,160	99,112	98,945	98,993	98,843
5	Mean	370.44	350.10	355.77	366.81	368.91	353.18	356.28	357.71
	N	81,854	81,689	81,691	81,839	81,770	81,619	81,673	81,547
6	Mean	373.22	350.93	352.65	369.64	371.69	352.06	357.68	357.76
	N	69,808	69,668	69,645	69,761	69,689	69,595	69,644	69,485
7	Mean	379.67	358.27	358.06	372.53	376.35	358.42	364.81	363.65
	N	66,637	66,504	66,472	66,563	66,492	66,419	66,470	66,273
8	Mean	385.21	365.21	362.53	375.05	380.43	364.15	371.33	368.87
	N	61,203	61,091	61,057	61,139	61,055	61,006	61,048	60,850
9	Mean	381.50	371.82	396.56	379.87	380.94	384.48	374.84	383.24
	N	68,157	68,155	68,021	68,095	67,779	67,928	68,023	67,535
10	Mean	386.20	373.86	398.42	381.18	383.96	386.43	377.70	385.55
	N	43,971	43,950	43,878	44,028	43,759	43,806	43,857	43,567
11	Mean	391.56	377.49	401.75	385.04	388.58	389.89	381.83	389.35
	N	34,928	34,904	34,857	34,960	34,727	34,789	34,826	34,569
12	Mean	393.09	378.60	402.09	387.87	390.78	390.64	383.16	390.61
	N	26,484	26,443	26,394	26,525	26,262	26,317	26,346	26,068

4.2.2.2 By Grade by Gender

Scale	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	F	Mean	272.05	195.35	215.41	303.26	287.87	205.63	218.35	230.09
		N	89,450	89,450	89,449	89,448	89,448	89,447	89,447	89,443
	M	Mean	265.03	189.27	205.01	296.80	281.15	197.39	212.00	222.31
		N	99,184	99,175	99,176	99,177	99,177	99,172	99,170	99,162
	Missing	Mean	245.54	174.04	190.53	272.28	259.07	182.53	195.50	205.31
		N	321	321	321	321	321	321	321	321
1	F	Mean	300.02	284.34	270.52	341.30	320.88	277.70	289.05	290.45
		N	84,682	84,632	84,616	84,635	84,595	84,567	84,620	84,510
	M	Mean	298.29	282.27	265.75	337.59	318.17	274.28	287.09	287.24
		N	93,462	93,390	93,379	93,414	93,367	93,322	93,365	93,247
	Missing	Mean	294.29	276.43	260.23	327.22	311.01	268.58	281.83	281.12
		N	278	278	278	279	278	277	278	277
2	F	Mean	327.62	311.62	289.56	360.14	344.20	300.83	316.55	313.63
		N	79,237	79,200	79,181	79,207	79,179	79,164	79,192	79,114
	M	Mean	325.89	308.95	284.37	356.99	341.75	296.90	314.16	310.15
		N	87,770	87,682	87,669	87,734	87,707	87,632	87,669	87,584
	Missing	Mean	309.72	297.91	274.49	351.26	330.76	286.45	301.56	299.41
		N	141	141	141	140	140	141	141	140
3	F	Mean	348.02	331.24	345.30	360.58	354.59	338.54	336.42	343.13
		N	67,171	67,082	67,090	67,158	67,127	67,046	67,070	67,001
	M	Mean	347.17	329.30	338.79	359.61	353.69	334.31	334.81	339.92
		N	75,625	75,476	75,469	75,607	75,563	75,411	75,462	75,352
	Missing	Mean	342.03	326.92	333.04	352.69	347.66	330.23	331.60	335.26
		N	230	230	229	229	229	229	230	228
4	F	Mean	360.65	342.15	352.95	363.05	362.09	347.80	347.82	351.91
		N	45,267	45,200	45,187	45,239	45,222	45,166	45,191	45,121
	M	Mean	360.52	340.22	346.50	362.94	361.98	343.62	346.45	348.96
		N	53,861	53,731	53,737	53,833	53,802	53,691	53,714	53,634
	Missing	Mean	356.65	336.01	338.65	356.85	356.98	337.59	342.28	343.16
		N	88	88	88	88	88	88	88	88
5	F	Mean	370.31	351.34	359.16	366.23	368.55	355.49	357.11	359.22
		N	37,080	37,021	37,034	37,079	37,053	37,001	37,016	36,976
	M	Mean	370.57	349.09	352.99	367.31	369.23	351.29	355.62	356.48
		N	44,646	44,540	44,529	44,631	44,589	44,490	44,529	44,443
	Missing	Mean	359.98	342.57	338.91	361.85	361.32	340.95	347.88	346.87
		N	128	128	128	129	128	128	128	128

Scale	Gender			List	Read	Writ	Spek	Oral	Litr	Cphn	Over
6	F	Mean		373.49	352.25	356.93	367.89	370.96	354.86	358.69	359.49
		N		31,036	31,002	30,993	31,017	30,988	30,977	30,992	30,930
	M	Mean		373.09	349.89	349.27	371.11	372.37	349.86	356.92	356.42
		N		38,565	38,459	38,445	38,537	38,494	38,411	38,445	38,348
	Missing	Mean		355.48	344.90	340.12	357.14	356.55	342.77	348.13	346.66
		N		207	207	207	207	207	207	207	207
7	F	Mean		379.83	359.68	362.17	370.66	375.50	361.19	365.84	365.31
		N		29,880	29,836	29,816	29,839	29,819	29,804	29,823	29,744
	M	Mean		379.55	357.11	354.72	374.03	377.05	356.18	363.98	362.30
		N		36,678	36,589	36,577	36,645	36,594	36,536	36,568	36,450
	Missing	Mean		373.65	358.85	350.70	379.73	376.87	354.99	363.41	361.34
		N		79	79	79	79	79	79	79	79
8	F	Mean		384.85	366.41	366.26	373.21	379.32	366.61	372.06	370.24
		N		27,328	27,280	27,275	27,293	27,261	27,252	27,269	27,193
	M	Mean		385.54	364.24	359.53	376.53	381.33	362.17	370.76	367.77
		N		33,785	33,721	33,692	33,756	33,704	33,664	33,689	33,567
	Missing	Mean		373.59	361.00	354.80	379.20	376.70	358.18	364.91	363.42
		N		90	90	90	90	90	90	90	90
9	F	Mean		382.31	373.65	400.80	379.40	381.13	387.52	376.36	385.41
		N		29,808	29,821	29,778	29,783	29,650	29,735	29,764	29,565
	M	Mean		381.03	370.45	393.36	380.47	380.99	382.20	373.74	381.66
		N		38,051	38,036	37,945	38,014	37,834	37,895	37,961	37,675
	Missing	Mean		361.92	363.07	380.33	350.58	356.27	371.97	362.82	367.15
		N		298	298	298	298	295	298	298	295
10	F	Mean		385.57	374.90	401.52	378.78	382.44	388.51	378.22	386.53
		N		19,714	19,716	19,697	19,736	19,628	19,669	19,676	19,558
	M	Mean		386.74	373.02	395.90	383.15	385.23	384.76	377.29	384.77
		N		24,205	24,182	24,129	24,240	24,082	24,085	24,129	23,960
	Missing	Mean		373.46	364.87	391.25	370.67	370.22	378.29	367.52	376.76
		N		52	52	52	52	49	52	52	49
11	F	Mean		391.16	378.40	404.55	382.39	387.07	391.73	382.36	390.18
		N		16,312	16,312	16,290	16,326	16,226	16,263	16,271	16,160
	M	Mean		391.91	376.70	399.30	387.40	389.91	388.29	381.38	388.63
		N		18,579	18,555	18,530	18,595	18,464	18,489	18,518	18,372
	Missing	Mean		387.59	370.54	397.89	375.00	381.49	384.54	375.70	383.38
		N		37	37	37	39	37	37	37	37
12	F	Mean		392.99	379.93	405.06	385.69	389.61	392.78	384.09	391.77
		N		12,623	12,601	12,587	12,633	12,520	12,553	12,555	12,431
	M	Mean		393.16	377.37	399.37	389.81	391.82	388.68	382.30	389.53
		N		13,841	13,822	13,787	13,871	13,722	13,744	13,771	13,617
	Missing	Mean		406.65	392.25	404.80	410.57	408.45	398.80	396.70	401.35
		N		20	20	20	21	20	20	20	20

4.2.2.3 By Grade by Ethnicity

Table 4.2.2.3

Mean Scale Scores by Grade by Ethnicity S301

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Non-Hispanic Asian	Mean	278.65	220.80	234.67	307.65	293.38	228.01	238.14	247.40
		N	24,291	24,289	24,290	24,290	24,290	24,288	24,288	24,287
	Non-Hispanic Pacific Islander	Mean	252.65	178.24	196.20	295.38	274.25	187.46	200.55	213.30
		N	1,671	1,671	1,671	1,671	1,671	1,671	1,671	1,671
	Non-Hispanic Black	Mean	279.57	209.44	225.43	319.49	299.75	217.69	230.47	242.09
		N	8,335	8,334	8,334	8,335	8,335	8,334	8,334	8,334
	Hispanic (Of Any Race)	Mean	264.24	185.00	203.55	295.80	280.25	194.52	208.77	220.04
		N	132,639	132,632	132,633	132,633	132,633	132,629	132,626	132,619
	Non-Hispanic American Indian	Mean	270.20	179.95	191.48	296.63	283.65	185.96	207.02	215.07
		N	3,482	3,482	3,480	3,482	3,482	3,480	3,482	3,480
	Non-Hispanic Multi-racial	Mean	301.96	220.59	235.34	330.79	316.60	228.24	244.98	254.52
		N	772	772	772	772	772	772	772	772
	Non-Hispanic White	Mean	283.47	203.49	223.74	313.30	298.61	213.87	227.47	239.09
		N	14,566	14,566	14,566	14,564	14,564	14,566	14,566	14,564
	Missing	Mean	258.73	186.38	203.19	292.41	275.77	195.02	208.08	219.06
		N	3,199	3,200	3,200	3,199	3,199	3,200	3,199	3,199
1	Non-Hispanic Asian	Mean	305.35	290.52	276.83	343.50	324.65	283.93	294.97	295.96
		N	22,545	22,528	22,533	22,539	22,527	22,519	22,521	22,500
	Non-Hispanic Pacific Islander	Mean	291.62	279.00	268.74	330.33	311.22	274.17	282.84	285.07
		N	1,461	1,460	1,459	1,461	1,460	1,458	1,458	1,457
	Non-Hispanic Black	Mean	300.53	285.24	270.05	347.57	324.25	277.92	289.85	291.64
		N	7,913	7,902	7,905	7,918	7,911	7,898	7,898	7,893
	Hispanic (Of Any Race)	Mean	297.53	281.69	266.04	337.16	317.57	274.13	286.45	286.95
		N	126,247	126,165	126,181	126,205	126,161	126,113	126,150	126,036
	Non-Hispanic American Indian	Mean	294.04	278.50	260.48	332.37	313.43	269.75	283.18	282.69
		N	3,350	3,350	3,295	3,300	3,290	3,290	3,346	3,275
	Non-Hispanic Multi-racial	Mean	313.39	292.01	276.13	360.31	336.97	284.31	298.39	299.89
		N	675	675	674	674	674	674	675	673
	Non-Hispanic White	Mean	304.57	286.33	272.30	350.33	327.66	279.58	291.82	293.82
		N	13,532	13,522	13,529	13,530	13,522	13,519	13,518	13,508
	Missing	Mean	295.70	280.09	265.73	335.30	315.74	273.16	284.79	285.73
		N	2,699	2,698	2,697	2,701	2,695	2,695	2,697	2,692

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
2	Non-Hispanic Asian	Mean	333.18	318.27	294.38	358.94	346.36	306.56	322.88	318.31
		N	21,199	21,187	21,185	21,186	21,180	21,182	21,183	21,165
	Non-Hispanic Pacific Islander	Mean	318.47	306.95	287.32	351.79	335.42	297.38	310.54	308.61
		N	1,438	1,439	1,438	1,437	1,437	1,438	1,438	1,436
	Non-Hispanic Black	Mean	325.80	310.52	286.99	361.01	343.74	298.98	315.22	312.20
		N	7,990	7,984	7,978	7,992	7,985	7,976	7,984	7,975
	Hispanic (Of Any Race)	Mean	325.26	308.39	285.24	357.74	341.81	297.05	313.58	310.27
		N	118,307	118,212	118,194	118,258	118,235	118,160	118,200	118,097
	Non-Hispanic American Indian	Mean	318.81	303.87	278.85	350.48	335.04	291.68	308.50	304.49
		N	3,217	3,211	3,205	3,216	3,203	3,198	3,210	3,192
	Non-Hispanic Multi-racial	Mean	338.74	319.94	292.29	370.81	355.09	306.41	325.76	320.82
		N	635	634	635	635	635	634	634	634
	Non-Hispanic White	Mean	333.78	316.02	291.33	366.83	350.61	303.91	321.47	317.72
		N	12,165	12,161	12,164	12,161	12,158	12,158	12,158	12,150
	Missing	Mean	319.14	305.59	283.16	350.80	335.26	294.62	309.78	306.59
		N	2,197	2,195	2,192	2,196	2,193	2,191	2,195	2,189
3	Non-Hispanic Asian	Mean	353.34	336.21	348.09	358.15	356.03	342.41	341.50	346.27
		N	17,499	17,491	17,487	17,497	17,488	17,485	17,490	17,476
	Non-Hispanic Pacific Islander	Mean	343.06	328.03	343.60	353.36	348.49	336.08	332.66	339.58
		N	1,267	1,266	1,267	1,266	1,266	1,266	1,266	1,265
	Non-Hispanic Black	Mean	346.78	328.76	340.69	360.96	354.18	335.00	334.30	340.56
		N	7,094	7,081	7,080	7,097	7,092	7,075	7,078	7,070
	Hispanic (Of Any Race)	Mean	346.65	329.20	340.87	360.12	353.68	335.30	334.58	340.60
		N	102159	101965	101978	102135	102090	101902	101952	101838
	Non-Hispanic American Indian	Mean	337.26	322.65	330.50	350.27	344.10	326.92	327.30	331.98
		N	3,222	3,207	3,205	3,217	3,210	3,191	3,203	3,179
	Non-Hispanic Multi-racial	Mean	359.61	339.07	346.97	372.80	366.46	343.25	345.40	349.96
		N	532	532	530	531	531	530	532	529
	Non-Hispanic White	Mean	352.09	334.36	346.05	367.38	360.03	340.48	339.84	346.14
		N	9,239	9,236	9,231	9,235	9,231	9,228	9,232	9,218
	Missing	Mean	341.61	325.92	336.84	352.66	347.46	331.64	330.79	336.22
		N	2,014	2,010	2,010	2,016	2,011	2,009	2,009	2,006
4	Non-Hispanic Asian	Mean	364.27	346.22	353.05	358.47	361.60	349.90	351.77	353.24
		N	12,220	12,215	12,206	12,215	12,208	12,201	12,211	12,191
	Non-Hispanic Pacific Islander	Mean	354.17	337.04	349.58	354.95	354.78	343.57	342.31	346.82
		N	919	921	920	919	918	919	918	917
	Non-Hispanic Black	Mean	358.77	338.99	347.13	362.51	360.90	343.30	345.09	348.44
		N	5,527	5,501	5,507	5,526	5,523	5,500	5,499	5,495
	Hispanic (Of Any Race)	Mean	360.49	340.53	349.23	364.05	362.52	345.13	346.65	350.17
		N	70,512	70,353	70,364	70,472	70,444	70,319	70,342	70,251
	Non-Hispanic American Indian	Mean	349.46	332.64	340.53	352.79	351.36	336.85	337.85	341.08
		N	2,460	2,460	2,447	2,452	2,446	2,443	2,456	2,431
	Non-Hispanic Multi-racial	Mean	367.47	348.29	355.82	370.80	369.24	352.30	354.18	357.15
		N	285	285	285	284	284	285	285	284
	Non-Hispanic White	Mean	363.79	345.10	352.24	368.28	366.28	348.93	350.85	353.94
		N	5,853	5,846	5,848	5,852	5,852	5,844	5,844	5,841
	Missing	Mean	348.89	332.69	340.08	350.24	349.91	336.66	337.68	340.48
		N	1,440	1,438	1,435	1,440	1,437	1,434	1,438	1,433

Grade	Ethnicity			List	Read	Writ	Spek	Oral	Litr	Cphn	Over
5	Non-Hispanic Asian	Mean		373.61	355.46	358.73	360.31	367.24	357.35	360.97	360.11
		N		10,170	10,163	10,161	10,169	10,161	10,156	10,160	10,148
	Non-Hispanic Pacific Islander	Mean		364.92	347.09	357.40	361.96	363.70	352.46	352.53	355.65
		N		866	864	864	863	863	863	864	861
	Non-Hispanic Black	Mean		368.67	348.62	353.41	365.64	367.46	351.26	354.71	355.91
		N		5,035	5,019	5,023	5,035	5,031	5,017	5,019	5,014
	Hispanic (Of Any Race)	Mean		370.48	349.47	355.85	368.51	369.78	352.91	355.86	357.78
		N		57,285	57,154	57,163	57,286	57,237	57,108	57,144	57,066
	Non-Hispanic American Indian	Mean		363.15	342.75	348.56	358.49	361.16	345.89	348.93	350.30
		N		2,423	2,420	2,414	2,418	2,411	2,414	2,420	2,406
	Non-Hispanic Multi-racial	Mean		375.13	356.88	359.03	371.45	373.65	357.92	362.37	362.27
		N		232	232	229	231	231	229	232	228
	Non-Hispanic White	Mean		372.39	353.42	356.48	369.85	371.41	355.21	359.21	359.88
		N		4,615	4,610	4,610	4,609	4,609	4,607	4,608	4,601
	Missing	Mean		359.40	343.63	346.68	353.73	356.83	345.46	348.47	348.68
		N		1,228	1,227	1,227	1,228	1,227	1,225	1,226	1,223
6	Non-Hispanic Asian	Mean		374.90	356.15	358.33	363.28	369.32	357.51	361.83	360.83
		N		8,817	8,817	8,813	8,811	8,805	8,810	8,812	8,796
	Non-Hispanic Pacific Islander	Mean		365.86	347.71	350.95	363.76	365.11	349.59	353.25	354.08
		N		810	809	809	810	808	808	808	806
	Non-Hispanic Black	Mean		370.33	350.35	351.40	369.66	370.26	351.15	356.43	356.71
		N		4,457	4,446	4,447	4,459	4,455	4,443	4,444	4,439
	Hispanic (Of Any Race)	Mean		373.69	350.05	351.73	371.11	372.66	351.16	357.21	357.42
		N		48,186	48,066	48,052	48,156	48,108	48,014	48,055	47,946
	Non-Hispanic American Indian	Mean		364.07	345.77	347.60	360.85	362.84	347.01	351.36	351.67
		N		2,186	2,181	2,179	2,174	2,169	2,175	2,179	2,163
	Non-Hispanic Multi-racial	Mean		379.63	355.95	358.62	383.55	381.79	357.28	363.11	364.30
		N		199	199	197	199	199	197	199	197
	Non-Hispanic White	Mean		377.08	354.92	356.79	375.02	376.34	356.12	361.63	362.00
		N		3,925	3,922	3,920	3,922	3,918	3,920	3,920	3,912
	Missing	Mean		360.75	347.62	348.31	357.44	359.59	348.23	351.64	351.43
		N		1,228	1,228	1,228	1,230	1,227	1,228	1,227	1,226
7	Non-Hispanic Asian	Mean		382.24	364.38	362.72	366.86	374.80	363.81	369.87	366.92
		N		9,040	9,039	9,039	9,037	9,030	9,036	9,033	9,023
	Non-Hispanic Pacific Islander	Mean		370.77	354.29	356.54	365.15	368.25	355.73	359.41	359.34
		N		978	975	973	974	974	972	975	971
	Non-Hispanic Black	Mean		377.70	357.71	356.66	372.05	375.13	357.43	363.86	362.61
		N		4,635	4,617	4,613	4,630	4,627	4,612	4,615	4,604
	Hispanic (Of Any Race)	Mean		379.74	357.08	357.30	373.92	377.09	357.45	364.00	363.18
		N		44,964	44,857	44,836	44,920	44,876	44,797	44,841	44,712
	Non-Hispanic American Indian	Mean		371.75	352.97	353.23	366.50	369.44	353.38	358.77	358.10
		N		2,106	2,106	2,097	2,093	2,080	2,093	2,101	2,067
	Non-Hispanic Multi-racial	Mean		390.51	367.16	364.11	388.64	389.78	365.89	374.24	372.88
		N		166	166	165	166	166	165	166	165
	Non-Hispanic White	Mean		384.67	363.20	362.08	378.44	381.78	362.91	369.78	368.40
		N		3,789	3,784	3,789	3,783	3,782	3,784	3,780	3,774
	Missing	Mean		366.48	353.45	351.67	357.28	362.12	352.79	357.49	355.41
		N		959	960	960	960	957	960	959	957

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
8	Non-Hispanic Asian	Mean	388.36	371.80	367.11	369.76	379.32	369.73	376.88	372.41
		N	8,928	8,927	8,926	8,922	8,918	8,921	8,921	8,907
	Non-Hispanic Pacific Islander	Mean	376.41	361.77	361.91	372.96	374.98	362.24	366.26	365.92
		N	858	857	855	854	853	854	856	850
	Non-Hispanic Black	Mean	380.71	363.11	360.85	373.09	377.21	362.24	368.58	366.60
		N	4,313	4,292	4,294	4,312	4,304	4,292	4,287	4,279
	Hispanic (Of Any Race)	Mean	385.15	363.95	361.66	376.36	381.04	363.08	370.42	368.31
		N	40,545	40,461	40,434	40,495	40,449	40,401	40,438	40,307
	Non-Hispanic American Indian	Mean	378.66	359.66	359.12	371.57	375.46	359.73	365.49	364.27
		N	1,970	1,973	1,971	1,970	1,953	1,963	1,966	1,943
	Non-Hispanic Multi-racial	Mean	396.04	374.24	368.06	383.50	390.25	371.40	380.84	377.09
		N	161	161	160	160	160	160	161	159
	Non-Hispanic White	Mean	392.28	370.64	366.46	381.45	387.16	368.82	377.27	374.17
		N	3,495	3,489	3,487	3,495	3,490	3,485	3,489	3,480
	Missing	Mean	372.49	359.21	356.14	361.74	367.56	357.93	363.33	360.64
		N	933	931	930	931	928	930	930	925
9	Non-Hispanic Asian	Mean	386.35	377.35	400.12	375.67	381.28	389.01	380.16	386.50
		N	10,247	10,248	10,237	10,239	10,214	10,230	10,239	10,195
	Non-Hispanic Pacific Islander	Mean	382.57	367.78	397.57	389.76	386.40	383.02	372.37	383.90
		N	917	922	914	916	910	913	917	905
	Non-Hispanic Black	Mean	377.27	369.40	393.43	376.18	376.94	381.69	371.81	380.02
		N	5,393	5,395	5,398	5,378	5,356	5,387	5,385	5,343
	Hispanic (Of Any Race)	Mean	380.66	370.97	396.21	381.04	381.11	383.88	374.00	382.88
		N	43,928	43,913	43,814	43,907	43,690	43,755	43,824	43,511
	Non-Hispanic American Indian	Mean	386.96	366.55	396.45	386.45	387.02	381.81	372.85	383.21
		N	2,050	2,050	2,036	2,030	2,015	2,029	2,044	2,003
	Non-Hispanic Multi-racial	Mean	396.93	383.17	405.21	394.35	395.78	394.44	387.28	394.61
		N	185	186	186	186	185	186	185	185
	Non-Hispanic White	Mean	389.78	377.68	401.07	387.02	388.70	389.67	381.41	389.18
		N	3973	3975	3977	3976	3954	3970	3968	3947
	Missing	Mean	355.79	360.16	379.81	351.07	353.56	370.26	358.89	364.98
		N	1,464	1,466	1,459	1,463	1,455	1,458	1,461	1,446
10	Non-Hispanic Asian	Mean	387.83	377.45	400.55	373.46	380.88	389.27	380.71	386.59
		N	7,744	7,749	7,743	7,752	7,728	7,739	7,735	7,713
	Non-Hispanic Pacific Islander	Mean	383.42	370.65	398.92	387.82	385.80	385.07	374.68	385.29
		N	559	560	557	564	559	557	557	554
	Non-Hispanic Black	Mean	385.50	372.73	397.00	380.53	383.29	385.13	376.68	384.39
		N	4,131	4,127	4,122	4,139	4,115	4,116	4,120	4,095
	Hispanic (Of Any Race)	Mean	384.99	372.63	397.76	382.65	384.12	385.49	376.48	384.95
		N	26,695	26,669	26,619	26,723	26,543	26,569	26,612	26,411
	Non-Hispanic American Indian	Mean	392.40	371.66	398.66	384.61	388.67	385.46	377.99	386.22
		N	1,331	1,330	1,329	1,332	1,317	1,323	1,326	1,308
	Non-Hispanic Multi-racial	Mean	398.26	381.01	405.63	389.82	394.30	393.59	386.31	393.55
		N	104	104	104	104	104	104	104	104
	Non-Hispanic White	Mean	394.76	379.92	402.42	389.79	392.54	391.45	384.48	391.60
		N	2,658	2,663	2,657	2,668	2,654	2,655	2,655	2,646
	Missing	Mean	374.95	370.37	391.99	369.24	372.21	381.67	371.83	378.73
		N	749	748	747	746	739	743	748	736

Grade	Ethnicity			List	Read	Writ	Spek	Oral	Litr	Cphn	Over
11	Non-Hispanic Asian	Mean		393.59	381.29	403.79	378.28	386.21	392.79	385.07	390.61
		N		6,836	6,841	6,837	6,842	6,816	6,833	6,829	6,805
	Non-Hispanic Pacific Islander	Mean		390.60	376.51	403.18	384.91	387.91	390.04	380.83	389.28
		N		334	334	335	337	332	333	333	330
	Non-Hispanic Black	Mean		389.76	374.52	400.34	382.40	386.33	387.68	379.25	387.09
		N		3,707	3,700	3,696	3,694	3,681	3,692	3,694	3,664
	Hispanic (Of Any Race)	Mean		390.28	376.17	400.87	386.93	388.89	388.80	380.53	388.69
		N		20,078	20,058	20,026	20,133	19,966	19,976	20,005	19,855
	Non-Hispanic American Indian	Mean		395.73	374.90	402.33	389.65	392.95	388.90	381.25	389.98
		N		1,108	1,106	1,098	1,101	1,090	1,094	1,106	1,079
	Non-Hispanic Multi-racial	Mean		402.34	389.42	407.06	399.00	400.96	398.51	393.37	398.97
		N		71	71	71	71	71	71	71	71
	Non-Hispanic White	Mean		400.37	384.31	406.63	393.20	397.16	395.71	389.24	396.08
		N		2,263	2,264	2,264	2,260	2,251	2,261	2,259	2,245
	Missing	Mean		379.05	374.56	394.96	372.73	376.00	385.07	376.00	382.19
		N		531	530	530	522	520	529	529	520
12	Non-Hispanic Asian	Mean		395.56	381.96	404.12	381.46	388.85	393.33	386.23	391.91
		N		5,297	5,298	5,295	5,303	5,266	5,285	5,282	5,244
	Non-Hispanic Pacific Islander	Mean		389.50	373.74	403.80	392.29	391.13	389.13	378.65	389.73
		N		263	264	261	260	260	260	263	257
	Non-Hispanic Black	Mean		390.38	375.67	400.53	387.31	389.08	388.39	380.35	388.59
		N		3,084	3,080	3,076	3,083	3,059	3,067	3,068	3,036
	Hispanic (Of Any Race)	Mean		391.90	377.51	401.10	389.35	390.94	389.60	382.03	389.92
		N		14,810	14,779	14,745	14,859	14,687	14,697	14,719	14,556
	Non-Hispanic American Indian	Mean		394.93	377.55	403.14	387.97	391.80	390.60	382.99	390.87
		N		918	910	910	905	897	906	910	891
	Non-Hispanic Multi-racial	Mean		405.42	386.94	412.77	396.82	401.42	400.18	392.58	400.24
		N		66	66	66	66	66	66	66	66
	Non-Hispanic White	Mean		402.38	385.10	406.60	396.90	399.92	396.11	390.51	397.13
		N		1,611	1,613	1,609	1,609	1,597	1,608	1,605	1,592
	Missing	Mean		384.83	375.36	400.10	381.69	383.34	388.07	378.42	386.64
		N		435	433	432	440	430	428	433	426

4.2.3 Correlations among Scale Scores by Grade-level Cluster

Table 4.2.3A

Correlations Among Scale Scores: K S301

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.542	.558	.783
	N	188,955	188,938	188,937	188,946
Reading	Pearson Correlation		1	.723	.506
	N		188,946	188,940	188,931
Writing	Pearson Correlation			1	.558
	N			188,946	188,931
Speaking	Pearson Correlation				1
	N				188,946

Table 4.2.3B

Correlations Among Scale Scores: 1-2 S301

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.698	.583	.465
	N	345,570	345,265	345,183	345,266
Reading	Pearson Correlation		1	.665	.449
	N		345,323	345,103	345,017
Writing	Pearson Correlation			1	.487
	N			345,264	345,060
Speaking	Pearson Correlation				1
	N				345,409

Table 4.2.3C

Correlations Among Scale Scores: 3-5 S301

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.709	.583	.449
	N	324,096	323,428	323,380	323,801
Reading	Pearson Correlation		1	.664	.455
	N		323,496	323,250	323,191
Writing	Pearson Correlation			1	.481
	N			323,491	323,241
Speaking	Pearson Correlation				1
	N				323,993

Table 4.2.3D

Correlations Among Scale Scores: 6-8 S301

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.664	.510	.527
	N	197,648	197,162	197,034	197,236
Reading	Pearson Correlation		1	.587	.471
	N		197,263	197,020	196,809
Writing	Pearson Correlation			1	.439
	N			197,174	196,776
Speaking	Pearson Correlation				1
	N				197,463

Table 4.2.3E

Correlations Among Scale Scores: 9-12 S301

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.693	.636	.575
	N	173,540	173,052	172,589	172,527
Reading	Pearson Correlation		1	.675	.511
	N		173,452	172,840	172,366
Writing	Pearson Correlation			1	.561
	N			173,150	172,177
Speaking	Pearson Correlation				1
	N				173,608

Note: all correlations are significant at the 0.01 level (2-tailed).

4.3 Proficiency Level Results

4.3.1 Listening

4.3.1.1 By Cluster by Tier

Table 4.3.1.1A

Proficiency Level by Cluster by Tier (Count): Listening S301

| Cluster | Tier | Listening Proficiency Range | | | | | | Total |
		1	2	3	4	5	6	
K (instructional)	-	23,601	11,113	19,918	30,728	57,659	45,936	188,955
K (accountability)	-	48,827	19,817	16,716	11,228	29,459	62,908	188,955
1-2	A	4,656	7,104	15,516	45,631	n/a	n/a	72,907
	B	1,147	5,331	22,321	21,919	137,407	n/a	188,125
	C	368	1,877	12,555	6,254	20,340	43,144	84,538
3-5	A	1,863	5,169	6,553	12,859	n/a	n/a	26,444
	B	508	4,451	17,489	24,934	93,921	n/a	141,303
	C	322	561	10,872	21,435	56,480	66,679	156,349
6-8	A	3,802	7,473	4,098	5,159	n/a	n/a	20,532
	B	1,324	9,750	21,525	19,616	25,163	n/a	77,378
	C	369	879	6,663	29,477	34,139	28,211	99,738
9-12	A	11,237	9,183	2,354	2,185	n/a	n/a	24,959
	B	2,184	7,401	15,932	20,224	22,966	n/a	68,707
	C	1,153	3,581	14,280	25,066	18,952	16,842	79,874

Table 4.3.1.1B

Proficiency Level by Cluster by Tier (Percent): Listening S301

| Cluster | Tier | Listening Proficiency Range | | | | | | Total |
		1	2	3	4	5	6	
K (instructional)	-	12.5%	5.9%	10.5%	16.3%	30.5%	24.3%	100.0%
K (accountability)	-	25.8%	10.5%	8.8%	5.9%	15.6%	33.3%	100.0%
1-2	A	6.4%	9.7%	21.3%	62.6%	n/a	n/a	100.0%
	B	0.6%	2.8%	11.9%	11.7%	73.0%	n/a	100.0%
	C	0.4%	2.2%	14.9%	7.4%	24.1%	51.0%	100.0%
3-5	A	7.0%	19.5%	24.8%	48.6%	n/a	n/a	100.0%
	B	0.4%	3.1%	12.4%	17.6%	66.5%	n/a	100.0%
	C	0.2%	0.4%	7.0%	13.7%	36.1%	42.6%	100.0%
6-8	A	18.5%	36.4%	20.0%	25.1%	n/a	n/a	100.0%
	B	1.7%	12.6%	27.8%	25.4%	32.5%	n/a	100.0%
	C	0.4%	0.9%	6.7%	29.6%	34.2%	28.3%	100.0%
9-12	A	45.0%	36.8%	9.4%	8.8%	n/a	n/a	100.0%
	B	3.2%	10.8%	23.2%	29.4%	33.4%	n/a	100.0%
	C	1.4%	4.5%	17.9%	31.4%	23.7%	21.1%	100.0%

4.3.1.2 By Grade by Tier

Table 4.3.1.2A

Proficiency Level by Grade by Tier (Count): Listening S301

Grade	Tier	Listening Proficiency Range						Total
		1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	-	23,601	11,113	19,918	30,728	57,659	45,936	188,955
K (accountability)	-	48,827	19,817	16,716	11,228	29,459	62,908	188,955
1	A	3,386	4,044	12,627	36,480	n/a	n/a	56,537
	B	647	3,382	13,264	7,829	67,809	n/a	92,931
	C	236	694	5,225	2,297	8,614	11,888	28,954
2	A	1,270	3,060	2,889	9,151	n/a	n/a	16,370
	B	500	1,949	9,057	14,090	69,598	n/a	95,194
	C	132	1,183	7,330	3,957	11,726	31,256	55,584
3	A	703	1,760	2,815	6,332	n/a	n/a	11,610
	B	231	1,181	7,071	11,022	48,723	n/a	68,228
	C	103	149	2,353	6,540	21,600	32,443	63,188
4	A	524	1,718	2,000	3,738	n/a	n/a	7,980
	B	153	1,760	5,926	7,873	25,025	n/a	40,737
	C	115	192	4,367	7,872	18,815	19,138	50,499
5	A	636	1,691	1,738	2,789	n/a	n/a	6,854
	B	124	1,510	4,492	6,039	20,173	n/a	32,338
	C	104	220	4,152	7,023	16,065	15,098	42,662
6	A	848	2,645	1,476	2,203	n/a	n/a	7,172
	B	300	2,720	6,738	6,648	10,654	n/a	27,060
	C	109	386	1,689	9,641	15,380	8,371	35,576
7	A	1,210	2,766	1,308	1,494	n/a	n/a	6,778
	B	427	3,228	7,412	6,803	8,736	n/a	26,606
	C	134	180	2,229	9,772	9,828	11,110	33,253
8	A	1,744	2,062	1,314	1,462	n/a	n/a	6,582
	B	597	3,802	7,375	6,165	5,773	n/a	23,712
	C	126	313	2,745	10,064	8,931	8,730	30,909
9	A	5,460	5,178	905	1,290	n/a	n/a	12,833
	B	387	2,195	4,978	8,870	9,156	n/a	25,586
	C	250	1,174	4,751	9,014	6,572	7,977	29,738
10	A	2,681	2,240	972	414	n/a	n/a	6,307
	B	384	2,708	4,506	4,158	6,190	n/a	17,946
	C	210	755	3,298	6,086	6,128	3,241	19,718
11	A	1,888	1,304	300	311	n/a	n/a	3,803
	B	630	1,585	3,700	3,445	5,229	n/a	14,589
	C	258	963	3,394	5,409	3,395	3,117	16,536
12	A	1,208	461	177	170	n/a	n/a	2,016
	B	783	913	2,748	3,751	2,391	n/a	10,586
	C	435	689	2,837	4,557	2,857	2,507	13,882

Table 4.3.1.2B
Proficiency Level by Grade by Tier (Percent): Listening S301

Grade	Tier	Listening Proficiency Range						Total
		1	**2**	**3**	**4**	**5**	**6**	
K (instructional)	-	12.5%	5.9%	10.5%	16.3%	30.5%	24.3%	100.0%
K (accountability)	-	25.8%	10.5%	8.8%	5.9%	15.6%	33.3%	100.0%
1	A	6.0%	7.2%	22.3%	64.5%	n/a	n/a	100.0%
	B	0.7%	3.6%	14.3%	8.4%	73.0%	n/a	100.0%
	C	0.8%	2.4%	18.0%	7.9%	29.8%	41.1%	100.0%
2	A	7.8%	18.7%	17.6%	55.9%	n/a	n/a	100.0%
	B	0.5%	2.0%	9.5%	14.8%	73.1%	n/a	100.0%
	C	0.2%	2.1%	13.2%	7.1%	21.1%	56.2%	100.0%
3	A	6.1%	15.2%	24.2%	54.5%	n/a	n/a	100.0%
	B	0.3%	1.7%	10.4%	16.2%	71.4%	n/a	100.0%
	C	0.2%	0.2%	3.7%	10.4%	34.2%	51.3%	100.0%
4	A	6.6%	21.5%	25.1%	46.8%	n/a	n/a	100.0%
	B	0.4%	4.3%	14.5%	19.3%	61.4%	n/a	100.0%
	C	0.2%	0.4%	8.6%	15.6%	37.3%	37.9%	100.0%
5	A	9.3%	24.7%	25.4%	40.7%	n/a	n/a	100.0%
	B	0.4%	4.7%	13.9%	18.7%	62.4%	n/a	100.0%
	C	0.2%	0.5%	9.7%	16.5%	37.7%	35.4%	100.0%
6	A	11.8%	36.9%	20.6%	30.7%	n/a	n/a	100.0%
	B	1.1%	10.1%	24.9%	24.6%	39.4%	n/a	100.0%
	C	0.3%	1.1%	4.7%	27.1%	43.2%	23.5%	100.0%
7	A	17.9%	40.8%	19.3%	22.0%	n/a	n/a	100.0%
	B	1.6%	12.1%	27.9%	25.6%	32.8%	n/a	100.0%
	C	0.4%	0.5%	6.7%	29.4%	29.6%	33.4%	100.0%
8	A	26.5%	31.3%	20.0%	22.2%	n/a	n/a	100.0%
	B	2.5%	16.0%	31.1%	26.0%	24.3%	n/a	100.0%
	C	0.4%	1.0%	8.9%	32.6%	28.9%	28.2%	100.0%
9	A	42.5%	40.3%	7.1%	10.1%	n/a	n/a	100.0%
	B	1.5%	8.6%	19.5%	34.7%	35.8%	n/a	100.0%
	C	0.8%	3.9%	16.0%	30.3%	22.1%	26.8%	100.0%
10	A	42.5%	35.5%	15.4%	6.6%	n/a	n/a	100.0%
	B	2.1%	15.1%	25.1%	23.2%	34.5%	n/a	100.0%
	C	1.1%	3.8%	16.7%	30.9%	31.1%	16.4%	100.0%
11	A	49.6%	34.3%	7.9%	8.2%	n/a	n/a	100.0%
	B	4.3%	10.9%	25.4%	23.6%	35.8%	n/a	100.0%
	C	1.6%	5.8%	20.5%	32.7%	20.5%	18.8%	100.0%
12	A	59.9%	22.9%	8.8%	8.4%	n/a	n/a	100.0%
	B	7.4%	8.6%	26.0%	35.4%	22.6%	n/a	100.0%
	C	3.1%	5.0%	20.4%	32.8%	20.6%	18.1%	100.0%

4.3.1.3 By Grade

Table 4.3.1.3A

Proficiency Level by Grade (Count): Listening S301

	Listening Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	23,601	11,113	19,918	30,728	57,659	45,936	188,955
K (accountability)	48,827	19,817	16,716	11,228	29,459	62,908	188,955
1	4,269	8,120	31,116	46,606	76,423	11,888	178,422
2	1,902	6,192	19,276	27,198	81,324	31,256	167,148
3	1,037	3,090	12,239	23,894	70,323	32,443	143,026
4	792	3,670	12,293	19,483	43,840	19,138	99,216
5	864	3,421	10,382	15,851	36,238	15,098	81,854
6	1,257	5,751	9,903	18,492	26,034	8,371	69,808
7	1,771	6,174	10,949	18,069	18,564	11,110	66,637
8	2,467	6,177	11,434	17,691	14,704	8,730	61,203
9	6,097	8,547	10,634	19,174	15,728	7,977	68,157
10	3,275	5,703	8,776	10,658	12,318	3,241	43,971
11	2,776	3,852	7,394	9,165	8,624	3,117	34,928
12	2,426	2,063	5,762	8,478	5,248	2,507	26,484

Table 4.3.1.3B

Proficiency Level by Grade (Percent): Listening S301

	Listening Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	12.5%	5.9%	10.5%	16.3%	30.5%	24.3%	100.0%
K (accountability)	25.8%	10.5%	8.8%	5.9%	15.6%	33.3%	100.0%
1	2.4%	4.6%	17.4%	26.1%	42.8%	6.7%	100.0%
2	1.1%	3.7%	11.5%	16.3%	48.7%	18.7%	100.0%
3	0.7%	2.2%	8.6%	16.7%	49.2%	22.7%	100.0%
4	0.8%	3.7%	12.4%	19.6%	44.2%	19.3%	100.0%
5	1.1%	4.2%	12.7%	19.4%	44.3%	18.4%	100.0%
6	1.8%	8.2%	14.2%	26.5%	37.3%	12.0%	100.0%
7	2.7%	9.3%	16.4%	27.1%	27.9%	16.7%	100.0%
8	4.0%	10.1%	18.7%	28.9%	24.0%	14.3%	100.0%
9	8.9%	12.5%	15.6%	28.1%	23.1%	11.7%	100.0%
10	7.4%	13.0%	20.0%	24.2%	28.0%	7.4%	100.0%
11	7.9%	11.0%	21.2%	26.2%	24.7%	8.9%	100.0%
12	9.2%	7.8%	21.8%	32.0%	19.8%	9.5%	100.0%

4.3.2 Reading

4.3.2.1 By Cluster by Tier

Table 4.3.2.1A

Proficiency Level by Cluster by Tier (Count): Reading S301

Cluster	Tier	Reading Proficiency Range 1	2	3	4	5	6	Total
K (instructional)	-	41,759	25,994	37,045	15,945	19,172	49,031	188,946
K (accountability)	-	126,738	13,177	8,726	10,238	30,067	0	188,946
1-2	A	17,909	17,887	14,158	22,820	n/a	n/a	72,774
	B	2,728	7,757	33,535	29,095	114,919	n/a	188,034
	C	878	3,063	12,361	10,311	23,197	34,705	84,515
3-5	A	8,245	8,450	3,737	5,838	n/a	n/a	26,270
	B	2,711	15,512	36,396	13,203	73,162	n/a	140,984
	C	575	3,979	20,094	15,525	51,196	64,873	156,242
6-8	A	7,851	8,299	2,466	1,760	n/a	n/a	20,376
	B	3,812	20,664	28,002	5,161	19,584	n/a	77,223
	C	1,297	17,224	32,749	13,776	18,725	15,893	99,664
9-12	A	9,133	9,305	3,117	3,301	n/a	n/a	24,856
	B	9,542	28,303	12,096	5,426	13,306	n/a	68,673
	C	1,594	11,610	12,567	11,329	17,438	25,385	79,923

Table 4.3.2.1B

Proficiency Level by Cluster by Tier (Percent): Reading S301

Cluster	Tier	Reading Proficiency Range 1	2	3	4	5	6	Total
K (instructional)	-	22.1%	13.8%	19.6%	8.4%	10.1%	25.9%	100.0%
K (accountability)	-	67.1%	7.0%	4.6%	5.4%	15.9%	0.0%	100.0%
1-2	A	24.6%	24.6%	19.5%	31.4%	n/a	n/a	100.0%
	B	1.5%	4.1%	17.8%	15.5%	61.1%	n/a	100.0%
	C	1.0%	3.6%	14.6%	12.2%	27.4%	41.1%	100.0%
3-5	A	31.4%	32.2%	14.2%	22.2%	n/a	n/a	100.0%
	B	1.9%	11.0%	25.8%	9.4%	51.9%	n/a	100.0%
	C	0.4%	2.5%	12.9%	9.9%	32.8%	41.5%	100.0%
6-8	A	38.5%	40.7%	12.1%	8.6%	n/a	n/a	100.0%
	B	4.9%	26.8%	36.3%	6.7%	25.4%	n/a	100.0%
	C	1.3%	17.3%	32.9%	13.8%	18.8%	15.9%	100.0%
9-12	A	36.7%	37.4%	12.5%	13.3%	n/a	n/a	100.0%
	B	13.9%	41.2%	17.6%	7.9%	19.4%	n/a	100.0%
	C	2.0%	14.5%	15.7%	14.2%	21.8%	31.8%	100.0%

4.3.2.2 By Grade by Tier

Table 4.3.2.2A

Proficiency Level by Grade by Tier (Count): Reading S301

Grade	Tier	Reading Proficiency Range						Total
		1	**2**	**3**	**4**	**5**	**6**	
K (instructional)	-	41,759	25,994	37,045	15,945	19,172	49,031	188,946
K (accountability)	-	126,738	13,177	8,726	10,238	30,067	0	188,946
1	A	12,515	13,909	12,269	17,784	n/a	n/a	56,477
	B	1,622	2,681	15,915	18,531	54,138	n/a	92,887
	C	463	997	4,587	4,620	8,008	10,261	28,936
2	A	5,394	3,978	1,889	5,036	n/a	n/a	16,297
	B	1,106	5,076	17,620	10,564	60,781	n/a	95,147
	C	415	2,066	7,774	5,691	15,189	24,444	55,579
3	A	2,509	4,041	1,926	3,058	n/a	n/a	11,534
	B	651	4,516	16,961	5,325	40,642	n/a	68,095
	C	133	959	4,233	4,126	25,504	28,204	63,159
4	A	2,728	2,630	942	1,625	n/a	n/a	7,925
	B	1,098	5,792	10,107	5,566	18,071	n/a	40,634
	C	205	1,046	6,485	8,149	13,105	21,470	50,460
5	A	3,008	1,779	869	1,155	n/a	n/a	6,811
	B	962	5,204	9,328	2,312	14,449	n/a	32,255
	C	237	1,974	9,376	3,250	12,587	15,199	42,623
6	A	1,914	3,456	1,072	686	n/a	n/a	7,128
	B	1,091	5,844	11,014	1,879	7,168	n/a	26,996
	C	376	4,841	13,111	6,174	6,315	4,727	35,544
7	A	2,638	2,603	894	587	n/a	n/a	6,722
	B	1,284	7,183	10,100	1,664	6,327	n/a	26,558
	C	434	6,038	10,714	5,151	5,887	5,000	33,224
8	A	3,299	2,240	500	487	n/a	n/a	6,526
	B	1,437	7,637	6,888	1,618	6,089	n/a	23,669
	C	487	6,345	8,924	2,451	6,523	6,166	30,896
9	A	4,809	4,859	1,771	1,372	n/a	n/a	12,811
	B	2,336	9,539	6,740	1,437	5,534	n/a	25,586
	C	278	3,154	4,840	4,151	6,370	10,965	29,758
10	A	2,101	2,495	668	1,018	n/a	n/a	6,282
	B	2,469	8,390	2,168	1,834	3,084	n/a	17,945
	C	374	3,044	3,943	2,905	3,921	5,536	19,723
11	A	1,340	1,401	465	572	n/a	n/a	3,778
	B	2,349	5,976	1,923	1,585	2,730	n/a	14,563
	C	472	2,923	2,047	2,376	3,401	5,344	16,563
12	A	883	550	213	339	n/a	n/a	1,985
	B	2,388	4,398	1,265	570	1,958	n/a	10,579
	C	470	2,489	1,737	1,897	3,746	3,540	13,879

Table 4.3.2.2B
Proficiency Level by Grade by Tier (Percent): Reading S301

Grade	Tier	Reading Proficiency Range						Total
		1	**2**	**3**	**4**	**5**	**6**	
K (instructional)	-	22.1%	13.8%	19.6%	8.4%	10.1%	25.9%	100.0%
K (accountability)	-	67.1%	7.0%	4.6%	5.4%	15.9%	0.0%	100.0%
1	A	22.2%	24.6%	21.7%	31.5%	n/a	n/a	100.0%
	B	1.7%	2.9%	17.1%	20.0%	58.3%	n/a	100.0%
	C	1.6%	3.4%	15.9%	16.0%	27.7%	35.5%	100.0%
2	A	33.1%	24.4%	11.6%	30.9%	n/a	n/a	100.0%
	B	1.2%	5.3%	18.5%	11.1%	63.9%	n/a	100.0%
	C	0.7%	3.7%	14.0%	10.2%	27.3%	44.0%	100.0%
3	A	21.8%	35.0%	16.7%	26.5%	n/a	n/a	100.0%
	B	1.0%	6.6%	24.9%	7.8%	59.7%	n/a	100.0%
	C	0.2%	1.5%	6.7%	6.5%	40.4%	44.7%	100.0%
4	A	34.4%	33.2%	11.9%	20.5%	n/a	n/a	100.0%
	B	2.7%	14.3%	24.9%	13.7%	44.5%	n/a	100.0%
	C	0.4%	2.1%	12.9%	16.1%	26.0%	42.5%	100.0%
5	A	44.2%	26.1%	12.8%	17.0%	n/a	n/a	100.0%
	B	3.0%	16.1%	28.9%	7.2%	44.8%	n/a	100.0%
	C	0.6%	4.6%	22.0%	7.6%	29.5%	35.7%	100.0%
6	A	26.9%	48.5%	15.0%	9.6%	n/a	n/a	100.0%
	B	4.0%	21.6%	40.8%	7.0%	26.6%	n/a	100.0%
	C	1.1%	13.6%	36.9%	17.4%	17.8%	13.3%	100.0%
7	A	39.2%	38.7%	13.3%	8.7%	n/a	n/a	100.0%
	B	4.8%	27.0%	38.0%	6.3%	23.8%	n/a	100.0%
	C	1.3%	18.2%	32.2%	15.5%	17.7%	15.0%	100.0%
8	A	50.6%	34.3%	7.7%	7.5%	n/a	n/a	100.0%
	B	6.1%	32.3%	29.1%	6.8%	25.7%	n/a	100.0%
	C	1.6%	20.5%	28.9%	7.9%	21.1%	20.0%	100.0%
9	A	37.5%	37.9%	13.8%	10.7%	n/a	n/a	100.0%
	B	9.1%	37.3%	26.3%	5.6%	21.6%	n/a	100.0%
	C	0.9%	10.6%	16.3%	13.9%	21.4%	36.8%	100.0%
10	A	33.4%	39.7%	10.6%	16.2%	n/a	n/a	100.0%
	B	13.8%	46.8%	12.1%	10.2%	17.2%	n/a	100.0%
	C	1.9%	15.4%	20.0%	14.7%	19.9%	28.1%	100.0%
11	A	35.5%	37.1%	12.3%	15.1%	n/a	n/a	100.0%
	B	16.1%	41.0%	13.2%	10.9%	18.7%	n/a	100.0%
	C	2.8%	17.6%	12.4%	14.3%	20.5%	32.3%	100.0%
12	A	44.5%	27.7%	10.7%	17.1%	n/a	n/a	100.0%
	B	22.6%	41.6%	12.0%	5.4%	18.5%	n/a	100.0%
	C	3.4%	17.9%	12.5%	13.7%	27.0%	25.5%	100.0%

4.3.2.3 By Grade

Table 4.3.2.3A
Proficiency Level by Grade (Count): Reading S301

	Reading Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	41,759	25,994	37,045	15,945	19,172	49,031	188,946
K (accountability)	126,738	13,177	8,726	10,238	30,067	0	188,946
1	14,600	17,587	32,771	40,935	62,146	10,261	178,300
2	6,915	11,120	27,283	21,291	75,970	24,444	167,023
3	3,293	9,516	23,120	12,509	66,146	28,204	142,788
4	4,031	9,468	17,534	15,340	31,176	21,470	99,019
5	4,207	8,957	19,573	6,717	27,036	15,199	81,689
6	3,381	14,141	25,197	8,739	13,483	4,727	69,668
7	4,356	15,824	21,708	7,402	12,214	5,000	66,504
8	5,223	16,222	16,312	4,556	12,612	6,166	61,091
9	7,423	17,552	13,351	6,960	11,904	10,965	68,155
10	4,944	13,929	6,779	5,757	7,005	5,536	43,950
11	4,161	10,300	4,435	4,533	6,131	5,344	34,904
12	3,741	7,437	3,215	2,806	5,704	3,540	26,443

Table 4.3.2.3B
Proficiency Level by Grade (Percent): Reading S301

	Reading Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	22.1%	13.8%	19.6%	8.4%	10.1%	25.9%	100.0%
K (accountability)	67.1%	7.0%	4.6%	5.4%	15.9%	0.0%	100.0%
1	8.2%	9.9%	18.4%	23.0%	34.9%	5.8%	100.0%
2	4.1%	6.7%	16.3%	12.7%	45.5%	14.6%	100.0%
3	2.3%	6.7%	16.2%	8.8%	46.3%	19.8%	100.0%
4	4.1%	9.6%	17.7%	15.5%	31.5%	21.7%	100.0%
5	5.2%	11.0%	24.0%	8.2%	33.1%	18.6%	100.0%
6	4.9%	20.3%	36.2%	12.5%	19.4%	6.8%	100.0%
7	6.5%	23.8%	32.6%	11.1%	18.4%	7.5%	100.0%
8	8.5%	26.6%	26.7%	7.5%	20.6%	10.1%	100.0%
9	10.9%	25.8%	19.6%	10.2%	17.5%	16.1%	100.0%
10	11.2%	31.7%	15.4%	13.1%	15.9%	12.6%	100.0%
11	11.9%	29.5%	12.7%	13.0%	17.6%	15.3%	100.0%
12	14.1%	28.1%	12.2%	10.6%	21.6%	13.4%	100.0%

4.3.3 Writing

4.3.3.1 By Cluster by Tier

Table 4.3.3.1A

Proficiency Level by Cluster by Tier (Count): Writing S301

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	34,166	57,004	30,936	26,431	35,097	5,312	188,946
K (accountability)	-	111,214	37,323	23,508	11,589	5,312	0	188,946
1-2	A	10,382	46,723	15,692	4	0	0	72,801
	B	12,787	80,456	93,578	1,135	1	0	187,957
	C	1,359	12,395	62,068	8,648	36	0	84,506
3-5	A	2,978	6,379	10,621	6,128	134	0	26,240
	B	1,606	8,882	38,722	83,954	7,775	55	140,994
	C	395	1,685	9,062	94,621	48,569	1,925	156,257
6-8	A	3,075	6,263	9,018	1,946	56	0	20,358
	B	3,817	10,455	41,978	20,452	489	8	77,199
	C	1,792	5,063	54,140	37,676	942	4	99,617
9-12	A	2,697	6,347	13,517	2,204	50	0	24,815
	B	3,249	4,175	22,412	30,182	7,902	614	68,534
	C	1,537	1,507	10,932	35,916	25,867	4,042	79,801

Table 4.3.3.1B

Proficiency Level by Cluster by Tier (Percent): Writing S301

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	18.1%	30.2%	16.4%	14.0%	18.6%	2.8%	100.0%
K (accountability)	-	58.9%	19.8%	12.4%	6.1%	2.8%	0.0%	100.0%
1-2	A	14.3%	64.2%	21.6%	0.0%	0.0%	0.0%	100.0%
	B	6.8%	42.8%	49.8%	0.6%	0.0%	0.0%	100.0%
	C	1.6%	14.7%	73.4%	10.2%	0.0%	0.0%	100.0%
3-5	A	11.3%	24.3%	40.5%	23.4%	0.5%	0.0%	100.0%
	B	1.1%	6.3%	27.5%	59.5%	5.5%	0.0%	100.0%
	C	0.3%	1.1%	5.8%	60.6%	31.1%	1.2%	100.0%
6-8	A	15.1%	30.8%	44.3%	9.6%	0.3%	0.0%	100.0%
	B	4.9%	13.5%	54.4%	26.5%	0.6%	0.0%	100.0%
	C	1.8%	5.1%	54.3%	37.8%	0.9%	0.0%	100.0%
9-12	A	10.9%	25.6%	54.5%	8.9%	0.2%	0.0%	100.0%
	B	4.7%	6.1%	32.7%	44.0%	11.5%	0.9%	100.0%
	C	1.9%	1.9%	13.7%	45.0%	32.4%	5.1%	100.0%

4.3.3.2 By Grade by Tier

Table 4.3.3.2A

Proficiency Level by Grade by Tier (Count): Writing S301

Grade	Tier	Writing Proficiency Range						Total
		1	**2**	**3**	**4**	**5**	**6**	
K (instructional)	-	34,166	57,004	30,936	26,431	35,097	5,312	188,946
K (accountability)	-	111,214	37,323	23,508	11,589	5,312	0	188,946
1	A	6,889	34,697	14,911	3	0	0	56,500
	B	8,305	38,363	45,504	658	1	0	92,831
	C	717	5,774	19,503	2,935	13	0	28,942
2	A	3,493	12,026	781	1	0	0	16,301
	B	4,482	42,093	48,074	477	0	0	95,126
	C	642	6,621	42,565	5,713	23	0	55,564
3	A	943	2,295	4,181	3,997	112	0	11,528
	B	552	4,580	14,772	42,387	5,783	35	68,109
	C	125	469	3,570	31,903	26,022	1,062	63,151
4	A	966	2,159	3,415	1,351	20	0	7,911
	B	478	2,711	11,728	24,219	1,480	16	40,632
	C	141	674	2,352	31,154	15,589	559	50,469
5	A	1,069	1,925	3,025	780	2	0	6,801
	B	576	1,591	12,222	17,348	512	4	32,253
	C	129	542	3,140	31,564	6,958	304	42,637
6	A	599	2,214	3,096	1,174	40	0	7,123
	B	1,243	2,977	11,843	10,601	324	3	26,991
	C	638	1,380	15,681	17,294	537	1	35,531
7	A	1,010	1,992	3,145	560	10	0	6,717
	B	1,282	3,516	15,167	6,448	129	4	26,546
	C	617	1,630	17,971	12,711	277	3	33,209
8	A	1,466	2,057	2,777	212	6	0	6,518
	B	1,292	3,962	14,968	3,403	36	1	23,662
	C	537	2,053	20,488	7,671	128	0	30,877
9	A	1,433	3,368	6,489	1,466	35	0	12,791
	B	745	1,202	5,772	12,029	5,352	409	25,509
	C	350	424	1,995	8,484	15,573	2,895	29,721
10	A	570	1,545	3,712	437	9	0	6,273
	B	892	1,057	5,863	8,434	1,546	128	17,920
	C	332	324	2,499	9,960	5,887	683	19,685
11	A	376	792	2,352	252	5	0	3,777
	B	819	827	5,905	6,167	766	59	14,543
	C	353	358	2,731	9,716	3,037	342	16,537
12	A	318	642	964	49	1	0	1,974
	B	793	1,089	4,872	3,552	238	18	10,562
	C	502	401	3,707	7,756	1,370	122	13,858

Table 4.3.3.2B
Proficiency Level by Grade by Tier (Percent): Writing S301

Grade	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	18.1%	30.2%	16.4%	14.0%	18.6%	2.8%	100.0%
K (accountability)	-	58.9%	19.8%	12.4%	6.1%	2.8%	0.0%	100.0%
1	A	12.2%	61.4%	26.4%	0.0%	0.0%	0.0%	100.0%
	B	8.9%	41.3%	49.0%	0.7%	0.0%	0.0%	100.0%
	C	2.5%	20.0%	67.4%	10.1%	0.0%	0.0%	100.0%
2	A	21.4%	73.8%	4.8%	0.0%	0.0%	0.0%	100.0%
	B	4.7%	44.2%	50.5%	0.5%	0.0%	0.0%	100.0%
	C	1.2%	11.9%	76.6%	10.3%	0.0%	0.0%	100.0%
3	A	8.2%	19.9%	36.3%	34.7%	1.0%	0.0%	100.0%
	B	0.8%	6.7%	21.7%	62.2%	8.5%	0.1%	100.0%
	C	0.2%	0.7%	5.7%	50.5%	41.2%	1.7%	100.0%
4	A	12.2%	27.3%	43.2%	17.1%	0.3%	0.0%	100.0%
	B	1.2%	6.7%	28.9%	59.6%	3.6%	0.0%	100.0%
	C	0.3%	1.3%	4.7%	61.7%	30.9%	1.1%	100.0%
5	A	15.7%	28.3%	44.5%	11.5%	0.0%	0.0%	100.0%
	B	1.8%	4.9%	37.9%	53.8%	1.6%	0.0%	100.0%
	C	0.3%	1.3%	7.4%	74.0%	16.3%	0.7%	100.0%
6	A	8.4%	31.1%	43.5%	16.5%	0.6%	0.0%	100.0%
	B	4.6%	11.0%	43.9%	39.3%	1.2%	0.0%	100.0%
	C	1.8%	3.9%	44.1%	48.7%	1.5%	0.0%	100.0%
7	A	15.0%	29.7%	46.8%	8.3%	0.1%	0.0%	100.0%
	B	4.8%	13.2%	57.1%	24.3%	0.5%	0.0%	100.0%
	C	1.9%	4.9%	54.1%	38.3%	0.8%	0.0%	100.0%
8	A	22.5%	31.6%	42.6%	3.3%	0.1%	0.0%	100.0%
	B	5.5%	16.7%	63.3%	14.4%	0.2%	0.0%	100.0%
	C	1.7%	6.6%	66.4%	24.8%	0.4%	0.0%	100.0%
9	A	11.2%	26.3%	50.7%	11.5%	0.3%	0.0%	100.0%
	B	2.9%	4.7%	22.6%	47.2%	21.0%	1.6%	100.0%
	C	1.2%	1.4%	6.7%	28.5%	52.4%	9.7%	100.0%
10	A	9.1%	24.6%	59.2%	7.0%	0.1%	0.0%	100.0%
	B	5.0%	5.9%	32.7%	47.1%	8.6%	0.7%	100.0%
	C	1.7%	1.6%	12.7%	50.6%	29.9%	3.5%	100.0%
11	A	10.0%	21.0%	62.3%	6.7%	0.1%	0.0%	100.0%
	B	5.6%	5.7%	40.6%	42.4%	5.3%	0.4%	100.0%
	C	2.1%	2.2%	16.5%	58.8%	18.4%	2.1%	100.0%
12	A	16.1%	32.5%	48.8%	2.5%	0.1%	0.0%	100.0%
	B	7.5%	10.3%	46.1%	33.6%	2.3%	0.2%	100.0%
	C	3.6%	2.9%	26.7%	56.0%	9.9%	0.9%	100.0%

4.3.3.3 By Grade

Table 4.3.3.3A

Proficiency Level by Grade (Count): Writing S301

	Writing Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
K (instructional)	34,166	57,004	30,936	26,431	35,097	5,312	188,946
K (accountability)	111,214	37,323	23,508	11,589	5,312	0	188,946
1	15,911	78,834	79,918	3,596	14	0	178,273
2	8,617	60,740	91,420	6,191	23	0	166,991
3	1,620	7,344	22,523	78,287	31,917	1,097	142,788
4	1,585	5,544	17,495	56,724	17,089	575	99,012
5	1,774	4,058	18,387	49,692	7,472	308	81,691
6	2,480	6,571	30,620	29,069	901	4	69,645
7	2,909	7,138	36,283	19,719	416	7	66,472
8	3,295	8,072	38,233	11,286	170	1	61,057
9	2,528	4,994	14,256	21,979	20,960	3,304	68,021
10	1,794	2,926	12,074	18,831	7,442	811	43,878
11	1,548	1,977	10,988	16,135	3,808	401	34,857
12	1,613	2,132	9,543	11,357	1,609	140	26,394

Table 4.3.3.3B

Proficiency Level by Grade (Percent): Writing S301

	Writing Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
K (instructional)	18.1%	30.2%	16.4%	14.0%	18.6%	2.8%	100.0%
K (accountability)	58.9%	19.8%	12.4%	6.1%	2.8%	0.0%	100.0%
1	8.9%	44.2%	44.8%	2.0%	0.0%	0.0%	100.0%
2	5.2%	36.4%	54.7%	3.7%	0.0%	0.0%	100.0%
3	1.1%	5.1%	15.8%	54.8%	22.4%	0.8%	100.0%
4	1.6%	5.6%	17.7%	57.3%	17.3%	0.6%	100.0%
5	2.2%	5.0%	22.5%	60.8%	9.1%	0.4%	100.0%
6	3.6%	9.4%	44.0%	41.7%	1.3%	0.0%	100.0%
7	4.4%	10.7%	54.6%	29.7%	0.6%	0.0%	100.0%
8	5.4%	13.2%	62.6%	18.5%	0.3%	0.0%	100.0%
9	3.7%	7.3%	21.0%	32.3%	30.8%	4.9%	100.0%
10	4.1%	6.7%	27.5%	42.9%	17.0%	1.8%	100.0%
11	4.4%	5.7%	31.5%	46.3%	10.9%	1.2%	100.0%
12	6.1%	8.1%	36.2%	43.0%	6.1%	0.5%	100.0%

4.3.4 Speaking

4.3.4.1 By Cluster by Tier

Table 4.3.4.1A

Proficiency Level by Cluster by Tier (Count): Speaking S301

Cluster	Tier	Speaking Proficiency Range						Total
		1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	-	46,000	15,241	29,639	31,319	20,085	46,662	188,946
K (accountability)	-	46,000	44,880	31,319	20,085	46,662	0	188,946
1-2	A	17,534	23,216	10,975	4,348	4,079	12,743	72,895
	B	6,975	34,112	35,345	18,099	19,105	74,382	188,018
	C	931	5,059	9,035	6,649	9,099	53,723	84,496
3-5	A	10,422	7,287	4,047	1,354	941	2,412	26,463
	B	4,610	22,330	37,510	19,074	15,962	41,754	141,240
	C	1,752	9,701	26,373	21,073	22,327	75,064	156,290
6-8	A	10,449	4,712	2,763	1,119	376	1,132	20,551
	B	3,508	10,021	18,379	15,449	7,044	22,881	77,282
	C	1,242	4,357	13,854	18,763	11,201	50,213	99,630
9-12	A	14,055	4,683	2,808	1,430	627	1,366	24,969
	B	5,626	8,882	12,988	11,947	7,376	21,891	68,710
	C	1,567	2,791	7,900	12,208	10,192	45,268	79,926

Table 4.3.4.1B

Proficiency Level by Cluster by Tier (Percent): Speaking S301

Cluster	Tier	Speaking Proficiency Range						Total
		1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	-	24.3%	8.1%	15.7%	16.6%	10.6%	24.7%	100.0%
K (accountability)	-	24.3%	23.8%	16.6%	10.6%	24.7%	0.0%	100.0%
1-2	A	24.1%	31.8%	15.1%	6.0%	5.6%	17.5%	100.0%
	B	3.7%	18.1%	18.8%	9.6%	10.2%	39.6%	100.0%
	C	1.1%	6.0%	10.7%	7.9%	10.8%	63.6%	100.0%
3-5	A	39.4%	27.5%	15.3%	5.1%	3.6%	9.1%	100.0%
	B	3.3%	15.8%	26.6%	13.5%	11.3%	29.6%	100.0%
	C	1.1%	6.2%	16.9%	13.5%	14.3%	48.0%	100.0%
6-8	A	50.8%	22.9%	13.4%	5.4%	1.8%	5.5%	100.0%
	B	4.5%	13.0%	23.8%	20.0%	9.1%	29.6%	100.0%
	C	1.2%	4.4%	13.9%	18.8%	11.2%	50.4%	100.0%
9-12	A	56.3%	18.8%	11.2%	5.7%	2.5%	5.5%	100.0%
	B	8.2%	12.9%	18.9%	17.4%	10.7%	31.9%	100.0%
	C	2.0%	3.5%	9.9%	15.3%	12.8%	56.6%	100.0%

4.3.4.2 By Grade by Tier

Table 4.3.4.2A

Proficiency Level by Grade by Tier (Count): Speaking S301

Grade	Tier	Speaking Proficiency Range 1	2	3	4	5	6	Total
K (instructional)	-	46,000	15,241	29,639	31,319	20,085	46,662	188,946
K (accountability)	-	46,000	44,880	31,319	20,085	46,662	0	188,946
1	A	11,715	19,352	8,806	3,457	3,268	9,939	56,537
	B	3,441	21,391	18,950	8,892	9,199	30,981	92,854
	C	388	2,641	3,948	2,584	3,387	15,989	28,937
2	A	5,819	3,864	2,169	891	811	2,804	16,358
	B	3,534	12,721	16,395	9,207	9,906	43,401	95,164
	C	543	2,418	5,087	4,065	5,712	37,734	55,559
3	A	3,683	3,776	1,970	646	429	1,110	11,614
	B	1,728	12,527	19,396	9,129	7,377	18,048	68,205
	C	560	4,728	11,934	8,815	8,853	28,285	63,175
4	A	3,574	1,915	1,137	396	297	672	7,991
	B	1,702	5,846	10,631	5,553	4,714	12,248	40,694
	C	682	3,050	8,522	6,894	7,419	23,908	50,475
5	A	3,165	1,596	940	312	215	630	6,858
	B	1,180	3,957	7,483	4,392	3,871	11,458	32,341
	C	510	1,923	5,917	5,364	6,055	22,871	42,640
6	A	3,177	1,804	1,160	493	120	424	7,178
	B	969	3,180	6,552	6,605	2,488	7,243	27,037
	C	416	1,601	5,312	8,478	4,097	15,642	35,546
7	A	3,751	1,160	944	443	140	347	6,785
	B	1,382	2,425	6,020	6,230	2,406	8,100	26,563
	C	446	992	4,005	7,188	3,710	16,874	33,215
8	A	3,521	1,748	659	183	116	361	6,588
	B	1,157	4,416	5,807	2,614	2,150	7,538	23,682
	C	380	1,764	4,537	3,097	3,394	17,697	30,869
9	A	8,016	1,754	1,334	851	294	568	12,817
	B	1,772	1,951	3,766	6,094	3,030	8,953	25,566
	C	463	529	1,669	5,403	3,965	17,682	29,711
10	A	3,400	1,465	693	262	153	340	6,313
	B	1,730	2,996	3,654	2,481	1,821	5,294	17,976
	C	408	832	2,420	2,726	2,495	10,857	19,738
11	A	1,835	934	485	190	105	267	3,816
	B	1,221	2,336	3,246	1,957	1,463	4,358	14,581
	C	326	759	2,027	2,224	2,061	9,165	16,562
12	A	804	530	296	127	75	191	2,023
	B	903	1,599	2,322	1,415	1,062	3,286	10,587
	C	370	671	1,784	1,855	1,671	7,564	13,915

Table 4.3.4.2B
Proficiency Level by Grade by Tier (Percent): Speaking S301

Grade	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	24.3%	8.1%	15.7%	16.6%	10.6%	24.7%	100.0%
K (accountability)	-	24.3%	23.8%	16.6%	10.6%	24.7%	0.0%	100.0%
1	A	20.7%	34.2%	15.6%	6.1%	5.8%	17.6%	100.0%
	B	3.7%	23.0%	20.4%	9.6%	9.9%	33.4%	100.0%
	C	1.3%	9.1%	13.6%	8.9%	11.7%	55.3%	100.0%
2	A	35.6%	23.6%	13.3%	5.4%	5.0%	17.1%	100.0%
	B	3.7%	13.4%	17.2%	9.7%	10.4%	45.6%	100.0%
	C	1.0%	4.4%	9.2%	7.3%	10.3%	67.9%	100.0%
3	A	31.7%	32.5%	17.0%	5.6%	3.7%	9.6%	100.0%
	B	2.5%	18.4%	28.4%	13.4%	10.8%	26.5%	100.0%
	C	0.9%	7.5%	18.9%	14.0%	14.0%	44.8%	100.0%
4	A	44.7%	24.0%	14.2%	5.0%	3.7%	8.4%	100.0%
	B	4.2%	14.4%	26.1%	13.6%	11.6%	30.1%	100.0%
	C	1.4%	6.0%	16.9%	13.7%	14.7%	47.4%	100.0%
5	A	46.2%	23.3%	13.7%	4.5%	3.1%	9.2%	100.0%
	B	3.6%	12.2%	23.1%	13.6%	12.0%	35.4%	100.0%
	C	1.2%	4.5%	13.9%	12.6%	14.2%	53.6%	100.0%
6	A	44.3%	25.1%	16.2%	6.9%	1.7%	5.9%	100.0%
	B	3.6%	11.8%	24.2%	24.4%	9.2%	26.8%	100.0%
	C	1.2%	4.5%	14.9%	23.9%	11.5%	44.0%	100.0%
7	A	55.3%	17.1%	13.9%	6.5%	2.1%	5.1%	100.0%
	B	5.2%	9.1%	22.7%	23.5%	9.1%	30.5%	100.0%
	C	1.3%	3.0%	12.1%	21.6%	11.2%	50.8%	100.0%
8	A	53.4%	26.5%	10.0%	2.8%	1.8%	5.5%	100.0%
	B	4.9%	18.6%	24.5%	11.0%	9.1%	31.8%	100.0%
	C	1.2%	5.7%	14.7%	10.0%	11.0%	57.3%	100.0%
9	A	62.5%	13.7%	10.4%	6.6%	2.3%	4.4%	100.0%
	B	6.9%	7.6%	14.7%	23.8%	11.9%	35.0%	100.0%
	C	1.6%	1.8%	5.6%	18.2%	13.3%	59.5%	100.0%
10	A	53.9%	23.2%	11.0%	4.2%	2.4%	5.4%	100.0%
	B	9.6%	16.7%	20.3%	13.8%	10.1%	29.5%	100.0%
	C	2.1%	4.2%	12.3%	13.8%	12.6%	55.0%	100.0%
11	A	48.1%	24.5%	12.7%	5.0%	2.8%	7.0%	100.0%
	B	8.4%	16.0%	22.3%	13.4%	10.0%	29.9%	100.0%
	C	2.0%	4.6%	12.2%	13.4%	12.4%	55.3%	100.0%
12	A	39.7%	26.2%	14.6%	6.3%	3.7%	9.4%	100.0%
	B	8.5%	15.1%	21.9%	13.4%	10.0%	31.0%	100.0%
	C	2.7%	4.8%	12.8%	13.3%	12.0%	54.4%	100.0%

4.3.4.3 By Grade

Table 4.3.4.3A

Proficiency Level by Grade (Count): Speaking S301

	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	46,000	15,241	29,639	31,319	20,085	46,662	188,946
K (accountability)	46,000	44,880	31,319	20,085	46,662	0	188,946
1	15,544	43,384	31,704	14,933	15,854	56,909	178,328
2	9,896	19,003	23,651	14,163	16,429	83,939	167,081
3	5,971	21,031	33,300	18,590	16,659	47,443	142,994
4	5,958	10,811	20,290	12,843	12,430	36,828	99,160
5	4,855	7,476	14,340	10,068	10,141	34,959	81,839
6	4,562	6,585	13,024	15,576	6,705	23,309	69,761
7	5,579	4,577	10,969	13,861	6,256	25,321	66,563
8	5,058	7,928	11,003	5,894	5,660	25,596	61,139
9	10,251	4,234	6,769	12,348	7,289	27,203	68,094
10	5,538	5,293	6,767	5,469	4,469	16,491	44,027
11	3,382	4,029	5,758	4,371	3,629	13,790	34,959
12	2,077	2,800	4,402	3,397	2,808	11,041	26,525

Table 4.3.4.3B

Proficiency Level by Grade (Percent): Speaking S301

	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	24.3%	8.1%	15.7%	16.6%	10.6%	24.7%	100.0%
K (accountability)	24.3%	23.8%	16.6%	10.6%	24.7%	0.0%	100.0%
1	8.7%	24.3%	17.8%	8.4%	8.9%	31.9%	100.0%
2	5.9%	11.4%	14.2%	8.5%	9.8%	50.2%	100.0%
3	4.2%	14.7%	23.3%	13.0%	11.7%	33.2%	100.0%
4	6.0%	10.9%	20.5%	13.0%	12.5%	37.1%	100.0%
5	5.9%	9.1%	17.5%	12.3%	12.4%	42.7%	100.0%
6	6.5%	9.4%	18.7%	22.3%	9.6%	33.4%	100.0%
7	8.4%	6.9%	16.5%	20.8%	9.4%	38.0%	100.0%
8	8.3%	13.0%	18.0%	9.6%	9.3%	41.9%	100.0%
9	15.1%	6.2%	9.9%	18.1%	10.7%	39.9%	100.0%
10	12.6%	12.0%	15.4%	12.4%	10.2%	37.5%	100.0%
11	9.7%	11.5%	16.5%	12.5%	10.4%	39.4%	100.0%
12	7.8%	10.6%	16.6%	12.8%	10.6%	41.6%	100.0%

4.3.5 Oral Language Composite

4.3.5.1 By Cluster by Tier

Table 4.3.5.1A
Proficiency Level by Cluster by Tier (Count): Oral S301

| Cluster | Tier | Oral Language Proficiency Range | | | | | | Total |
		1	2	3	4	5	6	
K (instructional)	-	28,902	17,766	23,533	39,201	41,242	38,302	188,946
K (accountability)	-	49,691	28,259	31,452	16,239	25,003	38,302	188,946
1-2	A	9,203	18,267	25,087	8,474	11,803	0	72,834
	B	1,946	11,338	51,726	34,330	88,604	0	187,944
	C	417	2,092	7,473	10,611	26,130	37,765	84,488
3-5	A	5,089	8,159	8,627	2,238	2,299	0	26,412
	B	1,207	6,752	32,429	46,948	53,814	0	141,150
	C	581	1,199	10,455	29,725	50,608	63,671	156,239
6-8	A	7,704	6,408	4,209	1,215	965	0	20,501
	B	1,594	8,600	19,884	24,091	23,014	0	77,183
	C	656	899	6,671	22,181	38,627	30,518	99,552
9-12	A	13,063	6,934	2,883	1,424	544	0	24,848
	B	2,596	7,973	14,757	20,671	22,295	0	68,292
	C	1,243	1,654	7,731	19,282	27,487	21,987	79,384

Table 4.3.5.1B
Proficiency Level by Cluster by Tier (Percent): Oral S301

| Cluster | Tier | Oral Language Proficiency Range | | | | | | Total |
		1	2	3	4	5	6	
K (instructional)	-	15.3%	9.4%	12.5%	20.7%	21.8%	20.3%	100.0%
K (accountability)	-	26.3%	15.0%	16.6%	8.6%	13.2%	20.3%	100.0%
1-2	A	12.6%	25.1%	34.4%	11.6%	16.2%	0.0%	100.0%
	B	1.0%	6.0%	27.5%	18.3%	47.1%	0.0%	100.0%
	C	0.5%	2.5%	8.8%	12.6%	30.9%	44.7%	100.0%
3-5	A	19.3%	30.9%	32.7%	8.5%	8.7%	0.0%	100.0%
	B	0.9%	4.8%	23.0%	33.3%	38.1%	0.0%	100.0%
	C	0.4%	0.8%	6.7%	19.0%	32.4%	40.8%	100.0%
6-8	A	37.6%	31.3%	20.5%	5.9%	4.7%	0.0%	100.0%
	B	2.1%	11.1%	25.8%	31.2%	29.8%	0.0%	100.0%
	C	0.7%	0.9%	6.7%	22.3%	38.8%	30.7%	100.0%
9-12	A	52.6%	27.9%	11.6%	5.7%	2.2%	0.0%	100.0%
	B	3.8%	11.7%	21.6%	30.3%	32.6%	0.0%	100.0%
	C	1.6%	2.1%	9.7%	24.3%	34.6%	27.7%	100.0%

4.3.5.2 By Grade by Tier

Table 4.3.5.2A

Proficiency Level by Grade by Tier (Count): Oral S301

Grade	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	28,902	17,766	23,533	39,201	41,242	38,302	188,946
K (accountability)	-	49,691	28,259	31,452	16,239	25,003	38,302	188,946
1	A	6,060	14,199	20,197	6,845	9,191	0	56,492
	B	1,071	6,620	29,731	17,225	38,167	0	92,814
	C	200	1,084	3,692	3,890	8,723	11,345	28,934
2	A	3,143	4,068	4,890	1,629	2,612	0	16,342
	B	875	4,718	21,995	17,105	50,437	0	95,130
	C	217	1,008	3,781	6,721	17,407	26,420	55,554
3	A	1,599	3,605	4,262	1,034	1,089	0	11,589
	B	451	2,657	17,215	23,504	24,349	0	68,176
	C	192	380	3,755	11,492	18,785	28,550	63,154
4	A	1,743	2,403	2,518	682	629	0	7,975
	B	443	2,073	8,735	13,630	15,794	0	40,675
	C	211	412	3,465	9,295	16,772	20,307	50,462
5	A	1,747	2,151	1,847	522	581	0	6,848
	B	313	2,022	6,479	9,814	13,671	0	32,299
	C	178	407	3,235	8,938	15,051	14,814	42,623
6	A	2,276	2,201	1,800	499	386	0	7,162
	B	484	2,623	6,625	9,324	7,948	0	27,004
	C	198	312	2,471	8,205	13,461	10,876	35,523
7	A	2,638	2,102	1,302	436	293	0	6,771
	B	542	2,821	7,126	7,874	8,173	0	26,536
	C	242	271	2,195	7,117	13,143	10,217	33,185
8	A	2,790	2,105	1,107	280	286	0	6,568
	B	568	3,156	6,133	6,893	6,893	0	23,643
	C	216	316	2,005	6,859	12,023	9,425	30,844
9	A	6,951	3,498	1,367	672	286	0	12,774
	B	713	2,047	4,414	7,778	10,488	0	25,440
	C	357	359	1,851	5,796	10,470	10,731	29,564
10	A	3,200	1,861	744	317	163	0	6,285
	B	676	2,354	3,943	5,406	5,478	0	17,857
	C	289	408	1,886	4,915	7,049	5,069	19,616
11	A	1,912	1,030	506	245	95	0	3,788
	B	602	2,026	3,656	4,251	3,965	0	14,500
	C	262	416	1,885	4,488	5,904	3,483	16,438
12	A	1,000	545	266	190	0	0	2,001
	B	605	1,546	2,744	3,236	2,364	0	10,495
	C	335	471	2,109	4,083	4,064	2,704	13,766

Table 4.3.5.2B
Proficiency Level by Grade by Tier (Percent): Oral S301

Grade	Tier	Oral Language Proficiency Range						Total
		1	**2**	**3**	**4**	**5**	**6**	
K (instructional)	-	15.3%	9.4%	12.5%	20.7%	21.8%	20.3%	100.0%
K (accountability)	-	26.3%	15.0%	16.6%	8.6%	13.2%	20.3%	100.0%
1	A	10.7%	25.1%	35.8%	12.1%	16.3%	0.0%	100.0%
	B	1.2%	7.1%	32.0%	18.6%	41.1%	0.0%	100.0%
	C	0.7%	3.7%	12.8%	13.4%	30.1%	39.2%	100.0%
2	A	19.2%	24.9%	29.9%	10.0%	16.0%	0.0%	100.0%
	B	0.9%	5.0%	23.1%	18.0%	53.0%	0.0%	100.0%
	C	0.4%	1.8%	6.8%	12.1%	31.3%	47.6%	100.0%
3	A	13.8%	31.1%	36.8%	8.9%	9.4%	0.0%	100.0%
	B	0.7%	3.9%	25.3%	34.5%	35.7%	0.0%	100.0%
	C	0.3%	0.6%	5.9%	18.2%	29.7%	45.2%	100.0%
4	A	21.9%	30.1%	31.6%	8.6%	7.9%	0.0%	100.0%
	B	1.1%	5.1%	21.5%	33.5%	38.8%	0.0%	100.0%
	C	0.4%	0.8%	6.9%	18.4%	33.2%	40.2%	100.0%
5	A	25.5%	31.4%	27.0%	7.6%	8.5%	0.0%	100.0%
	B	1.0%	6.3%	20.1%	30.4%	42.3%	0.0%	100.0%
	C	0.4%	1.0%	7.6%	21.0%	35.3%	34.8%	100.0%
6	A	31.8%	30.7%	25.1%	7.0%	5.4%	0.0%	100.0%
	B	1.8%	9.7%	24.5%	34.5%	29.4%	0.0%	100.0%
	C	0.6%	0.9%	7.0%	23.1%	37.9%	30.6%	100.0%
7	A	39.0%	31.0%	19.2%	6.4%	4.3%	0.0%	100.0%
	B	2.0%	10.6%	26.9%	29.7%	30.8%	0.0%	100.0%
	C	0.7%	0.8%	6.6%	21.4%	39.6%	30.8%	100.0%
8	A	42.5%	32.0%	16.9%	4.3%	4.4%	0.0%	100.0%
	B	2.4%	13.3%	25.9%	29.2%	29.2%	0.0%	100.0%
	C	0.7%	1.0%	6.5%	22.2%	39.0%	30.6%	100.0%
9	A	54.4%	27.4%	10.7%	5.3%	2.2%	0.0%	100.0%
	B	2.8%	8.0%	17.4%	30.6%	41.2%	0.0%	100.0%
	C	1.2%	1.2%	6.3%	19.6%	35.4%	36.3%	100.0%
10	A	50.9%	29.6%	11.8%	5.0%	2.6%	0.0%	100.0%
	B	3.8%	13.2%	22.1%	30.3%	30.7%	0.0%	100.0%
	C	1.5%	2.1%	9.6%	25.1%	35.9%	25.8%	100.0%
11	A	50.5%	27.2%	13.4%	6.5%	2.5%	0.0%	100.0%
	B	4.2%	14.0%	25.2%	29.3%	27.3%	0.0%	100.0%
	C	1.6%	2.5%	11.5%	27.3%	35.9%	21.2%	100.0%
12	A	50.0%	27.2%	13.3%	9.5%	0.0%	0.0%	100.0%
	B	5.8%	14.7%	26.1%	30.8%	22.5%	0.0%	100.0%
	C	2.4%	3.4%	15.3%	29.7%	29.5%	19.6%	100.0%

4.3.5.3 By Grade

Table 4.3.5.3A

Proficiency Level by Grade (Count): Oral S301

	Oral Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	28,902	17,766	23,533	39,201	41,242	38,302	188,946
K (accountability)	49,691	28,259	31,452	16,239	25,003	38,302	188,946
1	7,331	21,903	53,620	27,960	56,081	11,345	178,240
2	4,235	9,794	30,666	25,455	70,456	26,420	167,026
3	2,242	6,642	25,232	36,030	44,223	28,550	142,919
4	2,397	4,888	14,718	23,607	33,195	20,307	99,112
5	2,238	4,580	11,561	19,274	29,303	14,814	81,770
6	2,958	5,136	10,896	18,028	21,795	10,876	69,689
7	3,422	5,194	10,623	15,427	21,609	10,217	66,492
8	3,574	5,577	9,245	14,032	19,202	9,425	61,055
9	8,021	5,904	7,632	14,246	21,244	10,731	67,778
10	4,165	4,623	6,573	10,638	12,690	5,069	43,758
11	2,776	3,472	6,047	8,984	9,964	3,483	34,726
12	1,940	2,562	5,119	7,509	6,428	2,704	26,262

Table 4.3.5.3B

Proficiency Level by Grade (Percent): Oral S301

	Oral Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	15.3%	9.4%	12.5%	20.7%	21.8%	20.3%	100.0%
K (accountability)	26.3%	15.0%	16.6%	8.6%	13.2%	20.3%	100.0%
1	4.1%	12.3%	30.1%	15.7%	31.5%	6.4%	100.0%
2	2.5%	5.9%	18.4%	15.2%	42.2%	15.8%	100.0%
3	1.6%	4.6%	17.7%	25.2%	30.9%	20.0%	100.0%
4	2.4%	4.9%	14.8%	23.8%	33.5%	20.5%	100.0%
5	2.7%	5.6%	14.1%	23.6%	35.8%	18.1%	100.0%
6	4.2%	7.4%	15.6%	25.9%	31.3%	15.6%	100.0%
7	5.1%	7.8%	16.0%	23.2%	32.5%	15.4%	100.0%
8	5.9%	9.1%	15.1%	23.0%	31.5%	15.4%	100.0%
9	11.8%	8.7%	11.3%	21.0%	31.3%	15.8%	100.0%
10	9.5%	10.6%	15.0%	24.3%	29.0%	11.6%	100.0%
11	8.0%	10.0%	17.4%	25.9%	28.7%	10.0%	100.0%
12	7.4%	9.8%	19.5%	28.6%	24.5%	10.3%	100.0%

4.3.6 Literacy Composite

4.3.6.1 By Cluster by Tier

Table 4.3.6.1A

Proficiency Level by Cluster by Tier (Count): Literacy S301

| Cluster | Tier | Literacy Proficiency Range | | | | | | Total |
		1	2	3	4	5	6	
K (instructional)	-	33,587	46,700	34,542	23,717	38,105	12,289	188,940
K (accountability)	-	121,482	22,515	22,417	15,657	6,869	0	188,940
1-2	A	12,781	34,488	25,460	4	0	0	72,733
	B	3,464	37,080	138,825	8,515	2	0	187,886
	C	574	6,509	31,461	28,281	15,480	2,179	84,484
3-5	A	4,199	8,910	8,715	4,362	25	0	26,211
	B	1,280	9,556	42,795	79,731	7,506	0	140,868
	C	331	790	12,129	50,930	70,056	21,935	156,171
6-8	A	4,591	8,465	6,207	1,067	18	0	20,348
	B	2,608	15,531	40,044	18,443	473	1	77,100
	C	823	7,250	49,000	31,772	9,165	1,562	99,572
9-12	A	4,468	10,326	8,199	1,771	18	0	24,782
	B	3,467	12,840	26,392	19,825	5,850	44	68,418
	C	1,045	2,569	13,613	25,131	25,618	11,664	79,640

Table 4.3.6.1B

Proficiency Level by Cluster by Tier (Percent): Literacy S301

| Cluster | Tier | Literacy Proficiency Range | | | | | | Total |
		1	2	3	4	5	6	
K (instructional)	-	17.8%	24.7%	18.3%	12.6%	20.2%	6.5%	100.0%
K (accountability)	-	64.3%	11.9%	11.9%	8.3%	3.6%	0.0%	100.0%
1-2	A	17.6%	47.4%	35.0%	0.0%	0.0%	0.0%	100.0%
	B	1.8%	19.7%	73.9%	4.5%	0.0%	0.0%	100.0%
	C	0.7%	7.7%	37.2%	33.5%	18.3%	2.6%	100.0%
3-5	A	16.0%	34.0%	33.2%	16.6%	0.1%	0.0%	100.0%
	B	0.9%	6.8%	30.4%	56.6%	5.3%	0.0%	100.0%
	C	0.2%	0.5%	7.8%	32.6%	44.9%	14.0%	100.0%
6-8	A	22.6%	41.6%	30.5%	5.2%	0.1%	0.0%	100.0%
	B	3.4%	20.1%	51.9%	23.9%	0.6%	0.0%	100.0%
	C	0.8%	7.3%	49.2%	31.9%	9.2%	1.6%	100.0%
9-12	A	18.0%	41.7%	33.1%	7.1%	0.1%	0.0%	100.0%
	B	5.1%	18.8%	38.6%	29.0%	8.6%	0.1%	100.0%
	C	1.3%	3.2%	17.1%	31.6%	32.2%	14.6%	100.0%

4.3.6.2 By Grade by Tier

Table 4.3.6.2A

Proficiency Level by Grade by Tier (Count): Literacy S301

Grade	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	33,587	46,700	34,542	23,717	38,105	12,289	188,940
K (accountability)	-	121,482	22,515	22,417	15,657	6,869	0	188,940
1	A	8,268	27,330	20,850	3	0	0	56,451
	B	1,766	19,176	66,824	5,022	1	0	92,789
	C	288	2,879	11,753	8,782	4,385	839	28,926
2	A	4,513	7,158	4,610	1	0	0	16,282
	B	1,698	17,904	72,001	3,493	1	0	95,097
	C	286	3,630	19,708	19,499	11,095	1,340	55,558
3	A	1,086	3,718	3,937	2,750	23	0	11,514
	B	395	3,554	17,509	41,077	5,511	0	68,046
	C	104	163	2,459	15,161	34,246	10,993	63,126
4	A	1,329	2,815	2,729	1,029	1	0	7,903
	B	387	3,032	12,831	22,965	1,384	0	40,599
	C	122	274	3,713	18,725	20,979	6,630	50,443
5	A	1,784	2,377	2,049	583	1	0	6,794
	B	498	2,970	12,455	15,689	611	0	32,223
	C	105	353	5,957	17,044	14,831	4,312	42,602
6	A	973	2,919	2,584	634	10	0	7,120
	B	596	4,503	13,340	8,243	280	1	26,963
	C	216	2,015	16,970	12,539	3,264	508	35,512
7	A	1,575	2,777	2,051	307	4	0	6,714
	B	881	5,269	13,744	6,471	149	0	26,514
	C	267	2,342	16,318	10,587	3,173	504	33,191
8	A	2,043	2,769	1,572	126	4	0	6,514
	B	1,131	5,759	12,960	3,729	44	0	23,623
	C	340	2,893	15,712	8,646	2,728	550	30,869
9	A	2,383	5,252	4,023	1,103	14	0	12,775
	B	843	3,056	9,616	8,343	3,594	30	25,482
	C	242	499	3,272	7,871	11,658	6,129	29,671
10	A	955	2,693	2,184	430	2	0	6,264
	B	866	3,658	6,903	5,111	1,340	8	17,886
	C	213	584	3,468	6,625	6,338	2,428	19,656
11	A	628	1,565	1,384	195	1	0	3,773
	B	841	3,191	5,775	3,993	708	5	14,513
	C	241	665	3,321	5,706	4,639	1,931	16,503
12	A	502	816	608	43	1	0	1,970
	B	917	2,935	4,098	2,378	208	1	10,537
	C	349	821	3,552	4,929	2,983	1,176	13,810

Table 4.3.6.2B
Proficiency Level by Grade by Tier (Percent): Literacy S301

Grade	Tier	Literacy Proficiency Range						Total
		1	**2**	**3**	**4**	**5**	**6**	
K (instructional)	-	17.8%	24.7%	18.3%	12.6%	20.2%	6.5%	100.0%
K (accountability)	-	64.3%	11.9%	11.9%	8.3%	3.6%	0.0%	100.0%
1	A	14.6%	48.4%	36.9%	0.0%	0.0%	0.0%	100.0%
	B	1.9%	20.7%	72.0%	5.4%	0.0%	0.0%	100.0%
	C	1.0%	10.0%	40.6%	30.4%	15.2%	2.9%	100.0%
2	A	27.7%	44.0%	28.3%	0.0%	0.0%	0.0%	100.0%
	B	1.8%	18.8%	75.7%	3.7%	0.0%	0.0%	100.0%
	C	0.5%	6.5%	35.5%	35.1%	20.0%	2.4%	100.0%
3	A	9.4%	32.3%	34.2%	23.9%	0.2%	0.0%	100.0%
	B	0.6%	5.2%	25.7%	60.4%	8.1%	0.0%	100.0%
	C	0.2%	0.3%	3.9%	24.0%	54.3%	17.4%	100.0%
4	A	16.8%	35.6%	34.5%	13.0%	0.0%	0.0%	100.0%
	B	1.0%	7.5%	31.6%	56.6%	3.4%	0.0%	100.0%
	C	0.2%	0.5%	7.4%	37.1%	41.6%	13.1%	100.0%
5	A	26.3%	35.0%	30.2%	8.6%	0.0%	0.0%	100.0%
	B	1.5%	9.2%	38.7%	48.7%	1.9%	0.0%	100.0%
	C	0.2%	0.8%	14.0%	40.0%	34.8%	10.1%	100.0%
6	A	13.7%	41.0%	36.3%	8.9%	0.1%	0.0%	100.0%
	B	2.2%	16.7%	49.5%	30.6%	1.0%	0.0%	100.0%
	C	0.6%	5.7%	47.8%	35.3%	9.2%	1.4%	100.0%
7	A	23.5%	41.4%	30.5%	4.6%	0.1%	0.0%	100.0%
	B	3.3%	19.9%	51.8%	24.4%	0.6%	0.0%	100.0%
	C	0.8%	7.1%	49.2%	31.9%	9.6%	1.5%	100.0%
8	A	31.4%	42.5%	24.1%	1.9%	0.1%	0.0%	100.0%
	B	4.8%	24.4%	54.9%	15.8%	0.2%	0.0%	100.0%
	C	1.1%	9.4%	50.9%	28.0%	8.8%	1.8%	100.0%
9	A	18.7%	41.1%	31.5%	8.6%	0.1%	0.0%	100.0%
	B	3.3%	12.0%	37.7%	32.7%	14.1%	0.1%	100.0%
	C	0.8%	1.7%	11.0%	26.5%	39.3%	20.7%	100.0%
10	A	15.2%	43.0%	34.9%	6.9%	0.0%	0.0%	100.0%
	B	4.8%	20.5%	38.6%	28.6%	7.5%	0.0%	100.0%
	C	1.1%	3.0%	17.6%	33.7%	32.2%	12.4%	100.0%
11	A	16.6%	41.5%	36.7%	5.2%	0.0%	0.0%	100.0%
	B	5.8%	22.0%	39.8%	27.5%	4.9%	0.0%	100.0%
	C	1.5%	4.0%	20.1%	34.6%	28.1%	11.7%	100.0%
12	A	25.5%	41.4%	30.9%	2.2%	0.1%	0.0%	100.0%
	B	8.7%	27.9%	38.9%	22.6%	2.0%	0.0%	100.0%
	C	2.5%	5.9%	25.7%	35.7%	21.6%	8.5%	100.0%

4.3.6.3 By Grade

Table 4.3.6.3A

Proficiency Level by Grade (Count): Literacy S301

	Literacy Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	33,587	46,700	34,542	23,717	38,105	12,289	188,940
K (accountability)	121,482	22,515	22,417	15,657	6,869	0	188,940
1	10,322	49,385	99,427	13,807	4,386	839	178,166
2	6,497	28,692	96,319	22,993	11,096	1,340	166,937
3	1,585	7,435	23,905	58,988	39,780	10,993	142,686
4	1,838	6,121	19,273	42,719	22,364	6,630	98,945
5	2,387	5,700	20,461	33,316	15,443	4,312	81,619
6	1,785	9,437	32,894	21,416	3,554	509	69,595
7	2,723	10,388	32,113	17,365	3,326	504	66,419
8	3,514	11,421	30,244	12,501	2,776	550	61,006
9	3,468	8,807	16,911	17,317	15,266	6,159	67,928
10	2,034	6,935	12,555	12,166	7,680	2,436	43,806
11	1,710	5,421	10,480	9,894	5,348	1,936	34,789
12	1,768	4,572	8,258	7,350	3,192	1,177	26,317

Table 4.3.6.3B

Proficiency Level by Grade (Percent): Literacy S301

	Literacy Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	17.8%	24.7%	18.3%	12.6%	20.2%	6.5%	100.0%
K (accountability)	64.3%	11.9%	11.9%	8.3%	3.6%	0.0%	100.0%
1	5.8%	27.7%	55.8%	7.7%	2.5%	0.5%	100.0%
2	3.9%	17.2%	57.7%	13.8%	6.6%	0.8%	100.0%
3	1.1%	5.2%	16.8%	41.3%	27.9%	7.7%	100.0%
4	1.9%	6.2%	19.5%	43.2%	22.6%	6.7%	100.0%
5	2.9%	7.0%	25.1%	40.8%	18.9%	5.3%	100.0%
6	2.6%	13.6%	47.3%	30.8%	5.1%	0.7%	100.0%
7	4.1%	15.6%	48.3%	26.1%	5.0%	0.8%	100.0%
8	5.8%	18.7%	49.6%	20.5%	4.6%	0.9%	100.0%
9	5.1%	13.0%	24.9%	25.5%	22.5%	9.1%	100.0%
10	4.6%	15.8%	28.7%	27.8%	17.5%	5.6%	100.0%
11	4.9%	15.6%	30.1%	28.4%	15.4%	5.6%	100.0%
12	6.7%	17.4%	31.4%	27.9%	12.1%	4.5%	100.0%

4.3.7 Comprehension Composite

4.3.7.1 By Cluster by Tier

Table 4.3.7.1A

Proficiency Level by Cluster by Tier (Count): Comprehension S301

| Cluster | Tier | Comprehension Proficiency Range | | | | | | Total |
		1	2	3	4	5	6	
K (instructional)	-	23,823	28,466	40,666	26,872	28,135	40,976	188,938
K (accountability)	-	111,964	13,214	13,142	12,691	23,255	14,672	188,938
1-2	A	8,122	19,715	25,302	19,609	n/a	n/a	72,748
	B	1,392	4,050	35,719	52,419	94,424	n/a	188,004
	C	285	1,761	9,611	10,151	26,092	36,613	84,513
3-5	A	4,050	10,107	6,820	5,273	n/a	n/a	26,250
	B	821	8,393	37,626	31,830	62,284	n/a	140,954
	C	312	865	13,293	17,609	58,863	65,282	156,224
6-8	A	5,733	9,189	4,025	1,407	n/a	n/a	20,354
	B	1,406	16,693	30,796	15,140	13,133	n/a	77,168
	C	430	3,932	26,946	22,175	29,911	16,246	99,640
9-12	A	10,234	9,606	3,801	1,166	n/a	n/a	24,807
	B	4,102	20,561	20,250	14,032	9,570	n/a	68,515
	C	965	5,699	15,713	16,850	20,535	19,968	79,730

Table 4.3.7.1B

Proficiency Level by Cluster by Tier (Percent): Comprehension S301

| Cluster | Tier | Comprehension Proficiency Range | | | | | | Total |
		1	2	3	4	5	6	
K (instructional)	-	12.6%	15.1%	21.5%	14.2%	14.9%	21.7%	100.0%
K (accountability)	-	59.3%	7.0%	7.0%	6.7%	12.3%	7.8%	100.0%
1-2	A	11.2%	27.1%	34.8%	27.0%	n/a	n/a	100.0%
	B	0.7%	2.2%	19.0%	27.9%	50.2%	n/a	100.0%
	C	0.3%	2.1%	11.4%	12.0%	30.9%	43.3%	100.0%
3-5	A	15.4%	38.5%	26.0%	20.1%	n/a	n/a	100.0%
	B	0.6%	6.0%	26.7%	22.6%	44.2%	n/a	100.0%
	C	0.2%	0.6%	8.5%	11.3%	37.7%	41.8%	100.0%
6-8	A	28.2%	45.1%	19.8%	6.9%	n/a	n/a	100.0%
	B	1.8%	21.6%	39.9%	19.6%	17.0%	n/a	100.0%
	C	0.4%	3.9%	27.0%	22.3%	30.0%	16.3%	100.0%
9-12	A	41.3%	38.7%	15.3%	4.7%	n/a	n/a	100.0%
	B	6.0%	30.0%	29.6%	20.5%	14.0%	n/a	100.0%
	C	1.2%	7.1%	19.7%	21.1%	25.8%	25.0%	100.0%

4.3.7.2 By Grade by Tier

Table 4.3.7.2A

Proficiency Level by Grade by Tier (Count): Comprehension S301

Grade	Tier	Comprehension Proficiency Range 1	2	3	4	5	6	Total
K (instructional)	-	23,823	28,466	40,666	26,872	28,135	40,976	188,938
K (accountability)	-	111,964	13,214	13,142	12,691	23,255	14,672	188,938
1	A	5,227	15,099	20,905	15227	n/a	n/a	56,458
	B	930	1,535	16,104	30,543	43,757	n/a	92,869
	C	171	582	3,859	4,211	9,652	10,461	28,936
2	A	2,895	4,616	4,397	4382	n/a	n/a	16,290
	B	462	2,515	19,615	21,876	50,667	n/a	95,135
	C	114	1,179	5,752	5,940	16,440	26,152	55,577
3	A	983	4,489	3,283	2,772	n/a	n/a	11,527
	B	282	2,284	15,238	15,679	34,600	n/a	68,083
	C	103	119	1,986	4,491	24,912	31,541	63,152
4	A	1,318	3,159	1,965	1,475	n/a	n/a	7,917
	B	242	2,839	12,692	9,633	15,217	n/a	40,623
	C	112	243	4,866	6,582	18,636	20,014	50,453
5	A	1,749	2,459	1,572	1,026	n/a	n/a	6,806
	B	297	3,270	9,696	6,518	12,467	n/a	32,248
	C	97	503	6,441	6,536	15,315	13,727	42,619
6	A	1,333	3,365	1,847	580	n/a	n/a	7,125
	B	254	4,686	11,828	5,254	4,960	n/a	26,982
	C	121	819	9,782	8,356	11,226	5,233	35,537
7	A	1,956	3,024	1,288	445	n/a	n/a	6,713
	B	430	5,681	10,846	5,565	4,017	n/a	26,539
	C	153	1,323	8,869	8,158	9,393	5,322	33,218
8	A	2,444	2,800	890	382	n/a	n/a	6,516
	B	722	6,326	8,122	4,321	4,156	n/a	23,647
	C	156	1,790	8,295	5,661	9,292	5,691	30,885
9	A	5,333	4,914	1,906	643	n/a	n/a	12,796
	B	828	6,202	9,306	5,172	4,027	n/a	25,535
	C	221	1,197	5,682	5,566	8,553	8,473	29,692
10	A	2,362	2,657	997	255	n/a	n/a	6,271
	B	934	5,887	5,161	3,586	2,334	n/a	17,902
	C	196	1,425	4,382	4,359	4,860	4,462	19,684
11	A	1,582	1,436	577	174	n/a	n/a	3,769
	B	1,090	4,951	3,543	2,870	2,086	n/a	14,540
	C	218	1,575	3,121	3,565	4,029	4,009	16,517
12	A	957	599	321	94	n/a	n/a	1,971
	B	1,250	3,521	2,240	2,404	1,123	n/a	10,538
	C	330	1,502	2,528	3,360	3,093	3,024	13,837

Table 4.3.7.2B

Proficiency Level by Grade by Tier (Percent): Comprehension S301

Grade	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	12.6%	15.1%	21.5%	14.2%	14.9%	21.7%	100.0%
K (accountability)	-	59.3%	7.0%	7.0%	6.7%	12.3%	7.8%	100.0%
1	A	9.3%	26.7%	37.0%	27.0%	n/a	n/a	100.0%
	B	1.0%	1.7%	17.3%	32.9%	47.1%	n/a	100.0%
	C	0.6%	2.0%	13.3%	14.6%	33.4%	36.2%	100.0%
2	A	17.8%	28.3%	27.0%	26.9%	n/a	n/a	100.0%
	B	0.5%	2.6%	20.6%	23.0%	53.3%	n/a	100.0%
	C	0.2%	2.1%	10.3%	10.7%	29.6%	47.1%	100.0%
3	A	8.5%	38.9%	28.5%	24.0%	n/a	n/a	100.0%
	B	0.4%	3.4%	22.4%	23.0%	50.8%	n/a	100.0%
	C	0.2%	0.2%	3.1%	7.1%	39.4%	49.9%	100.0%
4	A	16.6%	39.9%	24.8%	18.6%	n/a	n/a	100.0%
	B	0.6%	7.0%	31.2%	23.7%	37.5%	n/a	100.0%
	C	0.2%	0.5%	9.6%	13.0%	36.9%	39.7%	100.0%
5	A	25.7%	36.1%	23.1%	15.1%	n/a	n/a	100.0%
	B	0.9%	10.1%	30.1%	20.2%	38.7%	n/a	100.0%
	C	0.2%	1.2%	15.1%	15.3%	35.9%	32.2%	100.0%
6	A	18.7%	47.2%	25.9%	8.1%	n/a	n/a	100.0%
	B	0.9%	17.4%	43.8%	19.5%	18.4%	n/a	100.0%
	C	0.3%	2.3%	27.5%	23.5%	31.6%	14.7%	100.0%
7	A	29.1%	45.0%	19.2%	6.6%	n/a	n/a	100.0%
	B	1.6%	21.4%	40.9%	21.0%	15.1%	n/a	100.0%
	C	0.5%	4.0%	26.7%	24.6%	28.3%	16.0%	100.0%
8	A	37.5%	43.0%	13.7%	5.9%	n/a	n/a	100.0%
	B	3.1%	26.8%	34.3%	18.3%	17.6%	n/a	100.0%
	C	0.5%	5.8%	26.9%	18.3%	30.1%	18.4%	100.0%
9	A	41.7%	38.4%	14.9%	5.0%	n/a	n/a	100.0%
	B	3.2%	24.3%	36.4%	20.3%	15.8%	n/a	100.0%
	C	0.7%	4.0%	19.1%	18.7%	28.8%	28.5%	100.0%
10	A	37.7%	42.4%	15.9%	4.1%	n/a	n/a	100.0%
	B	5.2%	32.9%	28.8%	20.0%	13.0%	n/a	100.0%
	C	1.0%	7.2%	22.3%	22.1%	24.7%	22.7%	100.0%
11	A	42.0%	38.1%	15.3%	4.6%	n/a	n/a	100.0%
	B	7.5%	34.1%	24.4%	19.7%	14.3%	n/a	100.0%
	C	1.3%	9.5%	18.9%	21.6%	24.4%	24.3%	100.0%
12	A	48.6%	30.4%	16.3%	4.8%	n/a	n/a	100.0%
	B	11.9%	33.4%	21.3%	22.8%	10.7%	n/a	100.0%
	C	2.4%	10.9%	18.3%	24.3%	22.4%	21.9%	100.0%

4.3.7.3 By Grade

Table 4.3.7.3A

Proficiency Level by Grade (Count): Comprehension S301

	Comprehension Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
K (instructional)	23,823	28,466	40,666	26,872	28,135	40,976	188,938
K (accountability)	111,964	13,214	13,142	12,691	23,255	14,672	188,938
1	6,328	17,216	40,868	49,981	53,409	10,461	178,263
2	3,471	8,310	29,764	32,198	67,107	26,152	167,002
3	1,368	6,892	20,507	22,942	59,512	31,541	142,762
4	1,672	6,241	19,523	17,690	33,853	20,014	98,993
5	2,143	6,232	17,709	14,080	27,782	13,727	81,673
6	1,708	8,870	23,457	14,190	16,186	5,233	69,644
7	2,539	10,028	21,003	14,168	13,410	5,322	66,470
8	3,322	10,916	17,307	10,364	13,448	5,691	61,048
9	6,382	12,313	16,894	11,381	12,580	8,473	68,023
10	3,492	9,969	10,540	8,200	7,194	4,462	43,857
11	2,890	7,962	7,241	6,609	6,115	4,009	34,826
12	2,537	5,622	5,089	5,858	4,216	3,024	26,346

Table 4.3.7.3B

Proficiency Level by Grade (Percent): Comprehension S301

	Comprehension Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
K (instructional)	12.6%	15.1%	21.5%	14.2%	14.9%	21.7%	100.0%
K (accountability)	59.3%	7.0%	7.0%	6.7%	12.3%	7.8%	100.0%
1	3.5%	9.7%	22.9%	28.0%	30.0%	5.9%	100.0%
2	2.1%	5.0%	17.8%	19.3%	40.2%	15.7%	100.0%
3	1.0%	4.8%	14.4%	16.1%	41.7%	22.1%	100.0%
4	1.7%	6.3%	19.7%	17.9%	34.2%	20.2%	100.0%
5	2.6%	7.6%	21.7%	17.2%	34.0%	16.8%	100.0%
6	2.5%	12.7%	33.7%	20.4%	23.2%	7.5%	100.0%
7	3.8%	15.1%	31.6%	21.3%	20.2%	8.0%	100.0%
8	5.4%	17.9%	28.3%	17.0%	22.0%	9.3%	100.0%
9	9.4%	18.1%	24.8%	16.7%	18.5%	12.5%	100.0%
10	8.0%	22.7%	24.0%	18.7%	16.4%	10.2%	100.0%
11	8.3%	22.9%	20.8%	19.0%	17.6%	11.5%	100.0%
12	9.6%	21.3%	19.3%	22.2%	16.0%	11.5%	100.0%

4.3.8 Overall Composite

4.3.8.1 By Cluster by Tier

Table 4.3.8.1A

Proficiency Level by Cluster by Tier (Count): Overall S301

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	27,484	38,690	37,378	33,174	39,785	12,415	188,926
K (accountability)	-	101,314	29,853	27,183	18,161	10,824	1,591	188,926
1-2	A	8,851	30,201	33,432	180	0	0	72,664
	B	1,569	19,991	113,732	52,450	22	0	187,764
	C	242	2,471	19,513	30,715	26,252	5,251	84,444
3-5	A	4,347	8,652	9,157	3,923	96	0	26,175
	B	670	6,663	39,479	77,918	16,003	0	140,733
	C	214	480	7,854	42,722	76,214	28,579	156,063
6-8	A	5,514	8,117	5,368	1,286	22	0	20,307
	B	1,304	11,654	34,325	27,341	2,268	0	76,892
	C	382	2,345	26,360	46,164	21,176	2,982	99,409
9-12	A	7,217	10,463	5,631	1,307	35	0	24,653
	B	2,214	10,342	24,090	22,654	8,654	21	67,975
	C	738	1,674	10,580	25,672	28,132	12,312	79,108

Table 4.3.8.1B

Proficiency Level by Cluster by Tier (Percent): Overall S301

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	14.5%	20.5%	19.8%	17.6%	21.1%	6.6%	100.0%
K (accountability)	-	53.6%	15.8%	14.4%	9.6%	5.7%	0.8%	100.0%
1-2	A	12.2%	41.6%	46.0%	0.2%	0.0%	0.0%	100.0%
	B	0.8%	10.6%	60.6%	27.9%	0.0%	0.0%	100.0%
	C	0.3%	2.9%	23.1%	36.4%	31.1%	6.2%	100.0%
3-5	A	16.6%	33.1%	35.0%	15.0%	0.4%	0.0%	100.0%
	B	0.5%	4.7%	28.1%	55.4%	11.4%	0.0%	100.0%
	C	0.1%	0.3%	5.0%	27.4%	48.8%	18.3%	100.0%
6-8	A	27.2%	40.0%	26.4%	6.3%	0.1%	0.0%	100.0%
	B	1.7%	15.2%	44.6%	35.6%	2.9%	0.0%	100.0%
	C	0.4%	2.4%	26.5%	46.4%	21.3%	3.0%	100.0%
9-12	A	29.3%	42.4%	22.8%	5.3%	0.1%	0.0%	100.0%
	B	3.3%	15.2%	35.4%	33.3%	12.7%	0.0%	100.0%
	C	0.9%	2.1%	13.4%	32.5%	35.6%	15.6%	100.0%

4.3.8.2 By Grade by Tier

Table 4.3.8.2A

Proficiency Level by Grade by Tier (Count): Overall S301

Grade	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	27,484	38,690	37,378	33,174	39,785	12,415	188,926
K (accountability)	-	101,314	29,853	27,183	18,161	10,824	1,591	188,926
1	A	5,395	23,651	27,193	162	0	0	56,401
	B	882	11,121	59,465	21,238	16	0	92,722
	C	112	1,238	8,474	10,127	7,289	1,671	28,911
2	A	3,456	6,550	6,239	18	0	0	16,263
	B	687	8,870	54,267	31,212	6	0	95,042
	C	130	1,233	11,039	20,588	18,963	3,580	55,533
3	A	1,225	3,556	4,363	2,269	83	0	11,496
	B	211	2,338	17,015	38,254	10,178	0	67,996
	C	65	113	1,612	12,427	34,373	14,499	63,089
4	A	1,424	2,744	2,722	990	12	0	7,892
	B	217	2,200	12,167	21,894	4,066	0	40,544
	C	84	169	2,683	15,295	23,652	8,524	50,407
5	A	1,698	2,352	2,072	664	1	0	6,787
	B	242	2,125	10,297	17,770	1,759	0	32,193
	C	65	198	3,559	15,000	18,189	5,556	42,567
6	A	1,415	2,788	2,227	664	17	0	7,111
	B	299	3,298	11,332	10,921	1,058	0	26,908
	C	108	604	8,486	17,701	7,508	1,059	35,466
7	A	1,884	2,691	1,759	364	3	0	6,701
	B	424	3,866	12,274	8,959	915	0	26,438
	C	127	744	9,220	14,661	7,409	973	33,134
8	A	2,215	2,638	1,382	258	2	0	6,495
	B	581	4,490	10,719	7,461	295	0	23,546
	C	147	997	8,654	13,802	6,259	950	30,809
9	A	4,037	5,143	2,786	728	25	0	12,719
	B	527	2,474	7,950	9,278	5,090	15	25,334
	C	186	321	2,301	7,574	12,524	6,575	29,481
10	A	1,536	2,888	1,467	338	7	0	6,236
	B	497	3,019	6,432	5,759	2,072	4	17,783
	C	138	371	2,634	6,816	7,026	2,562	19,547
11	A	997	1,670	917	166	2	0	3,752
	B	541	2,636	5,494	4,637	1,117	1	14,426
	C	159	430	2,698	5,947	5,170	1,986	16,390
12	A	647	762	461	75	1	0	1,946
	B	649	2,213	4,214	2,980	375	1	10,432
	C	255	552	2,947	5,335	3,412	1,189	13,690

Table 4.3.8.2B
Proficiency Level by Grade by Tier (Percent): Overall S301

Grade	Tier	Overall Proficiency Range						Total
		1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	-	14.5%	20.5%	19.8%	17.6%	21.1%	6.6%	100.0%
K (accountability)	-	53.6%	15.8%	14.4%	9.6%	5.7%	0.8%	100.0%
1	A	9.6%	41.9%	48.2%	0.3%	0.0%	0.0%	100.0%
	B	1.0%	12.0%	64.1%	22.9%	0.0%	0.0%	100.0%
	C	0.4%	4.3%	29.3%	35.0%	25.2%	5.8%	100.0%
2	A	21.3%	40.3%	38.4%	0.1%	0.0%	0.0%	100.0%
	B	0.7%	9.3%	57.1%	32.8%	0.0%	0.0%	100.0%
	C	0.2%	2.2%	19.9%	37.1%	34.1%	6.4%	100.0%
3	A	10.7%	30.9%	38.0%	19.7%	0.7%	0.0%	100.0%
	B	0.3%	3.4%	25.0%	56.3%	15.0%	0.0%	100.0%
	C	0.1%	0.2%	2.6%	19.7%	54.5%	23.0%	100.0%
4	A	18.0%	34.8%	34.5%	12.5%	0.2%	0.0%	100.0%
	B	0.5%	5.4%	30.0%	54.0%	10.0%	0.0%	100.0%
	C	0.2%	0.3%	5.3%	30.3%	46.9%	16.9%	100.0%
5	A	25.0%	34.7%	30.5%	9.8%	0.0%	0.0%	100.0%
	B	0.8%	6.6%	32.0%	55.2%	5.5%	0.0%	100.0%
	C	0.2%	0.5%	8.4%	35.2%	42.7%	13.1%	100.0%
6	A	19.9%	39.2%	31.3%	9.3%	0.2%	0.0%	100.0%
	B	1.1%	12.3%	42.1%	40.6%	3.9%	0.0%	100.0%
	C	0.3%	1.7%	23.9%	49.9%	21.2%	3.0%	100.0%
7	A	28.1%	40.2%	26.2%	5.4%	0.0%	0.0%	100.0%
	B	1.6%	14.6%	46.4%	33.9%	3.5%	0.0%	100.0%
	C	0.4%	2.2%	27.8%	44.2%	22.4%	2.9%	100.0%
8	A	34.1%	40.6%	21.3%	4.0%	0.0%	0.0%	100.0%
	B	2.5%	19.1%	45.5%	31.7%	1.3%	0.0%	100.0%
	C	0.5%	3.2%	28.1%	44.8%	20.3%	3.1%	100.0%
9	A	31.7%	40.4%	21.9%	5.7%	0.2%	0.0%	100.0%
	B	2.1%	9.8%	31.4%	36.6%	20.1%	0.1%	100.0%
	C	0.6%	1.1%	7.8%	25.7%	42.5%	22.3%	100.0%
10	A	24.6%	46.3%	23.5%	5.4%	0.1%	0.0%	100.0%
	B	2.8%	17.0%	36.2%	32.4%	11.7%	0.0%	100.0%
	C	0.7%	1.9%	13.5%	34.9%	35.9%	13.1%	100.0%
11	A	26.6%	44.5%	24.4%	4.4%	0.1%	0.0%	100.0%
	B	3.8%	18.3%	38.1%	32.1%	7.7%	0.0%	100.0%
	C	1.0%	2.6%	16.5%	36.3%	31.5%	12.1%	100.0%
12	A	33.2%	39.2%	23.7%	3.9%	0.1%	0.0%	100.0%
	B	6.2%	21.2%	40.4%	28.6%	3.6%	0.0%	100.0%
	C	1.9%	4.0%	21.5%	39.0%	24.9%	8.7%	100.0%

4.3.8.3 By Grade

Table 4.3.8.3A

Proficiency Level by Grade (Count): Overall S301

	Overall Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
K (instructional)	27,484	38,690	37,378	33,174	39,785	12,415	188,926
K (accountability)	101,314	29,853	27,183	18,161	10,824	1,591	188,926
1	6,389	36,010	95,132	31,527	7,305	1,671	178,034
2	4,273	16,653	71,545	51,818	18,969	3,580	166,838
3	1,501	6,007	22,990	52,950	44,634	14,499	142,581
4	1,725	5,113	17,572	38,179	27,730	8,524	98,843
5	2,005	4,675	15,928	33,434	19,949	5,556	81,547
6	1,822	6,690	22,045	29,286	8,583	1,059	69,485
7	2,435	7,301	23,253	23,984	8,327	973	66,273
8	2,943	8,125	20,755	21,521	6,556	950	60,850
9	4,750	7,938	13,037	17,580	17,639	6,590	67,534
10	2,171	6,278	10,533	12,913	9,105	2,566	43,566
11	1,697	4,736	9,109	10,750	6,289	1,987	34,568
12	1,551	3,527	7,622	8,390	3,788	1,190	26,068

Table 4.3.8.3B

Proficiency Level by Grade (Percent): Overall S301

	Overall Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
K (instructional)	14.5%	20.5%	19.8%	17.6%	21.1%	6.6%	100.0%
K (accountability)	53.6%	15.8%	14.4%	9.6%	5.7%	0.8%	100.0%
1	3.6%	20.2%	53.4%	17.7%	4.1%	0.9%	100.0%
2	2.6%	10.0%	42.9%	31.1%	11.4%	2.1%	100.0%
3	1.1%	4.2%	16.1%	37.1%	31.3%	10.2%	100.0%
4	1.7%	5.2%	17.8%	38.6%	28.1%	8.6%	100.0%
5	2.5%	5.7%	19.5%	41.0%	24.5%	6.8%	100.0%
6	2.6%	9.6%	31.7%	42.1%	12.4%	1.5%	100.0%
7	3.7%	11.0%	35.1%	36.2%	12.6%	1.5%	100.0%
8	4.8%	13.4%	34.1%	35.4%	10.8%	1.6%	100.0%
9	7.0%	11.8%	19.3%	26.0%	26.1%	9.8%	100.0%
10	5.0%	14.4%	24.2%	29.6%	20.9%	5.9%	100.0%
11	4.9%	13.7%	26.4%	31.1%	18.2%	5.7%	100.0%
12	5.9%	13.5%	29.2%	32.2%	14.5%	4.6%	100.0%

4.4 Mean Raw Score Results by Standards

4.4.1 Comprehension Composite

4.4.1.1 By Cluster

Table 4.4.1.1A

Mean Raw Score by Cluster by Tier by Standard: Comprehension S301

Cluster	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
1-2	A	Social Instructional Language	12	7.55	62.90%
		Language of Language Arts	9	5.99	66.58%
		Language of Math	9	5.10	56.71%
		Language of Science	6	3.97	66.24%
		Language of Social Studies	6	3.21	53.46%
	B	Social Instructional Language	6	3.88	64.59%
		Language of Language Arts	12	6.59	54.89%
		Language of Math	12	7.66	63.84%
		Language of Science	9	5.91	65.66%
		Language of Social Studies	9	5.63	62.54%
	C	Social Instructional Language	6	4.88	81.27%
		Language of Language Arts	12	8.40	70.02%
		Language of Math	12	7.43	61.95%
		Language of Science	9	6.05	67.27%
		Language of Social Studies	9	5.72	63.54%
3-5	A	Social Instructional Language	12	6.95	57.89%
		Language of Language Arts	9	4.72	52.41%
		Language of Math	9	4.60	51.13%
		Language of Science	6	3.86	64.28%
		Language of Social Studies	6	3.43	57.19%
	B	Social Instructional Language	6	4.15	69.10%
		Language of Language Arts	12	8.19	68.23%
		Language of Math	12	6.75	56.28%
		Language of Science	9	5.26	58.50%
		Language of Social Studies	9	4.77	53.01%
	C	Social Instructional Language	6	3.32	55.27%
		Language of Language Arts	12	7.63	63.58%
		Language of Math	12	6.05	50.42%
		Language of Science	9	4.76	52.88%
		Language of Social Studies	9	4.48	49.77%

Cluster	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
6-8	A	Social Instructional Language	12	6.54	54.46%
		Language of Language Arts	9	5.48	60.89%
		Language of Math	9	4.54	50.39%
		Language of Science	6	3.61	60.09%
		Language of Social Studies	6	3.18	52.94%
	B	Social Instructional Language	6	3.35	55.85%
		Language of Language Arts	12	8.61	71.75%
		Language of Math	12	7.01	58.40%
		Language of Science	9	4.78	53.13%
		Language of Social Studies	9	4.85	53.92%
	C	Social Instructional Language	6	3.46	57.69%
		Language of Language Arts	12	7.73	64.44%
		Language of Math	12	7.21	60.06%
		Language of Science	9	4.70	52.17%
		Language of Social Studies	9	3.55	39.43%
9-12	A	Social Instructional Language	12	7.23	60.24%
		Language of Language Arts	9	5.30	58.89%
		Language of Math	9	4.70	52.20%
		Language of Science	6	3.23	53.88%
		Language of Social Studies	6	3.87	64.54%
	B	Social Instructional Language	6	4.35	72.48%
		Language of Language Arts	12	8.68	72.29%
		Language of Math	12	6.74	56.19%
		Language of Science	9	5.13	57.04%
		Language of Social Studies	9	5.09	56.57%
	C	Social Instructional Language	6	4.14	69.00%
		Language of Language Arts	12	6.57	54.77%
		Language of Math	12	6.93	57.73%
		Language of Science	9	5.35	59.42%
		Language of Social Studies	9	4.77	52.95%

4.4.1.2 By Grade

Table 4.4.1.2A

Mean Raw Score by Grade by Tier by Standard: Comprehension S301

Grade	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
1	A	Social Instructional Language	12	7.47	62.24%
		Language of Language Arts	9	5.90	65.52%
		Language of Math	9	4.97	55.20%
		Language of Science	6	3.92	65.35%
		Language of Social Studies	6	3.15	52.42%
	B	Social Instructional Language	6	3.51	58.43%
		Language of Language Arts	12	5.71	47.60%
		Language of Math	12	7.03	58.54%
		Language of Science	9	5.26	58.42%
		Language of Social Studies	9	4.97	55.26%
	C	Social Instructional Language	6	4.45	74.10%
		Language of Language Arts	12	7.22	60.15%
		Language of Math	12	6.43	53.61%
		Language of Science	9	5.46	60.70%
		Language of Social Studies	9	4.69	52.08%
2	A	Social Instructional Language	12	7.82	65.21%
		Language of Language Arts	9	6.32	70.25%
		Language of Math	9	5.57	61.92%
		Language of Science	6	4.16	69.29%
		Language of Social Studies	6	3.42	57.05%
	B	Social Instructional Language	6	4.24	70.61%
		Language of Language Arts	12	7.44	62.00%
		Language of Math	12	8.28	69.02%
		Language of Science	9	6.55	72.73%
		Language of Social Studies	9	6.27	69.64%
	C	Social Instructional Language	6	5.10	85.00%
		Language of Language Arts	12	9.02	75.16%
		Language of Math	12	7.96	66.30%
		Language of Science	9	6.36	70.69%
		Language of Social Studies	9	6.26	69.51%

Grade	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
3	A	Social Instructional Language	12	6.77	56.44%
		Language of Language Arts	9	4.60	51.11%
		Language of Math	9	4.43	49.21%
		Language of Science	6	3.73	62.10%
		Language of Social Studies	6	3.23	53.87%
	B	Social Instructional Language	6	3.93	65.44%
		Language of Language Arts	12	7.67	63.94%
		Language of Math	12	6.26	52.17%
		Language of Science	9	4.91	54.58%
		Language of Social Studies	9	4.41	49.02%
	C	Social Instructional Language	6	2.99	49.82%
		Language of Language Arts	12	7.15	59.60%
		Language of Math	12	5.52	45.97%
		Language of Science	9	4.40	48.94%
		Language of Social Studies	9	3.99	44.37%
4	A	Social Instructional Language	12	6.96	58.03%
		Language of Language Arts	9	4.70	52.27%
		Language of Math	9	4.64	51.56%
		Language of Science	6	3.88	64.58%
		Language of Social Studies	6	3.47	57.76%
	B	Social Instructional Language	6	4.22	70.41%
		Language of Language Arts	12	8.37	69.76%
		Language of Math	12	6.93	57.76%
		Language of Science	9	5.38	59.78%
		Language of Social Studies	9	4.86	54.00%
	C	Social Instructional Language	6	3.38	56.28%
		Language of Language Arts	12	7.70	64.19%
		Language of Math	12	6.15	51.28%
		Language of Science	9	4.80	53.39%
		Language of Social Studies	9	4.55	50.57%

Grade	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
5	A	Social Instructional Language	12	7.22	60.19%
		Language of Language Arts	9	4.93	54.77%
		Language of Math	9	4.85	53.88%
		Language of Science	6	4.06	67.59%
		Language of Social Studies	6	3.73	62.15%
	B	Social Instructional Language	6	4.51	75.17%
		Language of Language Arts	12	9.04	75.32%
		Language of Math	12	7.57	63.08%
		Language of Science	9	5.86	65.15%
		Language of Social Studies	9	5.42	60.18%
	C	Social Instructional Language	6	3.73	62.15%
		Language of Language Arts	12	8.25	68.74%
		Language of Math	12	6.72	55.97%
		Language of Science	9	5.23	58.11%
		Language of Social Studies	9	5.12	56.85%
6	A	Social Instructional Language	12	6.49	54.08%
		Language of Language Arts	9	5.32	59.14%
		Language of Math	9	4.40	48.94%
		Language of Science	6	3.49	58.19%
		Language of Social Studies	6	3.06	50.94%
	B	Social Instructional Language	6	3.16	52.63%
		Language of Language Arts	12	8.27	68.88%
		Language of Math	12	6.63	55.21%
		Language of Science	9	4.53	50.31%
		Language of Social Studies	9	4.47	49.64%
	C	Social Instructional Language	6	3.20	53.32%
		Language of Language Arts	12	7.05	58.71%
		Language of Math	12	6.69	55.76%
		Language of Science	9	4.26	47.37%
		Language of Social Studies	9	3.10	34.43%

Grade	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
7	A	Social Instructional Language	12	6.50	54.16%
		Language of Language Arts	9	5.47	60.78%
		Language of Math	9	4.55	50.52%
		Language of Science	6	3.63	60.54%
		Language of Social Studies	6	3.19	53.13%
	B	Social Instructional Language	6	3.36	55.99%
		Language of Language Arts	12	8.67	72.25%
		Language of Math	12	7.05	58.76%
		Language of Science	9	4.83	53.67%
		Language of Social Studies	9	4.88	54.20%
	C	Social Instructional Language	6	3.49	58.24%
		Language of Language Arts	12	7.80	65.01%
		Language of Math	12	7.25	60.38%
		Language of Science	9	4.72	52.41%
		Language of Social Studies	9	3.56	39.59%
8	A	Social Instructional Language	12	6.62	55.18%
		Language of Language Arts	9	5.66	62.90%
		Language of Math	9	4.67	51.83%
		Language of Science	6	3.70	61.70%
		Language of Social Studies	6	3.29	54.91%
	B	Social Instructional Language	6	3.56	59.38%
		Language of Language Arts	12	8.93	74.45%
		Language of Math	12	7.40	61.65%
		Language of Science	9	5.02	55.77%
		Language of Social Studies	9	5.27	58.50%
	C	Social Instructional Language	6	3.73	62.11%
		Language of Language Arts	12	8.45	70.43%
		Language of Math	12	7.76	64.66%
		Language of Science	9	5.17	57.44%
		Language of Social Studies	9	4.05	45.01%

Grade	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
9	A	Social Instructional Language	12	6.88	57.32%
		Language of Language Arts	9	5.02	55.79%
		Language of Math	9	4.49	49.85%
		Language of Science	6	3.08	51.38%
		Language of Social Studies	6	3.72	62.06%
	B	Social Instructional Language	6	4.38	73.08%
		Language of Language Arts	12	8.69	72.39%
		Language of Math	12	6.68	55.66%
		Language of Science	9	5.13	57.00%
		Language of Social Studies	9	5.10	56.71%
	C	Social Instructional Language	6	4.25	70.83%
		Language of Language Arts	12	6.57	54.72%
		Language of Math	12	6.90	57.47%
		Language of Science	9	5.37	59.68%
		Language of Social Studies	9	4.70	52.21%
10	A	Social Instructional Language	12	7.50	62.52%
		Language of Language Arts	9	5.49	61.00%
		Language of Math	9	4.86	53.97%
		Language of Science	6	3.33	55.50%
		Language of Social Studies	6	4.00	66.59%
	B	Social Instructional Language	6	4.32	72.01%
		Language of Language Arts	12	8.58	71.47%
		Language of Math	12	6.75	56.28%
		Language of Science	9	5.09	56.55%
		Language of Social Studies	9	5.05	56.16%
	C	Social Instructional Language	6	4.11	68.49%
		Language of Language Arts	12	6.49	54.10%
		Language of Math	12	6.83	56.88%
		Language of Science	9	5.30	58.88%
		Language of Social Studies	9	4.70	52.26%

Grade	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
11	A	Social Instructional Language	12	7.76	64.68%
		Language of Language Arts	9	5.71	63.45%
		Language of Math	9	5.09	56.61%
		Language of Science	6	3.48	57.96%
		Language of Social Studies	6	4.13	68.83%
	B	Social Instructional Language	6	4.36	72.67%
		Language of Language Arts	12	8.79	73.21%
		Language of Math	12	6.86	57.15%
		Language of Science	9	5.21	57.84%
		Language of Social Studies	9	5.12	56.93%
	C	Social Instructional Language	6	4.10	68.33%
		Language of Language Arts	12	6.67	55.58%
		Language of Math	12	7.10	59.19%
		Language of Science	9	5.42	60.18%
		Language of Social Studies	9	4.90	54.40%
12	A	Social Instructional Language	12	7.58	63.14%
		Language of Language Arts	9	5.69	63.24%
		Language of Math	9	4.79	53.26%
		Language of Science	6	3.41	56.89%
		Language of Social Studies	6	3.94	65.68%
	B	Social Instructional Language	6	4.30	71.58%
		Language of Language Arts	12	8.66	72.18%
		Language of Math	12	6.72	56.00%
		Language of Science	9	5.12	56.86%
		Language of Social Studies	9	5.08	56.39%
	C	Social Instructional Language	6	4.00	66.63%
		Language of Language Arts	12	6.58	54.86%
		Language of Math	12	6.93	57.78%
		Language of Science	9	5.29	58.73%
		Language of Social Studies	9	4.84	53.77%

4.4.2 Writing

4.4.2.1 By Cluster

Table 4.4.2.1A

Mean Raw Score by Cluster by Tier by Standard: Writing S301

Cluster	Tier	Standard	Linguistic Complexity	Vocabulary Usage	Language Control	Total	Percent of Maximum
1-2	A	Social Instructional Language	5.16	4.79	4.22	14.17	19.68%
	B	Social Instructional Language	1.18	1.91	1.19	4.28	23.80%
		Language of Math / Science	2.37	2.22	1.93	6.53	36.27%
		Language of Language Arts / Social Studies	2.60	2.42	1.86	6.88	38.24%
	C	Social Instructional Language	2.77	2.52	2.19	7.47	41.50%
		Language of Math / Science	2.94	2.79	2.31	8.04	44.65%
		Language of Language Arts / Social Studies	3.10	2.81	2.36	8.27	45.92%
3-5	A	Social Instructional Language	2.13	1.91	1.58	5.63	31.27%
		Language of Math / Science	2.19	2.25	1.76	6.20	34.42%
		Language of Language Arts	2.20	1.96	1.67	5.82	32.36%
	B	Social Instructional Language	3.00	3.18	2.58	8.76	48.66%
		Language of Math / Science	2.98	3.09	2.66	8.73	48.50%
		Language of Language Arts / Social Studies	2.75	2.50	2.29	7.54	41.91%
	C	Social Instructional Language	3.28	3.52	2.92	9.72	53.98%
		Language of Math / Science	3.17	2.99	2.79	8.94	49.68%
		Language of Language Arts / Social Studies	3.12	2.78	2.65	8.55	47.49%
6-8	A	Social Instructional Language	2.32	2.01	1.93	6.25	34.72%
		Language of Math / Science	2.27	1.75	1.89	5.91	32.84%
		Language of Language Arts	2.44	2.16	1.88	6.48	36.00%
	B	Social Instructional Language	3.32	2.96	2.82	9.09	50.51%
		Language of Math / Science	3.18	3.40	2.84	9.43	52.36%
		Language of Language Arts / Social Studies	3.41	2.89	2.81	9.12	50.65%
	C	Social Instructional Language	3.66	3.18	3.22	10.05	55.86%
		Language of Math / Science	3.69	3.80	3.28	10.77	59.84%
		Language of Language Arts / Social Studies	3.82	3.25	3.26	10.33	57.38%
9-12	A	Social Instructional Language	2.26	2.11	1.96	6.33	35.19%
		Language of Math / Science	2.41	2.30	1.84	6.55	36.37%
		Language of Language Arts	2.53	2.20	1.85	6.58	36.56%
	B	Social Instructional Language	3.52	2.98	2.96	9.46	52.56%
		Language of Math / Science	3.36	2.98	2.86	9.20	51.11%
		Language of Language Arts / Social Studies	3.32	3.16	2.75	9.23	51.27%
	C	Social Instructional Language	3.86	3.35	3.39	10.61	58.92%
		Language of Math / Science	3.40	3.62	3.03	10.04	55.80%
		Language of Language Arts / Social Studies	3.79	3.56	3.23	10.59	58.81%

4.4.2.2 By Grade

Table 4.4.2.2A

Mean Raw Score by Grade by Tier by Standard: Writing S301

Grade	Tier	Standard	Mean Raw Score			Total	Percent of Maximum
			Linguistic Complexity	Vocabulary Usage	Language Control		
1	A	Social Instructional Language	5.10	4.73	4.12	13.95	19.37%
	B	Social Instructional Language	1.20	1.86	1.11	4.18	23.20%
		Language of Math / Science	2.14	2.04	1.73	5.91	32.83%
		Language of Language Arts / Social Studies	2.32	2.20	1.60	6.12	33.97%
	C	Social Instructional Language	2.58	2.27	1.93	6.78	37.65%
		Language of Math / Science	2.61	2.61	2.00	7.22	40.11%
		Language of Language Arts / Social Studies	2.80	2.57	2.04	7.40	41.12%
2	A	Social Instructional Language	5.40	4.98	4.56	14.94	20.75%
	B	Social Instructional Language	1.16	1.96	1.28	4.39	24.38%
		Language of Math / Science	2.60	2.40	2.13	7.13	39.62%
		Language of Language Arts / Social Studies	2.88	2.64	2.12	7.63	42.39%
	C	Social Instructional Language	2.87	2.64	2.32	7.83	43.51%
		Language of Math / Science	3.11	2.88	2.47	8.46	47.01%
		Language of Language Arts / Social Studies	3.25	2.94	2.53	8.72	48.43%
3	A	Social Instructional Language	2.06	1.84	1.50	5.40	30.02%
		Language of Math / Science	2.13	2.21	1.70	6.05	33.59%
		Language of Language Arts	2.13	1.92	1.61	5.67	31.49%
	B	Social Instructional Language	2.88	3.04	2.45	8.37	46.50%
		Language of Math / Science	2.90	2.94	2.55	8.39	46.60%
		Language of Language Arts / Social Studies	2.56	2.36	2.12	7.04	39.12%
	C	Social Instructional Language	3.15	3.39	2.77	9.31	51.75%
		Language of Math / Science	3.04	2.92	2.67	8.63	47.95%
		Language of Language Arts / Social Studies	2.90	2.64	2.46	8.00	44.46%
4	A	Social Instructional Language	2.14	1.92	1.59	5.64	31.35%
		Language of Math / Science	2.18	2.24	1.76	6.18	34.33%
		Language of Language Arts	2.19	1.95	1.66	5.80	32.24%
	B	Social Instructional Language	3.04	3.23	2.62	8.89	49.39%
		Language of Math / Science	3.01	3.15	2.70	8.86	49.21%
		Language of Language Arts / Social Studies	2.82	2.54	2.35	7.71	42.83%
	C	Social Instructional Language	3.30	3.55	2.93	9.78	54.34%
		Language of Math / Science	3.19	2.99	2.79	8.97	49.86%
		Language of Language Arts / Social Studies	3.15	2.80	2.68	8.63	47.92%

Grade	Tier	Standard	Mean Raw Score				Percent of Maximum
			Linguistic Complexity	Vocabulary Usage	Language Control	Total	
5	A	Social Instructional Language	2.25	2.03	1.71	5.99	33.29%
		Language of Math / Science	2.28	2.32	1.86	6.47	35.94%
		Language of Language Arts	2.31	2.04	1.77	6.11	33.96%
	B	Social Instructional Language	3.19	3.41	2.81	9.41	52.30%
		Language of Math / Science	3.11	3.32	2.85	9.29	51.60%
		Language of Language Arts / Social Studies	3.07	2.73	2.59	8.39	46.63%
	C	Social Instructional Language	3.45	3.68	3.10	10.24	56.87%
		Language of Math / Science	3.34	3.07	2.95	9.36	52.03%
		Language of Language Arts / Social Studies	3.39	2.96	2.91	9.26	51.47%
6	A	Social Instructional Language	2.26	1.96	1.87	6.09	33.82%
		Language of Math / Science	2.21	1.72	1.84	5.76	32.00%
		Language of Language Arts	2.40	2.14	1.84	6.39	35.49%
	B	Social Instructional Language	3.18	2.85	2.66	8.69	48.30%
		Language of Math / Science	3.05	3.27	2.71	9.02	50.14%
		Language of Language Arts / Social Studies	3.26	2.75	2.66	8.67	48.15%
	C	Social Instructional Language	3.50	3.06	3.04	9.59	53.30%
		Language of Math / Science	3.52	3.68	3.10	10.29	57.19%
		Language of Language Arts / Social Studies	3.61	3.05	3.05	9.71	53.95%
7	A	Social Instructional Language	2.31	2.01	1.92	6.24	34.68%
		Language of Math / Science	2.26	1.75	1.88	5.90	32.77%
		Language of Language Arts	2.43	2.15	1.87	6.44	35.80%
	B	Social Instructional Language	3.33	2.97	2.84	9.14	50.77%
		Language of Math / Science	3.20	3.42	2.86	9.48	52.68%
		Language of Language Arts / Social Studies	3.43	2.90	2.83	9.17	50.92%
	C	Social Instructional Language	3.68	3.18	3.24	10.09	56.06%
		Language of Math / Science	3.70	3.81	3.30	10.82	60.09%
		Language of Language Arts / Social Studies	3.84	3.27	3.28	10.39	57.70%
8	A	Social Instructional Language	2.38	2.06	1.99	6.43	35.73%
		Language of Math / Science	2.33	1.80	1.96	6.09	33.85%
		Language of Language Arts	2.49	2.20	1.92	6.62	36.75%
	B	Social Instructional Language	3.46	3.06	2.97	9.49	52.74%
		Language of Math / Science	3.32	3.53	2.97	9.82	54.55%
		Language of Language Arts / Social Studies	3.58	3.03	2.97	9.57	53.19%
	C	Social Instructional Language	3.83	3.31	3.41	10.55	58.59%
		Language of Math / Science	3.88	3.92	3.48	11.27	62.62%
		Language of Language Arts / Social Studies	4.02	3.48	3.48	10.98	60.98%

| Grade | Tier | Standard | Mean Raw Score | | | | |
			Linguistic Complexity	Vocabulary Usage	Language Control	Total	Percent of Maximum
9	A	Social Instructional Language	2.15	2.04	1.85	6.05	33.61%
		Language of Math / Science	2.26	2.16	1.72	6.13	34.07%
		Language of Language Arts	2.39	2.09	1.73	6.21	34.48%
	B	Social Instructional Language	3.53	2.96	2.99	9.47	52.63%
		Language of Math / Science	3.37	2.96	2.89	9.22	51.24%
		Language of Language Arts / Social Studies	3.34	3.14	2.79	9.27	51.52%
	C	Social Instructional Language	3.87	3.33	3.44	10.65	59.16%
		Language of Math / Science	3.43	3.65	3.09	10.17	56.48%
		Language of Language Arts / Social Studies	3.81	3.56	3.29	10.66	59.20%
10	A	Social Instructional Language	2.36	2.17	2.04	6.56	36.47%
		Language of Math / Science	2.52	2.39	1.93	6.83	37.97%
		Language of Language Arts	2.65	2.29	1.94	6.88	38.22%
	B	Social Instructional Language	3.49	2.95	2.92	9.37	52.07%
		Language of Math / Science	3.34	2.97	2.83	9.15	50.82%
		Language of Language Arts / Social Studies	3.28	3.13	2.71	9.12	50.67%
	C	Social Instructional Language	3.85	3.33	3.38	10.57	58.70%
		Language of Math / Science	3.37	3.60	3.01	9.98	55.45%
		Language of Language Arts / Social Studies	3.78	3.55	3.22	10.55	58.61%
11	A	Social Instructional Language	2.43	2.24	2.12	6.79	37.70%
		Language of Math / Science	2.66	2.55	2.06	7.27	40.38%
		Language of Language Arts	2.75	2.35	2.05	7.16	39.75%
	B	Social Instructional Language	3.56	3.04	2.96	9.56	53.09%
		Language of Math / Science	3.40	3.02	2.87	9.29	51.60%
		Language of Language Arts / Social Studies	3.36	3.22	2.77	9.35	51.92%
	C	Social Instructional Language	3.89	3.39	3.39	10.68	59.31%
		Language of Math / Science	3.42	3.64	3.03	10.09	56.07%
		Language of Language Arts / Social Studies	3.81	3.60	3.23	10.64	59.13%
12	A	Social Instructional Language	2.34	2.16	2.05	6.55	36.37%
		Language of Math / Science	2.52	2.42	1.96	6.90	38.32%
		Language of Language Arts	2.65	2.26	2.00	6.91	38.41%
	B	Social Instructional Language	3.52	3.01	2.93	9.45	52.52%
		Language of Math / Science	3.34	2.97	2.81	9.12	50.65%
		Language of Language Arts / Social Studies	3.29	3.14	2.71	9.14	50.80%
	C	Social Instructional Language	3.82	3.37	3.30	10.49	58.29%
		Language of Math / Science	3.33	3.54	2.94	9.82	54.53%
		Language of Language Arts / Social Studies	3.73	3.55	3.14	10.42	57.90%

4.4.3 Speaking

4.4.3.1 By Cluster

Table 4.4.3.1A

Mean Raw Score by Cluster by Tier by Standard: Speaking S301

Cluster	Tier	Standard	Maximum Score	Mean Raw Score	Percentage of Maximum
1-2	A	Social and Instructional Language	3	2.29	76.27%
	A	Language of Language Arts/Social Studies	5	2.79	55.72%
	A	Language of Mathematics/Science	5	2.63	52.65%
	B	Social and Instructional Language	3	2.83	94.27%
	B	Language of Language Arts/Social Studies	5	3.93	78.70%
	B	Language of Mathematics/Science	5	3.82	76.49%
	C	Social and Instructional Language	3	2.93	97.69%
	C	Language of Language Arts/Social Studies	5	4.51	90.20%
	C	Language of Mathematics/Science	5	4.45	89.05%
3-5	A	Social and Instructional Language	3	2.15	71.53%
	A	Language of Language Arts/Social Studies	5	2.15	42.90%
	A	Language of Mathematics/Science	5	2.14	42.82%
	B	Social and Instructional Language	3	2.90	96.72%
	B	Language of Language Arts/Social Studies	5	3.75	75.01%
	B	Language of Mathematics/Science	5	3.78	75.54%
	C	Social and Instructional Language	3	2.95	98.46%
	C	Language of Language Arts/Social Studies	5	4.27	85.46%
	C	Language of Mathematics/Science	5	4.30	86.03%
6-8	A	Social and Instructional Language	3	1.83	61.11%
	A	Language of Language Arts/Social Studies	5	1.88	37.64%
	A	Language of Mathematics/Science	5	1.57	31.32%
	B	Social and Instructional Language	3	2.85	94.93%
	B	Language of Language Arts/Social Studies	5	3.97	79.43%
	B	Language of Mathematics/Science	5	3.34	66.80%
	C	Social and Instructional Language	3	2.94	97.98%
	C	Language of Language Arts/Social Studies	5	4.48	89.52%
	C	Language of Mathematics/Science	5	4.04	80.73%
9-12	A	Social and Instructional Language	3	1.95	64.98%
	A	Language of Language Arts/Social Studies	5	1.86	37.21%
	A	Language of Mathematics/Science	5	1.50	29.95%
	B	Social and Instructional Language	3	2.83	94.25%
	B	Language of Language Arts/Social Studies	5	4.01	80.22%
	B	Language of Mathematics/Science	5	3.33	66.63%
	C	Social and Instructional Language	3	2.92	97.26%
	C	Language of Language Arts/Social Studies	5	4.63	92.68%
	C	Language of Mathematics/Science	5	4.15	82.96%

4.4.3.2 By Grade

Table 4.4.3.2A

Mean Raw Score by Grade by Tier by Standard: Speaking S301

Grade	Tier	Standard	Maximum Score	Mean Raw Score	Percent of Maximum
1	A	Social and Instructional Language	3	2.32	77.50%
		Language of Language Arts/Social Studies	5	2.84	56.74%
		Language of Mathematics/Science	5	2.68	53.68%
	B	Social and Instructional Language	3	2.79	93.11%
		Language of Language Arts/Social Studies	5	3.76	75.23%
		Language of Mathematics/Science	5	3.65	72.91%
	C	Social and Instructional Language	3	2.90	96.76%
		Language of Language Arts/Social Studies	5	4.35	86.97%
		Language of Mathematics/Science	5	4.28	85.66%
2	A	Social and Instructional Language	3	2.16	72.03%
		Language of Language Arts/Social Studies	5	2.61	52.20%
		Language of Mathematics/Science	5	2.45	49.09%
	B	Social and Instructional Language	3	2.86	95.40%
		Language of Language Arts/Social Studies	5	4.10	82.07%
		Language of Mathematics/Science	5	4.00	79.98%
	C	Social and Instructional Language	3	2.95	98.18%
		Language of Language Arts/Social Studies	5	4.59	91.88%
		Language of Mathematics/Science	5	4.54	90.81%
3	A	Social and Instructional Language	3	2.23	74.41%
		Language of Language Arts/Social Studies	5	2.24	44.82%
		Language of Mathematics/Science	5	2.23	44.62%
	B	Social and Instructional Language	3	2.90	96.80%
		Language of Language Arts/Social Studies	5	3.67	73.32%
		Language of Mathematics/Science	5	3.69	73.87%
	C	Social and Instructional Language	3	2.95	98.44%
		Language of Language Arts/Social Studies	5	4.20	84.05%
		Language of Mathematics/Science	5	4.24	84.72%
4	A	Social and Instructional Language	3	2.09	69.73%
		Language of Language Arts/Social Studies	5	2.09	41.74%
		Language of Mathematics/Science	5	2.08	41.62%
	B	Social and Instructional Language	3	2.89	96.44%
		Language of Language Arts/Social Studies	5	3.77	75.38%
		Language of Mathematics/Science	5	3.80	75.90%
	C	Social and Instructional Language	3	2.95	98.47%
		Language of Language Arts/Social Studies	5	4.26	85.28%
		Language of Mathematics/Science	5	4.30	85.94%

Grade	Tier	Standard	Maximum Score	Mean Raw Score	Percent of Maximum
5	A	Social and Instructional Language	3	2.06	68.76%
		Language of Language Arts/Social Studies	5	2.05	41.00%
		Language of Mathematics/Science	5	2.06	41.18%
	B	Social and Instructional Language	3	2.91	96.92%
		Language of Language Arts/Social Studies	5	3.91	78.11%
		Language of Mathematics/Science	5	3.93	78.60%
	C	Social and Instructional Language	3	2.95	98.48%
		Language of Language Arts/Social Studies	5	4.39	87.76%
		Language of Mathematics/Science	5	4.40	88.07%
6	A	Social and Instructional Language	3	1.87	62.20%
		Language of Language Arts/Social Studies	5	1.95	39.03%
		Language of Mathematics/Science	5	1.60	32.05%
	B	Social and Instructional Language	3	2.84	94.67%
		Language of Language Arts/Social Studies	5	3.93	78.64%
		Language of Mathematics/Science	5	3.25	64.96%
	C	Social and Instructional Language	3	2.93	97.78%
		Language of Language Arts/Social Studies	5	4.40	87.93%
		Language of Mathematics/Science	5	3.86	77.28%
7	A	Social and Instructional Language	3	1.80	60.14%
		Language of Language Arts/Social Studies	5	1.83	36.54%
		Language of Mathematics/Science	5	1.52	30.50%
	B	Social and Instructional Language	3	2.85	95.07%
		Language of Language Arts/Social Studies	5	3.97	79.48%
		Language of Mathematics/Science	5	3.37	67.31%
	C	Social and Instructional Language	3	2.94	98.03%
		Language of Language Arts/Social Studies	5	4.48	89.57%
		Language of Mathematics/Science	5	4.06	81.12%
8	A	Social and Instructional Language	3	1.83	60.93%
		Language of Language Arts/Social Studies	5	1.86	37.27%
		Language of Mathematics/Science	5	1.57	31.36%
	B	Social and Instructional Language	3	2.85	95.07%
		Language of Language Arts/Social Studies	5	4.01	80.28%
		Language of Mathematics/Science	5	3.42	68.34%
	C	Social and Instructional Language	3	2.94	98.15%
		Language of Language Arts/Social Studies	5	4.57	91.30%
		Language of Mathematics/Science	5	4.22	84.30%
9	A	Social and Instructional Language	3	1.80	59.90%
		Language of Language Arts/Social Studies	5	1.65	32.91%
		Language of Mathematics/Science	5	1.35	27.03%
	B	Social and Instructional Language	3	2.85	94.91%
		Language of Language Arts/Social Studies	5	4.12	82.46%
		Language of Mathematics/Science	5	3.48	69.56%
	C	Social and Instructional Language	3	2.93	97.68%
		Language of Language Arts/Social Studies	5	4.69	93.77%
		Language of Mathematics/Science	5	4.25	85.02%

Grade	Tier	Standard	Maximum Score	Mean Raw Score	Percent of Maximum
10	A	Social and Instructional Language	3	2.04	67.96%
		Language of Language Arts/Social Studies	5	1.94	38.74%
		Language of Mathematics/Science	5	1.55	31.07%
	B	Social and Instructional Language	3	2.81	93.78%
		Language of Language Arts/Social Studies	5	3.92	78.50%
		Language of Mathematics/Science	5	3.22	64.48%
	C	Social and Instructional Language	3	2.92	97.44%
		Language of Language Arts/Social Studies	5	4.63	92.59%
		Language of Mathematics/Science	5	4.12	82.38%
11	A	Social and Instructional Language	3	2.16	71.86%
		Language of Language Arts/Social Studies	5	2.14	42.76%
		Language of Mathematics/Science	5	1.70	33.96%
	B	Social and Instructional Language	3	2.83	94.30%
		Language of Language Arts/Social Studies	5	3.97	79.36%
		Language of Mathematics/Science	5	3.25	65.10%
	C	Social and Instructional Language	3	2.92	97.48%
		Language of Language Arts/Social Studies	5	4.62	92.49%
		Language of Mathematics/Science	5	4.11	82.18%
12	A	Social and Instructional Language	3	2.24	74.69%
		Language of Language Arts/Social Studies	5	2.44	48.89%
		Language of Mathematics/Science	5	1.86	37.17%
	B	Social and Instructional Language	3	2.80	93.42%
		Language of Language Arts/Social Studies	5	3.95	78.93%
		Language of Mathematics/Science	5	3.27	65.33%
	C	Social and Instructional Language	3	2.88	95.90%
		Language of Language Arts/Social Studies	5	4.54	90.71%
		Language of Mathematics/Science	5	4.02	80.36%

World-Class Instructional Design and Assessment

**Annual Technical Report for
ACCESS for ELLs English Language Proficiency Test,
Series 301, 2012-2013 Administration**

**Annual Technical Report No. 9
Volume 2 of 3: Analyses of Test Forms**

Prepared by:

CAL/WIDA Partnership Activities
Psychometrics/Research Team

Center for Applied Linguistics

May 30, 2014

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 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 ACCESS for ELLs is the Rasch measurement model (Wright & Stone, 1979). Additional information on its use in the development of the test is available in WIDA Technical Report 1, *Development and Field Test of ACCESS for ELLs*. 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. Thus, for example, data based on Rasch fit statistics guided the inclusion, revision, or deletion of items during the development and field testing of the test forms, and will continue to guide the refinement and further development of the test.

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

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

where

P_{ni1} = probability of a correct response "1" by person "n" on item "i"

P_{ni0} = probability of an incorrect response "0" by person "n" on item "i"

B_n = ability of person "n"

D_i = difficulty of item "i"

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

For the Writing tasks, a Rasch Rating Scale model was used. Mathematically, this can be represented as

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

where

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

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

B_n = ability of person "n"

D_i = difficulty of task "i"

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

All Rasch analyses were conducted using the Rasch measurement software program *Winsteps* (Linacre, 2006). Rasch statistics are presented in several of the tables that follow. When speaking of the measure of examinee ability, we use the term *ability measure* (rather than *theta* used commonly when discussing models based on Item Response Theory). When speaking of the measure of how hard an item was, we use the term *item difficulty measure* (rather than the *b parameter* 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 ACCESS for ELLs score scale for reporting purposes (see 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 where on the ability distribution they are, 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).

Also included in some of the tables are fit statistics for the Rasch model. 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 ACCESS for ELLs that needed to be deleted or revised and are presented in the appropriate tables. Outfit mean square statistics are influenced by outliers. For example, a difficult item that for some reason 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 1.00 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.

Linacre (2002, Autumn), the author of the Winsteps program, provides more guidance on how to interpret these statistics for test items. He writes:

- 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 Linacre calls "less productive for measurement, but not degrading"

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

Because conservative guidelines were followed in the development of 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 fit 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 301 of ACCESS for ELLs in the academic year 2012–2013. Exceptions are Tables E, G, H, and I. The equating summary tables (Table E) use data from a sample of about 1,000 students rather than the entire population of students, because the equating was done in the midst of the operational scoring. The item or task analysis summary tables (Table G), the complete item analysis tables (Table H), and the raw score to scale score conversion tables (Table I) use item and task difficulties from this equating.

5.1.3 Equating and Scaling

Complete information on the horizontal and vertical scaling of ACCESS for ELLs scores is provided in Technical Report 1, *Development and Field Test of ACCESS for ELLs*[®]. In brief, this scaling was accomplished during the field test based on an elaborate common item design, both across tiers and 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 ACCESS for ELLs scale scores used for reporting results on the test. Table D in Section 6 for each form provides the equation for converting Rasch ability measures in logits to ACCESS for ELLs scale scores.

The operational test forms in Series 301 represent a partial refreshment of Series 203. That is, while many items were common on both forms, certain folders on Series 203 were replaced with new items. Thus, to place results on Series 301 onto the ACCESS for ELLs scale score, items that were not revised or otherwise changed were anchored to the difficulty values from Series 203, which itself had been anchored to Series 202. Table E in Section 6 for each test form provides explicit information on the anchor items used for equating Series 301 results to those of Series 203.

5.1.4 DIF Analyses

Differential item analyses (DIF) attempt to investigate whether performances on items 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 that may be functioning differently for different groups based on criteria irrelevant to what is being tested. The performance of students on the ACCESS for ELLs items was compared by dividing students into two different groupings: first, males versus females; second, students of Hispanic ethnic background versus students of all other backgrounds. (For both analyses, students for whom gender or ethnicity was missing were excluded.) Two commonly used procedures for detecting

DIF were used: one for dichotomously scored items (Listening, Reading, and Speaking) and one for polytomously scored items (Writing).

5.1.4.1 Dichotomous Items

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

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

- A (no DIF), when the absolute value of delta was less than 1.0

- B (weak DIF), when the absolute value of delta was between 1.0 and 1.5

- C (strong DIF), when the absolute value of the delta was greater than 1.5

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

(For information on procedures for dealing with items with C-level DIF, see Section 1.4.5.)

5.1.4.2 Polytomous Items

For polytomous items (i.e., the Writing tasks), a similar approach is used. It is based on the Mantel Chi-square statistic and the standardized mean difference following procedures again developed by ETS. As with dichotomous items, the underlying assumption is that students who performed similarly overall on the test should perform similarly on the individual tasks. To test this assumption, students are placed into 6 groups based on their total raw score on the Writing test. We determined these categories by calculating what the total raw score of a student scoring WIDA Proficiency Levels 1, 2, 3, 4, 5, or 6 in each category would be. For example, a student consistently scoring 1 would have a total score of 18 on a Tier B or Tier C form. A student consistently scoring 2 would score a 36.

To divide the students into performance groups in this way, cut points were made halfway between the above totals, such that students in Group 1 would have a total score of 0 to 27; Group 2 totaled 28 to 45; Group 3 totaled 46 to 63; Group 4 totaled 64 to 81; and Group 5 totaled

82 to 108. (Note that Group 5 contained students scoring in the 6 range. These two groups were combined because there are so few students in that category.)

For each Writing task, performance was similarly categorized according to the scoring rubric. Thus, raw scores of 0 to 4 were category 1 (i.e., up to a score totaling 4, such as 2-1-1, which is a high 1 but not yet a 2); the raw scores of 5 to 7 were category 2; the raw scores of 8 to 10 were category 3; the raw scores of 11 to 13 were category 4; the raw scores of 14 to 16 were category 5; and the raw scores of 17 to 18 were category 6. (The only exception to this was Kindergarten Writing tasks, where there was much smaller spread of scores on the Writing tasks. In such cases, total raw scores were used to determine categories.)

Following formulae provided by Zwick, Donoghue, and Grima (1993), an Excel spreadsheet was programmed to take cross-tabulated data output by SPSS and 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 standardized mean difference (SMD) between the performances of the two groups being compared is calculated. The standardized mean difference compares the means of the two groups, adjusting for differences in the distribution of the two groups being compared across the values of the matching variable. To standardize the outcome, this difference is divided by the standard deviation (SD) of the item for the total group. This calculation is also programmed into the Excel spreadsheet.

Following guidance proposed by ETS, polytomously scaled items are classified into DIF levels as follows:

- AA (no DIF), when the Mantel Chi-square statistic is not significant; or, when it is significant and the absolute value of (SMD/SD) is less than or equal to .17

- BB (weak DIF), when the Mantel Chi-square statistic is significant and the absolute value of (SMD/SD) is greater than .17 but less than or equal to .25

- CC (strong DIF), when the Mantel Chi-square statistic is significant and the absolute value of (SMD/SD) is greater than .25

5.2 Descriptions

The following paragraphs describe the tables that follow and are repeated for each test form in each domain.

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. Listening, Reading, and Speaking were scored dichotomously (i.e., right or wrong). Thus, the highest possible score was the number of items on the test form. Each Writing task, however, could be awarded up to 18 points. Additionally, certain Writing tasks are weighted because of their potential to elicit higher levels of writing ability. For cluster 1–2, Tier A has a weight of 3 for the fourth task. For clusters 1–2, 3–5, 6–8, and 9–12, Tiers B and C have a weight of 2 for the second task and a weight of 3 for the third task. Thus, the maximum number of points on each Writing test form varies from 54 for the Tier A forms for clusters 3–5, 6–8, and 9–12 to 108 for the Tier B and C forms and cluster 1–2 Tier A.

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, 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 ACCESS for ELLs *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 given as the last table—Table I—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. To provide full perspective, it extends somewhat below and above the range of possible or observed scale scores. The vertical axis shows the number of students (count). Each bar shows how many students were awarded each scale score.

Table B shows, 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

Note that scale scores for Tier A and Tier B in Listening and Reading are capped. Within each grade, the highest possible scale score for Tier A is the scale score corresponding to the cut score for Proficiency Level 4 (i.e., proficiency level score of 4.0). For Tier B, the highest possible scale score within each grade is the score corresponding to the cut score for Proficiency Level 5 (i.e., proficiency level score of 5.0). Because of these grade-level cut scores, the scale score associated with a given proficiency level score increases by grade level within a cluster, and so the cap also increases by grade level. For example, for 3–5A Listening, the scale score is capped at 325 for Grade 3, 338 for Grade 4, and 350 for Grade 5 (see Table 6.3.1.1B). Thus, a third grade student with a raw score of 20 (out of 20) on that test will have a scale score of 325, a fourth grader with the same raw score will have a scale score of 338, and a fifth grader with the same raw score will have a scale score of 350. However, all three students would have a proficiency level score of 4.0.

Also note that, because the scale is vertically equated, the range of scale scores moves up the scale from one cluster to the next. Thus, a second grade student with a raw score of 0 on the Tier A Listening test would have a scale score of 108, while a fifth grade student with a raw score of 0 on the Tier A Listening test would have a scale score of 120.

Similarly, scale scores at the lower end may be truncated so that the lowest achievable proficiency level score is 1.0. Again, this results in a lower minimum scale score for students in lower grade levels within a grade-level cluster.

The influence of these cuts will also be noticed in Figure B, as well as in many other tables throughout the report.

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 the six WIDA proficiency levels. 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.

Each row of Table C shows, by each grade in the cluster and by total for the cluster:

- The WIDA proficiency level designation (1 to 6)

- 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

(Note that for some domains for Kindergarten and Tier A tests, it was not possible to place into all proficiency levels. Figure C and Table C also clearly show the effect of the scoring cap on Tiers A and B.)

For Kindergarten this information is provided for scores based on both the Accountability cut scores and the Instructional cut scores.

5.2.4 Scaling Equation Table (Table D)

For each test form, Table D provides the scaling equation for that domain. This equation is used to convert an examinee's ability measure into the scale score. Because ACCESS for ELLs is vertically equated (see 5.1.3 above), though each domain has its own equation, the same equation is used across all tiers and grade-level clusters within each domain.

5.2.5 Equating Summary (Table E)

Each year a certain percentage of items on each ACCESS for ELLs test form are refreshed. A post-equating procedure known as common item equating is used to equate results on new forms to the older forms. This means that the difficulty measure of items appearing on the new form that are the same as those on the older form are kept constant across both forms. Thus, performances on the newer form may be interpreted in the same frame of reference.

Many items appearing on ACCESS for ELLs Series 301 also appeared on Series 203. All items common to both forms were anchored in the first equating run. After the first equating run, some items that were originally anchored proved to have changed in their difficulty measure. This change is measured by the "Displacement" statistic. This statistic shows the difference between the difficulty value of the anchored item and what its difficulty value would have been had it not been anchored. For Listening and Reading items, and for Writing and Speaking tasks, if this value was large (i.e., usually above .30 or below -.30), that item was unanchored in the final equating run (i.e., it was treated as if it were a new item).

Table E presents a summary of the common item equating procedures. The first section of the table compares the current test (i.e., the Series 301 version of that test form) to the previous year's test (i.e., the Series 203 version of that test form). The number of items, the average item difficulty, the standard deviation of the item difficulty values, as well as the difficulty value of the easiest and hardest item on each test form is presented. These values are in terms of logits used in the Rasch measurement model.

The second section of the table presents information on the anchoring items. The total number of possible anchors (i.e., all common items) is shown, as well as the standard deviation of those items. Next, the number of items that were actually anchored (i.e., in general, those items whose displacement values were below .30 or above -.30) in the final equating run is shown, again with the average item difficulty and standard deviation. Finally, the percentage of items that served as anchors and the average displacement value is given. Generally speaking, the greater the number of tasks anchored and the closer the average displacement is to 0.00, the more trustworthy the equating results will be.

The final section of Table E shows the location of the anchor items or tasks, both by order on the test form and by order of difficulty. It is desirable that the anchored items appear throughout the test form in order to ensure that no systematic bias affects performance on them (e.g., if they all appear at the end of a test form, there may be a fatigue effect). It is also desirable that the anchor items represent a wide range of difficulties across the entire spectrum of the item difficulty values on a test form. The greater the representation across the difficulty range, the more trustworthy the equating results will be. This section also provides information on displacement; that is, the difference between the difficulty value of the anchored item and what that difficulty value would have been had the item not been anchored. Smaller displacement statistics indicate more consistency between the item's difficulty value on the Series 301 test form and on the

Return to Summary of Assessment Records Claims, Actions, and Evidence

Series 203 test form. Typically, random displacements of less than 0.5 logits are unlikely to have much impact on measurement in a test instrument (Displacement measures, 2006, January 29).

Note that for the Writing tasks, this table also provides the anchored step measures for the total score on each task. For the ACCESS Writing tasks, a rating scale model is used (see 5.1.1 above). Because a single generic rubric based on the generic WIDA performance level definitions is used to score all of the Writing tasks across all of the grade-level clusters, we modeled a rating scale that has the same step difficulty values across all Writing tasks across all grade-level clusters. Thus, these values are the same for every Writing task on ACCESS. These constant step difficulty values help to provide anchors in the calibration of new Writing tasks onto the common WIDA score scale each year.

Note that because the Kindergarten test form was newly created for Series 200, it was not equated to the Series 103 test. Therefore, Table E is not included for Kindergarten. For technical details on the Kindergarten test, see MacGregor, Kenyon, Gibson, and Evans, (2009). In addition, in the other grade-level clusters, scores for the Speaking test are based on a content analysis rather than on equating to previous forms; therefore, Table E is included only to verify that the raw score to scale score conversion remains within reasonable parameters.

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. Five vertical lines indicate the five cut scores for the highest grade in the cluster for the test form, dividing the figure into six sections for each of the WIDA proficiency levels (Levels 1–6) for the domain being tested. (Note that for some domains for Kindergarten and Tier A tests, it was not possible to place into all six language proficiency levels. 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 items on that form must be answered correctly (or points on the Writing section 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 Figures B and D, five vertical lines in Figure E indicate the five cut scores for the highest grade in the cluster for the test form, dividing the figure into six sections for each of the WIDA language proficiency levels (1–6) for the domain being tested. (Note that for some

domains for Kindergarten and Tier A tests, it was not possible to place into all six language proficiency levels. Note also that, although Listening and Reading scores on Tiers A and B were capped, all 5 horizontal lines indicating the cut points remain in this figure.) 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 1/2 and 2/3 for Tier A test forms; between 2/3, 3/4, and 4/5 for Tier B test forms; and between 3/4, 4/5, and 5/6 for Tier C test forms.

To help the reader interpret the figure, a horizontal line has been included at the level of information that corresponds to a conditional reliability of 0.8.

5.2.8 Reliability (Table F)

In contrast to Figure E, which is based on the Rasch measurement model, Table F presents reliability and accuracy information based on Classical Test Theory. It shows:

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

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

Cronbach's alpha is also affected by the distribution of ability within the group of students tested. All things being equal, the greater the heterogeneity of abilities within the group of students tested (i.e., the more widely the scores are distributed), the higher the reliability. In this sense, Cronbach's alpha is *sample dependent.* It is widely recognized that reliability can be as much a function of the test as of the sample of students tested. That is, the exact same test can produce widely disparate reliability indices based on ability distribution of the group of students tested. Because ACCESS for ELLs is a tiered test (that is, because each form in Tier A, B, or C targets only a certain range of the entire ability distribution), results for reliability on any one form, particularly for the shorter Listening test, may at times be lower than typically expected.

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 F 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

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

Traditionally, SEM has been used to create a band around an examinee's observed score, with the assertion in the view of classical test theory, 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. Statistically speaking, then, there is an expectation 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.

For the Writing tests (except Kindergarten, which is scored by the test administrator), information on inter-rater reliability is also provided in Table F. This portion of the table shows, for each of the three or four Writing tasks, the percent of agreement between two raters in terms of the three features being rated: Linguistic Complexity (LX), Vocabulary Usage (VU), and Language Control (LC). In this part of the table, the first column shows the Writing task (i.e., the first, second, third, or fourth, if applicable). The second column shows the number of Writing papers that were double scored. This number is generally 25% of all papers scored, chosen at random during the operational scoring process. The next column shows the feature, while the following columns show the rates of agreement: exact, adj (adjacent), and total sum of exact and adjacent. When the two raters agreed on the score, an exact agreement was counted. If the two raters were different in that feature by one point, an adjacent agreement was counted.

All operational Speaking tests are scored by the test administrator. In this report, information on inter-rater reliability for Speaking provided in Table F (except for Kindergarten) is based on data from the field test of the Speaking test, reported on fully in ACCESS for ELLs Technical Report 1, Development and Field Test of ACCESS for ELLs. This portion of the table shows, for each of the 13 Speaking tasks, the number of individuals in the sample responding to the task, the number of agreements between two raters as to the rating of the task, and the percent agreement of the rating.

5.2.9 Item/Task Analysis Summary (Table G)

Table G provides a summary of the analyses of the items (for Listening and Reading) or the tasks (for Writing and Speaking). The top part of the table gives an item or task summary. The first column in this part states the type of item (MC for multiple choice or ECR for extended constructed response). The next column shows the number of items or tasks on the test form. The next column gives the average item or task difficulty value in logits. For the multiple-choice items, the next column shows the average p-value. This is the average percent of correct 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 next section of Table G provides a summary of the findings of the DIF analyses (see 5.1.4). The first column gives the DIF level: A, B, or C for dichotomous items or AA, BB, or CC for polytomous tasks (i.e., Writing tasks). 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 A or 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 A or 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.

For the Writing tasks, the last part of this table shows the distribution of the raw scores on each task by total score category. (Recall that the total score for a task equals the sum of three feature scores, which are scored from 1 to 6, for a maximum total of 18; however, papers that are written in languages other than English or are totally incomprehensible may receive a score of 0, while papers that demonstrate the ability to copy or write a few words in English may be awarded a score of 1. The total score of 2 is impossible to achieve.)

5.2.10 Complete Item Analysis Table (Table H)

Table H 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 or task. The item or task names vary slightly across domains and grade-level clusters, but they usually consist of characters that represent the domain (e.g., "R" for Reading), the grade-level cluster (e.g., "g91" for Grades 9–12), the tier (e.g., C, if applicable), the unique number in the item database (e.g., 3820), the WIDA Standard (e.g., "MA" for the Language of Mathematics), the language proficiency level targeted (e.g., "p3"), the thematic folder name (e.g., "Cafeteria"), and the test series (e.g., 203). Note that for Writing, "IT" stands for the "integrated" task, which requires more extensive writing and that integrates model performance indicators for SI, LA, and SS. Also, note that for some Speaking and Kindergarten tasks, the naming system is a bit simpler, e.g., "1.S_A1_K_203", which contains the item order, domain, the folder, the proficiency level, the grade-level cluster, and the test series.

The second column in Table H 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. For dichotomously scored items (Listening, Reading, and Speaking), the fourth column shows the p-value (percent of correct answers on that item or, in the case of Speaking, percent of students meeting the expectations of that task). The next two columns show the Rasch fit statistics for the item or task, while the following columns show the results of the two DIF analyses for that item or task. These last columns are interpreted just as in Table G.

Note that in previous years, many of the Speaking tasks had high outfit values. This was especially true for the easier tasks that appeared early in a folder. An investigation into the response patterns to the Speaking test revealed a number of cases where either the test was administered incorrectly or one or more ratings were wrongly recorded. As explained in 1.2.5, if a student cannot meet the expectations of a task in a folder, the remaining tasks in that folder are not administered, and are assigned a score of 0. However, we found many cases in which students received a score of 0 for one task in a folder, and a score of 1 for a later task in that same folder. As a result, it appears that some students who would be expected to meet the

Return to Summary of Assessment Records Claims, Actions, and Evidence

expectations of certain low-level tasks did not meet those expectations; the existence of these outliers would increase the outfit value. Because these patterns indicate that either the test administrator did not follow the administration procedures, or that one or more responses were incorrectly recorded, we removed these responses from the data set when analyzing fit for the Speaking test. Table 5.2.10 shows how many such cases were removed from the analysis for each cluster.

Table 5.2.10

Rate of Speaking responses removed from fit analysis S301

Cluster	No. of responses	No. of responses removed	Percent of responses removed
1-2	345,409	13,860	4.0%
3-5	323,993	12,117	3.7%
6-8	197,463	7,481	3.8%
9-12	173,605	5,569	3.2%

Removing these items from the analysis helped to lower the outfit value for many of the Speaking items. However, there are still some items with high outfit values. We continue to investigate potential sources for these high outfit values.

Note also that the Kindergarten test used a new format starting with Series 200 (2008-2009). It was equated to Series 103 through a separate study, reported on in MacGregor, Kenyon, Gibson, and Evans (2009). Thus, the column labeled "Anchored?" is not included in Table H for the Kindergarten test.

5.2.11 Complete Raw Score to Scale Score Conversion Chart (Table I)

The next table in this section, Table I, presents the raw score to scale score conversion table for the test form. The first column shows all possible raw scores. The next one to four columns show the corresponding scale score for each grade level in the cluster. Note that for Listening and Reading items on Tier A, these have been capped to the scale score that represents the proficiency level score of 4.0. On Tier B, these have been capped to the scale score representing the proficiency level score of 5.0.

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 is below 100, which has been set as the lowest score on the scale. In those cases, the lower bound has been set at 100.

As can be clearly seen from the table, on any dichotomously scored test form, standard errors are very large at the lowest and highest ends of the raw score scale. Because of this phenomenon and because the scale scores are combined to form composite scores, the top scale scores for the Listening and Reading forms were often adjusted for an end-of-scale effect on Tier C 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.

Because on Tiers A and B the highest possible scores have been capped before the escalation of scale scores due to large standard errors at the highest end of the raw score scale inflates them, there has been no need to make any other adjustment to the scale scores for these tiers at the extreme high end of the raw score range. Since the point at which scale scores are capped depends on the proficiency level associated with the score, the caps take effect at lower scores for lower grades within a cluster. In this case the scores have been marked in Table I as capped, and the standard error, and low and high bound for the capped scale score, has been repeated in the final rows of the table.

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 1.0. As with the adjusted scores, the standard error and the lower and upper bounds reported in Table I reflect the true scale score, not the truncated score.

5.2.12 Raw Score to Proficiency Level Score Conversion Table (Table J)

The final table, Table J, shows the interpretive proficiency level score associated with each raw score. (Note that in previous Annual Technical Reports some of this information was included in Table I; however, with the grade-level cut scores in effect, we have put this information in a separate table for ease of reading.) The first column in Table J 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, along with the percentage of students in that grade who scored at 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. In Series 100 and 101, cut scores between proficiency levels were determined at the cluster level; thus, for example, in the 3–5 grade-level cluster, a given scale score was associated with the same proficiency level score for students in Grades 3, 4, and 5. Such a system, however, fails to take into account that older children can be expected to perform better on the test due to general cognitive growth over and above growth in English language proficiency. This effect can clearly be seen in Tables A and B, where average scores on any test form tend to rise, albeit slightly, by grade level. In order words, we would expect a fifth grader to perform better on the 3–5 grade-level cluster test form than a third grader at the same underlying level of English proficiency. To account for this effect, the WIDA Consortium adopted grade-level cut scores beginning with Series 102 so that, for any given raw score/scale score, the proficiency level score now associated with it differs according to the grade level of the student. (For details on how grade-level cut scores were determined, see Kenyon et al., 2013.) The effect of this for Table J is to require a separate column for each grade.

Second, because scale scores are capped on Listening and Reading for Tiers A and B at the scale score corresponding to the proficiency level score of 4.0 (for Tier A) and 5.0 (for Tier B), beginning with Series 102, this capped score is now dependent on the grade level (rather than

dependent on the cluster level). These differences in the cap are also shown in Table J on Tiers A and B for Listening and Reading.

For Kindergarten the proficiency level scores are provided based on both the Accountability cut scores and the Instructional cut scores.

6.   Analyses of Test Forms: Results

Chapter 6 contains proprietary test information and is not publicly available. State educational agencies (SEAs) may request this information; please contact us at help@wida.us.

World-Class Instructional Design and Assessment

Annual Technical Report for ACCESS for ELLs English Language Proficiency Test, Series 301, 2012-2013 Administration

Annual Technical Report No. 9
Volume 3 of 3: Analyses Across Tiers

Prepared by:

CAL/WIDA Partnership Activities
Psychometrics/Research Team

Center for Applied Linguistics

May 30, 2014

7. Analysis Across Tiers: Overview

7.1 Background

7.1.1 Reliability of Composites

Four composite scores are reported for ACCESS for ELLs: 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

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

where

k = number of components j

w_j = weight of component j

σ_j^2 = variance of component j

σ_c^2 = variance of composite.

The data to compute the stratified Cronbach's alpha is provided in the appropriate tables in Chapter 8.

7.1.2 Accuracy and Consistency of Classification

For each domain across tiers, as well as for the four composite scores, we have produced tables that indicate estimates of the accuracy and consistency of classification of examinees into the WIDA 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 ACCESS for ELLs, it is important to know how well these classifications are made. The analyses that we used 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.

True scores are, of course, unknown. The approach taken by Livingston and Lewis (1995) and implemented here 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 WIDA language proficiency level 5 ("bridging") or not, one can look at the accuracy index provided in the table for the cut score 4/5.

7.2 Descriptions

7.2.1 Scale Score Information (Figure A and Table A)

Figure A and Table A relate to the ACCESS for ELLs *scale scores* that were achieved by students in the grade-level cluster. Figure A shows the distribution of the scale scores. The horizontal axis shows the full range of all scale scores observed for the grade-level cluster. To provide full perspective, it extends somewhat below and above the range of observed scale scores. The vertical axis shows the number of students (count). Each bar shows how many students were awarded each scale score. Note that for Listening and Reading, the effects of capping the scores for Tier A and Tier B can often be clearly detected in this figure.

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

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

- Minimum observed scale score

- Maximum observed scale score

- The mean (average) scale score

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

7.2.2 Proficiency Level Information (Figure B and Table B)

Figure B and Table B provide information on the proficiency level distribution of the students in the grade-level cluster. Figure B shows the distribution of the proficiency levels. The horizontal axis shows the six WIDA proficiency levels. The vertical axis shows the percent of students. Each bar shows the percent of students who were placed into each language proficiency level.

Each row of Table B shows, by each grade in the cluster and by total for the cluster:

- The WIDA proficiency level designation (1 to 6)

- The number of students (count) whose performance on the test form placed them into that proficiency level in the domain being tested (the number students who were not absent, invalid, refused, exempt, or in the wrong cluster)

- The percent of students, out of the total number of students taking the form within a grade or within the total of students in the grade-level cluster, who were placed into that proficiency level in the domain being tested

For Kindergarten this information is provided for scores based on both the Accountability cut scores and the Instructional cut scores.

7.2.3 Conditional Standard Error of Measurement (Table C and Figures C and D)

Table C and Figures C and D provide information across the three overlapping tier forms within a grade-level cluster and on the comparative conditional standard error of measurement. (Note that this information applies only to the domain scores; this information is not applicable to the composite scores.)

Table C presents information on the conditional standard error of measurement at the most important points at which decisions are made about students on the basis of performances on ACCESS for ELLs, the cut points between language proficiency levels. Because the cut points depend on the grade level, information is provided for each grade level within the cluster. The leftmost column shows the cut (e.g., 1/2, which is the cut score between level 1 and level 2). The next column shows the grade level. The next column shows the cut score in the scale score metric (e.g., 305). In the last column(s), the corresponding conditional standard error of measurement is given for each cut score in the scale score metric. For Kindergarten, the SEMs are provided in separate tables for the accountability and instructional cut scores. For each of the other grade-level clusters, the SEMs for the cut scores are provided in one table for the Tiers (A, B, and C).

From this table it is possible to examine how well the different tiers are targeted for making decisions about students at the various cut scores. For example, Tier A is intended for students at the lowest end of the language proficiency continuum. Optimally, Tier A forms should have the lowest conditional SEM of any Tier at the 1/2 cut point, and a relatively low one at the 2/3 cut point. At the other end, Tier C forms should optimally have the lowest conditional SEM at the 5/6 cut point, and also a relatively low one at the 4/5 cut point. Tier B should have low SEM in the mid range. Information from this table provides easily comparable information on how well the three Tier forms are targeted to provide the most accurate measure to place their intended examinees into the language proficiency levels that they target. (Note that because of the capping of scores on Tiers A and B, there is no information given for some of the cuts.)

Figure C shows the test characteristic curve across the entire test for Kindergarten and across the three tiers for the other grade-level clusters. This figure reflects the "raw score" column in Table C. It shows graphically how the tiers differ in difficulty. A is represented by a dotted curve, Tier B by a light solid curve, and Tier C by a dark solid curve. Note that not all tiers have the same number of items. Thus, some curves for Listening and Reading in this figure may not end at the top horizontal line. Five vertical lines in the graphic indicate the cut scores at the highest grade in each cluster only.

Figure D compares the test information function across the entire test for Kindergarten and across the three tiers for the other grade-level clusters. This figure reflects the "SEM" column in Table C. Again, Tier A is represented by a dotted curve, Tier B by a light solid curve, and Tier C by a dark solid curve. As in Figure C, the cut scores at the highest grade in each cluster are indicated by vertical lines. These lines make it easy to see which form measures most accurately at which cut score.

7.2.4 Reliability Information (Table D)

In order to produce accuracy and consistency of classification tables, it was necessary to produce a single reliability estimate across the three tiers. For the domains, this was a weighted reliability estimate (Cronbach's alpha). In other words, it is the average reliability weighted by the number of students who were administered that tier form. Thus, Table D, based on the information from Table F in Chapter 6, provides the number of students and the reliability estimate for each tier. The final column presents the weighted reliability, an estimate of the reliability of the scale scores across the tiers.

For the composite scores, Table D presents the data used to calculate an estimate of the reliability of the composite using stratified Cronbach's alpha (see above). The first column shows the components forming the composite, the second column the weight of the composite in the total score, the third the variance of the scale scores, and the fourth the reliability of the composite. (Note that these are the weighted reliabilities across the tiers.) Unlike the weighted composite, which is an average, the stratified alpha reflects the fact that there are two or four measures being combined into one single measure. Thus, the reliability of the composite score will be higher than the reliability of any single subscore within the composite.

7.2.5 Accuracy and Consistency of Classification Tables (Table E)

Table E presents three rows of information related to the accuracy and consistency of placement into proficiency categories based on WIDA ACCESS (see above). With the adoption of grade-level cut scores with Series 102, placement within a proficiency level now depends on the grade

level of the student. Therefore, we provide a separate table for each grade level in a cluster. 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 (i.e., asking the question which students have reached level six and which have not). 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 ACCESS for ELLs.

Note that because of the scoring caps now imposed on Tier A and Tier B in Listening and Reading, in several cases only a very small percentage of test takers get placed into level 6. This outcome, combined with the range of observed scale scores, (which may be very close to the 5/6 cut), and the reliability of the test, means that the accuracy conditional on level for level 6 cannot be estimated. In such cases a hyphen (-) has been placed in the table. For Writing, this result can also occur for both levels 5 and 6.

For Kindergarten these tables are provided for both the Accountability cut scores and the Instructional cut scores.

8. Analysis Across Tiers: Results

8.1 Grade: K

8.1.1 Listening K

Figure 8.1.1A
Scale Scores: List K S301
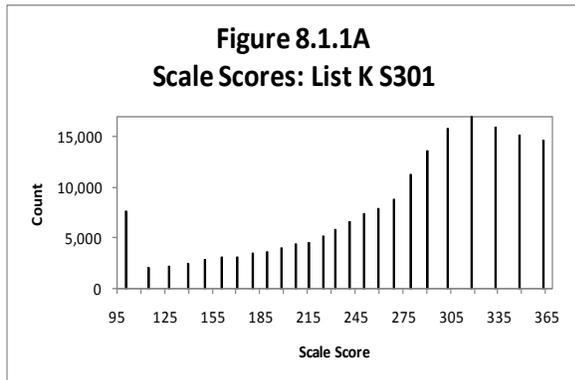
Table 8.1.1A

Scale Score Descriptive Statistics: List K S301

No. of Students	Min.	Max.	Mean	Std. Dev.
188,955	100	363	268.32	70.89

Figure 8.1.1Bi
Proficiency Level: List K S301
(Accountability)
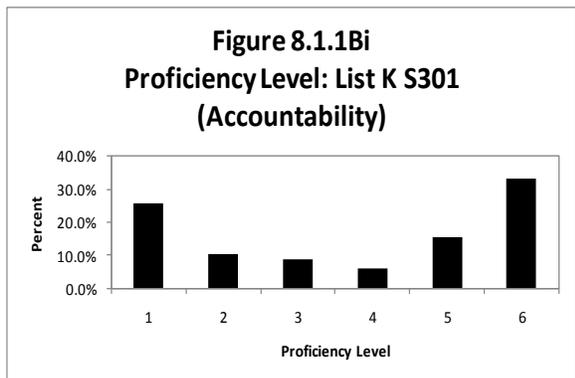
Table 8.1.1Bi

Proficiency Level Distribution: List K S301 (Accountability)

Level	Count	Percent
1	48,827	25.8%
2	19,817	10.5%
3	16,716	8.8%
4	11,228	5.9%
5	29,459	15.6%
6	62,908	33.3%
Total	188,955	100.0%

Figure 8.1.1Bii
Proficiency Level: List K S301
(Instructional)
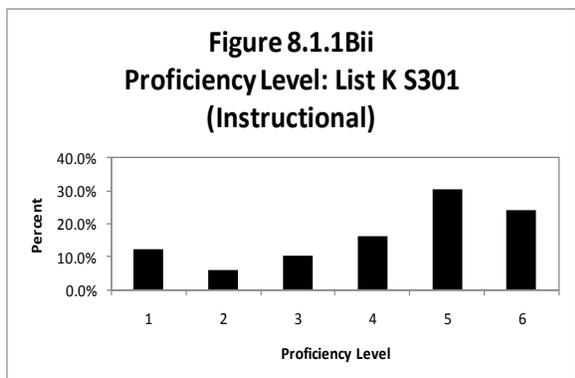
Table 8.1.1Bii

Proficiency Level Distribution: List K S301 (Instructional)

Level	Count	Percent
K1	23,601	12.5%
K2	11,113	5.9%
K3	19,918	10.5%
K4	30,728	16.3%
K5	57,659	30.5%
K6	45,936	24.3%
Total	188,955	100.0%

Table 8.1.1Ci
Conditional Standard Error of
Measurement at Cut Scores: List K
S301 (Accountability)

Proficiency Level	Cut Score	SEM
1/2	229	17.28
2/3	251	18.41
3/4	278	20.66
4/5	286	21.42
5/6	308	24.80

Table 8.1.1Cii
Conditional Standard Error of
Measurement at Cut Scores: List K
S301 (Instructional)

Proficiency Level	Cut Score	SEM
1/2	175	17.28
2/3	204	16.91
3/4	240	17.66
4/5	279	20.66
5/6	322	27.43

Figure 8.1.1D
Test Characteristic Curve: List K S301
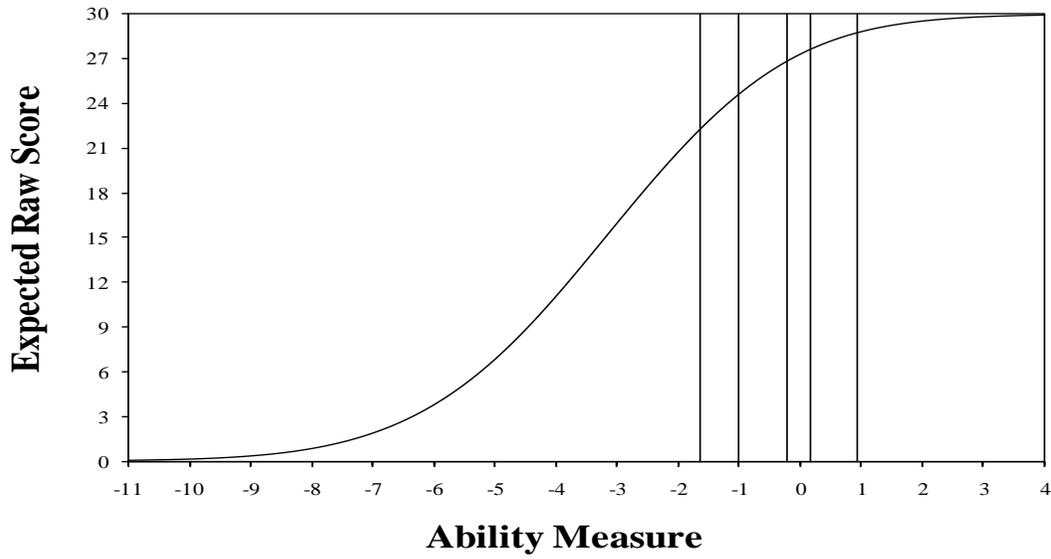
Figure 8.1.1D
Test Information Function: List K S301
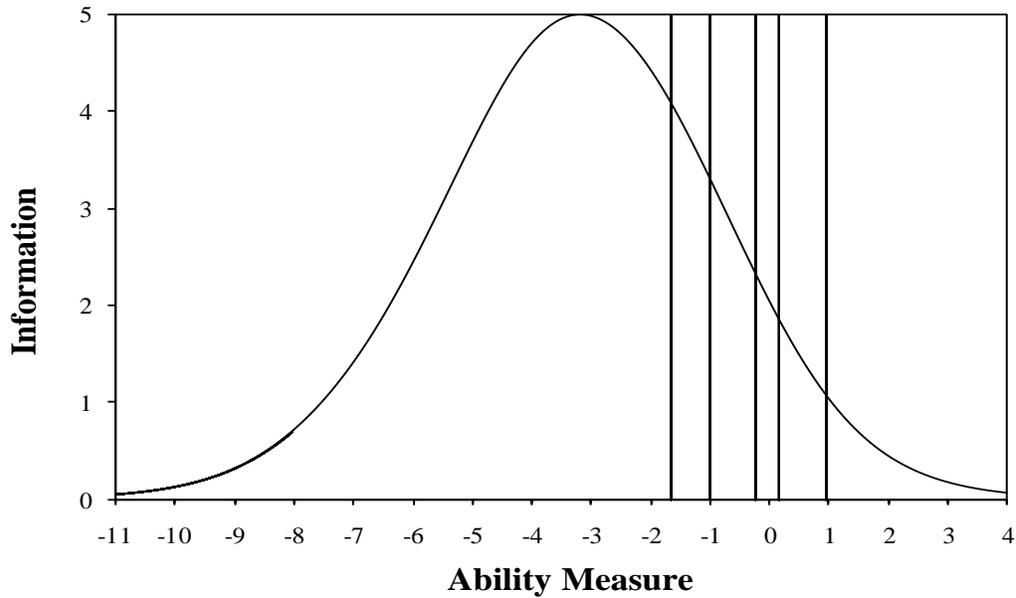
Table 8.1.1D

Reliability: List K S301

Tiers	No. of Students	Reliability
-	188,955	0.933

Table 8.1.1E

Accuracy and Consistency of Classification Indices: List (Grade K) S301
(Accountability)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.673	0.608		0.489	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.871		0.815	
	2	0.463		0.346	
	3	0.321		0.242	
	4	0.212		0.156	
	5	0.465		0.356	
	6	0.819		0.762	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.939	0.034	0.027	0.914
	2/3	0.926	0.028	0.045	0.898
	3/4	0.915	0.050	0.035	0.881
	4/5	0.906	0.045	0.049	0.873
	5/6	0.898	0.036	0.066	0.859

Table 8.1.1E

Accuracy and Consistency of Classification Indices: List (Grade K) S301
(Instructional)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.671	0.568		0.456	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.887		0.811	
	2	0.429		0.314	
	3	0.526		0.402	
	4	0.565		0.446	
	5	0.695		0.574	
	6	0.727		0.645	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.967	0.014	0.019	0.953
	2/3	0.956	0.022	0.022	0.937
	3/4	0.938	0.031	0.031	0.912
	4/5	0.915	0.042	0.043	0.882
	5/6	0.883	0.041	0.076	0.838

8.1.2 Reading K

Figure 8.1.2A
Scale Scores: Read K S301
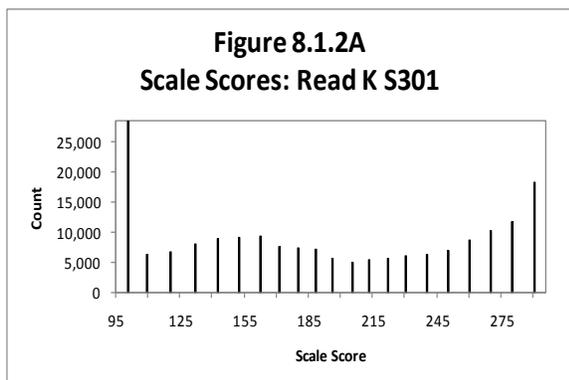
Table 8.1.2A
Scale Score Descriptive Statistics: Read K S301

No. of Students	Min.	Max.	Mean	Std. Dev.
188,946	100	290	192.12	66.04

Figure 8.1.2Bi
Proficiency Level: Read K S301 (Accountability)
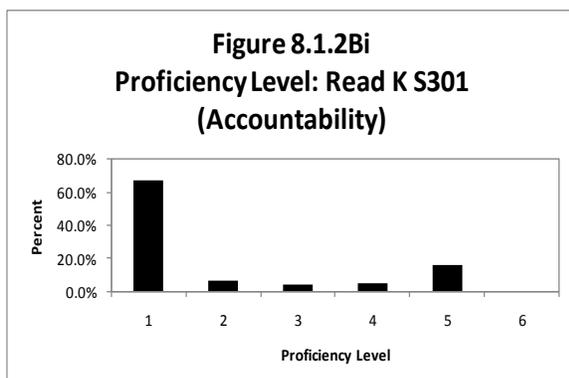
Table 8.1.2Bi
Proficiency Level Distribution: Read K S301 (Accountability)

Level	Count	Percent
1	126,738	67.1%
2	13,177	7.0%
3	8,726	4.6%
4	10,238	5.4%
5	30,067	15.9%
6	0	0.0%
Total	188,946	100.0%

Figure 8.1.2Bii
Proficiency Level: Read K S301 (Instructional)
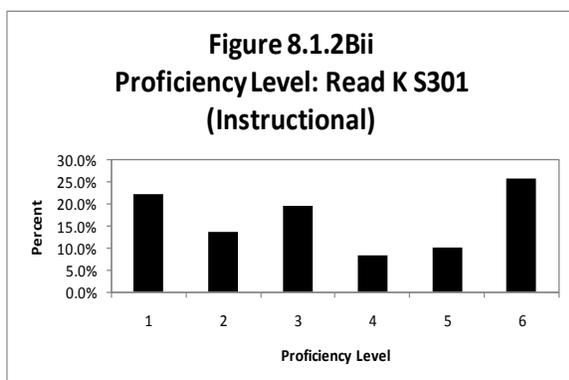
Table 8.1.2Bii
Proficiency Level Distribution: Read K S301 (Instructional)

Level	Count	Percent
K1	41,759	22.1%
K2	25,994	13.8%
K3	37,045	19.6%
K4	15,945	8.4%
K5	19,172	10.1%
K6	49,031	25.9%
Total	188,946	100.0%

Table 8.1.2Ci
Conditional Standard Error of Measurement
at Cut Scores: Read K S301 (Accountability)

Proficiency Level	Cut Score	SEM
1/2	238	15.08
2/3	251	16.90
3/4	261	18.98
4/5	274	22.10
5/6	295	30.68

Table 8.1.2Cii
Conditional Standard Error of Measurement
at Cut Scores: Read K S301 (Instructional)

Proficiency Level	Cut Score	SEM
1/2	121	14.04
2/3	159	13.52
3/4	204	13.00
4/5	228	14.04
5/6	255	17.68

Figure 8.1.2C
Test Characteristic Curve: Read K S301
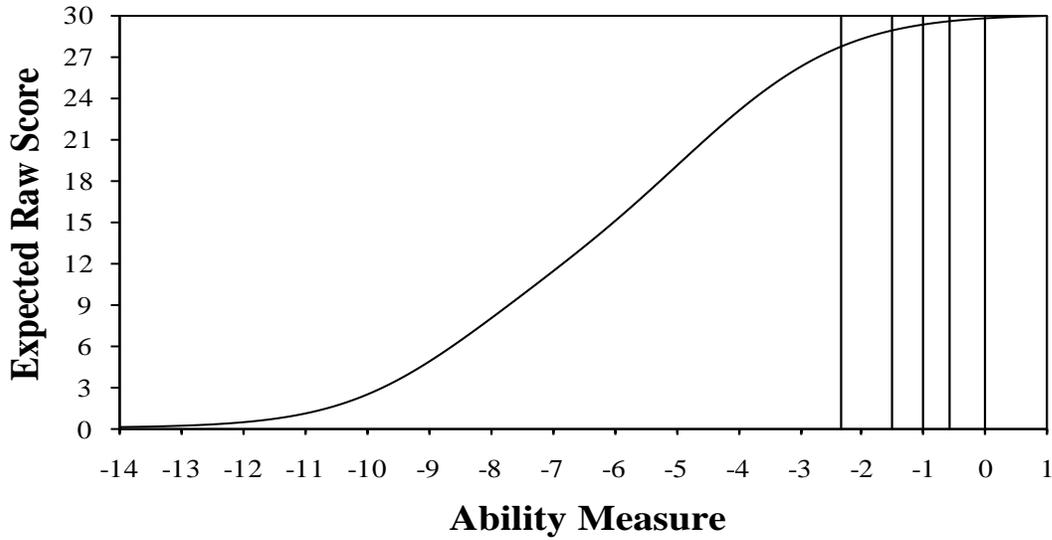
Figure 8.1.2D
Test Information Function: Read K S301
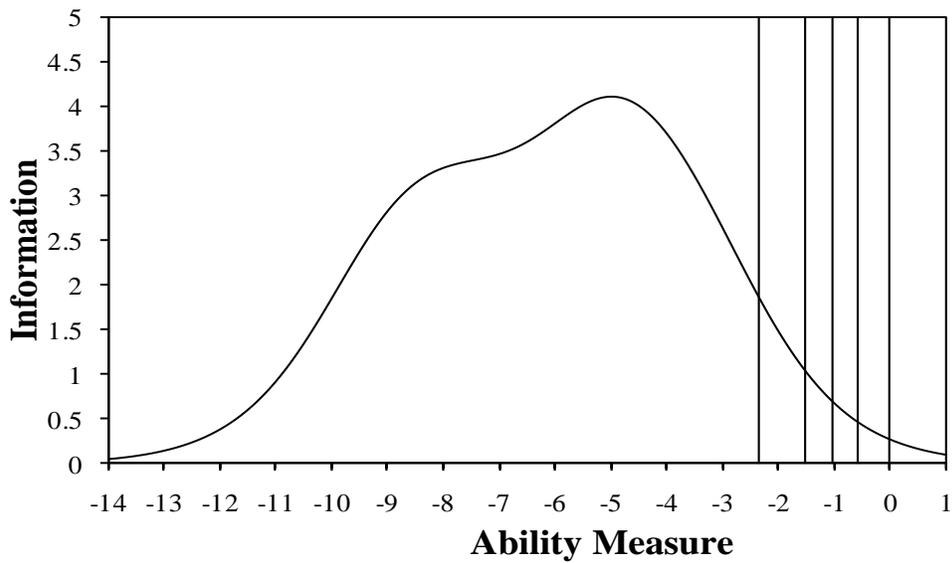
Table 8.1.2D

Reliability: Read K S301

Tiers	No. of Students	Reliability
-	188,946	0.947

Table 8.1.2E

Accuracy and Consistency of Classification Indices: Read (Grade K) S301 (Accountability)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.820	0.786		0.581	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.943		0.929	
	2	0.337		0.252	
	3	0.239		0.175	
	4	0.296		0.215	
	5	0.868		0.766	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.934	0.039	0.027	0.910
	2/3	0.937	0.032	0.031	0.913
	3/4	0.943	0.029	0.028	0.918
	4/5	0.948	0.033	0.019	0.925

Table 8.1.2E

Accuracy and Consistency of Classification Indices: Read (Grade K) S301 (Instructional)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.771	0.699		0.603	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.903		0.836	
	2	0.572		0.458	
	3	0.700		0.590	
	4	0.389		0.290	
	5	0.922		0.880	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.944	0.020	0.036	0.922
	2/3	0.940	0.033	0.026	0.915
	3/4	0.936	0.030	0.035	0.910
	4/5	0.940	0.032	0.028	0.914

8.1.3 Writing K

Figure 8.1.3A
Scale Scores: Writ K S301
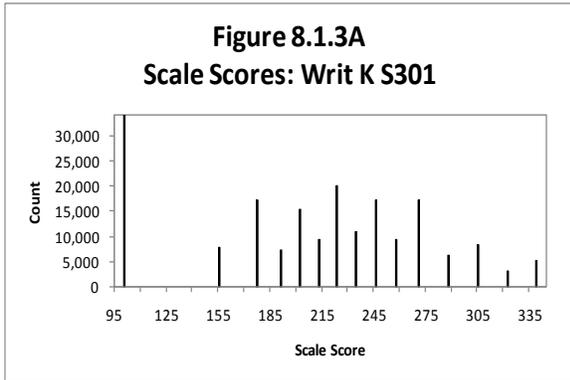
Table 8.1.3A
Scale Score Descriptive Statistics: Writ K S301

No. of Students	Min.	Max.	Mean	Std. Dev.
188,946	100	339	209.91	65.83

Figure 8.1.3Bi
Proficiency Level: Writ K S301
(Accountability)
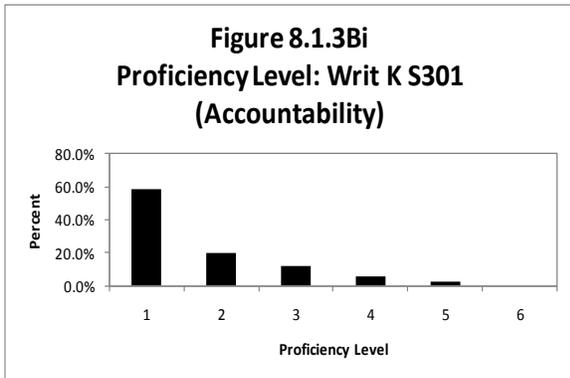
Table 8.1.3Bi
Proficiency Level Distribution: Writ K S301
(Accountability)

Level	Count	Percent
1	111,214	58.9%
2	37,323	19.8%
3	23,508	12.4%
4	11,589	6.1%
5	5,312	2.8%
6	0	0.0%
Total	188,946	100.0%

Figure 8.1.3Bii
Proficiency Level: Writ K S301
(Instructional)
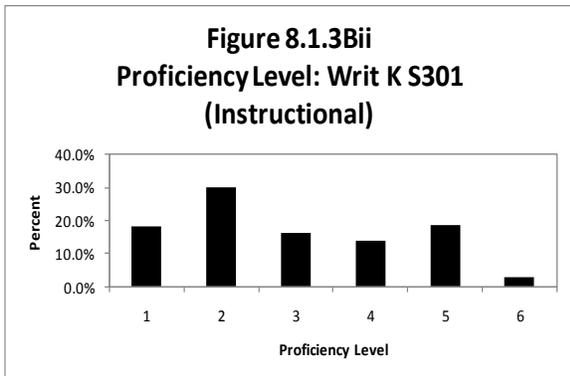
Table 8.1.3Bii
Proficiency Level Distribution: Writ K S301
(Instructional)

| Level	Count	Percent
K1	34,166	18.1%
K2	57,004	30.2%
K3	30,936	16.4%
K4	26,431	14.0%
K5	35,097	18.6%
K6	5,312	2.8%
Total	188,946	100.0%

Table 8.1.3Ci
Conditional Standard Error of
Measurement at Cut Scores: Writ K
S301 (Accountability)

| Proficiency Level	Cut Score	SEM
1/2	225	18.35
2/3	259	19.90
3/4	295	26.43
4/5	323	33.90
5/6	350	38.87

Table 8.1.3Cii
Conditional Standard Error of
Measurement at Cut Scores: Writ K
S301 (Instructional)

Proficiency Level	Cut Score	SEM
1/2	145	31.10
2/3	218	18.04
3/4	244	19.28
4/5	269	20.83
5/6	326	34.52

Figure 8.1.3C
Test Characteristic Curve: Writ K S301
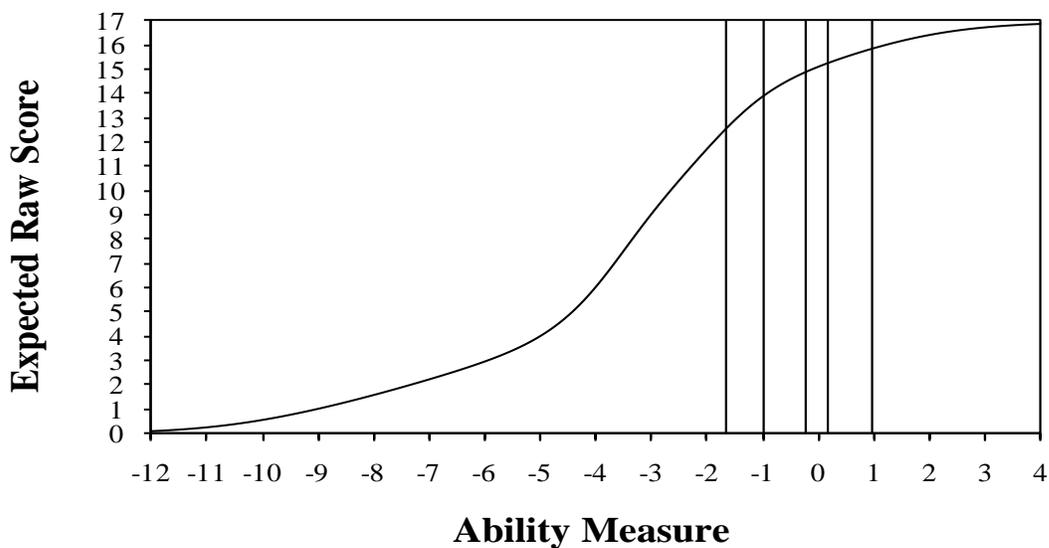
Figure 8.1.3D
Test Information Function: Writ K S301
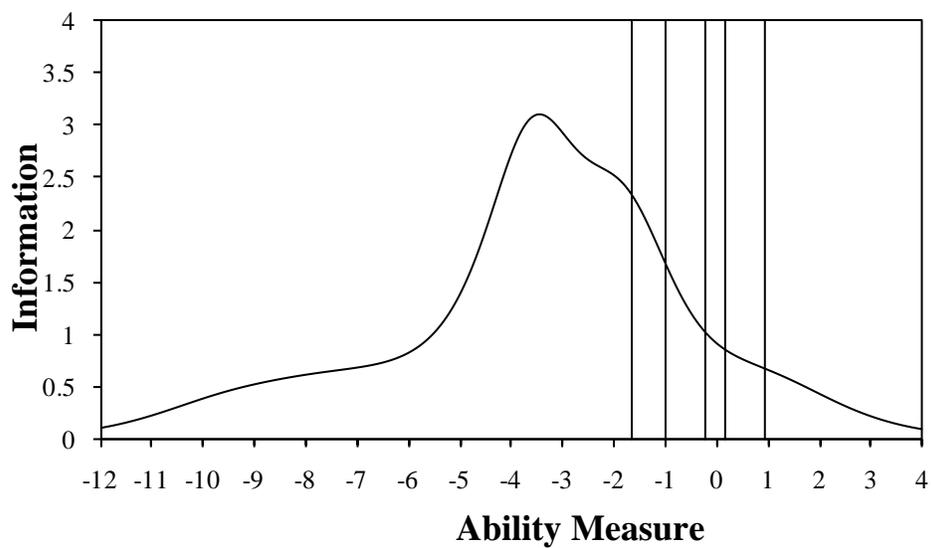
Table 8.1.3D

Reliability: Writ K S301

| Tiers	No. of Students	Reliability
-	188,946	0.923

Table 8.1.3E

Accuracy and Consistency of Classification Indices: Writ (Grade K) S301 (Accountability)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.743	0.692		0.484	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.942		0.915	
	2	0.608		0.463	
	3	0.391		0.349	
	4	-		0.255	
	5	-		0.145	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.923	0.034	0.044	0.894
	2/3	0.903	0.022	0.075	0.865
	3/4	0.911	0.089	0.000	0.899
	4/5	0.972	0.028	0.000	0.971

Table 8.1.3E

Accuracy and Consistency of Classification Indices: Writ (Grade K) S301 (Instructional)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.683	0.592		0.480	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.865		0.793	
	2	0.795		0.711	
	3	0.509		0.383	
	4	0.367		0.282	
	5	0.244		0.606	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.955	0.025	0.020	0.934
	2/3	0.917	0.040	0.043	0.885
	3/4	0.900	0.031	0.069	0.862
	4/5	0.881	0.044	0.075	0.834

8.1.4 Speaking K

Figure 8.1.4A
Scale Scores: Spek K S301
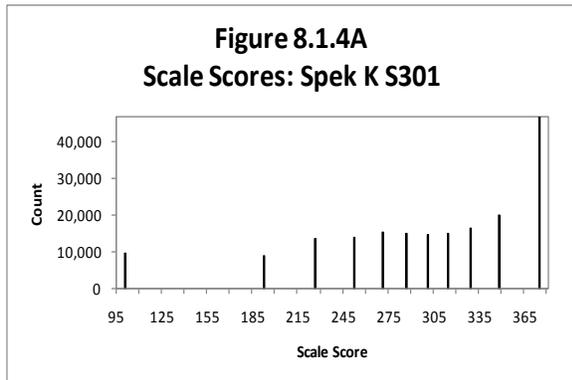
Table 8.1.4A
Scale Score Descriptive Statistics: Spek K S301

No. of Students	Min.	Max.	Mean	Std. Dev.
188,946	100	375	299.82	70.44

Figure 8.1.4Bi
Proficiency Level: Spek K S301
(Accountability)
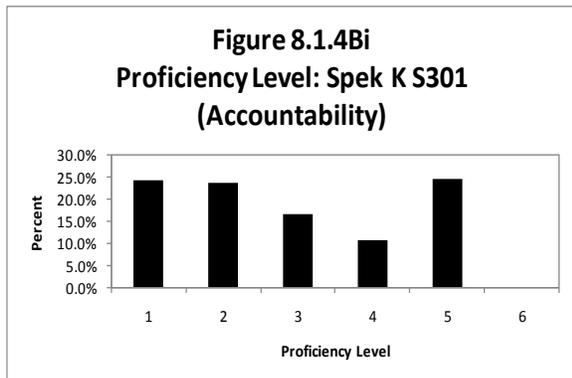
Table 8.1.4Bi
Proficiency Level Distribution: Spek K S301 (Accountability)

Level	Count	Percent
1	46,000	24.3%
2	44,880	23.8%
3	31,319	16.6%
4	20,085	10.6%
5	46,662	24.7%
6	0	0.0%
Total	188,946	100.0%

Figure 8.1.4Bii
Proficiency Level: Spek K S301
(Instructional)
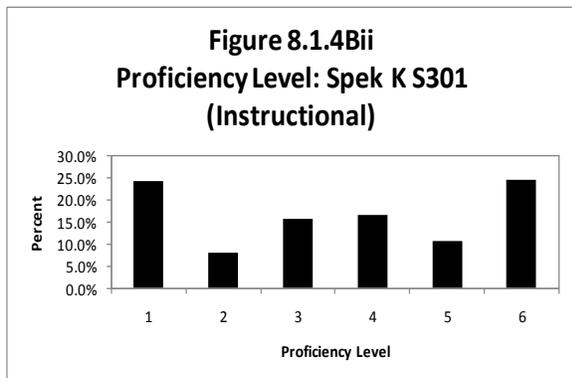
Table 8.1.4Bii
Proficiency Level Distribution: Spek K S301 (Instructional)

Level	Count	Percent
K1	46,000	24.3%
K2	15,241	8.1%
K3	29,639	15.7%
K4	31,319	16.6%
K5	20,085	10.6%
K6	46,662	24.7%
Total	188,946	100.0%

Table 8.1.4Ci
Conditional Standard Error of
Measurement at Cut Scores: Spek K
S301 (Accountability)

| Proficiency Level	Cut Score	SEM
1/2	269	18.68
2/3	314	16.27
3/4	343	20.89
4/5	366	31.33
5/6	383	44.99

Table 8.1.4Cii
Conditional Standard Error of
Measurement at Cut Scores: Spek K
S301 (Instructional)

Proficiency Level	Cut Score	SEM
1/2	256	20.89
2/3	285	17.07
3/4	308	16.27
4/5	342	20.49
5/6	365	30.53

Figure 8.1.4C
Test Characteristic Curve: Spek K S301
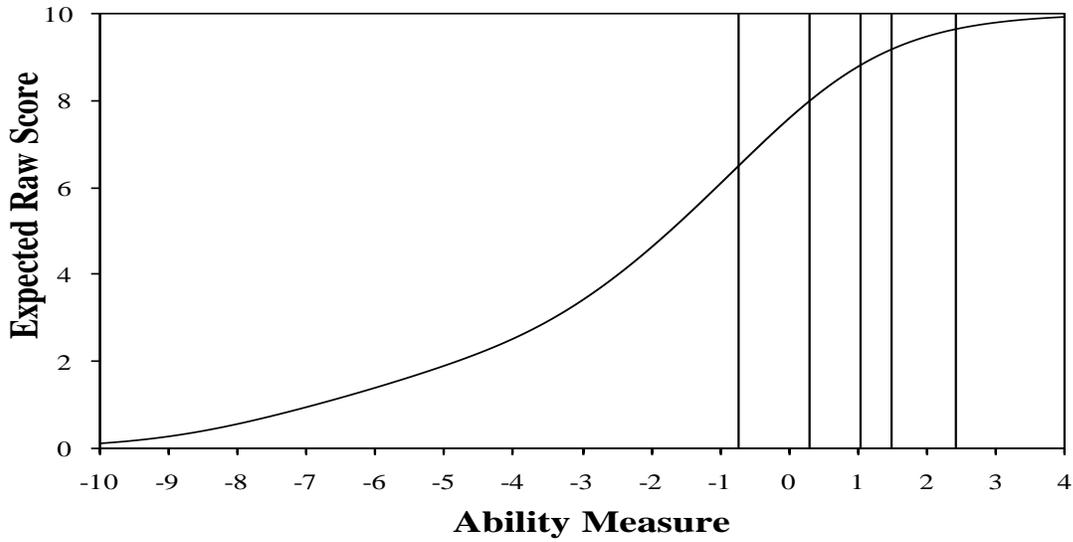
Figure 8.1.4D
Test Information Function: Spek K S301
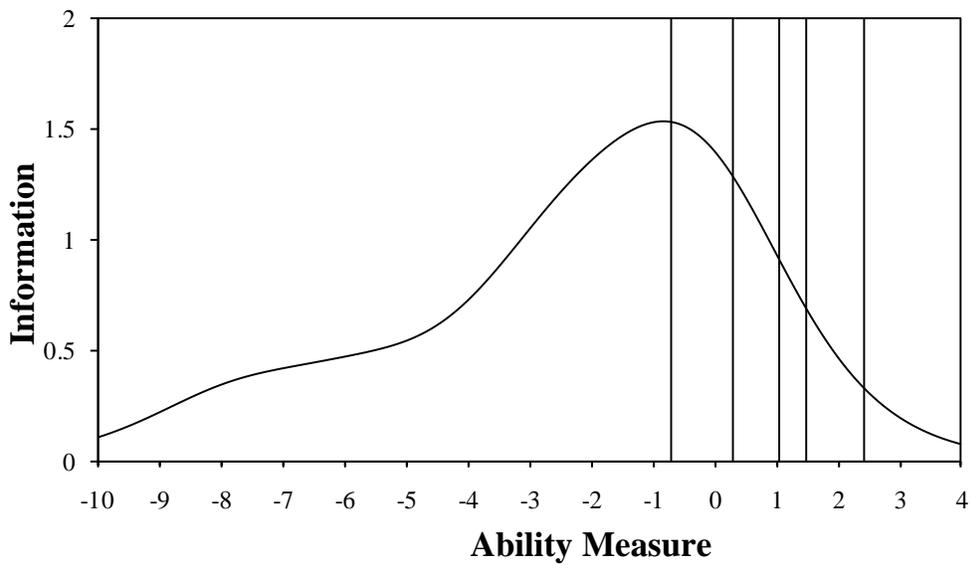
Table 8.1.4D

Reliability: Spek K S301

| Tiers	No. of Students	Reliability
-	188,946	0.894

Table 8.1.4E-1

Accuracy and Consistency of Classification Indices: Spek (Grade K) S301
(Accountability)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.478	0.455		0.324	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.834		0.765	
	2	0.663		0.531	
	3	0.371		0.259	
	4	0.209		0.191	
	5	-		0.552	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.926	0.042	0.032	0.896
	2/3	0.888	0.033	0.079	0.847
	3/4	0.864	0.049	0.087	0.786
	4/5	0.753	0.247	0.000	0.764

Table 8.1.4E-1

Accuracy and Consistency of Classification Indices: Spek (Grade K) S301
(Instructional)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.646	0.557		0.416	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.874		0.803	
	2	0.310		0.233	
	3	0.477		0.359	
	4	0.354		0.262	
	5	0.183		0.707	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.936	0.030	0.034	0.907
	2/3	0.912	0.048	0.040	0.881
	3/4	0.886	0.031	0.083	0.847
	4/5	0.858	0.063	0.079	0.790

8.1.5 Oral Language Composite K

Figure 8.1.5A
Scale Scores: Oral K S301
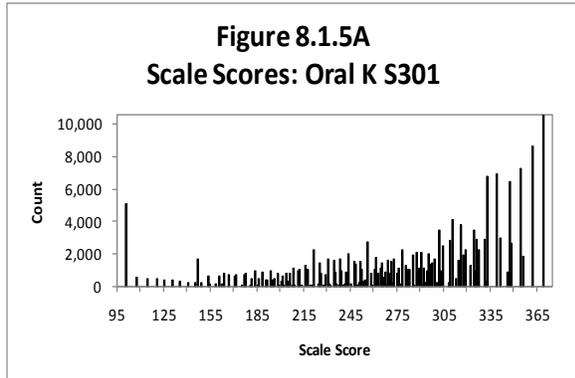
Table 8.1.5A

Scale Score Descriptive Statistics: Oral K S301

No. of Students	Min.	Max.	Mean	Std. Dev.
188,946	100	369	284.29	66.73

Figure 8.1.5Bi
Proficiency Level: Oral K S301
(Accountability)
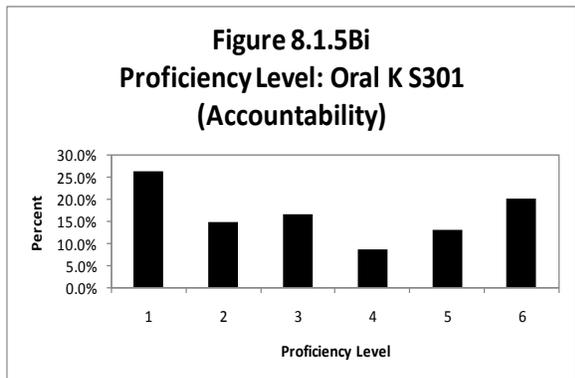
Table 8.1.5Bi

Proficiency Level Distribution: Oral K S301 (Accountability)

Level	Count	Percent
1	49,691	26.3%
2	28,259	15.0%
3	31,452	16.6%
4	16,239	8.6%
5	25,003	13.2%
6	38,302	20.3%
Total	188,946	100.0%

Figure 8.1.5Bii
Proficiency Level: Oral K S301
(Instructional)
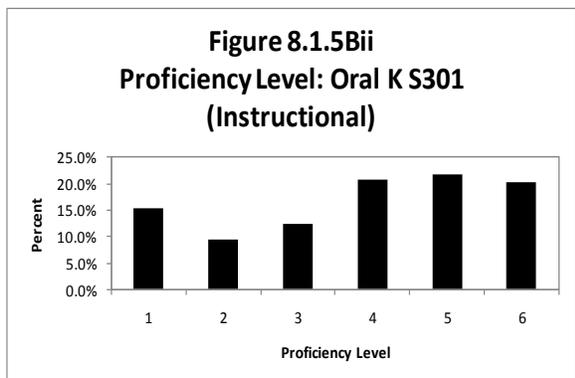
Table 8.1.5Bii

Proficiency Level Distribution: Oral K S301 (Instructional)

Level	Count	Percent
K1	28,902	15.3%
K2	17,766	9.4%
K3	23,533	12.5%
K4	39,201	20.7%
K5	41,242	21.8%
K6	38,302	20.3%
Total	188,946	100.0%

Table 8.1.5C

n/a

Figure 8.1.5C

n/a

Figure 8.1.5D

n/a

Table 8.1.5D

Oral Composite Reliability: Oral K S301

Component	Weight	Variance	Reliability
Listening	0.50	5023.787	0.933
Speaking	0.50	4959.953	0.894
Oral		4452.401	0.952

Table 8.1.5E

Accuracy and Consistency of Classification Indices: Oral (Grade K) S301 (Accountability)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.610	0.545		0.445	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.916		0.873	
	2	0.660		0.541	
	3	0.614		0.485	
	4	0.312		0.204	
	5	0.328		0.288	
	6	0.702		0.593	

| Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.957	0.022	0.020	0.939
	2/3	0.936	0.028	0.036	0.912
	3/4	0.926	0.023	0.051	0.895
	4/5	0.917	0.028	0.055	0.869
	5/6	0.840	0.128	0.032	0.824

Table 8.1.5E

Accuracy and Consistency of Classification Indices: Oral (Grade K) S301
(Instructional)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.652	0.558		0.460	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.895		0.839	
	2	0.608		0.480	
	3	0.589		0.470	
	4	0.699		0.574	
	5	0.511		0.432	
	6	0.685		0.586	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.970	0.016	0.014	0.957
	2/3	0.953	0.022	0.025	0.934
	3/4	0.940	0.026	0.034	0.917
	4/5	0.926	0.023	0.051	0.895
	5/6	0.858	0.090	0.052	0.825

8.1.6 Literacy Composite K

Figure 8.1.6A
Scale Scores: Litr K S301
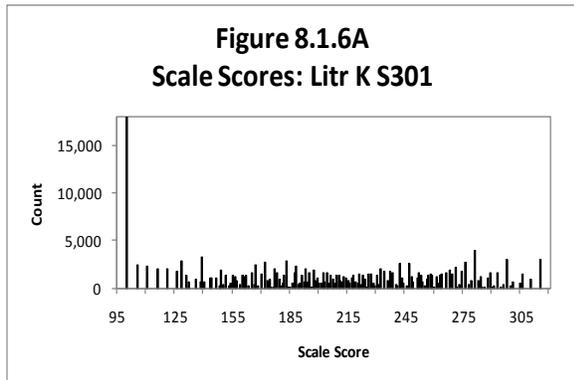
Table 8.1.6A
Scale Score Descriptive Statistics: Litr K S301

No. of Students	Min.	Max.	Mean	Std. Dev.
188,940	100	315	201.26	61.24

Figure 8.1.6Bi
Proficiency Level: Litr K S301 (Accountability)
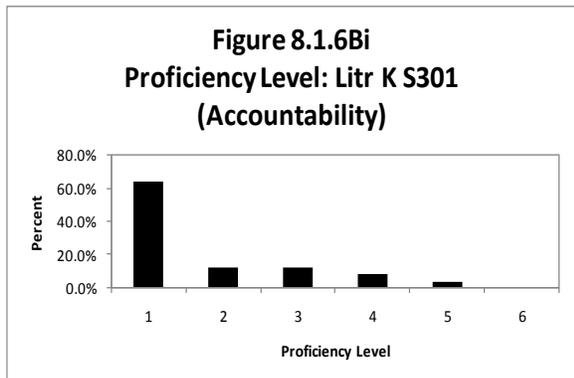
Table 8.1.6Bi
Proficiency Level Distribution: Litr K S301 (Accountability)

Level	Count	Percent
1	121,482	64.3%
2	22,515	11.9%
3	22,417	11.9%
4	15,657	8.3%
5	6,869	3.6%
6	0	0.0%
Total	188,940	100.0%

Figure 8.1.6Bii
Proficiency Level: Litr K S301 (Instructional)
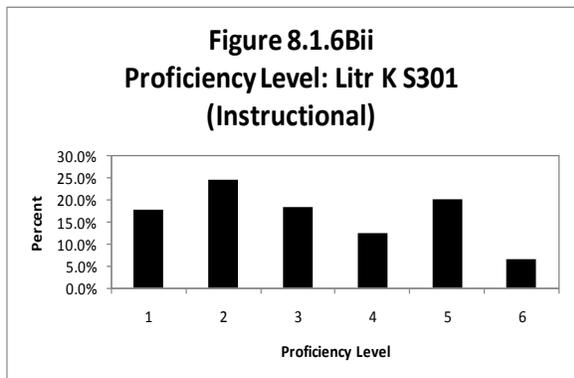
Table 8.1.6Bii
Proficiency Level Distribution: Litr K S301 (Instructional)

Level	Count	Percent
K1	33,587	17.8%
K2	46,700	24.7%
K3	34,542	18.3%
K4	23,717	12.6%
K5	38,105	20.2%
K6	12,289	6.5%
Total	188,940	100.0%

Table 8.1.6C
n/a

Figure 8.1.6C
n/a

Figure 8.1.6D
n/a

Table 8.1.6D
Literacy Composite Reliability: Litr K S301

Component	Weight	Variance	Reliability
Reading	0.50	4360.649	0.947
Writing	0.50	4333.709	0.923
Literacy		3750.255	0.962

Table 8.1.6E
Accuracy and Consistency of Classification Indices: Litr (Grade K) S301
(Instructional)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.748	0.672		0.596	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.919		0.873	
	2	0.839		0.774	
	3	0.729		0.623	
	4	0.576		0.450	
	5	0.647		0.615	
	6	-		0.403	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.970	0.014	0.016	0.957
	2/3	0.953	0.024	0.023	0.933
	3/4	0.944	0.024	0.032	0.922
	4/5	0.945	0.022	0.033	0.921
	5/6	0.935	0.065	0.000	0.926

Table 8.1.6E

Accuracy and Consistency of Classification Indices: Litr (Grade K) S301 (Accountability)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.795	0.746		0.539	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.961		0.943	
	2	0.560		0.430	
	3	0.499		0.389	
	4	0.451		0.389	
	5	-		0.259	
	6	-		0.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.947	0.025	0.029	0.925
	2/3	0.942	0.024	0.034	0.917
	3/4	0.931	0.035	0.033	0.904
	4/5	0.964	0.036	0.000	0.954
	5/6	1.000	0.000	0.000	1.000

8.1.7 Comprehension Composite K

Figure 8.1.7A
Scale Scores: Cphn K S301
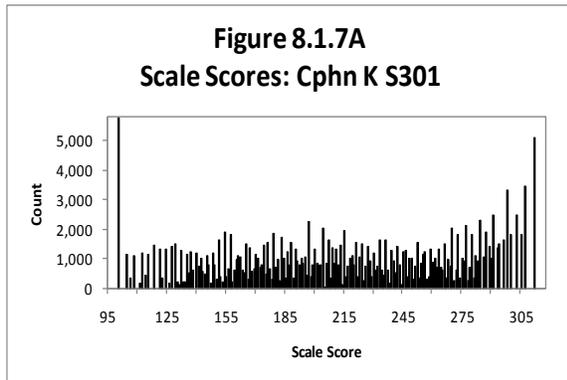
Table 8.1.7A

Scale Score Descriptive Statistics: Cphn K S301

No. of Students	Min.	Max.	Mean	Std. Dev.
188,938	100	312	214.98	60.43

Figure 8.1.7Bi
Proficiency Level: Cphn K S301 (Accountability)
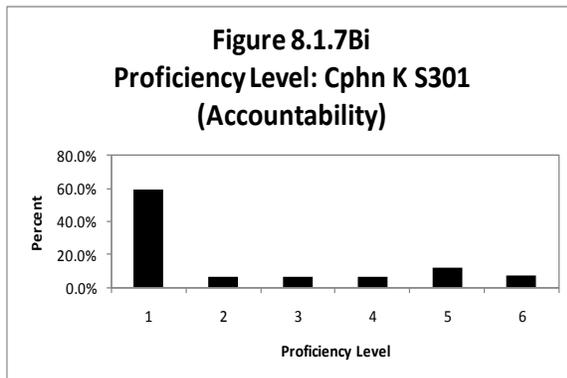
Table 8.1.7Bi

Proficiency Level Distribution: Cphn K S301 (Accountability)

Level	Count	Percent
1	111,964	59.3%
2	13,214	7.0%
3	13,142	7.0%
4	12,691	6.7%
5	23,255	12.3%
6	14,672	7.8%
Total	188,938	100.0%

Figure 8.1.7Bii
Proficiency Level: Cphn K S301 (Instructional)
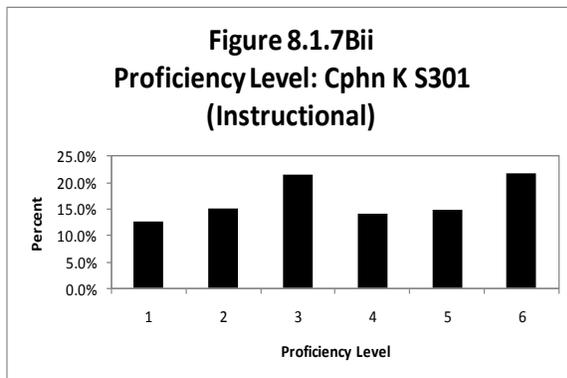
Table 8.1.7Bii

Proficiency Level Distribution: Cphn K S301 (Instructional)

Level	Count	Percent
K1	23,823	12.6%
K2	28,466	15.1%
K3	40,666	21.5%
K4	26,872	14.2%
K5	28,135	14.9%
K6	40,976	21.7%
Total	188,938	100.0%

Table 8.1.7C

n/a

Figure 8.1.7C

n/a

Figure 8.1.7D

n/a

Table 8.1.7D

Comprehension Composite Reliability: Cphn K S301

| Component	Weight	Variance	Reliability
Listening	0.30	5023.787	0.933
Reading	0.70	4360.649	0.947
Comprehension		3651.991	0.961

Table 8.1.7E-1

Accuracy and Consistency of Classification Indices: Cphn (Grade K) S301 (Accountability)

| Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.759	0.705		0.522	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.962		0.944	
	2	0.397		0.290	
	3	0.381		0.276	
	4	0.344		0.251	
	5	0.540		0.450	
	6	0.645		0.500	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.949	0.022	0.029	0.928
	2/3	0.950	0.025	0.025	0.928
	3/4	0.945	0.030	0.026	0.921
	4/5	0.938	0.032	0.030	0.914
	5/6	0.942	0.034	0.024	0.922

Table 8.1.7E-1

Accuracy and Consistency of Classification Indices: Cphn (Grade K) S301 (Instructional)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.768	0.679		0.612	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.899		0.840	
	2	0.765		0.670	
	3	0.782		0.694	
	4	0.638		0.517	
	5	0.625		0.503	
	6	0.875		0.822	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.973	0.012	0.015	0.961
	2/3	0.957	0.021	0.022	0.939
	3/4	0.947	0.024	0.029	0.926
	4/5	0.949	0.023	0.028	0.927
	5/6	0.941	0.032	0.026	0.917

8.1.8 Overall Composite K

Figure 8.1.8A
Scale Scores: Over K S301
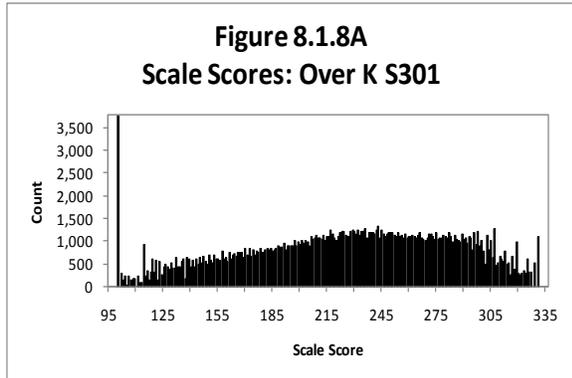
Table 8.1.8A

Scale Score Descriptive Statistics: Over K S301

No. of Students	Min.	Max.	Mean	Std. Dev.
188,926	100	331	225.97	57.38

Figure 8.1.8Bi
Proficiency Level: Over K S301
(Accountability)
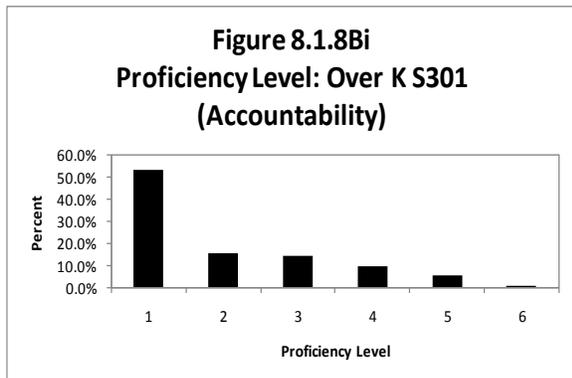
Table 8.1.8Bi

Proficiency Level Distribution: Over K S301 (Accountability)

Level	Count	Percent
1	101,314	53.6%
2	29,853	15.8%
3	27,183	14.4%
4	18,161	9.6%
5	10,824	5.7%
6	1,591	0.8%
Total	188,926	100.0%

Figure 8.1.8Bii
Proficiency Level: Over K S301
(Instructional)
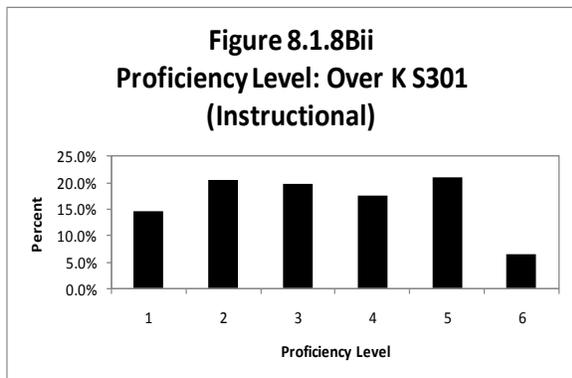
Table 8.1.8Bii

Proficiency Level Distribution: Over K S301 (Instructional)

Level	Count	Percent
K1	27,484	14.5%
K2	38,690	20.5%
K3	37,378	19.8%
K4	33,174	17.6%
K5	39,785	21.1%
K6	12,415	6.6%
Total	188,926	100.0%

Table 8.1.8C
n/a

Figure 8.1.8C
n/a

Figure 8.1.8D
n/a

Table 8.1.8D
Overall Composite Reliability: Over K S301

Component	Weight	Variance	Reliability
Listening	0.15	5023.787	0.933
Reading	0.35	4360.649	0.947
Speaking	0.15	4959.953	0.894
Writing	0.35	4333.709	0.923
Overall Composite		3293.012	0.973

Table 8.1.8E
Accuracy and Consistency of Classification Indices: Over (Grade K) S301 (Accountability)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.805	0.745		0.610	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.957		0.938	
	2	0.703		0.592	
	3	0.666		0.544	
	4	0.515		0.432	
	5	0.269		0.456	
	6	-		0.129	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.953	0.023	0.024	0.934
	2/3	0.951	0.022	0.027	0.931
	3/4	0.953	0.022	0.025	0.932
	4/5	0.951	0.037	0.011	0.939
	5/6	0.992	0.008	0.000	0.991

Table 8.1.8E

Accuracy and Consistency of Classification Indices: Over (Grade K) S301 (Instructional)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.788	0.711		0.646	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.926		0.885	
	2	0.841		0.776	
	3	0.779		0.688	
	4	0.723		0.617	
	5	0.728		0.673	
	6	0.692		0.529	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.977	0.011	0.012	0.967
	2/3	0.960	0.021	0.020	0.717
	3/4	0.950	0.023	0.027	0.524
	4/5	0.952	0.021	0.027	0.932
	5/6	0.949	0.039	0.012	0.937

8.2   Grades: 1–2

8.2.1 Listening 1-2

Figure 8.2.1A
Scale Scores: List 1-2 S301

Figure 8.2.1B
Proficiency Level: List 1-2 S301
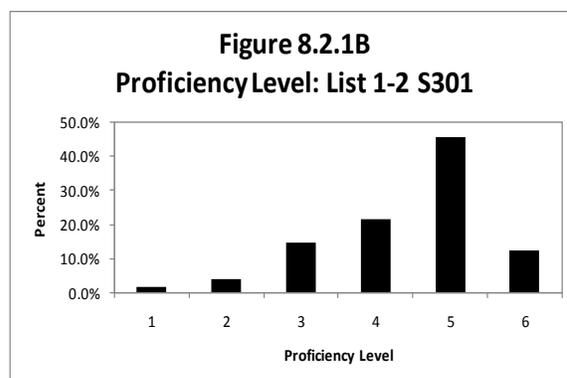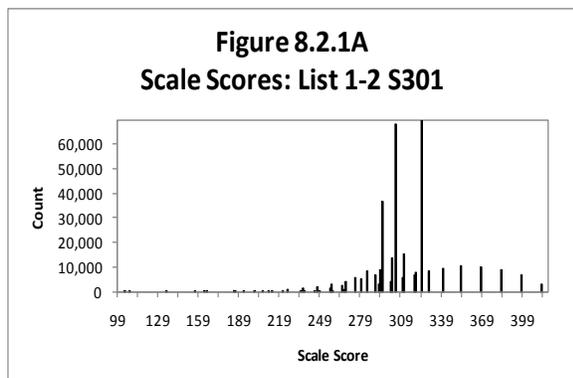
Table 8.2.1A

Scale Score Descriptive Statistics: List 1-2 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	178,422	104	413	299.11	27.06
2	167,148	108	413	326.70	32.18
Total	345,570	104	413	312.45	32.70

Table 8.2.1B

Proficiency Level Distribution: List 1-2 S301

Level	Grade 1 Count	Grade 1 Percent	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	4,269	2.4%	1,902	1.1%	6,171	1.8%
2	8,120	4.6%	6,192	3.7%	14,312	4.1%
3	31,116	17.4%	19,276	11.5%	50,392	14.6%
4	46,606	26.1%	27,198	16.3%	73,804	21.4%
5	76,423	42.8%	81,324	48.7%	157,747	45.6%
6	11,888	6.7%	31,256	18.7%	43,144	12.5%
Total	178,422	100.0%	167,148	100.0%	345,570	100.0%

Table 8.2.1C

Conditional Standard Error of Measurement at Cut Scores: List 1-2 S301

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	1	238	20.29	21.04	20.29
	2	247	19.91	19.91	19.54
2/3	1	267	19.91	18.79	18.41
	2	281	20.29	18.41	18.03
3/4	1	295	21.04	18.41	18.03
	2	311	22.54	18.79	18.41
4/5	1	305	n/a	18.41	18.03
	2	324	n/a	19.54	19.16
5/6	1	330	n/a	n/a	19.91
	2	350	n/a	n/a	22.17

Figure 8.2.1C
Test Characteristic Curve: List 1-2ABC S301
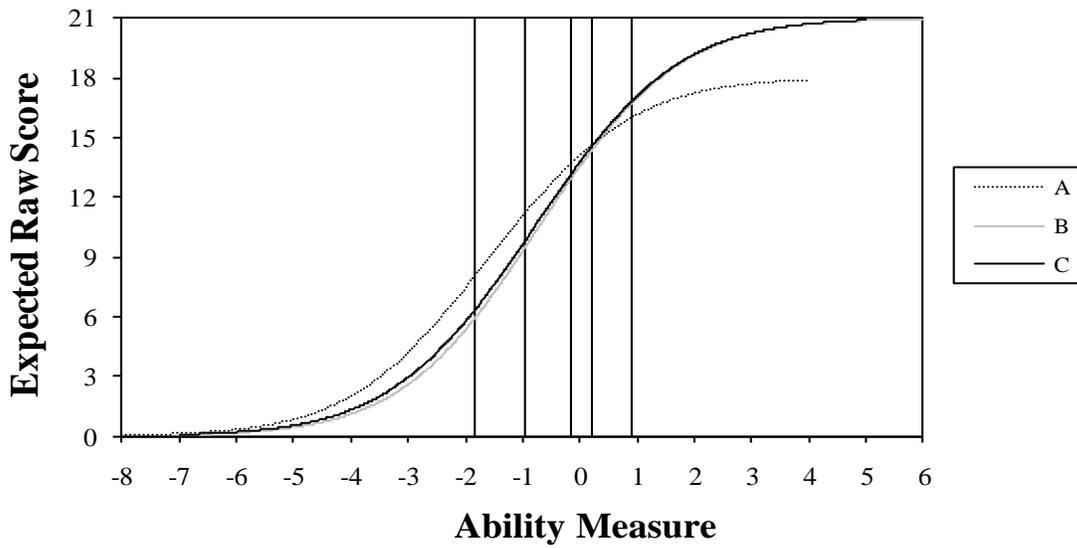
Figure 8.2.1D
Test Information Function: List 1-2ABC S301
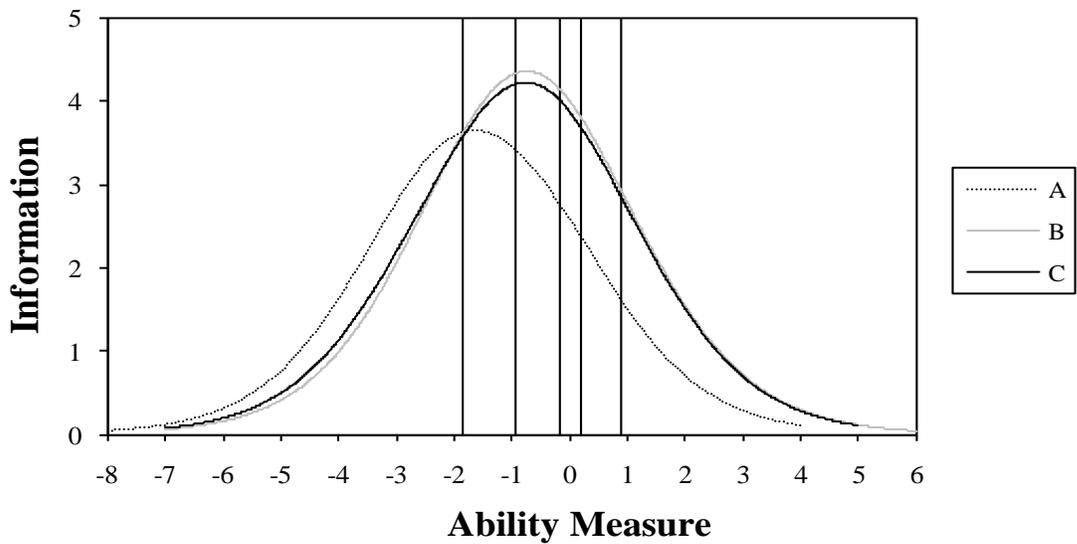
Table 8.2.1D

Weighted Reliability: List 1-2 S301

Tiers	No. of Students	Reliability	Reliability
A	72,907	0.753	
B	188,125	0.689	0.711
C	84,538	0.724	

Table 8.2.1E-1

Accuracy and Consistency of Classification Indices: List (Grade 1) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.530	0.399		0.184	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.855		0.642	
	2	0.459		0.282	
	3	0.446		0.317	
	4	0.406		0.328	
	5	0.619		0.554	
	6	0.378		0.180	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.986	0.002	0.012	0.981
	2/3	0.963	0.015	0.023	0.935
	3/4	0.834	0.104	0.062	0.762
	4/5	0.756	0.106	0.138	0.690
	5/6	0.928	0.058	0.015	0.860

Table 8.2.1E-2

Accuracy and Consistency of Classification Indices: List (Grade 2) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.538	0.419		0.208	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.791		0.435	
	2	0.459		0.259	
	3	0.382		0.252	
	4	0.295		0.230	
	5	0.665		0.594	
	6	0.606		0.440	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.991	0.001	0.009	0.988
	2/3	0.967	0.010	0.024	0.942
	3/4	0.872	0.079	0.049	0.806
	4/5	0.787	0.128	0.086	0.723
	5/6	0.854	0.070	0.076	0.782

8.2.2 Reading 1-2

Figure 8.2.2A
Scale Scores: Read 1-2 S301

Figure 8.2.2B
Proficiency Level: Read 1-2 S301
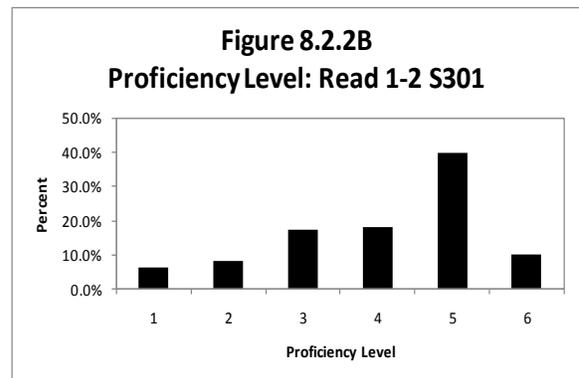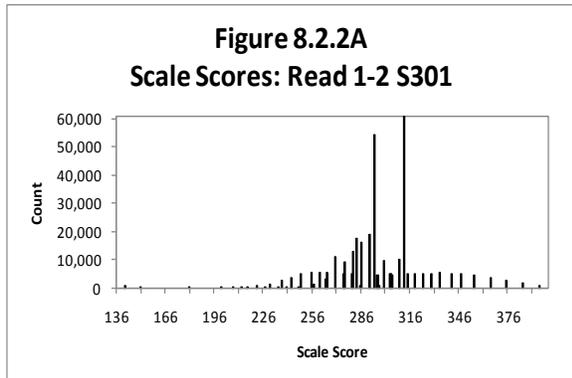
Table 8.2.2A

Scale Score Descriptive Statistics: Read 1-2 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	178,300	141	395	283.24	24.04
2	167,023	150	395	310.21	26.00
Total	345,323	141	395	296.29	28.41

Table 8.2.2B

Proficiency Level Distribution: Read 1-2 S301

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
1	14,600	8.2%	6,915	4.1%	21,515	6.2%
2	17,587	9.9%	11,120	6.7%	28,707	8.3%
3	32,771	18.4%	27,283	16.3%	60,054	17.4%
4	40,935	23.0%	21,291	12.7%	62,226	18.0%
5	62,146	34.9%	75,970	45.5%	138,116	40.0%
6	10,261	5.8%	24,444	14.6%	34,705	10.1%
Total	178,300	100.0%	167,023	100.0%	345,323	100.0%

Table 8.2.2C
Conditional Standard Error of Measurement at Cut Scores: Read 1-2 S301

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	1	253	11.96	15.34	14.30
	2	267	11.44	13.00	12.74
2/3	1	269	11.44	12.74	12.48
	2	286	11.70	11.18	11.18
3/4	1	283	11.70	11.44	11.44
	2	303	13.00	10.66	10.92
4/5	1	294	n/a	10.92	10.92
	2	312	n/a	10.66	10.92
5/6	1	314	n/a	n/a	11.18
	2	331	n/a	n/a	11.96

Figure 8.2.2C
Test Characteristic Curve: Read 1-2ABC S301
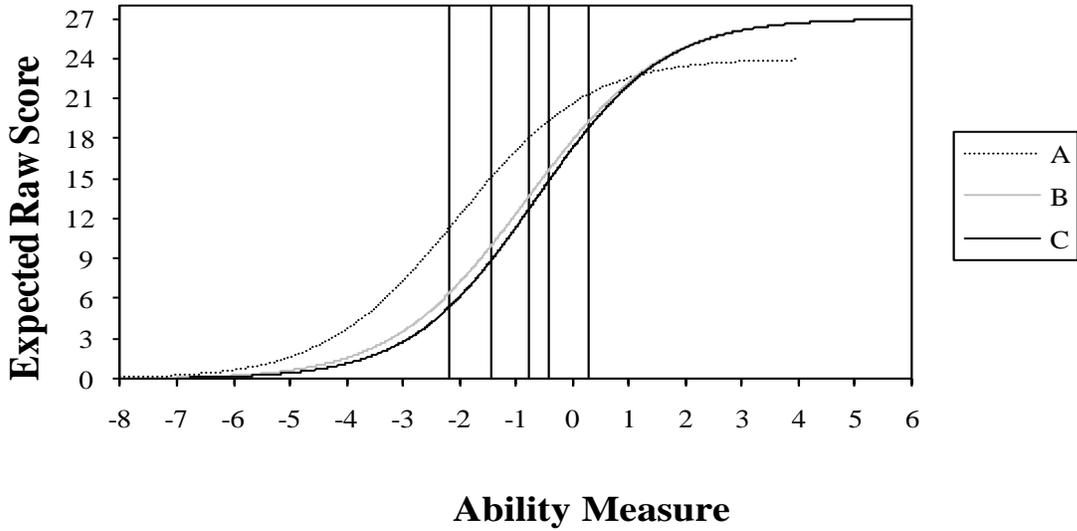
Figure 8.2.2D
Test Information Function: Read 1-2ABC S301
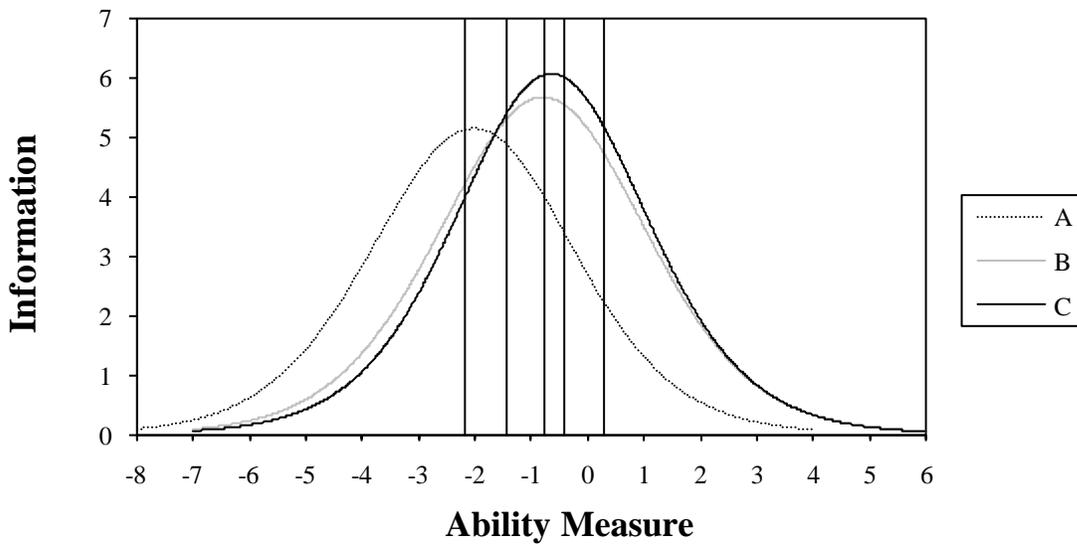
Table 8.2.2D

Weighted Reliability: Read 1-2 S301

Tiers	No. of Students	Reliability	Reliability
A	72,774	0.806	
B	188,034	0.829	0.825
C	84,515	0.831	

Table 8.2.2E-1

Accuracy and Consistency of Classification Indices: Read (Grade 1) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.555	0.437		0.270	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.823		0.689	
	2	0.505		0.358	
	3	0.488		0.355	
	4	0.411		0.315	
	5	0.617		0.553	
	6	-		0.156	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.967	0.014	0.019	0.951
	2/3	0.928	0.031	0.040	0.895
	3/4	0.864	0.055	0.081	0.808
	4/5	0.808	0.078	0.114	0.736
	5/6	0.942	0.058	0.000	0.908

Table 8.2.2E-2

Accuracy and Consistency of Classification Indices: Read (Grade 2) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.570	0.453		0.282	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.799		0.594	
	2	0.448		0.301	
	3	0.485		0.351	
	4	0.262		0.202	
	5	0.729		0.643	
	6	0.601		0.451	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.977	0.006	0.017	0.966
	2/3	0.944	0.027	0.029	0.912
	3/4	0.876	0.067	0.057	0.820
	4/5	0.827	0.108	0.065	0.769
	5/6	0.888	0.044	0.068	0.833

8.2.3 Writing 1-2

Figure 8.2.3A
Scale Scores: Writ 1-2 S301

Figure 8.2.3B
Proficiency Level: Writ 1-2 S301
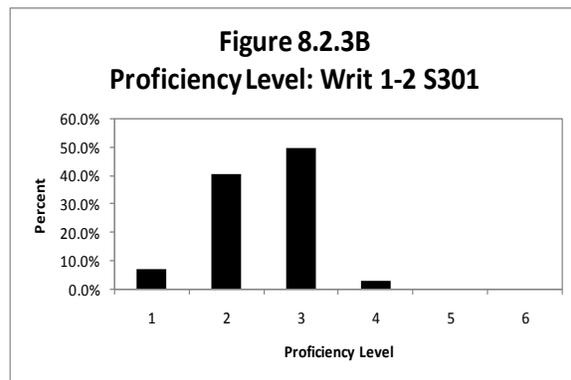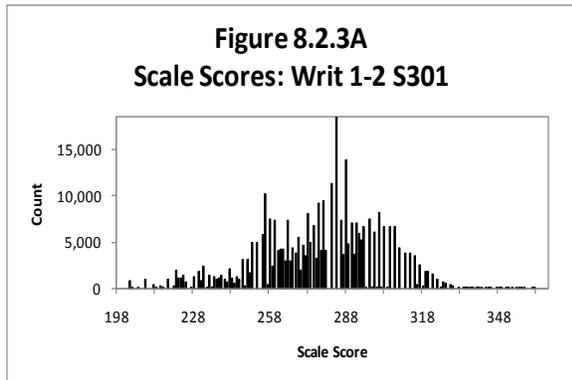
Table 8.2.3A
Scale Score Descriptive Statistics: Writ 1-2 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	178,273	203	347	268.01	21.09
2	166,991	209	362	286.82	21.05
Total	345,264	203	362	277.11	23.07

Table 8.2.3B
Proficiency Level Distribution: Writ 1-2 S301

Level	Grade 1 Count	Grade 1 Percent	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	15,911	8.9%	8,617	5.2%	24,528	7.1%
2	78,834	44.2%	60,740	36.4%	139,574	40.4%
3	79,918	44.8%	91,420	54.7%	171,338	49.6%
4	3,596	2.0%	6,191	3.7%	9,787	2.8%
5	14	0.0%	23	0.0%	37	0.0%
6	0	0.0%	0	0.0%	0	0.0%
Total	178,273	100.0%	166,991	100.0%	345,264	100.0%

Table 8.2.3C

Conditional Standard Error of Measurement at Cut Scores: Writ 1-2 S301

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	1	238	7.15	6.84	6.53
	2	251	6.84	7.77	7.46
2/3	1	272	7.77	8.09	8.40
	2	285	8.40	8.09	8.40
3/4	1	308	8.09	7.77	7.77
	2	320	7.77	7.46	7.46
4/5	1	336	n/a	6.84	6.84
	2	348	n/a	6.53	6.53
5/6	1	362	n/a	n/a	6.53
	2	373	n/a	n/a	7.46

Figure 8.2.3C

Test Characteristic Curve: Writ 1-2ABC S301
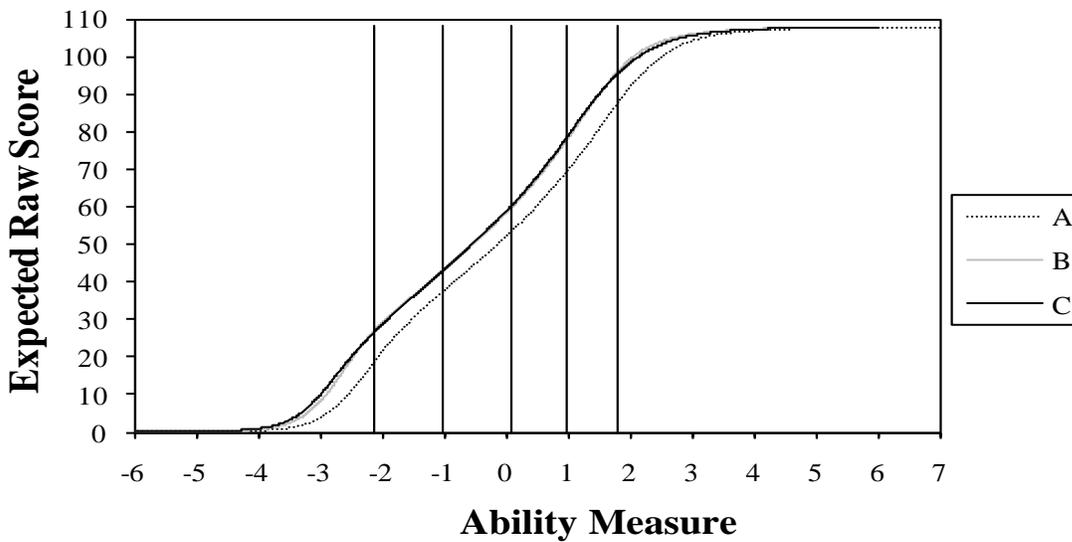
Figure 8.2.3D

Test Information Function: Writ 1-2ABC S301
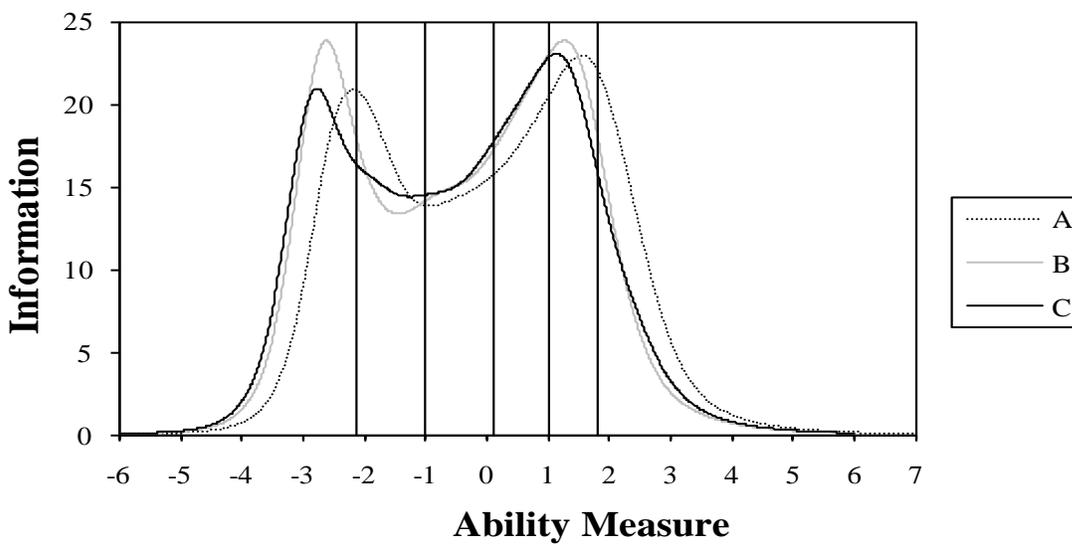
Table 8.2.3D

Weighted Reliability: Writ 1-2 S301

Tiers	No. of Students	Reliability	Reliability
A	72,801	0.897	
B	187,957	0.925	0.921
C	84,506	0.933	

Table 8.2.3E-1

Accuracy and Consistency of Classification Indices: Writ (Grade 1) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.864	0.814		0.684	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.851		0.752	
	2	0.869		0.818	
	3	0.863		0.829	
	4	-		0.296	
	5	-		1.000	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.970	0.013	0.018	0.957
	2/3	0.915	0.041	0.045	0.880
	3/4	0.980	0.020	0.000	0.977
	4/5	1.000	0.000	0.000	1.000

Table 8.2.3E-2

Accuracy and Consistency of Classification Indices: Writ (Grade 2) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.866	0.820		0.673	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.818		0.708	
	2	0.880		0.823	
	3	0.862		0.840	
	4	-		0.246	
	5	-		1.000	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.981	0.009	0.010	0.972
	2/3	0.922	0.033	0.046	0.890
	3/4	0.963	0.037	0.000	0.957
	4/5	1.000	0.000	0.000	1.000

8.2.4 Speaking 1-2

Figure 8.2.4A
Scale Scores: Spek 1-2 S301

Figure 8.2.4B
Proficiency Level: Spek 1-2 S301
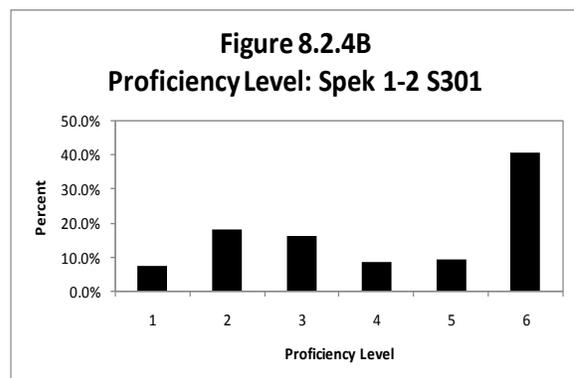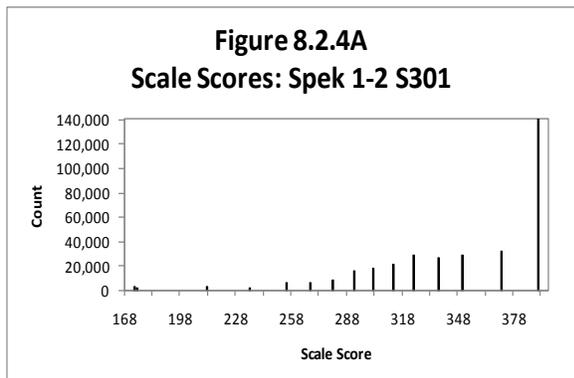
Table 8.2.4A

Scale Score Descriptive Statistics: Spek 1-2 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	178,328	173	391	339.34	48.98
2	167,081	174	391	358.48	43.04
Total	345,409	173	391	348.60	47.18

Table 8.2.4B

Proficiency Level Distribution: Spek 1-2 S301

Level	Grade 1 Count	Grade 1 Percent	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	15,544	8.7%	9,896	5.9%	25,440	7.4%
2	43,384	24.3%	19,003	11.4%	62,387	18.1%
3	31,704	17.8%	23,651	14.2%	55,355	16.0%
4	14,933	8.4%	14,163	8.5%	29,096	8.4%
5	15,854	8.9%	16,429	9.8%	32,283	9.3%
6	56,909	31.9%	83,939	50.2%	140,848	40.8%
Total	178,328	100.0%	167,081	100.0%	345,409	100.0%

Table 8.2.4C
Conditional Standard Error of Measurement at
Cut Scores: Spek 1-2 S301

Proficiency Level	Grade	Cut Score	SEM
1/2	1	278	20.89
	2	286	19.88
2/3	1	318	18.28
	2	322	18.28
3/4	1	344	19.08
	2	345	19.08
4/5	1	367	20.08
	2	368	20.08
5/6	1	385	20.69
	2	386	20.69

Figure 8.2.4C
Test Characteristic Curve: Spek 1-2 S301

Figure 8.2.4D
Test Information Function: Spek 1-2 S301
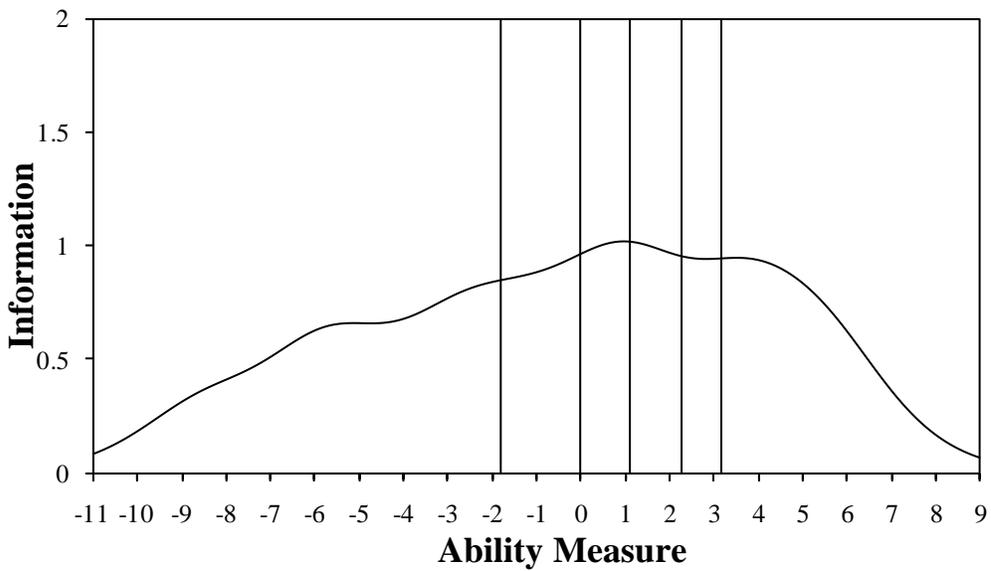
Table 8.2.4D

Reliability: Spek 1-2 S301

Tiers	No. of Students	Reliability
--	345,409	0.897

Table 8.2.4E-1

Accuracy and Consistency of Classification Indices: Spek (Grade 1) S301

Overall Indices	Accuracy	Consistency		Kappa (k)		
	0.616	0.528		0.419		
Conditional on Level	**Level**	**Accuracy**		**Consistency**		
	1	0.654		0.523		
	2	0.706		0.609		
	3	0.531		0.430		
	4	0.348		0.238		
	5	0.311		0.234		
	6	0.938		0.871		
Indices at Cut Points			**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**	
	1/2	0.945	0.036	0.019	0.918	
	2/3	0.895	0.044	0.062	0.861	
	3/4	0.922	0.023	0.055	0.886	
	4/5	0.945	0.031	0.023	0.910	
	5/6	0.885	0.101	0.014	0.865	

Table 8.2.4E-2

Accuracy and Consistency of Classification Indices: Spek (Grade 2) S301

Overall Indices	Accuracy	Consistency		Kappa (k)		
	0.667	0.600		0.448		
Conditional on Level	**Level**	**Accuracy**		**Consistency**		
	1	0.710		0.566		
	2	0.568		0.459		
	3	0.551		0.455		
	4	0.355		0.258		
	5	0.322		0.237		
	6	0.953		0.906		
Indices at Cut Points			**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**	
	1/2	0.967	0.018	0.014	0.951	
	2/3	0.929	0.035	0.037	0.906	
	3/4	0.923	0.021	0.057	0.896	
	4/5	0.944	0.027	0.029	0.910	
	5/6	0.879	0.101	0.020	0.853	

8.2.5 Oral Language Composite 1-2

Figure 8.2.5A
Scale Scores: Oral 1-2 S301

Figure 8.2.5B
Proficiency Level: Oral 1-2 S301
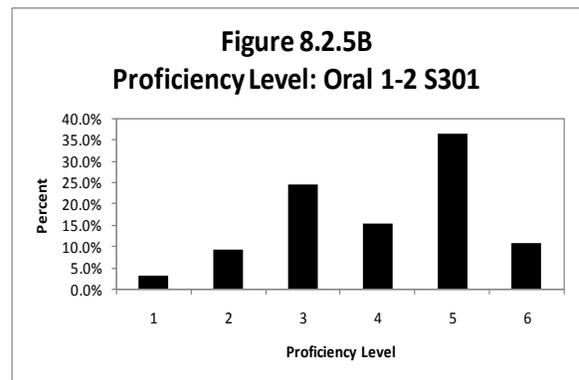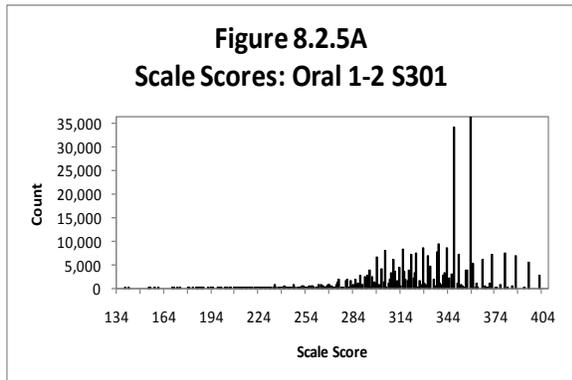
Table 8.2.5A

Scale Score Descriptive Statistics: Oral 1-2 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	178,240	139	402	319.44	32.58
2	167,026	141	402	342.90	32.04
Total	345,266	139	402	330.79	34.38

Table 8.2.5B

Proficiency Level Distribution: Oral 1-2 S301

Level	Grade 1 Count	Grade 1 Percent	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	7,331	4.1%	4,235	2.5%	11,566	3.3%
2	21,903	12.3%	9,794	5.9%	31,697	9.2%
3	53,620	30.1%	30,666	18.4%	84,286	24.4%
4	27,960	15.7%	25,455	15.2%	53,415	15.5%
5	56,081	31.5%	70,456	42.2%	126,537	36.6%
6	11,345	6.4%	26,420	15.8%	37,765	10.9%
Total	178,240	100.0%	167,026	100.0%	345,266	100.0%

Table 8.2.5C

n/a

Figure 8.2.5C

n/a

Figure 8.2.5D

n/a

Table 8.2.5D

Oral Composite Reliability: Oral 1-2 S301

Component	Weight	Variance	Reliability
Listening	0.50	1066.525	0.711
Speaking	0.50	2220.950	0.897
Oral		1179.390	0.886

*Variances from students who had results in all four domains

Table 8.2.5E-1

Accuracy and Consistency of Classification Indices: Oral (Grade 1) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.646	0.526		0.391	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.823		0.689	
	2	0.669		0.535	
	3	0.746		0.635	
	4	0.391		0.295	
	5	0.712		0.630	
	6	0.499		0.299	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.983	0.007	0.011	0.975
	2/3	0.938	0.032	0.030	0.910
	3/4	0.892	0.038	0.070	0.848
	4/5	0.885	0.067	0.049	0.836
	5/6	0.936	0.045	0.018	0.904

Table 8.2.5E-2

Accuracy and Consistency of Classification Indices: Oral (Grade 2) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.634	0.517		0.361	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.843		0.712	
	2	0.592		0.445	
	3	0.717		0.586	
	4	0.432		0.325	
	5	0.727		0.638	
	6	0.558		0.431	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.990	0.003	0.007	0.985
	2/3	0.964	0.020	0.016	0.946
	3/4	0.919	0.028	0.053	0.887
	4/5	0.890	0.058	0.052	0.842
	5/6	0.863	0.056	0.081	0.814

8.2.6 Literacy Composite 1-2

Figure 8.2.6A
Scale Scores: Litr 1-2 S301

Figure 8.2.6B
Proficiency Level: Litr 1-2 S301
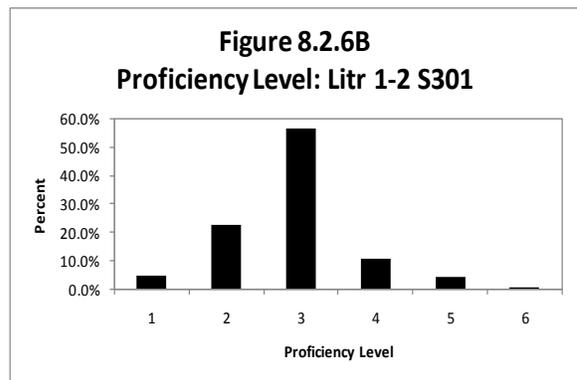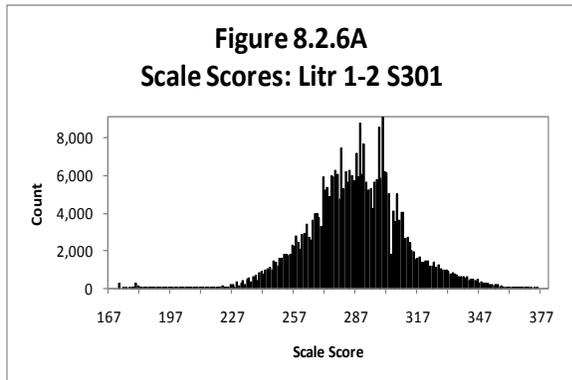
Table 8.2.6A
Scale Score Descriptive Statistics: Litr 1-2 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	178,166	172	368	275.90	19.72
2	166,937	180	375	298.75	21.34
Total	345,103	172	375	286.95	23.48

Table 8.2.6B
Proficiency Level Distribution: Litr 1-2 S301

Level	Grade 1 Count	Grade 1 Percent	Grade 2 Count	Grade 2 Percent	Total Count	Total Percent
1	10,322	5.8%	6,497	3.9%	16,819	4.9%
2	49,385	27.7%	28,692	17.2%	78,077	22.6%
3	99,427	55.8%	96,319	57.7%	195,746	56.7%
4	13,807	7.7%	22,993	13.8%	36,800	10.7%
5	4,386	2.5%	11,096	6.6%	15,482	4.5%
6	839	0.5%	1,340	0.8%	2,179	0.6%
Total	178,166	100.0%	166,937	100.0%	345,103	100.0%

Table 8.2.6C
n/a

Figure 8.2.6C
n/a

Figure 8.2.6D
n/a

Table 8.2.6D
Literacy Composite Reliability: Litr 1-2 S301

Component	Weight	Variance	Reliability
Reading	0.50	806.125	0.825
Writing	0.50	531.779	0.921
Literacy		551.177	0.917

Table 8.2.6E-1
Accuracy and Consistency of Classification Indices: Litr (Grade 1) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.809	0.740		0.582	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.768		0.704	
	2	0.796		0.710	
	3	0.883		0.846	
	4	0.504		0.396	
	5	0.884		0.668	
	6	-		1.000	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.974	0.014	0.013	0.967
	2/3	0.910	0.049	0.041	0.877
	3/4	0.939	0.027	0.034	0.916
	4/5	0.972	0.027	0.000	0.978
	5/6	0.995	0.005	0.000	1.000

Table 8.2.6E-2

Accuracy and Consistency of Classification Indices: Litr (Grade 2) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.788	0.711		0.549	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.794		0.738	
	2	0.741		0.633	
	3	0.896		0.856	
	4	0.540		0.437	
	5	0.757		0.624	
	6	-		0.986	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.983	0.008	0.010	0.979
	2/3	0.930	0.041	0.030	0.901
	3/4	0.916	0.029	0.055	0.883
	4/5	0.956	0.038	0.006	0.950
	5/6	0.992	0.008	0.000	0.995

8.2.7 Comprehension Composite 1-2

Figure 8.2.7A
Scale Scores: Cphn 1-2 S301

Figure 8.2.7B
Proficiency Level: Cphn 1-2 S301
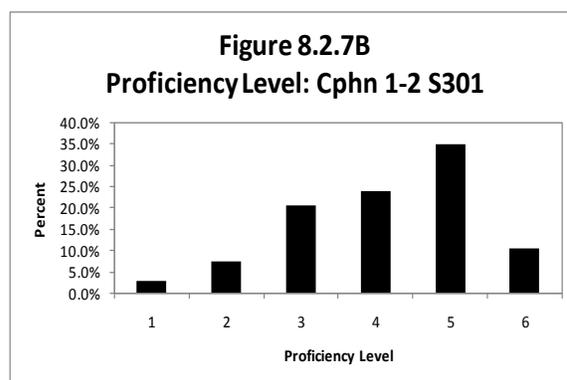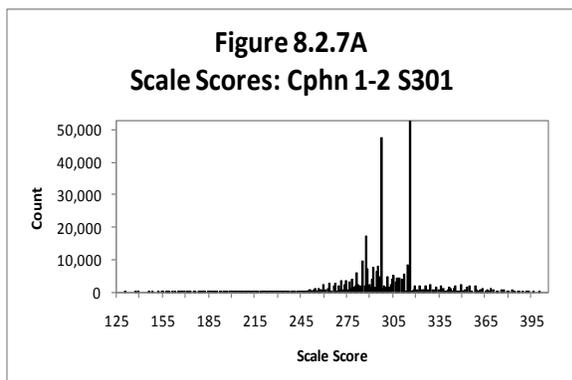
Table 8.2.7A

Scale Score Descriptive Statistics: Cphn 1-2 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	178,263	130	400	288.02	22.42
2	167,002	137	400	315.29	25.65
Total	345,265	130	400	301.21	27.63

Table 8.2.7B

Proficiency Level Distribution: Cphn 1-2 S301

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
1	6,328	3.5%	3,471	2.1%	9,799	2.8%
2	17,216	9.7%	8,310	5.0%	25,526	7.4%
3	40,868	22.9%	29,764	17.8%	70,632	20.5%
4	49,981	28.0%	32,198	19.3%	82,179	23.8%
5	53,409	30.0%	67,107	40.2%	120,516	34.9%
6	10,461	5.9%	26,152	15.7%	36,613	10.6%
Total	178,263	100.0%	167,002	100.0%	345,265	100.0%

Table 8.2.7C

n/a

Figure 8.2.7C

n/a

Figure 8.2.7D

n/a

Table 8.2.7D

Comprehension Composite Reliability: Cphn 1-2 S301

Component	Weight	Variance	Reliability
Listening	0.30	1066.525	0.711
Reading	0.70	806.125	0.825
Comprehension		763.199	0.873

*Variances from students who had results in all four domains

Table 8.2.7E-1

Accuracy and Consistency of Classification Indices: Cphn (Grade 1) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.623	0.511		0.362	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.788		0.634	
	2	0.650		0.496	
	3	0.598		0.483	
	4	0.577		0.472	
	5	0.645		0.563	
	6	0.657		0.400	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.983	0.007	0.010	0.975
	2/3	0.948	0.023	0.029	0.923
	3/4	0.872	0.074	0.054	0.824
	4/5	0.856	0.055	0.089	0.806
	5/6	0.950	0.041	0.009	0.930

Table 8.2.7E-2

Accuracy and Consistency of Classification Indices: Cphn (Grade 2) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.641	0.527		0.374	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.855		0.682	
	2	0.535		0.377	
	3	0.607		0.470	
	4	0.446		0.349	
	5	0.727		0.639	
	6	0.722		0.582	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.989	0.002	0.009	0.985
	2/3	0.965	0.018	0.017	0.944
	3/4	0.895	0.057	0.048	0.850
	4/5	0.857	0.075	0.067	0.806
	5/6	0.916	0.039	0.045	0.875

8.2.8 Overall Composite 1-2

Figure 8.2.8A
Scale Scores: Over 1-2 S301

Figure 8.2.8B
Proficiency Level: Over 1-2 S301
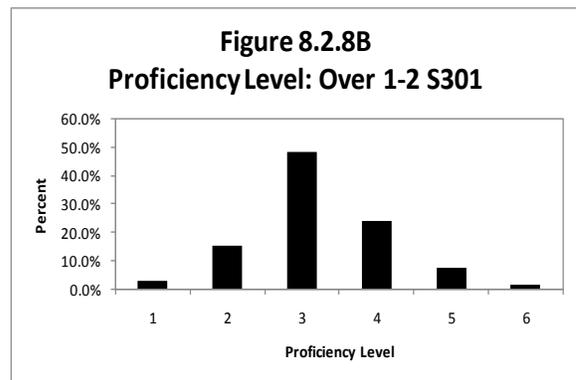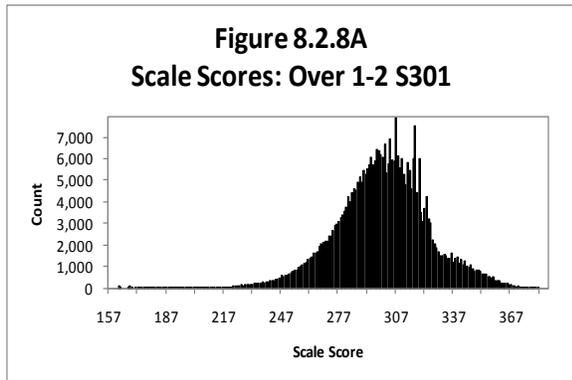
Table 8.2.8A

Scale Score Descriptive Statistics: Over 1-2 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	178,034	162	376	288.75	21.17
2	166,838	168	381	311.79	22.53
Total	344,872	162	381	299.90	24.69

Table 8.2.8B

Proficiency Level Distribution: Over 1-2 S301

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
1	6,389	3.6%	4,273	2.6%	10,662	3.1%
2	36,010	20.2%	16,653	10.0%	52,663	15.3%
3	95,132	53.4%	71,545	42.9%	166,677	48.3%
4	31,527	17.7%	51,818	31.1%	83,345	24.2%
5	7,305	4.1%	18,969	11.4%	26,274	7.6%
6	1,671	0.9%	3,580	2.1%	5,251	1.5%
Total	178,034	100.0%	166,838	100.0%	344,872	100.0%

Table 8.2.8C
n/a

Figure 8.2.8C
n/a

Figure 8.2.8D
n/a

Table 8.2.8D
Overall Composite Reliability: Over 1-2 S301

Component	Weight	Variance	Reliability
Listening	0.15	1066.525	0.711
Reading	0.35	806.125	0.825
Speaking	0.15	2220.950	0.897
Writing	0.35	531.779	0.921
Overall Composite		609.572	0.943

*Variances from students who had results in all four domains

Table 8.2.8E-1
Accuracy and Consistency of Classification Indices: Over (Grade 1) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.828	0.767		0.642	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.713		0.754	
	2	0.820		0.737	
	3	0.888		0.853	
	4	0.706		0.615	
	5	0.791		0.589	
	6	-		1.000	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.981	0.012	0.007	0.983
	2/3	0.936	0.037	0.027	0.917
	3/4	0.923	0.037	0.040	0.902
	4/5	0.966	0.027	0.006	0.967
	5/6	0.991	0.009	0.000	0.998

Table 8.2.8E-2

Accuracy and Consistency of Classification Indices: Over (Grade 2) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.796	0.723		0.605	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.814		0.811	
	2	0.759		0.653	
	3	0.879		0.827	
	4	0.760		0.680	
	5	0.654		0.561	
	6	-		0.950	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.990	0.005	0.005	0.990
	2/3	0.961	0.023	0.016	0.946
	3/4	0.915	0.036	0.049	0.884
	4/5	0.942	0.032	0.025	0.921
	5/6	0.978	0.022	0.000	0.981

8.3   Grades: 3–5

8.3.1 Listening 3-5

Figure 8.3.1A
Scale Scores: List 3-5 S301

Figure 8.3.1B
Proficiency Level: List 3-5 S301
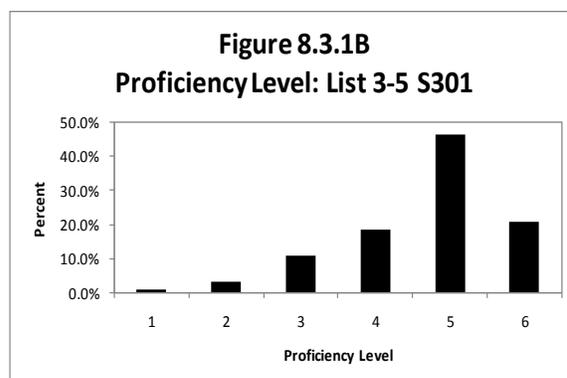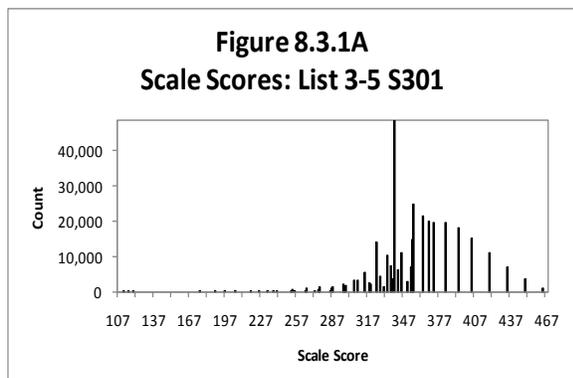
Table 8.3.1A
Scale Score Descriptive Statistics: List 3-5 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	143,026	112	464	347.56	32.24
4	99,216	116	464	360.57	35.21
5	81,854	120	464	370.44	35.75
Total	324,096	112	464	357.32	35.35

Table 8.3.1B
Proficiency Level Distribution: List 3-5 S301

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,037	0.7%	792	0.8%	864	1.1%	2,693	0.8%
2	3,090	2.2%	3,670	3.7%	3,421	4.2%	10,181	3.1%
3	12,239	8.6%	12,293	12.4%	10,382	12.7%	34,914	10.8%
4	23,894	16.7%	19,483	19.6%	15,851	19.4%	59,228	18.3%
5	70,323	49.2%	43,840	44.2%	36,238	44.3%	150,401	46.4%
6	32,443	22.7%	19,138	19.3%	15,098	18.4%	66,679	20.6%
Total	143,026	100.0%	99,216	100.0%	81,854	100.0%	324,096	100.0%

Table 8.3.1C
Conditional Standard Error of Measurement at Cut Scores: List 3-5 S301

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	3	255	21.42	22.17	27.05
	4	264	21.04	21.04	25.17
	5	274	20.29	20.29	23.67
2/3	3	295	19.91	19.16	20.66
	4	307	19.91	18.79	19.54
	5	318	20.29	18.41	18.79
3/4	3	325	20.66	18.41	18.41
	4	338	21.42	18.79	18.03
	5	350	22.54	19.16	18.03
4/5	3	340	n/a	18.79	18.03
	4	355	n/a	19.54	18.03
	5	368	n/a	20.29	18.41
5/6	3	367	n/a	n/a	18.41
	4	383	n/a	n/a	19.54
	5	397	n/a	n/a	21.04

Figure 8.3.1C

Test Characteristic Curve: List 3-5ABC S301
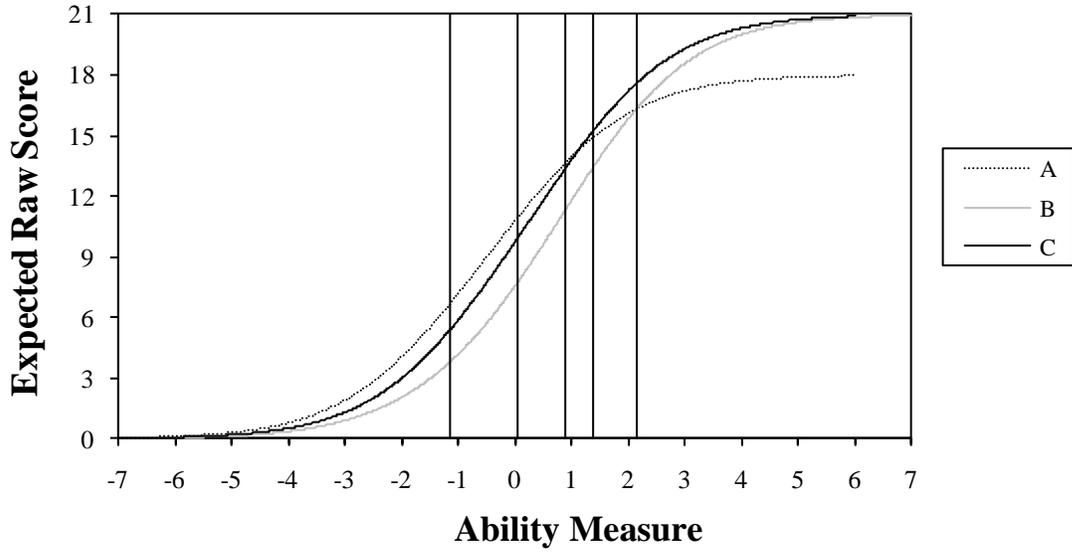
Figure 8.3.1D

Test Information Function: List 3-5ABC S301

Return to Visual Table of Tables and Figures
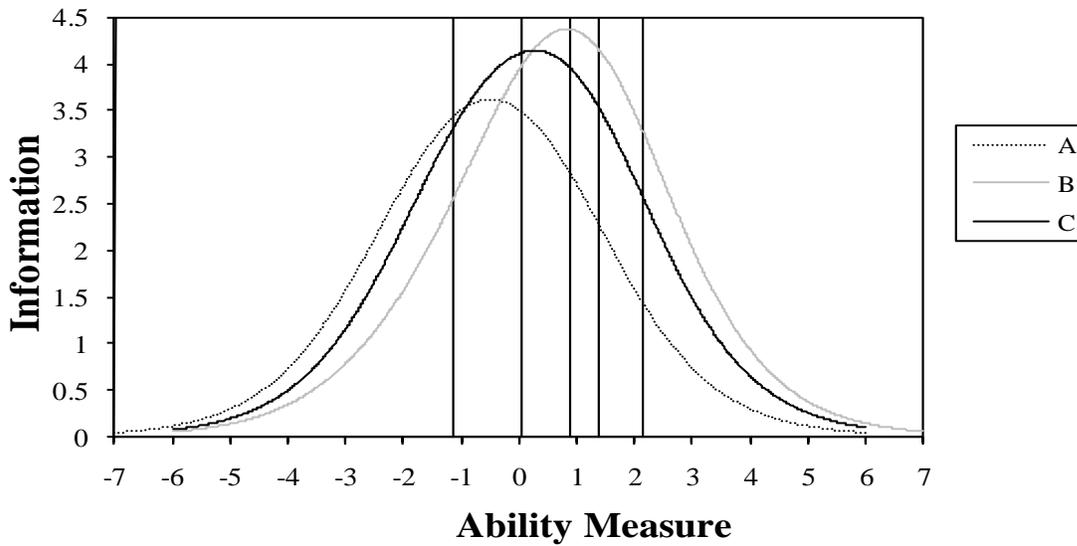
Table 8.3.1D

Weighted Reliability: List 3-5 S301

Tiers	No. of Students	Reliability	Weighted Reliability
A	26,444	0.780	
B	141,303	0.669	0.634
C	156,349	0.577	

Table 8.3.1E-1

Accuracy and Consistency of Classification Indices: List (Grade 3) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.523	0.399		0.163	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.855		0.504	
	2	0.404		0.196	
	3	0.351		0.204	
	4	0.313		0.234	
	5	0.619		0.558	
	6	0.561		0.416	
Indices at Cut Points		**Accuracy**			
			False Positives	**False Negatives**	
	Cut Point	**Accuracy**			**Consistency**
	1/2	0.994	0.000	0.006	0.993
	2/3	0.979	0.006	0.015	0.958
	3/4	0.896	0.058	0.046	0.828
	4/5	0.788	0.110	0.102	0.717
	5/6	0.802	0.093	0.105	0.722

Table 8.3.1E-2

Accuracy and Consistency of Classification Indices: List (Grade 4) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.481	0.366		0.140	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.719		0.396	
	2	0.487		0.265	
	3	0.427		0.274	
	4	0.342		0.256	
	5	0.551		0.501	
	6	0.447		0.323	
Indices at Cut Points		**Accuracy**			
			False Positives	**False Negatives**	
	Cut Point	**Accuracy**			**Consistency**
	1/2	0.994	0.001	0.005	0.990
	2/3	0.967	0.012	0.021	0.942
	3/4	0.879	0.052	0.069	0.819
	4/5	0.778	0.082	0.140	0.705
	5/6	0.785	0.099	0.116	0.708

Table 8.3.1E-3
Accuracy and Consistency of Classification Indices: List (Grade 5) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.486	0.366		0.139	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.742		0.436	
	2	0.491		0.280	
	3	0.423		0.270	
	4	0.335		0.248	
	5	0.551		0.507	
	6	0.447		0.307	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.992	0.001	0.006	0.988
	2/3	0.964	0.013	0.024	0.938
	3/4	0.877	0.052	0.071	0.814
	4/5	0.779	0.088	0.134	0.700
	5/6	0.803	0.129	0.068	0.722

8.3.2 Reading 3-5

Figure 8.3.2A
Scale Scores: Read 3-5 S301

Figure 8.3.2B
Proficiency Level: Read 3-5 S301
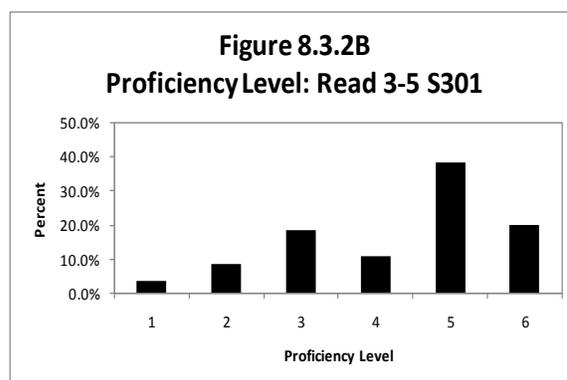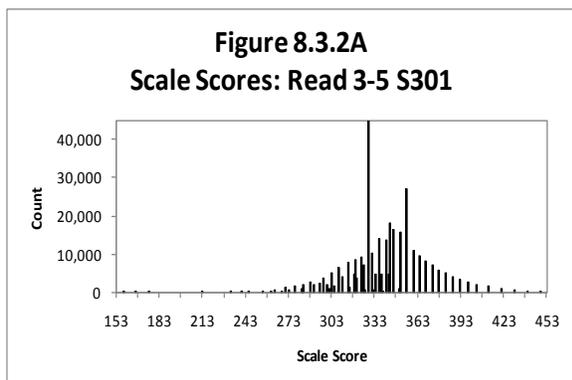
Table 8.3.2A

Scale Score Descriptive Statistics: Read 3-5 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	142,788	158	448	330.21	25.19
4	99,019	166	448	341.10	27.74
5	81,689	175	448	350.10	29.10
Total	323,496	158	448	338.56	28.21

Table 8.3.2B

Proficiency Level Distribution: Read 3-5 S301

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	3,293	2.3%	4,031	4.1%	4,207	5.2%	11,531	3.6%
2	9,516	6.7%	9,468	9.6%	8,957	11.0%	27,941	8.6%
3	23,120	16.2%	17,534	17.7%	19,573	24.0%	60,227	18.6%
4	12,509	8.8%	15,340	15.5%	6,717	8.2%	34,566	10.7%
5	66,146	46.3%	31,176	31.5%	27,036	33.1%	124,358	38.4%
6	28,204	19.8%	21,470	21.7%	15,199	18.6%	64,873	20.1%
Total	142,788	100.0%	99,019	100.0%	81,689	100.0%	323,496	100.0%

Table 8.3.2C
Conditional Standard Error of Measurement at Cut Scores: Read 3-5 S301

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	3	279	12.48	15.34	22.88
	4	291	11.96	13.52	19.24
	5	302	11.70	12.48	16.38
2/3	3	302	11.70	12.48	16.38
	4	316	11.96	11.44	14.04
	5	328	12.48	10.92	12.48
3/4	3	320	11.96	11.18	13.52
	4	336	13.00	10.92	11.70
	5	350	14.56	11.18	10.92
4/5	3	328	n/a	10.92	12.48
	4	343	n/a	10.92	11.18
	5	355	n/a	11.44	10.66
5/6	3	347	n/a	n/a	10.92
	4	360	n/a	n/a	10.66
	5	372	n/a	n/a	10.66

Figure 8.3.2C
Test Characteristic Curve: Read 3-5ABC S301
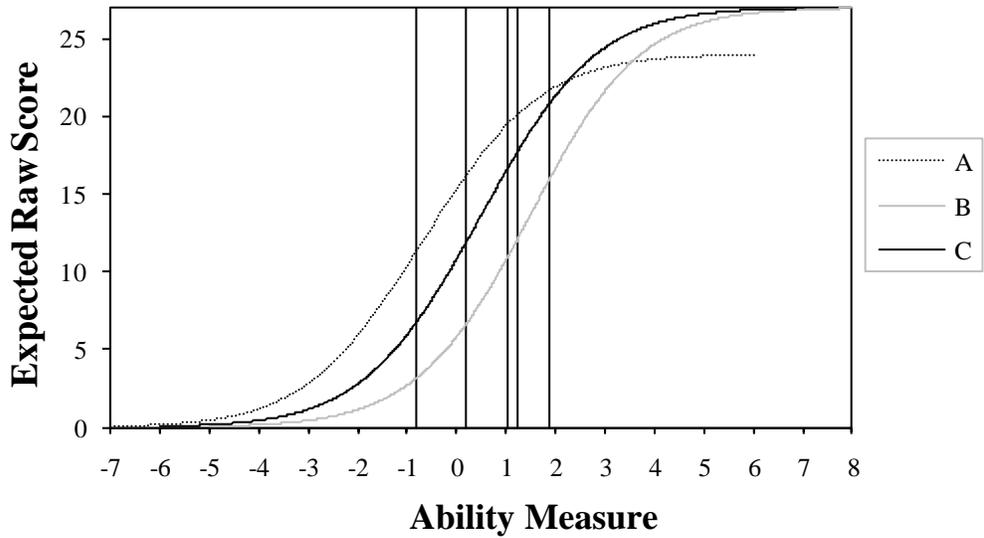
Figure 8.3.2D
Test Information Function: Read 3-5ABC S301
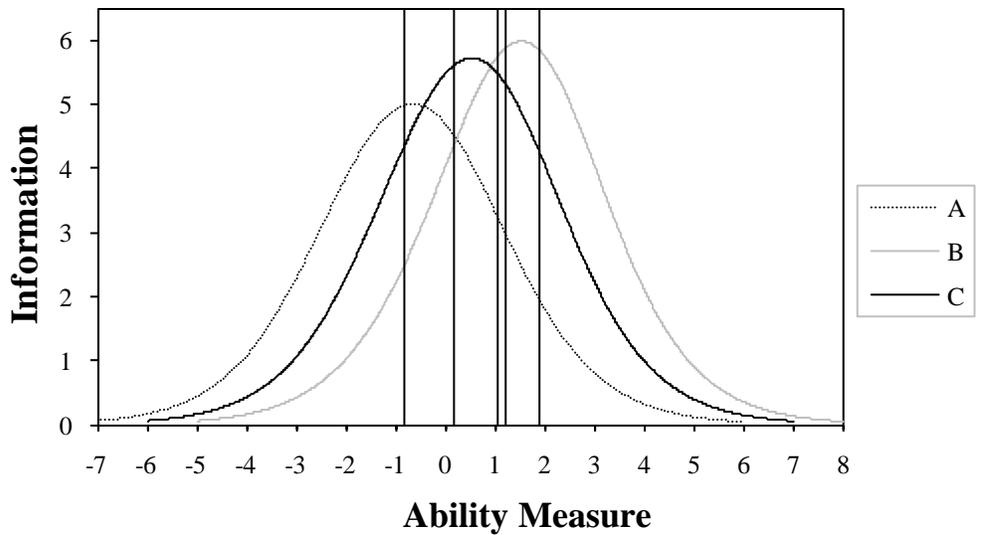
Table 8.3.2D
Weighted Reliability: Read 3-5 S301

Tiers	No. of Students	Reliability	Weighted Reliability
A	26,270	0.841	
B	140,984	0.802	0.782
C	156,242	0.753	

Table 8.3.2E-1
Accuracy and Consistency of Classification Indices: Read (Grade 3) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.562	0.443		0.252	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.763		0.545	
	2	0.542		0.366	
	3	0.519		0.368	
	4	0.190		0.140	
	5	0.682		0.598	
	6	0.560		0.436	
Indices at Cut Points		**Accuracy**			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.987	0.004	0.009	0.979
	2/3	0.951	0.021	0.028	0.924
	3/4	0.883	0.050	0.067	0.828
	4/5	0.841	0.087	0.071	0.777
	5/6	0.831	0.065	0.104	0.768

Table 8.3.2E-2
Accuracy and Consistency of Classification Indices: Read (Grade 4) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.504	0.400		0.240	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.778		0.588	
	2	0.560		0.390	
	3	0.449		0.333	
	4	0.296		0.227	
	5	0.495		0.420	
	6	0.602		0.469	
Indices at Cut Points		**Accuracy**			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.978	0.007	0.015	0.967
	2/3	0.935	0.026	0.039	0.901
	3/4	0.852	0.077	0.071	0.794
	4/5	0.811	0.076	0.114	0.751
	5/6	0.831	0.076	0.093	0.767

Table 8.3.2E-3

Accuracy and Consistency of Classification Indices: Read (Grade 5) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.518	0.409		0.249	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.782		0.587	
	2	0.523		0.367	
	3	0.514		0.398	
	4	0.152		0.118	
	5	0.541		0.457	
	6	0.580		0.439	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.972	0.009	0.019	0.958
	2/3	0.920	0.036	0.044	0.880
	3/4	0.833	0.083	0.084	0.771
	4/5	0.814	0.093	0.093	0.752
	5/6	0.846	0.070	0.084	0.785

8.3.3 Writing 3-5

Figure 8.3.3A
Scale Scores: Writ 3-5 S301

Figure 8.3.3B
Proficiency Level: Writ 3-5 S301
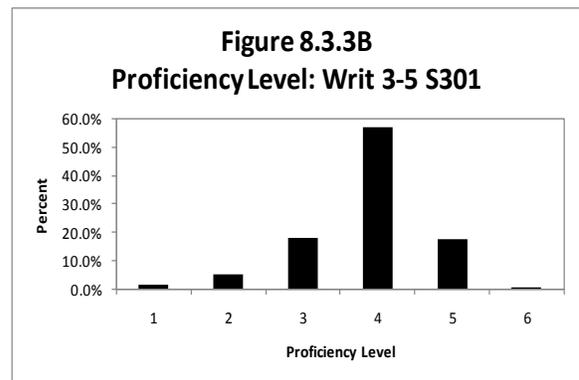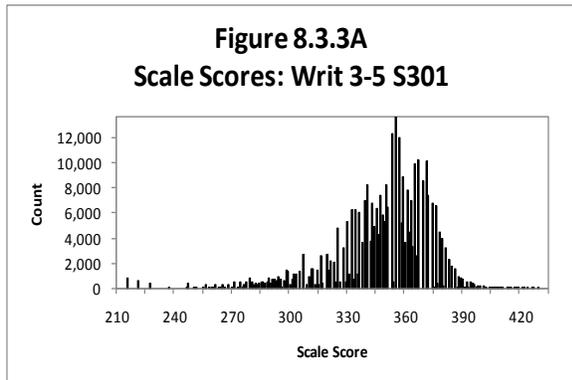
Table 8.3.3A

Scale Score Descriptive Statistics: Writ 3-5 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	142,788	215	418	341.84	25.73
4	99,012	221	429	349.44	25.35
5	81,691	227	427	355.77	24.78
Total	323,491	215	429	347.68	26.01

Table 8.3.3B

Proficiency Level Distribution: Writ 3-5 S301

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	1,620	1.1%	1,585	1.6%	1,774	2.2%	4,979	1.5%
2	7,344	5.1%	5,544	5.6%	4,058	5.0%	16,946	5.2%
3	22,523	15.8%	17,495	17.7%	18,387	22.5%	58,405	18.1%
4	78,287	54.8%	56,724	57.3%	49,692	60.8%	184,703	57.1%
5	31,917	22.4%	17,089	17.3%	7,472	9.1%	56,478	17.5%
6	1,097	0.8%	575	0.6%	308	0.4%	1,980	0.6%
Total	142,788	100.0%	99,012	100.0%	81,691	100.0%	323,491	100.0%

Table 8.3.3C
Conditional Standard Error of Measurement at Cut Scores: Writ 3-5 S301

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	3	264	10.88	7.77	11.82
	4	275	9.02	6.84	8.09
	5	287	9.64	6.84	6.53
2/3	3	297	11.19	7.77	6.53
	4	308	11.82	8.40	7.77
	5	319	11.82	8.40	8.40
3/4	3	330	11.51	8.09	8.40
	4	340	11.51	8.09	8.40
	5	350	11.19	7.77	8.09
4/5	3	360	n/a	7.46	7.77
	4	371	n/a	7.15	7.77
	5	381	n/a	6.84	7.15
5/6	3	384	n/a	n/a	7.15
	4	394	n/a	n/a	6.84
	5	403	n/a	n/a	6.53

Figure 8.3.3C

Test Characteristic Curve: Writ 3-5ABC S301
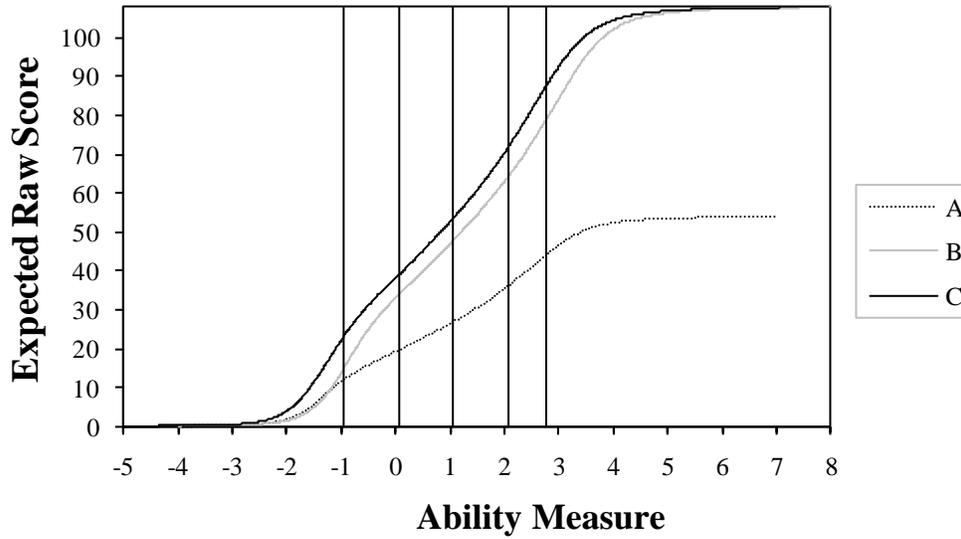
Figure 8.3.3D

Test Information Function: Writ 3-5ABC S301
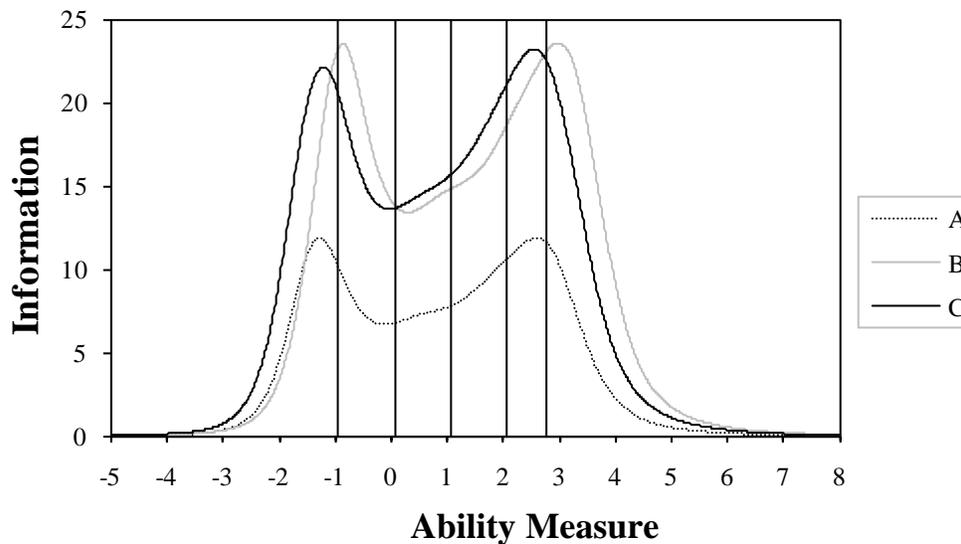
Table 8.3.3D
Weighted Reliability: Writ 3-5 S301

Tiers	No. of Students	Reliability	Weighted Reliability
A	26,240	0.915	
B	140,994	0.924	0.912
C	156,257	0.901	

Table 8.3.3E-3
Accuracy and Consistency of Classification Indices: Writ (Grade 3) S301

Overall Indices	Accuracy	Consistency			Kappa (k)	
	0.726	0.648			0.439	
Conditional on Level	Level	Accuracy			Consistency	
	1	0.827			0.733	
	2	0.807			0.708	
	3	0.788			0.684	
	4	0.791			0.719	
	5	0.541			0.457	
	6	-			0.512	
Indices at Cut Points		Accuracy				
	Cut Point	Accuracy	False Positives	False Negatives	Consistency	
	1/2	0.996	0.002	0.002	0.994	
	2/3	0.983	0.008	0.009	0.976	
	3/4	0.948	0.024	0.028	0.926	
	4/5	0.807	0.081	0.112	0.757	
	5/6	0.992	0.008	0.000	0.992	

Table 8.3.3E-2
Accuracy and Consistency of Classification Indices: Writ (Grade 4) S301

Overall Indices	Accuracy	Consistency			Kappa (k)	
	0.747	0.677			0.445	
Conditional on Level	Level	Accuracy			Consistency	
	1	0.810			0.769	
	2	0.798			0.697	
	3	0.829			0.735	
	4	0.725			0.724	
	5	-			0.331	
	6	-			0.981	
Indices at Cut Points		Accuracy				
	Cut Point	Accuracy	False Positives	False Negatives	Consistency	
	1/2	0.994	0.003	0.003	0.993	
	2/3	0.981	0.010	0.010	0.974	
	3/4	0.948	0.019	0.033	0.927	
	4/5	0.822	0.178	0.000	0.784	
	5/6	0.994	0.006	0.000	0.995	

Table 8.3.3E-3

Accuracy and Consistency of Classification Indices: Writ (Grade 5) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.814	0.752		0.541	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.770		0.824	
	2	0.727		0.607	
	3	0.851		0.759	
	4	0.811		0.797	
	5	-		0.195	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.991	0.005	0.003	0.992
	2/3	0.977	0.014	0.009	0.971
	3/4	0.931	0.024	0.045	0.909
	4/5	0.905	0.095	0.000	0.881
	5/6	0.996	0.004	0.000	1.000

8.3.4 Speaking 3-5

Figure 8.3.4A
Scale Scores: Spek 3-5 S301

Figure 8.3.4B
Proficiency Level: Spek 3-5 S301
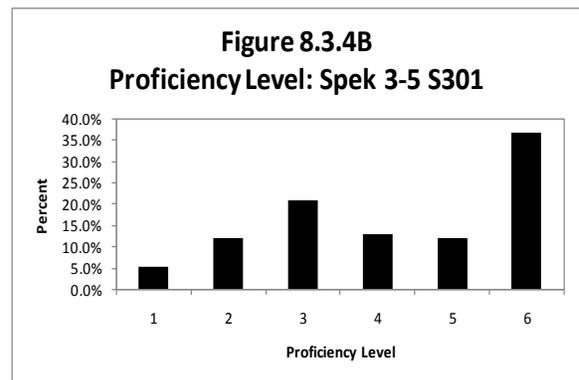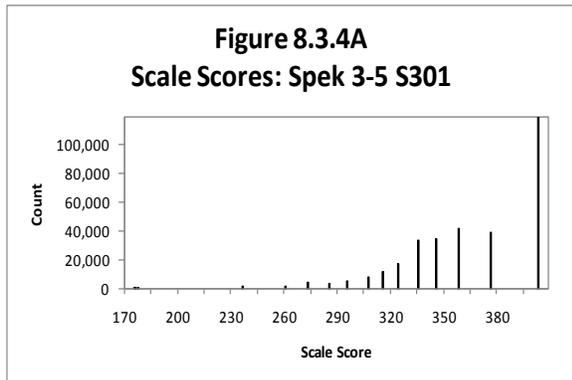
Table 8.3.4A

Scale Score Descriptive Statistics: Spek 3-5 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	142,994	175	403	360.05	40.17
4	99,160	176	403	362.98	41.29
5	81,839	177	403	366.81	41.33
Total	323,993	175	403	362.66	40.90

Table 8.3.4B

Proficiency Level Distribution: Spek 3-5 S301

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	5,971	4.2%	5,958	6.0%	4,855	5.9%	16,784	5.2%
2	21,031	14.7%	10,811	10.9%	7,476	9.1%	39,318	12.1%
3	33,300	23.3%	20,290	20.5%	14,340	17.5%	67,930	21.0%
4	18,590	13.0%	12,843	13.0%	10,068	12.3%	41,501	12.8%
5	16,659	11.7%	12,430	12.5%	10,141	12.4%	39,230	12.1%
6	47,443	33.2%	36,828	37.1%	34,959	42.7%	119,230	36.8%
Total	142,994	100.0%	99,160	100.0%	81,839	100.0%	323,993	100.0%

Table 8.3.4C
Conditional Standard Error of Measurement at
Cut Scores: Spek 3-5 S301

Proficiency Level	Grade	Cut Score	SEM
1/2	3	293	19.08
	4	299	19.48
	5	305	19.68
2/3	3	326	20.89
	4	329	21.09
	5	333	21.49
3/4	3	346	22.29
	4	348	22.49
	5	350	22.69
4/5	3	369	24.90
	4	371	25.31
	5	374	25.71
5/6	3	389	27.31
	4	391	27.52
	5	394	27.52

Figure 8.3.4C

Test Characteristic Curve: Spek 3-5 S301
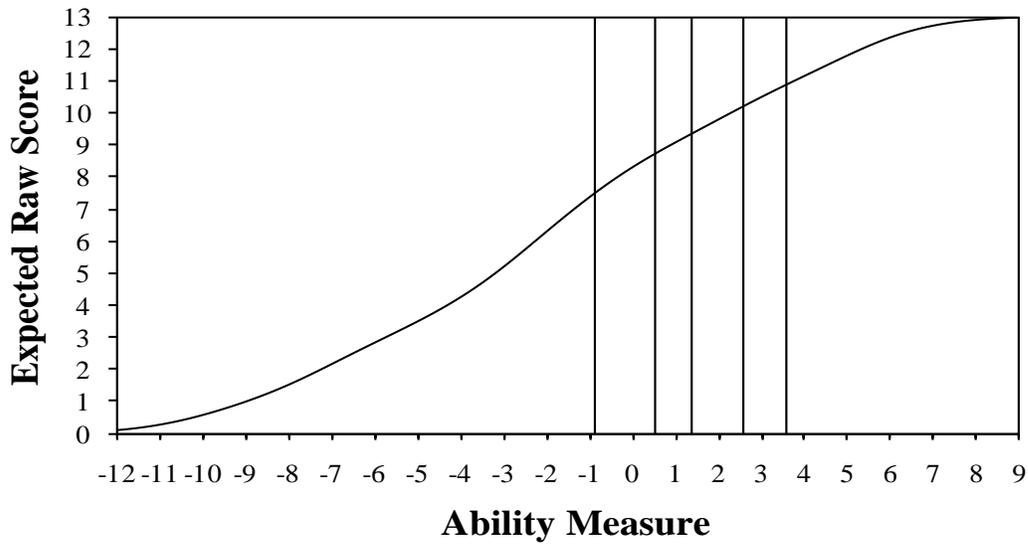
Figure 8.3.4D

Test Information Function: Spek 3-5 S301
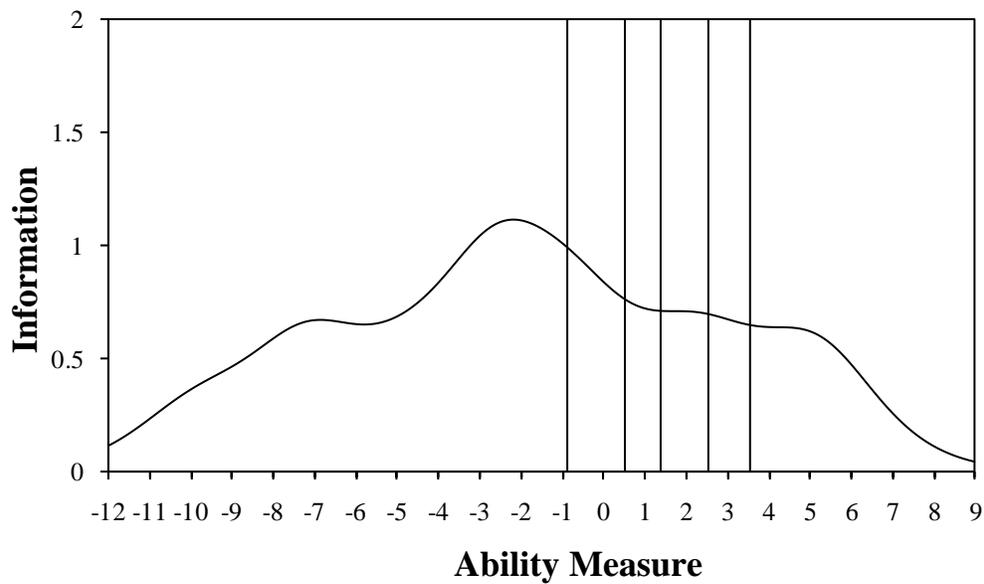
Table 8.3.4D

Reliability: Spek 3-5 S301

Tiers	No. of Students	Reliability
--	323,993	0.870

Table 8.3.4E-1

Accuracy and Consistency of Classification Indices: Spek (Grade 3) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.587	0.495		0.371	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.595		0.402	
	2	0.544		0.430	
	3	0.583		0.496	
	4	0.354		0.270	
	5	0.364		0.260	
	6	0.914		0.836	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.967	0.019	0.013	0.945
	2/3	0.893	0.056	0.051	0.860
	3/4	0.871	0.031	0.098	0.833
	4/5	0.922	0.035	0.044	0.874
	5/6	0.897	0.079	0.024	0.864

Table 8.3.4E-2

Accuracy and Consistency of Classification Indices: Spek (Grade 4) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.582	0.493		0.362	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.695		0.516	
	2	0.447		0.344	
	3	0.570		0.480	
	4	0.353		0.272	
	5	0.342		0.242	
	6	0.904		0.825	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.964	0.019	0.017	0.942
	2/3	0.903	0.054	0.043	0.873
	3/4	0.877	0.030	0.093	0.843
	4/5	0.915	0.035	0.050	0.865
	5/6	0.878	0.093	0.030	0.836

Table 8.3.4E-3

Accuracy and Consistency of Classification Indices: Spek (Grade 5) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.581	0.495		0.347	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.689		0.516	
	2	0.409		0.314	
	3	0.545		0.455	
	4	0.349		0.269	
	5	0.309		0.217	
	6	0.906		0.831	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.965	0.020	0.016	0.943
	2/3	0.911	0.050	0.038	0.884
	3/4	0.884	0.028	0.088	0.854
	4/5	0.917	0.035	0.048	0.867
	5/6	0.855	0.113	0.033	0.805

8.3.5 Oral Language Composite 3-5

Figure 8.3.5A
Scale Scores: Oral 3-5 S301

Figure 8.3.5B
Proficiency Level: Oral 3-5 S301
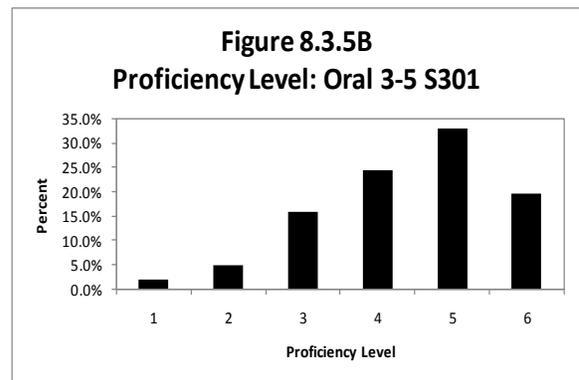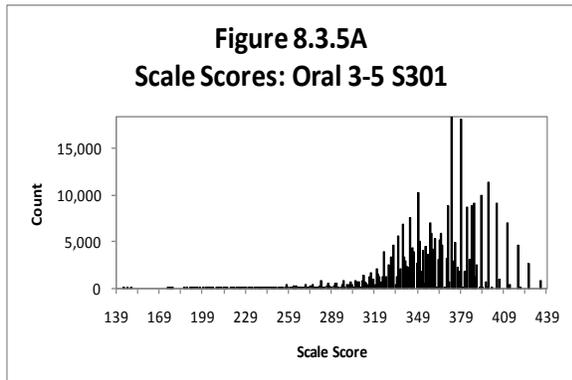
Table 8.3.5A
Scale Score Descriptive Statistics: Oral 3-5 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	142,919	144	434	354.10	30.66
4	99,112	146	434	362.03	32.63
5	81,770	149	434	368.91	33.15
Total	323,801	144	434	360.27	32.48

Table 8.3.5B
Proficiency Level Distribution: Oral 3-5 S301

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	2,242	1.6%	2,397	2.4%	2,238	2.7%	6,877	2.1%
2	6,642	4.6%	4,888	4.9%	4,580	5.6%	16,110	5.0%
3	25,232	17.7%	14,718	14.8%	11,561	14.1%	51,511	15.9%
4	36,030	25.2%	23,607	23.8%	19,274	23.6%	78,911	24.4%
5	44,223	30.9%	33,195	33.5%	29,303	35.8%	106,721	33.0%
6	28,550	20.0%	20,307	20.5%	14,814	18.1%	63,671	19.7%
Total	142,919	100.0%	99,112	100.0%	81,770	100.0%	323,801	100.0%

Table 8.3.5C

n/a

Figure 8.3.5C

n/a

Figure 8.3.5D

n/a

Table 8.3.5D

Oral Composite Reliability: Oral 3-5 S301

Component	Weight	Variance	Reliability
Listening	0.50	1244.288	0.634
Speaking	0.50	1664.578	0.870
Oral		1049.737	0.840

Table 8.3.5E-1

Accuracy and Consistency of Classification Indices: Oral (Grade 3) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.601	0.484		0.330	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.846		0.673	
	2	0.539		0.372	
	3	0.626		0.487	
	4	0.548		0.441	
	5	0.572		0.475	
	6	0.705		0.564	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.992	0.002	0.006	0.989
	2/3	0.965	0.019	0.016	0.945
	3/4	0.894	0.045	0.061	0.855
	4/5	0.857	0.054	0.088	0.804
	5/6	0.877	0.067	0.055	0.828

Table 8.3.5E-2

Accuracy and Consistency of Classification Indices: Oral (Grade 4) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.577	0.463		0.298	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.849		0.697	
	2	0.532		0.373	
	3	0.586		0.446	
	4	0.546		0.431	
	5	0.558		0.475	
	6	0.631		0.491	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.989	0.003	0.008	0.985
	2/3	0.965	0.019	0.017	0.945
	3/4	0.905	0.041	0.054	0.869
	4/5	0.858	0.050	0.092	0.804
	5/6	0.844	0.086	0.069	0.791

Table 8.3.5E-3

Accuracy and Consistency of Classification Indices: Oral (Grade 5) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.563	0.460		0.286	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.843		0.705	
	2	0.567		0.410	
	3	0.586		0.446	
	4	0.550		0.423	
	5	0.552		0.500	
	6	0.552		0.407	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.989	0.004	0.008	0.984
	2/3	0.963	0.019	0.018	0.944
	3/4	0.911	0.035	0.054	0.877
	4/5	0.857	0.046	0.096	0.802
	5/6	0.829	0.130	0.041	0.786

8.3.6 Literacy Composite 3-5

Figure 8.3.6A
Scale Scores: Litr 3-5 S301

Figure 8.3.6B
Proficiency Level: Litr 3-5 S301
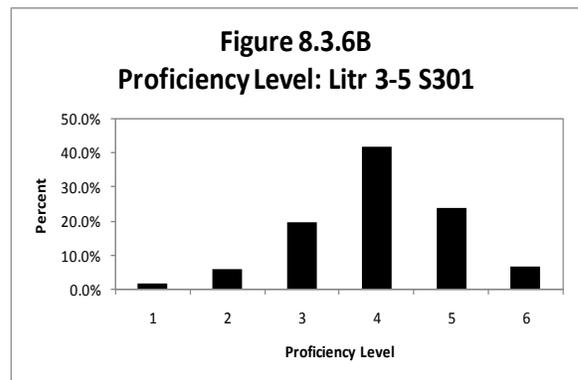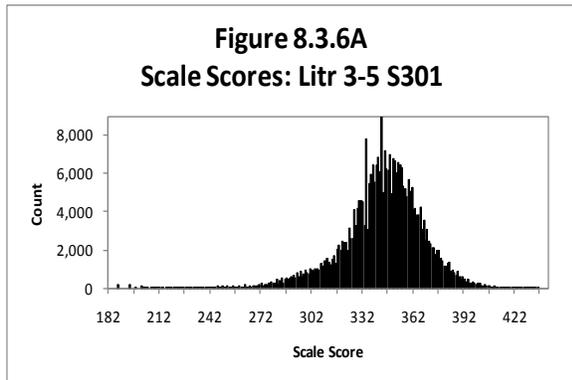
Table 8.3.6A
Scale Score Descriptive Statistics: Litr 3-5 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	142,686	187	429	336.29	22.89
4	98,945	194	434	345.52	24.11
5	81,619	201	436	353.18	24.72
Total	323,250	187	436	343.38	24.73

Table 8.3.6B
Proficiency Level Distribution: Litr 3-5 S301

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	1,585	1.1%	1,838	1.9%	2,387	2.9%	5,810	1.8%
2	7,435	5.2%	6,121	6.2%	5,700	7.0%	19,256	6.0%
3	23,905	16.8%	19,273	19.5%	20,461	25.1%	63,639	19.7%
4	58,988	41.3%	42,719	43.2%	33,316	40.8%	135,023	41.8%
5	39,780	27.9%	22,364	22.6%	15,443	18.9%	77,587	24.0%
6	10,993	7.7%	6,630	6.7%	4,312	5.3%	21,935	6.8%
Total	142,686	100.0%	98,945	100.0%	81,619	100.0%	323,250	100.0%

Table 8.3.6C

n/a

Figure 8.3.6C

n/a

Figure 8.3.6D

n/a

Table 8.3.6D

Literacy Composite Reliability: Litr 3-5 S301

Component	Weight	Variance	Reliability
Reading	0.50	794.667	0.782
Writing	0.50	675.502	0.912
Literacy		611.086	0.905

*Variances from students who had results in all four domains

Table 8.3.6E-1

Accuracy and Consistency of Classification Indices: Litr (Grade 3) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.704	0.614		0.464	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.836		0.713	
	2	0.766		0.642	
	3	0.705		0.582	
	4	0.807		0.720	
	5	0.601		0.549	
	6	-		0.414	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.996	0.002	0.003	0.994
	2/3	0.979	0.009	0.012	0.970
	3/4	0.926	0.040	0.034	0.893
	4/5	0.880	0.035	0.085	0.835
	5/6	0.923	0.077	0.000	0.917

Table 8.3.6E-2
Accuracy and Consistency of Classification Indices: Litr (Grade 4) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.684	0.589		0.434	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.857		0.755	
	2	0.750		0.627	
	3	0.723		0.602	
	4	0.795		0.695	
	5	0.532		0.472	
	6	-		0.312	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.994	0.002	0.004	0.991
	2/3	0.975	0.011	0.013	0.964
	3/4	0.918	0.042	0.040	0.883
	4/5	0.863	0.035	0.102	0.813
	5/6	0.933	0.067	0.000	0.925

Table 8.3.6E-3
Accuracy and Consistency of Classification Indices: Litr (Grade 5) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.677	0.578		0.426	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.881		0.795	
	2	0.702		0.573	
	3	0.756		0.645	
	4	0.750		0.642	
	5	0.502		0.435	
	6	-		0.269	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.991	0.003	0.006	0.987
	2/3	0.969	0.016	0.015	0.954
	3/4	0.904	0.044	0.051	0.866
	4/5	0.863	0.038	0.098	0.811
	5/6	0.947	0.053	0.000	0.943

8.3.7 Comprehension Composite 3-5

Figure 8.3.7A
Scale Scores: Cphn 3-5 S301

Figure 8.3.7B
Proficiency Level: Cphn 3-5 S301
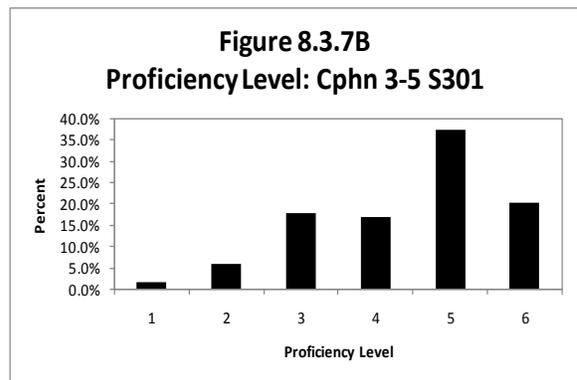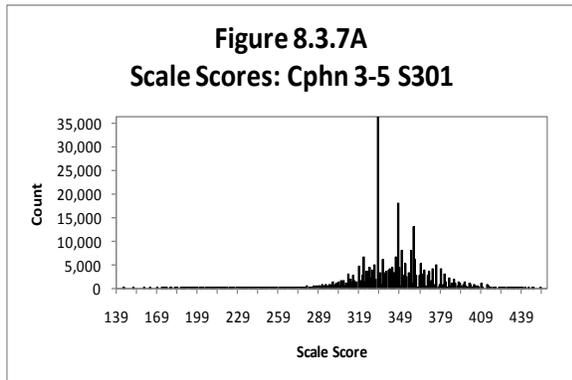
Table 8.3.7A
Scale Score Descriptive Statistics: Cphn 3-5 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	142,762	144	453	335.56	25.23
4	98,993	151	453	347.07	27.70
5	81,673	159	453	356.28	28.82
Total	323,428	144	453	344.32	28.25

Table 8.3.7B
Proficiency Level Distribution: Cphn 3-5 S301

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	1,368	1.0%	1,672	1.7%	2,143	2.6%	5,183	1.6%
2	6,892	4.8%	6,241	6.3%	6,232	7.6%	19,365	6.0%
3	20,507	14.4%	19,523	19.7%	17,709	21.7%	57,739	17.9%
4	22,942	16.1%	17,690	17.9%	14,080	17.2%	54,712	16.9%
5	59,512	41.7%	33,853	34.2%	27,782	34.0%	121,147	37.5%
6	31,541	22.1%	20,014	20.2%	13,727	16.8%	65,282	20.2%
Total	142,762	100.0%	98,993	100.0%	81,673	100.0%	323,428	100.0%

Table 8.3.7C

n/a

Figure 8.3.7C

n/a

Figure 8.3.7D

n/a

Table 8.3.7D

Comprehension Composite Reliability: Cphn 3-5 S301

Component	Weight	Variance	Reliability
Listening	0.30	1244.288	0.634
Reading	0.70	794.667	0.782
Comprehension		796.914	0.842

*Variances from students who had results in all four domains

Table 8.3.7E-1

Accuracy and Consistency of Classification Indices: Cphn (Grade 3) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.631	0.519		0.350	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.795		0.583	
	2	0.646		0.466	
	3	0.569		0.422	
	4	0.398		0.301	
	5	0.688		0.604	
	6	0.718		0.593	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.994	0.001	0.004	0.992
	2/3	0.971	0.012	0.018	0.955
	3/4	0.905	0.050	0.044	0.861
	4/5	0.856	0.070	0.073	0.804
	5/6	0.879	0.057	0.064	0.827

Table 8.3.7E-2
Accuracy and Consistency of Classification Indices: Cphn (Grade 4) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.591	0.479		0.324	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.797		0.613	
	2	0.616		0.443	
	3	0.607		0.474	
	4	0.400		0.309	
	5	0.599		0.507	
	6	0.679		0.547	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.991	0.003	0.006	0.987
	2/3	0.960	0.019	0.021	0.939
	3/4	0.882	0.058	0.059	0.836
	4/5	0.847	0.063	0.091	0.794
	5/6	0.874	0.058	0.068	0.820

Table 8.3.7E-3
Accuracy and Consistency of Classification Indices: Cphn (Grade 5) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.583	0.470		0.317	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.823		0.655	
	2	0.604		0.441	
	3	0.595		0.469	
	4	0.372		0.288	
	5	0.603		0.513	
	6	0.660		0.509	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.987	0.004	0.010	0.981
	2/3	0.953	0.022	0.025	0.928
	3/4	0.870	0.068	0.062	0.821
	4/5	0.841	0.067	0.092	0.788
	5/6	0.886	0.057	0.058	0.835

8.3.8 Overall Composite 3-5

Figure 8.3.8A
Scale Scores: Over 3-5 S301

Figure 8.3.8B
Proficiency Level: Over 3-5 S301
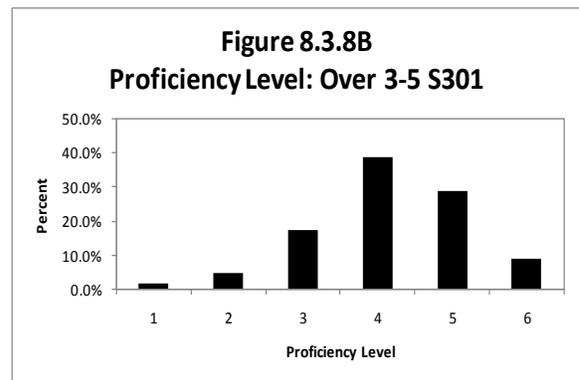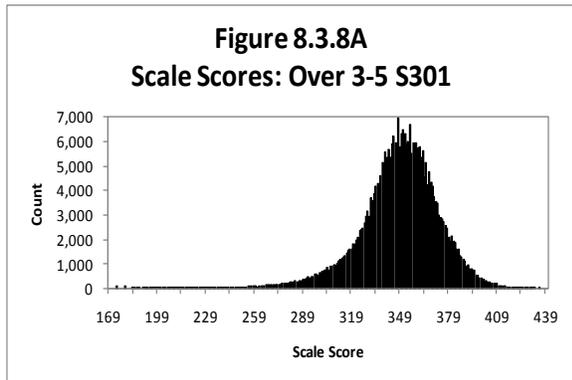
Table 8.3.8A

Scale Score Descriptive Statistics: Over 3-5 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	142,581	174	427	341.42	23.25
4	98,843	179	429	350.30	24.71
5	81,547	185	435	357.71	25.37
Total	322,971	174	435	348.25	25.15

Table 8.3.8B

Proficiency Level Distribution: Over 3-5 S301

Level	Grade 3 Count	Grade 3 Percent	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	1,501	1.1%	1,725	1.7%	2,005	2.5%	5,231	1.6%
2	6,007	4.2%	5,113	5.2%	4,675	5.7%	15,795	4.9%
3	22,990	16.1%	17,572	17.8%	15,928	19.5%	56,490	17.5%
4	52,950	37.1%	38,179	38.6%	33,434	41.0%	124,563	38.6%
5	44,634	31.3%	27,730	28.1%	19,949	24.5%	92,313	28.6%
6	14,499	10.2%	8,524	8.6%	5,556	6.8%	28,579	8.8%
Total	142,581	100.0%	98,843	100.0%	81,547	100.0%	322,971	100.0%

Table 8.3.8C
n/a

Figure 8.3.8C
n/a

Figure 8.3.8D
n/a

Table 8.3.8D
Overall Composite Reliability: Over 3-5 S301

Component	Weight	Variance	Reliability
Listening	0.15	1244.288	0.634
Reading	0.35	794.667	0.782
Speaking	0.15	1664.578	0.870
Writing	0.35	675.502	0.912
Overall Composite		632.541	0.931

Table 8.3.8E-1
Accuracy and Consistency of Classification Indices: Over (Grade 3) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.747	0.661		0.533	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.894		0.810	
	2	0.761		0.648	
	3	0.753		0.643	
	4	0.821		0.742	
	5	0.677		0.618	
	6	0.748		0.562	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.997	0.001	0.003	0.995
	2/3	0.984	0.008	0.008	0.976
	3/4	0.939	0.033	0.028	0.912
	4/5	0.905	0.032	0.064	0.868
	5/6	0.923	0.065	0.012	0.907

Table 8.3.8E-2
Accuracy and Consistency of Classification Indices: Over (Grade 4) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.717	0.636		0.501	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.889		0.812	
	2	0.755		0.643	
	3	0.754		0.645	
	4	0.820		0.734	
	5	0.607		0.562	
	6	-		0.415	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.995	0.002	0.003	0.993
	2/3	0.981	0.010	0.010	0.972
	3/4	0.934	0.035	0.031	0.905
	4/5	0.894	0.029	0.077	0.855
	5/6	0.914	0.086	0.000	0.908

Table 8.3.8E-3
Accuracy and Consistency of Classification Indices: Over (Grade 5) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.716	0.631		0.494	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.899		0.833	
	2	0.739		0.625	
	3	0.758		0.649	
	4	0.818		0.728	
	5	0.582		0.529	
	6	-		0.351	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.993	0.002	0.004	0.991
	2/3	0.978	0.012	0.010	0.968
	3/4	0.928	0.037	0.035	0.897
	4/5	0.884	0.029	0.087	0.841
	5/6	0.932	0.068	0.000	0.928

8.4　Grades: 6–8

8.4.1　Listening 6-8

Figure 8.4.1A
Scale Scores: List 6-8 S301

Figure 8.4.1B
Proficiency Level: List 6-8 S301

Table 8.4.1A

Scale Score Descriptive Statistics: List 6-8 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	69,808	124	503	373.22	38.98
7	66,637	128	503	379.67	40.73
8	61,203	132	503	385.21	42.82
Total	197,648	124	503	379.11	41.08

Table 8.4.1B

Proficiency Level Distribution: List 6-8 S301

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	1,257	1.8%	1,771	2.7%	2,467	4.0%	5,495	2.8%
2	5,751	8.2%	6,174	9.3%	6,177	10.1%	18,102	9.2%
3	9,903	14.2%	10,949	16.4%	11,434	18.7%	32,286	16.3%
4	18,492	26.5%	18,069	27.1%	17,691	28.9%	54,252	27.4%
5	26,034	37.3%	18,564	27.9%	14,704	24.0%	59,302	30.0%
6	8,371	12.0%	11,110	16.7%	8,730	14.3%	28,211	14.3%
Total	69,808	100.0%	66,637	100.0%	61,203	100.0%	197,648	100.0%

Table 8.4.1C

Conditional Standard Error of Measurement at Cut Scores: List 6-8 S301

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	6	283	21.04	20.66	27.43
	7	293	20.66	19.91	25.92
	8	302	20.66	19.54	24.42
2/3	6	328	21.04	19.16	21.42
	7	337	21.42	19.16	20.66
	8	345	21.79	19.54	20.29
3/4	6	359	23.29	19.91	19.54
	7	368	24.42	20.29	19.16
	8	375	25.55	21.04	19.16
4/5	6	380	n/a	21.42	19.16
	7	390	n/a	22.17	19.16
	8	399	n/a	23.29	19.16
5/6	6	409	n/a	n/a	19.54
	7	418	n/a	n/a	20.29
	8	426	n/a	n/a	20.66

Figure 8.4.1C
Test Characteristic Curve: List 6-8ABC S301
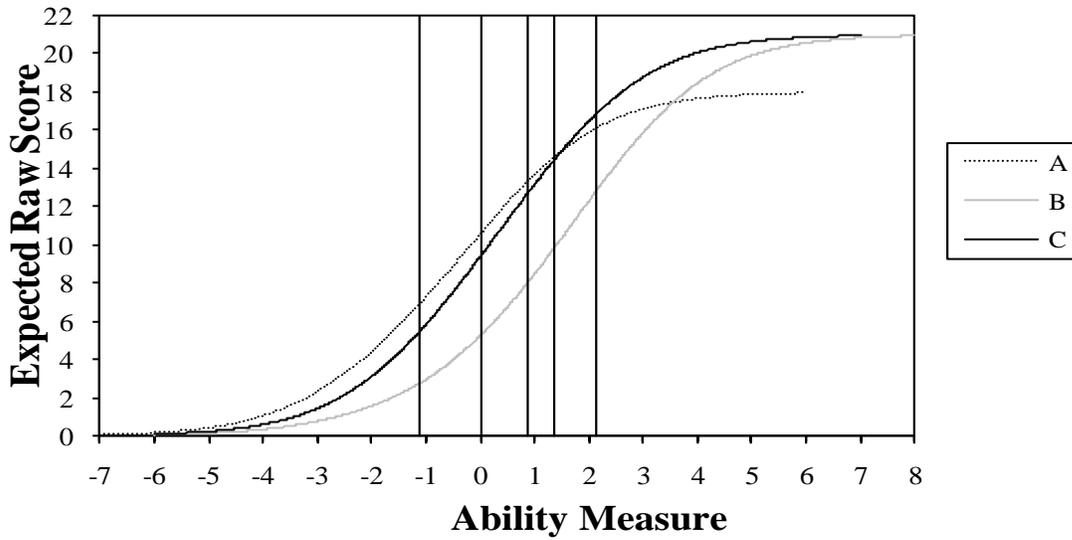
Figure 8.4.1D
Test Information Function: List 6-8ABC S301
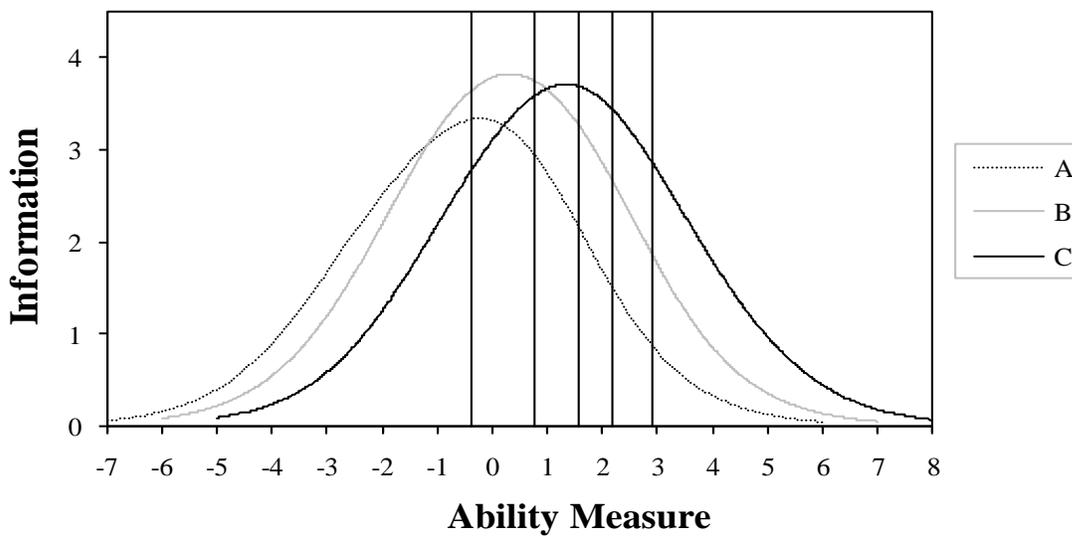
Table 8.4.1D

Weighted Reliability: List 6-8 S301

Tiers	No. of Students	Reliability	Reliability
A	20,532	0.744	
B	77,378	0.671	0.606
C	99,738	0.528	

Table 8.4.1E-1

Accuracy and Consistency of Classification Indices: List (Grade 6) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.454	0.325		0.111	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.670		0.453	
	2	0.636		0.411	
	3	0.366		0.215	
	4	0.359		0.290	
	5	0.475		0.442	
	6	-		0.167	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.988	0.006	0.007	0.980
	2/3	0.943	0.015	0.042	0.916
	3/4	0.853	0.042	0.104	0.767
	4/5	0.694	0.080	0.226	0.612
	5/6	0.880	0.120	0.000	0.779

Table 8.4.1E-2
Accuracy and Consistency of Classification Indices: List (Grade 7) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.385	0.301		0.108	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.704		0.491	
	2	0.589		0.374	
	3	0.377		0.229	
	4	0.355		0.298	
	5	0.363		0.335	
	6	-		0.233	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.983	0.007	0.010	0.973
	2/3	0.932	0.018	0.050	0.900
	3/4	0.832	0.044	0.125	0.742
	4/5	0.684	0.085	0.230	0.609
	5/6	0.833	0.167	0.000	0.757

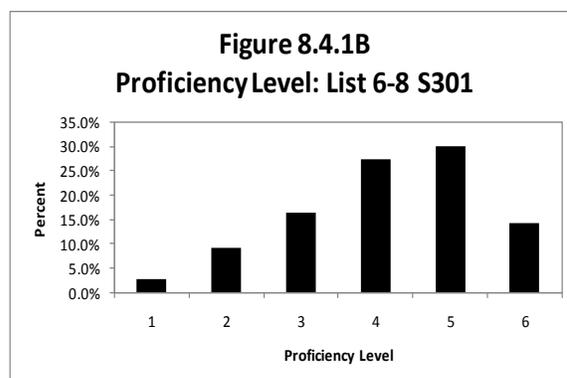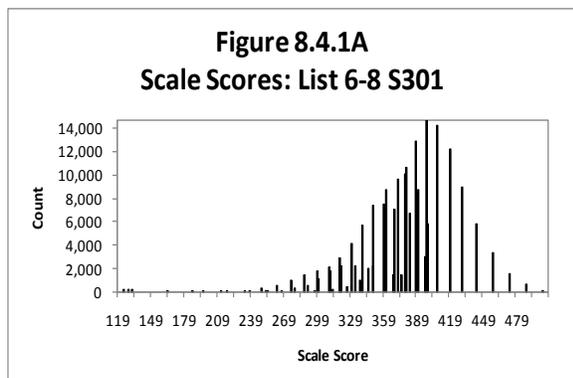**Table 8.4.1E-3**
Accuracy and Consistency of Classification Indices: List (Grade 8) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.376	0.296		0.113	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.745		0.529	
	2	0.512		0.322	
	3	0.380		0.249	
	4	0.370		0.321	
	5	0.334		0.300	
	6	-		0.212	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.976	0.008	0.016	0.962
	2/3	0.918	0.024	0.058	0.878
	3/4	0.808	0.051	0.141	0.722
	4/5	0.684	0.094	0.222	0.620
	5/6	0.857	0.143	0.000	0.776

8.4.2 Reading 6-8

Figure 8.4.2A
Scale Scores: Read 6-8 S301

Figure 8.4.2B
Proficiency Level: Read 6-8 S301
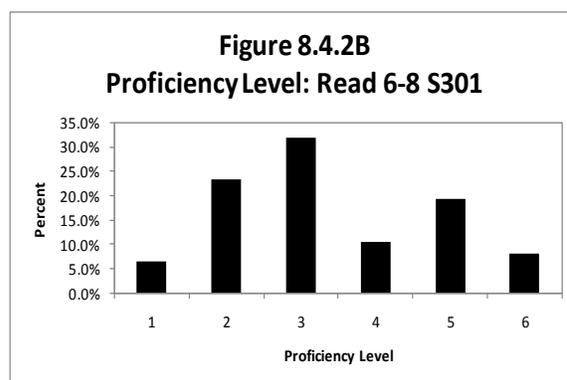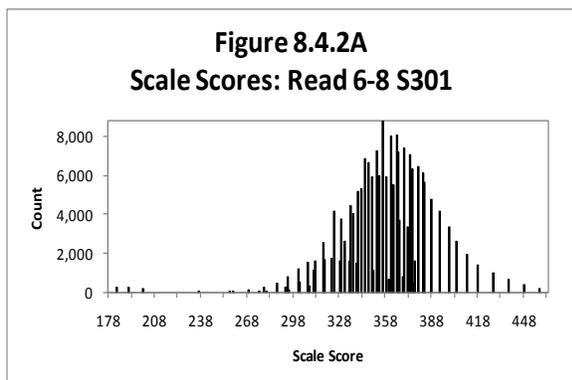
Table 8.4.2A

Scale Score Descriptive Statistics: Read 6-8 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	69,668	183	458	350.93	24.42
7	66,504	191	458	358.27	26.48
8	61,091	200	458	365.21	28.49
Total	197,263	183	458	357.82	27.06

Table 8.4.2B

Proficiency Level Distribution: Read 6-8 S301

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	3,381	4.9%	4,356	6.5%	5,223	8.5%	12,960	6.6%
2	14,141	20.3%	15,824	23.8%	16,222	26.6%	46,187	23.4%
3	25,197	36.2%	21,708	32.6%	16,312	26.7%	63,217	32.0%
4	8,739	12.5%	7,402	11.1%	4,556	7.5%	20,697	10.5%
5	13,483	19.4%	12,214	18.4%	12,612	20.6%	38,309	19.4%
6	4,727	6.8%	5,000	7.5%	6,166	10.1%	15,893	8.1%
Total	69,668	100.0%	66,504	100.0%	61,091	100.0%	197,263	100.0%

Table 8.4.2C

Conditional Standard Error of Measurement at Cut Scores: Read 6-8 S301

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	6	312	11.96	13.78	15.60
	7	321	11.70	12.74	14.04
	8	329	11.70	11.96	12.74
2/3	6	340	11.70	11.18	11.70
	7	349	12.22	10.92	11.18
	8	358	12.74	10.92	10.92
3/4	6	360	13.00	10.92	10.92
	7	369	13.78	10.92	10.66
	8	376	14.82	11.44	10.92
4/5	6	366	n/a	10.92	10.66
	7	375	n/a	11.18	10.92
	8	382	n/a	11.70	10.92
5/6	6	382	n/a	n/a	10.92
	7	391	n/a	n/a	11.44
	8	398	n/a	n/a	11.96

Figure 8.4.2C
Test Characteristic Curve: Read 6-8ABC S301
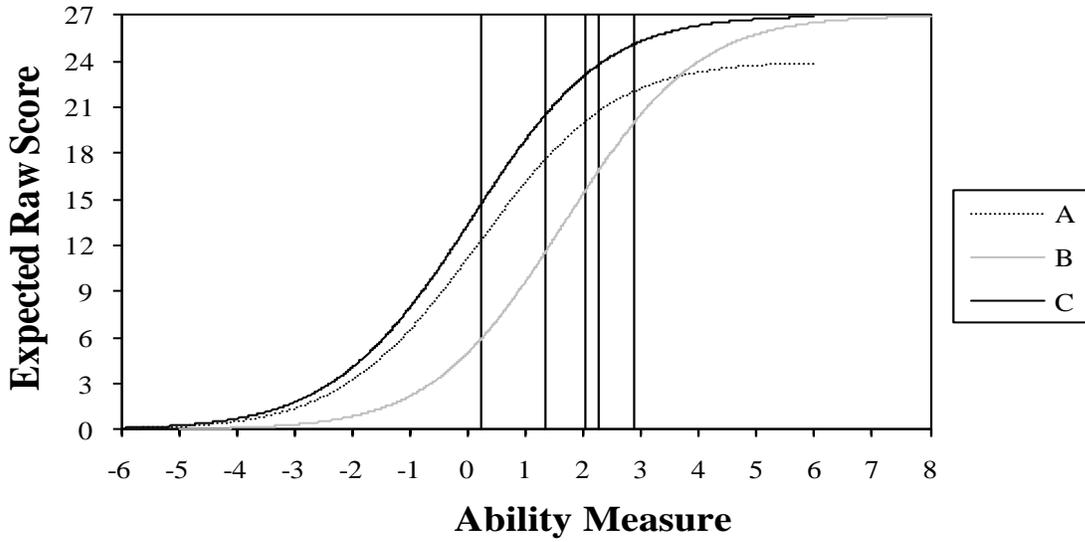
Figure 8.4.2D
Test Information Function: Read 6-8ABC S301
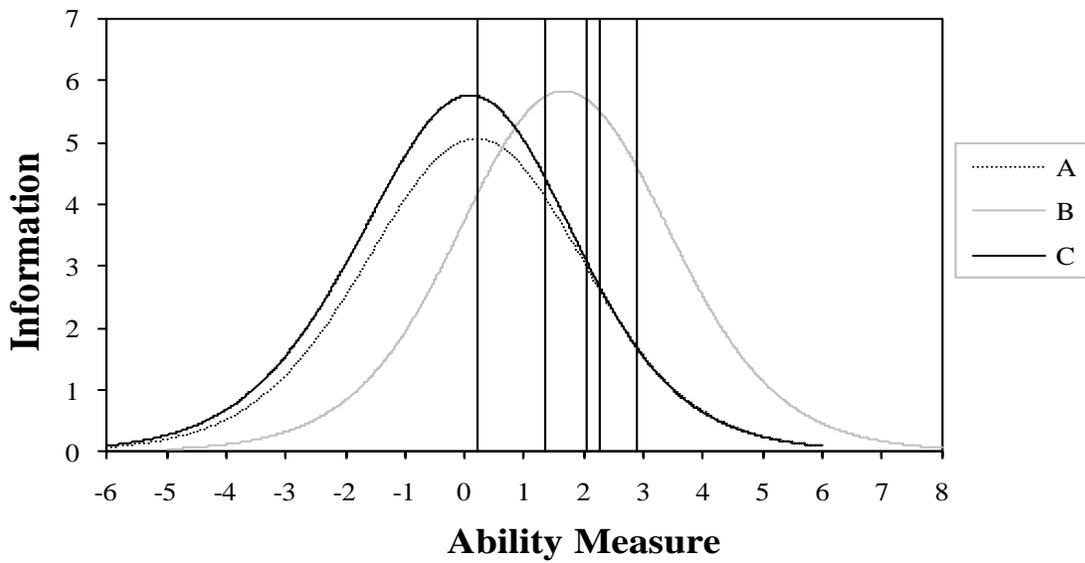
Table 8.4.2D

Weighted Reliability: Read 6-8 S301

Tiers	No. of Students	Reliability	Reliability
A	20,376	0.773	
B	77,223	0.773	0.765
C	99,664	0.757	

Table 8.4.2E-1

Accuracy and Consistency of Classification Indices: Read (Grade 6) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.464	0.387		0.203	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.735		0.567	
	2	0.706		0.535	
	3	0.556		0.443	
	4	0.192		0.168	
	5	0.373		0.310	
	6	-		0.142	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.974	0.013	0.013	0.960
	2/3	0.889	0.035	0.076	0.843
	3/4	0.763	0.075	0.162	0.687
	4/5	0.762	0.119	0.119	0.704
	5/6	0.932	0.068	0.000	0.902

Table 8.4.2E-2
Accuracy and Consistency of Classification Indices: Read (Grade 7) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.481	0.387		0.218	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.725		0.549	
	2	0.680		0.526	
	3	0.521		0.418	
	4	0.188		0.156	
	5	0.390		0.318	
	6	-		0.190	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.963	0.017	0.019	0.942
	2/3	0.868	0.045	0.087	0.817
	3/4	0.793	0.079	0.128	0.719
	4/5	0.797	0.095	0.107	0.732
	5/6	0.925	0.075	0.000	0.888

Table 8.4.2E-3
Accuracy and Consistency of Classification Indices: Read (Grade 8) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.484	0.389		0.235	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.727		0.548	
	2	0.653		0.519	
	3	0.465		0.368	
	4	0.139		0.108	
	5	0.432		0.359	
	6	0.484		0.297	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.950	0.022	0.028	0.924
	2/3	0.853	0.056	0.091	0.798
	3/4	0.822	0.076	0.102	0.752
	4/5	0.825	0.093	0.082	0.757
	5/6	0.899	0.100	0.001	0.860

8.4.3 Writing 6-8

Figure 8.4.3A
Scale Scores: Writ 6-8 S301

Figure 8.4.3B
Proficiency Level: Writ 6-8 S301
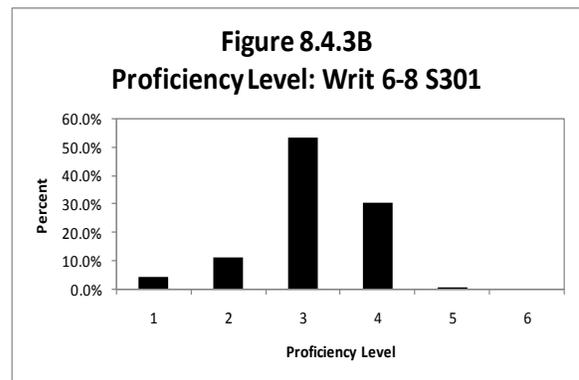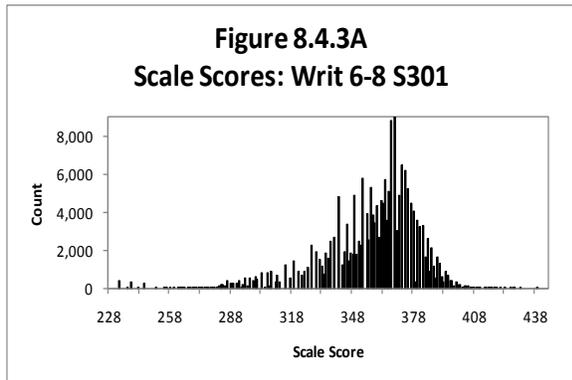
Table 8.4.3A
Scale Score Descriptive Statistics: Writ 6-8 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	69,645	233	428	352.65	23.80
7	66,472	239	431	358.06	23.64
8	61,057	245	439	362.53	23.47
Total	197,174	233	439	357.53	23.98

Table 8.4.3B
Proficiency Level Distribution: Writ 6-8 S301

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	2,480	3.6%	2,909	4.4%	3,295	5.4%	8,684	4.4%
2	6,571	9.4%	7,138	10.7%	8,072	13.2%	21,781	11.0%
3	30,620	44.0%	36,283	54.6%	38,233	62.6%	105,136	53.3%
4	29,069	41.7%	19,719	29.7%	11,286	18.5%	60,074	30.5%
5	901	1.3%	416	0.6%	170	0.3%	1,487	0.8%
6	4	0.0%	7	0.0%	1	0.0%	12	0.0%
Total	69,645	100.0%	66,472	100.0%	61,057	100.0%	197,174	100.0%

Table 8.4.3C
Conditional Standard Error of Measurement at Cut Scores: Writ 6-8 S301

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	6	298	9.02	7.77	8.09
	7	308	9.64	8.40	8.40
	8	318	11.19	8.40	8.40
2/3	6	329	11.82	8.40	8.09
	7	339	12.13	8.09	8.09
	8	348	11.82	8.09	7.77
3/4	6	361	11.51	7.77	7.46
	7	371	11.19	7.15	7.15
	8	381	10.57	6.84	6.84
4/5	6	391	n/a	6.53	6.53
	7	399	n/a	6.22	6.22
	8	408	n/a	6.53	6.84
5/6	6	412	n/a	n/a	7.15
	7	420	n/a	n/a	8.40
	8	428	n/a	n/a	10.57

Figure 8.4.3C
Test Characteristic Curve: Writ 6-8ABC S301
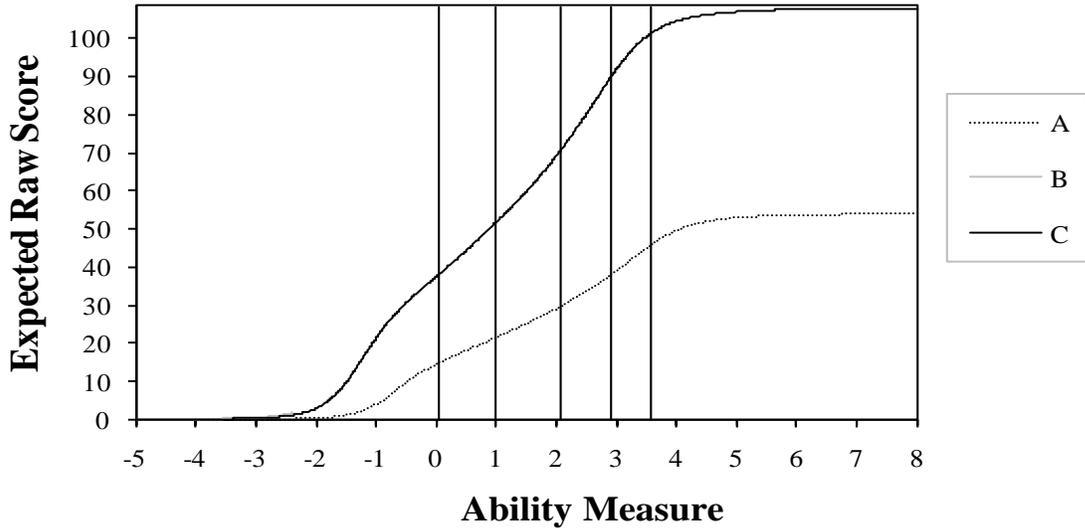
Figure 8.4.3D
Test Information Function: Writ 6-8ABC S301
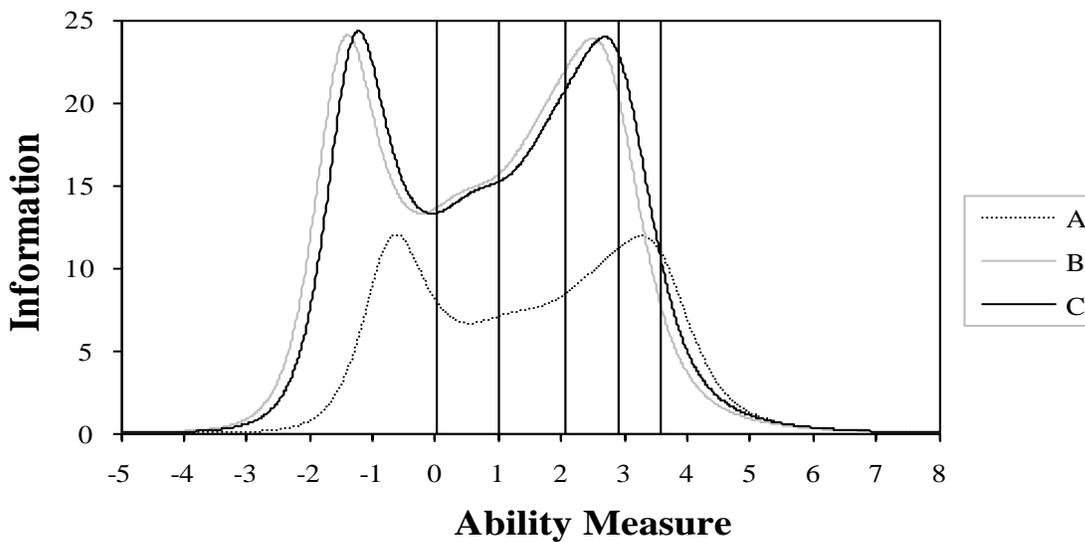
Table 8.4.3D

Weighted Reliability: Writ 6-8 S301

Tiers	No. of Students	Reliability	Reliability
A	20,358	0.890	
B	77,199	0.927	0.911
C	99,617	0.904	

Table 8.4.3E-1

Accuracy and Consistency of Classification Indices: Writ (Grade 6) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.798	0.722		0.551	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.882		0.806	
	2	0.768		0.659	
	3	0.847		0.733	
	4	0.761		0.720	
	5	-		0.034	
	6	-		1.000	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.990	0.004	0.006	0.986
	2/3	0.967	0.017	0.017	0.952
	3/4	0.854	0.041	0.105	0.796
	4/5	0.987	0.013	0.000	0.987
	5/6	1.000	0.000	0.000	1.000

Table 8.4.3E-2
Accuracy and Consistency of Classification Indices: Writ (Grade 7) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.737	0.660		0.442	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.885		0.814	
	2	0.780		0.674	
	3	0.832		0.723	
	4	0.602		0.540	
	5	-		0.000	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.989	0.005	0.006	0.984
	2/3	0.964	0.017	0.019	0.948
	3/4	0.791	0.058	0.151	0.732
	4/5	0.994	0.006	0.000	0.994
	5/6	1.000	0.000	0.000	1.000

Table 8.3.3E-3
Accuracy and Consistency of Classification Indices: Writ (Grade 8) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.757	0.686		0.428	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.883		0.810	
	2	0.801		0.701	
	3	0.742		0.762	
	4	-		0.363	
	5	-		-	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.986	0.006	0.007	0.980
	2/3	0.958	0.018	0.024	0.941
	3/4	0.812	0.188	0.000	0.766
	4/5	0.997	0.003	0.000	0.997
	5/6	1.000	0.000	0.000	1.000

8.4.4 Speaking 6-8

Figure 8.4.4A
Scale Scores: Spek 6-8 S301

Figure 8.4.4B
Proficiency Level: Spek 6-8 S301
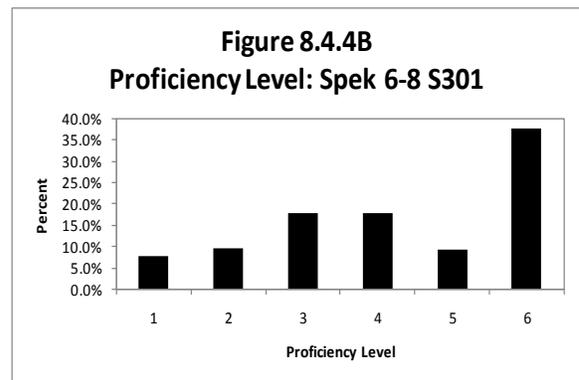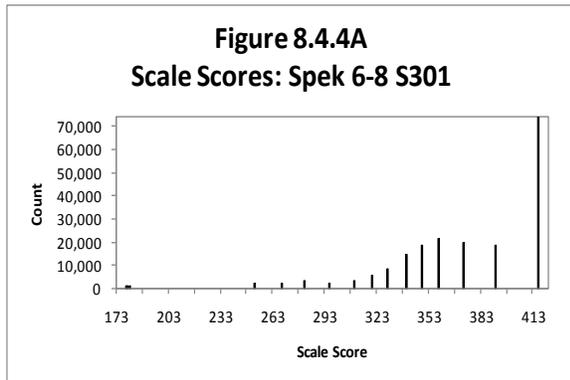
Table 8.4.4A
Scale Score Descriptive Statistics: Spek 6-8 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	69,761	178	416	369.64	45.54
7	66,563	179	416	372.53	46.55
8	61,139	180	416	375.05	46.90
Total	197,463	178	416	372.29	46.36

Table 8.4.4B
Proficiency Level Distribution: Spek 6-8 S301

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	4,562	6.5%	5,579	8.4%	5,058	8.3%	15,199	7.7%
2	6,585	9.4%	4,577	6.9%	7,928	13.0%	19,090	9.7%
3	13,024	18.7%	10,969	16.5%	11,003	18.0%	34,996	17.7%
4	15,576	22.3%	13,861	20.8%	5,894	9.6%	35,331	17.9%
5	6,705	9.6%	6,256	9.4%	5,660	9.3%	18,621	9.4%
6	23,309	33.4%	25,321	38.0%	25,596	41.9%	74,226	37.6%
Total	69,761	100.0%	66,563	100.0%	61,139	100.0%	197,463	100.0%

Table 8.4.4C
Conditional Standard Error of Measurement at
Cut Scores: Spek 6-8 S301

Proficiency Level	Grade	Cut Score	SEM
1/2	6	310	22.09
	7	314	22.29
	8	317	22.69
2/3	6	337	23.50
	7	340	23.50
	8	344	23.70
3/4	6	353	23.50
	7	358	23.30
	8	361	23.30
4/5	6	377	22.69
	7	380	22.29
	8	384	22.09
5/6	6	397	21.49
	7	400	21.49
	8	404	21.49

Figure 8.4.4C
Test Characteristic Curve: Spek 6-8 S301
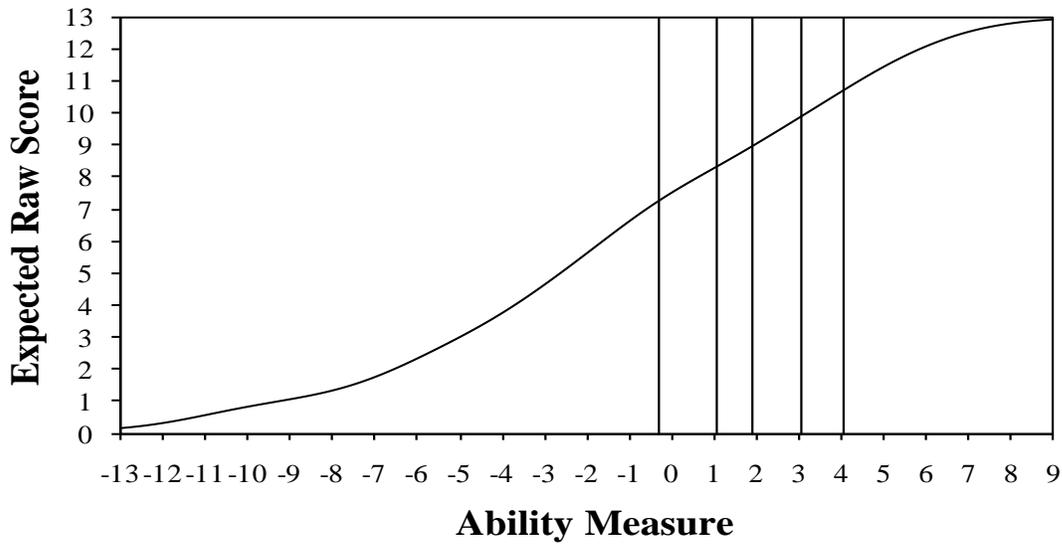
Figure 8.4.4D
Test Information Function: Spek 6-8 S301
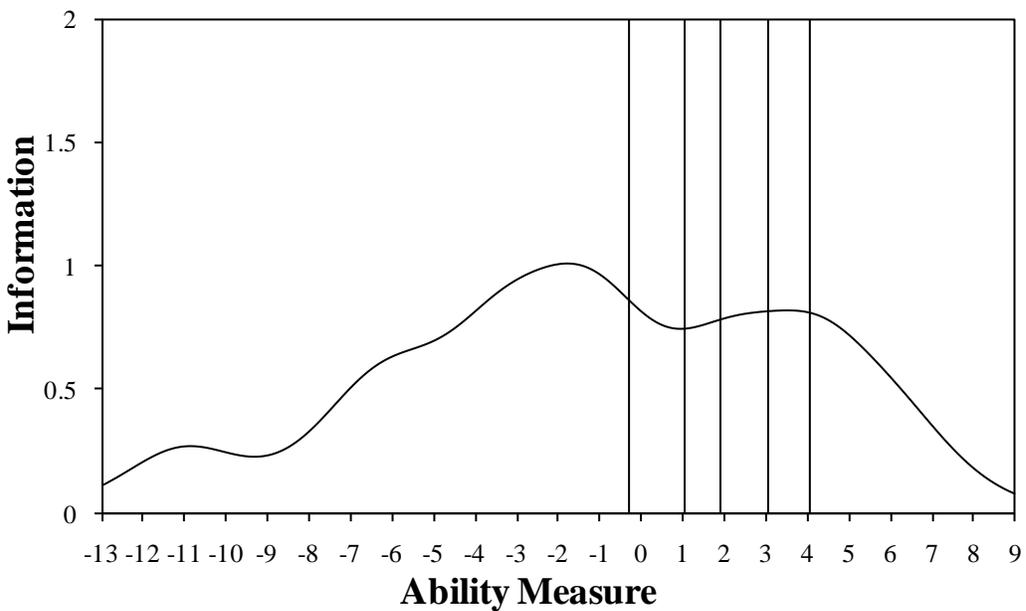
Table 8.4.4D

Reliability: Spek 6-8 S301

Tiers	No. of Students	Reliability
--	197,463	0.895

Table 8.4.4E-1

Accuracy and Consistency of Classification Indices: Spek (Grade 6) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.587	0.497		0.370	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.700		0.529	
	2	0.388		0.288	
	3	0.498		0.413	
	4	0.512		0.425	
	5	0.279		0.192	
	6	0.899		0.814	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.964	0.022	0.014	0.940
	2/3	0.902	0.062	0.036	0.869
	3/4	0.864	0.048	0.089	0.829
	4/5	0.896	0.026	0.077	0.847
	5/6	0.909	0.060	0.031	0.863

Table 8.4.4E-2

Accuracy and Consistency of Classification Indices: Spek (Grade 7) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.599	0.510		0.375	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.773		0.619	
	2	0.319		0.230	
	3	0.484		0.395	
	4	0.511		0.420	
	5	0.250		0.173	
	6	0.902		0.818	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.963	0.019	0.018	0.940
	2/3	0.917	0.054	0.029	0.885
	3/4	0.874	0.050	0.076	0.842
	4/5	0.893	0.027	0.080	0.845
	5/6	0.896	0.071	0.034	0.841

Table 8.4.4E-3

Accuracy and Consistency of Classification Indices: Spek (Grade 8) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.603	0.517		0.381	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.675		0.535	
	2	0.474		0.383	
	3	0.517		0.439	
	4	0.301		0.227	
	5	0.285		0.186	
	6	0.921		0.861	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.951	0.032	0.017	0.925
	2/3	0.894	0.054	0.052	0.866
	3/4	0.883	0.024	0.093	0.851
	4/5	0.933	0.026	0.042	0.888
	5/6	0.886	0.085	0.029	0.838

8.4.5 Oral Language Composite 6-8

Figure 8.4.5A
Scale Scores: Oral 6-8 S301

Figure 8.4.5B
Proficiency Level: Oral 6-8 S301
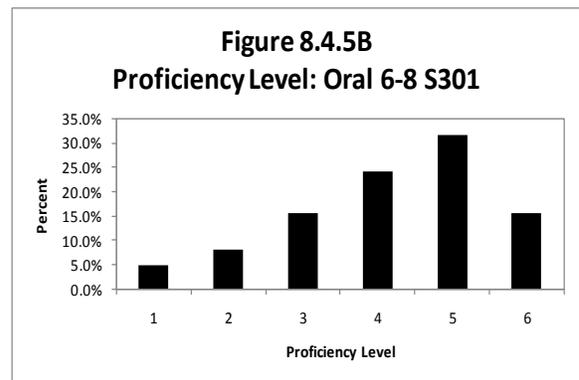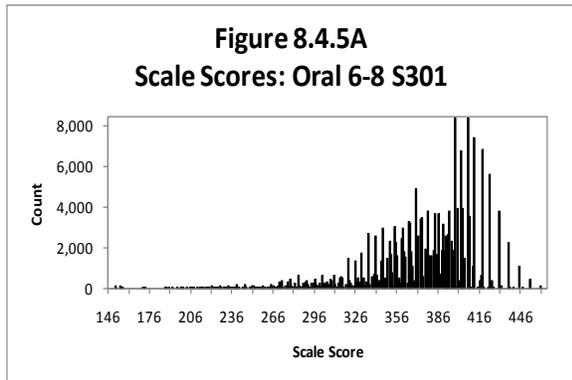
Table 8.4.5A

Scale Score Descriptive Statistics: Oral 6-8 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	69,689	151	460	371.69	36.67
7	66,492	154	460	376.35	38.10
8	61,055	156	460	380.43	39.42
Total	197,236	151	460	375.97	38.18

Table 8.4.5B

Proficiency Level Distribution: Oral 6-8 S301

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,958	4.2%	3,422	5.1%	3,574	5.9%	9,954	5.0%
2	5,136	7.4%	5,194	7.8%	5,577	9.1%	15,907	8.1%
3	10,896	15.6%	10,623	16.0%	9,245	15.1%	30,764	15.6%
4	18,028	25.9%	15,427	23.2%	14,032	23.0%	47,487	24.1%
5	21,795	31.3%	21,609	32.5%	19,202	31.5%	62,606	31.7%
6	10,876	15.6%	10,217	15.4%	9,425	15.4%	30,518	15.5%
Total	69,689	100.0%	66,492	100.0%	61,055	100.0%	197,236	100.0%

Table 8.4.5C
n/a

Figure 8.4.5C
n/a

Figure 8.4.5D
n/a

Table 8.4.5D
Oral Composite Reliability: Oral 6-8 S301

Component	Weight	Variance	Reliability
Listening	0.50	1677.033	0.606
Speaking	0.50	2135.697	0.895
Oral		1449.276	0.847

*Variances from students who had results in all four domains

Table 8.4.5E-1
Accuracy and Consistency of Classification Indices: Oral (Grade 6) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.544	0.441		0.282	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.842		0.710	
	2	0.556		0.406	
	3	0.560		0.427	
	4	0.546		0.424	
	5	0.515		0.465	
	6	0.531		0.373	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.983	0.006	0.011	0.975
	2/3	0.951	0.025	0.024	0.928
	3/4	0.898	0.039	0.063	0.861
	4/5	0.847	0.050	0.104	0.787
	5/6	0.846	0.135	0.018	0.804

Table 8.4.5E-2

Accuracy and Consistency of Classification Indices: Oral (Grade 7) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.537	0.433		0.275	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.841		0.712	
	2	0.543		0.398	
	3	0.557		0.424	
	4	0.496		0.375	
	5	0.522		0.479	
	6	-		0.353	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.980	0.007	0.013	0.970
	2/3	0.948	0.026	0.026	0.923
	3/4	0.897	0.039	0.064	0.859
	4/5	0.845	0.054	0.101	0.783
	5/6	0.846	0.154	0.000	0.800

Table 8.4.5E-3

Accuracy and Consistency of Classification Indices: Oral (Grade 8) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.533	0.430		0.277	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.828		0.696	
	2	0.560		0.420	
	3	0.525		0.396	
	4	0.492		0.370	
	5	0.516		0.474	
	6	-		0.358	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.976	0.009	0.015	0.965
	2/3	0.942	0.027	0.031	0.916
	3/4	0.899	0.037	0.064	0.860
	4/5	0.848	0.060	0.093	0.783
	5/6	0.846	0.154	0.000	0.802

8.4.6 Literacy Composite 6-8

Figure 8.4.6A
Scale Scores: Litr 6-8 S301

Figure 8.4.6B
Proficiency Level: Litr 6-8 S301
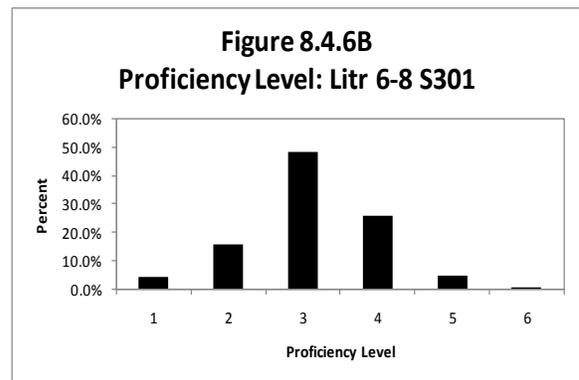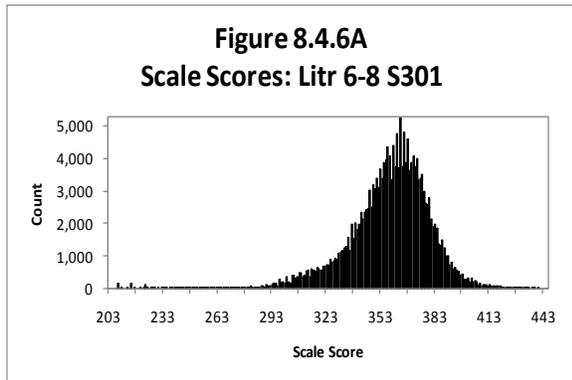
Table 8.4.6A
Scale Score Descriptive Statistics: Litr 6-8 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	69,595	208	433	352.06	21.18
7	66,419	215	435	358.42	22.22
8	61,006	223	440	364.15	23.30
Total	197,020	208	440	357.95	22.74

Table 8.4.6B
Proficiency Level Distribution: Litr 6-8 S301

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	1,785	2.6%	2,723	4.1%	3,514	5.8%	8,022	4.1%
2	9,437	13.6%	10,388	15.6%	11,421	18.7%	31,246	15.9%
3	32,894	47.3%	32,113	48.3%	30,244	49.6%	95,251	48.3%
4	21,416	30.8%	17,365	26.1%	12,501	20.5%	51,282	26.0%
5	3,554	5.1%	3,326	5.0%	2,776	4.6%	9,656	4.9%
6	509	0.7%	504	0.8%	550	0.9%	1,563	0.8%
Total	69,595	100.0%	66,419	100.0%	61,006	100.0%	197,020	100.0%

Table 8.4.6C

n/a

Figure 8.4.6C

n/a

Figure 8.4.6D

n/a

Table 8.4.6D

Literacy Composite Reliability: Litr 6-8 S301

Component	Weight	Variance	Reliability
Reading	0.50	730.965	0.765
Writing	0.50	573.734	0.911
Literacy		516.568	0.892

*Variances from students who had results in all four domains

Table 8.4.6E-1

Accuracy and Consistency of Classification Indices: Litr (Grade 6) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.733	0.650		0.471	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.646		0.689	
	2	0.790		0.673	
	3	0.829		0.726	
	4	0.626		0.573	
	5	-		0.182	
	6	-		1.000	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.984	0.011	0.005	0.985
	2/3	0.944	0.029	0.027	0.927
	3/4	0.847	0.050	0.102	0.794
	4/5	0.942	0.058	0.000	0.934
	5/6	0.993	0.007	0.000	0.999

Table 8.4.6E-2
Accuracy and Consistency of Classification Indices: Litr (Grade 7) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.719	0.634		0.456	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.710		0.721	
	2	0.777		0.661	
	3	0.817		0.715	
	4	0.579		0.520	
	5	-		0.184	
	6	-		1.000	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.979	0.014	0.008	0.978
	2/3	0.935	0.033	0.032	0.916
	3/4	0.845	0.054	0.102	0.790
	4/5	0.942	0.058	0.000	0.936
	5/6	0.992	0.008	0.000	0.999

Table 8.4.6E-3
Accuracy and Consistency of Classification Indices: Litr (Grade 8) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.717	0.630		0.455	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.753		0.732	
	2	0.760		0.647	
	3	0.809		0.723	
	4	0.536		0.461	
	5	-		0.202	
	6	-		0.998	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.973	0.015	0.012	0.970
	2/3	0.923	0.039	0.039	0.898
	3/4	0.856	0.057	0.088	0.807
	4/5	0.945	0.055	0.000	0.939
	5/6	0.991	0.009	0.000	0.998

8.4.7 Comprehension Composite 6-8

Figure 8.4.7A
Scale Scores: Cphn 6-8 S301

Figure 8.4.7B
Proficiency Level: Cphn 6-8 S301
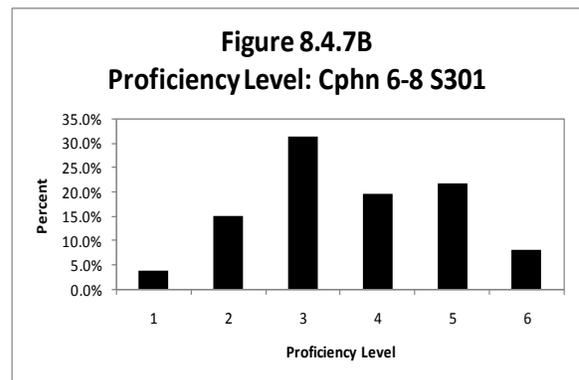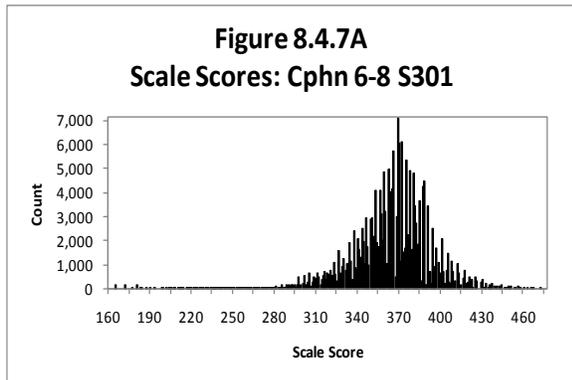
Table 8.4.7A
Scale Score Descriptive Statistics: Cphn 6-8 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	69,644	165	472	357.68	26.07
7	66,470	172	472	364.81	28.08
8	61,048	180	472	371.33	30.26
Total	197,162	165	472	364.31	28.64

Table 8.4.7B
Proficiency Level Distribution: Cphn 6-8 S301

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	1,708	2.5%	2,539	3.8%	3,322	5.4%	7,569	3.8%
2	8,870	12.7%	10,028	15.1%	10,916	17.9%	29,814	15.1%
3	23,457	33.7%	21,003	31.6%	17,307	28.3%	61,767	31.3%
4	14,190	20.4%	14,168	21.3%	10,364	17.0%	38,722	19.6%
5	16,186	23.2%	13,410	20.2%	13,448	22.0%	43,044	21.8%
6	5,233	7.5%	5,322	8.0%	5,691	9.3%	16,246	8.2%
Total	69,644	100.0%	66,470	100.0%	61,048	100.0%	197,162	100.0%

Table 8.4.7C

n/a

Figure 8.4.7C

n/a

Figure 8.4.7D

n/a

Table 8.4.7D

Comprehension Composite Reliability: Cphn 6-8 S301

Component	Weight	Variance	Reliability
Listening	0.30	1677.033	0.606
Reading	0.70	730.965	0.765
Comprehension		819.314	0.825

*Variances from students who had results in all four domains

Table 8.4.7E-1

Accuracy and Consistency of Classification Indices: Cphn (Grade 6) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.539	0.439		0.270	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.744		0.583	
	2	0.713		0.560	
	3	0.691		0.553	
	4	0.349		0.276	
	5	0.462		0.400	
	6	-		0.200	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.988	0.006	0.006	0.981
	2/3	0.935	0.028	0.037	0.906
	3/4	0.830	0.050	0.120	0.770
	4/5	0.804	0.061	0.135	0.734
	5/6	0.925	0.075	0.000	0.902

Table 8.4.7E-2
Accuracy and Consistency of Classification Indices: Cphn (Grade 7) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.531	0.428		0.271	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.760		0.598	
	2	0.695		0.545	
	3	0.639		0.514	
	4	0.384		0.302	
	5	0.435		0.369	
	6	-		0.244	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.981	0.009	0.010	0.971
	2/3	0.922	0.033	0.045	0.888
	3/4	0.836	0.060	0.104	0.776
	4/5	0.822	0.063	0.115	0.756
	5/6	0.920	0.080	0.000	0.893

Table 8.4.7E-3
Accuracy and Consistency of Classification Indices: Cphn (Grade 8) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.526	0.424		0.280	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.774		0.612	
	2	0.671		0.530	
	3	0.586		0.470	
	4	0.334		0.257	
	5	0.472		0.403	
	6	-		0.315	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.973	0.011	0.016	0.959
	2/3	0.906	0.040	0.054	0.866
	3/4	0.845	0.061	0.094	0.788
	4/5	0.843	0.067	0.090	0.780
	5/6	0.907	0.093	0.000	0.876

Return to Visual Table of Tables and Figures

8.4.8 Overall Composite 6-8

Figure 8.4.8A
Scale Scores: Over 6-8 S301

Figure 8.4.8B
Proficiency Level: Over 6-8 S301
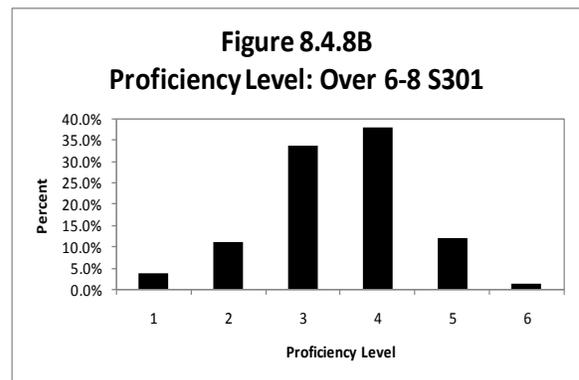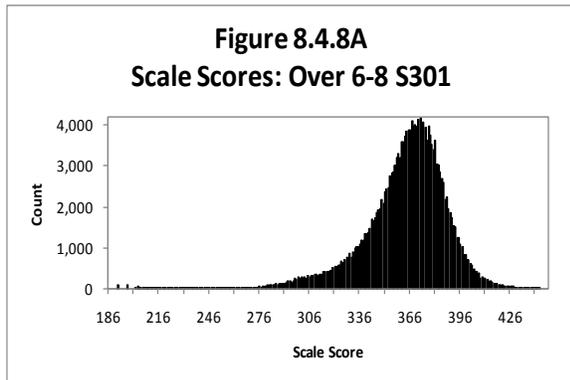
Table 8.4.8A
Scale Score Descriptive Statistics: Over 6-8 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	69,485	191	438	357.76	23.37
7	66,273	197	441	363.65	24.59
8	60,850	203	443	368.87	25.97
Total	196,608	191	443	363.18	25.02

Table 8.4.8B
Proficiency Level Distribution: Over 6-8 S301

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	1,822	2.6%	2,435	3.7%	2,943	4.8%	7,200	3.7%
2	6,690	9.6%	7,301	11.0%	8,125	13.4%	22,116	11.2%
3	22,045	31.7%	23,253	35.1%	20,755	34.1%	66,053	33.6%
4	29,286	42.1%	23,984	36.2%	21,521	35.4%	74,791	38.0%
5	8,583	12.4%	8,327	12.6%	6,556	10.8%	23,466	11.9%
6	1,059	1.5%	973	1.5%	950	1.6%	2,982	1.5%
Total	69,485	100.0%	66,273	100.0%	60,850	100.0%	196,608	100.0%

Table 8.4.8C

n/a

Figure 8.4.8C

n/a

Figure 8.4.8D

n/a

Table 8.4.8D

Overall Composite Reliability: Over 6-8 S301

Component	Weight	Variance	Reliability
Listening	0.15	1677.033	0.606
Reading	0.35	730.965	0.765
Speaking	0.15	2135.697	0.895
Writing	0.35	573.734	0.911
Overall Composite		626.054	0.925

*Variances from students who had results in all four domains

Table 8.4.8E-1

Accuracy and Consistency of Classification Indices: Over (Grade 6) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.732	0.657		0.505	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.687		0.785	
	2	0.777		0.671	
	3	0.838		0.749	
	4	0.952		0.660	
	5	0.088		0.372	
	6	-		0.996	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.986	0.010	0.004	0.989
	2/3	0.959	0.024	0.016	0.950
	3/4	0.904	0.037	0.059	0.875
	4/5	1.000	0.000	0.014	0.845
	5/6	0.985	0.015	0.000	0.992

Table 8.4.8E-2
Accuracy and Consistency of Classification Indices: Over (Grade 7) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.719	0.638		0.494	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.750		0.791	
	2	0.769		0.659	
	3	0.834		0.747	
	4	0.719		0.603	
	5	0.321		0.391	
	6	-		0.995	
Indices at Cut Points		Accuracy			
			False Positives	False Negatives	Consistency
	Cut Point	Accuracy			
	1/2	0.984	0.010	0.006	0.985
	2/3	0.955	0.026	0.019	0.943
	3/4	0.896	0.039	0.064	0.863
	4/5	0.914	0.058	0.028	0.847
	5/6	0.985	0.015	0.000	0.991

Table 8.4.8E-3
Accuracy and Consistency of Classification Indices: Over (Grade 8) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.723	0.635		0.497	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.812		0.786	
	2	0.767		0.659	
	3	0.808		0.716	
	4	0.678		0.614	
	5	0.469		0.373	
	6	-		0.980	
Indices at Cut Points		Accuracy			
			False Positives	False Negatives	Consistency
	Cut Point	Accuracy			
	1/2	0.983	0.009	0.008	0.980
	2/3	0.949	0.027	0.024	0.932
	3/4	0.898	0.040	0.062	0.861
	4/5	0.895	0.079	0.026	0.864
	5/6	0.984	0.016	0.000	0.988

8.5   Grades: 9–12

8.5.1 Listening 9-12

Figure 8.5.1A
Scale Scores: List 9-12 S301

Figure 8.5.1B
Proficiency Level: List 9-12 S301

Table 8.5.1A

Scale Score Descriptive Statistics: List 9-12 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	68,157	136	501	381.50	50.90
10	43,971	140	501	386.20	46.95
11	34,928	144	501	391.56	46.20
12	26,484	148	501	393.09	47.86
Total	173,540	136	501	386.48	48.75

Table 8.5.1B

Proficiency Level Distribution: List 9-12 S301

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	6,097	8.9%	3,275	7.4%	2,776	7.9%	2,426	9.2%	14,574	8.4%
2	8,547	12.5%	5,703	13.0%	3,852	11.0%	2,063	7.8%	20,165	11.6%
3	10,634	15.6%	8,776	20.0%	7,394	21.2%	5,762	21.8%	32,566	18.8%
4	19,174	28.1%	10,658	24.2%	9,165	26.2%	8,478	32.0%	47,475	27.4%
5	15,728	23.1%	12,318	28.0%	8,624	24.7%	5,248	19.8%	41,918	24.2%
6	7,977	11.7%	3,241	7.4%	3,117	8.9%	2,507	9.5%	16,842	9.7%
Total	68,157	100.0%	43,971	100.0%	34,928	100.0%	26,484	100.0%	173,540	100.0%

Table 8.5.1C
Conditional Standard Error of Measurement at Cut Scores: List 9-12 S301

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	9	312	19.91	19.54	24.42
	10	322	20.29	18.79	22.54
	11	332	21.04	18.41	21.42
	12	343	22.17	18.03	19.91
2/3	9	352	23.29	18.03	19.16
	10	358	24.42	18.03	18.79
	11	363	25.17	18.03	18.41
	12	366	25.92	18.03	18.41
3/4	9	381	29.31	18.79	18.03
	10	386	30.81	19.16	17.66
	11	389	31.56	19.16	17.66
	12	391	32.31	19.54	18.03
4/5	9	406	n/a	20.66	18.41
	10	412	n/a	21.42	18.79
	11	416	n/a	22.17	18.79
	12	418	n/a	22.54	19.16
5/6	9	432	n/a	n/a	20.29
	10	436	n/a	n/a	20.66
	11	438	n/a	n/a	21.04
	12	439	n/a	n/a	21.04

Figure 8.5.1C
Test Characteristic Curve: List 9-12ABC S301
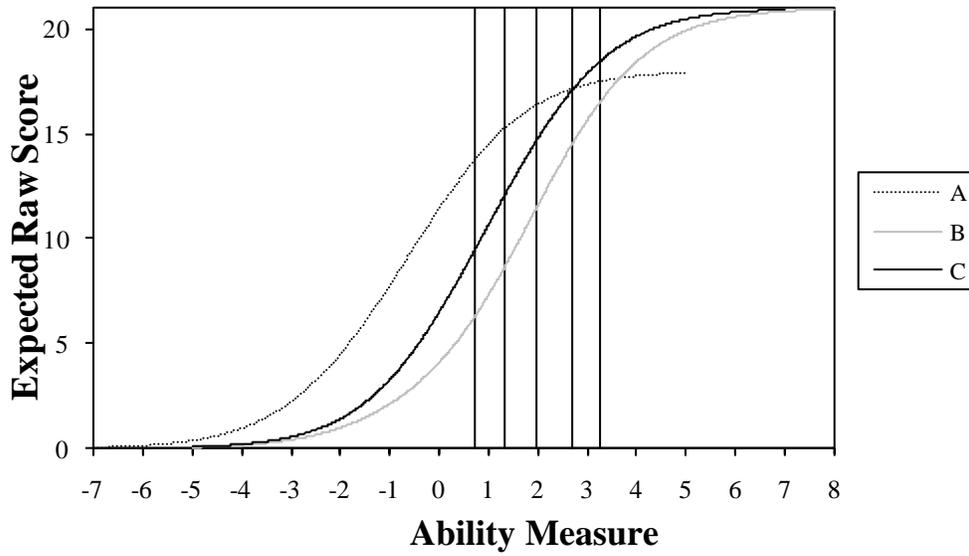
Figure 8.5.1D
Test Information Function: List 9-12ABC S301
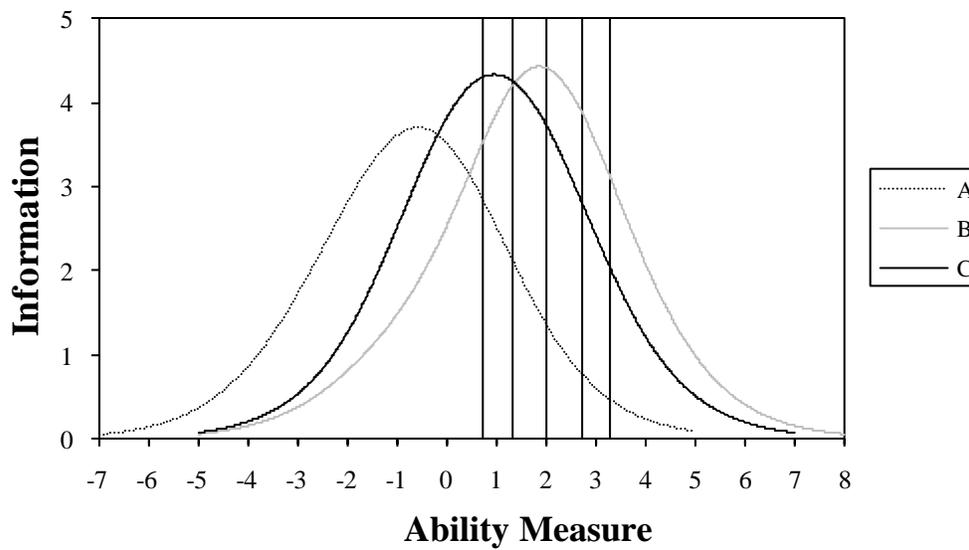
Table 8.5.1D

Weighted Reliability: List 9-12 S301

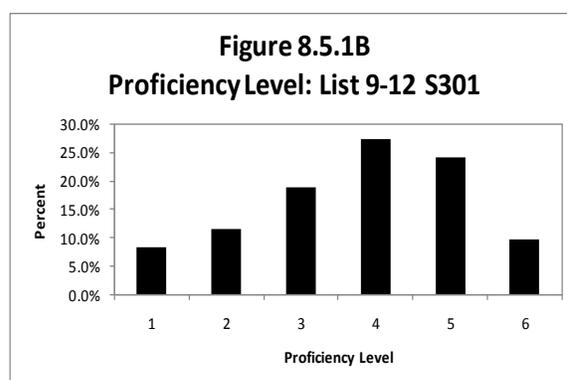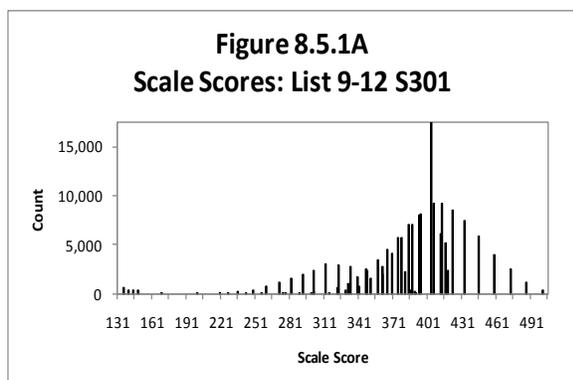| Tiers	No. of Students	Reliability	Reliability
A	24,959	0.740	
B	68,707	0.726	0.704
C	79,874	0.673	

Table 8.5.1E-1

Accuracy and Consistency of Classification Indices: List (Grade 9) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.391	0.312		0.147	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.817		0.684	
	2	0.524		0.327	
	3	0.297		0.191	
	4	0.353		0.329	
	5	0.330		0.300	
	6	-		0.169	

Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.962	0.015	0.023	0.945
	2/3	0.904	0.022	0.074	0.859
	3/4	0.810	0.047	0.143	0.709
	4/5	0.669	0.173	0.158	0.628
	5/6	0.883	0.117	0.000	0.813

Table 8.5.1E-2

Accuracy and Consistency of Classification Indices: List (Grade 10) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.353	0.311		0.135	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.789		0.650	
	2	0.549		0.333	
	3	0.355		0.235	
	4	0.292		0.281	
	5	-		0.359	
	6	-		0.103	

Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.967	0.015	0.018	0.951
	2/3	0.900	0.020	0.080	0.856
	3/4	0.778	0.037	0.185	0.679
	4/5	0.646	0.354	0.000	0.623
	5/6	0.926	0.074	0.000	0.865

Table 8.5.1E-3

Accuracy and Consistency of Classification Indices: List (Grade 11) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.362	0.305		0.125	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.822		0.695	
	2	0.487		0.266	
	3	0.368		0.242	
	4	0.309		0.299	
	5	-		0.304	
	6	-		0.116	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.969	0.013	0.017	0.954
	2/3	0.904	0.017	0.078	0.861
	3/4	0.758	0.027	0.216	0.665
	4/5	0.664	0.336	0.000	0.616
	5/6	0.911	0.089	0.000	0.847

Table 8.5.1E-4

Accuracy and Consistency of Classification Indices: List (Grade 12) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.401	0.304		0.111	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.891		0.739	
	2	0.314		0.126	
	3	0.354		0.238	
	4	0.361		0.355	
	5	-		0.228	
	6	-		0.112	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.970	0.009	0.022	0.952
	2/3	0.911	0.012	0.078	0.853
	3/4	0.728	0.019	0.253	0.629
	4/5	0.707	0.293	0.000	0.615
	5/6	0.905	0.095	0.000	0.841

8.5.2 Reading 9-12

Figure 8.5.2A
Scale Scores: Read 9-12 S301

Figure 8.5.2B
Proficiency Level: Read 9-12 S301
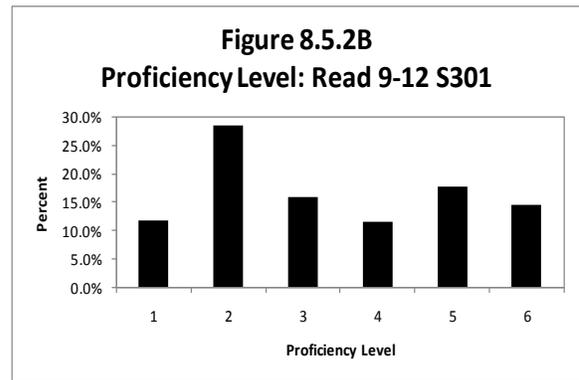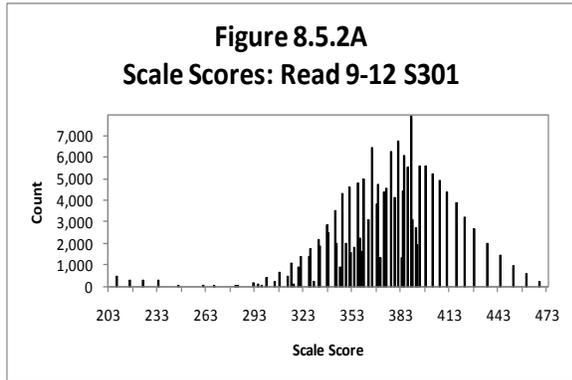
Table 8.5.2A

Scale Score Descriptive Statistics: Read 9-12 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	68,155	208	468	371.82	32.36
10	43,950	216	468	373.86	30.53
11	34,904	224	468	377.49	30.91
12	26,443	233	468	378.60	31.70
Total	173,452	208	468	374.51	31.63

Table 8.5.2B

Proficiency Level Distribution: Read 9-12 S301

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	7,423	10.9%	4,944	11.2%	4,161	11.9%	3,741	14.1%	20,269	11.7%
2	17,552	25.8%	13,929	31.7%	10,300	29.5%	7,437	28.1%	49,218	28.4%
3	13,351	19.6%	6,779	15.4%	4,435	12.7%	3,215	12.2%	27,780	16.0%
4	6,960	10.2%	5,757	13.1%	4,533	13.0%	2,806	10.6%	20,056	11.6%
5	11,904	17.5%	7,005	15.9%	6,131	17.6%	5,704	21.6%	30,744	17.7%
6	10,965	16.1%	5,536	12.6%	5,344	15.3%	3,540	13.4%	25,385	14.6%
Total	68,155	100.0%	43,950	100.0%	34,904	100.0%	26,443	100.0%	173,452	100.0%

Table 8.5.2C

Conditional Standard Error of Measurement at Cut Scores: Read 9-12 S301

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	9	336	11.96	11.70	14.82
	10	341	11.70	11.44	14.04
	11	346	11.70	11.18	13.26
	12	350	11.70	10.92	12.74
2/3	9	364	11.96	10.92	11.44
	10	370	12.48	10.92	10.92
	11	374	12.74	11.18	10.66
	12	376	13.00	11.18	10.66
3/4	9	381	13.52	11.44	10.40
	10	383	13.78	11.70	10.40
	11	384	13.78	11.70	10.40
	12	385	14.04	11.70	10.40
4/5	9	387	n/a	11.96	10.40
	10	390	n/a	12.22	10.40
	11	392	n/a	12.48	10.40
	12	393	n/a	12.48	10.40
5/6	9	402	n/a	n/a	10.66
	10	406	n/a	n/a	10.92
	11	407	n/a	n/a	10.92
	12	408	n/a	n/a	11.18

Figure 8.5.2C
Test Characteristic Curve: Read 9-12ABC S301
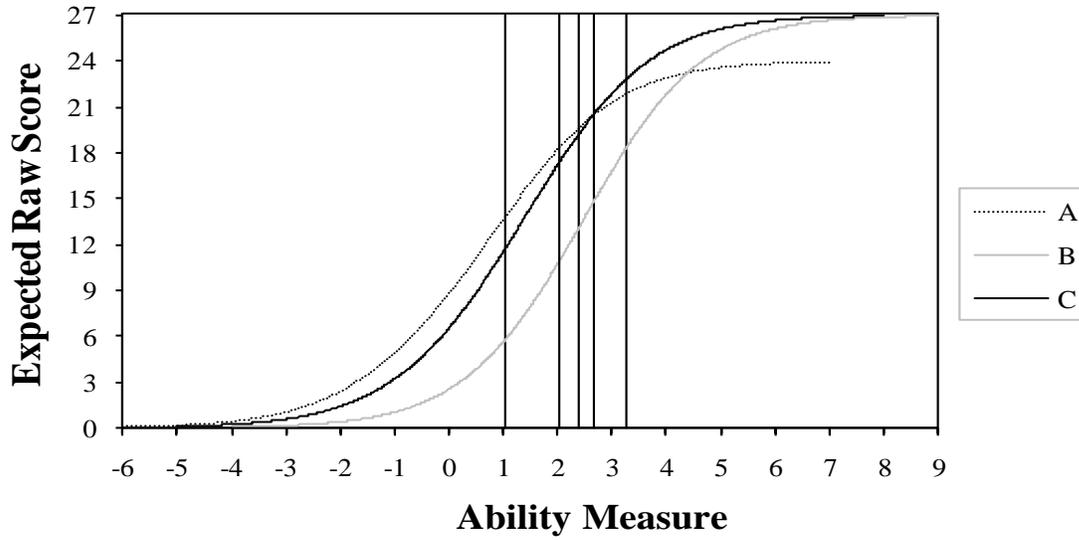
Figure 8.5.2D
Test Information Function: Read 9-12ABC S301
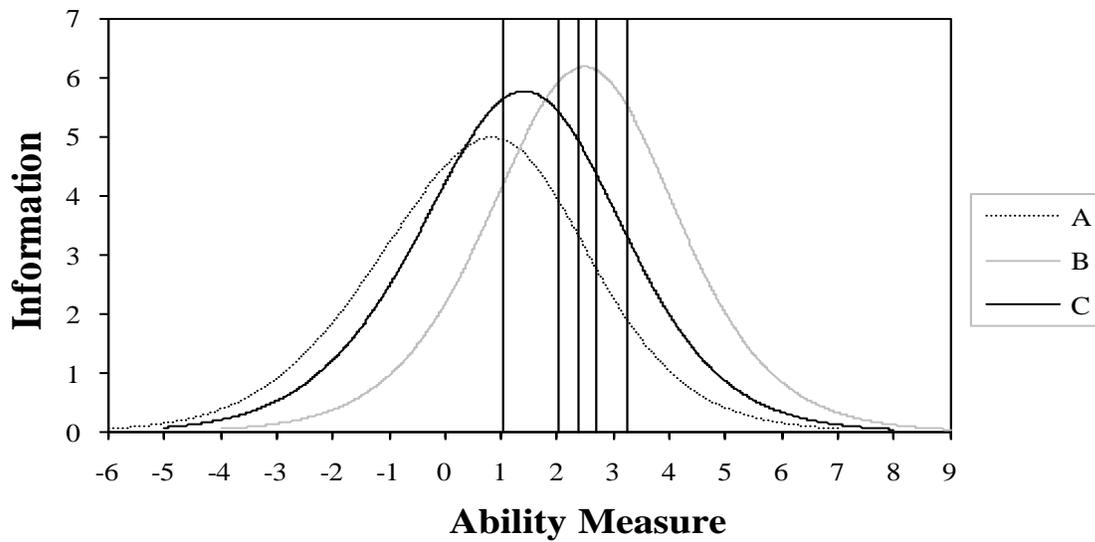
Table 8.5.2D

Weighted Reliability: Read 9-12 S301

Tiers	No. of Students	Reliability	Reliability
A	24,856	0.798	
B	68,673	0.809	0.796
C	79,923	0.785	

Table 8.5.2E-1

Accuracy and Consistency of Classification Indices: Read (Grade 9) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.477	0.388		0.253	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.741		0.586	
	2	0.649		0.517	
	3	0.377		0.291	
	4	0.199		0.151	
	5	0.348		0.286	
	6	0.608		0.427	

Indices at Cut Points		Accuracy			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.942	0.028	0.030	0.913
	2/3	0.861	0.050	0.089	0.810
	3/4	0.835	0.075	0.090	0.771
	4/5	0.830	0.072	0.098	0.765
	5/6	0.863	0.095	0.043	0.816

Table 8.5.2E-2

Accuracy and Consistency of Classification Indices: Read (Grade 10) 301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.483	0.395		0.255	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.727		0.568	
	2	0.690		0.568	
	3	0.296		0.229	
	4	0.263		0.199	
	5	0.352		0.284	
	6	0.621		0.402	

Indices at Cut Points		Accuracy			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.938	0.030	0.032	0.906
	2/3	0.847	0.055	0.098	0.792
	3/4	0.836	0.078	0.086	0.773
	4/5	0.841	0.066	0.094	0.778
	5/6	0.889	0.087	0.024	0.850

Table 8.5.2E-3
Accuracy and Consistency of Classification Indices: Read (Grade 11) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.475	0.389		0.252	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.739		0.586	
	2	0.662		0.536	
	3	0.248		0.189	
	4	0.248		0.189	
	5	0.354		0.289	
	6	0.608		0.419	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.937	0.031	0.032	0.905
	2/3	0.847	0.057	0.096	0.792
	3/4	0.834	0.071	0.095	0.770
	4/5	0.831	0.070	0.098	0.766
	5/6	0.868	0.094	0.038	0.822

Table 8.5.2E-4
Accuracy and Consistency of Classification Indices: Read (Grade 12) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.460	0.380		0.242	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.764		0.627	
	2	0.646		0.510	
	3	0.230		0.172	
	4	0.188		0.146	
	5	0.399		0.343	
	6	-		0.314	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.933	0.033	0.034	0.900
	2/3	0.845	0.051	0.104	0.790
	3/4	0.825	0.059	0.116	0.758
	4/5	0.813	0.075	0.112	0.742
	5/6	0.866	0.134	0.000	0.811

8.5.3 Writing 9-12

Figure 8.5.3A
Scale Scores: Writ 9-12 S301

Figure 8.5.3B
Proficiency Level: Writ 9-12 S301
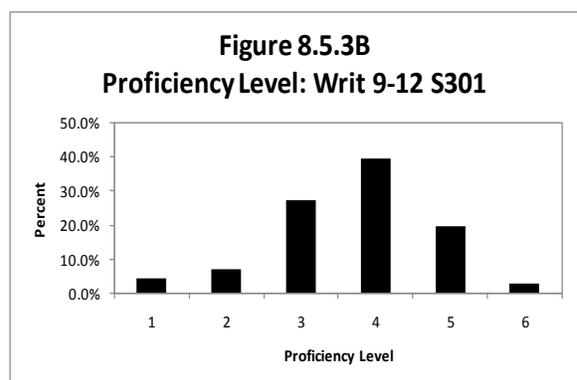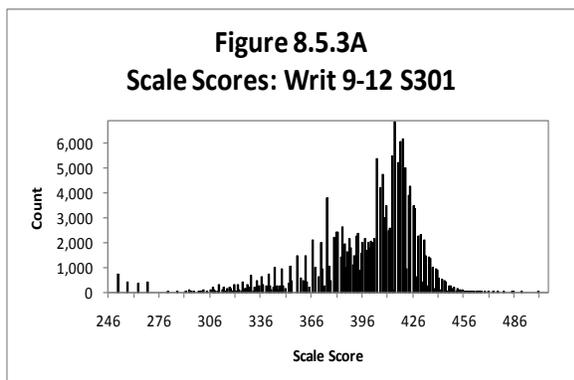
Table 8.5.3A

Scale Score Descriptive Statistics: Writ 9-12 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	68,021	251	476	396.56	32.81
10	43,878	257	479	398.42	30.16
11	34,857	263	490	401.75	29.16
12	26,394	269	500	402.09	30.39
Total	173,150	251	500	398.92	31.16

Table 8.5.3B

Proficiency Level Distribution: Writ 9-12 S301

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	2,528	3.7%	1,794	4.1%	1,548	4.4%	1,613	6.1%	7,483	4.3%
2	4,994	7.3%	2,926	6.7%	1,977	5.7%	2,132	8.1%	12,029	6.9%
3	14,256	21.0%	12,074	27.5%	10,988	31.5%	9,543	36.2%	46,861	27.1%
4	21,979	32.3%	18,831	42.9%	16,135	46.3%	11,357	43.0%	68,302	39.4%
5	20,960	30.8%	7,442	17.0%	3,808	10.9%	1,609	6.1%	33,819	19.5%
6	3,304	4.9%	811	1.8%	401	1.2%	140	0.5%	4,656	2.7%
Total	68,021	100.0%	43,878	100.0%	34,857	100.0%	26,394	100.0%	173,150	100.0%

Table 8.5.3C
Conditional Standard Error of Measurement at Cut Scores: Writ 9-12 S301

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	9	327	9.02	6.53	6.53
	10	336	10.26	7.46	6.84
	11	344	11.19	8.09	7.77
	12	352	11.82	8.40	8.40
2/3	9	356	12.13	8.40	8.40
	10	363	12.13	8.40	8.40
	11	370	11.82	8.40	8.40
	12	377	11.51	8.09	8.09
3/4	9	389	11.19	8.09	8.09
	10	397	11.19	7.77	7.77
	11	404	10.88	7.46	7.77
	12	410	10.57	7.15	7.46
4/5	9	415	n/a	7.15	7.15
	10	422	n/a	6.84	7.15
	11	428	n/a	6.53	6.84
	12	434	n/a	6.53	6.53
5/6	9	435	n/a	n/a	6.53
	10	441	n/a	n/a	6.53
	11	447	n/a	n/a	6.22
	12	452	n/a	n/a	6.53

Figure 8.5.3C
Test Characteristic Curve: Writ 9-12ABC S301
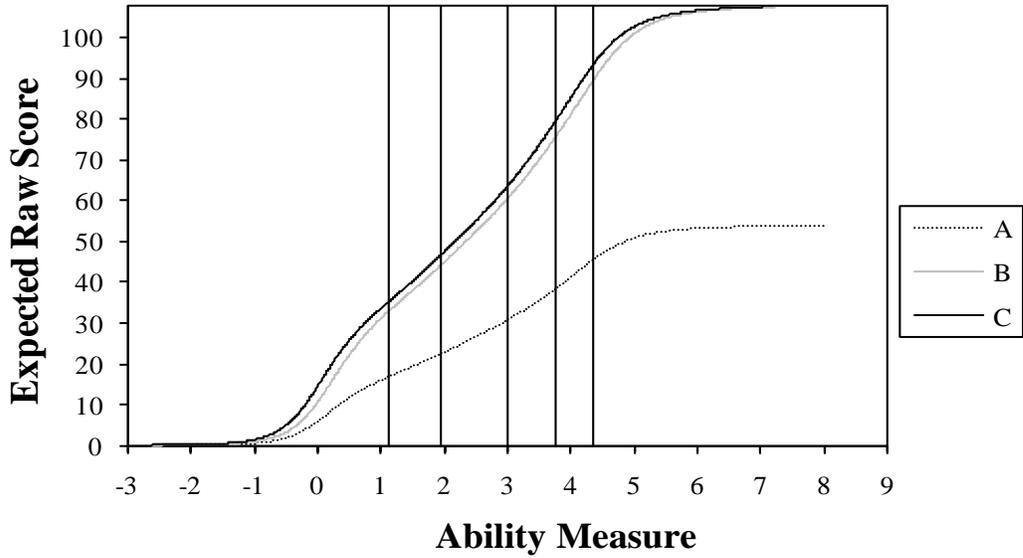
Figure 8.5.3D
Test Information Function: Writ 9-12ABC S301

Return to Visual Table of Tables and Figures
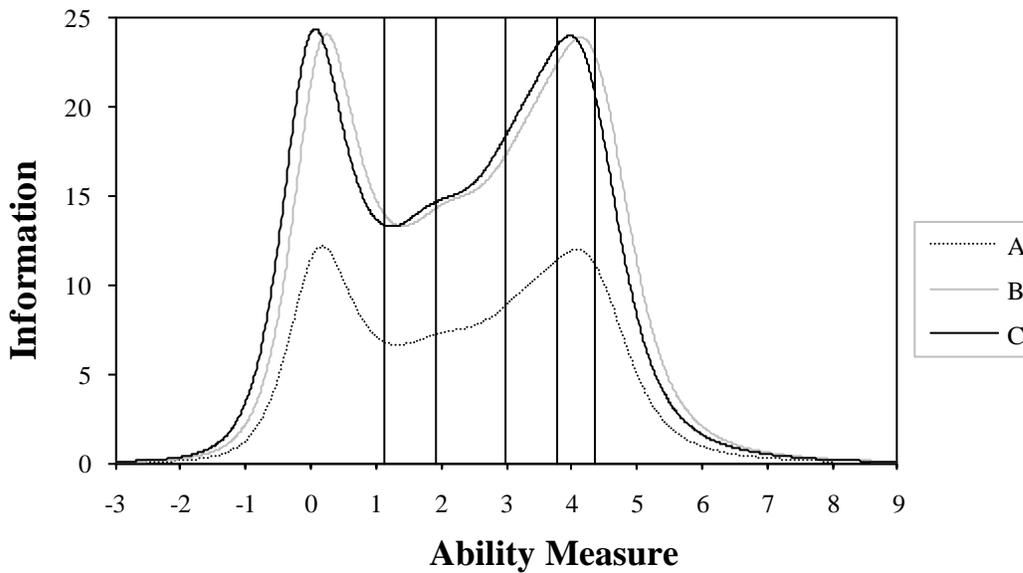
Table 8.5.3D
Weighted Reliability: Writ 9-12 S301

Tiers	No. of Students	Reliability	Reliability
A	24,815	0.853	
B	68,534	0.933	0.910
C	79,801	0.908	

Table 8.5.3E-1
Accuracy and Consistency of Classification Indices: Writ (Grade 9) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.660	0.553		0.394	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.838		0.740	
	2	0.709		0.579	
	3	0.796		0.692	
	4	0.621		0.479	
	5	0.600		0.552	
	6	-		0.115	

Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.988	0.006	0.006	0.983
	2/3	0.968	0.015	0.017	0.954
	3/4	0.932	0.021	0.047	0.906
	4/5	0.816	0.068	0.117	0.744
	5/6	0.951	0.049	0.000	0.937

Table 8.5.3E-2
Accuracy and Consistency of Classification Indices: Writ (Grade 10) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.683	0.595		0.416	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.865		0.781	
	2	0.680		0.545	
	3	0.831		0.728	
	4	0.618		0.603	
	5	-		0.351	
	6	-		0.060	

Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.989	0.005	0.006	0.983
	2/3	0.969	0.016	0.015	0.955
	3/4	0.913	0.025	0.062	0.879
	4/5	0.812	0.188	0.000	0.777
	5/6	0.982	0.018	0.000	0.980

Table 8.5.3E-3

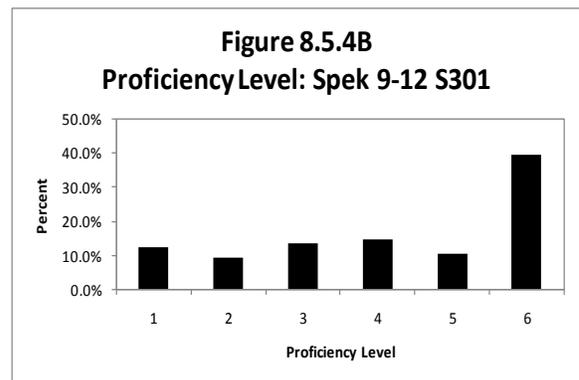Accuracy and Consistency of Classification Indices: Writ (Grade 11) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.729	0.633		0.437	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.890		0.822	
	2	0.657		0.518	
	3	0.849		0.717	
	4	0.678		0.652	
	5	-		0.216	
	6	-		0.192	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.990	0.005	0.006	0.985
	2/3	0.972	0.015	0.012	0.959
	3/4	0.887	0.025	0.087	0.842
	4/5	0.879	0.121	0.000	0.837
	5/6	0.988	0.012	0.000	0.988

Table 8.5.3E-4

Accuracy and Consistency of Classification Indices: Writ (Grade 12) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.698	0.604		0.393	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.876		0.807	
	2	0.738		0.611	
	3	0.800		0.583	
	4	0.637		0.614	
	5	-		0.108	
	6	-		0.000	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.986	0.008	0.006	0.980
	2/3	0.968	0.013	0.018	0.955
	3/4	0.810	0.028	0.162	0.736
	4/5	0.934	0.066	0.000	0.915
	5/6	0.995	0.005	0.000	0.995

8.5.4 Speaking 9-12

Figure 8.5.4A
Scale Scores: Spek 9-12 S301

Figure 8.5.4B
Proficiency Level: Spek 9-12 S301
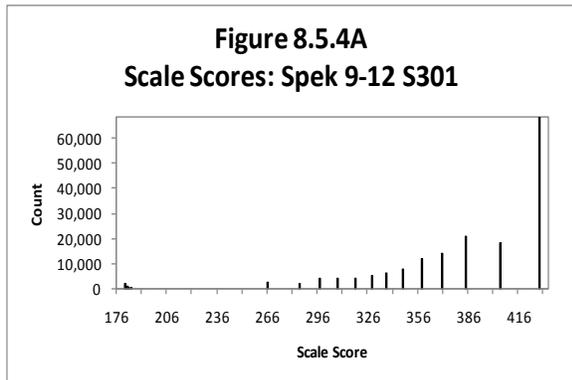
Table 8.5.4A

Scale Score Descriptive Statistics: Spek 9-12 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	68,094	181	428	379.87	57.57
10	44,027	182	428	381.18	51.80
11	34,959	183	428	385.05	48.83
12	26,525	184	428	387.87	48.12
Total	173,605	181	428	382.47	53.13

Table 8.5.4B

Proficiency Level Distribution: Spek 9-12 S301

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	10,251	15.1%	5,538	12.6%	3,382	9.7%	2,077	7.8%	21,248	12.2%
2	4,234	6.2%	5,293	12.0%	4,029	11.5%	2,800	10.6%	16,356	9.4%
3	6,769	9.9%	6,767	15.4%	5,758	16.5%	4,402	16.6%	23,696	13.6%
4	12,348	18.1%	5,469	12.4%	4,371	12.5%	3,397	12.8%	25,585	14.7%
5	7,289	10.7%	4,469	10.2%	3,629	10.4%	2,808	10.6%	18,195	10.5%
6	27,203	39.9%	16,491	37.5%	13,790	39.4%	11,041	41.6%	68,525	39.5%
Total	68,094	100.0%	44,027	100.0%	34,959	100.0%	26,525	100.0%	173,605	100.0%

Table 8.5.4C
Conditional Standard Error of Measurement at
Cut Scores: Spek 9-12 S301

Proficiency Level	Grade	Cut Score	SEM
1/2	9	319	20.49
	10	321	20.69
	11	322	20.69
	12	323	20.89
2/3	9	347	22.49
	10	351	22.69
	11	354	22.90
	12	357	23.10
3/4	9	366	23.90
	10	371	24.10
	11	377	24.90
	12	384	25.91
4/5	9	388	26.51
	10	393	27.92
	11	399	29.32
	12	405	30.53
5/6	9	407	30.93
	10	412	32.13
	11	416	32.94
	12	421	33.54

Figure 8.5.4C
Test Characteristic Curve: Spek 9-12 S301
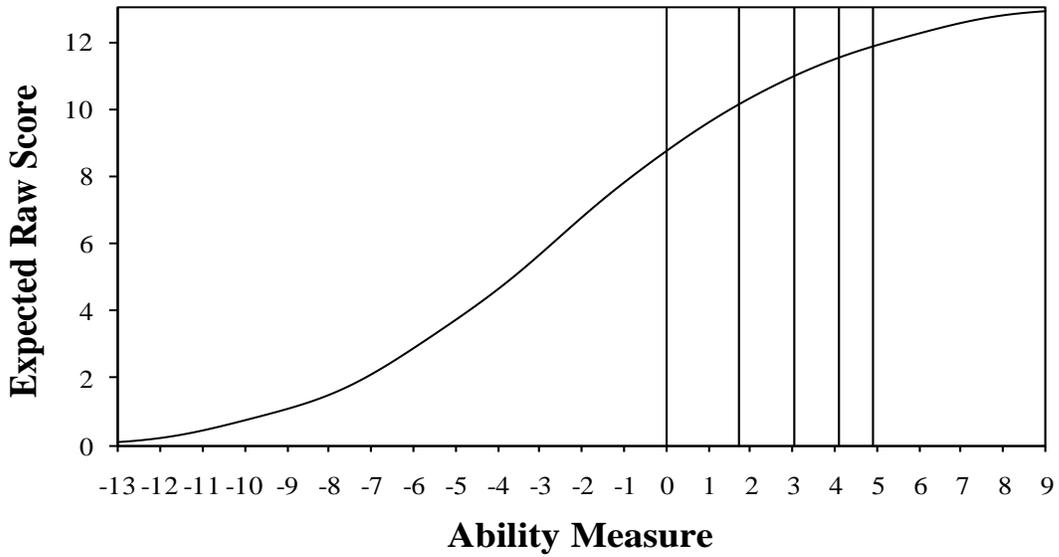
Figure 8.5.4D
Test Information Function: Spek 9-12 S301
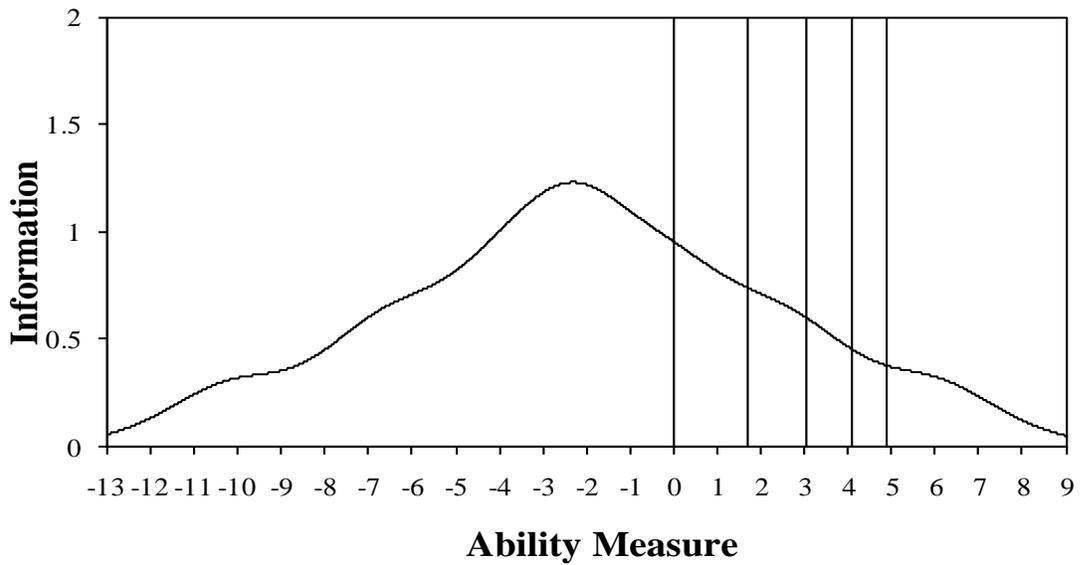
Table 8.5.4D

Reliability: Spek 9-12 S301

Tiers	No. of Students	Reliability
--	173,605	0.918

Table 8.5.4E-1

Accuracy and Consistency of Classification Indices: Spek (Grade 9) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.641	0.555		0.417	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.870		0.783	
	2	0.325		0.233	
	3	0.368		0.286	
	4	0.523		0.419	
	5	0.256		0.180	
	6	0.839		0.769	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.959	0.019	0.022	0.938
	2/3	0.933	0.040	0.027	0.905
	3/4	0.903	0.047	0.050	0.875
	4/5	0.891	0.027	0.082	0.853
	5/6	0.886	0.047	0.068	0.820

Table 8.5.4E-2

Accuracy and Consistency of Classification Indices: Spek (Grade 10) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.637	0.545		0.425	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.805		0.701	
	2	0.496		0.390	
	3	0.498		0.407	
	4	0.407		0.310	
	5	0.318		0.212	
	6	0.891		0.827	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.954	0.025	0.021	0.932
	2/3	0.914	0.044	0.042	0.886
	3/4	0.902	0.031	0.067	0.870
	4/5	0.926	0.026	0.048	0.886
	5/6	0.900	0.062	0.038	0.849

Table 8.5.4E-3

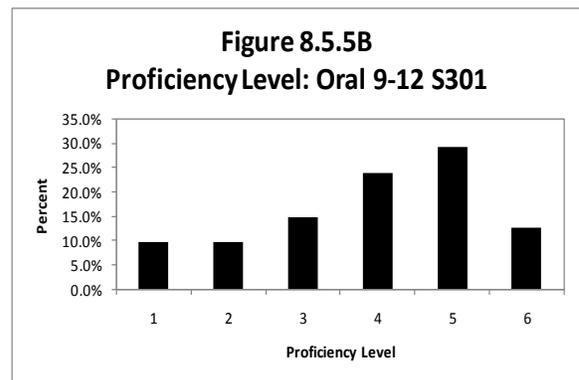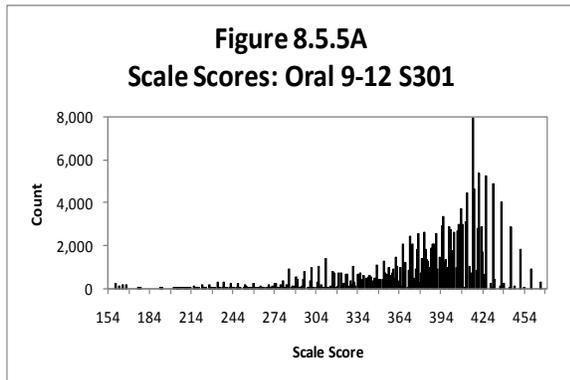Accuracy and Consistency of Classification Indices: Spek (Grade 11) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.633	0.531		0.405	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.792		0.680	
	2	0.532		0.419	
	3	0.559		0.461	
	4	0.424		0.314	
	5	0.294		0.195	
	6	0.888		0.824	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.962	0.021	0.017	0.944
	2/3	0.923	0.039	0.038	0.897
	3/4	0.910	0.028	0.062	0.880
	4/5	0.928	0.026	0.046	0.886
	5/6	0.880	0.081	0.040	0.822

Table 8.5.4E-4

Accuracy and Consistency of Classification Indices: Spek (Grade 12) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.349	0.433		0.303	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.784		0.674	
	2	0.572		0.453	
	3	0.614		0.508	
	4	0.415		0.267	
	5	0.165		0.160	
	6	-		0.756	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.969	0.018	0.012	0.954
	2/3	0.935	0.032	0.033	0.913
	3/4	0.920	0.023	0.057	0.892
	4/5	0.915	0.026	0.059	0.853
	5/6	0.584	0.416	0.000	0.702

8.5.5 Oral Language Composite 9-12

Figure 8.5.5A
Scale Scores: Oral 9-12 S301

Figure 8.5.5B
Proficiency Level: Oral 9-12 S301

Table 8.5.5A
Scale Score Descriptive Statistics: Oral 9-12 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	67,778	159	465	380.95	49.17
10	43,758	161	465	383.96	43.42
11	34,726	164	465	388.58	41.18
12	26,262	166	465	390.78	41.40
Total	172,524	159	465	384.75	45.22

Table 8.5.5B
Proficiency Level Distribution: Oral 9-12 S301

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	8,021	11.8%	4,165	9.5%	2,776	8.0%	1,940	7.4%	16,902	9.8%
2	5,904	8.7%	4,623	10.6%	3,472	10.0%	2,562	9.8%	16,561	9.6%
3	7,632	11.3%	6,573	15.0%	6,047	17.4%	5,119	19.5%	25,371	14.7%
4	14,246	21.0%	10,638	24.3%	8,984	25.9%	7,509	28.6%	41,377	24.0%
5	21,244	31.3%	12,690	29.0%	9,964	28.7%	6,428	24.5%	50,326	29.2%
6	10,731	15.8%	5,069	11.6%	3,483	10.0%	2,704	10.3%	21,987	12.7%
Total	67,778	100.0%	43,758	100.0%	34,726	100.0%	26,262	100.0%	172,524	100.0%

Table 8.5.5C

n/a

Figure 8.5.5C

n/a

Figure 8.5.5D

n/a

Table 8.5.5D

Oral Composite Reliability: Oral 9-12 S301

Component	Weight	Variance	Reliability
Listening	0.50	2357.730	0.704
Speaking	0.50	2802.444	0.918
Oral		2030.310	0.886

*Variances from students who had results in all four domains

Table 8.5.5E-1

Accuracy and Consistency of Classification Indices: Oral (Grade 9) S301

Overall Indices	Accuracy		Consistency		Kappa (k)	
	0.544		0.441		0.302	
Conditional on Level	**Level**		**Accuracy**		**Consistency**	
	1		0.886		0.807	
	2		0.525		0.393	
	3		0.449		0.324	
	4		0.491		0.356	
	5		0.512		0.474	
	6		-		0.339	

Indices at Cut Points		Accuracy			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.967	0.013	0.021	0.953
	2/3	0.947	0.024	0.029	0.923
	3/4	0.914	0.038	0.048	0.877
	4/5	0.849	0.053	0.099	0.788
	5/6	0.842	0.158	0.000	0.783

Table 8.5.5E-2

Accuracy and Consistency of Classification Indices: Oral (Grade 10) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.567	0.453		0.317	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.857		0.761	
	2	0.582		0.449	
	3	0.527		0.400	
	4	0.528		0.400	
	5	0.537		0.487	
	6	-		0.297	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.969	0.013	0.018	0.955
	2/3	0.940	0.027	0.033	0.915
	3/4	0.906	0.038	0.056	0.868
	4/5	0.850	0.059	0.091	0.788
	5/6	0.884	0.116	0.000	0.836

Table 8.5.5E-3

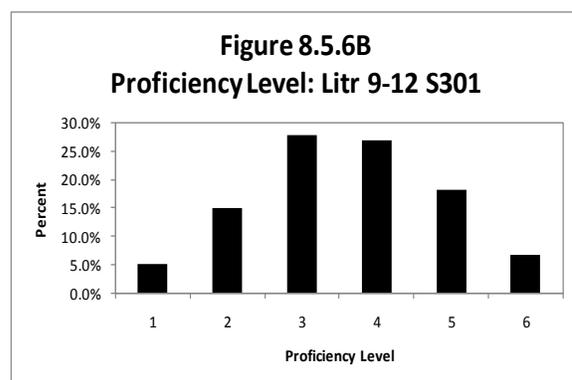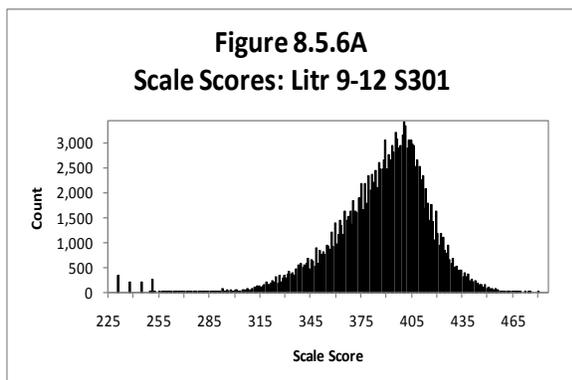Accuracy and Consistency of Classification Indices: Oral (Grade 11) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.565	0.447		0.300	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.844		0.743	
	2	0.588		0.455	
	3	0.599		0.465	
	4	0.508		0.380	
	5	0.531		0.482	
	6	-		0.226	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.974	0.012	0.014	0.961
	2/3	0.942	0.027	0.031	0.918
	3/4	0.904	0.030	0.066	0.867
	4/5	0.828	0.063	0.109	0.755
	5/6	0.900	0.100	0.000	0.851

Table 8.5.5E-4

Accuracy and Consistency of Classification Indices: Oral (Grade 12) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.504	0.418		0.256	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.818		0.719	
	2	0.615		0.479	
	3	0.626		0.463	
	4	0.444		0.370	
	5	0.422		0.384	
	6	-		0.187	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.975	0.014	0.010	0.963
	2/3	0.944	0.024	0.033	0.923
	3/4	0.889	0.022	0.089	0.846
	4/5	0.755	0.095	0.150	0.691
	5/6	0.897	0.103	0.000	0.866

8.5.6 Literacy Composite 9-12

Figure 8.5.6A
Scale Scores: Litr 9-12 S301

Figure 8.5.6B
Proficiency Level: Litr 9-12 S301

Table 8.5.6A

Scale Score Descriptive Statistics: Litr 9-12 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	67,928	230	472	384.48	30.04
10	43,806	237	467	386.43	27.58
11	34,789	244	475	389.89	27.21
12	26,317	251	480	390.64	28.22
Total	172,840	230	480	387.00	28.70

Table 8.5.6B

Proficiency Level Distribution: Litr 9-12 S301

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	3,468	5.1%	2,034	4.6%	1,710	4.9%	1,768	6.7%	8,980	5.2%
2	8,807	13.0%	6,935	15.8%	5,421	15.6%	4,572	17.4%	25,735	14.9%
3	16,911	24.9%	12,555	28.7%	10,480	30.1%	8,258	31.4%	48,204	27.9%
4	17,317	25.5%	12,166	27.8%	9,894	28.4%	7,350	27.9%	46,727	27.0%
5	15,266	22.5%	7,680	17.5%	5,348	15.4%	3,192	12.1%	31,486	18.2%
6	6,159	9.1%	2,436	5.6%	1,936	5.6%	1,177	4.5%	11,708	6.8%
Total	67,928	100.0%	43,806	100.0%	34,789	100.0%	26,317	100.0%	172,840	100.0%

Table 8.5.6C

n/a

Figure 8.5.6C

n/a

Figure 8.5.6D

n/a

Table 8.5.6D

Literacy Composite Reliability: Litr 9-12 S301

Component	Weight	Variance	Reliability
Reading	0.50	996.935	0.796
Writing	0.50	966.316	0.910
Literacy		822.131	0.912

*Variances from students who had results in all four domains

Table 8.5.6E-1

Accuracy and Consistency of Classification Indices: Litr (Grade 9) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.631	0.530		0.406	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.824		0.716	
	2	0.742		0.622	
	3	0.713		0.602	
	4	0.602		0.474	
	5	0.526		0.472	
	6	-		0.349	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.982	0.009	0.009	0.973
	2/3	0.948	0.023	0.029	0.926
	3/4	0.905	0.040	0.055	0.868
	4/5	0.881	0.043	0.077	0.830
	5/6	0.909	0.091	0.000	0.893

Table 8.5.6E-2
Accuracy and Consistency of Classification Indices: Litr (Grade 10) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.661	0.557		0.431	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.794		0.674	
	2	0.762		0.646	
	3	0.718		0.613	
	4	0.637		0.520	
	5	0.539		0.467	
	6	-		0.345	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.981	0.010	0.009	0.972
	2/3	0.939	0.027	0.035	0.913
	3/4	0.896	0.045	0.059	0.856
	4/5	0.897	0.043	0.060	0.852
	5/6	0.944	0.056	0.000	0.938

Table 8.5.6E-3
Accuracy and Consistency of Classification Indices: Litr (Grade 11) S301
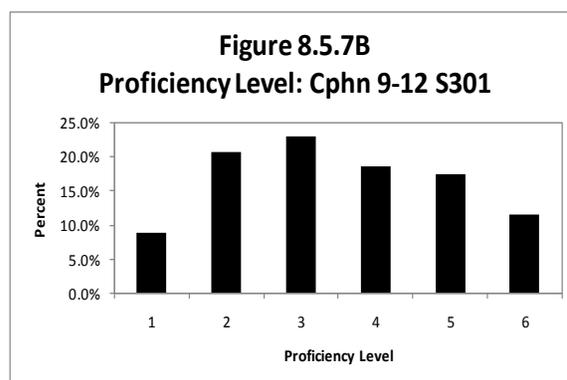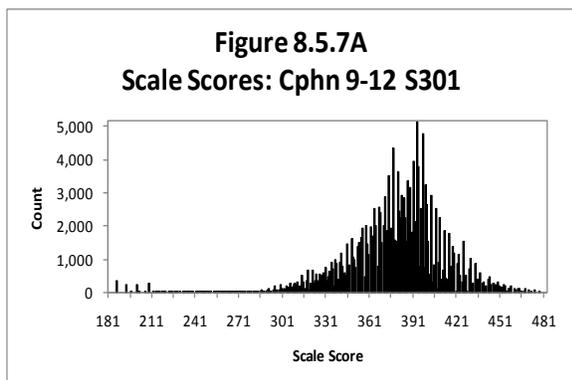
Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.637	0.536		0.401	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.801		0.690	
	2	0.761		0.644	
	3	0.745		0.635	
	4	0.591		0.476	
	5	0.454		0.389	
	6	-		0.239	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.982	0.010	0.008	0.973
	2/3	0.940	0.028	0.033	0.914
	3/4	0.891	0.037	0.072	0.851
	4/5	0.871	0.053	0.076	0.818
	5/6	0.944	0.056	0.000	0.939

Table 8.5.6E-4

Accuracy and Consistency of Classification Indices: Litr (Grade 12) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.615	0.518		0.371	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.813		0.715	
	2	0.775		0.659	
	3	0.730		0.589	
	4	0.482		0.449	
	5	-		0.281	
	6	-		0.156	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.978	0.013	0.009	0.967
	2/3	0.935	0.027	0.038	0.909
	3/4	0.866	0.030	0.103	0.816
	4/5	0.834	0.166	0.000	0.806
	5/6	0.955	0.045	0.000	0.953

8.5.7 Comprehension Composite 9-12

Figure 8.5.7A
Scale Scores: Cphn 9-12 S301

Figure 8.5.7B
Proficiency Level: Cphn 9-12 S301

Table 8.5.7A

Scale Score Descriptive Statistics: Cphn 9-12 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	68,023	186	478	374.84	35.24
10	43,857	193	478	377.70	32.53
11	34,826	200	478	381.83	32.45
12	26,346	208	478	383.16	33.47
Total	173,052	186	478	378.24	33.91

Table 8.5.7B

Proficiency Level Distribution: Cphn 9-12 S301

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	6,382	9.4%	3,492	8.0%	2,890	8.3%	2,537	9.6%	15,301	8.8%
2	12,313	18.1%	9,969	22.7%	7,962	22.9%	5,622	21.3%	35,866	20.7%
3	16,894	24.8%	10,540	24.0%	7,241	20.8%	5,089	19.3%	39,764	23.0%
4	11,381	16.7%	8,200	18.7%	6,609	19.0%	5,858	22.2%	32,048	18.5%
5	12,580	18.5%	7,194	16.4%	6,115	17.6%	4,216	16.0%	30,105	17.4%
6	8,473	12.5%	4,462	10.2%	4,009	11.5%	3,024	11.5%	19,968	11.5%
Total	68,023	100.0%	43,857	100.0%	34,826	100.0%	26,346	100.0%	173,052	100.0%

Table 8.5.7C

n/a

Figure 8.5.7C

n/a

Figure 8.5.7D

n/a

Table 8.5.7D

Comprehension Composite Reliability: Cphn 9-12 S301

Component	Weight	Variance	Reliability
Listening	0.30	2357.730	0.704
Reading	0.70	996.935	0.796
Comprehension		1148.476	0.859

*Variances from students who had results in all four domains

Table 8.5.7E-1

Accuracy and Consistency of Classification Indices: Cphn (Grade 9) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.508	0.422		0.296	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.812		0.689	
	2	0.645		0.511	
	3	0.559		0.442	
	4	0.351		0.266	
	5	0.398		0.343	
	6	0.596		0.386	
Indices at Cut Points			**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.963	0.017	0.020	0.945
	2/3	0.905	0.042	0.053	0.867
	3/4	0.857	0.053	0.089	0.806
	4/5	0.853	0.054	0.093	0.792
	5/6	0.877	0.119	0.004	0.851

Table 8.5.7E-2

Accuracy and Consistency of Classification Indices: Cphn (Grade 10) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.521	0.430		0.300	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.766		0.625	
	2	0.696		0.569	
	3	0.524		0.418	
	4	0.403		0.308	
	5	0.403		0.341	
	6	-		0.396	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.963	0.019	0.019	0.944
	2/3	0.891	0.046	0.063	0.850
	3/4	0.859	0.055	0.087	0.807
	4/5	0.869	0.053	0.078	0.810
	5/6	0.898	0.102	0.000	0.883

Table 8.5.7E-3

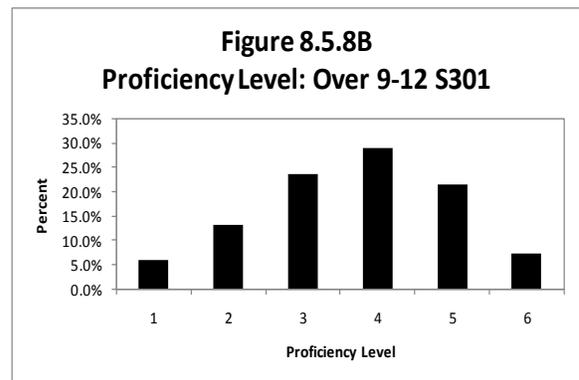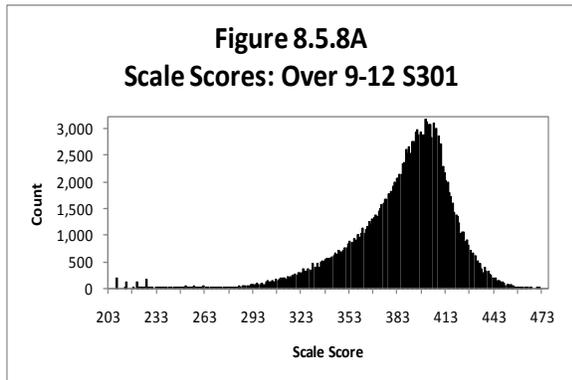Accuracy and Consistency of Classification Indices: Cphn (Grade 11) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.496	0.409		0.279	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.757		0.621	
	2	0.700		0.573	
	3	0.468		0.363	
	4	0.381		0.290	
	5	0.388		0.333	
	6	-		0.346	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.962	0.021	0.017	0.942
	2/3	0.889	0.044	0.067	0.849
	3/4	0.856	0.047	0.097	0.804
	4/5	0.848	0.052	0.099	0.784
	5/6	0.885	0.115	0.000	0.859

Table 8.5.7E-4

Accuracy and Consistency of Classification Indices: Cphn (Grade 12) S301

Overall Indices	Accuracy	Consistency		Kappa (k)		
	0.454	0.377		0.242		
Conditional on Level	**Level**	**Accuracy**		**Consistency**		
	1	0.761		0.639		
	2	0.693		0.555		
	3	0.414		0.305		
	4	0.374		0.305		
	5	0.321		0.278		
	6	-		0.257		
Indices at Cut Points		**Accuracy**				
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**	
	1/2	0.959	0.025	0.016	0.937	
	2/3	0.888	0.036	0.076	0.852	
	3/4	0.839	0.038	0.123	0.780	
	4/5	0.794	0.060	0.146	0.731	
	5/6	0.885	0.115	0.000	0.852	

8.5.8 Overall Composite 9-12

Figure 8.5.8A
Scale Scores: Over 9-12 S301

Figure 8.5.8B
Proficiency Level: Over 9-12 S301

Table 8.5.8A

Scale Score Descriptive Statistics: Over 9-12 S301

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	67,534	208	466	383.24	33.71
10	43,566	214	466	385.55	29.94
11	34,568	220	470	389.35	28.93
12	26,068	226	471	390.61	29.67
Total	171,736	208	471	386.17	31.38

Table 8.5.8B

Proficiency Level Distribution: Over 9-12 S301

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	4,750	7.0%	2,171	5.0%	1,697	4.9%	1,551	5.9%	10,169	5.9%
2	7,938	11.8%	6,278	14.4%	4,736	13.7%	3,527	13.5%	22,479	13.1%
3	13,037	19.3%	10,533	24.2%	9,109	26.4%	7,622	29.2%	40,301	23.5%
4	17,580	26.0%	12,913	29.6%	10,750	31.1%	8,390	32.2%	49,633	28.9%
5	17,639	26.1%	9,105	20.9%	6,289	18.2%	3,788	14.5%	36,821	21.4%
6	6,590	9.8%	2,566	5.9%	1,987	5.7%	1,190	4.6%	12,333	7.2%
Total	67,534	100.0%	43,566	100.0%	34,568	100.0%	26,068	100.0%	171,736	100.0%

Table 8.5.8C
n/a

Figure 8.5.8C
n/a

Figure 8.5.8D
n/a

Table 8.5.8D
Overall Composite Reliability: Over 9-12 S301

Component	Weight	Variance	Reliability
Listening	0.15	2357.730	0.704
Reading	0.35	996.935	0.796
Speaking	0.15	2802.444	0.918
Writing	0.35	966.316	0.910
Overall Composite		984.824	0.943

Table 8.5.8E-1
Accuracy and Consistency of Classification Indices: Over (Grade 9) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.678	0.585		0.477	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.884		0.815	
	2	0.763		0.658	
	3	0.713		0.602	
	4	0.698		0.576	
	5	0.586		0.543	
	6	-		0.390	
Indices at Cut Points			Accuracy		
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.982	0.008	0.010	0.975
	2/3	0.961	0.018	0.021	0.944
	3/4	0.929	0.034	0.037	0.900
	4/5	0.902	0.035	0.063	0.862
	5/6	0.902	0.098	0.000	0.888

Table 8.5.8E-2

Accuracy and Consistency of Classification Indices: Over (Grade 10) S301

Overall	Accuracy	Consistency		Kappa (k)	
Indices	0.714	0.622		0.515	
Conditional	Level	Accuracy		Consistency	
on Level	1	0.842		0.753	
	2	0.802		0.708	
	3	0.746		0.646	
	4	0.729		0.622	
	5	0.607		0.549	
	6	-		0.381	
Indices at		Accuracy			
Cut Points			False	False	
	Cut Point	Accuracy	Positives	Negatives	Consistency
	1/2	0.985	0.008	0.007	0.978
	2/3	0.955	0.020	0.025	0.936
	3/4	0.922	0.036	0.042	0.890
	4/5	0.911	0.035	0.054	0.873
	5/6	0.941	0.059	0.000	0.936

Table 8.5.8E-3

Accuracy and Consistency of Classification Indices: Over (Grade 11) S301

Overall	Accuracy	Consistency		Kappa (k)	
Indices	0.696	0.604		0.488	
Conditional	Level	Accuracy		Consistency	
on Level	1	0.839		0.753	
	2	0.793		0.692	
	3	0.781		0.685	
	4	0.710		0.585	
	5	0.533		0.477	
	6	-		0.298	
Indices at		Accuracy			
Cut Points			False	False	
	Cut Point	Accuracy	Positives	Negatives	Consistency
	1/2	0.985	0.008	0.006	0.979
	2/3	0.955	0.021	0.024	0.936
	3/4	0.920	0.031	0.049	0.889
	4/5	0.892	0.037	0.072	0.844
	5/6	0.943	0.057	0.000	0.940

Table 8.5.8E-4

Accuracy and Consistency of Classification Indices: Over (Grade 12) S301

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.655	0.576		0.438	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.837		0.764	
	2	0.785		0.678	
	3	0.814		0.711	
	4	0.535		0.516	
	5	-		0.341	
	6	-		0.555	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.984	0.010	0.006	0.976
	2/3	0.953	0.022	0.025	0.934
	3/4	0.908	0.021	0.071	0.875
	4/5	0.809	0.191	0.000	0.798
	5/6	0.954	0.046	0.000	0.954

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.

Bauman, J., Boals, T., Cranley, E., Gottlieb, M., and Kenyon, D.M. (2007). The Newly Developed English Language Tests (World-Class Instructional Design and Assessment – WIDA). In Abedi, Jamal (Ed.), *English Language Proficiency Assessment in the Nation: Current Status and Future Practice.* Davis: University of California.

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

Cook, H. G. (2007). Alignment Study Report: The WIDA Consortium's English Language Proficiency Standards for English Language Learners in Kindergarten through Grade 12 to ACCESS for ELLs® Assessment. Madison, WI: WIDA Consortium.

Cook, H. G., Boals, T., Wilmes, C., and Santos, M. (2007). Issues in the Development of Annual Measurable Achievement Objectives (AMAOs) for WIDA Consortium States. Madison, WI: WIDA Consortium.

Cronbach, L.J., Schonemann, P., & McKie, D. (1965). Alpha coefficients for stratified-parallel tests. *Educational and Psychological Measurement, 25,* 291-312.

Displacement measures. (2006, January 29). Retrieved December 4, 2006, from http://www.winsteps.com/winman/displacement.htm.

Fairbairn, S., & Fox, J. (2009). Inclusive achievement testing for linguistically and culturally diverse test takers: Essential considerations for test developers and decision makers**.** *Educational Measurement: Issues and Practice*, *28*(1), 10–24.

Fox, J. (2011). Test review: ACCESS for ELLs®. *Language Testing 28* (3): 425-431.

Gottlieb, M. (2004). English Language Proficiency Standards for English Language Learners in Kindergarten through Grade 12: Framework for Large-Scale State and Classroom Assessment. Madison, WI: WIDA Consortium.

Gottlieb, M. (2012). *ACCESS for ELLs® Interpretive Guide for Score Reports* (WIDA Consortium). Madison, WI: The Board of Regents of the University of Wisconsin System.

Gottlieb, M., and Boals, T. (2005). Considerations in Reconfiguring Cohorts and Resetting Annual Measurable Achievement Objectives (AMAOs) based on ACCESS for ELLs® Data (WIDA Consortium Technical Report No. 3).

Gottlieb, M. and Kenyon, D.M. (2006). The Bridge Study between Tests of English Language Proficiency and ACCESS for ELLs® (WIDA Consortium Technical Report No. 2).

Kane, M., & Case, S.M. (2004). The reliability and validity of weighted composite scores. *Applied Measurement in Education, 17, 221-240.*

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.

Keng, L., Miller, G.E., O'Malley, K. & Turhan, A. (2008). A Generalization of Stratified Alpha that Allows for Correlated Measurement Errors between Subtests. Paper presented at the annual meeting of the American Educational Research Association, New York City, New York.

Kenyon, D.M. (2006). *Development and Field Test of ACCESS for ELLs*[®] (WIDA Consortium Technical Report No. 1).

Kenyon, D.M., MacGregor, D., Li, D., and Cook, H. G. (2011). Issues in vertical scaling of a K-12 English language proficiency test. *Language Testing 28* (3): 383-400.

Kenyon, D.M., MacGregor, D., Louguit, M. Cho, B., and Ryu, J.R. (Willow), (2007). *Annual Technical Report for ACCESS for ELLs*[®] *English Language Proficiency Test, Series 101, 2005-2006 Administration* (WIDA Consortium Annual Technical Report No. 2).

Kenyon, D.M., MacGregor, D., Ryu, J.R. (Willow), Cho, B., and Louguit, M. (2006). *Annual Technical Report for ACCESS for ELLs*[®] *English Language Proficiency Test, Series 100, 2004-2005 Administration* (WIDA Consortium Annual Technical Report No. 1)

Kenyon, D.M., Ryu, J.R. (Willow), and MacGregor, D. (2013). *Setting Grade Level Cut Scores for ACCESS for ELLs*[®]. (WIDA Consortium Technical Report No. 4).

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 Rasch analysis* ver. 3.60.1, date 02-02-2006 [Computer Program]. www.winsteps.com

Livingston, S.A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32, 179-197.*

MacGregor, D., Kenyon, D.M., Gibson, S., and Evans, E. (2009). *Development and Field Test of Kindergarten ACCESS for ELLs*[®]. (WIDA Consortium).

MacGregor, D., Louguit, M., Yanosky, T., Fidelman, C. G., Pan, M., Huang, X., and Kenyon, D.M. (2010). *Annual Technical Report for ACCESS for ELLs*[®] *English Language Proficiency Test, Series 200, 2008-2009 Administration* (WIDA Consortium Annual Technical Report No. 5).

MacGregor, D., Louguit, M., Huang, X., and Kenyon, D.M. (2009). Annual Technical Report for ACCESS for ELLs[®] English Language Proficiency Test, Series 103, 2007-2008 Administration (WIDA Consortium Annual Technical Report No. 4).

MacGregor, D., Louguit, M., Ryu, J.R. (Willow), Kenyon, D.M., and Li, D. (2008). *Annual Technical Report for ACCESS for ELLs*[®] *English Language Proficiency Test, Series 102, 2006-2007 Administration* (WIDA Consortium Annual Technical Report No. 3).

Rudner, L. (2001, Spring). Informed test component weighting. *Educational Measurement: Issues and Practice, 20:1, 16-19.*

National Research Council. (2011). Allocating federal funds for state programs for English language learners. Panel to Review Alternative Data Sources for the Limited-English Proficiency Allocation Formula under Title III, Part A, Elementary and Secondary Education Act, Committee on National Statistics and Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Parker, C. E., Louie, J., and O'Dwyer, L. (2009). New measures of English language proficiency and their relationship to performance on large-scale content assessments (Issues & Answers Report, REL 2009–No. 066). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands. Retrieved from http://ies.ed.gov/ncee/edlabs, January 29, 2009.

Römhild, A., Kenyon, D. M., and MacGregor, D. (2011). Exploring domain-general and domain-specific linguistic knowledge in the assessment of academic English language proficiency. *Language Assessment Quarterly*, 8:213-228.

Waller, N.G. (n.d.). EZDIF: a computer program for detecting uniform and nonuniform differential item functioning with the Mantel-Haenszel and logistic regression procedures. [Computer Software]. Davis, CA: University of California – Davis.

Wang, H., Choi, I., Schmidgall, J., & Bachman, L. F. (2012). Review of Pearson Test of English Academic: Building an assessment use argument. *Language Testing, 29*(4), 603-619.

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). *2012 Amplification of the English Language Development Standards Kindergarten–Grade 12*. Madison, Wisconsin: Board of Regents of the University of Wisconsin System.

WIDA Consortium (2012b). *WIDA ACCESS for ELLs™ Test Administration Manual.* Retrieved from http://www.wida.us/assessment/ACCESS/#about.

WIDA Consortium. (2013). *Interpretive Guide for Score Reports Spring 2013* (WIDA Consortium). Madison, WI: The Board of Regents of the University of Wisconsin System.

Wright, B.D., & Stone, M.H. (1979). *Best test design: Rasch measurement.* Chicago, IL: MESA Press.

Wolf, M., Kao, J., Griffin, N., Herman, J., Bachman, P., Chang, S., and Farnsworth, T. (2008).

Yanosky, T., Amos, M., Louguit, M., Olson, Cameron, C., Louguit, M., MacGregor, D., Yen, S., and Kenyon, D.M. (2013). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test, Series 203, 2011-2012 Administration* (WIDA Consortium Annual Technical Report No. 8).

Yanosky, T., Chong, A., Louguit, M., Olson, E., Choi, Y., MacGregor, D., Yen, S., Cameron, C., and Kenyon, D.M. (2012). *Annual Technical Report for ACCESS for ELLs*[®] *English Language Proficiency Test, Series 202, 2010-2011 Administration* (WIDA Consortium Annual Technical Report No. 7).

Yanosky, T., Yen, S., Louguit, M., MacGregor, D., Zhang, Y., and Kenyon, D.M. (2011). *Annual Technical Report for ACCESS for ELLs*[®] *English Language Proficiency Test, Series 201, 2009-2010 Administration* (WIDA Consortium Annual Technical Report No. 6).

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, 233-251.*

Acknowledgements

We would like to extend our appreciation to the many CAL staff members who have supported this work, including:

Melissa Amos, M.A.
Catherine Cameron, M.A.
Deepak Ebenezer, M.S.
Simon Fong, B.A.
Dale Ingram, B.A.
Mohammed Louguit, Ph.D.
Dorry M. Kenyon, Ph.D.
David MacGregor, Ph.D.
April Maddy, Ed.M.
Jennifer Renn, Ph.D.
Shauna Sweet, M.S.
Shu Jing Yen, Ph.D.
Xin Yu, M.A.