

World-Class Instructional Design and Assessment



**Annual Technical Report for
ACCESS for ELLs[®] English Language Proficiency Test,
Series 303, 2014–2015 Administration**

**Annual Technical Report No. 11
Volume 1 of 3: Description, Validity, and Student Results**

Prepared by:

Center for Applied Linguistics

CAL/WIDA Partnership Activities
Psychometrics/Research Team

March 13, 2016

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.
- Elizabeth Peña, Ph.D., Professor, Department of Communication Sciences and Disorders, University of Texas at Austin.

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 11th annual technical report on Assessing Comprehension and Communication in English State-to-State for English Language Learners (hereafter, 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, 2014).

ACCESS for ELLs serves two purposes: 1) To assess reliably and validly the English language development (ELD) of English language learners (ELLs) in Grades K–12 according to WIDA *2012 Amplification of the English Language Development Standards Kindergarten–Grade 12* (WIDA, 2012b); 2) To place students appropriately into the 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 11th series of the test, Series 303. Series 303 was administered during the academic year 2014–2015 in 36 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.

Beginning with ACCESS Series 302 (operational year 2013-2014), the ACCESS Listening test transitioned from a traditional test administrator-read script to a media-delivered format, played either from CD or from streaming audio available online, for all grade-level clusters except Kindergarten. For more information, see the *ACCESS for ELLs Series 302 Media-Based Listening Field Test Technical Brief* (Center for Applied Linguistics, 2015).

As in the previous annual technical reports, this report provides background to the test (Chapter 1). The current report has been modified for Series 303 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 303, 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 303 test forms to those of Series 302. 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 consists 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 interpret and use ACCESS for ELLs. A larger, forthcoming (as of 2016), validation framework for the complete assessment from its inception to its consequences is currently under development by WIDA.

Demographic data

The Series 303 data set for analyses included the results of 1,569,136 students. The largest grade was Grade 1 with 223,097 students, while the smallest was Grade 12 with 34,526 students. Of the participating WIDA states, the largest was Illinois with 188,798 students, while the smallest was Northern Mariana Islands with 1,360 students. Technical analyses in this report are based on the performance of all students who were administered Series 303 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. (p. 3)

Results indicate that the reliability of the Overall Composite score for Series 303, presented in Chapter 8 Table D, is very high across all grade-level clusters. For Kindergarten it was .974; for Grades 1–2, .941; for Grades 3–5, .935; for Grades 6–8, .934; and for Grades 9–12, .946. 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 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)	0.978	0.961	0.951	0.952	0.946
K (accountability)	0.954	0.951	0.953	0.949	0.991
1	0.982	0.938	0.918	0.961	0.991
2	0.990	0.963	0.913	0.931	0.980
3	0.996	0.984	0.941	0.907	0.921
4	0.993	0.979	0.933	0.900	0.906
5	0.991	0.974	0.925	0.894	0.923
6	0.989	0.963	0.915	0.893	0.977
7	0.986	0.960	0.910	0.883	0.981
8	0.982	0.955	0.908	0.884	0.984
9	0.976	0.956	0.932	0.911	0.922
10	0.982	0.955	0.926	0.912	0.928
11	0.983	0.955	0.922	0.910	0.926
12	0.984	0.954	0.918	0.866	0.943

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 one-third to one-half of all ACCESS for ELLs test items. In academic year 2014–2015, Series 303 was administered in 36 WIDA Consortium states. The 44 test forms in Listening, Reading, Writing, and Speaking from Kindergarten to Grade 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, excluding 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 303. Data come from the 1,569,136 students who were administered the test operationally in the winter and spring of 2015.

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 303 (2014–2015) 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: 2014-2015

Technical Reports

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

Center for Applied Linguistics (2015). *ACCESS for ELLs Series 302 Media-Based Listening Field Test Technical Brief*. (WIDA Consortium).

This report provides detailed information on the conceptualization, development, and field testing of the ACCESS for ELLs Media-Based Listening Test.

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., & 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 ELLs, listed by year of publication. These reports provide extensive analysis of the results from the operational administrations 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., Cho, B., & 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., & Ryu, J. R. (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., Kenyon, D. M., & 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., & 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., & 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., & 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., . . . 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., Cameron, C., Louguit, M., MacGregor, D., Yen, S., & 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).
- Center for Applied Linguistics (2014). *Annual Technical Report for ACCESS for ELLs[®] English Language Proficiency Test, Series 301, 2012–2013 Administration* (WIDA Consortium Annual Technical Report No. 9).
- Center for Applied Linguistics (2015). *Annual Technical Report for ACCESS for ELLs[®] English Language Proficiency Test, Series 302, 2013–2014 Administration* (WIDA Consortium Annual Technical Report No. 10).

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., & Kenyon, D. M. (2007). The Newly Developed English Language Tests (World-Class Instructional Design and Assessment – WIDA). In J. Abedi (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.K. & 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.

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

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.

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., & 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. & Fairbairn, S. (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., & 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*. 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., & 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., & 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 2014. 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., & Farnsworth, T. (2008). *Issues in assessing English language learners: English language proficiency measures and accommodation uses—Practice review (Part 2 of 3)* (CRESST Report 732). Los Angeles, CA: 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 2005–2006, 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 (Eds.), *Differential item functioning* (pp. 337-347). Hillsdale, NJ: Lawrence Erlbaum Associates.

This book chapter describes procedures for conducting DIF analysis.

Table of Contents
Volume 1

<i>The WIDA ACCESS for ELLs Technical Advisory Committee</i>	<i>ii</i>
<i>Executive Summary</i>	<i>iii</i>
<i>Summary Highlights</i>	<i>iii</i>
<i>Annotated Bibliography: 2014-2015</i>	<i>vii</i>
<i>Technical Reports</i>	<i>vii</i>
<i>Annual Technical Reports for ACCESS for ELLs</i>	<i>viii</i>
<i>Other Documentation</i>	<i>ix</i>
<i>1. Description of ACCESS for ELLs English Language Proficiency Test</i>	<i>1</i>
<i>1.1 Purpose of ACCESS for ELLs</i>	<i>1</i>
<i>1.2 Format of ACCESS for ELLs</i>	<i>1</i>
1.2.1 Integration with the Standards	<i>1</i>
1.2.2 Grade-Level Clusters	<i>2</i>
1.2.3 Language Domains	<i>2</i>
1.2.4 Language Proficiency Levels.....	<i>2</i>
1.2.5 Tiers	<i>4</i>
<i>1.3 Test Development</i>	<i>5</i>
1.3.1 Field Test	<i>5</i>
1.3.2 Equating and Scaling	<i>7</i>
1.3.3 Standard Setting.....	<i>8</i>
<i>1.4 Ongoing Item Development</i>	<i>10</i>
1.4.1 Item Writing and Editing	<i>11</i>
1.4.2 Item Content and Bias and Sensitivity Reviews	<i>11</i>
1.4.3 Item Field Testing.....	<i>11</i>
1.4.4 Item Calibration and Analysis	<i>11</i>
1.4.5 DIF Items	<i>12</i>
<i>1.5 Reporting of Results</i>	<i>13</i>
1.5.1 Scale Scores	<i>13</i>
1.5.2 Language Proficiency Level Scores.....	<i>14</i>
1.5.3 Results by English Language Development Standards	<i>18</i>
<i>1.6 Test Administration</i>	<i>18</i>
1.6.1 Test Administrator Training	<i>18</i>
1.6.2 Test Security	<i>18</i>
1.6.3 Test Accommodations	<i>19</i>
<i>1.7 Scoring</i>	<i>19</i>
1.7.1 Listening and Reading	<i>19</i>
1.7.2 Writing.....	<i>19</i>
1.7.2.1 Scoring Procedures for Writing	<i>22</i>

1.7.3	Speaking.....	24
1.7.3.1	Training Procedures for Scoring Speaking.....	25
2.	<i>An Assessment Use Argument for ACCESS for ELLs: Focus on Assessment Records</i>	27
2.1	<i>The Generic Validation Framework for ACCESS</i>	28
2.2	<i>Focus on Assessment Records</i>	29
2.2.1	Breakdown of Claims for the <i>Assessment Records</i> Produced in the ACCESS for ELLs Assessment Program.....	30
2.3	<i>Evidence for Assessment Records Claims of ACCESS for ELLs</i>	32
2.4	<i>Summary of Assessment Records Claims, Actions, and Evidence</i>	39
2.5	<i>Visual Guide to Tables and Figures</i>	41
2.5.1	Chapter 4 Visual Guide to Tables and Figures.....	42
2.5.2	Chapter 6 Visual Guide to Tables and Figures.....	43
2.5.3	Chapter 8 Visual Guide to Tables and Figures.....	44
3.	<i>Descriptions of Student Results</i>	45
3.1	<i>Participation</i>	45
3.1.1	Grade-Level Cluster.....	45
3.1.2	Grade.....	45
3.1.3	Tier.....	45
3.2	<i>Scale Score Results</i>	47
3.2.1	Mean Scale Scores Across Domain and Composite Scores Section.....	47
3.2.2	Correlations.....	47
3.3	<i>Proficiency Level Results</i>	48
3.4	<i>Mean Raw Score Results by Standards</i>	48
3.4.1	Comprehension Composite.....	48
3.4.2	Writing.....	49
3.4.3	Speaking.....	49
4.	<i>Student Results</i>	50
4.1	<i>Participation</i>	50
4.1.1	Participation by Grade-Level Cluster.....	50
4.1.1.1	By State.....	50
4.1.1.2	By Gender.....	51
4.1.1.3	By Ethnicity.....	51
4.1.2	Participation by Grade.....	52
4.1.2.1	By State.....	52
4.1.2.2	By Gender.....	53
4.1.2.3	By Ethnicity.....	54
4.1.3	Participation by Tier.....	55
4.1.3.1	By Cluster by Domain (Test Form).....	55
4.1.3.2	By Grade by Domain (Test Form).....	56
4.1.3.3	By Cluster by Gender.....	57
4.1.3.4	By Cluster by Ethnicity.....	58

4.2	<i>Scale Score Results</i>	59
4.2.1	Mean Scale Scores by Grade-Level Cluster Across Domain and Composite Scores	59
4.2.1.1	By Cluster	59
4.2.1.2	By Cluster by Gender.....	60
4.2.1.3	By Cluster by Ethnicity.....	61
4.2.2	Mean Scale Scores by Grade Across Domain and Composite Scores.....	64
4.2.2.1	By Grade	64
4.2.2.2	By Grade by Gender	65
4.2.2.3	By Grade by Ethnicity	67
4.2.3	Correlations among Scale Scores by Grade-Level Cluster.....	72
4.3	<i>Proficiency Level Results</i>	74
4.3.1	Listening	74
4.3.1.1	By Cluster by Tier.....	74
4.3.1.2	By Grade by Tier	75
4.3.1.3	By Grade	77
4.3.2	Reading	78
4.3.2.1	By Cluster by Tier.....	78
4.3.2.2	By Grade by Tier	79
4.3.2.3	By Grade	81
4.3.3	Writing.....	82
4.3.3.1	By Cluster by Tier.....	82
4.3.3.2	By Grade by Tier	83
4.3.3.3	By Grade	85
4.3.4	Speaking.....	86
4.3.4.1	By Cluster by Tier.....	86
4.3.4.2	By Grade by Tier	87
4.3.4.3	By Grade	89
4.3.5	Oral Language Composite	90
4.3.5.1	By Cluster by Tier.....	90
4.3.5.2	By Grade by Tier	91
4.3.5.3	By Grade	93
4.3.6	Literacy Composite.....	94
4.3.6.1	By Cluster by Tier.....	94
4.3.6.2	By Grade by Tier	95
4.3.6.3	By Grade	97
4.3.7	Comprehension Composite.....	98
4.3.7.1	By Cluster by Tier.....	98
4.3.7.2	By Grade by Tier	99
4.3.7.3	By Grade	101
4.3.8	Overall Composite	102
4.3.8.1	By Cluster by Tier.....	102
4.3.8.2	By Grade by Tier	103
4.3.8.3	By Grade	105
4.4	<i>Mean Raw Score Results by Standards</i>	106
4.4.1	Comprehension Composite.....	106
4.4.1.1	By Cluster	106

4.4.1.2	By Grade	108
4.4.2	Writing	114
4.4.2.1	By Cluster	114
4.4.2.2	By Grade	115
4.4.3	Speaking.....	118
4.4.3.1	By Cluster	118
4.4.3.2	By Grade	119

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 (ELLs) in Grades K–12 in the United States as defined by the multi-state WIDA Consortium, first in the English Language Proficiency Standards (2004, 2007), then in the amplified 2012 English Language Development Standards (hereafter, 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 ELLs and they 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 the social and academic language proficiency of ELLs 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 language instructional educational 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—ELLs communicate in English for **Social and Instructional** purposes within the school setting.

Standard 2—ELLs communicate information, ideas, and concepts necessary for academic success in the content area of **Language Arts**.

Standard 3—ELLs communicate information, ideas, and concepts necessary for academic success in the content area of **Mathematics**.

Standard 4—ELLs communicate information, ideas, and concepts necessary for academic success in the content area of **Science**.

Standard 5—ELLs communicate information, ideas, and concepts necessary for academic success in the content area of **Social Studies**.

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

- Social and Instructional language: 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. In the case of some test items and tasks, the standards are combined as follows:

- Integrated Language of Science, Language of Language Arts, and Language of Social Studies: IT
- Language of Math and Language of Science: MS
- Language of English Language Arts and Language of Social Studies: LS

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 one section assessing each of these four language domains on the ACCESS for ELLs test.

1.2.4 Language Proficiency Levels

The WIDA ELD Standards document fully delineates the continuum of language development with five language proficiency levels. 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.

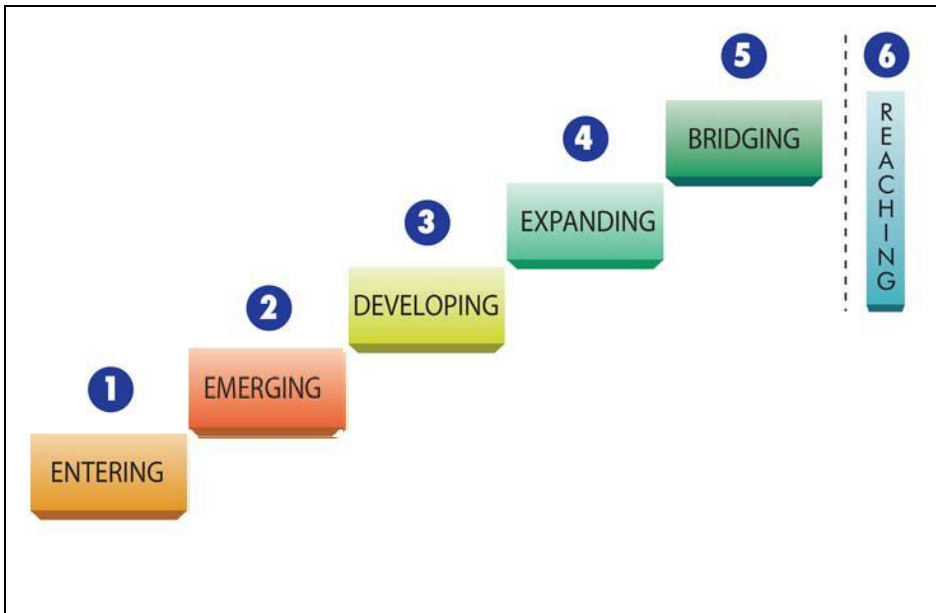


Figure 1.2.4A. The language proficiency levels of the WIDA ELD Standards

These language proficiency levels are 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. As general definitions that are 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; and the final criterion is the increasing development of phonological, syntactic, and semantic understanding in receptive skills or control of 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 in each of the five **Standards**, at five different **grade-level clusters**, across four **language domains**. 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. The grouping of five PIs in logical progression is called a “strand.” There is a total of 80 strands.

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

1.2.5 Tiers

Obviously, test items and tasks that allow Entering (Level 1) or Emerging (Level 2) students to demonstrate accomplishment of the PIs at their proficiency level will not allow Expanding (Level 4) or Bridging (Level 5) students to demonstrate the full extent of their language proficiency. Likewise, items and tasks that 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; items that are far too difficult for test takers may be frustrating and discourage them from performing their best. But more importantly, items that are too easy or too hard for a student add very little to the accuracy or quality of the measurement of that student’s proficiency. Tests need to be at the appropriate difficulty level for individual test takers in order to be valid and reliable.

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

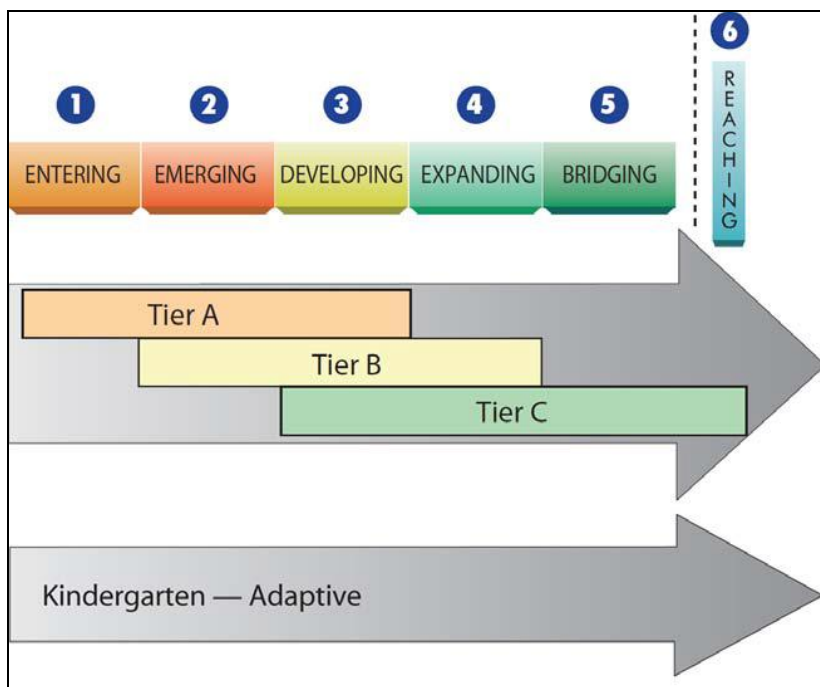


Figure 1.2.5A. Tier structure of ACCESS for ELLs

Thus, Tier A has items and tasks designed to allow students at the lowest language proficiency levels (Levels 1 and 2) to meet the WIDA ELD Standards at their language proficiency levels, and it includes some items targeted to Level 3. Likewise, Tier C has items and tasks designed to allow students at the highest language proficiency levels (Levels 4 and 5) to meet the WIDA

ELD Standards at their language proficiency levels, while also containing some items targeted to Level 3. In this test design, the tiers overlap: while Tier A and Tier C have 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 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 meet 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 meet 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 for 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. Within the sample, 72.3% 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 Section 1.2.5, for a total of 52 tasks. Field testing for Speaking was conducted in Wisconsin and Washington, D.C. 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 Washington, D.C. 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 ordered from easiest to hardest on the Kindergarten assessment. The Speaking portion of the Kindergarten assessment was the same as that of the 1–2 grade-level cluster, except it included only SI and LA/SS folders, in order to reduce testing time. The Writing portion of the Kindergarten assessment included very simple writing tasks that 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 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 test.)

A total of 154 students participated in the Kindergarten field test. Of those, 55% were boys (84 students) and 45% were girls (70 students). Spanish speakers comprised 90.2% (139) of the sample; 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 must be reported on a standard scale that is familiar to test users and that retains its 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 have the same meaning regardless of which test items the test taker is exposed to on the assessment.

A 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 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 must fall between these extremes.

It should be noted 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 its criticisms—as one climbs the scale from grade level to grade level (or grade-level cluster to grade-level cluster in the case of ACCESS for ELLs), the scale adjusts for 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, which is discussed in Section 1.3.3. We focus on the creation of the ACCESS for ELLs scale score here.

The procedure for developing the scale was complex but the steps involved were fairly basic. 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 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 contained common folders of items (in the case of Listening and Reading) or common tasks (in the case of Writing), 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 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 means of scaling the Writing tasks, and performances on the Listening items were used as a means of scaling the Speaking tasks.

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

Developing the Logit Scale: A calibration of the ability of the students and the difficulty of the items using Rasch procedures was then applied to these data matrices, putting student ability and item difficulty 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 negative infinity to positive 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 transformed into 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, Wisconsin between April 20 and 27,

2005. The purpose of the study was not to set new standards on ACCESS for ELLs, per se. Rather, the purpose was to use the WIDA ELD¹ Standards together with empirical information from the field test data to conduct a defensible and replicable approach to determining 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. 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. There were 20–22 teachers or administrators on each panel who were deemed qualified to participate in the study by WIDA. 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 work independently, reading through the items and bookmarking the item that they determined a student at Language Proficiency Level 1 would have a 50% chance of answering correctly. They were then asked to repeat this procedure for all levels up to Level 5.

After the initial round of bookmarking, the results were compiled and discussed with the panelists as a group. The panelists were then 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 the written responses from a single student's test. Student portfolios were selected from each tier, and an attempt was made to select 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 student work that had received the lowest total raw score across the four writing responses, and the last portfolio represented student work with a very high total raw score on the four pieces of writing responses.

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 responses represented a student at a given language proficiency level. For example, if they felt the portfolio represented the work of a 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 were then given the opportunity to reconsider and adjust their judgments, 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

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.

at which at least 50% of the panelists would be expected to agree that the writing responses would represent the work of the next higher proficiency level. The results from this analysis were used to set the cut scores for the language proficiency levels for Writing.

The procedure for Speaking was similar, requiring that the panelists listen to audio-recorded portfolios of spoken student responses, and record their judgments.

1.4 Ongoing Item Development

To ensure that ACCESS for ELLs remains secure, as well as to incorporate a program of continual refinement of the assessment, new items are developed and field tested every year. The field test is performed on separate forms during the operational administration of ACCESS for ELLs. In fact, one-third to one-half of the items are replaced yearly. The intent of ongoing item development is to replace all items or tasks in each test form completely over a three-year period.

The schedule for refreshing ACCESS for ELLs items is as follows²:

- Listening and Reading—All of the LA and MA items are replaced in alternating years, while the SC, SS, and SI items are replaced in a three-year cycle.
- Speaking—The SI task is replaced yearly, while the MS and LS tasks are replaced in alternating years.
- Writing—The IT task is replaced yearly, while the MS and LA tasks are replaced in a two-year cycle for Tier A, and the MS and SI tasks were replaced in a two-year cycle for Tiers B and C. (Note that starting with Series 201, the separate MA and SC folders were replaced with a combined MS folder.)

For Series 303, because all of the items on the Reading were refreshed, Listening was not refreshed. All of the items on the Writing were refreshed, except for Writing Tier A for grade cluster 1–2. Speaking was not refreshed.

Some changes have been adapted over the history of ACCESS for ELLs. Between Series 100 and Series 101, the Writing IT task and the Speaking MA/SC task were replaced. In Listening and Reading, various item folders were replaced following analysis of the field test and operational Series 100. Since ACCESS for ELLs was 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.

Beginning with Series 302, the Listening test transitioned from a traditional test administrator-read script to a media-delivered format, played either from CD or from streaming audio available online, for all grade-level clusters except for Kindergarten. For more information, see the *ACCESS for ELLs Series 302 Media-Based Listening Field Test Technical Brief* (Center for Applied Linguistics, 2015).

Starting in 2015-2016, an online-version of the test will be administered, and new items will be developed. Detailed description of the new item development will be reported in the forthcoming Annual Technical Report (Series 400, 2015-2016).

² This schedule applies to all grade-level clusters except Kindergarten, which was redesigned for Series 200 and is not refreshed annually.

1.4.1 Item Writing and Editing

The initial item writing is done by participants in an online item writing course conducted by CAL. Then, the items generated by that course are reviewed internally and selected 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. The chosen items are refined by CAL staff before undergoing an item content and bias and sensitivity reviews (see Section 1.4.2). Usually, some items require further revision before being sent to MetriTech and WIDA for final review. Once returned to CAL, the items are prepared for the field test.

1.4.2 Item Content and Bias and Sensitivity Reviews

After items are internally refined, they are reviewed by two panels: a content review panel and a bias and sensitivity review panel. 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 targeted grade-level cluster, and that each item or task matches the model performance indicators from the WIDA ELD Standards that it is intended to assess. After the items are revised based on comments from the content panel members, they are submitted to the bias and sensitivity review panel, which inspects the items for potential bias that may unfairly disadvantage some students over others. Bias and sensitivity 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 districts that accept the invitation 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 (nine 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 folders and one SI folder, while the other form may have one MA folder and two SI folders. For Writing, four tasks are prepared per grade-level cluster: one task at each tier for the grade's standard, and one IT task. Students are presented with just one task, and at the appropriate tier when possible. 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 on the next year's operational form. Only folders of items meeting all technical requirements are placed on the operational test 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 conducted, in addition to the DIF analysis conducted for the Annual Technical Report. Each operational 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 source is all student data available a week before the equating sample is pulled, called Early Return; the second source is the 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. For Series 303, four items were identified as having a C-level ethnicity DIF in the Early Return data: two favoring Hispanics and two favoring Non-Hispanics. The former exhibited A-level ethnicity DIF in the Equating Sample; therefore, no further action was required. However, the other two items favoring Non-Hispanics continued to exhibit C-level DIF in the Equating Sample and were reviewed by panelists. After a thorough investigation, it was decided that these two items would be scored operationally.

Phase II is conducted using all student data available in early May. During Phase II analysis, ethnicity and gender DIF are 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 for the next operational year. For Series 303, Phase II DIF analysis was not conducted because administration of ACCESS for ELLs 2.0 will begin in the next operational year (2015-2016, Series 400) with new items being developed.

For the Annual Technical Report, an ethnicity and gender DIF analysis is conducted using all student data. For Series 303, eight items showed DIF. Out of 270 Listening items, three (1.1%) showed C-level DIF based on ethnicity, favoring Hispanics. Out of 342 Reading items, two (0.6%) showed C-level DIF based on ethnicity, favoring Non-Hispanics. Out of 43 Writing tasks, two (4.7%) showed C-level DIF based on ethnicity, favoring Non-Hispanics. Out of 62 Speaking items, one (1.6%) showed C-level DIF based on ethnicity, favoring 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 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.

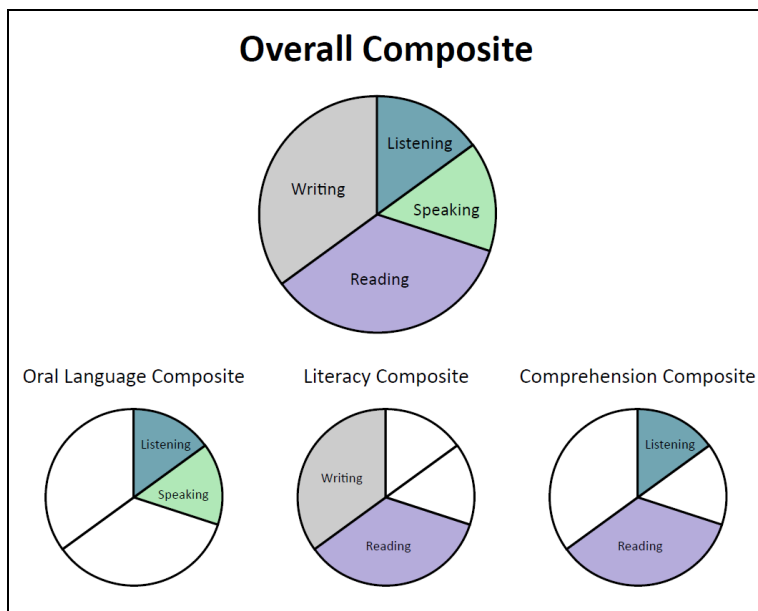


Figure 1.5.1A. Domain Composites

1.5.2 Language Proficiency Level Scores

In addition to the ACCESS for ELLs scale scores, test score users 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

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	323	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	346	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 the scale scores, which form an interval scale and are continuous across the grades from Kindergarten to Grade 12, 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.)

Because the width between cut scores varies, proficiency level scores should not be considered as forming an interval scale. That is, the distance between proficiency level scores 1.5 and 2.5 cannot be assumed to be equal to the distance between proficiency level scores 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., in Grade 3, the distance between 3.1 and 3.2 is the same as the distance between 3.7 and 3.8), but they do not form an interval scale across 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 ELD Standards (see Section 1.2.1 for more information about the Standards).

For Comprehension (combined Listening and Reading), the five ELD scores (Social and Instructional Language [SI], Language of Language Arts [LA], Language of Math [MA], Language of Science [SC], and Language of Social Studies [SS]) are reported as the number correct out of the 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, ELD 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 SI is 3; the maximum for LS and for MS is 5.

For Writing tasks, three ELD 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 303 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 be certified, participants had to pass a quiz after completing the course.

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 established policies and procedures regarding the security of the test, and

every individual involved in the administration of ACCESS for ELLs, from the district level 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 test taker also has an IEP, to the extent possible, the recommendations in the student's IEP are to be followed. The extent to which this was accomplished for Series 303 was a local decision during administration.

During the 2011–2012 testing cycle, the WIDA Consortium made available *Alternate ACCESS for ELLs* (hereafter, Alternate ACCESS). Alternate ACCESS is intended only for English language learners who have cognitive disabilities that are severe enough to prevent meaningful participation in ACCESS for ELLs testing, even 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 enter their answers directly into the test booklets, so each page is scanned into an electronic database.

1.7.2 Writing

Student responses to the Writing tasks are centrally scored at MetriTech by raters who are trained to follow the WIDA Consortium's Writing Rubric (see Section 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:

6 - Reaching	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• 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 on the generic rubric, scorers also receive training as to the 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 as one moves up the grade-level clusters. Thus, a single generic rubric rooted in the WIDA ELD Standards lies at the core of scoring for Writing, but developmental differences between grade-level clusters are part of the additional training that each rater receives.

During training, 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 excused from scoring operational Writing tests.

In applying the Writing Rubric, the following method of scoring 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 the writing displayed any particular weakness in any category (i.e., was lower in one of the three) or displayed any particular strength in any category (i.e., was higher in one of the three). 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 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, respectively. 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 online scoring system, MTscore. As with all aspects of ACCESS for ELLs, a key concern is the security of student data and the items and forms that elicit student responses. Some of the strict security measures implemented as part of MTscore include:

- All identifying or biographical data (including name, ID number, gender, etc.) of students will be stripped from scorer images and will not be included in data transferred into MTscore.
- 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, unless by master scorers, who may do so in cases of alert papers.

- Restrictions are in place for scoring session access, requiring scorer login during predetermined times and dates only.

Before scorer training begins, master scoring staff prepare group leaders and senior scorers for each content area, ensuring that they are familiar with the rubrics, annotated anchor papers, training sets, calibration sets, and scoring procedures. To qualify as a master scorer, MetriTech staff must have experience working with scoring protocols for various programs and states for a minimum of seven years. This core group works closely with the CAL test development staff, and annually augments the original training materials for ACCESS for ELLs scoring.

With scoring centers located near several universities, MetriTech has a large pool of qualified scoring applicants from which to choose. 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. Many applicants have backgrounds in education and are active or retired teachers.

Each potential scorer is selected to train on a particular grade-level cluster. The training process begins with an online training session, where each scorer reviews the rubrics, the elements of analytic scoring, and the anchor papers. Rubric score points are defined, and the scorer is presented with examples of student work that meet the criteria for each score point. Following this presentation, the scorers work through selected modules of sample papers. Each paper in a training module has already been reviewed and scored by a master scorer. As scorers finish the training module, their recorded scores and scoring rationale are systematically reviewed. Discrepancies are noted and feedback and additional modules are presented to the scorer as further training. Finally, each scorer is given a post-test module containing sample student responses. The scorer rates these modules independently, and the final scores that they assign are compared with those assigned by the master scorer. Scorers must reach the criterion of 70% exact agreement with the master scorer's rating to complete training and be approved to score operational test material. Training sessions utilizing online, self-paced training modules that are interactive; printed training manuals; and master scorer feedback; the session typically includes eight hours of material.

Once scorers are selected, they are supervised thusly during scoring sessions:

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 individual 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 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 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, for reporting in the technical report, and for 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 administered. 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. These 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 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).

1.7.3.1 Training Procedures for Scoring Speaking

The Speaking Test is the only portion of ACCESS for ELLs that is scored locally. Test administrators must complete the online ACCESS for ELLs Test Administrator training module for the Speaking Test and pass 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 rationale that are provided for each sample demonstrate how and why a particular score is assigned. To be certified to administer the ACCESS for ELLs Grades 1–12 Speaking test, test administrators must pass 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

One important factor in developing an assessment as a measurement tool is considering how to determine its validity. Validity is “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 2014, p. 11). Evaluations of test validity assess the evidence that supports the the interpretations and decisions made about test takers on the basis of their performance on a test, and the appropriateness and adequacy of such interpretations. A fully developed validation framework, including an Assessment Use Argument (AUA) (Bachman & Palmer, 2010), consists of several steps (described in Section 2.1 below) that connect test design and administration to intended and actual score interpretation and consequences. This chapter 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.

An argument-based approach to the ACCESS for ELLs validation framework organizes the information in the present report to support claims about Assessment Records (i.e., test scores and proficiency level descriptions collected via ACCESS for ELLs). Specifically, tables and figures from this report are explicitly linked to questions related to assessment data. Chapelle, Enright, & Jamieson (2010) support using such a structure to present information to assessment users because, “based on an analysis of four points of comparison—framing the intended score interpretation, outlining the essential research, structuring research results into a validity argument, and challenging the validity argument—we conclude that an argument-based approach to validity introduces some new and useful concepts and practices” (p. 3). A larger, though yet undocumented (as of 2014), validity argument for the complete assessment from its inception to its consequences is currently under development by WIDA.

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

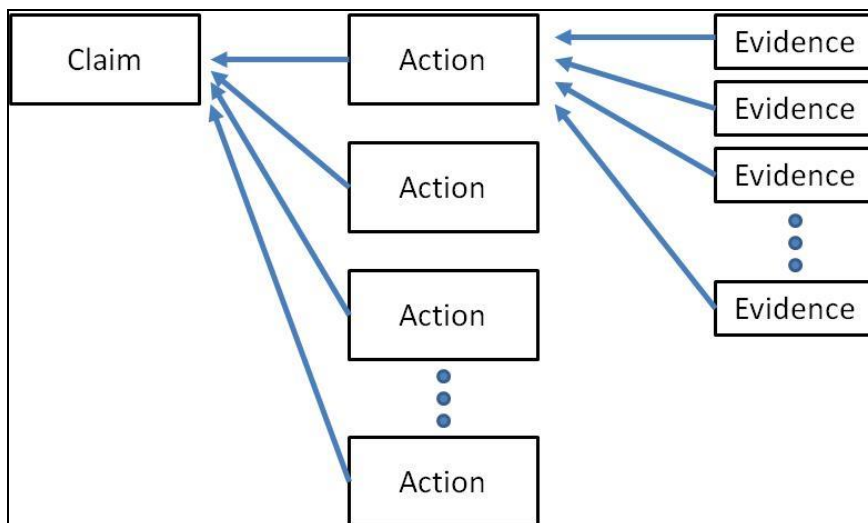


Figure 2A: General Argument Structure for Assessment Validation (simplified from Toulmin, 2003)

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) and is hereafter referred to as CAL’s Validation Framework. CAL’s Validation Framework, shown in Figure 2.1A, combines models for both test development (i.e., Evidence-Centered Design [Mislevy, Almond, & Lukas, 2004]) and assessment validation (i.e., the AUA from Bachman and Palmer [2010]) to cover the assessment development and implementation process from initial conceptualization to the score interpretations and consequences of using the assessment. This framework constantly looks both forward and backward, and each subsequent step depends upon the strength of the step below it; for this reason, the steps are numbered from seven to one. For example, during the initial *Plan* step, test developers state the anticipated decisions and consequences of implementing the assessment program, which are eventually investigated in *Decisions*, and *Consequences* represents the culmination of all previous steps. This structure highlights the fact that any weakness in a lower step affects the steps above it.

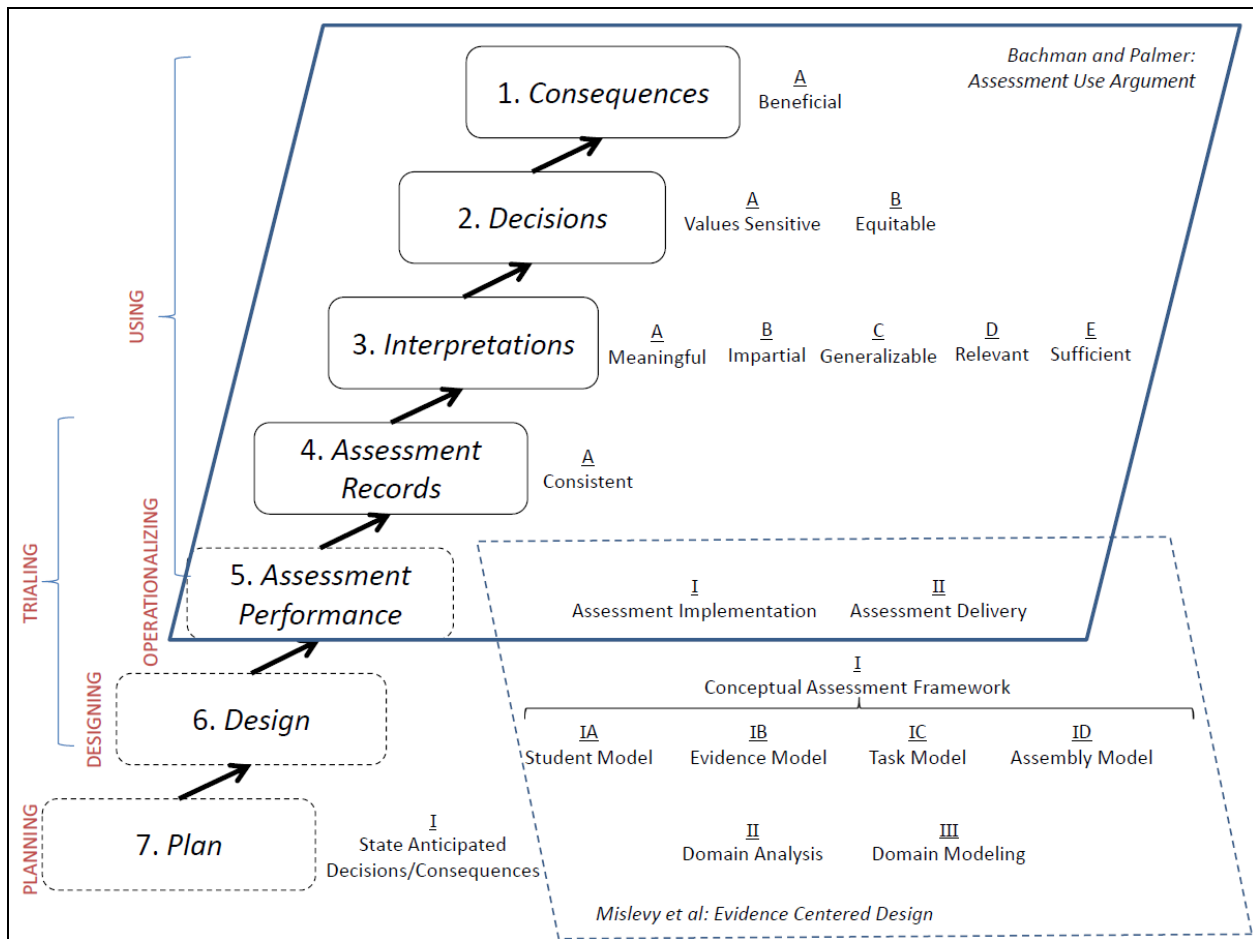


Figure 2.1A: CAL's Validation Framework (based on Bachman & Palmer, 2010; Mislevy, Almond, & Lukas, 2004)

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

The WIDA Consortium is using CAL's Validation Framework to present a complete validity argument, which will be updated as needed, for ACCESS for ELLs. To date, information related to *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 (see Figure 2.1A), the data presented in this document cover only *Assessment Records*. By focusing on Assessment Records (i.e., test scores and proficiency level descriptions), the information in the Annual Technical Report will be used to support claims related to the quality and consistency

of the assessment data gathered and analyzed using ACCESS for ELLs. The claims in this step of the AUA all pertain to the general question, “How do we know that the reported language domain scores and composite scores on ACCESS for ELLs are consistent and dependable?” Other questions about the development, administration, and outcomes 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 *Assessment Records* (Step 4), will fit into the complete, generic validation framework. 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.

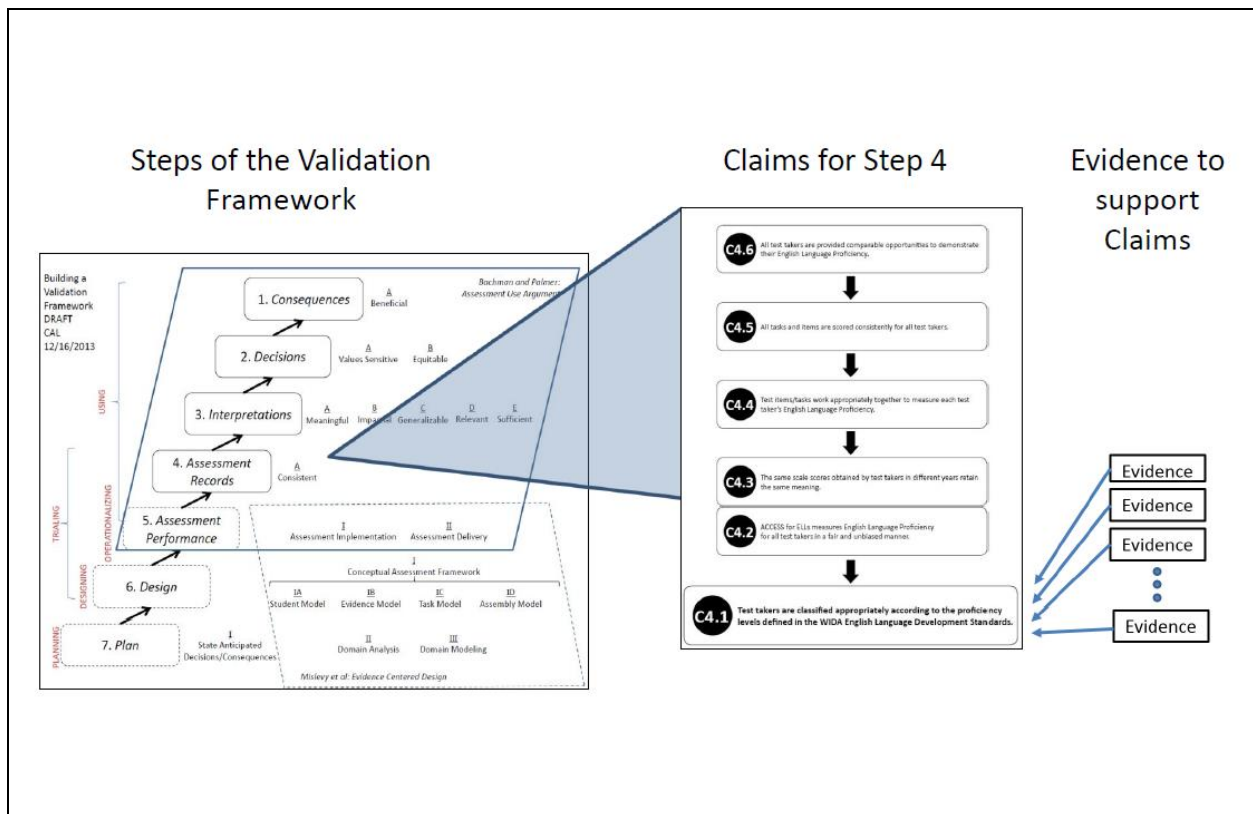


Figure 2.2A: Structure of the Argument-Based Approach Supporting Step 4 Contained in this Chapter

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

Assessment Records (Step 4) of the complete ACCESS for ELLs validation framework, is broken down into the following six claims:

- C4.6. All test takers are provided comparable opportunities to demonstrate their English Language Proficiency.
- C4.5. All tasks and items are scored consistently for all test takers.
- C4.4. Test items/tasks work appropriately together to measure each test taker’s English Language Proficiency.

- C4.3. The same scale scores obtained by test takers in different years retain the same meaning.
- C4.2. ACCESS for ELLs measures English Language Proficiency for all test takers in a fair and unbiased manner.
- C4.1. Test takers are classified appropriately according to the 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 (C4.6) down to (C4.1). 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.

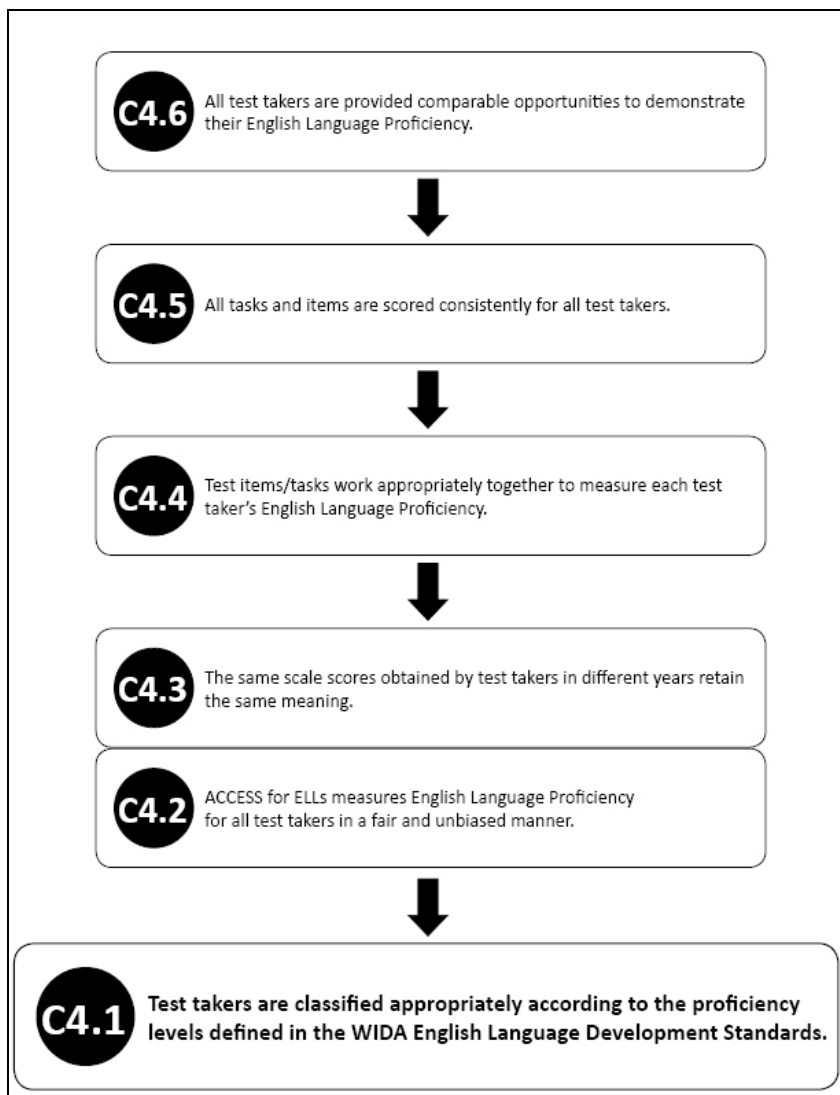


Figure 2.2.1A: Progression of Claims for Step 4: Assessment Records

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 *Assessment Records*, which is 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 corresponding 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 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 helps ensure that test takers within a range of the targeted proficiency level perform as expected. It is not as sensitive to outliers as outfit. Outfit can be skewed if test takers with extreme (i.e., high-level or low-level) proficiency do not perform as expected. High infit is a bigger threat to validity, but is more difficult to explain than high outfit (Linacre, 2002). The infit and outfit mean square statistics are part of the evaluation criteria used to select the items and tasks that appear on the final operational forms.

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 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, an item exhibiting B level DIF displays a moderate amount of bias, and an item exhibiting C level DIF is considered to display considerable evidence for potential bias and should be closely examined by test developers to identify any construct irrelevant factors that may contribute to DIF.

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

Evidence: As described in Chapter 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 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, that region of the score distribution usually has smaller conditional standard error values and more reliable scores.

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 intended 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 percentage 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.	a. Test Administration Manual
	b. Test administrators document and report any irregularities that may occur so that appropriate action may be taken.	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.	a. Chapter 1.7.2 (Scoring - Writing) ; Chapter 1.7.3 (Scoring - Speaking)
	b. Listening and Reading items are scored electronically onsite at MetriTech.	b. Chapter 1.7.1 (Scoring - Listening and Reading)
	c. Raters of performance-based tasks are certified, demonstrating that they can score appropriately.	c. Chapter 1.7.2 (Scoring - Writing) ; Chapter 1.7.3 (Scoring - Speaking)
	d. Raters of Writing tasks are monitored daily to ensure that they are scoring appropriately.	d. Chapter 1.7.2.1 (Scoring Procedures for Writing)
	e. Scoring data for Writing tasks are analyzed for rater agreement to understand how closely raters agree.	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.	a. Table 6F (Reliability)
	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.	b. Table 8D (Reliability)
	c. Analyses of Rasch model fit statistics are conducted to show that individual tasks perform appropriately	c. Table 6H (Complete Item Analysis)

3. The same scale scores obtained by test takers in different years retain the same meaning.	<p>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.</p> <p>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.</p> <p>c. The same scaling equation is applied from year to year to ensure that scale scores are obtained consistently over time.</p>	<p>a. Table 6E (<i>Equating Summary</i>)</p> <p>b. Table 6D (<i>Scaling Equation</i>)</p> <p>c. Table 6H (<i>Complete Item Analysis</i>)</p>
2. ACCESS for ELLs measures English Language Proficiency for all test takers in a fair and unbiased manner.	<p>a. Differential item functioning (DIF) analyses are conducted to determine whether any items or tasks are biased against certain subgroups.</p> <p>b. Items that show evidence of DIF are carefully reviewed so that any that indicate bias are not used for scoring and are removed from future test forms.</p>	<p>a. Table 6H (<i>Complete Item Analysis</i>); Table 6G (<i>Item/Task Analysis Summary</i>)</p> <p>b. Chapter 1.4.5 (<i>DIF Items</i>)</p>
1. Test takers are classified appropriately according to the proficiency levels defined in the WIDA English Language Development Standards.	<p>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.</p> <p>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.</p> <p>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.</p> <p>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.</p> <p>e. Classification and accuracy analyses are conducted by grade-level to confirm that proficiency level classifications are reliable for all domain and composite scores.</p>	<p>a. Figure 6A (<i>Raw Scores</i>) & Table 6A (<i>Raw Score Descriptive Statistics</i>); Figure 6B (<i>Scale Scores</i>) & Table 6B (<i>Scale Score Descriptive Statistics</i>); Figure 6C (<i>Proficiency Level</i>) & Table 6C (<i>Proficiency Level Distribution</i>); Table 6J (<i>Raw Score to Proficiency Level Score Conversion Chart</i>); Figure 6D (<i>Test Characteristic Curve</i>)</p> <p>b. Figure 8A (<i>Scale Scores</i>) & Table 8A (<i>Scale Score Descriptive Statistics</i>); Figure 8B (<i>Proficiency Level</i>) & Table 8B (<i>Proficiency Level Distribution</i>); Figure 8C (<i>Test Characteristic Curve</i>)</p> <p>c. Figure 6E (<i>Test Information Function</i>); Table 6I (<i>Raw Score to Scale Score Conversion Chart</i>)</p> <p>d. Table 8C (<i>Conditional Standard Error of Measurement</i>) & Figure 8D (<i>Test Information Function</i>)</p> <p>e. Table 8E (<i>Accuracy and Consistency of Classification Indices</i>)</p>

2.5 Visual Guide to Tables and Figures

This section provides navigational support for the tables and figures contained in the 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, 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. Detailed descriptions of the information in each of the tables and figures is included in the preceding chapter (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 from 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, the 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: (1) select Grade; (2) select Domains; (3) select Demographic Characteristics; (4) 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: (1) select Grade Cluster 9–12; (2) select Tier C; (3) select Reading under Domains. This reveals that 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: (1) select Grade Cluster 6–8; (2) select Domain; (3) select Writing. Information for 6–8 Writing is shown in section 8.5.3. Finally, look at the key and find the 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

		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				
		Listening	Reading	Writing	Speaking	
Grade-Level Cluster and Tier	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
Domain Composites	Overall	<u>8.1.8</u>	<u>8.2.8</u>	<u>8.3.8</u>	<u>8.4.8</u>	<u>8.5.8</u>
	Oral Language	<u>8.1.5</u>	<u>8.2.5</u>	<u>8.3.5</u>	<u>8.4.5</u>	<u>8.5.5</u>
	Literacy	<u>8.1.6</u>	<u>8.2.6</u>	<u>8.3.6</u>	<u>8.4.6</u>	<u>8.5.6</u>
	Comprehension	<u>8.1.7</u>	<u>8.2.7</u>	<u>8.3.7</u>	<u>8.4.7</u>	<u>8.5.7</u>
Domains	Listening	<u>8.1.1</u>	<u>8.2.1</u>	<u>8.3.1</u>	<u>8.4.1</u>	<u>8.5.1</u>
	Reading	<u>8.1.2</u>	<u>8.2.2</u>	<u>8.3.2</u>	<u>8.4.2</u>	<u>8.5.2</u>
	Writing	<u>8.1.3</u>	<u>8.2.3</u>	<u>8.3.3</u>	<u>8.4.3</u>	<u>8.5.3</u>
	Speaking	<u>8.1.4</u>	<u>8.2.4</u>	<u>8.3.4</u>	<u>8.4.4</u>	<u>8.5.4</u>

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: by grade-level cluster; by grade, and by 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 36 WIDA states that participated in the operational testing program in 2014–2015. The first row shows the grade-level cluster, the next 36 rows show the number of students in that grade-level cluster who took the test by state, and the final row shows the total number of participants across all 36 states.

Table 4.1.1.2 shows participation by grade-level cluster by gender across all 36 states combined, while Table 4.1.1.3 shows participation by grade-level cluster by ethnicity across all 36 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 information on participation by *tier*.

Table 4.1.3.1 shows this information by cluster, tier, and domain. Since ACCESS for ELLs is broken down into numerous test forms (i.e., Listening in the 1–2 grade-level cluster for Tier A represents one 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 the Speaking test for that grade-level cluster.

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 grade-level cluster and tier for gender. When reviewing data on 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 member states use the Census Bureau categories for student ethnicity. Again, this data may be useful when reviewing analyses of DIF in tables G and H 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 (i.e., a student who was reported to be in Kindergarten was administered a test in the 1–2 grade-level cluster). In all, 472 students were administered a test form that was from a grade-level 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 forms 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	Cluster and Tier												Total	
	K	1-2A	1-2B	1-2C	3-5A	3-5B	3-5C	6-8A	6-8B	6-8C	9-12A	9-12B		9-12C
K		48	26	4	4	3	4	1	0	1	3	0	1	95
1	0				7	12	2	0	2	1	0	0	1	25
2	0				11	19	21	1	1	1	0	0	0	54
3	0	29	49	37				0	2	2	0	0	1	120
4	0	7	6	3				1	3	2	0	0	0	22
5	0	6	3	1				6	10	16	0	1	0	43
6	0	3	2	0	8	10	8				0	0	0	31
7	0	2	3	0	0	2	0				1	1	2	11
8	0	2	2	1	1	1	2				11	9	9	38
9	0	0	0	0	0	0	0	6	4	15				25
10	0	0	0	0	0	0	0	0	1	2				3
11	0	2	1	0	0	0	0	0	0	0				3
12	0	0	0	1	0	0	0	0	0	1				2
Total	0	99	92	47	31	47	37	15	23	41	15	11	14	472

Note that the apparent number of Kindergarten students who were reported to have taken tests in higher grade-level clusters is, at least in part, spurious. In some states, when a grade level has not been defined for a student prior to sending out identification labels for the operational test, the “Grade” field is filled with a placeholder of 0, which is the same code that is used for Kindergarten. If grade information is never updated, the grade for the operational data is recorded as Kindergarten. Thus, many of the students who are reported 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 cluster* across the eight scores awarded on ACCESS, first for the four domains (Listening, Speaking, Reading, and Writing) and then for the four composites (Oral Language, Literacy, Comprehension, and Overall). 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 ethnicity and race. 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 are labeled as Hispanic are included in the Hispanic (Of Any Race) category, regardless of how many racial categories they are included in. Students who are identified as one of the racial categories (e.g., Asian) and have not been identified as Hispanic are identified in only one racial category; if they are identified in more than one racial category, and have not been identified as Hispanic, then they are labeled Non-Hispanic Multi-racial.

Section 4.2.2 shows the mean scale 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 and race.

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 that 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 *grade-level cluster* and tier in Section 4.3.*.1 (note that the * indicates a subsection variable). Tables 4.3.*.1A shows the number of students who were classified into each language proficiency level, while Table 4.3.*.1B shows the percentage 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 *grade* and tier 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 [SI], Language of Language Arts [LA], Language of Math [MA], Language of Science [SC], and Language of Social Studies [SS]; for more information about the Standards, see Section 1.2.1). 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 [SI], Language of Language Arts [LA], Language of Math [MA], Language of Science [SC], and Language of Social Studies [SS]). 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 [SI], Language of Language Arts [LA], Language of Math [MA], Language of Science [SC], and Language of Social Studies [SS]; for more information about the Standards, see Section 1.2.1). The fourth column shows the maximum possible score, the fifth column 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 S303

State	Cluster					Total
	K	1-2	3-5	6-8	9-12	
AK	1,564	3,362	4,303	3,013	2,883	15,125
AL	3,773	6,749	4,260	2,196	2,266	19,244
CO	12,132	25,748	29,945	21,112	17,365	106,302
DC	1,058	1,932	1,327	878	1,126	6,321
DE	1,848	3,173	2,005	919	1,011	8,956
GA	18,011	34,427	25,618	11,635	9,325	99,016
HI	1,612	3,982	3,647	2,559	3,174	14,974
IL	29,329	62,196	54,269	22,658	20,346	188,798
IN	7,985	14,639	15,658	11,702	11,036	61,020
KY	3,237	6,603	5,361	3,005	3,118	21,324
MA	9,917	20,422	20,843	12,970	14,698	78,850
MD	10,458	18,865	14,192	8,407	10,022	61,944
ME	481	1,038	1,253	1,048	1,257	5,077
MI	10,372	20,118	23,607	16,779	17,001	87,877
MN	8,686	16,886	19,768	13,354	12,198	70,892
MO	4,720	8,231	8,164	4,562	3,780	29,457
MP	35	136	531	432	226	1,360
MS	1,303	2,540	2,631	1,491	1,125	9,090
MT	324	739	1,058	684	385	3,190
NC	13,768	28,442	26,591	15,484	14,366	98,651
ND	411	786	797	746	710	3,450
NH	421	1,039	1,102	674	917	4,153
NJ	12,339	19,899	13,870	8,541	11,569	66,218
NM	6,070	13,241	15,016	10,170	8,366	52,863
NV	8,870	19,905	23,349	13,767	8,891	74,782
OK	7,818	13,421	11,736	6,818	5,367	45,160
PA	4,696	11,110	13,043	10,359	12,538	51,746
RI	1,169	2,645	2,578	1,782	1,987	10,161
SC	3,454	9,792	12,918	9,225	7,393	42,782
SD	671	1,187	1,072	760	950	4,640
TN	5,844	12,400	10,025	4,805	4,084	37,158
UT	5,628	11,521	9,601	6,232	5,268	38,250
VA	14,470	29,045	23,907	13,769	16,990	98,181
VT	184	404	352	215	344	1,499
WI	5,875	12,172	14,049	8,406	7,360	47,862
WY	460	731	737	438	397	2,763
Total	218,993	439,526	419,183	251,595	239,839	1,569,136

4.1.1.2 By Gender

Table 4.1.1.2
Participation by Cluster by Gender S303

Cluster		Gender			Total
		F	M	Missing	
K	Count	103,518	114,314	1,161	218,993
	% within Cluster	47.3%	52.2%	0.5%	100.0%
1-2	Count	207,476	229,977	2,073	439,526
	% within Cluster	47.2%	52.3%	0.5%	100.0%
3-5	Count	192,479	224,455	2,249	419,183
	% within Cluster	45.9%	53.5%	0.5%	100.0%
6-8	Count	110,783	139,377	1,435	251,595
	% within Cluster	44.0%	55.4%	0.6%	100.0%
9-12	Count	104,979	132,976	1,884	239,839
	% within Cluster	43.8%	55.4%	0.8%	100.0%
Total	Count	719,235	841,099	8,802	1,569,136
	% within Cluster	45.8%	53.6%	0.6%	100.0%

4.1.1.3 By Ethnicity

Table 4.1.1.3
Participation by Cluster by Ethnicity S303

Cluster		Hispanic/Non-Hispanic			Total
		Hispanic	Other	Missing	
K	Count	148,008	67,121	3,864	218,993
	% within Cluster	67.6%	30.6%	1.8%	100.0%
1-2	Count	304,293	129,159	6,074	439,526
	% within Cluster	69.2%	29.4%	1.4%	100.0%
3-5	Count	292,529	120,711	5,943	419,183
	% within Cluster	69.8%	28.8%	1.4%	100.0%
6-8	Count	171,857	75,757	3,981	251,595
	% within Cluster	68.3%	30.1%	1.6%	100.0%
9-12	Count	151,186	83,530	5,123	239,839
	% within Cluster	63.0%	34.8%	2.1%	100.0%
Total	Count	1,067,873	476,278	24,985	1,569,136
	% within Cluster	68.1%	30.4%	1.6%	100.0%

4.1.2 Participation by Grade

4.1.2.1 By State

Table 4.1.2.1

Participation by Grade by State S303

State	Grade												Total	
	K	1	2	3	4	5	6	7	8	9	10	11		12
AK	1,564	1,562	1,800	1,770	1,371	1,162	979	1,004	1,030	969	772	619	523	15,125
AL	3,773	3,506	3,243	2,392	1,092	776	651	715	830	1,098	547	373	248	19,244
CO	12,132	12,773	12,975	12,619	9,370	7,956	7,137	7,186	6,789	6,567	4,293	3,307	3,198	106,302
DC	1,058	1,069	863	692	360	275	288	264	326	646	223	147	110	6,321
DE	1,848	1,814	1,359	1,046	609	350	284	338	297	467	250	148	146	8,956
GA	18,011	17,780	16,647	13,829	7,088	4,701	3,619	3,979	4,037	5,189	2,221	1,082	833	99,016
HI	1,612	2,014	1,968	1,715	1,098	834	769	882	908	1,415	816	484	459	14,974
IL	29,329	30,855	31,341	29,557	15,171	9,541	7,416	7,185	8,057	9,027	5,392	3,375	2,552	188,798
IN	7,985	8,174	6,465	5,811	5,045	4,802	3,924	3,973	3,805	3,251	3,191	2,593	2,001	61,020
KY	3,237	3,465	3,138	2,642	1,602	1,117	925	994	1,086	1,391	764	534	429	21,324
MA	9,917	10,486	9,936	9,022	6,519	5,302	4,450	4,212	4,308	5,658	3,766	3,030	2,244	78,850
MD	10,458	9,737	9,128	7,439	3,902	2,851	2,633	2,851	2,923	5,645	2,414	1,200	763	61,944
ME	481	502	536	460	439	354	307	349	392	402	305	286	264	5,077
MI	10,372	10,258	9,860	9,254	7,492	6,861	5,974	5,648	5,157	5,166	4,637	3,565	3,633	87,877
MN	8,686	8,577	8,309	7,899	6,351	5,518	4,475	4,512	4,367	4,338	3,050	2,631	2,179	70,892
MO	4,720	4,264	3,967	3,699	2,535	1,930	1,618	1,512	1,432	1,521	931	772	556	29,457
MP	35	57	79	176	199	156	133	162	137	82	96	18	30	1,360
MS	1,303	1,284	1,256	1,172	789	670	544	482	465	476	356	186	107	9,090
MT	324	372	367	395	317	346	262	238	184	159	91	67	68	3,190
NC	13,768	14,114	14,328	13,736	7,399	5,456	4,760	5,284	5,440	7,282	3,484	2,152	1,448	98,651
ND	411	399	387	330	231	236	234	258	254	250	176	146	138	3,450
NH	421	503	536	564	332	206	210	221	243	345	248	188	136	4,153
NJ	12,339	10,533	9,366	7,117	3,910	2,843	2,640	2,820	3,081	4,240	3,220	2,522	1,587	66,218
NM	6,070	6,406	6,835	6,416	4,628	3,972	3,382	3,421	3,367	3,796	2,130	1,423	1,017	52,863
NV	8,870	9,838	10,067	9,856	7,734	5,759	4,672	4,611	4,484	3,625	2,242	1,594	1,430	74,782
OK	7,818	6,772	6,649	6,118	3,093	2,525	2,253	2,334	2,231	2,471	1,360	895	641	45,160
PA	4,696	5,374	5,736	5,314	4,155	3,574	3,434	3,459	3,466	3,960	3,578	2,723	2,277	51,746
RI	1,169	1,302	1,343	1,108	851	619	568	605	609	789	514	417	267	10,161
SC	3,454	4,823	4,969	4,674	4,369	3,875	3,264	3,226	2,735	3,120	2,019	1,329	925	42,782
SD	671	598	589	537	323	212	225	279	256	403	234	183	130	4,640
TN	5,844	6,631	5,769	4,107	3,394	2,524	1,729	1,601	1,475	1,893	1,030	675	486	37,158
UT	5,628	5,826	5,695	4,804	2,310	2,487	2,251	2,116	1,865	1,809	1,269	1,248	942	38,250
VA	14,470	14,816	14,229	12,256	6,574	5,077	4,132	4,669	4,968	8,082	4,155	3,306	1,447	98,181
VT	184	201	203	176	105	71	66	68	81	125	69	80	70	1,499
WI	5,875	6,049	6,123	5,911	4,786	3,352	2,691	2,708	3,007	3,082	1,746	1,351	1,181	47,862
WY	460	363	368	363	222	152	120	156	162	176	84	76	61	2,763
Total	218,993	223,097	216,429	194,976	125,765	98,442	83,019	84,322	84,254	98,915	61,673	44,725	34,526	1,569,136

4.1.2.2 By Gender

Table 4.1.2.2

Participation by Grade by Gender S303

Grade		Gender			Total
		F	M	Missing	
K	Count	103,518	114,314	1,161	218,993
	% within Grade	47.3%	52.2%	0.5%	100.0%
1	Count	105,243	116,634	1,220	223,097
	% within Grade	47.2%	52.3%	0.5%	100.0%
2	Count	102,233	113,343	853	216,429
	% within Grade	47.2%	52.4%	0.4%	100.0%
3	Count	91,311	102,467	1,198	194,976
	% within Grade	46.8%	52.6%	0.6%	100.0%
4	Count	57,132	68,005	628	125,765
	% within Grade	45.4%	54.1%	0.5%	100.0%
5	Count	44,036	53,983	423	98,442
	% within Grade	44.7%	54.8%	0.4%	100.0%
6	Count	36,343	46,000	676	83,019
	% within Grade	43.8%	55.4%	0.8%	100.0%
7	Count	37,398	46,559	365	84,322
	% within Grade	44.4%	55.2%	0.4%	100.0%
8	Count	37,042	46,818	394	84,254
	% within Grade	44.0%	55.6%	0.5%	100.0%
9	Count	42,252	55,611	1,052	98,915
	% within Grade	42.7%	56.2%	1.1%	100.0%
10	Count	26,749	34,552	372	61,673
	% within Grade	43.4%	56.0%	0.6%	100.0%
11	Count	20,017	24,448	260	44,725
	% within Grade	44.8%	54.7%	0.6%	100.0%
12	Count	15,961	18,365	200	34,526
	% within Grade	46.2%	53.2%	0.6%	100.0%
Total	Count	719,235	841,099	8,802	1,569,136
	% within Grade	45.8%	53.6%	0.6%	100.0%

4.1.2.3 By Ethnicity

Table 4.1.2.3

Participation by Grade by Ethnicity S303

Grade		Hispanic/Non-Hispanic			Total
		Hispanic	Other	Missing	
K	Count	148,008	67,121	3,864	218,993
	% within Grade	67.6%	30.6%	1.8%	100.0%
1	Count	153,035	66,705	3,357	223,097
	% within Grade	68.6%	29.9%	1.5%	100.0%
2	Count	151,258	62,454	2,717	216,429
	% within Grade	69.9%	28.9%	1.3%	100.0%
3	Count	136,970	55,190	2,816	194,976
	% within Grade	70.2%	28.3%	1.4%	100.0%
4	Count	87,980	36,026	1,759	125,765
	% within Grade	70.0%	28.6%	1.4%	100.0%
5	Count	67,579	29,495	1,368	98,442
	% within Grade	68.6%	30.0%	1.4%	100.0%
6	Count	56,426	25,036	1,557	83,019
	% within Grade	68.0%	30.2%	1.9%	100.0%
7	Count	57,654	25,423	1,245	84,322
	% within Grade	68.4%	30.1%	1.5%	100.0%
8	Count	57,777	25,298	1,179	84,254
	% within Grade	68.6%	30.0%	1.4%	100.0%
9	Count	67,140	29,403	2,372	98,915
	% within Grade	67.9%	29.7%	2.4%	100.0%
10	Count	39,254	21,225	1,194	61,673
	% within Grade	63.6%	34.4%	1.9%	100.0%
11	Count	26,005	17,857	863	44,725
	% within Grade	58.1%	39.9%	1.9%	100.0%
12	Count	18,787	15,045	694	34,526
	% within Grade	54.4%	43.6%	2.0%	100.0%
Total	Count	1,067,873	476,278	24,985	1,569,136
	% within Grade	68.1%	30.4%	1.6%	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 S303

Cluster			Domain			
			Listening	Reading	Writing	Speaking
K	Tier	-	218,898	218,888	218,890	218,897
1-2	Tier	A	94,664	94,655	94,637	94,622
		B	222,009	221,997	221,949	221,923
		C	122,601	122,604	122,595	122,588
	Total		439,274	439,256	439,181	439,133
3-5	Tier	A	38,343	38,344	38,317	38,319
		B	165,580	165,572	165,530	165,516
		C	215,003	215,005	214,967	214,926
	Total		418,926	418,921	418,814	418,761
6-8	Tier	A	31,770	31,770	31,757	31,772
		B	87,148	87,146	87,095	87,081
		C	132,218	132,216	132,160	132,113
	Total		251,136	251,132	251,012	250,966
9-12	Tier	A	39,218	39,260	39,193	39,138
		B	83,793	83,854	83,726	83,635
		C	113,721	113,797	113,581	113,598
	Total		236,732	236,911	236,500	236,371

4.1.3.2 By Grade by Domain (Test Form)

Table 4.1.3.2
Participation by Grade by Tier by Domain S303

Grade	Tier		Domain			
			Listening	Reading	Writing	Speaking
K	Tier	-	218,898	218,888	218,890	218,897
1	Tier	A	73,374	73,366	73,358	73,347
		B	106,975	106,967	106,947	106,937
		C	42,611	42,616	42,610	42,614
	Total			222,960	222,949	222,915
2	Tier	A	21,290	21,289	21,279	21,275
		B	115,034	115,030	115,002	114,986
		C	79,990	79,988	79,985	79,974
	Total			216,314	216,307	216,266
3	Tier	A	15,857	15,860	15,846	15,844
		B	86,134	86,131	86,113	86,105
		C	92,877	92,878	92,865	92,842
	Total			194,868	194,869	194,824
4	Tier	A	11,920	11,919	11,912	11,909
		B	45,627	45,623	45,614	45,612
		C	68,134	68,136	68,122	68,107
	Total			125,681	125,678	125,648
5	Tier	A	10,566	10,565	10,559	10,566
		B	33,819	33,818	33,803	33,799
		C	53,992	53,991	53,980	53,977
	Total			98,377	98,374	98,342
6	Tier	A	10,614	10,611	10,605	10,607
		B	28,275	28,275	28,261	28,263
		C	44,018	44,014	43,997	43,979
	Total			82,907	82,900	82,863
7	Tier	A	10,430	10,431	10,428	10,433
		B	30,024	30,027	30,009	30,002
		C	43,720	43,721	43,704	43,688
	Total			84,174	84,179	84,141
8	Tier	A	10,726	10,728	10,724	10,732
		B	28,849	28,844	28,825	28,816
		C	44,480	44,481	44,459	44,446
	Total			84,055	84,053	84,008
9	Tier	A	22,334	22,349	22,308	22,280
		B	32,047	32,071	32,023	31,993
		C	43,473	43,490	43,433	43,445
	Total			97,854	97,910	97,764
10	Tier	A	9,282	9,292	9,273	9,269
		B	22,800	22,819	22,775	22,747
		C	28,847	28,872	28,801	28,792
	Total			60,929	60,983	60,849
11	Tier	A	5,029	5,039	5,034	5,021
		B	16,208	16,210	16,194	16,166
		C	22,882	22,895	22,865	22,851
	Total			44,119	44,144	44,093
12	Tier	A	2,573	2,580	2,578	2,568
		B	12,738	12,754	12,734	12,729
		C	18,519	18,540	18,482	18,510
	Total			33,830	33,874	33,794

4.1.3.3 By Cluster by Gender

Table 4.1.3.3

Participation by Cluster by Tier by Gender S303

Cluster	Tier		Gender			Total
			F	M	Missing	
K	-	Count	103,518	114,314	1,161	218,993
		% within Tier	47.3%	52.2%	0.5%	100.0%
1-2	A	Count	42,446	51,763	537	94,746
		% within Tier	44.8%	54.6%	0.6%	100.0%
	B	Count	104,456	116,676	985	222,117
		% within Tier	47.0%	52.5%	0.4%	100.0%
	C	Count	60,574	61,538	551	122,663
		% within Tier	49.4%	50.2%	0.4%	100.0%
3-5	A	Count	17,079	20,979	335	38,393
		% within Tier	44.5%	54.6%	0.9%	100.0%
	B	Count	73,586	91,344	748	165,678
		% within Tier	44.4%	55.1%	0.5%	100.0%
	C	Count	101,814	112,132	1,166	215,112
		% within Tier	47.3%	52.1%	0.5%	100.0%
6-8	A	Count	14,174	17,408	265	31,847
		% within Tier	44.5%	54.7%	0.8%	100.0%
	B	Count	36,839	50,021	480	87,340
		% within Tier	42.2%	57.3%	0.5%	100.0%
	C	Count	59,770	71,948	690	132,408
		% within Tier	45.1%	54.3%	0.5%	100.0%
9-12	A	Count	16,642	22,644	355	39,641
		% within Tier	42.0%	57.1%	0.9%	100.0%
	B	Count	36,725	47,595	604	84,924
		% within Tier	43.2%	56.0%	0.7%	100.0%
	C	Count	51,612	62,737	925	115,274
		% within Tier	44.8%	54.4%	0.8%	100.0%

4.1.3.4 By Cluster by Ethnicity

Table 4.1.3.4

Participation by Cluster by Tier by Ethnicity S303

Cluster	Tier		Hispanic/Non-Hispanic			Total
			Hispanic	Other	Missing	
K	-	Count	148,008	67,121	3,864	218,993
		% within Tier	67.6%	30.6%	1.8%	100.0%
1-2	A	Count	66,824	26,096	1,826	94,746
		% within Tier	70.5%	27.5%	1.9%	100.0%
	B	Count	159,358	59,969	2,790	222,117
		% within Tier	71.7%	27.0%	1.3%	100.0%
	C	Count	78,111	43,094	1,458	122,663
		% within Tier	63.7%	35.1%	1.2%	100.0%
3-5	A	Count	23,434	13,595	1,364	38,393
		% within Tier	61.0%	35.4%	3.6%	100.0%
	B	Count	116,587	46,941	2,150	165,678
		% within Tier	70.4%	28.3%	1.3%	100.0%
	C	Count	152,508	60,175	2,429	215,112
		% within Tier	70.9%	28.0%	1.1%	100.0%
6-8	A	Count	20,478	10,229	1,140	31,847
		% within Tier	64.3%	32.1%	3.6%	100.0%
	B	Count	57,302	28,674	1,364	87,340
		% within Tier	65.6%	32.8%	1.6%	100.0%
	C	Count	94,077	36,854	1,477	132,408
		% within Tier	71.1%	27.8%	1.1%	100.0%
9-12	A	Count	26,773	11,385	1,483	39,641
		% within Tier	67.5%	28.7%	3.7%	100.0%
	B	Count	50,520	32,593	1,811	84,924
		% within Tier	59.5%	38.4%	2.1%	100.0%
	C	Count	73,893	39,552	1,829	115,274
		% within Tier	64.1%	34.3%	1.6%	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 S303

Cluster		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Mean	272.20	193.61	212.61	304.57	288.62	203.36	217.18	228.75
	N	218,288	218,268	218,272	218,222	218,213	218,252	218,256	218,174
1-2	Mean	312.55	297.54	278.47	348.13	330.60	288.24	302.10	300.76
	N	438,609	438,412	438,453	438,394	438,181	438,156	438,295	437,732
3-5	Mean	360.07	340.16	345.24	359.77	360.18	342.96	346.27	347.93
	N	418,276	417,952	417,985	418,023	417,843	417,580	417,816	417,136
6-8	Mean	384.82	358.47	356.07	372.42	378.91	357.53	366.45	363.74
	N	250,354	250,247	250,199	250,160	249,857	249,895	250,056	249,408
9-12	Mean	385.14	376.51	394.68	380.47	383.10	385.90	379.21	384.88
	N	235,149	235,288	234,846	234,787	233,665	234,365	234,669	232,864

4.2.1.2 By Cluster by Gender

Table 4.2.1.2
Mean Scale Scores by Cluster by Gender S303

Cluster	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	F	Mean	276.67	196.79	218.29	308.63	292.89	207.79	220.74	233.13
		N	103,187	103,179	103,174	103,156	103,153	103,169	103,177	103,137
	M	Mean	268.29	190.92	207.66	301.03	284.89	199.53	214.13	224.95
		N	113,941	113,929	113,938	113,906	113,900	113,923	113,919	113,877
	Missing	Mean	258.99	175.56	194.24	290.90	275.17	185.13	200.58	211.95
		N	1,160	1,160	1,160	1,160	1,160	1,160	1,160	1,160
1-2	F	Mean	313.74	298.66	281.33	349.72	331.98	290.23	303.24	302.56
		N	207,063	206,994	207,007	206,965	206,881	206,882	206,936	206,684
	M	Mean	311.54	296.59	275.94	346.78	329.42	286.51	301.14	299.19
		N	229,484	229,358	229,396	229,376	229,253	229,225	229,300	229,004
	Missing	Mean	305.70	289.89	273.09	338.90	322.53	281.77	294.68	293.81
		N	2,062	2,060	2,050	2,053	2,047	2,049	2,059	2,044
3-5	F	Mean	360.64	341.13	348.39	360.01	360.59	345.01	347.12	349.48
		N	192,091	191,982	192,001	191,980	191,911	191,856	191,928	191,666
	M	Mean	359.65	339.37	342.59	359.66	359.92	341.25	345.59	346.66
		N	223,946	223,731	223,748	223,809	223,699	223,488	223,650	223,239
	Missing	Mean	353.07	335.28	340.09	349.54	351.56	337.97	340.75	341.82
		N	2,239	2,239	2,236	2,234	2,233	2,236	2,238	2,231
6-8	F	Mean	386.31	360.10	359.91	371.09	378.99	360.26	368.04	365.67
		N	110,229	110,190	110,192	110,134	110,012	110,070	110,121	109,859
	M	Mean	383.77	357.23	353.08	373.59	378.97	355.42	365.27	362.28
		N	138,706	138,636	138,585	138,606	138,429	138,405	138,517	138,135
	Missing	Mean	371.59	353.24	349.72	359.92	366.23	351.72	358.80	355.92
		N	1,419	1,421	1,422	1,420	1,416	1,420	1,418	1,414
9-12	F	Mean	385.77	378.86	398.65	379.51	382.94	389.05	381.03	387.02
		N	103,117	103,190	103,040	102,971	102,536	102,868	102,939	102,253
	M	Mean	384.83	374.74	391.67	381.44	383.43	383.52	377.89	383.32
		N	130,218	130,274	129,997	130,002	129,336	129,690	129,921	128,829
	Missing	Mean	370.91	369.50	385.03	365.02	368.32	377.61	370.06	374.62
		N	1,814	1,824	1,809	1,814	1,793	1,807	1,809	1,782

4.2.1.3 By Cluster by Ethnicity

Table 4.2.1.3
Mean Scale Scores by Cluster by Ethnicity S303

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Non-Hispanic Asian	Mean	281.80	221.33	235.93	310.99	296.62	228.89	239.46	249.00
		N	29,016	29,014	29,015	29,017	29,016	29,013	29,013	29,013
	Non-Hispanic Pacific Islander	Mean	266.47	187.61	206.58	306.25	286.59	197.33	211.27	223.94
		N	1,624	1,623	1,623	1,624	1,624	1,623	1,622	1,622
	Non-Hispanic Black	Mean	280.37	208.36	225.72	319.61	300.22	217.30	229.96	241.99
		N	9,621	9,621	9,621	9,619	9,619	9,620	9,620	9,617
	Hispanic (Of Any Race)	Mean	268.23	185.98	205.87	300.80	284.75	196.17	210.65	222.56
		N	147,671	147,660	147,662	147,612	147,605	147,648	147,651	147,581
	Non-Hispanic American Indian	Mean	271.32	179.31	193.01	300.37	286.07	186.37	206.91	216.08
		N	3,278	3,277	3,276	3,278	3,278	3,275	3,277	3,275
	Non-Hispanic Multi-racial	Mean	303.98	216.22	232.47	332.48	318.46	224.60	242.52	252.54
		N	910	910	910	910	910	910	910	910
	Non-Hispanic White	Mean	284.08	204.62	225.23	315.82	300.18	215.18	228.44	240.49
		N	21,629	21,627	21,628	21,624	21,624	21,627	21,627	21,622
	Missing	Mean	262.50	189.21	207.05	297.35	280.14	198.38	211.19	222.72
		N	4,539	4,536	4,537	4,538	4,537	4,536	4,536	4,535
1-2	Non-Hispanic Asian	Mean	316.67	304.87	286.63	350.77	333.96	295.98	308.47	307.18
		N	53,848	53,829	53,836	53,823	53,807	53,808	53,819	53,769
	Non-Hispanic Pacific Islander	Mean	306.79	293.84	279.38	343.80	325.62	286.83	297.83	298.32
		N	3,559	3,560	3,556	3,546	3,546	3,556	3,556	3,542
	Non-Hispanic Black	Mean	312.45	298.80	279.39	352.59	332.78	289.34	302.95	302.19
		N	19,423	19,412	19,419	19,422	19,413	19,403	19,404	19,388
	Hispanic (Of Any Race)	Mean	311.68	295.91	276.58	346.67	329.44	286.49	300.70	299.18
		N	303,837	303,703	303,739	303,689	303,546	303,533	303,624	303,234
	Non-Hispanic American Indian	Mean	308.80	293.45	273.09	339.68	324.51	283.52	298.12	295.63
		N	6,912	6,904	6,900	6,909	6,896	6,893	6,901	6,882
	Non-Hispanic Multi-racial	Mean	321.44	305.38	283.77	362.21	342.11	294.83	310.29	308.88
		N	1,722	1,720	1,720	1,723	1,721	1,718	1,719	1,717
	Non-Hispanic White	Mean	315.69	300.87	282.36	355.80	336.00	291.85	305.38	304.91
		N	41,814	41,796	41,798	41,806	41,789	41,776	41,786	41,750
	Missing	Mean	305.06	292.56	275.64	340.59	323.09	284.34	296.37	295.77
		N	7,494	7,488	7,485	7,476	7,463	7,469	7,486	7,450

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
3-5	Non-Hispanic Asian	Mean	366.19	346.65	350.63	358.40	362.54	348.89	352.65	352.78
		N	48,036	48,023	48,013	48,004	47,994	47,993	48,009	47,947
	Non-Hispanic Pacific Islander	Mean	352.79	336.20	345.65	352.72	353.01	341.17	341.30	344.53
		N	3,613	3,614	3,610	3,609	3,607	3,609	3,612	3,605
	Non-Hispanic Black	Mean	360.01	339.42	343.62	360.55	360.55	341.80	345.74	347.25
		N	21,396	21,360	21,390	21,395	21,382	21,347	21,347	21,325
	Hispanic (Of Any Race)	Mean	359.20	339.20	344.58	359.78	359.75	342.15	345.34	347.23
		N	292,105	291,837	291,897	291,907	291,799	291,596	291,759	291,295
	Non-Hispanic American Indian	Mean	352.06	333.96	338.17	352.91	352.78	336.39	339.55	341.19
		N	8,251	8,245	8,208	8,249	8,233	8,201	8,238	8,184
	Non-Hispanic Multi-racial	Mean	368.86	346.03	348.33	369.77	369.56	347.48	352.99	353.92
		N	1,376	1,376	1,374	1,376	1,376	1,373	1,376	1,373
	Non-Hispanic White	Mean	362.98	342.57	347.36	365.23	364.37	345.22	348.84	350.77
		N	36,425	36,419	36,416	36,411	36,394	36,396	36,402	36,357
	Missing	Mean	350.82	333.27	337.33	347.81	349.60	335.56	338.66	339.59
		N	7,074	7,078	7,077	7,072	7,058	7,065	7,073	7,050
6-8	Non-Hispanic Asian	Mean	389.46	365.75	360.80	370.30	380.17	363.54	372.94	368.31
		N	29,783	29,777	29,779	29,763	29,738	29,754	29,764	29,711
	Non-Hispanic Pacific Islander	Mean	375.36	354.40	356.83	368.36	372.32	355.89	360.86	360.67
		N	2,438	2,445	2,440	2,440	2,425	2,436	2,432	2,416
	Non-Hispanic Black	Mean	383.81	358.51	354.07	373.25	378.81	356.55	366.18	363.03
		N	15,368	15,340	15,361	15,356	15,341	15,321	15,329	15,295
	Hispanic (Of Any Race)	Mean	384.59	357.00	355.40	372.80	378.98	356.46	365.35	363.01
		N	171,225	171,144	171,132	171,074	170,885	170,929	171,021	170,592
	Non-Hispanic American Indian	Mean	377.89	354.07	351.75	369.77	374.13	353.22	361.29	359.36
		N	6,250	6,254	6,213	6,248	6,221	6,203	6,238	6,178
	Non-Hispanic Multi-racial	Mean	390.21	363.27	357.71	378.72	384.84	360.73	371.37	367.80
		N	746	746	743	744	742	743	745	740
	Non-Hispanic White	Mean	387.81	363.38	359.21	377.33	382.86	361.55	370.79	367.73
		N	20,147	20,139	20,136	20,138	20,124	20,121	20,132	20,102
	Missing	Mean	366.38	351.37	348.05	351.29	359.13	349.96	355.91	352.49
		N	4,397	4,402	4,395	4,397	4,381	4,388	4,395	4,374

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
9-12	Non-Hispanic Asian	Mean	391.34	383.22	400.57	379.57	385.75	392.18	385.76	390.05
		N	34,279	34,309	34,273	34,250	34,164	34,242	34,250	34,107
	Non-Hispanic Pacific Islander	Mean	378.11	368.67	394.27	382.39	380.58	381.84	371.67	381.27
		N	2,451	2,455	2,437	2,454	2,438	2,434	2,443	2,422
	Non-Hispanic Black	Mean	383.12	376.56	394.22	381.62	382.68	385.68	378.65	384.59
		N	18,897	18,878	18,887	18,867	18,806	18,822	18,823	18,726
	Hispanic (Of Any Race)	Mean	383.12	374.67	393.19	379.46	381.57	384.24	377.31	383.26
		N	148,236	148,332	148,026	148,022	147,214	147,689	147,926	146,674
	Non-Hispanic American Indian	Mean	391.83	371.69	396.16	384.75	388.61	384.32	377.85	385.39
		N	4,752	4,751	4,709	4,744	4,693	4,700	4,734	4,648
	Non-Hispanic Multi-racial	Mean	391.99	382.20	398.16	391.31	392.02	390.53	385.20	390.72
		N	650	654	651	654	648	651	650	647
	Non-Hispanic White	Mean	394.75	382.54	398.60	392.14	393.74	390.85	386.30	391.51
		N	20,447	20,459	20,444	20,447	20,391	20,427	20,421	20,357
	Missing	Mean	368.45	368.63	383.29	359.43	364.26	376.33	368.68	372.62
		N	5,437	5,450	5,419	5,349	5,311	5,400	5,422	5,283

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 S303

Grade		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Mean	272.20	193.61	212.61	304.57	288.62	203.36	217.18	228.75
	N	218,288	218,268	218,272	218,222	218,213	218,252	218,256	218,174
1	Mean	299.04	284.55	268.88	338.90	319.19	276.96	288.85	289.45
	N	222,599	222,466	222,512	222,506	222,379	222,327	222,397	222,105
2	Mean	326.47	310.92	288.35	357.64	342.35	299.86	315.75	312.41
	N	216,010	215,946	215,941	215,888	215,802	215,829	215,898	215,627
3	Mean	351.27	333.33	341.16	357.58	354.71	337.48	338.88	342.46
	N	194,648	194,545	194,531	194,539	194,462	194,385	194,489	194,196
4	Mean	363.36	342.74	346.69	360.35	362.09	345.00	349.05	349.92
	N	125,460	125,339	125,363	125,384	125,328	125,223	125,305	125,094
5	Mean	373.31	350.41	351.49	363.36	368.61	351.23	357.36	356.24
	N	98,168	98,068	98,091	98,100	98,053	97,972	98,022	97,846
6	Mean	376.70	351.15	351.42	369.26	373.25	351.54	358.87	357.86
	N	82,654	82,599	82,586	82,580	82,500	82,485	82,546	82,338
7	Mean	385.51	358.50	356.44	373.04	379.55	357.72	366.70	364.07
	N	83,927	83,899	83,875	83,870	83,783	83,794	83,841	83,653
8	Mean	392.14	365.65	360.29	374.90	383.84	363.24	373.68	369.21
	N	83,773	83,749	83,738	83,710	83,574	83,616	83,669	83,417
9	Mean	378.95	372.45	390.84	374.51	377.02	381.96	374.51	380.30
	N	97,451	97,466	97,308	97,313	96,920	97,128	97,272	96,620
10	Mean	385.92	376.48	394.89	380.69	383.59	386.00	379.41	385.09
	N	60,572	60,612	60,482	60,446	60,193	60,359	60,472	59,990
11	Mean	391.42	381.04	398.66	385.95	389.03	390.15	384.26	389.65
	N	43,795	43,822	43,769	43,721	43,497	43,671	43,688	43,351
12	Mean	393.56	382.44	400.26	390.26	392.24	391.66	385.93	391.69
	N	33,331	33,388	33,287	33,307	33,055	33,207	33,237	32,903

4.2.2.2 By Grade by Gender

Table 4.2.2.2
Mean Scale Scores by Grade by Gender S303

Scale	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	F	Mean	276.67	196.79	218.29	308.63	292.89	207.79	220.74	233.13
		N	103,187	103,179	103,174	103,156	103,153	103,169	103,177	103,137
	M	Mean	268.29	190.92	207.66	301.03	284.89	199.53	214.13	224.95
		N	113,941	113,929	113,938	113,906	113,900	113,923	113,919	113,877
	Missing	Mean	258.99	175.56	194.24	290.90	275.17	185.13	200.58	211.95
		N	1,160	1,160	1,160	1,160	1,160	1,160	1,160	1,160
1	F	Mean	300.27	285.51	271.53	340.75	320.72	278.77	289.89	291.17
		N	105,027	104,972	104,994	104,982	104,932	104,913	104,938	104,807
	M	Mean	297.99	283.74	266.53	337.33	317.89	275.39	287.97	287.95
		N	116,358	116,282	116,313	116,320	116,245	116,210	116,248	116,097
	Missing	Mean	293.55	279.19	264.95	330.22	312.04	272.35	283.47	284.08
		N	1,214	1,212	1,205	1,204	1,202	1,204	1,211	1,201
2	F	Mean	327.61	312.19	291.41	358.95	343.58	302.02	316.99	314.29
		N	102,036	102,022	102,013	101,983	101,949	101,969	101,998	101,877
	M	Mean	325.47	309.81	285.62	356.51	341.29	297.94	314.67	310.75
		N	113,126	113,076	113,083	113,056	113,008	113,015	113,052	112,907
	Missing	Mean	323.09	305.17	284.69	351.21	337.45	295.18	310.69	307.69
		N	848	848	845	849	845	845	848	843
3	F	Mean	352.22	334.39	344.35	358.35	355.56	339.60	339.91	344.19
		N	91,165	91,142	91,147	91,121	91,091	91,103	91,117	91,019
	M	Mean	350.44	332.40	338.34	356.97	353.98	335.61	337.99	340.95
		N	102,290	102,211	102,192	102,229	102,182	102,090	102,180	101,988
	Missing	Mean	349.51	332.05	338.00	352.16	351.16	335.27	337.49	339.78
		N	1,193	1,192	1,192	1,189	1,189	1,192	1,192	1,189
4	F	Mean	364.01	343.86	349.95	360.55	362.51	347.19	350.03	351.58
		N	57,005	56,960	56,967	56,972	56,952	56,913	56,948	56,860
	M	Mean	362.86	341.83	343.98	360.28	361.81	343.19	348.26	348.58
		N	67,830	67,754	67,772	67,788	67,752	67,686	67,732	67,611
	Missing	Mean	358.35	338.47	342.91	349.83	354.27	341.01	344.52	344.78
		N	625	625	624	624	624	624	625	623
5	F	Mean	373.76	351.58	354.73	362.79	368.54	353.43	358.32	357.76
		N	43,921	43,880	43,887	43,887	43,868	43,840	43,863	43,787
	M	Mean	373.09	349.53	348.92	364.00	368.82	349.51	356.68	355.10
		N	53,826	53,766	53,784	53,792	53,765	53,712	53,738	53,640
	Missing	Mean	355.30	339.66	341.85	341.71	348.65	341.13	344.36	343.20
		N	421	422	420	421	420	420	421	419

Scale	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
6	F	Mean	378.32	352.56	355.50	368.03	373.46	354.29	360.35	359.85
		N	36,167	36,151	36,138	36,138	36,101	36,101	36,133	36,042
	M	Mean	375.65	350.09	348.28	370.48	373.33	349.45	357.82	356.42
		N	45,819	45,779	45,778	45,771	45,731	45,716	45,746	45,630
	Missing	Mean	360.64	347.29	345.19	352.33	356.87	346.46	351.35	349.38
		N	668	669	670	671	668	668	667	666
7	F	Mean	386.91	360.23	360.21	371.47	379.46	360.46	368.33	365.95
		N	37,227	37,209	37,216	37,195	37,160	37,176	37,189	37,113
	M	Mean	384.49	357.16	353.46	374.38	379.71	355.56	365.46	362.61
		N	46,339	46,329	46,298	46,315	46,263	46,257	46,291	46,180
	Missing	Mean	373.52	351.65	349.89	362.73	368.67	351.01	358.30	356.12
		N	361	361	361	360	360	361	361	360
8	F	Mean	393.55	367.37	363.94	373.71	383.94	365.91	375.29	371.11
		N	36,835	36,830	36,838	36,801	36,751	36,793	36,799	36,704
	M	Mean	391.05	364.30	357.42	375.88	383.79	361.14	372.41	367.73
		N	46,548	46,528	46,509	46,520	46,435	46,432	46,480	46,325
	Missing	Mean	388.57	364.88	357.32	370.42	380.09	361.36	372.01	366.95
		N	390	391	391	389	388	391	390	388
9	F	Mean	380.54	375.18	395.53	374.71	377.91	385.66	376.90	383.14
		N	41,711	41,721	41,670	41,662	41,517	41,609	41,646	41,418
	M	Mean	378.04	370.52	387.47	374.69	376.66	379.31	372.89	378.34
		N	54,724	54,723	54,623	54,633	54,399	54,506	54,613	54,204
	Missing	Mean	362.59	364.37	379.79	356.97	360.06	372.48	363.92	368.45
		N	1,016	1,022	1,015	1,018	1,004	1,013	1,013	998
10	F	Mean	385.79	378.59	398.38	379.15	382.76	388.79	380.85	386.78
		N	26,300	26,311	26,262	26,241	26,149	26,218	26,260	26,075
	M	Mean	386.12	374.88	392.26	381.98	384.32	383.88	378.35	383.83
		N	33,921	33,949	33,872	33,856	33,699	33,793	33,863	33,573
	Missing	Mean	376.98	372.86	388.36	370.21	374.06	380.76	374.36	378.59
		N	351	352	348	349	345	348	349	342
11	F	Mean	390.89	382.67	401.87	383.38	387.49	392.56	385.20	390.86
		N	19,649	19,676	19,643	19,624	19,527	19,613	19,621	19,471
	M	Mean	391.96	379.77	396.09	388.22	390.42	388.23	383.54	388.74
		N	23,894	23,892	23,873	23,845	23,719	23,805	23,815	23,630
	Missing	Mean	382.46	375.53	392.04	372.46	377.78	384.04	377.78	382.02
		N	252	254	253	252	251	253	252	250
12	F	Mean	393.38	384.40	403.40	388.13	391.08	394.18	387.22	393.08
		N	15,457	15,482	15,465	15,444	15,343	15,428	15,412	15,289
	M	Mean	393.76	380.73	397.53	392.15	393.29	389.47	384.82	390.49
		N	17,679	17,710	17,629	17,668	17,519	17,586	17,630	17,422
	Missing	Mean	388.41	382.47	397.42	388.11	388.72	390.47	384.31	390.00
		N	195	196	193	195	193	193	195	192

4.2.2.3 By Grade by Ethnicity

Table 4.2.2.3

Mean Scale Scores by Grade by Ethnicity S303

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over	
K	Non-Hispanic Asian	Mean	281.80	221.33	235.93	310.99	296.62	228.89	239.46	249.00	
		N	29,016	29,014	29,015	29,017	29,016	29,013	29,013	29,013	29,012
	Non-Hispanic Pacific Islander	Mean	266.47	187.61	206.58	306.25	286.59	197.33	211.27	223.94	
		N	1,624	1,623	1,623	1,624	1,624	1,623	1,622	1,622	1,622
	Non-Hispanic Black	Mean	280.37	208.36	225.72	319.61	300.22	217.30	229.96	241.99	
		N	9,621	9,621	9,621	9,619	9,619	9,620	9,620	9,617	
	Hispanic (Of Any Race)	Mean	268.23	185.98	205.87	300.80	284.75	196.17	210.65	222.56	
		N	147,671	147,660	147,662	147,612	147,605	147,648	147,651	147,581	
	Non-Hispanic American Indian	Mean	271.32	179.31	193.01	300.37	286.07	186.37	206.91	216.08	
		N	3,278	3,277	3,276	3,278	3,278	3,275	3,277	3,275	
	Non-Hispanic Multi-racial	Mean	303.98	216.22	232.47	332.48	318.46	224.60	242.52	252.54	
		N	910	910	910	910	910	910	910	910	910
	Non-Hispanic White	Mean	284.08	204.62	225.23	315.82	300.18	215.18	228.44	240.49	
		N	21,629	21,627	21,628	21,624	21,624	21,627	21,627	21,627	21,622
	Missing	Mean	262.50	189.21	207.05	297.35	280.14	198.38	211.19	222.72	
		N	4,539	4,536	4,537	4,538	4,537	4,536	4,536	4,535	
	1	Non-Hispanic Asian	Mean	303.65	291.55	277.92	342.96	323.52	284.98	295.14	296.35
			N	28,038	28,023	28,029	28,025	28,014	28,011	28,017	27,989
Non-Hispanic Pacific Islander		Mean	292.79	280.44	269.98	334.17	313.76	275.46	284.16	286.78	
		N	1,796	1,797	1,796	1,791	1,791	1,796	1,795	1,790	
Non-Hispanic Black		Mean	298.95	286.42	270.41	344.94	322.17	278.68	290.15	291.54	
		N	9,738	9,729	9,736	9,739	9,732	9,724	9,724	9,714	
Hispanic (Of Any Race)		Mean	297.93	282.87	266.60	336.70	317.54	274.98	287.34	287.57	
		N	152,774	152,683	152,712	152,716	152,629	152,588	152,638	152,437	
Non-Hispanic American Indian		Mean	295.51	281.97	263.64	331.45	313.68	273.06	285.97	285.06	
		N	3,377	3,374	3,372	3,376	3,368	3,369	3,372	3,361	
Non-Hispanic Multi-racial		Mean	308.82	292.14	274.87	357.15	333.16	283.80	297.17	298.50	
		N	885	884	886	885	885	884	883	883	
Non-Hispanic White		Mean	302.67	287.60	273.26	348.24	325.65	280.68	292.08	293.99	
		N	21,882	21,872	21,877	21,882	21,873	21,861	21,865	21,850	
Missing		Mean	293.17	280.91	267.39	333.25	313.40	274.42	284.56	285.91	
		N	4,109	4,104	4,104	4,092	4,087	4,094	4,103	4,081	

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
2	Non-Hispanic Asian	Mean	330.81	319.34	296.09	359.25	345.30	307.93	322.94	318.93
		N	25,810	25,806	25,807	25,798	25,793	25,797	25,802	25,780
	Non-Hispanic Pacific Islander	Mean	321.05	307.49	288.96	353.63	337.71	298.45	311.76	310.11
		N	1,763	1,763	1,760	1,755	1,755	1,760	1,761	1,752
	Non-Hispanic Black	Mean	326.04	311.23	288.41	360.27	343.44	300.05	315.82	312.88
		N	9,685	9,683	9,683	9,683	9,681	9,679	9,680	9,674
	Hispanic (Of Any Race)	Mean	325.59	309.10	286.67	356.76	341.47	298.11	314.21	310.92
		N	151,063	151,020	151,027	150,973	150,917	150,945	150,986	150,797
	Non-Hispanic American Indian	Mean	321.49	304.42	282.13	347.55	334.84	293.52	309.73	305.73
		N	3,535	3,530	3,528	3,533	3,528	3,524	3,529	3,521
	Non-Hispanic Multi-racial	Mean	334.78	319.38	293.23	367.55	351.59	306.54	324.15	319.87
		N	837	836	834	838	836	834	836	834
	Non-Hispanic White	Mean	329.99	315.43	292.35	364.10	347.36	304.11	319.97	316.89
		N	19,932	19,924	19,921	19,924	19,916	19,915	19,921	19,900
Missing	Mean	319.50	306.69	285.64	349.47	334.83	296.39	310.69	307.72	
	N	3,385	3,384	3,381	3,384	3,376	3,375	3,383	3,369	
3	Non-Hispanic Asian	Mean	358.17	339.21	347.34	358.36	358.53	343.50	345.07	347.82
		N	22,218	22,212	22,204	22,206	22,199	22,198	22,207	22,178
	Non-Hispanic Pacific Islander	Mean	344.29	329.63	341.33	349.49	347.17	335.74	334.17	338.99
		N	1,638	1,639	1,636	1,635	1,635	1,636	1,638	1,633
	Non-Hispanic Black	Mean	351.61	332.53	339.91	359.08	355.62	336.49	338.42	342.05
		N	9,043	9,034	9,041	9,041	9,037	9,029	9,030	9,022
	Hispanic (Of Any Race)	Mean	350.11	332.42	340.19	357.01	353.84	336.54	337.90	341.54
		N	136819	136733	136742	136746	136698	136636	136704	136520
	Non-Hispanic American Indian	Mean	340.27	325.48	331.65	347.48	344.16	328.89	330.13	333.35
		N	3,277	3,273	3,254	3,273	3,269	3,251	3,269	3,241
	Non-Hispanic Multi-racial	Mean	363.99	341.48	346.28	370.00	367.26	344.12	348.38	350.88
		N	676	676	675	676	676	675	676	675
	Non-Hispanic White	Mean	354.68	335.74	344.24	364.25	359.75	340.21	341.61	345.90
		N	17,516	17,515	17,513	17,506	17,499	17,505	17,505	17,483
Missing	Mean	345.62	328.78	335.26	348.46	347.41	332.25	334.00	336.62	
	N	3,461	3,463	3,466	3,456	3,449	3,455	3,460	3,444	
4	Non-Hispanic Asian	Mean	368.67	349.08	351.46	357.48	363.29	350.54	355.09	354.15
		N	14,242	14,238	14,238	14,236	14,233	14,230	14,234	14,219
	Non-Hispanic Pacific Islander	Mean	355.09	338.53	347.38	353.01	354.28	343.17	343.63	346.33
		N	1,124	1,124	1,122	1,122	1,121	1,122	1,123	1,121
	Non-Hispanic Black	Mean	361.95	341.00	344.12	360.99	361.71	342.86	347.42	348.35
		N	6,614	6,592	6,608	6,611	6,607	6,588	6,591	6,582
	Hispanic (Of Any Race)	Mean	362.87	341.98	346.29	360.77	362.06	344.43	348.36	349.51
		N	87,833	87,741	87,766	87,772	87,740	87,665	87,718	87,572
	Non-Hispanic American Indian	Mean	354.43	335.56	339.22	353.83	354.36	337.69	341.35	342.57
		N	2,622	2,621	2,609	2,617	2,614	2,607	2,620	2,602
	Non-Hispanic Multi-racial	Mean	367.98	345.29	348.43	367.20	367.81	347.22	352.22	353.26
		N	396	396	394	396	396	394	396	394
	Non-Hispanic White	Mean	366.04	345.08	348.32	364.70	365.59	346.99	351.50	352.35
		N	10,581	10,578	10,579	10,580	10,572	10,571	10,575	10,561
Missing	Mean	353.56	335.58	338.91	348.67	351.31	337.53	341.10	341.50	
	N	2,048	2,049	2,047	2,050	2,045	2,046	2,048	2,043	

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
5	Non-Hispanic Asian	Mean	378.53	357.93	355.94	359.62	369.33	357.20	364.19	360.61
		N	11,576	11,573	11,571	11,562	11,562	11,565	11,568	11,550
	Non-Hispanic Pacific Islander	Mean	366.10	345.79	351.67	358.53	362.54	348.96	351.94	352.81
		N	851	851	852	852	851	851	851	851
	Non-Hispanic Black	Mean	371.01	348.47	348.89	362.36	366.96	348.97	355.35	354.20
		N	5,739	5,734	5,741	5,743	5,738	5,730	5,726	5,721
	Hispanic (Of Any Race)	Mean	372.87	349.36	351.25	364.10	368.75	350.58	356.49	355.83
		N	67,453	67,363	67,389	67,389	67,361	67,295	67,337	67,203
	Non-Hispanic American Indian	Mean	365.85	343.98	346.06	359.41	363.00	345.36	350.66	350.52
		N	2,352	2,351	2,345	2,359	2,350	2,343	2,349	2,341
	Non-Hispanic Multi-racial	Mean	380.83	357.09	352.75	372.61	376.97	355.26	364.25	361.51
		N	304	304	305	304	304	304	304	304
Non-Hispanic White	Mean	376.58	353.73	352.69	367.96	372.55	353.49	360.66	359.00	
	N	8,328	8,326	8,324	8,325	8,323	8,320	8,322	8,313	
Missing	Mean	358.72	340.17	339.88	345.24	352.21	340.31	345.79	343.65	
	N	1,565	1,566	1,564	1,566	1,564	1,564	1,565	1,563	
6	Non-Hispanic Asian	Mean	380.78	357.44	356.11	366.20	373.77	357.05	364.48	361.85
		N	9,833	9,830	9,830	9,829	9,822	9,822	9,829	9,814
	Non-Hispanic Pacific Islander	Mean	367.29	347.77	352.34	364.59	366.21	350.33	353.74	354.96
		N	758	763	760	757	756	759	757	751
	Non-Hispanic Black	Mean	374.33	350.95	348.92	369.57	372.23	350.20	358.02	356.61
		N	4,913	4,903	4,908	4,906	4,901	4,897	4,900	4,888
	Hispanic (Of Any Race)	Mean	376.74	349.86	350.71	370.00	373.64	350.54	357.99	357.28
		N	56,236	56,191	56,194	56,183	56,133	56,123	56,157	56,024
	Non-Hispanic American Indian	Mean	368.79	347.09	346.91	365.92	367.53	347.27	353.63	353.25
		N	2,075	2,075	2,063	2,073	2,066	2,060	2,070	2,051
	Non-Hispanic Multi-racial	Mean	386.01	356.38	354.39	377.17	382.28	355.62	365.31	363.49
		N	252	252	252	250	249	252	252	249
Non-Hispanic White	Mean	379.44	355.57	355.19	373.69	376.84	355.64	362.77	361.79	
	N	6,872	6,868	6,864	6,864	6,863	6,860	6,867	6,855	
Missing	Mean	359.88	346.08	344.23	348.88	354.66	345.39	350.25	347.99	
	N	1,715	1,717	1,715	1,718	1,710	1,712	1,714	1,706	
7	Non-Hispanic Asian	Mean	390.31	366.18	361.19	371.01	380.94	363.94	373.52	368.82
		N	9,810	9,808	9,806	9,795	9,790	9,799	9,805	9,784
	Non-Hispanic Pacific Islander	Mean	374.47	353.91	355.76	367.39	371.38	355.07	360.18	359.82
		N	843	845	847	851	838	844	840	836
	Non-Hispanic Black	Mean	385.27	358.05	354.13	373.82	379.81	356.32	366.33	363.18
		N	5,164	5,158	5,163	5,162	5,158	5,152	5,154	5,146
	Hispanic (Of Any Race)	Mean	385.22	356.96	355.81	373.41	379.59	356.64	365.53	363.32
		N	57,448	57,427	57,420	57,412	57,360	57,367	57,391	57,275
	Non-Hispanic American Indian	Mean	379.00	354.37	352.34	371.06	375.30	353.69	361.85	360.02
		N	2,198	2,198	2,179	2,193	2,187	2,178	2,195	2,170
	Non-Hispanic Multi-racial	Mean	389.11	364.28	357.40	375.11	382.30	361.05	371.77	367.24
		N	237	236	236	236	236	236	236	236
Non-Hispanic White	Mean	388.51	363.91	359.57	377.83	383.44	361.98	371.43	368.21	
	N	6,852	6,849	6,851	6,848	6,845	6,846	6,845	6,838	
Missing	Mean	366.02	351.02	348.47	351.57	359.12	350.03	355.57	352.52	
	N	1,375	1,378	1,373	1,373	1,369	1,372	1,375	1,368	

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
8	Non-Hispanic Asian	Mean	397.05	373.37	364.97	373.59	385.63	369.44	380.56	374.10
		N	10,140	10,139	10,143	10,139	10,126	10,133	10,130	10,113
	Non-Hispanic Pacific Islander	Mean	383.55	360.94	362.02	372.77	378.82	361.78	368.01	366.71
		N	837	837	833	832	831	833	835	829
	Non-Hispanic Black	Mean	391.19	365.97	358.79	376.11	383.93	362.68	373.62	368.85
		N	5,291	5,279	5,290	5,288	5,282	5,272	5,275	5,261
	Hispanic (Of Any Race)	Mean	391.62	364.02	359.57	374.93	383.59	362.06	372.37	368.31
		N	57,541	57,526	57,518	57,479	57,392	57,439	57,473	57,293
	Non-Hispanic American Indian	Mean	386.21	361.04	356.18	372.38	379.76	358.93	368.70	365.04
		N	1,977	1,981	1,971	1,982	1,968	1,965	1,973	1,957
	Non-Hispanic Multi-racial	Mean	395.33	369.07	361.28	383.51	389.65	365.49	376.95	372.51
		N	257	258	255	258	257	255	257	255
Non-Hispanic White	Mean	396.02	371.17	363.12	380.69	388.69	367.41	378.68	373.56	
	N	6,423	6,422	6,421	6,426	6,416	6,415	6,420	6,409	
Missing	Mean	375.29	358.68	352.62	354.16	365.01	355.89	363.69	358.37	
	N	1,307	1,307	1,307	1,306	1,302	1,304	1,306	1,300	
9	Non-Hispanic Asian	Mean	389.51	382.13	399.32	379.37	384.76	391.00	384.45	388.91
		N	11,827	11,825	11,821	11,823	11,800	11,810	11,815	11,783
	Non-Hispanic Pacific Islander	Mean	372.35	364.45	391.15	378.25	375.70	378.18	366.95	377.25
		N	1,010	1,011	1,007	1,011	1,005	1,005	1,007	1,001
	Non-Hispanic Black	Mean	377.16	373.12	390.31	376.23	376.98	382.03	374.42	380.29
		N	6,698	6,678	6,692	6,687	6,672	6,665	6,668	6,643
	Hispanic (Of Any Race)	Mean	377.05	370.71	389.43	372.89	375.26	380.38	372.72	378.66
		N	66,186	66,208	66,086	66,090	65,794	65,959	66,067	65,587
	Non-Hispanic American Indian	Mean	388.48	368.28	393.88	382.28	385.52	381.48	374.51	382.46
		N	1,862	1,865	1,846	1,859	1,842	1,844	1,858	1,824
	Non-Hispanic Multi-racial	Mean	384.27	377.06	392.90	383.96	384.55	385.30	379.29	384.90
		N	281	281	280	283	281	280	281	280
Non-Hispanic White	Mean	386.48	377.77	394.81	385.99	386.50	386.59	380.51	386.37	
	N	7062	7065	7060	7061	7044	7056	7056	7034	
Missing	Mean	357.73	362.00	375.93	348.91	353.54	369.38	360.78	364.48	
	N	2,525	2,533	2,516	2,499	2,482	2,509	2,520	2,468	
10	Non-Hispanic Asian	Mean	389.71	381.72	399.45	377.55	383.88	390.88	384.22	388.55
		N	8,563	8,573	8,562	8,552	8,539	8,557	8,560	8,527
	Non-Hispanic Pacific Islander	Mean	379.41	368.75	394.20	380.97	380.48	381.71	372.09	381.13
		N	661	662	656	657	656	655	659	651
	Non-Hispanic Black	Mean	384.44	376.43	394.39	382.11	383.57	385.71	378.94	384.86
		N	4,785	4,782	4,782	4,781	4,768	4,765	4,770	4,746
	Hispanic (Of Any Race)	Mean	384.32	375.01	393.73	380.02	382.45	384.68	377.89	383.83
		N	38,522	38,551	38,460	38,445	38,250	38,374	38,457	38,112
	Non-Hispanic American Indian	Mean	392.00	372.75	397.04	386.38	389.47	385.26	378.63	386.36
		N	1,241	1,239	1,231	1,237	1,227	1,227	1,235	1,217
	Non-Hispanic Multi-racial	Mean	397.83	384.35	401.32	393.36	396.01	393.29	388.20	393.68
		N	166	167	166	166	165	166	166	165
Non-Hispanic White	Mean	394.51	381.62	397.81	391.21	393.13	389.97	385.58	390.69	
	N	5,369	5,368	5,361	5,363	5,357	5,358	5,364	5,347	
Missing	Mean	374.00	370.96	386.25	364.47	369.46	378.91	372.02	375.97	
	N	1,265	1,270	1,264	1,245	1,231	1,257	1,261	1,225	

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
11	Non-Hispanic Asian	Mean	393.97	385.51	402.30	380.79	387.72	394.19	388.13	392.07
		N	7,629	7,640	7,627	7,629	7,604	7,621	7,623	7,591
	Non-Hispanic Pacific Islander	Mean	387.27	374.94	398.64	389.74	388.84	387.28	378.91	387.47
		N	430	431	425	431	428	425	429	425
	Non-Hispanic Black	Mean	386.90	379.50	397.04	384.45	385.99	388.56	381.82	387.61
		N	4,042	4,037	4,041	4,033	4,021	4,029	4,027	4,008
	Hispanic (Of Any Race)	Mean	389.86	379.32	397.44	386.09	388.27	388.67	382.57	388.39
		N	25,419	25,436	25,403	25,366	25,216	25,332	25,350	25,116
	Non-Hispanic American Indian	Mean	395.19	374.84	397.40	389.53	393.02	386.49	381.00	388.34
		N	926	925	920	929	918	919	923	911
	Non-Hispanic Multi-racial	Mean	395.22	384.83	399.38	394.53	395.16	392.49	388.28	393.10
		N	116	117	116	116	116	116	116	116
	Non-Hispanic White	Mean	402.67	387.66	402.83	397.64	400.56	395.54	392.26	396.83
		N	4,299	4,302	4,303	4,305	4,284	4,297	4,289	4,277
Missing	Mean	378.84	376.25	391.14	370.30	375.21	383.97	377.18	381.38	
	N	934	934	934	912	910	932	931	907	
12	Non-Hispanic Asian	Mean	393.81	384.52	402.37	381.22	387.80	393.73	387.45	391.78
		N	6,260	6,271	6,263	6,246	6,221	6,254	6,252	6,206
	Non-Hispanic Pacific Islander	Mean	381.00	372.98	398.04	387.91	384.68	386.02	375.64	385.59
		N	350	351	349	355	349	349	348	345
	Non-Hispanic Black	Mean	388.58	379.99	398.33	388.27	388.79	389.44	382.80	389.15
		N	3,372	3,381	3,372	3,366	3,345	3,363	3,358	3,329
	Hispanic (Of Any Race)	Mean	393.26	381.87	399.83	392.92	393.41	391.17	385.44	391.69
		N	18,109	18,137	18,077	18,121	17,954	18,024	18,052	17,859
	Non-Hispanic American Indian	Mean	395.85	374.65	398.94	382.13	389.47	387.25	381.06	387.47
		N	723	722	712	719	706	710	718	696
	Non-Hispanic Multi-racial	Mean	401.47	390.93	407.20	406.70	404.56	399.30	394.45	400.78
		N	87	89	89	89	86	89	87	86
	Non-Hispanic White	Mean	401.65	386.99	402.03	398.76	400.48	394.79	391.49	396.30
		N	3,717	3,724	3,720	3,718	3,706	3,716	3,712	3,699
Missing	Mean	382.98	378.09	393.88	374.00	379.13	386.38	379.67	384.42	
	N	713	713	705	693	688	702	710	683	

4.2.3 Correlations among Scale Scores by Grade-Level Cluster

Table 4.2.3A

Correlations Among Scale Scores: K S303

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.536	.556	.784
	N	218,288	218,256	218,256	218,213
Reading	Pearson Correlation		1	.724	.498
	N		218,268	218,252	218,195
Writing	Pearson Correlation			1	.555
	N			218,272	218,197
Speaking	Pearson Correlation				1
	N				218,222

Table 4.2.3B

Correlations Among Scale Scores: 1-2 S303

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.703	.584	.518
	N	438,609	438,295	438,287	438,181
Reading	Pearson Correlation		1	.669	.460
	N		438,412	438,156	437,980
Writing	Pearson Correlation			1	.489
	N			438,453	438,056
Speaking	Pearson Correlation				1
	N				438,394

Table 4.2.3C

Correlations Among Scale Scores: 3-5 S303

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.744	.624	.506
	N	418,276	417,816	417,811	417,843
Reading	Pearson Correlation		1	.683	.501
	N		417,952	417,580	417,478
Writing	Pearson Correlation			1	.539
	N			417,985	417,566
Speaking	Pearson Correlation				1
	N				418,023

Table 4.2.3D

Correlations Among Scale Scores: 6-8 S303

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.699	.572	.599
	N	250,354	250,056	249,966	249,857
Reading	Pearson Correlation		1	.619	.514
	N		250,247	249,895	249,723
Writing	Pearson Correlation			1	.526
	N			250,199	249,748
Speaking	Pearson Correlation				1
	N				250,160

Table 4.2.3E

Correlations Among Scale Scores: 9-12 S303

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.718	.660	.625
	N	235,149	234,669	234,177	233,665
Reading	Pearson Correlation		1	.683	.555
	N		235,288	234,365	233,693
Writing	Pearson Correlation			1	.622
	N			234,846	233,498
Speaking	Pearson Correlation				1
	N				234,787

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 S303

Cluster	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	25,329	11,903	21,613	34,086	67,508	57,849	218,288
K (accountability)	-	52,336	21,616	18,979	12,635	34,465	78,257	218,288
1-2	A	8,179	13,081	22,973	50,192	n/a	n/a	94,425
	B	1,028	2,848	9,196	8,840	199,792	n/a	221,704
	C	451	2,767	16,899	10,864	32,417	59,082	122,480
3-5	A	3,046	11,652	10,789	12,705	n/a	n/a	38,192
	B	768	6,344	22,887	23,006	112,276	n/a	165,281
	C	318	1,668	13,522	18,018	54,631	126,646	214,803
6-8	A	8,347	12,792	5,946	4,534	n/a	n/a	31,619
	B	1,247	11,825	20,255	22,852	30,663	n/a	86,842
	C	340	790	8,451	19,358	44,683	58,271	131,893
9-12	A	19,201	14,004	3,104	2,631	n/a	n/a	38,940
	B	3,540	10,423	22,108	21,106	26,021	n/a	83,198
	C	1,399	4,708	15,739	33,522	30,659	26,984	113,011

Table 4.3.1.1B

Proficiency Level by Cluster by Tier (Percent): Listening S303

Cluster	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	11.6%	5.5%	9.9%	15.6%	30.9%	26.5%	100.0%
K (accountability)	-	24.0%	9.9%	8.7%	5.8%	15.8%	35.9%	100.0%
1-2	A	8.7%	13.9%	24.3%	53.2%	n/a	n/a	100.0%
	B	0.5%	1.3%	4.1%	4.0%	90.1%	n/a	100.0%
	C	0.4%	2.3%	13.8%	8.9%	26.5%	48.2%	100.0%
3-5	A	8.0%	30.5%	28.2%	33.3%	n/a	n/a	100.0%
	B	0.5%	3.8%	13.8%	13.9%	67.9%	n/a	100.0%
	C	0.1%	0.8%	6.3%	8.4%	25.4%	59.0%	100.0%
6-8	A	26.4%	40.5%	18.8%	14.3%	n/a	n/a	100.0%
	B	1.4%	13.6%	23.3%	26.3%	35.3%	n/a	100.0%
	C	0.3%	0.6%	6.4%	14.7%	33.9%	44.2%	100.0%
9-12	A	49.3%	36.0%	8.0%	6.8%	n/a	n/a	100.0%
	B	4.3%	12.5%	26.6%	25.4%	31.3%	n/a	100.0%
	C	1.2%	4.2%	13.9%	29.7%	27.1%	23.9%	100.0%

4.3.1.2 By Grade by Tier

Table 4.3.1.2A

Proficiency Level by Grade by Tier (Count): Listening S303

Grade	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	25,329	11,903	21,613	34,086	67,508	57,849	218,288
K (accountability)	-	52,336	21,616	18,979	12,635	34,465	78,257	218,288
1	A	5,124	8,403	18,904	40,795	n/a	n/a	73,226
	B	620	1,936	4,293	4,332	95,633	n/a	106,814
	C	227	1,094	8,469	4,403	10,961	17,405	42,559
2	A	3,055	4,678	4,069	9,397	n/a	n/a	21,199
	B	408	912	4,903	4,508	104,159	n/a	114,890
	C	224	1,673	8,430	6,461	21,456	41,677	79,921
3	A	783	4,285	4,708	6,019	n/a	n/a	15,795
	B	286	2,945	11,845	7,472	63,489	n/a	86,037
	C	119	518	4,680	3,790	21,800	61,909	92,816
4	A	998	3,702	3,374	3,794	n/a	n/a	11,868
	B	232	1,933	6,366	8,677	28,314	n/a	45,522
	C	109	595	3,841	6,838	20,549	36,138	68,070
5	A	1,265	3,665	2,707	2,892	n/a	n/a	10,529
	B	250	1,466	4,676	6,857	20,473	n/a	33,722
	C	90	555	5,001	7,390	12,282	28,599	53,917
6	A	1,750	4,395	2,533	1,890	n/a	n/a	10,568
	B	238	3,306	7,127	7,105	10,417	n/a	28,193
	C	109	226	3,528	6,290	16,264	17,476	43,893
7	A	2,720	4,254	2,090	1,314	n/a	n/a	10,378
	B	399	3,863	8,018	8,011	9,625	n/a	29,916
	C	118	257	3,111	5,515	16,881	17,751	43,633
8	A	3,877	4,143	1,323	1,330	n/a	n/a	10,673
	B	610	4,656	5,110	7,736	10,621	n/a	28,733
	C	113	307	1,812	7,553	11,538	23,044	44,367
9	A	10,321	9,212	1,134	1,553	n/a	n/a	22,220
	B	785	3,792	8,375	7,527	11,435	n/a	31,914
	C	299	923	4,066	11,286	16,521	10,222	43,317
10	A	4,505	3,126	1,083	504	n/a	n/a	9,218
	B	888	2,601	6,466	5,231	7,475	n/a	22,661
	C	276	1,341	4,424	9,052	6,874	6,726	28,693
11	A	2,731	1,174	723	353	n/a	n/a	4,981
	B	811	2,577	3,203	5,535	3,965	n/a	16,091
	C	339	1,021	3,467	6,975	5,268	5,653	22,723
12	A	1,644	492	164	221	n/a	n/a	2,521
	B	1,056	1,453	4,064	2,813	3,146	n/a	12,532
	C	485	1,423	3,782	6,209	1,996	4,383	18,278

Table 4.3.1.2B

Proficiency Level by Grade by Tier (Percent): Listening S303

Grade	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	11.6%	5.5%	9.9%	15.6%	30.9%	26.5%	100.0%
K (accountability)	-	24.0%	9.9%	8.7%	5.8%	15.8%	35.9%	100.0%
1	A	7.0%	11.5%	25.8%	55.7%	n/a	n/a	100.0%
	B	0.6%	1.8%	4.0%	4.1%	89.5%	n/a	100.0%
	C	0.5%	2.6%	19.9%	10.3%	25.8%	40.9%	100.0%
2	A	14.4%	22.1%	19.2%	44.3%	n/a	n/a	100.0%
	B	0.4%	0.8%	4.3%	3.9%	90.7%	n/a	100.0%
	C	0.3%	2.1%	10.5%	8.1%	26.8%	52.1%	100.0%
3	A	5.0%	27.1%	29.8%	38.1%	n/a	n/a	100.0%
	B	0.3%	3.4%	13.8%	8.7%	73.8%	n/a	100.0%
	C	0.1%	0.6%	5.0%	4.1%	23.5%	66.7%	100.0%
4	A	8.4%	31.2%	28.4%	32.0%	n/a	n/a	100.0%
	B	0.5%	4.2%	14.0%	19.1%	62.2%	n/a	100.0%
	C	0.2%	0.9%	5.6%	10.0%	30.2%	53.1%	100.0%
5	A	12.0%	34.8%	25.7%	27.5%	n/a	n/a	100.0%
	B	0.7%	4.3%	13.9%	20.3%	60.7%	n/a	100.0%
	C	0.2%	1.0%	9.3%	13.7%	22.8%	53.0%	100.0%
6	A	16.6%	41.6%	24.0%	17.9%	n/a	n/a	100.0%
	B	0.8%	11.7%	25.3%	25.2%	36.9%	n/a	100.0%
	C	0.2%	0.5%	8.0%	14.3%	37.1%	39.8%	100.0%
7	A	26.2%	41.0%	20.1%	12.7%	n/a	n/a	100.0%
	B	1.3%	12.9%	26.8%	26.8%	32.2%	n/a	100.0%
	C	0.3%	0.6%	7.1%	12.6%	38.7%	40.7%	100.0%
8	A	36.3%	38.8%	12.4%	12.5%	n/a	n/a	100.0%
	B	2.1%	16.2%	17.8%	26.9%	37.0%	n/a	100.0%
	C	0.3%	0.7%	4.1%	17.0%	26.0%	51.9%	100.0%
9	A	46.4%	41.5%	5.1%	7.0%	n/a	n/a	100.0%
	B	2.5%	11.9%	26.2%	23.6%	35.8%	n/a	100.0%
	C	0.7%	2.1%	9.4%	26.1%	38.1%	23.6%	100.0%
10	A	48.9%	33.9%	11.7%	5.5%	n/a	n/a	100.0%
	B	3.9%	11.5%	28.5%	23.1%	33.0%	n/a	100.0%
	C	1.0%	4.7%	15.4%	31.5%	24.0%	23.4%	100.0%
11	A	54.8%	23.6%	14.5%	7.1%	n/a	n/a	100.0%
	B	5.0%	16.0%	19.9%	34.4%	24.6%	n/a	100.0%
	C	1.5%	4.5%	15.3%	30.7%	23.2%	24.9%	100.0%
12	A	65.2%	19.5%	6.5%	8.8%	n/a	n/a	100.0%
	B	8.4%	11.6%	32.4%	22.4%	25.1%	n/a	100.0%
	C	2.7%	7.8%	20.7%	34.0%	10.9%	24.0%	100.0%

4.3.1.3 By Grade

Table 4.3.1.3A

Proficiency Level by Grade (Count): Listening S303

	Listening Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	25,329	11,903	21,613	34,086	67,508	57,849	218,288
K (accountability)	52,336	21,616	18,979	12,635	34,465	78,257	218,288
1	5,971	11,433	31,666	49,530	106,594	17,405	222,599
2	3,687	7,263	17,402	20,366	125,615	41,677	216,010
3	1,188	7,748	21,233	17,281	85,289	61,909	194,648
4	1,339	6,230	13,581	19,309	48,863	36,138	125,460
5	1,605	5,686	12,384	17,139	32,755	28,599	98,168
6	2,097	7,927	13,188	15,285	26,681	17,476	82,654
7	3,237	8,374	13,219	14,840	26,506	17,751	83,927
8	4,600	9,106	8,245	16,619	22,159	23,044	83,773
9	11,405	13,927	13,575	20,366	27,956	10,222	97,451
10	5,669	7,068	11,973	14,787	14,349	6,726	60,572
11	3,881	4,772	7,393	12,863	9,233	5,653	43,795
12	3,185	3,368	8,010	9,243	5,142	4,383	33,331

Table 4.3.1.3B

Proficiency Level by Grade (Percent): Listening S303

	Listening Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	11.6%	5.5%	9.9%	15.6%	30.9%	26.5%	100.0%
K (accountability)	24.0%	9.9%	8.7%	5.8%	15.8%	35.9%	100.0%
1	2.7%	5.1%	14.2%	22.3%	47.9%	7.8%	100.0%
2	1.7%	3.4%	8.1%	9.4%	58.2%	19.3%	100.0%
3	0.6%	4.0%	10.9%	8.9%	43.8%	31.8%	100.0%
4	1.1%	5.0%	10.8%	15.4%	38.9%	28.8%	100.0%
5	1.6%	5.8%	12.6%	17.5%	33.4%	29.1%	100.0%
6	2.5%	9.6%	16.0%	18.5%	32.3%	21.1%	100.0%
7	3.9%	10.0%	15.8%	17.7%	31.6%	21.2%	100.0%
8	5.5%	10.9%	9.8%	19.8%	26.5%	27.5%	100.0%
9	11.7%	14.3%	13.9%	20.9%	28.7%	10.5%	100.0%
10	9.4%	11.7%	19.8%	24.4%	23.7%	11.1%	100.0%
11	8.9%	10.9%	16.9%	29.4%	21.1%	12.9%	100.0%
12	9.6%	10.1%	24.0%	27.7%	15.4%	13.1%	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 S303

Cluster	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	47,510	29,483	41,583	18,226	22,306	59,160	218,268
K (accountability)	-	143,727	15,381	10,489	12,155	36,516	0	218,268
1-2	A	20,697	28,236	22,184	23,255	n/a	n/a	94,372
	B	2,908	11,736	26,564	39,707	140,673	n/a	221,588
	C	1,045	4,047	16,010	19,202	34,595	47,553	122,452
3-5	A	11,027	13,551	7,993	5,550	n/a	n/a	38,121
	B	2,253	16,897	36,108	18,693	91,132	n/a	165,083
	C	613	2,950	21,509	18,975	77,910	92,791	214,748
6-8	A	13,041	12,711	3,991	1,867	n/a	n/a	31,610
	B	4,397	26,519	25,402	7,600	22,867	n/a	86,785
	C	1,808	22,097	41,285	14,149	30,892	21,621	131,852
9-12	A	15,491	15,240	4,168	4,077	n/a	n/a	38,976
	B	11,278	31,161	16,132	6,900	17,765	n/a	83,236
	C	1,826	14,490	17,389	12,275	25,383	41,713	113,076

Table 4.3.2.1B

Proficiency Level by Cluster by Tier (Percent): Reading S303

Cluster	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	21.8%	13.5%	19.1%	8.4%	10.2%	27.1%	100.0%
K (accountability)	-	65.8%	7.0%	4.8%	5.6%	16.7%	0.0%	100.0%
1-2	A	21.9%	29.9%	23.5%	24.6%	n/a	n/a	100.0%
	B	1.3%	5.3%	12.0%	17.9%	63.5%	n/a	100.0%
	C	0.9%	3.3%	13.1%	15.7%	28.3%	38.8%	100.0%
3-5	A	28.9%	35.5%	21.0%	14.6%	n/a	n/a	100.0%
	B	1.4%	10.2%	21.9%	11.3%	55.2%	n/a	100.0%
	C	0.3%	1.4%	10.0%	8.8%	36.3%	43.2%	100.0%
6-8	A	41.3%	40.2%	12.6%	5.9%	n/a	n/a	100.0%
	B	5.1%	30.6%	29.3%	8.8%	26.3%	n/a	100.0%
	C	1.4%	16.8%	31.3%	10.7%	23.4%	16.4%	100.0%
9-12	A	39.7%	39.1%	10.7%	10.5%	n/a	n/a	100.0%
	B	13.5%	37.4%	19.4%	8.3%	21.3%	n/a	100.0%
	C	1.6%	12.8%	15.4%	10.9%	22.4%	36.9%	100.0%

4.3.2.2 By Grade by Tier

Table 4.3.2.2A

Proficiency Level by Grade by Tier (Count): Reading S303

Grade	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	47,510	29,483	41,583	18,226	22,306	59,160	218,268
K (accountability)	-	143,727	15,381	10,489	12,155	36,516	0	218,268
1	A	14,569	21,886	18,498	18,238	n/a	n/a	73,191
	B	1,943	3,587	10,379	19,654	71,186	n/a	106,749
	C	511	1,023	3,193	9,939	13,193	14,667	42,526
2	A	6,128	6,350	3,686	5,017	n/a	n/a	21,181
	B	965	8,149	16,185	20,053	69,487	n/a	114,839
	C	534	3,024	12,817	9,263	21,402	32,886	79,926
3	A	3,047	5,787	3,828	3,101	n/a	n/a	15,763
	B	655	4,459	13,002	12,689	55,168	n/a	85,973
	C	285	507	3,910	4,624	34,520	48,963	92,809
4	A	3,610	4,452	2,271	1,516	n/a	n/a	11,849
	B	574	6,668	11,963	3,509	22,732	n/a	45,446
	C	152	645	6,694	10,417	24,051	26,085	68,044
5	A	4,370	3,312	1,894	933	n/a	n/a	10,509
	B	1,024	5,770	11,143	2,495	13,232	n/a	33,664
	C	176	1,798	10,905	3,934	19,339	17,743	53,895
6	A	2,824	4,963	1,858	915	n/a	n/a	10,560
	B	904	7,732	8,752	3,709	7,068	n/a	28,165
	C	467	6,802	15,826	3,737	10,896	6,146	43,874
7	A	4,720	3,805	1,329	522	n/a	n/a	10,376
	B	1,168	9,428	8,650	2,020	8,632	n/a	29,898
	C	645	8,606	13,491	3,418	10,817	6,648	43,625
8	A	5,497	3,943	804	430	n/a	n/a	10,674
	B	2,325	9,359	8,000	1,871	7,167	n/a	28,722
	C	696	6,689	11,968	6,994	9,179	8,827	44,353
9	A	8,979	8,832	2,672	1,742	n/a	n/a	22,225
	B	3,022	10,514	9,256	1,855	7,264	n/a	31,911
	C	419	2,955	8,108	5,395	8,865	17,588	43,330
10	A	3,402	3,700	954	1,168	n/a	n/a	9,224
	B	3,230	9,781	3,006	2,509	4,150	n/a	22,676
	C	471	4,297	4,822	1,836	7,519	9,767	28,712
11	A	1,936	1,936	262	859	n/a	n/a	4,993
	B	2,839	5,700	2,250	1,819	3,485	n/a	16,093
	C	370	4,128	2,481	2,839	4,375	8,543	22,736
12	A	1,174	772	280	308	n/a	n/a	2,534
	B	2,187	5,166	1,620	717	2,866	n/a	12,556
	C	566	3,110	1,978	2,205	4,624	5,815	18,298

Table 4.3.2.2B

Proficiency Level by Grade by Tier (Percent): Reading S303

Grade	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	21.8%	13.5%	19.1%	8.4%	10.2%	27.1%	100.0%
K (accountability)	-	65.8%	7.0%	4.8%	5.6%	16.7%	0.0%	100.0%
1	A	19.9%	29.9%	25.3%	24.9%	n/a	n/a	100.0%
	B	1.8%	3.4%	9.7%	18.4%	66.7%	n/a	100.0%
	C	1.2%	2.4%	7.5%	23.4%	31.0%	34.5%	100.0%
2	A	28.9%	30.0%	17.4%	23.7%	n/a	n/a	100.0%
	B	0.8%	7.1%	14.1%	17.5%	60.5%	n/a	100.0%
	C	0.7%	3.8%	16.0%	11.6%	26.8%	41.1%	100.0%
3	A	19.3%	36.7%	24.3%	19.7%	n/a	n/a	100.0%
	B	0.8%	5.2%	15.1%	14.8%	64.2%	n/a	100.0%
	C	0.3%	0.5%	4.2%	5.0%	37.2%	52.8%	100.0%
4	A	30.5%	37.6%	19.2%	12.8%	n/a	n/a	100.0%
	B	1.3%	14.7%	26.3%	7.7%	50.0%	n/a	100.0%
	C	0.2%	0.9%	9.8%	15.3%	35.3%	38.3%	100.0%
5	A	41.6%	31.5%	18.0%	8.9%	n/a	n/a	100.0%
	B	3.0%	17.1%	33.1%	7.4%	39.3%	n/a	100.0%
	C	0.3%	3.3%	20.2%	7.3%	35.9%	32.9%	100.0%
6	A	26.7%	47.0%	17.6%	8.7%	n/a	n/a	100.0%
	B	3.2%	27.5%	31.1%	13.2%	25.1%	n/a	100.0%
	C	1.1%	15.5%	36.1%	8.5%	24.8%	14.0%	100.0%
7	A	45.5%	36.7%	12.8%	5.0%	n/a	n/a	100.0%
	B	3.9%	31.5%	28.9%	6.8%	28.9%	n/a	100.0%
	C	1.5%	19.7%	30.9%	7.8%	24.8%	15.2%	100.0%
8	A	51.5%	36.9%	7.5%	4.0%	n/a	n/a	100.0%
	B	8.1%	32.6%	27.9%	6.5%	25.0%	n/a	100.0%
	C	1.6%	15.1%	27.0%	15.8%	20.7%	19.9%	100.0%
9	A	40.4%	39.7%	12.0%	7.8%	n/a	n/a	100.0%
	B	9.5%	32.9%	29.0%	5.8%	22.8%	n/a	100.0%
	C	1.0%	6.8%	18.7%	12.5%	20.5%	40.6%	100.0%
10	A	36.9%	40.1%	10.3%	12.7%	n/a	n/a	100.0%
	B	14.2%	43.1%	13.3%	11.1%	18.3%	n/a	100.0%
	C	1.6%	15.0%	16.8%	6.4%	26.2%	34.0%	100.0%
11	A	38.8%	38.8%	5.2%	17.2%	n/a	n/a	100.0%
	B	17.6%	35.4%	14.0%	11.3%	21.7%	n/a	100.0%
	C	1.6%	18.2%	10.9%	12.5%	19.2%	37.6%	100.0%
12	A	46.3%	30.5%	11.0%	12.2%	n/a	n/a	100.0%
	B	17.4%	41.1%	12.9%	5.7%	22.8%	n/a	100.0%
	C	3.1%	17.0%	10.8%	12.1%	25.3%	31.8%	100.0%

4.3.2.3 By Grade

Table 4.3.2.3A

Proficiency Level by Grade (Count): Reading S303

	Reading Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	47,510	29,483	41,583	18,226	22,306	59,160	218,268
K (accountability)	143,727	15,381	10,489	12,155	36,516	0	218,268
1	17,023	26,496	32,070	47,831	84,379	14,667	222,466
2	7,627	17,523	32,688	34,333	90,889	32,886	215,946
3	3,987	10,753	20,740	20,414	89,688	48,963	194,545
4	4,336	11,765	20,928	15,442	46,783	26,085	125,339
5	5,570	10,880	23,942	7,362	32,571	17,743	98,068
6	4,195	19,497	26,436	8,361	17,964	6,146	82,599
7	6,533	21,839	23,470	5,960	19,449	6,648	83,899
8	8,518	19,991	20,772	9,295	16,346	8,827	83,749
9	12,420	22,301	20,036	8,992	16,129	17,588	97,466
10	7,103	17,778	8,782	5,513	11,669	9,767	60,612
11	5,145	11,764	4,993	5,517	7,860	8,543	43,822
12	3,927	9,048	3,878	3,230	7,490	5,815	33,388

Table 4.3.2.3B

Proficiency Level by Grade (Percent): Reading S303

	Reading Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	21.8%	13.5%	19.1%	8.4%	10.2%	27.1%	100.0%
K (accountability)	65.8%	7.0%	4.8%	5.6%	16.7%	0.0%	100.0%
1	7.7%	11.9%	14.4%	21.5%	37.9%	6.6%	100.0%
2	3.5%	8.1%	15.1%	15.9%	42.1%	15.2%	100.0%
3	2.0%	5.5%	10.7%	10.5%	46.1%	25.2%	100.0%
4	3.5%	9.4%	16.7%	12.3%	37.3%	20.8%	100.0%
5	5.7%	11.1%	24.4%	7.5%	33.2%	18.1%	100.0%
6	5.1%	23.6%	32.0%	10.1%	21.7%	7.4%	100.0%
7	7.8%	26.0%	28.0%	7.1%	23.2%	7.9%	100.0%
8	10.2%	23.9%	24.8%	11.1%	19.5%	10.5%	100.0%
9	12.7%	22.9%	20.6%	9.2%	16.5%	18.0%	100.0%
10	11.7%	29.3%	14.5%	9.1%	19.3%	16.1%	100.0%
11	11.7%	26.8%	11.4%	12.6%	17.9%	19.5%	100.0%
12	11.8%	27.1%	11.6%	9.7%	22.4%	17.4%	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 S303

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	37,235	63,948	36,315	31,246	42,997	6,531	218,272
K (accountability)	-	124,445	44,299	28,448	14,549	6,531	0	218,272
1-2	A	14,677	60,309	19,425	0	0	0	94,411
	B	12,972	88,157	117,138	3,315	1	0	221,583
	C	2,738	21,608	82,593	15,488	32	0	122,459
3-5	A	5,416	10,197	15,498	6,878	110	0	38,099
	B	1,999	10,452	51,253	95,866	5,575	30	165,175
	C	517	2,269	14,539	140,994	55,008	1,384	214,711
6-8	A	6,189	12,137	12,007	1,259	20	0	31,612
	B	3,623	12,612	48,899	21,335	289	0	86,758
	C	1,990	6,916	78,953	43,518	449	3	131,829
9-12	A	6,340	13,821	16,850	1,859	24	1	38,895
	B	4,759	5,755	30,920	35,304	6,157	231	83,126
	C	2,272	2,402	17,487	56,763	31,197	2,704	112,825

Table 4.3.3.1B

Proficiency Level by Cluster by Tier (Percent): Writing S303

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	17.1%	29.3%	16.6%	14.3%	19.7%	3.0%	100.0%
K (accountability)	-	57.0%	20.3%	13.0%	6.7%	3.0%	0.0%	100.0%
1-2	A	15.5%	63.9%	20.6%	0.0%	0.0%	0.0%	100.0%
	B	5.9%	39.8%	52.9%	1.5%	0.0%	0.0%	100.0%
	C	2.2%	17.6%	67.4%	12.6%	0.0%	0.0%	100.0%
3-5	A	14.2%	26.8%	40.7%	18.1%	0.3%	0.0%	100.0%
	B	1.2%	6.3%	31.0%	58.0%	3.4%	0.0%	100.0%
	C	0.2%	1.1%	6.8%	65.7%	25.6%	0.6%	100.0%
6-8	A	19.6%	38.4%	38.0%	4.0%	0.1%	0.0%	100.0%
	B	4.2%	14.5%	56.4%	24.6%	0.3%	0.0%	100.0%
	C	1.5%	5.2%	59.9%	33.0%	0.3%	0.0%	100.0%
9-12	A	16.3%	35.5%	43.3%	4.8%	0.1%	0.0%	100.0%
	B	5.7%	6.9%	37.2%	42.5%	7.4%	0.3%	100.0%
	C	2.0%	2.1%	15.5%	50.3%	27.7%	2.4%	100.0%

4.3.3.2 By Grade by Tier

Table 4.3.3.2A

Proficiency Level by Grade by Tier (Count): Writing S303

Grade	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	37,235	63,948	36,315	31,246	42,997	6,531	218,272
K (accountability)	-	124,445	44,299	28,448	14,549	6,531	0	218,272
1	A	9,800	44,857	18,562	0	0	0	73,219
	B	7,671	47,148	49,551	2,376	0	0	106,746
	C	1,335	10,010	26,090	5,095	17	0	42,547
2	A	4,877	15,452	863	0	0	0	21,192
	B	5,301	41,009	67,587	939	1	0	114,837
	C	1,403	11,598	56,503	10,393	15	0	79,912
3	A	1,779	3,464	5,855	4,567	102	0	15,767
	B	760	5,211	18,092	57,246	4,645	24	85,978
	C	170	613	5,042	50,270	35,616	1,075	92,786
4	A	1,717	3,503	5,061	1,545	4	0	11,830
	B	594	3,211	15,994	24,917	773	3	45,492
	C	172	1,000	3,372	50,181	13,111	205	68,041
5	A	1,920	3,230	4,582	766	4	0	10,502
	B	645	2,030	17,167	13,703	157	3	33,705
	C	175	656	6,125	40,543	6,281	104	53,884
6	A	1,018	3,825	4,892	806	16	0	10,557
	B	982	3,184	13,940	9,861	197	0	28,164
	C	623	1,691	19,488	21,740	322	1	43,865
7	A	1,973	4,073	3,963	364	4	0	10,377
	B	1,182	3,542	16,377	8,707	77	0	29,885
	C	702	2,301	26,209	14,309	90	2	43,613
8	A	3,198	4,239	3,152	89	0	0	10,678
	B	1,459	5,886	18,582	2,767	15	0	28,709
	C	665	2,924	33,256	7,469	37	0	44,351
9	A	3,374	8,374	9,045	1,369	17	0	22,179
	B	1,259	2,217	7,973	15,837	4,425	156	31,867
	C	582	792	2,802	16,844	20,183	2,059	43,262
10	A	1,393	2,768	4,659	380	5	0	9,205
	B	1,459	1,565	8,745	9,648	1,174	53	22,644
	C	514	573	3,855	16,249	7,010	432	28,633
11	A	921	1,763	2,204	95	1	0	4,984
	B	1,014	1,102	7,177	6,380	399	15	16,087
	C	545	493	5,164	13,459	2,883	154	22,698
12	A	652	916	942	15	1	1	2,527
	B	1,027	871	7,025	3,439	159	7	12,528
	C	631	544	5,666	10,211	1,121	59	18,232

Table 4.3.3.2B

Proficiency Level by Grade by Tier (Percent): Writing S303

Grade	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	17.1%	29.3%	16.6%	14.3%	19.7%	3.0%	100.0%
K (accountability)	-	57.0%	20.3%	13.0%	6.7%	3.0%	0.0%	100.0%
1	A	13.4%	61.3%	25.4%	0.0%	0.0%	0.0%	100.0%
	B	7.2%	44.2%	46.4%	2.2%	0.0%	0.0%	100.0%
	C	3.1%	23.5%	61.3%	12.0%	0.0%	0.0%	100.0%
2	A	23.0%	72.9%	4.1%	0.0%	0.0%	0.0%	100.0%
	B	4.6%	35.7%	58.9%	0.8%	0.0%	0.0%	100.0%
	C	1.8%	14.5%	70.7%	13.0%	0.0%	0.0%	100.0%
3	A	11.3%	22.0%	37.1%	29.0%	0.6%	0.0%	100.0%
	B	0.9%	6.1%	21.0%	66.6%	5.4%	0.0%	100.0%
	C	0.2%	0.7%	5.4%	54.2%	38.4%	1.2%	100.0%
4	A	14.5%	29.6%	42.8%	13.1%	0.0%	0.0%	100.0%
	B	1.3%	7.1%	35.2%	54.8%	1.7%	0.0%	100.0%
	C	0.3%	1.5%	5.0%	73.8%	19.3%	0.3%	100.0%
5	A	18.3%	30.8%	43.6%	7.3%	0.0%	0.0%	100.0%
	B	1.9%	6.0%	50.9%	40.7%	0.5%	0.0%	100.0%
	C	0.3%	1.2%	11.4%	75.2%	11.7%	0.2%	100.0%
6	A	9.6%	36.2%	46.3%	7.6%	0.2%	0.0%	100.0%
	B	3.5%	11.3%	49.5%	35.0%	0.7%	0.0%	100.0%
	C	1.4%	3.9%	44.4%	49.6%	0.7%	0.0%	100.0%
7	A	19.0%	39.3%	38.2%	3.5%	0.0%	0.0%	100.0%
	B	4.0%	11.9%	54.8%	29.1%	0.3%	0.0%	100.0%
	C	1.6%	5.3%	60.1%	32.8%	0.2%	0.0%	100.0%
8	A	29.9%	39.7%	29.5%	0.8%	0.0%	0.0%	100.0%
	B	5.1%	20.5%	64.7%	9.6%	0.1%	0.0%	100.0%
	C	1.5%	6.6%	75.0%	16.8%	0.1%	0.0%	100.0%
9	A	15.2%	37.8%	40.8%	6.2%	0.1%	0.0%	100.0%
	B	4.0%	7.0%	25.0%	49.7%	13.9%	0.5%	100.0%
	C	1.3%	1.8%	6.5%	38.9%	46.7%	4.8%	100.0%
10	A	15.1%	30.1%	50.6%	4.1%	0.1%	0.0%	100.0%
	B	6.4%	6.9%	38.6%	42.6%	5.2%	0.2%	100.0%
	C	1.8%	2.0%	13.5%	56.7%	24.5%	1.5%	100.0%
11	A	18.5%	35.4%	44.2%	1.9%	0.0%	0.0%	100.0%
	B	6.3%	6.9%	44.6%	39.7%	2.5%	0.1%	100.0%
	C	2.4%	2.2%	22.8%	59.3%	12.7%	0.7%	100.0%
12	A	25.8%	36.2%	37.3%	0.6%	0.0%	0.0%	100.0%
	B	8.2%	7.0%	56.1%	27.5%	1.3%	0.1%	100.0%
	C	3.5%	3.0%	31.1%	56.0%	6.1%	0.3%	100.0%

4.3.3.3 By Grade

Table 4.3.3.3A

Proficiency Level by Grade (Count): Writing S303

	Writing Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	37,235	63,948	36,315	31,246	42,997	6,531	218,272
K (accountability)	124,445	44,299	28,448	14,549	6,531	0	218,272
1	18,806	102,015	94,203	7,471	17	0	222,512
2	11,581	68,059	124,953	11,332	16	0	215,941
3	2,709	9,288	28,989	112,083	40,363	1,099	194,531
4	2,483	7,714	24,427	76,643	13,888	208	125,363
5	2,740	5,916	27,874	55,012	6,442	107	98,091
6	2,623	8,700	38,320	32,407	535	1	82,586
7	3,857	9,916	46,549	23,380	171	2	83,875
8	5,322	13,049	54,990	10,325	52	0	83,738
9	5,215	11,383	19,820	34,050	24,625	2,215	97,308
10	3,366	4,906	17,259	26,277	8,189	485	60,482
11	2,480	3,358	14,545	19,934	3,283	169	43,769
12	2,310	2,331	13,633	13,665	1,281	67	33,287

Table 4.3.3.3B

Proficiency Level by Grade (Percent): Writing S303

	Writing Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	17.1%	29.3%	16.6%	14.3%	19.7%	3.0%	100.0%
K (accountability)	57.0%	20.3%	13.0%	6.7%	3.0%	0.0%	100.0%
1	8.5%	45.8%	42.3%	3.4%	0.0%	0.0%	100.0%
2	5.4%	31.5%	57.9%	5.2%	0.0%	0.0%	100.0%
3	1.4%	4.8%	14.9%	57.6%	20.7%	0.6%	100.0%
4	2.0%	6.2%	19.5%	61.1%	11.1%	0.2%	100.0%
5	2.8%	6.0%	28.4%	56.1%	6.6%	0.1%	100.0%
6	3.2%	10.5%	46.4%	39.2%	0.6%	0.0%	100.0%
7	4.6%	11.8%	55.5%	27.9%	0.2%	0.0%	100.0%
8	6.4%	15.6%	65.7%	12.3%	0.1%	0.0%	100.0%
9	5.4%	11.7%	20.4%	35.0%	25.3%	2.3%	100.0%
10	5.6%	8.1%	28.5%	43.4%	13.5%	0.8%	100.0%
11	5.7%	7.7%	33.2%	45.5%	7.5%	0.4%	100.0%
12	6.9%	7.0%	41.0%	41.1%	3.8%	0.2%	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 S303

Cluster	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	47,924	16,688	33,273	36,538	24,409	59,390	218,222
K (accountability)	-	47,924	49,961	36,538	24,409	59,390	0	218,222
1-2	A	21,914	32,333	14,836	5,573	4,344	15,366	94,366
	B	7,890	41,550	43,744	22,793	20,051	85,541	221,569
	C	1,505	8,538	14,039	10,303	11,691	76,383	122,459
3-5	A	20,447	8,999	3,634	1,295	1,032	2,759	38,166
	B	9,181	32,389	36,161	19,717	17,609	50,127	165,184
	C	3,706	18,197	31,357	25,775	29,864	105,774	214,673
6-8	A	17,389	6,532	3,482	1,829	660	1,717	31,609
	B	4,218	9,478	17,303	18,656	9,463	27,655	86,773
	C	1,492	4,477	13,803	25,536	18,360	68,110	131,778
9-12	A	25,374	6,273	3,034	1,505	730	1,950	38,866
	B	8,641	11,859	13,924	11,923	7,828	28,878	83,053
	C	2,506	4,128	10,317	14,451	13,355	68,111	112,868

Table 4.3.4.1B

Proficiency Level by Cluster by Tier (Percent): Speaking S303

Cluster	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	22.0%	7.6%	15.2%	16.7%	11.2%	27.2%	100.0%
K (accountability)	-	22.0%	22.9%	16.7%	11.2%	27.2%	0.0%	100.0%
1-2	A	23.2%	34.3%	15.7%	5.9%	4.6%	16.3%	100.0%
	B	3.6%	18.8%	19.7%	10.3%	9.0%	38.6%	100.0%
	C	1.2%	7.0%	11.5%	8.4%	9.5%	62.4%	100.0%
3-5	A	53.6%	23.6%	9.5%	3.4%	2.7%	7.2%	100.0%
	B	5.6%	19.6%	21.9%	11.9%	10.7%	30.3%	100.0%
	C	1.7%	8.5%	14.6%	12.0%	13.9%	49.3%	100.0%
6-8	A	55.0%	20.7%	11.0%	5.8%	2.1%	5.4%	100.0%
	B	4.9%	10.9%	19.9%	21.5%	10.9%	31.9%	100.0%
	C	1.1%	3.4%	10.5%	19.4%	13.9%	51.7%	100.0%
9-12	A	65.3%	16.1%	7.8%	3.9%	1.9%	5.0%	100.0%
	B	10.4%	14.3%	16.8%	14.4%	9.4%	34.8%	100.0%
	C	2.2%	3.7%	9.1%	12.8%	11.8%	60.3%	100.0%

4.3.4.2 By Grade by Tier

Table 4.3.4.2A

Proficiency Level by Grade by Tier (Count): Speaking S303

Grade	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	47,924	16,688	33,273	36,538	24,409	59,390	218,222
K (accountability)	-	47,924	49,961	36,538	24,409	59,390	0	218,222
1	A	13,797	27,125	12,087	4,569	3,522	12,091	73,191
	B	3,807	25,356	22,693	10,867	9,209	34,822	106,754
	C	493	4,500	6,071	3,971	4,219	23,307	42,561
2	A	8,117	5,208	2,749	1,004	822	3,275	21,175
	B	4,083	16,194	21,051	11,926	10,842	50,719	114,815
	C	1,012	4,038	7,968	6,332	7,472	53,076	79,898
3	A	7,650	4,565	1,542	503	415	1,105	15,780
	B	4,262	19,309	19,632	10,247	8,930	23,605	85,985
	C	1,507	9,630	14,589	11,242	12,736	43,070	92,774
4	A	6,737	2,424	1,126	420	323	828	11,858
	B	2,980	7,978	9,879	5,479	4,938	14,247	45,501
	C	1,312	5,302	10,123	8,563	9,741	32,984	68,025
5	A	6,060	2,010	966	372	294	826	10,528
	B	1,939	5,102	6,650	3,991	3,741	12,275	33,698
	C	887	3,265	6,645	5,970	7,387	29,720	53,874
6	A	5,006	2,532	1,395	809	227	585	10,554
	B	1,070	3,085	5,872	7,425	2,895	7,833	28,180
	C	522	1,688	5,093	10,738	5,999	19,806	43,846
7	A	6,060	1,641	1,186	701	211	587	10,386
	B	1,591	2,408	5,341	7,412	3,277	9,858	29,887
	C	516	1,006	3,625	9,514	6,112	22,824	43,597
8	A	6,323	2,359	901	319	222	545	10,669
	B	1,557	3,985	6,090	3,819	3,291	9,964	28,706
	C	454	1,783	5,085	5,284	6,249	25,480	44,335
9	A	15,828	2,670	1,564	964	326	818	22,170
	B	3,224	2,733	4,511	6,432	3,293	11,665	31,858
	C	887	796	2,537	6,899	5,328	26,838	43,285
10	A	5,609	1,891	746	288	186	491	9,211
	B	2,751	4,109	4,140	2,359	2,012	7,246	22,617
	C	624	1,365	3,188	3,067	3,419	16,955	28,618
11	A	2,742	1,098	462	158	126	389	4,975
	B	1,591	2,936	2,987	1,748	1,398	5,393	16,053
	C	514	1,075	2,500	2,561	2,605	13,438	22,693
12	A	1,195	614	262	95	92	252	2,510
	B	1,075	2,081	2,286	1,384	1,125	4,574	12,525
	C	481	892	2,092	1,924	2,003	10,880	18,272

Table 4.3.4.2B
Proficiency Level by Grade by Tier (Percent): Speaking S303

Grade	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	22.0%	7.6%	15.2%	16.7%	11.2%	27.2%	100.0%
K (accountability)	-	22.0%	22.9%	16.7%	11.2%	27.2%	0.0%	100.0%
1	A	18.9%	37.1%	16.5%	6.2%	4.8%	16.5%	100.0%
	B	3.6%	23.8%	21.3%	10.2%	8.6%	32.6%	100.0%
	C	1.2%	10.6%	14.3%	9.3%	9.9%	54.8%	100.0%
2	A	38.3%	24.6%	13.0%	4.7%	3.9%	15.5%	100.0%
	B	3.6%	14.1%	18.3%	10.4%	9.4%	44.2%	100.0%
	C	1.3%	5.1%	10.0%	7.9%	9.4%	66.4%	100.0%
3	A	48.5%	28.9%	9.8%	3.2%	2.6%	7.0%	100.0%
	B	5.0%	22.5%	22.8%	11.9%	10.4%	27.5%	100.0%
	C	1.6%	10.4%	15.7%	12.1%	13.7%	46.4%	100.0%
4	A	56.8%	20.4%	9.5%	3.5%	2.7%	7.0%	100.0%
	B	6.5%	17.5%	21.7%	12.0%	10.9%	31.3%	100.0%
	C	1.9%	7.8%	14.9%	12.6%	14.3%	48.5%	100.0%
5	A	57.6%	19.1%	9.2%	3.5%	2.8%	7.8%	100.0%
	B	5.8%	15.1%	19.7%	11.8%	11.1%	36.4%	100.0%
	C	1.6%	6.1%	12.3%	11.1%	13.7%	55.2%	100.0%
6	A	47.4%	24.0%	13.2%	7.7%	2.2%	5.5%	100.0%
	B	3.8%	10.9%	20.8%	26.3%	10.3%	27.8%	100.0%
	C	1.2%	3.8%	11.6%	24.5%	13.7%	45.2%	100.0%
7	A	58.3%	15.8%	11.4%	6.7%	2.0%	5.7%	100.0%
	B	5.3%	8.1%	17.9%	24.8%	11.0%	33.0%	100.0%
	C	1.2%	2.3%	8.3%	21.8%	14.0%	52.4%	100.0%
8	A	59.3%	22.1%	8.4%	3.0%	2.1%	5.1%	100.0%
	B	5.4%	13.9%	21.2%	13.3%	11.5%	34.7%	100.0%
	C	1.0%	4.0%	11.5%	11.9%	14.1%	57.5%	100.0%
9	A	71.4%	12.0%	7.1%	4.3%	1.5%	3.7%	100.0%
	B	10.1%	8.6%	14.2%	20.2%	10.3%	36.6%	100.0%
	C	2.0%	1.8%	5.9%	15.9%	12.3%	62.0%	100.0%
10	A	60.9%	20.5%	8.1%	3.1%	2.0%	5.3%	100.0%
	B	12.2%	18.2%	18.3%	10.4%	8.9%	32.0%	100.0%
	C	2.2%	4.8%	11.1%	10.7%	11.9%	59.2%	100.0%
11	A	55.1%	22.1%	9.3%	3.2%	2.5%	7.8%	100.0%
	B	9.9%	18.3%	18.6%	10.9%	8.7%	33.6%	100.0%
	C	2.3%	4.7%	11.0%	11.3%	11.5%	59.2%	100.0%
12	A	47.6%	24.5%	10.4%	3.8%	3.7%	10.0%	100.0%
	B	8.6%	16.6%	18.3%	11.0%	9.0%	36.5%	100.0%
	C	2.6%	4.9%	11.4%	10.5%	11.0%	59.5%	100.0%

4.3.4.3 By Grade

Table 4.3.4.3A

Proficiency Level by Grade (Count): Speaking S303

	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	47,924	16,688	33,273	36,538	24,409	59,390	218,222
K (accountability)	47,924	49,961	36,538	24,409	59,390	0	218,222
1	18,097	56,981	40,851	19,407	16,950	70,220	222,506
2	13,212	25,440	31,768	19,262	19,136	107,070	215,888
3	13,419	33,504	35,763	21,992	22,081	67,780	194,539
4	11,029	15,704	21,128	14,462	15,002	48,059	125,384
5	8,886	10,377	14,261	10,333	11,422	42,821	98,100
6	6,598	7,305	12,360	18,972	9,121	28,224	82,580
7	8,167	5,055	10,152	17,627	9,600	33,269	83,870
8	8,334	8,127	12,076	9,422	9,762	35,989	83,710
9	19,939	6,199	8,612	14,295	8,947	39,321	97,313
10	8,984	7,365	8,074	5,714	5,617	24,692	60,446
11	4,847	5,109	5,949	4,467	4,129	19,220	43,721
12	2,751	3,587	4,640	3,403	3,220	15,706	33,307

Table 4.3.4.3B

Proficiency Level by Grade (Percent): Speaking S303

	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	22.0%	7.6%	15.2%	16.7%	11.2%	27.2%	100.0%
K (accountability)	22.0%	22.9%	16.7%	11.2%	27.2%	0.0%	100.0%
1	8.1%	25.6%	18.4%	8.7%	7.6%	31.6%	100.0%
2	6.1%	11.8%	14.7%	8.9%	8.9%	49.6%	100.0%
3	6.9%	17.2%	18.4%	11.3%	11.4%	34.8%	100.0%
4	8.8%	12.5%	16.9%	11.5%	12.0%	38.3%	100.0%
5	9.1%	10.6%	14.5%	10.5%	11.6%	43.7%	100.0%
6	8.0%	8.8%	15.0%	23.0%	11.0%	34.2%	100.0%
7	9.7%	6.0%	12.1%	21.0%	11.4%	39.7%	100.0%
8	10.0%	9.7%	14.4%	11.3%	11.7%	43.0%	100.0%
9	20.5%	6.4%	8.8%	14.7%	9.2%	40.4%	100.0%
10	14.9%	12.2%	13.4%	9.5%	9.3%	40.8%	100.0%
11	11.1%	11.7%	13.6%	10.2%	9.4%	44.0%	100.0%
12	8.3%	10.8%	13.9%	10.2%	9.7%	47.2%	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 S303

Cluster	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	30,390	18,938	25,671	44,254	49,752	49,208	218,213
K (accountability)	-	52,619	30,904	35,730	19,219	30,533	49,208	218,213
1-2	A	13,950	23,013	33,226	9,671	14,431	0	94,291
	B	1,860	9,191	63,967	42,709	103,758	0	221,485
	C	611	3,152	12,865	15,883	39,759	50,135	122,405
3-5	A	11,342	13,111	8,833	2,273	2,570	0	38,129
	B	1,694	11,295	39,573	48,498	64,052	0	165,112
	C	641	2,958	14,901	31,439	67,956	96,707	214,602
6-8	A	14,765	8,829	4,645	2,029	1,287	0	31,555
	B	2,031	8,738	17,879	28,975	29,028	0	86,651
	C	620	990	6,053	20,704	42,629	60,655	131,651
9-12	A	22,549	10,326	3,242	1,688	903	0	38,708
	B	3,751	12,356	17,102	20,990	28,435	0	82,634
	C	1,703	2,826	9,728	24,745	39,224	34,097	112,323

Table 4.3.5.1B

Proficiency Level by Cluster by Tier (Percent): Oral S303

Cluster	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	13.9%	8.7%	11.8%	20.3%	22.8%	22.6%	100.0%
K (accountability)	-	24.1%	14.2%	16.4%	8.8%	14.0%	22.6%	100.0%
1-2	A	14.8%	24.4%	35.2%	10.3%	15.3%	0.0%	100.0%
	B	0.8%	4.1%	28.9%	19.3%	46.8%	0.0%	100.0%
	C	0.5%	2.6%	10.5%	13.0%	32.5%	41.0%	100.0%
3-5	A	29.7%	34.4%	23.2%	6.0%	6.7%	0.0%	100.0%
	B	1.0%	6.8%	24.0%	29.4%	38.8%	0.0%	100.0%
	C	0.3%	1.4%	6.9%	14.6%	31.7%	45.1%	100.0%
6-8	A	46.8%	28.0%	14.7%	6.4%	4.1%	0.0%	100.0%
	B	2.3%	10.1%	20.6%	33.4%	33.5%	0.0%	100.0%
	C	0.5%	0.8%	4.6%	15.7%	32.4%	46.1%	100.0%
9-12	A	58.3%	26.7%	8.4%	4.4%	2.3%	0.0%	100.0%
	B	4.5%	15.0%	20.7%	25.4%	34.4%	0.0%	100.0%
	C	1.5%	2.5%	8.7%	22.0%	34.9%	30.4%	100.0%

4.3.5.2 By Grade by Tier

Table 4.3.5.2A

Proficiency Level by Grade by Tier (Count): Oral S303

Grade	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	30,390	18,938	25,671	44,254	49,752	49,208	218,213
K (accountability)	-	52,619	30,904	35,730	19,219	30,533	49,208	218,213
1	A	8,705	17,795	27,392	7,810	11,436	0	73,138
	B	1,079	5,453	36,080	20,849	43,247	0	106,708
	C	256	1,568	6,051	5,726	12,455	16,477	42,533
2	A	5,245	5,218	5,834	1,861	2,995	0	21,153
	B	781	3,738	27,887	21,860	60,511	0	114,777
	C	355	1,584	6,814	10,157	27,304	33,658	79,872
3	A	3,864	5,865	4,057	918	1,060	0	15,764
	B	709	5,576	22,476	25,953	31,238	0	85,952
	C	225	1,202	6,139	13,617	25,650	45,913	92,746
4	A	3,524	4,043	2,803	695	784	0	11,849
	B	505	3,063	10,506	13,369	18,034	0	45,477
	C	199	897	5,225	9,901	21,806	29,974	68,002
5	A	3,954	3,203	1,973	660	726	0	10,516
	B	480	2,656	6,591	9,176	14,780	0	33,683
	C	217	859	3,537	7,921	20,500	20,820	53,854
6	A	4,099	3,254	1,851	842	495	0	10,541
	B	501	2,506	5,905	10,250	8,981	0	28,143
	C	203	374	2,148	7,699	15,300	18,092	43,816
7	A	4,981	2,793	1,506	621	467	0	10,368
	B	659	3,060	6,263	9,113	10,759	0	29,854
	C	199	314	1,988	6,705	13,206	21,149	43,561
8	A	5,685	2,782	1,288	566	325	0	10,646
	B	871	3,172	5,711	9,612	9,288	0	28,654
	C	218	302	1,917	6,300	14,123	21,414	44,274
9	A	13,406	5,880	1,519	767	513	0	22,085
	B	1,212	3,802	5,262	8,526	12,922	0	31,724
	C	503	613	2,255	7,356	15,977	16,407	43,111
10	A	5,030	2,564	973	363	246	0	9,176
	B	1,094	3,956	4,673	5,417	7,366	0	22,506
	C	412	698	2,632	6,430	10,541	7,798	28,511
11	A	2,719	1,310	453	326	144	0	4,952
	B	748	2,538	3,990	3,866	4,825	0	15,967
	C	366	740	2,274	6,020	6,830	6,348	22,578
12	A	1,394	572	297	232	0	0	2,495
	B	697	2,060	3,177	3,181	3,322	0	12,437
	C	422	775	2,567	4,939	5,876	3,544	18,123

Table 4.3.5.2B

Proficiency Level by Grade by Tier (Percent): Oral S303

Grade	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	13.9%	8.7%	11.8%	20.3%	22.8%	22.6%	100.0%
K (accountability)	-	24.1%	14.2%	16.4%	8.8%	14.0%	22.6%	100.0%
1	A	11.9%	24.3%	37.5%	10.7%	15.6%	0.0%	100.0%
	B	1.0%	5.1%	33.8%	19.5%	40.5%	0.0%	100.0%
	C	0.6%	3.7%	14.2%	13.5%	29.3%	38.7%	100.0%
2	A	24.8%	24.7%	27.6%	8.8%	14.2%	0.0%	100.0%
	B	0.7%	3.3%	24.3%	19.0%	52.7%	0.0%	100.0%
	C	0.4%	2.0%	8.5%	12.7%	34.2%	42.1%	100.0%
3	A	24.5%	37.2%	25.7%	5.8%	6.7%	0.0%	100.0%
	B	0.8%	6.5%	26.1%	30.2%	36.3%	0.0%	100.0%
	C	0.2%	1.3%	6.6%	14.7%	27.7%	49.5%	100.0%
4	A	29.7%	34.1%	23.7%	5.9%	6.6%	0.0%	100.0%
	B	1.1%	6.7%	23.1%	29.4%	39.7%	0.0%	100.0%
	C	0.3%	1.3%	7.7%	14.6%	32.1%	44.1%	100.0%
5	A	37.6%	30.5%	18.8%	6.3%	6.9%	0.0%	100.0%
	B	1.4%	7.9%	19.6%	27.2%	43.9%	0.0%	100.0%
	C	0.4%	1.6%	6.6%	14.7%	38.1%	38.7%	100.0%
6	A	38.9%	30.9%	17.6%	8.0%	4.7%	0.0%	100.0%
	B	1.8%	8.9%	21.0%	36.4%	31.9%	0.0%	100.0%
	C	0.5%	0.9%	4.9%	17.6%	34.9%	41.3%	100.0%
7	A	48.0%	26.9%	14.5%	6.0%	4.5%	0.0%	100.0%
	B	2.2%	10.2%	21.0%	30.5%	36.0%	0.0%	100.0%
	C	0.5%	0.7%	4.6%	15.4%	30.3%	48.6%	100.0%
8	A	53.4%	26.1%	12.1%	5.3%	3.1%	0.0%	100.0%
	B	3.0%	11.1%	19.9%	33.5%	32.4%	0.0%	100.0%
	C	0.5%	0.7%	4.3%	14.2%	31.9%	48.4%	100.0%
9	A	60.7%	26.6%	6.9%	3.5%	2.3%	0.0%	100.0%
	B	3.8%	12.0%	16.6%	26.9%	40.7%	0.0%	100.0%
	C	1.2%	1.4%	5.2%	17.1%	37.1%	38.1%	100.0%
10	A	54.8%	27.9%	10.6%	4.0%	2.7%	0.0%	100.0%
	B	4.9%	17.6%	20.8%	24.1%	32.7%	0.0%	100.0%
	C	1.4%	2.4%	9.2%	22.6%	37.0%	27.4%	100.0%
11	A	54.9%	26.5%	9.1%	6.6%	2.9%	0.0%	100.0%
	B	4.7%	15.9%	25.0%	24.2%	30.2%	0.0%	100.0%
	C	1.6%	3.3%	10.1%	26.7%	30.3%	28.1%	100.0%
12	A	55.9%	22.9%	11.9%	9.3%	0.0%	0.0%	100.0%
	B	5.6%	16.6%	25.5%	25.6%	26.7%	0.0%	100.0%
	C	2.3%	4.3%	14.2%	27.3%	32.4%	19.6%	100.0%

4.3.5.3 By Grade

Table 4.3.5.3A

Proficiency Level by Grade (Count): Oral S303

	Oral Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	30,390	18,938	25,671	44,254	49,752	49,208	218,213
K (accountability)	52,619	30,904	35,730	19,219	30,533	49,208	218,213
1	10,040	24,816	69,523	34,385	67,138	16,477	222,379
2	6,381	10,540	40,535	33,878	90,810	33,658	215,802
3	4,798	12,643	32,672	40,488	57,948	45,913	194,462
4	4,228	8,003	18,534	23,965	40,624	29,974	125,328
5	4,651	6,718	12,101	17,757	36,006	20,820	98,053
6	4,803	6,134	9,904	18,791	24,776	18,092	82,500
7	5,839	6,167	9,757	16,439	24,432	21,149	83,783
8	6,774	6,256	8,916	16,478	23,736	21,414	83,574
9	15,121	10,295	9,036	16,649	29,412	16,407	96,920
10	6,536	7,218	8,278	12,210	18,153	7,798	60,193
11	3,833	4,588	6,717	10,212	11,799	6,348	43,497
12	2,513	3,407	6,041	8,352	9,198	3,544	33,055

Table 4.3.5.3B

Proficiency Level by Grade (Percent): Oral S303

	Oral Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	13.9%	8.7%	11.8%	20.3%	22.8%	22.6%	100.0%
K (accountability)	24.1%	14.2%	16.4%	8.8%	14.0%	22.6%	100.0%
1	4.5%	11.2%	31.3%	15.5%	30.2%	7.4%	100.0%
2	3.0%	4.9%	18.8%	15.7%	42.1%	15.6%	100.0%
3	2.5%	6.5%	16.8%	20.8%	29.8%	23.6%	100.0%
4	3.4%	6.4%	14.8%	19.1%	32.4%	23.9%	100.0%
5	4.7%	6.9%	12.3%	18.1%	36.7%	21.2%	100.0%
6	5.8%	7.4%	12.0%	22.8%	30.0%	21.9%	100.0%
7	7.0%	7.4%	11.6%	19.6%	29.2%	25.2%	100.0%
8	8.1%	7.5%	10.7%	19.7%	28.4%	25.6%	100.0%
9	15.6%	10.6%	9.3%	17.2%	30.3%	16.9%	100.0%
10	10.9%	12.0%	13.8%	20.3%	30.2%	13.0%	100.0%
11	8.8%	10.5%	15.4%	23.5%	27.1%	14.6%	100.0%
12	7.6%	10.3%	18.3%	25.3%	27.8%	10.7%	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 S303

Cluster	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	37,193	52,667	39,268	28,012	45,808	15,304	218,252
K (accountability)	-	137,028	26,573	26,943	19,202	8,506	0	218,252
1-2	A	15,123	48,398	30,795	0	0	0	94,316
	B	3,297	41,219	159,039	17,887	2	0	221,444
	C	795	10,281	45,006	40,240	22,669	3,405	122,396
3-5	A	6,713	13,883	12,654	4,750	14	0	38,014
	B	1,409	10,571	51,051	96,463	5,453	0	164,947
	C	419	905	15,804	69,176	98,553	29,762	214,619
6-8	A	8,619	14,438	7,733	778	3	0	31,571
	B	2,609	18,317	44,825	20,574	292	0	86,617
	C	908	10,146	64,644	42,133	11,832	2,044	131,707
9-12	A	9,011	17,907	10,603	1,295	7	0	38,823
	B	4,545	15,506	32,430	25,159	5,266	21	82,927
	C	1,436	3,676	19,561	33,470	37,096	17,376	112,615

Table 4.3.6.1B

Proficiency Level by Cluster by Tier (Percent): Literacy S303

Cluster	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	17.0%	24.1%	18.0%	12.8%	21.0%	7.0%	100.0%
K (accountability)	-	62.8%	12.2%	12.3%	8.8%	3.9%	0.0%	100.0%
1-2	A	16.0%	51.3%	32.7%	0.0%	0.0%	0.0%	100.0%
	B	1.5%	18.6%	71.8%	8.1%	0.0%	0.0%	100.0%
	C	0.6%	8.4%	36.8%	32.9%	18.5%	2.8%	100.0%
3-5	A	17.7%	36.5%	33.3%	12.5%	0.0%	0.0%	100.0%
	B	0.9%	6.4%	30.9%	58.5%	3.3%	0.0%	100.0%
	C	0.2%	0.4%	7.4%	32.2%	45.9%	13.9%	100.0%
6-8	A	27.3%	45.7%	24.5%	2.5%	0.0%	0.0%	100.0%
	B	3.0%	21.1%	51.8%	23.8%	0.3%	0.0%	100.0%
	C	0.7%	7.7%	49.1%	32.0%	9.0%	1.6%	100.0%
9-12	A	23.2%	46.1%	27.3%	3.3%	0.0%	0.0%	100.0%
	B	5.5%	18.7%	39.1%	30.3%	6.4%	0.0%	100.0%
	C	1.3%	3.3%	17.4%	29.7%	32.9%	15.4%	100.0%

4.3.6.2 By Grade by Tier

Table 4.3.6.2A

Proficiency Level by Grade by Tier (Count): Literacy S303

Grade	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	37,193	52,667	39,268	28,012	45,808	15,304	218,252
K (accountability)	-	137,028	26,573	26,943	19,202	8,506	0	218,252
1	A	9,327	38,140	25,683	0	0	0	73,150
	B	1,435	20,106	77,379	7,752	1	0	106,673
	C	347	4,124	17,155	13,189	6,547	1,142	42,504
2	A	5,796	10,258	5,112	0	0	0	21,166
	B	1,862	21,113	81,660	10,135	1	0	114,771
	C	448	6,157	27,851	27,051	16,122	2,263	79,892
3	A	1,786	5,346	5,629	2,949	13	0	15,723
	B	448	3,759	19,713	57,477	4,507	0	85,904
	C	156	220	2,860	19,046	53,554	16,922	92,758
4	A	2,076	4,509	3,927	1,299	0	0	11,811
	B	422	3,352	15,681	25,209	744	0	45,408
	C	138	267	4,498	27,590	27,590	7,921	68,004
5	A	2,851	4,028	3,098	502	1	0	10,480
	B	539	3,460	15,657	13,777	202	0	33,635
	C	125	418	8,446	22,540	17,409	4,919	53,857
6	A	1,520	5,210	3,331	480	3	0	10,544
	B	502	4,895	13,746	8,789	180	0	28,112
	C	232	2,667	20,191	15,928	4,062	749	43,829
7	A	2,907	4,618	2,611	231	0	0	10,367
	B	861	6,129	15,524	7,235	93	0	29,842
	C	327	3,265	21,184	14,078	4,033	698	43,585
8	A	4,192	4,610	1,791	67	0	0	10,660
	B	1,246	7,293	15,555	4,550	19	0	28,663
	C	349	4,214	23,269	12,127	3,737	597	44,293
9	A	5,124	10,208	5,902	901	5	0	22,140
	B	1,209	4,169	11,997	10,836	3,570	16	31,797
	C	385	791	4,930	11,208	17,397	8,480	43,191
10	A	1,951	4,302	2,639	297	1	0	9,190
	B	1,277	4,465	9,054	6,657	1,129	2	22,584
	C	316	879	5,181	8,863	9,337	4,009	28,585
11	A	1,147	2,294	1,450	82	0	0	4,973
	B	1,002	3,520	6,479	4,622	419	3	16,045
	C	338	910	4,773	7,299	6,320	3,013	22,653
12	A	789	1,103	612	15	1	0	2,520
	B	1,057	3,352	4,900	3,044	148	0	12,501
	C	397	1,096	4,677	6,100	4,042	1,874	18,186

Table 4.3.6.2B

Proficiency Level by Grade by Tier (Percent): Literacy S303

Grade	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	17.0%	24.1%	18.0%	12.8%	21.0%	7.0%	100.0%
K (accountability)	-	62.8%	12.2%	12.3%	8.8%	3.9%	0.0%	100.0%
1	A	12.8%	52.1%	35.1%	0.0%	0.0%	0.0%	100.0%
	B	1.3%	18.8%	72.5%	7.3%	0.0%	0.0%	100.0%
	C	0.8%	9.7%	40.4%	31.0%	15.4%	2.7%	100.0%
2	A	27.4%	48.5%	24.2%	0.0%	0.0%	0.0%	100.0%
	B	1.6%	18.4%	71.2%	8.8%	0.0%	0.0%	100.0%
	C	0.6%	7.7%	34.9%	33.9%	20.2%	2.8%	100.0%
3	A	11.4%	34.0%	35.8%	18.8%	0.1%	0.0%	100.0%
	B	0.5%	4.4%	22.9%	66.9%	5.2%	0.0%	100.0%
	C	0.2%	0.2%	3.1%	20.5%	57.7%	18.2%	100.0%
4	A	17.6%	38.2%	33.2%	11.0%	0.0%	0.0%	100.0%
	B	0.9%	7.4%	34.5%	55.5%	1.6%	0.0%	100.0%
	C	0.2%	0.4%	6.6%	40.6%	40.6%	11.6%	100.0%
5	A	27.2%	38.4%	29.6%	4.8%	0.0%	0.0%	100.0%
	B	1.6%	10.3%	46.5%	41.0%	0.6%	0.0%	100.0%
	C	0.2%	0.8%	15.7%	41.9%	32.3%	9.1%	100.0%
6	A	14.4%	49.4%	31.6%	4.6%	0.0%	0.0%	100.0%
	B	1.8%	17.4%	48.9%	31.3%	0.6%	0.0%	100.0%
	C	0.5%	6.1%	46.1%	36.3%	9.3%	1.7%	100.0%
7	A	28.0%	44.5%	25.2%	2.2%	0.0%	0.0%	100.0%
	B	2.9%	20.5%	52.0%	24.2%	0.3%	0.0%	100.0%
	C	0.8%	7.5%	48.6%	32.3%	9.3%	1.6%	100.0%
8	A	39.3%	43.2%	16.8%	0.6%	0.0%	0.0%	100.0%
	B	4.3%	25.4%	54.3%	15.9%	0.1%	0.0%	100.0%
	C	0.8%	9.5%	52.5%	27.4%	8.4%	1.3%	100.0%
9	A	23.1%	46.1%	26.7%	4.1%	0.0%	0.0%	100.0%
	B	3.8%	13.1%	37.7%	34.1%	11.2%	0.1%	100.0%
	C	0.9%	1.8%	11.4%	25.9%	40.3%	19.6%	100.0%
10	A	21.2%	46.8%	28.7%	3.2%	0.0%	0.0%	100.0%
	B	5.7%	19.8%	40.1%	29.5%	5.0%	0.0%	100.0%
	C	1.1%	3.1%	18.1%	31.0%	32.7%	14.0%	100.0%
11	A	23.1%	46.1%	29.2%	1.6%	0.0%	0.0%	100.0%
	B	6.2%	21.9%	40.4%	28.8%	2.6%	0.0%	100.0%
	C	1.5%	4.0%	21.1%	32.2%	27.9%	13.3%	100.0%
12	A	31.3%	43.8%	24.3%	0.6%	0.0%	0.0%	100.0%
	B	8.5%	26.8%	39.2%	24.4%	1.2%	0.0%	100.0%
	C	2.2%	6.0%	25.7%	33.5%	22.2%	10.3%	100.0%

4.3.6.3 By Grade

Table 4.3.6.3A

Proficiency Level by Grade (Count): Literacy S303

	Literacy Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	37,193	52,667	39,268	28,012	45,808	15,304	218,252
K (accountability)	137,028	26,573	26,943	19,202	8,506	0	218,252
1	11,109	62,370	120,217	20,941	6,548	1,142	222,327
2	8,106	37,528	114,623	37,186	16,123	2,263	215,829
3	2,390	9,325	28,202	79,472	58,074	16,922	194,385
4	2,636	8,128	24,106	54,098	28,334	7,921	125,223
5	3,515	7,906	27,201	36,819	17,612	4,919	97,972
6	2,254	12,772	37,268	25,197	4,245	749	82,485
7	4,095	14,012	39,319	21,544	4,126	698	83,794
8	5,787	16,117	40,615	16,744	3,756	597	83,616
9	6,718	15,168	22,829	22,945	20,972	8,496	97,128
10	3,544	9,646	16,874	15,817	10,467	4,011	60,359
11	2,487	6,724	12,702	12,003	6,739	3,016	43,671
12	2,243	5,551	10,189	9,159	4,191	1,874	33,207

Table 4.3.6.3B

Proficiency Level by Grade (Percent): Literacy S303

	Literacy Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	17.0%	24.1%	18.0%	12.8%	21.0%	7.0%	100.0%
K (accountability)	62.8%	12.2%	12.3%	8.8%	3.9%	0.0%	100.0%
1	5.0%	28.1%	54.1%	9.4%	2.9%	0.5%	100.0%
2	3.8%	17.4%	53.1%	17.2%	7.5%	1.0%	100.0%
3	1.2%	4.8%	14.5%	40.9%	29.9%	8.7%	100.0%
4	2.1%	6.5%	19.3%	43.2%	22.6%	6.3%	100.0%
5	3.6%	8.1%	27.8%	37.6%	18.0%	5.0%	100.0%
6	2.7%	15.5%	45.2%	30.5%	5.1%	0.9%	100.0%
7	4.9%	16.7%	46.9%	25.7%	4.9%	0.8%	100.0%
8	6.9%	19.3%	48.6%	20.0%	4.5%	0.7%	100.0%
9	6.9%	15.6%	23.5%	23.6%	21.6%	8.7%	100.0%
10	5.9%	16.0%	28.0%	26.2%	17.3%	6.6%	100.0%
11	5.7%	15.4%	29.1%	27.5%	15.4%	6.9%	100.0%
12	6.8%	16.7%	30.7%	27.6%	12.6%	5.6%	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 S303

Cluster	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	26,098	31,765	46,435	30,647	33,062	50,249	218,256
K (accountability)	-	125,923	15,211	15,453	15,004	28,336	18,329	218,256
1-2	A	11,418	24,083	39,925	18,896	n/a	n/a	94,322
	B	1,501	3,581	29,521	54,078	132,872	n/a	221,553
	C	398	1,861	13,058	17,110	40,107	49,886	122,420
3-5	A	6,198	16,280	10,892	4,709	n/a	n/a	38,079
	B	845	8,666	39,398	38,169	77,942	n/a	165,020
	C	348	895	13,863	19,175	75,998	104,438	214,717
6-8	A	10,891	14,111	5,199	1,373	n/a	n/a	31,574
	B	1,369	17,948	34,048	17,770	15,571	n/a	86,706
	C	437	4,650	29,078	24,852	40,882	31,877	131,776
9-12	A	16,529	16,125	4,914	1,308	n/a	n/a	38,876
	B	4,998	25,034	23,720	17,187	12,059	n/a	82,998
	C	1,194	7,303	18,981	22,097	29,960	33,260	112,795

Table 4.3.7.1B

Proficiency Level by Cluster by Tier (Percent): Comprehension S303

Cluster	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	12.0%	14.6%	21.3%	14.0%	15.1%	23.0%	100.0%
K (accountability)	-	57.7%	7.0%	7.1%	6.9%	13.0%	8.4%	100.0%
1-2	A	12.1%	25.5%	42.3%	20.0%	n/a	n/a	100.0%
	B	0.7%	1.6%	13.3%	24.4%	60.0%	n/a	100.0%
	C	0.3%	1.5%	10.7%	14.0%	32.8%	40.7%	100.0%
3-5	A	16.3%	42.8%	28.6%	12.4%	n/a	n/a	100.0%
	B	0.5%	5.3%	23.9%	23.1%	47.2%	n/a	100.0%
	C	0.2%	0.4%	6.5%	8.9%	35.4%	48.6%	100.0%
6-8	A	34.5%	44.7%	16.5%	4.3%	n/a	n/a	100.0%
	B	1.6%	20.7%	39.3%	20.5%	18.0%	n/a	100.0%
	C	0.3%	3.5%	22.1%	18.9%	31.0%	24.2%	100.0%
9-12	A	42.5%	41.5%	12.6%	3.4%	n/a	n/a	100.0%
	B	6.0%	30.2%	28.6%	20.7%	14.5%	n/a	100.0%
	C	1.1%	6.5%	16.8%	19.6%	26.6%	29.5%	100.0%

4.3.7.2 By Grade by Tier

Table 4.3.7.2A

Proficiency Level by Grade by Tier (Count): Comprehension S303

Grade	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	26,098	31,765	46,435	30,647	33,062	50,249	218,256
K (accountability)	-	125,923	15,211	15,453	15,004	28,336	18,329	218,256
1	A	7,051	17,407	34,236	14,462	n/a	n/a	73,156
	B	1,047	1,307	10,752	27,728	65,897	n/a	106,731
	C	229	490	4,506	6,304	15,821	15,160	42,510
2	A	4,367	6,676	5,689	4,434	n/a	n/a	21,166
	B	454	2,274	18,769	26,350	66,975	n/a	114,822
	C	169	1,371	8,552	10,806	24,286	34,726	79,910
3	A	1,385	6,673	5,106	2,576	n/a	n/a	15,740
	B	313	2,170	16,442	19,323	47,706	n/a	85,954
	C	145	137	2,088	4,494	32,663	53,268	92,795
4	A	1,919	5,323	3,268	1,331	n/a	n/a	11,841
	B	223	3,001	12,156	11,169	18,879	n/a	45,428
	C	107	250	4,900	7,766	24,224	30,789	68,036
5	A	2,894	4,284	2,518	802	n/a	n/a	10,498
	B	309	3,495	10,800	7,677	11,357	n/a	33,638
	C	96	508	6,875	6,915	19,111	20,381	53,886
6	A	2,324	5,346	2,212	668	n/a	n/a	10,550
	B	218	4,890	11,581	5,791	5,666	n/a	28,146
	C	125	1,143	10,704	8,857	13,874	9,147	43,850
7	A	3,503	4,775	1,710	379	n/a	n/a	10,367
	B	421	5,949	11,954	6,460	5,086	n/a	29,870
	C	157	1,513	9,695	8,489	13,321	10,429	43,604
8	A	5,064	3,990	1,277	326	n/a	n/a	10,657
	B	730	7,109	10,513	5,519	4,819	n/a	28,690
	C	155	1,994	8,679	7,506	13,687	12,301	44,322
9	A	9,621	9,218	2,671	672	n/a	n/a	22,182
	B	1,091	7,884	11,247	6,396	5,227	n/a	31,845
	C	313	1,429	6,885	7,789	13,037	13,792	43,245
10	A	3,509	4,100	1,290	305	n/a	n/a	9,204
	B	1,218	7,458	6,409	4,493	3,039	n/a	22,617
	C	263	1,847	5,442	5,785	7,633	7,681	28,651
11	A	2,107	2,001	639	228	n/a	n/a	4,975
	B	1,184	5,644	3,371	3,756	2,084	n/a	16,039
	C	261	2,182	3,647	4,733	5,278	6,573	22,674
12	A	1,292	806	314	103	n/a	n/a	2,515
	B	1,505	4,048	2,693	2,542	1,709	n/a	12,497
	C	357	1,845	3,007	3,790	4,012	5,214	18,225

Table 4.3.7.2B

Proficiency Level by Grade by Tier (Percent): Comprehension S303

Grade	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	12.0%	14.6%	21.3%	14.0%	15.1%	23.0%	100.0%
K (accountability)	-	57.7%	7.0%	7.1%	6.9%	13.0%	8.4%	100.0%
1	A	9.6%	23.8%	46.8%	19.8%	n/a	n/a	100.0%
	B	1.0%	1.2%	10.1%	26.0%	61.7%	n/a	100.0%
	C	0.5%	1.2%	10.6%	14.8%	37.2%	35.7%	100.0%
2	A	20.6%	31.5%	26.9%	20.9%	n/a	n/a	100.0%
	B	0.4%	2.0%	16.3%	22.9%	58.3%	n/a	100.0%
	C	0.2%	1.7%	10.7%	13.5%	30.4%	43.5%	100.0%
3	A	8.8%	42.4%	32.4%	16.4%	n/a	n/a	100.0%
	B	0.4%	2.5%	19.1%	22.5%	55.5%	n/a	100.0%
	C	0.2%	0.1%	2.3%	4.8%	35.2%	57.4%	100.0%
4	A	16.2%	45.0%	27.6%	11.2%	n/a	n/a	100.0%
	B	0.5%	6.6%	26.8%	24.6%	41.6%	n/a	100.0%
	C	0.2%	0.4%	7.2%	11.4%	35.6%	45.3%	100.0%
5	A	27.6%	40.8%	24.0%	7.6%	n/a	n/a	100.0%
	B	0.9%	10.4%	32.1%	22.8%	33.8%	n/a	100.0%
	C	0.2%	0.9%	12.8%	12.8%	35.5%	37.8%	100.0%
6	A	22.0%	50.7%	21.0%	6.3%	n/a	n/a	100.0%
	B	0.8%	17.4%	41.1%	20.6%	20.1%	n/a	100.0%
	C	0.3%	2.6%	24.4%	20.2%	31.6%	20.9%	100.0%
7	A	33.8%	46.1%	16.5%	3.7%	n/a	n/a	100.0%
	B	1.4%	19.9%	40.0%	21.6%	17.0%	n/a	100.0%
	C	0.4%	3.5%	22.2%	19.5%	30.5%	23.9%	100.0%
8	A	47.5%	37.4%	12.0%	3.1%	n/a	n/a	100.0%
	B	2.5%	24.8%	36.6%	19.2%	16.8%	n/a	100.0%
	C	0.3%	4.5%	19.6%	16.9%	30.9%	27.8%	100.0%
9	A	43.4%	41.6%	12.0%	3.0%	n/a	n/a	100.0%
	B	3.4%	24.8%	35.3%	20.1%	16.4%	n/a	100.0%
	C	0.7%	3.3%	15.9%	18.0%	30.1%	31.9%	100.0%
10	A	38.1%	44.5%	14.0%	3.3%	n/a	n/a	100.0%
	B	5.4%	33.0%	28.3%	19.9%	13.4%	n/a	100.0%
	C	0.9%	6.4%	19.0%	20.2%	26.6%	26.8%	100.0%
11	A	42.4%	40.2%	12.8%	4.6%	n/a	n/a	100.0%
	B	7.4%	35.2%	21.0%	23.4%	13.0%	n/a	100.0%
	C	1.2%	9.6%	16.1%	20.9%	23.3%	29.0%	100.0%
12	A	51.4%	32.0%	12.5%	4.1%	n/a	n/a	100.0%
	B	12.0%	32.4%	21.5%	20.3%	13.7%	n/a	100.0%
	C	2.0%	10.1%	16.5%	20.8%	22.0%	28.6%	100.0%

4.3.7.3 By Grade

Table 4.3.7.3A

Proficiency Level by Grade (Count): Comprehension S303

	Comprehension Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	26,098	31,765	46,435	30,647	33,062	50,249	218,256
K (accountability)	125,923	15,211	15,453	15,004	28,336	18,329	218,256
1	8,327	19,204	49,494	48,494	81,718	15,160	222,397
2	4,990	10,321	33,010	41,590	91,261	34,726	215,898
3	1,843	8,980	23,636	26,393	80,369	53,268	194,489
4	2,249	8,574	20,324	20,266	43,103	30,789	125,305
5	3,299	8,287	20,193	15,394	30,468	20,381	98,022
6	2,667	11,379	24,497	15,316	19,540	9,147	82,546
7	4,081	12,237	23,359	15,328	18,407	10,429	83,841
8	5,949	13,093	20,469	13,351	18,506	12,301	83,669
9	11,025	18,531	20,803	14,857	18,264	13,792	97,272
10	4,990	13,405	13,141	10,583	10,672	7,681	60,472
11	3,552	9,827	7,657	8,717	7,362	6,573	43,688
12	3,154	6,699	6,014	6,435	5,721	5,214	33,237

Table 4.3.7.3B

Proficiency Level by Grade (Percent): Comprehension S303

	Comprehension Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	12.0%	14.6%	21.3%	14.0%	15.1%	23.0%	100.0%
K (accountability)	57.7%	7.0%	7.1%	6.9%	13.0%	8.4%	100.0%
1	3.7%	8.6%	22.3%	21.8%	36.7%	6.8%	100.0%
2	2.3%	4.8%	15.3%	19.3%	42.3%	16.1%	100.0%
3	0.9%	4.6%	12.2%	13.6%	41.3%	27.4%	100.0%
4	1.8%	6.8%	16.2%	16.2%	34.4%	24.6%	100.0%
5	3.4%	8.5%	20.6%	15.7%	31.1%	20.8%	100.0%
6	3.2%	13.8%	29.7%	18.6%	23.7%	11.1%	100.0%
7	4.9%	14.6%	27.9%	18.3%	22.0%	12.4%	100.0%
8	7.1%	15.6%	24.5%	16.0%	22.1%	14.7%	100.0%
9	11.3%	19.1%	21.4%	15.3%	18.8%	14.2%	100.0%
10	8.3%	22.2%	21.7%	17.5%	17.6%	12.7%	100.0%
11	8.1%	22.5%	17.5%	20.0%	16.9%	15.0%	100.0%
12	9.5%	20.2%	18.1%	19.4%	17.2%	15.7%	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 S303

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	29,509	42,742	42,889	38,718	48,717	15,599	218,174
K (accountability)	-	112,412	34,939	32,749	22,475	13,637	1,962	218,174
1-2	A	12,147	39,589	42,272	167	0	0	94,175
	B	1,542	19,281	132,999	67,403	20	0	221,245
	C	355	3,965	30,058	45,335	36,387	6,212	122,312
3-5	A	8,207	13,840	11,902	3,903	81	0	37,933
	B	817	8,601	47,701	92,056	15,594	0	164,769
	C	259	641	10,760	54,585	103,477	44,712	214,434
6-8	A	11,328	12,374	6,446	1,334	11	0	31,493
	B	1,477	12,552	37,462	32,875	2,072	0	86,438
	C	374	2,760	29,923	59,245	34,269	4,906	131,477
9-12	A	13,632	16,969	6,620	1,351	11	0	38,583
	B	3,168	13,533	28,547	27,431	9,647	13	82,339
	C	983	2,512	14,440	33,856	40,915	19,236	111,942

Table 4.3.8.1B

Proficiency Level by Cluster by Tier (Percent): Overall S303

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	13.5%	19.6%	19.7%	17.7%	22.3%	7.1%	100.0%
K (accountability)	-	51.5%	16.0%	15.0%	10.3%	6.3%	0.9%	100.0%
1-2	A	12.9%	42.0%	44.9%	0.2%	0.0%	0.0%	100.0%
	B	0.7%	8.7%	60.1%	30.5%	0.0%	0.0%	100.0%
	C	0.3%	3.2%	24.6%	37.1%	29.7%	5.1%	100.0%
3-5	A	21.6%	36.5%	31.4%	10.3%	0.2%	0.0%	100.0%
	B	0.5%	5.2%	29.0%	55.9%	9.5%	0.0%	100.0%
	C	0.1%	0.3%	5.0%	25.5%	48.3%	20.9%	100.0%
6-8	A	36.0%	39.3%	20.5%	4.2%	0.0%	0.0%	100.0%
	B	1.7%	14.5%	43.3%	38.0%	2.4%	0.0%	100.0%
	C	0.3%	2.1%	22.8%	45.1%	26.1%	3.7%	100.0%
9-12	A	35.3%	44.0%	17.2%	3.5%	0.0%	0.0%	100.0%
	B	3.8%	16.4%	34.7%	33.3%	11.7%	0.0%	100.0%
	C	0.9%	2.2%	12.9%	30.2%	36.6%	17.2%	100.0%

4.3.8.2 By Grade by Tier

Table 4.3.8.2A

Proficiency Level by Grade by Tier (Count): Overall S303

Grade	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	29,509	42,742	42,889	38,718	48,717	15,599	218,174
K (accountability)	-	112,412	34,939	32,749	22,475	13,637	1,962	218,174
1	A	7,076	30,703	35,126	153	0	0	73,058
	B	912	9,833	70,160	25,656	11	0	106,572
	C	169	1,723	12,803	15,187	10,607	1,986	42,475
2	A	5,071	8,886	7,146	14	0	0	21,117
	B	630	9,448	62,839	41,747	9	0	114,673
	C	186	2,242	17,255	30,148	25,780	4,226	79,837
3	A	2,401	5,538	5,549	2,118	76	0	15,682
	B	264	3,050	20,917	49,661	11,933	0	85,825
	C	103	157	2,433	16,341	48,287	25,368	92,689
4	A	2,639	4,448	3,617	1,083	2	0	11,789
	B	257	2,871	14,426	25,080	2,725	0	45,359
	C	84	216	3,645	20,032	32,196	11,773	67,946
5	A	3,167	3,854	2,736	702	3	0	10,462
	B	296	2,680	12,358	17,315	936	0	33,585
	C	72	268	4,682	18,212	22,994	7,571	53,799
6	A	2,712	4,388	2,694	713	8	0	10,515
	B	283	3,336	11,680	11,623	1,145	0	28,067
	C	106	749	9,478	20,507	10,995	1,921	43,756
7	A	3,777	4,057	2,121	393	3	0	10,351
	B	471	4,185	13,090	11,294	742	0	29,782
	C	125	903	10,151	18,631	12,098	1,612	43,520
8	A	4,839	3,929	1,631	228	0	0	10,627
	B	723	5,031	12,692	9,958	185	0	28,589
	C	143	1,108	10,294	20,107	11,176	1,373	44,201
9	A	8,208	9,465	3,555	781	9	0	22,018
	B	920	3,697	9,694	11,505	5,790	9	31,615
	C	282	550	3,229	10,524	18,568	9,834	42,987
10	A	2,770	4,312	1,719	343	1	0	9,145
	B	851	4,085	7,982	7,076	2,432	1	22,427
	C	211	621	3,658	9,043	10,568	4,317	28,418
11	A	1,684	2,168	903	183	0	0	4,938
	B	678	3,124	5,915	5,109	1,077	3	15,906
	C	212	637	3,692	7,712	7,055	3,199	22,507
12	A	970	1,024	443	44	1	0	2,482
	B	719	2,627	4,956	3,741	348	0	12,391
	C	278	704	3,861	6,577	4,724	1,886	18,030

Table 4.3.8.2B

Proficiency Level by Grade by Tier (Percent): Overall S303

Grade	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	13.5%	19.6%	19.7%	17.7%	22.3%	7.1%	100.0%
K (accountability)	-	51.5%	16.0%	15.0%	10.3%	6.3%	0.9%	100.0%
1	A	9.7%	42.0%	48.1%	0.2%	0.0%	0.0%	100.0%
	B	0.9%	9.2%	65.8%	24.1%	0.0%	0.0%	100.0%
	C	0.4%	4.1%	30.1%	35.8%	25.0%	4.7%	100.0%
2	A	24.0%	42.1%	33.8%	0.1%	0.0%	0.0%	100.0%
	B	0.5%	8.2%	54.8%	36.4%	0.0%	0.0%	100.0%
	C	0.2%	2.8%	21.6%	37.8%	32.3%	5.3%	100.0%
3	A	15.3%	35.3%	35.4%	13.5%	0.5%	0.0%	100.0%
	B	0.3%	3.6%	24.4%	57.9%	13.9%	0.0%	100.0%
	C	0.1%	0.2%	2.6%	17.6%	52.1%	27.4%	100.0%
4	A	22.4%	37.7%	30.7%	9.2%	0.0%	0.0%	100.0%
	B	0.6%	6.3%	31.8%	55.3%	6.0%	0.0%	100.0%
	C	0.1%	0.3%	5.4%	29.5%	47.4%	17.3%	100.0%
5	A	30.3%	36.8%	26.2%	6.7%	0.0%	0.0%	100.0%
	B	0.9%	8.0%	36.8%	51.6%	2.8%	0.0%	100.0%
	C	0.1%	0.5%	8.7%	33.9%	42.7%	14.1%	100.0%
6	A	25.8%	41.7%	25.6%	6.8%	0.1%	0.0%	100.0%
	B	1.0%	11.9%	41.6%	41.4%	4.1%	0.0%	100.0%
	C	0.2%	1.7%	21.7%	46.9%	25.1%	4.4%	100.0%
7	A	36.5%	39.2%	20.5%	3.8%	0.0%	0.0%	100.0%
	B	1.6%	14.1%	44.0%	37.9%	2.5%	0.0%	100.0%
	C	0.3%	2.1%	23.3%	42.8%	27.8%	3.7%	100.0%
8	A	45.5%	37.0%	15.3%	2.1%	0.0%	0.0%	100.0%
	B	2.5%	17.6%	44.4%	34.8%	0.6%	0.0%	100.0%
	C	0.3%	2.5%	23.3%	45.5%	25.3%	3.1%	100.0%
9	A	37.3%	43.0%	16.1%	3.5%	0.0%	0.0%	100.0%
	B	2.9%	11.7%	30.7%	36.4%	18.3%	0.0%	100.0%
	C	0.7%	1.3%	7.5%	24.5%	43.2%	22.9%	100.0%
10	A	30.3%	47.2%	18.8%	3.8%	0.0%	0.0%	100.0%
	B	3.8%	18.2%	35.6%	31.6%	10.8%	0.0%	100.0%
	C	0.7%	2.2%	12.9%	31.8%	37.2%	15.2%	100.0%
11	A	34.1%	43.9%	18.3%	3.7%	0.0%	0.0%	100.0%
	B	4.3%	19.6%	37.2%	32.1%	6.8%	0.0%	100.0%
	C	0.9%	2.8%	16.4%	34.3%	31.3%	14.2%	100.0%
12	A	39.1%	41.3%	17.8%	1.8%	0.0%	0.0%	100.0%
	B	5.8%	21.2%	40.0%	30.2%	2.8%	0.0%	100.0%
	C	1.5%	3.9%	21.4%	36.5%	26.2%	10.5%	100.0%

4.3.8.3 By Grade

Table 4.3.8.3A

Proficiency Level by Grade (Count): Overall S303

	Overall Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	29,509	42,742	42,889	38,718	48,717	15,599	218,174
K (accountability)	112,412	34,939	32,749	22,475	13,637	1,962	218,174
1	8,157	42,259	118,089	40,996	10,618	1,986	222,105
2	5,887	20,576	87,240	71,909	25,789	4,226	215,627
3	2,768	8,745	28,899	68,120	60,296	25,368	194,196
4	2,980	7,535	21,688	46,195	34,923	11,773	125,094
5	3,535	6,802	19,776	36,229	23,933	7,571	97,846
6	3,101	8,473	23,852	32,843	12,148	1,921	82,338
7	4,373	9,145	25,362	30,318	12,843	1,612	83,653
8	5,705	10,068	24,617	30,293	11,361	1,373	83,417
9	9,410	13,712	16,478	22,810	24,367	9,843	96,620
10	3,832	9,018	13,359	16,462	13,001	4,318	59,990
11	2,574	5,929	10,510	13,004	8,132	3,202	43,351
12	1,967	4,355	9,260	10,362	5,073	1,886	32,903

Table 4.3.8.3B

Proficiency Level by Grade (Percent): Overall S303

	Overall Proficiency Range						Total
	1	2	3	4	5	6	
K (instructional)	13.5%	19.6%	19.7%	17.7%	22.3%	7.1%	100.0%
K (accountability)	51.5%	16.0%	15.0%	10.3%	6.3%	0.9%	100.0%
1	3.7%	19.0%	53.2%	18.5%	4.8%	0.9%	100.0%
2	2.7%	9.5%	40.5%	33.3%	12.0%	2.0%	100.0%
3	1.4%	4.5%	14.9%	35.1%	31.0%	13.1%	100.0%
4	2.4%	6.0%	17.3%	36.9%	27.9%	9.4%	100.0%
5	3.6%	7.0%	20.2%	37.0%	24.5%	7.7%	100.0%
6	3.8%	10.3%	29.0%	39.9%	14.8%	2.3%	100.0%
7	5.2%	10.9%	30.3%	36.2%	15.4%	1.9%	100.0%
8	6.8%	12.1%	29.5%	36.3%	13.6%	1.6%	100.0%
9	9.7%	14.2%	17.1%	23.6%	25.2%	10.2%	100.0%
10	6.4%	15.0%	22.3%	27.4%	21.7%	7.2%	100.0%
11	5.9%	13.7%	24.2%	30.0%	18.8%	7.4%	100.0%
12	6.0%	13.2%	28.1%	31.5%	15.4%	5.7%	100.0%

4.4 Mean Raw Score Results by Standards

4.4.1 Comprehension Composite

4.4.1.1 By Cluster

Table 4.4.1.1

Mean Raw Score by Cluster by Tier by Standard: Comprehension S303

Cluster	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
1-2	A	Social Instructional Language	12	7.72	64.32%
		Language of Language Arts	9	5.04	56.04%
		Language of Math	9	5.25	58.28%
		Language of Science	6	3.62	60.37%
		Language of Social Studies	6	3.72	62.00%
	B	Social Instructional Language	6	4.54	75.65%
		Language of Language Arts	12	8.06	67.16%
		Language of Math	12	7.54	62.86%
		Language of Science	9	5.37	59.66%
		Language of Social Studies	9	6.24	69.29%
	C	Social Instructional Language	6	4.98	82.92%
		Language of Language Arts	12	8.36	69.68%
		Language of Math	12	7.10	59.15%
		Language of Science	9	5.49	60.97%
		Language of Social Studies	9	6.36	70.69%
3-5	A	Social Instructional Language	12	7.10	59.15%
		Language of Language Arts	9	4.93	54.82%
		Language of Math	9	4.73	52.53%
		Language of Science	6	2.66	44.30%
		Language of Social Studies	6	3.25	54.25%
	B	Social Instructional Language	6	4.81	80.17%
		Language of Language Arts	12	7.56	62.99%
		Language of Math	12	6.90	57.51%
		Language of Science	9	5.23	58.11%
		Language of Social Studies	9	5.47	60.81%
	C	Social Instructional Language	6	3.86	64.30%
		Language of Language Arts	12	7.71	64.27%
		Language of Math	12	5.20	43.31%
		Language of Science	9	4.79	53.19%
		Language of Social Studies	9	4.37	48.61%

Cluster	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
6-8	A	Social Instructional Language	12	5.91	49.27%
		Language of Language Arts	9	3.99	44.38%
		Language of Math	9	4.64	51.56%
		Language of Science	6	2.88	47.97%
		Language of Social Studies	6	3.04	50.61%
	B	Social Instructional Language	6	3.90	65.07%
		Language of Language Arts	12	8.65	72.09%
		Language of Math	12	7.13	59.44%
		Language of Science	9	4.93	54.82%
		Language of Social Studies	9	5.53	61.48%
	C	Social Instructional Language	6	4.29	71.50%
		Language of Language Arts	12	8.16	67.97%
		Language of Math	12	7.45	62.06%
		Language of Science	9	5.33	59.22%
Language of Social Studies		9	4.66	51.74%	
9-12	A	Social Instructional Language	12	5.80	48.30%
		Language of Language Arts	9	5.01	55.63%
		Language of Math	9	4.81	53.47%
		Language of Science	6	3.45	57.49%
		Language of Social Studies	6	3.67	61.14%
	B	Social Instructional Language	6	4.32	71.94%
		Language of Language Arts	12	7.30	60.81%
		Language of Math	12	7.14	59.49%
		Language of Science	9	4.87	54.06%
		Language of Social Studies	9	5.14	57.07%
	C	Social Instructional Language	6	3.85	64.12%
		Language of Language Arts	12	7.52	62.68%
		Language of Math	12	7.51	62.57%
		Language of Science	9	5.35	59.39%
Language of Social Studies		9	4.88	54.24%	

4.4.1.2 By Grade

Table 4.4.1.2

Mean Raw Score by Grade by Tier by Standard: Comprehension S303

Grade	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
1	A	Social Instructional Language	12	7.65	63.78%
		Language of Language Arts	9	4.96	55.11%
		Language of Math	9	5.15	57.18%
		Language of Science	6	3.59	59.87%
		Language of Social Studies	6	3.67	61.22%
	B	Social Instructional Language	6	4.24	70.62%
		Language of Language Arts	12	7.28	60.68%
		Language of Math	12	6.75	56.25%
		Language of Science	9	4.90	54.47%
		Language of Social Studies	9	5.62	62.41%
	C	Social Instructional Language	6	4.57	76.17%
		Language of Language Arts	12	7.45	62.07%
		Language of Math	12	6.18	51.51%
		Language of Science	9	4.81	53.45%
		Language of Social Studies	9	5.52	61.28%
2	A	Social Instructional Language	12	7.94	66.19%
		Language of Language Arts	9	5.33	59.22%
		Language of Math	9	5.59	62.07%
		Language of Science	6	3.73	62.09%
		Language of Social Studies	6	3.88	64.66%
	B	Social Instructional Language	6	4.82	80.33%
		Language of Language Arts	12	8.78	73.18%
		Language of Math	12	8.28	69.00%
		Language of Science	9	5.80	64.48%
		Language of Social Studies	9	6.81	75.70%
	C	Social Instructional Language	6	5.19	86.52%
		Language of Language Arts	12	8.85	73.74%
		Language of Math	12	7.59	63.21%
		Language of Science	9	5.85	64.98%
		Language of Social Studies	9	6.81	75.70%

Grade	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
3	A	Social Instructional Language	12	6.86	57.19%
		Language of Language Arts	9	4.76	52.92%
		Language of Math	9	4.52	50.17%
		Language of Science	6	2.49	41.47%
		Language of Social Studies	6	3.02	50.39%
	B	Social Instructional Language	6	4.66	77.68%
		Language of Language Arts	12	7.24	60.36%
		Language of Math	12	6.57	54.71%
		Language of Science	9	5.00	55.57%
		Language of Social Studies	9	5.18	57.59%
	C	Social Instructional Language	6	3.64	60.61%
		Language of Language Arts	12	7.24	60.36%
		Language of Math	12	4.74	39.49%
		Language of Science	9	4.45	49.48%
Language of Social Studies		9	3.98	44.18%	
4	A	Social Instructional Language	12	7.16	59.63%
		Language of Language Arts	9	4.95	54.94%
		Language of Math	9	4.80	53.37%
		Language of Science	6	2.68	44.64%
		Language of Social Studies	6	3.29	54.80%
	B	Social Instructional Language	6	4.89	81.46%
		Language of Language Arts	12	7.67	63.94%
		Language of Math	12	7.05	58.79%
		Language of Science	9	5.32	59.08%
		Language of Social Studies	9	5.59	62.09%
	C	Social Instructional Language	6	3.89	64.90%
		Language of Language Arts	12	7.79	64.91%
		Language of Math	12	5.26	43.87%
		Language of Science	9	4.82	53.56%
Language of Social Studies		9	4.43	49.24%	

Grade	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
5	A	Social Instructional Language	12	7.39	61.55%
		Language of Language Arts	9	5.18	57.52%
		Language of Math	9	4.96	55.12%
		Language of Science	6	2.89	48.16%
		Language of Social Studies	6	3.56	59.41%
	B	Social Instructional Language	6	5.09	84.77%
		Language of Language Arts	12	8.21	68.41%
		Language of Math	12	7.55	62.92%
		Language of Science	9	5.70	63.29%
		Language of Social Studies	9	6.06	67.29%
	C	Social Instructional Language	6	4.19	69.88%
		Language of Language Arts	12	8.42	70.20%
		Language of Math	12	5.90	49.19%
		Language of Science	9	5.32	59.09%
Language of Social Studies		9	4.99	55.43%	
6	A	Social Instructional Language	12	5.83	48.62%
		Language of Language Arts	9	4.01	44.51%
		Language of Math	9	4.49	49.93%
		Language of Science	6	2.87	47.91%
		Language of Social Studies	6	2.91	48.50%
	B	Social Instructional Language	6	3.71	61.77%
		Language of Language Arts	12	8.28	68.98%
		Language of Math	12	6.74	56.15%
		Language of Science	9	4.57	50.83%
		Language of Social Studies	9	5.08	56.41%
	C	Social Instructional Language	6	3.98	66.36%
		Language of Language Arts	12	7.35	61.28%
		Language of Math	12	6.87	57.26%
		Language of Science	9	4.91	54.54%
Language of Social Studies		9	4.05	45.02%	

Grade	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
7	A	Social Instructional Language	12	5.91	49.27%
		Language of Language Arts	9	3.97	44.06%
		Language of Math	9	4.65	51.68%
		Language of Science	6	2.88	47.95%
		Language of Social Studies	6	3.06	50.96%
	B	Social Instructional Language	6	3.93	65.50%
		Language of Language Arts	12	8.69	72.45%
		Language of Math	12	7.15	59.61%
		Language of Science	9	4.97	55.17%
		Language of Social Studies	9	5.56	61.74%
	C	Social Instructional Language	6	4.32	71.97%
		Language of Language Arts	12	8.20	68.36%
		Language of Math	12	7.46	62.16%
		Language of Science	9	5.34	59.29%
Language of Social Studies		9	4.67	51.85%	
8	A	Social Instructional Language	12	5.99	49.92%
		Language of Language Arts	9	4.01	44.57%
		Language of Math	9	4.78	53.06%
		Language of Science	6	2.88	48.05%
		Language of Social Studies	6	3.14	52.36%
	B	Social Instructional Language	6	4.07	67.85%
		Language of Language Arts	12	8.97	74.76%
		Language of Math	12	7.50	62.49%
		Language of Science	9	5.25	58.38%
		Language of Social Studies	9	5.96	66.18%
	C	Social Instructional Language	6	4.57	76.13%
		Language of Language Arts	12	8.91	74.21%
		Language of Math	12	8.00	66.70%
		Language of Science	9	5.74	63.79%
Language of Social Studies		9	5.25	58.29%	

Grade	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
9	A	Social Instructional Language	12	5.44	45.31%
		Language of Language Arts	9	4.74	52.70%
		Language of Math	9	4.62	51.34%
		Language of Science	6	3.32	55.35%
		Language of Social Studies	6	3.54	58.95%
	B	Social Instructional Language	6	4.35	72.49%
		Language of Language Arts	12	7.31	60.88%
		Language of Math	12	7.03	58.61%
		Language of Science	9	4.80	53.36%
		Language of Social Studies	9	5.11	56.72%
	C	Social Instructional Language	6	3.92	65.27%
		Language of Language Arts	12	7.61	63.42%
		Language of Math	12	7.46	62.13%
		Language of Science	9	5.32	59.08%
Language of Social Studies		9	4.79	53.23%	
10	A	Social Instructional Language	12	6.10	50.80%
		Language of Language Arts	9	5.25	58.28%
		Language of Math	9	4.99	55.42%
		Language of Science	6	3.57	59.47%
		Language of Social Studies	6	3.80	63.37%
	B	Social Instructional Language	6	4.26	70.96%
		Language of Language Arts	12	7.20	60.00%
		Language of Math	12	7.08	59.02%
		Language of Science	9	4.82	53.58%
		Language of Social Studies	9	5.08	56.50%
	C	Social Instructional Language	6	3.83	63.78%
		Language of Language Arts	12	7.46	62.19%
		Language of Math	12	7.46	62.14%
		Language of Science	9	5.31	58.97%
Language of Social Studies		9	4.87	54.07%	

Grade	Tier	Standard	Maximum Score	Mean Score	Percent of Maximum
11	A	Social Instructional Language	12	6.44	53.66%
		Language of Language Arts	9	5.47	60.78%
		Language of Math	9	5.19	57.72%
		Language of Science	6	3.69	61.45%
		Language of Social Studies	6	3.92	65.33%
	B	Social Instructional Language	6	4.32	71.94%
		Language of Language Arts	12	7.35	61.24%
		Language of Math	12	7.30	60.87%
		Language of Science	9	4.97	55.25%
		Language of Social Studies	9	5.20	57.77%
	C	Social Instructional Language	6	3.81	63.50%
		Language of Language Arts	12	7.51	62.57%
		Language of Math	12	7.64	63.66%
		Language of Science	9	5.44	60.45%
Language of Social Studies		9	4.98	55.33%	
12	A	Social Instructional Language	12	6.58	54.82%
		Language of Language Arts	9	5.53	61.40%
		Language of Math	9	5.10	56.61%
		Language of Science	6	3.67	61.12%
		Language of Social Studies	6	3.84	64.02%
	B	Social Instructional Language	6	4.34	72.30%
		Language of Language Arts	12	7.38	61.50%
		Language of Math	12	7.30	60.82%
		Language of Science	9	4.96	55.17%
		Language of Social Studies	9	5.23	58.08%
	C	Social Instructional Language	6	3.76	62.72%
		Language of Language Arts	12	7.42	61.82%
		Language of Math	12	7.55	62.88%
		Language of Science	9	5.36	59.50%
Language of Social Studies		9	4.99	55.49%	

4.4.2 Writing

4.4.2.1 By Cluster

Table 4.4.2.1

Mean Raw Score by Cluster by Tier by Standard: Writing S303

Cluster	Tier	Standard	Mean Raw Score				Percent of Maximum
			Linguistic Complexity	Vocabulary Usage	Language Control	Total	
1-2	A	Social Instructional Language	5.09	4.78	4.23	14.10	19.59%
		Social Instructional Language	1.19	1.92	1.30	4.41	24.51%
	B	Language of Math / Science	2.39	2.26	1.96	6.62	36.76%
		Language of Language Arts / Social Studies	2.50	2.30	1.77	6.57	36.50%
	C	Social Instructional Language	2.95	2.65	2.31	7.90	43.90%
		Language of Math / Science	2.96	2.78	2.29	8.03	44.63%
		Language of Language Arts / Social Studies	3.03	2.70	2.30	8.04	44.64%
3-5	A	Social Instructional Language	2.11	1.84	1.57	5.51	30.61%
		Language of Math / Science	2.09	2.21	1.69	5.99	33.25%
		Language of Language Arts	2.07	1.75	1.51	5.33	29.62%
	B	Social Instructional Language	2.88	2.61	2.40	7.89	43.84%
		Language of Math / Science	2.91	3.06	2.54	8.50	47.25%
		Language of Language Arts / Social Studies	2.84	2.36	2.35	7.55	41.95%
	C	Social Instructional Language	3.07	2.86	2.70	8.63	47.95%
		Language of Math / Science	3.07	2.97	2.68	8.72	48.45%
		Language of Language Arts / Social Studies	3.11	2.68	2.68	8.47	47.06%
6-8	A	Social Instructional Language	2.18	1.84	1.71	5.73	31.81%
		Language of Math / Science	2.12	1.66	1.71	5.49	30.51%
		Language of Language Arts	2.20	1.94	1.65	5.79	32.17%
	B	Social Instructional Language	3.25	2.64	2.64	8.54	47.42%
		Language of Math / Science	3.17	3.37	2.67	9.21	51.16%
		Language of Language Arts / Social Studies	3.22	2.57	2.57	8.36	46.44%
	C	Social Instructional Language	3.57	2.93	2.96	9.46	52.57%
		Language of Math / Science	3.67	3.75	3.06	10.48	58.25%
		Language of Language Arts / Social Studies	3.63	2.90	2.93	9.46	52.56%
9-12	A	Social Instructional Language	2.16	1.97	1.75	5.88	32.64%
		Language of Math / Science	2.20	2.01	1.63	5.84	32.42%
		Language of Language Arts	2.49	2.05	1.80	6.35	35.26%
	B	Social Instructional Language	3.24	2.87	2.72	8.82	49.01%
		Language of Math / Science	3.20	2.88	2.68	8.76	48.64%
		Language of Language Arts / Social Studies	3.09	3.08	2.48	8.65	48.06%
	C	Social Instructional Language	3.62	3.18	3.16	9.95	55.30%
		Language of Math / Science	3.33	3.63	2.94	9.89	54.96%
		Language of Language Arts / Social Studies	3.61	3.50	3.01	10.13	56.26%

4.4.2.2 By Grade

Table 4.4.2.2

Mean Raw Score by Grade by Tier by Standard: Writing S303

Grade	Tier	Standard	Mean Raw Score				Percent of Maximum
			Linguistic Complexity	Vocabulary Usage	Language Control	Total	
1	A	Social Instructional Language	5.03	4.73	4.14	13.90	19.30%
		Social Instructional Language	1.22	1.88	1.20	4.29	23.86%
	B	Language of Math / Science	2.17	2.09	1.76	6.03	33.47%
		Language of Language Arts / Social Studies	2.18	2.06	1.49	5.74	31.88%
	C	Social Instructional Language	2.72	2.47	2.05	7.23	40.19%
		Language of Math / Science	2.62	2.61	1.98	7.22	40.11%
		Language of Language Arts / Social Studies	2.67	2.44	1.95	7.06	39.24%
2	A	Social Instructional Language	5.30	4.96	4.55	14.81	20.58%
		Social Instructional Language	1.17	1.96	1.40	4.52	25.12%
	B	Language of Math / Science	2.60	2.42	2.14	7.17	39.81%
		Language of Language Arts / Social Studies	2.80	2.52	2.03	7.34	40.80%
	C	Social Instructional Language	3.07	2.74	2.45	8.26	45.88%
		Language of Math / Science	3.14	2.87	2.46	8.47	47.03%
		Language of Language Arts / Social Studies	3.22	2.84	2.49	8.55	47.52%
3	A	Social Instructional Language	2.02	1.77	1.47	5.26	29.22%
		Language of Math / Science	1.99	2.14	1.60	5.73	31.85%
		Language of Language Arts	1.97	1.69	1.41	5.07	28.18%
	B	Social Instructional Language	2.82	2.54	2.32	7.67	42.63%
		Language of Math / Science	2.86	2.94	2.46	8.26	45.88%
			Language of Language Arts / Social Studies	2.74	2.23	2.25	7.23
C	Social Instructional Language	3.00	2.77	2.62	8.39	46.61%	
	Language of Math / Science	2.98	2.92	2.59	8.50	47.20%	
	Language of Language Arts / Social Studies	3.00	2.51	2.57	8.08	44.88%	
4	A	Social Instructional Language	2.12	1.85	1.58	5.54	30.79%
		Language of Math / Science	2.10	2.22	1.71	6.04	33.55%
		Language of Language Arts	2.09	1.76	1.52	5.38	29.88%
	B	Social Instructional Language	2.90	2.64	2.42	7.97	44.25%
		Language of Math / Science	2.93	3.12	2.57	8.62	47.87%
			Language of Language Arts / Social Studies	2.87	2.42	2.39	7.68
C	Social Instructional Language	3.08	2.87	2.71	8.66	48.13%	
	Language of Math / Science	3.09	2.98	2.69	8.76	48.66%	
		Language of Language Arts / Social Studies	3.13	2.71	2.69	8.53	47.39%

Grade	Tier	Standard	Mean Raw Score				Percent of Maximum
			Linguistic Complexity	Vocabulary Usage	Language Control	Total	
5	A	Social Instructional Language	2.23	1.93	1.69	5.85	32.50%
		Language of Math / Science	2.20	2.30	1.80	6.30	35.02%
		Language of Language Arts	2.20	1.83	1.63	5.67	31.49%
	B	Social Instructional Language	3.02	2.76	2.57	8.34	46.34%
		Language of Math / Science	3.01	3.28	2.69	8.98	49.89%
		Language of Language Arts / Social Studies	3.02	2.63	2.55	8.20	45.57%
	C	Social Instructional Language	3.18	2.99	2.83	9.00	50.02%
		Language of Math / Science	3.20	3.05	2.81	9.06	50.34%
		Language of Language Arts / Social Studies	3.27	2.95	2.85	9.07	50.41%
6	A	Social Instructional Language	2.13	1.80	1.66	5.59	31.04%
		Language of Math / Science	2.05	1.62	1.66	5.34	29.66%
		Language of Language Arts	2.13	1.88	1.59	5.60	31.10%
	B	Social Instructional Language	3.13	2.52	2.52	8.18	45.42%
		Language of Math / Science	3.03	3.21	2.55	8.79	48.84%
		Language of Language Arts / Social Studies	3.09	2.44	2.44	7.96	44.24%
	C	Social Instructional Language	3.44	2.79	2.82	9.05	50.26%
		Language of Math / Science	3.51	3.63	2.90	10.04	55.76%
		Language of Language Arts / Social Studies	3.46	2.73	2.77	8.96	49.76%
7	A	Social Instructional Language	2.18	1.84	1.71	5.73	31.84%
		Language of Math / Science	2.12	1.67	1.72	5.50	30.58%
		Language of Language Arts	2.21	1.95	1.66	5.82	32.31%
	B	Social Instructional Language	3.26	2.65	2.65	8.56	47.55%
		Language of Math / Science	3.18	3.38	2.68	9.25	51.38%
		Language of Language Arts / Social Studies	3.23	2.58	2.58	8.39	46.59%
	C	Social Instructional Language	3.58	2.94	2.97	9.49	52.71%
		Language of Math / Science	3.68	3.77	3.07	10.52	58.45%
		Language of Language Arts / Social Studies	3.64	2.91	2.94	9.49	52.70%
8	A	Social Instructional Language	2.22	1.88	1.76	5.86	32.55%
		Language of Math / Science	2.17	1.70	1.76	5.63	31.28%
		Language of Language Arts	2.26	1.99	1.70	5.96	33.09%
	B	Social Instructional Language	3.35	2.76	2.75	8.86	49.25%
		Language of Math / Science	3.30	3.50	2.78	9.58	53.22%
		Language of Language Arts / Social Studies	3.34	2.69	2.69	8.72	48.44%
	C	Social Instructional Language	3.69	3.08	3.08	9.85	54.73%
		Language of Math / Science	3.83	3.86	3.21	10.89	60.51%
		Language of Language Arts / Social Studies	3.79	3.07	3.08	9.94	55.21%

Grade	Tier	Standard	Mean Raw Score				Percent of Maximum
			Linguistic Complexity	Vocabulary Usage	Language Control	Total	
9	A	Social Instructional Language	2.05	1.89	1.65	5.59	31.08%
		Language of Math / Science	2.05	1.91	1.51	5.47	30.39%
		Language of Language Arts	2.35	1.94	1.68	5.98	33.21%
	B	Social Instructional Language	3.23	2.84	2.73	8.80	48.90%
		Language of Math / Science	3.18	2.84	2.68	8.71	48.38%
		Language of Language Arts / Social Studies	3.08	3.05	2.49	8.62	47.87%
	C	Social Instructional Language	3.63	3.16	3.19	9.98	55.46%
		Language of Math / Science	3.36	3.66	2.99	10.01	55.59%
		Language of Language Arts / Social Studies	3.62	3.49	3.04	10.16	56.42%
10	A	Social Instructional Language	2.27	2.04	1.84	6.15	34.15%
		Language of Math / Science	2.33	2.12	1.73	6.18	34.31%
		Language of Language Arts	2.63	2.16	1.92	6.71	37.28%
	B	Social Instructional Language	3.20	2.83	2.68	8.71	48.41%
		Language of Math / Science	3.18	2.87	2.64	8.68	48.24%
		Language of Language Arts / Social Studies	3.04	3.05	2.43	8.52	47.36%
	C	Social Instructional Language	3.61	3.16	3.14	9.91	55.06%
		Language of Math / Science	3.31	3.61	2.92	9.84	54.67%
		Language of Language Arts / Social Studies	3.60	3.50	3.01	10.11	56.17%
11	A	Social Instructional Language	2.35	2.11	1.91	6.37	35.37%
		Language of Math / Science	2.45	2.21	1.83	6.50	36.11%
		Language of Language Arts	2.73	2.24	2.02	6.98	38.79%
	B	Social Instructional Language	3.28	2.90	2.73	8.91	49.48%
		Language of Math / Science	3.25	2.93	2.70	8.87	49.30%
		Language of Language Arts / Social Studies	3.13	3.14	2.50	8.77	48.74%
	C	Social Instructional Language	3.63	3.20	3.14	9.97	55.41%
		Language of Math / Science	3.33	3.62	2.92	9.87	54.83%
		Language of Language Arts / Social Studies	3.62	3.53	3.00	10.14	56.35%
12	A	Social Instructional Language	2.36	2.10	1.91	6.38	35.42%
		Language of Math / Science	2.44	2.20	1.84	6.48	35.99%
		Language of Language Arts	2.72	2.25	2.02	6.99	38.84%
	B	Social Instructional Language	3.29	2.94	2.74	8.96	49.78%
		Language of Math / Science	3.24	2.92	2.70	8.85	49.18%
		Language of Language Arts / Social Studies	3.15	3.15	2.51	8.81	48.93%
	C	Social Instructional Language	3.60	3.21	3.12	9.93	55.15%
		Language of Math / Science	3.28	3.57	2.88	9.74	54.09%
		Language of Language Arts / Social Studies	3.59	3.51	2.97	10.06	55.91%

4.4.3 Speaking

4.4.3.1 By Cluster

Table 4.4.3.1

Mean Raw Score by Cluster by Tier by Standard: Speaking S303

Cluster	Tier	Standard	Maximum Score	Mean Raw Score	Percentage of Maximum
1-2	A	Social and Instructional Language	3	2.33	77.72%
		Language of Language Arts/Social Studies	5	2.88	57.57%
		Language of Mathematics/Science	5	2.44	48.86%
	B	Social and Instructional Language	3	2.85	94.96%
		Language of Language Arts/Social Studies	5	4.03	80.69%
		Language of Mathematics/Science	5	3.65	73.10%
	C	Social and Instructional Language	3	2.94	97.96%
		Language of Language Arts/Social Studies	5	4.54	90.87%
		Language of Mathematics/Science	5	4.32	86.42%
3-5	A	Social and Instructional Language	3	1.70	56.70%
		Language of Language Arts/Social Studies	5	1.95	39.08%
		Language of Mathematics/Science	5	1.57	31.36%
	B	Social and Instructional Language	3	2.84	94.58%
		Language of Language Arts/Social Studies	5	3.78	75.53%
		Language of Mathematics/Science	5	3.56	71.15%
	C	Social and Instructional Language	3	2.93	97.82%
		Language of Language Arts/Social Studies	5	4.32	86.30%
		Language of Mathematics/Science	5	4.20	83.96%
6-8	A	Social and Instructional Language	3	1.68	56.13%
		Language of Language Arts/Social Studies	5	1.76	35.13%
		Language of Mathematics/Science	5	1.45	29.08%
	B	Social and Instructional Language	3	2.83	94.21%
		Language of Language Arts/Social Studies	5	3.98	79.54%
		Language of Mathematics/Science	5	3.57	71.48%
	C	Social and Instructional Language	3	2.93	97.78%
		Language of Language Arts/Social Studies	5	4.50	89.98%
		Language of Mathematics/Science	5	4.21	84.28%
9-12	A	Social and Instructional Language	3	1.65	55.10%
		Language of Language Arts/Social Studies	5	1.60	31.98%
		Language of Mathematics/Science	5	1.45	29.07%
	B	Social and Instructional Language	3	2.77	92.29%
		Language of Language Arts/Social Studies	5	3.83	76.53%
		Language of Mathematics/Science	5	3.43	68.66%
	C	Social and Instructional Language	3	2.91	97.06%
		Language of Language Arts/Social Studies	5	4.58	91.61%
		Language of Mathematics/Science	5	4.27	85.36%

4.4.3.2 By Grade

Table 4.4.3.2

Mean Raw Score by Grade by Tier by Standard: Speaking S303

Grade	Tier	Standard	Maximum Score	Mean Raw Score	Percent of Maximum
1	A	Social and Instructional Language	3	2.39	79.55%
		Language of Language Arts/Social Studies	5	2.96	59.10%
		Language of Mathematics/Science	5	2.51	50.25%
	B	Social and Instructional Language	3	2.82	93.85%
		Language of Language Arts/Social Studies	5	3.87	77.31%
		Language of Mathematics/Science	5	3.47	69.34%
	C	Social and Instructional Language	3	2.92	97.32%
		Language of Language Arts/Social Studies	5	4.40	87.93%
		Language of Mathematics/Science	5	4.14	82.83%
2	A	Social and Instructional Language	3	2.14	71.39%
		Language of Language Arts/Social Studies	5	2.61	52.29%
		Language of Mathematics/Science	5	2.20	44.05%
	B	Social and Instructional Language	3	2.88	95.99%
		Language of Language Arts/Social Studies	5	4.19	83.84%
		Language of Mathematics/Science	5	3.83	76.59%
	C	Social and Instructional Language	3	2.95	98.30%
		Language of Language Arts/Social Studies	5	4.62	92.43%
		Language of Mathematics/Science	5	4.42	88.33%
3	A	Social and Instructional Language	3	1.74	58.03%
		Language of Language Arts/Social Studies	5	1.99	39.77%
		Language of Mathematics/Science	5	1.55	31.05%
	B	Social and Instructional Language	3	2.83	94.39%
		Language of Language Arts/Social Studies	5	3.71	74.23%
		Language of Mathematics/Science	5	3.46	69.12%
	C	Social and Instructional Language	3	2.93	97.63%
		Language of Language Arts/Social Studies	5	4.25	85.08%
		Language of Mathematics/Science	5	4.11	82.26%
4	A	Social and Instructional Language	3	1.68	56.01%
		Language of Language Arts/Social Studies	5	1.93	38.61%
		Language of Mathematics/Science	5	1.57	31.40%
	B	Social and Instructional Language	3	2.84	94.68%
		Language of Language Arts/Social Studies	5	3.79	75.89%
		Language of Mathematics/Science	5	3.60	72.06%
	C	Social and Instructional Language	3	2.94	97.92%
		Language of Language Arts/Social Studies	5	4.31	86.25%
		Language of Mathematics/Science	5	4.21	84.15%

Grade	Tier	Standard	Maximum Score	Mean Raw Score	Percent of Maximum
5	A	Social and Instructional Language	3	1.66	55.47%
		Language of Language Arts/Social Studies	5	1.93	38.58%
		Language of Mathematics/Science	5	1.59	31.78%
	B	Social and Instructional Language	3	2.85	94.92%
		Language of Language Arts/Social Studies	5	3.92	78.37%
		Language of Mathematics/Science	5	3.75	75.07%
	C	Social and Instructional Language	3	2.94	98.03%
		Language of Language Arts/Social Studies	5	4.42	88.45%
		Language of Mathematics/Science	5	4.33	86.64%
6	A	Social and Instructional Language	3	1.74	57.94%
		Language of Language Arts/Social Studies	5	1.84	36.73%
		Language of Mathematics/Science	5	1.53	30.66%
	B	Social and Instructional Language	3	2.82	93.83%
		Language of Language Arts/Social Studies	5	3.90	78.03%
		Language of Mathematics/Science	5	3.46	69.23%
	C	Social and Instructional Language	3	2.92	97.23%
		Language of Language Arts/Social Studies	5	4.39	87.86%
		Language of Mathematics/Science	5	4.05	81.05%
7	A	Social and Instructional Language	3	1.67	55.75%
		Language of Language Arts/Social Studies	5	1.74	34.79%
		Language of Mathematics/Science	5	1.43	28.58%
	B	Social and Instructional Language	3	2.83	94.41%
		Language of Language Arts/Social Studies	5	3.99	79.80%
		Language of Mathematics/Science	5	3.61	72.12%
	C	Social and Instructional Language	3	2.94	97.97%
		Language of Language Arts/Social Studies	5	4.51	90.16%
		Language of Mathematics/Science	5	4.24	84.88%
8	A	Social and Instructional Language	3	1.64	54.72%
		Language of Language Arts/Social Studies	5	1.69	33.88%
		Language of Mathematics/Science	5	1.40	28.00%
	B	Social and Instructional Language	3	2.83	94.38%
		Language of Language Arts/Social Studies	5	4.04	80.76%
		Language of Mathematics/Science	5	3.65	73.02%
	C	Social and Instructional Language	3	2.94	98.13%
		Language of Language Arts/Social Studies	5	4.59	91.89%
		Language of Mathematics/Science	5	4.34	86.88%
9	A	Social and Instructional Language	3	1.51	50.29%
		Language of Language Arts/Social Studies	5	1.40	28.03%
		Language of Mathematics/Science	5	1.29	25.79%
	B	Social and Instructional Language	3	2.78	92.65%
		Language of Language Arts/Social Studies	5	3.92	78.31%
		Language of Mathematics/Science	5	3.51	70.16%
	C	Social and Instructional Language	3	2.92	97.43%
		Language of Language Arts/Social Studies	5	4.63	92.67%
		Language of Mathematics/Science	5	4.32	86.48%

Grade	Tier	Standard	Maximum Score	Mean Raw Score	Percent of Maximum
10	A	Social and Instructional Language	3	1.78	59.21%
		Language of Language Arts/Social Studies	5	1.74	34.82%
		Language of Mathematics/Science	5	1.57	31.49%
	B	Social and Instructional Language	3	2.74	91.43%
		Language of Language Arts/Social Studies	5	3.72	74.31%
		Language of Mathematics/Science	5	3.33	66.51%
	C	Social and Instructional Language	3	2.91	97.06%
		Language of Language Arts/Social Studies	5	4.57	91.46%
		Language of Mathematics/Science	5	4.24	84.81%
11	A	Social and Instructional Language	3	1.88	62.61%
		Language of Language Arts/Social Studies	5	1.93	38.58%
		Language of Mathematics/Science	5	1.73	34.58%
	B	Social and Instructional Language	3	2.78	92.60%
		Language of Language Arts/Social Studies	5	3.78	75.59%
		Language of Mathematics/Science	5	3.40	67.98%
	C	Social and Instructional Language	3	2.91	97.06%
		Language of Language Arts/Social Studies	5	4.55	91.00%
		Language of Mathematics/Science	5	4.24	84.87%
12	A	Social and Instructional Language	3	2.02	67.29%
		Language of Language Arts/Social Studies	5	2.16	43.12%
		Language of Mathematics/Science	5	1.90	37.98%
	B	Social and Instructional Language	3	2.78	92.56%
		Language of Language Arts/Social Studies	5	3.86	77.23%
		Language of Mathematics/Science	5	3.48	69.58%
	C	Social and Instructional Language	3	2.89	96.18%
		Language of Language Arts/Social Studies	5	4.51	90.11%
		Language of Mathematics/Science	5	4.21	84.16%

World-Class Instructional Design and Assessment



**Annual Technical Report for
ACCESS for ELLs English Language Proficiency Test,
Series 303, 2014-2015 Administration**

**Annual Technical Report No. 11
Volume 2 of 3: Analyses of Test Forms**

Prepared by:

Center for Applied Linguistics

CAL/WIDA Partnership Activities
Psychometrics/Research Team

February 22, 2016

Table of Contents

Volume 2

5. Analyses of Test Forms: Overview	122
5.1 Background.....	122
5.1.1 Measurement Models Used.....	122
5.1.2 Sampling	124
5.1.3 Equating and Scaling	124
5.1.4 DIF Analyses	124
5.1.4.1 Dichotomous Items	125
5.1.4.2 Polytomous Items.....	125
5.2 Descriptions.....	127
5.2.1 Raw Score Information (Figure A and Table A)	127
5.2.2 Scale Score Information (Figure B and Table B)	127
5.2.3 Proficiency Level Information (Figure C and Table C).....	128
5.2.4 Scaling Equation Table (Table D)	129
5.2.5 Equating Summary (Table E)	129
5.2.6 Test Characteristic Curve (Figure D).....	130
5.2.7 Test Information Function (Figure E).....	130
5.2.8 Reliability (Table F).....	131
5.2.9 Item/Task Analysis Summary (Table G)	132
5.2.10 Complete Item Analysis Table (Table H).....	133
5.2.11 Complete Raw Score to Scale Score Conversion Chart (Table I)	134
5.2.12 Raw Score to Proficiency Level Score Conversion Table (Table J)	135

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* (Kenyon, 2006). 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 Annual 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), the author of the *Winsteps* software 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;” and
- values below 0.5 Linacre calls “less productive for measurement, but not degrading.”

Linacre also states in his 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 303 of ACCESS for ELLs in the academic year 2014–2015. 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 Annual Technical Report No.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 to report 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 303 represent a partial refreshment of Series 302. That is, while many items were common on both forms, certain folders on Series 302 were replaced with new items (see Chapter 1.4). Thus, to place the results of Series 303 onto the ACCESS for ELLs score scale, items that were not revised or otherwise changed were anchored to the difficulty values from Series 302, which itself had been anchored to Series 301. Table E in Section 6 for each test form provides explicit information on the anchor items used for equating Series 303 results to those of Series 302.

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 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 that were originally proposed by 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 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 statistic to be appropriately calculated. (Note that this software program uses a two-step purification process; that is, items with C-level DIF in the first pass are removed from the matching variable in the second stage, and the DIF is then recalculated for the remaining items.)

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., Writing tasks), a similar approach is used. It is based on the Mantel-Haenszel 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 six groups based on their total raw score on the Writing test. These categories are determined by calculating what the total raw score would be for a student scoring WIDA Proficiency Levels 1, 2, 3, 4, 5, or 6 in each category. 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 on the same forms.

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 few students at Proficiency Level 6.)

For each Writing task, performance was similarly categorized according to the scoring rubric. Thus, raw scores of 0 to 4 fell under 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 a 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-Haenszel 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 comparison groups is calculated. The standardized mean difference compares the means of the two groups, adjusting for differences in the distribution of the two comparison groups 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-Haenszel 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-Haenszel 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-Haenszel 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., correct or incorrect). 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 grade-level cluster 1–2, Tier A has a weight of 3 for the fourth task. For grade-level 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 grade-level clusters 3–5, 6–8, and 9–12 to 108 for the Tier B and C forms and for Tier A for grade-level cluster 1–2 (see Chapter 1.7.2).

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 and by total for the grade-level cluster:

- the number of students in the analyses (the number of students who were not absent, invalid, refused, exempt, or in the wrong grade-level cluster),
- the minimum observed raw score,
- the maximum observed raw score,
- the mean (average) raw score, and
- 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 presented 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 and by total for the grade-level cluster:

- the number of students in the analyses,
- the minimum observed scale score,

- the maximum observed scale score,
- the mean (average) scale score, and
- 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, as well as the cap, increase by grade within a grade-level cluster. For example, for Listening 3–5 Tier A, 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 18 (out of 18) 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. For more information, see WIDA Technical Report 1, *Development and Field Test of ACCESS for ELLs* (Kenyon, 2006).

Also note that, because the scale is vertically equated, the range of scale scores moves up the scale from one grade-level 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 grades 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 grade-level cluster as a whole. The horizontal axis shows the six WIDA Proficiency Levels. The vertical axis shows the percentage of students. Each bar shows the percentage of students who were placed into each Proficiency Level in the domain being tested on this test form.

Each row of Table C shows, by grade and by total for the grade-level 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, and
- the percentage of students, out of the total number of students taking the form who were placed into that Proficiency Level in the domain being tested.

(Note that for Kindergarten and Tier A tests in some domains, 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 is refreshed. A post-equating procedure known as common-item equating is used to equate the results on new forms to the older forms. This means that the difficulty measure for items that appear on both the new and the old forms are kept constant across both forms. Thus, performances on the newer form may be interpreted with the same frame of reference.

Many items appearing on ACCESS for ELLs Series 303 also appeared on Series 302. 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 303 version of that test form) to the previous year's test (i.e., the Series 302 version of that test form). The number of items, the average item difficulty, the standard deviation of the item difficulty values, and the difficulty value of the easiest and hardest item on each test form are shown. 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 these items (e.g., if anchor items 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 303 test form and on the Series 302 test form. Typically, random displacements of less than 0.5 logits are unlikely to have much impact on measurement in a test instrument (Linacre, n.d.).

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). Additionally, 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.

Note that for Series 303, no equating was performed for Writing Tier A for grade cluster 1–2. Also, no equating was performed for all Listening and Speaking grade-level clusters. The results of the unequated tests needed for certain tables were taken from the results of Series 302.

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 grade-level 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 Kindergarten and Tier A tests in some domains, 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 how many 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 Figure D, five vertical lines in Figure E indicate the five cut scores for the highest grade in the grade-level cluster for the test form, dividing the figure into six sections for each of the WIDA Proficiency Levels (1–6) for the domain being tested. (Note that for Kindergarten and Tier A tests in some domains, it was not possible to place into all six language proficiency levels. Note also that, although Listening and Reading scores on Tiers A and B were capped, all five horizontal lines indicating the cut points remain in this figure.) It is important that each test form measures 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.

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), and
- 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 examinees (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 the ability distribution of the group of examinees. 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

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

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 interrater 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 responses 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, adjacent (adj), 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.

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 Section 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 Section 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 or Hispanic versus other ethnicities. Even though DIF may be negligible (category A or AA), this table shows the number of items that favored 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, responses that are written in languages other than English or are totally incomprehensible will receive a score of 0, while papers that demonstrate the ability to copy or write a few words in English will be awarded a score of 1.)

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., 303).

Note that for Writing, the “integrated” task (“IT”), which requires more extensive writing, integrates Model Performance Indicators for WIDA ELD Standards 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_303”, 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 (except for Kindergarten), 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 (percentage of correct answers on that item or, in the case of Speaking, percentage 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 Section 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, in many cases, examinees 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 examinees who would be expected to meet the 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, these responses were removed from the data set when analyzing fit for the Speaking test. Table 5.2.10 shows the number of such cases that were removed from the analysis for each cluster.

Table 5.2.10

Rate of Speaking responses removed from fit analysis S303

Cluster	No. of responses	No. of responses removed	Percent of responses removed
1-2	438,394	17,225	3.9%
3-5	418,023	20,299	4.9%
6-8	250,160	6,208	2.5%
9-12	234,787	7,172	3.1%

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. Potential sources for these high outfit values are continually investigated.

Note also that the Kindergarten test used a new format starting with Series 200 (academic year 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 following column(s) show the corresponding scale score for each grade level in the grade-level cluster. Note that Tier A Listening and Reading items have been capped to the scale score that represents the Proficiency Level score of 4.0. Tier B Listening and Reading items 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 scale scores on the upper end of the raw score scale increased 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 always *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 the highest possible scores have been capped for Tiers A and B, preventing the inflation of scale scores due to large standard errors at the highest end of the raw score scale, 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. Because 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 grade-level cluster. In this case, the scores have been marked in Table I as capped, and the standard error, as well as the low and high bound for the capped scale score, have been repeated in the final rows of the table. In addition, at the lower end of the raw score scale, scale scores are truncated where 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 grade-level 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 by grade-level cluster (i.e., 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 above and beyond 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 other 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, its associated Proficiency Level score

now 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.

The second note is, because scale scores for Tiers A and B in Listening and Reading are capped 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 grade-level cluster). These differences 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 303, 2014-2015 Administration**

**Annual Technical Report No. 11
Volume 3 of 3: Analyses Across Tiers**

Prepared by:

Center for Applied Linguistics

CAL/WIDA Partnership Activities
Psychometrics/Research Team

February 22, 2016

Table of Contents

Volume 3

7. Analysis Across Tiers: Overview	453
7.1 Background	453
7.1.1 Reliability of Composites	453
7.1.2 Accuracy and Consistency of Classification.....	453
7.2 Descriptions.....	454
7.2.1 Scale Score Information (Figure A and Table A).....	454
7.2.2 Proficiency Level Information (Figure B and Table B).....	455
7.2.3 Conditional Standard Error of Measurement (Table C and Figures C and D)	455
7.2.4 Reliability Information (Table D).....	456
7.2.5 Accuracy and Consistency of Classification Tables (Table E).....	457
8. Analysis Across Tiers: Results.....	458
8.1 Grade: K.....	458
8.1.1 Listening K.....	458
8.1.2 Reading K	462
8.1.3 Writing K	466
8.1.4 Speaking K.....	470
8.1.5 Oral Language Composite K	474
8.1.6 Literacy Composite K.....	477
8.1.7 Comprehension Composite K.....	480
8.1.8 Overall Composite K	483
8.2 Grades: 1–2	486
8.2.1 Listening 1–2	486
8.2.2 Reading 1–2	490
8.2.3 Writing 1–2.....	494
8.2.4 Speaking 1–2.....	498
8.2.5 Oral Language Composite 1–2	502
8.2.6 Literacy Composite 1–2.....	505
8.2.7 Comprehension Composite 1–2.....	508
8.2.8 Overall Composite 1–2	511
8.3 Grades: 3–5	514
8.3.1 Listening 3–5	514
8.3.2 Reading 3–5	519
8.3.3 Writing 3–5	525
8.3.4 Speaking 3–5.....	530
8.3.5 Oral 3–5	535
8.3.6 Literacy Composite 3–5.....	538
8.3.7 Comprehension Composite 3–5.....	541
8.3.8 Overall Composite 3–5	544

8.4	Grades: 6–8	547
8.4.1	Listening 6–8	547
8.4.2	Reading 6–8	552
8.4.3	Writing 6–8	557
8.4.4	Speaking 6–8	562
8.4.5	Oral Language Composite 6–8	567
8.4.6	Literacy Composite 6–8	570
8.4.7	Comprehension Composite 6–8	573
8.4.8	Overall Composite 6–8	576
8.5	Grades: 9–12	579
8.5.1	Listening 9–12	579
8.5.2	Reading 9–12	584
8.5.3	Writing 9–12	589
8.5.4	Speaking 9–12	594
8.5.5	Oral Language Composite 9–12	599
8.5.6	Literacy Composite 9–12	603
8.5.7	Comprehension Composite 9–12	607
8.5.8	Overall Composite 9–12	611
References		615
Acknowledgements		619

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; Kane & Case, 2004; Rudner, 2001) is computed, weighted by the contribution of each domain score into the composite. Specifically, the formula is

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

where

k = number of components j

w_j = weight of component j

σ_j^2 = variance of component j

σ_c^2 = variance of composite

ρ_j = reliability coefficient of component j .

The data to compute the stratified Cronbach's alpha is provided in the appropriate tables in Chapter 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 performance 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 according to their performance on ACCESS for ELLs, it is important to know how well these classifications are made. The analyses that were used utilize 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 Level cut score (based on their “true score”), but were classified based on the observed score as being above the cut score, are considered to be false positives. Students who were above the Proficiency Level cut score (based on their “true score”), but were classified as being below a cut score based on the observed 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 Proficiency Level divided by all students classified into that Level according either to the true score distribution (accuracy) or 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 (i.e., accuracy), we provide the percentage 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 percentage 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 or not students are being accurately classified into WIDA language Proficiency Level 5 (“Bridging”), 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 a 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 and total for the grade-level cluster:

- The number of students in the analyses (the number students who were not absent, invalid, refused, exempt, or in the wrong cluster),
- the minimum observed scale score,
- the maximum observed scale score,
- the mean (average) scale score, and
- 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 Language Proficiency Levels. The vertical axis shows the percentage of students. Each bar shows the percentage of students who were placed into each Proficiency Level.

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

- The WIDA Language 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, and
- the percentage 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, based on performance 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 grade-level cluster. The leftmost column shows the cut (e.g., 1/2, which is the cut score between Proficiency Level 1 and Proficiency 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 SEM at the 2/3 cut point. At the other end of the continuum, Tier C forms should optimally have the lowest conditional SEM at the 5/6 cut point, and a relatively low SEM at the 4/5 cut point. Tier B should have low SEM in the middle range. Information from this table provides comparable information on how well the three Tier forms are targeted to provide the most accurate measure in order 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. It shows graphically how the tiers differ in difficulty. Tier 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 columns 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 grade-level 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 Section 7.1.1). The first column shows the components forming the composite, the second column shows the weight of the composite in the total score, the third shows the variance of the scale scores, and the fourth shows 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 to four measures being combined into one single measure. Thus, the reliability of the composite score will be higher than the reliability of any single sub-score 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 the WIDA Language Proficiency Levels (see Section 7.1.2). 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 grade-level 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 per Proficiency 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., determining which students have reached Proficiency Level 6). Note that the consistency is generally higher at the cut points than over the Proficiency Levels. For practical purposes, the primary score used for such decisions are the Overall Composite scores.

Note that because of the scoring caps imposed on Tier A and Tier B in Listening and Reading, in several cases, only a very small percentage of examinees get placed into Proficiency 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 for Proficiency 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

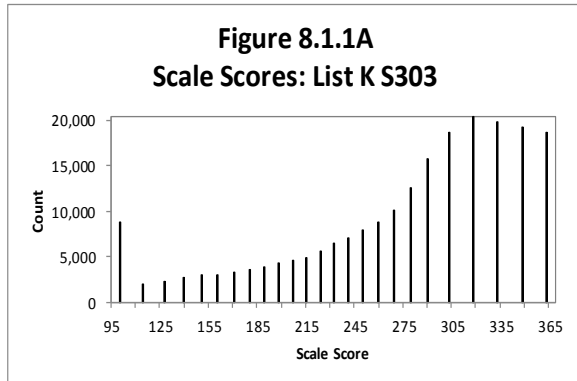


Table 8.1.1A
Scale Score Descriptive Statistics: List K S303

No. of Students	Min.	Max.	Mean	Std. Dev.
218,288	100	363	272.20	70.60

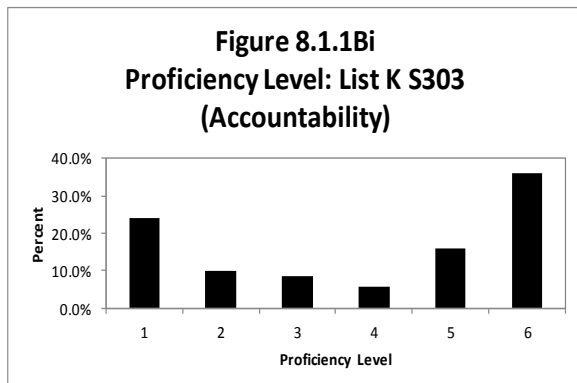


Table 8.1.1Bi
Proficiency Level Distribution: List K S303
(Accountability)

Level	Count	Percent
1	52,336	24.0%
2	21,616	9.9%
3	18,979	8.7%
4	12,635	5.8%
5	34,465	15.8%
6	78,257	35.9%
Total	218,288	100.0%

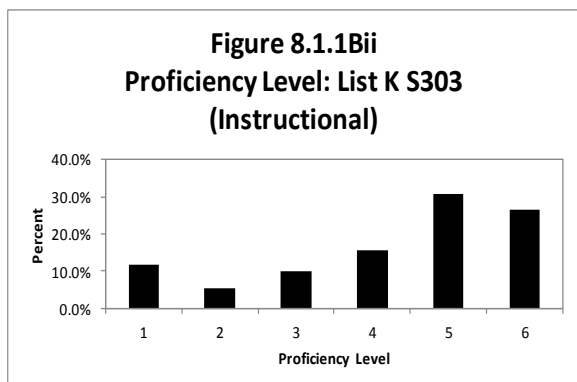


Table 8.1.1Bii
Proficiency Level Distribution: List K S303
(Instructional)

Level	Count	Percent
K1	25,329	11.6%
K2	11,903	5.5%
K3	21,613	9.9%
K4	34,086	15.6%
K5	67,508	30.9%
K6	57,849	26.5%
Total	218,288	100.0%

Table 8.1.1 Ci

Conditional Standard Error of
Measurement at Cut Scores: List K
S303 (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 8.1.1 Cii

Conditional Standard Error of
Measurement at Cut Scores: List K
S303 (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 8.1.1C
Test Characteristic Curve: List K S303

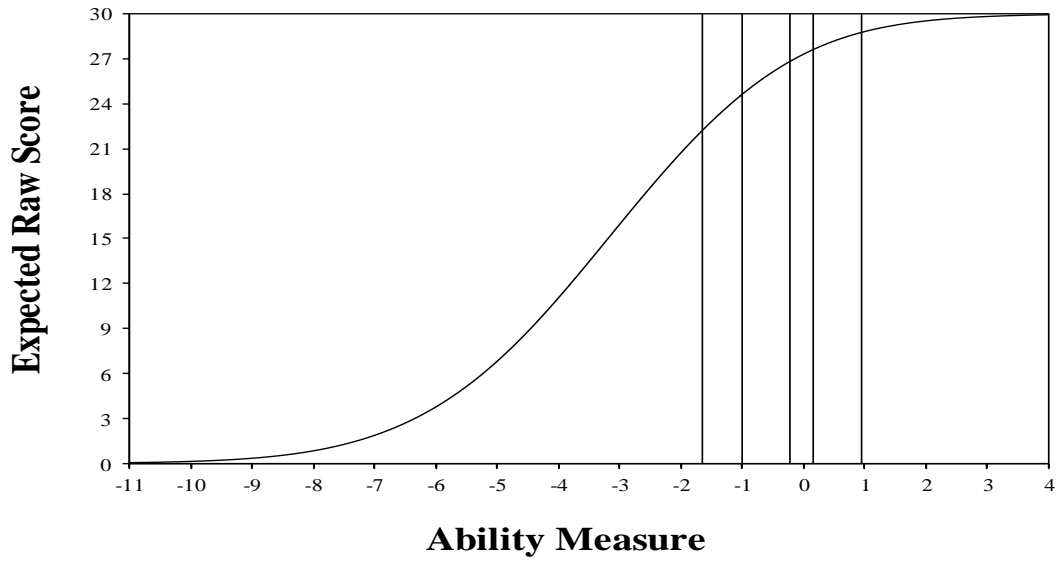


Figure 8.1.1D
Test Information Function: List K S303

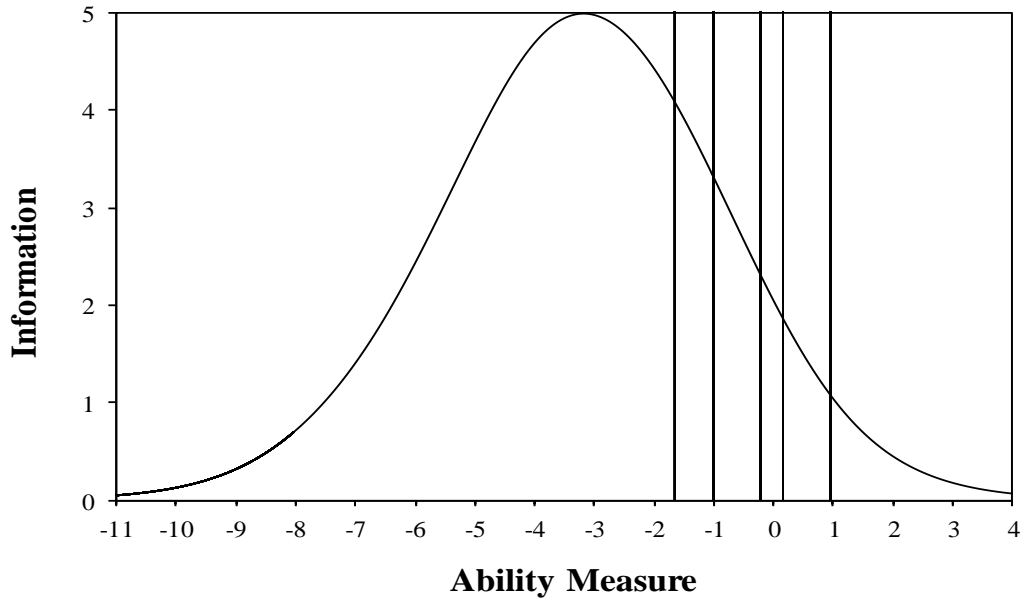


Table 8.1.1D

Reliability: List K S303

Tiers	No. of Students	Reliability
-	218,288	0.934

Table 8.1.1E-1Accuracy and Consistency of Classification Indices: List (Grade K) S303
(Accountability)

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.680	0.616	0.493		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.869	0.812		
	2	0.461	0.343		
	3	0.325	0.246		
	4	0.210	0.154		
	5	0.470	0.359		
	6	0.825	0.774		
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.942	0.032	0.025	0.919
	2/3	0.930	0.027	0.043	0.902
	3/4	0.917	0.048	0.035	0.884
	4/5	0.907	0.044	0.049	0.875
	5/6	0.897	0.035	0.068	0.857

Table 8.1.1E-2Accuracy and Consistency of Classification Indices: List (Grade K) S303
(Instructional)

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.676	0.574	0.458		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.887	0.811		
	2	0.429	0.313		
	3	0.526	0.400		
	4	0.565	0.445		
	5	0.698	0.573		
	6	0.732	0.657		
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.970	0.013	0.018	0.957
	2/3	0.959	0.020	0.021	0.941
	3/4	0.941	0.029	0.030	0.916
	4/5	0.917	0.041	0.042	0.885
	5/6	0.877	0.040	0.082	0.830

8.1.2 Reading K

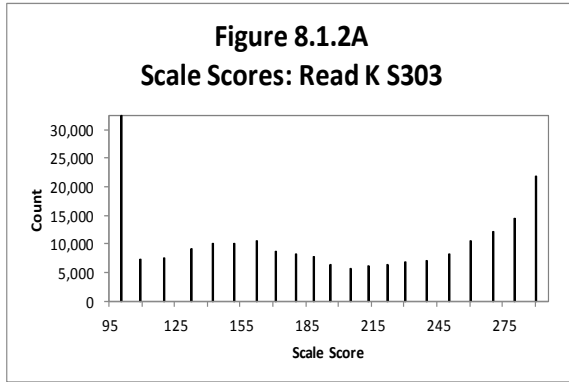


Table 8.1.2A
Scale Score Descriptive Statistics: Read K S303

No. of Students	Min.	Max.	Mean	Std. Dev.
218,268	100	290	193.61	66.46

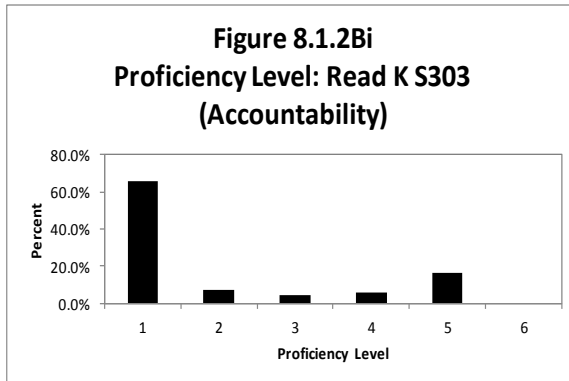


Table 8.1.2Bi
Proficiency Level Distribution: Read K S303
(Accountability)

Level	Count	Percent
1	143,727	65.8%
2	15,381	7.0%
3	10,489	4.8%
4	12,155	5.6%
5	36,516	16.7%
6	0	0.0%
Total	218,268	100.0%

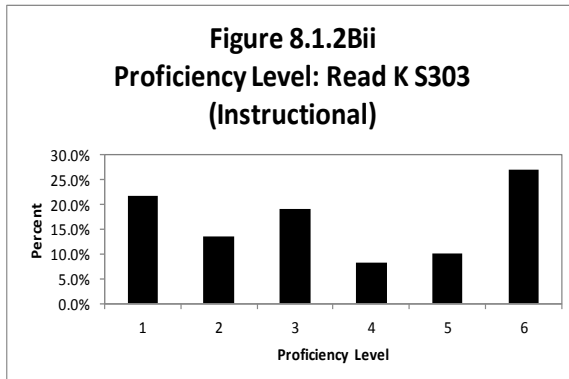


Table 8.1.2Bii
Proficiency Level Distribution: Read K S303
(Instructional)

Level	Count	Percent
K1	47,510	21.8%
K2	29,483	13.5%
K3	41,583	19.1%
K4	18,226	8.4%
K5	22,306	10.2%
K6	59,160	27.1%
Total	218,268	100.0%

Table 8.1.2Ci

Conditional Standard Error of Measurement
at Cut Scores: Read K S303 (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 S303 (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 S303

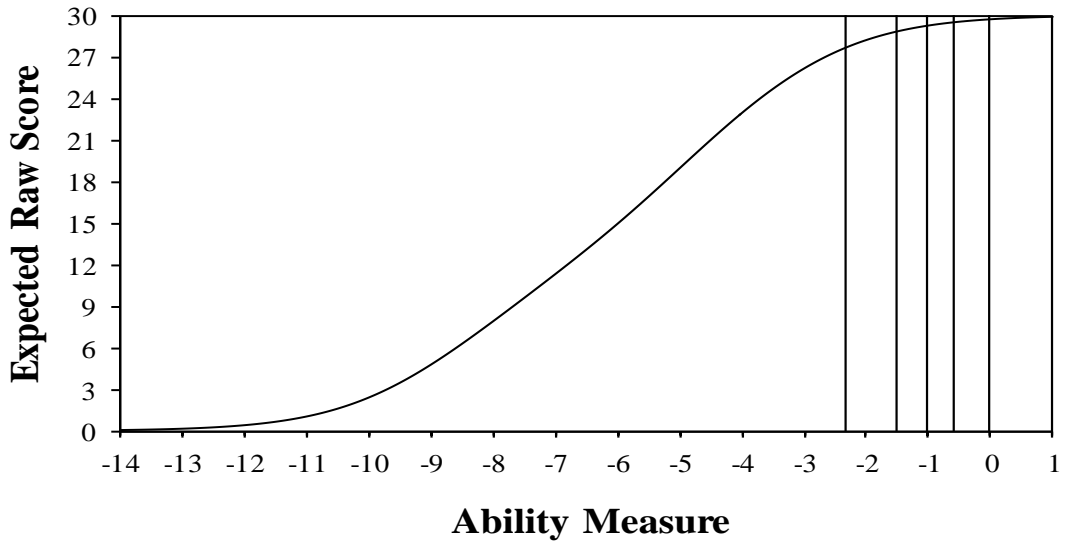


Figure 8.1.2D
 Test Information Function: Read K S303

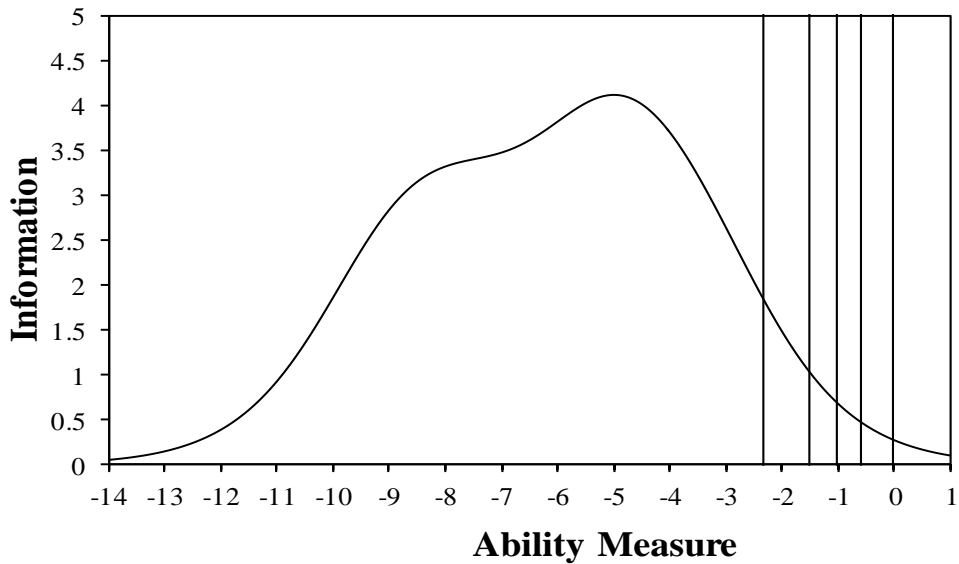


Table 8.1.2D

Reliability: Read K S303

Tiers	No. of Students	Reliability
-	218,268	0.948

Table 8.1.2E-1Accuracy and Consistency of Classification Indices: Read (Grade K) S303
(Accountability)

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.816	0.781	0.582		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.942	0.927		
	2	0.337	0.252		
	3	0.243	0.179		
	4	0.294	0.213		
	5	0.867	0.768		
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.934	0.039	0.027	0.910
	2/3	0.936	0.032	0.032	0.912
	3/4	0.942	0.029	0.029	0.916
	4/5	0.946	0.034	0.020	0.922

Table 8.1.2E-2Accuracy and Consistency of Classification Indices: Read (Grade K) S303
(Instructional)

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.775	0.704	0.608		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.903	0.837		
	2	0.575	0.460		
	3	0.698	0.587		
	4	0.390	0.291		
	5	0.925	0.885		
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.945	0.020	0.035	0.924
	2/3	0.942	0.032	0.026	0.917
	3/4	0.937	0.029	0.034	0.912
	4/5	0.940	0.032	0.028	0.915

8.1.3 Writing K

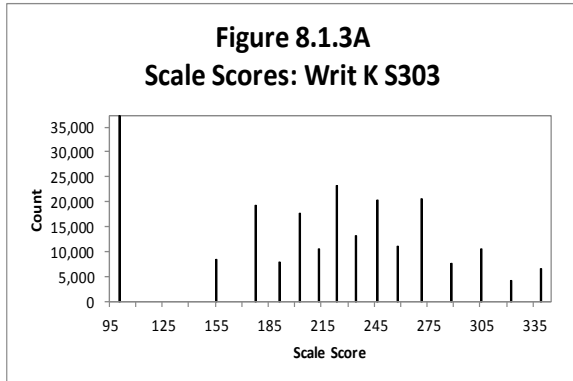


Table 8.1.3A
Scale Score Descriptive Statistics: Writ K S303

No. of Students	Min.	Max.	Mean	Std. Dev.
218,272	100	339	212.61	65.68

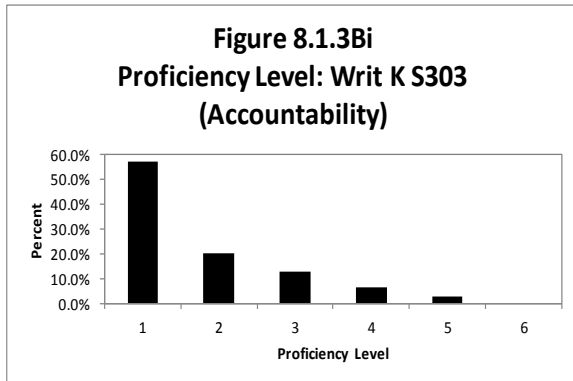


Table 8.1.3Bi
Proficiency Level Distribution: Writ K S303
(Accountability)

Level	Count	Percent
1	124,445	57.0%
2	44,299	20.3%
3	28,448	13.0%
4	14,549	6.7%
5	6,531	3.0%
6	0	0.0%
Total	218,272	100.0%

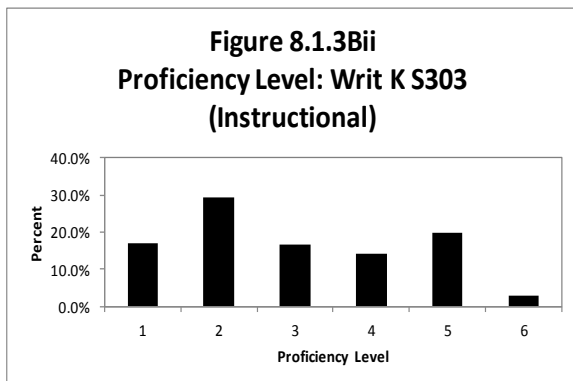


Table 8.1.3Bii
Proficiency Level Distribution: Writ K S303
(Instructional)

Level	Count	Percent
K1	37,235	17.1%
K2	63,948	29.3%
K3	36,315	16.6%
K4	31,246	14.3%
K5	42,997	19.7%
K6	6,531	3.0%
Total	218,272	100.0%

Table 8.1.3Ci

Conditional Standard Error of
Measurement at Cut Scores: Writ K
S303 (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
S303 (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 S303

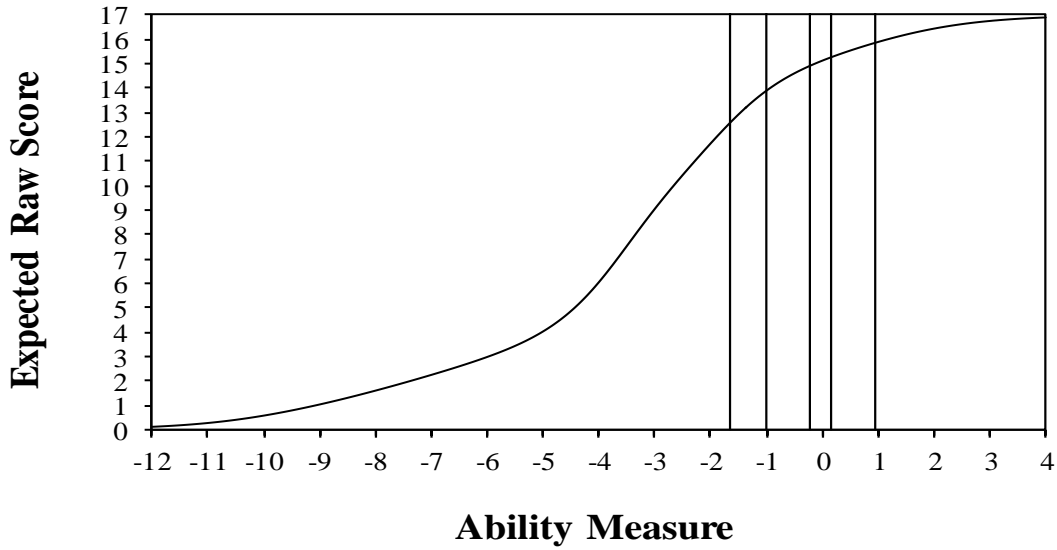


Figure 8.1.3D
 Test Information Function: Writ K S303

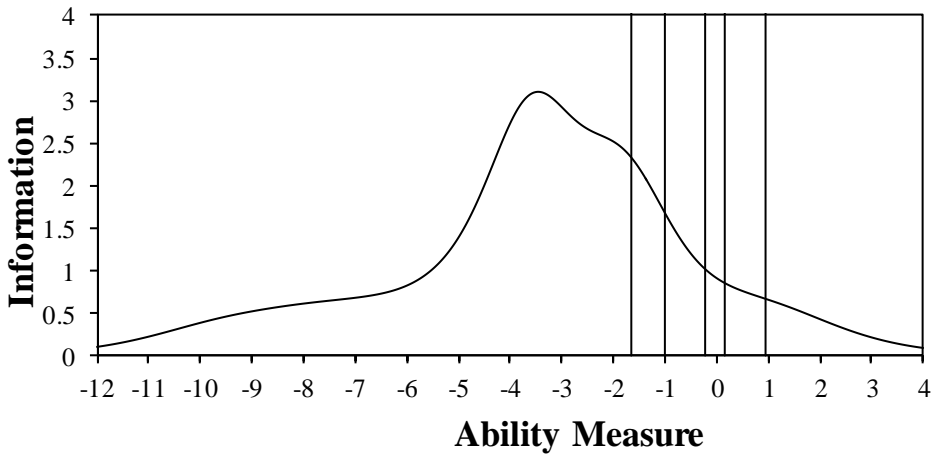


Table 8.1.3D

Reliability: Writ K S303

Tiers	No. of Students	Reliability
-	218,272	0.925

Table 8.1.3E-1

Accuracy and Consistency of Classification Indices: Writ (Grade K) S303

(Accountability)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.737	0.685		0.485	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.939		0.911	
	2	0.624		0.480	
	3	0.395		0.353	
	4	-		0.272	
	5	-		0.152	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.923	0.035	0.042	0.895
	2/3	0.904	0.021	0.074	0.866
	3/4	0.903	0.097	0.000	0.893
4/5	0.970	0.030	0.000	0.969	

Table 8.1.3E-2

Accuracy and Consistency of Classification Indices: Writ (Grade K) S303

(Instructional)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.688	0.596		0.486	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.864		0.791	
	2	0.790		0.706	
	3	0.522		0.397	
	4	0.380		0.287	
	5	0.233		0.632	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.957	0.024	0.020	0.937
	2/3	0.918	0.041	0.041	0.886
	3/4	0.901	0.031	0.067	0.864
4/5	0.884	0.042	0.074	0.836	

8.1.4 Speaking K

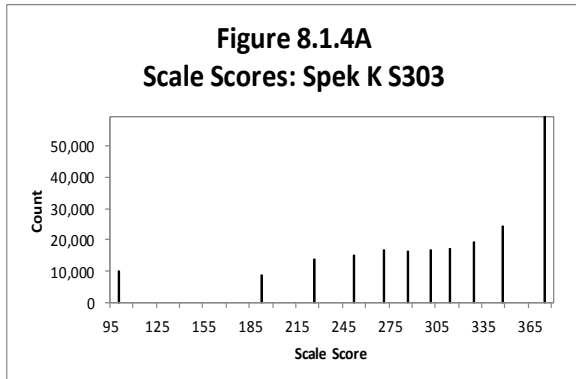


Table 8.1.4A

Scale Score Descriptive Statistics: Spek K S303

No. of Students	Min.	Max.	Mean	Std. Dev.
218,222	100	375	304.57	69.22

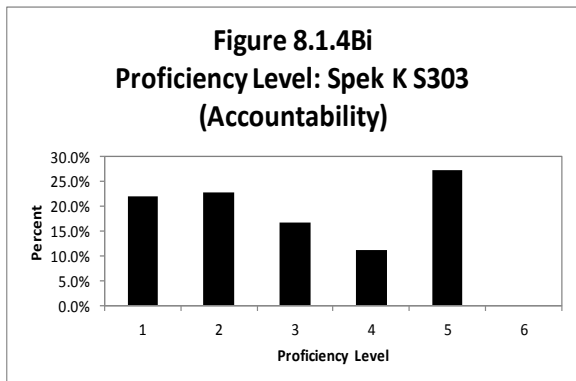


Table 8.1.4Bi

Proficiency Level Distribution: Spek K S303
(Accountability)

Level	Count	Percent
1	47,924	22.0%
2	49,961	22.9%
3	36,538	16.7%
4	24,409	11.2%
5	59,390	27.2%
6	0	0.0%
Total	218,222	100.0%

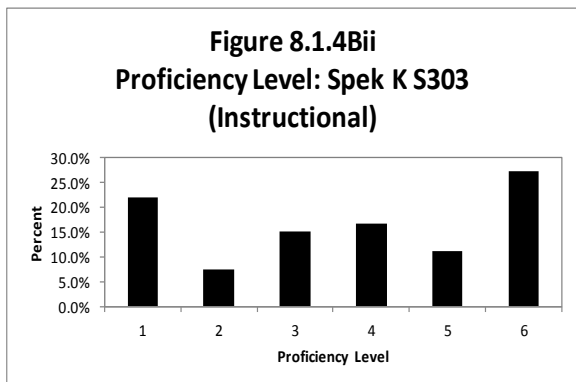


Table 8.1.4Bii

Proficiency Level Distribution: Spek K S303
(Instructional)

Level	Count	Percent
K1	47,924	22.0%
K2	16,688	7.6%
K3	33,273	15.2%
K4	36,538	16.7%
K5	24,409	11.2%
K6	59,390	27.2%
Total	218,222	100.0%

Table 8.1.4Ci

Conditional Standard Error of
Measurement at Cut Scores: Spek K
S303 (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
S303 (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 S303

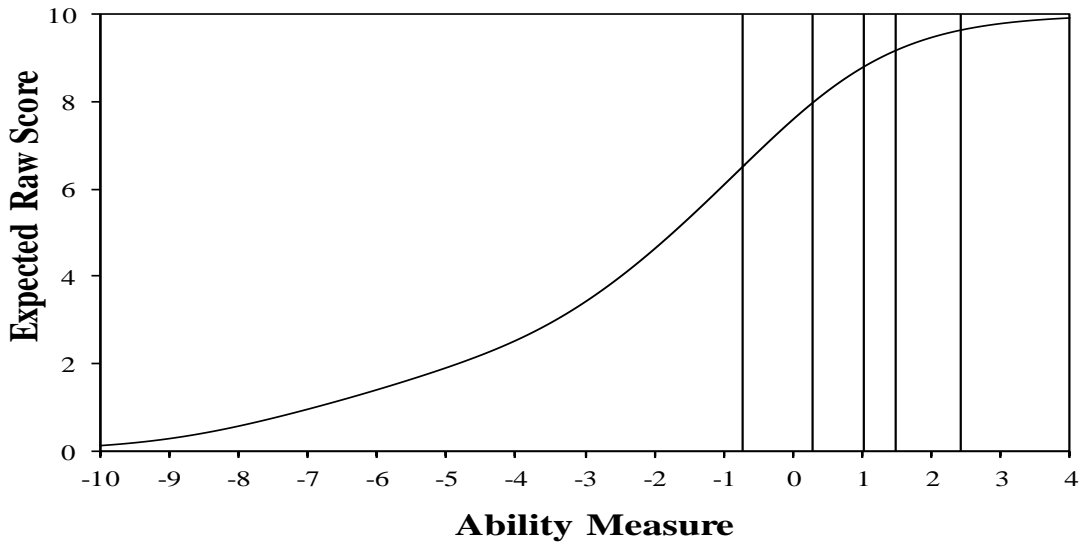


Figure 8.1.4D
 Test Information Function: Spek K S303

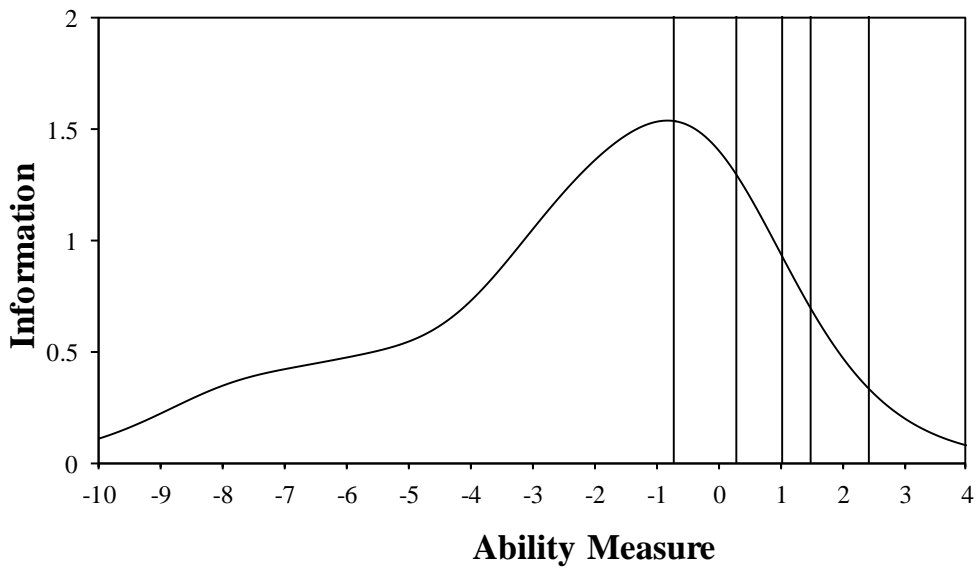


Table 8.1.4D

Reliability: Spek K S303

Tiers	No. of Students	Reliability
-	218,222	0.895

Table 8.1.4E-1Accuracy and Consistency of Classification Indices: Spek (Grade K) S303
(Accountability)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.459	0.447		0.318	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.825		0.753	
	2	0.667		0.538	
	3	0.381		0.261	
	4	0.208		0.191	
	5	-		0.576	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.930	0.040	0.030	0.902
	2/3	0.892	0.032	0.076	0.853
	3/4	0.866	0.044	0.090	0.787
4/5	0.728	0.272	0.000	0.747	

Table 8.1.4E-2Accuracy and Consistency of Classification Indices: Spek (Grade K) S303
(Instructional)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.656	0.567		0.421	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.868		0.792	
	2	0.310		0.232	
	3	0.480		0.364	
	4	0.362		0.264	
	5	0.800		0.729	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.940	0.029	0.031	0.913
	2/3	0.915	0.047	0.038	0.886
	3/4	0.889	0.031	0.080	0.852
4/5	0.860	0.059	0.081	0.790	

8.1.5 Oral Language Composite K

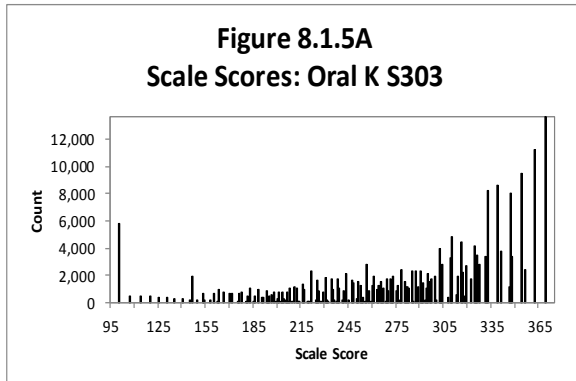


Table 8.1.5A
Scale Score Descriptive Statistics: Oral K S303

No. of Students	Min.	Max.	Mean	Std. Dev.
218,213	100	369	288.62	66.05

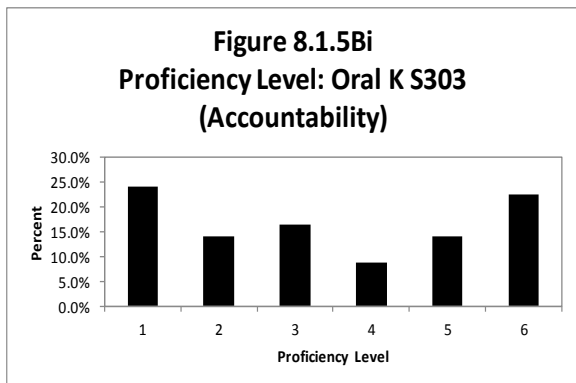


Table 8.1.5Bi
Proficiency Level Distribution: Oral K S303
(Accountability)

Level	Count	Percent
1	52,619	24.1%
2	30,904	14.2%
3	35,730	16.4%
4	19,219	8.8%
5	30,533	14.0%
6	49,208	22.6%
Total	218,213	100.0%

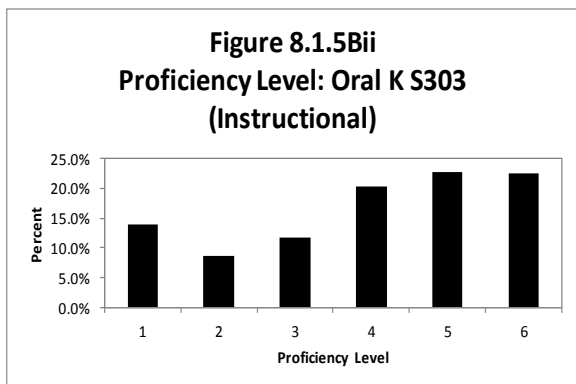


Table 8.1.5Bii
Proficiency Level Distribution: Oral K S303
(Instructional)

Level	Count	Percent
K1	30,390	13.9%
K2	18,938	8.7%
K3	25,671	11.8%
K4	44,254	20.3%
K5	49,752	22.8%
K6	49,208	22.6%
Total	218,213	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 S303

Component	Weight	Variance	Reliability
Listening	0.50	4982.084	0.934
Speaking	0.50	4790.384	0.895
Oral		4362.070	0.952

*Variances from students who had results in all four domains

Table 8.1.5E-1

Accuracy and Consistency of Classification Indices: Oral (Grade K) S303

(Accountability)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.613	0.539		0.439	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.914		0.870	
	2	0.660		0.541	
	3	0.619		0.491	
	4	0.322		0.209	
	5	0.333		0.287	
	6	0.706		0.609	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.960	0.021	0.019	0.943
	2/3	0.939	0.027	0.034	0.916
	3/4	0.926	0.023	0.051	0.897
	4/5	0.917	0.027	0.057	0.869
	5/6	0.836	0.120	0.044	0.810

Table 8.1.5E-2Accuracy and Consistency of Classification Indices: Oral (Grade K) S303
(Instructional)

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.655	0.557		0.455
Conditional on Level	Level	Accuracy		Consistency	
	1	0.894		0.837	
	2	0.606		0.477	
	3	0.590		0.470	
	4	0.704		0.579	
	5	0.517		0.429	
	6	0.687		0.599	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.972	0.015	0.012	0.960
	2/3	0.956	0.021	0.023	0.938
	3/4	0.943	0.025	0.032	0.921
	4/5	0.927	0.022	0.051	0.897
	5/6	0.850	0.086	0.063	0.811

8.1.6 Literacy Composite K

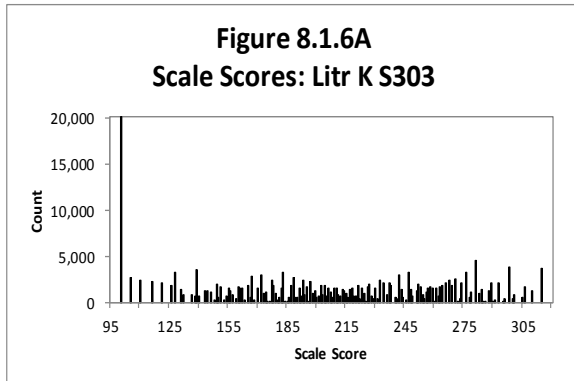


Table 8.1.6A

Scale Score Descriptive Statistics: Litr K S303

No. of Students	Min.	Max.	Mean	Std. Dev.
218,252	100	315	203.36	61.39

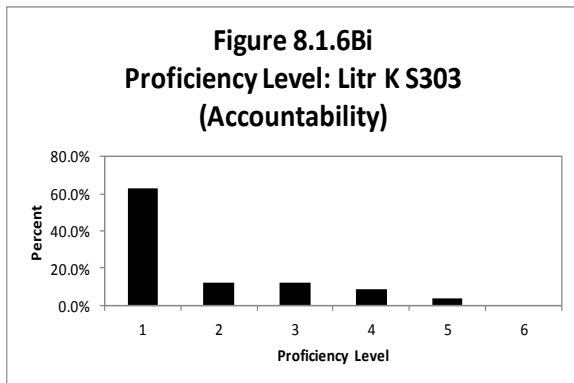


Table 8.1.6Bi

Proficiency Level Distribution: Litr K S303
(Accountability)

Level	Count	Percent
1	137,028	62.8%
2	26,573	12.2%
3	26,943	12.3%
4	19,202	8.8%
5	8,506	3.9%
6	0	0.0%
Total	218,252	100.0%

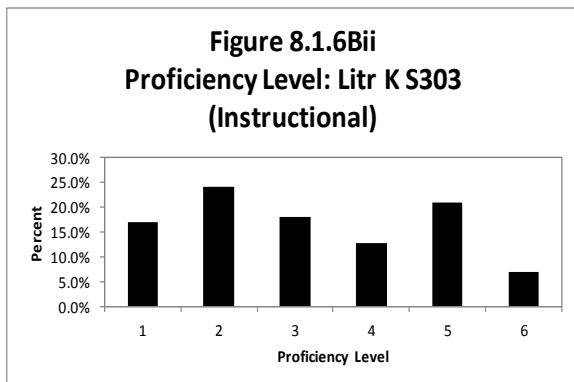


Table 8.1.6Bii

Proficiency Level Distribution: Litr K S303
(Instructional)

Level	Count	Percent
K1	37,193	17.0%
K2	52,667	24.1%
K3	39,268	18.0%
K4	28,012	12.8%
K5	45,808	21.0%
K6	15,304	7.0%
Total	218,252	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 S303

Component	Weight	Variance	Reliability
Reading	0.50	4416.660	0.948
Writing	0.50	4311.938	0.925
Literacy		3768.172	0.963

*Variances from students who had results in all four domains

Table 8.1.6E-1

Accuracy and Consistency of Classification Indices: Litr (Grade K) S303

(Accountability)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.790	0.739		0.540	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.961		0.942	
	2	0.569		0.439	
	3	0.506		0.393	
	4	0.454		0.394	
	5	-		0.263	
	6	-		0.000	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.947	0.024	0.028	0.926
	2/3	0.941	0.024	0.035	0.917
	3/4	0.929	0.035	0.036	0.901
	4/5	0.961	0.039	0.000	0.951
	5/6	1.000	0.000	0.000	1.000

Table 8.1.6E-2

Accuracy and Consistency of Classification Indices: Litr (Grade K) S303

(Instructional)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.747	0.673		0.596	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.919		0.874	
	2	0.843		0.778	
	3	0.732		0.626	
	4	0.588		0.462	
	5	0.645		0.616	
	6	-		0.412	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.971	0.014	0.015	0.959
	2/3	0.955	0.023	0.023	0.936
	3/4	0.945	0.024	0.031	0.924
	4/5	0.945	0.022	0.033	0.921
5/6	0.930	0.070	0.000	0.921	

8.1.7 Comprehension Composite K

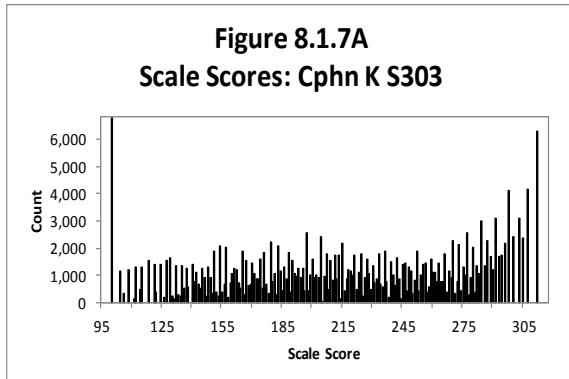


Table 8.1.7A
Scale Score Descriptive Statistics: Cphn K S303

No. of Students	Min.	Max.	Mean	Std. Dev.
218,256	100	312	217.18	60.56

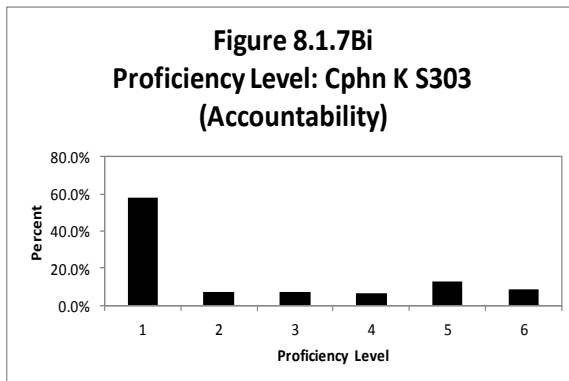


Table 8.1.7Bi
Proficiency Level Distribution: Cphn K S303
(Accountability)

Level	Count	Percent
1	125,923	57.7%
2	15,211	7.0%
3	15,453	7.1%
4	15,004	6.9%
5	28,336	13.0%
6	18,329	8.4%
Total	218,256	100.0%

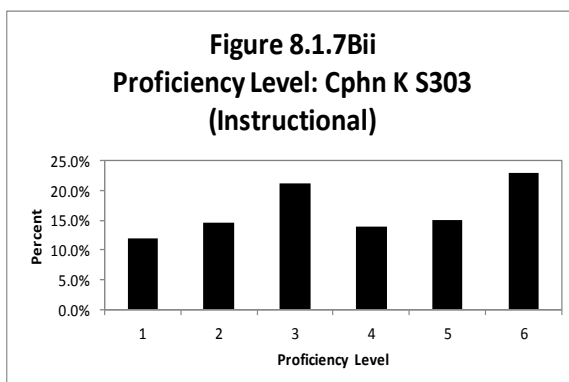


Table 8.1.7Bii
Proficiency Level Distribution: Cphn K S303
(Instructional)

Level	Count	Percent
K1	26,098	12.0%
K2	31,765	14.6%
K3	46,435	21.3%
K4	30,647	14.0%
K5	33,062	15.1%
K6	50,249	23.0%
Total	218,256	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 S303

Component	Weight	Variance	Reliability
Listening	0.30	4982.084	0.934
Reading	0.70	4416.660	0.948
Comprehension		3666.938	0.961

*Variances from students who had results in all four domains

Table 8.1.7E-1Accuracy and Consistency of Classification Indices: Cphn (Grade K) S303
(Accountability)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.750	0.695		0.517	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.962		0.943	
	2	0.395		0.288	
	3	0.382		0.275	
	4	0.339		0.246	
	5	0.533		0.447	
	6	0.636		0.495	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.950	0.022	0.029	0.929
	2/3	0.950	0.024	0.026	0.927
	3/4	0.943	0.030	0.027	0.919
	4/5	0.936	0.033	0.032	0.909
5/6	0.936	0.038	0.026	0.915	

Table 8.1.7E-2

Accuracy and Consistency of Classification Indices: Cphn (Grade K) S303
(Instructional)

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.770	0.682	0.614		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.900	0.841		
	2	0.770	0.675		
	3	0.788	0.701		
	4	0.637	0.516		
	5	0.624	0.499		
	6	0.875	0.824		
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.974	0.012	0.014	0.964
	2/3	0.959	0.020	0.021	0.942
	3/4	0.948	0.023	0.029	0.927
	4/5	0.949	0.023	0.028	0.927
5/6	0.939	0.033	0.028	0.913	

8.1.8 Overall Composite K

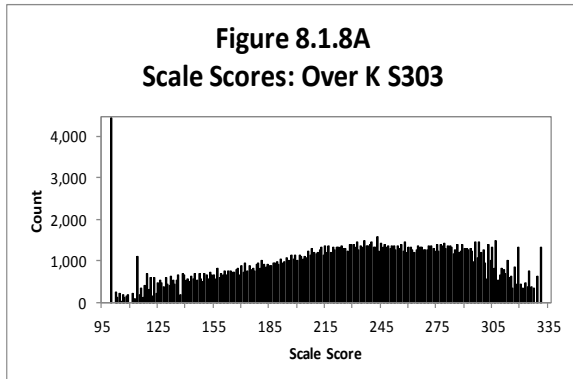


Table 8.1.8A
Scale Score Descriptive Statistics: Over K S303

No. of Students	Min.	Max.	Mean	Std. Dev.
218,174	100	331	228.75	57.23

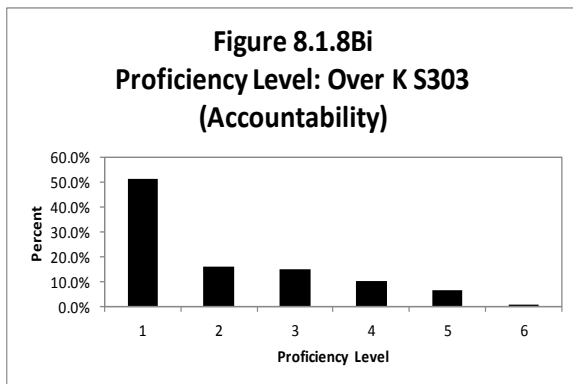


Table 8.1.8Bi
Proficiency Level Distribution: Over K S303
(Accountability)

Level	Count	Percent
1	112,412	51.5%
2	34,939	16.0%
3	32,749	15.0%
4	22,475	10.3%
5	13,637	6.3%
6	1,962	0.9%
Total	218,174	100.0%

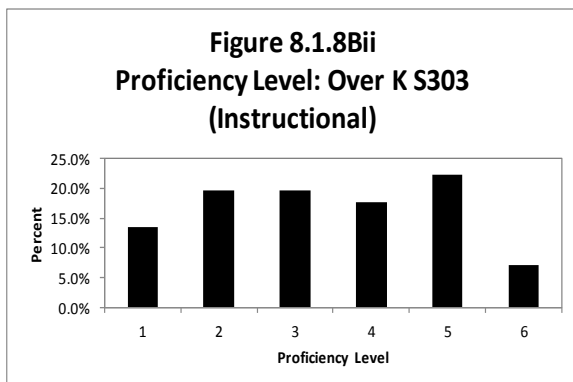


Table 8.1.8Bii
Proficiency Level Distribution: Over K S303
(Instructional)

Level	Count	Percent
K1	29,509	13.5%
K2	42,742	19.6%
K3	42,889	19.7%
K4	38,718	17.7%
K5	48,717	22.3%
K6	15,599	7.1%
Total	218,174	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 S303

Component	Weight	Variance	Reliability
Listening	0.15	4982.084	0.934
Reading	0.35	4416.660	0.948
Speaking	0.15	4790.384	0.895
Writing	0.35	4311.938	0.925
Overall Composite		3275.445	0.974

*Variances from students who had results in all four domains

Table 8.1.8E-1Accuracy and Consistency of Classification Indices: Over (Grade K) S303
(Accountability)

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.802	0.741		0.614	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.956		0.936	
	2	0.711		0.601	
	3	0.681		0.561	
	4	0.527		0.442	
	5	0.580		0.471	
	6	-		0.126	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.954	0.023	0.023	0.935
	2/3	0.951	0.022	0.027	0.931
	3/4	0.953	0.022	0.025	0.932
	4/5	0.949	0.039	0.012	0.935
5/6	0.991	0.009	0.000	0.990	

Table 8.1.8E-2

Accuracy and Consistency of Classification Indices: Over (Grade K) S303
(Instructional)

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.788	0.713		0.648
Conditional on Level	Level	Accuracy		Consistency	
	1	0.925		0.884	
	2	0.840		0.775	
	3	0.784		0.695	
	4	0.730		0.625	
	5	0.736		0.682	
	6	0.697		0.541	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.978	0.010	0.012	0.969
	2/3	0.961	0.020	0.019	0.945
	3/4	0.951	0.022	0.027	0.931
	4/5	0.952	0.021	0.027	0.933
	5/6	0.946	0.041	0.013	0.933

8.2 Grades: 1–2

8.2.1 Listening 1–2

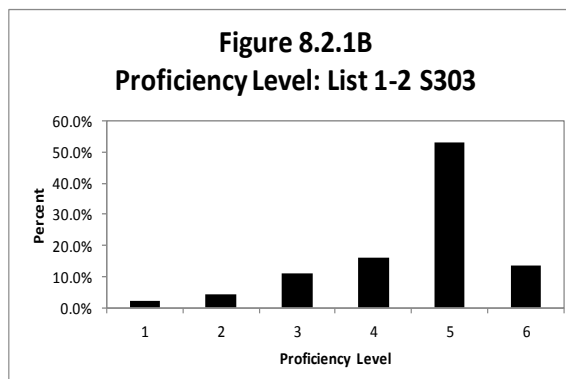
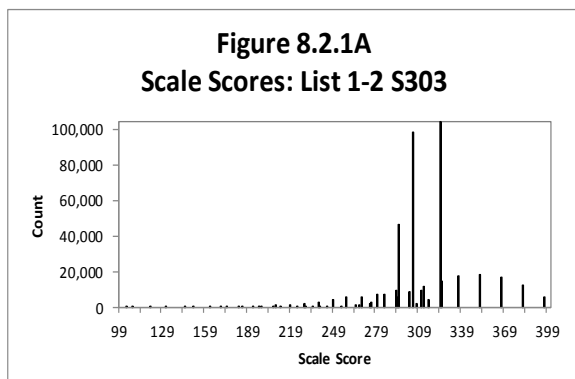


Table 8.2.1A
Scale Score Descriptive Statistics: List 1-2 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	222,599	104	397	299.04	27.70
2	216,010	108	397	326.47	29.54
Total	438,609	104	397	312.55	31.74

Table 8.2.1B
Proficiency Level Distribution: List 1-2 S303

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
1	5,971	2.7%	3,687	1.7%	9,658	2.2%
2	11,433	5.1%	7,263	3.4%	18,696	4.3%
3	31,666	14.2%	17,402	8.1%	49,068	11.2%
4	49,530	22.3%	20,366	9.4%	69,896	15.9%
5	106,594	47.9%	125,615	58.2%	232,209	52.9%
6	17,405	7.8%	41,677	19.3%	59,082	13.5%
Total	222,599	100.0%	216,010	100.0%	438,609	100.0%

Table 8.2.1C

Conditional Standard Error of Measurement at Cut Scores: List 1-2 S303*

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	1	238	19.16	19.91	19.16
	2	247	19.16	19.54	18.79
2/3	1	267	19.91	19.16	18.03
	2	281	21.04	19.16	18.41
3/4	1	295	22.92	19.54	18.79
	2	311	25.55	20.29	19.91
4/5	1	305	n/a	19.91	19.16
	2	324	n/a	21.42	21.04
5/6	1	330	n/a	n/a	21.79
	2	350	n/a	n/a	24.80

* No equating was performed for S303

Figure 8.2.1C
 Test Characteristic Curve: List 1-2ABC S303

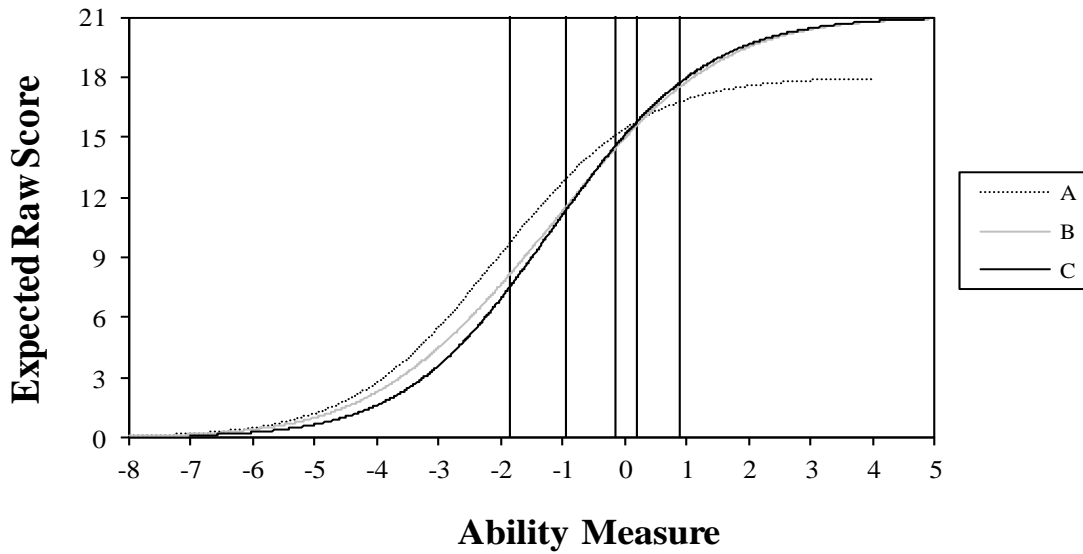


Figure 8.2.1D
 Test Information Function: List 1-2ABC S303

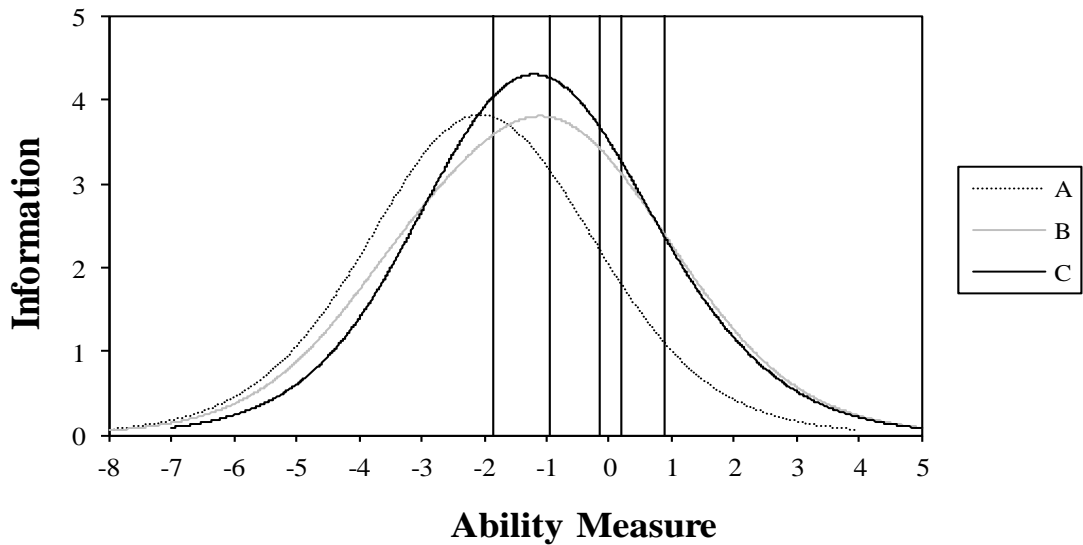


Table 8.2.1D

Weighted Reliability: List 1-2 S303

Tiers	No. of Students	Reliability	Reliability
A	94,425	0.793	0.691
B	221,704	0.672	
C	122,480	0.646	

Table 8.2.1E-1

Accuracy and Consistency of Classification Indices: List (Grade 1) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.544	0.398		0.155	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.819		0.643	
	2	0.537		0.346	
	3	0.415		0.252	
	4	0.337		0.255	
	5	0.618		0.580	
	6	-		0.127	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.986	0.004	0.010	0.980
	2/3	0.963	0.012	0.025	0.941
	3/4	0.860	0.071	0.069	0.771
	4/5	0.739	0.097	0.165	0.651
5/6	0.922	0.078	0.000	0.847	

Table 8.2.1E-2

Accuracy and Consistency of Classification Indices: List (Grade 2) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.572	0.430		0.151	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.834		0.634	
	2	0.509		0.337	
	3	0.408		0.237	
	4	0.214		0.138	
	5	0.670		0.647	
	6	0.433		0.303	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.990	0.002	0.009	0.987
	2/3	0.975	0.007	0.018	0.961
	3/4	0.922	0.040	0.038	0.865
	4/5	0.840	0.092	0.068	0.758
5/6	0.797	0.161	0.042	0.713	

8.2.2 Reading 1–2

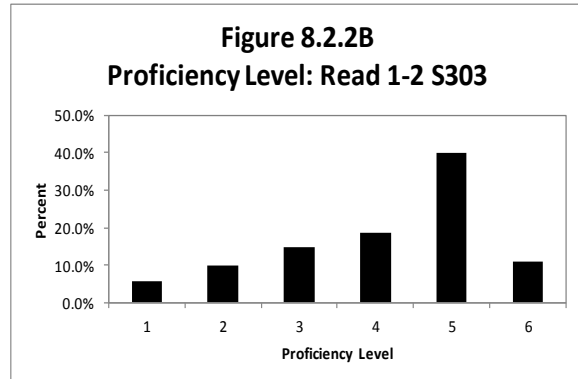
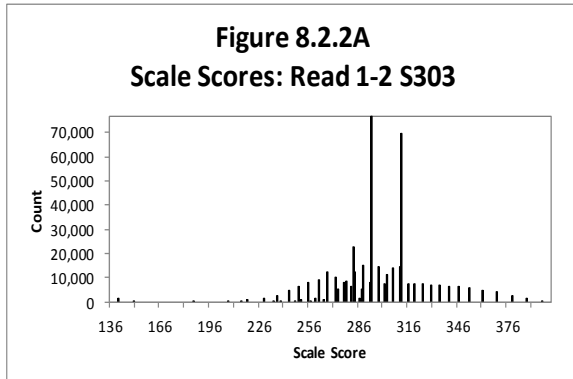


Table 8.2.2A

Scale Score Descriptive Statistics: Read 1-2 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	222,466	141	397	284.55	24.16
2	215,946	150	397	310.92	26.10
Total	438,412	141	397	297.54	28.38

Table 8.2.2B

Proficiency Level Distribution: Read 1-2 S303

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
1	17,023	7.7%	7,627	3.5%	24,650	5.6%
2	26,496	11.9%	17,523	8.1%	44,019	10.0%
3	32,070	14.4%	32,688	15.1%	64,758	14.8%
4	47,831	21.5%	34,333	15.9%	82,164	18.7%
5	84,379	37.9%	90,889	42.1%	175,268	40.0%
6	14,667	6.6%	32,886	15.2%	47,553	10.8%
Total	222,466	100.0%	215,946	100.0%	438,412	100.0%

Table 8.2.2C

Conditional Standard Error of Measurement at Cut Scores: Read 1-2 S303

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	1	253	12.48	15.86	15.60
	2	267	11.70	13.26	13.52
2/3	1	269	11.44	13.00	13.26
	2	286	11.44	11.18	11.44
3/4	1	283	11.44	11.44	11.70
	2	303	12.22	10.40	10.92
4/5	1	294	n/a	10.92	11.18
	2	312	n/a	10.40	10.66
5/6	1	314	n/a	n/a	10.92
	2	331	n/a	n/a	11.44

Figure 8.2.2C
 Test Characteristic Curve: Read 1-2ABC S303

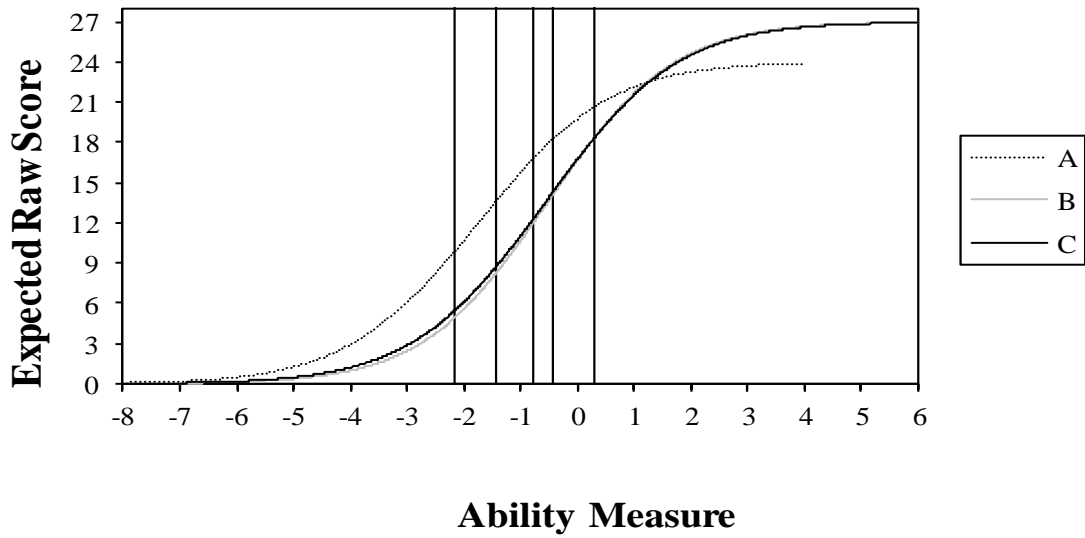


Figure 8.2.2D
 Test Information Function: Read 1-2ABC S303

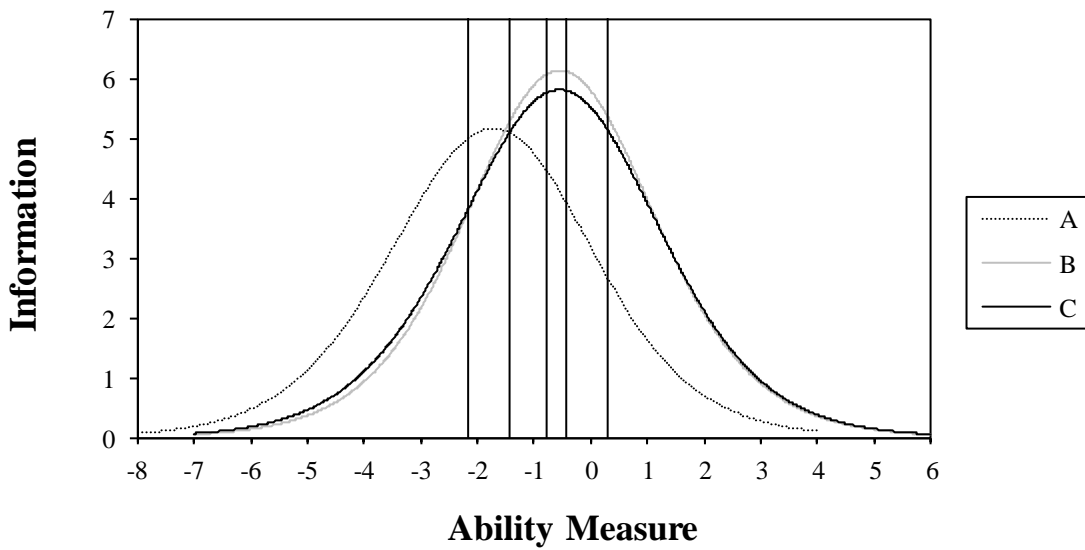


Table 8.2.2D

Weighted Reliability: Read 1-2 S303

Tiers	No. of Students	Reliability	Reliability
A	94,372	0.755	0.805
B	221,588	0.817	
C	122,452	0.822	

Table 8.2.2E-1

Accuracy and Consistency of Classification Indices: Read (Grade 1) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.553	0.428		0.257	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.780		0.637	
	2	0.567		0.413	
	3	0.396		0.280	
	4	0.397		0.298	
	5	0.634		0.575	
	6	-		0.169	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.964	0.016	0.019	0.947
	2/3	0.922	0.026	0.051	0.890
	3/4	0.875	0.056	0.070	0.815
	4/5	0.808	0.083	0.109	0.736
	5/6	0.934	0.066	0.000	0.888

Table 8.2.2E-2

Accuracy and Consistency of Classification Indices: Read (Grade 2) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.544	0.429		0.261	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.732		0.493	
	2	0.492		0.327	
	3	0.412		0.299	
	4	0.306		0.240	
	5	0.680		0.592	
	6	0.617		0.462	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.977	0.007	0.017	0.964
	2/3	0.935	0.027	0.038	0.900
	3/4	0.863	0.080	0.057	0.803
	4/5	0.814	0.108	0.078	0.756
	5/6	0.888	0.046	0.066	0.831

8.2.3 Writing 1–2

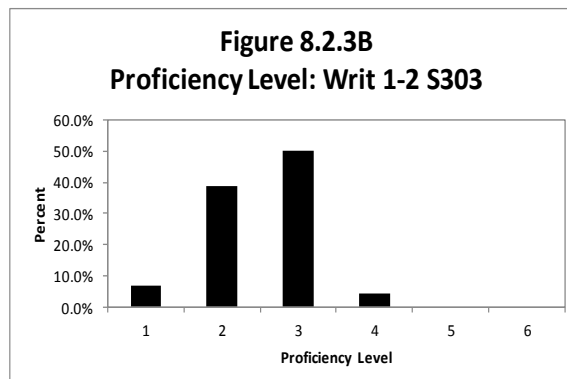
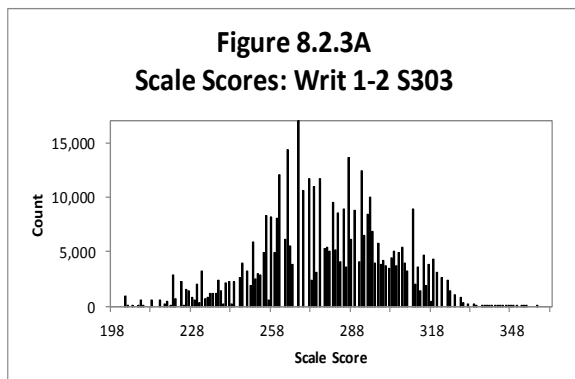


Table 8.2.3A

Scale Score Descriptive Statistics: Writ 1-2 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	222,512	203	348	268.88	21.41
2	215,941	209	358	288.35	22.25
Total	438,453	203	358	278.47	23.90

Table 8.2.3B

Proficiency Level Distribution: Writ 1-2 S303

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
1	18,806	8.5%	11,581	5.4%	30,387	6.9%
2	102,015	45.8%	68,059	31.5%	170,074	38.8%
3	94,203	42.3%	124,953	57.9%	219,156	50.0%
4	7,471	3.4%	11,332	5.2%	18,803	4.3%
5	17	0.0%	16	0.0%	33	0.0%
6	0	0.0%	0	0.0%	0	0.0%
Total	222,512	100.0%	215,941	100.0%	438,453	100.0%

Table 8.2.3C

Conditional Standard Error of Measurement at Cut Scores: Writ 1-2 S303*

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	7.46	7.46
2/3	1	272	7.77	8.09	8.40
	2	285	8.40	8.40	8.40
3/4	1	308	8.09	8.09	7.77
	2	320	7.77	7.46	7.46
4/5	1	336	7.46	7.15	6.84
	2	348	6.84	6.53	6.53
5/6	1	362	6.53	6.53	6.53
	2	373	6.53	7.15	7.46

* No equating was performed for Writing Tier A S303

Figure 8.2.3C
 Test Characteristic Curve: Writ 1-2ABC S303

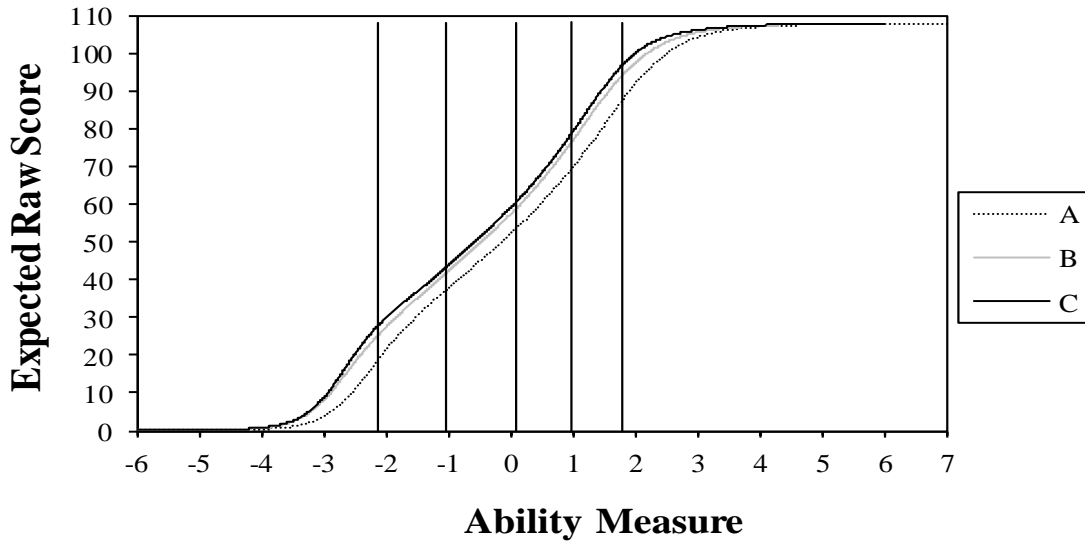


Figure 8.2.3D
 Test Information Function: Writ 1-2ABC S303

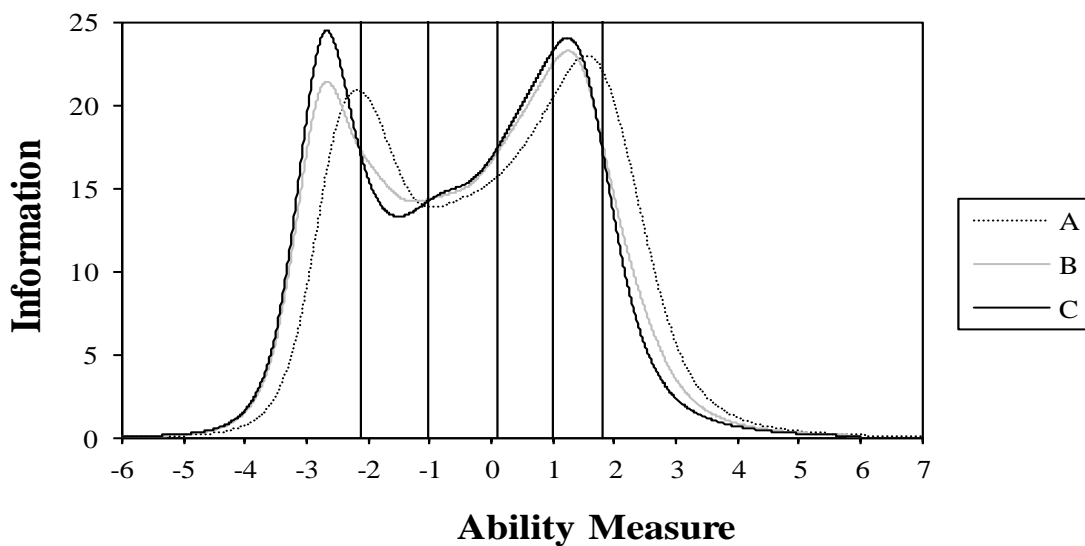


Table 8.2.3D

Weighted Reliability: Writ 1-2 S303

Tiers	No. of Students	Reliability	Reliability
A	94,411	0.899	0.924
B	221,583	0.925	
C	122,459	0.942	

Table 8.2.3E-1

Accuracy and Consistency of Classification Indices: Writ (Grade 1) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.860	0.806		0.676	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.843		0.734	
	2	0.873		0.827	
	3	0.852		0.814	
	4	0.773		0.533	
	5	-		1.000	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.969	0.012	0.019	0.956
	2/3	0.916	0.040	0.044	0.882
	3/4	0.974	0.022	0.003	0.968
	4/5	1.000	0.000	0.000	1.000

Table 8.2.3E-2

Accuracy and Consistency of Classification Indices: Writ (Grade 2) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.855	0.804		0.644	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.830		0.723	
	2	0.860		0.796	
	3	0.855		0.843	
	4	-		0.311	
	5	-		1.000	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.981	0.009	0.010	0.972
	2/3	0.927	0.033	0.040	0.897
	3/4	0.947	0.053	0.000	0.935
	4/5	1.000	0.000	0.000	1.000

8.2.4 Speaking 1–2

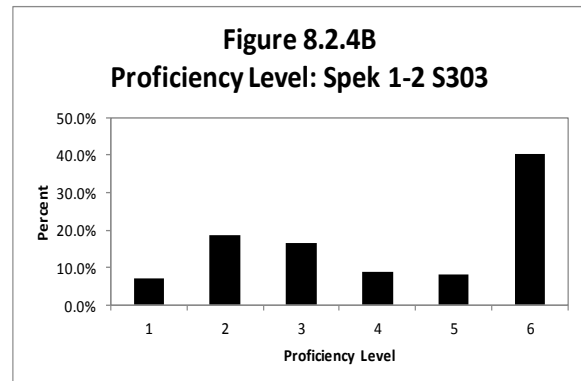
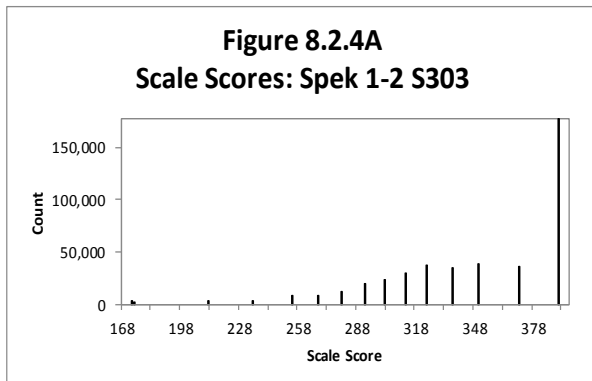


Table 8.2.4A
Scale Score Descriptive Statistics: Spek 1-2 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	222,506	173	391	338.90	48.18
2	215,888	174	391	357.64	43.10
Total	438,394	173	391	348.13	46.70

Table 8.2.4B
Proficiency Level Distribution: Spek 1-2 S303

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
1	18,097	8.1%	13,212	6.1%	31,309	7.1%
2	56,981	25.6%	25,440	11.8%	82,421	18.8%
3	40,851	18.4%	31,768	14.7%	72,619	16.6%
4	19,407	8.7%	19,262	8.9%	38,669	8.8%
5	16,950	7.6%	19,136	8.9%	36,086	8.2%
6	70,220	31.6%	107,070	49.6%	177,290	40.4%
Total	222,506	100.0%	215,888	100.0%	438,394	100.0%

Table 8.2.4CConditional Standard Error of Measurement at
Cut Scores: Spek 1-2 S303*

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

* No equating was performed for S303

Figure 8.2.4C
 Test Characteristic Curve: Spek 1-2 S303

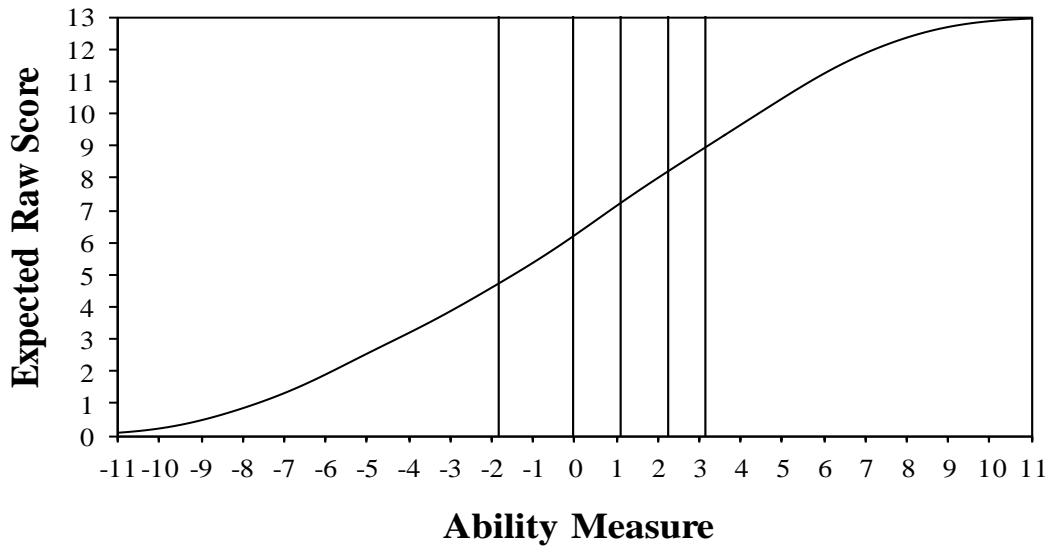


Figure 8.2.4D
 Test Information Function: Spek 1-2 S303

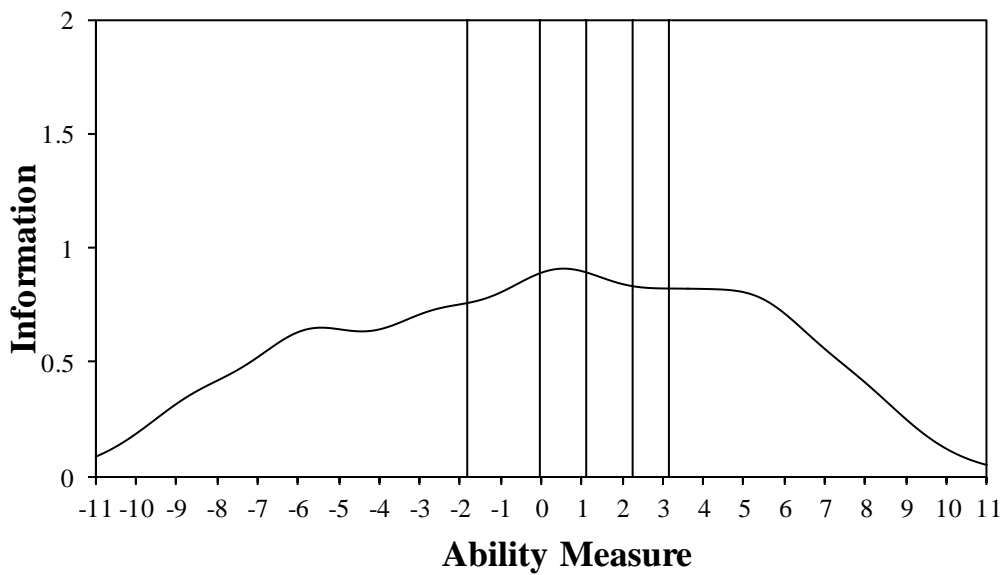


Table 8.2.4D

Reliability: Spek 1-2 S303

Tiers	No. of Students	Reliability
--	438,394	0.894

Table 8.2.4E-1

Accuracy and Consistency of Classification Indices: Spek (Grade 1) S303

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.614	0.525	0.415		
Conditional on Level	Level	Accuracy		Consistency	
	1	0.620		0.486	
	2	0.714		0.619	
	3	0.533		0.433	
	4	0.365		0.250	
	5	0.279		0.208	
	6	0.947		0.885	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.943	0.038	0.019	0.915
	2/3	0.891	0.045	0.063	0.857
	3/4	0.920	0.024	0.056	0.884
	4/5	0.949	0.028	0.024	0.914
5/6	0.887	0.100	0.012	0.869	

Table 8.2.4E-2

Accuracy and Consistency of Classification Indices: Spek (Grade 2) S303

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.663	0.596	0.446		
Conditional on Level	Level	Accuracy		Consistency	
	1	0.702		0.558	
	2	0.563		0.455	
	3	0.553		0.458	
	4	0.367		0.267	
	5	0.297		0.217	
	6	0.957		0.911	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.965	0.020	0.015	0.948
	2/3	0.926	0.037	0.037	0.902
	3/4	0.920	0.022	0.058	0.893
	4/5	0.945	0.025	0.030	0.910
5/6	0.881	0.102	0.018	0.856	

8.2.5 Oral Language Composite 1–2

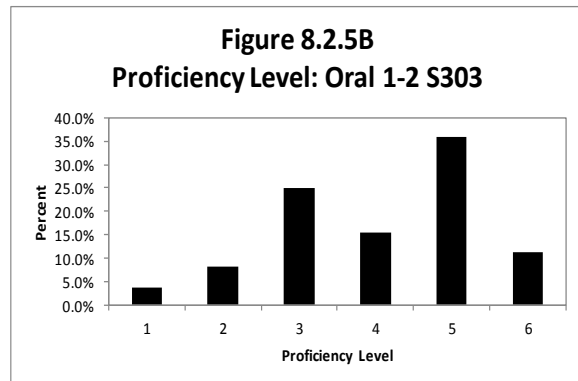
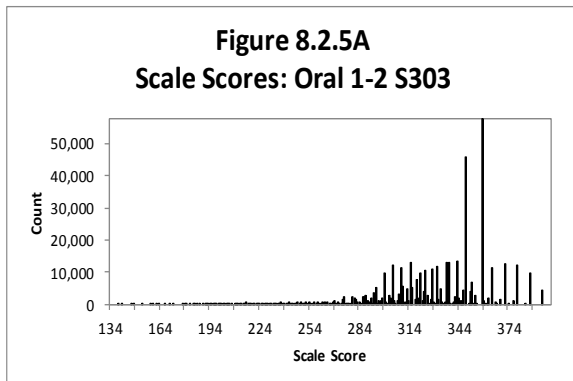


Table 8.2.5A

Scale Score Descriptive Statistics: Oral 1-2 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	222,379	139	394	319.19	33.08
2	215,802	141	394	342.35	31.54
Total	438,181	139	394	330.60	34.34

Table 8.2.5B

Proficiency Level Distribution: Oral 1-2 S303

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
1	10,040	4.5%	6,381	3.0%	16,421	3.7%
2	24,816	11.2%	10,540	4.9%	35,356	8.1%
3	69,523	31.3%	40,535	18.8%	110,058	25.1%
4	34,385	15.5%	33,878	15.7%	68,263	15.6%
5	67,138	30.2%	90,810	42.1%	157,948	36.0%
6	16,477	7.4%	33,658	15.6%	50,135	11.4%
Total	222,379	100.0%	215,802	100.0%	438,181	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 S303

Component	Weight	Variance	Reliability
Listening	0.50	1005.833	0.691
Speaking	0.50	2177.737	0.894
Oral		1178.240	0.885

*Variances from students who had results in all four domains

Table 8.2.5E-1

Accuracy and Consistency of Classification Indices: Oral (Grade 1) S303

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.632	0.515	0.377		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.837	0.717		
	2	0.640	0.502		
	3	0.766	0.656		
	4	0.374	0.276		
	5	0.652	0.589		
	6	-	0.279		
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.983	0.007	0.010	0.975
	2/3	0.939	0.033	0.028	0.912
	3/4	0.887	0.034	0.079	0.844
	4/5	0.882	0.061	0.057	0.826
5/6	0.926	0.074	0.000	0.897	

Table 8.2.5E-2

Accuracy and Consistency of Classification Indices: Oral (Grade 2) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.628	0.510		0.339
Conditional on Level	Level	Accuracy		Consistency	
	1	0.889		0.803	
	2	0.585		0.440	
	3	0.775		0.651	
	4	0.454		0.325	
	5	0.631		0.604	
	6	-		0.334	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.991	0.003	0.006	0.988
	2/3	0.970	0.018	0.012	0.954
	3/4	0.924	0.022	0.054	0.896
	4/5	0.891	0.043	0.066	0.837
	5/6	0.844	0.156	0.000	0.790

8.2.6 Literacy Composite 1–2

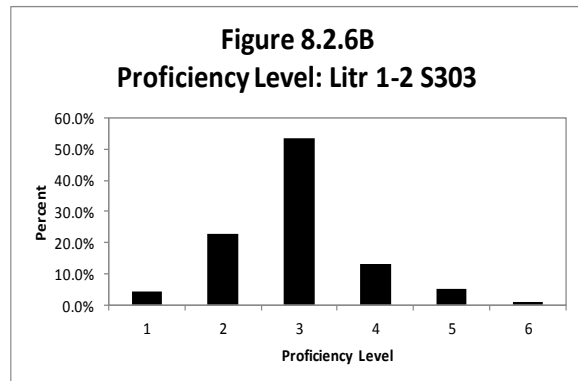
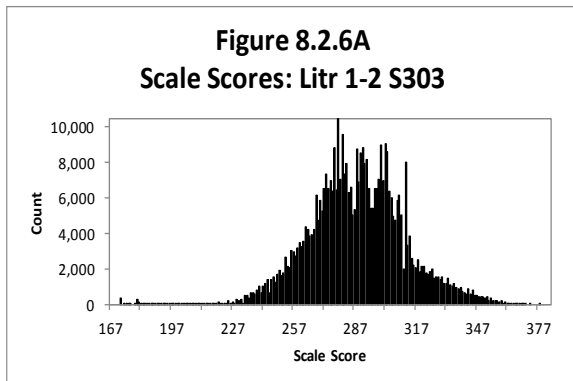


Table 8.2.6A

Scale Score Descriptive Statistics: Litr 1-2 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	222,327	172	367	276.96	19.98
2	215,829	180	378	299.86	21.91
Total	438,156	172	378	288.24	23.87

Table 8.2.6B

Proficiency Level Distribution: Litr 1-2 S303

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
1	11,109	5.0%	8,106	3.8%	19,215	4.4%
2	62,370	28.1%	37,528	17.4%	99,898	22.8%
3	120,217	54.1%	114,623	53.1%	234,840	53.6%
4	20,941	9.4%	37,186	17.2%	58,127	13.3%
5	6,548	2.9%	16,123	7.5%	22,671	5.2%
6	1,142	0.5%	2,263	1.0%	3,405	0.8%
Total	222,327	100.0%	215,829	100.0%	438,156	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 S303

Component	Weight	Variance	Reliability
Reading	0.50	805.189	0.805
Writing	0.50	570.580	0.924
Literacy		569.721	0.912

*Variances from students who had results in all four domains

Table 8.2.6E-1

Accuracy and Consistency of Classification Indices: Litr (Grade 1) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.802	0.728		0.571	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.752		0.667	
	2	0.800		0.713	
	3	0.869		0.827	
	4	0.544		0.431	
	5	0.842		0.625	
	6	-		0.998	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.976	0.013	0.012	0.968
	2/3	0.908	0.048	0.043	0.874
	3/4	0.935	0.029	0.036	0.909
	4/5	0.973	0.026	0.001	0.975
	5/6	0.995	0.005	0.000	0.999

Table 8.2.6E-2

Accuracy and Consistency of Classification Indices: Litr (Grade 2) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.775	0.690		0.537	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.825		0.715	
	2	0.746		0.636	
	3	0.870		0.822	
	4	0.597		0.488	
	5	0.684		0.567	
	6	-		0.765	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.984	0.006	0.010	0.978
	2/3	0.931	0.037	0.032	0.901
	3/4	0.911	0.034	0.055	0.874
	4/5	0.957	0.033	0.010	0.943
	5/6	0.989	0.011	0.000	0.990

8.2.7 Comprehension Composite 1–2

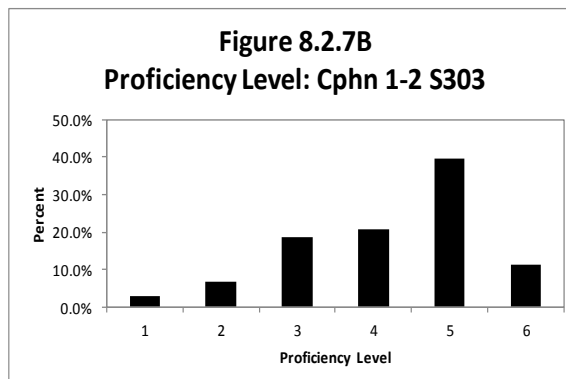
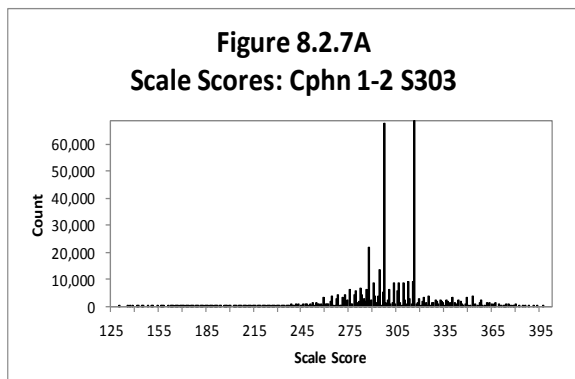


Table 8.2.7A

Scale Score Descriptive Statistics: Cphn 1-2 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	222,397	130	397	288.85	22.83
2	215,898	137	397	315.75	25.00
Total	438,295	130	397	302.10	27.44

Table 8.2.7B

Proficiency Level Distribution: Cphn 1-2 S303

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
1	8,327	3.7%	4,990	2.3%	13,317	3.0%
2	19,204	8.6%	10,321	4.8%	29,525	6.7%
3	49,494	22.3%	33,010	15.3%	82,504	18.8%
4	48,494	21.8%	41,590	19.3%	90,084	20.6%
5	81,718	36.7%	91,261	42.3%	172,979	39.5%
6	15,160	6.8%	34,726	16.1%	49,886	11.4%
Total	222,397	100.0%	215,898	100.0%	438,295	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 S303

Component	Weight	Variance	Reliability
Listening	0.30	1005.833	0.691
Reading	0.70	805.189	0.805
Comprehension		753.180	0.861

*Variances from students who had results in all four domains

Table 8.2.7E-1

Accuracy and Consistency of Classification Indices: Cphn (Grade 1) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.623	0.511		0.350	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.799		0.648	
	2	0.614		0.458	
	3	0.621		0.496	
	4	0.475		0.371	
	5	0.681		0.617	
	6	0.587		0.327	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.983	0.007	0.010	0.974
	2/3	0.949	0.023	0.028	0.925
	3/4	0.882	0.060	0.059	0.834
	4/5	0.854	0.065	0.081	0.800
	5/6	0.935	0.058	0.007	0.910

Table 8.2.7E-2

Accuracy and Consistency of Classification Indices: Cphn (Grade 2) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.634	0.518		0.355
Conditional on Level	Level	Accuracy		Consistency	
	1	0.874		0.731	
	2	0.561		0.403	
	3	0.584		0.438	
	4	0.455		0.352	
	5	0.715		0.632	
	6	0.680		0.536	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.989	0.002	0.009	0.986
	2/3	0.969	0.015	0.016	0.952
	3/4	0.903	0.053	0.044	0.860
	4/5	0.854	0.073	0.073	0.802
	5/6	0.900	0.047	0.054	0.853

8.2.8 Overall Composite 1–2

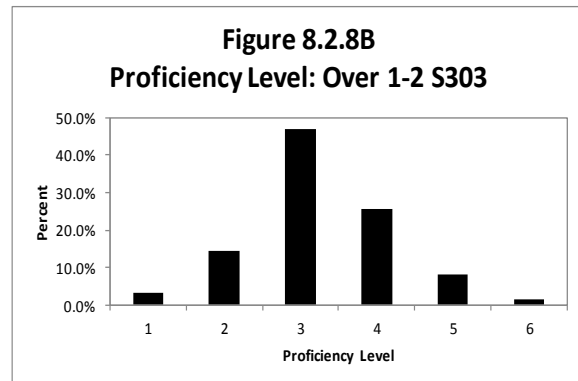
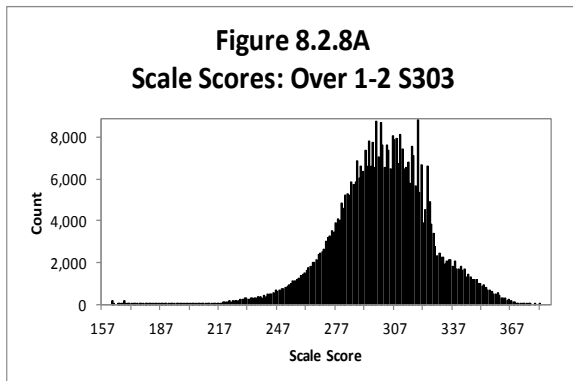


Table 8.2.8A

Scale Score Descriptive Statistics: Over 1-2 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	222,105	162	375	289.45	21.52
2	215,627	168	382	312.41	22.66
Total	437,732	162	382	300.76	24.89

Table 8.2.8B

Proficiency Level Distribution: Over 1-2 S303

Level	Grade 1		Grade 2		Total	
	Count	Percent	Count	Percent	Count	Percent
1	8,157	3.7%	5,887	2.7%	14,044	3.2%
2	42,259	19.0%	20,576	9.5%	62,835	14.4%
3	118,089	53.2%	87,240	40.5%	205,329	46.9%
4	40,996	18.5%	71,909	33.3%	112,905	25.8%
5	10,618	4.8%	25,789	12.0%	36,407	8.3%
6	1,986	0.9%	4,226	2.0%	6,212	1.4%
Total	222,105	100.0%	215,627	100.0%	437,732	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 S303

Component	Weight	Variance	Reliability
Listening	0.15	1005.833	0.691
Reading	0.35	805.189	0.805
Speaking	0.15	2177.737	0.894
Writing	0.35	570.580	0.924
Overall Composite		619.753	0.941

*Variances from students who had results in all four domains

Table 8.2.8E-1

Accuracy and Consistency of Classification Indices: Over (Grade 1) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.819	0.755		0.627	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.726		0.765	
	2	0.815		0.727	
	3	0.888		0.850	
	4	0.684		0.592	
	5	0.763		0.569	
	6	-		1.000	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.982	0.011	0.007	0.983
	2/3	0.938	0.036	0.026	0.919
	3/4	0.918	0.037	0.044	0.895
	4/5	0.961	0.031	0.007	0.960
5/6	0.991	0.009	0.000	0.998	

Table 8.2.8E-2

Accuracy and Consistency of Classification Indices: Over (Grade 2) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.788	0.712		0.592
Conditional on Level	Level	Accuracy		Consistency	
	1	0.811		0.825	
	2	0.761		0.656	
	3	0.873		0.815	
	4	0.757		0.678	
	5	0.641		0.540	
	6	-		0.969	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.990	0.005	0.005	0.990
	2/3	0.963	0.022	0.015	0.949
	3/4	0.913	0.036	0.050	0.882
	4/5	0.931	0.038	0.031	0.906
	5/6	0.980	0.020	0.000	0.983

8.3 Grades: 3–5

8.3.1 Listening 3–5

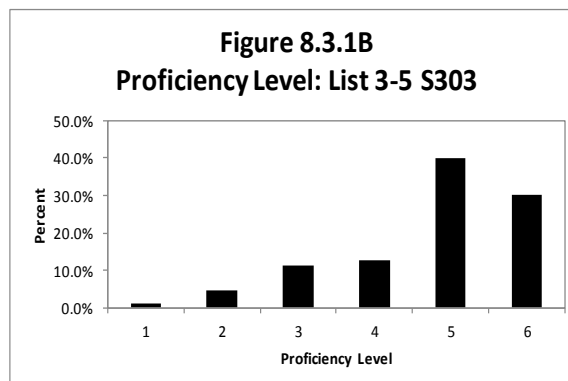
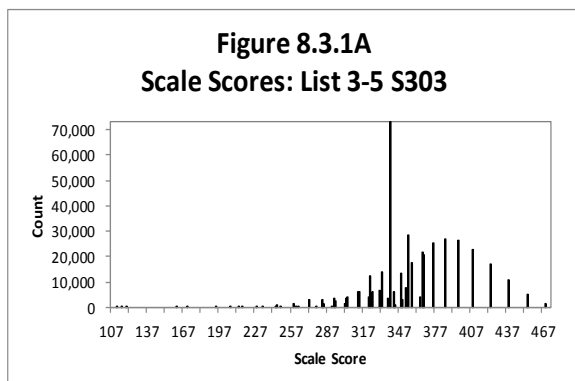


Table 8.3.1A
Scale Score Descriptive Statistics: List 3-5 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	194,648	112	469	351.27	36.03
4	125,460	116	469	363.36	38.25
5	98,168	120	469	373.31	40.37
Total	418,276	112	469	360.07	38.81

Table 8.3.1B
Proficiency Level Distribution: List 3-5 S303

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,188	0.6%	1,339	1.1%	1,605	1.6%	4,132	1.0%
2	7,748	4.0%	6,230	5.0%	5,686	5.8%	19,664	4.7%
3	21,233	10.9%	13,581	10.8%	12,384	12.6%	47,198	11.3%
4	17,281	8.9%	19,309	15.4%	17,139	17.5%	53,729	12.8%
5	85,289	43.8%	48,863	38.9%	32,755	33.4%	166,907	39.9%
6	61,909	31.8%	36,138	28.8%	28,599	29.1%	126,646	30.3%
Total	194,648	100.0%	125,460	100.0%	98,168	100.0%	418,276	100.0%

Table 8.3.1C

Conditional Standard Error of Measurement at Cut Scores: List 3-5 S303*

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	3	255	22.54	22.17	27.05
	4	264	21.79	21.42	25.17
	5	274	20.66	20.66	23.29
2/3	3	295	19.54	19.54	20.66
	4	307	19.16	19.16	19.54
	5	318	19.54	19.16	18.79
3/4	3	325	19.54	19.16	18.41
	4	338	20.29	19.16	18.03
	5	350	21.42	19.91	18.03
4/5	3	340	n/a	19.54	18.03
	4	355	n/a	20.29	18.41
	5	368	n/a	21.42	18.79
5/6	3	367	n/a	n/a	18.41
	4	383	n/a	n/a	19.54
	5	397	n/a	n/a	20.66

* No equating was performed for S303

Figure 8.3.1C
 Test Characteristic Curve: List 3-5ABC S303

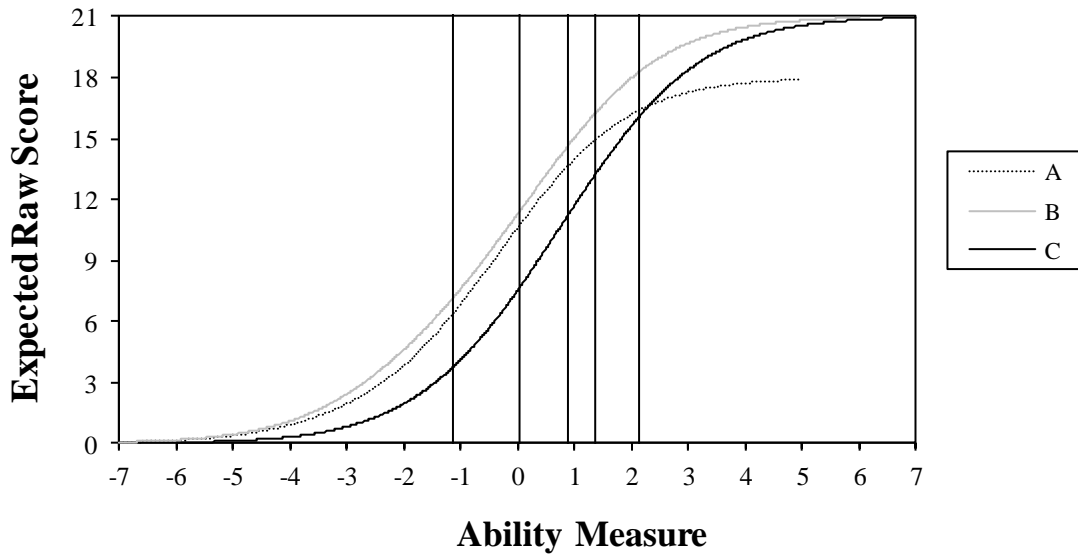


Figure 8.3.1D
 Test Information Function: List 3-5ABC S303

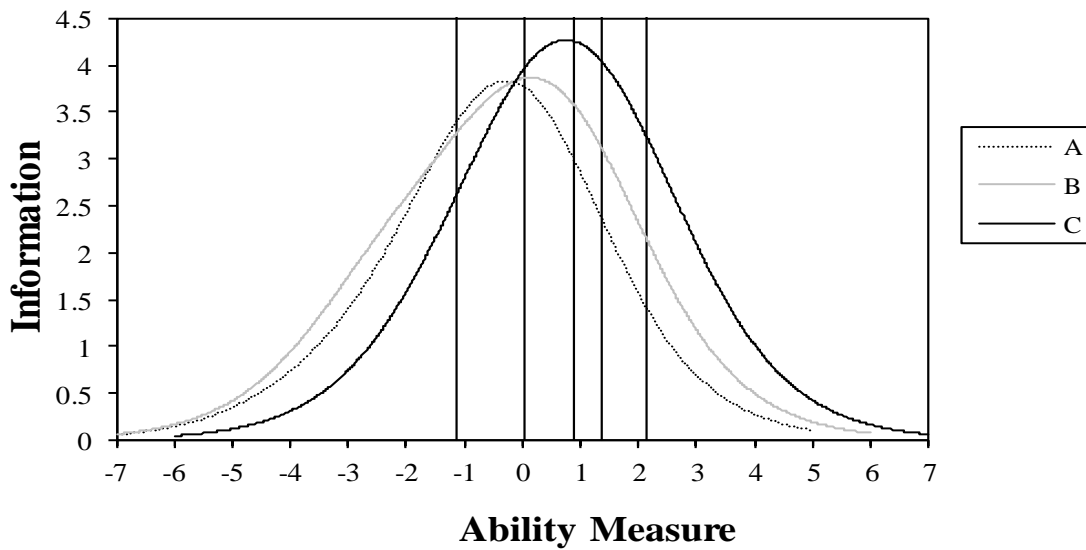


Table 8.3.1D

Weighted Reliability: List 3-5 S303

Tiers	No. of Students	Reliability	Weighted Reliability
A	38,192	0.753	0.661
B	165,281	0.662	
C	214,803	0.644	

Table 8.3.1E-1

Accuracy and Consistency of Classification Indices: List (Grade 3) S303

Overall Indices	Accuracy	Consistency			Kappa (k)
	0.541	0.436			0.219
Conditional on Level	Level	Accuracy			Consistency
	1	0.571			0.204
	2	0.485			0.251
	3	0.384			0.247
	4	0.178			0.133
	5	0.615			0.536
	6	0.693			0.570
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.994	0.000	0.006	0.992
	2/3	0.961	0.006	0.032	0.937
	3/4	0.885	0.053	0.062	0.819
	4/5	0.813	0.122	0.066	0.743
5/6	0.804	0.099	0.097	0.732	

Table 8.3.1E-2

Accuracy and Consistency of Classification Indices: List (Grade 4) S303

Overall Indices	Accuracy	Consistency			Kappa (k)
	0.514	0.408			0.205
Conditional on Level	Level	Accuracy			Consistency
	1	0.747			0.366
	2	0.504			0.289
	3	0.355			0.233
	4	0.293			0.223
	5	0.548			0.474
	6	0.665			0.531
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.991	0.001	0.009	0.987
	2/3	0.956	0.011	0.033	0.929
	3/4	0.879	0.059	0.062	0.815
	4/5	0.801	0.103	0.096	0.734
5/6	0.806	0.097	0.096	0.733	

Table 8.3.1E-3

Accuracy and Consistency of Classification Indices: List (Grade 5) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.487	0.385		0.188
Conditional on Level	Level	Accuracy		Consistency	
	1	0.735		0.419	
	2	0.487		0.295	
	3	0.386		0.263	
	4	0.326		0.246	
	5	0.466		0.398	
	6	0.649		0.511	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.987	0.002	0.011	0.981
	2/3	0.950	0.016	0.034	0.920
	3/4	0.871	0.058	0.071	0.812
	4/5	0.800	0.083	0.117	0.731
	5/6	0.790	0.114	0.096	0.716

8.3.2 Reading 3–5

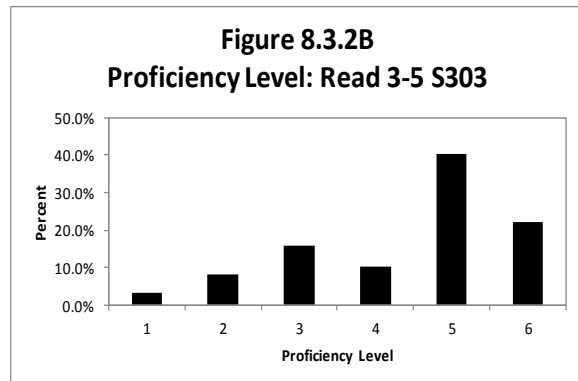
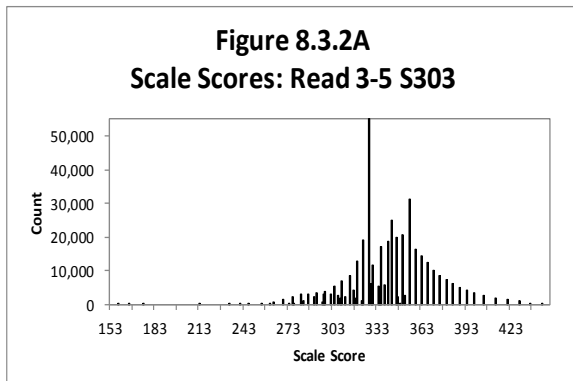


Table 8.3.2A

Scale Score Descriptive Statistics: Read 3-5 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	194,545	158	445	333.33	25.75
4	125,339	166	445	342.74	27.56
5	98,068	175	445	350.41	29.95
Total	417,952	158	445	340.16	28.20

Table 8.3.2B

Proficiency Level Distribution: Read 3-5 S303

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	3,987	2.0%	4,336	3.5%	5,570	5.7%	13,893	3.3%
2	10,753	5.5%	11,765	9.4%	10,880	11.1%	33,398	8.0%
3	20,740	10.7%	20,928	16.7%	23,942	24.4%	65,610	15.7%
4	20,414	10.5%	15,442	12.3%	7,362	7.5%	43,218	10.3%
5	89,688	46.1%	46,783	37.3%	32,571	33.2%	169,042	40.4%
6	48,963	25.2%	26,085	20.8%	17,743	18.1%	92,791	22.2%
Total	194,545	100.0%	125,339	100.0%	98,068	100.0%	417,952	100.0%

Table 8.3.2C

Conditional Standard Error of Measurement at Cut Scores: Read 3-5 S303

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	3	279	12.22	15.60	25.74
	4	291	11.70	13.78	21.06
	5	302	11.44	12.48	17.68
2/3	3	302	11.44	12.48	17.68
	4	316	11.70	11.44	14.56
	5	328	12.22	10.92	12.74
3/4	3	320	11.70	11.18	13.78
	4	336	13.00	10.92	11.70
	5	350	14.82	11.18	10.66
4/5	3	328	n/a	10.92	12.74
	4	343	n/a	10.92	11.18
	5	355	n/a	11.18	10.40
5/6	3	347	n/a	n/a	10.92
	4	360	n/a	n/a	10.40
	5	372	n/a	n/a	10.40

Figure 8.3.2C
 Test Characteristic Curve: Read 3-5ABC S303

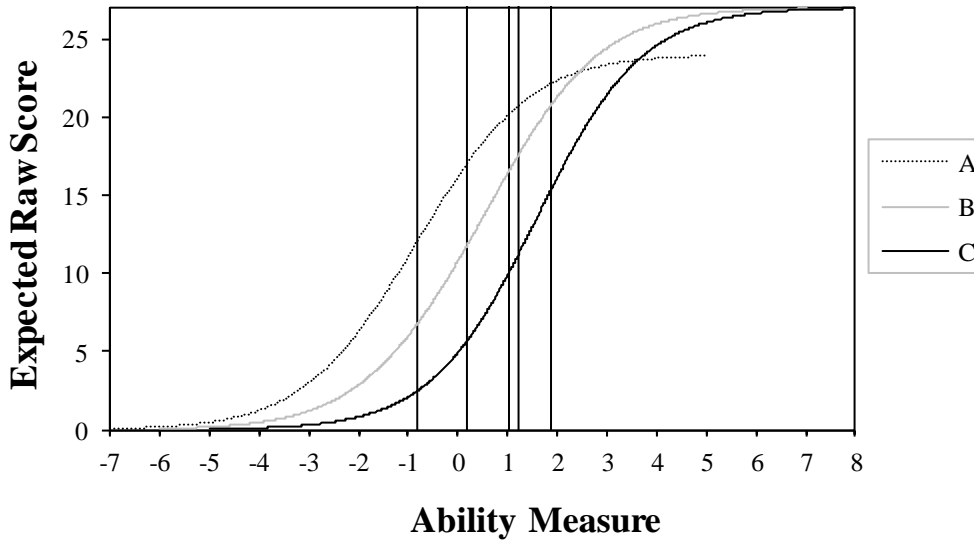


Figure 8.3.2D
 Test Information Function: Read 3-5ABC S303

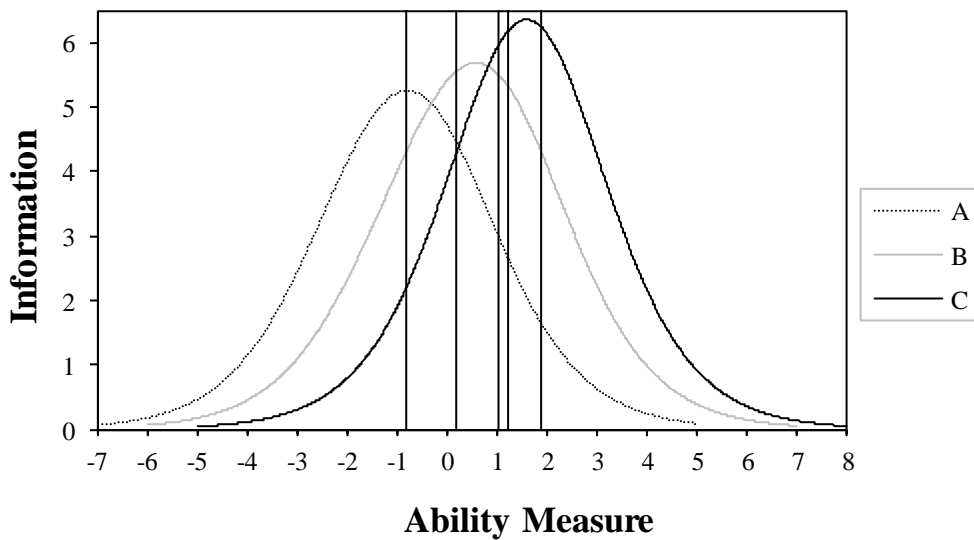


Table 8.3.2D

Weighted Reliability: Read 3-5 S303

Tiers	No. of Students	Reliability	Weighted Reliability
A	38,121	0.799	0.764
B	165,083	0.779	
C	214,748	0.747	

Table 8.3.2E-1

Accuracy and Consistency of Classification Indices: Read (Grade 3) S303

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.559	0.445	0.243		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.769	0.540		
	2	0.537	0.344		
	3	0.393	0.260		
	4	0.234	0.171		
	5	0.661	0.580		
	6	0.604	0.494		
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.987	0.003	0.009	0.981
	2/3	0.959	0.015	0.026	0.934
	3/4	0.895	0.056	0.049	0.839
	4/5	0.836	0.088	0.075	0.775
	5/6	0.811	0.069	0.120	0.744

Table 8.3.2E-2

Accuracy and Consistency of Classification Indices: Read (Grade 4) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.509	0.401		0.233
Conditional on Level	Level	Accuracy		Consistency	
	1	0.753		0.535	
	2	0.556		0.380	
	3	0.426		0.311	
	4	0.232		0.179	
	5	0.570		0.489	
	6	0.573		0.449	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.979	0.007	0.015	0.968
	2/3	0.934	0.025	0.041	0.899
	3/4	0.853	0.078	0.069	0.791
	4/5	0.810	0.094	0.096	0.748
	5/6	0.828	0.066	0.106	0.763

Table 8.3.2E-3

Accuracy and Consistency of Classification Indices: Read (Grade 5) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.506	0.398		0.238
Conditional on Level	Level	Accuracy		Consistency	
	1	0.778		0.567	
	2	0.484		0.336	
	3	0.497		0.387	
	4	0.133		0.105	
	5	0.541		0.456	
	6	0.572		0.430	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.967	0.010	0.023	0.951
	2/3	0.912	0.040	0.048	0.867
	3/4	0.823	0.089	0.088	0.759
	4/5	0.805	0.102	0.093	0.742
	5/6	0.848	0.065	0.087	0.785

8.3.3 Writing 3–5

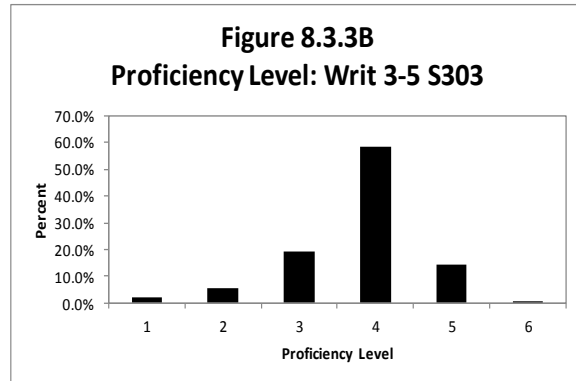
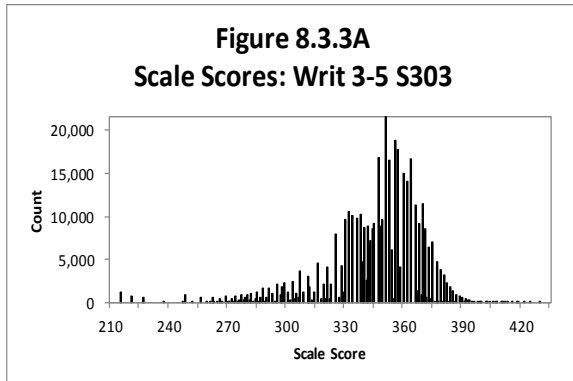


Table 8.3.3A
Scale Score Descriptive Statistics: Writ 3-5 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	194,531	215	412	341.16	25.20
4	125,363	221	425	346.69	25.13
5	98,091	227	430	351.49	25.36
Total	417,985	215	430	345.24	25.56

Table 8.3.3B
Proficiency Level Distribution: Writ 3-5 S303

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,709	1.4%	2,483	2.0%	2,740	2.8%	7,932	1.9%
2	9,288	4.8%	7,714	6.2%	5,916	6.0%	22,918	5.5%
3	28,989	14.9%	24,427	19.5%	27,874	28.4%	81,290	19.4%
4	112,083	57.6%	76,643	61.1%	55,012	56.1%	243,738	58.3%
5	40,363	20.7%	13,888	11.1%	6,442	6.6%	60,693	14.5%
6	1,099	0.6%	208	0.2%	107	0.1%	1,414	0.3%
Total	194,531	100.0%	125,363	100.0%	98,091	100.0%	417,985	100.0%

Table 8.3.3C

Conditional Standard Error of Measurement at Cut Scores: Writ 3-5 S303

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	3	264	10.88	7.77	12.44
	4	275	9.02	6.53	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	7.46
	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	10.57	7.46	8.09
	4	371	9.95	7.15	7.77
	5	381	9.64	6.84	7.15
5/6	3	384	9.33	6.53	7.15
	4	394	9.02	6.53	6.84
	5	403	9.02	6.53	6.53

Figure 8.3.3C
 Test Characteristic Curve: Writ 3-5ABC S303

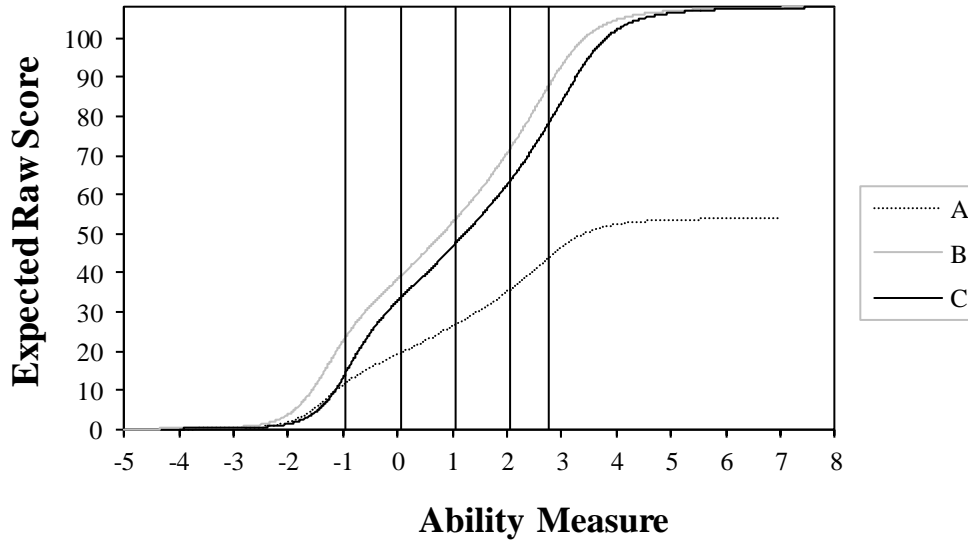


Figure 8.3.3D
 Test Information Function: Writ 3-5ABC S303

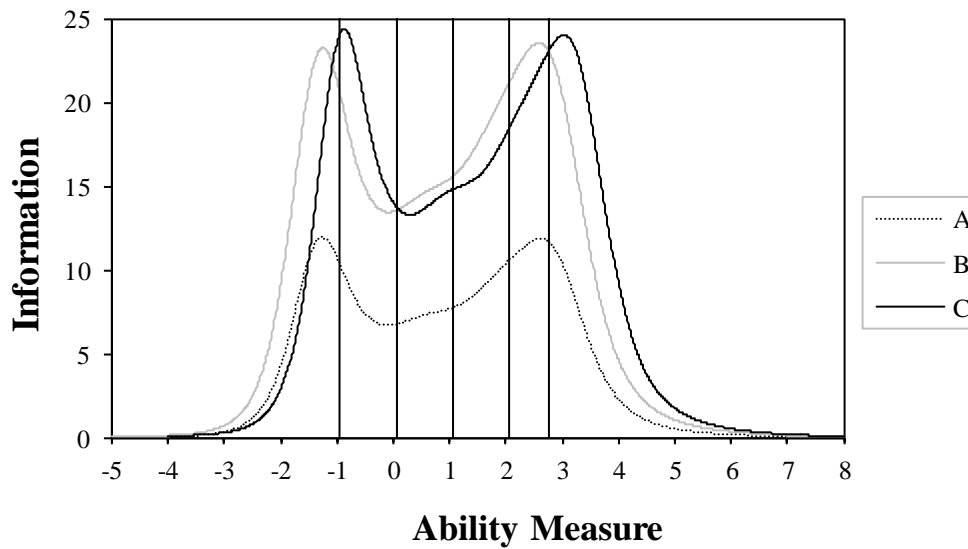


Table 8.3.3D

Weighted Reliability: Writ 3-5 S303

Tiers	No. of Students	Reliability	Weighted Reliability
A	38,099	0.925	0.919
B	165,175	0.931	
C	214,711	0.908	

Table 8.3.3E-1

Accuracy and Consistency of Classification Indices: Writ (Grade 3) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.723	0.657		0.425	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.811		0.786	
	2	0.805		0.710	
	3	0.808		0.710	
	4	0.701		0.724	
	5	-		0.398	
	6	-		0.981	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.995	0.003	0.002	0.994
	2/3	0.984	0.008	0.008	0.979
	3/4	0.954	0.021	0.026	0.935
	4/5	0.787	0.213	0.000	0.751
5/6	0.994	0.006	0.000	0.995	

Table 8.3.3E-2

Accuracy and Consistency of Classification Indices: Writ (Grade 4) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.808	0.744		0.529	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.820		0.797	
	2	0.805		0.708	
	3	0.839		0.746	
	4	0.801		0.791	
	5	-		0.229	
	6	-		0.993	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.993	0.004	0.003	0.992
	2/3	0.980	0.010	0.010	0.973
	3/4	0.944	0.020	0.036	0.923
	4/5	0.888	0.112	0.000	0.856
5/6	0.998	0.002	0.000	1.000	

Table 8.3.3E-3

Accuracy and Consistency of Classification Indices: Writ (Grade 5) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.822	0.770		0.593
Conditional on Level	Level	Accuracy		Consistency	
	1	0.868		0.835	
	2	0.734		0.618	
	3	0.871		0.780	
	4	0.811		0.793	
	5	-		0.152	
	6	-		1.000	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.992	0.004	0.004	0.991
	2/3	0.976	0.014	0.010	0.966
	3/4	0.918	0.022	0.060	0.888
	4/5	0.933	0.067	0.000	0.925
	5/6	0.999	0.001	0.000	1.000

8.3.4 Speaking 3–5

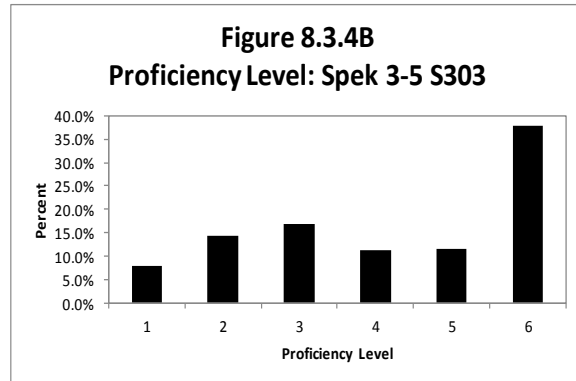
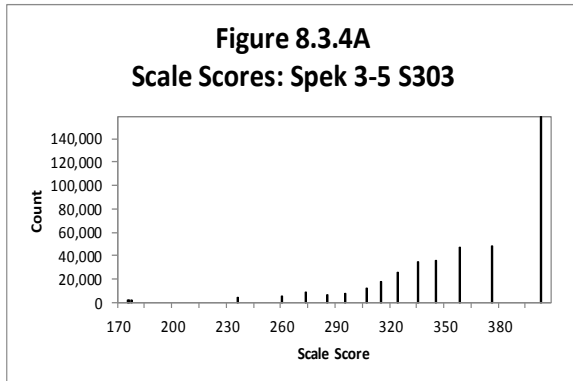


Table 8.3.4A
Scale Score Descriptive Statistics: Spek 3-5 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	194,539	175	403	357.58	45.19
4	125,384	176	403	360.35	46.23
5	98,100	177	403	363.36	47.64
Total	418,023	175	403	359.77	46.15

Table 8.3.4B
Proficiency Level Distribution: Spek 3-5 S303

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	13,419	6.9%	11,029	8.8%	8,886	9.1%	33,334	8.0%
2	33,504	17.2%	15,704	12.5%	10,377	10.6%	59,585	14.3%
3	35,763	18.4%	21,128	16.9%	14,261	14.5%	71,152	17.0%
4	21,992	11.3%	14,462	11.5%	10,333	10.5%	46,787	11.2%
5	22,081	11.4%	15,002	12.0%	11,422	11.6%	48,505	11.6%
6	67,780	34.8%	48,059	38.3%	42,821	43.7%	158,660	38.0%
Total	194,539	100.0%	125,384	100.0%	98,100	100.0%	418,023	100.0%

Table 8.3.4CConditional Standard Error of Measurement at
Cut Scores: Spek 3-5 S303*

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

*No equating was performed for S303

Figure 8.3.4C
 Test Characteristic Curve: Spek 3-5 S303

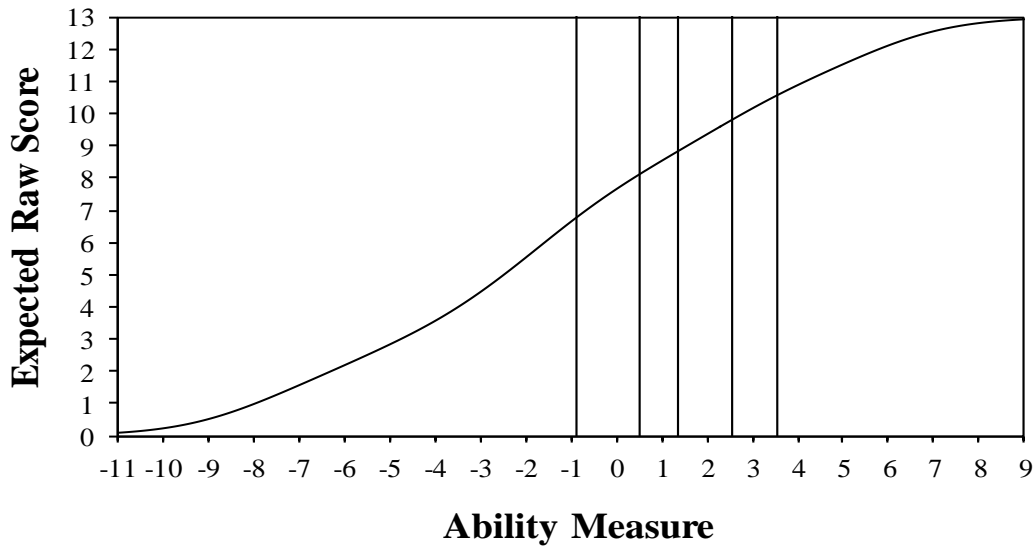


Figure 8.3.4D
 Test Information Function: Spek 3-5 S303

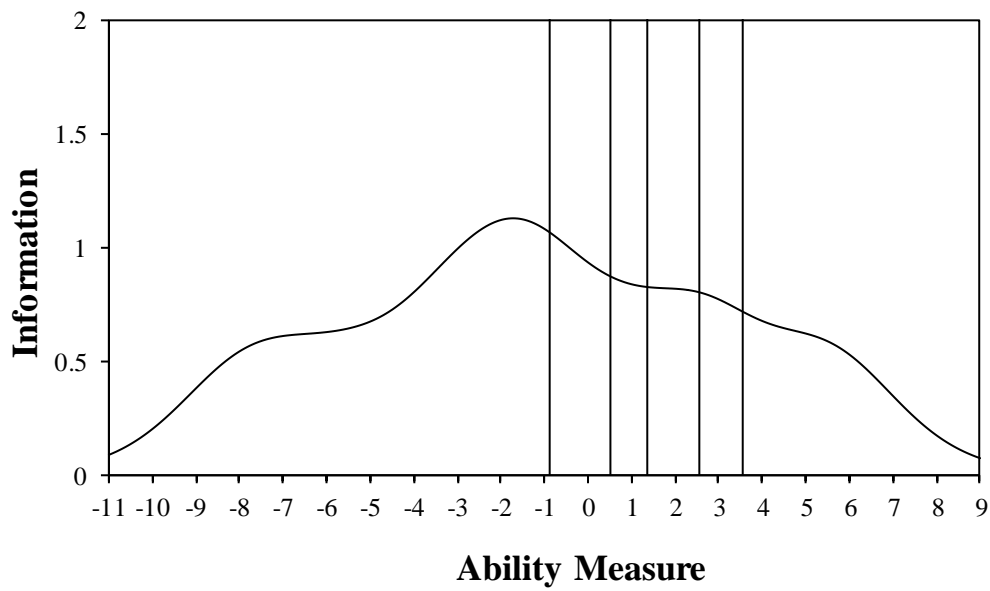


Table 8.3.4D

Reliability: Spek 3-5 S303

Tiers	No. of Students	Reliability
--	418,023	0.897

Table 8.3.4E-1

Accuracy and Consistency of Classification Indices: Spek (Grade 3) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.618	0.529		0.412	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.669		0.516	
	2	0.606		0.497	
	3	0.520		0.432	
	4	0.361		0.269	
	5	0.384		0.277	
	6	0.918		0.853	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.958	0.026	0.016	0.934
	2/3	0.897	0.047	0.056	0.866
	3/4	0.894	0.028	0.078	0.858
	4/5	0.930	0.032	0.037	0.889
	5/6	0.904	0.071	0.025	0.872

Table 8.3.4E-2

Accuracy and Consistency of Classification Indices: Spek (Grade 4) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.609	0.518		0.390	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.739		0.602	
	2	0.500		0.397	
	3	0.522		0.432	
	4	0.364		0.274	
	5	0.345		0.238	
	6	0.899		0.831	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.957	0.025	0.018	0.934
	2/3	0.906	0.048	0.046	0.878
	3/4	0.897	0.027	0.076	0.864
	4/5	0.926	0.031	0.044	0.882
	5/6	0.883	0.084	0.034	0.836

Table 8.3.4E-3

Accuracy and Consistency of Classification Indices: Spek (Grade 5) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.592	0.491		0.346
Conditional on Level	Level	Accuracy		Consistency	
	1	0.736		0.607	
	2	0.451		0.359	
	3	0.496		0.410	
	4	0.344		0.255	
	5	0.278		0.188	
	6	0.882		0.815	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.956	0.026	0.018	0.934
	2/3	0.911	0.047	0.042	0.887
	3/4	0.901	0.024	0.075	0.873
	4/5	0.927	0.028	0.044	0.881
	5/6	0.844	0.112	0.043	0.777

8.3.5 Oral 3–5

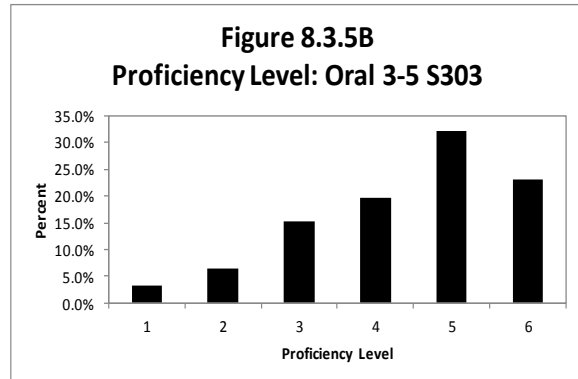
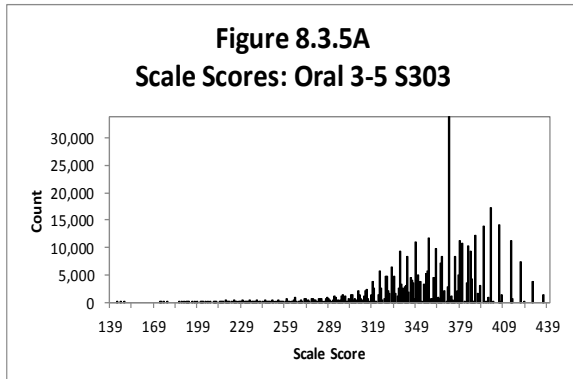


Table 8.3.5A
Scale Score Descriptive Statistics: Oral 3-5 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	194,462	144	436	354.71	35.00
4	125,328	146	436	362.09	36.85
5	98,053	149	436	368.61	38.76
Total	417,843	144	436	360.18	36.90

Table 8.3.5B
Proficiency Level Distribution: Oral 3-5 S303

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	4,798	2.5%	4,228	3.4%	4,651	4.7%	13,677	3.3%
2	12,643	6.5%	8,003	6.4%	6,718	6.9%	27,364	6.5%
3	32,672	16.8%	18,534	14.8%	12,101	12.3%	63,307	15.2%
4	40,488	20.8%	23,965	19.1%	17,757	18.1%	82,210	19.7%
5	57,948	29.8%	40,624	32.4%	36,006	36.7%	134,578	32.2%
6	45,913	23.6%	29,974	23.9%	20,820	21.2%	96,707	23.1%
Total	194,462	100.0%	125,328	100.0%	98,053	100.0%	417,843	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 S303

Component	Weight	Variance	Reliability
Listening	0.50	1502.166	0.661
Speaking	0.50	2122.801	0.897
Oral		1357.734	0.866

*Variances from students who had results in all four domains

Table 8.3.5E-1

Accuracy and Consistency of Classification Indices: Oral (Grade 3) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.618	0.506		0.367	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.833		0.672	
	2	0.579		0.423	
	3	0.603		0.470	
	4	0.501		0.397	
	5	0.597		0.492	
	6	0.746		0.635	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.988	0.003	0.009	0.983
	2/3	0.958	0.021	0.021	0.936
	3/4	0.901	0.043	0.056	0.863
	4/5	0.871	0.057	0.073	0.821
5/6	0.883	0.055	0.062	0.834	

Table 8.3.5E-2

Accuracy and Consistency of Classification Indices: Oral (Grade 4) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.601	0.487		0.340	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.838		0.695	
	2	0.556		0.406	
	3	0.578		0.444	
	4	0.488		0.380	
	5	0.597		0.493	
	6	0.688		0.576	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.985	0.005	0.010	0.979
	2/3	0.958	0.021	0.021	0.937
	3/4	0.909	0.039	0.052	0.874
	4/5	0.875	0.053	0.072	0.826
	5/6	0.856	0.064	0.079	0.801

Table 8.3.5E-3

Accuracy and Consistency of Classification Indices: Oral (Grade 5) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.578	0.467		0.308	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.852		0.729	
	2	0.553		0.410	
	3	0.526		0.393	
	4	0.477		0.362	
	5	0.610		0.519	
	6	0.588		0.473	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.982	0.006	0.012	0.974
	2/3	0.957	0.021	0.022	0.936
	3/4	0.917	0.036	0.048	0.883
	4/5	0.874	0.053	0.073	0.823
	5/6	0.829	0.074	0.097	0.774

8.3.6 Literacy Composite 3–5

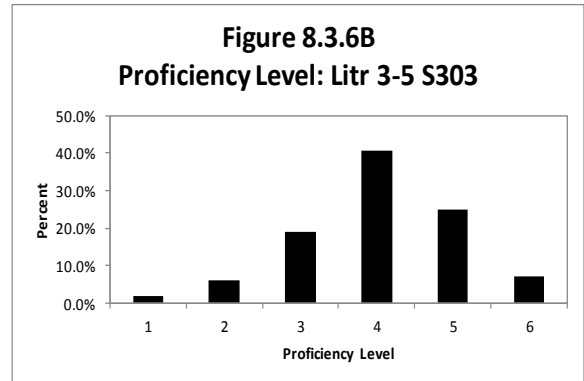
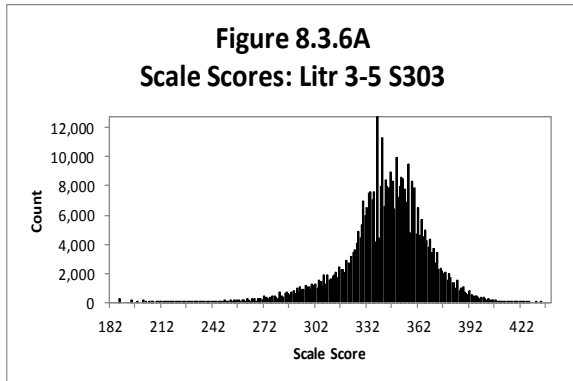


Table 8.3.6A
Scale Score Descriptive Statistics: Litr 3-5 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	194,385	187	419	337.48	23.09
4	125,223	194	434	345.00	24.13
5	97,972	201	434	351.23	25.60
Total	417,580	187	434	342.96	24.65

Table 8.3.6B
Proficiency Level Distribution: Litr 3-5 S303

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,390	1.2%	2,636	2.1%	3,515	3.6%	8,541	2.0%
2	9,325	4.8%	8,128	6.5%	7,906	8.1%	25,359	6.1%
3	28,202	14.5%	24,106	19.3%	27,201	27.8%	79,509	19.0%
4	79,472	40.9%	54,098	43.2%	36,819	37.6%	170,389	40.8%
5	58,074	29.9%	28,334	22.6%	17,612	18.0%	104,020	24.9%
6	16,922	8.7%	7,921	6.3%	4,919	5.0%	29,762	7.1%
Total	194,385	100.0%	125,223	100.0%	97,972	100.0%	417,580	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 S303

Component	Weight	Variance	Reliability
Reading	0.50	793.701	0.764
Writing	0.50	651.754	0.919
Literacy		607.042	0.901

*Variances from students who had results in all four domains

Table 8.3.6E-1

Accuracy and Consistency of Classification Indices: Litr (Grade 3) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.695	0.605		0.450	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.863		0.754	
	2	0.767		0.647	
	3	0.680		0.550	
	4	0.809		0.719	
	5	0.598		0.550	
	6	-		0.407	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.995	0.001	0.003	0.994
	2/3	0.981	0.008	0.011	0.973
	3/4	0.930	0.039	0.030	0.898
	4/5	0.875	0.036	0.090	0.829
	5/6	0.913	0.087	0.000	0.904

Table 8.3.6E-2

Accuracy and Consistency of Classification Indices: Litr (Grade 4) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.687	0.588		0.434	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.865		0.764	
	2	0.744		0.622	
	3	0.702		0.580	
	4	0.791		0.695	
	5	0.545		0.479	
	6	-		0.328	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.993	0.003	0.005	0.990
	2/3	0.974	0.011	0.015	0.962
	3/4	0.915	0.045	0.040	0.878
	4/5	0.868	0.038	0.094	0.817
	5/6	0.937	0.063	0.000	0.928

Table 8.3.6E-3

Accuracy and Consistency of Classification Indices: Litr (Grade 5) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.671	0.569		0.422	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.879		0.789	
	2	0.686		0.554	
	3	0.759		0.651	
	4	0.713		0.603	
	5	0.506		0.434	
	6	-		0.268	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.988	0.004	0.008	0.984
	2/3	0.963	0.019	0.018	0.945
	3/4	0.897	0.046	0.058	0.856
	4/5	0.871	0.042	0.087	0.819
	5/6	0.950	0.050	0.000	0.945

8.3.7 Comprehension Composite 3–5

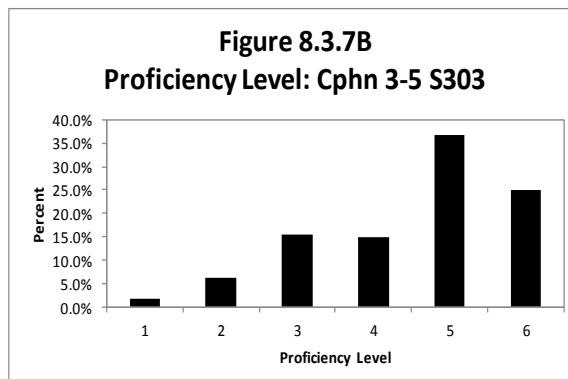
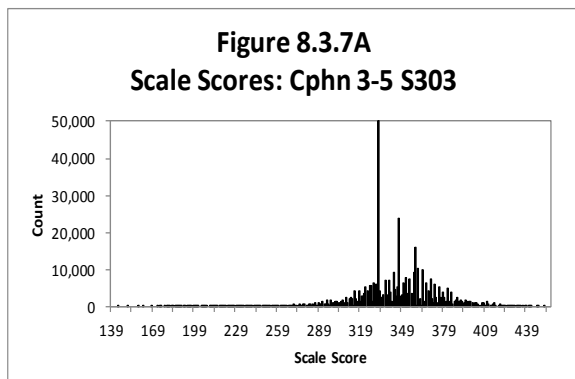


Table 8.3.7A
Scale Score Descriptive Statistics: Cphn 3-5 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	194,489	144	452	338.88	26.89
4	125,305	151	452	349.05	28.69
5	98,022	159	452	357.36	30.99
Total	417,816	144	452	346.27	29.41

Table 8.3.7B
Proficiency Level Distribution: Cphn 3-5 S303

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	1,843	0.9%	2,249	1.8%	3,299	3.4%	7,391	1.8%
2	8,980	4.6%	8,574	6.8%	8,287	8.5%	25,841	6.2%
3	23,636	12.2%	20,324	16.2%	20,193	20.6%	64,153	15.4%
4	26,393	13.6%	20,266	16.2%	15,394	15.7%	62,053	14.9%
5	80,369	41.3%	43,103	34.4%	30,468	31.1%	153,940	36.8%
6	53,268	27.4%	30,789	24.6%	20,381	20.8%	104,438	25.0%
Total	194,489	100.0%	125,305	100.0%	98,022	100.0%	417,816	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 S303

Component	Weight	Variance	Reliability
Listening	0.30	1502.166	0.661
Reading	0.70	793.701	0.764
Comprehension		863.790	0.841

*Variances from students who had results in all four domains

Table 8.3.7E-1

Accuracy and Consistency of Classification Indices: Cphn (Grade 3) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.636	0.528		0.356	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.792		0.560	
	2	0.638		0.455	
	3	0.527		0.378	
	4	0.354		0.264	
	5	0.689		0.602	
	6	0.745		0.639	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.994	0.001	0.005	0.991
	2/3	0.971	0.010	0.018	0.956
	3/4	0.912	0.047	0.041	0.868
	4/5	0.860	0.073	0.066	0.810
5/6	0.868	0.058	0.074	0.814	

Table 8.3.7E-2

Accuracy and Consistency of Classification Indices: Cphn (Grade 4) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.601	0.494		0.341	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.794		0.591	
	2	0.635		0.466	
	3	0.540		0.408	
	4	0.377		0.290	
	5	0.613		0.521	
	6	0.737		0.620	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.990	0.003	0.008	0.985
	2/3	0.959	0.016	0.025	0.938
	3/4	0.890	0.060	0.050	0.843
	4/5	0.849	0.071	0.080	0.798
	5/6	0.875	0.059	0.067	0.822

Table 8.3.7E-3

Accuracy and Consistency of Classification Indices: Cphn (Grade 5) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.580	0.473		0.329	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.825		0.651	
	2	0.584		0.427	
	3	0.562		0.440	
	4	0.347		0.269	
	5	0.577		0.484	
	6	0.718		0.585	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.983	0.004	0.013	0.976
	2/3	0.947	0.024	0.029	0.919
	3/4	0.870	0.069	0.061	0.821
	4/5	0.845	0.068	0.087	0.793
	5/6	0.884	0.057	0.059	0.832

8.3.8 Overall Composite 3–5

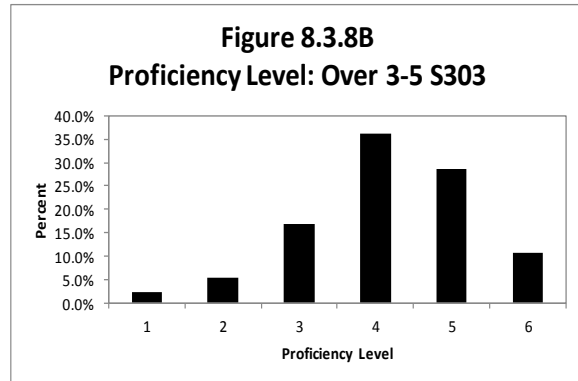
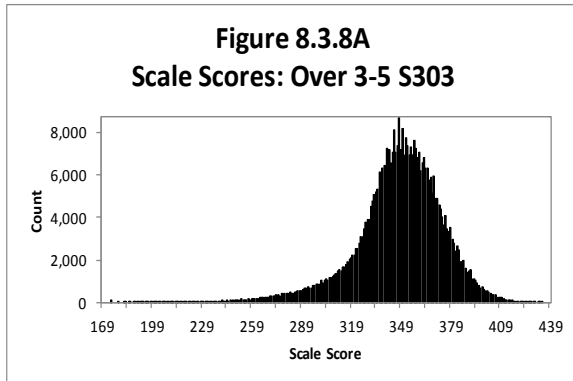


Table 8.3.8A
Scale Score Descriptive Statistics: Over 3-5 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	194,196	174	422	342.46	24.83
4	125,094	179	434	349.92	26.13
5	97,846	185	432	356.24	27.80
Total	417,136	174	434	347.93	26.54

Table 8.3.8B
Proficiency Level Distribution: Over 3-5 S303

Level	Grade 3		Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,768	1.4%	2,980	2.4%	3,535	3.6%	9,283	2.2%
2	8,745	4.5%	7,535	6.0%	6,802	7.0%	23,082	5.5%
3	28,899	14.9%	21,688	17.3%	19,776	20.2%	70,363	16.9%
4	68,120	35.1%	46,195	36.9%	36,229	37.0%	150,544	36.1%
5	60,296	31.0%	34,923	27.9%	23,933	24.5%	119,152	28.6%
6	25,368	13.1%	11,773	9.4%	7,571	7.7%	44,712	10.7%
Total	194,196	100.0%	125,094	100.0%	97,846	100.0%	417,136	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 S303

Component	Weight	Variance	Reliability
Listening	0.15	1502.166	0.661
Reading	0.35	793.701	0.764
Speaking	0.15	2122.801	0.897
Writing	0.35	651.754	0.919
Overall Composite		704.341	0.935

*Variances from students who had results in all four domains

Table 8.3.8E-1

Accuracy and Consistency of Classification Indices: Over (Grade 3) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.749	0.656		0.535	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.909		0.837	
	2	0.766		0.659	
	3	0.738		0.625	
	4	0.819		0.737	
	5	0.685		0.608	
	6	0.759		0.618	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.996	0.001	0.003	0.994
	2/3	0.984	0.008	0.008	0.976
	3/4	0.941	0.033	0.026	0.915
	4/5	0.907	0.031	0.062	0.871
5/6	0.921	0.056	0.024	0.896	

Table 8.3.8E-2

Accuracy and Consistency of Classification Indices: Over (Grade 4) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.712	0.634		0.508
Conditional on Level	Level	Accuracy		Consistency	
	1	0.900		0.828	
	2	0.762		0.656	
	3	0.737		0.626	
	4	0.814		0.729	
	5	0.603		0.567	
	6	1.000		0.477	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.993	0.002	0.005	0.991
	2/3	0.979	0.010	0.011	0.969
	3/4	0.933	0.036	0.030	0.904
	4/5	0.900	0.030	0.070	0.862
	5/6	0.906	0.094	0.000	0.903

Table 8.3.8E-3

Accuracy and Consistency of Classification Indices: Over (Grade 5) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.706	0.620		0.494
Conditional on Level	Level	Accuracy		Consistency	
	1	0.905		0.840	
	2	0.729		0.616	
	3	0.748		0.639	
	4	0.796		0.702	
	5	0.583		0.532	
	6	-		0.394	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.991	0.003	0.006	0.987
	2/3	0.974	0.014	0.013	0.961
	3/4	0.925	0.039	0.036	0.894
	4/5	0.894	0.029	0.077	0.854
	5/6	0.923	0.077	0.000	0.917

8.4 Grades: 6–8

8.4.1 Listening 6–8

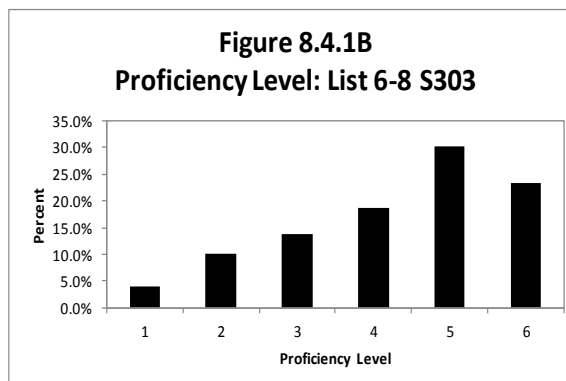
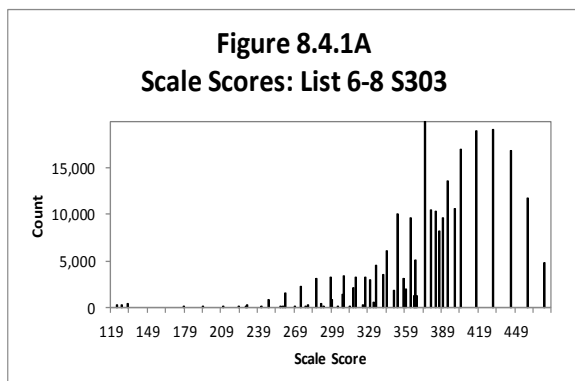


Table 8.4.1A
Scale Score Descriptive Statistics: List 6-8 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	82,654	124	473	376.70	44.51
7	83,927	128	473	385.51	47.01
8	83,773	132	473	392.14	49.72
Total	250,354	124	473	384.82	47.56

Table 8.4.1B
Proficiency Level Distribution: List 6-8 S303

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,097	2.5%	3,237	3.9%	4,600	5.5%	9,934	4.0%
2	7,927	9.6%	8,374	10.0%	9,106	10.9%	25,407	10.1%
3	13,188	16.0%	13,219	15.8%	8,245	9.8%	34,652	13.8%
4	15,285	18.5%	14,840	17.7%	16,619	19.8%	46,744	18.7%
5	26,681	32.3%	26,506	31.6%	22,159	26.5%	75,346	30.1%
6	17,476	21.1%	17,751	21.2%	23,044	27.5%	58,271	23.3%
Total	82,654	100.0%	83,927	100.0%	83,773	100.0%	250,354	100.0%

Table 8.4.1C

Conditional Standard Error of Measurement at Cut Scores: List 6-8 S303*

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	6	283	21.04	19.54	23.67
	7	293	20.66	18.79	21.79
	8	302	20.29	18.41	20.66
2/3	6	328	20.29	18.03	18.41
	7	337	20.66	18.41	18.03
	8	345	21.04	18.79	17.66
3/4	6	359	22.17	19.54	17.66
	7	368	23.29	20.29	18.03
	8	375	24.05	21.04	18.03
4/5	6	380	n/a	21.79	18.41
	7	390	n/a	23.29	19.16
	8	399	n/a	24.42	19.91
5/6	6	409	n/a	n/a	21.04
	7	418	n/a	n/a	22.54
	8	426	n/a	n/a	23.67

* No equating was performed for S303

Figure 8.4.1C
 Test Characteristic Curve: List 6-8ABC S303

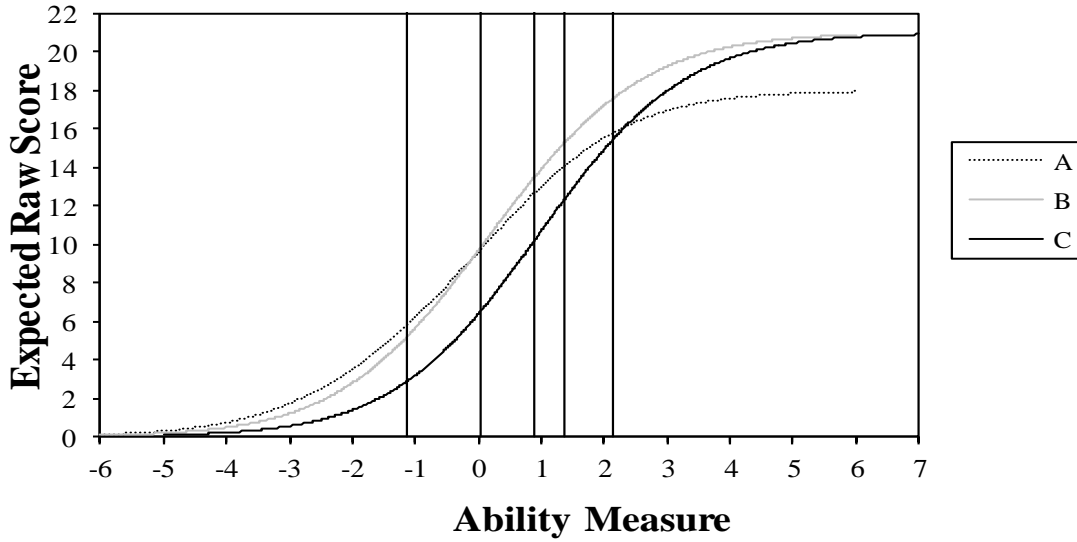


Figure 8.4.1D
 Test Information Function: List 6-8ABC S303

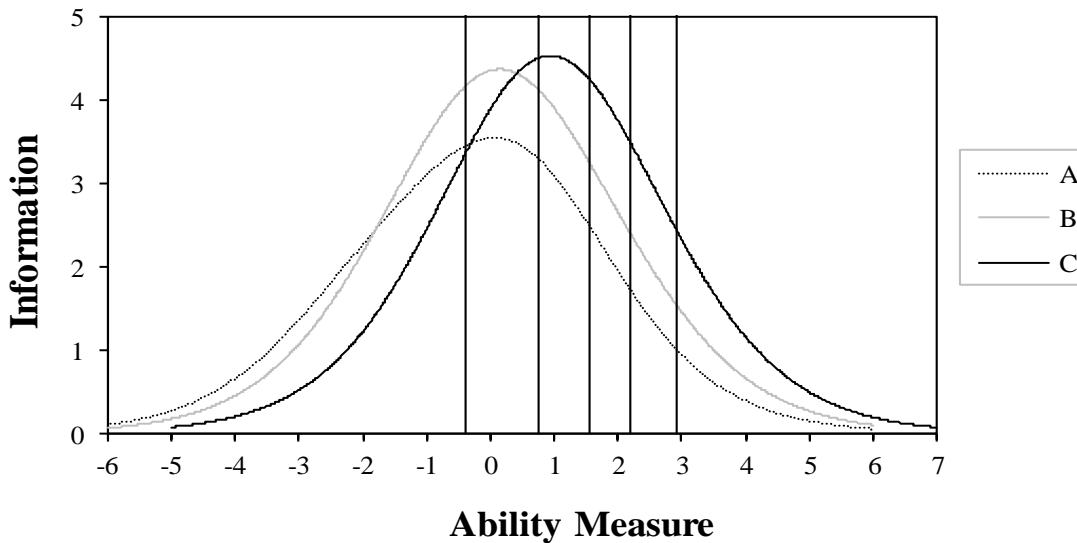


Table 8.4.1D

Weighted Reliability: List 6-8 S303

Tiers	No. of Students	Reliability	Reliability
A	31,619	0.739	0.664
B	86,842	0.677	
C	131,893	0.638	

Table 8.4.1E-1

Accuracy and Consistency of Classification Indices: List (Grade 6) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.447	0.347		0.169	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.660		0.380	
	2	0.522		0.344	
	3	0.391		0.281	
	4	0.308		0.238	
	5	0.461		0.399	
	6	0.554		0.399	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.980	0.006	0.014	0.967
	2/3	0.922	0.026	0.052	0.884
	3/4	0.846	0.061	0.092	0.783
	4/5	0.786	0.092	0.122	0.711
5/6	0.808	0.112	0.080	0.738	

Table 8.4.1E-2

Accuracy and Consistency of Classification Indices: List (Grade 7) S303

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.429	0.333	0.160		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.697	0.435		
	2	0.468	0.312		
	3	0.372	0.268		
	4	0.286	0.221		
	5	0.446	0.388		
	6	0.523	0.380		
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.972	0.008	0.020	0.956
	2/3	0.912	0.030	0.058	0.872
	3/4	0.842	0.060	0.098	0.778
	4/5	0.780	0.094	0.126	0.703
	5/6	0.797	0.113	0.090	0.726

Table 8.4.1E-3

Accuracy and Consistency of Classification Indices: List (Grade 8) S303

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.409	0.333	0.159		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.723	0.499		
	2	0.483	0.332		
	3	0.248	0.170		
	4	0.330	0.245		
	5	0.361	0.320		
	6	0.579	0.444		
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.963	0.011	0.027	0.945
	2/3	0.910	0.025	0.065	0.872
	3/4	0.864	0.060	0.075	0.797
	4/5	0.779	0.091	0.129	0.697
	5/6	0.756	0.160	0.084	0.694

8.4.2 Reading 6–8

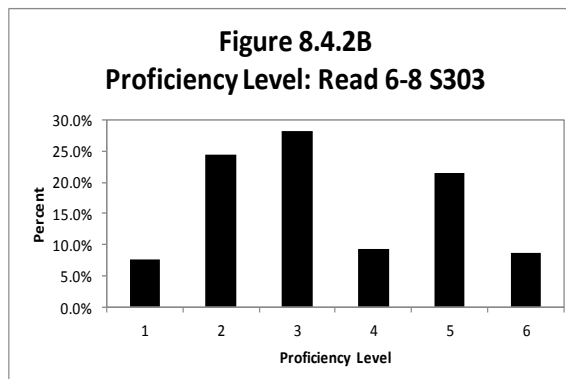
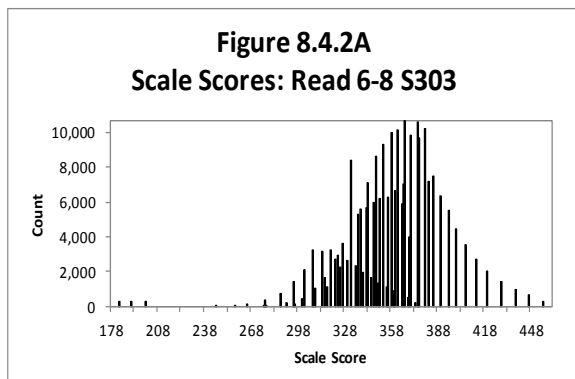


Table 8.4.2A
Scale Score Descriptive Statistics: Read 6-8 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	82,599	183	456	351.15	25.25
7	83,899	191	456	358.50	27.58
8	83,749	200	456	365.65	29.55
Total	250,247	183	456	358.47	28.15

Table 8.4.2B
Proficiency Level Distribution: Read 6-8 S303

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	4,195	5.1%	6,533	7.8%	8,518	10.2%	19,246	7.7%
2	19,497	23.6%	21,839	26.0%	19,991	23.9%	61,327	24.5%
3	26,436	32.0%	23,470	28.0%	20,772	24.8%	70,678	28.2%
4	8,361	10.1%	5,960	7.1%	9,295	11.1%	23,616	9.4%
5	17,964	21.7%	19,449	23.2%	16,346	19.5%	53,759	21.5%
6	6,146	7.4%	6,648	7.9%	8,827	10.5%	21,621	8.6%
Total	82,599	100.0%	83,899	100.0%	83,749	100.0%	250,247	100.0%

Table 8.4.2C

Conditional Standard Error of Measurement at Cut Scores: Read 6-8 S303

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	6	312	12.22	13.26	15.86
	7	321	11.70	12.48	14.30
	8	329	11.44	11.70	13.00
2/3	6	340	11.44	11.18	11.96
	7	349	11.70	10.92	11.44
	8	358	12.22	10.92	10.92
3/4	6	360	12.22	10.92	10.92
	7	369	13.26	11.18	10.66
	8	376	14.04	11.44	10.92
4/5	6	366	n/a	10.92	10.92
	7	375	n/a	11.44	10.92
	8	382	n/a	11.96	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 S303

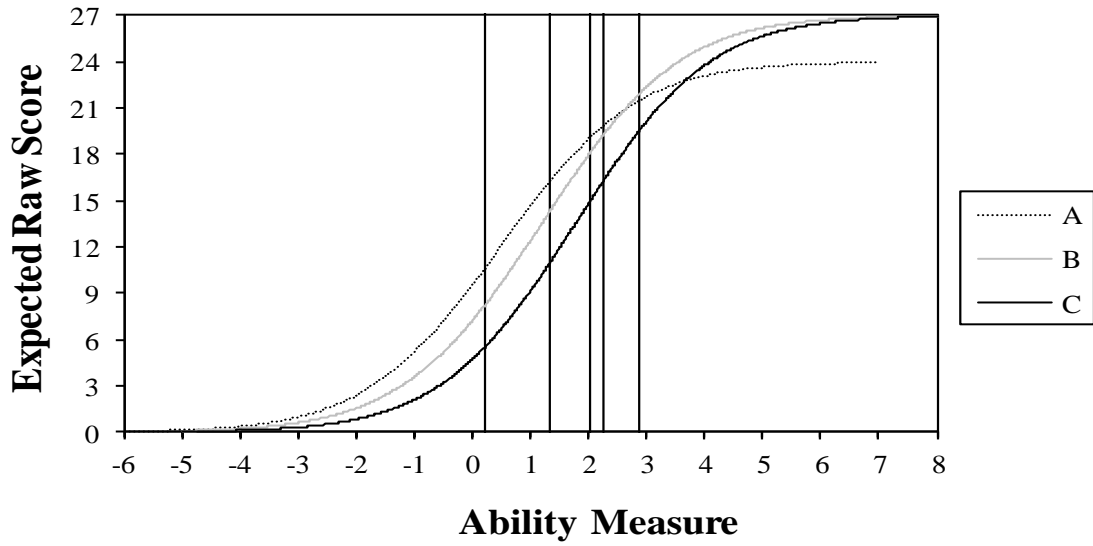


Figure 8.4.2D
 Test Information Function: Read 6-8ABC S303

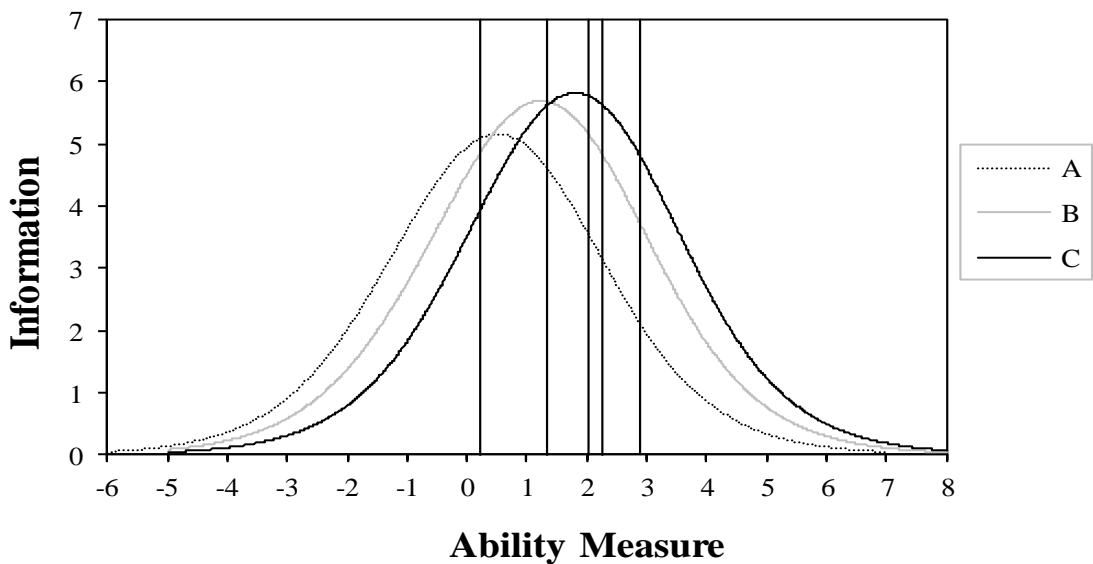


Table 8.4.2D

Weighted Reliability: Read 6-8 S303

Tiers	No. of Students	Reliability	Reliability
A	31,610	0.744	0.775
B	86,785	0.792	
C	131,852	0.771	

Table 8.4.2E-1

Accuracy and Consistency of Classification Indices: Read (Grade 6) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.516	0.409		0.244	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.697		0.505	
	2	0.700		0.555	
	3	0.546		0.440	
	4	0.184		0.145	
	5	0.470		0.389	
	6	-		0.227	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.969	0.015	0.017	0.950
	2/3	0.876	0.045	0.080	0.827
	3/4	0.818	0.076	0.107	0.748
	4/5	0.819	0.089	0.092	0.753
	5/6	0.926	0.074	0.000	0.886

Table 8.4.2E-2

Accuracy and Consistency of Classification Indices: Read (Grade 7) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.517	0.409		0.253
Conditional on Level	Level	Accuracy		Consistency	
	1	0.721		0.539	
	2	0.675		0.542	
	3	0.497		0.397	
	4	0.138		0.107	
	5	0.510		0.422	
	6	0.483		0.263	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.954	0.020	0.026	0.929
	2/3	0.863	0.049	0.088	0.812
	3/4	0.833	0.076	0.091	0.765
	4/5	0.831	0.097	0.072	0.769
	5/6	0.920	0.073	0.007	0.878

Table 8.4.2E-3

Accuracy and Consistency of Classification Indices: Read (Grade 8) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.490	0.390		0.246
Conditional on Level	Level	Accuracy		Consistency	
	1	0.763		0.588	
	2	0.601		0.472	
	3	0.446		0.357	
	4	0.212		0.165	
	5	0.423		0.339	
	6	0.583		0.359	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.944	0.021	0.035	0.918
	2/3	0.861	0.061	0.078	0.804
	3/4	0.823	0.091	0.086	0.761
	4/5	0.830	0.084	0.086	0.769
	5/6	0.905	0.068	0.027	0.861

8.4.3 Writing 6–8

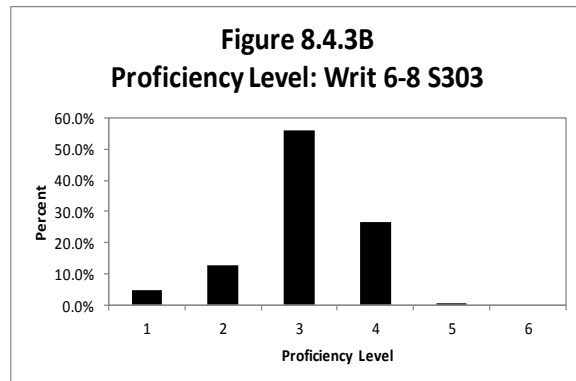
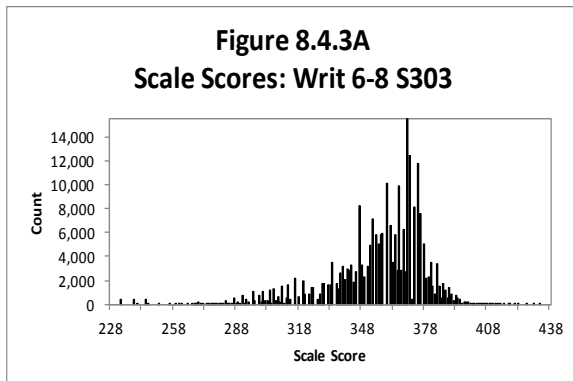


Table 8.4.3A
Scale Score Descriptive Statistics: Writ 6-8 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	82,586	233	430	351.42	23.50
7	83,875	239	433	356.44	23.05
8	83,738	245	427	360.29	22.97
Total	250,199	233	433	356.07	23.45

Table 8.4.3B
Proficiency Level Distribution: Writ 6-8 S303

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,623	3.2%	3,857	4.6%	5,322	6.4%	11,802	4.7%
2	8,700	10.5%	9,916	11.8%	13,049	15.6%	31,665	12.7%
3	38,320	46.4%	46,549	55.5%	54,990	65.7%	139,859	55.9%
4	32,407	39.2%	23,380	27.9%	10,325	12.3%	66,112	26.4%
5	535	0.6%	171	0.2%	52	0.1%	758	0.3%
6	1	0.0%	2	0.0%	0	0.0%	3	0.0%
Total	82,586	100.0%	83,875	100.0%	83,738	100.0%	250,199	100.0%

Table 8.4.3C

Conditional Standard Error of Measurement at Cut Scores: Writ 6-8 S303

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	6	298	9.02	6.84	7.46
	7	308	9.64	7.77	8.09
	8	318	11.19	8.40	8.40
2/3	6	329	11.82	8.40	8.40
	7	339	12.13	8.09	8.09
	8	348	11.82	8.09	8.09
3/4	6	361	11.51	7.77	7.77
	7	371	11.19	7.46	7.15
	8	381	10.57	7.15	6.84
4/5	6	391	10.26	6.84	6.53
	7	399	9.64	6.53	6.53
	8	408	9.33	6.53	6.53
5/6	6	412	9.02	6.53	6.84
	7	420	9.02	6.84	7.46
	8	428	9.33	8.09	9.02

Figure 8.4.3C
 Test Characteristic Curve: Writ 6-8ABC S303

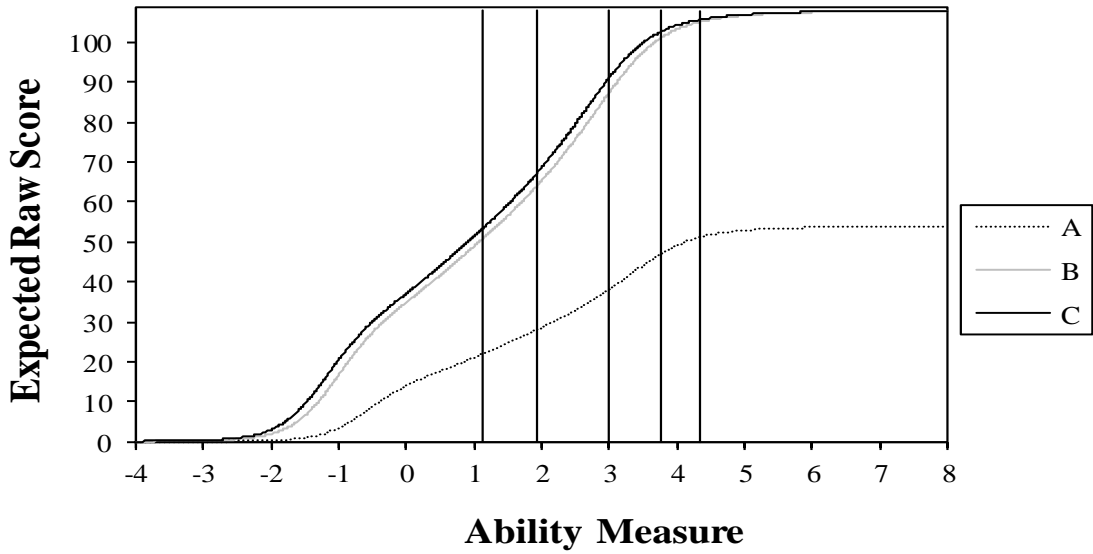


Figure 8.4.3D
 Test Information Function: Writ 6-8ABC S303

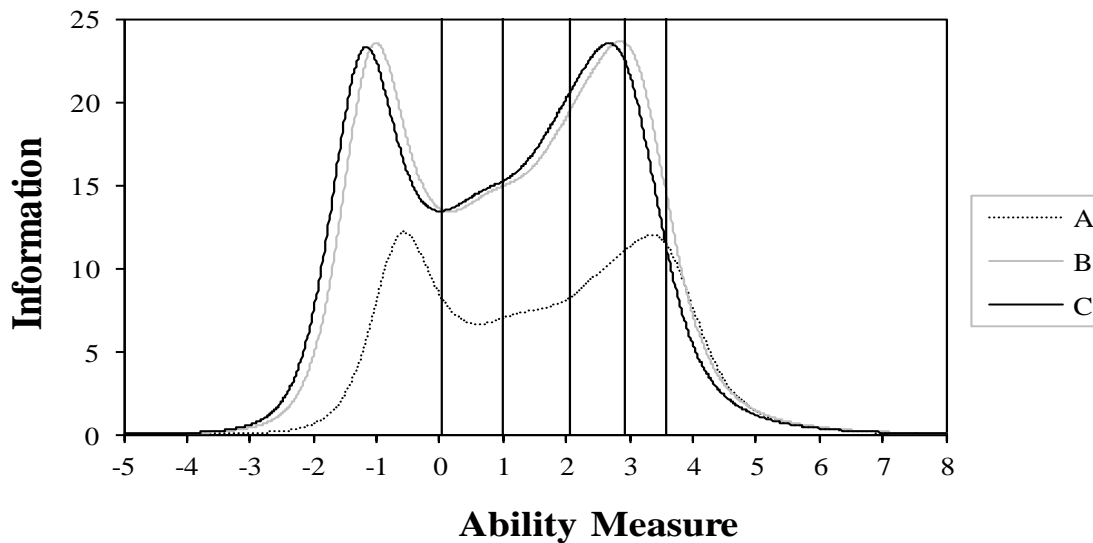


Table 8.4.3D

Weighted Reliability: Writ 6-8 S303

Tiers	No. of Students	Reliability	Reliability
A	31,612	0.884	0.915
B	86,758	0.933	
C	131,829	0.910	

Table 8.4.3E-1

Accuracy and Consistency of Classification Indices: Writ (Grade 6) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.804	0.727		0.561	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.867		0.781	
	2	0.789		0.686	
	3	0.848		0.745	
	4	0.763		0.717	
	5	-		0.000	
	6	-		1.000	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.991	0.004	0.005	0.987
	2/3	0.965	0.017	0.018	0.949
	3/4	0.855	0.044	0.102	0.798
	4/5	0.994	0.006	0.000	0.993
5/6	1.000	0.000	0.000	1.000	

Table 8.4.3E-2

Accuracy and Consistency of Classification Indices: Writ (Grade 7) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.742	0.667		0.446	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.890		0.821	
	2	0.801		0.704	
	3	0.807		0.720	
	4	0.603		0.527	
	5	-		-	
	6	-		1.000	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.989	0.005	0.007	0.984
	2/3	0.963	0.017	0.020	0.948
	3/4	0.792	0.078	0.130	0.737
	4/5	0.998	0.002	0.000	0.998
	5/6	1.000	0.000	0.000	1.000

Table 8.4.3E-3

Accuracy and Consistency of Classification Indices: Writ (Grade 8) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.816	0.744		0.497	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.877		0.815	
	2	0.825		0.735	
	3	0.809		0.808	
	4	-		0.255	
	5	-		1.000	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.984	0.008	0.008	0.978
	2/3	0.955	0.018	0.027	0.937
	3/4	0.876	0.124	0.000	0.829
	4/5	0.999	0.001	0.000	1.000

Proficiency level 6 was not observed with 8th graders.

8.4.4 Speaking 6–8

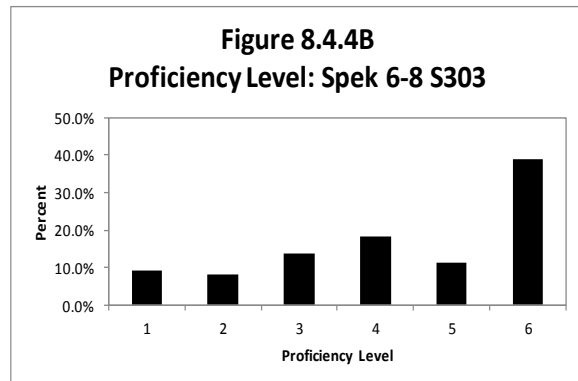
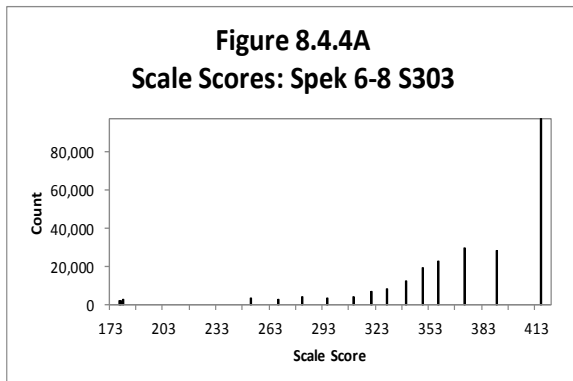


Table 8.4.4A
Scale Score Descriptive Statistics: Spek 6-8 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	82,580	178	416	369.26	50.05
7	83,870	179	416	373.04	51.09
8	83,710	180	416	374.90	52.15
Total	250,160	178	416	372.42	51.16

Table 8.4.4B
Proficiency Level Distribution: Spek 6-8 S303

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	6,598	8.0%	8,167	9.7%	8,334	10.0%	23,099	9.2%
2	7,305	8.8%	5,055	6.0%	8,127	9.7%	20,487	8.2%
3	12,360	15.0%	10,152	12.1%	12,076	14.4%	34,588	13.8%
4	18,972	23.0%	17,627	21.0%	9,422	11.3%	46,021	18.4%
5	9,121	11.0%	9,600	11.4%	9,762	11.7%	28,483	11.4%
6	28,224	34.2%	33,269	39.7%	35,989	43.0%	97,482	39.0%
Total	82,580	100.0%	83,870	100.0%	83,710	100.0%	250,160	100.0%

Table 8.4.4C

Conditional Standard Error of Measurement at
Cut Scores: Spek 6-8 S303*

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

*No equating was performed for Series 303

Figure 8.4.4C
 Test Characteristic Curve: Spek 6-8 S303

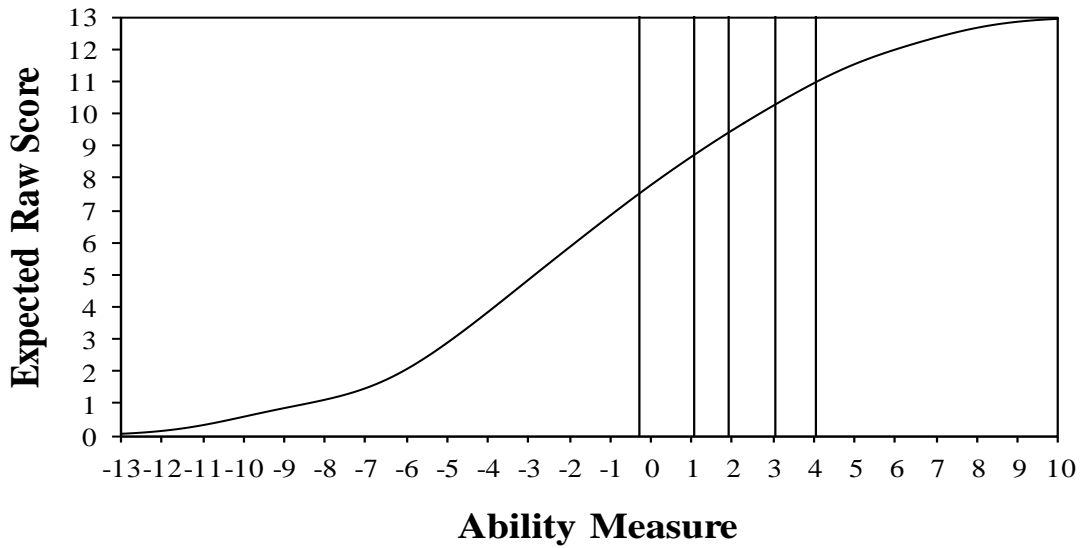


Figure 8.4.4D
 Test Information Function: Spek 6-8 S303

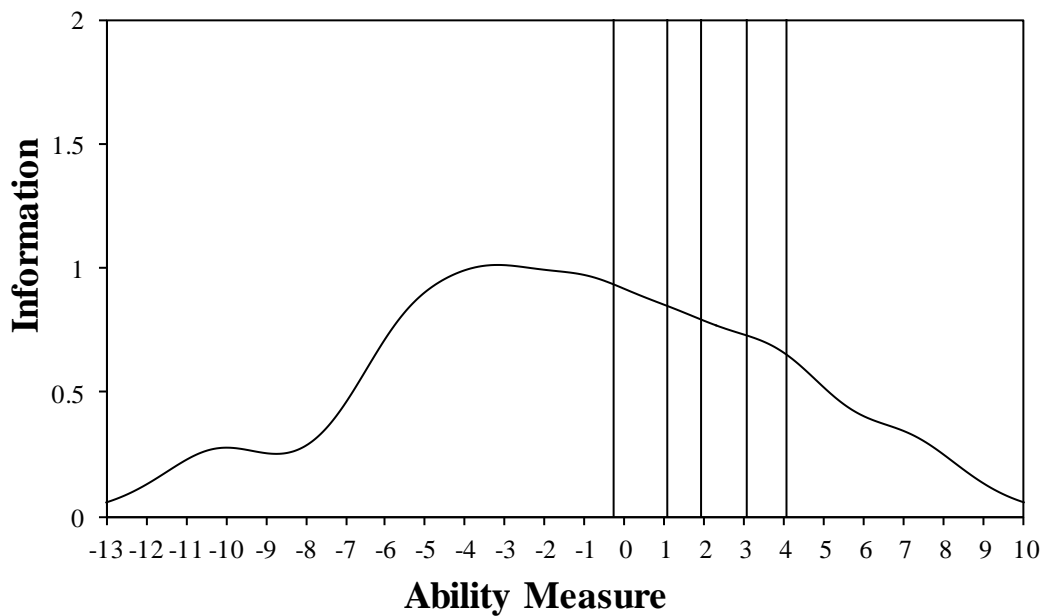


Table 8.4.4D

Reliability: Spek 6-8 S303

Tiers	No. of Students	Reliability
--	250,160	0.914

Table 8.4.4E-1

Accuracy and Consistency of Classification Indices: Spek (Grade 6) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.582	0.474		0.341	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.748		0.612	
	2	0.408		0.304	
	3	0.456		0.374	
	4	0.547		0.449	
	5	0.249		0.174	
	6	0.816		0.733	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.965	0.023	0.012	0.944
	2/3	0.915	0.053	0.033	0.886
	3/4	0.881	0.044	0.075	0.850
	4/5	0.885	0.020	0.095	0.839
5/6	0.878	0.059	0.064	0.805	

Table 8.4.4E-2

Accuracy and Consistency of Classification Indices: Spek (Grade 7) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.600	0.486		0.340	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.820		0.706	
	2	0.341		0.246	
	3	0.442		0.355	
	4	0.555		0.442	
	5	0.228		0.165	
	6	0.811		0.734	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.968	0.018	0.014	0.949
	2/3	0.933	0.042	0.025	0.908
	3/4	0.900	0.043	0.058	0.873
	4/5	0.882	0.020	0.098	0.838
	5/6	0.855	0.068	0.077	0.768

Table 8.4.4E-3

Accuracy and Consistency of Classification Indices: Spek (Grade 8) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.580	0.464		0.317	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.754		0.643	
	2	0.441		0.352	
	3	0.512		0.429	
	4	0.341		0.243	
	5	0.232		0.170	
	6	0.819		0.756	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.958	0.028	0.014	0.937
	2/3	0.916	0.046	0.038	0.894
	3/4	0.902	0.022	0.076	0.877
	4/5	0.917	0.019	0.063	0.866
	5/6	0.824	0.104	0.072	0.737

8.4.5 Oral Language Composite 6–8

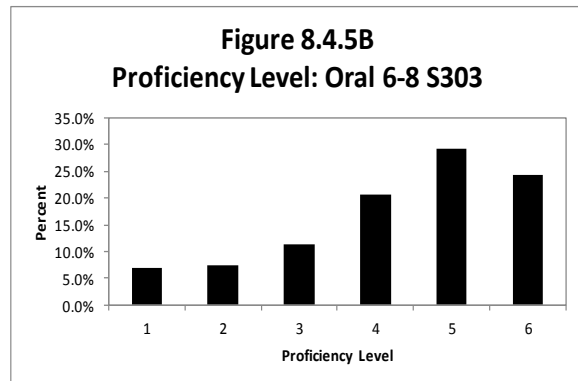
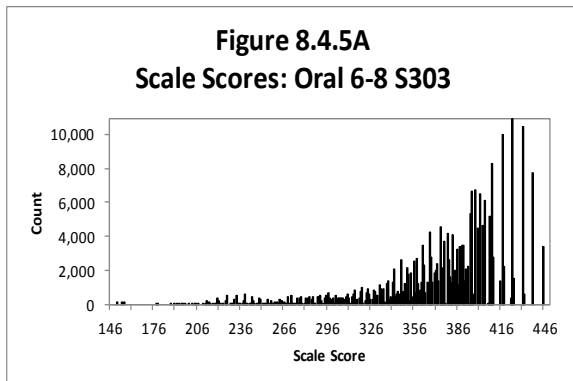


Table 8.4.5A
Scale Score Descriptive Statistics: Oral 6-8 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	82,500	151	445	373.25	41.95
7	83,783	154	445	379.55	43.90
8	83,574	156	445	383.84	45.90
Total	249,857	151	445	378.91	44.17

Table 8.4.5B
Proficiency Level Distribution: Oral 6-8 S303

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	4,803	5.8%	5,839	7.0%	6,774	8.1%	17,416	7.0%
2	6,134	7.4%	6,167	7.4%	6,256	7.5%	18,557	7.4%
3	9,904	12.0%	9,757	11.6%	8,916	10.7%	28,577	11.4%
4	18,791	22.8%	16,439	19.6%	16,478	19.7%	51,708	20.7%
5	24,776	30.0%	24,432	29.2%	23,736	28.4%	72,944	29.2%
6	18,092	21.9%	21,149	25.2%	21,414	25.6%	60,655	24.3%
Total	82,500	100.0%	83,783	100.0%	83,574	100.0%	249,857	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 S303

Component	Weight	Variance	Reliability
Listening	0.50	2259.054	0.664
Speaking	0.50	2614.955	0.914
Oral		1949.684	0.874

*Variances from students who had results in all four domains

Table 8.4.5E-1

Accuracy and Consistency of Classification Indices: Oral (Grade 6) S303

Overall	Accuracy	Consistency	Kappa (k)		
Indices	0.567	0.458	0.311		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.852	0.736		
	2	0.539	0.395		
	3	0.479	0.357		
	4	0.553	0.433		
	5	0.529	0.441		
6	0.621	0.503			
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.979	0.008	0.014	0.969
	2/3	0.952	0.024	0.024	0.928
	3/4	0.910	0.042	0.048	0.876
	4/5	0.866	0.048	0.086	0.816
5/6	0.837	0.075	0.088	0.782	

Table 8.4.5E-2

Accuracy and Consistency of Classification Indices: Oral (Grade 7) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.559	0.453		0.306	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.854		0.743	
	2	0.512		0.375	
	3	0.471		0.351	
	4	0.505		0.385	
	5	0.506		0.418	
	6	0.636		0.529	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.975	0.009	0.016	0.964
	2/3	0.950	0.024	0.025	0.927
	3/4	0.912	0.040	0.048	0.879
	4/5	0.871	0.046	0.083	0.821
	5/6	0.821	0.082	0.097	0.763

Table 8.4.5E-3

Accuracy and Consistency of Classification Indices: Oral (Grade 8) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.544	0.443		0.296	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.865		0.765	
	2	0.510		0.376	
	3	0.453		0.333	
	4	0.505		0.372	
	5	0.474		0.401	
	6	0.610		0.507	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.973	0.010	0.017	0.961
	2/3	0.950	0.024	0.026	0.928
	3/4	0.917	0.036	0.047	0.885
	4/5	0.867	0.046	0.087	0.812
	5/6	0.801	0.097	0.102	0.747

8.4.6 Literacy Composite 6–8

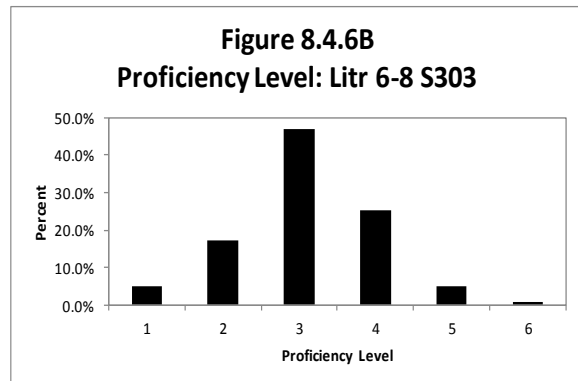
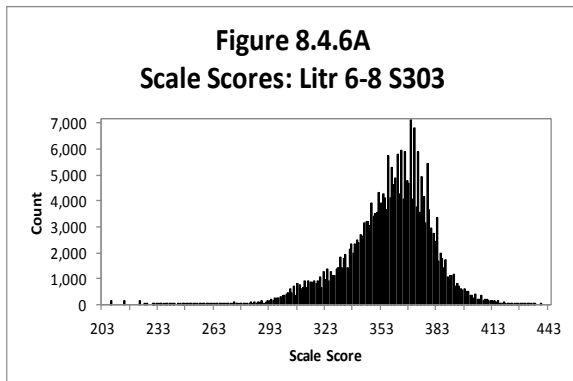


Table 8.4.6A
Scale Score Descriptive Statistics: Litr 6-8 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	82,485	208	430	351.54	21.65
7	83,794	215	439	357.72	22.75
8	83,616	223	435	363.24	23.75
Total	249,895	208	439	357.53	23.23

Table 8.4.6B
Proficiency Level Distribution: Litr 6-8 S303

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,254	2.7%	4,095	4.9%	5,787	6.9%	12,136	4.9%
2	12,772	15.5%	14,012	16.7%	16,117	19.3%	42,901	17.2%
3	37,268	45.2%	39,319	46.9%	40,615	48.6%	117,202	46.9%
4	25,197	30.5%	21,544	25.7%	16,744	20.0%	63,485	25.4%
5	4,245	5.1%	4,126	4.9%	3,756	4.5%	12,127	4.9%
6	749	0.9%	698	0.8%	597	0.7%	2,044	0.8%
Total	82,485	100.0%	83,794	100.0%	83,616	100.0%	249,895	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 S303

Component	Weight	Variance	Reliability
Reading	0.50	791.689	0.775
Writing	0.50	548.659	0.915
Literacy		539.376	0.896

*Variances from students who had results in all four domains

Table 8.4.6E-1

Accuracy and Consistency of Classification Indices: Litr (Grade 6) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.746	0.659		0.496	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.660		0.674	
	2	0.791		0.681	
	3	0.817		0.732	
	4	0.659		0.598	
	5	-		0.232	
	6	-		0.998	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.983	0.011	0.006	0.983
	2/3	0.938	0.031	0.031	0.919
	3/4	0.869	0.053	0.078	0.823
	4/5	0.939	0.061	0.000	0.927
	5/6	0.991	0.009	0.000	0.997

Table 8.4.6E-2

Accuracy and Consistency of Classification Indices: Litr (Grade 7) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.731	0.643		0.479	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.749		0.742	
	2	0.766		0.653	
	3	0.811		0.725	
	4	0.609		0.540	
	5	-		0.220	
	6	-		0.998	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.977	0.013	0.010	0.975
	2/3	0.931	0.035	0.034	0.910
	3/4	0.862	0.055	0.083	0.815
	4/5	0.942	0.058	0.000	0.932
	5/6	0.992	0.008	0.000	0.998

Table 8.4.6E-3

Accuracy and Consistency of Classification Indices: Litr (Grade 8) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.725	0.636		0.471	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.792		0.756	
	2	0.751		0.639	
	3	0.809		0.727	
	4	0.550		0.472	
	5	-		0.224	
	6	-		0.998	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.971	0.015	0.014	0.967
	2/3	0.922	0.039	0.039	0.896
	3/4	0.867	0.055	0.078	0.820
	4/5	0.948	0.052	0.000	0.940
	5/6	0.993	0.007	0.000	0.998

8.4.7 Comprehension Composite 6–8

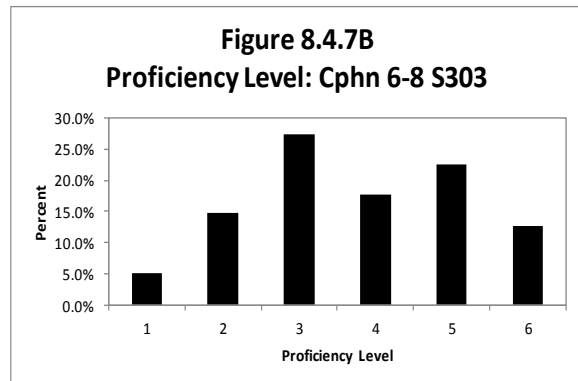
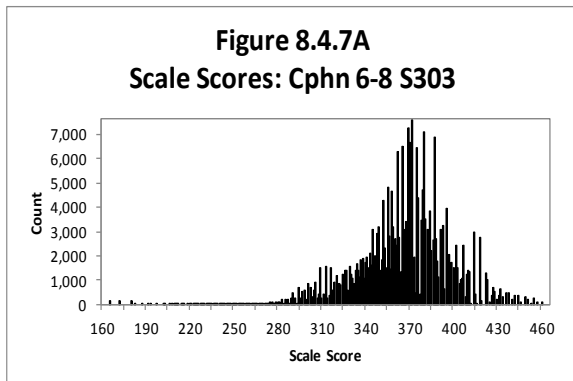


Table 8.4.7A
Scale Score Descriptive Statistics: Cphn 6-8 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	82,546	165	461	358.87	28.32
7	83,841	172	461	366.70	30.77
8	83,669	180	461	373.68	33.09
Total	250,056	165	461	366.45	31.38

Table 8.4.7B
Proficiency Level Distribution: Cphn 6-8 S303

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,667	3.2%	4,081	4.9%	5,949	7.1%	12,697	5.1%
2	11,379	13.8%	12,237	14.6%	13,093	15.6%	36,709	14.7%
3	24,497	29.7%	23,359	27.9%	20,469	24.5%	68,325	27.3%
4	15,316	18.6%	15,328	18.3%	13,351	16.0%	43,995	17.6%
5	19,540	23.7%	18,407	22.0%	18,506	22.1%	56,453	22.6%
6	9,147	11.1%	10,429	12.4%	12,301	14.7%	31,877	12.7%
Total	82,546	100.0%	83,841	100.0%	83,669	100.0%	250,056	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 S303

Component	Weight	Variance	Reliability
Listening	0.30	2259.054	0.664
Reading	0.70	791.689	0.775
Comprehension		984.454	0.842

*Variances from students who had results in all four domains

Table 8.4.7E-1

Accuracy and Consistency of Classification Indices: Cphn (Grade 6) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.566	0.458		0.313	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.774		0.602	
	2	0.681		0.533	
	3	0.644		0.533	
	4	0.389		0.301	
	5	0.506		0.424	
	6	0.648		0.428	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.983	0.006	0.010	0.975
	2/3	0.927	0.033	0.040	0.894
	3/4	0.855	0.064	0.081	0.803
	4/5	0.853	0.056	0.090	0.797
	5/6	0.907	0.073	0.021	0.874

Table 8.4.7E-2

Accuracy and Consistency of Classification Indices: Cphn (Grade 7) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.548	0.440		0.303	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.786		0.623	
	2	0.646		0.502	
	3	0.610		0.498	
	4	0.383		0.296	
	5	0.477		0.392	
	6	0.653		0.448	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.976	0.009	0.015	0.964
	2/3	0.920	0.036	0.044	0.884
	3/4	0.855	0.065	0.080	0.803
	4/5	0.854	0.060	0.086	0.797
	5/6	0.900	0.073	0.027	0.863

Table 8.4.7E-3

Accuracy and Consistency of Classification Indices: Cphn (Grade 8) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.540	0.433		0.305	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.814		0.666	
	2	0.613		0.476	
	3	0.558		0.446	
	4	0.343		0.264	
	5	0.477		0.387	
	6	0.674		0.492	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.967	0.011	0.021	0.953
	2/3	0.911	0.041	0.047	0.873
	3/4	0.858	0.064	0.078	0.807
	4/5	0.856	0.063	0.081	0.800
	5/6	0.893	0.071	0.037	0.851

8.4.8 Overall Composite 6–8

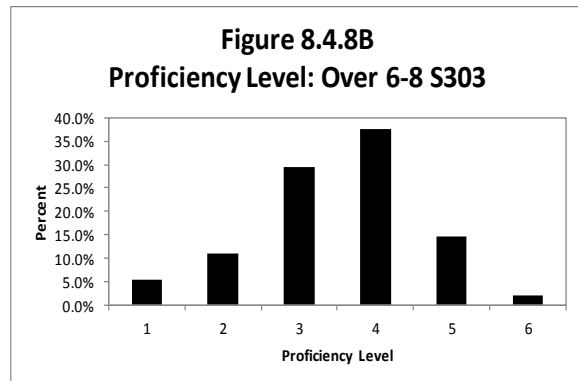
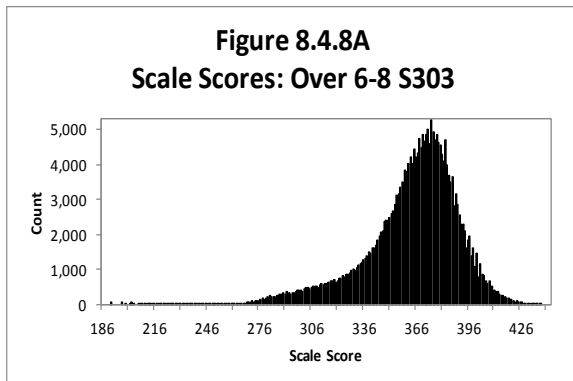


Table 8.4.8A
Scale Score Descriptive Statistics: Over 6-8 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	82,338	191	432	357.86	25.45
7	83,653	197	438	364.07	26.94
8	83,417	203	436	369.21	28.40
Total	249,408	191	438	363.74	27.36

Table 8.4.8B
Proficiency Level Distribution: Over 6-8 S303

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	3,101	3.8%	4,373	5.2%	5,705	6.8%	13,179	5.3%
2	8,473	10.3%	9,145	10.9%	10,068	12.1%	27,686	11.1%
3	23,852	29.0%	25,362	30.3%	24,617	29.5%	73,831	29.6%
4	32,843	39.9%	30,318	36.2%	30,293	36.3%	93,454	37.5%
5	12,148	14.8%	12,843	15.4%	11,361	13.6%	36,352	14.6%
6	1,921	2.3%	1,612	1.9%	1,373	1.6%	4,906	2.0%
Total	82,338	100.0%	83,653	100.0%	83,417	100.0%	249,408	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 S303

Component	Weight	Variance	Reliability
Listening	0.15	2259.054	0.664
Reading	0.35	791.689	0.775
Speaking	0.15	2614.955	0.914
Writing	0.35	548.659	0.915
Overall Composite		748.367	0.934

*Variances from students who had results in all four domains

Table 8.4.8E-1

Accuracy and Consistency of Classification Indices: Over (Grade 6) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.742	0.652		0.520	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.863		0.818	
	2	0.767		0.663	
	3	0.809		0.720	
	4	0.763		0.674	
	5	0.556		0.465	
	6	-		0.792	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.989	0.005	0.006	0.986
	2/3	0.963	0.020	0.017	0.948
	3/4	0.915	0.036	0.049	0.882
	4/5	0.893	0.047	0.060	0.851
5/6	0.977	0.023	0.000	0.978	

Table 8.4.8E-2

Accuracy and Consistency of Classification Indices: Over (Grade 7) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.724	0.631		0.502
Conditional on Level	Level	Accuracy		Consistency	
	1	0.876		0.830	
	2	0.753		0.646	
	3	0.806		0.715	
	4	0.718		0.620	
	5	0.553		0.461	
	6	-		0.842	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.986	0.006	0.008	0.981
	2/3	0.960	0.021	0.019	0.944
	3/4	0.910	0.038	0.052	0.875
	4/5	0.883	0.051	0.066	0.840
	5/6	0.981	0.019	0.000	0.982

Table 8.4.8E-3

Accuracy and Consistency of Classification Indices: Over (Grade 8) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.718	0.624		0.495
Conditional on Level	Level	Accuracy		Consistency	
	1	0.882		0.835	
	2	0.737		0.628	
	3	0.790		0.693	
	4	0.705		0.618	
	5	0.525		0.425	
	6	-		0.877	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.982	0.008	0.011	0.976
	2/3	0.955	0.024	0.021	0.937
	3/4	0.908	0.038	0.054	0.872
	4/5	0.884	0.061	0.055	0.844
	5/6	0.984	0.016	0.000	0.985

8.5 Grades: 9–12

8.5.1 Listening 9–12

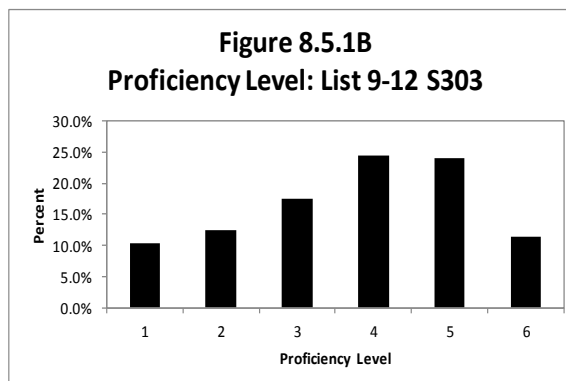
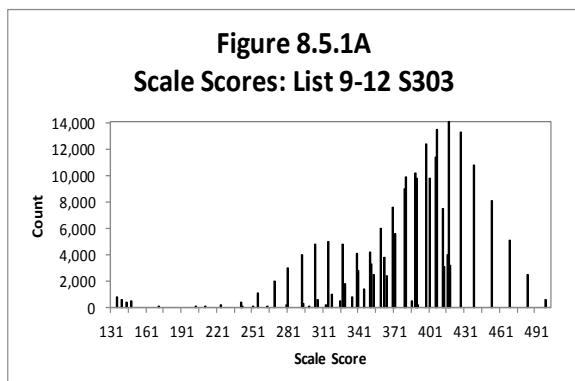


Table 8.5.1A
Scale Score Descriptive Statistics: List 9-12 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	97,451	136	499	378.95	53.17
10	60,572	140	499	385.92	48.69
11	43,795	144	499	391.42	47.03
12	33,331	148	499	393.56	47.30
Total	235,149	136	499	385.14	50.45

Table 8.5.1B
Proficiency Level Distribution: List 9-12 S303

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	11,405	11.7%	5,669	9.4%	3,881	8.9%	3,185	9.6%	24,140	10.3%
2	13,927	14.3%	7,068	11.7%	4,772	10.9%	3,368	10.1%	29,135	12.4%
3	13,575	13.9%	11,973	19.8%	7,393	16.9%	8,010	24.0%	40,951	17.4%
4	20,366	20.9%	14,787	24.4%	12,863	29.4%	9,243	27.7%	57,259	24.4%
5	27,956	28.7%	14,349	23.7%	9,233	21.1%	5,142	15.4%	56,680	24.1%
6	10,222	10.5%	6,726	11.1%	5,653	12.9%	4,383	13.1%	26,984	11.5%
Total	97,451	100.0%	60,572	100.0%	43,795	100.0%	33,331	100.0%	235,149	100.0%

Table 8.5.1C

Conditional Standard Error of Measurement at Cut Scores: List 9-12 S303*

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	9	312	20.66	21.04	22.17
	10	322	20.66	20.29	21.04
	11	332	21.04	19.91	19.91
	12	343	21.42	19.54	19.16
2/3	9	352	22.17	19.54	18.41
	10	358	22.92	19.54	18.41
	11	363	23.29	19.54	18.03
	12	366	23.67	19.54	18.03
3/4	9	381	25.92	19.54	18.03
	10	386	26.68	19.91	18.03
	11	389	27.05	19.91	18.03
	12	391	27.80	19.91	18.03
4/5	9	406	n/a	20.66	18.79
	10	412	n/a	21.42	19.16
	11	416	n/a	21.79	19.54
	12	418	n/a	21.79	19.54
5/6	9	432	n/a	n/a	21.04
	10	436	n/a	n/a	21.42
	11	438	n/a	n/a	21.79
	12	439	n/a	n/a	22.17

* No equating was performed for S303

Figure 8.5.1C
 Test Characteristic Curve: List 9-12ABC S303

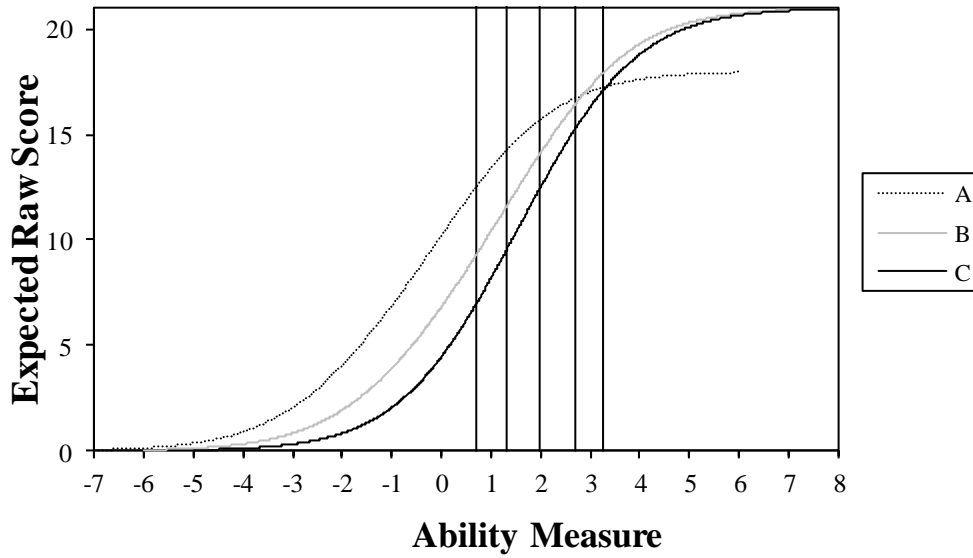


Figure 8.5.1D
 Test Information Function: List 9-12ABC S303

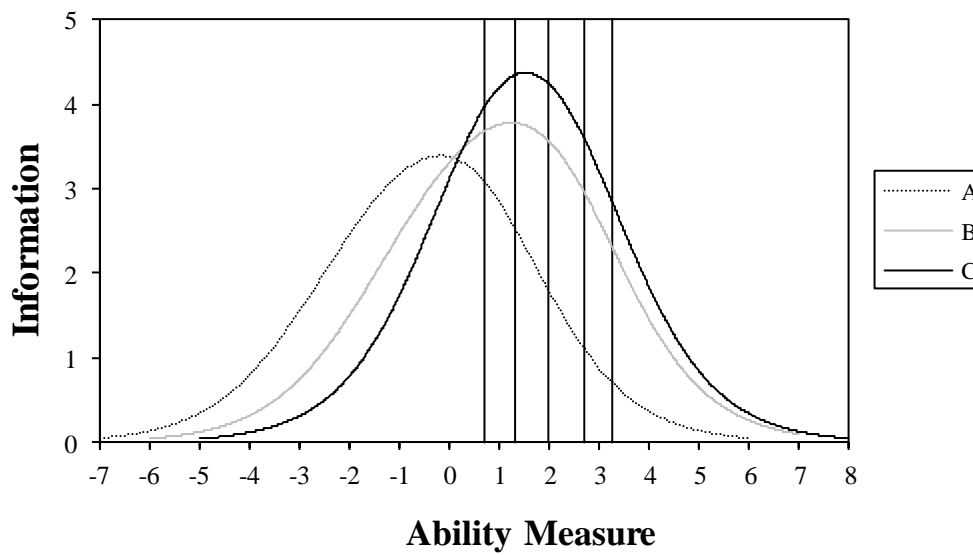


Table 8.5.1D

Weighted Reliability: List 9-12 S303

Tiers	No. of Students	Reliability	Reliability
A	38,940	0.667	0.694
B	83,198	0.732	
C	113,011	0.676	

Table 8.5.1E-1

Accuracy and Consistency of Classification Indices: List (Grade 9) S303

Overall Indices	Accuracy	Consistency			Kappa (k)
	0.412	0.316			0.163
Conditional on Level	Level	Accuracy			Consistency
	1	0.813			0.670
	2	0.489			0.323
	3	0.264			0.176
	4	0.282			0.251
	5	0.438			0.391
	6	-			0.166
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.945	0.019	0.036	0.922
	2/3	0.889	0.027	0.085	0.838
	3/4	0.819	0.058	0.123	0.720
	4/5	0.704	0.157	0.139	0.646
5/6	0.895	0.105	0.000	0.815	

Table 8.5.1E-2

Accuracy and Consistency of Classification Indices: List (Grade 10) S303

Overall Indices	Accuracy	Consistency			Kappa (k)
	0.362	0.307			0.142
Conditional on Level	Level	Accuracy			Consistency
	1	0.818			0.675
	2	0.459			0.280
	3	0.363			0.239
	4	0.298			0.286
	5	-			0.310
	6	-			0.162
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.959	0.015	0.026	0.940
	2/3	0.901	0.026	0.074	0.852
	3/4	0.786	0.044	0.170	0.691
	4/5	0.652	0.348	0.000	0.632
5/6	0.889	0.111	0.000	0.824	

Table 8.5.1E-3

Accuracy and Consistency of Classification Indices: List (Grade 11) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.391	0.306		0.132	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.813		0.672	
	2	0.454		0.262	
	3	0.321		0.203	
	4	0.349		0.338	
	5	-		0.267	
	6	-		0.178	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.962	0.015	0.023	0.944
	2/3	0.904	0.021	0.075	0.858
	3/4	0.799	0.038	0.163	0.699
	4/5	0.660	0.340	0.000	0.623
	5/6	0.871	0.129	0.000	0.806

Table 8.5.1E-4

Accuracy and Consistency of Classification Indices: List (Grade 12) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.373	0.292		0.115	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.829		0.675	
	2	0.372		0.185	
	3	0.376		0.267	
	4	0.323		0.313	
	5	-		0.186	
	6	-		0.166	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.961	0.015	0.024	0.940
	2/3	0.893	0.018	0.090	0.843
	3/4	0.714	0.024	0.263	0.633
	4/5	0.714	0.286	0.000	0.624
	5/6	0.869	0.132	0.000	0.805

8.5.2 Reading 9–12

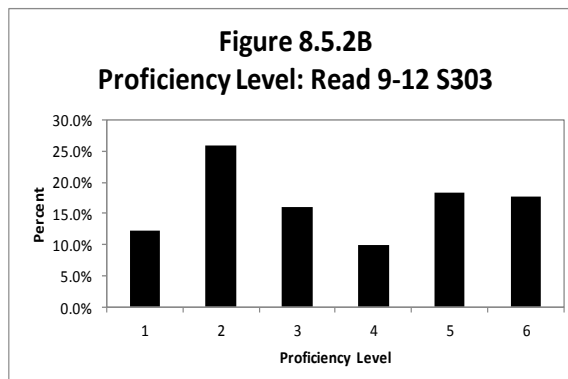
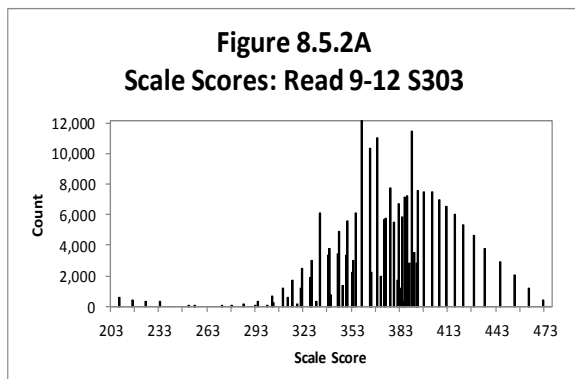


Table 8.5.2A
Scale Score Descriptive Statistics: Read 9-12 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	97,466	208	472	372.45	33.60
10	60,612	216	472	376.48	32.34
11	43,822	224	472	381.04	32.46
12	33,388	233	472	382.44	32.95
Total	235,288	208	472	376.51	33.21

Table 8.5.2B
Proficiency Level Distribution: Read 9-12 S303

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	12,420	12.7%	7,103	11.7%	5,145	11.7%	3,927	11.8%	28,595	12.2%
2	22,301	22.9%	17,778	29.3%	11,764	26.8%	9,048	27.1%	60,891	25.9%
3	20,036	20.6%	8,782	14.5%	4,993	11.4%	3,878	11.6%	37,689	16.0%
4	8,992	9.2%	5,513	9.1%	5,517	12.6%	3,230	9.7%	23,252	9.9%
5	16,129	16.5%	11,669	19.3%	7,860	17.9%	7,490	22.4%	43,148	18.3%
6	17,588	18.0%	9,767	16.1%	8,543	19.5%	5,815	17.4%	41,713	17.7%
Total	97,466	100.0%	60,612	100.0%	43,822	100.0%	33,388	100.0%	235,288	100.0%

Table 8.5.2C

Conditional Standard Error of Measurement at Cut Scores: Read 9-12 S303

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	9	336	11.70	12.22	15.08
	10	341	11.70	11.96	14.30
	11	346	11.70	11.70	13.26
	12	350	11.70	11.44	12.74
2/3	9	364	11.96	11.18	11.18
	10	370	12.48	11.18	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	11.44	10.40
	11	384	13.78	11.70	10.40
	12	385	14.04	11.70	10.40
4/5	9	387	n/a	11.70	10.40
	10	390	n/a	11.96	10.40
	11	392	n/a	11.96	10.40
	12	393	n/a	12.22	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	10.92

Figure 8.5.2C
 Test Characteristic Curve: Read 9-12ABC S303

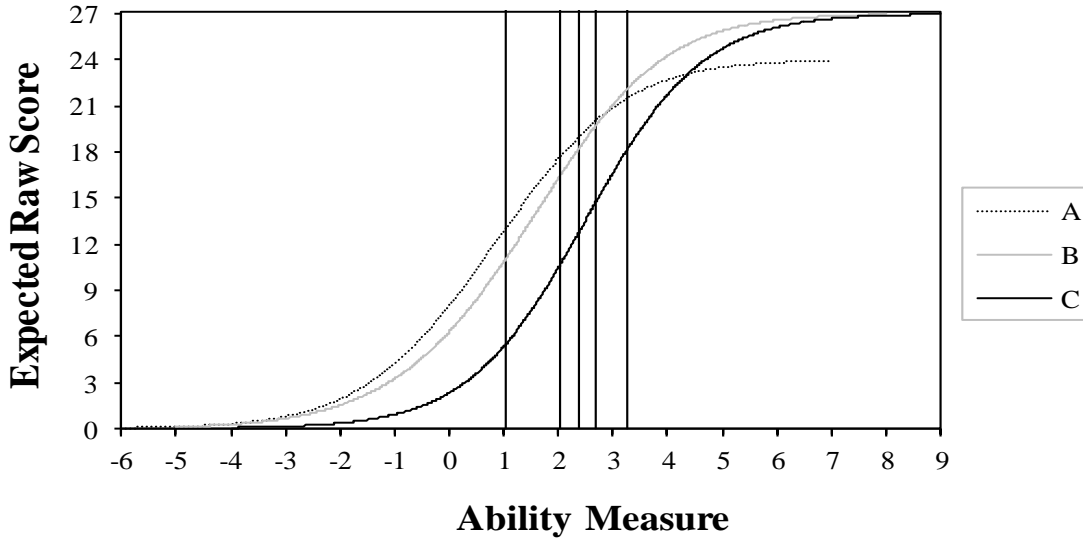


Figure 8.5.2D
 Test Information Function: Read 9-12ABC S303

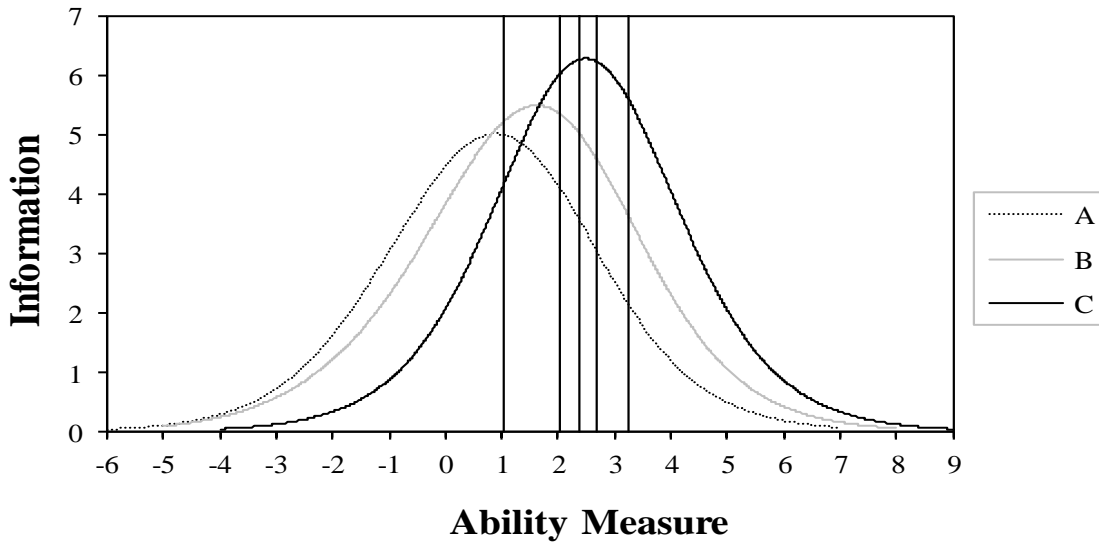


Table 8.5.2D

Weighted Reliability: Read 9-12 S303

Tiers	No. of Students	Reliability	Reliability
A	38,976	0.767	0.798
B	83,236	0.799	
C	113,076	0.808	

Table 8.5.2E-1

Accuracy and Consistency of Classification Indices: Read (Grade 9) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.501	0.408		0.278	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.771		0.614	
	2	0.576		0.452	
	3	0.405		0.318	
	4	0.189		0.144	
	5	0.354		0.277	
	6	0.689		0.522	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.935	0.026	0.039	0.904
	2/3	0.863	0.061	0.076	0.810
	3/4	0.838	0.079	0.083	0.780
	4/5	0.841	0.074	0.085	0.782
	5/6	0.879	0.073	0.049	0.828

Table 8.5.2E-2

Accuracy and Consistency of Classification Indices: Read (Grade 10) 303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.510	0.414		0.280	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.729		0.568	
	2	0.663		0.543	
	3	0.295		0.227	
	4	0.192		0.144	
	5	0.422		0.332	
	6	0.670		0.500	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.934	0.030	0.036	0.901
	2/3	0.854	0.056	0.090	0.800
	3/4	0.847	0.073	0.080	0.786
	4/5	0.849	0.076	0.075	0.789
	5/6	0.886	0.068	0.046	0.839

Table 8.5.2E-3

Accuracy and Consistency of Classification Indices: Read (Grade 11) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.503	0.411		0.280	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.730		0.565	
	2	0.619		0.498	
	3	0.232		0.179	
	4	0.253		0.194	
	5	0.378		0.294	
	6	0.704		0.549	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.934	0.030	0.037	0.900
	2/3	0.852	0.063	0.085	0.797
	3/4	0.839	0.079	0.082	0.780
	4/5	0.840	0.074	0.087	0.780
	5/6	0.879	0.068	0.054	0.826

Table 8.5.2E-4

Accuracy and Consistency of Classification Indices: Read (Grade 12) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.484	0.391		0.254	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.711		0.554	
	2	0.627		0.500	
	3	0.230		0.176	
	4	0.182		0.140	
	5	0.425		0.342	
	6	0.584		0.440	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.933	0.035	0.032	0.898
	2/3	0.845	0.059	0.096	0.792
	3/4	0.834	0.065	0.101	0.772
	4/5	0.829	0.077	0.094	0.764
	5/6	0.857	0.067	0.076	0.800

8.5.3 Writing 9–12

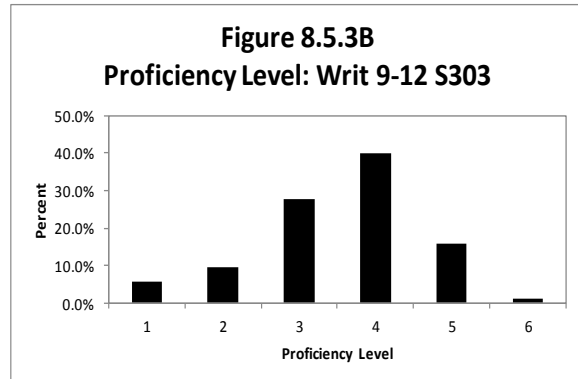
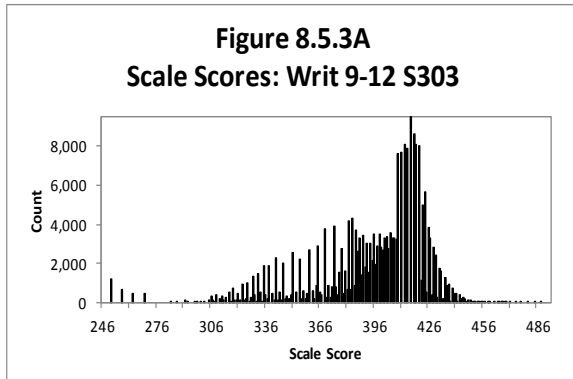


Table 8.5.3A
Scale Score Descriptive Statistics: Writ 9-12 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	97,308	251	477	390.84	34.71
10	60,482	257	485	394.89	31.01
11	43,769	263	488	398.66	29.07
12	33,287	269	469	400.26	29.00
Total	234,846	251	488	394.68	32.20

Table 8.5.3B
Proficiency Level Distribution: Writ 9-12 S303

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	5,215	5.4%	3,366	5.6%	2,480	5.7%	2,310	6.9%	13,371	5.7%
2	11,383	11.7%	4,906	8.1%	3,358	7.7%	2,331	7.0%	21,978	9.4%
3	19,820	20.4%	17,259	28.5%	14,545	33.2%	13,633	41.0%	65,257	27.8%
4	34,050	35.0%	26,277	43.4%	19,934	45.5%	13,665	41.1%	93,926	40.0%
5	24,625	25.3%	8,189	13.5%	3,283	7.5%	1,281	3.8%	37,378	15.9%
6	2,215	2.3%	485	0.8%	169	0.4%	67	0.2%	2,936	1.3%
Total	97,308	100.0%	60,482	100.0%	43,769	100.0%	33,287	100.0%	234,846	100.0%

Table 8.5.3C

Conditional Standard Error of Measurement at Cut Scores: Writ 9-12 S303

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	9	327	9.02	6.53	6.22
	10	336	10.26	7.15	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	8.40
3/4	9	389	11.19	8.09	8.09
	10	397	11.19	7.77	8.09
	11	404	10.88	7.77	7.77
	12	410	10.57	7.46	7.46
4/5	9	415	10.26	7.15	7.46
	10	422	9.95	6.84	7.15
	11	428	9.33	6.84	6.84
	12	434	9.33	6.53	6.53
5/6	9	435	9.33	6.53	6.53
	10	441	9.02	6.53	6.53
	11	447	9.02	6.53	6.22
	12	452	9.33	6.53	6.53

Figure 8.5.3C
 Test Characteristic Curve: Writ 9-12ABC S303

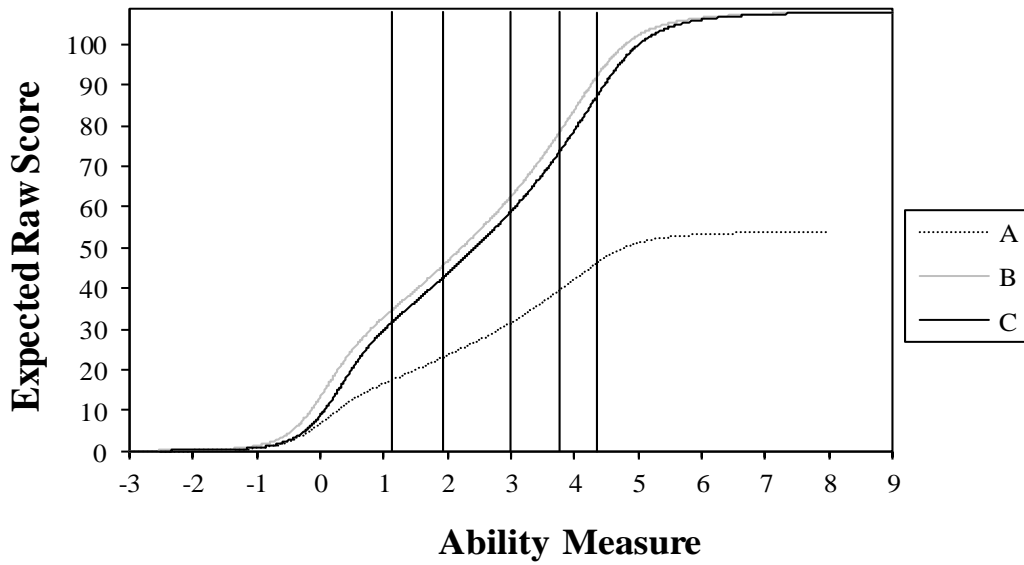


Figure 8.5.3D
 Test Information Function: Writ 9-12ABC S303

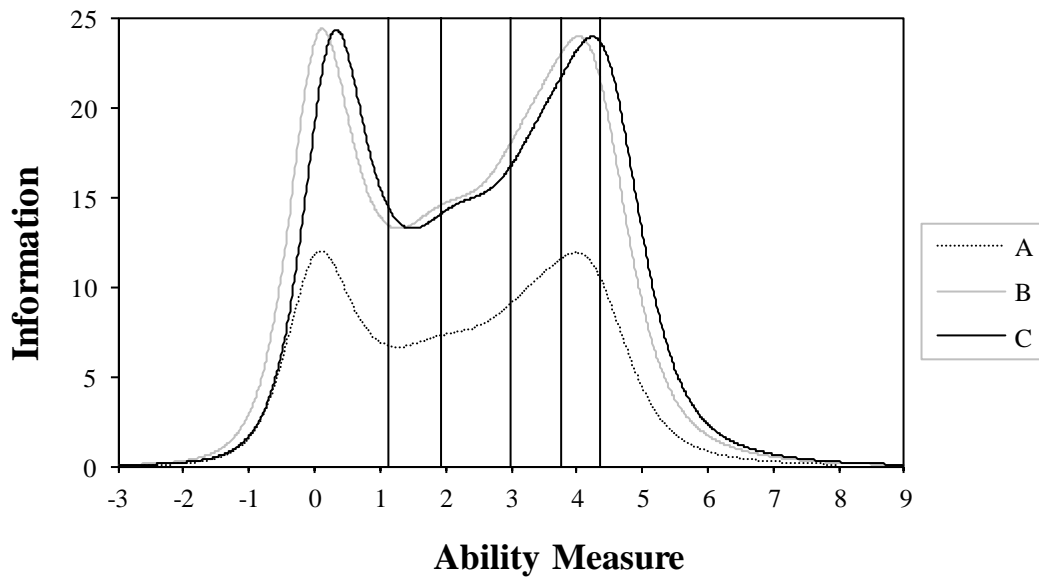


Table 8.5.3D

Weighted Reliability: Writ 9-12 S303

Tiers	No. of Students	Reliability	Reliability
A	38,895	0.869	0.914
B	83,126	0.935	
C	112,825	0.915	

Table 8.5.3E-1

Accuracy and Consistency of Classification Indices: Writ (Grade 9) S303

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.665	0.563	0.417		
Conditional on Level	Level	Accuracy		Consistency	
	1	0.819		0.718	
	2	0.774		0.663	
	3	0.729		0.613	
	4	0.636		0.529	
	5	0.588		0.517	
	6	-		0.064	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.982	0.010	0.008	0.973
	2/3	0.958	0.016	0.026	0.942
	3/4	0.928	0.027	0.045	0.898
	4/5	0.815	0.084	0.101	0.761
	5/6	0.977	0.023	0.000	0.967

Table 8.5.3E-2

Accuracy and Consistency of Classification Indices: Writ (Grade 10) S303

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.715	0.620	0.448		
Conditional on Level	Level	Accuracy		Consistency	
	1	0.879		0.804	
	2	0.707		0.580	
	3	0.819		0.705	
	4	0.659		0.636	
	5	-		0.296	
	6	-		0.036	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.986	0.007	0.007	0.980
	2/3	0.966	0.016	0.018	0.952
	3/4	0.906	0.027	0.067	0.868
	4/5	0.857	0.143	0.000	0.815
	5/6	0.992	0.008	0.000	0.991

Table 8.5.3E-3

Accuracy and Consistency of Classification Indices: Writ (Grade 11) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.752	0.661		0.479
Conditional on Level	Level	Accuracy		Consistency	
	1	0.887		0.820	
	2	0.712		0.585	
	3	0.831		0.686	
	4	0.710		0.682	
	5	-		0.161	
	6	-		0.583	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.987	0.006	0.007	0.981
	2/3	0.969	0.015	0.016	0.955
	3/4	0.876	0.029	0.096	0.823
	4/5	0.921	0.079	0.000	0.892
	5/6	0.996	0.004	0.000	0.996

Table 8.5.3E-4

Accuracy and Consistency of Classification Indices: Writ (Grade 12) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.709	0.622		0.411
Conditional on Level	Level	Accuracy		Consistency	
	1	0.903		0.844	
	2	0.679		0.547	
	3	0.814		0.621	
	4	0.641		0.610	
	5	-		0.074	
	6	-		-	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.986	0.007	0.008	0.979
	2/3	0.969	0.015	0.016	0.956
	3/4	0.794	0.036	0.170	0.722
	4/5	0.960	0.040	0.000	0.953
	5/6	0.998	0.002	0.000	0.998

8.5.4 Speaking 9–12

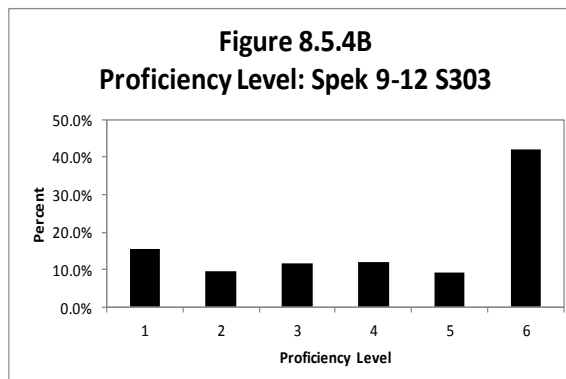
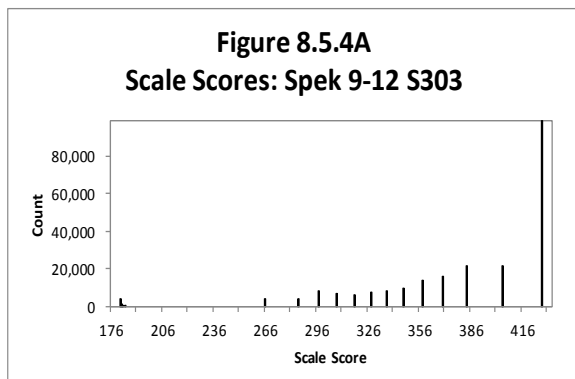


Table 8.5.4A
Scale Score Descriptive Statistics: Spek 9-12 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	97,313	181	428	374.51	62.52
10	60,446	182	428	380.69	54.20
11	43,721	183	428	385.95	50.92
12	33,307	184	428	390.26	48.64
Total	234,787	181	428	380.47	56.82

Table 8.5.4B
Proficiency Level Distribution: Spek 9-12 S303

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	19,939	20.5%	8,984	14.9%	4,847	11.1%	2,751	8.3%	36,521	15.6%
2	6,199	6.4%	7,365	12.2%	5,109	11.7%	3,587	10.8%	22,260	9.5%
3	8,612	8.8%	8,074	13.4%	5,949	13.6%	4,640	13.9%	27,275	11.6%
4	14,295	14.7%	5,714	9.5%	4,467	10.2%	3,403	10.2%	27,879	11.9%
5	8,947	9.2%	5,617	9.3%	4,129	9.4%	3,220	9.7%	21,913	9.3%
6	39,321	40.4%	24,692	40.8%	19,220	44.0%	15,706	47.2%	98,939	42.1%
Total	97,313	100.0%	60,446	100.0%	43,721	100.0%	33,307	100.0%	234,787	100.0%

Table 8.5.4C

Conditional Standard Error of Measurement at
Cut Scores: Spek 9-12 S303*

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

*No equating was performed for Series 303

Figure 8.5.4C
 Test Characteristic Curve: Spek 9-12 S303

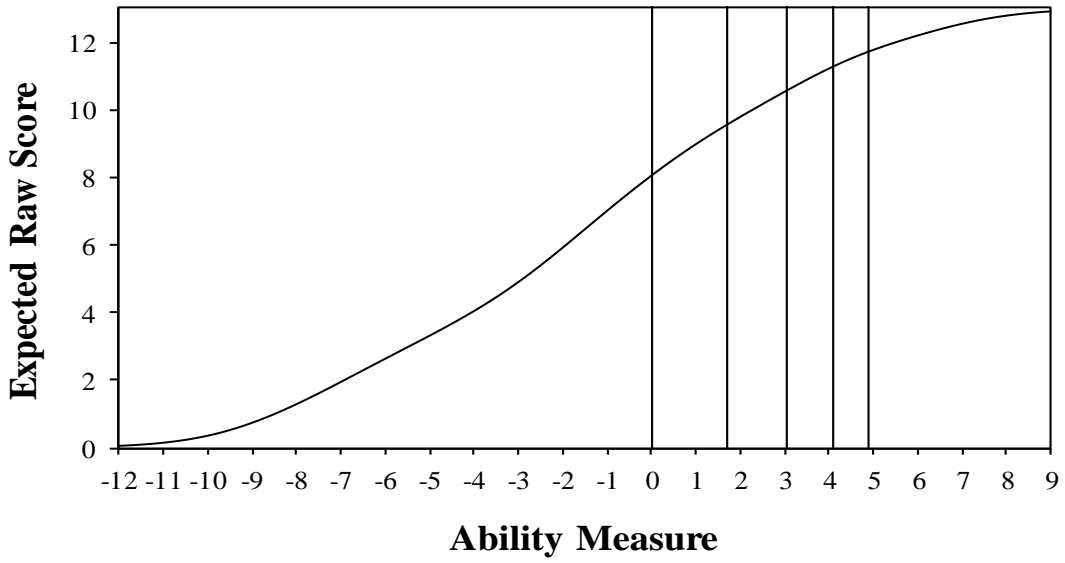


Figure 8.5.4D
 Test Information Function: Spek 9-12 S303

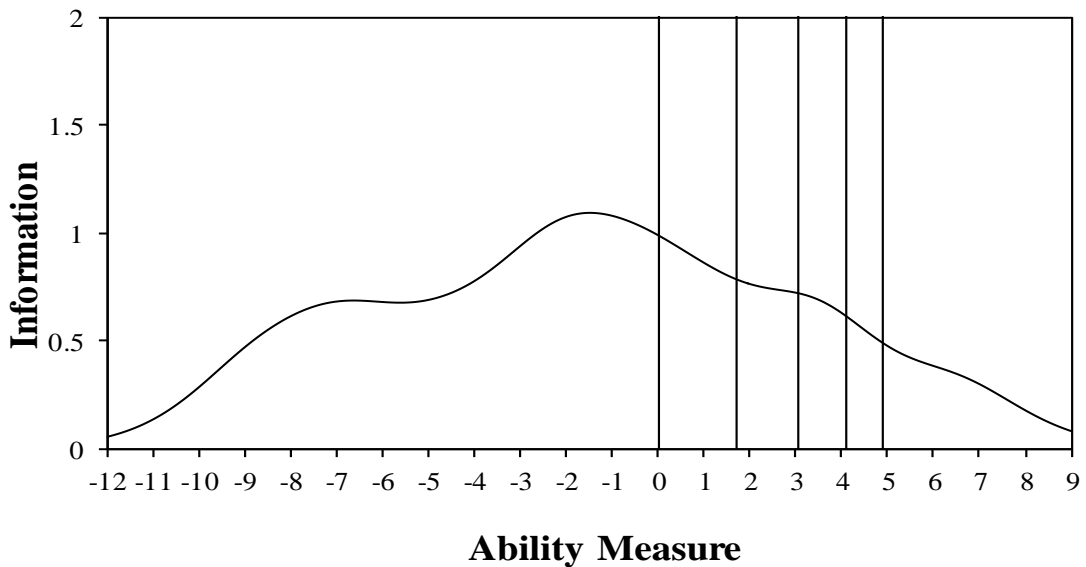


Table 8.5.4D

Reliability: Spek 9-12 S303

Tiers	No. of Students	Reliability
--	234,787	0.930

Table 8.5.4E-1

Accuracy and Consistency of Classification Indices: Spek (Grade 9) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.687	0.618		0.494	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.899		0.834	
	2	0.327		0.238	
	3	0.362		0.277	
	4	0.499		0.395	
	5	0.294		0.210	
	6	0.895		0.835	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.953	0.020	0.027	0.931
	2/3	0.936	0.037	0.027	0.909
	3/4	0.916	0.042	0.042	0.888
	4/5	0.912	0.029	0.059	0.878
5/6	0.913	0.045	0.042	0.868	

Table 8.5.4E-2

Accuracy and Consistency of Classification Indices: Spek (Grade 10) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.683	0.609		0.492	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.827		0.736	
	2	0.508		0.406	
	3	0.491		0.395	
	4	0.375		0.279	
	5	0.370		0.262	
	6	0.932		0.885	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.950	0.026	0.024	0.928
	2/3	0.919	0.040	0.041	0.893
	3/4	0.919	0.025	0.056	0.890
	4/5	0.939	0.027	0.034	0.906
5/6	0.920	0.054	0.026	0.888	

Table 8.5.4E-3

Accuracy and Consistency of Classification Indices: Spek (Grade 11) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.690	0.611		0.485	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.813		0.716	
	2	0.562		0.450	
	3	0.532		0.432	
	4	0.417		0.308	
	5	0.349		0.243	
	6	0.931		0.886	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.961	0.021	0.018	0.943
	2/3	0.930	0.033	0.037	0.907
	3/4	0.926	0.024	0.050	0.898
	4/5	0.939	0.027	0.034	0.906
	5/6	0.910	0.062	0.028	0.873

Table 8.5.4E-4

Accuracy and Consistency of Classification Indices: Spek (Grade 12) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.649	0.533		0.393	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.796		0.695	
	2	0.604		0.487	
	3	0.586		0.483	
	4	0.445		0.314	
	5	0.240		0.160	
	6	0.900		0.849	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.969	0.018	0.013	0.955
	2/3	0.940	0.027	0.033	0.919
	3/4	0.934	0.021	0.045	0.909
	4/5	0.942	0.022	0.036	0.905
	5/6	0.844	0.117	0.040	0.769

8.5.5 Oral Language Composite 9–12

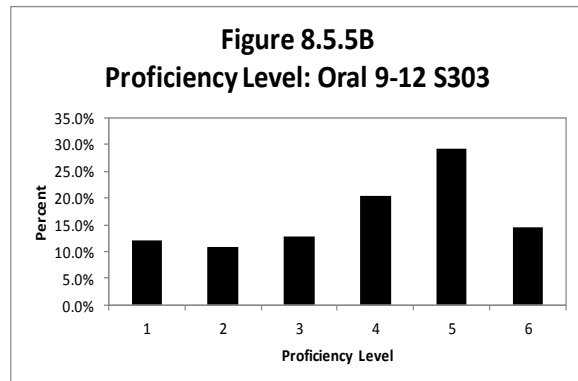
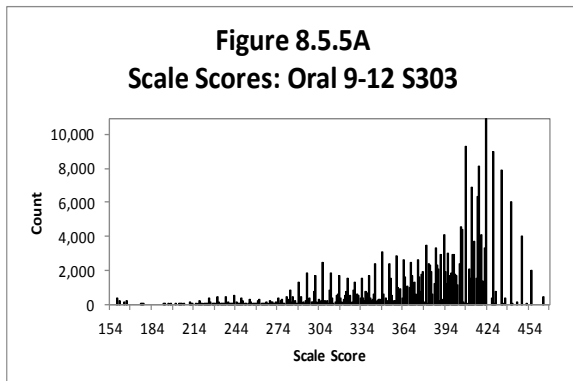


Table 8.5.5A
Scale Score Descriptive Statistics: Oral 9-12 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	96,920	159	464	377.02	53.09
10	60,193	161	464	383.59	46.06
11	43,497	164	464	389.03	43.12
12	33,055	166	464	392.24	41.60
Total	233,665	159	464	383.10	48.38

Table 8.5.5B
Proficiency Level Distribution: Oral 9-12 S303

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	15,121	15.6%	6,536	10.9%	3,833	8.8%	2,513	7.6%	28,003	12.0%
2	10,295	10.6%	7,218	12.0%	4,588	10.5%	3,407	10.3%	25,508	10.9%
3	9,036	9.3%	8,278	13.8%	6,717	15.4%	6,041	18.3%	30,072	12.9%
4	16,649	17.2%	12,210	20.3%	10,212	23.5%	8,352	25.3%	47,423	20.3%
5	29,412	30.3%	18,153	30.2%	11,799	27.1%	9,198	27.8%	68,562	29.3%
6	16,407	16.9%	7,798	13.0%	6,348	14.6%	3,544	10.7%	34,097	14.6%
Total	96,920	100.0%	60,193	100.0%	43,497	100.0%	33,055	100.0%	233,665	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 S303

Component	Weight	Variance	Reliability
Listening	0.50	2541.941	0.694
Speaking	0.50	3223.215	0.930
Oral		2339.037	0.893

*Variances from students who had results in all four domains

Table 8.5.5E-1

Accuracy and Consistency of Classification Indices: Oral (Grade 9) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.565	0.463		0.341	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.886		0.813	
	2	0.535		0.408	
	3	0.371		0.263	
	4	0.449		0.334	
	5	0.572		0.490	
6	0.540		0.423		
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.955	0.016	0.028	0.938
	2/3	0.941	0.024	0.036	0.914
	3/4	0.918	0.043	0.039	0.880
	4/5	0.868	0.065	0.067	0.818
5/6	0.846	0.065	0.089	0.798	

Table 8.5.5E-2

Accuracy and Consistency of Classification Indices: Oral (Grade 10) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.569	0.459		0.333	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.850		0.756	
	2	0.601		0.472	
	3	0.495		0.373	
	4	0.493		0.370	
	5	0.573		0.509	
	6	0.481		0.347	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.964	0.016	0.020	0.949
	2/3	0.936	0.027	0.037	0.911
	3/4	0.912	0.037	0.051	0.875
	4/5	0.867	0.061	0.071	0.813
	5/6	0.867	0.085	0.048	0.823

Table 8.5.5E-3

Accuracy and Consistency of Classification Indices: Oral (Grade 11) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.549	0.449		0.315	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.843		0.744	
	2	0.591		0.460	
	3	0.556		0.430	
	4	0.528		0.393	
	5	0.492		0.452	
	6	-		0.351	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.971	0.013	0.016	0.957
	2/3	0.940	0.027	0.033	0.916
	3/4	0.911	0.033	0.056	0.876
	4/5	0.857	0.056	0.087	0.793
	5/6	0.854	0.146	0.000	0.814

Table 8.5.5E-4

Accuracy and Consistency of Classification Indices: Oral (Grade 12) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.545	0.434		0.285
Conditional on Level	Level	Accuracy		Consistency	
	1	0.814		0.711	
	2	0.613		0.480	
	3	0.621		0.484	
	4	0.471		0.351	
	5	0.497		0.456	
	6	-		0.217	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.974	0.015	0.012	0.961
	2/3	0.941	0.025	0.034	0.920
	3/4	0.905	0.025	0.070	0.868
	4/5	0.809	0.065	0.126	0.730
	5/6	0.893	0.107	0.000	0.850

8.5.6 Literacy Composite 9–12

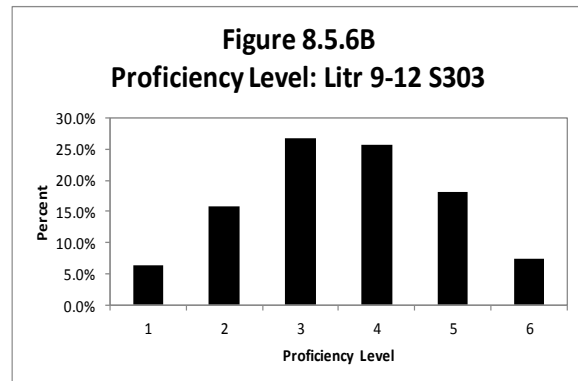
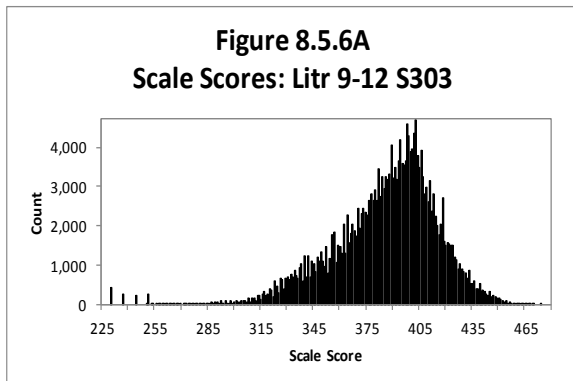


Table 8.5.6A
Scale Score Descriptive Statistics: Litr 9-12 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	97,128	230	470	381.96	31.60
10	60,359	237	474	386.00	28.86
11	43,671	244	469	390.15	27.88
12	33,207	251	466	391.66	27.97
Total	234,365	230	474	385.90	29.98

Table 8.5.6B
Proficiency Level Distribution: Litr 9-12 S303

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	6,718	6.9%	3,544	5.9%	2,487	5.7%	2,243	6.8%	14,992	6.4%
2	15,168	15.6%	9,646	16.0%	6,724	15.4%	5,551	16.7%	37,089	15.8%
3	22,829	23.5%	16,874	28.0%	12,702	29.1%	10,189	30.7%	62,594	26.7%
4	22,945	23.6%	15,817	26.2%	12,003	27.5%	9,159	27.6%	59,924	25.6%
5	20,972	21.6%	10,467	17.3%	6,739	15.4%	4,191	12.6%	42,369	18.1%
6	8,496	8.7%	4,011	6.6%	3,016	6.9%	1,874	5.6%	17,397	7.4%
Total	97,128	100.0%	60,359	100.0%	43,671	100.0%	33,207	100.0%	234,365	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 S303

Component	Weight	Variance	Reliability
Reading	0.50	1101.096	0.798
Writing	0.50	1032.202	0.914
Literacy		897.695	0.913

*Variances from students who had results in all four domains

Table 8.5.6E-1

Accuracy and Consistency of Classification Indices: Litr (Grade 9) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.639	0.533		0.419	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.829		0.722	
	2	0.744		0.629	
	3	0.667		0.555	
	4	0.594		0.475	
	5	0.554		0.484	
	6	0.627		0.413	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.975	0.012	0.013	0.964
	2/3	0.940	0.024	0.036	0.916
	3/4	0.905	0.045	0.049	0.867
	4/5	0.892	0.046	0.061	0.848
5/6	0.920	0.070	0.010	0.900	

Table 8.5.6E-2

Accuracy and Consistency of Classification Indices: Litr (Grade 10) S303

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.647	0.545	0.423		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.823	0.715		
	2	0.749	0.633		
	3	0.714	0.607		
	4	0.617	0.496		
	5	0.509	0.443		
	6	-	0.355		
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.979	0.010	0.011	0.969
	2/3	0.938	0.028	0.034	0.912
	3/4	0.897	0.044	0.060	0.857
	4/5	0.895	0.041	0.064	0.848
	5/6	0.934	0.066	0.000	0.925

Table 8.5.6E-3

Accuracy and Consistency of Classification Indices: Litr (Grade 11) S303

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.642	0.540	0.414		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.820	0.711		
	2	0.736	0.616		
	3	0.725	0.619		
	4	0.628	0.506		
	5	0.468	0.405		
	6	-	0.361		
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.979	0.010	0.010	0.970
	2/3	0.938	0.029	0.033	0.911
	3/4	0.894	0.043	0.063	0.854
	4/5	0.895	0.039	0.066	0.847
	5/6	0.931	0.069	0.000	0.924

Table 8.5.6E-4

Accuracy and Consistency of Classification Indices: Litr (Grade 12) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.621	0.524		0.387
Conditional on Level	Level	Accuracy		Consistency	
	1	0.817		0.716	
	2	0.743		0.623	
	3	0.735		0.619	
	4	0.553		0.460	
	5	0.390		0.327	
	6	-		0.231	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.977	0.013	0.010	0.966
	2/3	0.933	0.031	0.036	0.905
	3/4	0.885	0.035	0.080	0.842
	4/5	0.866	0.070	0.065	0.818
	5/6	0.944	0.056	0.000	0.940

8.5.7 Comprehension Composite 9–12

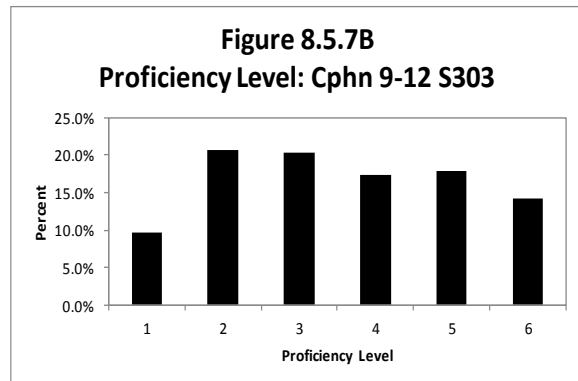
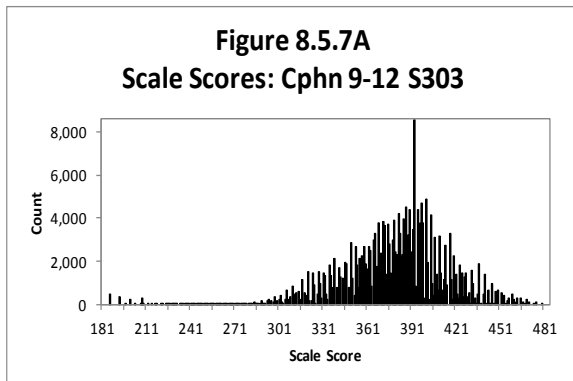


Table 8.5.7A
Scale Score Descriptive Statistics: Cphn 9-12 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	97,272	186	480	374.51	36.95
10	60,472	193	480	379.41	34.51
11	43,688	200	480	384.26	33.98
12	33,237	208	480	385.93	34.24
Total	234,669	186	480	379.21	35.70

Table 8.5.7B
Proficiency Level Distribution: Cphn 9-12 S303

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	11,025	11.3%	4,990	8.3%	3,552	8.1%	3,154	9.5%	22,721	9.7%
2	18,531	19.1%	13,405	22.2%	9,827	22.5%	6,699	20.2%	48,462	20.7%
3	20,803	21.4%	13,141	21.7%	7,657	17.5%	6,014	18.1%	47,615	20.3%
4	14,857	15.3%	10,583	17.5%	8,717	20.0%	6,435	19.4%	40,592	17.3%
5	18,264	18.8%	10,672	17.6%	7,362	16.9%	5,721	17.2%	42,019	17.9%
6	13,792	14.2%	7,681	12.7%	6,573	15.0%	5,214	15.7%	33,260	14.2%
Total	97,272	100.0%	60,472	100.0%	43,688	100.0%	33,237	100.0%	234,669	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 S303

Component	Weight	Variance	Reliability
Listening	0.30	2541.941	0.694
Reading	0.70	1101.096	0.798
Comprehension		1274.074	0.860

*Variances from students who had results in all four domains

Table 8.5.7E-1

Accuracy and Consistency of Classification Indices: Cphn (Grade 9) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.543	0.440		0.322	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.819		0.696	
	2	0.630		0.502	
	3	0.506		0.398	
	4	0.351		0.267	
	5	0.444		0.356	
	6	0.667		0.493	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.954	0.019	0.027	0.933
	2/3	0.902	0.043	0.055	0.862
	3/4	0.869	0.060	0.071	0.819
	4/5	0.867	0.060	0.074	0.814
5/6	0.896	0.066	0.038	0.856	

Table 8.5.7E-2

Accuracy and Consistency of Classification Indices: Cphn (Grade 10) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.529	0.428		0.303	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.761		0.619	
	2	0.686		0.558	
	3	0.495		0.391	
	4	0.387		0.295	
	5	0.419		0.342	
	6	0.649		0.452	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.961	0.020	0.019	0.941
	2/3	0.892	0.046	0.062	0.851
	3/4	0.863	0.055	0.082	0.813
	4/5	0.868	0.053	0.080	0.812
	5/6	0.895	0.080	0.026	0.862

Table 8.5.7E-3

Accuracy and Consistency of Classification Indices: Cphn (Grade 11) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.522	0.421		0.298	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.740		0.594	
	2	0.687		0.557	
	3	0.411		0.320	
	4	0.431		0.331	
	5	0.387		0.309	
	6	0.673		0.497	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.959	0.022	0.019	0.938
	2/3	0.889	0.045	0.065	0.847
	3/4	0.863	0.058	0.079	0.813
	4/5	0.861	0.048	0.090	0.806
	5/6	0.889	0.073	0.037	0.849

Table 8.5.7E-4

Accuracy and Consistency of Classification Indices: Cphn (Grade 12) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.458	0.392		0.266
Conditional on Level	Level	Accuracy		Consistency	
	1	0.753		0.621	
	2	0.651		0.520	
	3	0.426		0.329	
	4	0.386		0.290	
	5	0.343		0.299	
	6	-		0.411	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.957	0.025	0.018	0.934
	2/3	0.889	0.043	0.067	0.852
	3/4	0.860	0.044	0.097	0.810
	4/5	0.844	0.048	0.108	0.779
	5/6	0.843	0.157	0.000	0.819

8.5.8 Overall Composite 9–12

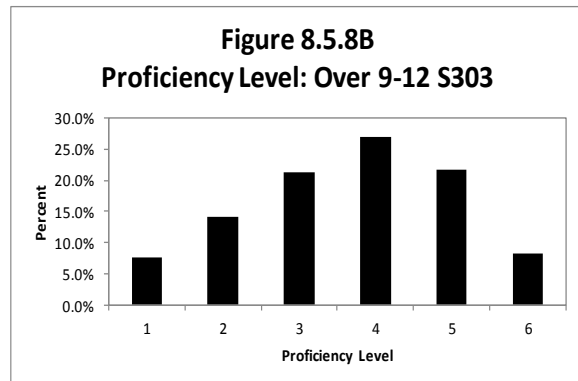
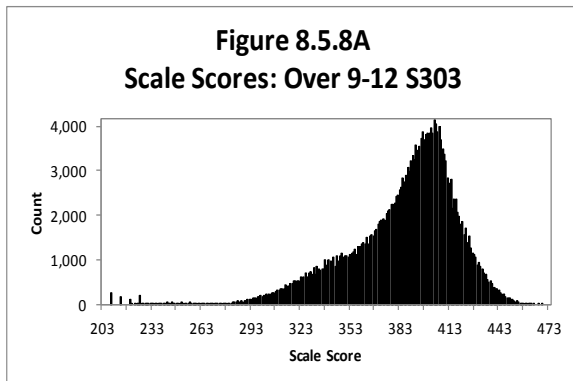


Table 8.5.8A
Scale Score Descriptive Statistics: Over 9-12 S303

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	96,620	208	464	380.30	36.16
10	59,990	214	469	385.09	31.90
11	43,351	220	464	389.65	30.20
12	32,903	226	463	391.69	29.62
Total	232,864	208	469	384.88	33.43

Table 8.5.8B
Proficiency Level Distribution: Over 9-12 S303

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	9,410	9.7%	3,832	6.4%	2,574	5.9%	1,967	6.0%	17,783	7.6%
2	13,712	14.2%	9,018	15.0%	5,929	13.7%	4,355	13.2%	33,014	14.2%
3	16,478	17.1%	13,359	22.3%	10,510	24.2%	9,260	28.1%	49,607	21.3%
4	22,810	23.6%	16,462	27.4%	13,004	30.0%	10,362	31.5%	62,638	26.9%
5	24,367	25.2%	13,001	21.7%	8,132	18.8%	5,073	15.4%	50,573	21.7%
6	9,843	10.2%	4,318	7.2%	3,202	7.4%	1,886	5.7%	19,249	8.3%
Total	96,620	100.0%	59,990	100.0%	43,351	100.0%	32,903	100.0%	232,864	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 S303

Component	Weight	Variance	Reliability
Listening	0.15	2541.941	0.694
Reading	0.35	1101.096	0.798
Speaking	0.15	3223.215	0.930
Writing	0.35	1032.202	0.914
Overall Composite		1117.700	0.946

*Variances from students who had results in all four domains

Table 8.5.8E-1

Accuracy and Consistency of Classification Indices: Over (Grade 9) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.699	0.595		0.501	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.889		0.823	
	2	0.770		0.672	
	3	0.659		0.544	
	4	0.684		0.569	
	5	0.653		0.566	
6	0.652		0.491		
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.976	0.010	0.014	0.966
	2/3	0.956	0.018	0.026	0.938
	3/4	0.932	0.037	0.031	0.903
	4/5	0.911	0.039	0.050	0.876
5/6	0.922	0.050	0.028	0.896	

Table 8.5.8E-2

Accuracy and Consistency of Classification Indices: Over (Grade 10) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.704	0.613		0.512	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.861		0.781	
	2	0.801		0.709	
	3	0.730		0.626	
	4	0.721		0.609	
	5	0.593		0.542	
	6	-		0.404	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.982	0.009	0.009	0.974
	2/3	0.955	0.020	0.025	0.936
	3/4	0.926	0.035	0.039	0.896
	4/5	0.912	0.033	0.054	0.876
	5/6	0.928	0.072	0.000	0.921

Table 8.5.8E-3

Accuracy and Consistency of Classification Indices: Over (Grade 11) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.697	0.608		0.501	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.857		0.776	
	2	0.781		0.680	
	3	0.754		0.654	
	4	0.738		0.626	
	5	0.541		0.489	
	6	-		0.406	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.983	0.009	0.008	0.976
	2/3	0.955	0.021	0.024	0.936
	3/4	0.922	0.034	0.043	0.891
	4/5	0.910	0.030	0.060	0.871
	5/6	0.926	0.074	0.000	0.922

Table 8.5.8E-4

Accuracy and Consistency of Classification Indices: Over (Grade 12) S303

Overall Indices	Accuracy	Consistency		Kappa (k)	
		0.671	0.584		0.458
Conditional on Level	Level	Accuracy		Consistency	
	1	0.844		0.769	
	2	0.773		0.666	
	3	0.807		0.714	
	4	0.657		0.539	
	5	0.444		0.393	
	6	-		0.351	
Indices at Cut Points	Cut Point	Accuracy			Consistency
		Accuracy	False Positives	False Negatives	
	1/2	0.984	0.010	0.007	0.976
	2/3	0.954	0.023	0.024	0.934
	3/4	0.918	0.025	0.056	0.888
	4/5	0.866	0.055	0.079	0.815
	5/6	0.943	0.057	0.000	0.942

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Bauman, J., Boals, T., Cranley, E., Gottlieb, M., & Kenyon, D.M. (2007). The Newly Developed English Language Tests (World-Class Instructional Design and Assessment – WIDA). In J. Abedi (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.
- Center for Applied Linguistics (2015). *ACCESS for ELLs Series 302 Media-Based Listening Field Test Technical Brief*. (WIDA Consortium).
- Chapelle, C. A., Enright, M. K., & Jamieson, J. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language*. London: Routledge
- Chapelle, C. A., Enright, M. K., & Jamieson, J. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice*, 29 (1), 3–13.
- 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., & 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.
- 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., & Fairbairn, S. (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., & 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. & Kenyon, D.M. (2006). *The Bridge Study between Tests of English Language Proficiency and ACCESS for ELLs®* (WIDA Consortium Technical Report No. 2).
- Kamata, A., Turhan, A., & Darandari, E. (2003, April). *Estimating reliability for multidimensional composite scale scores*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.
- Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practices*, 21(1), 31–41.
- Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, 42(4), 448–457.
- Kane, M., & Case, S.M. (2004). The reliability and validity of weighted composite scores. *Applied Measurement in Education*, 17(3), 221–240.
- Keng, L., Miller, G.E., O'Malley, K. & Turhan, A. (2008, March). *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., & 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., & Ryu, J.R. (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., Cho, B., & 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., & 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. (n.d). Displacement measures. Retrieved from <http://www.winsteps.com/winman/displacement.htm>.
- Linacre, J.M. (2002, Autumn). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16(2), 878. Retrieved April 10, 2006, from www.rasch.org/rmt/rmt162f.htm
- Linacre, J.M. (2006). Winsteps Rasch analysis (Version 3.60.1) [Computer Program]. Retrieved from 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.

- Llosa, L. (2008). Building and supporting a validity argument for a standards-based classroom assessment of English proficiency based on teacher judgments. *Educational Measurement: Issues and Practice*, 27(3), 32–42.
- MacGregor, D., Kenyon, D.M., Gibson, S., & Evans, E. (2009). *Development and Field Test of Kindergarten ACCESS for ELLs®*. (WIDA Consortium).
- MacGregor, D., Louguit, M., Huang, X., & 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., Kenyon, D.M., & 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., Yanosky, T., Fidelman, C. G., Pan, M., Huang, X., & 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).
- 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.
- National Research Council. (2011). *Allocating federal funds for state programs for English language learners*. 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.
- Rudner, L. (2001, Spring). Informed test component weighting. *Educational Measurement: Issues and Practice*, 20(1), 16–19.
- Toulmin, S. E. (2003). *The uses of argument* (Updated ed.). Cambridge: Cambridge University Press.
- 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.
- 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.
- Wolf, M., Kao, J., Griffin, N., Herman, J., Bachman, P., Chang, S., & Farnsworth, T. (2008). *Issues In Assessing English Language Learners: English Language Proficiency Measures And Accommodation Uses—Practice Review (Part 2 Of 3)* (CRESST Report 732). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.
- Wright, B.D., & Stone, M.H. (1979). *Best test design: Rasch measurement*. Chicago, IL: MESA Press.
- Yanosky, T., Amos, M., Cameron, C., Louguit, M., MacGregor, D., Yen, S., & 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., . . . 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., & 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* (pp. 337–347). 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.S.

Keira Ballantyne, Ph.D.

Marisa Gomez, M.A.

Daniel Lee, M.S.

Mohammed Louguit, Ph.D.

Dorry M. Kenyon, Ph.D.

David MacGregor, Ph.D.

James Marcus, Ph.D.

Jennifer Renn, Ph.D.

Yu-Chia Wu, M.Ed.

Shu Jing Yen, Ph.D.

Xin Yu, M.A.