World-Class Instructional Design and Assessment

Annual Technical Report for ACCESS for ELLs® 2.0 Paper English Language Proficiency Test, Series 400, 2015–2016 Administration

Annual Technical Report No. 12B

Prepared by:

Center for Applied Linguistics

Language Assessment Division
Psychometrics and Quantitative Research Team

May 2017

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, Graduate School of Education, 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, University of Illinois at Chicago.

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

Executive Summary

This is the 12th annual technical report on the ACCESS for ELLs® English Language Proficiency Test, and the first report on the ACCESS for ELLs 2.0 assessment. ACCESS for ELLs 2.0 measures the same constructs and uses the same scale as ACCESS for ELLs, but for the first time, the assessment is offered in an online, multi-stage adaptive format.

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 is intended 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 Consortium, 2012). Results on ACCESS for ELLs are used by WIDA Consortium states for monitoring the progress of students, for making decisions about exiting students from language support services, and for accountability.

ACCESS for ELLs 2.0 Series 400 was administered in school year 2015–16 in 36 states, the District of Columbia, and the Commonwealth of the Northern Marianas, for a total of 38 state entities (henceforth "states"). ACCESS for ELLs 2.0 Series 400 was offered in two administrative formats, an online format (grades 1–12) and a paper format (kindergarten–grade 12). Table 0.1 summarizes the numbers of students, by state, who participated in the grades 1–12 assessment online, in the grades 1–12 assessment on paper, the total number of students who participated in the grades 1–12 assessment, the total number who participated in the Kindergarten assessment (only offered in the paper format), and the total participants in ACCESS K–12. The current report (WIDA ACCESS Technical Report 12B) provides technical information pertaining to ACCESS for ELLs 2.0 Series 400 Paper, including the Kindergarten assessment. A second report (WIDA ACCESS Technical Report 12A) provides technical information for the ACCESS for ELLs Series 2.0 Series 400 Online assessment.

Table 0.1

Participation in ACCESS for ELLs Online and Paper, Series 400

State	Participants in ACCESS for ELLs Grades 1–12			Participants in Kindergarten	Total Participants in ACCESS for ELLs Grades K–12
	Participants in ACCESS for ELLs Online	Participants in ACCESS for ELLs Paper	Total Participants in ACCESS for ELLs		
AK	9,696	3,266	12,962	1,450	14,412
AL	12,112	4,330	16,442	3,641	20,083
CO	63,313	28,860	92,173	11,155	103,328
DC	5,498	82	5,580	1,023	6,603
DE	8,341	18	8,359	1,922	10,281
FL	0	224,490	224,490	34,806	259,296
GA	69,114	16,255	85,369	17,236	102,605
HI	0	11,746	11,746	1,992	13,738
ID	11,498	39	11,537	2,274	13,811
IL	119,961	41,230	161,191	27,203	188,394
IN	44,981	996	45,977	7,346	53,323
KY	18,378	541	18,919	3,227	22,146
MA	33,221	38,819	72,040	10,187	82,227
MD	54,350	235	54,585	10,305	64,890
ME	3,782	885	4,667	463	5,130
MI	76,134	5,787	81,921	10,326	92,247
MN	59,449	904	60,353	8,349	68,702
MO	25,185	122	25,307	4,736	30,043
MP	1,094	0	1,094	44	1,138
MT	2,470	11	2,481	150	2,631
NC	81,695	1,463	83,158	12,664	95,822
ND	2,698	80	2,778	419	3,197
NH	3,007	623	3,630	459	4,089
NJ	55,397	1,832	57,229	11,990	69,219
NM	40,236	3,947	44,183	5,453	49,636
NV	68,505	23	68,528	8,000	76,528
OK	17,713	20,114	37,827	7,534	45,361
PA	37,036	11,887	48,923	4,898	53,821
RI	5,646	1,864	7,510	1,199	8,709
SC	31,864	6,882	38,746	3,999	42,745
SD	3,400	246	3,646	748	4,394
TN	35,935	17	35,952	5,507	41,459
UT	33,081	174	33,255	5,028	38,283
VA	67,987	17,390	85,377	13,857	99,234
VI	743	0	743	124	867
VT	1,243	15	1,258	180	1,438
WI	41,378	215	41,593	5,601	47,194
WY	2,196	163	2,359	425	2,784
Total	1,148,337	445,551	1,593,888	245,920	1,839,808

This report follows the same structure as the ACCESS 1.0 technical reports. The report first provides background to the test (Chapter 1), followed by 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 test forms that constitute ACCESS for ELLs 2.0 (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 ACCESS for ELLs 2.0 Series 400 Paper, which cannot be fully summarized 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 and a separate table highlighting conditional standard errors around the cut scores. 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.

Launch of ACCESS 2.0

Series 400 Paper is the first series of the ACCESS 2.0 assessment. ACCESS 2.0 is now offered in two formats. The Paper format is available for grades K–12, and the Online format is available for grades 1–12. The cluster structure of ACCESS 2.0 Paper has been updated from the cluster structure of ACCESS 1.0. ACCESS for ELLs Paper Series 400 has the following grade-level clusters: K, 1, 2, 3, 4–5, 6–8, 9–12.

The Listening and Reading assessments were developed within the scope of ACCESS 1.0. The majority of Listening and Reading forms are the same forms as were used on ACCESS for ELLs Series 302. Writing and Speaking were developed as a part of ACCESS 2.0, and the Writing and Speaking tasks are paperized versions of the Online Writing and Speaking tasks.

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.

Demographic data

The Series 400 Paper data set for analyses included the results of 691,471 students. The largest grade was Kindergarten with 245,920 students, while the smallest was Grade 12 with 10,563 students. Of the participating WIDA states, the largest was Florida with 259,296 students, while the smallest was Northern Mariana Islands with 44 students.

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.

Results indicate that the reliability (stratified Cronbach's alpha, see 7.2.6 in Volume 3) of the Overall Composite score for Series 400 Paper, presented in Chapter 8 Table D, is very high across all grade-level clusters. For Kindergarten it was .974; for Grade 1, .932; for Grade 2, .937, for Grade 3, .939, for Grades 4–5, .939; for Grades 6–8, .937; and for Grades 9–12, .943. Likewise, as Table 0.1 shows, the accuracy of classification for 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 0.2
Accuracy of Classification 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)	0.975	0.957	0.949	0.955	0.953
K (accountability)	0.952	0.952	0.961	0.959	0.991
1	0.978	0.936	0.909	0.952	0.988
2	0.978	0.956	0.905	0.887	0.984
3	0.991	0.974	0.938	0.907	0.906
4	0.992	0.980	0.942	0.905	0.885
5	0.989	0.974	0.935	0.896	0.911
6	0.986	0.966	0.924	0.858	0.982
7	0.980	0.957	0.912	0.878	0.988
8	0.976	0.950	0.908	0.894	0.994
9	0.972	0.954	0.931	0.909	0.925
10	0.975	0.951	0.927	0.917	0.940
11	0.978	0.952	0.923	0.913	0.926
12	0.982	0.956	0.921	0.870	0.944

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, 2004–05). This is the twelfth technical report.

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

Volume I contains 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 400 Paper (2015–2016) 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, 4–5, 6–8, 9–12); then by
- Domain (Listening, Reading, Writing, and Speaking, abbreviated List, Read, Writ, and Spek, respectively); then by
- Tier (A, B, C)

Thus, all of the results for Kindergarten are presented before the results for Grade 1, and all of the results for Grade 1 Listening are presented before results for Grade 1 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, 4–5, 6–8, 9–12); then by
- 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

Technical Reports

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). 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).

Center for Applied Linguistics (2016). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test, Series 303, 2014–2015 Administration* (WIDA Consortium Annual Technical Report No. 11).

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. (2012). *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. (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.

Volume 1

1. Description of ACCESS for ELLs English Language Proficiency Test

1.1 Purpose of ACCESS for ELLs

The overarching purpose of ACCESS for ELLs 2.0 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 (Gottlieb, 2004; WIDA Consortium, 2007), then in the amplified 2012 English Language Development (ELD) Standards (WIDA Consortium, 2012). The WIDA ELD Standards, which correspond to the academic language identified in state academic content standards, describe six levels of developing English language proficiency and form the core of the WIDA Consortium's approach to instructing and testing ELLs. ACCESS 2.0 may thus be described as a standards-based English language proficiency test designed to measure the social and academic language proficiency of ELLs in English. It assesses social and instructional English as well as the academic language associated with language arts, mathematics, science, and social studies within the school context across the four language domains (Listening, Reading, Writing, and Speaking).

Other major purposes of ACCESS 2.0 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.

ACCESS 2.0 is offered in two formats: ACCESS 2.0 Paper, described in this report, and ACCESS 2.0 Online, described in a companion report.

1.2 Format of ACCESS 2.0 Paper

1.2.1 Integration with the Standards

The original ACCESS test design, 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: SIL
- Language of English Language Arts: LoLA
- Language of Math: LoMA
- Language of Science: LoSC
- Language of Social Studies: LoSS

Every selected response item and every performance-based task on ACCESS for ELLs targets at least one of these five Standards. In the case of some test items and tasks, the standards are combined as follows:

- Integrated Language of Science (LoSC), Language of Language Arts (LoLA), and Language of Social Studies (LoSS): IT
- Language of Math (LoMA) and Language of Science (LoSC): MS
- Language of English Language Arts (LoLA) and Language of Social Studies (LoSS): LS

1.2.2 Grade-Level Clusters

The grade-level cluster structure for ACCESS 2.0 Paper is as follows: K, 1, 2, 3, 4–5, 6–8, 9–12.

In the lower grades (grades 1–5), test forms may be shared across grade-level clusters. As described in Sections 1.3.3 and 1.3.4 below, the development of the Listening and Reading tests was conducted as part of ACCESS 1.0, which has a cluster structure that differs from that of

ACCESS 2.0 in the lower grades. The Speaking and Writing tests were developed using the ACCESS 2.0 Online cluster structure. ACCESS 2.0 Paper clusters, therefore, bridge the cluster structure of ACCESS 1.0 and ACCESS 2.0 Online. For example, the Cluster 2 tests in the domains of Reading and Listening are the same test forms as the Cluster 1 tests. The Cluster 2 tests in the domains of Speaking and Writing are the same test forms as the Cluster 3 tests in these domains. Table 1.2.2A details the grade-level cluster structure of ACCESS2.0 Paper and the shared forms across clusters.

Table 1.2.2A
ACCESS 2.0 Paper Grade-level Clusters and Shared Forms Across Clusters

ACCESS 2.0 Paper Grade-level Clusters	Shared Test Forms (Listening and Reading)	Shared Test Forms (Speaking and Writing)	Grade
K	K	K	K
1	Cluster 1 and Cluster 2	Cluster 1	1
2		Cluster 2 and Cluster 3	2
3	Cluster 3 and Cluster 4–5		3
4–5		Cluster 4–5	4
			5
6–8	Cluster 6–8	Cluster 6–8	6
			7
			8
9–12	Cluster 9–12	Cluster 9–12	9
			10
			11
			12

1.2.3 Language Domains

The WIDA ELD Standards describe developing English language proficiency for each of the four language domains: Listening, Reading, Writing, and Speaking. Thus, ACCESS 2.0 Paper contains four sections, each assessing an individual language domain.

1.2.4 Language Proficiency Levels

The WIDA ELD Standards document fully delineates the continuum of language development via five language proficiency levels (PLs) that are fully delineated in the WIDA ELD Standards document (WIDA, 2012), with scores indicating progression through each level. These levels are *Entering*, *Emerging*, *Developing*, *Expanding*, and *Bridging*. There is also a final stage known as *Reaching*, which is used to describe students who have progressed across the entire WIDA English language proficiency continuum; as such, scores do not indicate progression through this level. The proficiency levels are shown graphically in Figure 1.2.4A.

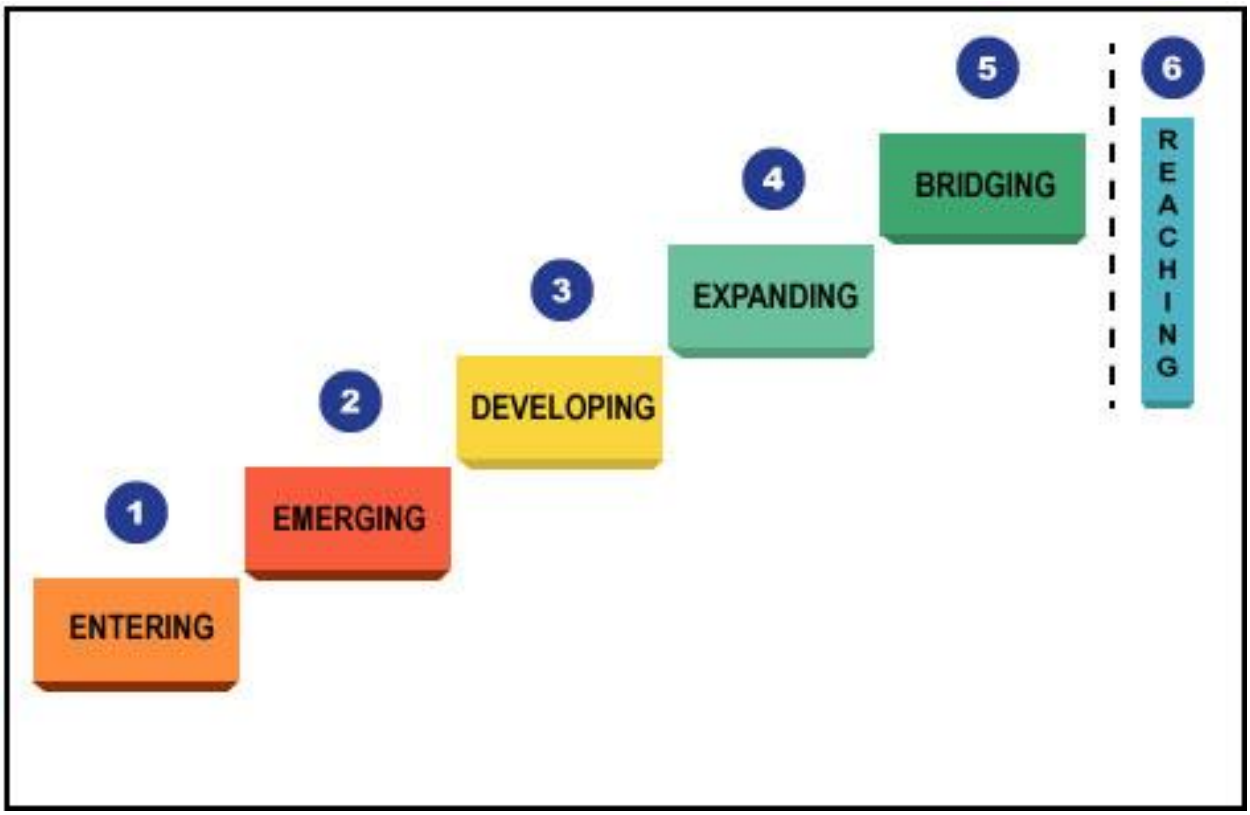


Figure 1.2.4A. The language proficiency levels of the WIDA ELD Standards.

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** (MPIs; see below) for each language proficiency level. The performance definitions are based on three criteria: (a) vocabulary usage at the word/phrase level; (b) language forms and conventions at the sentence level; and (c) linguistic complexity at the discourse level. Vocabulary usage refers to students' increasing comprehension and production of the technical language required for success in the academic content areas. Language forms and conventions refers to the increasing development of phonological, syntactic, and semantic understanding in receptive skills or control of usage in productive language skills. Linguistic complexity refers to students' demonstration of oral interaction or writing of increasing quantity and variety.

Second, the language proficiency levels of the WIDA ELD Standards are fully embedded in the accompanying MPIs, which exemplify the Standards. The MPIs describe the expectations for ELL students in each of the five Standards**,** at five[1] different grade-level clusters, across the four

[1] The grade-level clusters are 1, 2–3, 4–5, 6–8, and 9–12. The Listening and Reading tests are based on MPIs aligned with the ACCESS 1.0 cluster structure. See Section 1.3 for further detail.

language domains. That is, an MPI at each of the five language proficiency levels can be found within each combination of Standard, grade-level cluster, and language. *Reaching* (PL 6), represents the end of the continuum rather than another level of language proficiency. The sequence of these five MPIs 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 MPIs in logical progression is called a "strand."

ACCESS 2.0 is based on individual MPIs organized into strands within the WIDA ELD Standards.[2] Each selected-response item or performance-based task on ACCESS for ELLs is carefully developed, reviewed, piloted, and field tested to ensure that it allows students to demonstrate accomplishment of the targeted MPI.

1.2.5 Tiers

Tests must be at the appropriate difficulty level for individual test takers in order to be valid and reliable. As one might expect, test items and tasks that allow *Entering* (PL 1) or *Emerging* (PL 2) students to demonstrate accomplishment of the MPIs at their proficiency level will not allow *Expanding* (PL 4) or *Bridging* (PL 5) students to demonstrate the full extent of their language proficiency. Likewise, items and tasks that allow *Expanding* (PL 4) and *Bridging* (PL 5) students to demonstrate accomplishment of the MPIs at their level would be far too challenging for *Entering* (PL 1) or *Emerging* (PL 2) students. Items that are far too easy for 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 language proficiency.

In order to make ACCESS 2.0 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.

[2] The ELD Standards, the MPIs, and sample items are available at the WIDA website, www.wida.us.

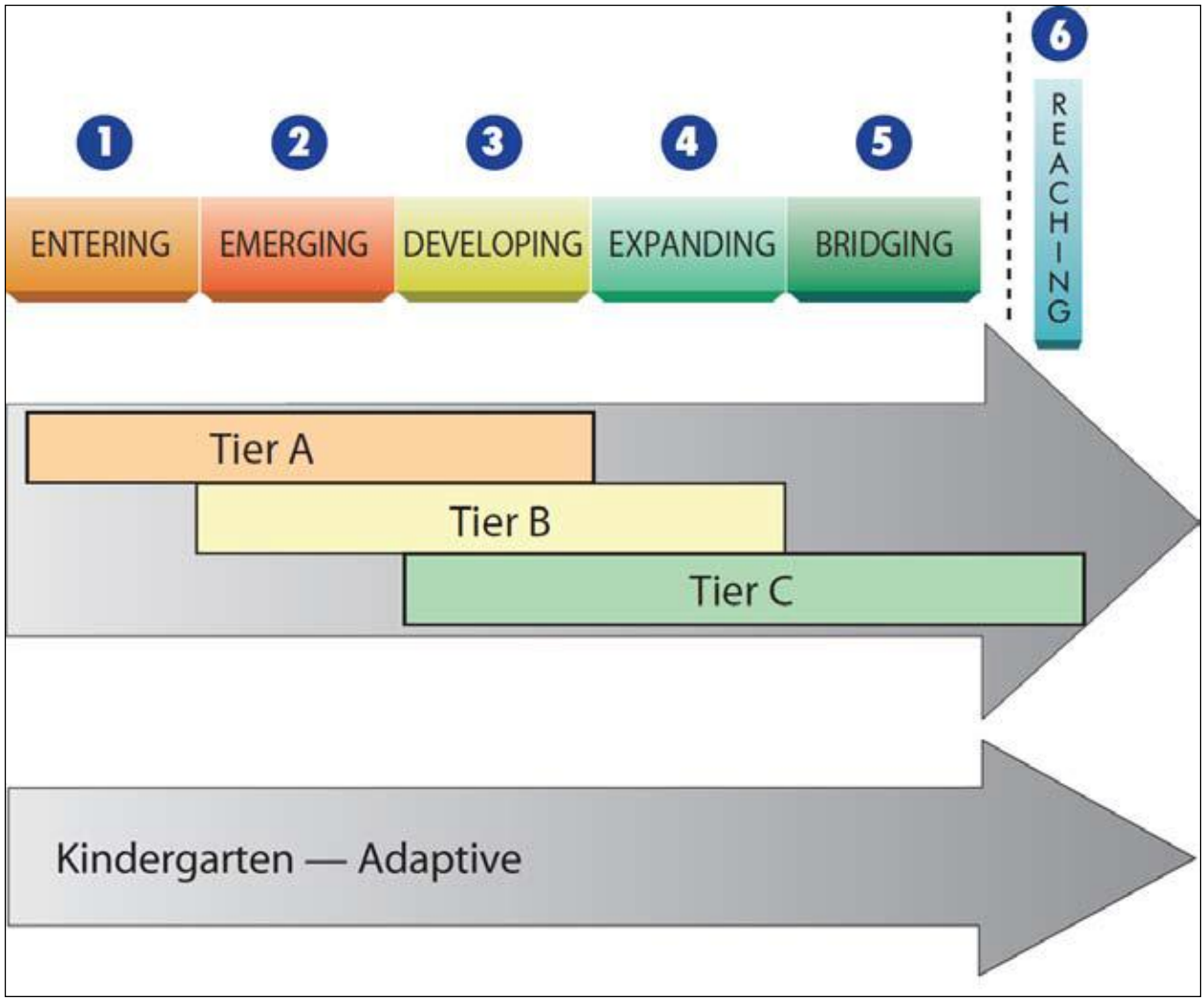


Figure 1.2.5.A.2.5A. Tier structure of ACCESS for ELLs

Each grade 1–12 test-taker takes either the Tier A, Tier B, or Tier C form of the assessment. The Kindergarten assessment is not tiered.

In ACCESS 2.0 Paper, the Listening and Reading tests have three forms (that is, one at each tier) for each grade-level cluster. Tier A has items and tasks designed to allow students at the lowest language proficiency levels (PLs 1 and 2) to meet the WIDA ELD Standards at their language proficiency levels, and it includes some items targeted to Level 3. Likewise, Tier C has items and tasks designed to allow students at the highest language proficiency levels (PLs 4 and 5) to meet the WIDA ELD Standards at their language proficiency levels, while also containing some items targeted to Level 3. (Note that, in order to assure that students are accurately measured to PL 6, Tier C also includes some items that are slightly more difficult than PL 5 items.) 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 (PL 2) and C (PL 4), as well as tasks from PL 3. This overlap of tiers ensures that all of the proficiency levels are assessed across the assessment as a whole; however,

each test booklet need not contain an unduly large number of test items. The overlap also ensures that the entire language proficiency range is covered. Thus, a test booklet at any given tier is primarily composed of items and tasks that span three targeted language proficiency levels.

In the domains of Writing and Speaking, for each grade-level cluster, there are two forms: a Tier A form, and a shared Tier B and Tier C form. The Tier A form of the Writing test has items targeting PLs 1, 2, and 3. The Writing test form that is shared by Tier B and Tier C has items targeting PLs 4 and 5. The Tier A form of the Speaking test has item targeting PLs 1 and 3, and the test form that is shared by Tier B and Tier C has items targeting PLs 3 and 5.

1.3 Test Development

Development of ACCESS 2.0 Series 400 Paper marked the transition point from the original ACCESS testing program, which was entirely paper-based, to the launch of ACCESS 2.0, which is offered both in Online and Paper formats. Development for ACCESS 2.0 Series 400 Paper reflects this transition. The Listening and Reading tests for ACCESS 2.0 Series 400 Paper were developed under the framework of ACCESS, while the Writing and Speaking tests were developed under the ACCESS 2.0 framework. The general process of item writing and editing, and of item content bias and sensitivity reviews, is similar from ACCESS to ACCESS 2.0; these processes are described in the sections below and apply to all four domains of the test. Details are also provided on the development of the Listening and Reading tests and then on development of the Writing and Speaking tests. Finally, we provide a brief overview of the development of the Kindergarten test.

1.3.1 Item Writing and Editing

Initial item writing is done by participants in an online item writing course or item writing workshop conducted by the Center for Applied Linguistics (CAL). Then, the items generated are reviewed internally and selected for further development based on how well they fit the Standards and MPIs, 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 item content and bias and sensitivity reviews.

1.3.2 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 WIDA Consortium states. Items are 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 MPI from the WIDA ELD Standards that it is intended to assess. The bias and sensitivity review panel inspects the items for potential bias that may unfairly disadvantage some students over others. Bias and sensitivity panelists represent a variety of language backgrounds and ethnicities. Based on their recommendations, the items are revised as necessary.

1.3.3 Development of Listening and Reading

The Listening and Reading components of ACCESS 2.0 Series 400 Paper were created during the original ACCESS development cycle. ACCESS was first field tested in 2004, and from 2004–2014, development continued for ACCESS, culminating in Series 303, operational in 2014–2015. For further detail on this original field test and on the processes for ongoing item development from 2004–2014, see the ACCESS for ELLs Technical Reports, particularly ACCESS for ELLs Technical Report No. 1, *Development and Field Test of ACCESS for ELLs* (2006) and *ACCESS for ELLs Technical Report No. 11* (2016).

The Listening and Reading tests for ACCESS 2.0 Series 400 Paper are composed of the same sets of items, across all grade-level clusters and tiers, as ACCESS Series 302, with minor exceptions. First, the grade-level cluster structure was updated for ACCESS 2.0. Second, there are two Reading test forms in which items are not the same between Series 302 and Series 400 Paper. In the form shared across Clusters 1 and 2 (Tier C), three items from the Series 302 form were substituted with three items from Series 203 to produce the Series 400 form. This substitution was made to avoid having very similar text appear in the key for different items on the same test form. Likewise, in the form shared across Clusters 3 and 4–5 (Tier B), three items from the Series 302 form were substituted with three items from Series 203 to produce the Series 400 form. This substitution was made to avoid a potential sensitivity issue in the wake of 2015 current events.

1.3.4 Development of Writing and Speaking

The Writing and Speaking tests for ACCESS 2.0 Series 400 Paper were developed to be shared across the Online and Paper versions of ACCESS 2.0 and followed the usual development cycle. In other words, the Online and Paper versions of the tests have the same tasks, by grade-level cluster and tier, with minor exceptions. In Writing, there are some differences in presentation between the Online and Paper test which results form the mode difference. In addition, the Paper test does not include the Speaking tier pre-A, which is included on the Online test.[3] Second, the Paper test maintains the tier structure of ACCESS for ELLs 1.0, which was provided in three tiers (A, B, and C), in order to accomodate the tier structure of Listening and Reading. Writing and Speaking tasks, however, were developed for ACCESS for ELLs 2.0 Online, which has two tiers in these domains (A and B/C). To bridge the structure of ACCESS for ELLs 1.0 and ACCESS for ELLs 2.0 Online, the same test form is shared across Tier B and Tier C Writing and Speaking tests. Table 1.3.2.A provides a graphic representation of this tier structure.

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

Table 1.3.2A

ACCESS 2.0 Paper Tier Structure and Shared Forms Across Tiers in Writing and Speaking

Domain	Tier	Shared forms
Writing	A	A
	B	B and C are shared
	C	
Speaking	A	A
	B	B and C are shared
	C	

1.3.4.1 Development of Tasks

For Writing tasks, after the external bias, sensitivity and content reviews, tasks are subject to two rounds of small-scale tryouts, the first led by CAL and the second by teachers in the field. In these tryouts, candidate folders are administered to students; student responses, as well as observations and interviews, inform further revisions to the folders. A small-scale field test of Writing folders is conducted, with responses scored at CAL, followed by a qualitative analysis of the collected responses. The main purpose of this small-scale field testing is to confirm that the tasks are functioning as intended, identify benchmark samples for rater training, and inform the rating of the tasks when they become operational.

The development of Speaking tasks is similar to that of Writing, but, as with Listening and Reading, all Speaking tasks undergo large-scale field testing. Thus, Speaking tasks undergo both quantitative and qualitative analyses following the field test to determine their appropriateness for inclusion in the next year's operational test.

1.3.4.2 Equipercentile Linking

In order to link the new ACCESS 2.0 scoring scales in Writing and Speaking to the ACCESS for ELLs 1.0 scoring scale, an equipercentile linking study was conducted in the spring of 2016. This method maintains the distribution of scale scores across two assessments by assigning scale scores to students based on their percentile ranking in the assessment.

The ACCESS 2.0 Series 400 Writing and Speaking tests were designed to measure the same constructs and had the same specifications as ACCESS 1.0 Series 303. However, several changes were made to the Series 400 Writing and Speaking scoring scales and scoring procedures such that the reporting scales cannot be adequately maintained through traditional scaling procedures (Mislevy, 1992). An equipercentile linking study (Kolen & Brennan, 2004) was conducted to link the Series 400 and Series 303 Writing and Speaking scale scores in order to maintain the ACCESS Writing and Speaking score distribution. The linking study adapted a process for

concordance that was proposed by Pommerich, Hanson, Harris, and Sconing (2004) and seen in Pommerich (2007). The main analysis involves linking the scale score distribution of Series 400 early return data to those of Series 303 population data. The computer software program LEGS (*L*inking with *E*quivalent *G*roups or *S*ingle Group Design) (Brennan, 2004a) was used in conducting the linking.

The Series 303 Writing test had three tiers (A, B, C) while the Series 400 Writing test combines Tiers B and C, therefore the Writing linking analysis was conducted by the Series 400 grade and tiers (A, B/C) so that the data being linked between two administrations would be comparable. To obtain Series 400 Writing scale score distribution for the early return data, student measures were derived using the field test parameters and transformed to the ACCESS scale score metric. During the ACCESS 2.0 Series 400 Writing field test, students took field test tasks after taking the operational ACCESS test. For the field test analysis, Writing field test tasks and rating scale parameters were estimated while anchoring on ACCESS Writing task parameters. These field test parameters were used to establish a temporary scale for conducting the equipercentile linking between series. Because the Series 303 test utilized separate Tier B and Tier C forms, population writing data from these tiers were first combined and then used in the equipercentile analyses. After the linking analyses were completed, each Series 400 scale score could be linked to a Series 303 equated scale score. Series 303 Tier B scale scores were used in creating the link. Then raw score to scale score tables were created by grade and tier. Finally, the conditional standard errors of measurement for the Series 303 equated scale scores were used to report out the conditional errors of measurement for Series 400 scale scores.

Since the Series 303 Speaking test was not tiered while the Series 400 Speaking test has three tiers (Pre-A, A, B/C), Speaking linking analyses were conducted by grade across Series 400 tiers. To obtain Series 400 Speaking scale score distribution for the early return data, a Rasch calibration was first conducted by grade, which puts task and person measure on the same logit scale by grade. Student measures were then derived and transformed to a temporary scale score metric and used in the equipercentile analyses. After the linking analyses were completed, each Series 400 scale score could be linked to a Series 303 (equated) scale score. Then raw score to scale score tables were created by grade and tier. Essentially, the raw score range of each Series 303 grade-level cluster test was separated into three sections, one for each Series 400 tiers. Because the Series 303 grade-level cluster raw score range is relatively short (0–13), not all of the proficiency levels are covered at the Series 400 grade and tier level. However, all proficiency levels are covered at the Series 400 grade-level cluster level. Finally, the conditional standard errors of measurement for the Series 303 equated scale scores were used to report out the conditional errors of measurement for Series 400 scale scores.

Since the goal of the equipercentile procedure is to preserve the scale score distribution of the ACCESS Series 303 Speaking and Writing test, the proportion of students at each observable scale score and WIDA proficiency level is constrained to be more or less the same between

series at the level where the linking was conducted. Such an approach provides stability for the ACCESS 2.0 Series 400 Writing and Speaking scores.

1.3.5 Development of Kindergarten Test

A separate field test was conducted for the Kindergarten test in 2008 in Washington, D.C. 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 Writing portion of the Kindergarten assessment included very simple writing tasks that were adapted from the SIL Writing folder on the original ACCESS Cluster 1–2 Tier A test form. The Speaking portion of the Kindergarten assessment was the same as that of the original ACCESS Cluster 1–2 test form, except it included only SIL and LoLA/LoSS folders, in order to reduce testing time.

The adaptive administration of the Kindergarten assessment includes stopping rules. 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.

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).

1.3.6 Reporting Scale

ACCESS has a vertically-equated scale (i.e., one that can measure progress across the grade levels from K to 12), as well as being horizontally equated across tiers within each grade-level cluster.

The scale runs from 100 to 600 scale score points. The scale has an interpretive center point across domains and composites. The centering value is 350, which represents, for original ACCESS, the cut score between PLs 3 and 4 for grade 5. The scale has 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 can go no lower than a scale score of 100 while students from Cluster 9–12 with the highest language proficiency in any domain can go no higher than 600. 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, as one moves from grade to grade, the scale adjusts for developmental growth. Thus, even if a student consistently receives 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 the meaning of the scale score , a standard-setting study was conducted, which is discussed in Section 1.3.7. We focus on the creation of the ACCESS for ELLs scale score here.

For details on the initial development of the ACCESS score scale, conducted subsequent to the first field test administration, see ACCESS for ELLs Technical Report No. 1, *Development and Field Test of ACCESS for ELLs* (2006), as well as Kenyon, MacGregor, Li, and Cook (2011).

Throughout the duration of ACCESS for ELLs 1.0, annual equating procedures were conducted to ensure that test results were reported on a consistent scale, year-to-year. This annual equating is the process used to maintain the ACCESS score scale.

The reporting scale for ACCESS 2.0 Series 400 maintains the same scale as ACCESS. The logit scale is transformed into a reporting scale by means of a linear transformation of the logit scores. There is a separate scale, and hence a separate transformation constant, for each of the four domains: Listening, Reading, Writing, and Speaking. For Series 400, the linear transformation of logit scores was used to generate scale scores for Listening and Reading only.

1.3.7 Standard Setting

In order to interpret appropriately what the ACCESS scale scores meant, a standard setting study was conducted in Madison, Wisconsin between April 20–27, 2005. The purpose of the study was not to set new standards, per se; rather, it was to use the WIDA ELD[4] 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 ACCESS 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 No. 1, *Development and Field Test of ACCESS for ELLs* (2006).

Four panels were convened, one for each grade-level cluster: 1–2, 3–5, 6–8, and 9–12. There were 20–22 teachers or school administrators on each panel who were deemed qualified by WIDA to participate in the study. 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 MPIs 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 PL 1 would have a 50% chance of answering correctly. They were then asked to repeat this procedure for all levels up to PL 5.

[4] 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. A new standard setting study was conducted in summer 2016, and new scale score cut points will be applied to ACCESS 2.0 (Paper and Online) beginning with the 2016–17 operational year of testing.

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 desired. 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. For Speaking, student portfolios consisted of audio-recorded spoken responses. 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 responses, and the last portfolio represented student work with a very high total raw score across responses.

After discussing the MPIs and the performance level descriptions as a group, the panelists were asked to read through the Writing portfolios or listen to the Speaking 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 PL 3, they would write 100% under the column "3" on their paper. If they felt that it was a borderline performance between PLs 2 and 3, they would write 50% under "2" and 50% under "3." They were allowed to indicate up to two language proficiency levels with 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 desired.

The final results were analyzed by CAL using a logistic regression procedure to determine the points along the underlying proficiency continuum at which at least 50% of the panelists would be expected to agree that the 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 and Speaking.

1.4 Reporting of Results

1.4.1 Scale Scores

ACCESS scores are reported as both scale scores and proficiency level scores. Scale scores, ranging from 100 to 600, are given for all four language domains. In addition, four composite scores, also ranging from 100 to 600, are given: Oral Language, Literacy, Comprehension, and Overall Composite.

The four composite scores are calculated using the following scale score weighting scheme:

- Oral Language (50% Listening + 50% Speaking)
- Literacy (50% Reading + 50% Writing)

- Comprehension (30% Listening + 70% Reading)
- Overall Composite (15% Listening + 15% Speaking + 35% Reading + 35% Writing)

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. This weighting resulted from a policy decision by the WIDA Board before the first operational administration of ACCESS, based on the view that literacy skills are paramount in developing academic language proficiency.

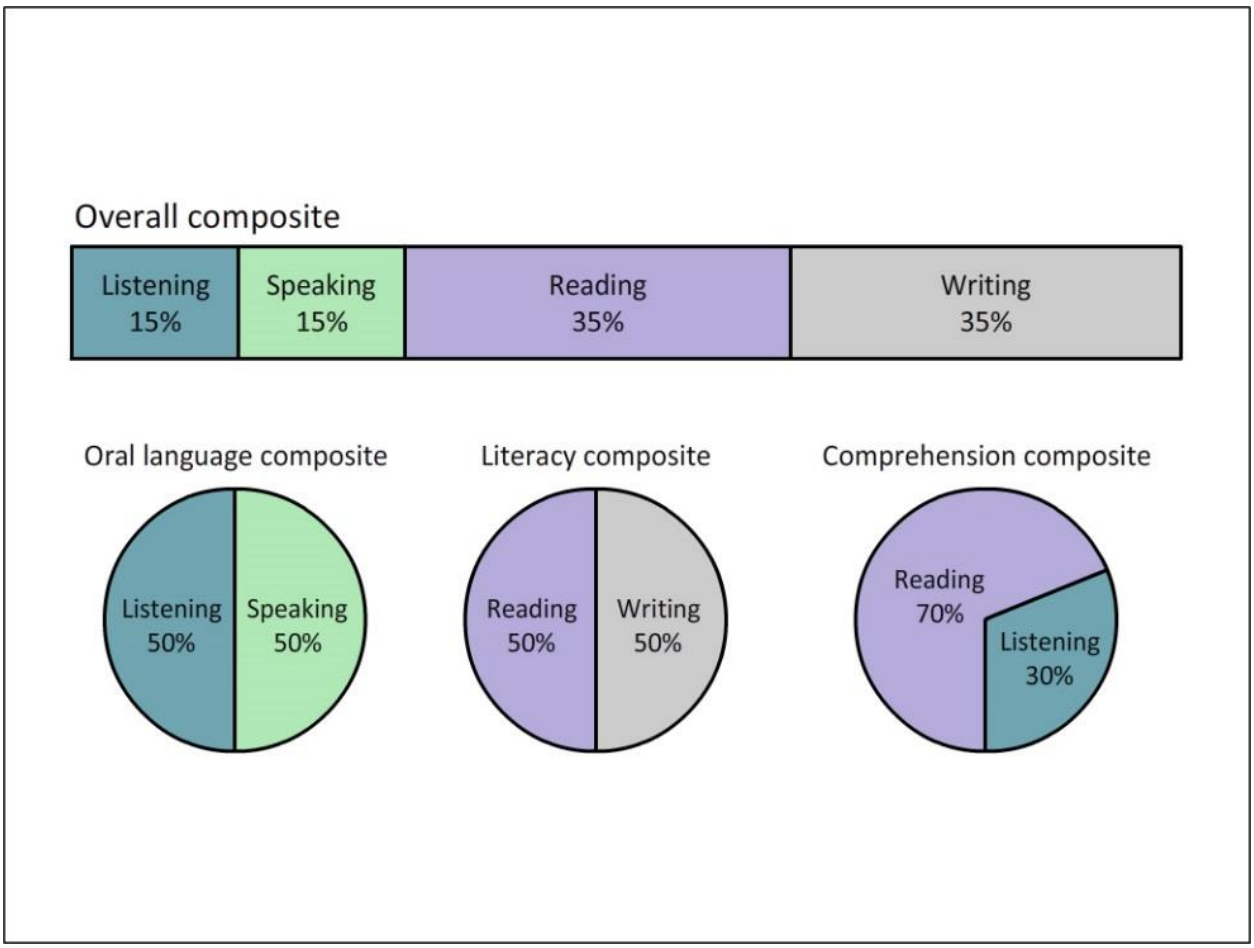


Figure 1.5.1A. Domain Composites

1.4.2 Language Proficiency Level Scores

In addition to the ACCESS 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

1.4.2A–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 1.4.2A
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 1.4.2B
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 1.4.2C
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 1.4.2D
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 1.4.2E
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 1.4.2F
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 1.4.2G
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 1.4.2H
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 PL 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 dependent upon which grade a student was in when ACCESS was administered. See, for example, the Listening cut scores in Table 1.5.2A. If a Grade 2 student receives a 350 in Listening, that would be a PL score of 6.0; if a Grade 5 student receives a 350 in Listening, that would be a 4.0; if a Grade 8 student receives a 350 in Listening, that would be a 3.2; and if a Grade 12 student receives a 350 in Listening, it would be a 2.3.

Because the bands between cut scores vary in width, proficiency level scores should not be considered to form an interval scale. That is, the distance between PL scores 1.5 and 2.5 cannot be assumed to be equal to the distance between PL scores 2.5 and 3.5. Only scale scores should be used as interval measures. Proficiency level scores are interval within a grade and proficiency level (e.g., in Grade 3, the distance between 3.1 and 3.2 is the same as the distance between 3.7 and 3.8), but they do not form an interval scale across proficiency levels.

1.5 Test Administration

1.5.1 Test Administrator Training

To prepare individuals to serve as test administrators, test administrator training for ACCESS 2.0 Series 400 Paper was conducted through an online course hosted on WIDA's website. Three certifications were offered to participants: a group test administration certification pertaining to the Listening, Reading, and Writing portions of ACCESS 2.0; a certification for the Speaking test; and a certification for the Kindergarten test. In order to be certified to administer the Listening, Reading, and Writing portions, participants had to complete and mark off tasks listed on a test administrator checklist. In order to be certified to administer the Speaking test and the Kindergarten test, participants had to pass a quiz after completing the course.

1.5.2 Test Security

Every effort is made to keep the test secure at all levels of development and administration. WIDA, CAL, and Data Recognition Corporation (DRC, the entity responsible for distribution, collection, and scoring of the printed tests) follow established policies and procedures regarding the security of the test, and every individual involved in the administration of ACCESS 2.0, from the district level to the classroom level, is trained in issues of test security.

1.5.3 Test Accommodations

If a test taker has an Individualized Education Plan (IEP), to the extent possible, the recommendations in the student's IEP are to be followed. The extent to which this was accomplished for ACCESS 2.0 Series 400 Paper was a local decision made during administration.

Starting with the 2011–2012 testing cycle, WIDA made available the *Alternate ACCESS for ELLs* test (hereafter, Alternate ACCESS). Alternate ACCESS is intended only for ELLs who

have cognitive disabilities[5] that are so significant as to prevent meaningful participation in ACCESS testing, even with accommodations. The results of the Alternate ACCESS operational administration will appear in a separate technical report.

1.6 Scoring

Test booklets are returned to DRC after testing, where they are electronically scanned in preparation for scoring. Listening, Reading, and Writing are scored by DRC. Speaking is locally scored by the test administrator. Details of the scoring methods are described below.

1.6.1 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 mark their answers directly in their test booklets, so each page is scanned into an electronic database.

1.6.2 Writing

Student responses to the Writing tasks are centrally scored at DRC. The ACCESS 2.0 Writing Scoring Scale is distinct from the WIDA Writing Rubric, which is a tool for evaluating student writing in classrooms and for interpreting student scores from ACCESS 2.0. The Writing Scoring Scale, however, was designed specifically as a scoring tool only and is not appropriate for any other purposes.

The ACCESS 2.0 Writing Scoring Scale has six whole score points that range from 1 through 6. For responses that fall in between the whole score points, plus score points are available. The scale descriptors include three different yet interrelated dimensions: discourse, sentence and word/phrase. The scale descriptors guide raters as they consider all three dimensions in order to make holistic judgments about which score points best suit a response. The dimensions are distinguished as follows:

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

When assigning a score, a rater needs to make an initial judgment about which score point (1 to 6) best describes a response and then determines whether the three descriptors for that score point suit for that response. If all three descriptors fit, a whole score point should be awarded. If

[5] Recommendations regarding physical disabilities, such as deafness or blindness, are available on the WIDA website.

there is clear evidence that one or two descriptors from an adjacent score point are a better fit, a plus score point is awarded. In addition to scale descriptors, scoring rules address special cases where responses are nonscorable, completely or partially off-task, and completely or partially off-topic. Both nonscorable and completely off-task responses are scored as 0. Completely off-topic responses receive a maximum score of 2+. Partially off-task and off-topic responses are scored in their entirety using the Scoring Scale.

To calculate a raw score for the Writing test, raters' scores for each Writing task are converted to whole numbers ranging from 0–11, as shown in Table 1.6.2A. On Tier A tests, for all grade-level clusters except for Grade 1, the scores from the three tasks are added to calculate a total raw score, which can range from 0–33. An exception to this rule is the Grade 1 Tier A test. On this form, there are four Writing tasks. The first two of these tasks use a modified version of the scoring scale and have score ranges of 0–1 and 0–3 respectively. The third and fourth task use the full scoring scale from 0–11; additionally the last task is weighted as 3. Therefore, the possible final raw scores for Grade 1 Tier A range from 0–48.

On a Tier B or Tier C test, results from the different tasks are given different weights. (Note that for ACCESS 2.0 Series 400 Paper, the Tier B Writing test is always identical to the Tier C test. The weighting rules are also identical for Tier B and Tier C tests). These weights are specified to reflect intended amounts of time that a student should spend on each task. The first task is given a weight of 1, the second task is given a weight of 2, and the third task is given a weight of 3. Thus, for example, a student with raw scores of 5, 6, and 7 on the three tasks would have a total raw score of 38 (5*1 + 2*6 + 3*7), while a student with raw scores of 7, 6, and 5 on the three tasks would have a total raw score of 34 (7*1 + 2*6 + 3*5). Raw scores on the Tier B/C tests can range from 0–66.

Table 1.6.2A
Rating to raw score conversion (Writing)

Rating	Raw Score
0	0
1	1
1+	2
2	3
2+	4
3	5
3+	6
4	7
4+	8
5	9
5+	10
6	11

1.6.2.1 Scoring Procedures for Writing

Writing tasks are scored by trained raters using the ACCESS 2.0 Writing Scoring Scale. According to documentation from DRC, raters are well-educated professionals, with at least a

four-year college degree in a relevant field and a demonstrated writing ability. Prior to scoring any live student responses, the raters undergo thorough training and qualifying. Training is task-specific in order to ensure that raters understand the nuances of each unique Writing task. Team Leaders, who are selected based on prior performance as raters and for their leadership skills, are assigned to small groups of raters; there are typically ten raters on each team. The Team Leaders are responsible for monitoring the performance of their team members and providing ongoing feedback to support accurate scoring. Scoring Directors are promoted from within DRC and earn their positions by demonstrating quality work as scorers and Team Leaders on previous projects. Scoring Directors are responsible for a specific set of tasks within a single grade-level cluster. The Scoring Directors train and oversee the teams of scorers assigned to these tasks. What follows are general scoring procedures utilized by DRC.

Rater Training and Qualifying

- Raters are seated at stations and are assigned unique ID numbers and passwords.
- The Scoring Director provides detailed directions for use of DRC's computerized scoring system.
- The Scoring Director trains the raters using task-specific anchor sets and training sets.
- Raters must demonstrate scoring proficiency on qualifying sets before scoring live responses. Scoring proficiency is defined as 70% agreement on two qualifying sets for Writing.
- Once raters are qualified, they are further trained for their grade-level cluster on the specific tasks for which they will rate responses. After this more specific training, they take calibration sets to ensure a consistent interrater understanding of how to apply the scoring scale to their particular tasks.
- DRC uses calibration sets to calibrate the raters to the actual tasks they will be scoring.

Routing Responses to Ensure "Blind" Second Ratings

- The DRC scoring system ensures that responses are routed to qualified raters until the prescribed number of ratings is performed for all responses.
- Raters do not know if they are the first or second rater.

Monitoring Scoring (Quality Control)

- Ongoing quality control checks and procedures help monitor and maintain the quality of the scoring sessions. DRC monitors rater reliability with a 20% read-behind protocol. Read-behind data are monitored daily.
- Responses can be retrieved on-demand (e.g., specific grade-level clusters, specific students) should the need arise during or subsequent to the scoring process.

- If needed, responses can be rescored based on task- or response-level information, such as task number, date, score value assigned, or scorer ID.
- For Writing, DRC uses validation sets. These are sets of items seeded into the operational sets that, on a daily basis, monitor how raters are doing when compared to the known ratings of the validity sets. The raters do not know which items are operational and which are from a validation set.

Handling Unusual Responses

- Raters can forward responses to Team Leaders for assistance.
- Responses requiring special attention, including nonscorable responses, are routed to Scoring Directors for review and resolution.

1.6.3 Speaking

The Speaking test is administered individually to each test taker. The test is media delivered. Students listen to an audio recording of the test input while following along in a test booklet. For each task on the Speaking test, a model student response exemplifies the task-level expectations for students and also serves as a scoring benchmark. The test administrator monitors and scores the test. Responses are immediately scored by the administrator while the test is administered. After listening to the student's responses, the administrator assigns a score. The Speaking test is scored using a scoring scale that is designed to evaluate student responses relative to the model student's response. As part of test administration, the test administrators hear the model student response before each student response, which supports them in assigning an appropriate score relative to the model response. The possible ratings are defined as follows:

- **Exemplary** use of oral language to provide an elaborated response. The student's language use is comparable to or going beyond the model in sophistication.
- **Strong** use of oral language to provide a detailed response. The student's language use is approaching that of the model in sophistication, though not as rich.
- **Adequate** use of oral language to provide a satisfactory response. The student's language use is not as sophisticated as that of the model.
- **Attempted** use of oral language to provide a response in English. The student's language use does not support an adequate response.
- **No response** in English.

Operationally, a score of 4 is given for every task with a score of Exemplary, 3 for Strong, 2 for Adequate, 1 for Attempted, and 0 for No Response. The sum of those scores is the total Speaking raw score for that student.

Table 1.6.3A presents the WIDA Consortium's Speaking Scoring Scale, which summarizes the scoring criteria for each score point. These criteria are applied relative to the target proficiency

level of the task (P1, P3, or P5), and the task-level expectations are embedded within the model student response. For P1 tasks, only scores of No Response (0), Attempted (1), or Adequate and above (2) are possible.

Table 1.6.3A
Speaking Scoring Scale

ACCESS for ELLs 2.0 Speaking Scoring Scale	
Score point	**Response characteristics**
Exemplary use of oral language to provide an elaborated response	• Language use comparable to or going beyond the model in sophistication • Clear, automatic, and fluent delivery • Precise and appropriate word choice
Strong use of oral language to provide a detailed response	• Language use approaching that of model in sophistication, though not as rich • Clear delivery • Appropriate word choice
Adequate use of oral language to provide a satisfactory response	• Language use not as sophisticated as that of model • Generally comprehensible use of oral language • Adequate word choice
Attempted use of oral language to provide a response in English	• Language use does not support an adequate response • Comprehensibility may be compromised • Word choice may not be fully adequate
No response (in English)	• Does not respond (in English)

To calculate a raw score for the Speaking test, the five score points are converted to whole numbers, as shown in Table 1.6.3B. To calculate a total raw score, the raw scores for each task are added together. Speaking tasks on Tier A target PL 1 and PL 3, and Speaking tasks on tiers B and C target PL 3 and PL 5. To compute raw scores for Tiers B and C, six points are added to the total raw score, representing a score of *Adequate and Above* for three tasks targeting language at PL 1. Though a Tier B or C student would not have been administered any tasks targeting the PL 1 level, it is assumed that a score of *Adequate and Above* would be applicable to such tasks. Thus, on the tier A form, scores range from 0–18; on the B/C test, from 6–30.

Table 1.6.3B
Score point to raw score conversion (Speaking).

Score Points	Raw Score
No Response (B, F, or I)*	0
Attempted	1
Adequate/Adequate and Above	2
Strong	3
Exemplary	4

* B= Blank response; F= Foreign language response; I = Indecipherable response

1.6.3.1 Training Procedures for Scoring Speaking

The Speaking Test is the only portion of ACCESS 2.0 that is scored locally. Test administrators must complete the relevant online ACCESS 2.0 Paper test administrator training module (either Grades 1–5 or Grades 6–12) for the Speaking test and pass the accompanying quiz. The training focuses on developing the test administrator's ability to score the test reliably. Separate training materials are available that address test administration and monitoring procedures. To reliably score the test, test administrators are then trained on the Speaking Scoring Scale (see Table 1.6.3A). Training materials are available for each grade-level cluster, and raters listen to anchor samples and view score justifications that provide detailed explanations for scores based on the scoring scale. Practice samples are also available so that raters can practice assigning scores. The course includes both required training material for each grade-level cluster as well as optional training material. Raters are required to complete training sections for each grade-level cluster they will administer and score. However, if a rater will score more than three grade-level clusters, they may complete rater training for only three. The quizzes include 12 items in which raters listen to and assign a score to a task response. The pass rate for the quiz is 80% correct.

2. An Assessment Use Argument for ACCESS 2.0: 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 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 2.0.

An argument-based approach to the ACCESS 2.0 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 2.0). 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).

The complete validity argument that will be employed to support the use of ACCESS 2.0 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 2.0.

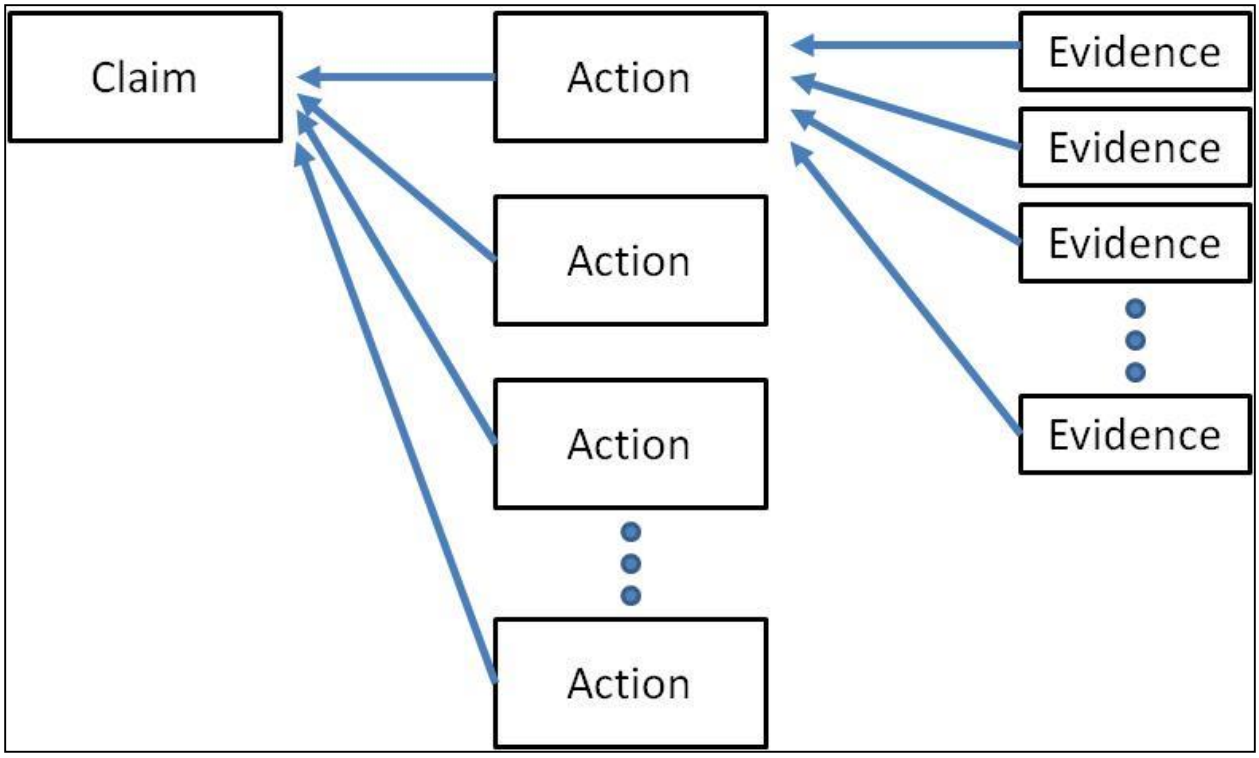


Figure 2A: General Argument Structure for Assessment Validation (simplified from Toulmin, 2003).

2.1 The Generic Validation Framework for ACCESS 2.0

The generic validation framework that will be applied to the entire ACCESS 2.0 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.

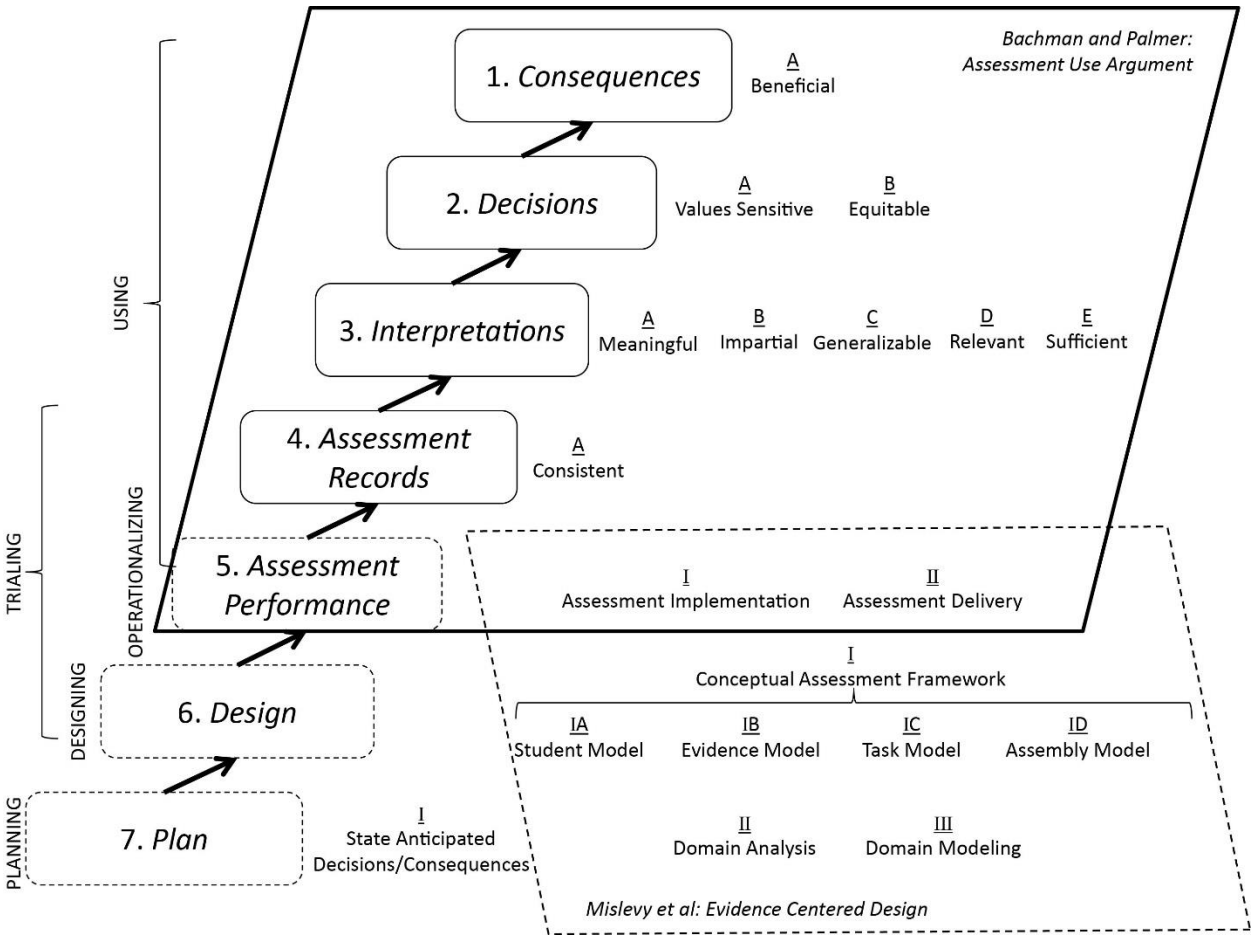


Figure 2.1A: CAL's Validation Framework (based on Bachman & Palmer, 2010; Mislevy, Almond, & Lukas, 2004).

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*).

2.2 Focus on Assessment Records

Although the complete validation framework for ACCESS 2.0 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 2.0. 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 2.0 are consistent and dependable?" Other questions about the development, administration, and outcomes of ACCESS 2.0 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 2.0.

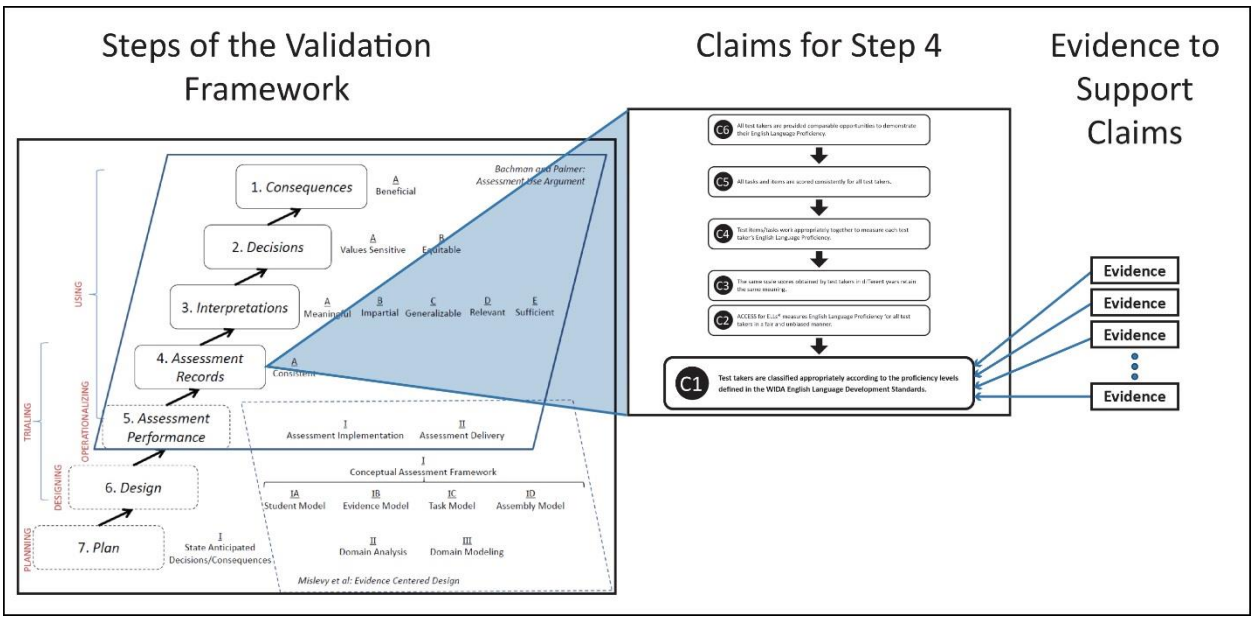


Figure 2.2A: Structure of the Argument-Based Approach Supporting *Assessment Records* (Step 4) contained in this chapter.

2.2.1 Breakdown of Claims for the *Assessment Records* Produced in the ACCESS 2.0 Assessment Program

Assessment Records (Step 4) of the complete ACCESS 2.0 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 2.0 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.

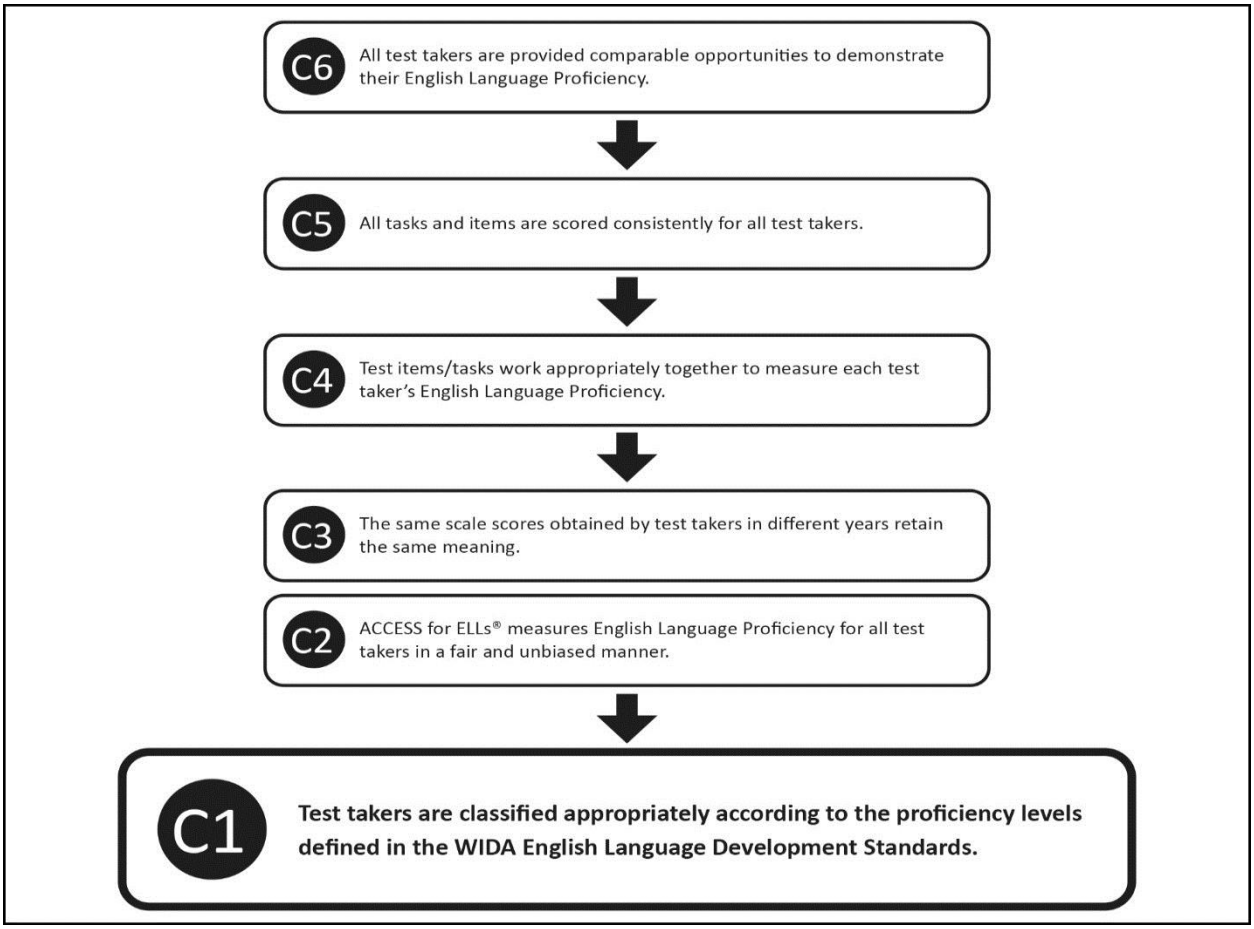


Figure 2.2.1A: Progression of Claims for Step 4: Assessment Records.

2.3 Evidence for Assessment Records Claims of ACCESS 2.0

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 sections of the report, and the tables, figures, and external sources that provide evidence related to each action. A summary table of the information presented in this section 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 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 2.0, 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 2.0 *Test Administrator Manual*.[1]

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

Evidence: General processes and procedures for test irregularities due to student condition, testing environment, or other unusual occurrences can be found in the *District and School Test Coordinator Test Administrator Manual*.[2] Specific testing situations, including where to start and stop the test, when breaks can be taken, material management protocol in the case of damaged testing material, and other detailed guidance, can be found in the *Test Administrator Manual*. States each have a specific policy for Test Administrators to follow in the case of a testing irregularity, which can include steps such as documentation to use or notification procedures to follow. These state specific steps can be found on the ACCESS 2.0 State Checklists, found on the state pages[3] and within the training course. Additionally, the ACCESS 2.0 Training Course highlights common testing irregularities and the resources to use in these circumstances.

[1] The Test Administrator Manual can be found at: https://www.wida.us/assessment/access%202.0/documents/2016TestAdministratorManual.pdf
[2] The District and School Test Coordinator Test Administration Manual can be found at: https://www.wida.us/assessment/access%202.0/documents/2016DistrictandSchoolTestCoordinatorManual.pdf
[3] WIDA state pages can be found at: https://www.wida.us/membership/states/index.aspx

In the case that the test administrator has additional questions about how to proceed in the event of a testing irregularity, the WIDA Client Services Center can be contacted via email at help@wida.us or toll free at 1-866-276-7735.

Action 4.6c: Procedures are in place to ensure that items and tasks do not have issues with bias or sensitivity.

Evidence: As detailed in Section 1.3.2, all test items and tasks are subject to bias and sensitivity reviews. These reviews examine items to ensure that they do not favor students from a particular SES, geographic area, educational background, or introduce other systematic biases.

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

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

Evidence: Section 1.6 of this report specifies the scoring procedure for ACCESS 2.0. Section 1.6.2 provides information rater training and qualifying protocols for the Writing domain, which is centrally scored by DRC. The Speaking test is locally scored. Section 1.6.3.1 details the training processes that should be followed by local schools and districts. Local schools and districts are responsible for ensuring that each rater is properly trained using these materials, for providing sufficient time and training to prepare raters for rating the speaking test, ensuring that that the appropriate resources needed to rate the Speaking test are provided, and for routinely monitoring the rating of speaking tests and evaluating inter-rater reliability indices.

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

Evidence: Section 1.6 of this report specifies the scoring procedure for ACCESS 2.0. Listening and Reading items are dichotomous and are scored electronically by DRC (see Section 1.6.1).

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

Evidence: Section 1.6 of this report specifies the scoring procedure for ACCESS 2.0. Writing tasks are centrally scored at DRC, and all raters are pre-screened, trained, and subject to qualifying scoring tests before becoming operational raters. Once raters are qualified, they then undergo additional training on the grade-level cluster and specific tasks they will be scoring. Following this more intense training, the raters are subject to calibration sets to ensure that they are properly calibrated to the grade cluster and task(s) (see Section 1.6.2).

Speaking is scored by the local test administrator after the completion of training on test administration and on the Speaking Rubric (see Section 1.6.3).

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

Evidence: DRC provides raters of performance-based tasks with specially prepared calibration sets each day to ensure that the scoring rubric is being applied consistently across scoring sessions (see Section 1.6.2). For the Writing test, pre-rated and vetted validation sets are seeded into the operational items for scoring. The validation sets are utilized to ensure that raters are scoring accurately and consistently and any drift is identified and promptly corrected.

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

Evidence: For a sample of 20% of responses to each task, interrater reliability is calculated for each of the Writing tasks (see Section 5.2.8; see Table 6F). During operational scoring, these data are monitored daily for quality control purposes.

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: Section 5.2.8 describes the ways in which test reliability is computed for the forms. Results are presented in 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 (see also Section 7.1.1.).

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

Evidence: The Complete Item or Task Analysis Summary table includes information on the Rasch fit statistics for each test item (see Section 5.2.9, Table 6G). 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.

Action 4.4d: Items and tasks of appropriate difficulty are chosen for each domain.

Evidence: The Complete Item or Task Analysis and Summary tables (see Section 5.2.9, Tables 6G) provide information on the difficulty of each item or task. Section 5.2.9 describes the construction of these tables. When the test is assembled, task difficulty is one of several criteria used to select appropriate items for operational assessment from the pool of field tested items.

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: For ACCESS 2.0 Series 400 Paper, the Listening and Reading test forms were reused forms from ACCESS Series 302. See Section 1.3.3 for further detail.

For ACCESS 2.0 Series 400 Paper Speaking and Writing, all tasks were new. Equipercentile linking was conducted to link the distribution of scores on ACCESS 2.0 Series 400 to the distribution of scores on ACCESS Series 303. See Section 1.3.4.2 for further information on equipercentile linking.

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: In typical years for the original ACCESS test, both new and previously used items and tasks (i.e., anchor items) are included on each test form (see Table 6G for a list of new and anchored test items/tasks).

For ACCESS 2.0 Series 400, which marks the transition between the original ACCESS test and ACCESS 2.0, there were no test forms which had a mix of both old and new items and tasks. Consistency with the original ACCESS scale was maintained in two ways. In the domains of Listening and Reading, the Series 400 Paper tests are reused forms of the ACCESS Series 302

test. In the domains of Writing and Speaking, equipercentile linking was conducted to maintain the distribution of scale scores across years (see Section 1.3.4.2).

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 following scaling equations are used to convert ability measures in logits to scale scores:

- L: (Ability Measure in Logits*37.571) + 316.637
- R: (Ability Measure in Logits*26.000) + 323.272

These equations have been in use from the first operational administration of ACCESS (Series 100).

For Writing and Speaking, because an equipercentile approach was used for scaling results, scaling equations were not used for ACESS 2.0 Series 400. Scaling was conducted during the ACESS 2.0 Series 400 operational year, and new scaling equations for these domains will be applied to ACCESS 2.0 Series 401. See Section 1.3.4.2 for further information on equipercentile linking.

Claim 4.2 – ACCESS 2.0 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: Results of DIF analyses are provided in Table 6H (see Section 5.2.10 for an overview of these tables). Analyses search for bias in contrasting groups based on gender (male versus female) and ethnicity (Hispanic versus non-Hispanic). Table H in Chapter 6 shows the number of items that favored one group or the other at all levels of 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: If an item shows C-level DIF, a content review panel is convened to examine the content of the item. The panel is composed of diverse members and is chosen carefully so that panelists include male and female members as well as bilingual individuals who speak either English and Spanish or English and another language. The panel then comes to a consensus on whether or not the item content is likely to favor or disfavor specific subgroups of students.

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 2.0 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' raw scores on ACCESS 2.0, organized by individual test form (e.g., Reading 3–5B), shows the extent to which ACCESS 2.0 effectively measures the performance of test takers across the range of ELD abilities that each form was designed to assess (see Section 5.2.1; see Table 6A; see Table 6B).

The distribution of test takers' scale scores on ACCESS 2.0, organized by test form (e.g., Reading 3–5B), shows that ACCESS 2.0 effectively measures the performance of test takers across the range of ELD abilities that each form was designed to assess (see Section 5.2.2; see Table 6B; see Figure 6B).

The proficiency level distribution of test takers' scores on ACCESS 2.0, organized by individual test form (e.g., Reading 3–5B), shows that ACCESS 2.0 effectively measures the performance of test takers across the range of proficiency levels that each form was designed to assess (see Section 5.2.3; 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 Section 5.2.12; see Table 6J). This distribution of scores shows that ACCESS 2.0 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 Section 5.2.6; 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 language proficiency levels. The curve shows that higher expected raw scores are required to be placed into higher language proficiency levels. Note that for Series 400, the test forms for Writing and Speaking were linked to ACCESS Series 303 using an equipercentile linking methodology. The Test Characteristic Curve is not appropriate for this year's assessment. See Section 1.3.4.2 for further information on equipercentile linking.

Action 4.1b: Distributions of scale scores and proficiency levels, organized by grade-level cluster, are analyzed to confirm that ACCESS 2.0 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 2.0, organized by grade-level cluster, shows that ACCESS 2.0 effectively measures the performance of test takers across the range of ELD abilities as described by the WIDA ELD Standards (see Section 7.2.1; Table 8A; see Figure 8A).

The proficiency level distribution of test takers' scores on ACCESS 2.0, organized by grade-level cluster, shows that ACCESS 2.0 effectively measures the performance of test takers across the range of proficiency levels as defined by the WIDA ELD Standards (see Section 7.2.2; see Table 8B; see Figure 8B).

The Test Characteristic Curve reflects test takers' mean raw scores by domain on ACCESS 2.0 across the entire test for Kindergarten and across the three tiers for the other grade-level clusters (see Section 7.2.4; Figure 8C). It also graphically illustrates how the tiers differ in difficulty, showing that ACCESS 2.0 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.

Note that for ACCESS 2.0 Series 400, the test forms for Writing and Speaking were linked to ACCESS Series 303 using an equipercentile linking methodology (described in Section 1.3.4.2). The Test Characteristic Curve is not appropriate for this year's assessment.

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 Section 5.1.1; 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 should occur within the desired range of the cut scores.

Note that for ACCESS 2.0 Series 400, Test Information Function figures are provided for Listening and Reading. Test Information Function figures are not provided for Writing and Speaking, as the equipercentile linking methodology means that the Speaking and Writing task parameters are not on the ACCESS logit scale.

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 (CSEM) at the cut point provides information on how precisely test takers' performances on ACCESS 2.0 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. The CSEM at the cut score point tables provide information on the conditional standard error of measurement at the cut scores by grade-level cluster and domain. Because the cut points depend on the grade, information for each domain is provided for each grade within a grade-level cluster (see Section 7.2.3; see Table 8C).

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).

Note that for ACCESS 2.0 Series 400, the CSEM values are estimated from the equipercentile linking (see Section 1.3.4.2 and Section 7.2.3).

The Test Information Function reflects the precision of measurement by graphically presenting the standard error of measurement across tiers for grade-level clusters (see Section 7.2.5, 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 (see Section 7.2.4), 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.

Note that for ACCESS 2.0 Series 400, Test Information Function figures are provided for Listening and Reading. Test Information Function figures are not provided for Writing and Speaking, as the equipercentile linking methodology means that the Speaking and Writing task parameters are not on the ACCESS logit scale.

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 Section 7.2.7; see Table 8E). A separate table is provided for each grade in a grade-level 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). Table 8E also shows accuracy and consistency information conditional on level and provides indices of classification accuracy and consistency at the cut points.

2.4 Summary of Assessment Records Claims, Actions, and Evidence

Table 2.4A
Summary of *Assessment Records* Claims, Actions, and Evidence

Claim	Actions	Evidence
6. All test takers are provided comparable opportunities to demonstrate their English Language Proficiency.	a. Well-specified procedures were developed for test administrators so that they are able to administer the test consistently. b. Test administrators document and report any irregularities that may occur so that appropriate action may be taken. c. Procedures are in place to ensure that items and tasks do not have issues with bias or sensitivity.	a. Test Administration Manual b. Evidence summarized with claim at 4.6b. c. Section 1.3.2
5. All items and tasks are scored consistently for all test takers.	a. Raters of performance-based tasks undergo training so that they know how to score appropriately. b. Listening and Reading items are scored electronically using a carefully checked key.. c. Raters are of performance-based tasks are certified, demonstrating that they can score appropriately. d. Raters of Writing tasks are monitored daily to ensure that they are scoring appropriately. e. Scoring data for Writing tasks are analyzed for rater agreement to understand how closely raters agree.	a. Section 1.6 b. Section 1.6 c. Section 1.6 d. Section 1.6.2 e. Section 5.2.8, Table 6F
4. Test items/tasks work appropriately together to measure each test taker's English Language Proficiency.	a. For each test form (e.g., Reading 6–8B), item and task analyses are performed and psychometric properties of the items and tasks are evaluated to confirm that scores are internally consistent. b. For each domain and composite score across tiers, item and task analyses are performed and psychometric properties of the items and tasks are evaluated to confirm that scores are internally consistent. c. Analyses of Rasch model fit statistics are conducted to show that individual tasks perform appropriately d. Items and tasks of appropriate difficulty are chosen for each domain.	a. Section 5.2.8, Table 6F b. Section 7.7.1., Table 8D c. Section 5.1.1., Table 6G d. Section 5.2.9, Table 6G

3. The same scale scores obtained by test takers in different years retain the same meaning.	a. A sufficient number of items and tasks are used as anchor items across adjacent years to maintain a consistent scale from year to year. 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. c. The same scaling equation is applied from year to year to ensure that scale scores are obtained consistently over time	a. Section 1.3.3, section 1.3.4.2 b. n/a for Series 400 Paper, see discussion with claim 4.3b c. Evidence summarized with claim at 4.3c, see also Section 1.3.4.2.
2. ACCESS 2.0 measures English Language Proficiency for all test takers in a fair and unbiased manner.	a. Differential item functioning (DIF) analyses are conducted to determine whether any items or tasks are biased against certain subgroups. 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	a. Section 5.2.10, Table 6H b. Evidence summarized with claim at 4.3b
1. Test takers are classified appropriately according to the proficiency levels defined in the WIDA English Language Development Standards.	a. Distributions of raw scores, scale scores and proficiency levels for each domain are analyzed to confirm that ACCESS 2.0 effectively measures the performance of test takers across the range of English Language Proficiency levels as defined by the WIDA English Language Development Standards. b. Distributions of scale scores and proficiency levels, organized by grade-level cluster, are analyzed to confirm that ACCESS 2.0 effectively measures the performance of test takers across the range of English Language Proficiency levels as defined by the WIDA English Language Development Standards. 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. 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 e. Classification and accuracy analyses are conducted by grade-level to confirm that proficiency level classifications are reliable for all domain and composite scores.	a. Sections 5.2.1.; 5.2.2.; 5.2.3; 5.2.6; 5.2.12; Tables 6A; 6B; 6C; 6J; Figures 6A; 6B; 6C; 6D. b. Sections 7.2.1; 7.2.2; 7.2.4; Tables 8A; 8B; Figures 8A; 8B; 8C. c. Section 5.1.1, Figure 6E d. Sections 7.2.3; 7.2.5; Table 8C; Figure 8D e. Section 7.2.7; Table 8E

2.5 Visual Guide to Tables and Figures

This section provides a visual overview to the tables and figures contained in this report. For readers who are reviewing this report in an electronic format, section headers are built into the document structure to assist the reader to navigate through the document.

2.5.1 Guide to Chapter 4, *Student Results*

Chapter 4 has three subsections:

- 4.1 Participation
- 4.2 Scale Score Results
- 4.3 Proficiency Level Results

Section 4.1, *Participation*, presents distributions of students' participation by grade and grade-level cluster. Table 2.5.1A provides an overview of the tables included in this section.

Table 2.5.1A

Table Numbering System for Section 4.1, Participation

4.1.1. Participation by Grade-level Cluster	
Table	*Title*
4.1.1.1	Participation by Grade-Level Cluster by State
4.1.1.2	Participation by Grade-Level Cluster by Gender
4.1.1.3	Participation by Grade-Level Cluster by Ethnicity
4.1.2. Participation by Grade	
Table	*Title*
4.1.2.1	Participation by Grade by State
4.1.2.2	Participation by Grade by Gender
4.1.2.3	Participation by Grade by Ethnicity
4.1.3. Participation by Tier	
Table	*Title*
4.1.3.1	Participation by Grade-Level Cluster by Tier and by Domain
4.1.3.2	Participation by Grade by Tier and by Domain
4.1.3.3	Participation by Grade-Level Cluster by Tier and by Gender
4.1.3.4	Participation by Grade-Level Cluster by Tier and by Ethnicity

Section 4.2, *Scale Score Results*, presents distributions of scale score results by grade and by grade-level cluster. These are further broken down by gender and ethnicity, and finally, correlations among scale score results are presented. Table 2.5.1.B presents the section numbering system for this section.

Table 2.5.1B

Section Numbering System for Section 4.2, Scale Score Results

Mean Scale Scores Across Domain and Composite		
	4.2.1. By Grade-level Cluster	**4.2.2. By Grade**
Alone	4.2.1.1	4.2.2.1
And by Gender	4.2.1.2	4.2.2.2
And by Ethnicity	4.2.1.3	4.2.2.3
4.2.3. Correlations Among Scale Scores by Grade-level Cluster		

Section 4.3, *Proficiency Level Results*, presents distributions of students' proficiency level results for the four domains and four composites, by grade and by grade-level cluster. Table 2.5.1C lists the numbering system for subsections. Each subsection contains a table expressing descriptive statistics as counts (Table A) and percentages (Table B).

Table 2.5.1C

Section Numbering System for Section 4.3, Proficiency Level Results

		By Grade-Level Cluster by Tier	**By Grade by Tier**	**By Grade**
		For each, distributions by count and by percent		
4.3.1	**Listening**	4.3.1.1	4.3.1.2	4.3.1.3
4.3.2	**Reading**	4.3.2.1	4.3.2.2	4.3.2.3
4.3.3	**Writing**	4.3.3.1	4.3.3.2	4.3.3.3
4.3.4	**Speaking**	4.3.4.1	4.3.4.2	4.3.4.3
4.3.5	**Oral Composite**	4.3.5.1	4.3.5.2	4.3.5.3
4.3.6	**Literacy Composite**	4.3.6.1	4.3.6.2	4.3.6.3
4.3.7	**Comprehension Composite**	4.3.7.1	4.3.7.2	4.3.7.3
4.3.8	**Overall Composite**	4.3.8.1	4.3.8.2	4.3.8.3

2.5.2. Guide to Chapter 6, *Analyses of Test Forms Results*

Chapter 6 is organized by grade-level cluster. Each grade-level cluster is divided into 4 subsections, one for each domain, as follows.

Table 2.5.2A

Section Numbering System for Chapter 6, Analysis of Test Forms Results

	Grade-level Cluster						
Domain or Composite	**K**	**1**	**2**	**3**	**4–5**	**6–8**	**9–12**
Listening	6.1.1	6.2.1	6.3.1	6.4.1	6.5.1	6.6.1	6.7.1
Reading	6.1.2	6.2.2	6.3.2	6.4.2	6.5.2	6.6.2	6.7.2
Writing	6.1.3	6.2.3	6.3.3	6.4.3	6.5.3	6.6.3	6.7.3
Speaking	6.1.4	6.2.4	6.3.4	6.4.4	6.5.4	6.6.4	6.7.4

The 28 subsections in Table 2.5.2A are further divided by tier. For each of the tier subsections, the following tables and figures are presented:

Table 2.5.2B

Table and Figure Numbering System for Chapter 8, Analysis Across Tiers Results

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

2.5.3 Guide to Chapter 8, *Analysis Across Tiers Results*

Chapter 8 is organized by grade-level cluster. Each grade-level cluster is divided into 8 subsections, one for each domain and one for each composite, as follows.

Table 2.5.3A

Section Numbering System for Chapter 8, Analysis Across Tiers Results

	Grade-level Cluster						
Domain or Composite	**K**	**1**	**2**	**3**	**4–5**	**6–8**	**9–12**
Listening	8.1.1	8.2.1	8.3.1	8.4.1	8.5.1	8.6.1	8.7.1
Reading	8.1.2	8.2.2	8.3.2	8.4.2	8.5.2	8.6.2	8.7.2
Writing	8.1.3	8.2.3	8.3.3	8.4.3	8.5.3	8.6.3	8.7.3
Speaking	8.1.4	8.2.4	8.3.4	8.4.4	8.5.4	8.6.4	8.7.4
Oral Composite	8.1.5	8.2.5	8.3.5	8.4.5	8.5.5	8.6.5	8.7.5
Literacy Composite	8.1.6	8.2.6	8.3.6	8.4.6	8.5.6	8.6.6	8.7.6
Comprehension Composite	8.1.7	8.2.7	8.3.7	8.4.7	8.5.7	8.6.7	8.7.7
Overall Composite	8.1.8	8.2.8	8.3.8	8.4.8	8.5.8	8.6.8	8.7.8

For each domain and composite subsection, the following tables and figures are presented:

Table 2.5.3B

Table and Figure Numbering System for Chapter 8, Analysis Across Tiers Results

	Figure	**Table**	**Applies to**
Scale Score Distributions	A	A	*Domains and Composites*
Proficiency Level Distributions	B	B	*Domains and Composites*
CSEM at Cut Scores		C	*Domains only*
Test Characteristic Curve	C		*Domains only*
Test Information Function	D		*Domains only*
Weighted Reliability		D	*Domains and Composites*
Accuracy and Consistency of Classification		E	*Domains and Composites*

3. Descriptions of Student Results

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

3.1 Participation

Participation in ACCESS 2.0 Paper is shown in three ways: by grade-level cluster; by grade, and by tier.

3.1.1 Grade-Level Cluster

Chapter 4.1.1 gives information on participation by grade-level cluster.

Table 4.1.1.1 shows participation across the 38 WIDA states that participated in the operational testing program of ACCESS 2.0 Paper in 2015–2016. The first row shows the grade-level cluster, the next 38 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 38 states.

Table 4.1.1.2 shows participation by grade-level cluster and by gender across all 38 states combined, while Table 4.1.1.3 shows participation by grade-level cluster and by ethnicity across all 38 states.

3.1.2 Grade

Section 4.1.2 provides similar data as the previous section, but it is broken out by grade rather than by grade-level cluster.

3.1.3 Tier

Section 4.1.3 gives information on participation by tier.

Table 4.1.3.1 shows this information by grade-level cluster, tier, and domain.

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.

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.

Note that in some circumstances there was a mismatch between a student's reported grade and the reported grade-level cluster of the test the student took (e.g., a student who was reported to be in Grade 5 was administered a test in the 6–8 grade-level cluster). In all, 586 students were administered a test form for a grade-level cluster other than their reported grade. Table 3.1 below shows the number of students in each grade who were administered out-of-grade 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/Grade-Level Cluster Mismatch

	Grade-Level Cluster and Tier																			
		1			2			3			4–5			6–8			9–12			
Grade	**K**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**	**Total**
K		27	9	4	5	7	1	7	3	1	4	2	3	5	1	1	7	3	1	91
1	131				11	8	7	6	1	1	3	3	2	4	1	2	4	1	0	185
2	10	19	23	10				10	13	6	4	1	0	0	1	1	2	0	4	104
3	10	1	1	0	7	8	8				12	17	6	1	0	1	0	1	1	74
4	5	1	0	0	0	0	1	7	4	1				0	0	3	0	1	0	23
5	5	0	0	0	0	0	2	3	3	2				4	5	7	0	0	0	31
6	1	2	0	0	0	0	0	0	2	0	4	2	1				0	1	0	13
7	2	0	0	0	0	1	1	2	3	0	1	1	1				1	0	0	13
8	1	1	1	0	0	0	0	0	0	0	0	0	1				4	5	3	16
9	1	0	0	0	0	0	0	0	0	0	0	0	0	1	7	3				12
10	2	1	1	0	0	0	1	1	0	0	0	0	0	2	0	2				10
11	2	5	2	0	0	0	0	0	0	0	1	0	0	1	0	0				11
12	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0				3
Total	171	57	37	14	23	26	21	36	29	11	29	26	14	18	15	20	18	12	9	586

3.2 Scale Score Results

3.2.1 Mean Scale Scores Across Domain and Composite Scores Section

Chapter 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 race and ethnicity. In 2010, the Census Bureau introduced a new approach to reporting race and ethnicity. Previously, race and ethnicity had been a single category with six values (Hispanic, Asian/Pacific Islander/Hawaiian, Black/African American, American Indian/Alaskan Native, White-Non Hispanic, and Multi-racial/Other). Under the new approach, ethnicity has become a binary category (Hispanic or Non-Hispanic), with five categories for race (American Indian/Alaskan Native, Asian, Black/African American, Pacific Islander/Hawaiian, and White) that are not mutually exclusive. Thus, for example, Student A may be labeled as Hispanic for ethnicity and Asian for race, while Student B may be labeled as Non-Hispanic for ethnicity and both American Indian/Alaskan Native and Black/African American for race. Starting with Series 202, students who 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 4.2.3G 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 grade-level cluster 1, Table 4.2.3C shows the results for grade-level cluster 2, Table 4.2.3D shows the results for grade-level cluster 3, Table 4.2.3E shows the results for grade-level cluster 4–5, Table 4.2.3F shows the results for grade-level cluster 6–8, and Table 4.2.3G shows the results for grade-level cluster 9–12. 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 4.2.3G 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:

Table 4.3.1 Listening

Table 4.3.2 Reading

Table 4.3.3 Writing

Table 4.3.4 Speaking

Table 4.3.5 Oral Language Composite

Table 4.3.6 Literacy Composite

Table 4.3.7 Comprehension Composite

Table 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 * 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.

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 S400 Paper

State	Cluster							Total
	K	1	2	3	4-5	6-8	9-12	
AK	1,450	313	316	350	742	862	683	4,716
AL	3,641	1,130	1,016	864	396	461	463	7,971
CO	11,155	3,723	3,936	3,883	5,742	6,279	5,297	40,015
DC	1,023	29	22	15	11	4	1	1,105
DE	1,922	4	7	2	1	2	2	1,940
FL	34,806	38,241	35,830	25,762	44,921	39,165	40,571	259,296
GA	17,236	5,184	5,050	3,768	1,489	425	339	33,491
HI	1,992	1,490	1,910	1,757	1,831	2,269	2,489	13,738
ID	2,274	6	2	7	8	7	9	2,313
IL	27,203	13,017	13,363	5,897	3,695	2,795	2,463	68,433
IN	7,346	170	244	141	196	196	49	8,342
KY	3,227	97	100	84	83	86	91	3,768
MA	10,187	6,273	6,272	5,484	6,808	6,329	7,653	49,006
MD	10,305	12	15	15	90	86	17	10,540
ME	463	56	55	43	78	246	407	1,348
MI	10,326	625	678	634	1,052	1,287	1,511	16,113
MN	8,349	152	92	80	166	208	206	9,253
MO	4,736	20	17	27	25	19	14	4,858
MP	44	0	0	0	0	0	0	44
MT	150	2	5	4	0	0	0	161
NC	12,664	314	314	316	238	139	142	14,127
ND	419	12	8	11	11	15	23	499
NH	459	108	93	98	103	103	118	1,082
NJ	11,990	576	230	187	206	295	338	13,822
NM	5,453	475	518	613	914	1,140	287	9,400
NV	8,000	0	0	1	0	4	18	8,023
OK	7,534	3,987	3,694	3,583	3,611	2,935	2,304	27,648
PA	4,898	1,453	1,480	1,367	2,310	2,978	2,299	16,785
RI	1,199	336	286	301	295	315	331	3,063
SC	3,999	808	1,055	1,034	1,501	1,569	915	10,881
SD	748	60	65	53	28	35	5	994
TN	5,507	10	3	1	1	1	1	5,524
UT	5,028	51	38	12	19	26	28	5,202
VA	13,857	5,376	5,024	2,515	1,488	1,245	1,742	31,247
VI	124	0	0	0	0	0	0	124
VT	180	3	3	3	2	2	2	195
WI	5,601	30	36	28	43	36	42	5,816
WY	425	35	18	20	25	35	30	588
Total	245,920	84,178	81,795	58,960	78,129	71,599	70,890	691,471

4.1.1.2 By Gender

Table 4.1.1.2
Participation by Cluster by Gender S400 Paper

Cluster		Gender			Total
		F	M	Missing	
K	Count	114,213	128,977	2,730	245,920
	% within Cluster	46.4%	52.4%	1.1%	100.0%
1	Count	39,655	44,102	421	84,178
	% within Cluster	47.1%	52.4%	0.5%	100.0%
2	Count	38,360	43,071	364	81,795
	% within Cluster	46.9%	52.7%	0.4%	100.0%
3	Count	26,770	31,842	348	58,960
	% within Cluster	45.4%	54.0%	0.6%	100.0%
4-5	Count	35,173	42,542	414	78,129
	% within Cluster	45.0%	54.5%	0.5%	100.0%
6-8	Count	32,119	39,036	444	71,599
	% within Cluster	44.9%	54.5%	0.6%	100.0%
9-12	Count	32,238	38,215	437	70,890
	% within Cluster	45.5%	53.9%	0.6%	100.0%
Total	Count	318,528	367,785	5,158	691,471
	% within Cluster	46.1%	53.2%	0.7%	100.0%

4.1.1.3 By Ethnicity

Table 4.1.1.3
Participation by Cluster by Ethnicity S400 Paper

Cluster		Hispanic/Non-Hispanic			Total
		Hispanic	Other	Unknown	
K	Count	165,084	73,437	7,399	245,920
	% within Cluster	67.1%	29.9%	3.0%	100.0%
1	Count	62,414	20,707	1,057	84,178
	% within Cluster	74.1%	24.6%	1.3%	100.0%
2	Count	60,765	20,015	1,015	81,795
	% within Cluster	74.3%	24.5%	1.2%	100.0%
3	Count	42,770	15,230	960	58,960
	% within Cluster	72.5%	25.8%	1.6%	100.0%
4-5	Count	57,317	19,686	1,126	78,129
	% within Cluster	73.4%	25.2%	1.4%	100.0%
6-8	Count	50,912	19,248	1,439	71,599
	% within Cluster	71.1%	26.9%	2.0%	100.0%
9-12	Count	48,512	20,959	1,419	70,890
	% within Cluster	68.4%	29.6%	2.0%	100.0%
Total	Count	487,774	189,282	14,415	691,471
	% within Cluster	70.5%	27.4%	2.1%	100.0%

4.1.2 Participation by Grade

4.1.2.1 By State

Table 4.1.2.1
Participation by Grade by State S400 Paper

State	Grade													Total
	K	1	2	3	4	5	6	7	8	9	10	11	12	
AK	1,450	313	316	350	402	340	328	289	245	288	177	123	95	4,716
AL	3,641	1,130	1,016	864	235	161	150	159	152	213	147	67	36	7,971
CO	11,155	3,723	3,936	3,883	3,308	2,434	2,094	2,084	2,101	2,469	1,245	795	788	40,015
DC	1,023	29	22	15	5	6	2	1	1	0	1	0	0	1,105
DE	1,922	4	7	2	1	0	0	2	0	1	0	0	1	1,940
FL	34,806	38,241	35,830	25,762	25,379	19,542	13,908	12,722	12,535	12,981	11,717	9,535	6,338	259,296
GA	17,236	5,184	5,050	3,768	852	637	161	127	137	188	83	38	30	33,491
HI	1,992	1,490	1,910	1,757	956	875	737	770	762	1,041	640	445	363	13,738
ID	2,274	6	2	7	2	6	0	3	4	5	1	1	2	2,313
IL	27,203	13,017	13,363	5,897	2,239	1,456	1,032	893	870	1,100	638	452	273	68,433
IN	7,346	170	244	141	103	93	72	59	65	20	13	12	4	8,342
KY	3,227	97	100	84	44	39	34	23	29	45	25	12	9	3,768
MA	10,187	6,273	6,272	5,484	3,801	3,007	2,264	2,067	1,998	2,807	2,014	1,650	1,182	49,006
MD	10,305	12	15	15	48	42	37	30	19	9	3	2	3	10,540
ME	463	56	55	43	36	42	18	109	119	112	102	110	83	1,348
MI	10,326	625	678	634	584	468	453	416	418	551	415	303	242	16,113
MN	8,349	152	92	80	86	80	81	67	60	43	43	46	74	9,253
MO	4,736	20	17	27	13	12	10	3	6	6	5	2	1	4,858
MP	44	0	0	0	0	0	0	0	0	0	0	0	0	44
MT	150	2	5	4	0	0	0	0	0	0	0	0	0	161
NC	12,664	314	314	316	170	68	40	46	53	64	26	30	22	14,127
ND	419	12	8	11	4	7	5	7	3	14	2	5	2	499
NH	459	108	93	98	56	47	29	39	35	43	33	25	17	1,082
NJ	11,990	576	230	187	120	86	101	95	99	177	90	42	29	13,822
NM	5,453	475	518	613	451	463	460	307	373	136	67	44	40	9,400
NV	8,000	0	0	1	0	0	0	0	4	3	5	6	4	8,023
OK	7,534	3,987	3,694	3,583	2,327	1,284	1,183	872	880	1,034	621	388	261	27,648
PA	4,898	1,453	1,480	1,367	1,242	1,068	1,003	1,039	936	763	702	467	367	16,785
RI	1,199	336	286	301	152	143	100	111	104	131	99	60	41	3,063
SC	3,999	808	1,055	1,034	778	723	549	539	481	423	194	191	107	10,881
SD	748	60	65	53	12	16	13	8	14	3	0	1	1	994
TN	5,507	10	3	1	1	0	0	0	1	0	0	1	0	5,524
UT	5,028	51	38	12	8	11	10	7	9	7	8	10	3	5,202
VA	13,857	5,376	5,024	2,515	874	614	385	419	441	858	413	339	132	31,247
VI	124	0	0	0	0	0	0	0	0	0	0	0	0	124
VT	180	3	3	3	1	1	0	1	1	1	0	1	0	195
WI	5,601	30	36	28	28	15	9	16	11	22	7	8	5	5,816
WY	425	35	18	20	10	15	18	8	9	11	6	5	8	588
Total	245,920	84,178	81,795	58,960	44,328	33,801	25,286	23,338	22,975	25,569	19,542	15,216	10,563	691,471

4.1.2.2 By Gender

Table 4.1.2.2

Participation by Grade by Gender S400 Paper

Grade		Gender			Total
		F	M	Missing	
K	Count	114,213	128,977	2,730	245,920
	% within Grade	46.4%	52.4%	1.1%	100.0%
1	Count	39,655	44,102	421	84,178
	% within Grade	47.1%	52.4%	0.5%	100.0%
2	Count	38,360	43,071	364	81,795
	% within Grade	46.9%	52.7%	0.4%	100.0%
3	Count	26,770	31,842	348	58,960
	% within Grade	45.4%	54.0%	0.6%	100.0%
4	Count	20,076	23,988	264	44,328
	% within Grade	45.3%	54.1%	0.6%	100.0%
5	Count	15,097	18,554	150	33,801
	% within Grade	44.7%	54.9%	0.4%	100.0%
6	Count	11,310	13,794	182	25,286
	% within Grade	44.7%	54.6%	0.7%	100.0%
7	Count	10,470	12,727	141	23,338
	% within Grade	44.9%	54.5%	0.6%	100.0%
8	Count	10,339	12,515	121	22,975
	% within Grade	45.0%	54.5%	0.5%	100.0%
9	Count	11,185	14,188	196	25,569
	% within Grade	43.7%	55.5%	0.8%	100.0%
10	Count	8,836	10,587	119	19,542
	% within Grade	45.2%	54.2%	0.6%	100.0%
11	Count	7,051	8,083	82	15,216
	% within Grade	46.3%	53.1%	0.5%	100.0%
12	Count	5,166	5,357	40	10,563
	% within Grade	48.9%	50.7%	0.4%	100.0%
Total	Count	318,528	367,785	5,158	691,471
	% within Grade	46.1%	53.2%	0.7%	100.0%

4.1.2.3 By Ethnicity

Table 4.1.2.3
Participation by Grade by Ethnicity S400 Paper

Grade		Hispanic/Non-Hispanic			Total
		Hispanic	Other	Unknown	
K	Count	165,084	73,437	7,399	245,920
	% within Grade	67.1%	29.9%	3.0%	100.0%
1	Count	62,414	20,707	1,057	84,178
	% within Grade	74.1%	24.6%	1.3%	100.0%
2	Count	60,765	20,015	1,015	81,795
	% within Grade	74.3%	24.5%	1.2%	100.0%
3	Count	42,770	15,230	960	58,960
	% within Grade	72.5%	25.8%	1.6%	100.0%
4	Count	32,747	10,926	655	44,328
	% within Grade	73.9%	24.6%	1.5%	100.0%
5	Count	24,570	8,760	471	33,801
	% within Grade	72.7%	25.9%	1.4%	100.0%
6	Count	17,967	6,779	540	25,286
	% within Grade	71.1%	26.8%	2.1%	100.0%
7	Count	16,612	6,277	449	23,338
	% within Grade	71.2%	26.9%	1.9%	100.0%
8	Count	16,333	6,192	450	22,975
	% within Grade	71.1%	27.0%	2.0%	100.0%
9	Count	18,289	6,723	557	25,569
	% within Grade	71.5%	26.3%	2.2%	100.0%
10	Count	13,566	5,584	392	19,542
	% within Grade	69.4%	28.6%	2.0%	100.0%
11	Count	10,159	4,823	234	15,216
	% within Grade	66.8%	31.7%	1.5%	100.0%
12	Count	6,498	3,829	236	10,563
	% within Grade	61.5%	36.2%	2.2%	100.0%
Total	Count	487,774	189,282	14,415	691,471
	% within Grade	70.5%	27.4%	2.1%	100.0%

4.1.3 Participation by Tier

4.1.3.1 By Cluster by Domain

Table 4.1.3.1

Participation by Cluster by Tier by Domain S400 Paper

Cluster			Domain			
			Listening	**Reading**	**Writing**	**Speaking**
K	Tier	-	245,920	245,920	245,920	245,920
1	Tier	A	30,871	30,871	30,871	30,871
		B	33,166	33,166	33,166	33,166
		C	20,141	20,141	20,141	20,141
	Total		84,178	84,178	84,178	84,178
2	Tier	A	10,443	10,443	10,443	10,443
		B	33,659	33,659	33,659	33,659
		C	37,693	37,693	37,693	37,693
	Total		81,795	81,795	81,795	81,795
3	Tier	A	10,579	10,579	10,579	10,579
		B	21,826	21,826	21,826	21,826
		C	26,555	26,555	26,555	26,555
	Total		58,960	58,960	58,960	58,960
4-5	Tier	A	13,495	13,495	13,495	13,495
		B	23,831	23,831	23,831	23,831
		C	40,803	40,803	40,803	40,803
	Total		78,129	78,129	78,129	78,129
6-8	Tier	A	17,007	17,007	17,007	17,007
		B	21,611	21,611	21,611	21,611
		C	32,981	32,981	32,981	32,981
	Total		71,599	71,599	71,599	71,599
9-12	Tier	A	19,245	19,245	19,245	19,245
		B	21,330	21,330	21,330	21,330
		C	30,315	30,315	30,315	30,315
	Total		70,890	70,890	70,890	70,890

4.1.3.2 By Grade by Domain

Table 4.1.3.2
Participation by Grade by Tier by Domain S400 Paper

Grade			Domain			
			Listening	**Reading**	**Writing**	**Speaking**
K	Tier	-	245,920	245,920	245,920	245,920
1	Tier	A	30,871	30,871	30,871	30,871
		B	33,166	33,166	33,166	33,166
		C	20,141	20,141	20,141	20,141
	Total		84,178	84,178	84,178	84,178
2	Tier	A	10,443	10,443	10,443	10,443
		B	33,659	33,659	33,659	33,659
		C	37,693	37,693	37,693	37,693
	Total		81,795	81,795	81,795	81,795
3	Tier	A	10,579	10,579	10,579	10,579
		B	21,826	21,826	21,826	21,826
		C	26,555	26,555	26,555	26,555
	Total		58,960	58,960	58,960	58,960
4	Tier	A	7,272	7,272	7,272	7,272
		B	14,862	14,862	14,862	14,862
		C	22,194	22,194	22,194	22,194
	Total		44,328	44,328	44,328	44,328
5	Tier	A	6,223	6,223	6,223	6,223
		B	8,969	8,969	8,969	8,969
		C	18,609	18,609	18,609	18,609
	Total		33,801	33,801	33,801	33,801
6	Tier	A	5,779	5,779	5,779	5,779
		B	7,803	7,803	7,803	7,803
		C	11,704	11,704	11,704	11,704
	Total		25,286	25,286	25,286	25,286
7	Tier	A	5,756	5,756	5,756	5,756
		B	6,982	6,982	6,982	6,982
		C	10,600	10,600	10,600	10,600
	Total		23,338	23,338	23,338	23,338
8	Tier	A	5,472	5,472	5,472	5,472
		B	6,826	6,826	6,826	6,826
		C	10,677	10,677	10,677	10,677
	Total		22,975	22,975	22,975	22,975
9	Tier	A	8,118	8,118	8,118	8,118
		B	7,111	7,111	7,111	7,111
		C	10,340	10,340	10,340	10,340
	Total		25,569	25,569	25,569	25,569
10	Tier	A	5,577	5,577	5,577	5,577
		B	5,983	5,983	5,983	5,983
		C	7,982	7,982	7,982	7,982
	Total		19,542	19,542	19,542	19,542
11	Tier	A	3,799	3,799	3,799	3,799
		B	4,785	4,785	4,785	4,785
		C	6,632	6,632	6,632	6,632
	Total		15,216	15,216	15,216	15,216
12	Tier	A	1,751	1,751	1,751	1,751
		B	3,451	3,451	3,451	3,451
		C	5,361	5,361	5,361	5,361
	Total		10,563	10,563	10,563	10,563

4.1.3.3 By Cluster by Gender

Table 4.1.3.3

Participation by Cluster by Tier by Gender S400 Paper

Cluster	Tier		Gender			Total
			F	M	Missing	
K	-	Count	114,213	128,977	2,730	245,920
		% within Tier	46.4%	52.4%	1.1%	100.0%
1	A	Count	14,004	16,711	156	30,871
		% within Tier	45.4%	54.1%	0.5%	100.0%
	B	Count	15,573	17,414	179	33,166
		% within Tier	47.0%	52.5%	0.5%	100.0%
	C	Count	10,078	9,977	86	20,141
		% within Tier	50.0%	49.5%	0.4%	100.0%
2	A	Count	4,570	5,798	75	10,443
		% within Tier	43.8%	55.5%	0.7%	100.0%
	B	Count	15,297	18,183	179	33,659
		% within Tier	45.4%	54.0%	0.5%	100.0%
	C	Count	18,493	19,090	110	37,693
		% within Tier	49.1%	50.6%	0.3%	100.0%
3	A	Count	4,552	5,967	60	10,579
		% within Tier	43.0%	56.4%	0.6%	100.0%
	B	Count	9,730	11,968	128	21,826
		% within Tier	44.6%	54.8%	0.6%	100.0%
	C	Count	12,488	13,907	160	26,555
		% within Tier	47.0%	52.4%	0.6%	100.0%
4-5	A	Count	5,995	7,391	109	13,495
		% within Tier	44.4%	54.8%	0.8%	100.0%
	B	Count	10,370	13,311	150	23,831
		% within Tier	43.5%	55.9%	0.6%	100.0%
	C	Count	18,808	21,840	155	40,803
		% within Tier	46.1%	53.5%	0.4%	100.0%
6-8	A	Count	7,814	9,071	122	17,007
		% within Tier	45.9%	53.3%	0.7%	100.0%
	B	Count	9,312	12,129	170	21,611
		% within Tier	43.1%	56.1%	0.8%	100.0%
	C	Count	14,993	17,836	152	32,981
		% within Tier	45.5%	54.1%	0.5%	100.0%
9-12	A	Count	8,749	10,349	147	19,245
		% within Tier	45.5%	53.8%	0.8%	100.0%
	B	Count	9,610	11,597	123	21,330
		% within Tier	45.1%	54.4%	0.6%	100.0%
	C	Count	13,879	16,269	167	30,315
		% within Tier	45.8%	53.7%	0.6%	100.0%

4.1.3.4 By Cluster by Ethnicity

Table 4.1.3.4

Participation by Cluster by Tier by Ethnicity S400 Paper

Cluster	Tier		Hispanic/Non-Hispanic			Total
			Hispanic	Other	Unknown	
K	-	Count	165,084	73,437	7,399	245,920
		% within Tier	67.1%	29.9%	3.0%	100.0%
1	A	Count	23,383	7,037	451	30,871
		% within Tier	75.7%	22.8%	1.5%	100.0%
	B	Count	24,161	8,625	380	33,166
		% within Tier	72.8%	26.0%	1.1%	100.0%
	C	Count	14,870	5,045	226	20,141
		% within Tier	73.8%	25.0%	1.1%	100.0%
2	A	Count	7,336	2,875	232	10,443
		% within Tier	70.2%	27.5%	2.2%	100.0%
	B	Count	25,105	8,091	463	33,659
		% within Tier	74.6%	24.0%	1.4%	100.0%
	C	Count	28,324	9,049	320	37,693
		% within Tier	75.1%	24.0%	0.8%	100.0%
3	A	Count	7,726	2,607	246	10,579
		% within Tier	73.0%	24.6%	2.3%	100.0%
	B	Count	15,930	5,569	327	21,826
		% within Tier	73.0%	25.5%	1.5%	100.0%
	C	Count	19,114	7,054	387	26,555
		% within Tier	72.0%	26.6%	1.5%	100.0%
4-5	A	Count	9,768	3,408	319	13,495
		% within Tier	72.4%	25.3%	2.4%	100.0%
	B	Count	16,694	6,741	396	23,831
		% within Tier	70.1%	28.3%	1.7%	100.0%
	C	Count	30,855	9,537	411	40,803
		% within Tier	75.6%	23.4%	1.0%	100.0%
6-8	A	Count	12,819	3,834	354	17,007
		% within Tier	75.4%	22.5%	2.1%	100.0%
	B	Count	14,159	6,978	474	21,611
		% within Tier	65.5%	32.3%	2.2%	100.0%
	C	Count	23,934	8,436	611	32,981
		% within Tier	72.6%	25.6%	1.9%	100.0%
9-12	A	Count	14,292	4,564	389	19,245
		% within Tier	74.3%	23.7%	2.0%	100.0%
	B	Count	13,836	7,155	339	21,330
		% within Tier	64.9%	33.5%	1.6%	100.0%
	C	Count	20,384	9,240	691	30,315
		% within Tier	67.2%	30.5%	2.3%	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 S400 Paper

Cluster		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Mean	272.32	194.15	211.38	306.09	289.43	203.02	217.59	228.73
	N	245,920	245,920	245,920	245,920	245,920	245,920	245,920	245,920
1	Mean	300.66	286.45	271.19	345.68	323.72	279.37	290.94	292.73
	N	81,435	77,556	84,178	83,486	80,880	77,556	75,575	75,085
2	Mean	328.17	313.22	287.59	368.72	348.88	300.87	317.95	315.21
	N	80,833	78,002	81,795	81,205	80,320	78,002	77,283	76,803
3	Mean	347.71	330.22	327.39	370.79	359.70	329.16	335.67	338.25
	N	58,389	56,094	58,960	58,487	57,989	56,094	55,691	55,326
4-5	Mean	369.13	347.51	352.02	364.10	367.01	350.21	354.17	355.14
	N	77,593	74,739	78,129	77,611	77,128	74,739	74,345	73,905
6-8	Mean	379.84	355.61	353.95	379.66	380.20	355.28	363.10	362.68
	N	70,986	67,276	71,599	70,858	70,350	67,276	66,882	66,290
9-12	Mean	382.08	378.46	390.96	390.42	386.69	385.35	379.72	385.67
	N	69,802	66,928	70,890	69,360	68,568	66,928	66,264	65,134

4.2.1.2 By Cluster by Gender

Table 4.2.1.2
Mean Scale Scores by Cluster by Gender S400 Paper

Cluster	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	F	Mean	277.43	197.48	217.14	310.94	294.41	207.56	221.45	233.40
		N	114,213	114,213	114,213	114,213	114,213	114,213	114,213	114,213
	M	Mean	267.74	191.39	206.48	301.80	284.99	199.17	214.28	224.72
		N	128,977	128,977	128,977	128,977	128,977	128,977	128,977	128,977
	Missing	Mean	275.09	185.92	202.11	306.04	290.78	194.25	212.66	223.01
		N	2,730	2,730	2,730	2,730	2,730	2,730	2,730	2,730
1	F	Mean	302.00	287.66	274.33	347.89	325.44	281.48	292.17	294.70
		N	38,526	36,580	39,655	39,342	38,270	36,580	35,743	35,514
	M	Mean	299.43	285.36	268.39	343.71	322.17	277.47	289.82	290.94
		N	42,518	40,603	44,102	43,745	42,226	40,603	39,470	39,214
	Missing	Mean	301.04	286.69	267.95	342.61	323.39	279.29	291.34	293.14
		N	391	373	421	399	384	373	362	357
2	F	Mean	329.26	314.77	291.58	369.69	349.89	303.60	319.35	317.40
		N	37,955	36,638	38,360	38,097	37,726	36,638	36,329	36,114
	M	Mean	327.23	311.84	284.07	367.92	348.02	298.45	316.72	313.29
		N	42,522	41,019	43,071	42,755	42,243	41,019	40,611	40,351
	Missing	Mean	323.48	312.59	283.11	362.54	343.25	298.63	315.86	311.71
		N	356	345	364	353	351	345	343	338
3	F	Mean	348.27	331.10	332.26	371.91	360.50	331.95	336.41	340.37
		N	26,541	25,507	26,770	26,563	26,364	25,507	25,340	25,180
	M	Mean	347.19	329.43	323.33	369.90	359.01	326.81	335.00	336.44
		N	31,508	30,269	31,842	31,580	31,286	30,269	30,035	29,831
	Missing	Mean	352.87	334.30	324.39	367.25	360.52	330.16	340.30	339.33
		N	340	318	348	344	339	318	316	315
4-5	F	Mean	369.31	348.41	355.15	364.01	367.03	352.19	354.83	356.50
		N	34,959	33,721	35,173	34,941	34,746	33,721	33,570	33,368
	M	Mean	369.08	346.81	349.51	364.34	367.12	348.62	353.68	354.09
		N	42,229	40,625	42,542	42,265	41,982	40,625	40,386	40,153
	Missing	Mean	358.04	343.73	343.76	347.57	353.00	344.85	348.34	347.23
		N	405	393	414	405	400	393	389	384
6-8	F	Mean	380.54	356.83	357.04	378.67	380.07	357.41	364.14	364.12
		N	31,863	30,227	32,119	31,772	31,561	30,227	30,072	29,791
	M	Mean	379.36	354.61	351.46	380.65	380.44	353.54	362.26	361.53
		N	38,691	36,651	39,036	38,654	38,361	36,651	36,415	36,108
	Missing	Mean	372.24	355.64	348.94	363.81	368.87	353.83	360.74	358.26
		N	432	398	444	432	428	398	395	391
9-12	F	Mean	381.46	380.37	394.20	390.23	386.24	387.89	380.86	387.27
		N	31,842	30,498	32,238	31,571	31,291	30,498	30,257	29,747
	M	Mean	382.64	376.85	388.30	390.69	387.13	383.23	378.76	384.34
		N	37,544	36,017	38,215	37,371	36,868	36,017	35,603	34,988
	Missing	Mean	379.33	377.68	384.11	381.10	381.32	383.56	378.50	383.16
		N	416	413	437	418	409	413	404	399

4.2.1.3 By Cluster by Ethnicity

Table 4.2.1.3

Mean Scale Scores by Cluster by Ethnicity S400 Paper

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Non-Hispanic Asian	Mean	281.92	220.25	235.84	312.56	297.46	228.31	238.73	248.84
		N	31,163	31,163	31,163	31,163	31,163	31,163	31,163	31,163
	Non-Hispanic Pacific Islander	Mean	265.62	184.03	203.90	308.71	287.40	194.20	208.50	221.95
		N	1,677	1,677	1,677	1,677	1,677	1,677	1,677	1,677
	Non-Hispanic Black	Mean	275.85	204.66	217.68	317.25	296.77	211.42	226.00	236.82
		N	13,033	13,033	13,033	13,033	13,033	13,033	13,033	13,033
	Hispanic (Of Any Race)	Mean	269.13	187.65	205.33	303.13	286.35	196.73	212.08	223.41
		N	165,084	165,084	165,084	165,084	165,084	165,084	165,084	165,084
	Non-Hispanic American Indian	Mean	271.90	181.50	194.69	298.01	285.18	188.33	208.61	217.19
		N	2,767	2,767	2,767	2,767	2,767	2,767	2,767	2,767
	Non-Hispanic Multi-racial	Mean	288.21	206.65	220.74	321.76	305.20	213.94	231.09	241.12
		N	1,199	1,199	1,199	1,199	1,199	1,199	1,199	1,199
	Non-Hispanic White	Mean	283.50	205.03	224.71	316.12	300.04	215.13	228.55	240.38
		N	23,598	23,598	23,598	23,598	23,598	23,598	23,598	23,598
	Unknown	Mean	260.26	181.22	196.16	293.29	276.98	188.92	204.92	215.15
		N	7,399	7,399	7,399	7,399	7,399	7,399	7,399	7,399
1	Non-Hispanic Asian	Mean	303.50	293.54	278.46	349.06	326.75	286.60	296.79	298.72
		N	6,874	6,585	7,065	7,005	6,829	6,585	6,455	6,416
	Non-Hispanic Pacific Islander	Mean	289.26	277.19	271.87	339.10	314.40	275.12	280.84	286.68
		N	702	687	714	708	701	687	679	678
	Non-Hispanic Black	Mean	299.97	286.13	269.83	349.00	325.29	278.60	290.47	292.62
		N	5,093	4,836	5,335	5,279	5,043	4,836	4,674	4,628
	Hispanic (Of Any Race)	Mean	300.38	285.53	270.37	344.47	322.95	278.45	290.19	291.82
		N	60,444	57,679	62,414	61,955	60,062	57,679	56,220	55,884
	Non-Hispanic American Indian	Mean	298.07	286.12	264.33	332.70	316.04	277.00	290.15	289.33
		N	1,025	881	1,084	1,064	1,008	881	851	838
	Non-Hispanic Multi-racial	Mean	304.37	289.97	275.21	349.73	328.10	283.24	294.63	296.70
		N	300	289	311	308	297	289	282	279
	Non-Hispanic White	Mean	302.93	289.07	273.87	354.59	329.48	282.07	293.61	296.51
		N	6,006	5,652	6,198	6,142	5,961	5,652	5,501	5,460
	Unknown	Mean	297.13	285.09	266.90	341.85	320.23	277.41	289.13	290.80
		N	991	947	1,057	1,025	979	947	913	902

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
2	Non-Hispanic Asian	Mean	331.57	321.55	294.25	371.44	351.84	308.34	324.76	321.22
		N	6,558	6,424	6,630	6,578	6,516	6,424	6,368	6,326
	Non-Hispanic Pacific Islander	Mean	317.18	305.39	287.87	368.13	343.14	297.23	309.19	311.00
		N	830	817	838	833	826	817	810	807
	Non-Hispanic Black	Mean	324.53	309.83	284.61	367.84	346.60	297.68	314.54	312.36
		N	5,285	5,040	5,354	5,328	5,261	5,040	4,988	4,965
	Hispanic (Of Any Race)	Mean	328.12	312.38	286.96	368.37	348.66	300.11	317.33	314.60
		N	60,098	57,995	60,765	60,357	59,731	57,995	57,487	57,142
	Non-Hispanic American Indian	Mean	322.30	307.88	287.25	360.42	341.94	298.61	312.58	311.72
		N	1,131	1,037	1,162	1,131	1,105	1,037	1,019	997
	Non-Hispanic Multi-racial	Mean	331.80	316.42	288.08	375.26	354.21	302.63	321.51	318.19
		N	323	317	331	330	322	317	311	310
	Non-Hispanic White	Mean	332.15	318.39	290.43	373.06	353.12	304.97	322.89	319.47
		N	5,623	5,414	5,700	5,659	5,591	5,414	5,355	5,327
	Unknown	Mean	320.26	308.71	280.98	360.34	341.40	295.69	312.46	309.56
		N	985	958	1,015	989	968	958	945	929
3	Non-Hispanic Asian	Mean	357.16	339.36	334.52	374.54	366.23	337.39	344.96	345.95
		N	4,642	4,523	4,682	4,644	4,612	4,523	4,495	4,467
	Non-Hispanic Pacific Islander	Mean	340.06	327.61	330.08	371.89	356.21	329.32	331.47	337.32
		N	773	751	780	775	771	751	747	745
	Non-Hispanic Black	Mean	345.52	327.37	323.11	371.74	359.27	325.59	333.06	335.71
		N	4,413	4,231	4,483	4,444	4,377	4,231	4,177	4,147
	Hispanic (Of Any Race)	Mean	346.65	329.19	326.83	370.03	358.75	328.32	334.60	337.34
		N	42,418	40,745	42,770	42,475	42,161	40,745	40,487	40,250
	Non-Hispanic American Indian	Mean	338.65	325.46	331.07	363.98	351.81	328.79	329.74	335.83
		N	1,166	1,067	1,187	1,140	1,124	1,067	1,054	1,016
	Non-Hispanic Multi-racial	Mean	351.06	332.11	325.36	379.07	365.84	329.00	338.30	340.19
		N	240	234	242	241	239	234	232	231
	Non-Hispanic White	Mean	355.59	335.96	330.99	377.58	367.17	333.91	342.09	343.85
		N	3,800	3,651	3,856	3,827	3,778	3,651	3,618	3,598
	Unknown	Mean	343.92	328.29	317.07	359.89	352.47	323.66	333.52	332.47
		N	937	892	960	941	927	892	881	872

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
4-5	Non-Hispanic Asian	Mean	371.82	353.22	355.70	363.67	368.19	354.99	358.97	358.87
		N	5,007	4,889	5,040	4,990	4,961	4,889	4,868	4,826
	Non-Hispanic Pacific Islander	Mean	352.94	336.56	351.22	359.52	356.49	344.17	341.54	347.64
		N	847	831	853	849	844	831	829	826
	Non-Hispanic Black	Mean	367.47	344.95	349.85	365.95	367.18	347.90	351.98	353.66
		N	6,840	6,493	6,920	6,861	6,785	6,493	6,428	6,377
	Hispanic (Of Any Race)	Mean	369.64	347.51	352.13	364.08	367.24	350.22	354.31	355.20
		N	56,986	54,929	57,317	56,993	56,692	54,929	54,680	54,399
	Non-Hispanic American Indian	Mean	359.27	341.26	348.66	353.79	356.79	345.91	346.78	349.19
		N	1,755	1,668	1,787	1,763	1,738	1,668	1,651	1,636
	Non-Hispanic Multi-racial	Mean	369.91	348.10	353.06	367.70	369.03	351.20	354.85	356.58
		N	364	346	369	365	360	346	341	337
	Non-Hispanic White	Mean	372.58	351.58	353.58	370.64	372.03	353.08	358.07	358.65
		N	4,685	4,501	4,717	4,682	4,653	4,501	4,477	4,446
	Unknown	Mean	353.98	338.17	342.31	347.05	350.84	341.02	343.11	343.94
		N	1,109	1,082	1,126	1,108	1,095	1,082	1,071	1,058
6-8	Non-Hispanic Asian	Mean	386.67	362.61	359.64	385.53	386.50	361.66	370.04	369.05
		N	5,031	4,801	5,066	5,013	4,986	4,801	4,780	4,740
	Non-Hispanic Pacific Islander	Mean	365.49	348.67	354.13	376.54	371.25	352.00	353.81	357.62
		N	937	896	950	925	916	896	887	867
	Non-Hispanic Black	Mean	379.51	354.04	352.16	383.22	381.80	353.66	362.02	362.18
		N	6,364	5,921	6,427	6,383	6,325	5,921	5,877	5,841
	Hispanic (Of Any Race)	Mean	379.02	354.68	353.43	378.21	379.07	354.49	362.17	361.75
		N	50,531	47,936	50,912	50,433	50,102	47,936	47,678	47,274
	Non-Hispanic American Indian	Mean	379.10	356.46	354.98	381.96	380.70	356.82	363.36	363.89
		N	1,987	1,831	2,029	1,973	1,947	1,831	1,812	1,778
	Non-Hispanic Multi-racial	Mean	392.96	359.10	359.23	394.93	394.28	359.74	369.24	369.76
		N	278	267	278	275	275	267	267	264
	Non-Hispanic White	Mean	386.00	361.82	356.98	386.78	386.88	359.95	369.38	367.99
		N	4,456	4,280	4,498	4,452	4,417	4,280	4,255	4,219
	Unknown	Mean	375.03	353.88	348.38	367.86	372.13	352.35	360.67	358.54
		N	1,402	1,344	1,439	1,404	1,382	1,344	1,326	1,307

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
9-12	Non-Hispanic Asian	Mean	389.06	385.08	399.15	401.87	395.86	392.83	386.53	393.71
		N	5,408	5,236	5,485	5,391	5,331	5,236	5,188	5,114
	Non-Hispanic Pacific Islander	Mean	380.69	369.92	394.67	403.50	392.34	382.79	373.26	385.54
		N	995	968	1,008	997	989	968	961	955
	Non-Hispanic Black	Mean	378.86	377.44	390.26	394.69	386.92	384.54	378.04	384.84
		N	8,579	8,038	8,734	8,366	8,242	8,038	7,940	7,631
	Hispanic (Of Any Race)	Mean	380.88	377.44	389.77	387.13	384.51	384.17	378.61	384.22
		N	47,809	45,956	48,512	47,693	47,178	45,956	45,518	44,936
	Non-Hispanic American Indian	Mean	383.54	371.68	387.94	385.02	384.14	381.11	375.46	381.63
		N	1,081	1,036	1,114	1,021	1,006	1,036	1,022	960
	Non-Hispanic Multi-racial	Mean	395.16	385.99	397.55	405.99	400.94	392.83	389.49	395.63
		N	218	210	221	214	212	210	208	202
	Non-Hispanic White	Mean	391.52	385.71	395.35	402.11	397.17	391.39	387.74	393.15
		N	4,334	4,137	4,397	4,293	4,247	4,137	4,102	4,023
	Unknown	Mean	384.44	381.63	389.42	389.20	387.74	386.94	382.95	387.46
		N	1,378	1,347	1,419	1,385	1,363	1,347	1,325	1,313

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 S400 Paper

Grade		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Mean	272.32	194.15	211.38	306.09	289.43	203.02	217.59	228.73
	N	245,920	245,920	245,920	245,920	245,920	245,920	245,920	245,920
1	Mean	300.66	286.45	271.19	345.68	323.72	279.37	290.94	292.73
	N	81,435	77,556	84,178	83,486	80,880	77,556	75,575	75,085
2	Mean	328.17	313.22	287.59	368.72	348.88	300.87	317.95	315.21
	N	80,833	78,002	81,795	81,205	80,320	78,002	77,283	76,803
3	Mean	347.71	330.22	327.39	370.79	359.70	329.16	335.67	338.25
	N	58,389	56,094	58,960	58,487	57,989	56,094	55,691	55,326
4	Mean	364.47	343.57	350.66	363.11	364.19	347.53	350.03	352.42
	N	43,996	42,232	44,328	44,032	43,724	42,232	41,983	41,726
5	Mean	375.22	352.64	353.81	365.40	370.69	353.69	359.54	358.67
	N	33,597	32,507	33,801	33,579	33,404	32,507	32,362	32,179
6	Mean	374.20	349.99	352.22	378.77	376.93	351.62	357.42	359.13
	N	25,093	23,619	25,286	25,053	24,892	23,619	23,496	23,311
7	Mean	379.66	355.49	353.71	379.29	379.90	355.08	362.97	362.46
	N	23,141	21,959	23,338	23,094	22,930	21,959	21,832	21,632
8	Mean	386.25	361.87	356.09	381.02	384.12	359.48	369.41	366.78
	N	22,752	21,698	22,975	22,711	22,528	21,698	21,554	21,347
9	Mean	376.56	373.41	386.56	381.29	379.46	380.69	374.51	380.24
	N	25,126	24,028	25,569	25,047	24,720	24,028	23,765	23,397
10	Mean	381.43	377.66	390.30	389.66	385.94	384.49	378.93	384.80
	N	19,303	18,436	19,542	19,133	18,958	18,436	18,282	17,967
11	Mean	385.74	382.91	394.26	396.49	391.49	389.19	383.96	389.82
	N	14,994	14,448	15,216	14,879	14,708	14,448	14,303	14,036
12	Mean	391.36	385.64	398.06	405.27	398.68	392.60	387.55	394.36
	N	10,379	10,016	10,563	10,301	10,182	10,016	9,914	9,734

4.2.2.2 By Grade by Gender

Table 4.2.2.2

Mean Scale Scores by Grade by Gender S400 Paper

Grade	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	F	Mean	277.43	197.48	217.14	310.94	294.41	207.56	221.45	233.40
		N	114,213	114,213	114,213	114,213	114,213	114,213	114,213	114,213
	M	Mean	267.74	191.39	206.48	301.80	284.99	199.17	214.28	224.72
		N	128,977	128,977	128,977	128,977	128,977	128,977	128,977	128,977
	Missing	Mean	275.09	185.92	202.11	306.04	290.78	194.25	212.66	223.01
		N	2,730	2,730	2,730	2,730	2,730	2,730	2,730	2,730
1	F	Mean	302.00	287.66	274.33	347.89	325.44	281.48	292.17	294.70
		N	38,526	36,580	39,655	39,342	38,270	36,580	35,743	35,514
	M	Mean	299.43	285.36	268.39	343.71	322.17	277.47	289.82	290.94
		N	42,518	40,603	44,102	43,745	42,226	40,603	39,470	39,214
	Missing	Mean	301.04	286.69	267.95	342.61	323.39	279.29	291.34	293.14
		N	391	373	421	399	384	373	362	357
2	F	Mean	329.26	314.77	291.58	369.69	349.89	303.60	319.35	317.40
		N	37,955	36,638	38,360	38,097	37,726	36,638	36,329	36,114
	M	Mean	327.23	311.84	284.07	367.92	348.02	298.45	316.72	313.29
		N	42,522	41,019	43,071	42,755	42,243	41,019	40,611	40,351
	Missing	Mean	323.48	312.59	283.11	362.54	343.25	298.63	315.86	311.71
		N	356	345	364	353	351	345	343	338
3	F	Mean	348.27	331.10	332.26	371.91	360.50	331.95	336.41	340.37
		N	26,541	25,507	26,770	26,563	26,364	25,507	25,340	25,180
	M	Mean	347.19	329.43	323.33	369.90	359.01	326.81	335.00	336.44
		N	31,508	30,269	31,842	31,580	31,286	30,269	30,035	29,831
	Missing	Mean	352.87	334.30	324.39	367.25	360.52	330.16	340.30	339.33
		N	340	318	348	344	339	318	316	315
4	F	Mean	364.85	344.55	353.83	363.15	364.40	349.57	350.82	353.86
		N	19,949	19,164	20,076	19,933	19,813	19,164	19,074	18,947
	M	Mean	364.27	342.79	348.12	363.28	364.18	345.90	349.44	351.31
		N	23,792	22,821	23,988	23,843	23,660	22,821	22,666	22,540
	Missing	Mean	352.95	338.49	339.68	345.01	348.98	340.38	343.06	342.64
		N	255	247	264	256	251	247	243	239
5	F	Mean	375.22	353.49	356.90	365.16	370.53	355.65	360.11	359.96
		N	15,010	14,557	15,097	15,008	14,933	14,557	14,496	14,421
	M	Mean	375.29	351.95	351.31	365.70	370.92	352.10	359.09	357.64
		N	18,437	17,804	18,554	18,422	18,322	17,804	17,720	17,613
	Missing	Mean	366.71	352.60	350.93	351.97	359.77	352.40	357.12	354.81
		N	150	146	150	149	149	146	146	145

Grade	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
6	F	Mean	375.31	351.10	355.64	378.32	377.29	353.84	358.54	360.81
		N	11,231	10,551	11,310	11,200	11,132	10,551	10,503	10,414
	M	Mean	373.42	349.10	349.54	379.45	376.84	349.84	356.56	357.85
		N	13,690	12,902	13,794	13,680	13,590	12,902	12,830	12,736
	Missing	Mean	363.61	348.87	343.26	353.68	359.95	348.30	353.37	351.65
		N	172	166	182	173	170	166	163	161
7	F	Mean	379.78	356.48	356.33	377.65	379.09	356.89	363.64	363.42
		N	10,386	9,885	10,470	10,344	10,278	9,885	9,838	9,734
	M	Mean	379.63	354.64	351.58	380.78	380.67	353.58	362.41	361.68
		N	12,615	11,951	12,727	12,609	12,512	11,951	11,871	11,775
	Missing	Mean	373.72	357.66	352.24	366.06	370.80	356.23	362.42	360.57
		N	140	123	141	141	140	123	123	123
8	F	Mean	387.03	363.36	359.31	380.08	384.10	361.79	370.70	368.40
		N	10,246	9,791	10,339	10,228	10,151	9,791	9,731	9,643
	M	Mean	385.64	360.61	353.46	381.84	384.19	357.56	368.33	365.43
		N	12,386	11,798	12,515	12,365	12,259	11,798	11,714	11,597
	Missing	Mean	382.87	363.68	353.63	375.96	379.43	359.56	369.86	365.55
		N	120	109	121	118	118	109	109	107
9	F	Mean	376.55	375.67	390.22	381.54	379.60	383.67	376.12	382.35
		N	11,023	10,521	11,185	10,977	10,858	10,521	10,425	10,278
	M	Mean	376.56	371.62	383.74	381.18	379.39	378.34	373.21	378.55
		N	13,919	13,324	14,188	13,886	13,684	13,324	13,162	12,945
	Missing	Mean	377.07	374.04	381.91	373.76	376.68	381.10	375.66	380.42
		N	184	183	196	184	178	183	178	174
10	F	Mean	380.97	379.53	393.78	389.33	385.45	387.13	380.08	386.45
		N	8,755	8,339	8,836	8,649	8,594	8,339	8,290	8,139
	M	Mean	381.90	376.11	387.46	390.02	386.43	382.30	378.01	383.47
		N	10,433	9,983	10,587	10,367	10,249	9,983	9,880	9,716
	Missing	Mean	374.42	376.47	384.97	382.54	379.29	382.38	375.82	381.19
		N	115	114	119	117	115	114	112	112
11	F	Mean	384.08	384.10	396.57	395.16	389.95	390.83	384.28	390.49
		N	6,972	6,709	7,051	6,909	6,850	6,709	6,653	6,536
	M	Mean	387.24	381.88	392.35	397.80	392.93	387.79	383.69	389.27
		N	7,944	7,661	8,083	7,892	7,781	7,661	7,573	7,424
	Missing	Mean	381.90	382.64	383.71	381.97	383.17	386.54	382.94	385.82
		N	78	78	82	78	77	78	77	76
12	F	Mean	389.34	386.77	400.29	403.92	396.95	394.21	387.66	394.83
		N	5,092	4,929	5,166	5,036	4,989	4,929	4,889	4,794
	M	Mean	393.26	384.51	395.94	406.54	400.31	391.04	387.42	393.88
		N	5,248	5,049	5,357	5,226	5,154	5,049	4,988	4,903
	Missing	Mean	399.38	388.68	393.15	409.64	404.79	392.79	391.03	396.59
		N	39	38	40	39	39	38	37	37

4.2.2.3 By Grade by Ethnicity

Table 4.2.2.3
Mean Scale Scores by Grade by Ethnicity S400 Paper

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Non-Hispanic Asian	Mean	281.92	220.25	235.84	312.56	297.46	228.31	238.73	248.84
		N	31,163	31,163	31,163	31,163	31,163	31,163	31,163	31,163
	Non-Hispanic Pacific Islander	Mean	265.62	184.03	203.90	308.71	287.40	194.20	208.50	221.95
		N	1,677	1,677	1,677	1,677	1,677	1,677	1,677	1,677
	Non-Hispanic Black	Mean	275.85	204.66	217.68	317.25	296.77	211.42	226.00	236.82
		N	13,033	13,033	13,033	13,033	13,033	13,033	13,033	13,033
	Hispanic (Of Any Race)	Mean	269.13	187.65	205.33	303.13	286.35	196.73	212.08	223.41
		N	165,084	165,084	165,084	165,084	165,084	165,084	165,084	165,084
	Non-Hispanic American Indian	Mean	271.90	181.50	194.69	298.01	285.18	188.33	208.61	217.19
		N	2,767	2,767	2,767	2,767	2,767	2,767	2,767	2,767
	Non-Hispanic Multi-racial	Mean	288.21	206.65	220.74	321.76	305.20	213.94	231.09	241.12
		N	1,199	1,199	1,199	1,199	1,199	1,199	1,199	1,199
	Non-Hispanic White	Mean	283.50	205.03	224.71	316.12	300.04	215.13	228.55	240.38
		N	23,598	23,598	23,598	23,598	23,598	23,598	23,598	23,598
	Unknown	Mean	260.26	181.22	196.16	293.29	276.98	188.92	204.92	215.15
		N	7,399	7,399	7,399	7,399	7,399	7,399	7,399	7,399
1	Non-Hispanic Asian	Mean	303.50	293.54	278.46	349.06	326.75	286.60	296.79	298.72
		N	6,874	6,585	7,065	7,005	6,829	6,585	6,455	6,416
	Non-Hispanic Pacific Islander	Mean	289.26	277.19	271.87	339.10	314.40	275.12	280.84	286.68
		N	702	687	714	708	701	687	679	678
	Non-Hispanic Black	Mean	299.97	286.13	269.83	349.00	325.29	278.60	290.47	292.62
		N	5,093	4,836	5,335	5,279	5,043	4,836	4,674	4,628
	Hispanic (Of Any Race)	Mean	300.38	285.53	270.37	344.47	322.95	278.45	290.19	291.82
		N	60,444	57,679	62,414	61,955	60,062	57,679	56,220	55,884
	Non-Hispanic American Indian	Mean	298.07	286.12	264.33	332.70	316.04	277.00	290.15	289.33
		N	1,025	881	1,084	1,064	1,008	881	851	838
	Non-Hispanic Multi-racial	Mean	304.37	289.97	275.21	349.73	328.10	283.24	294.63	296.70
		N	300	289	311	308	297	289	282	279
	Non-Hispanic White	Mean	302.93	289.07	273.87	354.59	329.48	282.07	293.61	296.51
		N	6,006	5,652	6,198	6,142	5,961	5,652	5,501	5,460
	Unknown	Mean	297.13	285.09	266.90	341.85	320.23	277.41	289.13	290.80
		N	991	947	1,057	1,025	979	947	913	902

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
2	Non-Hispanic Asian	Mean	331.57	321.55	294.25	371.44	351.84	308.34	324.76	321.22
		N	6,558	6,424	6,630	6,578	6,516	6,424	6,368	6,326
	Non-Hispanic Pacific Islander	Mean	317.18	305.39	287.87	368.13	343.14	297.23	309.19	311.00
		N	830	817	838	833	826	817	810	807
	Non-Hispanic Black	Mean	324.53	309.83	284.61	367.84	346.60	297.68	314.54	312.36
		N	5,285	5,040	5,354	5,328	5,261	5,040	4,988	4,965
	Hispanic (Of Any Race)	Mean	328.12	312.38	286.96	368.37	348.66	300.11	317.33	314.60
		N	60,098	57,995	60,765	60,357	59,731	57,995	57,487	57,142
	Non-Hispanic American Indian	Mean	322.30	307.88	287.25	360.42	341.94	298.61	312.58	311.72
		N	1,131	1,037	1,162	1,131	1,105	1,037	1,019	997
	Non-Hispanic Multi-racial	Mean	331.80	316.42	288.08	375.26	354.21	302.63	321.51	318.19
		N	323	317	331	330	322	317	311	310
	Non-Hispanic White	Mean	332.15	318.39	290.43	373.06	353.12	304.97	322.89	319.47
		N	5,623	5,414	5,700	5,659	5,591	5,414	5,355	5,327
	Unknown	Mean	320.26	308.71	280.98	360.34	341.40	295.69	312.46	309.56
		N	985	958	1,015	989	968	958	945	929
3	Non-Hispanic Asian	Mean	357.16	339.36	334.52	374.54	366.23	337.39	344.96	345.95
		N	4,642	4,523	4,682	4,644	4,612	4,523	4,495	4,467
	Non-Hispanic Pacific Islander	Mean	340.06	327.61	330.08	371.89	356.21	329.32	331.47	337.32
		N	773	751	780	775	771	751	747	745
	Non-Hispanic Black	Mean	345.52	327.37	323.11	371.74	359.27	325.59	333.06	335.71
		N	4,413	4,231	4,483	4,444	4,377	4,231	4,177	4,147
	Hispanic (Of Any Race)	Mean	346.65	329.19	326.83	370.03	358.75	328.32	334.60	337.34
		N	42418	40745	42770	42475	42161	40745	40487	40250
	Non-Hispanic American Indian	Mean	338.65	325.46	331.07	363.98	351.81	328.79	329.74	335.83
		N	1,166	1,067	1,187	1,140	1,124	1,067	1,054	1,016
	Non-Hispanic Multi-racial	Mean	351.06	332.11	325.36	379.07	365.84	329.00	338.30	340.19
		N	240	234	242	241	239	234	232	231
	Non-Hispanic White	Mean	355.59	335.96	330.99	377.58	367.17	333.91	342.09	343.85
		N	3,800	3,651	3,856	3,827	3,778	3,651	3,618	3,598
	Unknown	Mean	343.92	328.29	317.07	359.89	352.47	323.66	333.52	332.47
		N	937	892	960	941	927	892	881	872
4	Non-Hispanic Asian	Mean	368.49	349.27	354.79	363.65	366.57	352.46	355.24	356.62
		N	2,823	2,754	2,838	2,811	2,796	2,754	2,743	2,717
	Non-Hispanic Pacific Islander	Mean	348.08	333.65	348.88	354.03	351.17	341.40	338.01	344.01
		N	461	451	465	464	460	451	449	448
	Non-Hispanic Black	Mean	360.73	339.47	347.96	362.81	362.31	344.17	346.12	349.59
		N	3,765	3,561	3,820	3,785	3,732	3,561	3,516	3,485
	Hispanic (Of Any Race)	Mean	365.09	343.72	350.86	363.37	364.61	347.67	350.31	352.62
		N	32,543	31,263	32,747	32,569	32,378	31,263	31,105	30,949
	Non-Hispanic American Indian	Mean	355.00	337.52	346.20	351.93	353.79	342.77	343.05	346.19
		N	916	858	932	917	904	858	850	840
	Non-Hispanic Multi-racial	Mean	364.63	346.12	351.13	366.49	365.77	349.35	351.66	354.33
		N	215	202	219	217	213	202	198	196
	Non-Hispanic White	Mean	367.24	346.98	352.01	369.68	368.84	350.01	353.24	355.51
		N	2,630	2,517	2,652	2,629	2,609	2,517	2,502	2,481
	Unknown	Mean	350.97	334.99	340.27	344.13	347.89	338.50	340.02	341.21
		N	643	626	655	640	632	626	620	610

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
5	Non-Hispanic Asian	Mean	376.12	358.31	356.88	363.68	370.29	358.25	363.79	361.76
		N	2,184	2,135	2,202	2,179	2,165	2,135	2,125	2,109
	Non-Hispanic Pacific Islander	Mean	358.75	340.01	354.03	366.13	362.86	347.45	345.70	351.94
		N	386	380	388	385	384	380	380	378
	Non-Hispanic Black	Mean	375.73	351.60	352.17	369.81	373.14	352.43	359.06	358.56
		N	3,075	2,932	3,100	3,076	3,053	2,932	2,912	2,892
	Hispanic (Of Any Race)	Mean	375.69	352.52	353.83	365.02	370.74	353.59	359.59	358.60
		N	24,443	23,666	24,570	24,424	24,314	23,666	23,575	23,450
	Non-Hispanic American Indian	Mean	363.92	345.23	351.34	355.80	360.06	349.25	350.74	352.37
		N	839	810	855	846	834	810	801	796
	Non-Hispanic Multi-racial	Mean	377.52	350.89	355.87	369.47	373.74	353.80	359.27	359.71
		N	149	144	150	148	147	144	143	141
	Non-Hispanic White	Mean	379.42	357.40	355.58	371.86	376.10	356.96	364.19	362.61
		N	2,055	1,984	2,065	2,053	2,044	1,984	1,975	1,965
	Unknown	Mean	358.14	342.54	345.15	351.04	354.86	344.47	347.36	347.66
		N	466	456	471	468	463	456	451	448
6	Non-Hispanic Asian	Mean	380.96	357.22	357.22	382.16	381.91	357.73	364.65	364.97
		N	1,784	1,703	1,792	1,782	1,775	1,703	1,698	1,690
	Non-Hispanic Pacific Islander	Mean	358.83	343.79	352.46	370.68	365.07	348.48	348.30	353.38
		N	316	305	318	317	315	305	303	302
	Non-Hispanic Black	Mean	373.28	348.26	349.77	381.11	377.71	349.66	356.06	358.15
		N	2,137	1,974	2,163	2,147	2,123	1,974	1,958	1,944
	Hispanic (Of Any Race)	Mean	373.87	349.14	352.04	378.68	376.70	351.03	356.68	358.59
		N	17,853	16,803	17,967	17,803	17,705	16,803	16,724	16,589
	Non-Hispanic American Indian	Mean	370.32	349.94	353.21	377.38	373.96	352.37	356.08	358.67
		N	764	711	776	767	758	711	705	700
	Non-Hispanic Multi-racial	Mean	394.36	355.97	361.04	391.48	393.26	359.06	367.72	369.25
		N	113	109	113	111	111	109	109	107
	Non-Hispanic White	Mean	378.42	355.59	354.56	382.71	381.11	355.63	362.69	363.26
		N	1,603	1,522	1,617	1,600	1,588	1,522	1,513	1,499
	Unknown	Mean	363.70	346.16	341.06	353.14	359.48	345.38	351.93	350.28
		N	523	492	540	526	517	492	486	480
7	Non-Hispanic Asian	Mean	386.34	362.33	359.22	385.15	386.25	361.45	369.69	368.76
		N	1,639	1,565	1,655	1,630	1,619	1,565	1,556	1,537
	Non-Hispanic Pacific Islander	Mean	364.46	347.85	354.98	375.44	369.97	351.97	353.04	357.18
		N	310	294	314	304	300	294	292	283
	Non-Hispanic Black	Mean	377.89	353.77	351.42	380.71	379.58	353.02	361.31	361.06
		N	2,121	1,974	2,132	2,119	2,109	1,974	1,964	1,953
	Hispanic (Of Any Race)	Mean	378.89	354.52	353.16	377.75	378.77	354.25	362.04	361.51
		N	16,496	15,657	16,612	16,459	16,357	15,657	15,579	15,444
	Non-Hispanic American Indian	Mean	379.76	357.79	353.61	384.98	382.56	357.36	364.50	365.20
		N	605	553	626	606	594	553	544	535
	Non-Hispanic Multi-racial	Mean	388.99	356.17	357.52	396.52	393.05	357.33	366.05	367.80
		N	96	93	96	96	96	93	93	93
	Non-Hispanic White	Mean	387.28	362.33	357.55	387.45	387.82	360.48	370.15	368.56
		N	1,439	1,399	1,454	1,436	1,424	1,399	1,391	1,378
	Unknown	Mean	375.80	353.56	350.68	372.55	374.72	352.98	360.59	359.46
		N	435	424	449	444	431	424	413	409

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
8	Non-Hispanic Asian	Mean	393.34	368.86	362.73	389.66	391.86	366.23	376.40	373.91
		N	1,608	1,533	1,619	1,601	1,592	1,533	1,526	1,513
	Non-Hispanic Pacific Islander	Mean	373.29	354.51	354.95	383.76	378.99	355.63	360.30	362.60
		N	311	297	318	304	301	297	292	282
	Non-Hispanic Black	Mean	387.46	360.09	355.32	387.87	388.19	358.30	368.68	367.34
		N	2,106	1,973	2,132	2,117	2,093	1,973	1,955	1,944
	Hispanic (Of Any Race)	Mean	384.83	360.87	355.22	378.16	382.00	358.49	368.26	365.43
		N	16,182	15,476	16,333	16,171	16,040	15,476	15,375	15,241
	Non-Hispanic American Indian	Mean	389.31	363.35	358.53	384.78	387.43	361.89	371.38	369.34
		N	618	567	627	600	595	567	563	543
	Non-Hispanic Multi-racial	Mean	396.19	368.55	358.65	398.34	397.66	364.32	376.35	373.45
		N	69	65	69	68	68	65	65	64
	Non-Hispanic White	Mean	393.28	368.27	359.14	390.70	392.44	364.25	376.06	372.67
		N	1,414	1,359	1,427	1,416	1,405	1,359	1,351	1,342
	Unknown	Mean	387.61	363.09	354.87	380.91	384.61	359.74	370.70	367.12
		N	444	428	450	434	434	428	427	418
9	Non-Hispanic Asian	Mean	385.58	382.56	396.11	398.18	392.40	390.14	383.80	390.91
		N	1,864	1,795	1,888	1,862	1,845	1,795	1,779	1,761
	Non-Hispanic Pacific Islander	Mean	377.61	367.10	392.56	398.99	388.57	380.35	370.29	382.66
		N	438	426	443	436	434	426	425	421
	Non-Hispanic Black	Mean	371.61	372.10	383.76	383.59	377.77	378.90	372.18	377.92
		N	2,346	2,176	2,395	2,287	2,248	2,176	2,152	2,064
	Hispanic (Of Any Race)	Mean	375.32	372.22	385.45	377.82	377.18	379.42	373.25	378.69
		N	17,977	17,242	18,289	17,988	17,750	17,242	17,041	16,834
	Non-Hispanic American Indian	Mean	379.14	364.84	384.39	373.34	376.10	375.73	369.21	375.28
		N	424	404	432	399	396	404	401	376
	Non-Hispanic Multi-racial	Mean	394.33	383.43	396.52	403.63	398.92	390.75	387.94	393.69
		N	84	80	85	84	83	80	79	78
	Non-Hispanic White	Mean	384.41	380.75	390.79	394.52	389.74	386.66	382.04	387.60
		N	1455	1376	1480	1452	1435	1376	1367	1348
	Unknown	Mean	381.46	377.87	386.95	381.22	382.15	384.35	379.30	383.77
		N	538	529	557	539	529	529	521	515
10	Non-Hispanic Asian	Mean	389.51	384.90	399.10	400.76	395.36	392.47	386.44	393.16
		N	1,400	1,352	1,415	1,393	1,382	1,352	1,341	1,324
	Non-Hispanic Pacific Islander	Mean	381.81	370.45	394.38	404.47	393.48	382.92	374.18	386.27
		N	246	240	250	248	245	240	237	236
	Non-Hispanic Black	Mean	382.29	377.71	391.63	397.94	390.45	385.20	379.14	386.39
		N	2,245	2,088	2,282	2,193	2,164	2,088	2,063	1,988
	Hispanic (Of Any Race)	Mean	379.67	376.48	388.74	385.98	383.25	383.06	377.55	383.02
		N	13,417	12,838	13,566	13,334	13,221	12,838	12,744	12,562
	Non-Hispanic American Indian	Mean	374.90	368.74	383.87	380.26	377.66	377.41	370.76	376.91
		N	326	312	341	312	305	312	306	290
	Non-Hispanic Multi-racial	Mean	391.64	388.37	398.26	402.63	397.27	394.61	390.18	395.60
		N	64	62	65	64	63	62	61	60
	Non-Hispanic White	Mean	391.37	385.25	395.90	400.71	396.40	391.39	387.40	392.94
		N	1,219	1,167	1,231	1,201	1,194	1,167	1,158	1,137
	Unknown	Mean	380.52	378.38	388.97	391.40	386.64	384.37	379.60	385.46
		N	386	377	392	388	384	377	372	370

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
11	Non-Hispanic Asian	Mean	391.92	388.30	401.46	403.84	398.34	395.57	389.68	396.38
		N	1,209	1,175	1,226	1,205	1,189	1,175	1,164	1,144
	Non-Hispanic Pacific Islander	Mean	382.29	372.72	397.77	408.77	395.53	385.44	375.69	388.19
		N	187	185	190	189	186	185	182	181
	Non-Hispanic Black	Mean	380.35	379.94	392.85	396.42	388.31	386.90	380.25	386.93
		N	2,140	2,028	2,172	2,081	2,054	2,028	2,004	1,923
	Hispanic (Of Any Race)	Mean	385.21	382.39	393.50	394.50	390.31	388.49	383.38	389.00
		N	10,015	9,681	10,159	9,989	9,881	9,681	9,588	9,464
	Non-Hispanic American Indian	Mean	395.34	381.77	395.57	403.56	398.99	390.22	386.15	392.69
		N	197	188	202	185	182	188	186	174
	Non-Hispanic Multi-racial	Mean	393.78	383.77	399.15	406.74	401.18	391.79	387.00	394.74
		N	40	39	40	39	39	39	39	38
	Non-Hispanic White	Mean	395.53	390.55	397.19	407.24	401.81	394.72	392.40	396.95
		N	982	935	993	965	956	935	928	903
	Unknown	Mean	377.74	381.35	385.26	382.38	381.18	385.49	381.26	384.82
		N	224	217	234	226	221	217	212	209
12	Non-Hispanic Asian	Mean	391.64	386.14	402.27	408.35	400.37	395.15	387.97	396.69
		N	935	914	956	931	915	914	904	885
	Non-Hispanic Pacific Islander	Mean	386.94	374.64	398.05	409.39	398.46	387.23	378.43	390.34
		N	124	117	125	124	124	117	117	117
	Non-Hispanic Black	Mean	382.15	380.88	393.88	402.82	392.58	388.05	381.50	389.18
		N	1,848	1,746	1,885	1,805	1,776	1,746	1,721	1,656
	Hispanic (Of Any Race)	Mean	392.27	386.24	398.22	404.26	398.63	392.94	388.20	394.57
		N	6,400	6,195	6,498	6,382	6,326	6,195	6,145	6,076
	Non-Hispanic American Indian	Mean	401.18	385.20	397.83	406.72	404.11	393.33	390.64	396.88
		N	134	132	139	125	123	132	129	120
	Non-Hispanic Multi-racial	Mean	406.80	390.97	396.81	420.22	415.37	396.14	395.62	402.85
		N	30	29	31	27	27	29	29	26
	Non-Hispanic White	Mean	401.24	390.03	401.50	413.61	407.95	396.56	393.69	399.91
		N	678	659	693	675	662	659	649	635
	Unknown	Mean	404.53	396.24	400.09	410.71	408.86	398.79	398.86	402.04
		N	230	224	236	232	229	224	220	219

4.2.3 Correlations Among Scale Scores by Grade Level Cluster

Table 4.2.3A

Correlations Among Scale Scores: K S400 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.536	.556	.791
	N	245,920	245,920	245,920	245,920
Reading	Pearson Correlation		1	.723	.494
	N		245,920	245,920	245,920
Writing	Pearson Correlation			1	.550
	N			245,920	245,920
Speaking	Pearson Correlation				1
	N				245,920

Table 4.2.3B

Correlations Among Scale Scores: 1 S400 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.650	.485	.492
	N	81,435	75,575	81,435	80,880
Reading	Pearson Correlation		1	.530	.424
	N		77,556	77,556	77,033
Writing	Pearson Correlation			1	.452
	N			84,178	83,486
Speaking	Pearson Correlation				1
	N				83,486

Table 4.2.3C

Correlations Among Scale Scores: 2 S400 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.664	.505	.511
	N	80,833	77,283	80,833	80,320
Reading	Pearson Correlation		1	.608	.449
	N		78,002	78,002	77,512
Writing	Pearson Correlation			1	.518
	N			81,795	81,205
Speaking	Pearson Correlation				1
	N				81,205

Table 4.2.3D

Correlations Among Scale Scores: 3 S400 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.746	.565	.482
	N	58,389	55,691	58,389	57,989
Reading	Pearson Correlation		1	.631	.476
	N		56,094	56,094	55,719
Writing	Pearson Correlation			1	.588
	N			58,960	58,487
Speaking	Pearson Correlation				1
	N				58,487

Table 4.2.3E

Correlations Among Scale Scores: 4-5 S400 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.763	.554	.567
	N	77,593	74,345	77,593	77,128
Reading	Pearson Correlation		1	.603	.560
	N		74,739	74,739	74,283
Writing	Pearson Correlation			1	.611
	N			78,129	77,611
Speaking	Pearson Correlation				1
	N				77,611

Table 4.2.3F

Correlations Among Scale Scores: 6-8 S400 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.695	.590	.620
	N	70,986	66,882	70,986	70,350
Reading	Pearson Correlation		1	.595	.546
	N		67,276	67,276	66,655
Writing	Pearson Correlation			1	.647
	N			71,599	70,858
Speaking	Pearson Correlation				1
	N				70,858

Table 4.2.3G

Correlations Among Scale Scores: 9-12 S400 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.737	.645	.638
	N	69,802	66,264	69,802	68,568
Reading	Pearson Correlation		1	.650	.573
	N		66,928	66,928	65,725
Writing	Pearson Correlation			1	.672
	N			70,890	69,360
Speaking	Pearson Correlation				1
	N				69,360

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 S400 Paper

Cluster	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	29,491	13,289	23,327	37,380	75,357	67,076	245,920
K (accountability)	-	59,186	23,544	20,757	14,070	38,431	89,932	245,920
1	A	1,785	3,300	7,484	16,879	n/a	n/a	29,448
	B	136	537	1,173	1,225	29,154	n/a	32,225
	C	55	482	3,637	1,983	5,145	8,460	19,762
2	A	1,433	2,174	1,899	4,670	n/a	n/a	10,176
	B	119	362	1,843	1,505	29,449	n/a	33,278
	C	42	633	3,702	2,942	10,089	19,971	37,379
3	A	263	1,879	2,419	5,821	n/a	n/a	10,382
	B	50	818	2,933	1,807	15,978	n/a	21,586
	C	7	192	1,631	1,230	6,761	16,600	26,421
4-5	A	836	3,073	3,282	6,090	n/a	n/a	13,281
	B	152	935	2,602	4,228	15,756	n/a	23,673
	C	7	264	2,165	3,753	9,870	24,580	40,639
6-8	A	2,940	6,334	3,944	3,578	n/a	n/a	16,796
	B	244	3,240	4,933	5,490	7,496	n/a	21,403
	C	8	186	1,865	4,403	10,722	15,603	32,787
9-12	A	8,023	6,973	2,177	1,719	n/a	n/a	18,892
	B	728	2,711	5,830	5,635	6,086	n/a	20,990
	C	118	1,055	3,883	8,961	8,215	7,688	29,920

Table 4.3.1.1B

Proficiency Level by Cluster By Tier (Percent): Listening S400 Paper

Cluster	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	12.0%	5.4%	9.5%	15.2%	30.6%	27.3%	100.0%
K (accountability)	-	24.1%	9.6%	8.4%	5.7%	15.6%	36.6%	100.0%
1	A	6.1%	11.2%	25.4%	57.3%	n/a	n/a	100.0%
	B	0.4%	1.7%	3.6%	3.8%	90.5%	n/a	100.0%
	C	0.3%	2.4%	18.4%	10.0%	26.0%	42.8%	100.0%
2	A	14.1%	21.4%	18.7%	45.9%	n/a	n/a	100.0%
	B	0.4%	1.1%	5.5%	4.5%	88.5%	n/a	100.0%
	C	0.1%	1.7%	9.9%	7.9%	27.0%	53.4%	100.0%
3	A	2.5%	18.1%	23.3%	56.1%	n/a	n/a	100.0%
	B	0.2%	3.8%	13.6%	8.4%	74.0%	n/a	100.0%
	C	0.0%	0.7%	6.2%	4.7%	25.6%	62.8%	100.0%
4-5	A	6.3%	23.1%	24.7%	45.9%	n/a	n/a	100.0%
	B	0.6%	3.9%	11.0%	17.9%	66.6%	n/a	100.0%
	C	0.0%	0.6%	5.3%	9.2%	24.3%	60.5%	100.0%
6-8	A	17.5%	37.7%	23.5%	21.3%	n/a	n/a	100.0%
	B	1.1%	15.1%	23.0%	25.7%	35.0%	n/a	100.0%
	C	0.0%	0.6%	5.7%	13.4%	32.7%	47.6%	100.0%
9-12	A	42.5%	36.9%	11.5%	9.1%	n/a	n/a	100.0%
	B	3.5%	12.9%	27.8%	26.8%	29.0%	n/a	100.0%
	C	0.4%	3.5%	13.0%	29.9%	27.5%	25.7%	100.0%

4.3.1.2 By Grade by Tier

Table 4.3.1.2A

Proficiency Level by Grade By Tier (Count): Listening S400 Paper

Grade	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	29,491	13,289	23,327	37,380	75,357	67,076	245,920
K (accountability)	-	59,186	23,544	20,757	14,070	38,431	89,932	245,920
1	A	1,785	3,300	7,484	16,879	n/a	n/a	29,448
	B	136	537	1,173	1,225	29,154	n/a	32,225
	C	55	482	3,637	1,983	5,145	8,460	19,762
2	A	1,433	2,174	1,899	4,670	n/a	n/a	10,176
	B	119	362	1,843	1,505	29,449	n/a	33,278
	C	42	633	3,702	2,942	10,089	19,971	37,379
3	A	263	1,879	2,419	5,821	n/a	n/a	10,382
	B	50	818	2,933	1,807	15,978	n/a	21,586
	C	7	192	1,631	1,230	6,761	16,600	26,421
4	A	344	1,542	1,742	3,515	n/a	n/a	7,143
	B	73	523	1,552	2,535	10,077	n/a	14,760
	C	4	144	918	1,702	5,874	13,451	22,093
5	A	492	1,531	1,540	2,575	n/a	n/a	6,138
	B	79	412	1,050	1,693	5,679	n/a	8,913
	C	3	120	1,247	2,051	3,996	11,129	18,546
6	A	656	1,964	1,580	1,499	n/a	n/a	5,699
	B	42	865	1,865	1,912	3,065	n/a	7,749
	C	0	63	765	1,412	4,059	5,346	11,645
7	A	964	2,223	1,451	1,055	n/a	n/a	5,693
	B	76	1,045	1,833	1,875	2,081	n/a	6,910
	C	7	48	693	1,298	3,941	4,551	10,538
8	A	1,320	2,147	913	1,024	n/a	n/a	5,404
	B	126	1,330	1,235	1,703	2,350	n/a	6,744
	C	1	75	407	1,693	2,722	5,706	10,604
9	A	2,915	3,565	591	867	n/a	n/a	7,938
	B	104	861	1,912	1,714	2,396	n/a	6,987
	C	2	192	933	2,623	3,940	2,511	10,201
10	A	2,186	2,110	806	398	n/a	n/a	5,500
	B	174	651	1,710	1,402	1,970	n/a	5,907
	C	15	289	1,022	2,493	1,993	2,084	7,896
11	A	1,859	910	651	315	n/a	n/a	3,735
	B	205	746	1,030	1,710	1,011	n/a	4,702
	C	42	218	896	1,943	1,660	1,798	6,557
12	A	1,063	388	129	139	n/a	n/a	1,719
	B	245	453	1,178	809	709	n/a	3,394
	C	59	356	1,032	1,902	622	1,295	5,266

Table 4.3.1.2B

Proficiency Level by Grade By Tier (Percent): Listening S400 Paper

Grade	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	12.0%	5.4%	9.5%	15.2%	30.6%	27.3%	100.0%
K (accountability)	-	24.1%	9.6%	8.4%	5.7%	15.6%	36.6%	100.0%
1	A	6.1%	11.2%	25.4%	57.3%	n/a	n/a	100.0%
	B	0.4%	1.7%	3.6%	3.8%	90.5%	n/a	100.0%
	C	0.3%	2.4%	18.4%	10.0%	26.0%	42.8%	100.0%
2	A	14.1%	21.4%	18.7%	45.9%	n/a	n/a	100.0%
	B	0.4%	1.1%	5.5%	4.5%	88.5%	n/a	100.0%
	C	0.1%	1.7%	9.9%	7.9%	27.0%	53.4%	100.0%
3	A	2.5%	18.1%	23.3%	56.1%	n/a	n/a	100.0%
	B	0.2%	3.8%	13.6%	8.4%	74.0%	n/a	100.0%
	C	0.0%	0.7%	6.2%	4.7%	25.6%	62.8%	100.0%
4	A	4.8%	21.6%	24.4%	49.2%	n/a	n/a	100.0%
	B	0.5%	3.5%	10.5%	17.2%	68.3%	n/a	100.0%
	C	0.0%	0.7%	4.2%	7.7%	26.6%	60.9%	100.0%
5	A	8.0%	24.9%	25.1%	42.0%	n/a	n/a	100.0%
	B	0.9%	4.6%	11.8%	19.0%	63.7%	n/a	100.0%
	C	0.0%	0.6%	6.7%	11.1%	21.5%	60.0%	100.0%
6	A	11.5%	34.5%	27.7%	26.3%	n/a	n/a	100.0%
	B	0.5%	11.2%	24.1%	24.7%	39.6%	n/a	100.0%
	C	0.0%	0.5%	6.6%	12.1%	34.9%	45.9%	100.0%
7	A	16.9%	39.0%	25.5%	18.5%	n/a	n/a	100.0%
	B	1.1%	15.1%	26.5%	27.1%	30.1%	n/a	100.0%
	C	0.1%	0.5%	6.6%	12.3%	37.4%	43.2%	100.0%
8	A	24.4%	39.7%	16.9%	18.9%	n/a	n/a	100.0%
	B	1.9%	19.7%	18.3%	25.3%	34.8%	n/a	100.0%
	C	0.0%	0.7%	3.8%	16.0%	25.7%	53.8%	100.0%
9	A	36.7%	44.9%	7.4%	10.9%	n/a	n/a	100.0%
	B	1.5%	12.3%	27.4%	24.5%	34.3%	n/a	100.0%
	C	0.0%	1.9%	9.1%	25.7%	38.6%	24.6%	100.0%
10	A	39.7%	38.4%	14.7%	7.2%	n/a	n/a	100.0%
	B	2.9%	11.0%	28.9%	23.7%	33.4%	n/a	100.0%
	C	0.2%	3.7%	12.9%	31.6%	25.2%	26.4%	100.0%
11	A	49.8%	24.4%	17.4%	8.4%	n/a	n/a	100.0%
	B	4.4%	15.9%	21.9%	36.4%	21.5%	n/a	100.0%
	C	0.6%	3.3%	13.7%	29.6%	25.3%	27.4%	100.0%
12	A	61.8%	22.6%	7.5%	8.1%	n/a	n/a	100.0%
	B	7.2%	13.3%	34.7%	23.8%	20.9%	n/a	100.0%
	C	1.1%	6.8%	19.6%	36.1%	11.8%	24.6%	100.0%

4.3.1.3 By Grade

Table 4.3.1.3A

Proficiency Level by Grade (Count): Listening S400 Paper

	Listening Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	29,491	13,289	23,327	37,380	75,357	67,076	245,920
K (accountability)	59,186	23,544	20,757	14,070	38,431	89,932	245,920
1	1,976	4,319	12,294	20,087	34,299	8,460	81,435
2	1,594	3,169	7,444	9,117	39,538	19,971	80,833
3	320	2,889	6,983	8,858	22,739	16,600	58,389
4	421	2,209	4,212	7,752	15,951	13,451	43,996
5	574	2,063	3,837	6,319	9,675	11,129	33,597
6	698	2,892	4,210	4,823	7,124	5,346	25,093
7	1,047	3,316	3,977	4,228	6,022	4,551	23,141
8	1,447	3,552	2,555	4,420	5,072	5,706	22,752
9	3,021	4,618	3,436	5,204	6,336	2,511	25,126
10	2,375	3,050	3,538	4,293	3,963	2,084	19,303
11	2,106	1,874	2,577	3,968	2,671	1,798	14,994
12	1,367	1,197	2,339	2,850	1,331	1,295	10,379

Table 4.3.1.3B

Proficiency Level by Grade (Percent): Listening S400 Paper

	Listening Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	12.0%	5.4%	9.5%	15.2%	30.6%	27.3%	100.0%
K (accountability)	24.1%	9.6%	8.4%	5.7%	15.6%	36.6%	100.0%
1	2.4%	5.3%	15.1%	24.7%	42.1%	10.4%	100.0%
2	2.0%	3.9%	9.2%	11.3%	48.9%	24.7%	100.0%
3	0.5%	4.9%	12.0%	15.2%	38.9%	28.4%	100.0%
4	1.0%	5.0%	9.6%	17.6%	36.3%	30.6%	100.0%
5	1.7%	6.1%	11.4%	18.8%	28.8%	33.1%	100.0%
6	2.8%	11.5%	16.8%	19.2%	28.4%	21.3%	100.0%
7	4.5%	14.3%	17.2%	18.3%	26.0%	19.7%	100.0%
8	6.4%	15.6%	11.2%	19.4%	22.3%	25.1%	100.0%
9	12.0%	18.4%	13.7%	20.7%	25.2%	10.0%	100.0%
10	12.3%	15.8%	18.3%	22.2%	20.5%	10.8%	100.0%
11	14.0%	12.5%	17.2%	26.5%	17.8%	12.0%	100.0%
12	13.2%	11.5%	22.5%	27.5%	12.8%	12.5%	100.0%

4.3.2 Reading

4.3.2.1 By Cluster by Tier

Table 4.3.2.1A

Proficiency Level by Cluster By Tier (Count): Reading S400 Paper

Cluster	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	54,136	32,382	46,058	19,956	25,147	68,241	245,920
K (accountability)	-	160,103	17,576	11,951	13,821	42,469	0	245,920
1	A	5,705	6,983	6,135	10,063	n/a	n/a	28,886
	B	42	482	4,201	5,485	19,727	n/a	29,937
	C	57	475	1,516	4,199	5,017	7,469	18,733
2	A	3,488	2,504	1,099	2,915	n/a	n/a	10,006
	B	205	1,973	6,266	3,487	19,935	n/a	31,866
	C	169	1,234	4,923	4,024	7,612	18,168	36,130
3	A	1,542	3,350	1,810	3,464	n/a	n/a	10,166
	B	93	1,838	3,773	3,024	12,054	n/a	20,782
	C	15	275	1,533	1,546	9,698	12,079	25,146
4-5	A	3,707	3,742	2,109	3,479	n/a	n/a	13,037
	B	577	2,882	5,500	2,551	11,266	n/a	22,776
	C	51	834	4,890	3,740	9,929	19,482	38,926
6-8	A	5,066	6,583	2,442	2,119	n/a	n/a	16,210
	B	835	4,814	6,497	1,397	6,195	n/a	19,738
	C	373	5,749	10,141	4,472	5,886	4,707	31,328
9-12	A	5,842	7,015	2,590	3,187	n/a	n/a	18,634
	B	2,011	7,104	3,478	1,834	5,542	n/a	19,969
	C	204	2,739	3,213	3,297	5,950	12,922	28,325

Table 4.3.2.1B

Proficiency Level by Cluster By Tier (Percent): Reading S400 Paper

Cluster	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	22.0%	13.2%	18.7%	8.1%	10.2%	27.7%	100.0%
K (accountability)	-	65.1%	7.1%	4.9%	5.6%	17.3%	0.0%	100.0%
1	A	19.8%	24.2%	21.2%	34.8%	n/a	n/a	100.0%
	B	0.1%	1.6%	14.0%	18.3%	65.9%	n/a	100.0%
	C	0.3%	2.5%	8.1%	22.4%	26.8%	39.9%	100.0%
2	A	34.9%	25.0%	11.0%	29.1%	n/a	n/a	100.0%
	B	0.6%	6.2%	19.7%	10.9%	62.6%	n/a	100.0%
	C	0.5%	3.4%	13.6%	11.1%	21.1%	50.3%	100.0%
3	A	15.2%	33.0%	17.8%	34.1%	n/a	n/a	100.0%
	B	0.4%	8.8%	18.2%	14.6%	58.0%	n/a	100.0%
	C	0.1%	1.1%	6.1%	6.1%	38.6%	48.0%	100.0%
4-5	A	28.4%	28.7%	16.2%	26.7%	n/a	n/a	100.0%
	B	2.5%	12.7%	24.1%	11.2%	49.5%	n/a	100.0%
	C	0.1%	2.1%	12.6%	9.6%	25.5%	50.0%	100.0%
6-8	A	31.3%	40.6%	15.1%	13.1%	n/a	n/a	100.0%
	B	4.2%	24.4%	32.9%	7.1%	31.4%	n/a	100.0%
	C	1.2%	18.4%	32.4%	14.3%	18.8%	15.0%	100.0%
9-12	A	31.4%	37.6%	13.9%	17.1%	n/a	n/a	100.0%
	B	10.1%	35.6%	17.4%	9.2%	27.8%	n/a	100.0%
	C	0.7%	9.7%	11.3%	11.6%	21.0%	45.6%	100.0%

4.3.2.2 By Grade by Tier

Table 4.3.2.2A
Proficiency Level by Grade By Tier (Count): Reading S400 Paper

Grade	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	54,136	32,382	46,058	19,956	25,147	68,241	245,920
K (accountability)	-	160,103	17,576	11,951	13,821	42,469	0	245,920
1	A	5,705	6,983	6,135	10,063	n/a	n/a	28,886
	B	42	482	4,201	5,485	19,727	n/a	29,937
	C	57	475	1,516	4,199	5,017	7,469	18,733
2	A	3,488	2,504	1,099	2,915	n/a	n/a	10,006
	B	205	1,973	6,266	3,487	19,935	n/a	31,866
	C	169	1,234	4,923	4,024	7,612	18,168	36,130
3	A	1,542	3,350	1,810	3,464	n/a	n/a	10,166
	B	93	1,838	3,773	3,024	12,054	n/a	20,782
	C	15	275	1,533	1,546	9,698	12,079	25,146
4	A	1,718	2,128	1,132	2,016	n/a	n/a	6,994
	B	226	1,425	2,930	1,964	7,594	n/a	14,139
	C	19	243	2,006	2,576	4,955	11,300	21,099
5	A	1,989	1,614	977	1,463	n/a	n/a	6,043
	B	351	1,457	2,570	587	3,672	n/a	8,637
	C	32	591	2,884	1,164	4,974	8,182	17,827
6	A	1,126	2,599	994	743	n/a	n/a	5,462
	B	203	1,351	2,698	501	2,372	n/a	7,125
	C	86	1,518	3,796	1,926	2,053	1,653	11,032
7	A	1,743	2,111	890	743	n/a	n/a	5,487
	B	277	1,628	2,154	471	1,847	n/a	6,377
	C	127	1,961	3,222	1,650	1,748	1,387	10,095
8	A	2,197	1,873	558	633	n/a	n/a	5,261
	B	355	1,835	1,645	425	1,976	n/a	6,236
	C	160	2,270	3,123	896	2,085	1,667	10,201
9	A	2,548	2,940	1,218	1,105	n/a	n/a	7,811
	B	453	2,069	1,726	426	1,979	n/a	6,653
	C	21	736	1,226	1,168	1,990	4,423	9,564
10	A	1,635	2,087	666	1,024	n/a	n/a	5,412
	B	584	2,232	727	659	1,388	n/a	5,590
	C	51	676	1,006	883	1,471	3,347	7,434
11	A	1,076	1,416	486	729	n/a	n/a	3,707
	B	482	1,604	581	547	1,252	n/a	4,466
	C	65	692	530	703	1,159	3,126	6,275
12	A	583	572	220	329	n/a	n/a	1,704
	B	492	1,199	444	202	923	n/a	3,260
	C	67	635	451	543	1,330	2,026	5,052

Table 4.3.2.2B

Proficiency Level by Grade By Tier (Percent):Reading S400 Paper

Grade	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	22.0%	13.2%	18.7%	8.1%	10.2%	27.7%	100.0%
K (accountability)	-	65.1%	7.1%	4.9%	5.6%	17.3%	0.0%	100.0%
1	A	19.8%	24.2%	21.2%	34.8%	n/a	n/a	100.0%
	B	0.1%	1.6%	14.0%	18.3%	65.9%	n/a	100.0%
	C	0.3%	2.5%	8.1%	22.4%	26.8%	39.9%	100.0%
2	A	34.9%	25.0%	11.0%	29.1%	n/a	n/a	100.0%
	B	0.6%	6.2%	19.7%	10.9%	62.6%	n/a	100.0%
	C	0.5%	3.4%	13.6%	11.1%	21.1%	50.3%	100.0%
3	A	15.2%	33.0%	17.8%	34.1%	n/a	n/a	100.0%
	B	0.4%	8.8%	18.2%	14.6%	58.0%	n/a	100.0%
	C	0.1%	1.1%	6.1%	6.1%	38.6%	48.0%	100.0%
4	A	24.6%	30.4%	16.2%	28.8%	n/a	n/a	100.0%
	B	1.6%	10.1%	20.7%	13.9%	53.7%	n/a	100.0%
	C	0.1%	1.2%	9.5%	12.2%	23.5%	53.6%	100.0%
5	A	32.9%	26.7%	16.2%	24.2%	n/a	n/a	100.0%
	B	4.1%	16.9%	29.8%	6.8%	42.5%	n/a	100.0%
	C	0.2%	3.3%	16.2%	6.5%	27.9%	45.9%	100.0%
6	A	20.6%	47.6%	18.2%	13.6%	n/a	n/a	100.0%
	B	2.8%	19.0%	37.9%	7.0%	33.3%	n/a	100.0%
	C	0.8%	13.8%	34.4%	17.5%	18.6%	15.0%	100.0%
7	A	31.8%	38.5%	16.2%	13.5%	n/a	n/a	100.0%
	B	4.3%	25.5%	33.8%	7.4%	29.0%	n/a	100.0%
	C	1.3%	19.4%	31.9%	16.3%	17.3%	13.7%	100.0%
8	A	41.8%	35.6%	10.6%	12.0%	n/a	n/a	100.0%
	B	5.7%	29.4%	26.4%	6.8%	31.7%	n/a	100.0%
	C	1.6%	22.3%	30.6%	8.8%	20.4%	16.3%	100.0%
9	A	32.6%	37.6%	15.6%	14.1%	n/a	n/a	100.0%
	B	6.8%	31.1%	25.9%	6.4%	29.7%	n/a	100.0%
	C	0.2%	7.7%	12.8%	12.2%	20.8%	46.2%	100.0%
10	A	30.2%	38.6%	12.3%	18.9%	n/a	n/a	100.0%
	B	10.4%	39.9%	13.0%	11.8%	24.8%	n/a	100.0%
	C	0.7%	9.1%	13.5%	11.9%	19.8%	45.0%	100.0%
11	A	29.0%	38.2%	13.1%	19.7%	n/a	n/a	100.0%
	B	10.8%	35.9%	13.0%	12.2%	28.0%	n/a	100.0%
	C	1.0%	11.0%	8.4%	11.2%	18.5%	49.8%	100.0%
12	A	34.2%	33.6%	12.9%	19.3%	n/a	n/a	100.0%
	B	15.1%	36.8%	13.6%	6.2%	28.3%	n/a	100.0%
	C	1.3%	12.6%	8.9%	10.7%	26.3%	40.1%	100.0%

4.3.2.3 By Grade

Table 4.3.2.3A

Proficiency Level by Grade (Count): Reading S400 Paper

	Reading Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	54,136	32,382	46,058	19,956	25,147	68,241	245,920
K (accountability)	160,103	17,576	11,951	13,821	42,469	0	245,920
1	5,804	7,940	11,852	19,747	24,744	7,469	77,556
2	3,862	5,711	12,288	10,426	27,547	18,168	78,002
3	1,650	5,463	7,116	8,034	21,752	12,079	56,094
4	1,963	3,796	6,068	6,556	12,549	11,300	42,232
5	2,372	3,662	6,431	3,214	8,646	8,182	32,507
6	1,415	5,468	7,488	3,170	4,425	1,653	23,619
7	2,147	5,700	6,266	2,864	3,595	1,387	21,959
8	2,712	5,978	5,326	1,954	4,061	1,667	21,698
9	3,022	5,745	4,170	2,699	3,969	4,423	24,028
10	2,270	4,995	2,399	2,566	2,859	3,347	18,436
11	1,623	3,712	1,597	1,979	2,411	3,126	14,448
12	1,142	2,406	1,115	1,074	2,253	2,026	10,016

Table 4.3.2.3B

Proficiency Level by Grade (Percent): Reading S400 Paper

	Reading Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	22.0%	13.2%	18.7%	8.1%	10.2%	27.7%	100.0%
K (accountability)	65.1%	7.1%	4.9%	5.6%	17.3%	0.0%	100.0%
1	7.5%	10.2%	15.3%	25.5%	31.9%	9.6%	100.0%
2	5.0%	7.3%	15.8%	13.4%	35.3%	23.3%	100.0%
3	2.9%	9.7%	12.7%	14.3%	38.8%	21.5%	100.0%
4	4.6%	9.0%	14.4%	15.5%	29.7%	26.8%	100.0%
5	7.3%	11.3%	19.8%	9.9%	26.6%	25.2%	100.0%
6	6.0%	23.2%	31.7%	13.4%	18.7%	7.0%	100.0%
7	9.8%	26.0%	28.5%	13.0%	16.4%	6.3%	100.0%
8	12.5%	27.6%	24.5%	9.0%	18.7%	7.7%	100.0%
9	12.6%	23.9%	17.4%	11.2%	16.5%	18.4%	100.0%
10	12.3%	27.1%	13.0%	13.9%	15.5%	18.2%	100.0%
11	11.2%	25.7%	11.1%	13.7%	16.7%	21.6%	100.0%
12	11.4%	24.0%	11.1%	10.7%	22.5%	20.2%	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 S400 Paper

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	43,495	72,812	40,205	34,140	47,962	7,306	245,920
K (accountability)	-	142,202	48,450	31,715	16,247	7,306	0	245,920
1	A	4,689	16,100	10,082	0	0	0	30,871
	B	2,787	11,993	17,084	1,302	0	0	33,166
	C	560	4,750	12,648	2,183	0	0	20,141
2	A	3,242	5,365	1,720	116	0	0	10,443
	B	1,919	12,332	18,975	433	0	0	33,659
	C	271	6,451	29,466	1,504	1	0	37,693
3	A	2,954	5,182	2,114	329	0	0	10,579
	B	410	1,379	4,874	13,330	1,821	12	21,826
	C	87	453	3,712	17,965	4,307	31	26,555
4-5	A	1,453	1,994	5,803	4,245	0	0	13,495
	B	354	837	3,841	17,155	1,642	2	23,831
	C	128	336	3,160	31,535	5,626	18	40,803
6-8	A	3,085	6,136	7,164	622	0	0	17,007
	B	1,157	2,042	12,758	5,631	23	0	21,611
	C	441	1,075	18,416	12,946	103	0	32,981
9-12	A	3,124	5,501	9,148	1,465	7	0	19,245
	B	1,472	1,105	6,281	10,081	2,336	55	21,330
	C	576	443	4,838	17,373	6,883	202	30,315

Table 4.3.3.1B

Proficiency Level by Cluster By Tier (Percent): Writing S400 Paper

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	17.7%	29.6%	16.3%	13.9%	19.5%	3.0%	100.0%
K (accountability)	-	57.8%	19.7%	12.9%	6.6%	3.0%	0.0%	100.0%
1	A	15.2%	52.2%	32.7%	0.0%	0.0%	0.0%	100.0%
	B	8.4%	36.2%	51.5%	3.9%	0.0%	0.0%	100.0%
	C	2.8%	23.6%	62.8%	10.8%	0.0%	0.0%	100.0%
2	A	31.0%	51.4%	16.5%	1.1%	0.0%	0.0%	100.0%
	B	5.7%	36.6%	56.4%	1.3%	0.0%	0.0%	100.0%
	C	0.7%	17.1%	78.2%	4.0%	0.0%	0.0%	100.0%
3	A	27.9%	49.0%	20.0%	3.1%	0.0%	0.0%	100.0%
	B	1.9%	6.3%	22.3%	61.1%	8.3%	0.1%	100.0%
	C	0.3%	1.7%	14.0%	67.7%	16.2%	0.1%	100.0%
4-5	A	10.8%	14.8%	43.0%	31.5%	0.0%	0.0%	100.0%
	B	1.5%	3.5%	16.1%	72.0%	6.9%	0.0%	100.0%
	C	0.3%	0.8%	7.7%	77.3%	13.8%	0.0%	100.0%
6-8	A	18.1%	36.1%	42.1%	3.7%	0.0%	0.0%	100.0%
	B	5.4%	9.4%	59.0%	26.1%	0.1%	0.0%	100.0%
	C	1.3%	3.3%	55.8%	39.3%	0.3%	0.0%	100.0%
9-12	A	16.2%	28.6%	47.5%	7.6%	0.0%	0.0%	100.0%
	B	6.9%	5.2%	29.4%	47.3%	11.0%	0.3%	100.0%
	C	1.9%	1.5%	16.0%	57.3%	22.7%	0.7%	100.0%

4.3.3.2 By Grade by Tier

Table 4.3.3.2A
Proficiency Level by Grade By Tier (Count): Writing S400 Paper

Grade	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	43,495	72,812	40,205	34,140	47,962	7,306	245,920
K (accountability)	-	142,202	48,450	31,715	16,247	7,306	0	245,920
1	A	4,689	16,100	10,082	0	0	0	30,871
	B	2,787	11,993	17,084	1,302	0	0	33,166
	C	560	4,750	12,648	2,183	0	0	20,141
2	A	3,242	5,365	1,720	116	0	0	10,443
	B	1,919	12,332	18,975	433	0	0	33,659
	C	271	6,451	29,466	1,504	1	0	37,693
3	A	2,954	5,182	2,114	329	0	0	10,579
	B	410	1,379	4,874	13,330	1,821	12	21,826
	C	87	453	3,712	17,965	4,307	31	26,555
4	A	633	1,092	2,387	3,160	0	0	7,272
	B	194	519	1,771	11,074	1,302	2	14,862
	C	71	187	1,241	16,578	4,099	18	22,194
5	A	820	902	3,416	1,085	0	0	6,223
	B	160	318	2,070	6,081	340	0	8,969
	C	57	149	1,919	14,957	1,527	0	18,609
6	A	647	1,906	2,758	468	0	0	5,779
	B	230	522	3,259	3,773	19	0	7,803
	C	102	250	3,382	7,872	98	0	11,704
7	A	1,076	1,857	2,692	131	0	0	5,756
	B	349	612	4,521	1,496	4	0	6,982
	C	133	292	6,291	3,879	5	0	10,600
8	A	1,362	2,373	1,714	23	0	0	5,472
	B	578	908	4,978	362	0	0	6,826
	C	206	533	8,743	1,195	0	0	10,677
9	A	996	2,916	3,558	641	7	0	8,118
	B	407	423	1,372	3,311	1,553	45	7,111
	C	151	168	776	4,705	4,385	155	10,340
10	A	1,026	1,014	2,999	538	0	0	5,577
	B	445	318	1,597	3,088	527	8	5,983
	C	119	114	1,006	5,140	1,564	39	7,982
11	A	683	1,078	1,798	240	0	0	3,799
	B	357	225	1,750	2,249	202	2	4,785
	C	126	74	1,354	4,353	717	8	6,632
12	A	419	493	793	46	0	0	1,751
	B	263	139	1,562	1,433	54	0	3,451
	C	180	87	1,702	3,175	217	0	5,361

Table 4.3.3.2B
Proficiency Level by Grade By Tier (Percent): Writing S400 Paper

Grade	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	17.7%	29.6%	16.3%	13.9%	19.5%	3.0%	100.0%
K (accountability)	-	57.8%	19.7%	12.9%	6.6%	3.0%	0.0%	100.0%
1	A	15.2%	52.2%	32.7%	0.0%	0.0%	0.0%	100.0%
	B	8.4%	36.2%	51.5%	3.9%	0.0%	0.0%	100.0%
	C	2.8%	23.6%	62.8%	10.8%	0.0%	0.0%	100.0%
2	A	31.0%	51.4%	16.5%	1.1%	0.0%	0.0%	100.0%
	B	5.7%	36.6%	56.4%	1.3%	0.0%	0.0%	100.0%
	C	0.7%	17.1%	78.2%	4.0%	0.0%	0.0%	100.0%
3	A	27.9%	49.0%	20.0%	3.1%	0.0%	0.0%	100.0%
	B	1.9%	6.3%	22.3%	61.1%	8.3%	0.1%	100.0%
	C	0.3%	1.7%	14.0%	67.7%	16.2%	0.1%	100.0%
4	A	8.7%	15.0%	32.8%	43.5%	0.0%	0.0%	100.0%
	B	1.3%	3.5%	11.9%	74.5%	8.8%	0.0%	100.0%
	C	0.3%	0.8%	5.6%	74.7%	18.5%	0.1%	100.0%
5	A	13.2%	14.5%	54.9%	17.4%	0.0%	0.0%	100.0%
	B	1.8%	3.5%	23.1%	67.8%	3.8%	0.0%	100.0%
	C	0.3%	0.8%	10.3%	80.4%	8.2%	0.0%	100.0%
6	A	11.2%	33.0%	47.7%	8.1%	0.0%	0.0%	100.0%
	B	2.9%	6.7%	41.8%	48.4%	0.2%	0.0%	100.0%
	C	0.9%	2.1%	28.9%	67.3%	0.8%	0.0%	100.0%
7	A	18.7%	32.3%	46.8%	2.3%	0.0%	0.0%	100.0%
	B	5.0%	8.8%	64.8%	21.4%	0.1%	0.0%	100.0%
	C	1.3%	2.8%	59.3%	36.6%	0.0%	0.0%	100.0%
8	A	24.9%	43.4%	31.3%	0.4%	0.0%	0.0%	100.0%
	B	8.5%	13.3%	72.9%	5.3%	0.0%	0.0%	100.0%
	C	1.9%	5.0%	81.9%	11.2%	0.0%	0.0%	100.0%
9	A	12.3%	35.9%	43.8%	7.9%	0.1%	0.0%	100.0%
	B	5.7%	5.9%	19.3%	46.6%	21.8%	0.6%	100.0%
	C	1.5%	1.6%	7.5%	45.5%	42.4%	1.5%	100.0%
10	A	18.4%	18.2%	53.8%	9.6%	0.0%	0.0%	100.0%
	B	7.4%	5.3%	26.7%	51.6%	8.8%	0.1%	100.0%
	C	1.5%	1.4%	12.6%	64.4%	19.6%	0.5%	100.0%
11	A	18.0%	28.4%	47.3%	6.3%	0.0%	0.0%	100.0%
	B	7.5%	4.7%	36.6%	47.0%	4.2%	0.0%	100.0%
	C	1.9%	1.1%	20.4%	65.6%	10.8%	0.1%	100.0%
12	A	23.9%	28.2%	45.3%	2.6%	0.0%	0.0%	100.0%
	B	7.6%	4.0%	45.3%	41.5%	1.6%	0.0%	100.0%
	C	3.4%	1.6%	31.7%	59.2%	4.0%	0.0%	100.0%

4.3.3.3 By Grade

Table 4.3.3.3A
Proficiency Level by Grade (Count): Writing S400 Paper

	Writing Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	43,495	72,812	40,205	34,140	47,962	7,306	245,920
K (accountability)	142,202	48,450	31,715	16,247	7,306	0	245,920
1	8,036	32,843	39,814	3,485	0	0	84,178
2	5,432	24,148	50,161	2,053	1	0	81,795
3	3,451	7,014	10,700	31,624	6,128	43	58,960
4	898	1,798	5,399	30,812	5,401	20	44,328
5	1,037	1,369	7,405	22,123	1,867	0	33,801
6	979	2,678	9,399	12,113	117	0	25,286
7	1,558	2,761	13,504	5,506	9	0	23,338
8	2,146	3,814	15,435	1,580	0	0	22,975
9	1,554	3,507	5,706	8,657	5,945	200	25,569
10	1,590	1,446	5,602	8,766	2,091	47	19,542
11	1,166	1,377	4,902	6,842	919	10	15,216
12	862	719	4,057	4,654	271	0	10,563

Table 4.3.3.3B
Proficiency Level by Grade (Percent): Writing S400 Paper

	Writing Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	17.7%	29.6%	16.3%	13.9%	19.5%	3.0%	100.0%
K (accountability)	57.8%	19.7%	12.9%	6.6%	3.0%	0.0%	100.0%
1	9.5%	39.0%	47.3%	4.1%	0.0%	0.0%	100.0%
2	6.6%	29.5%	61.3%	2.5%	0.0%	0.0%	100.0%
3	5.9%	11.9%	18.1%	53.6%	10.4%	0.1%	100.0%
4	2.0%	4.1%	12.2%	69.5%	12.2%	0.0%	100.0%
5	3.1%	4.1%	21.9%	65.5%	5.5%	0.0%	100.0%
6	3.9%	10.6%	37.2%	47.9%	0.5%	0.0%	100.0%
7	6.7%	11.8%	57.9%	23.6%	0.0%	0.0%	100.0%
8	9.3%	16.6%	67.2%	6.9%	0.0%	0.0%	100.0%
9	6.1%	13.7%	22.3%	33.9%	23.3%	0.8%	100.0%
10	8.1%	7.4%	28.7%	44.9%	10.7%	0.2%	100.0%
11	7.7%	9.0%	32.2%	45.0%	6.0%	0.1%	100.0%
12	8.2%	6.8%	38.4%	44.1%	2.6%	0.0%	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 S400 Paper

Cluster	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	52,694	17,614	35,845	40,763	28,218	70,786	245,920
K (accountability)	-	52,694	53,459	40,763	28,218	70,786	0	245,920
1	A	8,620	7,318	4,637	0	0	9,995	30,570
	B	903	6,471	2,552	3,320	2,683	16,976	32,905
	C	210	2,170	1,002	1,545	1,379	13,705	20,011
2	A	3,404	1,801	1,558	0	0	3,567	10,330
	B	1,209	3,413	1,919	2,217	3,387	21,231	33,376
	C	312	1,375	1,032	1,432	2,439	30,909	37,499
3	A	3,123	2,281	1,229	0	1,154	2,695	10,482
	B	1,209	2,450	1,952	1,556	1,830	12,613	21,610
	C	391	1,415	1,484	1,449	1,749	19,907	26,395
4-5	A	6,771	2,988	1,168	0	1,066	1,391	13,384
	B	1,957	3,552	1,691	1,916	2,067	12,480	23,663
	C	733	3,119	1,929	2,583	3,000	29,200	40,564
6-8	A	7,740	1,629	2,391	1,525	0	3,531	16,816
	B	959	1,803	1,796	2,225	1,552	13,008	21,343
	C	141	628	1,053	1,852	1,702	27,323	32,699
9-12	A	9,414	2,001	1,506	853	1,665	3,464	18,903
	B	933	3,008	1,466	1,637	1	13,793	20,838
	C	123	789	718	1,141	0	26,848	29,619

Table 4.3.4.1B
Proficiency Level by Cluster By Tier (Percent): Speaking S400 Paper

Cluster	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	21.4%	7.2%	14.6%	16.6%	11.5%	28.8%	100.0%
K (accountability)	-	21.4%	21.7%	16.6%	11.5%	28.8%	0.0%	100.0%
1	A	28.2%	23.9%	15.2%	0.0%	0.0%	32.7%	100.0%
	B	2.7%	19.7%	7.8%	10.1%	8.2%	51.6%	100.0%
	C	1.0%	10.8%	5.0%	7.7%	6.9%	68.5%	100.0%
2	A	33.0%	17.4%	15.1%	0.0%	0.0%	34.5%	100.0%
	B	3.6%	10.2%	5.7%	6.6%	10.1%	63.6%	100.0%
	C	0.8%	3.7%	2.8%	3.8%	6.5%	82.4%	100.0%
3	A	29.8%	21.8%	11.7%	0.0%	11.0%	25.7%	100.0%
	B	5.6%	11.3%	9.0%	7.2%	8.5%	58.4%	100.0%
	C	1.5%	5.4%	5.6%	5.5%	6.6%	75.4%	100.0%
4-5	A	50.6%	22.3%	8.7%	0.0%	8.0%	10.4%	100.0%
	B	8.3%	15.0%	7.1%	8.1%	8.7%	52.7%	100.0%
	C	1.8%	7.7%	4.8%	6.4%	7.4%	72.0%	100.0%
6-8	A	46.0%	9.7%	14.2%	9.1%	0.0%	21.0%	100.0%
	B	4.5%	8.4%	8.4%	10.4%	7.3%	60.9%	100.0%
	C	0.4%	1.9%	3.2%	5.7%	5.2%	83.6%	100.0%
9-12	A	49.8%	10.6%	8.0%	4.5%	8.8%	18.3%	100.0%
	B	4.5%	14.4%	7.0%	7.9%	0.0%	66.2%	100.0%
	C	0.4%	2.7%	2.4%	3.9%	0.0%	90.6%	100.0%

4.3.4.2 By Grade by Tier

Table 4.3.4.2A

Proficiency Level by Grade By Tier (Count): Speaking S400 Paper

Grade	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	52,694	17,614	35,845	40,763	28,218	70,786	245,920
K (accountability)	-	52,694	53,459	40,763	28,218	70,786	0	245,920
1	A	8,620	7,318	4,637	0	0	9,995	30,570
	B	903	6,471	2,552	3,320	2,683	16,976	32,905
	C	210	2,170	1,002	1,545	1,379	13,705	20,011
2	A	3,404	1,801	1,558	0	0	3,567	10,330
	B	1,209	3,413	1,919	2,217	3,387	21,231	33,376
	C	312	1,375	1,032	1,432	2,439	30,909	37,499
3	A	3,123	2,281	1,229	0	1,154	2,695	10,482
	B	1,209	2,450	1,952	1,556	1,830	12,613	21,610
	C	391	1,415	1,484	1,449	1,749	19,907	26,395
4	A	3,543	1,667	659	0	578	760	7,207
	B	1,266	2,310	1,080	1,246	1,348	7,505	14,755
	C	478	1,876	1,134	1,500	1,720	15,362	22,070
5	A	3,228	1,321	509	0	488	631	6,177
	B	691	1,242	611	670	719	4,975	8,908
	C	255	1,243	795	1,083	1,280	13,838	18,494
6	A	2,595	556	838	500	0	1,218	5,707
	B	326	396	711	1,106	592	4,599	7,730
	C	60	150	420	1,022	685	9,279	11,616
7	A	2,687	531	756	498	0	1,219	5,691
	B	313	570	723	574	494	4,223	6,897
	C	43	199	412	413	554	8,885	10,506
8	A	2,458	542	797	527	0	1,094	5,418
	B	320	837	362	545	466	4,186	6,716
	C	38	279	221	417	463	9,159	10,577
9	A	5,232	0	0	853	622	1,258	7,965
	B	609	792	340	470	0	4,738	6,949
	C	88	250	194	382	0	9,219	10,133
10	A	2,643	606	740	0	498	993	5,480
	B	175	1,153	355	491	1	3,663	5,838
	C	22	288	149	302	0	7,054	7,815
11	A	1,214	863	505	0	339	813	3,734
	B	127	632	446	397	0	3,082	4,684
	C	9	129	207	245	0	5,871	6,461
12	A	325	532	261	0	206	400	1,724
	B	22	431	325	279	0	2,310	3,367
	C	4	122	168	212	0	4,704	5,210

Table 4.3.4.2B

Proficiency Level by Grade By Tier (Percent): Speaking S400 Paper

Grade	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	21.4%	7.2%	14.6%	16.6%	11.5%	28.8%	100.0%
K (accountability)	-	21.4%	21.7%	16.6%	11.5%	28.8%	0.0%	100.0%
1	A	28.2%	23.9%	15.2%	0.0%	0.0%	32.7%	100.0%
	B	2.7%	19.7%	7.8%	10.1%	8.2%	51.6%	100.0%
	C	1.0%	10.8%	5.0%	7.7%	6.9%	68.5%	100.0%
2	A	33.0%	17.4%	15.1%	0.0%	0.0%	34.5%	100.0%
	B	3.6%	10.2%	5.7%	6.6%	10.1%	63.6%	100.0%
	C	0.8%	3.7%	2.8%	3.8%	6.5%	82.4%	100.0%
3	A	29.8%	21.8%	11.7%	0.0%	11.0%	25.7%	100.0%
	B	5.6%	11.3%	9.0%	7.2%	8.5%	58.4%	100.0%
	C	1.5%	5.4%	5.6%	5.5%	6.6%	75.4%	100.0%
4	A	49.2%	23.1%	9.1%	0.0%	8.0%	10.5%	100.0%
	B	8.6%	15.7%	7.3%	8.4%	9.1%	50.9%	100.0%
	C	2.2%	8.5%	5.1%	6.8%	7.8%	69.6%	100.0%
5	A	52.3%	21.4%	8.2%	0.0%	7.9%	10.2%	100.0%
	B	7.8%	13.9%	6.9%	7.5%	8.1%	55.8%	100.0%
	C	1.4%	6.7%	4.3%	5.9%	6.9%	74.8%	100.0%
6	A	45.5%	9.7%	14.7%	8.8%	0.0%	21.3%	100.0%
	B	4.2%	5.1%	9.2%	14.3%	7.7%	59.5%	100.0%
	C	0.5%	1.3%	3.6%	8.8%	5.9%	79.9%	100.0%
7	A	47.2%	9.3%	13.3%	8.8%	0.0%	21.4%	100.0%
	B	4.5%	8.3%	10.5%	8.3%	7.2%	61.2%	100.0%
	C	0.4%	1.9%	3.9%	3.9%	5.3%	84.6%	100.0%
8	A	45.4%	10.0%	14.7%	9.7%	0.0%	20.2%	100.0%
	B	4.8%	12.5%	5.4%	8.1%	6.9%	62.3%	100.0%
	C	0.4%	2.6%	2.1%	3.9%	4.4%	86.6%	100.0%
9	A	65.7%	0.0%	0.0%	10.7%	7.8%	15.8%	100.0%
	B	8.8%	11.4%	4.9%	6.8%	0.0%	68.2%	100.0%
	C	0.9%	2.5%	1.9%	3.8%	0.0%	91.0%	100.0%
10	A	48.2%	11.1%	13.5%	0.0%	9.1%	18.1%	100.0%
	B	3.0%	19.7%	6.1%	8.4%	0.0%	62.7%	100.0%
	C	0.3%	3.7%	1.9%	3.9%	0.0%	90.3%	100.0%
11	A	32.5%	23.1%	13.5%	0.0%	9.1%	21.8%	100.0%
	B	2.7%	13.5%	9.5%	8.5%	0.0%	65.8%	100.0%
	C	0.1%	2.0%	3.2%	3.8%	0.0%	90.9%	100.0%
12	A	18.9%	30.9%	15.1%	0.0%	11.9%	23.2%	100.0%
	B	0.7%	12.8%	9.7%	8.3%	0.0%	68.6%	100.0%
	C	0.1%	2.3%	3.2%	4.1%	0.0%	90.3%	100.0%

4.3.4.3 By Grade

Table 4.3.4.3A

Proficiency Level by Grade (Count): Speaking S400 Paper

	Speaking Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	52,694	17,614	35,845	40,763	28,218	70,786	245,920
K (accountability)	52,694	53,459	40,763	28,218	70,786	0	245,920
1	9,733	15,959	8,191	4,865	4,062	40,676	83,486
2	4,925	6,589	4,509	3,649	5,826	55,707	81,205
3	4,723	6,146	4,665	3,005	4,733	35,215	58,487
4	5,287	5,853	2,873	2,746	3,646	23,627	44,032
5	4,174	3,806	1,915	1,753	2,487	19,444	33,579
6	2,981	1,102	1,969	2,628	1,277	15,096	25,053
7	3,043	1,300	1,891	1,485	1,048	14,327	23,094
8	2,816	1,658	1,380	1,489	929	14,439	22,711
9	5,929	1,042	534	1,705	622	15,215	25,047
10	2,840	2,047	1,244	793	499	11,710	19,133
11	1,350	1,624	1,158	642	339	9,766	14,879
12	351	1,085	754	491	206	7,414	10,301

Table 4.3.4.3B

Proficiency Level by Grade (Percent): Speaking S400 Paper

	Speaking Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	21.4%	7.2%	14.6%	16.6%	11.5%	28.8%	100.0%
K (accountability)	21.4%	21.7%	16.6%	11.5%	28.8%	0.0%	100.0%
1	11.7%	19.1%	9.8%	5.8%	4.9%	48.7%	100.0%
2	6.1%	8.1%	5.6%	4.5%	7.2%	68.6%	100.0%
3	8.1%	10.5%	8.0%	5.1%	8.1%	60.2%	100.0%
4	12.0%	13.3%	6.5%	6.2%	8.3%	53.7%	100.0%
5	12.4%	11.3%	5.7%	5.2%	7.4%	57.9%	100.0%
6	11.9%	4.4%	7.9%	10.5%	5.1%	60.3%	100.0%
7	13.2%	5.6%	8.2%	6.4%	4.5%	62.0%	100.0%
8	12.4%	7.3%	6.1%	6.6%	4.1%	63.6%	100.0%
9	23.7%	4.2%	2.1%	6.8%	2.5%	60.7%	100.0%
10	14.8%	10.7%	6.5%	4.1%	2.6%	61.2%	100.0%
11	9.1%	10.9%	7.8%	4.3%	2.3%	65.6%	100.0%
12	3.4%	10.5%	7.3%	4.8%	2.0%	72.0%	100.0%

4.3.5 Oral Composite

4.3.5.1 By Cluster by Tier

Table 4.3.5.1A

Proficiency Level by Cluster By Tier (Count): Oral S400 Paper

Cluster	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	34,668	20,454	27,496	48,227	56,670	58,405	245,920
K (accountability)	-	58,608	33,616	38,621	21,739	34,931	58,405	245,920
1	A	3,770	8,739	7,126	837	8,747	0	29,219
	B	166	1,897	7,688	3,372	18,889	0	32,012
	C	41	721	2,168	1,969	5,710	9,040	19,649
2	A	2,454	2,042	2,182	440	2,976	0	10,094
	B	142	992	5,450	2,582	23,867	0	33,033
	C	29	428	1,731	3,115	13,437	18,453	37,193
3	A	1,714	2,413	2,446	1,214	2,518	0	10,305
	B	158	1,088	3,179	3,498	13,488	0	21,411
	C	9	250	1,181	2,140	6,472	16,221	26,273
4-5	A	3,571	3,789	3,425	1,103	1,296	0	13,184
	B	343	1,518	4,465	3,616	13,588	0	23,530
	C	18	419	1,959	4,284	12,044	21,690	40,414
6-8	A	5,778	3,751	3,049	1,880	2,176	0	16,634
	B	469	1,675	2,609	5,106	11,325	0	21,184
	C	20	120	632	2,226	8,892	20,642	32,532
9-12	A	8,011	4,526	2,663	2,280	1,131	0	18,611
	B	309	2,389	3,053	4,649	10,196	0	20,596
	C	20	353	1,174	4,115	12,316	11,383	29,361

Table 4.3.5.1B

Proficiency Level by Cluster By Tier (Percent): Oral S400 Paper

Cluster	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	14.1%	8.3%	11.2%	19.6%	23.0%	23.7%	100.0%
K (accountability)	-	23.8%	13.7%	15.7%	8.8%	14.2%	23.7%	100.0%
1	A	12.9%	29.9%	24.4%	2.9%	29.9%	0.0%	100.0%
	B	0.5%	5.9%	24.0%	10.5%	59.0%	0.0%	100.0%
	C	0.2%	3.7%	11.0%	10.0%	29.1%	46.0%	100.0%
2	A	24.3%	20.2%	21.6%	4.4%	29.5%	0.0%	100.0%
	B	0.4%	3.0%	16.5%	7.8%	72.3%	0.0%	100.0%
	C	0.1%	1.2%	4.7%	8.4%	36.1%	49.6%	100.0%
3	A	16.6%	23.4%	23.7%	11.8%	24.4%	0.0%	100.0%
	B	0.7%	5.1%	14.8%	16.3%	63.0%	0.0%	100.0%
	C	0.0%	1.0%	4.5%	8.1%	24.6%	61.7%	100.0%
4-5	A	27.1%	28.7%	26.0%	8.4%	9.8%	0.0%	100.0%
	B	1.5%	6.5%	19.0%	15.4%	57.7%	0.0%	100.0%
	C	0.0%	1.0%	4.8%	10.6%	29.8%	53.7%	100.0%
6-8	A	34.7%	22.6%	18.3%	11.3%	13.1%	0.0%	100.0%
	B	2.2%	7.9%	12.3%	24.1%	53.5%	0.0%	100.0%
	C	0.1%	0.4%	1.9%	6.8%	27.3%	63.5%	100.0%
9-12	A	43.0%	24.3%	14.3%	12.3%	6.1%	0.0%	100.0%
	B	1.5%	11.6%	14.8%	22.6%	49.5%	0.0%	100.0%
	C	0.1%	1.2%	4.0%	14.0%	41.9%	38.8%	100.0%

4.3.5.2 By Grade by Tier

Table 4.3.5.2A

Proficiency Level by Grade By Tier (Count): Oral S400 Paper

Grade	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	34,668	20,454	27,496	48,227	56,670	58,405	245,920
K (accountability)	-	58,608	33,616	38,621	21,739	34,931	58,405	245,920
1	A	3,770	8,739	7,126	837	8,747	0	29,219
	B	166	1,897	7,688	3,372	18,889	0	32,012
	C	41	721	2,168	1,969	5,710	9,040	19,649
2	A	2,454	2,042	2,182	440	2,976	0	10,094
	B	142	992	5,450	2,582	23,867	0	33,033
	C	29	428	1,731	3,115	13,437	18,453	37,193
3	A	1,714	2,413	2,446	1,214	2,518	0	10,305
	B	158	1,088	3,179	3,498	13,488	0	21,411
	C	9	250	1,181	2,140	6,472	16,221	26,273
4	A	1,723	2,080	1,965	585	729	0	7,082
	B	180	944	2,889	2,237	8,416	0	14,666
	C	8	262	1,207	2,443	5,605	12,451	21,976
5	A	1,848	1,709	1,460	518	567	0	6,102
	B	163	574	1,576	1,379	5,172	0	8,864
	C	10	157	752	1,841	6,439	9,239	18,438
6	A	1,831	1,160	1,136	616	895	0	5,638
	B	133	474	949	1,677	4,456	0	7,689
	C	8	50	215	893	3,102	7,297	11,565
7	A	2,013	1,239	1,008	591	787	0	5,638
	B	154	545	814	1,590	3,737	0	6,840
	C	7	31	224	641	2,806	6,743	10,452
8	A	1,934	1,352	905	673	494	0	5,358
	B	182	656	846	1,839	3,132	0	6,655
	C	5	39	193	692	2,984	6,602	10,515
9	A	3,815	1,682	895	817	608	0	7,817
	B	133	807	792	1,055	4,074	0	6,861
	C	6	107	291	832	3,978	4,828	10,042
10	A	2,181	1,455	816	624	338	0	5,414
	B	80	740	823	1,283	2,860	0	5,786
	C	7	83	321	909	3,552	2,886	7,758
11	A	1,380	965	617	531	185	0	3,678
	B	62	495	801	1,202	2,061	0	4,621
	C	4	88	255	1,115	2,493	2,454	6,409
12	A	635	424	335	308	0	0	1,702
	B	34	347	637	1,109	1,201	0	3,328
	C	3	75	307	1,259	2,293	1,215	5,152

Table 4.3.5.2B

Proficiency Level by Grade By Tier (Percent): Oral S400 Paper

Grade	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	14.1%	8.3%	11.2%	19.6%	23.0%	23.7%	100.0%
K (accountability)	-	23.8%	13.7%	15.7%	8.8%	14.2%	23.7%	100.0%
1	A	12.9%	29.9%	24.4%	2.9%	29.9%	0.0%	100.0%
	B	0.5%	5.9%	24.0%	10.5%	59.0%	0.0%	100.0%
	C	0.2%	3.7%	11.0%	10.0%	29.1%	46.0%	100.0%
2	A	24.3%	20.2%	21.6%	4.4%	29.5%	0.0%	100.0%
	B	0.4%	3.0%	16.5%	7.8%	72.3%	0.0%	100.0%
	C	0.1%	1.2%	4.7%	8.4%	36.1%	49.6%	100.0%
3	A	16.6%	23.4%	23.7%	11.8%	24.4%	0.0%	100.0%
	B	0.7%	5.1%	14.8%	16.3%	63.0%	0.0%	100.0%
	C	0.0%	1.0%	4.5%	8.1%	24.6%	61.7%	100.0%
4	A	24.3%	29.4%	27.7%	8.3%	10.3%	0.0%	100.0%
	B	1.2%	6.4%	19.7%	15.3%	57.4%	0.0%	100.0%
	C	0.0%	1.2%	5.5%	11.1%	25.5%	56.7%	100.0%
5	A	30.3%	28.0%	23.9%	8.5%	9.3%	0.0%	100.0%
	B	1.8%	6.5%	17.8%	15.6%	58.3%	0.0%	100.0%
	C	0.1%	0.9%	4.1%	10.0%	34.9%	50.1%	100.0%
6	A	32.5%	20.6%	20.1%	10.9%	15.9%	0.0%	100.0%
	B	1.7%	6.2%	12.3%	21.8%	58.0%	0.0%	100.0%
	C	0.1%	0.4%	1.9%	7.7%	26.8%	63.1%	100.0%
7	A	35.7%	22.0%	17.9%	10.5%	14.0%	0.0%	100.0%
	B	2.3%	8.0%	11.9%	23.2%	54.6%	0.0%	100.0%
	C	0.1%	0.3%	2.1%	6.1%	26.8%	64.5%	100.0%
8	A	36.1%	25.2%	16.9%	12.6%	9.2%	0.0%	100.0%
	B	2.7%	9.9%	12.7%	27.6%	47.1%	0.0%	100.0%
	C	0.0%	0.4%	1.8%	6.6%	28.4%	62.8%	100.0%
9	A	48.8%	21.5%	11.4%	10.5%	7.8%	0.0%	100.0%
	B	1.9%	11.8%	11.5%	15.4%	59.4%	0.0%	100.0%
	C	0.1%	1.1%	2.9%	8.3%	39.6%	48.1%	100.0%
10	A	40.3%	26.9%	15.1%	11.5%	6.2%	0.0%	100.0%
	B	1.4%	12.8%	14.2%	22.2%	49.4%	0.0%	100.0%
	C	0.1%	1.1%	4.1%	11.7%	45.8%	37.2%	100.0%
11	A	37.5%	26.2%	16.8%	14.4%	5.0%	0.0%	100.0%
	B	1.3%	10.7%	17.3%	26.0%	44.6%	0.0%	100.0%
	C	0.1%	1.4%	4.0%	17.4%	38.9%	38.3%	100.0%
12	A	37.3%	24.9%	19.7%	18.1%	0.0%	0.0%	100.0%
	B	1.0%	10.4%	19.1%	33.3%	36.1%	0.0%	100.0%
	C	0.1%	1.5%	6.0%	24.4%	44.5%	23.6%	100.0%

4.3.5.3 By Grade

Table 4.3.5.3A

Proficiency Level by Grade (Count): Oral S400 Paper

	Oral Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	34,668	20,454	27,496	48,227	56,670	58,405	245,920
K (accountability)	58,608	33,616	38,621	21,739	34,931	58,405	245,920
1	3,977	11,357	16,982	6,178	33,346	9,040	80,880
2	2,625	3,462	9,363	6,137	40,280	18,453	80,320
3	1,881	3,751	6,806	6,852	22,478	16,221	57,989
4	1,911	3,286	6,061	5,265	14,750	12,451	43,724
5	2,021	2,440	3,788	3,738	12,178	9,239	33,404
6	1,972	1,684	2,300	3,186	8,453	7,297	24,892
7	2,174	1,815	2,046	2,822	7,330	6,743	22,930
8	2,121	2,047	1,944	3,204	6,610	6,602	22,528
9	3,954	2,596	1,978	2,704	8,660	4,828	24,720
10	2,268	2,278	1,960	2,816	6,750	2,886	18,958
11	1,446	1,548	1,673	2,848	4,739	2,454	14,708
12	672	846	1,279	2,676	3,494	1,215	10,182

Table 4.3.5.3B

Proficiency Level by Grade (Percent): Oral S400 Paper

	Oral Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	14.1%	8.3%	11.2%	19.6%	23.0%	23.7%	100.0%
K (accountability)	23.8%	13.7%	15.7%	8.8%	14.2%	23.7%	100.0%
1	4.9%	14.0%	21.0%	7.6%	41.2%	11.2%	100.0%
2	3.3%	4.3%	11.7%	7.6%	50.1%	23.0%	100.0%
3	3.2%	6.5%	11.7%	11.8%	38.8%	28.0%	100.0%
4	4.4%	7.5%	13.9%	12.0%	33.7%	28.5%	100.0%
5	6.1%	7.3%	11.3%	11.2%	36.5%	27.7%	100.0%
6	7.9%	6.8%	9.2%	12.8%	34.0%	29.3%	100.0%
7	9.5%	7.9%	8.9%	12.3%	32.0%	29.4%	100.0%
8	9.4%	9.1%	8.6%	14.2%	29.3%	29.3%	100.0%
9	16.0%	10.5%	8.0%	10.9%	35.0%	19.5%	100.0%
10	12.0%	12.0%	10.3%	14.9%	35.6%	15.2%	100.0%
11	9.8%	10.5%	11.4%	19.4%	32.2%	16.7%	100.0%
12	6.6%	8.3%	12.6%	26.3%	34.3%	11.9%	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 S400 Paper

Cluster	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	42,982	58,617	43,798	31,536	51,688	17,299	245,920
K (accountability)	-	154,227	30,075	30,360	21,645	9,613	0	245,920
1	A	4,110	12,383	12,393	0	0	0	28,886
	B	323	4,504	21,678	3,432	0	0	29,937
	C	71	1,423	7,298	6,155	3,338	448	18,733
2	A	3,281	3,743	2,870	112	0	0	10,006
	B	460	6,086	22,866	2,454	0	0	31,866
	C	64	2,067	13,634	13,164	6,680	521	36,130
3	A	2,053	4,256	3,594	263	0	0	10,166
	B	155	1,243	4,745	13,039	1,600	0	20,782
	C	15	108	1,855	8,332	12,275	2,561	25,146
4-5	A	1,655	3,482	5,188	2,712	0	0	13,037
	B	302	1,081	5,632	13,871	1,890	0	22,776
	C	39	160	3,657	13,706	16,654	4,710	38,926
6-8	A	3,449	7,034	5,372	355	0	0	16,210
	B	664	3,160	10,858	5,036	20	0	19,738
	C	140	1,840	15,835	10,894	2,454	165	31,328
9-12	A	3,804	7,225	6,523	1,082	0	0	18,634
	B	1,079	2,912	7,066	7,268	1,644	0	19,969
	C	176	685	4,126	8,648	9,927	4,763	28,325

Table 4.3.6.1B
Proficiency Level by Cluster By Tier (Percent): Literacy S400 Paper

Cluster	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	17.5%	23.8%	17.8%	12.8%	21.0%	7.0%	100.0%
K (accountability)	-	62.7%	12.2%	12.3%	8.8%	3.9%	0.0%	100.0%
1	A	14.2%	42.9%	42.9%	0.0%	0.0%	0.0%	100.0%
	B	1.1%	15.0%	72.4%	11.5%	0.0%	0.0%	100.0%
	C	0.4%	7.6%	39.0%	32.9%	17.8%	2.4%	100.0%
2	A	32.8%	37.4%	28.7%	1.1%	0.0%	0.0%	100.0%
	B	1.4%	19.1%	71.8%	7.7%	0.0%	0.0%	100.0%
	C	0.2%	5.7%	37.7%	36.4%	18.5%	1.4%	100.0%
3	A	20.2%	41.9%	35.4%	2.6%	0.0%	0.0%	100.0%
	B	0.7%	6.0%	22.8%	62.7%	7.7%	0.0%	100.0%
	C	0.1%	0.4%	7.4%	33.1%	48.8%	10.2%	100.0%
4-5	A	12.7%	26.7%	39.8%	20.8%	0.0%	0.0%	100.0%
	B	1.3%	4.7%	24.7%	60.9%	8.3%	0.0%	100.0%
	C	0.1%	0.4%	9.4%	35.2%	42.8%	12.1%	100.0%
6-8	A	21.3%	43.4%	33.1%	2.2%	0.0%	0.0%	100.0%
	B	3.4%	16.0%	55.0%	25.5%	0.1%	0.0%	100.0%
	C	0.4%	5.9%	50.5%	34.8%	7.8%	0.5%	100.0%
9-12	A	20.4%	38.8%	35.0%	5.8%	0.0%	0.0%	100.0%
	B	5.4%	14.6%	35.4%	36.4%	8.2%	0.0%	100.0%
	C	0.6%	2.4%	14.6%	30.5%	35.0%	16.8%	100.0%

4.3.6.2 By Grade by Tier

Table 4.3.6.2A

Proficiency Level by Grade By Tier (Count): Literacy S400 Paper

Grade	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	42,982	58,617	43,798	31,536	51,688	17,299	245,920
K (accountability)	-	154,227	30,075	30,360	21,645	9,613	0	245,920
1	A	4,110	12,383	12,393	0	0	0	28,886
	B	323	4,504	21,678	3,432	0	0	29,937
	C	71	1,423	7,298	6,155	3,338	448	18,733
2	A	3,281	3,743	2,870	112	0	0	10,006
	B	460	6,086	22,866	2,454	0	0	31,866
	C	64	2,067	13,634	13,164	6,680	521	36,130
3	A	2,053	4,256	3,594	263	0	0	10,166
	B	155	1,243	4,745	13,039	1,600	0	20,782
	C	15	108	1,855	8,332	12,275	2,561	25,146
4	A	654	1,726	2,701	1,913	0	0	6,994
	B	177	505	2,880	8,995	1,582	0	14,139
	C	22	71	1,354	6,860	9,789	3,003	21,099
5	A	1,001	1,756	2,487	799	0	0	6,043
	B	125	576	2,752	4,876	308	0	8,637
	C	17	89	2,303	6,846	6,865	1,707	17,827
6	A	777	2,243	2,201	241	0	0	5,462
	B	112	768	3,419	2,811	15	0	7,125
	C	25	342	4,530	4,897	1,144	94	11,032
7	A	1,176	2,444	1,769	98	0	0	5,487
	B	211	1,032	3,532	1,598	4	0	6,377
	C	49	540	5,222	3,422	811	51	10,095
8	A	1,496	2,347	1,402	16	0	0	5,261
	B	341	1,360	3,907	627	1	0	6,236
	C	66	958	6,083	2,575	499	20	10,201
9	A	1,495	3,081	2,755	480	0	0	7,811
	B	317	644	2,180	2,460	1,052	0	6,653
	C	35	124	956	2,729	3,838	1,882	9,564
10	A	1,101	2,067	1,882	362	0	0	5,412
	B	333	815	1,924	2,131	387	0	5,590
	C	37	169	1,019	2,240	2,722	1,247	7,434
11	A	770	1,394	1,343	200	0	0	3,707
	B	253	727	1,685	1,638	163	0	4,466
	C	41	153	1,040	1,949	2,031	1,061	6,275
12	A	438	683	543	40	0	0	1,704
	B	176	726	1,277	1,039	42	0	3,260
	C	63	239	1,111	1,730	1,336	573	5,052

Table 4.3.6.2B

Proficiency Level by Grade By Tier (Percent): Literacy S400 Paper

Grade	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	17.5%	23.8%	17.8%	12.8%	21.0%	7.0%	100.0%
K (accountability)	-	62.7%	12.2%	12.3%	8.8%	3.9%	0.0%	100.0%
1	A	14.2%	42.9%	42.9%	0.0%	0.0%	0.0%	100.0%
	B	1.1%	15.0%	72.4%	11.5%	0.0%	0.0%	100.0%
	C	0.4%	7.6%	39.0%	32.9%	17.8%	2.4%	100.0%
2	A	32.8%	37.4%	28.7%	1.1%	0.0%	0.0%	100.0%
	B	1.4%	19.1%	71.8%	7.7%	0.0%	0.0%	100.0%
	C	0.2%	5.7%	37.7%	36.4%	18.5%	1.4%	100.0%
3	A	20.2%	41.9%	35.4%	2.6%	0.0%	0.0%	100.0%
	B	0.7%	6.0%	22.8%	62.7%	7.7%	0.0%	100.0%
	C	0.1%	0.4%	7.4%	33.1%	48.8%	10.2%	100.0%
4	A	9.4%	24.7%	38.6%	27.4%	0.0%	0.0%	100.0%
	B	1.3%	3.6%	20.4%	63.6%	11.2%	0.0%	100.0%
	C	0.1%	0.3%	6.4%	32.5%	46.4%	14.2%	100.0%
5	A	16.6%	29.1%	41.2%	13.2%	0.0%	0.0%	100.0%
	B	1.4%	6.7%	31.9%	56.5%	3.6%	0.0%	100.0%
	C	0.1%	0.5%	12.9%	38.4%	38.5%	9.6%	100.0%
6	A	14.2%	41.1%	40.3%	4.4%	0.0%	0.0%	100.0%
	B	1.6%	10.8%	48.0%	39.5%	0.2%	0.0%	100.0%
	C	0.2%	3.1%	41.1%	44.4%	10.4%	0.9%	100.0%
7	A	21.4%	44.5%	32.2%	1.8%	0.0%	0.0%	100.0%
	B	3.3%	16.2%	55.4%	25.1%	0.1%	0.0%	100.0%
	C	0.5%	5.3%	51.7%	33.9%	8.0%	0.5%	100.0%
8	A	28.4%	44.6%	26.6%	0.3%	0.0%	0.0%	100.0%
	B	5.5%	21.8%	62.7%	10.1%	0.0%	0.0%	100.0%
	C	0.6%	9.4%	59.6%	25.2%	4.9%	0.2%	100.0%
9	A	19.1%	39.4%	35.3%	6.1%	0.0%	0.0%	100.0%
	B	4.8%	9.7%	32.8%	37.0%	15.8%	0.0%	100.0%
	C	0.4%	1.3%	10.0%	28.5%	40.1%	19.7%	100.0%
10	A	20.3%	38.2%	34.8%	6.7%	0.0%	0.0%	100.0%
	B	6.0%	14.6%	34.4%	38.1%	6.9%	0.0%	100.0%
	C	0.5%	2.3%	13.7%	30.1%	36.6%	16.8%	100.0%
11	A	20.8%	37.6%	36.2%	5.4%	0.0%	0.0%	100.0%
	B	5.7%	16.3%	37.7%	36.7%	3.6%	0.0%	100.0%
	C	0.7%	2.4%	16.6%	31.1%	32.4%	16.9%	100.0%
12	A	25.7%	40.1%	31.9%	2.3%	0.0%	0.0%	100.0%
	B	5.4%	22.3%	39.2%	31.9%	1.3%	0.0%	100.0%
	C	1.2%	4.7%	22.0%	34.2%	26.4%	11.3%	100.0%

4.3.6.3 By Grade

Table 4.3.6.3A

Proficiency Level by Grade (Count): Literacy S400 Paper

	Literacy Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	42,982	58,617	43,798	31,536	51,688	17,299	245,920
K (accountability)	154,227	30,075	30,360	21,645	9,613	0	245,920
1	4,504	18,310	41,369	9,587	3,338	448	77,556
2	3,805	11,896	39,370	15,730	6,680	521	78,002
3	2,223	5,607	10,194	21,634	13,875	2,561	56,094
4	853	2,302	6,935	17,768	11,371	3,003	42,232
5	1,143	2,421	7,542	12,521	7,173	1,707	32,507
6	914	3,353	10,150	7,949	1,159	94	23,619
7	1,436	4,016	10,523	5,118	815	51	21,959
8	1,903	4,665	11,392	3,218	500	20	21,698
9	1,847	3,849	5,891	5,669	4,890	1,882	24,028
10	1,471	3,051	4,825	4,733	3,109	1,247	18,436
11	1,064	2,274	4,068	3,787	2,194	1,061	14,448
12	677	1,648	2,931	2,809	1,378	573	10,016

Table 4.3.6.3B

Proficiency Level by Grade (Percent): Literacy S400 Paper

	Literacy Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	17.5%	23.8%	17.8%	12.8%	21.0%	7.0%	100.0%
K (accountability)	62.7%	12.2%	12.3%	8.8%	3.9%	0.0%	100.0%
1	5.8%	23.6%	53.3%	12.4%	4.3%	0.6%	100.0%
2	4.9%	15.3%	50.5%	20.2%	8.6%	0.7%	100.0%
3	4.0%	10.0%	18.2%	38.6%	24.7%	4.6%	100.0%
4	2.0%	5.5%	16.4%	42.1%	26.9%	7.1%	100.0%
5	3.5%	7.4%	23.2%	38.5%	22.1%	5.3%	100.0%
6	3.9%	14.2%	43.0%	33.7%	4.9%	0.4%	100.0%
7	6.5%	18.3%	47.9%	23.3%	3.7%	0.2%	100.0%
8	8.8%	21.5%	52.5%	14.8%	2.3%	0.1%	100.0%
9	7.7%	16.0%	24.5%	23.6%	20.4%	7.8%	100.0%
10	8.0%	16.5%	26.2%	25.7%	16.9%	6.8%	100.0%
11	7.4%	15.7%	28.2%	26.2%	15.2%	7.3%	100.0%
12	6.8%	16.5%	29.3%	28.0%	13.8%	5.7%	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 S400 Paper

Cluster	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	30,075	35,081	51,150	34,428	37,075	58,111	245,920
K (accountability)	-	140,502	17,239	17,351	16,789	32,396	21,643	245,920
1	A	2,439	7,275	10,500	7,656	n/a	n/a	27,870
	B	7	176	3,261	7,599	18,247	n/a	29,290
	C	6	163	1,972	2,728	5,947	7,599	18,415
2	A	1,983	3,008	2,325	2,519	n/a	n/a	9,835
	B	22	576	6,571	5,576	18,838	n/a	31,583
	C	21	477	3,258	4,300	10,822	16,987	35,865
3	A	448	3,367	3,160	3,058	n/a	n/a	10,033
	B	19	730	4,408	4,837	10,622	n/a	20,616
	C	2	33	746	1,730	9,137	13,394	25,042
4-5	A	1,731	4,407	3,720	3,025	n/a	n/a	12,883
	B	116	1,667	5,788	5,424	9,662	n/a	22,657
	C	3	189	3,120	3,476	12,177	19,840	38,805
6-8	A	3,681	7,120	3,848	1,424	n/a	n/a	16,073
	B	256	3,839	7,201	4,495	3,822	n/a	19,613
	C	16	1,107	7,093	5,491	10,391	7,098	31,196
9-12	A	6,162	7,838	3,317	1,081	n/a	n/a	18,398
	B	899	5,267	5,642	4,570	3,380	n/a	19,758
	C	46	1,289	3,784	5,038	7,859	10,092	28,108

Table 4.3.7.1B

Proficiency Level by Cluster By Tier (Percent): Comprehension S400 Paper

Cluster	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	12.2%	14.3%	20.8%	14.0%	15.1%	23.6%	100.0%
K (accountability)	-	57.1%	7.0%	7.1%	6.8%	13.2%	8.8%	100.0%
1	A	8.8%	26.1%	37.7%	27.5%	n/a	n/a	100.0%
	B	0.0%	0.6%	11.1%	25.9%	62.3%	n/a	100.0%
	C	0.0%	0.9%	10.7%	14.8%	32.3%	41.3%	100.0%
2	A	20.2%	30.6%	23.6%	25.6%	n/a	n/a	100.0%
	B	0.1%	1.8%	20.8%	17.7%	59.6%	n/a	100.0%
	C	0.1%	1.3%	9.1%	12.0%	30.2%	47.4%	100.0%
3	A	4.5%	33.6%	31.5%	30.5%	n/a	n/a	100.0%
	B	0.1%	3.5%	21.4%	23.5%	51.5%	n/a	100.0%
	C	0.0%	0.1%	3.0%	6.9%	36.5%	53.5%	100.0%
4-5	A	13.4%	34.2%	28.9%	23.5%	n/a	n/a	100.0%
	B	0.5%	7.4%	25.5%	23.9%	42.6%	n/a	100.0%
	C	0.0%	0.5%	8.0%	9.0%	31.4%	51.1%	100.0%
6-8	A	22.9%	44.3%	23.9%	8.9%	n/a	n/a	100.0%
	B	1.3%	19.6%	36.7%	22.9%	19.5%	n/a	100.0%
	C	0.1%	3.5%	22.7%	17.6%	33.3%	22.8%	100.0%
9-12	A	33.5%	42.6%	18.0%	5.9%	n/a	n/a	100.0%
	B	4.6%	26.7%	28.6%	23.1%	17.1%	n/a	100.0%
	C	0.2%	4.6%	13.5%	17.9%	28.0%	35.9%	100.0%

4.3.7.2 By Grade by Tier

Table 4.3.7.2A

Proficiency Level by Grade By Tier (Count): Comprehension S400 Paper

Grade	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	30,075	35,081	51,150	34,428	37,075	58,111	245,920
K (accountability)	-	140,502	17,239	17,351	16,789	32,396	21,643	245,920
1	A	2,439	7,275	10,500	7,656	n/a	n/a	27,870
	B	7	176	3,261	7,599	18,247	n/a	29,290
	C	6	163	1,972	2,728	5,947	7,599	18,415
2	A	1,983	3,008	2,325	2,519	n/a	n/a	9,835
	B	22	576	6,571	5,576	18,838	n/a	31,583
	C	21	477	3,258	4,300	10,822	16,987	35,865
3	A	448	3,367	3,160	3,058	n/a	n/a	10,033
	B	19	730	4,408	4,837	10,622	n/a	20,616
	C	2	33	746	1,730	9,137	13,394	25,042
4	A	651	2,422	2,081	1,746	n/a	n/a	6,900
	B	39	826	3,276	3,448	6,475	n/a	14,064
	C	1	53	1,353	1,830	6,417	11,365	21,019
5	A	1,080	1,985	1,639	1,279	n/a	n/a	5,983
	B	77	841	2,512	1,976	3,187	n/a	8,593
	C	2	136	1,767	1,646	5,760	8,475	17,786
6	A	766	2,452	1,658	537	n/a	n/a	5,413
	B	26	1,010	2,796	1,710	1,550	n/a	7,092
	C	1	247	2,489	1,919	3,817	2,518	10,991
7	A	1,361	2,357	1,262	467	n/a	n/a	5,447
	B	86	1,297	2,389	1,504	1,059	n/a	6,335
	C	8	350	2,485	1,757	3,155	2,295	10,050
8	A	1,554	2,311	928	420	n/a	n/a	5,213
	B	144	1,532	2,016	1,281	1,213	n/a	6,186
	C	7	510	2,119	1,815	3,419	2,285	10,155
9	A	2,645	3,199	1,361	493	n/a	n/a	7,698
	B	150	1,334	2,318	1,457	1,317	n/a	6,576
	C	3	211	1,304	1,627	2,980	3,366	9,491
10	A	1,638	2,409	1,016	291	n/a	n/a	5,354
	B	224	1,541	1,612	1,207	960	n/a	5,544
	C	5	313	1,049	1,308	2,096	2,613	7,384
11	A	1,180	1,599	674	204	n/a	n/a	3,657
	B	244	1,403	964	1,156	646	n/a	4,413
	C	12	356	765	1,151	1,526	2,423	6,233
12	A	699	631	266	93	n/a	n/a	1,689
	B	281	989	748	750	457	n/a	3,225
	C	26	409	666	952	1,257	1,690	5,000

Table 4.3.7.2B

Proficiency Level by Grade By Tier (Percent): Comprehension S400 Paper

Grade	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	12.2%	14.3%	20.8%	14.0%	15.1%	23.6%	100.0%
K (accountability)	-	57.1%	7.0%	7.1%	6.8%	13.2%	8.8%	100.0%
1	A	8.8%	26.1%	37.7%	27.5%	n/a	n/a	100.0%
	B	0.0%	0.6%	11.1%	25.9%	62.3%	n/a	100.0%
	C	0.0%	0.9%	10.7%	14.8%	32.3%	41.3%	100.0%
2	A	20.2%	30.6%	23.6%	25.6%	n/a	n/a	100.0%
	B	0.1%	1.8%	20.8%	17.7%	59.6%	n/a	100.0%
	C	0.1%	1.3%	9.1%	12.0%	30.2%	47.4%	100.0%
3	A	4.5%	33.6%	31.5%	30.5%	n/a	n/a	100.0%
	B	0.1%	3.5%	21.4%	23.5%	51.5%	n/a	100.0%
	C	0.0%	0.1%	3.0%	6.9%	36.5%	53.5%	100.0%
4	A	9.4%	35.1%	30.2%	25.3%	n/a	n/a	100.0%
	B	0.3%	5.9%	23.3%	24.5%	46.0%	n/a	100.0%
	C	0.0%	0.3%	6.4%	8.7%	30.5%	54.1%	100.0%
5	A	18.1%	33.2%	27.4%	21.4%	n/a	n/a	100.0%
	B	0.9%	9.8%	29.2%	23.0%	37.1%	n/a	100.0%
	C	0.0%	0.8%	9.9%	9.3%	32.4%	47.6%	100.0%
6	A	14.2%	45.3%	30.6%	9.9%	n/a	n/a	100.0%
	B	0.4%	14.2%	39.4%	24.1%	21.9%	n/a	100.0%
	C	0.0%	2.2%	22.6%	17.5%	34.7%	22.9%	100.0%
7	A	25.0%	43.3%	23.2%	8.6%	n/a	n/a	100.0%
	B	1.4%	20.5%	37.7%	23.7%	16.7%	n/a	100.0%
	C	0.1%	3.5%	24.7%	17.5%	31.4%	22.8%	100.0%
8	A	29.8%	44.3%	17.8%	8.1%	n/a	n/a	100.0%
	B	2.3%	24.8%	32.6%	20.7%	19.6%	n/a	100.0%
	C	0.1%	5.0%	20.9%	17.9%	33.7%	22.5%	100.0%
9	A	34.4%	41.6%	17.7%	6.4%	n/a	n/a	100.0%
	B	2.3%	20.3%	35.2%	22.2%	20.0%	n/a	100.0%
	C	0.0%	2.2%	13.7%	17.1%	31.4%	35.5%	100.0%
10	A	30.6%	45.0%	19.0%	5.4%	n/a	n/a	100.0%
	B	4.0%	27.8%	29.1%	21.8%	17.3%	n/a	100.0%
	C	0.1%	4.2%	14.2%	17.7%	28.4%	35.4%	100.0%
11	A	32.3%	43.7%	18.4%	5.6%	n/a	n/a	100.0%
	B	5.5%	31.8%	21.8%	26.2%	14.6%	n/a	100.0%
	C	0.2%	5.7%	12.3%	18.5%	24.5%	38.9%	100.0%
12	A	41.4%	37.4%	15.7%	5.5%	n/a	n/a	100.0%
	B	8.7%	30.7%	23.2%	23.3%	14.2%	n/a	100.0%
	C	0.5%	8.2%	13.3%	19.0%	25.1%	33.8%	100.0%

4.3.7.3 By Grade

Table 4.3.7.3A

Proficiency Level by Grade (Count): Comprehension S400 Paper

	Comprehension Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	30,075	35,081	51,150	34,428	37,075	58,111	245,920
K (accountability)	140,502	17,239	17,351	16,789	32,396	21,643	245,920
1	2,452	7,614	15,733	17,983	24,194	7,599	75,575
2	2,026	4,061	12,154	12,395	29,660	16,987	77,283
3	469	4,130	8,314	9,625	19,759	13,394	55,691
4	691	3,301	6,710	7,024	12,892	11,365	41,983
5	1,159	2,962	5,918	4,901	8,947	8,475	32,362
6	793	3,709	6,943	4,166	5,367	2,518	23,496
7	1,455	4,004	6,136	3,728	4,214	2,295	21,832
8	1,705	4,353	5,063	3,516	4,632	2,285	21,554
9	2,798	4,744	4,983	3,577	4,297	3,366	23,765
10	1,867	4,263	3,677	2,806	3,056	2,613	18,282
11	1,436	3,358	2,403	2,511	2,172	2,423	14,303
12	1,006	2,029	1,680	1,795	1,714	1,690	9,914

Table 4.3.7.3B

Proficiency Level by Grade (Percent): Comprehension S400 Paper

	Comprehension Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	12.2%	14.3%	20.8%	14.0%	15.1%	23.6%	100.0%
K (accountability)	57.1%	7.0%	7.1%	6.8%	13.2%	8.8%	100.0%
1	3.2%	10.1%	20.8%	23.8%	32.0%	10.1%	100.0%
2	2.6%	5.3%	15.7%	16.0%	38.4%	22.0%	100.0%
3	0.8%	7.4%	14.9%	17.3%	35.5%	24.1%	100.0%
4	1.6%	7.9%	16.0%	16.7%	30.7%	27.1%	100.0%
5	3.6%	9.2%	18.3%	15.1%	27.6%	26.2%	100.0%
6	3.4%	15.8%	29.5%	17.7%	22.8%	10.7%	100.0%
7	6.7%	18.3%	28.1%	17.1%	19.3%	10.5%	100.0%
8	7.9%	20.2%	23.5%	16.3%	21.5%	10.6%	100.0%
9	11.8%	20.0%	21.0%	15.1%	18.1%	14.2%	100.0%
10	10.2%	23.3%	20.1%	15.3%	16.7%	14.3%	100.0%
11	10.0%	23.5%	16.8%	17.6%	15.2%	16.9%	100.0%
12	10.1%	20.5%	16.9%	18.1%	17.3%	17.0%	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 S400 Paper

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	33,777	47,902	47,405	43,817	54,984	18,035	245,920
K (accountability)	-	126,036	39,370	37,062	25,417	15,756	2,279	245,920
1	A	2,861	9,884	14,443	472	0	0	27,660
	B	82	2,094	15,709	11,211	12	0	29,108
	C	18	600	4,638	6,701	5,400	960	18,317
2	A	2,485	3,199	3,502	572	0	0	9,758
	B	122	2,222	14,384	14,630	0	0	31,358
	C	19	401	6,678	14,927	12,319	1,343	35,687
3	A	1,646	3,296	4,544	477	0	0	9,963
	B	85	825	3,739	10,809	4,999	0	20,457
	C	2	41	895	5,068	13,467	5,433	24,906
4-5	A	2,274	3,435	4,890	2,193	0	0	12,792
	B	236	999	4,442	12,278	4,569	0	22,524
	C	11	116	1,854	9,111	19,547	7,950	38,589
6-8	A	4,253	5,487	4,821	1,358	0	0	15,919
	B	426	1,902	7,036	9,519	532	0	19,415
	C	21	338	4,820	14,787	10,110	880	30,956
9-12	A	5,005	6,646	4,785	1,697	11	0	18,144
	B	612	2,151	5,547	7,442	3,639	0	19,391
	C	28	323	2,219	7,659	11,826	5,544	27,599

Table 4.3.8.1B

Proficiency Level by Cluster By Tier (Percent): Overall S400 Paper

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	13.7%	19.5%	19.3%	17.8%	22.4%	7.3%	100.0%
K (accountability)	-	51.3%	16.0%	15.1%	10.3%	6.4%	0.9%	100.0%
1	A	10.3%	35.7%	52.2%	1.7%	0.0%	0.0%	100.0%
	B	0.3%	7.2%	54.0%	38.5%	0.0%	0.0%	100.0%
	C	0.1%	3.3%	25.3%	36.6%	29.5%	5.2%	100.0%
2	A	25.5%	32.8%	35.9%	5.9%	0.0%	0.0%	100.0%
	B	0.4%	7.1%	45.9%	46.7%	0.0%	0.0%	100.0%
	C	0.1%	1.1%	18.7%	41.8%	34.5%	3.8%	100.0%
3	A	16.5%	33.1%	45.6%	4.8%	0.0%	0.0%	100.0%
	B	0.4%	4.0%	18.3%	52.8%	24.4%	0.0%	100.0%
	C	0.0%	0.2%	3.6%	20.3%	54.1%	21.8%	100.0%
4-5	A	17.8%	26.9%	38.2%	17.1%	0.0%	0.0%	100.0%
	B	1.0%	4.4%	19.7%	54.5%	20.3%	0.0%	100.0%
	C	0.0%	0.3%	4.8%	23.6%	50.7%	20.6%	100.0%
6-8	A	26.7%	34.5%	30.3%	8.5%	0.0%	0.0%	100.0%
	B	2.2%	9.8%	36.2%	49.0%	2.7%	0.0%	100.0%
	C	0.1%	1.1%	15.6%	47.8%	32.7%	2.8%	100.0%
9-12	A	27.6%	36.6%	26.4%	9.4%	0.1%	0.0%	100.0%
	B	3.2%	11.1%	28.6%	38.4%	18.8%	0.0%	100.0%
	C	0.1%	1.2%	8.0%	27.8%	42.8%	20.1%	100.0%

4.3.8.2 By Grade by Tier

Table 4.3.8.2A

Proficiency Level by Grade By Tier (Count): Overall S400 Paper

Grade	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	33,777	47,902	47,405	43,817	54,984	18,035	245,920
K (accountability)	-	126,036	39,370	37,062	25,417	15,756	2,279	245,920
1	A	2,861	9,884	14,443	472	0	0	27,660
	B	82	2,094	15,709	11,211	12	0	29,108
	C	18	600	4,638	6,701	5,400	960	18,317
2	A	2,485	3,199	3,502	572	0	0	9,758
	B	122	2,222	14,384	14,630	0	0	31,358
	C	19	401	6,678	14,927	12,319	1,343	35,687
3	A	1,646	3,296	4,544	477	0	0	9,963
	B	85	825	3,739	10,809	4,999	0	20,457
	C	2	41	895	5,068	13,467	5,433	24,906
4	A	1,016	1,731	2,755	1,342	0	0	6,844
	B	137	514	2,476	7,459	3,392	0	13,978
	C	5	58	847	4,401	10,619	4,974	20,904
5	A	1,258	1,704	2,135	851	0	0	5,948
	B	99	485	1,966	4,819	1,177	0	8,546
	C	6	58	1,007	4,710	8,928	2,976	17,685
6	A	1,168	1,734	1,786	667	0	0	5,355
	B	97	454	2,102	4,027	357	0	7,037
	C	3	69	1,158	5,054	4,172	463	10,919
7	A	1,452	1,858	1,644	439	0	0	5,393
	B	132	632	2,398	2,961	149	0	6,272
	C	10	99	1,670	4,580	3,338	270	9,967
8	A	1,633	1,895	1,391	252	0	0	5,171
	B	197	816	2,536	2,531	26	0	6,106
	C	8	170	1,992	5,153	2,600	147	10,070
9	A	2,295	2,644	1,867	773	11	0	7,590
	B	193	544	1,448	2,400	1,874	0	6,459
	C	6	72	437	2,021	4,455	2,357	9,348
10	A	1,321	2,039	1,415	501	0	0	5,276
	B	175	667	1,516	2,020	1,052	0	5,430
	C	7	73	508	1,991	3,195	1,487	7,261
11	A	890	1,373	1,014	328	0	0	3,605
	B	148	525	1,405	1,702	557	0	4,337
	C	6	81	550	1,859	2,490	1,108	6,094
12	A	499	590	489	95	0	0	1,673
	B	96	415	1,178	1,320	156	0	3,165
	C	9	97	724	1,788	1,686	592	4,896

Table 4.3.8.2B

Proficiency Level by Grade By Tier (Percent): Overall S400 Paper

Grade	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K (instructional)	-	13.7%	19.5%	19.3%	17.8%	22.4%	7.3%	100.0%
K (accountability)	-	51.3%	16.0%	15.1%	10.3%	6.4%	0.9%	100.0%
1	A	10.3%	35.7%	52.2%	1.7%	0.0%	0.0%	100.0%
	B	0.3%	7.2%	54.0%	38.5%	0.0%	0.0%	100.0%
	C	0.1%	3.3%	25.3%	36.6%	29.5%	5.2%	100.0%
2	A	25.5%	32.8%	35.9%	5.9%	0.0%	0.0%	100.0%
	B	0.4%	7.1%	45.9%	46.7%	0.0%	0.0%	100.0%
	C	0.1%	1.1%	18.7%	41.8%	34.5%	3.8%	100.0%
3	A	16.5%	33.1%	45.6%	4.8%	0.0%	0.0%	100.0%
	B	0.4%	4.0%	18.3%	52.8%	24.4%	0.0%	100.0%
	C	0.0%	0.2%	3.6%	20.3%	54.1%	21.8%	100.0%
4	A	14.8%	25.3%	40.3%	19.6%	0.0%	0.0%	100.0%
	B	1.0%	3.7%	17.7%	53.4%	24.3%	0.0%	100.0%
	C	0.0%	0.3%	4.1%	21.1%	50.8%	23.8%	100.0%
5	A	21.1%	28.6%	35.9%	14.3%	0.0%	0.0%	100.0%
	B	1.2%	5.7%	23.0%	56.4%	13.8%	0.0%	100.0%
	C	0.0%	0.3%	5.7%	26.6%	50.5%	16.8%	100.0%
6	A	21.8%	32.4%	33.4%	12.5%	0.0%	0.0%	100.0%
	B	1.4%	6.5%	29.9%	57.2%	5.1%	0.0%	100.0%
	C	0.0%	0.6%	10.6%	46.3%	38.2%	4.2%	100.0%
7	A	26.9%	34.5%	30.5%	8.1%	0.0%	0.0%	100.0%
	B	2.1%	10.1%	38.2%	47.2%	2.4%	0.0%	100.0%
	C	0.1%	1.0%	16.8%	46.0%	33.5%	2.7%	100.0%
8	A	31.6%	36.6%	26.9%	4.9%	0.0%	0.0%	100.0%
	B	3.2%	13.4%	41.5%	41.5%	0.4%	0.0%	100.0%
	C	0.1%	1.7%	19.8%	51.2%	25.8%	1.5%	100.0%
9	A	30.2%	34.8%	24.6%	10.2%	0.1%	0.0%	100.0%
	B	3.0%	8.4%	22.4%	37.2%	29.0%	0.0%	100.0%
	C	0.1%	0.8%	4.7%	21.6%	47.7%	25.2%	100.0%
10	A	25.0%	38.6%	26.8%	9.5%	0.0%	0.0%	100.0%
	B	3.2%	12.3%	27.9%	37.2%	19.4%	0.0%	100.0%
	C	0.1%	1.0%	7.0%	27.4%	44.0%	20.5%	100.0%
11	A	24.7%	38.1%	28.1%	9.1%	0.0%	0.0%	100.0%
	B	3.4%	12.1%	32.4%	39.2%	12.8%	0.0%	100.0%
	C	0.1%	1.3%	9.0%	30.5%	40.9%	18.2%	100.0%
12	A	29.8%	35.3%	29.2%	5.7%	0.0%	0.0%	100.0%
	B	3.0%	13.1%	37.2%	41.7%	4.9%	0.0%	100.0%
	C	0.2%	2.0%	14.8%	36.5%	34.4%	12.1%	100.0%

4.3.8.3 By Grade

Table 4.3.8.3A

Proficiency Level by Grade (Count): Overall S400 Paper

	Overall Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	33,777	47,902	47,405	43,817	54,984	18,035	245,920
K (accountability)	126,036	39,370	37,062	25,417	15,756	2,279	245,920
1	2,961	12,578	34,790	18,384	5,412	960	75,085
2	2,626	5,822	24,564	30,129	12,319	1,343	76,803
3	1,733	4,162	9,178	16,354	18,466	5,433	55,326
4	1,158	2,303	6,078	13,202	14,011	4,974	41,726
5	1,363	2,247	5,108	10,380	10,105	2,976	32,179
6	1,268	2,257	5,046	9,748	4,529	463	23,311
7	1,594	2,589	5,712	7,980	3,487	270	21,632
8	1,838	2,881	5,919	7,936	2,626	147	21,347
9	2,494	3,260	3,752	5,194	6,340	2,357	23,397
10	1,503	2,779	3,439	4,512	4,247	1,487	17,967
11	1,044	1,979	2,969	3,889	3,047	1,108	14,036
12	604	1,102	2,391	3,203	1,842	592	9,734

Table 4.3.8.3B

Proficiency Level by Grade (Percent): Overall S400 Paper

	Overall Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K (instructional)	13.7%	19.5%	19.3%	17.8%	22.4%	7.3%	100.0%
K (accountability)	51.3%	16.0%	15.1%	10.3%	6.4%	0.9%	100.0%
1	3.9%	16.8%	46.3%	24.5%	7.2%	1.3%	100.0%
2	3.4%	7.6%	32.0%	39.2%	16.0%	1.7%	100.0%
3	3.1%	7.5%	16.6%	29.6%	33.4%	9.8%	100.0%
4	2.8%	5.5%	14.6%	31.6%	33.6%	11.9%	100.0%
5	4.2%	7.0%	15.9%	32.3%	31.4%	9.2%	100.0%
6	5.4%	9.7%	21.6%	41.8%	19.4%	2.0%	100.0%
7	7.4%	12.0%	26.4%	36.9%	16.1%	1.2%	100.0%
8	8.6%	13.5%	27.7%	37.2%	12.3%	0.7%	100.0%
9	10.7%	13.9%	16.0%	22.2%	27.1%	10.1%	100.0%
10	8.4%	15.5%	19.1%	25.1%	23.6%	8.3%	100.0%
11	7.4%	14.1%	21.2%	27.7%	21.7%	7.9%	100.0%
12	6.2%	11.3%	24.6%	32.9%	18.9%	6.1%	100.0%

World-Class Instructional Design and Assessment

Annual Technical Report for ACCESS for ELLs 2.0 Paper English Language Proficiency Test, Series 400, 2015–2016 Administration

Annual Technical Report No. 12B
Volume 2 of 3: Analyses of Test Forms

Prepared by:

Center for Applied Linguistics

Language Assessment Division
Psychometrics and Quantitative Research Team

April 2017

Volume 2

5. Analyses of Test Forms: Overview

This chapter contains two parts. The first part provides background information on the technical measurement and statistical tools used to analyze ACCESS 2.0 Paper. 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 ACCESS for ELLs Technical Report No. 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.

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

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

where:

P_{ni1} = probability of 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 and Speaking 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 Item Response Theory). *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 score scale for reporting purposes (see ACCESS for ELLs Technical Report No. 1, *Development and Field Test of ACCESS for ELLs* [2006] 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 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, 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 ACCESS 2.0 Series 400 Paper in the academic year 2015–2016, with the exception of some students as described in Section 3.1.

5.1.3 Equating and Scaling

Complete information on the horizontal and vertical scaling of ACCESS scores is provided in ACCESS for ELLs Technical Report No. 1, *Development and Field Test of ACCESS for ELLs* (2006). 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 scale scores used to report results on the test. Table D in Section 6 provides the equation for converting Rasch ability measures in logits to ACCESS scale scores for Listening and Reading. In the domains of Writing and Speaking, no scaling equation is provided, as a temporary logit scale was created solely for

the purpose of conducting the linking analysis, and scaling constants were not used to derive scale scores for Writing and Speaking for ACCESS 2.0 Series 400 Paper.

No equating summaries are provided for ACCESS 2.0 Series 400 Paper. In the domains of Listening and Reading, items were drawn from the ACCESS Series 302 assessment (see Section 1.3.3). In the domains of Writing and Speaking, the tasks were linked to ACCESS tasks via equipercentile linking (see Section 1.3.4.2).

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, they attempt 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 2.0 Paper 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 and Reading) and one for polytomously scored items (Writing and Speaking).

5.1.4.1 Dichotomous Items

Following procedures that were originally proposed by Educational Testing Service (ETS), the Mantel-Haenszel (M-H) Chi-square statistic (Mantel & Haenszel, 1959) 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 M-H Chi-square statistic is used to check the probability that the two groups performed similarly 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 and Speaking tasks), a similar approach is used. It is based on the M-H Chi-square statistic and the standardized mean difference following procedures again developed by ETS. The DIF procedures developed by the ETS (Zwick, Donoghue, & Grima, 1993; Allen, Carlson, & Zalanak, 1999) for polytomous items were used for identifying tasks that exhibit DIF. JMetrik (Meyer, 2014), an open source computer program for psychometric analysis, was used in conducting the analyses. The procedures implemented in JMetrik first calculate the Cochran-Mantel-Haenszel Chi-square statistic and determine its probability of significance. This statistic gives an indication of the probability that observed differences are the result of chance but does not indicate how significant that difference is. To indicate how significant the difference is, the standardized mean difference (SMD) between the performances of the two groups being compared is calculated. The SMD compares the means of the two groups, adjusting for differences in the distribution of the two groups being compared across the values of the total raw scores. To standardize the outcome, this difference is divided by the item score range and serves as an effect size measure for the Cochran-Mantel-Haenszel Chi-square statistic. This effect size measure (reported as standardized P-DIF in JMetrik) ranges from -1 to 1, which may present some challenges when interpreting it. To mitigate this, the absolute value is taken in JMetrik (Meyer, 2014), thereby restricting the range of the rescaled effect size (standardized P-DIF*) to fall between 0 and 1. The effect size flagging criterion for polytomous items, proposed by ETS (Allen, Carlson, & Zalanak, 1999), is also rescaled to the standardized P-DIF* metric (Meyer, 2014).

Following guidance proposed by ETS for the NAEP assessment (Allen, Carlson, & Zalanak, 1999), ACCESS 2.0 Writing and Speaking tasks are classified into three DIF levels as follows:

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

Table B provides a summary of the findings of the DIF analyses at the top, followed by detailed information for each item or task. The first column gives the DIF level: A, B, or C for dichotomous items or AA, BB, or CC for polytomous tasks (i.e., Writing and Speaking tasks). The next columns show the contrasting groups in the DIF analyses: either male versus female or Hispanic versus 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.

Items and tasks which show C-level (or CC-level) DIF are investigated by a team of content experts to determine if any construct-irrelevant factors can be identified that may contribute to DIF. If such a factor is identified, that item or task will be removed from the test for the next operational year.

5.1.5 Analysis of Writing and Speaking Forms

As noted in Chapter 1, Section 1.2.5, ACCESS 2.0 Series 400 Paper has three tiers, and students' test booklets are printed with the tier designation (A, B, or C). In the Listening and Reading domains, Tiers A, B, and C are each different test forms. In Writing and Speaking, however, Tier B and Tier C tests are identical for every grade-level cluster.

In the analyses in Chapter 6, descriptive statistics for Writing and Speaking (distribution of raw score, scale score, and proficiency level) are computed separately for students who took the Tier B designation and students who took the Tier C designation. Statistics which apply to the test as a whole are computed by pooling all of the Tier B and Tier C students for the grade or grade-level cluster, so that students who took the same form are included in the same analysis. Statistics presented in this fashion are included in Table F (Reliability), Table G (Complete Item Analysis and Summary), Table H (DIF Analysis and Summary), Table I (Raw Score to Scale Score Conversion), and Table J (Raw Score to Proficiency Level Conversion).

5.2 Descriptions of Tables and Figures

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 and Reading are scored dichotomously (i.e., correct or incorrect), thus, the highest possible score was the number of items on the test form.

The range for raw scores on the Writing test depends on the test form. For Grade 1 Tier A, the range is 0–44. For all other grade-level clusters, the range on the Tier A test is 0–33. For all Tier B or C tests, the range is 0–66 (see Section 1.6.2 for more information on raw scores for Writing).

The range for raw scores on the Speaking test is either 0–18 (Tier A) or 6–30 (Tier B or Tier C) (see Section 1.6.3 for more information on raw scores for Speaking).

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 2.0 Paper 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 second to 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 PL 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 PL 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 Reading 6–8 Tier A, the scale score is capped at 360 for Grade 6, 369 for Grade 7, and 376 for Grade 8 (see Table 6.6.2.1B). Thus, a Grade 6 student with a raw score of 24 (out of 24) on that test will have a scale score of 360, a Grade 7 student with the same raw score will have a scale score of 369, and a Grade 8 student with the same raw score will have a scale score of 376. However, all three students would have a proficiency level score of 4.0. For more information, see ACCESS for ELLs Technical Report No. 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 Grade 2 student with a raw score of 0 on the Tier A Listening test would have a scale score of 108, while a Grade 5 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 caps 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 (1–6),
- 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 Listening and Reading 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 Section 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. For Speaking and Writing, the scaling equation is not provided for Series 400, as scale scores were not derived from the logit scale directly (see Section 5.1.3 for further detail, as well as Section 1.3.4.2 for detail on the equipercentile linking methods).

5.2.5 Equating Summary (Table E)

No equating summaries are provided for Series 400 Paper. In the domains of Listening and Reading, items were drawn from the Series 302 assessment (see Section 1.3.3). In the domains of Speaking and Writing, the tasks were linked to ACCESS 1.0 tasks via equipercentile linking (see Section 1.3.4.2).

5.2.6 Test Characteristic Curve (Figure D)

For each test form in Listening and Reading, 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 (PLs 1–6) for the domain being tested. (Note that for Kindergarten and Tier A tests in some domains, it was not possible to place into all six proficiency levels.) As would be expected, higher raw scores are required to be placed into higher proficiency levels. The relative width of each section between the cut score lines, however, gives an indication of how many items on that form must be answered correctly (or how many points on the Writing section must be earned) to be placed into a WIDA proficiency level.

As ACCESS 2.0 Series 400 Paper Listening and Reading forms are the same forms as were provided to students in the ACCESS Series 302 administration (see Section 1.3.3), no Test Characteristic Curve is presented.

For the Writing and Speaking domains, no Test Characteristic Curve is presented for Series 400. As described in Section 1.3.4.2, a temporary logit scale was created for Writing and Speaking tests solely for the purpose of conducting the linking analyses. These Writing and Speaking temporary logit scales are not on the ACCESS scales. Therefore, it is not informative to present the Test Characteristic Curve for Series 400 Writing and Speaking domains.

5.2.7 Test Information Function (Figure E)

With the Rasch measurement model, as with any measurement model following Item Response Theory, 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 receives a perfect or near-perfect score), 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 in Listening and Reading, 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 language 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 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.

As the Listening and Reading test forms for ACCESS 2.0 Series 400 Paper were the same forms as were provided to students in the ACCESS Series 302 administration (see Section 1.3.3), new test information function curves are not presented.

For ACCESS 2.0 Series 400 Paper, Figure E is provided only for Listening and Reading since Writing and Speaking task parameters are not on the ACCESS logit scale. As described in Section 1.3.4.2, a temporary logit scale was created for the Writing and Speaking tests solely for the purpose of conducting the linking analyses. These temporary logit scales and the ACCESS cut scores are not on the same scale. Therefore, it is not appropriate to present the test information function curves for the Writing and Speaking domains of ACCESS 2.0 Series 400 Paper.

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 same test can produce widely disparate reliability indices based on the ability distribution of the group of examinees. Because ACCESS 2.0 Paper 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 SEM based on classical test theory. Unlike Item Response Theory, in this approach, SEM is seen as a constant across the spread of test scores (ability continuum). Thus, it is not conditional on ability being measured. It is, however, a function of two statistics: the reliability of the test and the (observed) standard deviation of the test scores. It is calculated as

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

Traditionally, SEM has been used to create a band around an examinee's observed score, with the assertion in the view of classical test theory, that the examinee's true score (i.e., what the examinee's score would be if it could be measured without error) would lie with a certain degree of probability within this band. Statistically speaking, then, there is an expectation that an

examinee's true score has a 68% probability of lying within the band extending from the observed score minus 1 SEM to the observed score plus 1 SEM.

For the Writing tests (except Kindergarten, which is scored by the test administrator), information on interrater reliability for a sample of 20% of the task raters, is also provided in Table F. This portion of the table shows, for each of the tasks, the percentage of agreement between two raters. In this part of the table, the first column shows the task and the second column shows the number of responses that were double scored. DRC selects a sample of 20% of all responses scored, chosen at random during the operational scoring process. The next column shows the rates of agreement: exact, adjacent, and non-adjacent. For Writing, with 0–6 as defined levels and the possibility of awarding a "plus" score between levels (e.g., 3, 3+, or 4 are all valid scores), scores that match or are contiguous are categorized as agreement (for example, if Rater 1 assigns a score of 3+, a Rater 2 score of 3, 3+ or 4 is categorized as agreement). Scores that are one whole score point apart are categorized as adjacent (for example, if Rater 1 assigns a score of 3+, a Rater 2 score of 2+ or 4+ is categorized as adjacent).

As the Speaking test is scored locally, it is not possible to provide interrater reliability data for Speaking. Section 1.6.3.1 describes training procedures that local raters must complete before being certified to administer and score the Speaking test.

5.2.9 Complete Item or Task Analysis and 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), along with analyses of each item or task. Table G has either two parts (Listening and Reading) or three parts (Speaking and Writing).

The first section of this table provides information about the total set of items or tasks, and includes the item type (selected response or constructed response), the average item difficulty (in logits), the number of items, the average p-value (for Listening and Reading only), the average infit mean square, and the average outfit mean square.

The second section of Table G presents results of the analyses of all of the items or tasks on the test form. For Listening, Reading, and Writing, the results for each tier are presented in a separate table. In these tables, first column provides the unique item name. The second column in this part of Table G presents the item difficulty in logits, while the third column indicates whether that item served as a common item across years (note that for ACCESS 2.0 Series 400 Paper, no items served as anchors, as explained in Section 5.2.5). For dichotomously scored items (Listening and Reading), the fourth column shows the p-value (percentage of correct answers on that item). The next two columns show the Rasch fit statistics for the item or task.

For the Speaking test, due to the design of the test, there are a number of items which are shared between tiers. In order to best present the results of the Speaking task analysis, all Speaking items in a grade-level cluster are presented in one single table; this table is repeated in its entirety for each tier in the text. A column to the right of the task name indicates whether the item is found on the Tier A or Tier B and C tests (recall that Tier B and Tier C are the same form for

Speaking). An asterisk indicates shared tasks. The intended proficiency level of the task (1, 3, or 5) is found in the next column. The remaining columns to the right—item difficulty, anchors, and fit statistics—are displayed as for Listening, Reading, and Writing, described above.

The final portion of Table G applies to Writing and Speaking only. This portion of the table provides raw score distributions by task.

5.2.10 DIF Analysis and Summary (Table H)

Table H presents results of DIF analysis by item or task, as well as a summary of items or tasks showing DIF for the form.

The top section of Table H presents a summary of DIF results, indicating how many items or tasks showed A/AA, B/BB, or C/CC level DIF in the form, and which groups were favored.

The second section of Table H presents the same information for each item or task in the form.

5.2.11 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 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, has been repeated in the final rows of the table. In addition, at the lower end of the raw score scale, scale scores are truncated 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.

Note additionally that for Writing and Speaking, for each grade-level cluster, the Tier B form is identical to the Tier C form. The tables are therefore simply repeated in the section for the Tier B form and the section for the Tier C form.

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. 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 ACCESS Series 100 and Series 101, cut scores between proficiency levels were determined by grade-level cluster (e.g.., in Cluster 3–5, a given scale score was associated with the same proficiency level score for Grades 3, 4, and 5 students). 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 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 order words, a Grade 5 student would be expected to perform better on the Cluster 3–5 test than a Grade 3 student at the same underlying level of English proficiency.

To account for this effect, the WIDA Consortium adopted grade-level cut scores beginning with ACCESS 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.) Hence, Table J includes 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 ACCESS 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.

Note additionally that for Writing and Speaking, for each grade-level cluster, the Tier B form is identical to the Tier C form. The tables are therefore simply repeated in the section for the Tier B form and the section for the Tier C form.

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 Agency Representatives (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 2.0 Paper English Language Proficiency Test, Series 400, 2015–2016 Administration

Annual Technical Report No. 12B
Volume 3 of 3: Analyses Across Tiers

Prepared by:

Center for Applied Linguistics

Language Assessment Division
Psychometrics and Quantitative Research Team

April 2017

Volume 3

7. Analysis Across Tiers: Overview

7.1 Background

7.1.1 Reliability of Composites

Four composite scores are reported for ACCESS 2.0 Paper: Oral Language (Oral), Literacy (Litr), Comprehension (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 input to compute the stratified Cronbach's alpha is provided in the appropriate tables in Chapter 8.

Note that when Cronbach's alpha is computed for domains, as described in Section 5.2.8, the alpha is computed for each test form. When Cronbach's alpha is computed for the composites, it is computed by grade-level cluster. For example, for Grade 2, Cronbach's alpha for the domains of Listening and Reading is computed for the Cluster 1–2 Tier A, Tier B, and Tier C test forms. For the domains of Writing and Speaking, Cronbach's alpha is computed for the Cluster 2–3 Tier A test form and for the Tier B and C test form (recall Tier B and Tier C are the same form for Writing and Speaking). To compute Cronbach's alpha for the Grade 2 composites, the analysis is run at the grade-level cluster. The variance and the reliability at the grade-level cluster for each of the domains is presented, along with the variance and reliability of the composite.

7.1.2 Accuracy and Consistency of Classification

For each domain across tiers, as well as for the four composite scores, tables are provided that indicate estimates of the accuracy and consistency of classification of examinees into the WIDA 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 proficiency levels according to their performance on ACCESS 2.0 Paper, 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, 2004b) (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 those made if each student could somehow be tested with all possible parallel forms of the assessments; that is, the examinees' "true score." Meanwhile, the consistency of a decision is the extent to which decisions made on the basis of the administered test would agree with those made if each student were to take a different but parallel form of the test. Thus, in every analysis of classification, two parallel analyses are made: accuracy (vis-à-vis "true scores") and consistency (vis-à-vis a parallel test).

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, the distribution of the true scores and of scores on a parallel form were modeled. 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 proficiency level by both the administered test and either the true score distribution (accuracy) or a parallel test (consistency). (These 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.)

Accuracy and consistency are also observed conditional on the 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, the most important set of indices may be the indices at the cut points. At every cut point, using the true score distribution (i.e., accuracy), the percentage of students who are consistently placed above and below the cut score is provided, 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 PL 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 2.0 Paper 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 grade and by 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 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 grade and by total for the grade-level cluster:

- The WIDA proficiency level designation (1–6),
- the number of students (count) whose performance on the test form placed them into that proficiency level in the domain being tested, 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 at Cut Scores (Table C)

Table C and Figures C and D provide information across the three overlapping tiers within a grade-level cluster and on the comparative conditional standard error of measurement (CSEM). 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 CSEM at the most important points at which decisions are made about students, based on performance on ACCESS 2.0 Paper—the cut points between proficiency levels. Because the cut points depend on grade level, information is provided for each grade within the grade-level cluster. The leftmost column shows the cut (e.g., 1/2, which is the cut score between PL 1 and PL 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 CSEM 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 CSEM of any tier at the 1/2 cut point, and a relatively low CSEM at the 2/3 cut point. At the other end of the continuum, Tier C forms should optimally have the lowest CSEM at the 5/6 cut point, and a relatively low CSEM at the 4/5 cut point. Tier B should have low CSEM in the middle range. Information from this table provides comparable information on how well the three tiers are targeted to provide the most accurate measure in order to place their intended examinees into the 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.)

As described in Section 1.3.4.2, the CSEM from the equated ACCESS Series 303 Writing and Speaking scale scores are used to approximate the CSEM for ACCESS 2.0 Series 400 scale scores. These CSEM are approximations since there were obtained through the equipercentile relationship between ACCESS Series 303 and ACCESS 2.0 Series 400 scale scores. Since the ACCESS Series 303 Speaking test was not tiered, the CSEM at the cut scores for Speaking are presented by grade-level cluster.

7.2.4 Test Characteristic Curve (Figure C)

Figure C shows the test characteristic curve across the entire test for Kindergarten and across the three tiers for the other grade-level clusters. (Note that this information applies only to the domain scores; this information is not applicable to the composite scores. For ACCESS 2.0 Series 400, it applies to Listening and Reading only.) 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.

7.2.5 Test Information Function (Figure D)

Figure D compares the test information function across the entire test for Kindergarten and across the three tiers for the other grade-level clusters. (Note that this information applies only to the domain scores; this information is not applicable to the composite scores. For ACCESS 2.0 Series 400, it applies to Listening and Reading only.) 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.

7.2.6 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 test 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, which is 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 2–4 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.

7.2.7 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 proficiency levels (see Section 7.1.2). With the adoption of grade-level cut scores with ACCESS Series 102, placement within a proficiency level now depends on the grade level of the student. Therefore, separate tables for each grade in a grade-level cluster are provided. 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 PL 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.

Because of the scoring caps placed on Listening and Reading, there are several cases where there were no test takers placed into the proficiency level and accuracy of classification conditional on that level cannot be computed. In these cases, 'NA' has been placed in the table. In addition, there are a few cases where due to the small percentage of test takers placed into the proficiency level and the range of observed scale scores, accuracy of classification conditional on that level cannot be estimated by BB-CLASS. In such cases, a hyphen (-) has been placed in the table. For Writing, these results can also occur in higher proficiency levels.

For Kindergarten, these tables are provided for both the accountability cut scores and the instructional cut scores.

8 Analyses Across Tiers: Results

8.1 Grade: K

8.1.1 Listening K

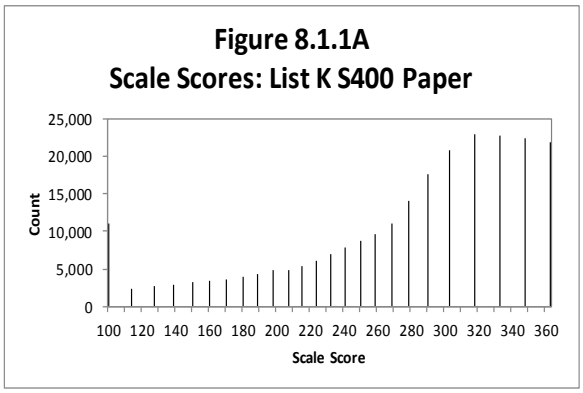


Table 8.1.1A
Scale Score Descriptive Statistics: List K S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	245,920	100	363	272.32	71.74

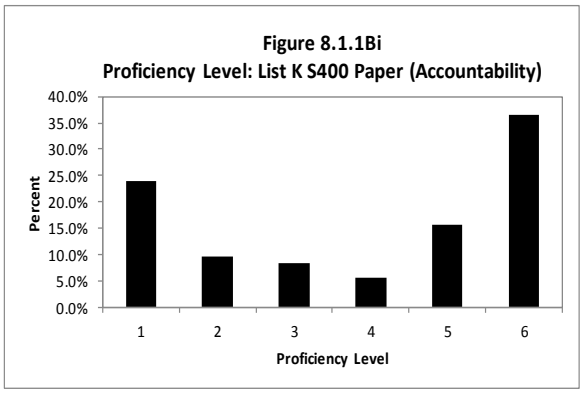


Table 8.1.1Bi
Proficiency Level Distribution: List K S400 Paper (Accountability)

Level	Count	Percent
1	59,186	24.1%
2	23,544	9.6%
3	20,757	8.4%
4	14,070	5.7%
5	38,431	15.6%
6	89,932	36.6%
Total	245,920	100.0%

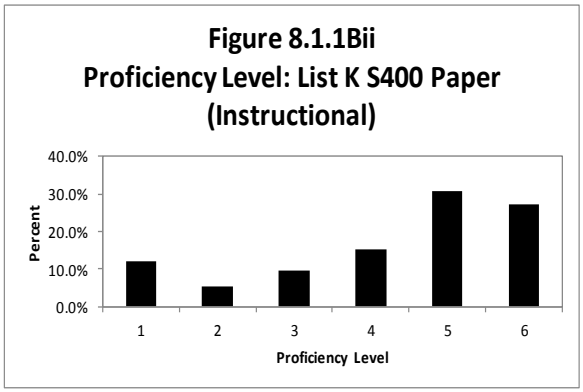


Table 8.1.1Bii
Proficiency Level Distribution: List K S400 Paper (Instructional)

Level	Count	Percent
K1	29,491	12.0%
K2	13,289	5.4%
K3	23,327	9.5%
K4	37,380	15.2%
K5	75,357	30.6%
K6	67,076	27.3%
Total	245,920	100.0%

Table 8.1.1ci

Conditional Standard Error of Measurement at Cut Scores: List K Accountability S400 Paper

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

Table 8.1.1cii

Conditional Standard Error of Measurement at Cut Scores: List K Instructional S400 Paper

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 S400 Paper

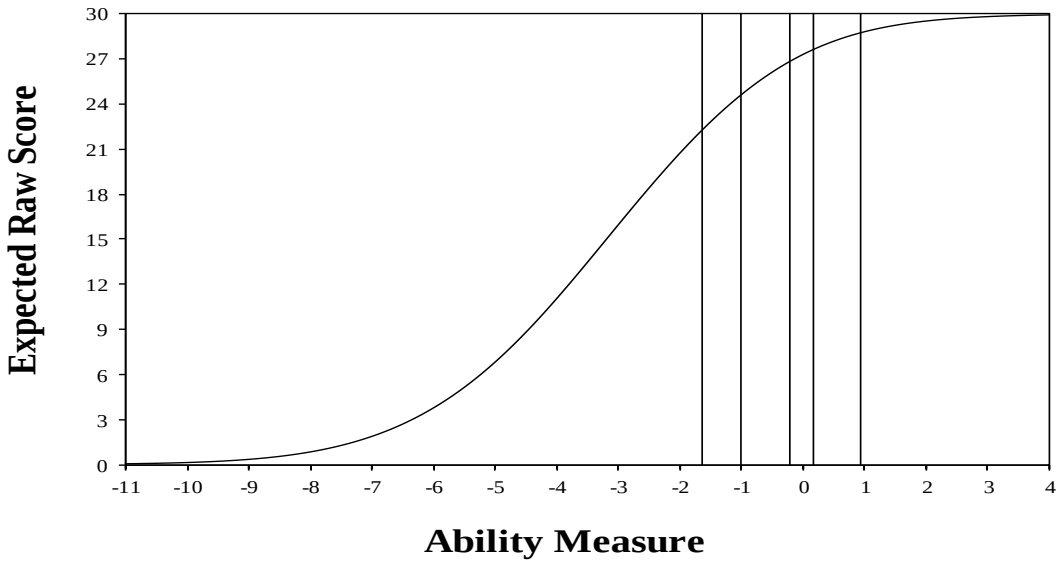


Figure 8.1.1D
Test Information Function: List K S400 Paper

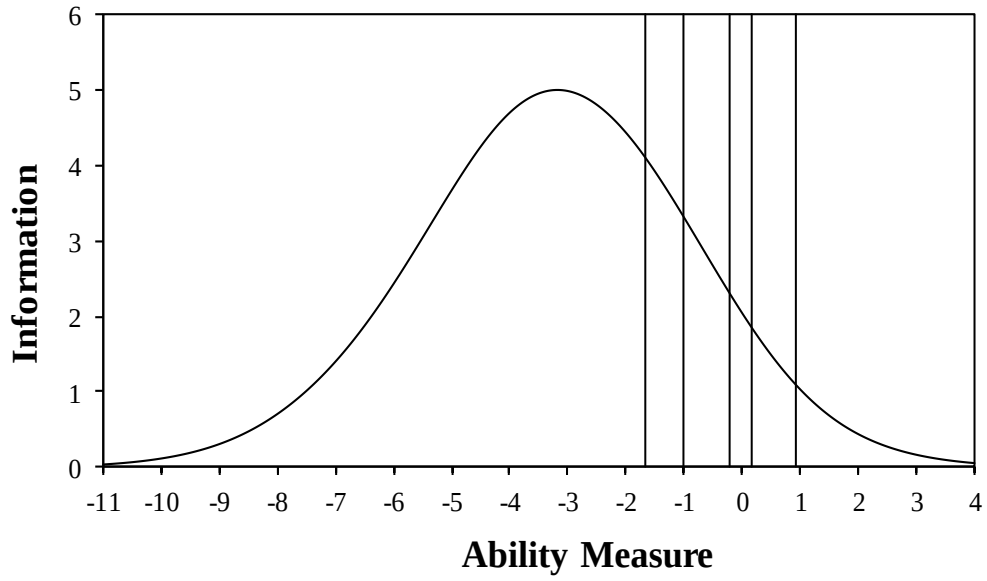


Table 8.1.1D
Reliability: List K S400 Paper

Tiers	No. of Students	Reliability
-	245,920	0.937

Table 8.1.1Ei
Accuracy and Consistency of Classification Indices: List (Grade K) S400 Paper (Accountability)

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.687	0.624		0.501	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.873		0.818	
	2	0.461		0.343	
	3	0.325		0.246	
	4	0.212		0.157	
	5	0.471		0.358	
	6	0.827		0.778	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.944	0.031	0.024	0.922
	2/3	0.932	0.027	0.041	0.905
	3/4	0.919	0.047	0.034	0.887
	4/5	0.909	0.043	0.048	0.878
	5/6	0.897	0.033	0.069	0.858

Table 8.1.1Eii
Accuracy and Consistency of Classification Indices: List (Grade K) S400 Paper (Instructional)

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.679	0.578		0.462	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.891		0.819	
	2	0.434		0.317	
	3	0.521		0.396	
	4	0.568		0.448	
	5	0.697		0.568	
	6	0.730		0.659	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.970	0.013	0.017	0.957
	2/3	0.960	0.020	0.020	0.943
	3/4	0.943	0.029	0.028	0.919
	4/5	0.920	0.039	0.041	0.889
	5/6	0.874	0.039	0.086	0.826

8.1.2 Reading K

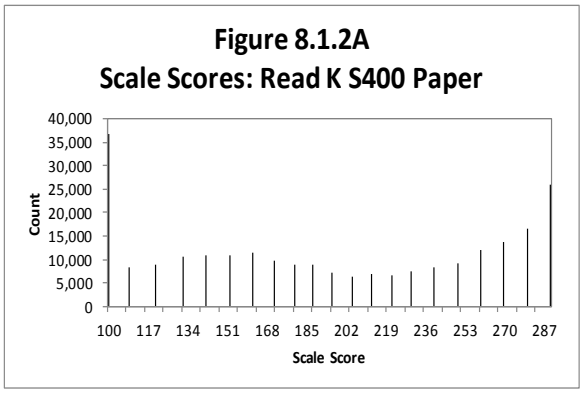


Table 8.1.2A

Scale Score Descriptive Statistics: Read K S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	245,920	100	290	194.15	66.93

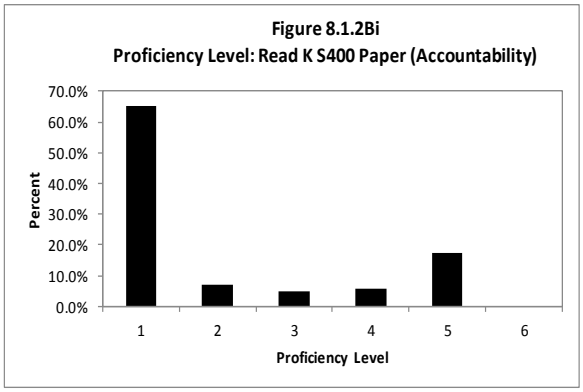


Table 8.1.2Bi

Proficiency Level Distribution: Read K S400 Paper (Accountability)

Level	Count	Percent
1	160,103	65.1%
2	17,576	7.1%
3	11,951	4.9%
4	13,821	5.6%
5	42,469	17.3%
6	0	0.0%
Total	245,920	100.0%

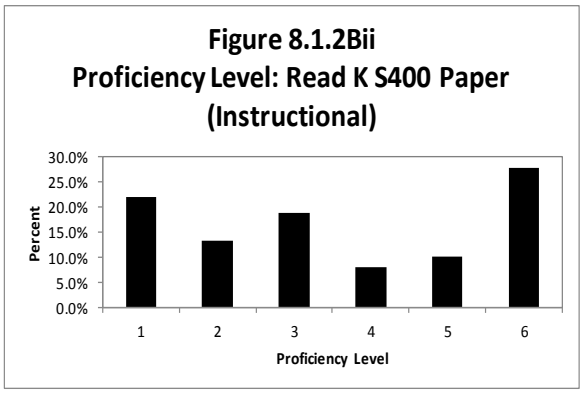


Table 8.1.2Bii

Proficiency Level Distribution: Read K S400 Paper (Instructional)

Level	Count	Percent
K1	54,136	22.0%
K2	32,382	13.2%
K3	46,058	18.7%
K4	19,956	8.1%
K5	25,147	10.2%
K6	68,241	27.7%
Total	245,920	100.0%

Table 8.1.2ci

Conditional Standard Error of Measurement at Cut Scores: Read K Accountability S400 Paper

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 Instructional S400 Paper

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 S400 Paper

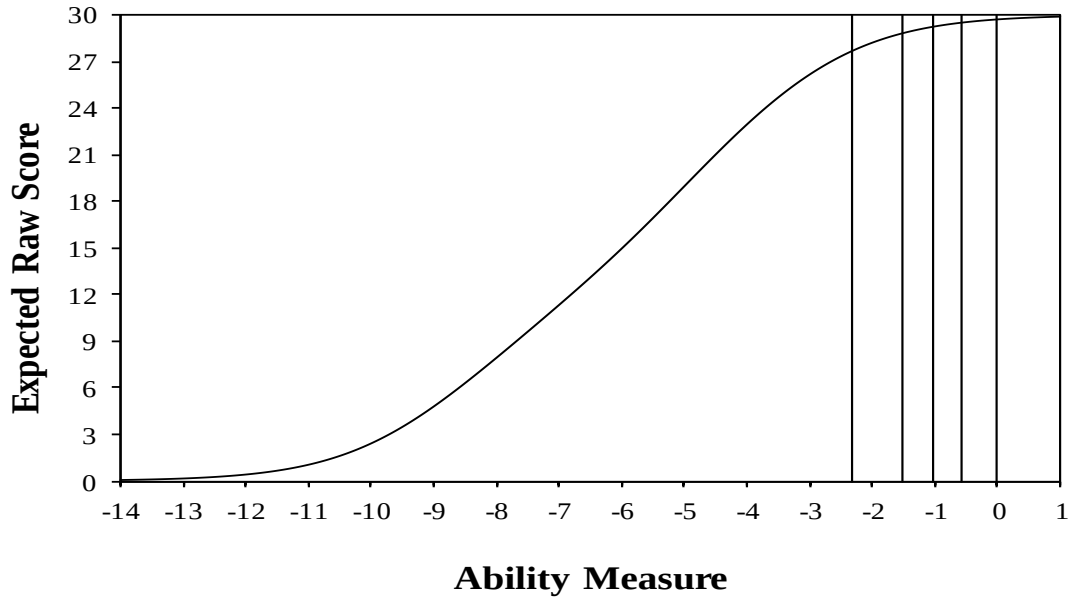


Figure 8.1.2D
Test Information Function: Read K S400 Paper

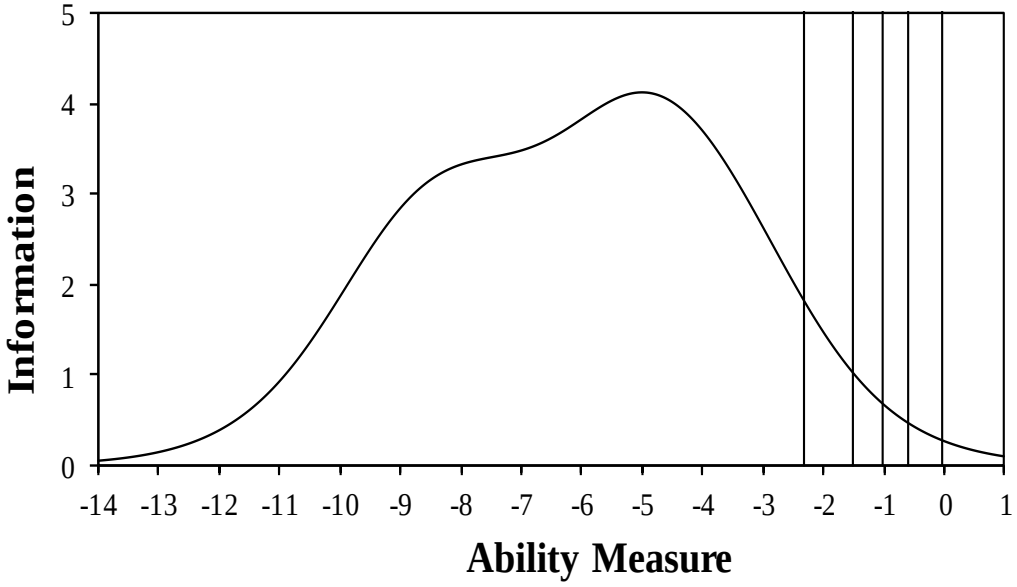


Table 8.1.2D
Reliability: Read K S400 Paper

Tiers	No. of Students	Reliability
-	245,920	0.949

Table 8.1.2Ei
Accuracy and Consistency of Classification Indices: Read (Grade K) S400 Paper (Accountability)

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.815	0.779		0.585	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.941		0.926	
	2	0.341		0.256	
	3	0.244		0.180	
	4	0.293		0.212	
	5	0.868		0.771	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.934	0.039	0.027	0.910
	2/3	0.936	0.032	0.033	0.911
	3/4	0.941	0.029	0.030	0.916
	4/5	0.945	0.034	0.021	0.920

Table 8.1.2Eii
Accuracy and Consistency of Classification Indices: Read (Grade K) S400 Paper (Instructional)

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.727	0.646		0.562	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.906		0.842	
	2	0.573		0.457	
	3	0.701		0.591	
	4	0.392		0.291	
	5	0.482		0.364	
	6	0.908		0.856	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.946	0.019	0.035	0.925
	2/3	0.943	0.032	0.025	0.919
	3/4	0.939	0.028	0.033	0.914
	4/5	0.943	0.030	0.027	0.918
	5/6	0.942	0.034	0.025	0.918

8.1.3 Writing K

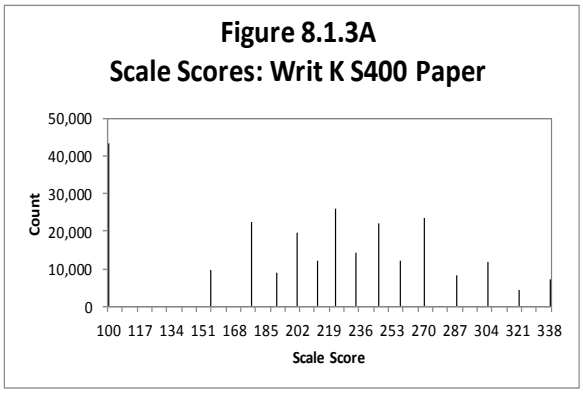


Table 8.1.3A

Scale Score Descriptive Statistics: Writ K S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	245,920	100	339	211.38	66.13

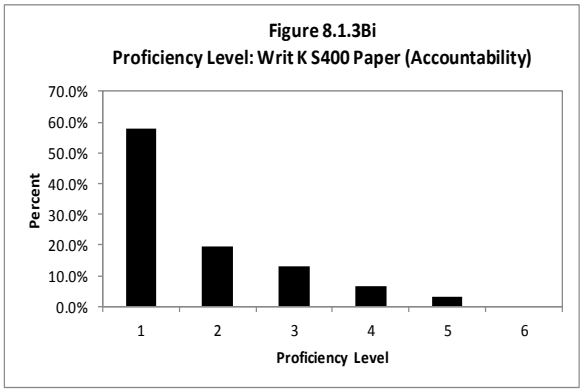


Table 8.1.3Bi

Proficiency Level Distribution: Writ K S400 Paper (Accountability)

Level	Count	Percent
1	142,202	57.8%
2	48,450	19.7%
3	31,715	12.9%
4	16,247	6.6%
5	7,306	3.0%
6	0	0.0%
Total	245,920	100.0%

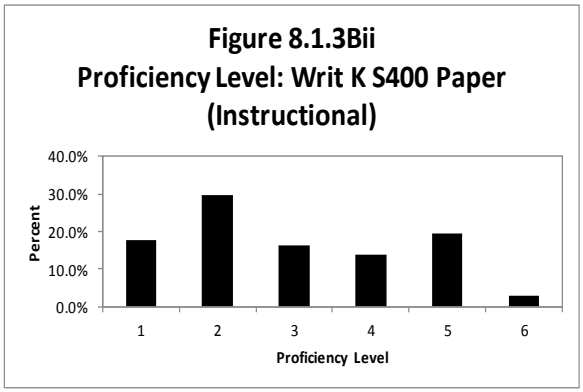


Table 8.1.3Bii

Proficiency Level Distribution: Writ K S400 Paper (Instructional)

Level	Count	Percent
K1	43,495	17.7%
K2	72,812	29.6%
K3	40,205	16.3%
K4	34,140	13.9%
K5	47,962	19.5%
K6	7,306	3.0%
Total	245,920	100.0%

Table 8.1.3ci

Conditional Standard Error of Measurement at Cut Scores: Writ K Accountability S400 Paper

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 Instructional S400 Paper

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 S400

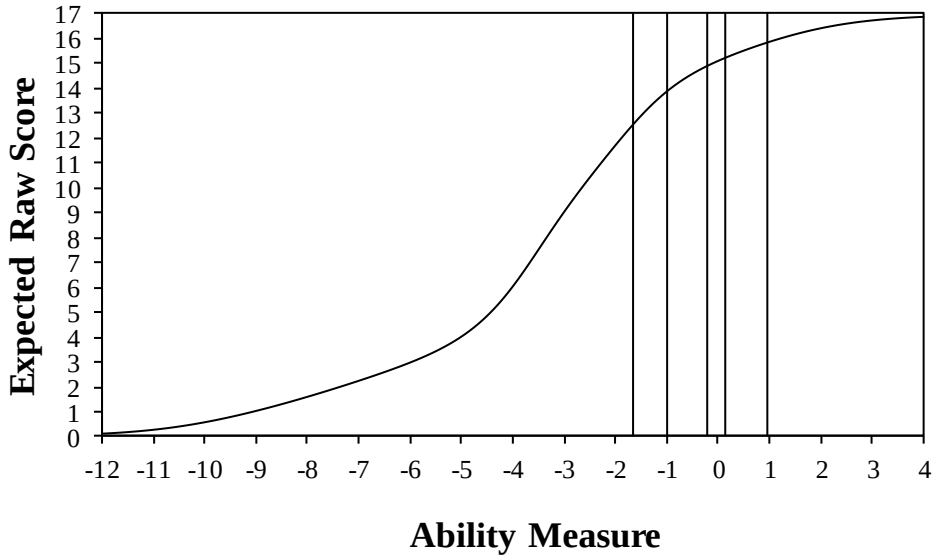


Figure 8.1.3D
Test Information Function: Writ K S400

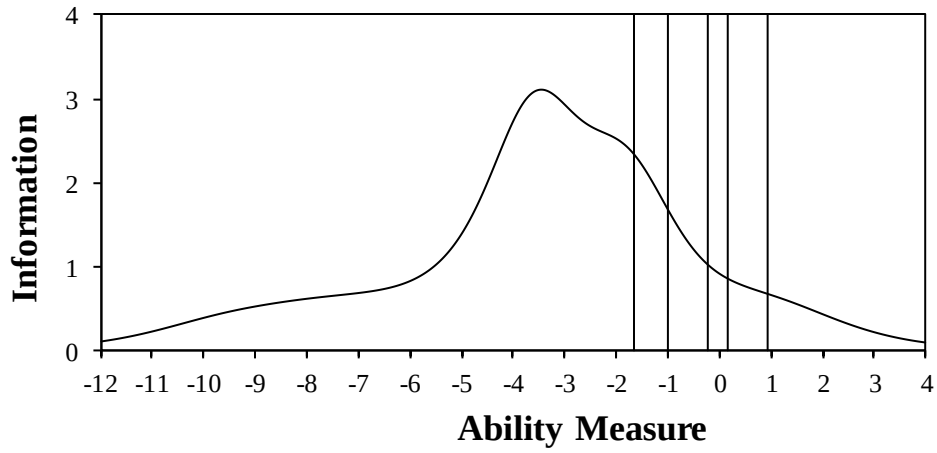


Table 8.1.3D
Reliability: Writ K S400 Paper

Tiers	No. of Students	Reliability
-	245,920	0.923

Table 8.1.3Ei
Accuracy and Consistency of Classification Indices: Writ (Grade K) S400 Paper (Accountability)

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.735	0.684		0.479	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.940		0.912	
	2	0.603		0.459	
	3	0.390		0.348	
	4	-		0.267	
	5	-		0.149	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.922	0.034	0.043	0.894
	2/3	0.902	0.022	0.076	0.863
	3/4	0.904	0.096	0.000	0.893
	4/5	0.970	0.030	0.000	0.969

Table 8.1.3Eii
Accuracy and Consistency of Classification Indices: Writ (Grade K) S400 Paper (Instructional)

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.659	0.571		0.459	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.865		0.791	
	2	0.789		0.704	
	3	0.507		0.383	
	4	0.374		0.278	
	5	0.600		0.532	
	6	-		0.135	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.955	0.024	0.020	0.935
	2/3	0.916	0.041	0.043	0.884
	3/4	0.899	0.031	0.070	0.862
	4/5	0.886	0.039	0.076	0.836
	5/6	0.970	0.030	0.000	0.969

8.1.4 Speaking K

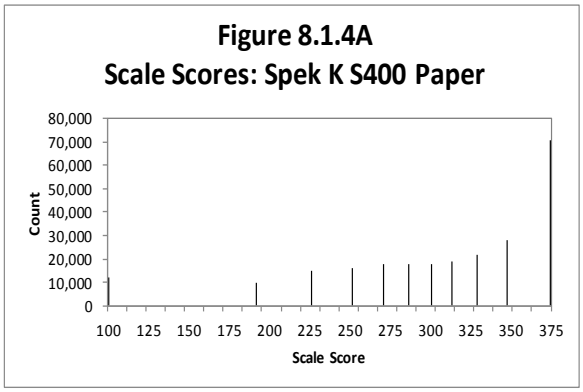


Table 8.1.4A

Scale Score Descriptive Statistics: Spek K S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	245,920	100	375	306.09	70.10

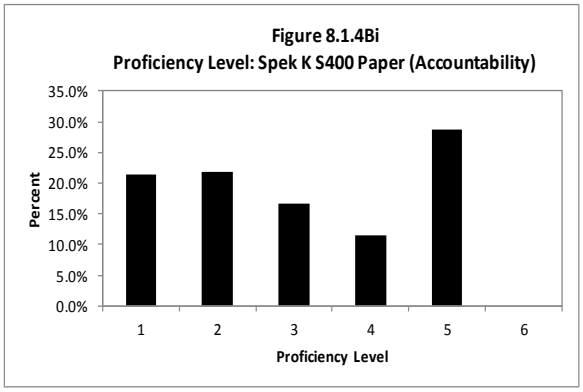


Table 8.1.4Bi

Proficiency Level Distribution: Spek K S400 Paper (Accountability)

Level	Count	Percent
1	52,694	21.4%
2	53,459	21.7%
3	40,763	16.6%
4	28,218	11.5%
5	70,786	28.8%
6	0	0.0%
Total	245,920	100.0%

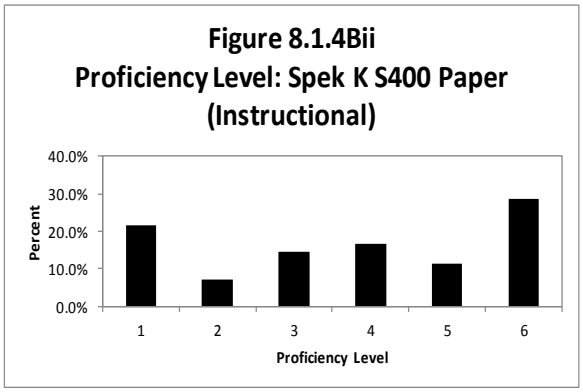


Table 8.1.4Bii

Proficiency Level Distribution: Spek K S400 Paper (Instructional)

Level	Count	Percent
K1	52,694	21.4%
K2	17,614	7.2%
K3	35,845	14.6%
K4	40,763	16.6%
K5	28,218	11.5%
K6	70,786	28.8%
Total	245,920	100.0%

Table 8.1.4ci
Conditional Standard Error of Measurement at Cut Scores: Spek K Accountability S400 Paper

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 Instructional S400 Paper

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 S400

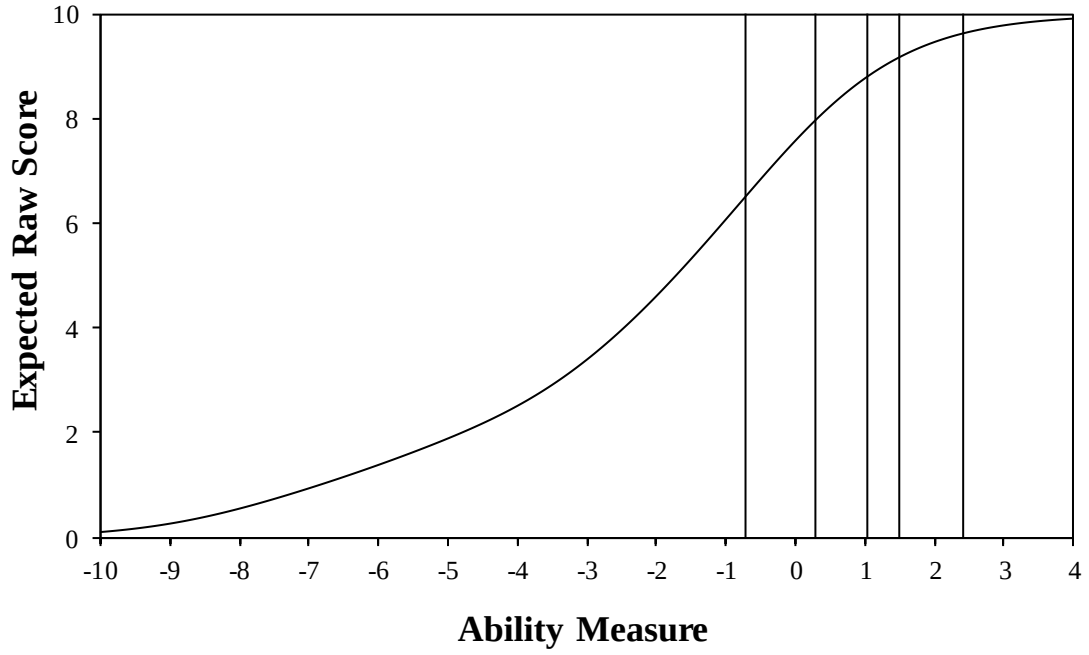


Figure 8.1.4D
Test Information Function: Spek K S400

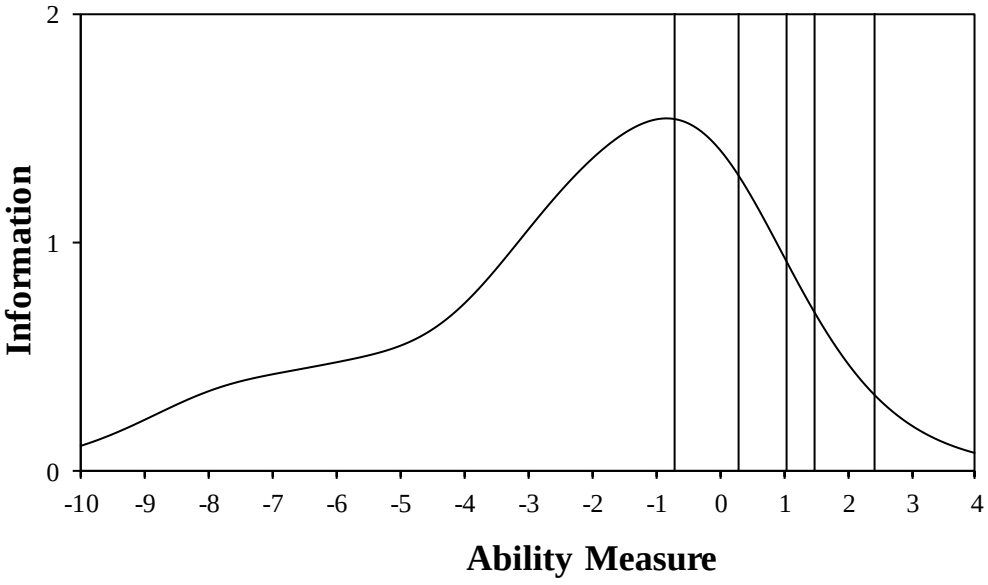


Table 8.1.4D
Reliability: Spek K S400 Paper

Tiers	No. of Students	Reliability
-	245,920	0.897

Table 8.1.4Ei
Accuracy and Consistency of Classification Indices: Spek (Grade K) S400 Paper (Accountability)

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.442	0.440		0.310	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.830		0.759	
	2	0.655		0.519	
	3	0.368		0.250	
	4	0.203		0.187	
	5	-		0.576	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.933	0.038	0.029	0.907
	2/3	0.892	0.031	0.078	0.852
	3/4	0.859	0.043	0.098	0.777
	4/5	0.712	0.288	0.000	0.734

Table 8.1.4Eii
Accuracy and Consistency of Classification Indices: S400 Paper (Grade K) S400 Paper (Instructional)

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.409	0.417		0.295	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.867		0.792	
	2	0.296		0.224	
	3	0.487		0.373	
	4	0.383		0.262	
	5	0.209		0.193	
	6	-		0.591	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.941	0.028	0.030	0.915
	2/3	0.916	0.046	0.037	0.888
	3/4	0.896	0.027	0.077	0.863
	4/5	0.871	0.041	0.087	0.792
	5/6	0.712	0.288	0.000	0.739

8.1.5 Oral Language Composite K

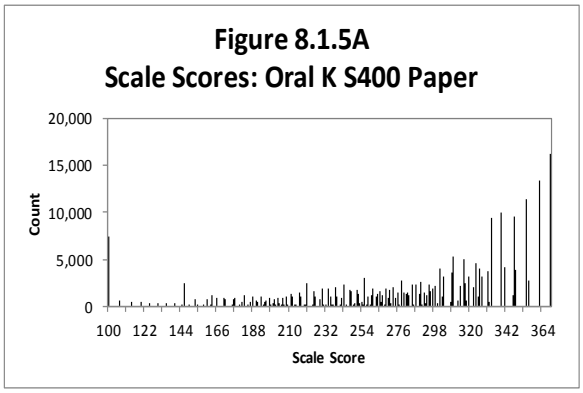


Table 8.1.5A

Scale Score Descriptive Statistics: Oral K S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
0	245,920	100	369	289.43	67.13

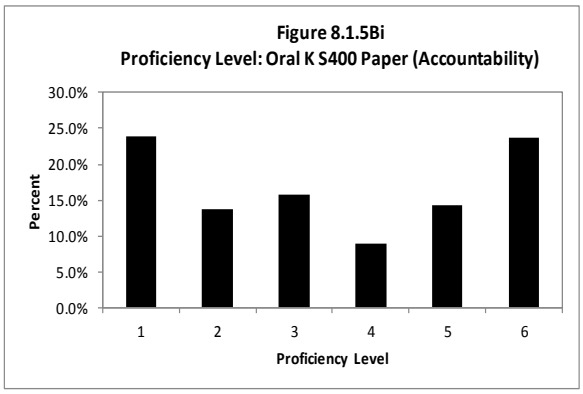


Table 8.1.5Bi

Proficiency Level Distribution: Oral K S400 Paper (Accountability)

Level	Count	Percent
1	58,608	23.8%
2	33,616	13.7%
3	38,621	15.7%
4	21,739	8.8%
5	34,931	14.2%
6	58,405	23.7%
Total	245,920	100.0%

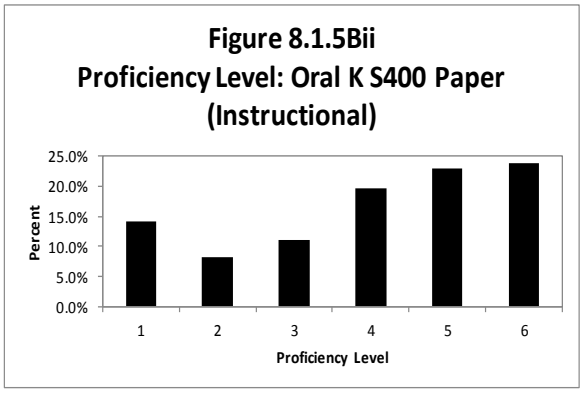


Table 8.1.5Bii

Proficiency Level Distribution: Oral K S400 Paper (Instructional)

Level	Count	Percent
K1	34,668	14.1%
K2	20,454	8.3%
K3	27,496	11.2%
K4	48,227	19.6%
K5	56,670	23.0%
K6	58,405	23.7%
Total	245,920	100.0%

Table 8.1.5Ci

n/a

Figure 8.1.5.Cii

n/a

Figure 8.1.5D

n/a

Table 8.1.5D
Oral Composite Reliability: Oral K S400 Paper

Component	Weight	Variance	Reliability
Listening	0.50	5145.919	0.937
Speaking	0.50	4913.404	0.897
Oral		4505.828	0.954

*Variances from students who had results in all four domains

Table 8.1.5Ei
Accuracy and Consistency of Classification Indices: Oral (Grade K) S400 Paper (Accountability)

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.642	0.550		0.451	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.901		0.853	
	2	0.619		0.500	
	3	0.597		0.481	
	4	0.360		0.247	
	5	0.380		0.298	
	6	0.757		0.667	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.955	0.024	0.021	0.936
	2/3	0.936	0.030	0.034	0.912
	3/4	0.929	0.026	0.045	0.902
	4/5	0.930	0.025	0.046	0.893
	5/6	0.865	0.086	0.049	0.825

Table 8.1.5Eii

Accuracy and Consistency of Classification Indices: Oral (Grade K) S400 Paper (Instructional)

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.668	0.563		0.462	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.880		0.817	
	2	0.547		0.422	
	3	0.539		0.425	
	4	0.690		0.573	
	5	0.568		0.450	
	6	0.730		0.648	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.968	0.017	0.014	0.954
	2/3	0.951	0.024	0.025	0.932
	3/4	0.939	0.028	0.033	0.915
	4/5	0.932	0.024	0.044	0.904
	5/6	0.871	0.066	0.063	0.825

8.1.6 Literacy Composite K

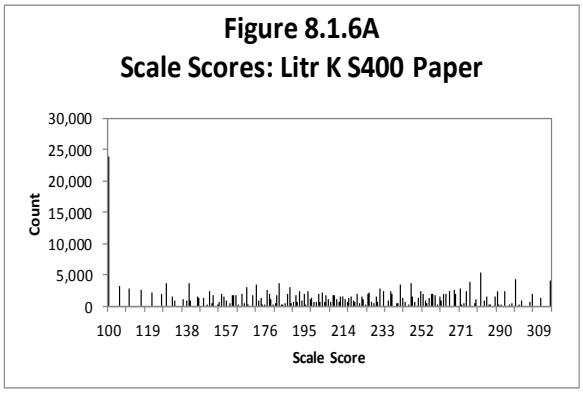


Table 8.1.6A

Scale Score Descriptive Statistics: Litr K S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	245,920	100	315	203.02	61.79

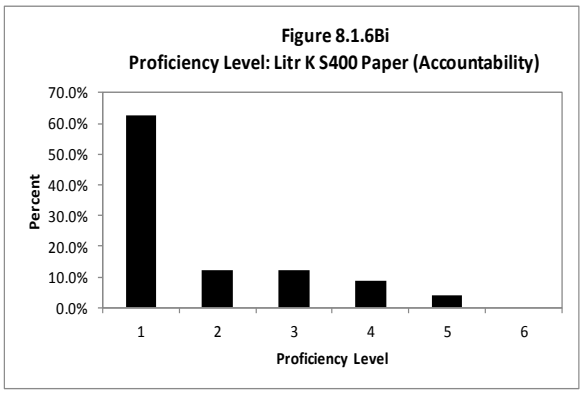


Table 8.1.6Bi

Proficiency Level Distribution: Litr K S400 Paper (Accountability)

Level	Count	Percent
1	154,227	62.7%
2	30,075	12.2%
3	30,360	12.3%
4	21,645	8.8%
5	9,613	3.9%
6	0	0.0%
Total	245,920	100.0%

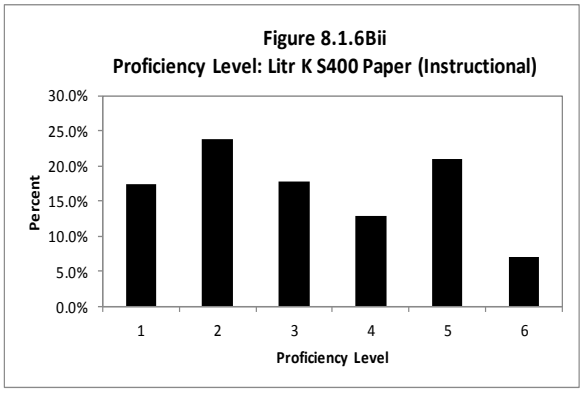


Table 8.1.6Bii

Proficiency Level Distribution: Litr K S400 Paper (Instructional)

Level	Count	Percent
K1	42,982	17.5%
K2	58,617	23.8%
K3	43,798	17.8%
K4	31,536	12.8%
K5	51,688	21.0%
K6	17,299	7.0%
Total	245,920	100.0%

Table 8.1.6Ci

n/a

Figure 8.1.6Cii

n/a

Figure 8.1.6D

n/a

Table 8.1.6D
Literacy Composite Reliability: Litr K S400 Paper

Component	Weight	Variance	Reliability
Reading	0.50	4479.701	0.949
Writing	0.50	4372.858	0.923
Literacy		3818.479	0.963

*Variances from students who had results in all four domains

Table 8.1.6Ei
Accuracy and Consistency of Classification Indices: Litr (Grade K) S400 Paper (Accountability)

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.806	0.755		0.567	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.956		0.939	
	2	0.582		0.461	
	3	0.582		0.446	
	4	0.501		0.442	
	5	-		0.337	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.946	0.028	0.027	0.924
	2/3	0.947	0.024	0.029	0.925
	3/4	0.948	0.024	0.028	0.922
	4/5	0.961	0.039	0.000	0.957

Table 8.1.6Eii

Accuracy and Consistency of Classification Indices: Litr (Grade K) S400 Paper (Instructional)

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.735	0.662	0.584
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.905	0.850
	2	0.809	0.735
	3	0.700	0.591
	4	0.593	0.471
	5	0.654	0.635
	6	-	0.477

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.964	0.016	0.020	0.949
	2/3	0.948	0.027	0.025	0.926
	3/4	0.942	0.027	0.030	0.920
	4/5	0.949	0.022	0.029	0.927
	5/6	0.930	0.070	0.000	0.927

8.1.7 Comprehension Composite K

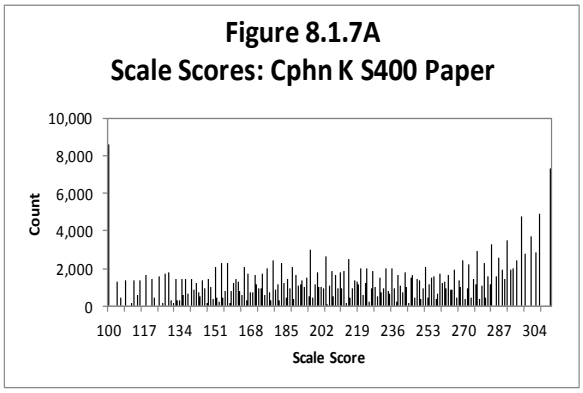


Table 8.1.7A

Scale Score Descriptive Statistics: Cphn K S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	245,920	100	312	217.59	61.13

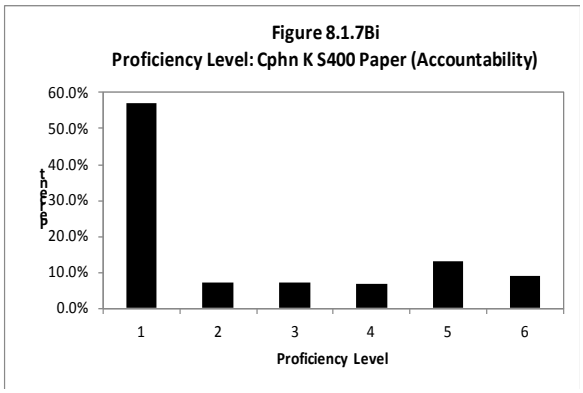


Table 8.1.7Bi

Proficiency Level Distribution: Cphn K S400 Paper (Accountability)

Level	Count	Percent
1	140,502	57.1%
2	17,239	7.0%
3	17,351	7.1%
4	16,789	6.8%
5	32,396	13.2%
6	21,643	8.8%
Total	245,920	100.0%

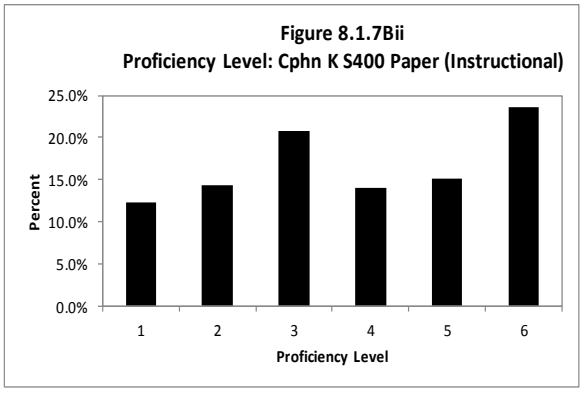


Table 8.1.7Bii

Proficiency Level Distribution: Cphn K S400 Paper (Instructional)

Level	Count	Percent
K1	30,075	12.2%
K2	35,081	14.3%
K3	51,150	20.8%
K4	34,428	14.0%
K5	37,075	15.1%
K6	58,111	23.6%
Total	245,920	100.0%

Table 8.1.7Ci

n/a

Figure 8.1.7Cii

n/a

Figure 8.1.7D

n/a

Table 8.1.7D
Comprehension Composite Reliability: Cphn K S400 Paper

Component	Weight	Variance	Reliability
Listening	0.30	5145.919	0.937
Reading	0.70	4479.701	0.949
Comprehension		3737.336	0.962

*Variances from students who had results in all four domains

Table 8.7.1Ei
Accuracy and Consistency of Classification Indices: Cphn (Grade K) S400 Paper (Accountability)

Overall Indices	Accuracy	Consistency	Kappa (k)
	0.785	0.730	0.575
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.956	0.938
	2	0.392	0.288
	3	0.403	0.298
	4	0.398	0.299
	5	0.673	0.565
	6	0.818	0.695

Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.947	0.025	0.028	0.925
	2/3	0.950	0.027	0.023	0.928
	3/4	0.949	0.030	0.021	0.929
	4/5	0.950	0.029	0.021	0.930
	5/6	0.963	0.022	0.015	0.947

Table 8.1.7Eii

Accuracy and Consistency of Classification Indices: Cphn (Grade K) S400 Paper (Instructional)

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.760	0.671	0.601
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.879	0.803
	2	0.706	0.598
	3	0.743	0.647
	4	0.612	0.493
	5	0.657	0.544
	6	0.914	0.864

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.967	0.014	0.018	0.953
	2/3	0.950	0.024	0.026	0.929
	3/4	0.942	0.027	0.031	0.919
	4/5	0.948	0.026	0.026	0.926
	5/6	0.951	0.029	0.019	0.932

8.1.8 Overall Composite K

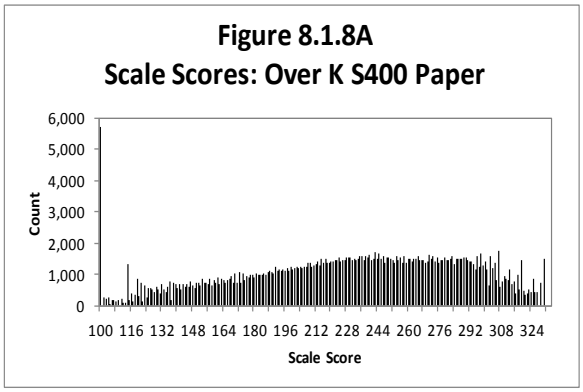


Table 8.1.8A

Scale Score Descriptive Statistics: Over K S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	245,920	100	331	228.73	57.72

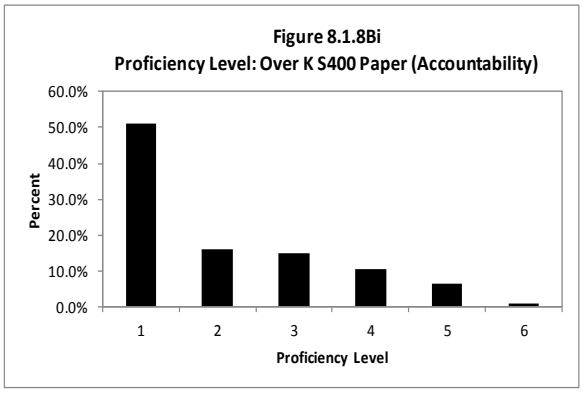


Table 8.1.8Bi

Proficiency Level Distribution: Over K S400 Paper (Accountability)

Level	Count	Percent
1	126,036	51.3%
2	39,370	16.0%
3	37,062	15.1%
4	25,417	10.3%
5	15,756	6.4%
6	2,279	0.9%
Total	245,920	100.0%

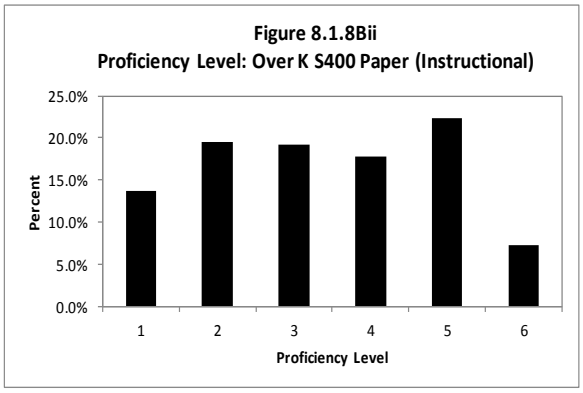


Table 8.1.8Bii

Proficiency Level Distribution: Over K S400 Paper (Instructional)

Level	Count	Percent
K1	33,777	13.7%
K2	47,902	19.5%
K3	47,405	19.3%
K4	43,817	17.8%
K5	54,984	22.4%
K6	18,035	7.3%
Total	245,920	100.0%

Table 8.1.8Ci

n/a

Figure 8.1.8Cii

n/a

Figure 8.1.8D

n/a

Table 8.1.8D
Overall Composite Reliability: Over K S400 Paper

Component	Weight	Variance	Reliability
Listening	0.15	5145.919	0.937
Reading	0.35	4479.701	0.949
Speaking	0.15	4913.404	0.897
Writing	0.35	4372.858	0.923
Overall Composite		3331.258	0.974

*Variances from students who had results in all four domains

Table 8.1.8Ei
Accuracy and Consistency of Classification Indices: Over (Grade K) S400 Paper (Accountability)

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.815	0.754		0.634	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.952		0.933	
	2	0.708		0.598	
	3	0.711		0.601	
	4	0.595		0.487	
	5	0.620		0.544	
	6	-		-	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.952	0.024	0.023	0.933
	2/3	0.952	0.022	0.025	0.932
	3/4	0.961	0.018	0.021	0.944
	4/5	0.959	0.030	0.011	0.944
	5/6	0.991	0.009	0.000	0.991

Table 8.1.8Eii

Accuracy and Consistency of Classification Indices: Over (Grade K) S400 Paper (Instructional)

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.789	0.711	0.645
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.913	0.866
	2	0.819	0.747
	3	0.762	0.668
	4	0.734	0.631
	5	0.764	0.704
	6	0.759	0.619

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.975	0.012	0.014	0.964
	2/3	0.957	0.022	0.021	0.939
	3/4	0.949	0.024	0.027	0.928
	4/5	0.955	0.020	0.024	0.936
	5/6	0.953	0.035	0.012	0.941

8.2 Grade: 1

8.2.1 Listening 1

Figure 8.2.1A
Scale Scores: List 1 S400 Paper

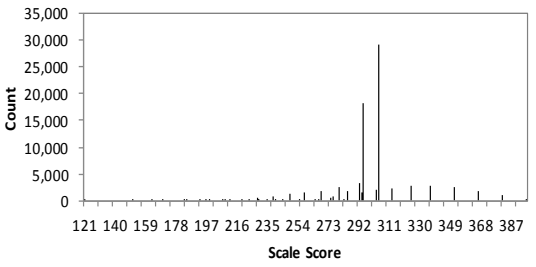


Figure 8.2.1B
Proficiency Level: List 1 S400 Paper

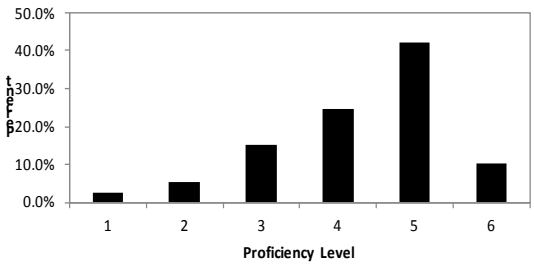


Table 8.2.1A
Scale Score Descriptive Statistics: List 1 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	81,435	121	397	300.66	27.51

Table 8.2.1B
Proficiency Level Distribution: List 1 S400 Paper

Level	Count	Percent
1	1,976	2.4%
2	4,319	5.3%
3	12,294	15.1%
4	20,087	24.7%
5	34,299	42.1%
6	8,460	10.4%
Total	81,435	100.0%

Table 8.2.1C
Conditional Standard Error of Measurement at Cut Scores: List 1 S400 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	1	238	19.16	19.91	19.16
2/3	1	267	19.91	19.16	18.03
3/4	1	295	22.92	19.54	18.79
4/5	1	305	n/a	19.91	19.16
5/6	1	330	n/a	n/a	21.79

Figure 8.2.1C
Test Characteristic Curve: List 1ABC S400 Paper

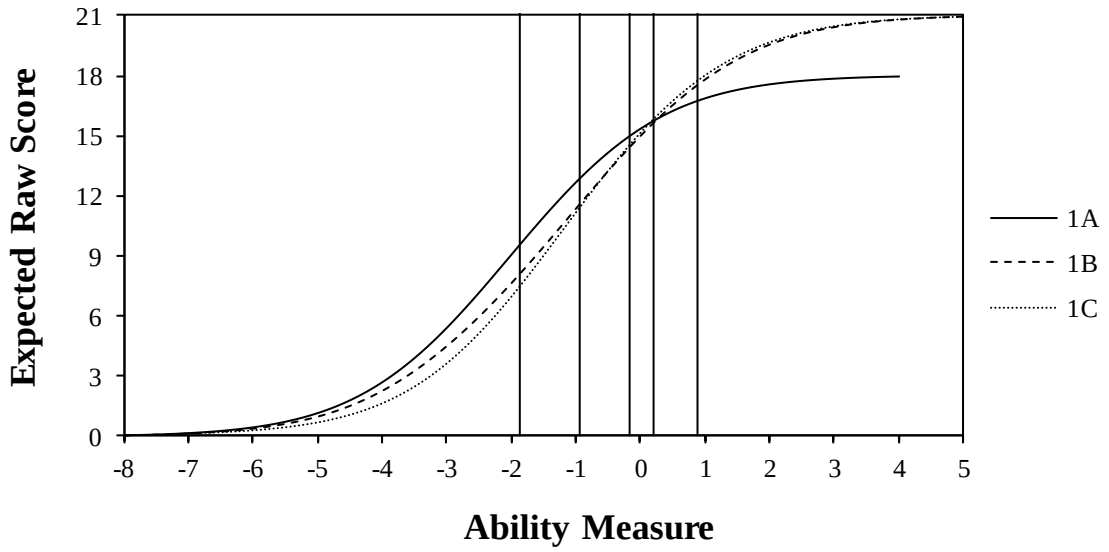


Figure 8.2.1D
Test Information Function: List 1ABC S400 Paper

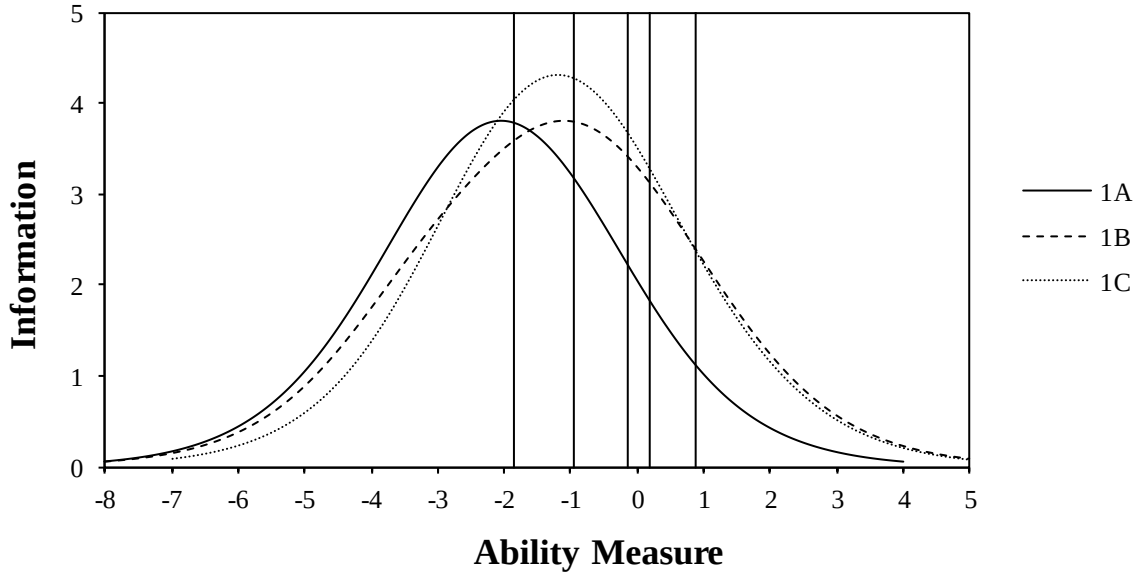


Table 8.2.1D
Weighted Reliability: List 1 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	29,448	0.750	0.668
B	32,225	0.630	
C	19,762	0.610	

Table 8.2.1E
Accuracy and Consistency of Classification Indices: List (Grade 1) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.481	0.374		0.177	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.850		0.388	
	2	0.371		0.231	
	3	0.337		0.249	
	4	0.371		0.311	
	5	0.629		0.546	
	6	0.581		0.357	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.976	0.000	0.024	0.974
	2/3	0.946	0.013	0.042	0.908
	3/4	0.802	0.150	0.048	0.733
	4/5	0.754	0.146	0.100	0.693
	5/6	0.911	0.050	0.039	0.852

8.2.2 Reading 1

Figure 8.2.2A
Scale Scores: Read 1 S400 Paper

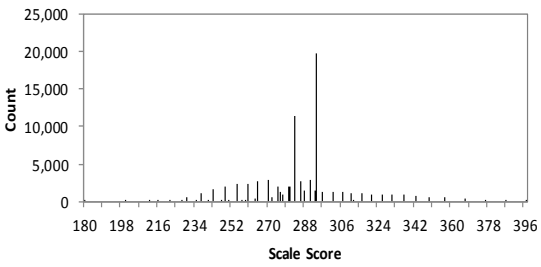


Figure 8.2.2B
Proficiency Level: Read 1 S400 Paper

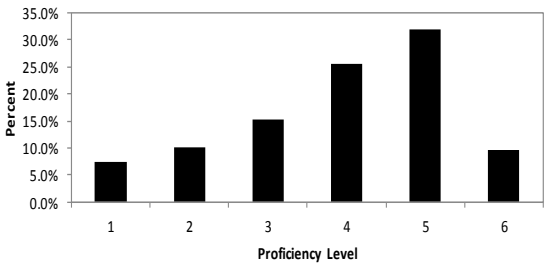


Table 8.2.2A
Scale Score Descriptive Statistics: Read 1 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	77,556	180	397	286.45	23.76

Table 8.2.2B
Proficiency Level Distribution: Read 1 S400 Paper

Level	Count	Percent
1	5,804	7.5%
2	7,940	10.2%
3	11,852	15.3%
4	19,747	25.5%
5	24,744	31.9%
6	7,469	9.6%
Total	77,556	100.0%

Table 8.2.2C
Conditional Standard Error of Measurement at Cut Scores: Read 1 S400 Paper

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

Figure 8.2.2C
Test Characteristic Curve: Read 1ABC S400 Paper

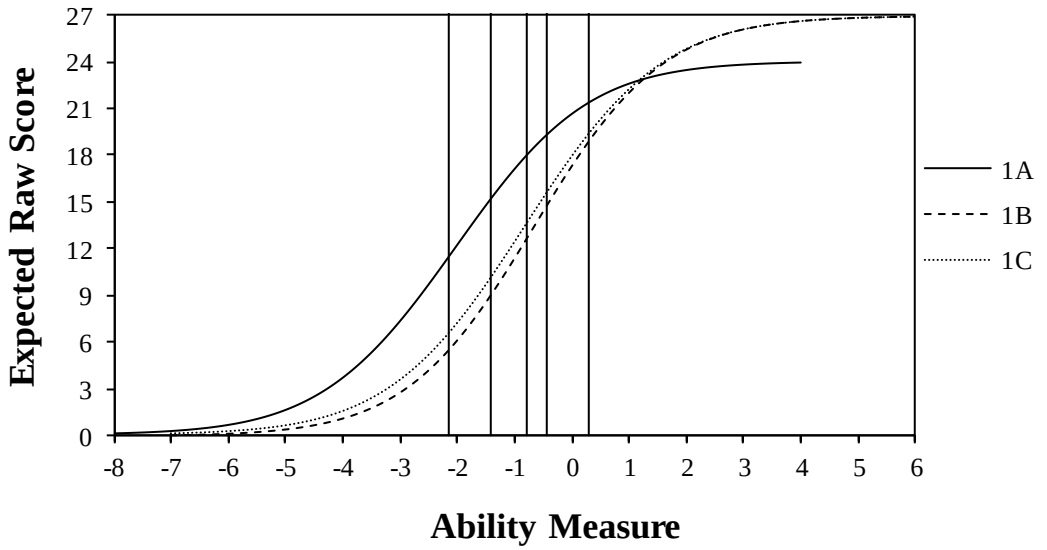


Figure 8.2.2D
Test Information Function: Read 1ABC S400 Paper

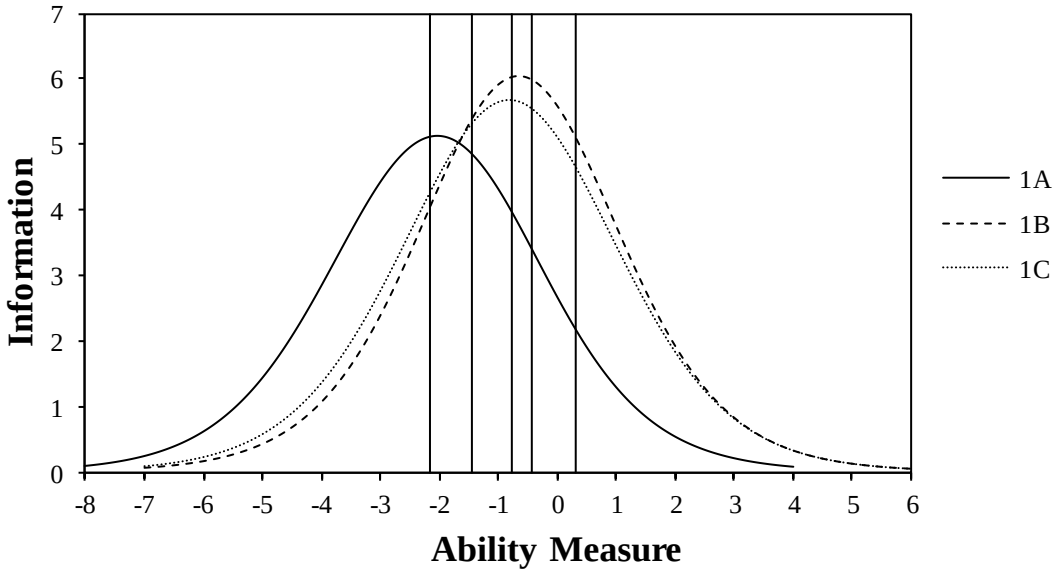


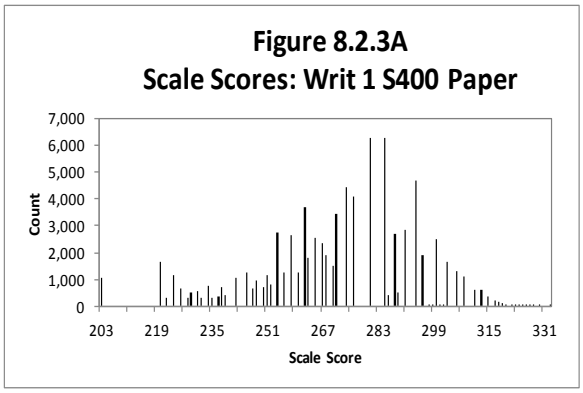
Table 8.2.2D
Weighted Reliability: Read 1 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	28,886	0.776	0.773
B	29,930	0.766	
C	18,732	0.783	

Table 8.2.2E
Accuracy and Consistency of Classification Indices: Read (Grade 1) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.468	0.377		0.226	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	-		0.351	
	2	0.332		0.252	
	3	0.311		0.238	
	4	0.437		0.359	
	5	0.657		0.543	
	6	0.668		0.487	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.925	0.000	0.075	0.913
	2/3	0.888	0.056	0.056	0.827
	3/4	0.820	0.137	0.043	0.761
	4/5	0.819	0.116	0.065	0.768
	5/6	0.937	0.030	0.033	0.900

8.2.3 Writing 1



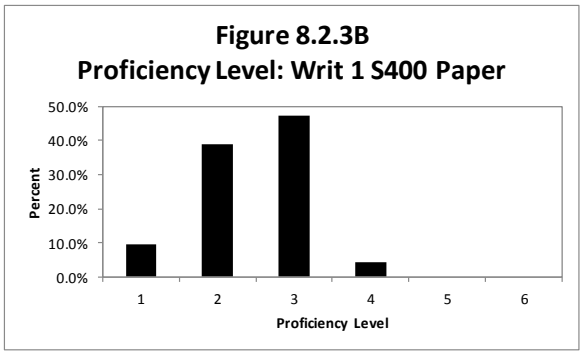


Table 8.2.3A

Scale Score Descriptive Statistics: Writ 1 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	84,178	203	333	271.19	23.22

Table 8.2.3B

Proficiency Level Distribution: Writ 1 S400 Paper

Level	Count	Percent
1	8,036	9.5%
2	32,843	39.0%
3	39,814	47.3%
4	3,485	4.1%
5	0	0.0%
6	0	0.0%
Total	84,178	100.0%

Table 8.2.3C

Conditional Standard Error of Measurement at Cut Scores: Writ 1 S400 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	1	238	7.15	6.84
2/3	1	272	7.77	8.09
3/4	1	308	8.09	8.09
4/5	1	336	7.46	7.15
5/6	1	362	6.53	6.53

Figure 8.2.3C

n/a

Figure 8.2.3D

n/a

Table 8.2.3D
Weighted Reliability: Writ 1 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	30,871	0.874	0.912
B/C	53,307	0.934	

Table 8.2.3E
Accuracy and Consistency of Classification Indices: Writ (Grade 1) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.839	0.783		0.640	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.839		0.735	
	2	0.857		0.797	
	3	0.827		0.798	
	4	-		0.280	
	5	N/A		N/A	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.967	0.015	0.019	0.952
	2/3	0.914	0.036	0.050	0.880
	3/4	0.959	0.041	0.000	0.952

8.2.4 Speaking 1

Figure 8.2.4A
Scale Scores: Spek 1 S400 Paper

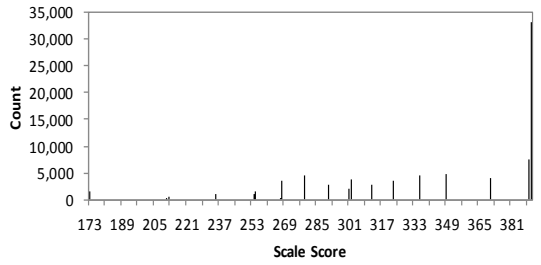


Figure 8.2.4B
Proficiency Level: Spek 1 S400 Paper

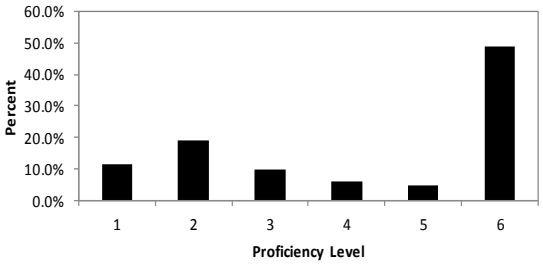


Table 8.2.4A
Scale Score Descriptive Statistics: Spek 1 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	83,486	173	391	345.68	54.72

Table 8.2.4B
Proficiency Level Distribution: Spek 1 S400 Paper

Level	Count	Percent
1	9,733	11.7%
2	15,959	19.1%
3	8,191	9.8%
4	4,865	5.8%
5	4,062	4.9%
6	40,676	48.7%
Total	83,486	100.0%

Table 8.2.4C
Conditional Standard Error of Measurement at Cut Scores: Spek 1 S400 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	1	278	21.43	21.43
2/3	1	318	20.41	20.41
3/4	1	344	19.39	19.39
4/5	1	367	19.39	19.39
5/6	1	385	19.39	19.39

Figure 8.2.4C

n/a

Figure 8.2.4D

n/a

Table 8.2.4D
Weighted Reliability: Spek 1 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	30,570	0.848	0.885
B/C	52,916	0.906	

Table 8.2.4E
Accuracy and Consistency of Classification Indices: Spek (Grade 1) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.659	0.579		0.432	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.650		0.555	
	2	0.645		0.543	
	3	0.384		0.296	
	4	0.292		0.191	
	5	0.165		0.102	
	6	0.956		0.915	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.927	0.050	0.023	0.900
	2/3	0.913	0.033	0.054	0.887
	3/4	0.937	0.020	0.043	0.906
	4/5	0.951	0.024	0.025	0.917
	5/6	0.887	0.095	0.018	0.846

8.2.5 Oral Language Composite 1

Figure 8.2.5A
Scale Scores: Oral 1 S400 Paper

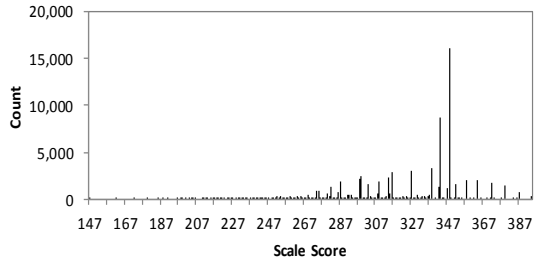


Figure 8.2.5B
Proficiency Level: Oral 1 S400 Paper

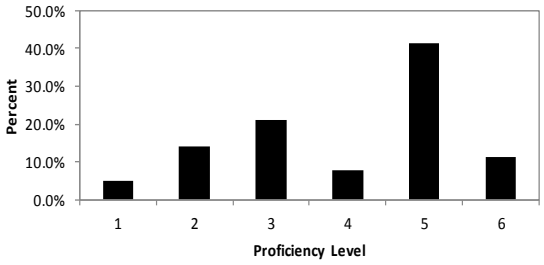


Table 8.2.5A
Scale Score Descriptive Statistics: Oral 1 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	80,880	147	394	323.72	35.85

Table 8.2.5B
Proficiency Level Distribution: Oral 1 S400 Paper

Level	Count	Percent
1	3,977	4.9%
2	11,357	14.0%
3	16,982	21.0%
4	6,178	7.6%
5	33,346	41.2%
6	9,040	11.2%
Total	80,880	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 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.50	747.597	0.668
Speaking	0.50	2912.176	0.885
Oral		1273.903	0.886

*Variances from students who had results in all four domains

Table 8.2.5E
Accuracy and Consistency of Classification Indices: Oral (Grade 1) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.628	0.508		0.369	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.784		0.643	
	2	0.709		0.588	
	3	0.679		0.549	
	4	0.230		0.161	
	5	0.759		0.685	
	6	0.482		0.339	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.978	0.010	0.012	0.967
	2/3	0.936	0.026	0.038	0.912
	3/4	0.919	0.024	0.057	0.880
	4/5	0.893	0.074	0.033	0.843
	5/6	0.884	0.059	0.057	0.839

8.2.6 Literacy Composite 1

Figure 8.2.6A
Scale Scores: Litr 1 S400 Paper

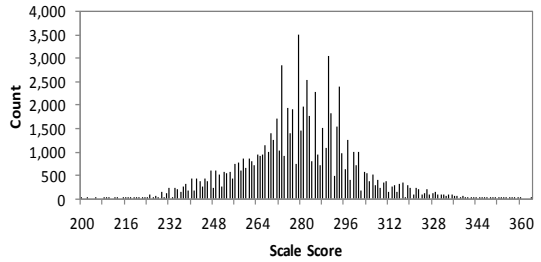


Figure 8.2.6B
Proficiency Level: Litr 1 S400 Paper

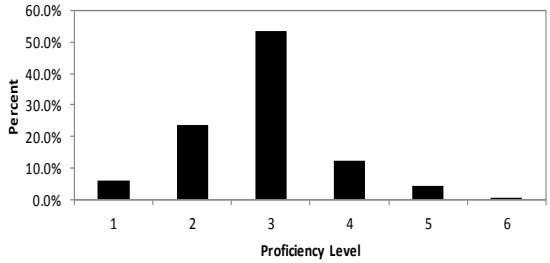


Table 8.2.6A
Scale Score Descriptive Statistics: Litr 1 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	77,556	200	364	279.37	20.33

Table 8.2.6B
Proficiency Level Distribution: Litr 1 S400 Paper

Level	Count	Percent
1	4,504	5.8%
2	18,310	23.6%
3	41,369	53.3%
4	9,587	12.4%
5	3,338	4.3%
6	448	0.6%
Total	77,556	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 S400 Paper

Component	Weight	Variance	Reliability
Reading	0.50	562.489	0.773
Writing	0.50	509.875	0.912
Literacy		409.334	0.895

*Variances from students who had results in all four domains

Table 8.2.6E
Accuracy and Consistency of Classification Indices: Litr (Grade 1) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.781	0.692		0.531	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.845		0.693	
	2	0.708		0.611	
	3	0.860		0.811	
	4	0.621		0.487	
	5	0.751		0.623	
	6	0.932		0.767	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.971	0.006	0.023	0.961
	2/3	0.904	0.055	0.041	0.863
	3/4	0.931	0.031	0.037	0.899
	4/5	0.979	0.016	0.005	0.971
	5/6	0.996	0.004	0.000	0.996

8.2.7 Comprehension Composite 1

Figure 8.2.7A
Scale Scores: Cphn 1 S400 Paper

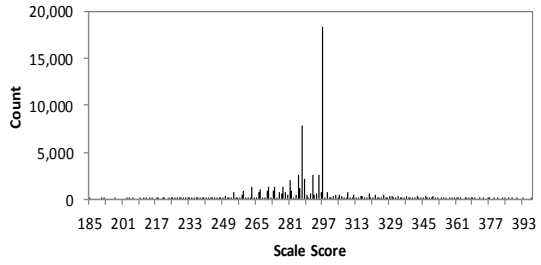


Figure 8.2.7B
Proficiency Level: Cphn 1 S400 Paper

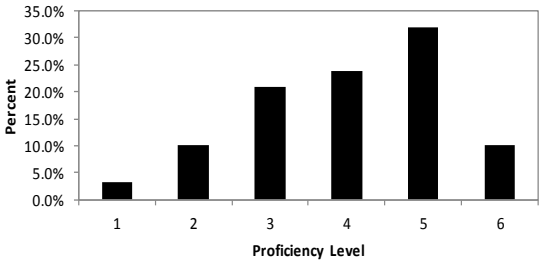


Table 8.2.7A

Scale Score Descriptive Statistics: Cphn 1 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	75,575	185	397	290.94	22.76

Table 8.2.7B

Proficiency Level Distribution: Cphn 1 S400 Paper

Level	Count	Percent
1	2,452	3.2%
2	7,614	10.1%
3	15,733	20.8%
4	17,983	23.8%
5	24,194	32.0%
6	7,599	10.1%
Total	75,575	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 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.30	747.597	0.668
Reading	0.70	562.489	0.773
Comprehension		518.169	0.836

*Variances from students who had results in all four domains

Table 8.2.7E
Accuracy and Consistency of Classification Indices: Cphn (Grade 1) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.575	0.472		0.324	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	-		0.287	
	2	0.484		0.380	
	3	0.478		0.381	
	4	0.487		0.393	
	5	0.708		0.603	
	6	0.748		0.592	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.968	0.000	0.032	0.965
	2/3	0.918	0.027	0.055	0.880
	3/4	0.856	0.106	0.039	0.805
	4/5	0.853	0.086	0.061	0.806
	5/6	0.949	0.026	0.025	0.922

8.2.8 Overall Composite 1

Figure 8.2.8A
Scale Scores: Over 1 S400 Paper

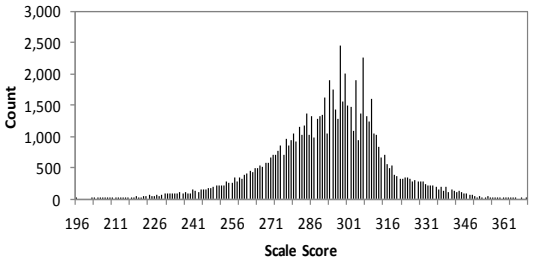


Figure 8.2.8B
Proficiency Level: Over 1 S400 Paper

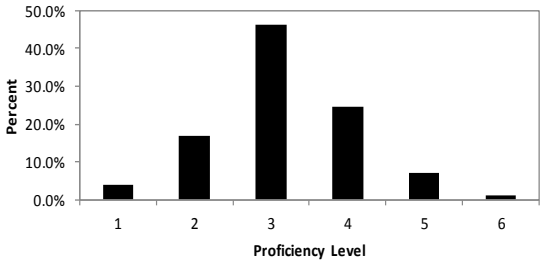


Table 8.2.8A

Scale Score Descriptive Statistics: Over 1 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	75,085	196	370	292.73	22.45

Table 8.2.8B

Proficiency Level Distribution: Over 1 S400 Paper

Level	Count	Percent
1	2,961	3.9%
2	12,578	16.8%
3	34,790	46.3%
4	18,384	24.5%
5	5,412	7.2%
6	960	1.3%
Total	75,085	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 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.15	747.597	0.668
Reading	0.35	562.489	0.773
Speaking	0.15	2912.176	0.885
Writing	0.35	509.875	0.912
Overall Composite		504.069	0.932

*Variances from students who had results in all four domains

Table 8.2.8E

Accuracy and Consistency of Classification Indices: Over (Grade 1) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.795	0.714		0.590	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.868		0.772	
	2	0.771		0.677	
	3	0.853		0.800	
	4	0.745		0.649	
	5	0.639		0.533	
	6	-		0.784	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.985	0.004	0.010	0.980
	2/3	0.943	0.031	0.027	0.918
	3/4	0.916	0.040	0.044	0.882
	4/5	0.962	0.021	0.017	0.946
	5/6	0.987	0.013	0.000	0.988

8.3 Grade: 2

8.3.1 Listening 2

Figure 8.3.1A
Scale Scores: List 2 S400 Paper

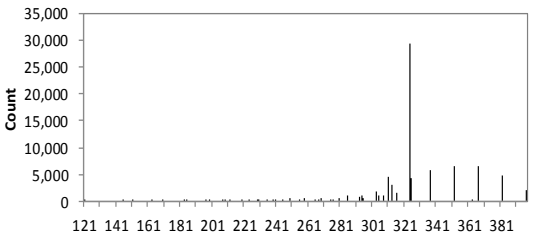


Figure 8.3.1B
Proficiency Level: List 2 S400 Paper

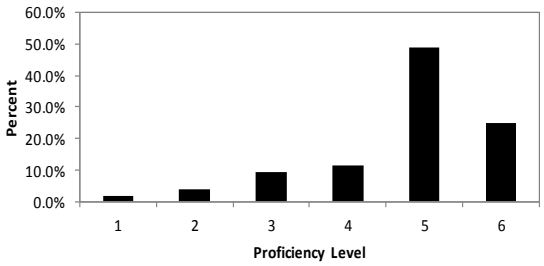


Table 8.3.1A
Scale Score Descriptive Statistics: List 2 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	80,833	121	397	328.17	31.03

Table 8.3.1B
Proficiency Level Distribution: List 2 S400 Paper

Level	Count	Percent
1	1,594	2.0%
2	3,169	3.9%
3	7,444	9.2%
4	9,117	11.3%
5	39,538	48.9%
6	19,971	24.7%
Total	80,833	100.0%

Table 8.3.1C
Conditional Standard Error of Measurement at Cut Scores: List 2 S400 Paper

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

Figure 8.3.1C
Test Characteristic Curve: List 2ABC S400 Paper

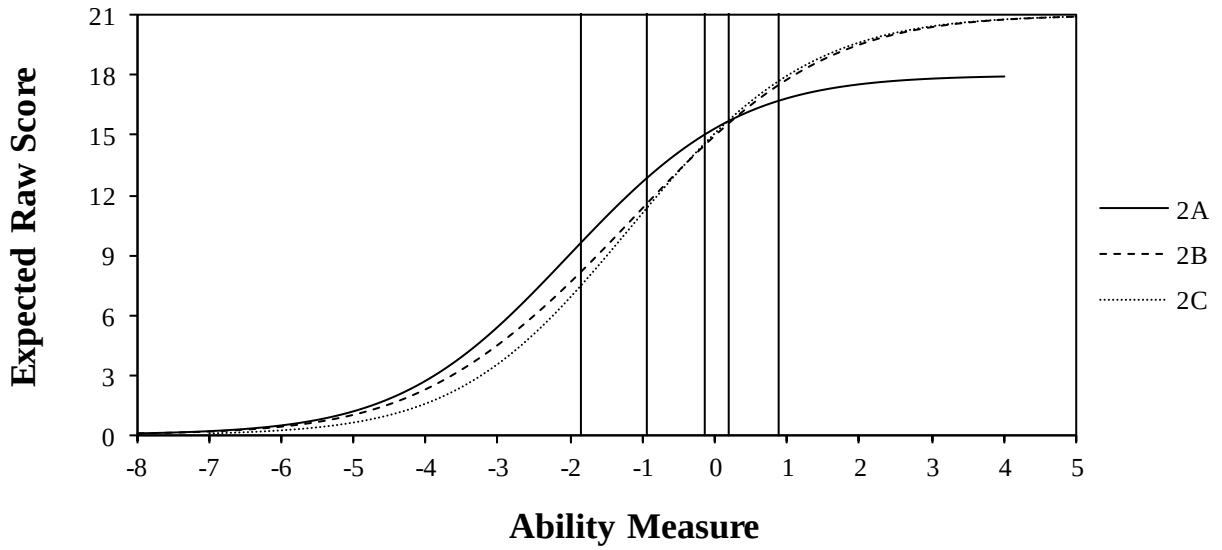


Figure 8.3.1D
Test Information Function: List 2ABC S400 Paper

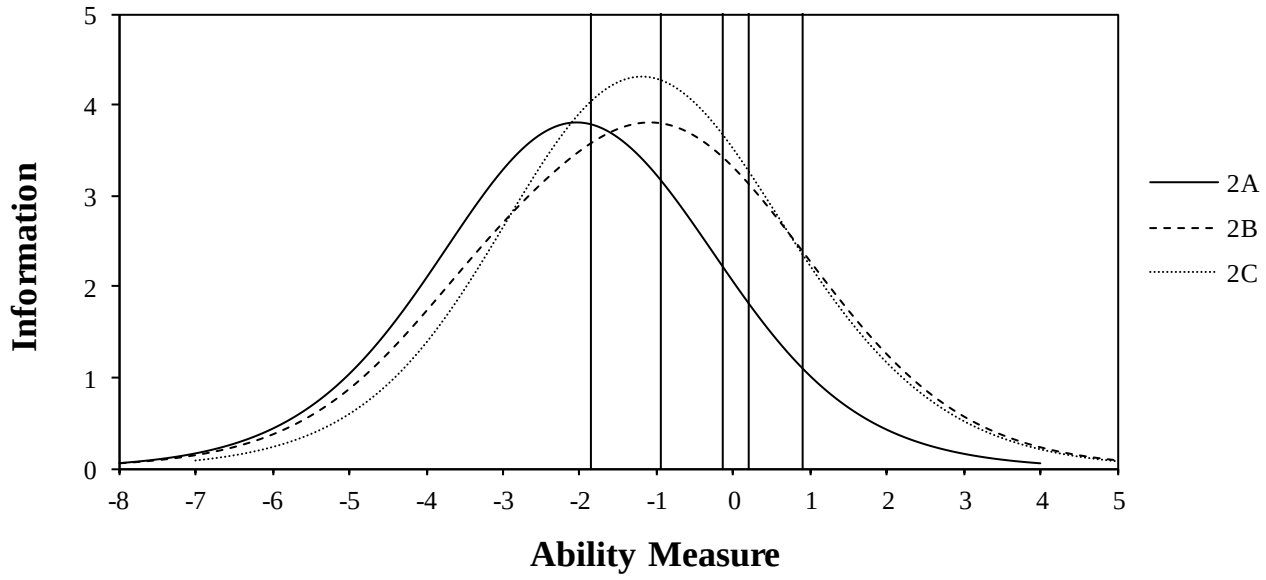


Table 8.3.1D
Weighted Reliability: List 2 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,176	0.815	0.617
B	33,278	0.626	
C	37,378	0.555	

Table 8.3.1E
Accuracy and Consistency of Classification Indices: List (Grade 2) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.513	0.400		0.174	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.845		0.469	
	2	0.371		0.224	
	3	0.309		0.193	
	4	0.207		0.155	
	5	0.628		0.567	
	6	0.638		0.467	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.982	0.000	0.017	0.980
	2/3	0.960	0.007	0.034	0.934
	3/4	0.881	0.069	0.050	0.808
	4/5	0.789	0.138	0.073	0.718
	5/6	0.811	0.114	0.076	0.734

8.3.2 Reading 2

Figure 8.3.2A
Scale Scores: Read 2 S400 Paper

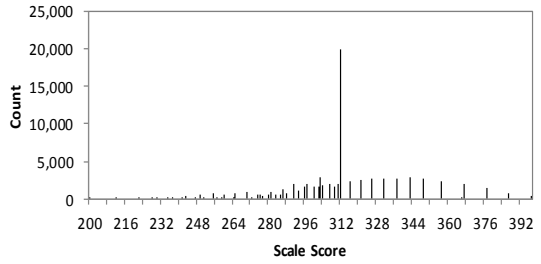


Figure 8.3.2B
Proficiency Level: Read 2 S400 Paper

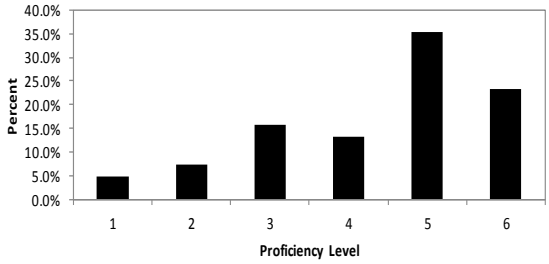


Table 8.3.2A
Scale Score Descriptive Statistics: Read 2 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	78,002	200	397	313.22	27.59

Table 8.3.2B
Proficiency Level Distribution: Read 2 S400 Paper

Level	Count	Percent
1	3,862	5.0%
2	5,711	7.3%
3	12,288	15.8%
4	10,426	13.4%
5	27,547	35.3%
6	18,168	23.3%
Total	78,002	100.0%

Table 8.3.2C
Conditional Standard Error of Measurement at Cut Scores: Read 2 S400 Paper

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

Figure 8.3.2C
Test Characteristic Curve: Read 2ABC S400 Paper

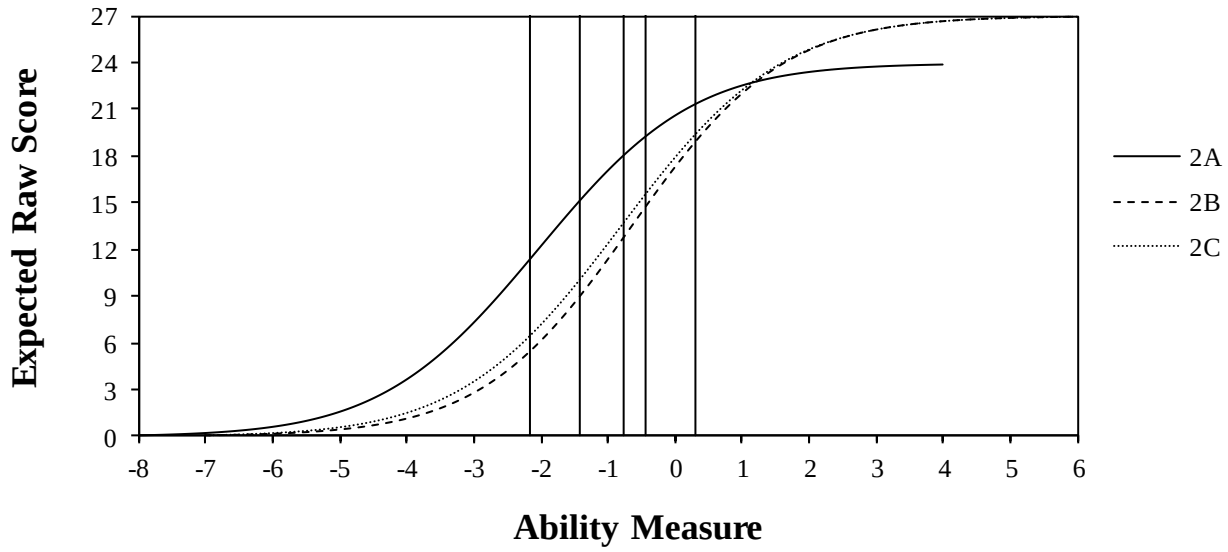


Figure 8.3.2D
Test Information Function: Read 2ABC S400 Paper

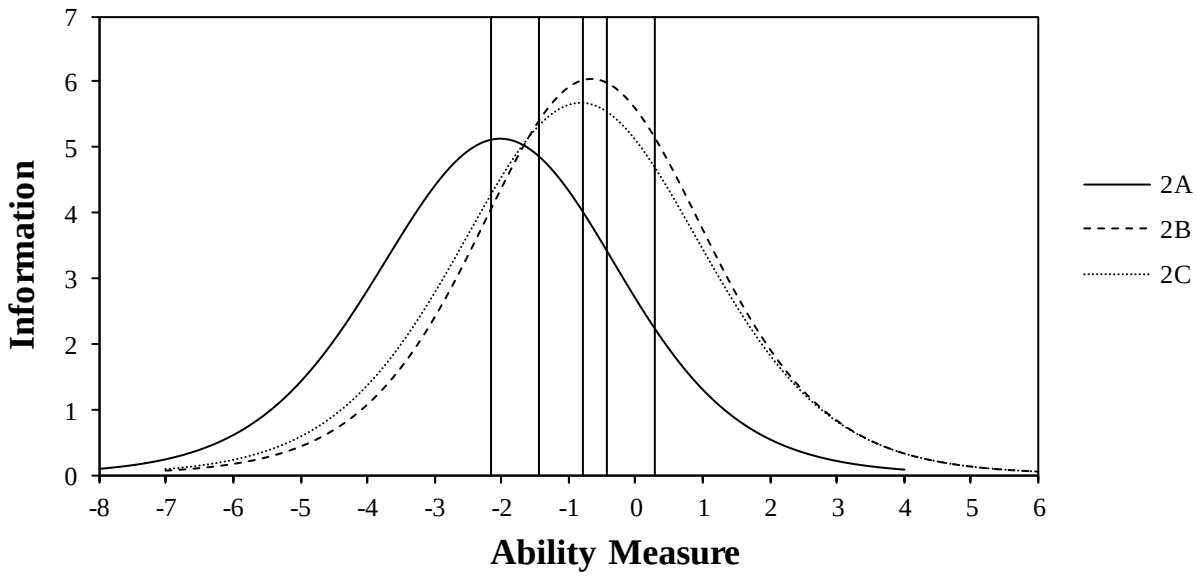


Table 8.3.2D
Weighted Reliability: Read 2 S400 Paper

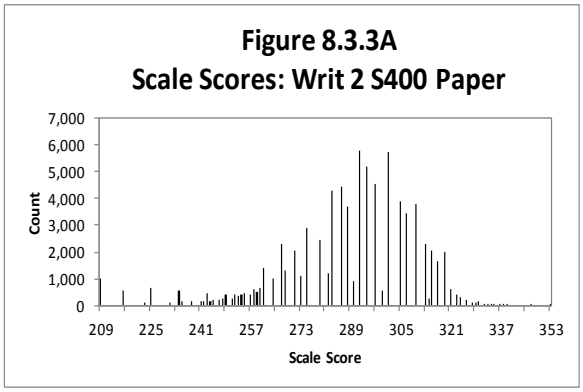
Tiers	No. of Students	Reliability	Weighted Reliability
A	10,006	0.828	0.811
B	31,865	0.834	
C	36,130	0.787	

Table 8.3.2E
Accuracy and Consistency of Classification Indices: Read (Grade 2) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.535	0.443		0.291	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.765		0.402	
	2	0.341		0.255	
	3	0.413		0.306	
	4	0.276		0.216	
	5	0.644		0.547	
	6	0.752		0.632	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.951	0.000	0.049	0.945
	2/3	0.925	0.034	0.041	0.881
	3/4	0.863	0.088	0.050	0.807
	4/5	0.831	0.109	0.060	0.779
	5/6	0.888	0.053	0.060	0.839

8.3.3 Writing 2

Figure 8.3.3.A Scale Scores: Writ 2 S400 Paper



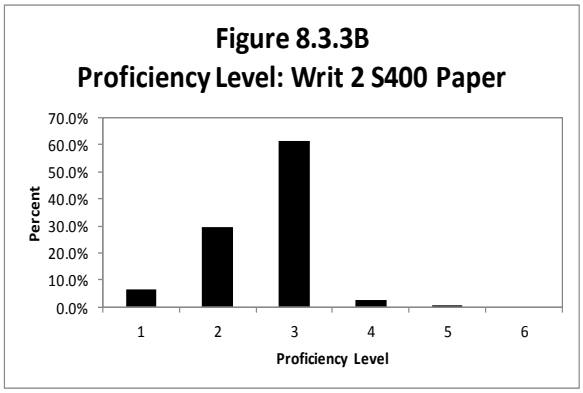


Table 8.3.3A

Scale Score Descriptive Statistics: Writ 2 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	81,795	209	353	287.59	22.98

Table 8.3.3B

Proficiency Level Distribution: Writ 2 S400 Paper

Level	Count	Percent
1	5,432	6.6%
2	24,148	29.5%
3	50,161	61.3%
4	2,053	2.5%
5	1	0.0%
6	0	0.0%
Total	81,795	100.0%

Table 8.3.3C

Conditional Standard Error of Measurement at Cut Scores: Writ 2 S400 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	2	251	6.84	7.46
2/3	2	285	8.40	8.40
3/4	2	320	7.77	7.46
4/5	2	348	6.84	6.53
5/6	2	373	6.53	7.15

Figure 8.3.3C

n/a

Figure 8.3.3D

n/a

Table 8.3.3D
Weighted Reliability: Writ 2 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,443	0.906	0.925
B/C	71,352	0.928	

Table 8.3.3E
Accuracy and Consistency of Classification Indices: Writ (Grade 2) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.887	0.845		0.701	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.843		0.752	
	2	0.869		0.802	
	3	0.899		0.881	
	4	-		0.091	
	5	N/A		N/A	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.980	0.011	0.009	0.971
	2/3	0.932	0.027	0.041	0.904
	3/4	0.975	0.025	0.000	0.970

4/5 cut could not be estimated because there was only one case after the cut and it was removed from the data

8.3.4 Speaking 2

Figure 8.3.4A
Scale Scores: Spek 2 S400 Paper

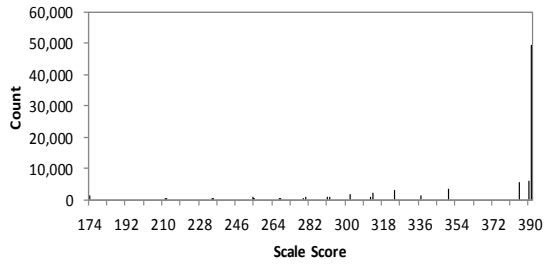


Figure 8.3.4B
Proficiency Level: Spek 2 S400 Paper

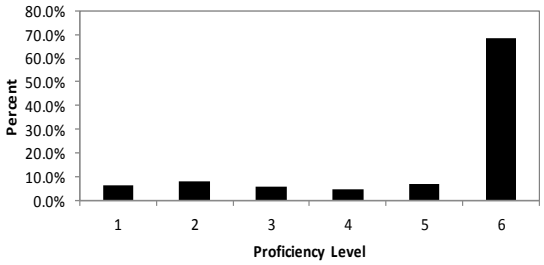


Table 8.3.4A
Scale Score Descriptive Statistics: Spek 2 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	81,205	174	391	368.72	44.49

Table 8.3.4B
Proficiency Level Distribution: Spek 2 S400 Paper

Level	Count	Percent
1	4,925	6.1%
2	6,589	8.1%
3	4,509	5.6%
4	3,649	4.5%
5	5,826	7.2%
6	55,707	68.6%
Total	81,205	100.0%

Table 8.3.4C
Conditional Standard Error of Measurement at Cut Scores: Spek 2 S400 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	2	286	21.43	21.43
2/3	2	322	20.41	20.41
3/4	2	345	19.39	19.39
4/5	2	368	19.39	19.39
5/6	2	386	19.39	19.39

Figure 8.3.4C

n/a

Figure 8.3.4D

n/a

Table 8.3.4D
Weighted Reliability: Spek 2 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,330	0.912	0.911
B/C	70,875	0.911	

Table 8.3.4E
Accuracy and Consistency of Classification Indices: Spek (Grade 2) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.751	0.632		0.368	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.730		0.624	
	2	0.593		0.489	
	3	0.364		0.289	
	4	0.304		0.208	
	5	0.185		0.100	
	6	0.931		0.905	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.970	0.018	0.012	0.958
	2/3	0.953	0.017	0.030	0.940
	3/4	0.955	0.013	0.032	0.936
	4/5	0.961	0.017	0.022	0.937
	5/6	0.878	0.076	0.045	0.774

8.3.5 Oral Language Composite 2

Figure 8.3.5A
Scale Scores: Oral 2 S400 Paper

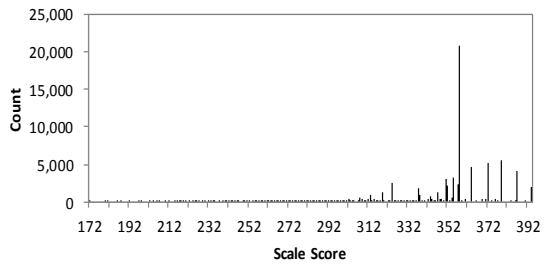


Figure 8.3.5B
Proficiency Level: Oral 2 S400 Paper

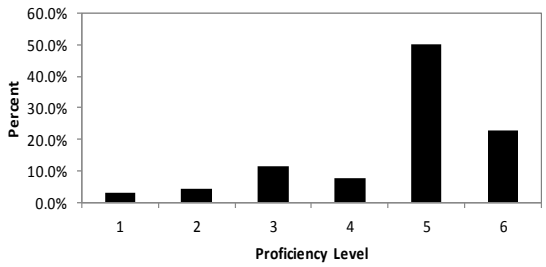


Table 8.3.5A
Scale Score Descriptive Statistics: Oral 2 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	80,320	172	394	348.88	32.78

Table 8.3.5B
Proficiency Level Distribution: Oral 2 S400 Paper

Level	Count	Percent
1	2,625	3.3%
2	3,462	4.3%
3	9,363	11.7%
4	6,137	7.6%
5	40,280	50.1%
6	18,453	23.0%
Total	80,320	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 2 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.50	957.194	0.617
Speaking	0.50	1923.229	0.911
Oral		1065.326	0.874

*Variances from students who had results in all four domains

Table 8.3.5E
Accuracy and Consistency of Classification Indices: Oral (Grade 2) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.612	0.484		0.261	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.865		0.762	
	2	0.554		0.419	
	3	0.719		0.578	
	4	0.304		0.191	
	5	0.632		0.620	
	6	-		0.351	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.989	0.004	0.006	0.985
	2/3	0.972	0.016	0.012	0.960
	3/4	0.947	0.013	0.039	0.928
	4/5	0.922	0.035	0.044	0.872
	5/6	0.770	0.230	0.000	0.688

8.3.6 Literacy Composite 2

Figure 8.3.6A
Scale Scores: Litr 2 S400 Paper

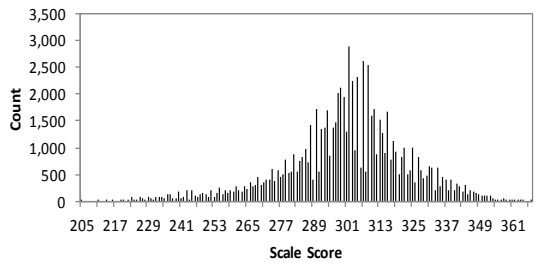


Figure 8.3.6B
Proficiency Level: Litr 2 S400 Paper

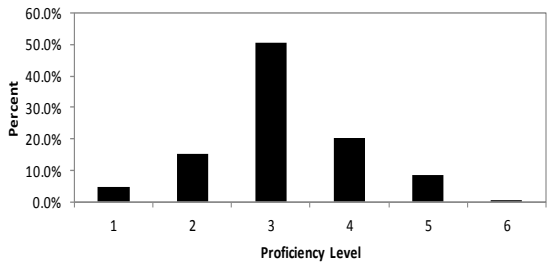


Table 8.3.6A
Scale Score Descriptive Statistics: Litr 2 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	78,002	205	368	300.87	22.52

Table 8.3.6B
Proficiency Level Distribution: Litr 2 S400 Paper

Level	Count	Percent
1	3,805	4.9%
2	11,896	15.3%
3	39,370	50.5%
4	15,730	20.2%
5	6,680	8.6%
6	521	0.7%
Total	78,002	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 2 S400 Paper

Component	Weight	Variance	Reliability
Reading	0.50	759.202	0.811
Writing	0.50	504.991	0.925
Literacy		502.920	0.910

*Variances from students who had results in all four domains

Table 8.3.6E

Accuracy and Consistency of Classification Indices: Litr (Grade 2) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.763	0.674		0.526	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.842		0.762	
	2	0.703		0.588	
	3	0.863		0.809	
	4	0.611		0.507	
	5	0.722		0.578	
	6	-		0.973	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.980	0.007	0.013	0.975
	2/3	0.933	0.038	0.029	0.904
	3/4	0.904	0.038	0.058	0.865
	4/5	0.947	0.040	0.013	0.930
	5/6	0.993	0.007	0.000	0.995

8.3.7 Comprehension Composite 2

Figure 8.3.7A
Scale Scores: Cphn 2 S400 Paper

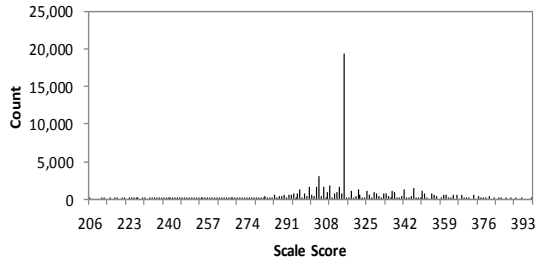


Figure 8.3.7B
Proficiency Level: Cphn 2 S400 Paper

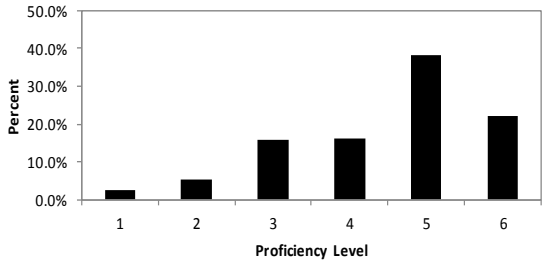


Table 8.3.7A
Scale Score Descriptive Statistics: Cphn 2 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	77,283	206	397	317.95	26.41

Table 8.3.7B
Proficiency Level Distribution: Cphn 2 S400 Paper

Level	Count	Percent
1	2,026	2.6%
2	4,061	5.3%
3	12,154	15.7%
4	12,395	16.0%
5	29,660	38.4%
6	16,987	22.0%
Total	77,283	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 2 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.30	957.194	0.617
Reading	0.70	759.202	0.811
Comprehension		697.564	0.852

*Variances from students who had results in all four domains

Table 8.3.7E

Accuracy and Consistency of Classification Indices: Cphn (Grade 2) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.616	0.511		0.360	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.887		0.602	
	2	0.428		0.327	
	3	0.546		0.409	
	4	0.381		0.294	
	5	0.703		0.611	
	6	0.770		0.652	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.976	0.000	0.024	0.976
	2/3	0.960	0.017	0.023	0.935
	3/4	0.896	0.060	0.044	0.850
	4/5	0.855	0.086	0.059	0.805
	5/6	0.901	0.048	0.051	0.856

8.3.8 Overall Composite 2

Figure 8.3.8A
Scale Scores: Over 2 S400 Paper

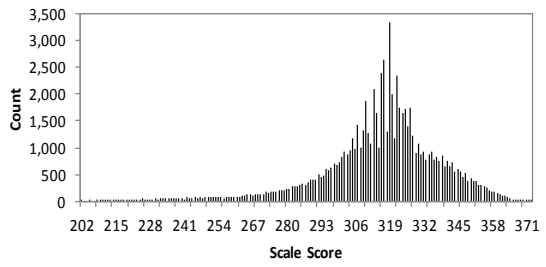


Figure 8.3.8B
Proficiency Level: Over 2 S400 Paper

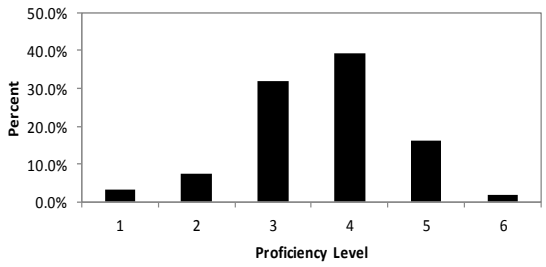


Table 8.3.8A
Scale Score Descriptive Statistics: Over 2 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	76,803	202	373	315.21	23.42

Table 8.3.8B
Proficiency Level Distribution: Over 2 S400 Paper

Level	Count	Percent
1	2,626	3.4%
2	5,822	7.6%
3	24,564	32.0%
4	30,129	39.2%
5	12,319	16.0%
6	1,343	1.7%
Total	76,803	100.0%

Table 8.3.8C

n/a

Figure 8.3.8C

n/a

Figure 8.3.8 D

n/a

Table 8.3.8D

Overall Composite Reliability: Over 2 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.15	957.194	0.617
Reading	0.35	759.202	0.811
Speaking	0.15	1923.229	0.911
Writing	0.35	504.991	0.925
Overall Composite		548.708	0.937

*Variances from students who had results in all four domains

Table 8.3.8E

Accuracy and Consistency of Classification Indices: Over (Grade 2) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.760	0.679		0.552	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.750		0.850	
	2	0.721		0.605	
	3	0.847		0.771	
	4	0.768		0.679	
	5	0.622		0.529	
	6	-		0.997	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.984	0.009	0.006	0.988
	2/3	0.964	0.024	0.012	0.956
	3/4	0.913	0.040	0.047	0.886
	4/5	0.893	0.049	0.058	0.857
	5/6	0.983	0.018	0.000	0.989

8.4 Grade: 3

8.4.1 Listening 3

Figure 8.4.1A
Scale Scores: List 3 S400 Paper

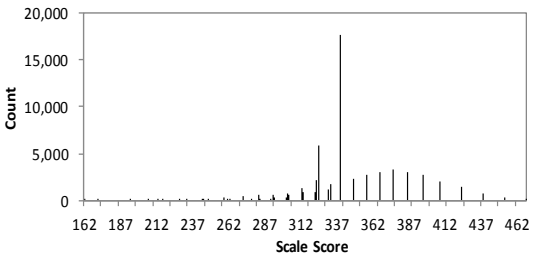


Figure 8.4.1B
Proficiency Level: List 3 S400 Paper

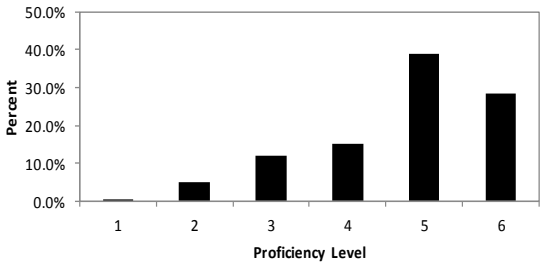


Table 8.4.1A
Scale Score Descriptive Statistics: List 3 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	58,389	162	469	347.71	34.84

Table 8.4.1B
Proficiency Level Distribution: List 3 S400 Paper

Level	Count	Percent
1	320	0.5%
2	2,889	4.9%
3	6,983	12.0%
4	8,858	15.2%
5	22,739	38.9%
6	16,600	28.4%
Total	58,389	100.0%

Table 8.4.1C
Conditional Standard Error of Measurement at Cut Scores: List 3 S400 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	3	255	22.54	22.17	27.05
2/3	3	295	19.54	19.54	20.66
3/4	3	325	19.54	19.16	18.41
4/5	3	340	n/a	19.54	18.03
5/6	3	367	n/a	n/a	18.41

Figure 8.4.1C
Test Characteristic Curve: List 3ABC S400 Paper

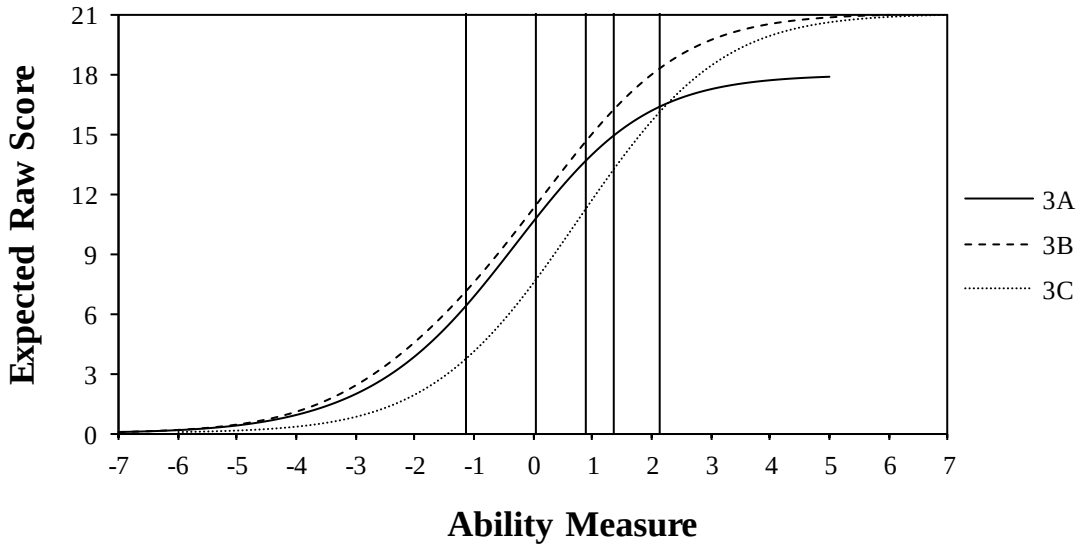


Figure 8.4.1D
Test Information Function: List 3ABC S400 Paper

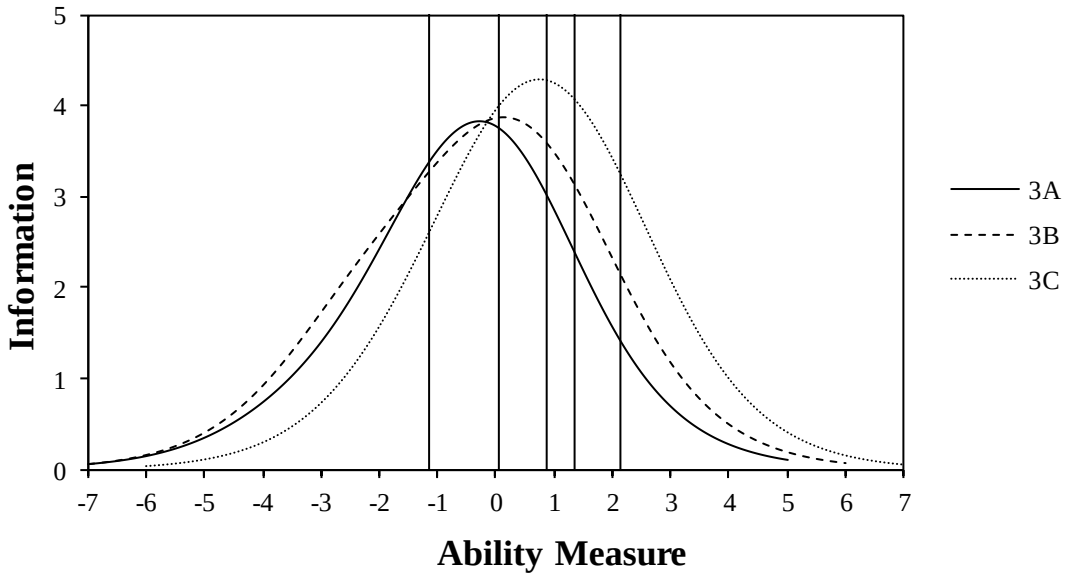


Table 8.4.1D
Weighted Reliability: List 3 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,382	0.752	0.640
B	21,586	0.618	
C	26,420	0.613	

Table 8.4.1E
Accuracy and Consistency of Classification Indices: List (Grade 3) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.487	0.406		0.209	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	-		0.022	
	2	-		0.148	
	3	0.299		0.221	
	4	0.247		0.205	
	5	0.562		0.474	
	6	0.764		0.616	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.995	0.000	0.005	0.994
	2/3	0.945	0.000	0.055	0.917
	3/4	0.824	0.104	0.072	0.754
	4/5	0.760	0.173	0.068	0.695
	5/6	0.847	0.094	0.059	0.785

8.4.2 Reading 3

Figure 8.4.2A
Scale Scores: Read 3 S400 Paper

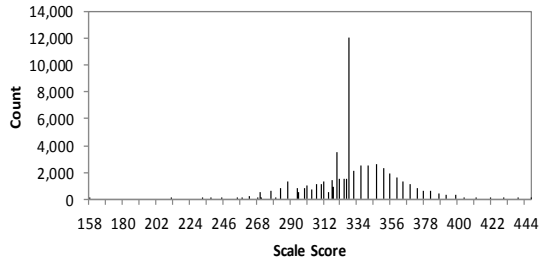


Figure 8.4.2B
Proficiency Level: Read 3 S400 Paper

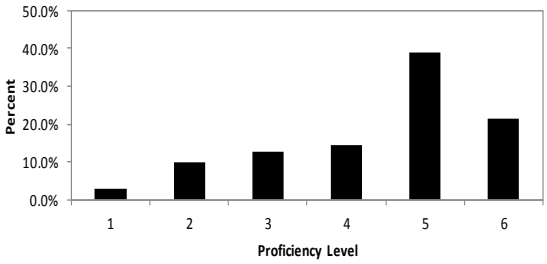


Table 8.4.2A
Scale Score Descriptive Statistics: Read 3 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	56,094	158	448	330.22	25.73

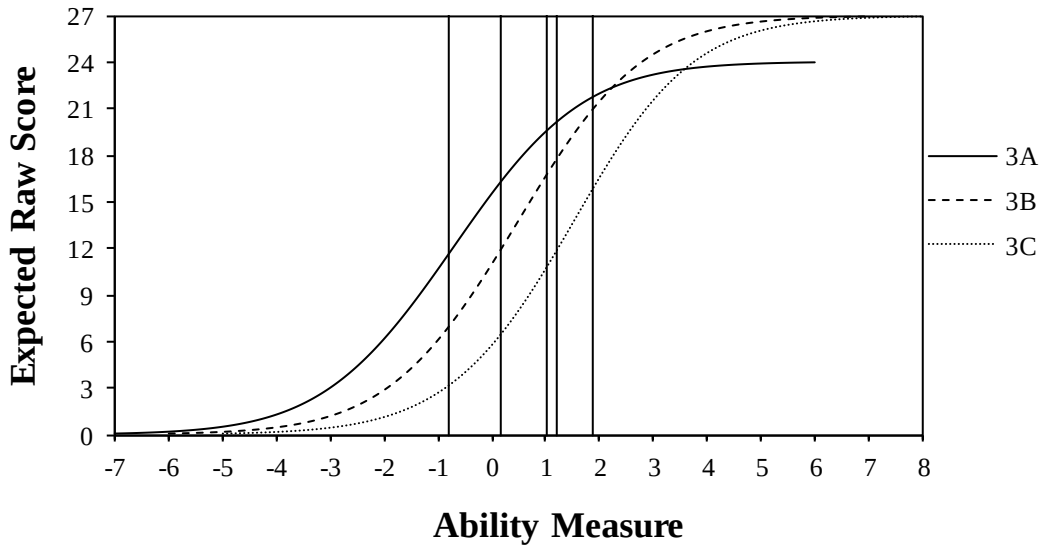
Table 8.4.2B
Proficiency Level Distribution: Read 3 S400 Paper

Level	Count	Percent
1	1,650	2.9%
2	5,463	9.7%
3	7,116	12.7%
4	8,034	14.3%
5	21,752	38.8%
6	12,079	21.5%
Total	56,094	100.0%

Table 8.4.2C
Conditional Standard Error of Measurement at Cut Scores: Read 3 S400 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	3	279	12.48	15.34	22.88
2/3	3	302	11.70	12.48	16.38
3/4	3	320	11.96	11.18	13.52
4/5	3	328	n/a	10.92	12.48
5/6	3	347	n/a	n/a	10.92

Figure 8.4.2C
Test Characteristic Curve: Read 3ABC S400 Paper



:

Figure 8.4.2D
Test Information Function: Read 3ABC S400 Paper

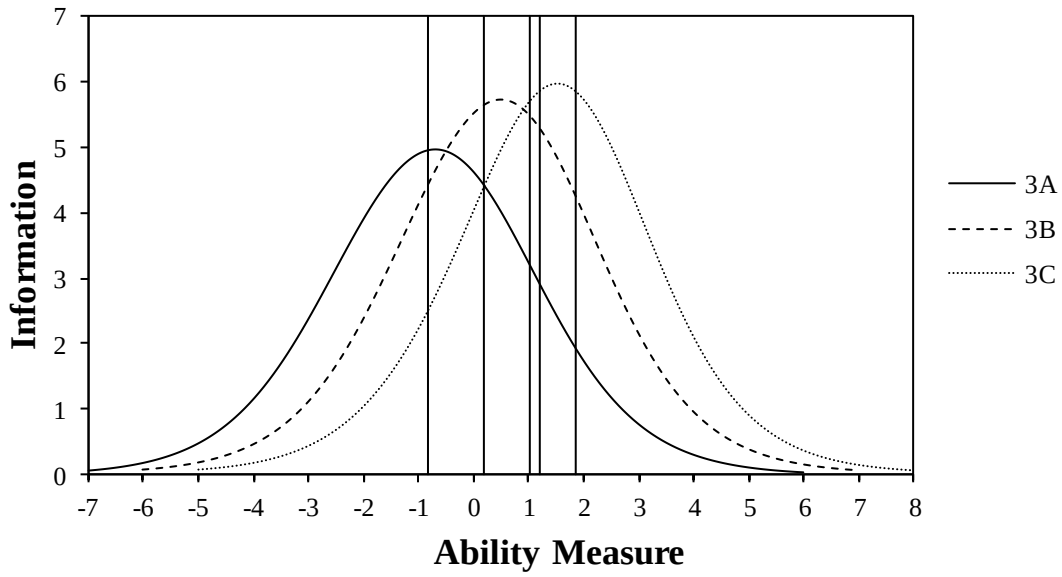


Table 8.4.2D
Weighted Reliability: Read 3 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,166	0.804	0.742
B	20,782	0.771	
C	25,145	0.693	

Table 8.4.2E
Accuracy and Consistency of Classification Indices: Read (Grade 3) S400 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)	
	0.507	0.412	0.244	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.800	0.308	
	2	0.485	0.344	
	3	0.308	0.226	
	4	0.263	0.207	
	5	0.623	0.534	
	6	0.704	0.560	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.971	0.000	0.029	0.965
	2/3	0.917	0.020	0.063	0.877
	3/4	0.846	0.103	0.051	0.780
	4/5	0.801	0.127	0.073	0.742
	5/6	0.873	0.064	0.064	0.815

8.4.3 Writing 3

Figure 8.4.3A
Scale Scores: Writ 3 S400 Paper

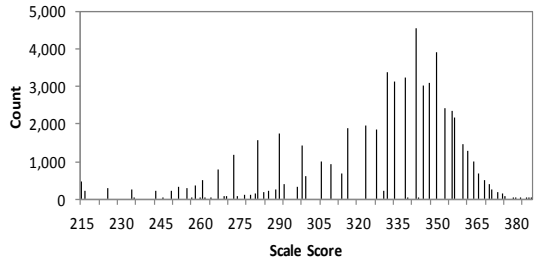


Figure 8.4.3B
Proficiency Level: Writ 3 S400 Paper

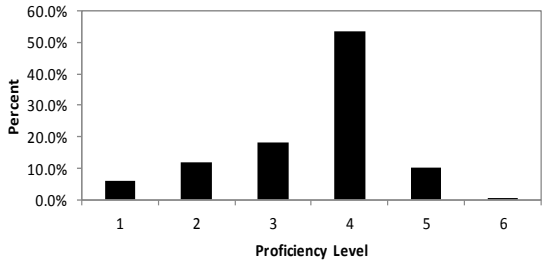


Table 8.4.3A
Scale Score Descriptive Statistics: Writ 3 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	58,960	215	386	327.39	33.14

Table 8.4.3B
Proficiency Level Distribution: Writ 3 S400 Paper

Level	Count	Percent
1	3,451	5.9%
2	7,014	11.9%
3	10,700	18.1%
4	31,624	53.6%
5	6,128	10.4%
6	43	0.1%
Total	58,960	100.0%

Table 8.4.3C
Conditional Standard Error of Measurement at Cut Scores: Writ 3 S400 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	3	264	10.88	7.77
2/3	3	297	11.19	7.77
3/4	3	330	11.51	8.09
4/5	3	360	10.57	7.46
5/6	3	384	9.33	6.53

Figure 8.4.3C

n/a

Figure 8.4.3D

n/a

Table 8.4.3D
Weighted Reliability: Writ 3 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,579	0.895	0.916
B/C	48,381	0.921	

Table 8.4.3E
Accuracy and Consistency of Classification Indices: Writ (Grade 3) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.767	0.669		0.486	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.852		0.761	
	2	0.788		0.690	
	3	0.696		0.561	
	4	0.776		0.759	
	5	-		0.228	
	6	-		-	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.982	0.009	0.009	0.974
	2/3	0.963	0.014	0.023	0.948
	3/4	0.927	0.032	0.041	0.893
	4/5	0.895	0.105	0.000	0.850
	5/6	0.999	0.001	0.000	0.999

8.4.4 Speaking 3

Figure 8.4.4A
Scale Scores: Spek 3 S400 Paper

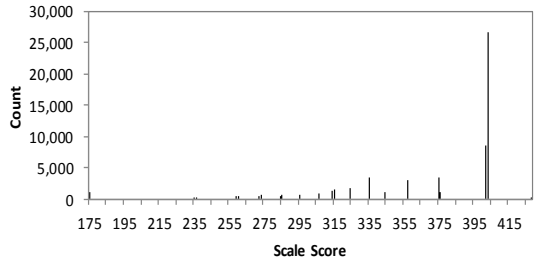


Figure 8.4.4B
Proficiency Level: Spek 3 S400 Paper

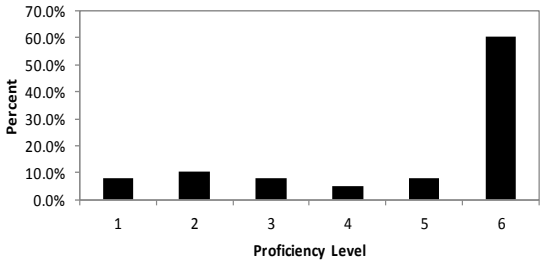


Table 8.4.4A
Scale Score Descriptive Statistics: Spek 3 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	58,487	175	428	370.79	49.99

Table 8.4.4B
Proficiency Level Distribution: Spek 3 S400 Paper

Level	Count	Percent
1	4,723	8.1%
2	6,146	10.5%
3	4,665	8.0%
4	3,005	5.1%
5	4,733	8.1%
6	35,215	60.2%
Total	58,487	100.0%

Table 8.4.4C
Conditional Standard Error of Measurement at Cut Scores: Spek 3 S400 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	3	293	18.88	18.88
2/3	3	326	22.45	22.45
3/4	3	346	23.98	23.98
4/5	3	369	23.98	23.98
5/6	3	389	24.49	24.49

Figure 8.4.4C

n/a

Figure 8.4.4D

n/a

Table 8.4.4D
Weighted Reliability: Spek 3 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,482	0.898	0.909
B/C	48,005	0.911	

Table 8.4.4E
Accuracy and Consistency of Classification Indices: Spek (Grade 3) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.741	0.605		0.403	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.794		0.700	
	2	0.654		0.551	
	3	0.463		0.374	
	4	0.291		0.176	
	5	0.231		0.123	
	6	0.902		0.880	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.971	0.018	0.011	0.958
	2/3	0.946	0.020	0.033	0.931
	3/4	0.950	0.010	0.041	0.932
	4/5	0.956	0.014	0.030	0.920
	5/6	0.885	0.056	0.059	0.774

8.4.5 Oral Language Composite 3

Figure 8.4.5A
Scale Scores: Oral 3 S400 Paper

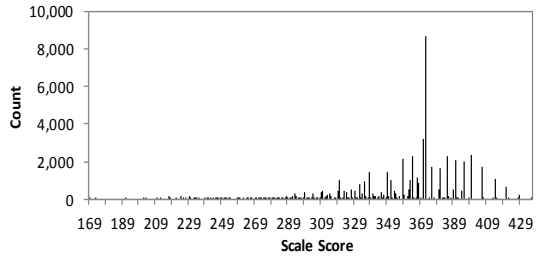


Figure 8.4.5B
Proficiency Level: Oral 3 S400 Paper

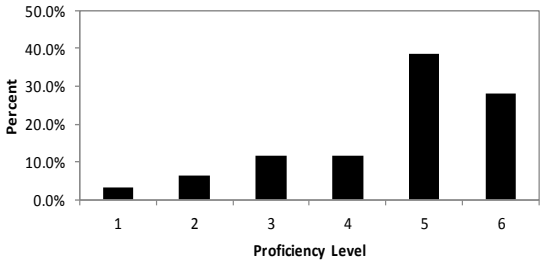


Table 8.4.5A

Scale Score Descriptive Statistics: Oral 3 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	57,989	169	436	359.70	36.57

Table 8.4.5B

Proficiency Level Distribution: Oral 3 S400 Paper

Level	Count	Percent
1	1,881	3.2%
2	3,751	6.5%
3	6,806	11.7%
4	6,852	11.8%
5	22,478	38.8%
6	16,221	28.0%
Total	57,989	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 3 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.50	1214.722	0.640
Speaking	0.50	2460.427	0.909
Oral		1337.093	0.876

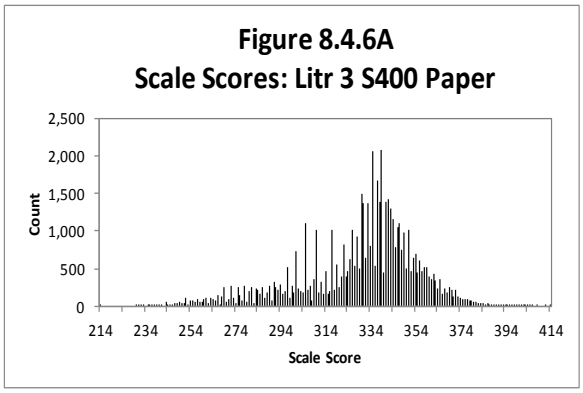
*Variances from students who had results in all four domains

Table 8.4.5E

Accuracy and Consistency of Classification Indices: Oral (Grade 3) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.623	0.516		0.356	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.847		0.724	
	2	0.631		0.491	
	3	0.602		0.456	
	4	0.389		0.284	
	5	0.691		0.574	
	6	0.643		0.566	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.987	0.004	0.008	0.982
	2/3	0.965	0.016	0.019	0.950
	3/4	0.935	0.025	0.040	0.906
	4/5	0.897	0.054	0.049	0.853
	5/6	0.823	0.049	0.128	0.767

8.4.6 Literacy Composite 3



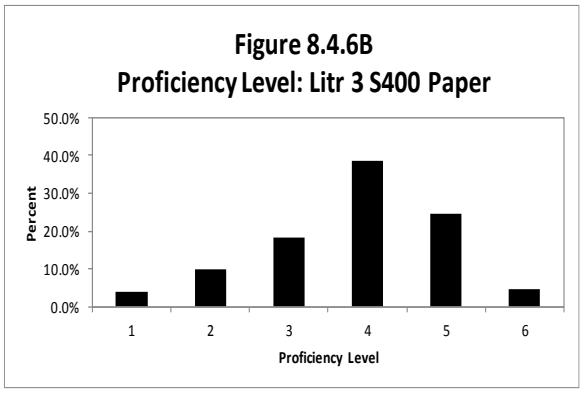


Table 8.4.6A

Scale Score Descriptive Statistics: Litr 3 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	56,094	214	414	329.16	26.47

Table 8.4.6B

Proficiency Level Distribution: Litr 3 S400 Paper

Level	Count	Percent
1	2,223	4.0%
2	5,607	10.0%
3	10,194	18.2%
4	21,634	38.6%
5	13,875	24.7%
6	2,561	4.6%
Total	56,094	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 3 S400 Paper

Component	Weight	Variance	Reliability
Reading	0.50	661.147	0.742
Writing	0.50	1062.725	0.916
Literacy		695.076	0.907

*Variances from students who had results in all four domains

Table 8.4.6E

Accuracy and Consistency of Classification Indices:Litr (Grade 3) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.697	0.592		0.457	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.874		0.774	
	2	0.760		0.654	
	3	0.649		0.527	
	4	0.767		0.667	
	5	0.618		0.546	
	6	-		0.270	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.986	0.004	0.009	0.981
	2/3	0.964	0.014	0.022	0.949
	3/4	0.914	0.049	0.036	0.877
	4/5	0.877	0.042	0.081	0.829
	5/6	0.954	0.046	0.000	0.942

8.4.7 Comprehension Composite 3

Figure 8.4.7A
Scale Scores: Cphn 3 S400 Paper

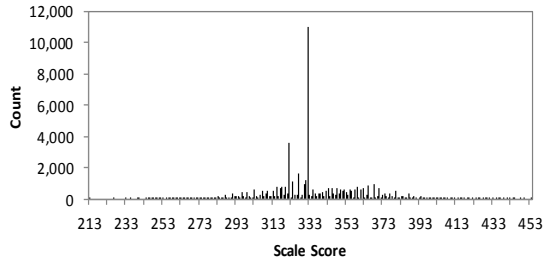


Figure 8.4.7B
Proficiency Level: Cphn 3 S400 Paper

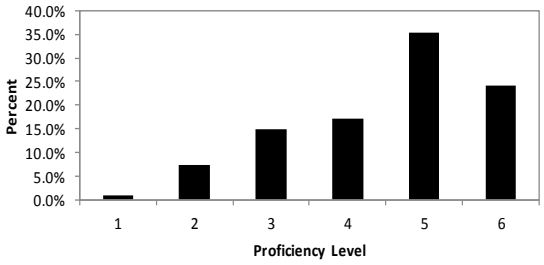


Table 8.4.7A
Scale Score Descriptive Statistics: Cphn 3 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	55,691	213	454	335.67	26.69

Table 8.4.7B
Proficiency Level Distribution: Cphn 3 S400 Paper

Level	Count	Percent
1	469	0.8%
2	4,130	7.4%
3	8,314	14.9%
4	9,625	17.3%
5	19,759	35.5%
6	13,394	24.1%
Total	55,691	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 3 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.30	1214.722	0.640
Reading	0.70	661.147	0.742
Comprehension		712.766	0.828

*Variances from students who had results in all four domains

Table 8.4.7E
Accuracy and Consistency of Classification Indices: Cphn (Grade 3) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.598	0.497		0.344	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	-		0.254	
	2	0.635		0.464	
	3	0.460		0.350	
	4	0.386		0.300	
	5	0.648		0.555	
	6	0.787		0.670	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.992	0.000	0.008	0.991
	2/3	0.946	0.011	0.043	0.925
	3/4	0.884	0.074	0.043	0.832
	4/5	0.843	0.089	0.068	0.792
	5/6	0.895	0.055	0.050	0.849

8.4.8 Overall Composite 3

Figure 8.4.8A
Scale Scores: Over 3 S400 Paper

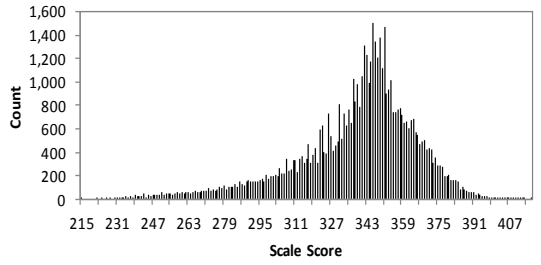


Figure 8.4.8B
Proficiency Level: Over 3 S400 Paper

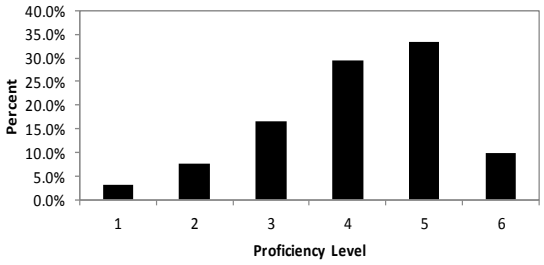


Table 8.4.8A

Scale Score Descriptive Statistics: Over 3 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	55,326	215	417	338.25	27.60

Table 8.4.8B

Proficiency Level Distribution: Over 3 S400 Paper

Level	Count	Percent
1	1,733	3.1%
2	4,162	7.5%
3	9,178	16.6%
4	16,354	29.6%
5	18,466	33.4%
6	5,433	9.8%
Total	55,326	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 3 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.15	1214.722	0.640
Reading	0.35	661.147	0.742
Speaking	0.15	2460.427	0.909
Writing	0.35	1062.725	0.916
Overall Composite		761.687	0.939

*Variances from students who had results in all four domains

Table 8.4.8E
Accuracy and Consistency of Classification Indices: Over (Grade 3) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.723	0.639		0.520	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.907		0.837	
	2	0.777		0.681	
	3	0.741		0.633	
	4	0.773		0.673	
	5	0.669		0.635	
	6	0.621		0.431	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.991	0.003	0.006	0.988
	2/3	0.976	0.011	0.013	0.966
	3/4	0.940	0.032	0.028	0.914
	4/5	0.910	0.033	0.057	0.875
	5/6	0.905	0.089	0.006	0.891

8.5 Grades: 4–5

8.5.1 Listening 4–5

Figure 8.5.1A
Scale Scores: List 4-5 S400 Paper

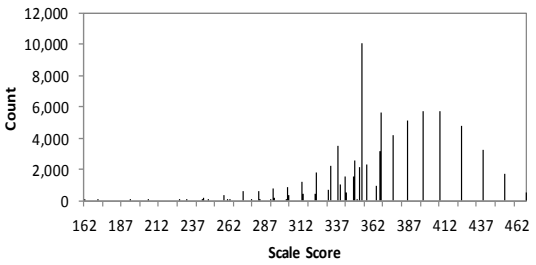


Figure 8.5.1B
Proficiency Level: List 4-5 S400 Paper

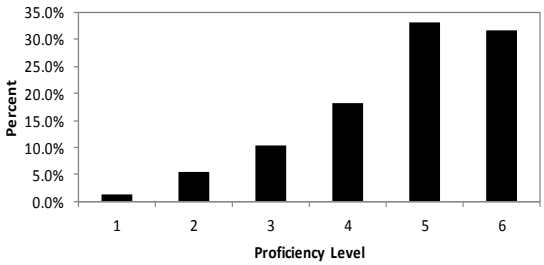


Table 8.5.1A
Scale Score Descriptive Statistics: List 4-5 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	43,996	162	469	364.47	38.07
5	33,597	162	469	375.22	40.57
Total	77,593	162	469	369.13	39.54

Table 8.5.1B
Proficiency Level Distribution: List 4-5 S400 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	421	1.0%	574	1.7%	995	1.3%
2	2,209	5.0%	2,063	6.1%	4,272	5.5%
3	4,212	9.6%	3,837	11.4%	8,049	10.4%
4	7,752	17.6%	6,319	18.8%	14,071	18.1%
5	15,951	36.3%	9,675	28.8%	25,626	33.0%
6	13,451	30.6%	11,129	33.1%	24,580	31.7%
Total	43,996	100.0%	33,597	100.0%	77,593	100.0%

Table 8.5.1C

Conditional Standard Error of Measurement at Cut Scores: List 4-5 S400 Paper

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

Figure 8.5.1C
Test Characteristic Curve: List 4-5ABC S400 Paper

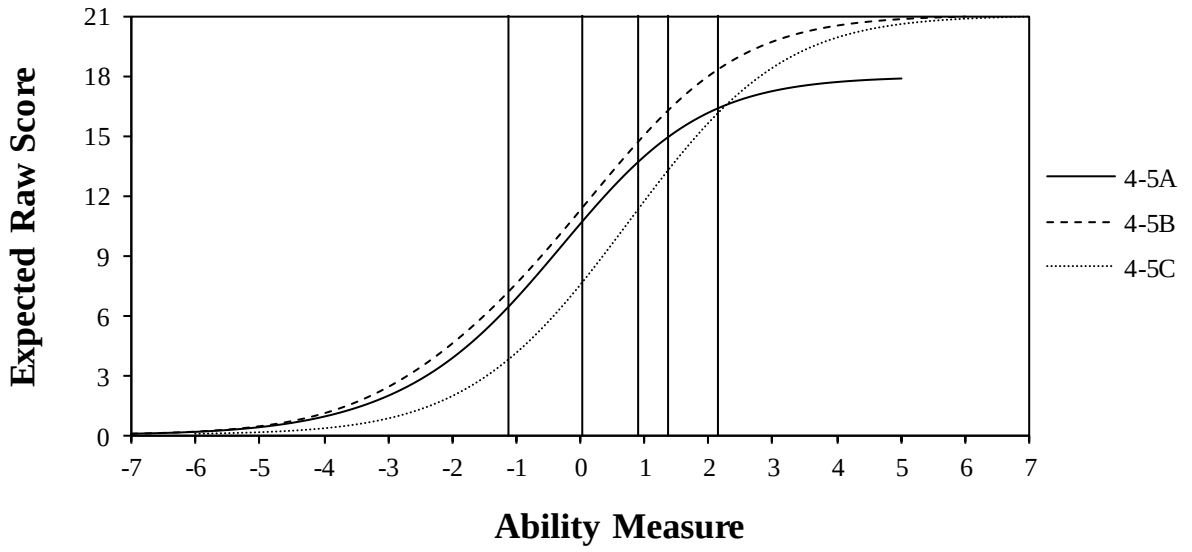


Figure 8.5.1D
Test Information Function: List 4-5ABC S400 Paper

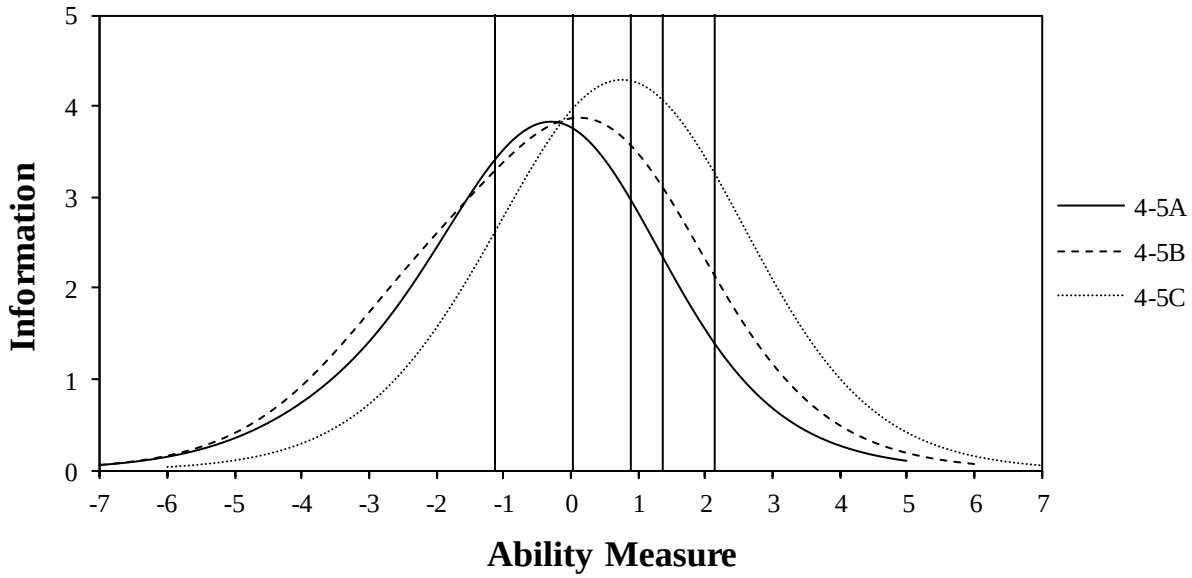


Table 8.5.1D
Weighted Reliability: List 4-5 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	13,281	0.768	0.659
B	23,672	0.668	
C	40,639	0.618	

Table 8.5.1Ei
Accuracy and Consistency of Classification Indices: List (Grade 4) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.502	0.415		0.220	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	-		0.085	
	2	-		0.204	
	3	0.270		0.194	
	4	0.311		0.248	
	5	0.536		0.455	
	6	0.748		0.613	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.990	0.000	0.010	0.989
	2/3	0.940	0.000	0.060	0.915
	3/4	0.857	0.088	0.055	0.788
	4/5	0.783	0.135	0.083	0.719
	5/6	0.835	0.093	0.071	0.769

Table 8.5.1Eii
Accuracy and Consistency of Classification Indices: List (Grade 5) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.499	0.411		0.225	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	-		0.260	
	2	0.432		0.265	
	3	0.316		0.225	
	4	0.347		0.272	
	5	0.443		0.371	
	6	0.769		0.632	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.983	0.000	0.017	0.980
	2/3	0.938	0.011	0.050	0.905
	3/4	0.859	0.077	0.063	0.795
	4/5	0.799	0.101	0.100	0.735
	5/6	0.824	0.109	0.067	0.757

8.5.2 Reading 4–5

Figure 8.5.2A
Scale Scores: Read 4-5 S400 Paper

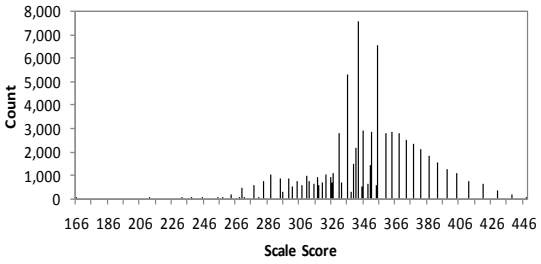


Figure 8.5.2B
Proficiency Level: Read 4-5 S400 Paper

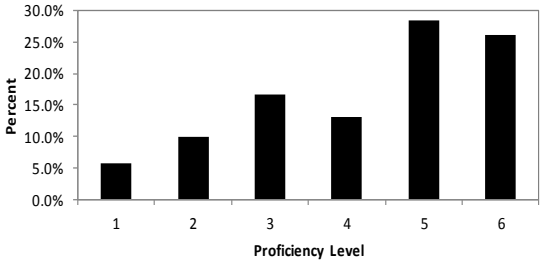


Table 8.5.2A
Scale Score Descriptive Statistics: Read 4-5 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	42,232	166	448	343.57	28.20
5	32,507	212	448	352.64	31.31
Total	74,739	166	448	347.51	29.94

Table 8.5.2B
Proficiency Level Distribution: Read 4-5 S400 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,963	4.6%	2,372	7.3%	4,335	5.8%
2	3,796	9.0%	3,662	11.3%	7,458	10.0%
3	6,068	14.4%	6,431	19.8%	12,499	16.7%
4	6,556	15.5%	3,214	9.9%	9,770	13.1%
5	12,549	29.7%	8,646	26.6%	21,195	28.4%
6	11,300	26.8%	8,182	25.2%	19,482	26.1%
Total	42,232	100.0%	32,507	100.0%	74,739	100.0%

Table 8.5.2C
Conditional Standard Error of Measurement at Cut Scores: Read 4-5 S400 Paper

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

Figure 8.5.2C
Test Characteristic Curve: Read 4-5ABC S400 Paper

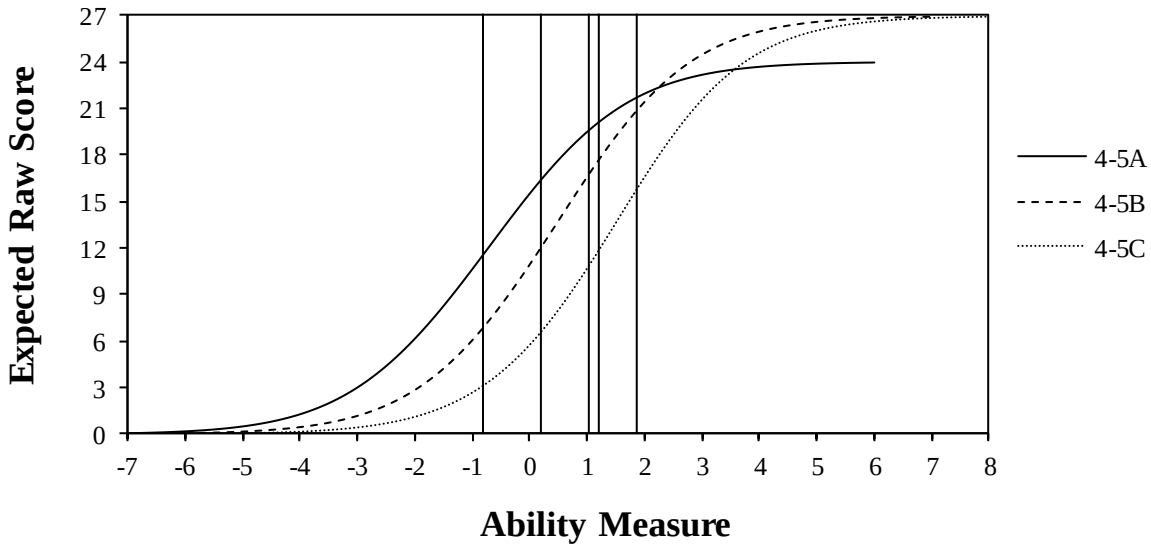


Figure 8.5.2D
Test Information Function: Read 4-5ABC S400 Paper

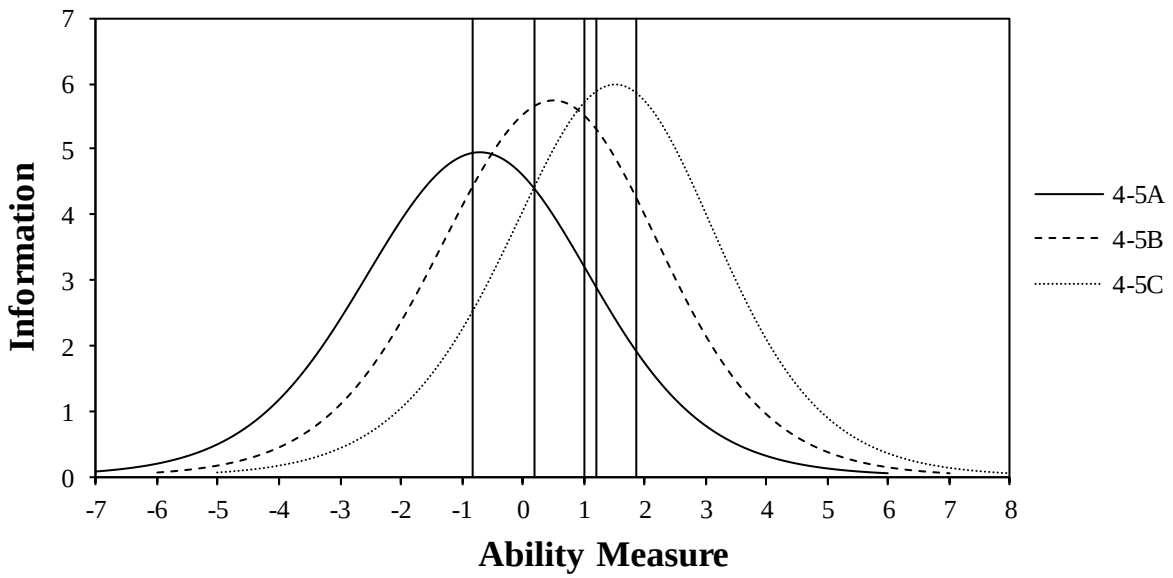


Table 8.5.2D
Weighted Reliability: Read 4-5 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	13,037	0.848	0.797
B	22,775	0.809	
C	38,926	0.773	

Table 8.5.2Ei
Accuracy and Consistency of Classification Indices: Read (Grade 4) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.529	0.438		0.289	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.825		0.560	
	2	0.464		0.338	
	3	0.370		0.275	
	4	0.311		0.247	
	5	0.531		0.446	
	6	0.759		0.638	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.963	0.003	0.034	0.957
	2/3	0.933	0.025	0.042	0.894
	3/4	0.850	0.103	0.047	0.795
	4/5	0.821	0.098	0.082	0.768
	5/6	0.870	0.066	0.064	0.815

Table 8.5.2Eii
Accuracy and Consistency of Classification Indices: Read (Grade 5) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.530	0.437		0.294	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.829		0.622	
	2	0.464		0.336	
	3	0.439		0.341	
	4	0.197		0.155	
	5	0.495		0.408	
	6	0.738		0.613	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.955	0.007	0.037	0.941
	2/3	0.915	0.037	0.048	0.870
	3/4	0.837	0.100	0.062	0.781
	4/5	0.824	0.095	0.081	0.770
	5/6	0.868	0.067	0.066	0.813

8.5.3 Writing 4–5

Figure 8.5.3A
Scale Scores: Writ 4-5 S400 Paper

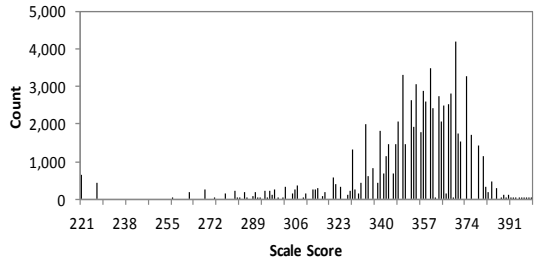


Figure 8.5.3B
Proficiency Level: Writ 4-5 S400 Paper

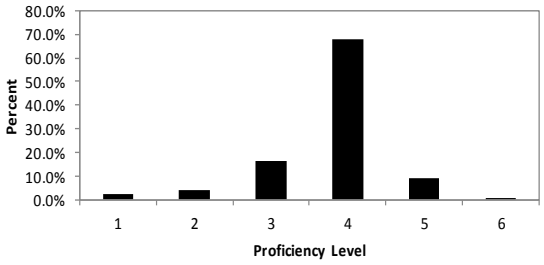


Table 8.5.3A
Scale Score Descriptive Statistics: Writ 4-5 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	44,328	221	398	350.66	24.98
5	33,801	227	400	353.81	25.33
Total	78,129	221	400	352.02	25.18

Table 8.5.3B
Proficiency Level Distribution: Writ 4-5 S400 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	898	2.0%	1,037	3.1%	1,935	2.5%
2	1,798	4.1%	1,369	4.1%	3,167	4.1%
3	5,399	12.2%	7,405	21.9%	12,804	16.4%
4	30,812	69.5%	22,123	65.5%	52,935	67.8%
5	5,401	12.2%	1,867	5.5%	7,268	9.3%
6	20	0.0%	0	0.0%	20	0.0%
Total	44,328	100.0%	33,801	100.0%	78,129	100.0%

Table 8.5.3C

Conditional Standard Error of Measurement at Cut Scores: Writ 4-5 S400 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	4	275	9.02	6.53
	5	287	9.64	6.84
2/3	4	308	11.82	8.40
	5	319	11.82	8.40
3/4	4	340	11.51	8.09
	5	350	11.19	7.77
4/5	4	371	9.95	7.15
	5	381	9.64	6.84
5/6	4	394	9.02	6.53
	5	403	9.02	6.53

Figure 8.5.3C

n/a

Figure 8.5.3D

n/a

Table 8.5.3D
Weighted Reliability: Writ 4-5 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	13,495	0.882	0.902
B/C	64,634	0.906	

Table 8.5.3Ei
Accuracy and Consistency of Classification Indices: Writ (Grade 4) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.818	0.753		0.443	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.874		0.796	
	2	0.779		0.653	
	3	0.860		0.706	
	4	0.813		0.810	
	5	-		0.168	
	6	-		-	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.995	0.003	0.002	0.993
	2/3	0.987	0.006	0.007	0.980
	3/4	0.964	0.006	0.030	0.942
	4/5	0.878	0.122	0.000	0.838
	5/6	1.000	0.000	0.000	1.000

Table 8.5.3Eii
Accuracy and Consistency of Classification Indices: Writ (Grade 5) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.829	0.776		0.530	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.905		0.847	
	2	0.669		0.538	
	3	0.854		0.686	
	4	0.830		0.819	
	5	-		0.086	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.994	0.003	0.004	0.991
	2/3	0.981	0.011	0.007	0.972
	3/4	0.909	0.015	0.075	0.873
	4/5	0.945	0.055	0.000	0.938

8.5.4 Speaking 4–5

Figure 8.5.4A
Scale Scores: Spek 4-5 S400 Paper

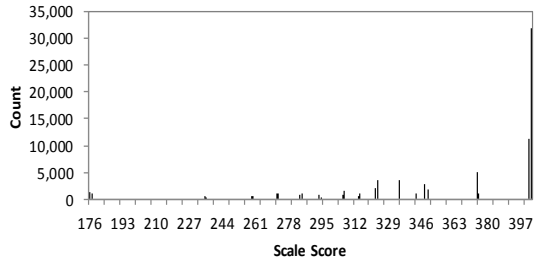


Figure 8.5.4B
Proficiency Level: Spek 4-5 S400 Paper

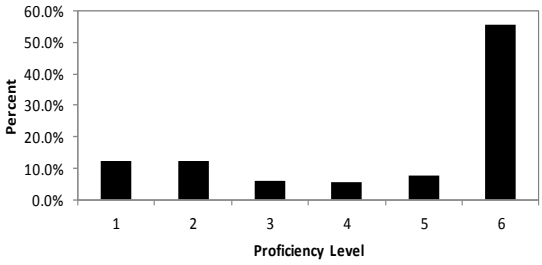


Table 8.5.4A
Scale Score Descriptive Statistics: Spek 4-5 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	44,032	176	403	363.11	54.20
5	33,579	177	403	365.40	55.92
Total	77,611	176	403	364.10	54.96

Table 8.5.4B
Proficiency Level Distribution: Spek 4-5 S400 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	5,287	12.0%	4,174	12.4%	9,461	12.2%
2	5,853	13.3%	3,806	11.3%	9,659	12.4%
3	2,873	6.5%	1,915	5.7%	4,788	6.2%
4	2,746	6.2%	1,753	5.2%	4,499	5.8%
5	3,646	8.3%	2,487	7.4%	6,133	7.9%
6	23,627	53.7%	19,444	57.9%	43,071	55.5%
Total	44,032	100.0%	33,579	100.0%	77,611	100.0%

Table 8.5.4C

Conditional Standard Error of Measurement at Cut Scores: Spek 4-5 S400 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	4	299	19.39	19.39
	5	305	19.90	19.90
2/3	4	329	22.45	22.45
	5	333	22.96	22.96
3/4	4	348	23.98	23.98
	5	350	23.98	23.98
4/5	4	371	23.98	23.98
	5	374	23.98	23.98
5/6	4	391	25.00	25.00
	5	394	25.00	25.00

Figure 8.5.4C

n/a

Figure 8.5.4D

n/a

Table 8.5.4D
Weighted Reliability: Spek 4-5 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	13,384	0.902	0.903
B/C	64,227	0.903	

Table 8.5.4Ei
Accuracy and Consistency of Classification Indices: Spek (Grade 4) S400 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)	
	0.696	0.633	0.465	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.747	0.654	
	2	0.585	0.483	
	3	0.277	0.217	
	4	0.279	0.193	
	5	0.292	0.200	
	6	0.939	0.900	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.945	0.033	0.022	0.924
	2/3	0.923	0.024	0.053	0.901
	3/4	0.932	0.021	0.047	0.903
	4/5	0.941	0.030	0.030	0.906
	5/6	0.897	0.073	0.030	0.857

Table 8.5.4Eii
Accuracy and Consistency of Classification Indices: Spek (Grade 5) S400 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)	
	0.699	0.623	0.432	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.758	0.676	
	2	0.542	0.446	
	3	0.263	0.206	
	4	0.240	0.161	
	5	0.223	0.133	
	6	0.935	0.896	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.946	0.033	0.021	0.927
	2/3	0.926	0.023	0.051	0.907
	3/4	0.934	0.017	0.049	0.908
	4/5	0.944	0.026	0.031	0.905
	5/6	0.881	0.085	0.035	0.821

8.5.5 Oral Language Composite 4–5

Figure 8.5.5A
Scale Scores: Oral 4-5 S400 Paper

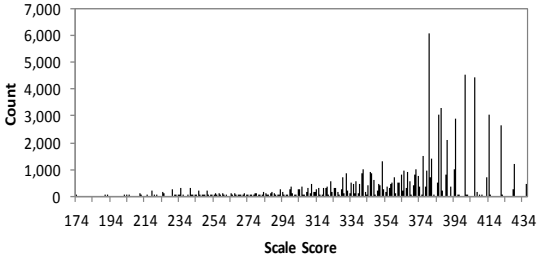


Figure 8.5.5B
Proficiency Level: Oral 4-5 S400 Paper

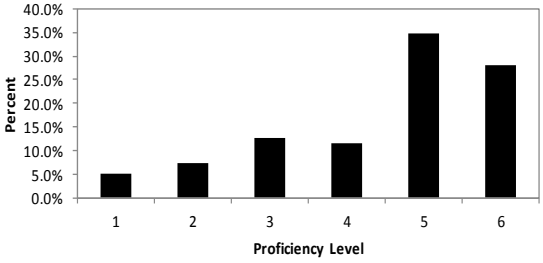


Table 8.5.5A
Scale Score Descriptive Statistics: Oral 4-5 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	43,724	191	436	364.19	40.53
5	33,404	174	436	370.69	43.18
Total	77,128	174	436	367.01	41.83

Table 8.5.5B
Proficiency Level Distribution: Oral 4-5 S400 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,911	4.4%	2,021	6.1%	3,932	5.1%
2	3,286	7.5%	2,440	7.3%	5,726	7.4%
3	6,061	13.9%	3,788	11.3%	9,849	12.8%
4	5,265	12.0%	3,738	11.2%	9,003	11.7%
5	14,750	33.7%	12,178	36.5%	26,928	34.9%
6	12,451	28.5%	9,239	27.7%	21,690	28.1%
Total	43,724	100.0%	33,404	100.0%	77,128	100.0%

Table 8.5.55C

n/a

Figure 8.5.55C

n/a

Figure 8.5.55D

n/a

Table 8.5.5D
Oral Composite Reliability: Oral 4-5 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.50	1561.651	0.659
Speaking	0.50	2983.671	0.903
Oral		1748.202	0.882

*Variances from students who had results in all four domains

Table 8.5.5Ei
Accuracy and Consistency of Classification Indices: Oral (Grade 4) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.615	0.507		0.360	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.835		0.707	
	2	0.588		0.450	
	3	0.603		0.466	
	4	0.376		0.278	
	5	0.644		0.521	
	6	0.677		0.598	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.983	0.006	0.011	0.975
	2/3	0.956	0.022	0.022	0.936
	3/4	0.924	0.028	0.049	0.893
	4/5	0.897	0.052	0.051	0.853
	5/6	0.836	0.053	0.110	0.780

Table 8.5.5Eii
Accuracy and Consistency of Classification Indices: Oral (Grade 5) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.596	0.489		0.333	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.854		0.744	
	2	0.552		0.420	
	3	0.537		0.401	
	4	0.362		0.263	
	5	0.647		0.528	
	6	0.630		0.547	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.978	0.008	0.014	0.969
	2/3	0.955	0.022	0.023	0.936
	3/4	0.929	0.027	0.044	0.899
	4/5	0.898	0.053	0.049	0.854
	5/6	0.813	0.060	0.128	0.754

8.5.6 Literacy Composite 4–5

Figure 8.5.6A
Scale Scores: Litr 4-5 S400 Paper

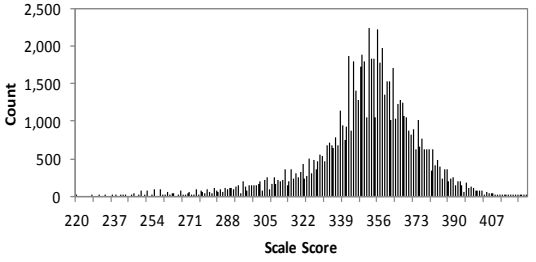


Figure 8.5.6B
Proficiency Level: Litr 4-5 S400 Paper

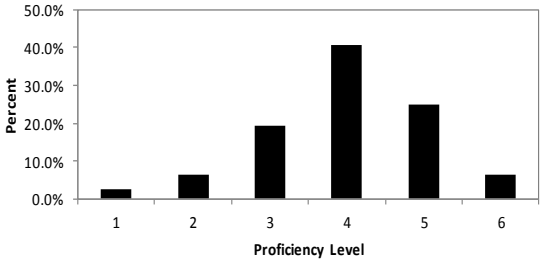


Table 8.5.6A
Scale Score Descriptive Statistics: Litr 4-5 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	42,232	227	423	347.53	23.37
5	32,507	220	423	353.69	25.28
Total	74,739	220	423	350.21	24.41

Table 8.5.6B
Proficiency Level Distribution: Litr 4-5 S400 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	853	2.0%	1,143	3.5%	1,996	2.7%
2	2,302	5.5%	2,421	7.4%	4,723	6.3%
3	6,935	16.4%	7,542	23.2%	14,477	19.4%
4	17,768	42.1%	12,521	38.5%	30,289	40.5%
5	11,371	26.9%	7,173	22.1%	18,544	24.8%
6	3,003	7.1%	1,707	5.3%	4,710	6.3%
Total	42,232	100.0%	32,507	100.0%	74,739	100.0%

Table 8.5.66C

n/a

Figure 8.5.66C

n/a

Figure 8.5.66D

n/a

Table 8.5.6D
Literacy Composite Reliability: Litr 4-5 S400 Paper

Component	Weight	Variance	Reliability
Reading	0.50	892.680	0.797
Writing	0.50	592.838	0.902
Literacy		590.684	0.899

*Variances from students who had results in all four domains

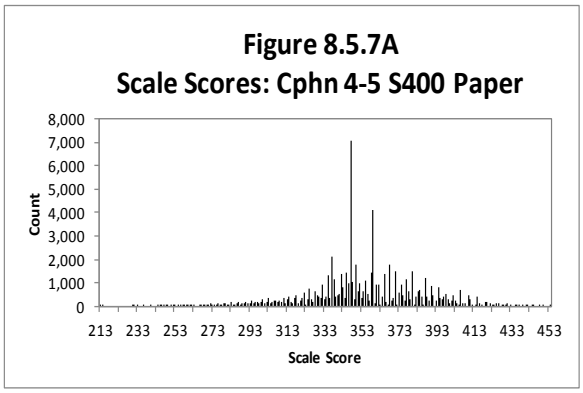
Table 8.5.6Ei
Accuracy and Consistency of Classification Indices: Litr (Grade 4) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.689	0.592		0.436	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.896		0.807	
	2	0.737		0.620	
	3	0.693		0.567	
	4	0.799		0.699	
	5	0.577		0.519	
	6	-		0.303	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.993	0.002	0.005	0.991
	2/3	0.979	0.009	0.012	0.969
	3/4	0.925	0.042	0.034	0.891
	4/5	0.863	0.037	0.100	0.813
	5/6	0.929	0.071	0.000	0.916

Table 8.5.6Eii
Accuracy and Consistency of Classification Indices: Litr (Grade 5) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.676	0.573		0.427	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.889		0.798	
	2	0.687		0.559	
	3	0.724		0.608	
	4	0.737		0.628	
	5	0.556		0.487	
	6	-		0.261	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.988	0.003	0.008	0.984
	2/3	0.967	0.016	0.017	0.951
	3/4	0.905	0.047	0.048	0.865
	4/5	0.867	0.041	0.092	0.815
	5/6	0.948	0.053	0.000	0.939

8.5.7 Comprehension Composite 4–5



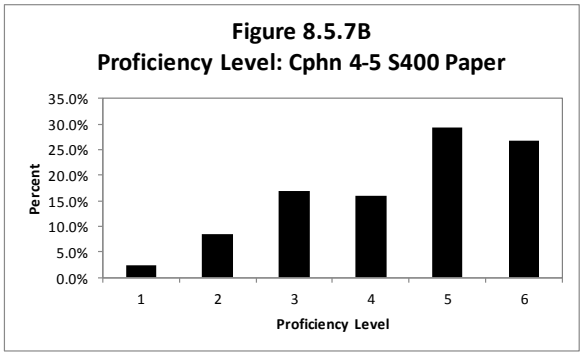


Table 8.5.7A

Scale Score Descriptive Statistics: Cphn 4-5 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	41,983	213	454	350.03	29.31
5	32,362	230	454	359.54	32.12
Total	74,345	213	454	354.17	30.93

Table 8.5.7B

Proficiency Level Distribution: Cphn 4-5 S400 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	691	1.6%	1,159	3.6%	1,850	2.5%
2	3,301	7.9%	2,962	9.2%	6,263	8.4%
3	6,710	16.0%	5,918	18.3%	12,628	17.0%
4	7,024	16.7%	4,901	15.1%	11,925	16.0%
5	12,892	30.7%	8,947	27.6%	21,839	29.4%
6	11,365	27.1%	8,475	26.2%	19,840	26.7%
Total	41,983	100.0%	32,362	100.0%	74,345	100.0%

Table 8.5.77C

n/a

Figure 8.5.77C

n/a

Figure 8.5.77D

n/a

Table 8.5.7D

Comprehension Composite Reliability: Cphn 4-5 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.30	1561.651	0.659
Reading	0.70	892.680	0.797
Comprehension		956.239	0.857

*Variances from students who had results in all four domains

Table 8.5.7Ei

Accuracy and Consistency of Classification Indices: Cphn (Grade 4) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.616	0.517		0.376	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.839		0.559	
	2	0.638		0.498	
	3	0.519		0.398	
	4	0.402		0.313	
	5	0.606		0.511	
	6	0.804		0.702	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.987	0.001	0.013	0.985
	2/3	0.955	0.014	0.031	0.934
	3/4	0.890	0.067	0.043	0.844
	4/5	0.856	0.073	0.071	0.808
	5/6	0.893	0.054	0.053	0.847

Table 8.5.7Eii

Accuracy and Consistency of Classification Indices: Cphn (Grade 5) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.603	0.504		0.371	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.857		0.678	
	2	0.602		0.461	
	3	0.534		0.416	
	4	0.362		0.281	
	5	0.563		0.468	
	6	0.799		0.692	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.980	0.003	0.017	0.974
	2/3	0.948	0.021	0.031	0.922
	3/4	0.879	0.071	0.050	0.833
	4/5	0.857	0.067	0.076	0.809
	5/6	0.892	0.056	0.052	0.846

8.5.8 Overall Composite 4–5

Figure 8.5.8A
Scale Scores: Over 4-5 S400 Paper

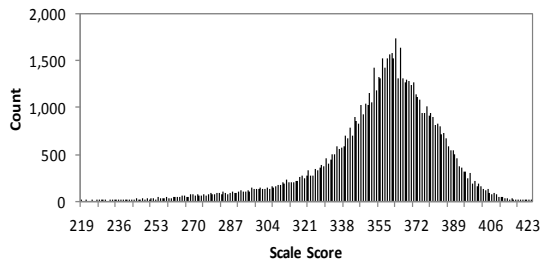


Figure 8.5.8B
Proficiency Level: Over 4-5 S400 Paper

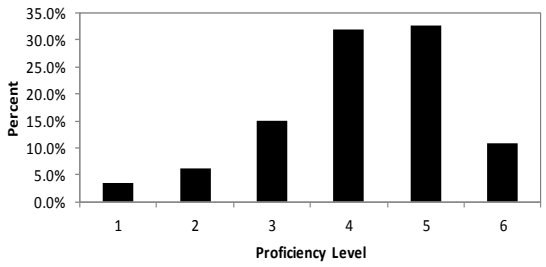


Table 8.5.8A
Scale Score Descriptive Statistics: Over 4-5 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	41,726	219	425	352.42	26.76
5	32,179	221	425	358.67	28.97
Total	73,905	219	425	355.14	27.91

Table 8.5.8B
Proficiency Level Distribution: Over 4-5 S400 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,158	2.8%	1,363	4.2%	2,521	3.4%
2	2,303	5.5%	2,247	7.0%	4,550	6.2%
3	6,078	14.6%	5,108	15.9%	11,186	15.1%
4	13,202	31.6%	10,380	32.3%	23,582	31.9%
5	14,011	33.6%	10,105	31.4%	24,116	32.6%
6	4,974	11.9%	2,976	9.2%	7,950	10.8%
Total	41,726	100.0%	32,179	100.0%	73,905	100.0%

Table 8.5.88C

n/a

Figure 8.5.88C

n/a

Figure 8.5.88D

n/a

Table 8.5.8D
Overall Composite Reliability: Over 4-5 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.15	1561.651	0.659
Reading	0.35	892.680	0.797
Speaking	0.15	2983.671	0.903
Writing	0.35	592.838	0.902
Overall Composite		779.104	0.939

*Variances from students who had results in all four domains

Table 8.5.8Ei
Accuracy and Consistency of Classification Indices: Over (Grade 4) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.706	0.631		0.504	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.921		0.862	
	2	0.757		0.656	
	3	0.733		0.620	
	4	0.804		0.710	
	5	0.626		0.604	
	6	-		0.452	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.993	0.002	0.005	0.991
	2/3	0.982	0.009	0.009	0.973
	3/4	0.944	0.031	0.025	0.919
	4/5	0.907	0.029	0.064	0.872
	5/6	0.881	0.119	0.000	0.872

Table 8.5.8Eii
Accuracy and Consistency of Classification Indices: Over (Grade 5) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.709	0.620		0.495	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.914		0.854	
	2	0.744		0.639	
	3	0.717		0.602	
	4	0.788		0.686	
	5	0.633		0.592	
	6	-		0.354	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.990	0.003	0.007	0.986
	2/3	0.976	0.012	0.012	0.966
	3/4	0.937	0.035	0.029	0.910
	4/5	0.898	0.030	0.072	0.860
	5/6	0.908	0.092	0.000	0.893

8.6 Grades: 6–8

8.6.1 Listening 6–8

Figure 8.6.1A
Scale Scores: List 6-8 S400 Paper

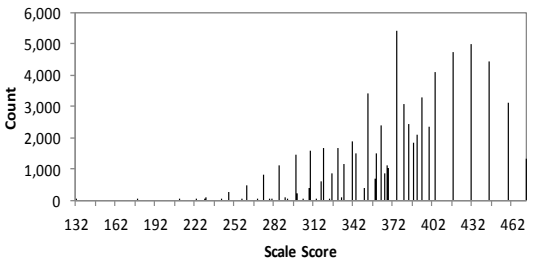


Figure 8.6.1B
Proficiency Level: List 6-8 S400 Paper

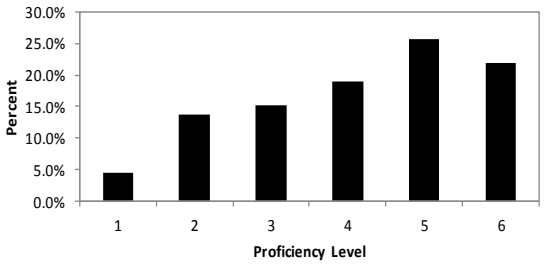


Table 8.6.1A
Scale Score Descriptive Statistics: List 6-8 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	25,093	178	473	374.20	44.51
7	23,141	178	473	379.66	47.87
8	22,752	132	473	386.25	49.49
Total	70,986	132	473	379.84	47.51

Table 8.6.1B
Proficiency Level Distribution: List 6-8 S400 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	698	2.8%	1,047	4.5%	1,447	6.4%	3,192	4.5%
2	2,892	11.5%	3,316	14.3%	3,552	15.6%	9,760	13.7%
3	4,210	16.8%	3,977	17.2%	2,555	11.2%	10,742	15.1%
4	4,823	19.2%	4,228	18.3%	4,420	19.4%	13,471	19.0%
5	7,124	28.4%	6,022	26.0%	5,072	22.3%	18,218	25.7%
6	5,346	21.3%	4,551	19.7%	5,706	25.1%	15,603	22.0%
Total	25,093	100.0%	23,141	100.0%	22,752	100.0%	70,986	100.0%

Table 8.6.1C

Conditional Standard Error of Measurement at Cut Scores: List 6-8 S400 Paper

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

Figure 8.6.1C
Test Characteristic Curve: List 6-8ABC S400 Paper

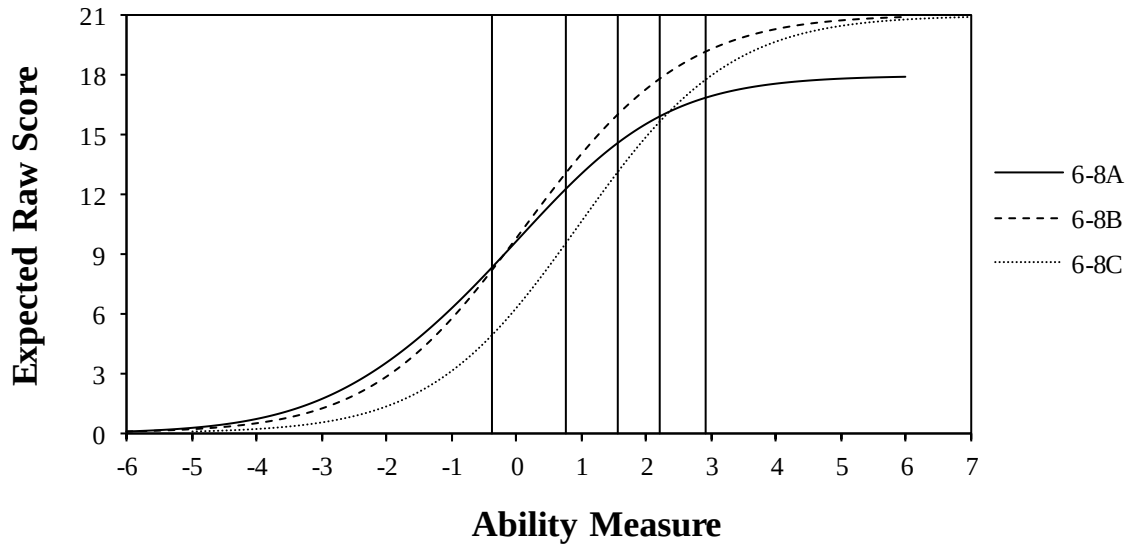


Figure 8.6.1D
Test Information Function: List 6-8ABC S400 Paper

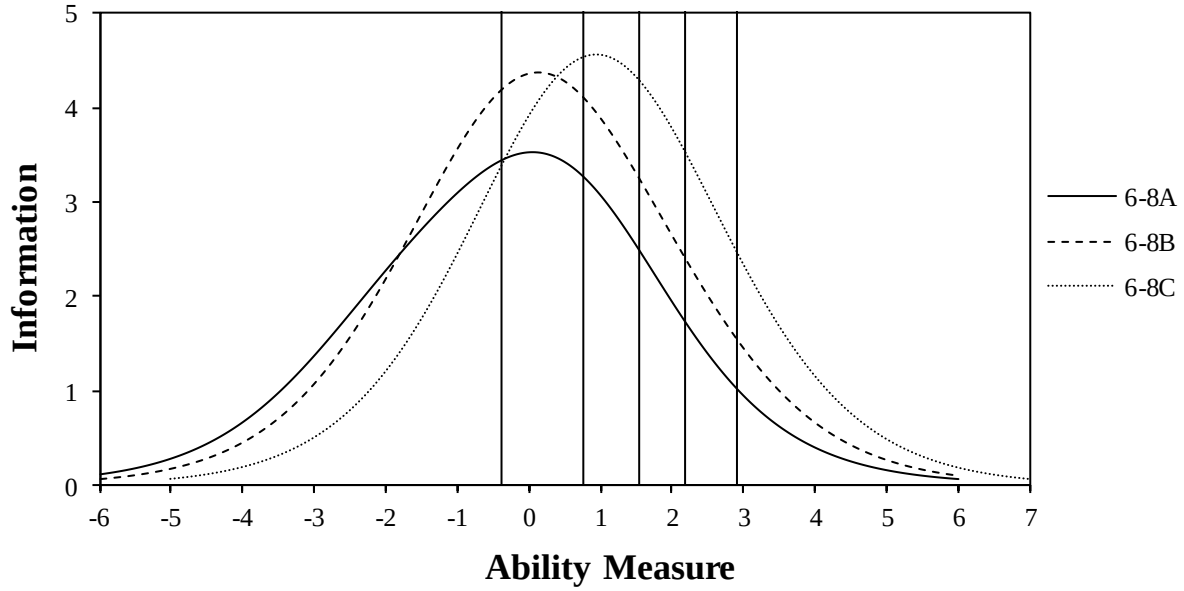


Table 8.6.1D
Weighted Reliability: List 6-8 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	16,796	0.734	0.651
B	21,402	0.653	
C	32,786	0.607	

Table 8.6.1Ei
Accuracy and Consistency of Classification Indices: List (Grade 6) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.446	0.354		0.192	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.655		0.262	
	2	0.480		0.321	
	3	0.343		0.261	
	4	0.309		0.249	
	5	0.460		0.378	
	6	0.694		0.509	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.973	0.001	0.026	0.961
	2/3	0.896	0.031	0.073	0.846
	3/4	0.818	0.094	0.087	0.748
	4/5	0.785	0.124	0.092	0.716
	5/6	0.855	0.091	0.054	0.791

Table 8.6.1Eii
Accuracy and Consistency of Classification Indices: List (Grade 7) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.430	0.341		0.188	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.633		0.299	
	2	0.466		0.332	
	3	0.326		0.253	
	4	0.294		0.237	
	5	0.446		0.360	
	6	0.676		0.488	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.958	0.004	0.038	0.937
	2/3	0.873	0.042	0.085	0.817
	3/4	0.812	0.099	0.089	0.741
	4/5	0.791	0.124	0.085	0.724
	5/6	0.862	0.085	0.054	0.798

Table 8.6.1Eiii
Accuracy and Consistency of Classification Indices: List (Grade 8) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.430	0.354		0.205	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.610		0.316	
	2	0.458		0.347	
	3	0.224		0.172	
	4	0.334		0.262	
	5	0.391		0.314	
	6	0.755		0.583	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.939	0.005	0.055	0.916
	2/3	0.864	0.041	0.094	0.810
	3/4	0.829	0.104	0.067	0.759
	4/5	0.804	0.121	0.075	0.739
	5/6	0.849	0.103	0.048	0.791

8.6.2 Reading 6–8

Figure 8.6.2A
Scale Scores: Read 6-8 S400 Paper

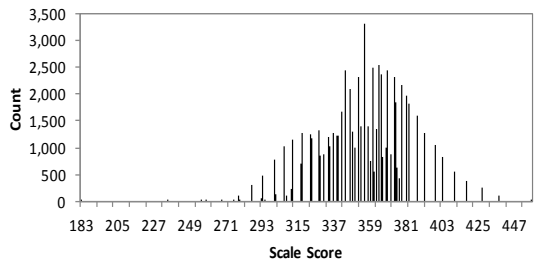


Figure 8.6.2B
Proficiency Level: Read 6-8 S400 Paper

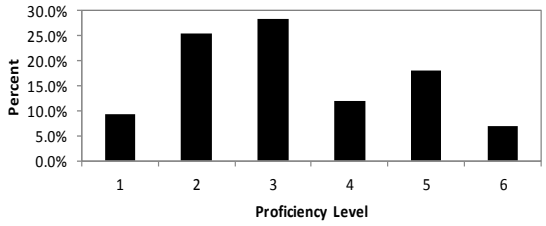


Table 8.6.2A
Scale Score Descriptive Statistics: Read 6-8 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	23,619	183	438	349.99	23.14
7	21,959	256	438	355.49	25.20
8	21,698	236	458	361.87	26.58
Total	67,276	183	458	355.61	25.43

Table 8.6.2B
Proficiency Level Distribution: Read 6-8 S400 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,415	6.0%	2,147	9.8%	2,712	12.5%	6,274	9.3%
2	5,468	23.2%	5,700	26.0%	5,978	27.6%	17,146	25.5%
3	7,488	31.7%	6,266	28.5%	5,326	24.5%	19,080	28.4%
4	3,170	13.4%	2,864	13.0%	1,954	9.0%	7,988	11.9%
5	4,425	18.7%	3,595	16.4%	4,061	18.7%	12,081	18.0%
6	1,653	7.0%	1,387	6.3%	1,667	7.7%	4,707	7.0%
Total	23,619	100.0%	21,959	100.0%	21,698	100.0%	67,276	100.0%

Table 8.6.2C

Conditional Standard Error of Measurement at Cut Scores: Read 6-8 S400 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	6	312	11.96	13.78	15.60
	7	321	11.70	12.74	14.04
	8	329	11.70	11.96	12.74
2/3	6	340	11.70	11.18	11.70
	7	349	12.22	10.92	11.18
	8	358	12.74	10.92	10.92
3/4	6	360	13.00	10.92	10.92
	7	369	13.78	10.92	10.66
	8	376	14.82	11.44	10.92
4/5	6	366	n/a	10.92	10.66
	7	375	n/a	11.18	10.92
	8	382	n/a	11.70	10.92
5/6	6	382	n/a	n/a	10.92
	7	391	n/a	n/a	11.44
	8	398	n/a	n/a	11.96

Figure 8.6.2C
Test Characteristic Curve: Read 6-8ABC S400 Paper

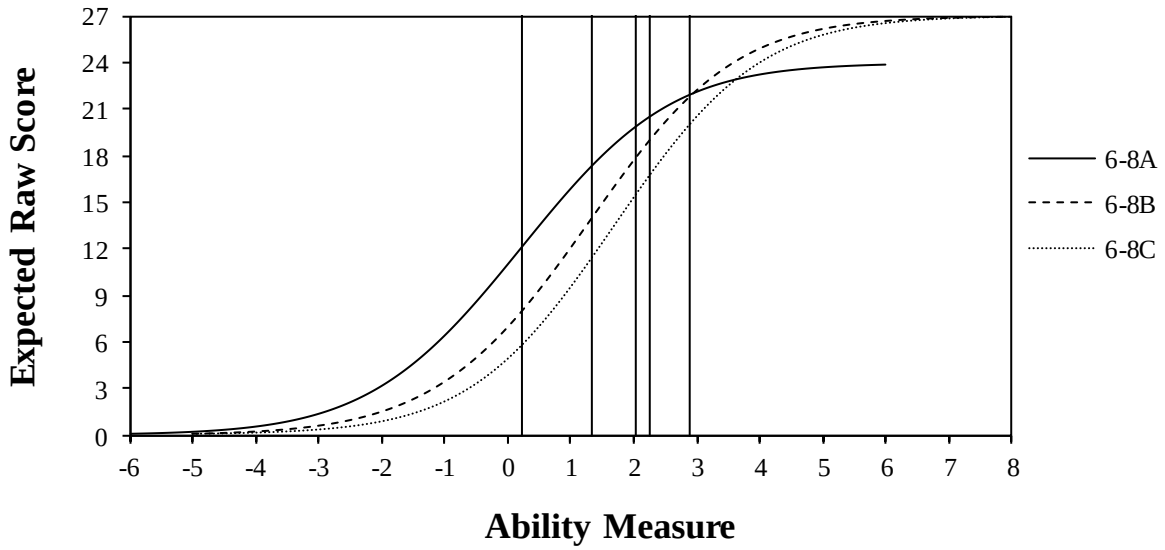


Figure 8.6.2D
Test Information Function: Read 6-8ABC S400 Paper

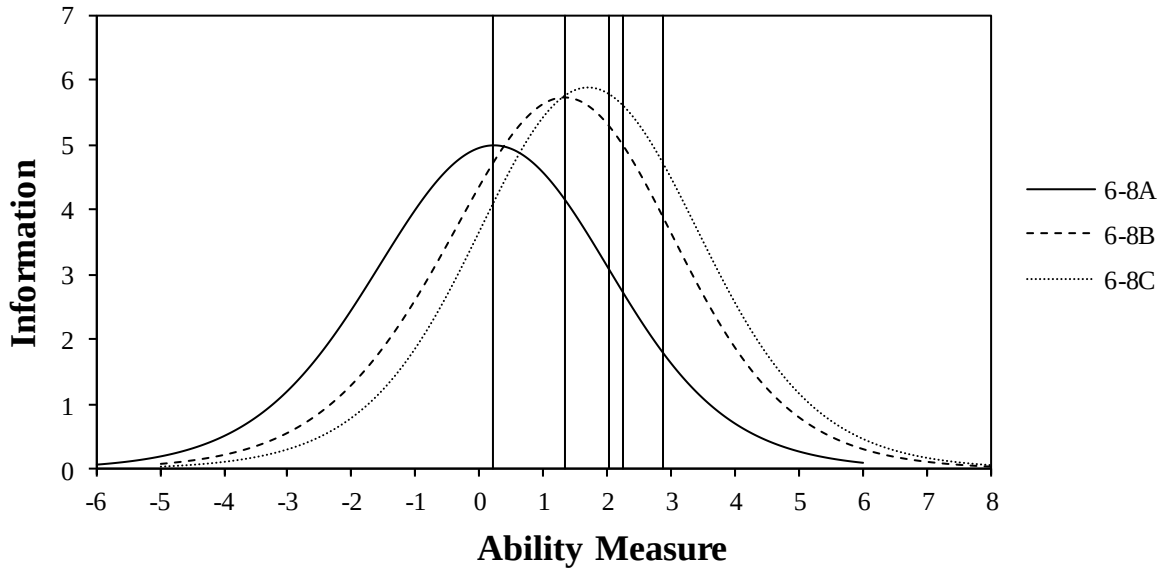


Table 8.6.2D
Weighted Reliability: Read 6-8 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	16,210	0.770	0.767
B	19,737	0.794	
C	31,327	0.748	

Table 8.6.2Ei
Accuracy and Consistency of Classification Indices: Read (Grade 6) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.543	0.434		0.275	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.762		0.530	
	2	0.634		0.507	
	3	0.543		0.455	
	4	0.281		0.219	
	5	0.509		0.392	
	6	0.683		0.420	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.960	0.009	0.031	0.943
	2/3	0.871	0.058	0.071	0.817
	3/4	0.831	0.107	0.062	0.776
	4/5	0.857	0.083	0.060	0.804
	5/6	0.946	0.040	0.014	0.918

Table 8.6.2Eii
Accuracy and Consistency of Classification Indices: Read (Grade 7) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.527	0.421		0.267	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.782		0.586	
	2	0.620		0.501	
	3	0.498		0.413	
	4	0.279		0.218	
	5	0.457		0.349	
	6	0.623		0.349	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.942	0.015	0.043	0.918
	2/3	0.863	0.064	0.073	0.805
	3/4	0.831	0.108	0.060	0.778
	4/5	0.858	0.080	0.062	0.806
	5/6	0.945	0.041	0.013	0.915

Table 8.6.2Eiii

Accuracy and Consistency of Classification Indices: Read (Grade 8) S400 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.524	0.422	0.275
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.797	0.623
	2	0.616	0.503
	3	0.457	0.367
	4	0.195	0.151
	5	0.492	0.387
	6	0.622	0.368

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.931	0.019	0.049	0.904
	2/3	0.859	0.066	0.075	0.800
	3/4	0.842	0.097	0.062	0.785
	4/5	0.853	0.094	0.053	0.802
	5/6	0.933	0.051	0.016	0.901

8.6.3 Writing 6–8

Figure 8.6.3A
Scale Scores: Writ 6-8 S400 Paper

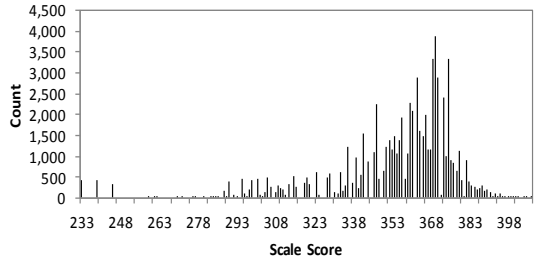


Figure 8.6.3B
Proficiency Level: Writ 6-8 S400 Paper

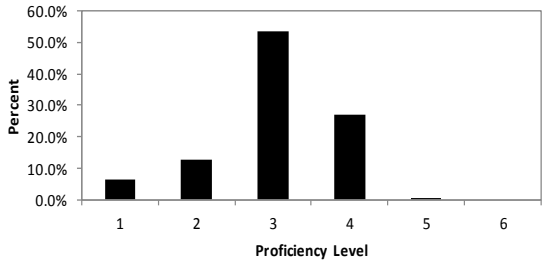


Table 8.6.3A

Scale Score Descriptive Statistics: Writ 6-8 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	25,286	233	401	352.22	26.40
7	23,338	239	402	353.71	26.19
8	22,975	245	407	356.09	25.50
Total	71,599	233	407	353.95	26.09

Table 8.6.3B

Proficiency Level Distribution: Writ 6-8 S400 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	979	3.9%	1,558	6.7%	2,146	9.3%	4,683	6.5%
2	2,678	10.6%	2,761	11.8%	3,814	16.6%	9,253	12.9%
3	9,399	37.2%	13,504	57.9%	15,435	67.2%	38,338	53.5%
4	12,113	47.9%	5,506	23.6%	1,580	6.9%	19,199	26.8%
5	117	0.5%	9	0.0%	0	0.0%	126	0.2%
6	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Total	25,286	100.0%	23,338	100.0%	22,975	100.0%	71,599	100.0%

Table 8.6.3C

Conditional Standard Error of Measurement at Cut Scores: Writ 6-8 S400 Paper

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

Figure 8.6.33C

n/a

Figure 8.6.3D

n/a

Table 8.6.3D
Weighted Reliability: Writ 6-8 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	17,007	0.859	0.901
B/C	54,592	0.914	

Table 8.6.3Ei
Accuracy and Consistency of Classification Indices: Writ (Grade 6) S400 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)	
	0.725	0.622	0.386	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.815	0.719	
	2	0.823	0.725	
	3	0.734	0.521	
	4	0.700	0.671	
	5	-	0.000	
	6	N/A	N/A	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.988	0.008	0.005	0.981
	2/3	0.967	0.012	0.021	0.954
	3/4	0.775	0.043	0.182	0.690
	4/5	0.995	0.005	0.000	0.995

Table 8.6.3Eii
Accuracy and Consistency of Classification Indices: Writ (Grade 7) S400 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)	
	0.709	0.628	0.353	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.878	0.807	
	2	0.809	0.698	
	3	0.683	0.685	
	4	-	0.350	
	5	N/A	N/A	
	6	N/A	N/A	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.984	0.008	0.008	0.977
	2/3	0.962	0.012	0.025	0.946
	3/4	0.764	0.236	0.000	0.704

Table 8.6.3Eiii

Accuracy and Consistency of Classification Indices: Writ (Grade 8) S400 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.854	0.792	0.571
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.885	0.819
	2	0.807	0.692
	3	0.859	0.847
	4	-	0.120
	5	N/A	N/A
	6	N/A	N/A

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.978	0.011	0.011	0.969
	2/3	0.944	0.017	0.039	0.921
	3/4	0.931	0.069	0.000	0.901

8.6.4 Speaking 6–8

Figure 8.6.4A
Scale Scores: Spek 6-8 S400 Paper

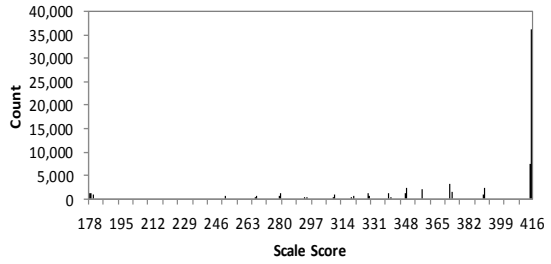


Figure 8.6.4B
Proficiency Level: Spek 6-8 S400 Paper

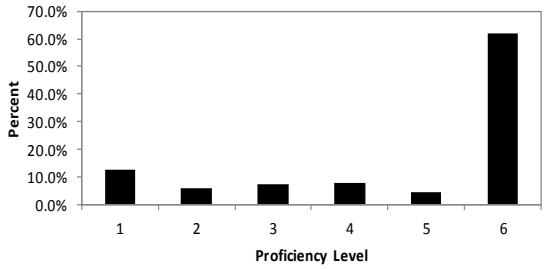


Table 8.6.4A
Scale Score Descriptive Statistics: Spek 6-8 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	25,053	178	416	378.77	61.79
7	23,094	179	416	379.29	61.40
8	22,711	180	416	381.02	59.65
Total	70,858	178	416	379.66	60.99

Table 8.6.4B
Proficiency Level Distribution: Spek 6-8 S400 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,981	11.9%	3,043	13.2%	2,816	12.4%	8,840	12.5%
2	1,102	4.4%	1,300	5.6%	1,658	7.3%	4,060	5.7%
3	1,969	7.9%	1,891	8.2%	1,380	6.1%	5,240	7.4%
4	2,628	10.5%	1,485	6.4%	1,489	6.6%	5,602	7.9%
5	1,277	5.1%	1,048	4.5%	929	4.1%	3,254	4.6%
6	15,096	60.3%	14,327	62.0%	14,439	63.6%	43,862	61.9%
Total	25,053	100.0%	23,094	100.0%	22,711	100.0%	70,858	100.0%

Table 8.6.4C

Conditional Standard Error of Measurement at Cut Scores: Spek 6-8 S400 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	6	310	21.43	21.43
	7	314	21.43	21.43
	8	317	21.94	21.94
2/3	6	337	23.47	23.47
	7	340	23.47	23.47
	8	344	23.47	23.47
3/4	6	353	23.47	23.47
	7	358	22.96	22.96
	8	361	22.96	22.96
4/5	6	377	21.94	21.94
	7	380	21.94	21.94
	8	384	21.94	21.94
5/6	6	397	21.94	21.94
	7	400	21.43	21.43
	8	404	21.94	21.94

Figure 8.6.4C

n/a

Figure 8.6.4D

n/a

Table 8.6.4D
Weighted Reliability: Spek 6-8 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	16,816	0.895	0.911
B/C	54,042	0.915	

Table 8.6.4Ei
Accuracy and Consistency of Classification Indices: Spek (Grade 6) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.725	0.606		0.375	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.824		0.745	
	2	0.280		0.215	
	3	0.386		0.299	
	4	0.345		0.229	
	5	0.092		0.064	
	6	0.867		0.830	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.964	0.022	0.014	0.947
	2/3	0.942	0.034	0.024	0.925
	3/4	0.924	0.022	0.055	0.905
	4/5	0.900	0.013	0.087	0.858
	5/6	0.876	0.038	0.086	0.762

Table 8.6.4Eii
Accuracy and Consistency of Classification Indices: Spek (Grade 7) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.745	0.637		0.417	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.816		0.741	
	2	0.331		0.260	
	3	0.399		0.315	
	4	0.251		0.169	
	5	0.103		0.062	
	6	0.915		0.878	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.957	0.026	0.017	0.940
	2/3	0.938	0.032	0.030	0.921
	3/4	0.926	0.018	0.056	0.907
	4/5	0.928	0.015	0.057	0.889
	5/6	0.900	0.047	0.053	0.796

Table 8.6.4Eiii

Accuracy and Consistency of Classification Indices: Spek (Grade 8) S400 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.746	0.638	0.415
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.795	0.721
	2	0.428	0.342
	3	0.308	0.244
	4	0.284	0.187
	5	0.102	0.057
	6	0.932	0.897

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.955	0.028	0.017	0.939
	2/3	0.938	0.026	0.036	0.922
	3/4	0.933	0.018	0.049	0.913
	4/5	0.936	0.015	0.048	0.898
	5/6	0.900	0.058	0.042	0.797

8.6.5 Oral Language Composite 6–8

Figure 8.6.5A
Scale Scores: Oral 6-8 S400 Paper

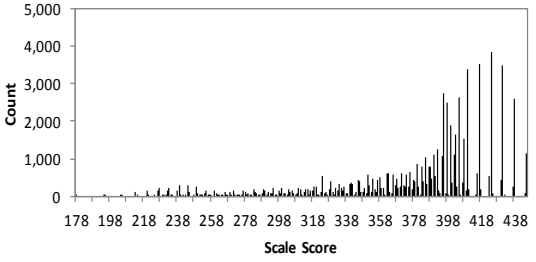


Figure 8.6.5B
Proficiency Level: Oral 6-8 S400 Paper

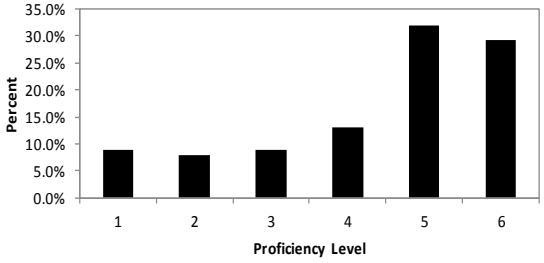


Table 8.6.5A
Scale Score Descriptive Statistics: Oral 6-8 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	24,892	178	445	376.93	47.61
7	22,930	195	445	379.90	49.36
8	22,528	195	445	384.12	49.27
Total	70,350	178	445	380.20	48.81

Table 8.6.5B
Proficiency Level Distribution: Oral 6-8 S400 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,972	7.9%	2,174	9.5%	2,121	9.4%	6,267	8.9%
2	1,684	6.8%	1,815	7.9%	2,047	9.1%	5,546	7.9%
3	2,300	9.2%	2,046	8.9%	1,944	8.6%	6,290	8.9%
4	3,186	12.8%	2,822	12.3%	3,204	14.2%	9,212	13.1%
5	8,453	34.0%	7,330	32.0%	6,610	29.3%	22,393	31.8%
6	7,297	29.3%	6,743	29.4%	6,602	29.3%	20,642	29.3%
Total	24,892	100.0%	22,930	100.0%	22,528	100.0%	70,350	100.0%

Table 8.6.5C

n/a

Figure 8.6.5C

n/a

Figure 8.6.5D

n/a

Table 8.6.5D
Oral Composite Reliability: Oral 6-8 S303

Component	Weight	Variance	Reliability
Listening	0.50	2274.197	0.651
Speaking	0.50	3662.433	0.911
Oral		2382.977	0.883

*Variances from students who had results in all four domains

Table 8.6.5Ei
Accuracy and Consistency of Classification Indices: Oral (Grade 6) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.573	0.473		0.313	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.874		0.782	
	2	0.515		0.384	
	3	0.468		0.340	
	4	0.403		0.290	
	5	0.580		0.469	
	6	0.600		0.529	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.975	0.009	0.015	0.965
	2/3	0.956	0.021	0.022	0.937
	3/4	0.932	0.028	0.040	0.904
	4/5	0.892	0.048	0.060	0.846
	5/6	0.785	0.059	0.156	0.727

Table 8.6.5Eii
Accuracy and Consistency of Classification Indices: Oral (Grade 7) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.570	0.470		0.317	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.864		0.772	
	2	0.524		0.395	
	3	0.426		0.308	
	4	0.388		0.274	
	5	0.555		0.449	
	6	0.613		0.538	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.970	0.012	0.018	0.957
	2/3	0.951	0.022	0.027	0.929
	3/4	0.930	0.030	0.039	0.900
	4/5	0.892	0.049	0.059	0.845
	5/6	0.790	0.066	0.144	0.731

Table 8.6.5Eiii

Accuracy and Consistency of Classification Indices: Oral (Grade 8) S400 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.568	0.465	0.318
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.842	0.742
	2	0.543	0.414
	3	0.386	0.280
	4	0.423	0.304
	5	0.523	0.421
	6	0.641	0.557

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.967	0.014	0.019	0.952
	2/3	0.945	0.023	0.032	0.922
	3/4	0.927	0.033	0.041	0.895
	4/5	0.886	0.052	0.062	0.838
	5/6	0.804	0.073	0.123	0.744

8.6.6 Literacy Composite 6–8

Figure 8.6.6A
Scale Scores: Litr 6-8 S400 Paper

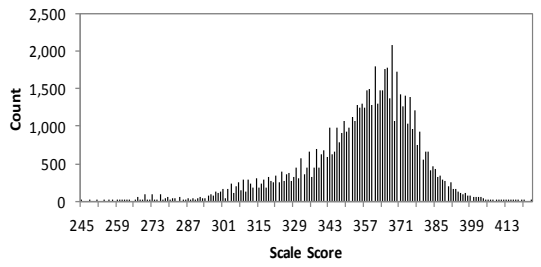


Figure 8.6.6B
Proficiency Level: Litr 6-8 S400 Paper

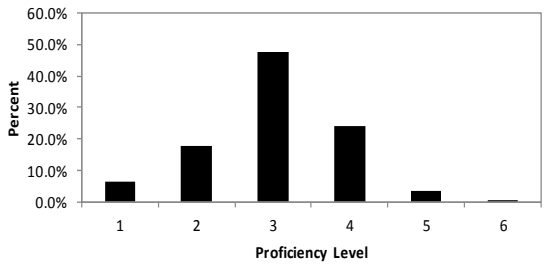


Table 8.7.8A
Scale Score Descriptive Statistics: Over 9-12 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	23,397	256	454	380.24	35.99
10	17,967	260	455	384.80	33.02
11	14,036	260	459	389.82	30.83
12	9,734	268	459	394.36	27.64
Total	65,134	256	459	385.67	33.33

Table 8.6.6B
Proficiency Level Distribution: Litr 6-8 S400 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	914	3.9%	1,436	6.5%	1,903	8.8%	4,253	6.3%
2	3,353	14.2%	4,016	18.3%	4,665	21.5%	12,034	17.9%
3	10,150	43.0%	10,523	47.9%	11,392	52.5%	32,065	47.7%
4	7,949	33.7%	5,118	23.3%	3,218	14.8%	16,285	24.2%
5	1,159	4.9%	815	3.7%	500	2.3%	2,474	3.7%
6	94	0.4%	51	0.2%	20	0.1%	165	0.2%
Total	23,619	100.0%	21,959	100.0%	21,698	100.0%	67,276	100.0%

Table 8.6.6C

n/a

Figure 8.6.6C

n/a

Figure 8.6.6D

n/a

Table 8.6.6D
Literacy Composite Reliability: Litr 6-8 S400 Paper

Component	Weight	Variance	Reliability
Reading	0.50	646.253	0.767
Writing	0.50	633.124	0.901
Literacy		510.632	0.896

*Variances from students who had results in all four domains

Table 8.6.6Ei
Accuracy and Consistency of Classification Indices: Litr (Grade 6) S400 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.750	0.658	0.495		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.797	0.757		
	2	0.773	0.664		
	3	0.810	0.711		
	4	0.681	0.626		
	5	-	0.183		
	6	-	0.996		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.984	0.008	0.008	0.981
	2/3	0.946	0.027	0.027	0.926
	3/4	0.865	0.051	0.084	0.813
	4/5	0.947	0.053	0.000	0.930
	5/6	0.996	0.004	0.000	0.999

Table 8.6.6Eii
Accuracy and Consistency of Classification Indices: Litr (Grade 7) S400 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.738	0.648	0.483		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.834	0.767		
	2	0.755	0.647		
	3	0.805	0.719		
	4	0.606	0.530		
	5	-	0.175		
	6	-	1.000		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.976	0.010	0.014	0.969
	2/3	0.931	0.034	0.035	0.904
	3/4	0.864	0.056	0.081	0.812
	4/5	0.961	0.039	0.000	0.953
	5/6	0.998	0.002	0.000	1.000

Table 8.6.6Eiii

Accuracy and Consistency of Classification Indices:Litr (Grade 8) S400 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.747	0.658	0.480
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.862	0.781
	2	0.750	0.644
	3	0.812	0.745
	4	0.509	0.409
	5	-	0.127
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.971	0.011	0.017	0.961
	2/3	0.921	0.038	0.041	0.888
	3/4	0.873	0.057	0.069	0.825
	4/5	0.976	0.024	0.000	0.973
	5/6	0.999	0.001	0.000	1.000

8.6.7 Comprehension Composite 6–8

Figure 8.6.7A
Scale Scores: Cphn 6-8 S400 Paper

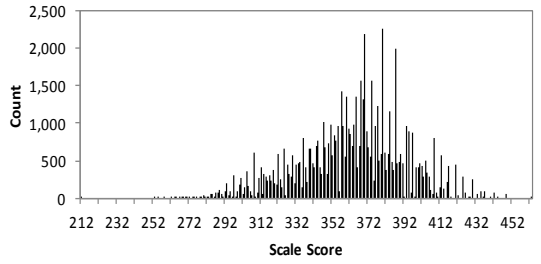


Figure 8.6.7B
Proficiency Level: Cphn 6-8 S400 Paper

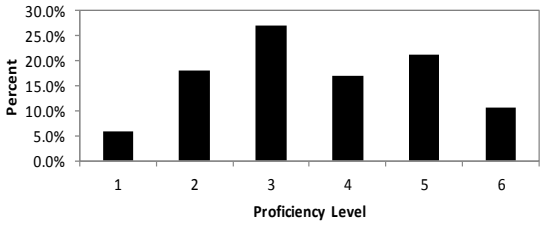


Table 8.6.7A

Scale Score Descriptive Statistics: Cphn 6-8 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	23,496	212	449	357.42	27.04
7	21,832	253	449	362.97	29.55
8	21,554	253	463	369.41	30.97
Total	66,882	212	463	363.10	29.59

Table 8.6.7B

Proficiency Level Distribution: Cphn 6-8 S400 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	793	3.4%	1,455	6.7%	1,705	7.9%	3,953	5.9%
2	3,709	15.8%	4,004	18.3%	4,353	20.2%	12,066	18.0%
3	6,943	29.5%	6,136	28.1%	5,063	23.5%	18,142	27.1%
4	4,166	17.7%	3,728	17.1%	3,516	16.3%	11,410	17.1%
5	5,367	22.8%	4,214	19.3%	4,632	21.5%	14,213	21.3%
6	2,518	10.7%	2,295	10.5%	2,285	10.6%	7,098	10.6%
Total	23,496	100.0%	21,832	100.0%	21,554	100.0%	66,882	100.0%

Table 8.6.7C

n/a

Figure 8.6.7C

n/a

Figure 8.6.7D

n/a

Table 8.6.7D

Comprehension Composite Reliability: Cphn 6-8 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.30	2274.197	0.651
Reading	0.70	646.253	0.767
Comprehension		875.832	0.834

*Variances from students who had results in all four domains

Table 8.6.7Ei

Accuracy and Consistency of Classification Indices: Cphn (Grade 6) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.592	0.482		0.344	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.807		0.594	
	2	0.671		0.536	
	3	0.614		0.512	
	4	0.391		0.306	
	5	0.574		0.463	
	6	0.752		0.561	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.979	0.004	0.017	0.972
	2/3	0.917	0.036	0.047	0.880
	3/4	0.858	0.079	0.063	0.806
	4/5	0.866	0.069	0.064	0.816
	5/6	0.936	0.043	0.021	0.906

Table 8.6.7Eii

Accuracy and Consistency of Classification Indices: Cphn (Grade 7) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.571	0.462		0.331	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.826		0.652	
	2	0.641		0.516	
	3	0.588		0.484	
	4	0.386		0.300	
	5	0.510		0.404	
	6	0.720		0.518	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.964	0.008	0.027	0.951
	2/3	0.906	0.043	0.051	0.865
	3/4	0.859	0.077	0.064	0.808
	4/5	0.870	0.066	0.063	0.821
	5/6	0.930	0.048	0.022	0.899

Table 8.6.7Eiii

Accuracy and Consistency of Classification Indices: Cphn (Grade 8) S400 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.556	0.448	0.320
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.820	0.661
	2	0.660	0.536
	3	0.521	0.415
	4	0.361	0.281
	5	0.523	0.424
	6	0.663	0.462

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.960	0.011	0.028	0.945
	2/3	0.903	0.043	0.054	0.861
	3/4	0.861	0.076	0.063	0.809
	4/5	0.861	0.073	0.066	0.811
	5/6	0.921	0.052	0.028	0.885

8.6.8 Overall Composite 6–8

Figure 8.6.8A
Scale Scores: Over 6-8 S400 Paper

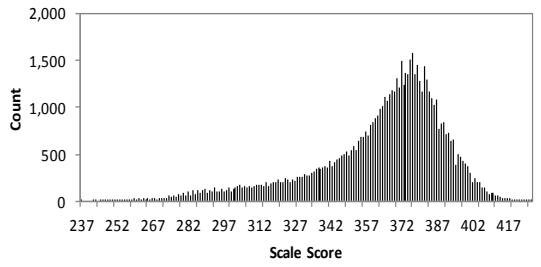


Figure 8.6.8B
Proficiency Level: Over 6-8 S400 Paper

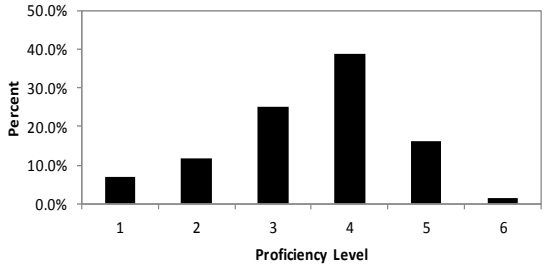


Table 8.6.8A

Scale Score Descriptive Statistics: Over 6-8 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	23,311	237	427	359.13	27.53
7	21,632	237	425	362.46	28.82
8	21,347	253	427	366.78	29.08
Total	66,290	237	427	362.68	28.63

Table 8.6.8B

Proficiency Level Distribution: Over 6-8 S400 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,268	5.4%	1,594	7.4%	1,838	8.6%	4,700	7.1%
2	2,257	9.7%	2,589	12.0%	2,881	13.5%	7,727	11.7%
3	5,046	21.6%	5,712	26.4%	5,919	27.7%	16,677	25.2%
4	9,748	41.8%	7,980	36.9%	7,936	37.2%	25,664	38.7%
5	4,529	19.4%	3,487	16.1%	2,626	12.3%	10,642	16.1%
6	463	2.0%	270	1.2%	147	0.7%	880	1.3%
Total	23,311	100.0%	21,632	100.0%	21,347	100.0%	66,290	100.0%

Table 8.6.8C

n/a

Figure 8.6.8C

n/a

Figure 8.6.8D

n/a

Table 8.6.8D

Overall Composite Reliability: Over 6-8 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.15	2274.197	0.651
Reading	0.35	646.253	0.767
Speaking	0.15	3662.433	0.911
Writing	0.35	633.124	0.901
Overall Composite		819.734	0.937

*Variances from students who had results in all four domains

Table 8.6.8Ei

Accuracy and Consistency of Classification Indices: Over (Grade 6) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.717	0.626		0.490	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.913		0.857	
	2	0.765		0.666	
	3	0.769		0.662	
	4	0.766		0.661	
	5	0.553		0.477	
	6	-		0.152	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.988	0.004	0.008	0.983
	2/3	0.969	0.016	0.015	0.955
	3/4	0.927	0.033	0.039	0.897
	4/5	0.853	0.048	0.099	0.803
	5/6	0.980	0.020	0.000	0.979

Table 8.6.8Eii

Accuracy and Consistency of Classification Indices: Over (Grade 7) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.717	0.622		0.498	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.904		0.845	
	2	0.750		0.647	
	3	0.771		0.669	
	4	0.724		0.625	
	5	0.553		0.459	
	6	-		0.500	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.982	0.007	0.011	0.976
	2/3	0.960	0.020	0.020	0.942
	3/4	0.914	0.039	0.047	0.879
	4/5	0.872	0.051	0.077	0.827
	5/6	0.988	0.012	0.000	0.988

Table 8.6.8Eiii

Accuracy and Consistency of Classification Indices: Over (Grade 8) S400 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**	
	0.718	0.630	0.501	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.902	0.845	
	2	0.754	0.653	
	3	0.780	0.675	
	4	0.649	0.629	
	5	-	0.364	
	6	-	0.886	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.980	0.008	0.012	0.972
	2/3	0.955	0.023	0.022	0.937
	3/4	0.912	0.035	0.053	0.876
	4/5	0.870	0.130	0.000	0.842
	5/6	0.993	0.007	0.000	0.993

8.7 Grades: 9–12

8.7.1 Listening 9–12

Figure 8.7.1A
Scale Scores: List 9-12 S400 Paper

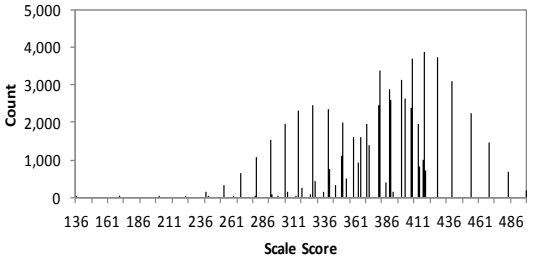


Figure 8.7.1B
Proficiency Level: List 9-12 S400 Paper

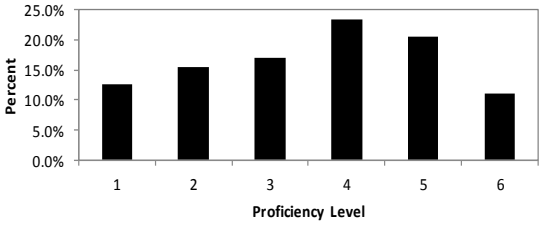


Table 8.7.1A
Scale Score Descriptive Statistics: List 9-12 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	25,126	136	499	376.56	49.49
10	19,303	203	499	381.43	46.39
11	14,994	224	499	385.74	44.91
12	10,379	224	499	391.36	41.22
Total	69,802	136	499	382.08	46.79

Table 8.7.1B
Proficiency Level Distribution: List 9-12 S400 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,021	12.0%	2,375	12.3%	2,106	14.0%	1,367	13.2%	8,869	12.7%
2	4,618	18.4%	3,050	15.8%	1,874	12.5%	1,197	11.5%	10,739	15.4%
3	3,436	13.7%	3,538	18.3%	2,577	17.2%	2,339	22.5%	11,890	17.0%
4	5,204	20.7%	4,293	22.2%	3,968	26.5%	2,850	27.5%	16,315	23.4%
5	6,336	25.2%	3,963	20.5%	2,671	17.8%	1,331	12.8%	14,301	20.5%
6	2,511	10.0%	2,084	10.8%	1,798	12.0%	1,295	12.5%	7,688	11.0%
Total	25,126	100.0%	19,303	100.0%	14,994	100.0%	10,379	100.0%	69,802	100.0%

Table 8.7.1C

Conditional Standard Error of Measurement at Cut Scores: List 9-12 S400 Paper

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

Figure 8.7.1C
Test Characteristic Curve: List 9-12ABC S400 Paper

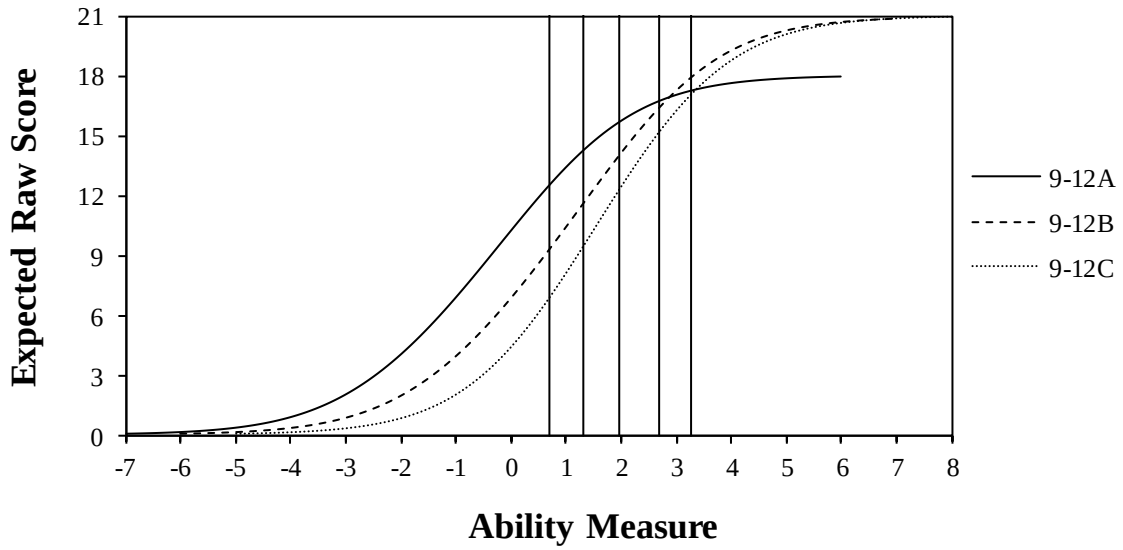


Figure 8.7.1D
Test Information Function: List 9-12ABC S400 Paper

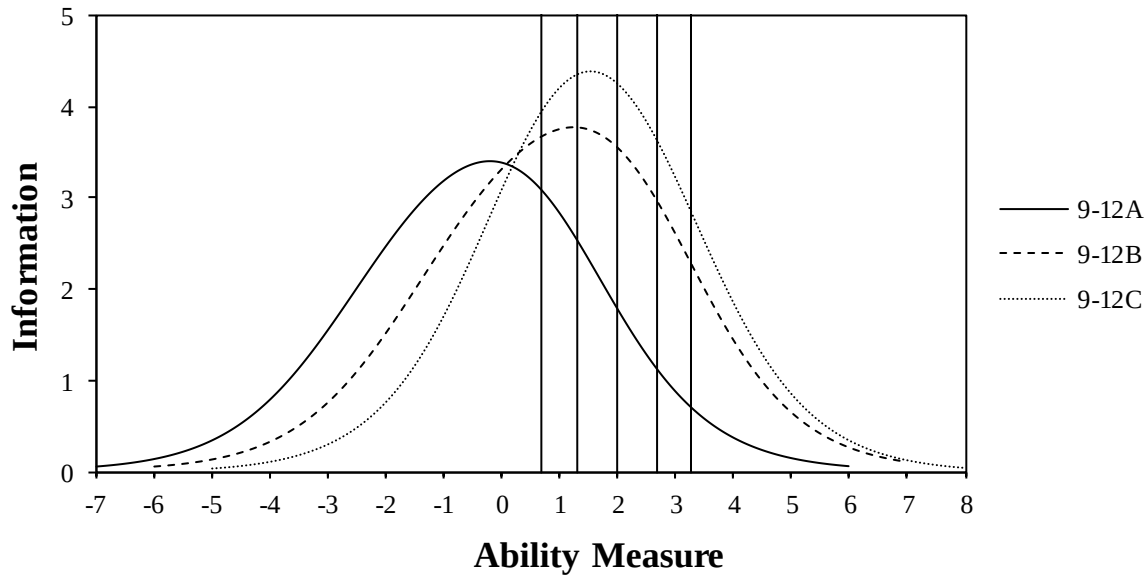


Table 8.7.1D
Weighted Reliability: List 9-12 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	18,892	0.639	0.631
B	20,988	0.651	
C	29,920	0.611	

Table 8.7.1Ei
Accuracy and Consistency of Classification Indices: List (Grade 9) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.400	0.309		0.164	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.719		0.501	
	2	0.452		0.338	
	3	0.246		0.184	
	4	0.323		0.264	
	5	0.467		0.387	
	6	0.447		0.252	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.913	0.021	0.065	0.880
	2/3	0.854	0.049	0.097	0.791
	3/4	0.813	0.108	0.079	0.734
	4/5	0.780	0.144	0.076	0.718
	5/6	0.896	0.083	0.021	0.834

Table 8.7.1Eii
Accuracy and Consistency of Classification Indices: List (Grade 10) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.398	0.313		0.167	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.764		0.530	
	2	0.394		0.287	
	3	0.321		0.246	
	4	0.345		0.285	
	5	0.405		0.325	
	6	0.538		0.295	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.917	0.018	0.066	0.884
	2/3	0.853	0.055	0.092	0.788
	3/4	0.802	0.106	0.092	0.727
	4/5	0.793	0.138	0.068	0.731
	5/6	0.895	0.088	0.018	0.842

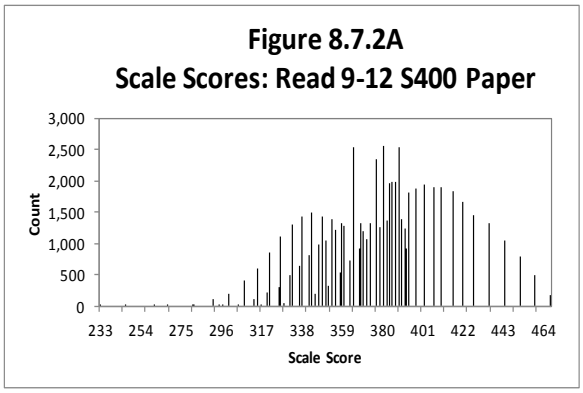
Table 8.7.1Eiii
Accuracy and Consistency of Classification Indices: List (Grade 11) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.393	0.312		0.165	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.791		0.567	
	2	0.324		0.229	
	3	0.305		0.230	
	4	0.396		0.332	
	5	0.340		0.275	
	6	0.541		0.299	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.911	0.018	0.071	0.876
	2/3	0.863	0.054	0.082	0.797
	3/4	0.799	0.109	0.092	0.722
	4/5	0.789	0.134	0.077	0.725
	5/6	0.883	0.103	0.014	0.829

Table 8.7.1Eiv
Accuracy and Consistency of Classification Indices: List (Grade 12) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.395	0.316		0.161	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.743		0.514	
	2	0.289		0.206	
	3	0.382		0.300	
	4	0.413		0.344	
	5	0.265		0.206	
	6	0.609		0.332	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.912	0.023	0.065	0.872
	2/3	0.854	0.056	0.090	0.792
	3/4	0.790	0.091	0.119	0.717
	4/5	0.814	0.105	0.081	0.741
	5/6	0.881	0.110	0.010	0.836

8.7.2 Reading 9–12



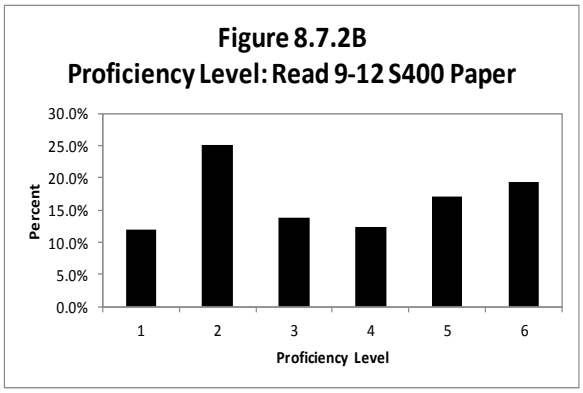


Table 8.7.2A

Scale Score Descriptive Statistics: Read 9-12 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	24,028	246	468	373.41	31.86
10	18,436	246	468	377.66	31.57
11	14,448	268	468	382.91	31.05
12	10,016	233	468	385.64	29.92
Total	66,928	233	468	378.46	31.66

Table 8.7.2B

Proficiency Level Distribution: Read 9-12 S400 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,022	12.6%	2,270	12.3%	1,623	11.2%	1,142	11.4%	8,057	12.0%
2	5,745	23.9%	4,995	27.1%	3,712	25.7%	2,406	24.0%	16,858	25.2%
3	4,170	17.4%	2,399	13.0%	1,597	11.1%	1,115	11.1%	9,281	13.9%
4	2,699	11.2%	2,566	13.9%	1,979	13.7%	1,074	10.7%	8,318	12.4%
5	3,969	16.5%	2,859	15.5%	2,411	16.7%	2,253	22.5%	11,492	17.2%
6	4,423	18.4%	3,347	18.2%	3,126	21.6%	2,026	20.2%	12,922	19.3%
Total	24,028	100.0%	18,436	100.0%	14,448	100.0%	10,016	100.0%	66,928	100.0%

Table 8.7.2C

Conditional Standard Error of Measurement at Cut Scores: Read 9-12 S400 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	9	336	11.96	11.70	14.82
	10	341	11.70	11.44	14.04
	11	346	11.70	11.18	13.26
	12	350	11.70	10.92	12.74
2/3	9	364	11.96	10.92	11.44
	10	370	12.48	10.92	10.92
	11	374	12.74	11.18	10.66
	12	376	13.00	11.18	10.66
3/4	9	381	13.52	11.44	10.40
	10	383	13.78	11.70	10.40
	11	384	13.78	11.70	10.40
	12	385	14.04	11.70	10.40
4/5	9	387	n/a	11.96	10.40
	10	390	n/a	12.22	10.40
	11	392	n/a	12.48	10.40
	12	393	n/a	12.48	10.40
5/6	9	402	n/a	n/a	10.66
	10	406	n/a	n/a	10.92
	11	407	n/a	n/a	10.92
	12	408	n/a	n/a	11.18

Figure 8.7.2C
Test Characteristic Curve: Read 9-12ABC S400 Paper

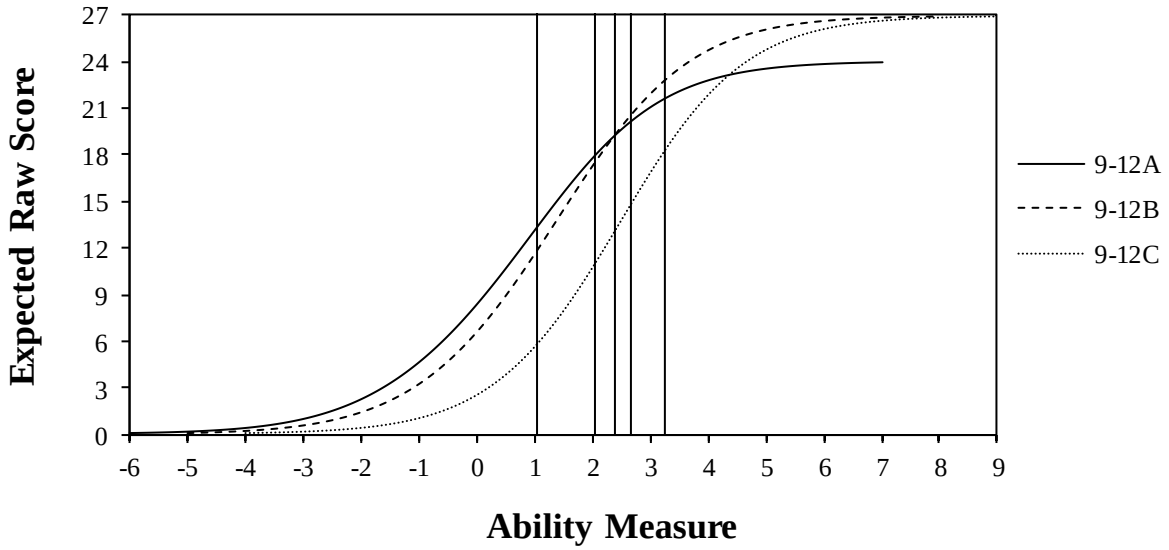


Figure 8.7.2D
Test Information Function: Read 9-12ABC S400 Paper

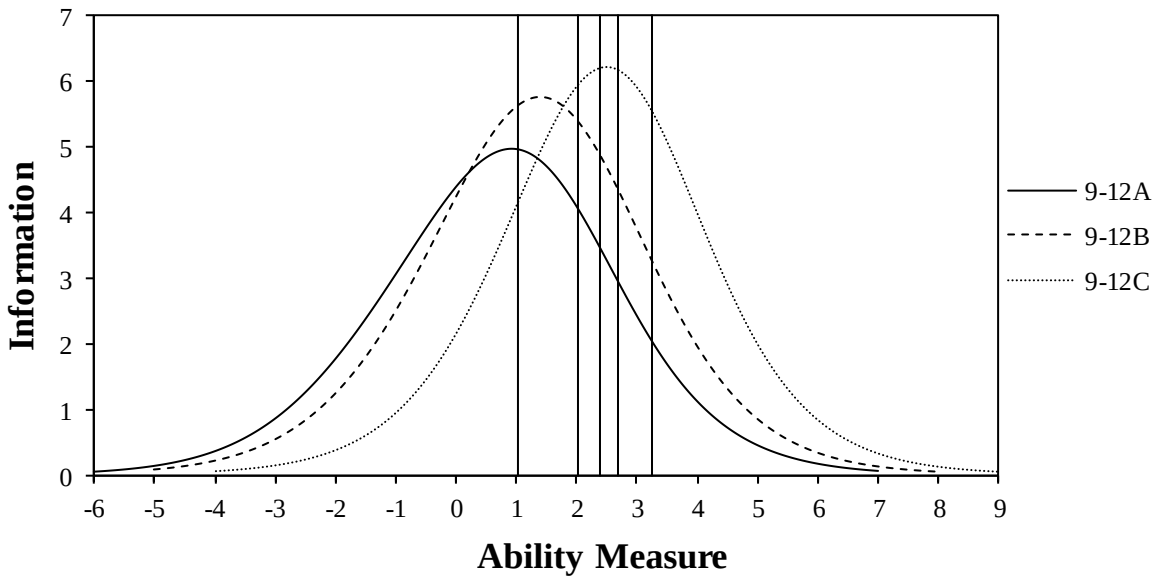


Table 8.7.2D
Weighted Reliability: Read 9-12 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	18,634	0.791	0.800
B	19,965	0.813	
C	28,324	0.798	

Table 8.7.2Ei
Accuracy and Consistency of Classification Indices: Read (Grade 9) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.516	0.423		0.295	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.734		0.547	
	2	0.555		0.448	
	3	0.357		0.280	
	4	0.251		0.194	
	5	0.404		0.314	
	6	0.765		0.620	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.917	0.024	0.058	0.885
	2/3	0.868	0.068	0.064	0.811
	3/4	0.845	0.103	0.052	0.795
	4/5	0.854	0.081	0.065	0.805
	5/6	0.906	0.054	0.040	0.864

Table 8.7.2Eii
Accuracy and Consistency of Classification Indices: Read (Grade 10) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.515	0.424		0.293	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.677		0.483	
	2	0.570		0.472	
	3	0.274		0.212	
	4	0.315		0.243	
	5	0.400		0.308	
	6	0.790		0.647	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.908	0.028	0.064	0.873
	2/3	0.861	0.077	0.061	0.803
	3/4	0.845	0.105	0.050	0.796
	4/5	0.861	0.072	0.067	0.814
	5/6	0.916	0.049	0.035	0.876

Table 8.7.2Eiii

Accuracy and Consistency of Classification Indices: Read (Grade 11) S400 Paper

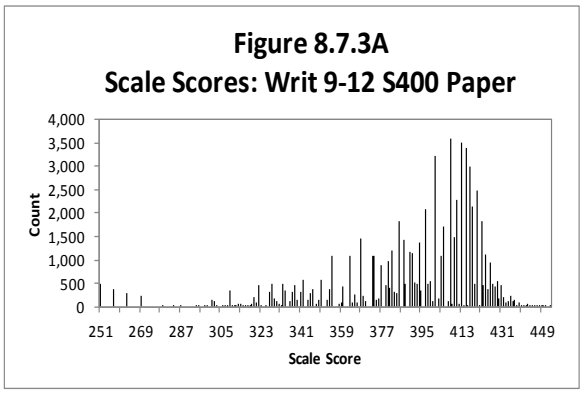
Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.515	0.427		0.294	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.652		0.448	
	2	0.547		0.454	
	3	0.235		0.180	
	4	0.298		0.231	
	5	0.406		0.314	
	6	0.795		0.668	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.910	0.025	0.065	0.876
	2/3	0.858	0.083	0.059	0.800
	3/4	0.844	0.101	0.055	0.793
	4/5	0.856	0.079	0.066	0.806
	5/6	0.905	0.053	0.042	0.862

Table 8.7.2Eiv

Accuracy and Consistency of Classification Indices: Read (Grade 12) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.512	0.421		0.289	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.663		0.463	
	2	0.531		0.434	
	3	0.238		0.180	
	4	0.227		0.175	
	5	0.512		0.404	
	6	0.754		0.625	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.912	0.026	0.062	0.878
	2/3	0.859	0.081	0.060	0.801
	3/4	0.848	0.093	0.059	0.792
	4/5	0.846	0.097	0.057	0.796
	5/6	0.900	0.050	0.050	0.856

8.7.3 Writing 9–12



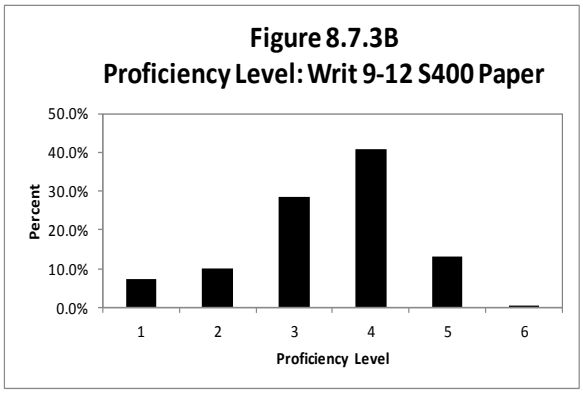


Table 8.7.3A

Scale Score Descriptive Statistics: Writ 9-12 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	25,569	251	446	386.56	36.02
10	19,542	257	451	390.30	34.32
11	15,216	263	453	394.26	32.91
12	10,563	269	451	398.06	31.58
Total	70,890	251	453	390.96	34.50

Table 8.7.3B

Proficiency Level Distribution: Writ 9-12 S400 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,554	6.1%	1,590	8.1%	1,166	7.7%	862	8.2%	5,172	7.3%
2	3,507	13.7%	1,446	7.4%	1,377	9.0%	719	6.8%	7,049	9.9%
3	5,706	22.3%	5,602	28.7%	4,902	32.2%	4,057	38.4%	20,267	28.6%
4	8,657	33.9%	8,766	44.9%	6,842	45.0%	4,654	44.1%	28,919	40.8%
5	5,945	23.3%	2,091	10.7%	919	6.0%	271	2.6%	9,226	13.0%
6	200	0.8%	47	0.2%	10	0.1%	0	0.0%	257	0.4%
Total	25,569	100.0%	19,542	100.0%	15,216	100.0%	10,563	100.0%	70,890	100.0%

Table 8.7.3C

Conditional Standard Error of Measurement at Cut Scores: Writ 9-12 S400 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	9	327	9.02	6.53
	10	336	10.26	7.15
	11	344	11.19	8.09
	12	352	11.82	8.40
2/3	9	356	12.13	8.40
	10	363	12.13	8.40
	11	370	11.82	8.40
	12	377	11.51	8.09
3/4	9	389	11.19	8.09
	10	397	11.19	7.77
	11	404	10.88	7.77
	12	410	10.57	7.46
4/5	9	415	10.26	7.15
	10	422	9.95	6.84
	11	428	9.33	6.84
	12	434	9.33	6.53
5/6	9	435	9.33	6.53
	10	441	9.02	6.53
	11	447	9.02	6.53
	12	452	9.33	6.53

Figure 8.7.3C

n/a

Figure 8.7.3D

n/a

Table 8.7.3D
Weighted Reliability: Writ 9-12 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	19,245	0.838	0.899
B/C	51,645	0.922	

Table 8.7.3Ei
Accuracy and Consistency of Classification Indices: Writ (Grade 9) S400 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)	
	0.591	0.527	0.368	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.766	0.651	
	2	0.747	0.629	
	3	0.694	0.558	
	4	0.508	0.493	
	5	-	0.461	
	6	-	0.019	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.975	0.016	0.010	0.962
	2/3	0.944	0.020	0.036	0.924
	3/4	0.911	0.026	0.063	0.872
	4/5	0.760	0.240	0.000	0.752
	5/6	0.992	0.008	0.000	0.990

Table 8.7.3Eii
Accuracy and Consistency of Classification Indices: Writ (Grade 10) S400 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)	
	0.696	0.586	0.391	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.887	0.815	
	2	0.609	0.471	
	3	0.759	0.570	
	4	0.661	0.634	
	5	-	0.198	
	6	-	0.000	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.980	0.009	0.011	0.971
	2/3	0.961	0.019	0.020	0.945
	3/4	0.863	0.032	0.105	0.801
	4/5	0.891	0.109	0.000	0.845
	5/6	0.998	0.002	0.000	0.997

Table 8.7.3Eiii
Accuracy and Consistency of Classification Indices: Writ (Grade 11) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.689	0.581		0.366	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.860		0.785	
	2	0.708		0.579	
	3	0.715		0.492	
	4	0.656		0.631	
	5	-		0.102	
	6	-		0.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.980	0.011	0.009	0.971
	2/3	0.961	0.014	0.024	0.946
	3/4	0.807	0.036	0.156	0.724
	4/5	0.939	0.061	0.000	0.916
	5/6	0.999	0.001	0.000	0.999

Table 8.7.3Eiv
Accuracy and Consistency of Classification Indices: Writ (Grade 12) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.652	0.558		0.307	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.894		0.837	
	2	0.680		0.544	
	3	0.679		0.482	
	4	0.609		0.583	
	5	-		0.039	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.983	0.009	0.008	0.976
	2/3	0.970	0.011	0.019	0.957
	3/4	0.724	0.051	0.226	0.641
	4/5	0.974	0.026	0.000	0.969

8.7.4 Speaking 9–12

Figure 8.7.4A
Scale Scores: Spek 9-12 S400 Paper

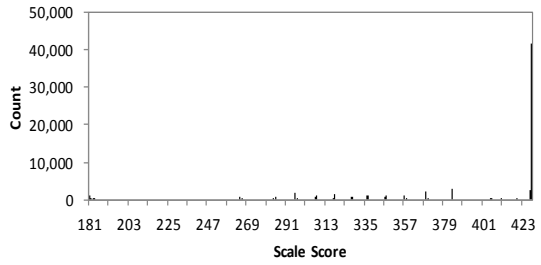


Figure 8.7.4B
Proficiency Level: Spek 9-12 S400 Paper

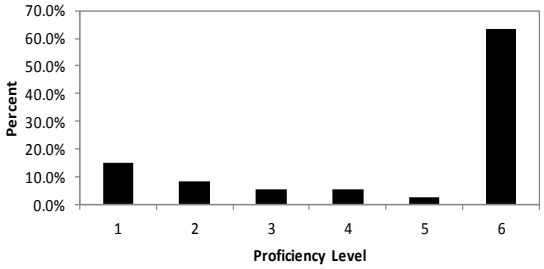


Table 8.7.4A
Scale Score Descriptive Statistics: Spek 9-12 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	25,047	181	428	381.29	70.00
10	19,133	182	428	389.66	58.68
11	14,879	183	428	396.49	51.95
12	10,301	184	428	405.27	44.64
Total	69,360	181	428	390.42	60.54

Table 8.7.4B
Proficiency Level Distribution: Spek 9-12 S400 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	5,929	23.7%	2,840	14.8%	1,350	9.1%	351	3.4%	10,470	15.1%
2	1,042	4.2%	2,047	10.7%	1,624	10.9%	1,085	10.5%	5,798	8.4%
3	534	2.1%	1,244	6.5%	1,158	7.8%	754	7.3%	3,690	5.3%
4	1,705	6.8%	793	4.1%	642	4.3%	491	4.8%	3,631	5.2%
5	622	2.5%	499	2.6%	339	2.3%	206	2.0%	1,666	2.4%
6	15,215	60.7%	11,710	61.2%	9,766	65.6%	7,414	72.0%	44,105	63.6%
Total	25,047	100.0%	19,133	100.0%	14,879	100.0%	10,301	100.0%	69,360	100.0%

Table 8.7.4C

Conditional Standard Error of Measurement at Cut Scores: Spek 9-12 S400 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	9	319	20.92	20.92
	10	321	20.92	20.92
	11	322	20.92	20.92
	12	323	21.43	21.43
2/3	9	347	22.45	22.45
	10	351	22.96	22.96
	11	354	23.47	23.47
	12	357	23.47	23.47
3/4	9	366	24.49	24.49
	10	371	24.49	24.49
	11	377	25.00	25.00
	12	384	26.02	26.02
4/5	9	388	26.53	26.53
	10	393	27.04	27.04
	11	399	28.06	28.06
	12	405	29.08	29.08
5/6	9	407	29.59	29.59
	10	412	30.61	30.61
	11	416	31.12	31.12
	12	421	32.14	32.14

Figure 8.7.4C

n/a

Figure 8.7.4D

n/a

Table 8.7.4D
Weighted Reliability: Spek 9-12 S400 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	18,903	0.877	0.916
B/C	50,457	0.931	

Table 8.7.4Ei
Accuracy and Consistency of Classification Indices: Spek (Grade 9) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.813	0.773		0.612	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.922		0.882	
	2	0.270		0.195	
	3	0.137		0.095	
	4	0.396		0.274	
	5	0.113		0.077	
	6	0.962		0.936	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.953	0.018	0.030	0.934
	2/3	0.955	0.019	0.026	0.934
	3/4	0.953	0.026	0.021	0.931
	4/5	0.945	0.020	0.035	0.920
	5/6	0.933	0.044	0.022	0.900

Table 8.7.4Eii
Accuracy and Consistency of Classification Indices: Spek (Grade 10) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.773	0.723		0.547	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.800		0.731	
	2	0.526		0.433	
	3	0.326		0.258	
	4	0.240		0.163	
	5	0.136		0.084	
	6	0.976		0.951	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.947	0.032	0.022	0.927
	2/3	0.932	0.024	0.044	0.913
	3/4	0.942	0.015	0.043	0.919
	4/5	0.959	0.016	0.025	0.931
	5/6	0.937	0.049	0.014	0.904

Table 8.7.4Eiii
Accuracy and Consistency of Classification Indices: Spek (Grade 11) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.778	0.726		0.522	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.727		0.635	
	2	0.585		0.489	
	3	0.400		0.325	
	4	0.261		0.177	
	5	0.121		0.073	
	6	0.980		0.959	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.957	0.029	0.015	0.940
	2/3	0.936	0.024	0.040	0.919
	3/4	0.946	0.013	0.041	0.924
	4/5	0.962	0.015	0.023	0.934
	5/6	0.935	0.052	0.012	0.902

Table 8.7.4Eiv
Accuracy and Consistency of Classification Indices: Spek (Grade 12) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.788	0.733		0.478	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.558		0.439	
	2	0.690		0.591	
	3	0.416		0.345	
	4	0.319		0.213	
	5	0.096		0.054	
	6	0.984		0.965	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.972	0.022	0.006	0.959
	2/3	0.943	0.017	0.039	0.930
	3/4	0.954	0.012	0.035	0.934
	4/5	0.963	0.012	0.024	0.936
	5/6	0.926	0.064	0.011	0.889

8.7.5 Oral Language Composite 9–12

Figure 8.7.5A
Scale Scores: Oral 9-12 S400 Paper

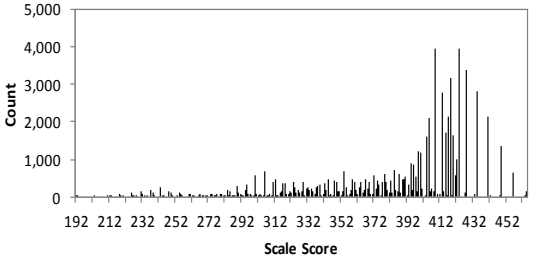


Figure 8.7.5B
Proficiency Level: Oral 9-12 S400 Paper

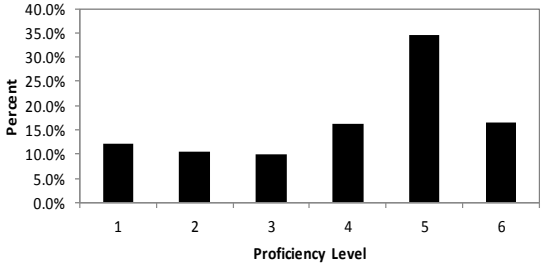


Table 8.7.5A
Scale Score Descriptive Statistics: Oral 9-12 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	24,720	192	464	379.46	54.78
10	18,958	193	464	385.94	47.36
11	14,708	212	464	391.49	43.22
12	10,182	213	464	398.68	37.37
Total	68,568	192	464	386.69	48.53

Table 8.7.5B
Proficiency Level Distribution: Oral 9-12 S400 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,954	16.0%	2,268	12.0%	1,446	9.8%	672	6.6%	8,340	12.2%
2	2,596	10.5%	2,278	12.0%	1,548	10.5%	846	8.3%	7,268	10.6%
3	1,978	8.0%	1,960	10.3%	1,673	11.4%	1,279	12.6%	6,890	10.0%
4	2,704	10.9%	2,816	14.9%	2,848	19.4%	2,676	26.3%	11,044	16.1%
5	8,660	35.0%	6,750	35.6%	4,739	32.2%	3,494	34.3%	23,643	34.5%
6	4,828	19.5%	2,886	15.2%	2,454	16.7%	1,215	11.9%	11,383	16.6%
Total	24,720	100.0%	18,958	100.0%	14,708	100.0%	10,182	100.0%	68,568	100.0%

Table 8.7.5C

n/a

Figure 8.7.5C

n/a

Figure 8.7.5D

n/a

Table 8.7.5D
Oral Composite Reliability: Oral 9-12 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.50	2204.334	0.631
Speaking	0.50	3640.630	0.916
Oral		2368.980	0.882

*Variances from students who had results in all four domains

Table 8.7.5Ei
Accuracy and Consistency of Classification Indices: Oral (Grade 9) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.539	0.449		0.313	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.869		0.801	
	2	0.555		0.430	
	3	0.369		0.252	
	4	0.314		0.201	
	5	0.575		0.521	
	6	0.442		0.371	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.955	0.020	0.025	0.937
	2/3	0.944	0.021	0.035	0.921
	3/4	0.933	0.027	0.040	0.899
	4/5	0.880	0.056	0.063	0.818
	5/6	0.776	0.086	0.138	0.740

Table 8.7.5Eii
Accuracy and Consistency of Classification Indices: Oral (Grade 10) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.552	0.447		0.316	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.846		0.755	
	2	0.595		0.471	
	3	0.420		0.302	
	4	0.393		0.281	
	5	0.628		0.555	
	6	0.451		0.351	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.959	0.017	0.024	0.942
	2/3	0.938	0.023	0.039	0.913
	3/4	0.925	0.033	0.043	0.888
	4/5	0.869	0.072	0.059	0.814
	5/6	0.829	0.065	0.105	0.784

Table 8.7.5Eiii
Accuracy and Consistency of Classification Indices: Oral (Grade 11) S400 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.554	0.448	0.314
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.853	0.759
	2	0.597	0.469
	3	0.473	0.348
	4	0.484	0.353
	5	0.558	0.500
	6	0.487	0.373

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.967	0.014	0.020	0.953
	2/3	0.944	0.022	0.034	0.922
	3/4	0.924	0.032	0.044	0.889
	4/5	0.864	0.066	0.071	0.805
	5/6	0.829	0.095	0.076	0.783

Table 8.7.5Eiv
Accuracy and Consistency of Classification Indices: Oral (Grade 12) S400 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.583	0.452	0.294
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.840	0.740
	2	0.620	0.484
	3	0.552	0.417
	4	0.566	0.412
	5	0.561	0.519
	6	-	0.242

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.978	0.010	0.012	0.967
	2/3	0.956	0.018	0.027	0.938
	3/4	0.926	0.030	0.044	0.894
	4/5	0.831	0.056	0.113	0.760
	5/6	0.881	0.119	0.000	0.809

8.7.6 Literacy Composite 9–12

Figure 8.7.6A
Scale Scores: Litr 9-12 S400 Paper

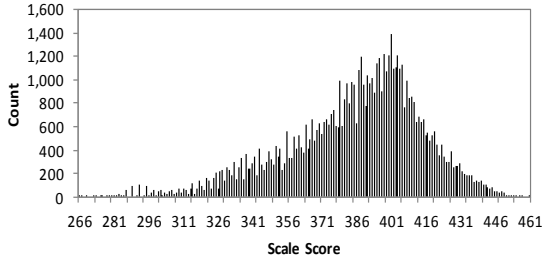


Figure 8.7.6B
Proficiency Level: Litr 9-12 S400 Paper

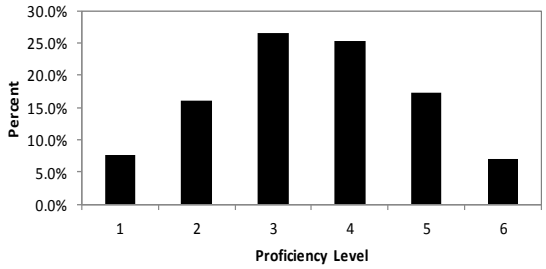


Table 8.7.6A
Scale Score Descriptive Statistics: Litr 9-12 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	24,028	266	454	380.69	30.49
10	18,436	269	457	384.49	29.50
11	14,448	266	461	389.19	28.23
12	10,016	281	458	392.60	26.32
Total	66,928	266	461	385.35	29.46

Table 8.7.6B
Proficiency Level Distribution: Litr 9-12 S400 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,847	7.7%	1,471	8.0%	1,064	7.4%	677	6.8%	5,059	7.6%
2	3,849	16.0%	3,051	16.5%	2,274	15.7%	1,648	16.5%	10,822	16.2%
3	5,891	24.5%	4,825	26.2%	4,068	28.2%	2,931	29.3%	17,715	26.5%
4	5,669	23.6%	4,733	25.7%	3,787	26.2%	2,809	28.0%	16,998	25.4%
5	4,890	20.4%	3,109	16.9%	2,194	15.2%	1,378	13.8%	11,571	17.3%
6	1,882	7.8%	1,247	6.8%	1,061	7.3%	573	5.7%	4,763	7.1%
Total	24,028	100.0%	18,436	100.0%	14,448	100.0%	10,016	100.0%	66,928	100.0%

Table 8.7.6C

n/a

Figure 8.7.6C

n/a

Figure 8.7.6D

n/a

Table 8.7.6D
Literacy Composite Reliability: Litr 9-12 S400 Paper

Component	Weight	Variance	Reliability
Reading	0.50	999.594	0.800
Writing	0.50	1080.746	0.899
Literacy		858.862	0.910

*Variances from students who had results in all four domains

Table 8.7.6Ei
Accuracy and Consistency of Classification Indices: Litr (Grade 9) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.658	0.546		0.436	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.848		0.739	
	2	0.717		0.604	
	3	0.664		0.555	
	4	0.603		0.490	
	5	0.604		0.503	
	6	0.674		0.477	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.972	0.011	0.018	0.960
	2/3	0.937	0.027	0.035	0.911
	3/4	0.902	0.052	0.046	0.863
	4/5	0.901	0.048	0.051	0.861
	5/6	0.941	0.041	0.018	0.918

Table 8.7.6Eii
Accuracy and Consistency of Classification Indices: Litr (Grade 10) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.657	0.548		0.435	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.852		0.746	
	2	0.710		0.596	
	3	0.672		0.565	
	4	0.636		0.523	
	5	0.550		0.457	
	6	0.704		0.474	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.971	0.011	0.018	0.959
	2/3	0.934	0.031	0.035	0.906
	3/4	0.896	0.054	0.050	0.856
	4/5	0.907	0.040	0.053	0.867
	5/6	0.944	0.047	0.009	0.930

Table 8.7.6Eiii
Accuracy and Consistency of Classification Indices: Litr (Grade 11) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.651	0.546		0.430	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.849		0.743	
	2	0.701		0.581	
	3	0.697		0.592	
	4	0.640		0.526	
	5	0.500		0.421	
	6	0.737		0.497	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.974	0.010	0.016	0.962
	2/3	0.934	0.033	0.034	0.905
	3/4	0.893	0.051	0.056	0.853
	4/5	0.910	0.037	0.052	0.870
	5/6	0.935	0.060	0.005	0.926

Table 8.7.6Eiv
Accuracy and Consistency of Classification Indices: Litr (Grade 12) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.656	0.552		0.428	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.836		0.728	
	2	0.723		0.605	
	3	0.713		0.608	
	4	0.652		0.536	
	5	0.477		0.405	
	6	-		0.374	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.976	0.011	0.014	0.965
	2/3	0.933	0.032	0.034	0.905
	3/4	0.894	0.047	0.059	0.853
	4/5	0.906	0.038	0.055	0.863
	5/6	0.943	0.057	0.000	0.937

8.7.7 Comprehension Composite 9–12

Figure 8.7.7A
Scale Scores: Cphn 9-12 S400 Paper

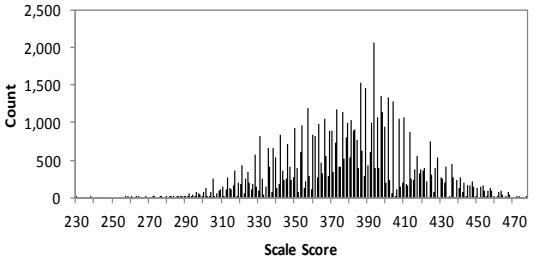


Figure 8.7.7B
Proficiency Level: Cphn 9-12 S400 Paper

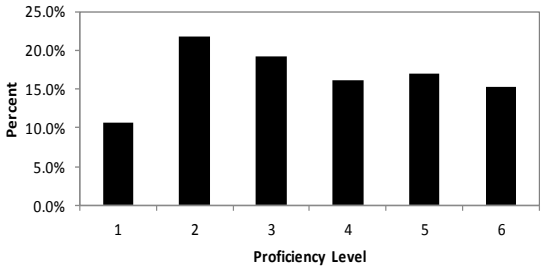


Table 8.7.7A
Scale Score Descriptive Statistics: Cphn 9-12 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	23,765	238	477	374.51	35.03
10	18,282	260	477	378.93	33.71
11	14,303	276	477	383.96	32.78
12	9,914	230	477	387.55	30.81
Total	66,264	230	477	379.72	33.92

Table 8.7.7B
Proficiency Level Distribution: Cphn 9-12 S400 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,798	11.8%	1,867	10.2%	1,436	10.0%	1,006	10.1%	7,107	10.7%
2	4,744	20.0%	4,263	23.3%	3,358	23.5%	2,029	20.5%	14,394	21.7%
3	4,983	21.0%	3,677	20.1%	2,403	16.8%	1,680	16.9%	12,743	19.2%
4	3,577	15.1%	2,806	15.3%	2,511	17.6%	1,795	18.1%	10,689	16.1%
5	4,297	18.1%	3,056	16.7%	2,172	15.2%	1,714	17.3%	11,239	17.0%
6	3,366	14.2%	2,613	14.3%	2,423	16.9%	1,690	17.0%	10,092	15.2%
Total	23,765	100.0%	18,282	100.0%	14,303	100.0%	9,914	100.0%	66,264	100.0%

Table 8.7.7C

n/a

Figure 8.7.7C

n/a

Figure 8.7.7D

n/a

Table 8.7.7D
Comprehension Composite Reliability: Cphn 9-12 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.30	2204.334	0.631
Reading	0.70	999.594	0.800
Comprehension		1149.945	0.851

*Variances from students who had results in all four domains

Table 8.7.7Ei
Accuracy and Consistency of Classification Indices: Cphn (Grade 9) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.562	0.459		0.345	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.824		0.676	
	2	0.591		0.475	
	3	0.487		0.385	
	4	0.363		0.283	
	5	0.503		0.395	
	6	0.762		0.611	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.942	0.016	0.042	0.920
	2/3	0.897	0.050	0.053	0.852
	3/4	0.869	0.076	0.054	0.821
	4/5	0.875	0.071	0.054	0.831
	5/6	0.927	0.042	0.031	0.893

Table 8.7.7Eii
Accuracy and Consistency of Classification Indices: Cphn (Grade 10) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.564	0.460		0.343	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.777		0.605	
	2	0.624		0.514	
	3	0.465		0.365	
	4	0.377		0.292	
	5	0.497		0.384	
	6	0.784		0.638	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.942	0.018	0.040	0.919
	2/3	0.887	0.057	0.056	0.840
	3/4	0.871	0.071	0.058	0.822
	4/5	0.885	0.065	0.050	0.841
	5/6	0.930	0.042	0.028	0.899

Table 8.7.7Eiii
Accuracy and Consistency of Classification Indices: Cphn (Grade 11) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.558	0.458		0.342	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.764		0.590	
	2	0.627		0.513	
	3	0.395		0.307	
	4	0.419		0.327	
	5	0.440		0.336	
	6	0.809		0.674	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.942	0.019	0.038	0.918
	2/3	0.885	0.057	0.058	0.837
	3/4	0.867	0.075	0.059	0.819
	4/5	0.883	0.060	0.057	0.837
	5/6	0.925	0.046	0.029	0.891

Table 8.7.7Eiv
Accuracy and Consistency of Classification Indices: Cphn (Grade 12) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.558	0.455		0.341	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.752		0.579	
	2	0.582		0.469	
	3	0.412		0.318	
	4	0.430		0.336	
	5	0.497		0.384	
	6	0.805		0.675	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.940	0.020	0.040	0.915
	2/3	0.888	0.055	0.056	0.843
	3/4	0.872	0.068	0.059	0.824
	4/5	0.884	0.065	0.051	0.839
	5/6	0.925	0.044	0.031	0.892

8.7.8 Overall Composite 9–12

Figure 8.7.8A
Scale Scores: Over 9-12 S400 Paper

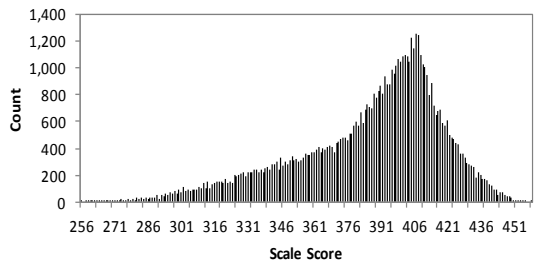


Figure 8.7.8B
Proficiency Level: Over 9-12 S400 Paper

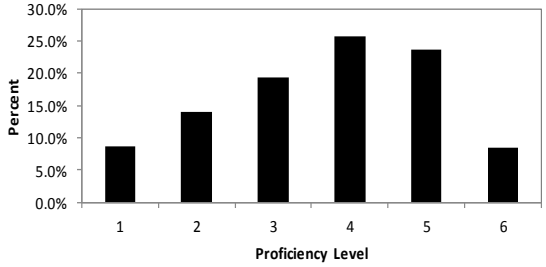


Table 8.7.8A
Scale Score Descriptive Statistics: Over 9-12 S400 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	23,397	256	454	380.24	35.99
10	17,967	260	455	384.80	33.02
11	14,036	260	459	389.82	30.83
12	9,734	268	459	394.36	27.64
Total	65,134	256	459	385.67	33.33

Table 8.7.8B
Proficiency Level Distribution: Over 9-12 S400 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,494	10.7%	1,503	8.4%	1,044	7.4%	604	6.2%	5,645	8.7%
2	3,260	13.9%	2,779	15.5%	1,979	14.1%	1,102	11.3%	9,120	14.0%
3	3,752	16.0%	3,439	19.1%	2,969	21.2%	2,391	24.6%	12,551	19.3%
4	5,194	22.2%	4,512	25.1%	3,889	27.7%	3,203	32.9%	16,798	25.8%
5	6,340	27.1%	4,247	23.6%	3,047	21.7%	1,842	18.9%	15,476	23.8%
6	2,357	10.1%	1,487	8.3%	1,108	7.9%	592	6.1%	5,544	8.5%
Total	23,397	100.0%	17,967	100.0%	14,036	100.0%	9,734	100.0%	65,134	100.0%

Table 8.7.8C

n/a

Figure 8.7.8C

n/a

Figure 8.7.8D

n/a

Table 8.7.8D

Overall Composite Reliability: Over 9-12 S400 Paper

Component	Weight	Variance	Reliability
Listening	0.15	2204.334	0.631
Reading	0.35	999.594	0.800
Speaking	0.15	3640.630	0.916
Writing	0.35	1080.746	0.899
Overall Composite		1110.741	0.943

*Variances from students who had results in all four domains

Table 8.7.8Ei

Accuracy and Consistency of Classification Indices: Over (Grade 9) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.694	0.587		0.492	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.896		0.831	
	2	0.752		0.652	
	3	0.644		0.525	
	4	0.661		0.543	
	5	0.678		0.588	
	6	0.606		0.457	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.973	0.010	0.016	0.963
	2/3	0.957	0.018	0.025	0.939
	3/4	0.933	0.038	0.030	0.905
	4/5	0.909	0.042	0.048	0.874
	5/6	0.919	0.044	0.037	0.890

Table 8.7.8Eii

Accuracy and Consistency of Classification Indices: Over (Grade 10) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.708	0.604		0.510	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.878		0.800	
	2	0.777		0.683	
	3	0.676		0.564	
	4	0.701		0.590	
	5	0.659		0.575	
	6	0.660		0.481	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.977	0.010	0.014	0.967
	2/3	0.953	0.020	0.027	0.934
	3/4	0.927	0.040	0.033	0.897
	4/5	0.915	0.038	0.047	0.880
	5/6	0.935	0.046	0.019	0.914

Table 8.7.8Eiii
Accuracy and Consistency of Classification Indices: Over (Grade 11) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.703	0.607		0.508	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.876		0.798	
	2	0.766		0.666	
	3	0.707		0.599	
	4	0.727		0.621	
	5	0.609		0.545	
	6	0.671		0.466	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.979	0.009	0.012	0.970
	2/3	0.954	0.021	0.025	0.935
	3/4	0.924	0.039	0.037	0.893
	4/5	0.916	0.034	0.050	0.881
	5/6	0.929	0.063	0.008	0.917

Table 8.7.8Eiv
Accuracy and Consistency of Classification Indices: Over (Grade 12) S400 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.714	0.623		0.513	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.886		0.814	
	2	0.745		0.638	
	3	0.765		0.666	
	4	0.763		0.659	
	5	0.567		0.508	
	6	-		0.368	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.984	0.007	0.009	0.977
	2/3	0.960	0.020	0.020	0.942
	3/4	0.922	0.037	0.041	0.890
	4/5	0.909	0.031	0.060	0.870
	5/6	0.939	0.061	0.000	0.934

References

Allen, N. L., Carlson, J. E., & Zalanak, C. A. (1999). *The NAEP 1996 technical report.* Washington, DC: National Center for Education Statistics.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.

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. (2004a). Linking with equivalent group or single group design (LEGS) (version 2.0). [Computer Software]. University of Iowa: Center for Advanced Studies in Measurement and Assessment (CASMA).

Brennan, R. L. (2004b). 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*. Madison, WI: WIDA Consortium.

Center for Applied Linguistics (2016). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test, Series 303, 2014–2015 Administration* (WIDA Consortium Annual Technical Report No. 11).

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.

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

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

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).

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.

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).

Kolen, M. J., & Brennan, R. L. (2004). *Test Equating: Methods and Practices* (2nd ed.). New York, NY: Springer-Verlag.

Lee, W., Hanson, B. A., & Brennan, R. L. (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement, 26,* 412–432.

Linacre, J.M. (2002, Autumn). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878. 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

Linacre, J. M. (n.d). Displacement measures. Retrieved from http://www.winsteps.com/winman/displacement.htm.

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).

Mantel, N. & Haenszel, W. (1959). Statistical aspect of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*, 719-748.

Meyer, J. P. (2014). jMetrik. [Computer Software]. http://www.jmetrik.com.

Mislevy, R. J. (1992). *Linking Educational Assessments: Concepts, Issues, Methods, and Prospects*. Princeton, NJ: Educational Testing Service, Policy Information Center.

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.

Pommerich, M. (2007). Concordance: The good, the bad, and the ugly. In N. J. Dorans, M. Pommerich, & P. W. Holland (Ed.), *Linking and Aligning Scores and Scales* (pp. 199-230). New York, NY: Springer-Verlag.

Pommerich, M. Hanson, B. A., Harris, D. J., & Sconing, J. A. (2004). Issues in conducting linkages between distinct tests. *Applied Psychological Measurement, 28(4),* 247–273.

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

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., and Kenyon, D. M. (2013). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test, Series 203, 2011–2012 Administration* (WIDA Consortium Annual Technical Report No. 8).

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

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

Young, M. J., & Yoon, B. (1998, April). *Estimating the consistency and accuracy of classifications in a standards-referenced assessment* (CSE Tech. Rep. 475). Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education and Information Studies.

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. Holland & H. Wainer (Eds.), *Differential Item Functioning* (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.
Fikru Hirpa, M.A.
Min Feng, B.S.
Justin Kelly, Ph.D.
Daniel Lee, M.S.
Chi-Kai (Cary) Lin, Ph.D.
Mohammed Louguit, Ph.D.
Michele Kawood, M.S.Ed.
Dorry M. Kenyon, Ph.D.
Meg Montee, M.A.T
Jennifer Norton, Ed.D.
Jennifer Renn, Ph.D.
Shu Jing Yen, Ph.D.
Jing Wei, Ph.D.
Xin Yu, M.A.
Yu-Chia Wu, M.Ed.