World-Class Instructional Design and Assessment

Annual Technical Report for ACCESS for ELLs® 2.0 Paper English Language Proficiency Test, Series 401, 2016–2017 Administration

Annual Technical Report No. 13B

Prepared by:

Center for Applied Linguistics

Language Assessment Division
Psychometrics and Quantitative Research Team

April 25, 2018

© 2018 Board of Regents of the University of Wisconsin System on behalf of the WIDA Consortium.

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., Professor Emerita, Educational Psychology, University of Illinois at Chicago

Executive Summary

This is the 13th annual technical report on the ACCESS for ELLs® English Language Proficiency Test, and the second 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 401 was administered in school year 2016–17 in 35 states, the District of Columbia, the Commonwealth of the Northern Marianas, and the Virgin Islands of the United States, for a total of 38 state entities (henceforth "states"). ACCESS for ELLs 2.0 Series 401 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 13B) provides technical information pertaining to ACCESS for ELLs 2.0 Series 401 Paper, including the Kindergarten assessment. A second report (WIDA ACCESS Technical Report 13A) provides technical information for the ACCESS for ELLs 2.0 Series 401 Online assessment.

Table 0.1

Participation in ACCESS for ELLs Online and Paper, Series 401

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	7,737	4,795	12,532	1,386	13,918
AL	11,649	5,741	17,390	3,487	20,877
CO	61,768	29,167	90,935	10,836	101,771
DC	5,610	75	5,685	1,089	6,774
DE	9,495	13	9,508	1,637	11,145
FL	0	243,736	243,736	35,774	279,510
GA	73,992	12,164	86,156	17,196	103,352
HI	7,807	3,014	10,821	1,876	12,697
ID	13,367	39	13,406	2,230	15,636
IL	130,134	31,024	161,158	26,670	187,828
IN	41,970	523	42,493	7,405	49,898
KY	19,390	347	19,737	3,377	23,114
MA	46,274	28,823	75,097	10,330	85,427
MD	60,015	122	60,137	10,675	70,812
ME	4,711	247	4,958	485	5,443
MI	84,524	3,774	88,298	10,371	98,669
MN	59,906	597	60,503	8,316	68,819
MO	26,122	69	26,191	4,899	31,090
MP	1,302	0	1,302	78	1,380
MT	2,581	0	2,581	137	2,718
NC	79,468	1,543	81,011	11,957	92,968
ND	2,725	41	2,766	384	3,150
NH	3,594	294	3,888	441	4,329
NJ	60,066	955	61,021	12,035	73,056
NM	38,249	2,796	41,045	4,717	45,762
NV	64,380	39	64,419	7,956	72,375
OK	24,430	13,433	37,863	6,902	44,765
PA	41,074	10,708	51,782	5,017	56,799
RI	8,744	1,117	9,861	1,092	10,953
SC	39,374	1,295	40,669	3,478	44,147
SD	3,653	192	3,845	742	4,587
TN	38,872	15	38,887	5,711	44,598
UT	34,945	6	34,951	4,975	39,926
VA	76,847	11,395	88,242	14,215	102,457
VI	1,023	0	1,023	96	1,119
VT	1,295	13	1,308	178	1,486
WI	42,080	180	42,260	5,531	47,791
WY	2,186	60	2,246	386	2,632
Total	1,231,359	408,352	1,639,711	244,067	1,883,778

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 401 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.

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 401 Paper data set for analyses included the results of 652,419 students. The largest grade was Kindergarten with 244,067 students, while the smallest was Grade 12 with 9,925 students. Of the participating WIDA states, the largest was Florida with 279,510 students, while the smallest was Northern Mariana Islands with 78 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 401 Paper, presented in Chapter 8 Table D, is very high across all grade-level clusters. For Kindergarten it was .973; for Grade 1, .939; for Grade 2, .952, for Grade 3, .940, for Grades 4–5, .945; for Grades 6–8, .949; and for Grades 9–12, .951. 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	0.952	0.954	0.963	0.982	N/A
1	0.970	0.929	0.949	0.990	0.999
2	0.984	0.956	0.920	0.965	0.999
3	0.985	0.963	0.908	0.948	0.996
4	0.980	0.963	0.915	0.927	0.992
5	0.981	0.965	0.921	0.906	0.993
6	0.978	0.954	0.921	0.972	0.999
7	0.973	0.950	0.923	0.969	0.998
8	0.970	0.950	0.924	0.956	0.998
9	0.967	0.950	0.931	0.954	0.998
10	0.966	0.945	0.933	0.968	0.999
11	0.967	0.943	0.935	0.976	0.999
12	0.974	0.944	0.937	0.982	1.000

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 thirteenth 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 401 Paper (2016–2017) 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)

Series 401 Paper: Special Considerations

Data Exclusion: State of Michigan

Data for the production of the Annual Technical Report were received by CAL in late September of 2017. Data for the state of Michigan were not available in the system at the time of the initial data pull. Michigan data were received in an additional data draw in mid-November. In order to allow for the timely production of the report, analyses which pertain to the technical properties of test forms (the analyses included in Chapter 6 and Chapter 8) were conducted using the original September data. These analyses do not include data from the state of Michigan. Students from the state of Michigan constitute 14,145 of 652,419 total students in the ACCESS Paper population, a proportion of 2.17%. Students from the state of Michigan are included in summary tables which pertain to the counts of students participating in the assessment (the tables in Chapter 4).

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

1.1 Purpose of ACCESS for ELLs 1

1.2 Format of ACCESS 2.0 Paper 2

1.2.1 Integration with the Standards 2

1.2.2 Grade-Level Clusters 2

1.2.3 Language Domains 3

1.2.4 Language Proficiency Levels 3

1.2.5 Tiers 5

1.3 Test Development 7

1.3.1 Item Writing and Editing 7

1.3.2 Item Content and Bias and Sensitivity Reviews 7

1.3.3 Development of Listening and Reading 8

1.3.4 Development of Writing and Speaking 8

1.3.5 Development of Kindergarten Test 9

1.3.6 Reporting Scale 10

1.3.7 Standard Setting 11

1.4 Reporting of Results 11

1.4.1 Scale Scores 11

1.4.2 Language Proficiency Level Scores 13

1.5 Test Administration 17

1.5.1 Test Administrator Training 17

1.5.2 Test Security 17

1.5.3 Test Accommodations 17

1.6 Scoring 18

1.6.1 Listening and Reading 18

1.6.2 Writing 18

1.6.3 Speaking 21

2. An Assessment Use Argument for ACCESS 2.0: Focus on Assessment Records 24

2.1 The Generic Validation Framework for ACCESS 2.0 25

2.2 Focus on Assessment Records 26

2.2.1 Breakdown of Claims for the *Assessment Records* Produced in the ACCESS 2.0 Assessment Program 27

2.3 Evidence for Assessment Records Claims of ACCESS 2.0 29
2.4 Summary of Assessment Records Claims, Actions, and Evidence 37
2.5 Visual Guide to Tables and Figures 39
2.5.1 Guide to Chapter 4, *Student Results* 39
2.5.2. Guide to Chapter 6, *Analyses of Test Forms Results* 41
2.5.3 Guide to Chapter 8, *Analysis Across Tiers Results* 42
3. Descriptions of Student Results 43
3.1 Participation 43
3.1.1 Grade-Level Cluster 43
3.1.2 Grade 43
3.1.3 Tier 43
3.2 Scale Score Results 43
3.2.1 Mean Scale Scores Across Domain and Composite Scores Section 43
3.2.2 Correlations 44
3.3 Proficiency Level Results 44
4 Student Results 46
4.1 Participation 46
4.1.1 Participation by Grade-Level Cluster 46
4.1.2 Participation by Grade 48
4.1.3 Participation by Tier 51
4.2 Scale Score Results 55
4.2.1 Mean Scale Scores by Grade Level Cluster Across Domain and Composite Scores 55
4.2.2 Mean Scale Scores by Grade Across Domain and Composite Scores 61
4.2.3 Correlations Among Scale Scores by Grade Level Cluster 69
4.3 Proficiency Level Results 72
4.3.1 Listening 72
4.3.2 Reading 78
4.3.3 Writing 83
4.3.4 Speaking 88
4.3.5 Oral Composite 93
4.3.6 Literacy Composite 98
4.3.7 Comprehension Composite 103
4.3.8 Overall Composite 108

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 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 Social and Instructional Language (SIL), Language of Language Arts (LoLA), and Language of Social Studies (LoSS): IT
- Language of Math (LoMA) and Language of Science (LoSC): MS
- Language of 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 ACCESS 2.0 Paper and the shared forms across clusters.

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

<table>
<tr><th>ACCESS 2.0 Paper Grade-level Clusters</th><th>Shared Test Forms (Listening and Reading)</th><th>Shared Test Forms (Speaking and Writing)</th><th>Grade</th></tr>
<tr><td>K</td><td>K</td><td>K</td><td>K</td></tr>
<tr><td>1</td><td rowspan="2">Cluster 1 and Cluster 2</td><td>Cluster 1</td><td>1</td></tr>
<tr><td>2</td><td rowspan="2">Cluster 2 and Cluster 3</td><td>2</td></tr>
<tr><td>3</td><td rowspan="2">Cluster 3 and Cluster 4–5</td><td>3</td></tr>
<tr><td rowspan="2">4–5</td><td rowspan="2">Cluster 4–5</td><td>4</td></tr>
<tr><td>5</td></tr>
<tr><td rowspan="3">6–8</td><td rowspan="3">Cluster 6–8</td><td rowspan="3">Cluster 6–8</td><td>6</td></tr>
<tr><td>7</td></tr>
<tr><td>8</td></tr>
<tr><td rowspan="4">9–12</td><td rowspan="4">Cluster 9–12</td><td rowspan="4">Cluster 9–12</td><td>9</td></tr>
<tr><td>10</td></tr>
<tr><td>11</td></tr>
<tr><td>12</td></tr>
</table>

Note that in our analyses of student participation in the assessment (analyses discussed and presented in chapters 3 and 4), analysis is conducted by cluster (K, 1, 2, 3, 4-5, 6-8, 9-12). In our analyses of test forms (chapters 5 and 6), analysis is conducted by test form (i.e. in Listening and Reading, a single analysis is conducted for the cluster 1 and cluster 2 form). These analyses are presented by cluster; if a table of results pertains to more than one cluster, it is repeated in each cluster.

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) (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.

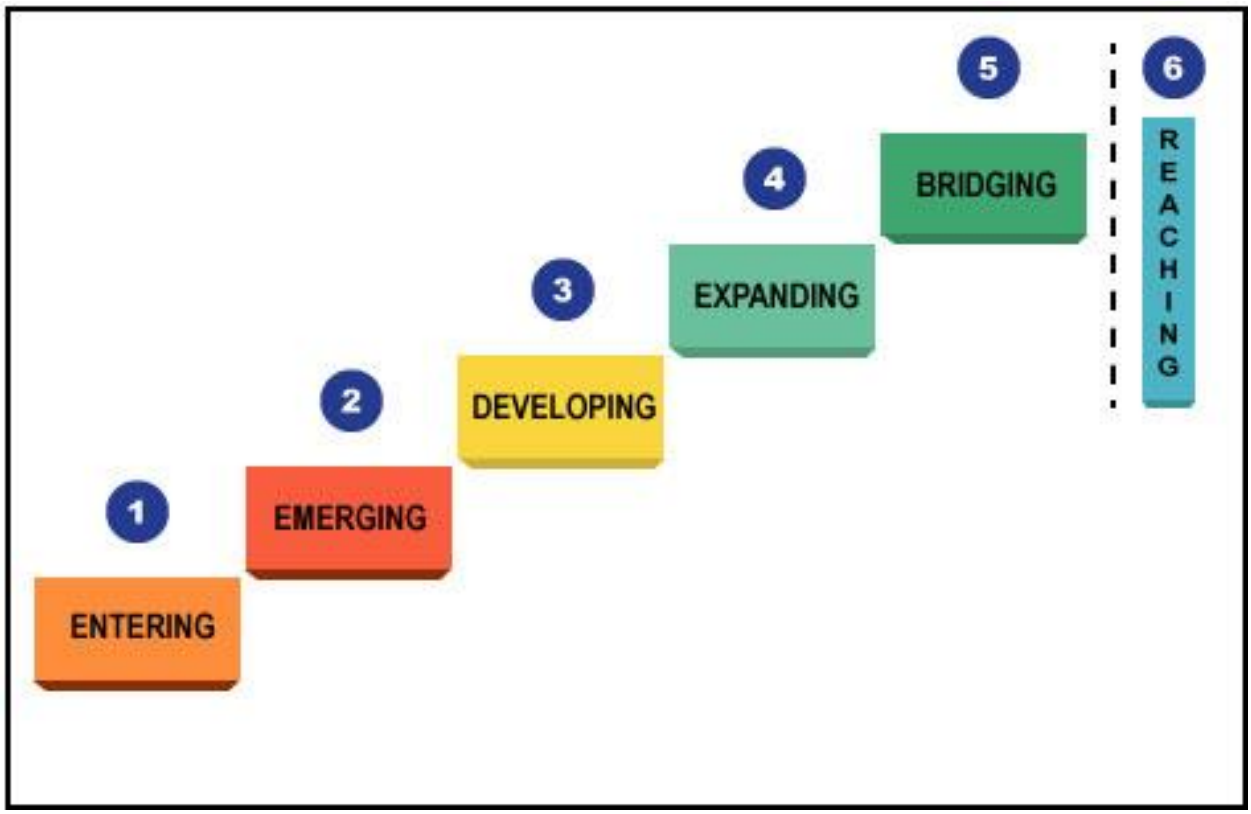


Figure 1.2.4A. The language proficiency levels of the WIDA ELD Standards.

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 dimension; (b) language forms and conventions at the sentence dimension; and (c) linguistic complexity at the discourse dimension. 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**,** by grade-level cluster, across the four 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 domain. *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.[1] 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.

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

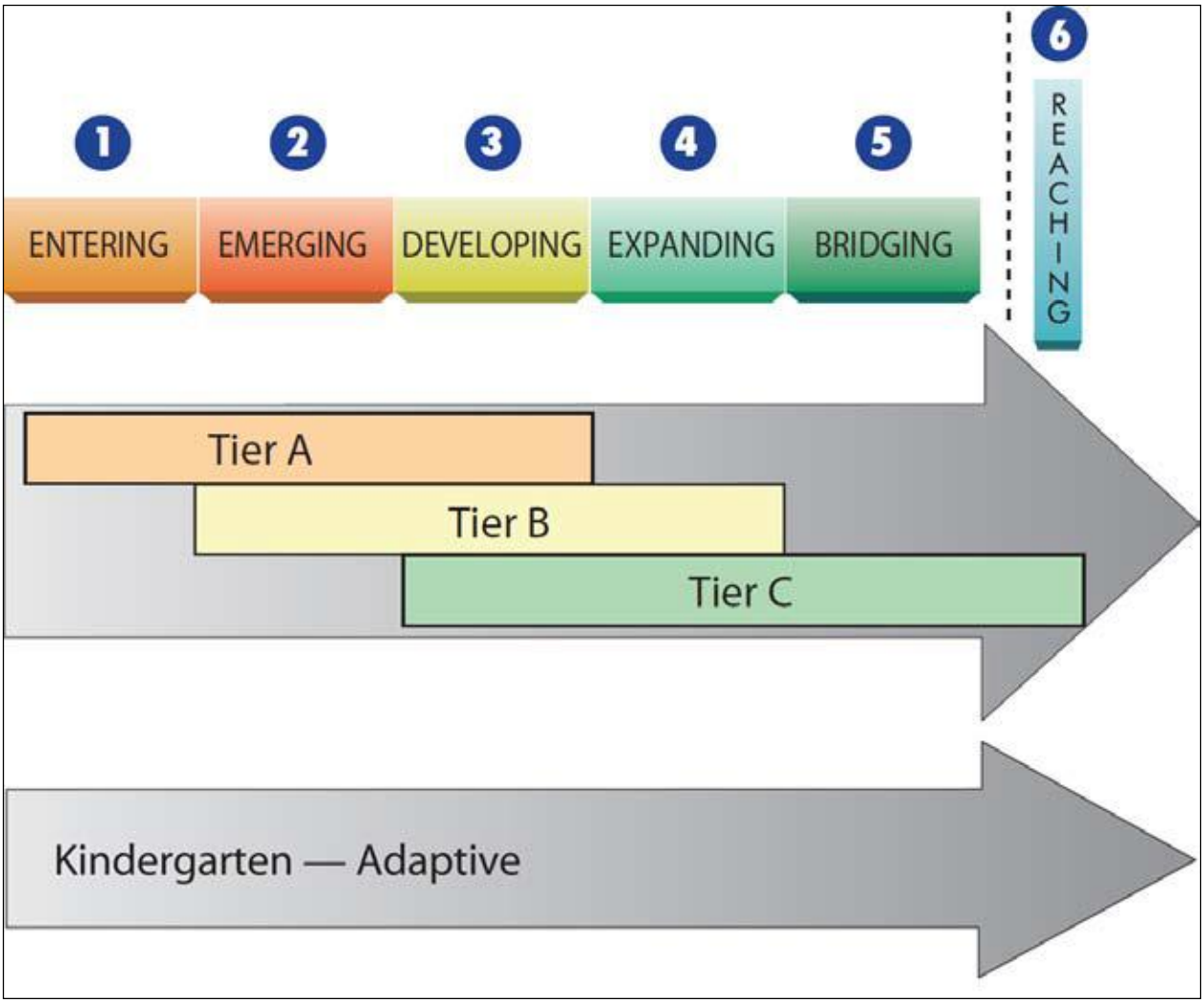


Figure 1.2.5.A. Tier structure of ACCESS for ELLs

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 PL 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 PL 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 401 Paper continues to reflects this transition. The Listening and Reading tests for ACCESS 2.0 Series 401 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 wass 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 advantage 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 401 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* (Kenyon, 2006) and *ACCESS for ELLs Technical Report No. 11* (2016).

The Listening and Reading tests for ACCESS 2.0 Series 401 Paper are composed of the same sets of items, across all grade-level clusters and tiers, as ACCESS Series 303, 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 303 and Series 401 Paper. In the form shared across Clusters 1 and 2 (Tier C), three items from the Series 303 form were substituted with three items from Series 203 to produce the Series 401 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 303 form were substituted with three items from Series 203 to produce the Series 401 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 401 Paper were developed to be shared across the Online and Paper versions of ACCESS 2.0. In other words, the Online and Paper versions of the tests have the same tasks, by grade-level cluster and tier, with minor differences. Writing items were developed using the processes for Online Writing task development (see the companion Annual Technical Report for Online ACCESS for further detail), and then converted in format so they could be used in the Paper test. There are therefore some differences in presentation between the Online and Paper test which result from the mode difference in the domain of Writing. Speaking tasks also have some differences in presentation between online and paper. In addition, the Paper test does not include the Speaking tier pre-A, which is included on the Online test.[2] Second, in order to accommodate the tier structure of Listening and Reading, the Paper test maintains the tier structure of ACCESS for ELLs 1.0, which was provided in three tiers (A, B, and C). 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.

[2] 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 small-scale tryouts, led by CAL staff. 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 stand-alone field test of Writing folders is conducted, with responses scored at CAL, followed by a qualitative analysis of the collected responses. The main purposes of this small-scale field testing are to 1) confirm that the tasks are functioning as intended, 2) identify preliminary exemplars that have potential to be turned into anchors in operational scoring, and 3) 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 using the computer-based test format. 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. After field testing the Speaking tasks are then produced in the paper-based format.

Many of the speaking folders used in Series 401 were previously field tested as part of Series 400 and then further revised and field tested again for Series 401. Much of the content of the Speaking items on Series 401 was adapted for ACCESS 2.0 from both operational items from previous paper-based test series and from materials that were not developed to finality for previous test series. Some folder content was created specifically for ACCESS 2.0.

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 tasks 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 tasks, 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. Scale scores are calculated by transforming the person ability estimate via a scaling equation. The scaling equations for each domain are provided in Chapter 6, Table D. In the domains of Listening and Reading, the ACCESS scale was maintained through the transition from ACCESS 1.0 to ACCESS 2.0 in Series 400, and is continued to Series 401 (evidence for scale maintenance from ACCES 1.0 to ACCESS 2.0 can be found in Center for Applied Linguistics [2016]). In the domains of Writing and Speaking, a study was conducted in the summer of 2016 to reconstruct the scoring scale (see Center for Applied Linguistics [2017]).

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 series of standard-setting studies were conducted, 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* (Kenyon, 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 401 Listening and Reading maintains the same scale as ACCESS. In the domains of Writing and Speaking, a study was conducted in the summer of 2016 to reconstruct the logit scale (see Center for Applied Linguistics [2017]).
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.

1.3.7 Standard Setting

Proficiency Level (PL) scores are interpretations of these scale scores in terms of the PLs described in the WIDA ELD Standards. These interpretations derive from a series of standard setting studies, in which educators reviewed evidence from the test, either in the form of items for the selected response sections (Listening and Reading) or student portfolios for the constructed response sections (Writing and Speaking), to establish cut scores between the PLs. The first standard setting study for ACCESS took place in 2005; it established cut scores for all four domains by grade-level cluster (Kenyon, 2006). The second cut score study took place in 2007; it established cut sores for all four domains by grade level (Kenyon, Ryu, & MacGregor, 2013). These cut scores were used to derive PL scores through Series 400 of ACCESS 2.0 Online. A third cut score study was conducted in summer 2016 (Cook and MacGregor, 2017). The purpose of this study was to re-examine cut scores for each of the PLs on the new ACCESS 2.0 assessment in light of the migration from the paper-and-pencil only assessment, the revision of the Speaking test, and the influence of college- and career-ready standards. Test Series 401 is the first series which employed these newly revised proficiency level cut scores. New cut scores apply to all grades K–12 of the Paper assessment.

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.

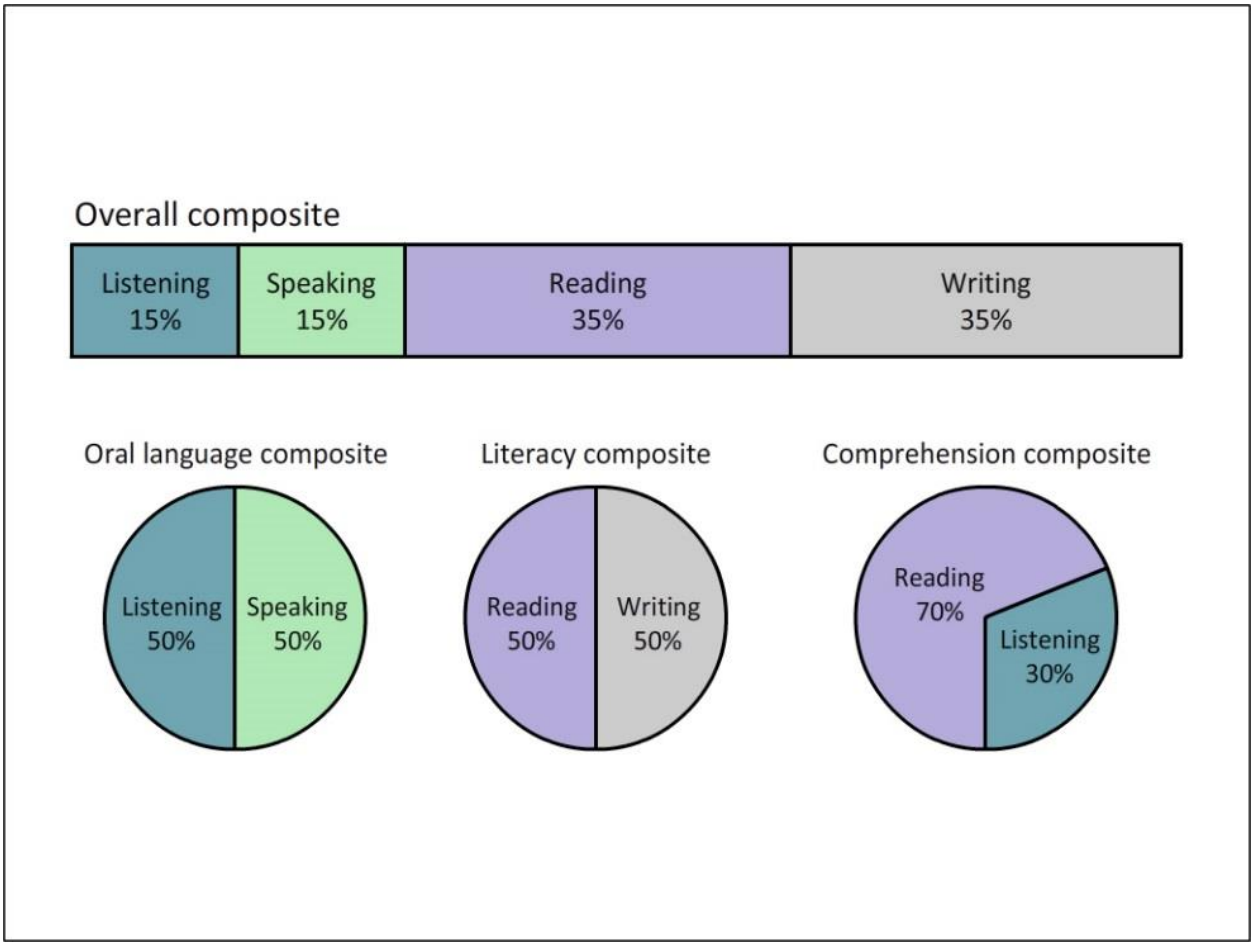


Figure 1.5.1A. Domain Composites

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 401 and beyond, as well as the 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	List	229	251	278	286	308
1	List	236	259	291	303	327
2	List	245	283	314	330	354
3	List	262	300	331	349	374
4	List	275	313	343	363	388
5	List	285	323	354	375	401
6	List	294	332	363	385	411
7	List	302	340	370	394	420
8	List	308	347	377	402	427
9	List	314	353	383	409	434
10	List	325	358	389	415	441
11	List	335	364	394	420	447
12	List	342	368	398	426	452

Table 1.4.2B
Cut Scores (Reading)

Grades	Domain	Cut				
		1/2	2/3	3/4	4/5	5/6
K	Read	241	259	279	289	310
1	Read	264	286	304	315	334
2	Read	283	307	326	337	355
3	Read	297	323	342	352	370
4	Read	307	335	354	364	382
5	Read	316	345	364	373	391
6	Read	323	353	373	382	399
7	Read	329	360	380	389	406
8	Read	335	366	386	395	412
9	Read	340	372	392	401	418
10	Read	344	377	397	406	423
11	Read	348	382	402	410	427
12	Read	352	386	407	414	432

Table 1.4.2C
Cut Scores (Writing)

Grades	Domain	Cut				
		1/2	2/3	3/4	4/5	5/6
K	Writ	234	271	311	367	389
1	Writ	238	275	337	382	405
2	Writ	242	279	341	388	411
3	Writ	247	283	346	394	418
4	Writ	266	288	351	401	425
5	Writ	267	293	356	407	433
6	Writ	268	298	361	413	441
7	Writ	273	305	367	419	450
8	Writ	281	311	372	424	459
9	Writ	289	319	378	430	469
10	Writ	298	326	385	436	479
11	Writ	308	335	391	441	490
12	Writ	318	344	398	447	501

Table 1.4.2D
Cut Scores (Speaking)

Grades	Domain	Cut				
		1/2	2/3	3/4	4/5	5/6
K	Spek	191	250	301	349	392
1	Spek	205	261	311	361	403
2	Spek	220	273	322	374	415
3	Spek	234	283	332	386	425
4	Spek	246	293	342	397	435
5	Spek	258	302	350	407	443
6	Spek	268	310	360	417	451
7	Spek	277	317	369	425	457
8	Spek	284	323	377	433	463
9	Spek	290	328	385	440	468
10	Spek	295	333	393	446	471
11	Spek	299	337	400	451	474
12	Spek	302	340	406	455	476

Table 1.4.2E
Cut Scores (Oral Language Composite)

Grades	Domain	Cut				
		1/2	2/3	3/4	4/5	5/6
K	Oral	210	251	290	318	350
1	Oral	221	260	301	332	365
2	Oral	233	278	318	352	385
3	Oral	248	292	332	368	400
4	Oral	261	303	343	380	412
5	Oral	272	313	352	391	422
6	Oral	281	321	362	401	431
7	Oral	290	329	370	410	439
8	Oral	296	335	377	418	445
9	Oral	302	341	384	425	451
10	Oral	310	346	391	431	456
11	Oral	317	351	397	436	461
12	Oral	322	354	402	441	464

Table 1.4.2F
Cut Scores (Literacy Composite)

Grades	Domain	Cut				
		1/2	2/3	3/4	4/5	5/6
K	Litr	238	265	295	328	350
1	Litr	251	281	321	349	370
2	Litr	263	293	334	363	383
3	Litr	272	303	344	373	394
4	Litr	287	312	353	383	404
5	Litr	292	319	360	390	412
6	Litr	296	326	367	398	420
7	Litr	301	333	374	404	428
8	Litr	308	339	379	410	436
9	Litr	315	346	385	416	444
10	Litr	321	352	391	421	451
11	Litr	328	359	397	426	459
12	Litr	335	365	403	431	467

Table 1.4.2G
Cut Scores (Comprehension Composite)

Grades	Domain	Cut				
		1/2	2/3	3/4	4/5	5/6
K	Cphn	237	257	279	288	309
1	Cphn	256	278	300	311	332
2	Cphn	272	300	322	335	355
3	Cphn	287	316	339	351	371
4	Cphn	297	328	351	364	384
5	Cphn	307	338	361	374	394
6	Cphn	314	347	370	383	403
7	Cphn	321	354	377	391	410
8	Cphn	327	360	383	397	417
9	Cphn	332	366	389	403	423
10	Cphn	338	371	395	409	428
11	Cphn	344	377	400	413	433
12	Cphn	349	381	404	418	438

Table 1.4.2H
Cut Scores (Overall Composite)

Grades	Domain	Cut				
		1/2	2/3	3/4	4/5	5/6
K	Over	229	261	293	325	350
1	Over	242	274	315	344	368
2	Over	254	289	329	359	383
3	Over	265	300	340	371	396
4	Over	279	309	350	382	406
5	Over	286	317	358	390	415
6	Over	291	324	365	399	423
7	Over	298	331	372	406	431
8	Over	304	337	378	412	438
9	Over	311	344	385	418	446
10	Over	318	350	391	424	453
11	Over	325	356	397	429	459
12	Over	331	362	402	434	466

A PL score consists of a two-digit decimal number (e.g., 4.5). The first digit represents the student's overall PL range based on the student's scale score. The number to the right of the decimal is an indication of the proportion of the range between cut scores that the student's scale score represents. A score of 4.5, for example, tells us that the student is in PL4 and that his/her 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 grades from Kindergarten to Grade 12, PL scores are dependent upon which grade a student was in when ACCESS 2.0 Online was administered. Using the cut scores newly in effect for Series 401, if a

Grade 2 student receives a 350 in Listening, it would be interpreted as a PL score of 5.8; if a Grade 5 student receives a 350 in Listening, it would be a 3.8; if a Grade 8 student receives a 350 in Listening, it would be a 3.1; and if a Grade 12 student receives a 350 in Listening, it would be a 2.3.

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

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 401 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 receive any of the three certifications, participants had to complete the relevant online course and 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 printing, 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 401 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[3] that are so significant as to prevent meaningful participation in

[3] Recommendations regarding physical disabilities, such as deafness or blindness, are available on the WIDA website, www.wida.us.

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 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 whole score point (1 to 6) best describes a response and then determines whether the three descriptors for that whole score point suit for that response. If all three descriptors fit, a whole score point should be awarded. If 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-topic responses are scored in their entirety using the Scoring Scale, while partially off-task responses are scored by ignoring the off-task portion and scoring only the on-task portion.

To calculate a raw score for the Writing test, raters' scores for each Writing task are converted to whole numbers ranging from 0–9, 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 0–27. 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 0–9; additionally the last task is weighted as 3. Therefore, the possible final raw scores for Grade 1 Tier A range from 0–40.

On a Tier B or Tier C test, results from the different tasks are given different weights. (Note that for ACCESS 2.0 Series 401 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 (1*5 + 2*6 + 3*7), while a student with raw scores of 7, 6, and 5 on the three tasks would have a total raw score of 34 (1*7 + 2*6 + 3*5). Raw scores on the Tier B and Tier Ctests can range from 0–54.

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+	9
6	9

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 and 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 by scoring at least 70% agreement on a qualifying set before scoring live responses.
- Once raters are qualified, they are further trained for their grade-level cluster on the specific tasks for which they will rate responses.
- Once raters have trained, qualified, and begun live scoring, DRC uses calibration sets to keep the raters calibrated on the actual tasks they are 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.

Calculating Score Agreement for Scoring Monitoring

- For monitoring and review purposes, agreement is defined as two adjacent scores. (See section 3.3.1.3 for a description of the writing scoring scale.) For example, using the writing scoring scale, a score of 2 and 2+ would be considered agreement as would scores of 2 and 2 or scores of 2+ and 3. Scores of 2 and 3 on the writing scoring scale would be considered adjacent and scores of 2 and 3+ would be considered non-adjacent.

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 both re-calibration and validation sets. For each of the first five days that a rater scores a task, he or she takes one re-calibration set of five responses per task. After the raters takes a recalibration set, the Scoring Director or Team Leader reviews it using descriptors from the Writing Scoring Scale and the anchor responses to explain the rationale behind each response's score. Starting on the 6th day of scoring, DRC uses validity sets to monitor rater performance. 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 for the Speaking test and pass the accompanying quiz (either Grades 1–5 or Grades 6–12). The training focuses on developing the test administrators' ability to score the test reliably. Separate training materials are available that address test administration and monitoring procedures. To 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.

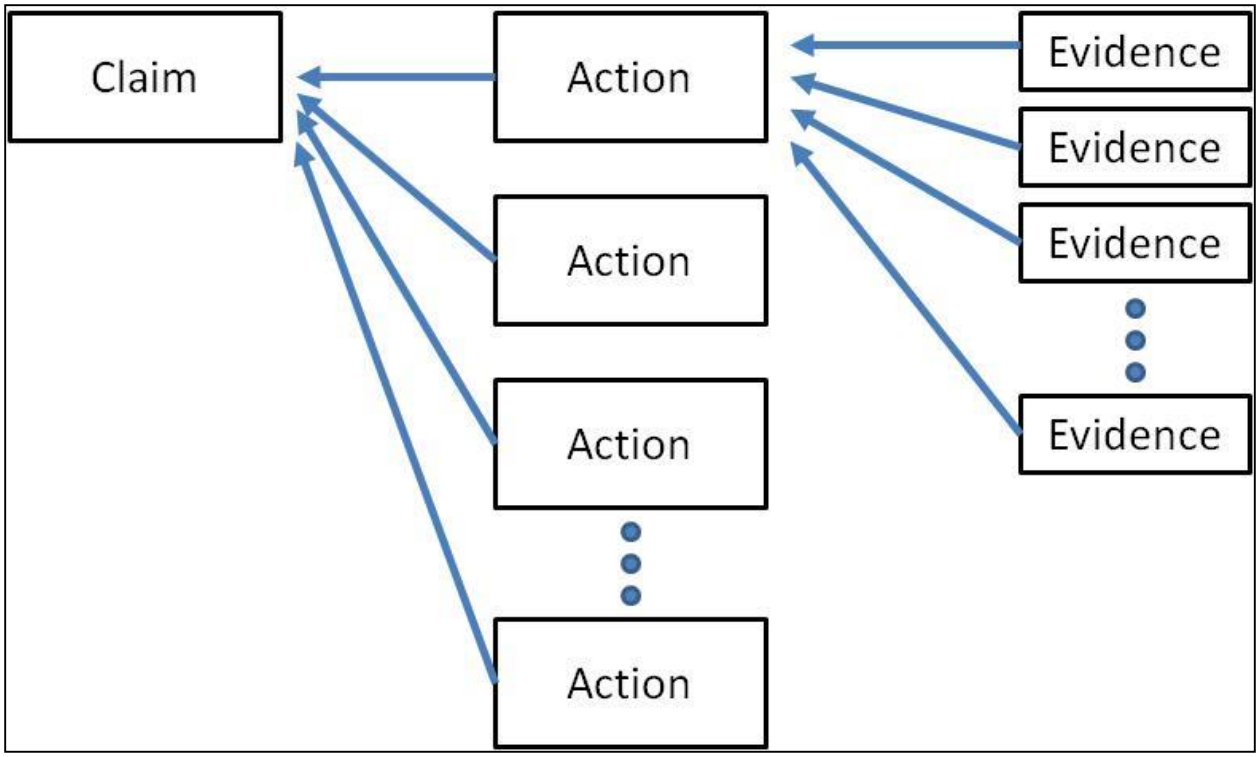


Figure 2A: General Argument Structure for Assessment Validation (simplified from Toulmin, 2003).

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.

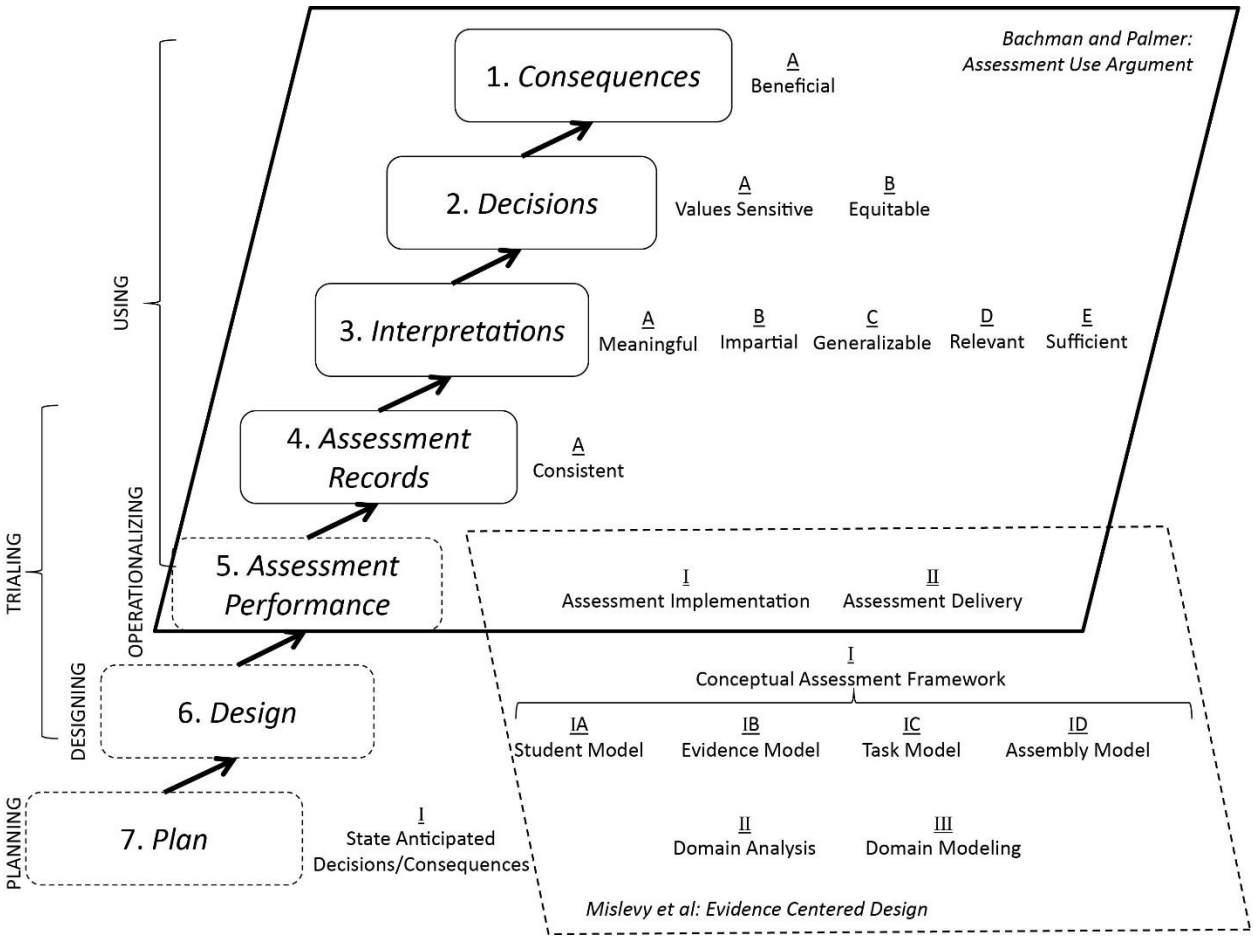


Figure 2.1A: CAL's Validation Framework (based on Bachman & Palmer, 2010; Mislevy, Almond, & Lukas, 2004).

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.

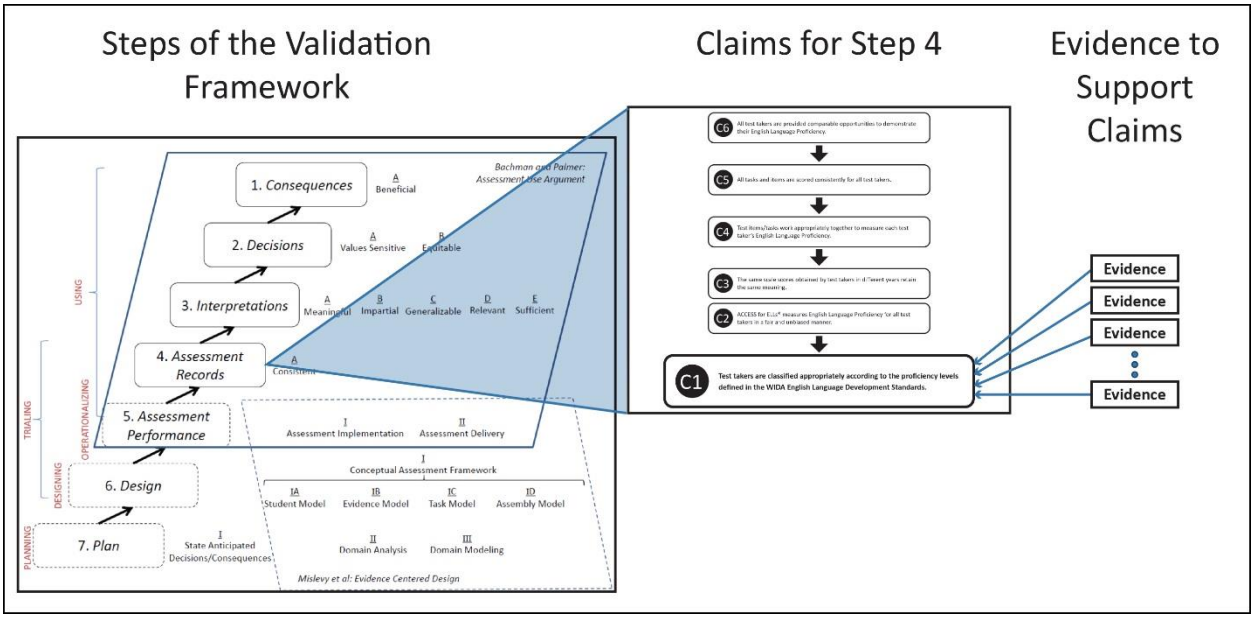


Figure 2.2A: Structure of the Argument-Based Approach Supporting *Assessment Records* (Step 4) contained in this chapter.

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.

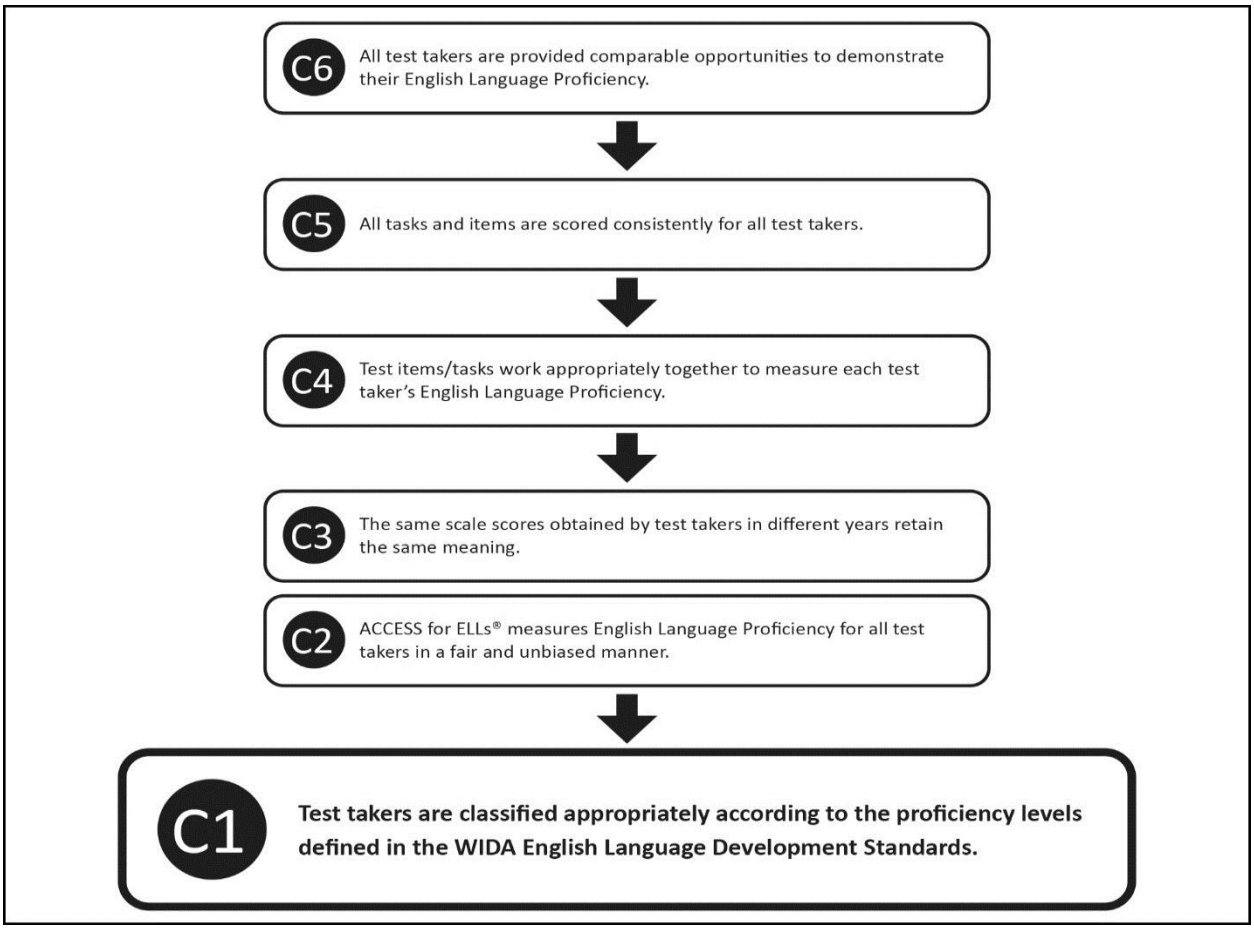


Figure 2.2.1A: Progression of Claims for Step 4: Assessment Records.

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*.

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*. 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*. Both the *District and School Test Coordinator Manual* and the *Test Administrator Manual* can be found on WIDA's website. 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[1] and within the training course. Additionally, the ACCESS 2.0 Training Course highlights common testing irregularities and the resources to use in these circumstances.

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 regarding 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

[1] WIDA state pages can be found at: https://www.wida.us/membership/states/index.aspx

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 scanned and then 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 electronically scored 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 5.1.1., 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 401 Paper, the Listening and Reading test forms were reused forms from ACCESS Series 303. See Section 1.3.3 for further detail.

For ACCESS 2.0 Series 401 paper, in the domains of Writing and Speaking a certain percentage of items from the Series 400 test form were refreshed, and a number of tasks were retained from the previous year's assessment for the purpose of scale maintenance. Section 5.2.5. of this report describes the equating procedures used, and Table E presents item-by-item information, including information on which tasks were used as anchor tasks.

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: For ACCESS 2.0 Series 401 Paper, year-to-year consistency with the ACCESS scale was maintained in two ways. In the domains of Listening and Reading, the Series 401 Paper tests are reused forms of the ACCESS Series 303 test.

For Writing and Speaking, Section 5.2.5. describes the equating summary included in this report, and Table E in Chapter 6 provides detailed information on which tasks were used as anchor tasks.

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

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

- L: (Ability Measure in Logits*37.571) + 316.637
- R: (Ability Measure in Logits*26.000) + 323.272
- W: (Ability Measure in Logits*26.851) + 303.332
- S: (Ability Measure in Logits*29.248) + 265.076

For Listening and Reading, these equations have been in use from the first operational administration of ACCESS (Series 100). Evidence for scale maintenance in Listening and Reading is detailed in the *ACCESS for ELLs Series 400 Listening and Reading Scale Maintenance: Technical Brief* (Center for Applied Linguistics, 2016).

For Writing and Speaking, scaling equations are new for Series 401. A scaling study was conducted in summer 2016 (see Center for Applied Linguistics [2017]). The equations derived from this scaling study were used for the first time in Series 401 (2016–17 operational year).

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 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 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.

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.

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.

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).

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.

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., 5.2.9., 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. Section 5.2.5., Table 6E. c. Evidence summarized with claim at 4.3c.
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 ELD 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. There were a total of 41 students excluded from the analyses due to mismatches in students' tiers across domains.

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.

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 S401 Paper

State	Cluster							Total
	K	1	2	3	4-5	6-8	9-12	
AK	1,386	438	475	461	945	1,337	1,139	6,181
AL	3,487	1,347	1,234	841	736	704	879	9,228
CO	10,836	3,388	3,661	3,766	6,098	6,769	5,485	40,003
DC	1,089	2	3	3	3	2	62	1,164
DE	1,637	4	0	3	1	5	0	1,650
FL	35,774	35,808	39,288	38,839	42,497	43,091	44,213	279,510
GA	17,196	4,126	3,729	2,984	1,116	114	95	29,360
HI	1,876	574	477	557	482	341	583	4,890
ID	2,230	2	4	2	8	11	12	2,269
IL	26,670	10,165	10,846	3,493	2,859	2,184	1,477	57,694
IN	7,405	102	103	112	95	98	13	7,928
KY	3,377	59	47	52	63	55	71	3,724
MA	10,330	5,017	4,544	4,034	4,071	5,042	6,115	39,153
MD	10,675	12	16	19	33	28	14	10,797
ME	485	26	38	35	64	68	16	732
MI	10,371	441	444	275	520	951	1,143	14,145
MN	8,316	132	134	96	111	85	39	8,913
MO	4,899	16	11	8	19	8	7	4,968
MP	78	0	0	0	0	0	0	78
MT	137	0	0	0	0	0	0	137
NC	11,957	369	330	337	256	127	124	13,500
ND	384	4	4	3	8	13	9	425
NH	441	37	44	61	60	44	48	735
NJ	12,035	276	141	93	91	153	201	12,990
NM	4,717	387	387	418	710	775	119	7,513
NV	7,956	0	0	0	1	4	34	7,995
OK	6,902	2,871	2,634	2,483	2,841	1,619	985	20,335
PA	5,017	1,363	1,335	1,278	2,142	2,442	2,148	15,725
RI	1,092	227	210	148	177	161	194	2,209
SC	3,478	173	145	191	252	285	249	4,773
SD	742	64	39	51	21	17	0	934
TN	5,711	5	4	3	2	1	0	5,726
UT	4,975	1	1	2	0	1	1	4,981
VA	14,215	4,304	3,614	970	893	789	825	25,610
VI	96	0	0	0	0	0	0	96
VT	178	3	1	3	3	0	3	191
WI	5,531	47	29	26	33	21	24	5,711
WY	386	4	12	8	8	23	5	446
Total	244,067	71,794	73,984	61,655	67,219	67,368	66,332	652,419

4.1.1.2 By Gender

Table 4.1.1.2

Participation by Cluster by Gender S401 Paper

Cluster		Gender			Total
		F	M	Missing	
K	Count	112,595	127,961	3,511	244,067
	% within Cluster	46.1%	52.4%	1.4%	100.0%
1	Count	33,480	38,152	162	71,794
	% within Cluster	46.6%	53.1%	0.2%	100.0%
2	Count	34,958	38,872	154	73,984
	% within Cluster	47.3%	52.5%	0.2%	100.0%
3	Count	28,696	32,791	168	61,655
	% within Cluster	46.5%	53.2%	0.3%	100.0%
4-5	Count	30,688	36,368	163	67,219
	% within Cluster	45.7%	54.1%	0.2%	100.0%
6-8	Count	30,349	36,809	210	67,368
	% within Cluster	45.0%	54.6%	0.3%	100.0%
9-12	Count	30,131	35,883	318	66,332
	% within Cluster	45.4%	54.1%	0.5%	100.0%
Total	Count	300,897	346,836	4,686	652,419
	% 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 S401 Paper

Cluster		Hispanic/Non-Hispanic			Total
		Hispanic	Other	Unknown	
K	Count	162,046	74,171	7,850	244,067
	% within Cluster	66.4%	30.4%	3.2%	100.0%
1	Count	53,748	17,372	674	71,794
	% within Cluster	74.9%	24.2%	0.9%	100.0%
2	Count	55,992	17,343	649	73,984
	% within Cluster	75.7%	23.4%	0.9%	100.0%
3	Count	46,712	14,465	478	61,655
	% within Cluster	75.8%	23.5%	0.8%	100.0%
4-5	Count	50,830	15,847	542	67,219
	% within Cluster	75.6%	23.6%	0.8%	100.0%
6-8	Count	50,508	16,334	526	67,368
	% within Cluster	75.0%	24.2%	0.8%	100.0%
9-12	Count	48,142	17,527	663	66,332
	% within Cluster	72.6%	26.4%	1.0%	100.0%
Total	Count	467,978	173,059	11,382	652,419
	% within Cluster	71.7%	26.5%	1.7%	100.0%

4.1.2 Participation by Grade

4.1.2.1 By State

Table 4.1.2.1

Participation by Grade by State S401 Paper

State	Grade													Total
	K	1	2	3	4	5	6	7	8	9	10	11	12	
AK	1,386	438	475	461	466	479	456	479	402	430	302	234	173	6,181
AL	3,487	1,347	1,234	841	459	277	215	247	242	354	283	158	84	9,228
CO	10,836	3,388	3,661	3,766	3,205	2,893	2,273	2,233	2,263	2,424	1,378	925	758	40,003
DC	1,089	2	3	3	3	0	1	0	1	29	12	5	16	1,164
DE	1,637	4	0	3	1	0	2	2	1	0	0	0	0	1,650
FL	35,774	35,808	39,288	38,839	20,626	21,871	15,783	13,730	13,578	14,054	12,804	10,621	6,734	279,510
GA	17,196	4,126	3,729	2,984	653	463	54	32	28	36	37	14	8	29,360
HI	1,876	574	477	557	267	215	133	95	113	253	149	93	88	4,890
ID	2,230	2	4	2	4	4	2	4	5	10	0	1	1	2,269
IL	26,670	10,165	10,846	3,493	1,721	1,138	815	725	644	538	417	290	232	57,694
IN	7,405	102	103	112	57	38	33	41	24	5	6	2	0	7,928
KY	3,377	59	47	52	32	31	23	19	13	37	15	14	5	3,724
MA	10,330	5,017	4,544	4,034	2,273	1,798	1,805	1,674	1,563	2,345	1,545	1,269	956	39,153
MD	10,675	12	16	19	18	15	14	4	10	7	4	1	2	10,797
ME	485	26	38	35	41	23	26	15	27	1	3	2	10	732
MI	10,371	441	444	275	260	260	334	316	301	479	255	240	169	14,145
MN	8,316	132	134	96	65	46	32	28	25	17	7	10	5	8,913
MO	4,899	16	11	8	14	5	4	3	1	2	2	2	1	4,968
MP	78	0	0	0	0	0	0	0	0	0	0	0	0	78
MT	137	0	0	0	0	0	0	0	0	0	0	0	0	137
NC	11,957	369	330	337	173	83	44	38	45	59	31	16	18	13,500
ND	384	4	4	3	6	2	8	2	3	1	3	1	4	425
NH	441	37	44	61	39	21	15	14	15	19	8	13	8	735
NJ	12,035	276	141	93	52	39	36	59	58	102	53	33	13	12,990
NM	4,717	387	387	418	375	335	284	267	224	54	29	24	12	7,513
NV	7,956	0	0	0	1	0	1	1	2	3	9	15	7	7,995
OK	6,902	2,871	2,634	2,483	1,774	1,067	571	537	511	434	247	177	127	20,335
PA	5,017	1,363	1,335	1,278	1,158	984	873	795	774	677	592	515	364	15,725
RI	1,092	227	210	148	117	60	54	51	56	59	55	55	25	2,209
SC	3,478	173	145	191	146	106	101	101	83	116	60	41	32	4,773
SD	742	64	39	51	17	4	8	7	2	0	0	0	0	934
TN	5,711	5	4	3	0	2	0	0	1	0	0	0	0	5,726
UT	4,975	1	1	2	0	0	0	1	0	1	0	0	0	4,981
VA	14,215	4,304	3,614	970	559	334	279	245	265	449	167	141	68	25,610
VI	96	0	0	0	0	0	0	0	0	0	0	0	0	96
VT	178	3	1	3	3	0	0	0	0	0	0	2	1	191
WI	5,531	47	29	26	16	17	6	6	9	8	5	8	3	5,711
WY	386	4	12	8	5	3	8	9	6	1	3	0	1	446
Total	244,067	71,794	73,984	61,655	34,606	32,613	24,293	21,780	21,295	23,004	18,481	14,922	9,925	652,419

4.1.2.2 By Gender

Table 4.1.2.2

Participation by Grade by Gender S401 Paper

Grade		Gender			Total
		F	M	Missing	
K	Count	112,595	127,961	3,511	244,067
	% within Grade	46.1%	52.4%	1.4%	100.0%
1	Count	33,480	38,152	162	71,794
	% within Grade	46.6%	53.1%	0.2%	100.0%
2	Count	34,958	38,872	154	73,984
	% within Grade	47.3%	52.5%	0.2%	100.0%
3	Count	28,696	32,791	168	61,655
	% within Grade	46.5%	53.2%	0.3%	100.0%
4	Count	15,616	18,880	110	34,606
	% within Grade	45.1%	54.6%	0.3%	100.0%
5	Count	15,072	17,488	53	32,613
	% within Grade	46.2%	53.6%	0.2%	100.0%
6	Count	10,890	13,295	108	24,293
	% within Grade	44.8%	54.7%	0.4%	100.0%
7	Count	9,769	11,957	54	21,780
	% within Grade	44.9%	54.9%	0.2%	100.0%
8	Count	9,690	11,557	48	21,295
	% within Grade	45.5%	54.3%	0.2%	100.0%
9	Count	10,003	12,783	218	23,004
	% within Grade	43.5%	55.6%	0.9%	100.0%
10	Count	8,390	10,041	50	18,481
	% within Grade	45.4%	54.3%	0.3%	100.0%
11	Count	6,989	7,906	27	14,922
	% within Grade	46.8%	53.0%	0.2%	100.0%
12	Count	4,749	5,153	23	9,925
	% within Grade	47.8%	51.9%	0.2%	100.0%
Total	Count	300,897	346,836	4,686	652,419
	% 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 S401 Paper

Grade		Hispanic/Non-Hispanic			Total
		Hispanic	Other	Unknown	
K	Count	162,046	74,171	7,850	244,067
	% within Grade	66.4%	30.4%	3.2%	100.0%
1	Count	53,748	17,372	674	71,794
	% within Grade	74.9%	24.2%	0.9%	100.0%
2	Count	55,992	17,343	649	73,984
	% within Grade	75.7%	23.4%	0.9%	100.0%
3	Count	46,712	14,465	478	61,655
	% within Grade	75.8%	23.5%	0.8%	100.0%
4	Count	26,111	8,168	327	34,606
	% within Grade	75.5%	23.6%	0.9%	100.0%
5	Count	24,719	7,679	215	32,613
	% within Grade	75.8%	23.5%	0.7%	100.0%
6	Count	18,141	5,902	250	24,293
	% within Grade	74.7%	24.3%	1.0%	100.0%
7	Count	16,274	5,361	145	21,780
	% within Grade	74.7%	24.6%	0.7%	100.0%
8	Count	16,093	5,071	131	21,295
	% within Grade	75.6%	23.8%	0.6%	100.0%
9	Count	17,167	5,501	336	23,004
	% within Grade	74.6%	23.9%	1.5%	100.0%
10	Count	13,786	4,548	147	18,481
	% within Grade	74.6%	24.6%	0.8%	100.0%
11	Count	10,593	4,231	98	14,922
	% within Grade	71.0%	28.4%	0.7%	100.0%
12	Count	6,596	3,247	82	9,925
	% within Grade	66.5%	32.7%	0.8%	100.0%
Total	Count	467,978	173,059	11,382	652,419
	% within Grade	71.7%	26.5%	1.7%	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 S401 Paper

Cluster			Domain							
			Listening		Reading		Writing		Speaking	
			N	%	N	%	N	%	N	%
K	Tier	-	244,059	-	244,059	-	244,057	-	244,061	-
1	Tier	A	28,255	39.4%	28,272	39.4%	28,270	39.4%	28,272	39.4%
		B	25,848	36.0%	25,854	36.0%	25,857	36.0%	25,856	36.0%
		C	17,647	24.6%	17,655	24.6%	17,657	24.6%	17,655	24.6%
	Total		71,750	100.0%	71,781	100.0%	71,784	100.0%	71,783	100.0%
2	Tier	A	10,375	14.0%	10,375	14.0%	10,375	14.0%	10,375	14.0%
		B	26,679	36.1%	26,690	36.1%	26,690	36.1%	26,685	36.1%
		C	36,901	49.9%	36,909	49.9%	36,913	49.9%	36,908	49.9%
	Total		73,955	100.0%	73,974	100.0%	73,978	100.0%	73,968	100.0%
3	Tier	A	6,871	11.1%	6,871	11.1%	6,869	11.1%	6,869	11.1%
		B	15,692	25.5%	15,692	25.5%	15,691	25.5%	15,690	25.5%
		C	39,089	63.4%	39,087	63.4%	39,088	63.4%	39,089	63.4%
	Total		61,652	100.0%	61,650	100.0%	61,648	100.0%	61,648	100.0%
4-5	Tier	A	10,624	15.8%	10,626	15.8%	10,626	15.8%	10,625	15.8%
		B	15,946	23.7%	15,945	23.7%	15,944	23.7%	15,945	23.7%
		C	40,639	60.5%	40,639	60.5%	40,637	60.5%	40,638	60.5%
	Total		67,209	100.0%	67,210	100.0%	67,207	100.0%	67,208	100.0%
6-8	Tier	A	14,662	21.8%	14,662	21.8%	14,661	21.8%	14,660	21.8%
		B	17,167	25.5%	17,166	25.5%	17,166	25.5%	17,167	25.5%
		C	35,534	52.8%	35,533	52.8%	35,532	52.8%	35,530	52.7%
	Total		67,363	100.0%	67,361	100.0%	67,359	100.0%	67,357	100.0%
9-12	Tier	A	16,687	25.2%	16,688	25.2%	16,687	25.2%	16,687	25.2%
		B	17,850	26.9%	17,852	26.9%	17,850	26.9%	17,850	26.9%
		C	31,773	47.9%	31,775	47.9%	31,772	47.9%	31,778	47.9%
	Total		66,310	100.0%	66,315	100.0%	66,309	100.0%	66,315	100.0%

4.1.3.2 By Grade by Domain

Table 4.1.3.2
Participation by Grade by Tier by Domain S401 Paper

Grade			Domain			
			Listening	**Reading**	**Writing**	**Speaking**
K	Tier	-	244,059	244,059	244,057	244,061
1	Tier	A	28,255	28,272	28,270	28,272
		B	25,848	25,854	25,857	25,856
		C	17,647	17,655	17,657	17,655
	Total		71,750	71,781	71,784	71,783
2	Tier	A	10,375	10,375	10,375	10,375
		B	26,679	26,690	26,690	26,685
		C	36,901	36,909	36,913	36,908
	Total		73,955	73,974	73,978	73,968
3	Tier	A	6,871	6,871	6,869	6,869
		B	15,692	15,692	15,691	15,690
		C	39,089	39,087	39,088	39,089
	Total		61,652	61,650	61,648	61,648
4	Tier	A	5,707	5,709	5,709	5,709
		B	8,812	8,811	8,811	8,811
		C	20,081	20,081	20,080	20,080
	Total		34,600	34,601	34,600	34,600
5	Tier	A	4,917	4,917	4,917	4,916
		B	7,134	7,134	7,133	7,134
		C	20,558	20,558	20,557	20,558
	Total		32,609	32,609	32,607	32,608
6	Tier	A	5,055	5,055	5,054	5,055
		B	5,662	5,662	5,661	5,663
		C	13,573	13,573	13,572	13,570
	Total		24,290	24,290	24,287	24,288
7	Tier	A	4,847	4,847	4,847	4,847
		B	5,896	5,896	5,897	5,897
		C	11,035	11,035	11,034	11,035
	Total		21,778	21,778	21,778	21,779
8	Tier	A	4,760	4,760	4,760	4,758
		B	5,609	5,608	5,608	5,607
		C	10,926	10,925	10,926	10,925
	Total		21,295	21,293	21,294	21,290
9	Tier	A	7,199	7,199	7,199	7,197
		B	5,683	5,684	5,685	5,683
		C	10,115	10,116	10,114	10,117
	Total		22,997	22,999	22,998	22,997
10	Tier	A	4,669	4,669	4,668	4,670
		B	5,076	5,076	5,075	5,076
		C	8,729	8,730	8,730	8,731
	Total		18,474	18,475	18,473	18,477
11	Tier	A	3,422	3,422	3,422	3,422
		B	4,224	4,225	4,224	4,225
		C	7,271	7,270	7,269	7,272
	Total		14,917	14,917	14,915	14,919
12	Tier	A	1,397	1,398	1,398	1,398
		B	2,867	2,867	2,866	2,866
		C	5,658	5,659	5,659	5,658
	Total		9,922	9,924	9,923	9,922

4.1.3.3 By Cluster by Gender

Table 4.1.3.3

Participation by Cluster by Tier by Gender S401 Paper

Cluster	Tier		Gender			Total
			F	M	Missing	
K	-	Count	112,595	127,961	3,511	244,067
		% within Tier	46.1%	52.4%	1.4%	100.0%
1	A	Count	12,791	15,423	60	28,274
		% within Tier	45.2%	54.5%	0.2%	100.0%
	B	Count	12,056	13,747	60	25,863
		% within Tier	46.6%	53.2%	0.2%	100.0%
	C	Count	8,633	8,982	42	17,657
		% within Tier	48.9%	50.9%	0.2%	100.0%
2	A	Count	4,716	5,623	38	10,377
		% within Tier	45.4%	54.2%	0.4%	100.0%
	B	Count	12,155	14,484	54	26,693
		% within Tier	45.5%	54.3%	0.2%	100.0%
	C	Count	18,087	18,765	62	36,914
		% within Tier	49.0%	50.8%	0.2%	100.0%
3	A	Count	3,136	3,695	40	6,871
		% within Tier	45.6%	53.8%	0.6%	100.0%
	B	Count	6,931	8,721	41	15,693
		% within Tier	44.2%	55.6%	0.3%	100.0%
	C	Count	18,629	20,375	87	39,091
		% within Tier	47.7%	52.1%	0.2%	100.0%
4-5	A	Count	4,933	5,644	50	10,627
		% within Tier	46.4%	53.1%	0.5%	100.0%
	B	Count	7,059	8,849	43	15,951
		% within Tier	44.3%	55.5%	0.3%	100.0%
	C	Count	18,696	21,875	70	40,641
		% within Tier	46.0%	53.8%	0.2%	100.0%
6-8	A	Count	6,704	7,858	102	14,664
		% within Tier	45.7%	53.6%	0.7%	100.0%
	B	Count	7,592	9,529	49	17,170
		% within Tier	44.2%	55.5%	0.3%	100.0%
	C	Count	16,053	19,422	59	35,534
		% within Tier	45.2%	54.7%	0.2%	100.0%
9-12	A	Count	7,479	9,072	142	16,693
		% within Tier	44.8%	54.3%	0.9%	100.0%
	B	Count	8,072	9,689	93	17,854
		% within Tier	45.2%	54.3%	0.5%	100.0%
	C	Count	14,580	17,122	83	31,785
		% within Tier	45.9%	53.9%	0.3%	100.0%

4.1.3.4 By Cluster by Ethnicity

Table 4.1.3.4

Participation by Cluster by Tier by Ethnicity S401 Paper

Cluster	Tier		Hispanic/Non-Hispanic			Total
			Hispanic	Other	Unknown	
K	-	Count	162,046	74,171	7,850	244,067
		% within Tier	66.4%	30.4%	3.2%	100.0%
1	A	Count	21,920	6,076	278	28,274
		% within Tier	77.5%	21.5%	1.0%	100.0%
	B	Count	19,049	6,573	241	25,863
		% within Tier	73.7%	25.4%	0.9%	100.0%
	C	Count	12,779	4,723	155	17,657
		% within Tier	72.4%	26.7%	0.9%	100.0%
2	A	Count	7,699	2,532	146	10,377
		% within Tier	74.2%	24.4%	1.4%	100.0%
	B	Count	20,226	6,223	244	26,693
		% within Tier	75.8%	23.3%	0.9%	100.0%
	C	Count	28,067	8,588	259	36,914
		% within Tier	76.0%	23.3%	0.7%	100.0%
3	A	Count	5,030	1,727	114	6,871
		% within Tier	73.2%	25.1%	1.7%	100.0%
	B	Count	11,641	3,923	129	15,693
		% within Tier	74.2%	25.0%	0.8%	100.0%
	C	Count	30,041	8,815	235	39,091
		% within Tier	76.8%	22.5%	0.6%	100.0%
4-5	A	Count	7,883	2,545	199	10,627
		% within Tier	74.2%	23.9%	1.9%	100.0%
	B	Count	11,465	4,345	141	15,951
		% within Tier	71.9%	27.2%	0.9%	100.0%
	C	Count	31,482	8,957	202	40,641
		% within Tier	77.5%	22.0%	0.5%	100.0%
6-8	A	Count	11,276	3,190	198	14,664
		% within Tier	76.9%	21.8%	1.4%	100.0%
	B	Count	11,928	5,110	132	17,170
		% within Tier	69.5%	29.8%	0.8%	100.0%
	C	Count	27,304	8,034	196	35,534
		% within Tier	76.8%	22.6%	0.6%	100.0%
9-12	A	Count	12,513	3,917	263	16,693
		% within Tier	75.0%	23.5%	1.6%	100.0%
	B	Count	12,494	5,172	188	17,854
		% within Tier	70.0%	29.0%	1.1%	100.0%
	C	Count	23,135	8,438	212	31,785
		% within Tier	72.8%	26.5%	0.7%	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 S401 Paper

Cluster		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Mean	269.81	190.14	207.14	272.67	271.50	198.88	214.03	220.45
	N	243,666	243,664	243,666	243,662	243,655	243,659	243,657	243,643
1	Mean	295.80	289.21	262.15	282.08	289.20	275.93	291.22	279.70
	N	71,713	71,749	71,743	71,741	71,684	71,726	71,698	71,651
2	Mean	329.39	317.21	297.91	298.24	314.07	307.81	320.94	309.48
	N	73,920	73,923	73,937	73,925	73,898	73,912	73,899	73,870
3	Mean	357.21	339.96	314.31	312.71	335.22	327.39	345.20	329.53
	N	61,632	61,612	61,630	61,601	61,595	61,600	61,609	61,564
4-5	Mean	370.67	349.35	338.26	339.04	355.11	344.07	355.84	347.15
	N	67,170	67,128	67,164	67,174	67,164	67,112	67,118	67,097
6-8	Mean	379.63	357.47	331.21	354.48	367.31	344.60	364.16	351.20
	N	67,289	67,284	67,258	67,287	67,272	67,242	67,275	67,217
9-12	Mean	381.21	378.88	357.84	349.23	365.46	368.62	379.65	367.46
	N	66,242	66,243	66,226	66,250	66,226	66,208	66,229	66,183

4.2.1.2 By Cluster by Gender

Table 4.2.1.2
Mean Scale Scores by Cluster by Gender S401 Paper

Cluster	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	F	Mean	274.81	192.39	212.02	280.89	278.11	202.45	217.10	224.92
		N	112,397	112,396	112,395	112,394	112,391	112,394	112,393	112,386
	M	Mean	265.10	187.73	202.42	264.98	265.30	195.32	210.93	216.09
		N	127,759	127,758	127,761	127,758	127,754	127,755	127,754	127,747
	Missing	Mean	281.32	205.86	222.68	288.96	285.39	214.52	228.49	235.55
		N	3,510	3,510	3,510	3,510	3,510	3,510	3,510	3,510
1	F	Mean	297.28	289.97	266.92	285.77	291.78	278.69	292.19	282.41
		N	33,455	33,465	33,461	33,457	33,440	33,454	33,448	33,425
	M	Mean	294.55	288.57	258.02	278.96	287.01	273.54	290.39	277.38
		N	38,096	38,122	38,120	38,122	38,082	38,110	38,088	38,064
	Missing	Mean	286.64	284.36	249.02	257.41	272.25	266.91	285.07	268.31
		N	162	162	162	162	162	162	162	162
2	F	Mean	330.72	318.36	303.98	300.00	315.62	311.43	322.15	312.47
		N	34,929	34,932	34,938	34,935	34,922	34,927	34,921	34,910
	M	Mean	328.24	316.20	292.49	296.72	312.73	304.60	319.89	306.82
		N	38,837	38,837	38,845	38,836	38,822	38,831	38,824	38,806
	Missing	Mean	318.90	311.16	286.18	281.23	300.30	298.92	313.56	299.14
		N	154	154	154	154	154	154	154	154
3	F	Mean	357.58	340.87	320.68	313.73	335.91	331.02	345.94	332.27
		N	28,681	28,675	28,681	28,670	28,667	28,671	28,673	28,656
	M	Mean	356.91	339.19	308.80	311.93	334.68	324.25	344.57	327.17
		N	32,783	32,769	32,781	32,763	32,760	32,761	32,768	32,740
	Missing	Mean	353.39	335.41	303.18	290.48	322.16	319.55	340.88	320.11
		N	168	168	168	168	168	168	168	168
4-5	F	Mean	370.53	350.16	343.16	339.77	355.40	346.92	356.36	349.24
		N	30,669	30,654	30,663	30,666	30,664	30,646	30,651	30,639
	M	Mean	370.84	348.70	334.19	338.54	354.94	341.71	355.43	345.45
		N	36,338	36,311	36,339	36,345	36,337	36,304	36,304	36,296
	Missing	Mean	359.12	342.60	322.37	313.40	336.49	332.77	347.65	333.57
		N	163	163	162	163	163	162	163	162
6-8	F	Mean	380.19	358.97	336.94	354.56	367.63	348.21	365.38	353.82
		N	30,312	30,309	30,305	30,315	30,308	30,296	30,304	30,287
	M	Mean	379.33	356.30	326.62	354.68	367.26	341.72	363.25	349.16
		N	36,768	36,766	36,745	36,763	36,755	36,738	36,762	36,722
	Missing	Mean	351.01	346.22	306.50	308.04	329.79	326.69	347.69	327.42
		N	209	209	208	209	209	208	209	208
9-12	F	Mean	381.21	381.28	363.27	348.61	365.15	372.53	381.33	370.10
		N	30,094	30,095	30,090	30,097	30,090	30,083	30,088	30,073
	M	Mean	381.43	376.96	353.43	350.02	365.97	365.46	378.36	365.40
		N	35,831	35,831	35,819	35,836	35,819	35,808	35,824	35,793
	Missing	Mean	357.43	368.78	340.03	318.31	338.10	354.66	365.43	349.47
		N	317	317	317	317	317	317	317	317

4.2.1.3 By Cluster by Ethnicity

Table 4.2.1.3

Mean Scale Scores by Cluster by Ethnicity S401 Paper

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Non-Hispanic Asian	Mean	280.47	217.14	231.90	280.36	280.68	224.78	236.12	241.32
		N	31,276	31,275	31,275	31,274	31,274	31,274	31,275	31,272
	Non-Hispanic Pacific Islander	Mean	264.09	176.68	195.83	268.47	266.53	186.49	202.89	210.28
		N	1,698	1,698	1,698	1,698	1,698	1,698	1,698	1,698
	Non-Hispanic Black	Mean	269.27	197.96	211.37	283.92	276.85	204.92	219.33	226.27
		N	13,926	13,925	13,927	13,926	13,925	13,925	13,925	13,924
	Hispanic (Of Any Race)	Mean	266.49	182.83	200.38	268.43	267.72	191.84	207.92	214.39
		N	161,805	161,802	161,802	161,801	161,798	161,798	161,798	161,789
	Non-Hispanic American Indian	Mean	270.67	176.48	190.66	263.03	267.12	183.80	204.73	208.57
		N	2,557	2,557	2,557	2,557	2,557	2,557	2,557	2,557
	Non-Hispanic Multi-racial	Mean	287.41	207.07	217.90	297.30	292.62	212.73	231.15	236.48
		N	1,430	1,430	1,430	1,430	1,430	1,430	1,430	1,430
	Non-Hispanic White	Mean	280.33	201.31	220.76	286.49	283.67	211.29	225.00	232.77
		N	23,140	23,140	23,140	23,139	23,139	23,140	23,140	23,139
	Unknown	Mean	263.49	190.76	206.09	268.08	266.02	198.65	212.58	218.66
		N	7,834	7,837	7,837	7,837	7,834	7,837	7,834	7,834
1	Non-Hispanic Asian	Mean	298.78	297.45	274.80	289.42	294.37	286.37	297.85	288.56
		N	5,524	5,523	5,526	5,524	5,522	5,523	5,521	5,519
	Non-Hispanic Pacific Islander	Mean	288.80	285.83	268.20	282.60	285.95	277.26	286.73	279.62
		N	351	351	352	352	351	351	351	351
	Non-Hispanic Black	Mean	292.98	288.44	259.02	282.21	287.86	273.98	289.83	277.93
		N	4,797	4,802	4,801	4,801	4,796	4,801	4,797	4,795
	Hispanic (Of Any Race)	Mean	295.53	288.05	260.67	280.35	288.19	274.61	290.33	278.47
		N	53,688	53,715	53,706	53,705	53,664	53,695	53,679	53,639
	Non-Hispanic American Indian	Mean	290.22	286.58	250.81	257.82	274.22	268.92	287.70	270.32
		N	1,100	1,099	1,100	1,101	1,100	1,098	1,098	1,097
	Non-Hispanic Multi-racial	Mean	301.61	296.17	270.17	293.50	297.87	283.40	297.77	287.56
		N	396	398	398	398	396	398	396	396
	Non-Hispanic White	Mean	299.76	293.26	268.63	297.05	298.67	281.19	295.24	286.23
		N	5,184	5,188	5,187	5,187	5,183	5,187	5,184	5,182
	Unknown	Mean	292.26	290.72	259.26	276.37	284.68	275.22	291.25	277.90
		N	673	673	673	673	672	673	672	672

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
2	Non-Hispanic Asian	Mean	333.56	327.91	310.15	305.94	320.01	319.28	329.68	319.28
		N	5,175	5,173	5,175	5,175	5,175	5,173	5,173	5,173
	Non-Hispanic Pacific Islander	Mean	325.12	318.13	307.01	299.43	312.52	312.74	320.25	312.45
		N	286	285	286	286	286	285	285	285
	Non-Hispanic Black	Mean	324.61	312.54	290.70	295.43	310.29	301.88	316.23	304.18
		N	5,140	5,140	5,140	5,141	5,139	5,138	5,139	5,136
	Hispanic (Of Any Race)	Mean	329.33	316.28	297.15	297.16	313.49	306.97	320.27	308.71
		N	55,943	55,952	55,958	55,946	55,922	55,943	55,930	55,904
	Non-Hispanic American Indian	Mean	318.86	307.93	286.51	267.61	293.49	297.49	311.25	296.06
		N	1,131	1,130	1,131	1,131	1,131	1,130	1,130	1,130
	Non-Hispanic Multi-racial	Mean	331.71	321.50	299.58	305.13	318.68	310.79	324.63	312.92
		N	373	373	373	373	373	373	373	373
	Non-Hispanic White	Mean	333.80	322.90	303.24	311.81	323.06	313.34	326.25	316.04
		N	5,223	5,221	5,225	5,224	5,223	5,221	5,220	5,220
	Unknown	Mean	323.25	316.87	294.10	291.57	307.67	305.75	318.85	306.09
		N	649	649	649	649	649	649	649	649
3	Non-Hispanic Asian	Mean	364.12	347.92	324.85	320.83	342.72	336.64	352.84	338.25
		N	3,827	3,826	3,827	3,826	3,826	3,826	3,826	3,825
	Non-Hispanic Pacific Islander	Mean	347.71	335.23	322.75	307.09	327.65	329.20	339.03	328.44
		N	308	308	307	308	308	307	308	307
	Non-Hispanic Black	Mean	353.12	335.61	306.10	309.31	331.47	321.12	340.93	323.99
		N	4,978	4,974	4,977	4,976	4,976	4,973	4,974	4,971
	Hispanic (Of Any Race)	Mean	357.03	339.64	314.01	311.87	334.70	327.08	344.92	329.16
		N	46,697	46,681	46,694	46,669	46,666	46,671	46,679	46,642
	Non-Hispanic American Indian	Mean	339.20	330.31	310.05	290.29	315.07	320.45	333.01	318.63
		N	1,098	1,098	1,100	1,100	1,098	1,098	1,098	1,098
	Non-Hispanic Multi-racial	Mean	358.05	339.16	312.83	316.48	337.53	326.25	344.89	329.44
		N	351	351	351	351	351	351	351	351
	Non-Hispanic White	Mean	364.15	345.21	319.61	327.81	346.24	332.65	350.96	336.54
		N	3,895	3,896	3,896	3,893	3,892	3,896	3,895	3,892
	Unknown	Mean	352.71	335.92	307.08	295.04	324.10	321.74	341.02	322.25
		N	478	478	478	478	478	478	478	478

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
4-5	Non-Hispanic Asian	Mean	373.12	355.29	344.43	341.78	357.71	350.12	360.74	352.17
		N	3,220	3,219	3,220	3,220	3,220	3,219	3,219	3,219
	Non-Hispanic Pacific Islander	Mean	357.75	341.41	337.54	321.29	339.77	339.69	346.40	339.49
		N	296	296	296	296	296	296	296	296
	Non-Hispanic Black	Mean	368.81	345.86	333.66	341.08	355.20	340.02	352.83	344.36
		N	6,015	6,006	6,013	6,015	6,014	6,004	6,005	6,003
	Hispanic (Of Any Race)	Mean	371.30	349.60	338.77	339.15	355.48	344.45	356.20	347.52
		N	50,807	50,778	50,803	50,809	50,802	50,765	50,770	50,752
	Non-Hispanic American Indian	Mean	357.73	340.67	329.59	312.87	335.55	335.38	345.88	335.23
		N	1,886	1,886	1,886	1,888	1,886	1,885	1,885	1,884
	Non-Hispanic Multi-racial	Mean	370.56	349.88	337.39	345.38	358.25	343.88	356.19	347.98
		N	393	393	393	393	393	393	393	393
	Non-Hispanic White	Mean	373.12	352.60	340.33	349.69	361.66	346.75	358.86	350.99
		N	4,011	4,008	4,011	4,011	4,011	4,008	4,008	4,008
	Unknown	Mean	351.45	339.17	320.88	307.15	329.54	330.28	342.97	329.84
		N	542	542	542	542	542	542	542	542
6-8	Non-Hispanic Asian	Mean	388.15	366.33	342.97	364.72	376.69	354.91	372.92	361.23
		N	3,158	3,158	3,152	3,156	3,156	3,152	3,158	3,150
	Non-Hispanic Pacific Islander	Mean	374.17	352.54	335.91	348.35	361.50	344.47	359.05	349.34
		N	231	231	231	231	231	231	231	231
	Non-Hispanic Black	Mean	377.00	353.60	326.31	352.61	365.07	340.21	360.67	347.47
		N	6,033	6,035	6,033	6,034	6,032	6,033	6,033	6,030
	Hispanic (Of Any Race)	Mean	379.21	356.88	330.64	353.64	366.68	344.02	363.62	350.60
		N	50,457	50,451	50,439	50,457	50,445	50,423	50,444	50,405
	Non-Hispanic American Indian	Mean	374.34	355.26	332.52	336.37	355.59	344.15	361.03	347.35
		N	2,304	2,304	2,304	2,303	2,303	2,304	2,304	2,303
	Non-Hispanic Multi-racial	Mean	382.75	359.72	332.11	359.28	371.31	346.16	366.66	353.49
		N	380	380	380	380	380	380	380	380
	Non-Hispanic White	Mean	386.70	365.22	336.71	371.61	379.42	351.21	371.71	359.45
		N	4,201	4,201	4,196	4,200	4,200	4,196	4,201	4,195
	Unknown	Mean	365.70	353.42	319.06	336.71	351.62	336.53	357.14	340.84
		N	525	524	523	526	525	523	524	523

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
9-12	Non-Hispanic Asian	Mean	391.70	387.76	370.57	361.28	376.73	379.42	389.02	378.40
		N	3,342	3,343	3,342	3,342	3,342	3,342	3,342	3,341
	Non-Hispanic Pacific Islander	Mean	379.51	371.68	359.16	369.14	374.57	365.68	374.09	368.11
		N	343	343	343	343	343	343	343	343
	Non-Hispanic Black	Mean	374.09	374.75	354.20	345.44	359.99	364.73	374.62	363.11
		N	7,983	7,985	7,978	7,986	7,981	7,975	7,982	7,970
	Hispanic (Of Any Race)	Mean	381.02	378.78	357.33	348.02	364.76	368.31	379.52	367.04
		N	48,085	48,082	48,077	48,087	48,071	48,064	48,074	48,046
	Non-Hispanic American Indian	Mean	383.75	367.33	353.44	339.24	361.73	360.65	372.32	360.76
		N	1,544	1,545	1,544	1,545	1,544	1,544	1,544	1,543
	Non-Hispanic Multi-racial	Mean	385.88	382.61	360.70	361.43	373.99	371.95	383.74	372.35
		N	314	315	314	315	314	314	314	314
	Non-Hispanic White	Mean	390.25	386.75	363.72	364.99	377.85	375.49	387.86	375.99
		N	3,969	3,968	3,966	3,970	3,969	3,964	3,968	3,964
	Unknown	Mean	366.70	372.79	347.42	334.58	350.88	360.36	371.03	357.31
		N	662	662	662	662	662	662	662	662

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 S401 Paper

Grade		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Mean	269.81	190.14	207.14	272.67	271.50	198.88	214.03	220.45
	N	243,666	243,664	243,666	243,662	243,655	243,659	243,657	243,643
1	Mean	295.80	289.21	262.15	282.08	289.20	275.93	291.22	279.70
	N	71,713	71,749	71,743	71,741	71,684	71,726	71,698	71,651
2	Mean	329.39	317.21	297.91	298.24	314.07	307.81	320.94	309.48
	N	73,920	73,923	73,937	73,925	73,898	73,912	73,899	73,870
3	Mean	357.21	339.96	314.31	312.71	335.22	327.39	345.20	329.53
	N	61,632	61,612	61,630	61,601	61,595	61,600	61,609	61,564
4	Mean	364.99	344.96	333.23	334.48	349.99	339.35	351.05	342.33
	N	34,579	34,561	34,578	34,584	34,575	34,552	34,554	34,542
5	Mean	376.69	354.01	343.59	343.88	360.53	349.07	360.91	352.27
	N	32,591	32,567	32,586	32,590	32,589	32,560	32,564	32,555
6	Mean	375.21	351.74	326.82	351.50	363.61	339.53	358.82	346.54
	N	24,272	24,268	24,264	24,270	24,267	24,259	24,268	24,254
7	Mean	379.32	357.19	330.74	353.33	366.57	344.22	363.89	350.70
	N	21,756	21,755	21,740	21,759	21,752	21,733	21,750	21,724
8	Mean	384.98	364.30	336.69	359.07	372.29	350.76	370.54	357.01
	N	21,261	21,261	21,254	21,258	21,253	21,250	21,257	21,239
9	Mean	374.88	373.44	351.08	341.73	358.55	362.52	373.93	361.12
	N	22,969	22,966	22,964	22,969	22,961	22,955	22,962	22,944
10	Mean	381.21	378.29	356.68	346.80	364.24	367.74	379.24	366.48
	N	18,464	18,464	18,457	18,468	18,463	18,454	18,462	18,452
11	Mean	384.55	383.03	362.59	353.06	369.06	373.07	383.55	371.65
	N	14,900	14,900	14,898	14,903	14,897	14,893	14,897	14,889
12	Mean	390.88	386.35	368.51	365.35	378.36	377.68	387.78	377.69
	N	9,909	9,913	9,907	9,910	9,905	9,906	9,908	9,898

4.2.2.2 By Grade by Gender

Table 4.2.2.2

Mean Scale Scores by Grade by Gender S401 Paper

Grade	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	F	Mean	274.81	192.39	212.02	280.89	278.11	202.45	217.10	224.92
		N	112,397	112,396	112,395	112,394	112,391	112,394	112,393	112,386
	M	Mean	265.10	187.73	202.42	264.98	265.30	195.32	210.93	216.09
		N	127,759	127,758	127,761	127,758	127,754	127,755	127,754	127,747
	Missing	Mean	281.32	205.86	222.68	288.96	285.39	214.52	228.49	235.55
		N	3,510	3,510	3,510	3,510	3,510	3,510	3,510	3,510
1	F	Mean	297.28	289.97	266.92	285.77	291.78	278.69	292.19	282.41
		N	33,455	33,465	33,461	33,457	33,440	33,454	33,448	33,425
	M	Mean	294.55	288.57	258.02	278.96	287.01	273.54	290.39	277.38
		N	38,096	38,122	38,120	38,122	38,082	38,110	38,088	38,064
	Missing	Mean	286.64	284.36	249.02	257.41	272.25	266.91	285.07	268.31
		N	162	162	162	162	162	162	162	162
2	F	Mean	330.72	318.36	303.98	300.00	315.62	311.43	322.15	312.47
		N	34,929	34,932	34,938	34,935	34,922	34,927	34,921	34,910
	M	Mean	328.24	316.20	292.49	296.72	312.73	304.60	319.89	306.82
		N	38,837	38,837	38,845	38,836	38,822	38,831	38,824	38,806
	Missing	Mean	318.90	311.16	286.18	281.23	300.30	298.92	313.56	299.14
		N	154	154	154	154	154	154	154	154
3	F	Mean	357.58	340.87	320.68	313.73	335.91	331.02	345.94	332.27
		N	28,681	28,675	28,681	28,670	28,667	28,671	28,673	28,656
	M	Mean	356.91	339.19	308.80	311.93	334.68	324.25	344.57	327.17
		N	32,783	32,769	32,781	32,763	32,760	32,761	32,768	32,740
	Missing	Mean	353.39	335.41	303.18	290.48	322.16	319.55	340.88	320.11
		N	168	168	168	168	168	168	168	168
4	F	Mean	364.71	345.41	337.84	334.80	350.01	341.88	351.28	344.10
		N	15,607	15,600	15,603	15,605	15,604	15,595	15,599	15,592
	M	Mean	365.29	344.61	329.52	334.35	350.08	337.32	350.90	340.93
		N	18,862	18,851	18,866	18,869	18,861	18,848	18,845	18,841
	Missing	Mean	354.05	340.06	316.09	311.80	333.15	328.40	344.35	329.43
		N	110	110	109	110	110	109	110	109
5	F	Mean	376.56	355.08	348.67	344.93	360.99	352.15	361.62	354.56
		N	15,062	15,054	15,060	15,061	15,060	15,051	15,052	15,047
	M	Mean	376.82	353.11	339.24	343.06	360.19	346.44	360.32	350.32
		N	17,476	17,460	17,473	17,476	17,476	17,456	17,459	17,455
	Missing	Mean	369.62	347.85	335.28	316.72	343.42	341.75	354.51	342.08
		N	53	53	53	53	53	53	53	53

Grade	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
6	F	Mean	376.08	353.26	332.80	352.05	364.31	343.28	360.14	349.37
		N	10,881	10,878	10,880	10,881	10,880	10,877	10,878	10,876
	M	Mean	374.77	350.60	322.14	351.44	363.36	336.62	357.89	344.43
		N	13,283	13,282	13,277	13,281	13,279	13,275	13,282	13,271
	Missing	Mean	342.82	340.22	299.53	303.73	323.54	320.22	341.05	321.08
		N	108	108	107	108	108	107	108	107
7	F	Mean	379.96	358.69	336.82	353.36	366.91	348.01	365.13	353.46
		N	9,760	9,760	9,757	9,762	9,758	9,753	9,757	9,748
	M	Mean	378.93	355.99	325.89	353.51	366.46	341.20	362.93	348.55
		N	11,943	11,942	11,930	11,944	11,941	11,927	11,940	11,923
	Missing	Mean	351.77	350.19	305.96	305.51	328.91	328.36	350.70	328.25
		N	53	53	53	53	53	53	53	53
8	F	Mean	385.05	365.67	341.73	358.59	372.09	353.96	371.53	359.20
		N	9,671	9,671	9,668	9,672	9,670	9,666	9,669	9,663
	M	Mean	384.99	363.19	332.52	359.64	372.57	348.13	369.77	355.25
		N	11,542	11,542	11,538	11,538	11,535	11,536	11,540	11,528
	Missing	Mean	368.58	355.35	322.65	320.54	344.83	339.27	359.33	340.65
		N	48	48	48	48	48	48	48	48
9	F	Mean	376.53	376.84	357.87	343.80	360.40	367.61	376.80	365.23
		N	9,989	9,986	9,987	9,990	9,989	9,982	9,985	9,981
	M	Mean	374.08	370.96	346.04	340.73	357.66	358.76	371.96	358.22
		N	12,762	12,762	12,759	12,761	12,754	12,755	12,759	12,745
	Missing	Mean	346.38	363.41	334.39	305.11	326.00	349.15	358.33	341.99
		N	218	218	218	218	218	218	218	218
10	F	Mean	380.70	379.96	361.32	345.14	363.15	370.89	380.26	368.36
		N	8,384	8,383	8,382	8,386	8,384	8,380	8,382	8,379
	M	Mean	381.69	376.96	352.88	348.21	365.18	365.18	378.45	364.97
		N	10,031	10,032	10,026	10,033	10,030	10,025	10,031	10,024
	Missing	Mean	370.51	366.55	339.10	343.69	357.31	353.06	367.80	354.12
		N	49	49	49	49	49	49	49	49
11	F	Mean	383.15	384.61	366.69	350.21	366.92	375.90	384.24	372.99
		N	6,980	6,982	6,980	6,980	6,978	6,980	6,980	6,977
	M	Mean	385.77	381.59	358.95	355.60	370.93	370.54	382.91	370.43
		N	7,893	7,891	7,891	7,896	7,892	7,886	7,890	7,885
	Missing	Mean	392.63	395.96	368.93	350.67	371.85	382.70	395.07	379.30
		N	27	27	27	27	27	27	27	27
12	F	Mean	389.14	388.07	373.05	362.53	376.07	380.81	388.47	379.20
		N	4,741	4,744	4,741	4,741	4,739	4,741	4,741	4,736
	M	Mean	392.47	384.73	364.35	368.01	380.50	374.80	387.13	376.31
		N	5,145	5,146	5,143	5,146	5,143	5,142	5,144	5,139
	Missing	Mean	393.00	392.57	361.57	351.39	372.35	377.30	392.83	375.52
		N	23	23	23	23	23	23	23	23

4.2.2.3 By Grade by Ethnicity

Table 4.2.2.3

Mean Scale Scores by Grade by Ethnicity S401 Paper

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
K	Non-Hispanic Asian	Mean	280.47	217.14	231.90	280.36	280.68	224.78	236.12	241.32
		N	31,276	31,275	31,275	31,274	31,274	31,274	31,275	31,272
	Non-Hispanic Pacific Islander	Mean	264.09	176.68	195.83	268.47	266.53	186.49	202.89	210.28
		N	1,698	1,698	1,698	1,698	1,698	1,698	1,698	1,698
	Non-Hispanic Black	Mean	269.27	197.96	211.37	283.92	276.85	204.92	219.33	226.27
		N	13,926	13,925	13,927	13,926	13,925	13,925	13,925	13,924
	Hispanic (Of Any Race)	Mean	266.49	182.83	200.38	268.43	267.72	191.84	207.92	214.39
		N	161,805	161,802	161,802	161,801	161,798	161,798	161,798	161,789
	Non-Hispanic American Indian	Mean	270.67	176.48	190.66	263.03	267.12	183.80	204.73	208.57
		N	2,557	2,557	2,557	2,557	2,557	2,557	2,557	2,557
	Non-Hispanic Multi-racial	Mean	287.41	207.07	217.90	297.30	292.62	212.73	231.15	236.48
		N	1,430	1,430	1,430	1,430	1,430	1,430	1,430	1,430
	Non-Hispanic White	Mean	280.33	201.31	220.76	286.49	283.67	211.29	225.00	232.77
		N	23,140	23,140	23,140	23,139	23,139	23,140	23,140	23,139
	Unknown	Mean	263.49	190.76	206.09	268.08	266.02	198.65	212.58	218.66
		N	7,834	7,837	7,837	7,837	7,834	7,837	7,834	7,834
1	Non-Hispanic Asian	Mean	298.78	297.45	274.80	289.42	294.37	286.37	297.85	288.56
		N	5,524	5,523	5,526	5,524	5,522	5,523	5,521	5,519
	Non-Hispanic Pacific Islander	Mean	288.80	285.83	268.20	282.60	285.95	277.26	286.73	279.62
		N	351	351	352	352	351	351	351	351
	Non-Hispanic Black	Mean	292.98	288.44	259.02	282.21	287.86	273.98	289.83	277.93
		N	4,797	4,802	4,801	4,801	4,796	4,801	4,797	4,795
	Hispanic (Of Any Race)	Mean	295.53	288.05	260.67	280.35	288.19	274.61	290.33	278.47
		N	53,688	53,715	53,706	53,705	53,664	53,695	53,679	53,639
	Non-Hispanic American Indian	Mean	290.22	286.58	250.81	257.82	274.22	268.92	287.70	270.32
		N	1,100	1,099	1,100	1,101	1,100	1,098	1,098	1,097
	Non-Hispanic Multi-racial	Mean	301.61	296.17	270.17	293.50	297.87	283.40	297.77	287.56
		N	396	398	398	398	396	398	396	396
	Non-Hispanic White	Mean	299.76	293.26	268.63	297.05	298.67	281.19	295.24	286.23
		N	5,184	5,188	5,187	5,187	5,183	5,187	5,184	5,182
	Unknown	Mean	292.26	290.72	259.26	276.37	284.68	275.22	291.25	277.90
		N	673	673	673	673	672	673	672	672

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
2	Non-Hispanic Asian	Mean	333.56	327.91	310.15	305.94	320.01	319.28	329.68	319.28
		N	5,175	5,173	5,175	5,175	5,175	5,173	5,173	5,173
	Non-Hispanic Pacific Islander	Mean	325.12	318.13	307.01	299.43	312.52	312.74	320.25	312.45
		N	286	285	286	286	286	285	285	285
	Non-Hispanic Black	Mean	324.61	312.54	290.70	295.43	310.29	301.88	316.23	304.18
		N	5,140	5,140	5,140	5,141	5,139	5,138	5,139	5,136
	Hispanic (Of Any Race)	Mean	329.33	316.28	297.15	297.16	313.49	306.97	320.27	308.71
		N	55,943	55,952	55,958	55,946	55,922	55,943	55,930	55,904
	Non-Hispanic American Indian	Mean	318.86	307.93	286.51	267.61	293.49	297.49	311.25	296.06
		N	1,131	1,130	1,131	1,131	1,131	1,130	1,130	1,130
	Non-Hispanic Multi-racial	Mean	331.71	321.50	299.58	305.13	318.68	310.79	324.63	312.92
		N	373	373	373	373	373	373	373	373
	Non-Hispanic White	Mean	333.80	322.90	303.24	311.81	323.06	313.34	326.25	316.04
		N	5,223	5,221	5,225	5,224	5,223	5,221	5,220	5,220
	Unknown	Mean	323.25	316.87	294.10	291.57	307.67	305.75	318.85	306.09
		N	649	649	649	649	649	649	649	649
3	Non-Hispanic Asian	Mean	364.12	347.92	324.85	320.83	342.72	336.64	352.84	338.25
		N	3,827	3,826	3,827	3,826	3,826	3,826	3,826	3,825
	Non-Hispanic Pacific Islander	Mean	347.71	335.23	322.75	307.09	327.65	329.20	339.03	328.44
		N	308	308	307	308	308	307	308	307
	Non-Hispanic Black	Mean	353.12	335.61	306.10	309.31	331.47	321.12	340.93	323.99
		N	4,978	4,974	4,977	4,976	4,976	4,973	4,974	4,971
	Hispanic (Of Any Race)	Mean	357.03	339.64	314.01	311.87	334.70	327.08	344.92	329.16
		N	46697	46681	46694	46669	46666	46671	46679	46642
	Non-Hispanic American Indian	Mean	339.20	330.31	310.05	290.29	315.07	320.45	333.01	318.63
		N	1,098	1,098	1,100	1,100	1,098	1,098	1,098	1,098
	Non-Hispanic Multi-racial	Mean	358.05	339.16	312.83	316.48	337.53	326.25	344.89	329.44
		N	351	351	351	351	351	351	351	351
	Non-Hispanic White	Mean	364.15	345.21	319.61	327.81	346.24	332.65	350.96	336.54
		N	3,895	3,896	3,896	3,893	3,892	3,896	3,895	3,892
	Unknown	Mean	352.71	335.92	307.08	295.04	324.10	321.74	341.02	322.25
		N	478	478	478	478	478	478	478	478
4	Non-Hispanic Asian	Mean	368.59	352.14	341.12	339.45	354.28	346.88	357.17	348.88
		N	1,750	1,750	1,750	1,750	1,750	1,750	1,750	1,750
	Non-Hispanic Pacific Islander	Mean	353.56	339.90	332.67	313.79	333.90	336.48	344.08	335.49
		N	165	165	165	165	165	165	165	165
	Non-Hispanic Black	Mean	363.83	341.83	328.77	335.99	350.17	335.54	348.52	339.73
		N	2,931	2,929	2,931	2,931	2,930	2,928	2,928	2,927
	Hispanic (Of Any Race)	Mean	365.43	345.06	333.58	334.67	350.30	339.58	351.26	342.57
		N	26,097	26,082	26,096	26,100	26,094	26,075	26,077	26,067
	Non-Hispanic American Indian	Mean	352.56	336.12	324.15	307.59	330.34	330.38	341.15	330.19
		N	976	976	976	978	976	975	975	974
	Non-Hispanic Multi-racial	Mean	366.65	347.09	333.79	343.75	355.47	340.68	353.06	344.93
		N	208	208	208	208	208	208	208	208
	Non-Hispanic White	Mean	367.52	347.76	335.59	344.06	356.05	341.94	353.77	345.94
		N	2,125	2,124	2,125	2,125	2,125	2,124	2,124	2,124
	Unknown	Mean	346.87	335.65	314.90	302.19	324.76	325.53	339.10	325.08
		N	327	327	327	327	327	327	327	327

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
5	Non-Hispanic Asian	Mean	378.52	359.05	348.36	344.55	361.80	353.98	365.00	356.09
		N	1,470	1,469	1,470	1,470	1,470	1,469	1,469	1,469
	Non-Hispanic Pacific Islander	Mean	363.02	343.32	343.68	330.73	347.16	343.73	349.31	344.52
		N	131	131	131	131	131	131	131	131
	Non-Hispanic Black	Mean	373.54	349.70	338.31	345.92	359.98	344.28	356.94	348.76
		N	3,084	3,077	3,082	3,084	3,084	3,076	3,077	3,076
	Hispanic (Of Any Race)	Mean	377.50	354.40	344.25	343.89	360.94	349.59	361.42	352.75
		N	24,710	24,696	24,707	24,709	24,708	24,690	24,693	24,685
	Non-Hispanic American Indian	Mean	363.28	345.54	335.41	318.53	341.14	340.73	350.95	340.62
		N	910	910	910	910	910	910	910	910
	Non-Hispanic Multi-racial	Mean	374.97	353.02	341.44	347.22	361.38	347.48	359.71	351.42
		N	185	185	185	185	185	185	185	185
	Non-Hispanic White	Mean	379.43	358.05	345.66	356.03	367.99	352.16	364.60	356.69
		N	1,886	1,884	1,886	1,886	1,886	1,884	1,884	1,884
	Unknown	Mean	358.43	344.53	329.98	314.71	336.82	337.49	348.86	337.07
		N	215	215	215	215	215	215	215	215
6	Non-Hispanic Asian	Mean	381.64	358.93	337.53	359.31	370.73	348.44	365.77	354.91
		N	1,099	1,099	1,098	1,099	1,099	1,098	1,099	1,098
	Non-Hispanic Pacific Islander	Mean	365.68	343.71	325.44	320.45	343.29	334.80	350.27	337.12
		N	85	85	85	85	85	85	85	85
	Non-Hispanic Black	Mean	372.77	348.42	321.19	348.76	361.01	335.05	355.76	342.62
		N	2,241	2,241	2,241	2,240	2,240	2,241	2,241	2,240
	Hispanic (Of Any Race)	Mean	375.52	351.52	326.96	352.03	364.02	339.49	358.76	346.64
		N	18,128	18,124	18,122	18,126	18,124	18,117	18,124	18,113
	Non-Hispanic American Indian	Mean	363.61	347.10	322.25	322.71	343.42	334.92	352.09	337.24
		N	801	801	801	801	801	801	801	801
	Non-Hispanic Multi-racial	Mean	372.29	350.58	324.90	352.04	362.47	337.96	357.11	345.12
		N	155	155	155	155	155	155	155	155
	Non-Hispanic White	Mean	380.01	357.83	330.57	364.14	372.33	344.45	364.52	352.60
		N	1,514	1,514	1,514	1,514	1,514	1,514	1,514	1,514
	Unknown	Mean	359.75	347.61	313.97	328.87	344.88	331.14	351.29	335.08
		N	249	249	248	250	249	248	249	248
7	Non-Hispanic Asian	Mean	388.91	366.44	342.55	363.13	376.28	354.79	373.25	361.01
		N	1,015	1,015	1,010	1,015	1,015	1,010	1,015	1,010
	Non-Hispanic Pacific Islander	Mean	372.18	353.15	337.01	350.99	361.82	345.38	358.93	350.06
		N	68	68	68	68	68	68	68	68
	Non-Hispanic Black	Mean	376.52	353.03	325.78	351.43	364.24	339.66	360.14	346.85
		N	1,941	1,942	1,941	1,942	1,941	1,941	1,941	1,940
	Hispanic (Of Any Race)	Mean	378.75	356.58	330.07	352.17	365.70	343.58	363.29	349.99
		N	16,255	16,254	16,250	16,258	16,252	16,243	16,250	16,236
	Non-Hispanic American Indian	Mean	373.90	354.74	331.30	336.49	355.40	343.28	360.55	346.69
		N	809	809	809	808	808	809	809	808
	Non-Hispanic Multi-racial	Mean	381.10	358.96	330.49	354.07	367.86	345.01	365.68	351.64
		N	125	125	125	125	125	125	125	125
	Non-Hispanic White	Mean	387.38	365.01	337.26	373.74	380.81	351.38	371.78	359.95
		N	1,398	1,398	1,393	1,398	1,398	1,393	1,398	1,393
	Unknown	Mean	368.25	355.02	322.17	337.70	353.21	338.80	359.03	342.85
		N	145	144	144	145	145	144	144	144

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
8	Non-Hispanic Asian	Mean	394.27	374.01	349.11	371.97	383.39	361.83	380.12	368.09
		N	1,044	1,044	1,044	1,042	1,042	1,044	1,044	1,042
	Non-Hispanic Pacific Islander	Mean	385.17	361.63	346.37	376.47	381.05	354.22	368.72	362.03
		N	78	78	78	78	78	78	78	78
	Non-Hispanic Black	Mean	382.63	360.48	333.06	358.51	370.85	347.04	367.17	354.01
		N	1,851	1,852	1,851	1,852	1,851	1,851	1,851	1,850
	Hispanic (Of Any Race)	Mean	383.83	363.24	335.37	356.95	370.65	349.57	369.46	355.69
		N	16,074	16,073	16,067	16,073	16,069	16,063	16,070	16,056
	Non-Hispanic American Indian	Mean	387.24	365.27	345.81	352.00	369.85	355.81	371.91	359.80
		N	694	694	694	694	694	694	694	694
	Non-Hispanic Multi-racial	Mean	401.01	374.83	345.31	377.01	389.33	360.32	382.69	368.78
		N	100	100	100	100	100	100	100	100
	Non-Hispanic White	Mean	393.84	374.14	343.31	378.10	386.26	358.98	380.08	366.95
		N	1,289	1,289	1,289	1,288	1,288	1,289	1,289	1,288
	Unknown	Mean	374.18	362.68	325.29	350.56	362.66	344.24	366.18	349.53
		N	131	131	131	131	131	131	131	131
9	Non-Hispanic Asian	Mean	385.66	383.02	365.36	357.44	371.79	374.44	383.85	373.44
		N	1,067	1,067	1,066	1,067	1,067	1,066	1,067	1,066
	Non-Hispanic Pacific Islander	Mean	372.56	369.16	351.48	369.06	371.02	360.55	370.21	363.46
		N	149	149	149	149	149	149	149	149
	Non-Hispanic Black	Mean	367.43	370.17	346.51	338.24	353.06	358.61	369.41	356.75
		N	2,278	2,279	2,278	2,278	2,277	2,277	2,278	2,275
	Hispanic (Of Any Race)	Mean	374.61	373.12	350.63	340.31	357.71	362.13	373.62	360.59
		N	17,146	17,143	17,141	17,145	17,139	17,135	17,140	17,126
	Non-Hispanic American Indian	Mean	379.63	361.93	348.60	329.49	354.80	355.54	367.31	355.06
		N	557	557	557	557	557	557	557	557
	Non-Hispanic Multi-racial	Mean	381.85	378.14	352.01	355.31	368.82	365.32	379.32	366.16
		N	94	94	94	94	94	94	94	94
	Non-Hispanic White	Mean	384.00	381.89	356.52	357.86	371.13	369.47	382.56	369.74
		N	1343	1342	1344	1344	1343	1342	1342	1342
	Unknown	Mean	359.91	367.93	341.12	327.53	343.97	354.78	365.60	351.32
		N	335	335	335	335	335	335	335	335
10	Non-Hispanic Asian	Mean	391.86	386.67	369.47	356.99	374.65	378.32	388.30	376.99
		N	893	893	893	893	893	893	893	893
	Non-Hispanic Pacific Islander	Mean	377.56	371.70	361.33	361.87	370.00	366.83	373.55	367.53
		N	89	89	89	89	89	89	89	89
	Non-Hispanic Black	Mean	376.40	374.37	354.25	346.04	361.44	364.57	375.06	363.42
		N	1,920	1,919	1,917	1,920	1,920	1,916	1,919	1,916
	Hispanic (Of Any Race)	Mean	380.72	378.14	355.84	345.15	363.16	367.24	378.99	365.81
		N	13,774	13,775	13,772	13,778	13,773	13,770	13,773	13,768
	Non-Hispanic American Indian	Mean	379.65	366.69	352.33	334.59	357.36	359.75	370.62	358.84
		N	439	439	438	439	439	438	439	438
	Non-Hispanic Multi-racial	Mean	386.86	381.63	360.20	362.91	375.15	371.15	383.25	372.17
		N	88	88	88	88	88	88	88	88
	Non-Hispanic White	Mean	389.68	385.80	363.10	364.34	377.24	374.72	387.04	375.28
		N	1,114	1,114	1,113	1,114	1,114	1,113	1,114	1,113
	Unknown	Mean	364.75	372.87	348.16	334.84	350.02	360.78	370.49	357.31
		N	147	147	147	147	147	147	147	147

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
11	Non-Hispanic Asian	Mean	395.90	391.12	373.98	367.03	381.70	382.80	392.60	382.26
		N	764	764	764	764	764	764	764	764
	Non-Hispanic Pacific Islander	Mean	388.75	372.04	367.06	366.57	377.89	369.81	377.13	371.98
		N	53	53	53	53	53	53	53	53
	Non-Hispanic Black	Mean	375.38	376.94	356.58	346.17	360.97	367.00	376.54	364.97
		N	2,058	2,058	2,058	2,060	2,058	2,057	2,058	2,057
	Hispanic (Of Any Race)	Mean	384.66	383.39	362.56	352.45	368.81	373.24	383.83	371.69
		N	10,578	10,577	10,579	10,578	10,575	10,575	10,575	10,571
	Non-Hispanic American Indian	Mean	385.92	370.90	356.42	342.95	364.67	363.93	375.44	363.90
		N	328	328	328	328	328	328	328	328
	Non-Hispanic Multi-racial	Mean	386.10	384.14	361.73	358.29	372.77	373.40	385.10	372.96
		N	78	79	78	79	78	78	78	78
	Non-Hispanic White	Mean	393.89	390.81	369.70	367.71	381.05	380.49	391.79	380.46
		N	943	943	940	943	943	940	943	940
	Unknown	Mean	379.40	380.73	354.85	336.61	358.26	368.07	380.42	364.99
		N	98	98	98	98	98	98	98	98
12	Non-Hispanic Asian	Mean	396.69	393.37	376.89	367.02	382.11	385.38	394.53	384.24
		N	618	619	619	618	618	619	618	618
	Non-Hispanic Pacific Islander	Mean	393.35	378.48	369.40	384.42	389.15	374.21	383.06	378.46
		N	52	52	52	52	52	52	52	52
	Non-Hispanic Black	Mean	378.76	378.60	361.45	353.40	366.34	370.29	378.72	368.93
		N	1,727	1,729	1,725	1,728	1,726	1,725	1,727	1,722
	Hispanic (Of Any Race)	Mean	392.52	387.50	369.44	366.95	379.97	378.73	389.08	378.89
		N	6,587	6,587	6,585	6,586	6,584	6,584	6,586	6,581
	Non-Hispanic American Indian	Mean	399.10	376.94	363.41	367.52	383.63	370.41	383.76	374.33
		N	220	221	221	221	220	221	220	220
	Non-Hispanic Multi-racial	Mean	390.94	389.76	375.13	374.28	382.89	382.70	390.24	382.52
		N	54	54	54	54	54	54	54	54
	Non-Hispanic White	Mean	400.08	393.36	372.04	378.65	389.60	382.95	395.43	384.74
		N	569	569	569	569	569	569	569	569
	Unknown	Mean	382.74	382.98	362.95	360.45	371.82	373.21	382.99	372.57
		N	82	82	82	82	82	82	82	82

4.2.3 Correlations Among Scale Scores by Grade Level Cluster

Table 4.2.3A

Correlations Among Scale Scores: K S401 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.532	.558	.772
	N	243,666	243,657	243,659	243,655
Reading	Pearson Correlation		1	.724	.489
	N		243,664	243,659	243,653
Writing	Pearson Correlation			1	.534
	N			243,666	243,655
Speaking	Pearson Correlation				1
	N				243,662

Table 4.2.3B

Correlations Among Scale Scores: 1 S401 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.567	.535	.523
	N	71,713	71,698	71,689	71,684
Reading	Pearson Correlation		1	.541	.440
	N		71,749	71,726	71,720
Writing	Pearson Correlation			1	.502
	N			71,743	71,715
Speaking	Pearson Correlation				1
	N				71,741

Table 4.2.3C

Correlations Among Scale Scores: 2 S401 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.615	.548	.543
	N	73,920	73,899	73,905	73,898
Reading	Pearson Correlation		1	.639	.517
	N		73,923	73,912	73,903
Writing	Pearson Correlation			1	.532
	N			73,937	73,910
Speaking	Pearson Correlation				1
	N				73,925

Table 4.2.3D

Correlations Among Scale Scores: 3 S401 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.692	.553	.533
	N	61,632	61,609	61,620	61,595
Reading	Pearson Correlation		1	.612	.531
	N		61,612	61,600	61,575
Writing	Pearson Correlation			1	.532
	N			61,630	61,593
Speaking	Pearson Correlation				1
	N				61,601

Table 4.2.3E

Correlations Among Scale Scores: 4-5 S401 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.732	.615	.603
	N	67,170	67,118	67,153	67,164
Reading	Pearson Correlation		1	.667	.599
	N		67,128	67,112	67,120
Writing	Pearson Correlation			1	.602
	N			67,164	67,157
Speaking	Pearson Correlation				1
	N				67,174

Table 4.2.3F

Correlations Among Scale Scores: 6-8 S401 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.700	.687	.656
	N	67,289	67,275	67,245	67,272
Reading	Pearson Correlation		1	.667	.589
	N		67,284	67,242	67,267
Writing	Pearson Correlation			1	.658
	N			67,258	67,242
Speaking	Pearson Correlation				1
	N				67,287

Table 4.2.3G

Correlations Among Scale Scores: 9-12 S401 Paper

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.721	.676	.667
	N	66,242	66,229	66,209	66,226
Reading	Pearson Correlation		1	.688	.625
	N		66,243	66,208	66,226
Writing	Pearson Correlation			1	.660
	N			66,226	66,211
Speaking	Pearson Correlation				1
	N				66,250

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 S401 Paper

Cluster	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	61,785	23,495	20,691	13,932	36,936	86,827	243,666
1	A	2,474	2,147	5,669	17,955	n/a	n/a	28,245
	B	198	349	1,634	1,213	22,435	n/a	25,829
	C	78	566	2,356	1,557	6,583	6,499	17,639
2	A	1,706	2,334	1,877	4,447	n/a	n/a	10,364
	B	84	383	1,618	3,315	21,263	n/a	26,663
	C	58	842	7,136	4,213	11,789	12,855	36,893
3	A	734	1,700	1,822	2,609	n/a	n/a	6,865
	B	55	845	2,484	3,164	9,135	n/a	15,683
	C	2	296	2,346	7,259	8,327	20,854	39,084
4-5	A	1,496	3,540	1,710	3,861	n/a	n/a	10,607
	B	119	818	2,257	3,037	9,703	n/a	15,934
	C	31	535	3,377	5,275	14,437	16,974	40,629
6-8	A	4,493	4,809	2,857	2,475	n/a	n/a	14,634
	B	450	3,218	4,721	3,351	5,412	n/a	17,152
	C	29	443	2,696	7,117	9,398	15,820	35,503
9-12	A	7,873	6,007	1,686	1,115	n/a	n/a	16,681
	B	798	2,808	5,751	4,288	4,185	n/a	17,830
	C	266	1,758	6,442	10,188	6,999	6,078	31,731

Table 4.3.1.1B

Proficiency Level by Cluster By Tier (Percent): Listening S401 Paper

Cluster	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	25.4%	9.6%	8.5%	5.7%	15.2%	35.6%	100.0%
1	A	8.8%	7.6%	20.1%	63.6%	n/a	n/a	100.0%
	B	0.8%	1.4%	6.3%	4.7%	86.9%	n/a	100.0%
	C	0.4%	3.2%	13.4%	8.8%	37.3%	36.8%	100.0%
2	A	16.5%	22.5%	18.1%	42.9%	n/a	n/a	100.0%
	B	0.3%	1.4%	6.1%	12.4%	79.7%	n/a	100.0%
	C	0.2%	2.3%	19.3%	11.4%	32.0%	34.8%	100.0%
3	A	10.7%	24.8%	26.5%	38.0%	n/a	n/a	100.0%
	B	0.4%	5.4%	15.8%	20.2%	58.2%	n/a	100.0%
	C	0.0%	0.8%	6.0%	18.6%	21.3%	53.4%	100.0%
4-5	A	14.1%	33.4%	16.1%	36.4%	n/a	n/a	100.0%
	B	0.7%	5.1%	14.2%	19.1%	60.9%	n/a	100.0%
	C	0.1%	1.3%	8.3%	13.0%	35.5%	41.8%	100.0%
6-8	A	30.7%	32.9%	19.5%	16.9%	n/a	n/a	100.0%
	B	2.6%	18.8%	27.5%	19.5%	31.6%	n/a	100.0%
	C	0.1%	1.2%	7.6%	20.0%	26.5%	44.6%	100.0%
9-12	A	47.2%	36.0%	10.1%	6.7%	n/a	n/a	100.0%
	B	4.5%	15.7%	32.3%	24.0%	23.5%	n/a	100.0%
	C	0.8%	5.5%	20.3%	32.1%	22.1%	19.2%	100.0%

4.3.1.2 By Grade by Tier

Table 4.3.1.2A

Proficiency Level by Grade By Tier (Count): Listening S401 Paper

Grade	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	61,785	23,495	20,691	13,932	36,936	86,827	243,666
1	A	2,474	2,147	5,669	17,955	n/a	n/a	28,245
	B	198	349	1,634	1,213	22,435	n/a	25,829
	C	78	566	2,356	1,557	6,583	6,499	17,639
2	A	1,706	2,334	1,877	4,447	n/a	n/a	10,364
	B	84	383	1,618	3,315	21,263	n/a	26,663
	C	58	842	7,136	4,213	11,789	12,855	36,893
3	A	734	1,700	1,822	2,609	n/a	n/a	6,865
	B	55	845	2,484	3,164	9,135	n/a	15,683
	C	2	296	2,346	7,259	8,327	20,854	39,084
4	A	717	1,847	964	2,172	n/a	n/a	5,700
	B	50	424	1,216	1,660	5,455	n/a	8,805
	C	9	156	1,823	2,639	6,986	8,461	20,074
5	A	779	1,693	746	1,689	n/a	n/a	4,907
	B	69	394	1,041	1,377	4,248	n/a	7,129
	C	22	379	1,554	2,636	7,451	8,513	20,555
6	A	1,257	1,552	1,236	1,003	n/a	n/a	5,048
	B	90	789	1,454	1,314	2,012	n/a	5,659
	C	10	138	974	3,142	3,534	5,767	13,565
7	A	1,620	1,492	1,011	717	n/a	n/a	4,840
	B	165	993	1,625	1,412	1,696	n/a	5,891
	C	14	194	795	2,488	3,078	4,456	11,025
8	A	1,616	1,765	610	755	n/a	n/a	4,746
	B	195	1,436	1,642	625	1,704	n/a	5,602
	C	5	111	927	1,487	2,786	5,597	10,913
9	A	2,994	3,092	762	348	n/a	n/a	7,196
	B	99	774	1,696	1,350	1,758	n/a	5,677
	C	31	265	1,851	3,264	2,439	2,246	10,096
10	A	2,148	1,593	590	338	n/a	n/a	4,669
	B	166	708	1,631	1,557	1,008	n/a	5,070
	C	38	526	1,427	2,716	1,988	2,030	8,725
11	A	1,851	1,041	232	295	n/a	n/a	3,419
	B	241	848	1,477	828	828	n/a	4,222
	C	58	416	1,862	2,375	1,475	1,073	7,259
12	A	880	281	102	134	n/a	n/a	1,397
	B	292	478	947	553	591	n/a	2,861
	C	139	551	1,302	1,833	1,097	729	5,651

Table 4.3.1.2B

Proficiency Level by Grade By Tier (Percent): Listening S401 Paper

Grade	Tier	Listening Proficiency Range						Total
		1	2	3	4	5	6	
K	-	25.4%	9.6%	8.5%	5.7%	15.2%	35.6%	100.0%
1	A	8.8%	7.6%	20.1%	63.6%	n/a	n/a	100.0%
	B	0.8%	1.4%	6.3%	4.7%	86.9%	n/a	100.0%
	C	0.4%	3.2%	13.4%	8.8%	37.3%	36.8%	100.0%
2	A	16.5%	22.5%	18.1%	42.9%	n/a	n/a	100.0%
	B	0.3%	1.4%	6.1%	12.4%	79.7%	n/a	100.0%
	C	0.2%	2.3%	19.3%	11.4%	32.0%	34.8%	100.0%
3	A	10.7%	24.8%	26.5%	38.0%	n/a	n/a	100.0%
	B	0.4%	5.4%	15.8%	20.2%	58.2%	n/a	100.0%
	C	0.0%	0.8%	6.0%	18.6%	21.3%	53.4%	100.0%
4	A	12.6%	32.4%	16.9%	38.1%	n/a	n/a	100.0%
	B	0.6%	4.8%	13.8%	18.9%	62.0%	n/a	100.0%
	C	0.0%	0.8%	9.1%	13.1%	34.8%	42.1%	100.0%
5	A	15.9%	34.5%	15.2%	34.4%	n/a	n/a	100.0%
	B	1.0%	5.5%	14.6%	19.3%	59.6%	n/a	100.0%
	C	0.1%	1.8%	7.6%	12.8%	36.2%	41.4%	100.0%
6	A	24.9%	30.7%	24.5%	19.9%	n/a	n/a	100.0%
	B	1.6%	13.9%	25.7%	23.2%	35.6%	n/a	100.0%
	C	0.1%	1.0%	7.2%	23.2%	26.1%	42.5%	100.0%
7	A	33.5%	30.8%	20.9%	14.8%	n/a	n/a	100.0%
	B	2.8%	16.9%	27.6%	24.0%	28.8%	n/a	100.0%
	C	0.1%	1.8%	7.2%	22.6%	27.9%	40.4%	100.0%
8	A	34.0%	37.2%	12.9%	15.9%	n/a	n/a	100.0%
	B	3.5%	25.6%	29.3%	11.2%	30.4%	n/a	100.0%
	C	0.0%	1.0%	8.5%	13.6%	25.5%	51.3%	100.0%
9	A	41.6%	43.0%	10.6%	4.8%	n/a	n/a	100.0%
	B	1.7%	13.6%	29.9%	23.8%	31.0%	n/a	100.0%
	C	0.3%	2.6%	18.3%	32.3%	24.2%	22.2%	100.0%
10	A	46.0%	34.1%	12.6%	7.2%	n/a	n/a	100.0%
	B	3.3%	14.0%	32.2%	30.7%	19.9%	n/a	100.0%
	C	0.4%	6.0%	16.4%	31.1%	22.8%	23.3%	100.0%
11	A	54.1%	30.4%	6.8%	8.6%	n/a	n/a	100.0%
	B	5.7%	20.1%	35.0%	19.6%	19.6%	n/a	100.0%
	C	0.8%	5.7%	25.7%	32.7%	20.3%	14.8%	100.0%
12	A	63.0%	20.1%	7.3%	9.6%	n/a	n/a	100.0%
	B	10.2%	16.7%	33.1%	19.3%	20.7%	n/a	100.0%
	C	2.5%	9.8%	23.0%	32.4%	19.4%	12.9%	100.0%

4.3.1.3 By Grade

Table 4.3.1.3A

Proficiency Level by Grade (Count): Listening S401 Paper

	Listening Proficiency Range						Total
	1	2	3	4	5	6	
K	61,785	23,495	20,691	13,932	36,936	86,827	243,666
1	2,750	3,062	9,659	20,725	29,018	6,499	71,713
2	1,848	3,559	10,631	11,975	33,052	12,855	73,920
3	791	2,841	6,652	13,032	17,462	20,854	61,632
4	776	2,427	4,003	6,471	12,441	8,461	34,579
5	870	2,466	3,341	5,702	11,699	8,513	32,591
6	1,357	2,479	3,664	5,459	5,546	5,767	24,272
7	1,799	2,679	3,431	4,617	4,774	4,456	21,756
8	1,816	3,312	3,179	2,867	4,490	5,597	21,261
9	3,124	4,131	4,309	4,962	4,197	2,246	22,969
10	2,352	2,827	3,648	4,611	2,996	2,030	18,464
11	2,150	2,305	3,571	3,498	2,303	1,073	14,900
12	1,311	1,310	2,351	2,520	1,688	729	9,909

Table 4.3.1.3B

Proficiency Level by Grade (Percent): Listening S401 Paper

	Listening Proficiency Range						Total
	1	2	3	4	5	6	
K	25.4%	9.6%	8.5%	5.7%	15.2%	35.6%	100.0%
1	3.8%	4.3%	13.5%	28.9%	40.5%	9.1%	100.0%
2	2.5%	4.8%	14.4%	16.2%	44.7%	17.4%	100.0%
3	1.3%	4.6%	10.8%	21.1%	28.3%	33.8%	100.0%
4	2.2%	7.0%	11.6%	18.7%	36.0%	24.5%	100.0%
5	2.7%	7.6%	10.3%	17.5%	35.9%	26.1%	100.0%
6	5.6%	10.2%	15.1%	22.5%	22.8%	23.8%	100.0%
7	8.3%	12.3%	15.8%	21.2%	21.9%	20.5%	100.0%
8	8.5%	15.6%	15.0%	13.5%	21.1%	26.3%	100.0%
9	13.6%	18.0%	18.8%	21.6%	18.3%	9.8%	100.0%
10	12.7%	15.3%	19.8%	25.0%	16.2%	11.0%	100.0%
11	14.4%	15.5%	24.0%	23.5%	15.5%	7.2%	100.0%
12	13.2%	13.2%	23.7%	25.4%	17.0%	7.4%	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 S401 Paper

Cluster	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	171,625	8,663	23,868	15,406	24,102	0	243,664
1	A	11,056	9,871	4,206	3,131	n/a	n/a	28,264
	B	703	5,414	10,004	3,261	6,453	n/a	25,835
	C	281	2,895	4,867	3,966	3,318	2,323	17,650
2	A	5,857	2,073	1,361	1,073	n/a	n/a	10,364
	B	1,652	6,871	7,374	2,872	7,892	n/a	26,661
	C	1,182	7,355	9,029	4,832	6,893	7,607	36,898
3	A	3,317	2,011	997	540	n/a	n/a	6,865
	B	1,159	3,832	5,874	1,849	2,956	n/a	15,670
	C	275	1,605	12,128	7,948	10,338	6,783	39,077
4-5	A	5,473	2,831	1,259	1,043	n/a	n/a	10,606
	B	1,531	5,235	4,333	1,898	2,922	n/a	15,919
	C	202	4,835	13,611	7,766	8,779	5,410	40,603
6-8	A	7,139	5,336	1,279	877	n/a	n/a	14,631
	B	2,065	6,939	4,192	1,457	2,498	n/a	17,151
	C	1,168	10,685	12,755	3,513	4,411	2,970	35,502
9-12	A	6,490	6,803	1,864	1,522	n/a	n/a	16,679
	B	2,930	8,278	3,630	1,029	1,965	n/a	17,832
	C	501	6,348	8,975	3,965	6,833	5,110	31,732

Table 4.3.2.1B

Proficiency Level by Cluster By Tier (Percent): Reading S401 Paper

Cluster	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	70.4%	3.6%	9.8%	6.3%	9.9%	0.0%	100.0%
1	A	39.1%	34.9%	14.9%	11.1%	n/a	n/a	100.0%
	B	2.7%	21.0%	38.7%	12.6%	25.0%	n/a	100.0%
	C	1.6%	16.4%	27.6%	22.5%	18.8%	13.2%	100.0%
2	A	56.5%	20.0%	13.1%	10.4%	n/a	n/a	100.0%
	B	6.2%	25.8%	27.7%	10.8%	29.6%	n/a	100.0%
	C	3.2%	19.9%	24.5%	13.1%	18.7%	20.6%	100.0%
3	A	48.3%	29.3%	14.5%	7.9%	n/a	n/a	100.0%
	B	7.4%	24.5%	37.5%	11.8%	18.9%	n/a	100.0%
	C	0.7%	4.1%	31.0%	20.3%	26.5%	17.4%	100.0%
4-5	A	51.6%	26.7%	11.9%	9.8%	n/a	n/a	100.0%
	B	9.6%	32.9%	27.2%	11.9%	18.4%	n/a	100.0%
	C	0.5%	11.9%	33.5%	19.1%	21.6%	13.3%	100.0%
6-8	A	48.8%	36.5%	8.7%	6.0%	n/a	n/a	100.0%
	B	12.0%	40.5%	24.4%	8.5%	14.6%	n/a	100.0%
	C	3.3%	30.1%	35.9%	9.9%	12.4%	8.4%	100.0%
9-12	A	38.9%	40.8%	11.2%	9.1%	n/a	n/a	100.0%
	B	16.4%	46.4%	20.4%	5.8%	11.0%	n/a	100.0%
	C	1.6%	20.0%	28.3%	12.5%	21.5%	16.1%	100.0%

4.3.2.2 By Grade by Tier

Table 4.3.2.2A

Proficiency Level by Grade By Tier (Count): Reading S401 Paper

Grade	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	171,625	8,663	23,868	15,406	24,102	0	243,664
1	A	11,056	9,871	4,206	3,131	n/a	n/a	28,264
	B	703	5,414	10,004	3,261	6,453	n/a	25,835
	C	281	2,895	4,867	3,966	3,318	2,323	17,650
2	A	5,857	2,073	1,361	1,073	n/a	n/a	10,364
	B	1,652	6,871	7,374	2,872	7,892	n/a	26,661
	C	1,182	7,355	9,029	4,832	6,893	7,607	36,898
3	A	3,317	2,011	997	540	n/a	n/a	6,865
	B	1,159	3,832	5,874	1,849	2,956	n/a	15,670
	C	275	1,605	12,128	7,948	10,338	6,783	39,077
4	A	2,791	1,534	780	596	n/a	n/a	5,701
	B	664	2,899	2,421	1,043	1,766	n/a	8,793
	C	71	1,962	6,395	4,900	4,225	2,514	20,067
5	A	2,682	1,297	479	447	n/a	n/a	4,905
	B	867	2,336	1,912	855	1,156	n/a	7,126
	C	131	2,873	7,216	2,866	4,554	2,896	20,536
6	A	2,258	2,017	433	339	n/a	n/a	5,047
	B	612	2,385	1,381	546	733	n/a	5,657
	C	366	4,002	5,665	1,497	1,309	725	13,564
7	A	2,369	1,791	392	284	n/a	n/a	4,836
	B	689	2,253	1,486	607	857	n/a	5,892
	C	391	3,698	3,331	1,307	1,397	903	11,027
8	A	2,512	1,528	454	254	n/a	n/a	4,748
	B	764	2,301	1,325	304	908	n/a	5,602
	C	411	2,985	3,759	709	1,705	1,342	10,911
9	A	3,229	2,814	529	623	n/a	n/a	7,195
	B	808	2,423	1,311	510	623	n/a	5,675
	C	69	1,639	2,775	1,294	2,536	1,783	10,096
10	A	1,635	1,951	617	465	n/a	n/a	4,668
	B	756	2,462	1,035	223	595	n/a	5,071
	C	126	1,638	2,547	1,113	1,997	1,304	8,725
11	A	1,112	1,457	560	289	n/a	n/a	3,418
	B	773	1,987	848	189	426	n/a	4,223
	C	169	1,506	1,972	883	1,364	1,365	7,259
12	A	514	581	158	145	n/a	n/a	1,398
	B	593	1,406	436	107	321	n/a	2,863
	C	137	1,565	1,681	675	936	658	5,652

Table 4.3.2.2B

Proficiency Level by Grade By Tier (Percent): Reading S401 Paper

Grade	Tier	Reading Proficiency Range						Total
		1	2	3	4	5	6	
K	-	70.4%	3.6%	9.8%	6.3%	9.9%	0.0%	100.0%
1	A	39.1%	34.9%	14.9%	11.1%	n/a	n/a	100.0%
	B	2.7%	21.0%	38.7%	12.6%	25.0%	n/a	100.0%
	C	1.6%	16.4%	27.6%	22.5%	18.8%	13.2%	100.0%
2	A	56.5%	20.0%	13.1%	10.4%	n/a	n/a	100.0%
	B	6.2%	25.8%	27.7%	10.8%	29.6%	n/a	100.0%
	C	3.2%	19.9%	24.5%	13.1%	18.7%	20.6%	100.0%
3	A	48.3%	29.3%	14.5%	7.9%	n/a	n/a	100.0%
	B	7.4%	24.5%	37.5%	11.8%	18.9%	n/a	100.0%
	C	0.7%	4.1%	31.0%	20.3%	26.5%	17.4%	100.0%
4	A	49.0%	26.9%	13.7%	10.5%	n/a	n/a	100.0%
	B	7.6%	33.0%	27.5%	11.9%	20.1%	n/a	100.0%
	C	0.4%	9.8%	31.9%	24.4%	21.1%	12.5%	100.0%
5	A	54.7%	26.4%	9.8%	9.1%	n/a	n/a	100.0%
	B	12.2%	32.8%	26.8%	12.0%	16.2%	n/a	100.0%
	C	0.6%	14.0%	35.1%	14.0%	22.2%	14.1%	100.0%
6	A	44.7%	40.0%	8.6%	6.7%	n/a	n/a	100.0%
	B	10.8%	42.2%	24.4%	9.7%	13.0%	n/a	100.0%
	C	2.7%	29.5%	41.8%	11.0%	9.7%	5.3%	100.0%
7	A	49.0%	37.0%	8.1%	5.9%	n/a	n/a	100.0%
	B	11.7%	38.2%	25.2%	10.3%	14.5%	n/a	100.0%
	C	3.5%	33.5%	30.2%	11.9%	12.7%	8.2%	100.0%
8	A	52.9%	32.2%	9.6%	5.3%	n/a	n/a	100.0%
	B	13.6%	41.1%	23.7%	5.4%	16.2%	n/a	100.0%
	C	3.8%	27.4%	34.5%	6.5%	15.6%	12.3%	100.0%
9	A	44.9%	39.1%	7.4%	8.7%	n/a	n/a	100.0%
	B	14.2%	42.7%	23.1%	9.0%	11.0%	n/a	100.0%
	C	0.7%	16.2%	27.5%	12.8%	25.1%	17.7%	100.0%
10	A	35.0%	41.8%	13.2%	10.0%	n/a	n/a	100.0%
	B	14.9%	48.6%	20.4%	4.4%	11.7%	n/a	100.0%
	C	1.4%	18.8%	29.2%	12.8%	22.9%	14.9%	100.0%
11	A	32.5%	42.6%	16.4%	8.5%	n/a	n/a	100.0%
	B	18.3%	47.1%	20.1%	4.5%	10.1%	n/a	100.0%
	C	2.3%	20.7%	27.2%	12.2%	18.8%	18.8%	100.0%
12	A	36.8%	41.6%	11.3%	10.4%	n/a	n/a	100.0%
	B	20.7%	49.1%	15.2%	3.7%	11.2%	n/a	100.0%
	C	2.4%	27.7%	29.7%	11.9%	16.6%	11.6%	100.0%

4.3.2.3 By Grade

Table 4.3.2.3A

Proficiency Level by Grade (Count): Reading S401 Paper

	Reading Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	171,625	8,663	23,868	15,406	24,102	0	243,664
1	12,040	18,180	19,077	10,358	9,771	2,323	71,749
2	8,691	16,299	17,764	8,777	14,785	7,607	73,923
3	4,751	7,448	18,999	10,337	13,294	6,783	61,612
4	3,526	6,395	9,596	6,539	5,991	2,514	34,561
5	3,680	6,506	9,607	4,168	5,710	2,896	32,567
6	3,236	8,404	7,479	2,382	2,042	725	24,268
7	3,449	7,742	5,209	2,198	2,254	903	21,755
8	3,687	6,814	5,538	1,267	2,613	1,342	21,261
9	4,106	6,876	4,615	2,427	3,159	1,783	22,966
10	2,517	6,051	4,199	1,801	2,592	1,304	18,464
11	2,054	4,950	3,380	1,361	1,790	1,365	14,900
12	1,244	3,552	2,275	927	1,257	658	9,913

Table 4.3.2.3B

Proficiency Level by Grade (Percent): Reading S401 Paper

	Reading Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	70.4%	3.6%	9.8%	6.3%	9.9%	0.0%	100.0%
1	16.8%	25.3%	26.6%	14.4%	13.6%	3.2%	100.0%
2	11.8%	22.0%	24.0%	11.9%	20.0%	10.3%	100.0%
3	7.7%	12.1%	30.8%	16.8%	21.6%	11.0%	100.0%
4	10.2%	18.5%	27.8%	18.9%	17.3%	7.3%	100.0%
5	11.3%	20.0%	29.5%	12.8%	17.5%	8.9%	100.0%
6	13.3%	34.6%	30.8%	9.8%	8.4%	3.0%	100.0%
7	15.9%	35.6%	23.9%	10.1%	10.4%	4.2%	100.0%
8	17.3%	32.0%	26.0%	6.0%	12.3%	6.3%	100.0%
9	17.9%	29.9%	20.1%	10.6%	13.8%	7.8%	100.0%
10	13.6%	32.8%	22.7%	9.8%	14.0%	7.1%	100.0%
11	13.8%	33.2%	22.7%	9.1%	12.0%	9.2%	100.0%
12	12.5%	35.8%	22.9%	9.4%	12.7%	6.6%	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 S401 Paper

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	147,086	46,105	39,586	10,889	0	0	243,666
1	A	9,859	16,959	1,437	0	0	0	28,255
	B	3,123	10,537	11,730	446	2	0	25,838
	C	609	4,969	11,252	807	11	2	17,650
2	A	3,683	3,334	3,349	1	0	0	10,367
	B	1,871	5,217	17,263	2,307	10	1	26,669
	C	445	3,101	26,740	6,571	40	4	36,901
3	A	2,190	2,757	1,913	4	0	0	6,864
	B	729	1,845	10,738	2,367	6	0	15,685
	C	256	1,510	27,722	9,509	80	4	39,081
4-5	A	2,227	2,358	5,927	95	1	0	10,608
	B	311	773	10,289	4,432	121	7	15,933
	C	102	503	21,320	17,887	763	48	40,623
6-8	A	6,164	5,222	3,210	26	0	0	14,622
	B	1,243	2,146	10,579	3,158	12	0	17,138
	C	505	1,723	21,520	11,681	64	5	35,498
9-12	A	4,938	5,655	5,747	335	0	0	16,675
	B	1,466	2,166	9,846	4,263	80	0	17,821
	C	603	1,070	15,344	14,431	280	2	31,730

Table 4.3.3.1B

Proficiency Level by Cluster By Tier (Percent): Writing S401 Paper

Cluster	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	60.4%	18.9%	16.2%	4.5%	0.0%	0.0%	100.0%
1	A	34.9%	60.0%	5.1%	0.0%	0.0%	0.0%	100.0%
	B	12.1%	40.8%	45.4%	1.7%	0.0%	0.0%	100.0%
	C	3.5%	28.2%	63.8%	4.6%	0.1%	0.0%	100.0%
2	A	35.5%	32.2%	32.3%	0.0%	0.0%	0.0%	100.0%
	B	7.0%	19.6%	64.7%	8.7%	0.0%	0.0%	100.0%
	C	1.2%	8.4%	72.5%	17.8%	0.1%	0.0%	100.0%
3	A	31.9%	40.2%	27.9%	0.1%	0.0%	0.0%	100.0%
	B	4.6%	11.8%	68.5%	15.1%	0.0%	0.0%	100.0%
	C	0.7%	3.9%	70.9%	24.3%	0.2%	0.0%	100.0%
4-5	A	21.0%	22.2%	55.9%	0.9%	0.0%	0.0%	100.0%
	B	2.0%	4.9%	64.6%	27.8%	0.8%	0.0%	100.0%
	C	0.3%	1.2%	52.5%	44.0%	1.9%	0.1%	100.0%
6-8	A	42.2%	35.7%	22.0%	0.2%	0.0%	0.0%	100.0%
	B	7.3%	12.5%	61.7%	18.4%	0.1%	0.0%	100.0%
	C	1.4%	4.9%	60.6%	32.9%	0.2%	0.0%	100.0%
9-12	A	29.6%	33.9%	34.5%	2.0%	0.0%	0.0%	100.0%
	B	8.2%	12.2%	55.2%	23.9%	0.4%	0.0%	100.0%
	C	1.9%	3.4%	48.4%	45.5%	0.9%	0.0%	100.0%

4.3.3.2 By Grade by Tier

Table 4.3.3.2A

Proficiency Level by Grade By Tier (Count): Writing S401 Paper

Grade	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	147,086	46,105	39,586	10,889	0	0	243,666
1	A	9,859	16,959	1,437	0	0	0	28,255
	B	3,123	10,537	11,730	446	2	0	25,838
	C	609	4,969	11,252	807	11	2	17,650
2	A	3,683	3,334	3,349	1	0	0	10,367
	B	1,871	5,217	17,263	2,307	10	1	26,669
	C	445	3,101	26,740	6,571	40	4	36,901
3	A	2,190	2,757	1,913	4	0	0	6,864
	B	729	1,845	10,738	2,367	6	0	15,685
	C	256	1,510	27,722	9,509	80	4	39,081
4	A	1,301	1,273	3,056	70	1	0	5,701
	B	176	417	5,807	2,332	71	3	8,806
	C	51	261	11,210	8,081	437	31	20,071
5	A	926	1,085	2,871	25	0	0	4,907
	B	135	356	4,482	2,100	50	4	7,127
	C	51	242	10,110	9,806	326	17	20,552
6	A	1,831	1,881	1,325	7	0	0	5,044
	B	324	799	3,363	1,168	3	0	5,657
	C	138	800	8,076	4,532	16	1	13,563
7	A	2,046	1,933	849	9	0	0	4,837
	B	409	788	3,628	1,047	8	0	5,880
	C	154	586	6,694	3,566	21	2	11,023
8	A	2,287	1,408	1,036	10	0	0	4,741
	B	510	559	3,588	943	1	0	5,601
	C	213	337	6,750	3,583	27	2	10,912
9	A	2,148	2,426	2,392	224	0	0	7,190
	B	309	488	3,037	1,793	51	0	5,678
	C	84	231	4,060	5,563	156	2	10,096
10	A	1,190	1,756	1,646	75	0	0	4,667
	B	429	654	2,807	1,163	12	0	5,065
	C	133	248	4,155	4,125	64	0	8,725
11	A	1,044	1,203	1,154	19	0	0	3,420
	B	390	642	2,245	934	10	0	4,221
	C	161	323	3,498	3,230	45	0	7,257
12	A	556	270	555	17	0	0	1,398
	B	338	382	1,757	373	7	0	2,857
	C	225	268	3,631	1,513	15	0	5,652

Table 4.3.3.2B

Proficiency Level by Grade By Tier (Percent): Writing S401 Paper

Grade	Tier	Writing Proficiency Range						Total
		1	2	3	4	5	6	
K	-	60.4%	18.9%	16.2%	4.5%	0.0%	0.0%	100.0%
1	A	34.9%	60.0%	5.1%	0.0%	0.0%	0.0%	100.0%
	B	12.1%	40.8%	45.4%	1.7%	0.0%	0.0%	100.0%
	C	3.5%	28.2%	63.8%	4.6%	0.1%	0.0%	100.0%
2	A	35.5%	32.2%	32.3%	0.0%	0.0%	0.0%	100.0%
	B	7.0%	19.6%	64.7%	8.7%	0.0%	0.0%	100.0%
	C	1.2%	8.4%	72.5%	17.8%	0.1%	0.0%	100.0%
3	A	31.9%	40.2%	27.9%	0.1%	0.0%	0.0%	100.0%
	B	4.6%	11.8%	68.5%	15.1%	0.0%	0.0%	100.0%
	C	0.7%	3.9%	70.9%	24.3%	0.2%	0.0%	100.0%
4	A	22.8%	22.3%	53.6%	1.2%	0.0%	0.0%	100.0%
	B	2.0%	4.7%	65.9%	26.5%	0.8%	0.0%	100.0%
	C	0.3%	1.3%	55.9%	40.3%	2.2%	0.2%	100.0%
5	A	18.9%	22.1%	58.5%	0.5%	0.0%	0.0%	100.0%
	B	1.9%	5.0%	62.9%	29.5%	0.7%	0.1%	100.0%
	C	0.2%	1.2%	49.2%	47.7%	1.6%	0.1%	100.0%
6	A	36.3%	37.3%	26.3%	0.1%	0.0%	0.0%	100.0%
	B	5.7%	14.1%	59.4%	20.6%	0.1%	0.0%	100.0%
	C	1.0%	5.9%	59.5%	33.4%	0.1%	0.0%	100.0%
7	A	42.3%	40.0%	17.6%	0.2%	0.0%	0.0%	100.0%
	B	7.0%	13.4%	61.7%	17.8%	0.1%	0.0%	100.0%
	C	1.4%	5.3%	60.7%	32.4%	0.2%	0.0%	100.0%
8	A	48.2%	29.7%	21.9%	0.2%	0.0%	0.0%	100.0%
	B	9.1%	10.0%	64.1%	16.8%	0.0%	0.0%	100.0%
	C	2.0%	3.1%	61.9%	32.8%	0.2%	0.0%	100.0%
9	A	29.9%	33.7%	33.3%	3.1%	0.0%	0.0%	100.0%
	B	5.4%	8.6%	53.5%	31.6%	0.9%	0.0%	100.0%
	C	0.8%	2.3%	40.2%	55.1%	1.5%	0.0%	100.0%
10	A	25.5%	37.6%	35.3%	1.6%	0.0%	0.0%	100.0%
	B	8.5%	12.9%	55.4%	23.0%	0.2%	0.0%	100.0%
	C	1.5%	2.8%	47.6%	47.3%	0.7%	0.0%	100.0%
11	A	30.5%	35.2%	33.7%	0.6%	0.0%	0.0%	100.0%
	B	9.2%	15.2%	53.2%	22.1%	0.2%	0.0%	100.0%
	C	2.2%	4.5%	48.2%	44.5%	0.6%	0.0%	100.0%
12	A	39.8%	19.3%	39.7%	1.2%	0.0%	0.0%	100.0%
	B	11.8%	13.4%	61.5%	13.1%	0.2%	0.0%	100.0%
	C	4.0%	4.7%	64.2%	26.8%	0.3%	0.0%	100.0%

4.3.3.3 By Grade

Table 4.3.3.3A

Proficiency Level by Grade (Count): Writing S401 Paper

	Writing Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	147,086	46,105	39,586	10,889	0	0	243,666
1	13,591	32,465	24,419	1,253	13	2	71,743
2	5,999	11,652	47,352	8,879	50	5	73,937
3	3,175	6,112	40,373	11,880	86	4	61,630
4	1,528	1,951	20,073	10,483	509	34	34,578
5	1,112	1,683	17,463	11,931	376	21	32,586
6	2,293	3,480	12,764	5,707	19	1	24,264
7	2,609	3,307	11,171	4,622	29	2	21,740
8	3,010	2,304	11,374	4,536	28	2	21,254
9	2,541	3,145	9,489	7,580	207	2	22,964
10	1,752	2,658	8,608	5,363	76	0	18,457
11	1,595	2,168	6,897	4,183	55	0	14,898
12	1,119	920	5,943	1,903	22	0	9,907

Table 4.3.3.3B

Proficiency Level by Grade (Percent): Writing S401 Paper

	Writing Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	60.4%	18.9%	16.2%	4.5%	0.0%	0.0%	100.0%
1	18.9%	45.3%	34.0%	1.7%	0.0%	0.0%	100.0%
2	8.1%	15.8%	64.0%	12.0%	0.1%	0.0%	100.0%
3	5.2%	9.9%	65.5%	19.3%	0.1%	0.0%	100.0%
4	4.4%	5.6%	58.1%	30.3%	1.5%	0.1%	100.0%
5	3.4%	5.2%	53.6%	36.6%	1.2%	0.1%	100.0%
6	9.5%	14.3%	52.6%	23.5%	0.1%	0.0%	100.0%
7	12.0%	15.2%	51.4%	21.3%	0.1%	0.0%	100.0%
8	14.2%	10.8%	53.5%	21.3%	0.1%	0.0%	100.0%
9	11.1%	13.7%	41.3%	33.0%	0.9%	0.0%	100.0%
10	9.5%	14.4%	46.6%	29.1%	0.4%	0.0%	100.0%
11	10.7%	14.6%	46.3%	28.1%	0.4%	0.0%	100.0%
12	11.3%	9.3%	60.0%	19.2%	0.2%	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 S401 Paper

Cluster	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	53,899	52,866	18,905	21,100	27,612	69,280	243,662
1	A	7,317	9,658	6,285	3,680	1,320	0	28,260
	B	744	6,234	9,168	7,430	1,794	460	25,830
	C	106	2,007	5,267	6,750	2,581	940	17,651
2	A	4,507	2,326	2,651	617	263	0	10,364
	B	2,233	6,928	10,805	4,797	1,297	603	26,663
	C	790	4,809	13,727	10,834	4,201	2,537	36,898
3	A	3,762	1,534	955	610	0	0	6,861
	B	1,373	4,024	6,384	2,849	487	557	15,674
	C	782	5,422	15,892	10,847	2,578	3,545	39,066
4-5	A	5,605	2,390	1,372	935	305	0	10,607
	B	1,009	2,819	4,945	4,913	1,591	660	15,937
	C	412	2,944	9,903	16,646	7,239	3,486	40,630
6-8	A	7,880	2,356	2,568	1,328	386	118	14,636
	B	1,616	3,797	4,825	5,120	1,199	595	17,152
	C	474	2,881	8,195	14,726	5,693	3,530	35,499
9-12	A	11,046	1,826	2,794	890	127	0	16,683
	B	4,525	3,601	5,297	2,678	626	1,102	17,829
	C	1,600	3,420	10,749	9,155	2,888	3,926	31,738

Table 4.3.4.1B

Proficiency Level by Cluster By Tier (Percent): Speaking S401 Paper

Cluster	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	22.1%	21.7%	7.8%	8.7%	11.3%	28.4%	100.0%
1	A	25.9%	34.2%	22.2%	13.0%	4.7%	0.0%	100.0%
	B	2.9%	24.1%	35.5%	28.8%	6.9%	1.8%	100.0%
	C	0.6%	11.4%	29.8%	38.2%	14.6%	5.3%	100.0%
2	A	43.5%	22.4%	25.6%	6.0%	2.5%	0.0%	100.0%
	B	8.4%	26.0%	40.5%	18.0%	4.9%	2.3%	100.0%
	C	2.1%	13.0%	37.2%	29.4%	11.4%	6.9%	100.0%
3	A	54.8%	22.4%	13.9%	8.9%	0.0%	0.0%	100.0%
	B	8.8%	25.7%	40.7%	18.2%	3.1%	3.6%	100.0%
	C	2.0%	13.9%	40.7%	27.8%	6.6%	9.1%	100.0%
4-5	A	52.8%	22.5%	12.9%	8.8%	2.9%	0.0%	100.0%
	B	6.3%	17.7%	31.0%	30.8%	10.0%	4.1%	100.0%
	C	1.0%	7.2%	24.4%	41.0%	17.8%	8.6%	100.0%
6-8	A	53.8%	16.1%	17.5%	9.1%	2.6%	0.8%	100.0%
	B	9.4%	22.1%	28.1%	29.9%	7.0%	3.5%	100.0%
	C	1.3%	8.1%	23.1%	41.5%	16.0%	9.9%	100.0%
9-12	A	66.2%	10.9%	16.7%	5.3%	0.8%	0.0%	100.0%
	B	25.4%	20.2%	29.7%	15.0%	3.5%	6.2%	100.0%
	C	5.0%	10.8%	33.9%	28.8%	9.1%	12.4%	100.0%

4.3.4.2 By Grade by Tier

Table 4.3.4.2A

Proficiency Level by Grade By Tier (Count): Speaking S401 Paper

Grade	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	53,899	52,866	18,905	21,100	27,612	69,280	243,662
1	A	7,317	9,658	6,285	3,680	1,320	0	28,260
	B	744	6,234	9,168	7,430	1,794	460	25,830
	C	106	2,007	5,267	6,750	2,581	940	17,651
2	A	4,507	2,326	2,651	617	263	0	10,364
	B	2,233	6,928	10,805	4,797	1,297	603	26,663
	C	790	4,809	13,727	10,834	4,201	2,537	36,898
3	A	3,762	1,534	955	610	0	0	6,861
	B	1,373	4,024	6,384	2,849	487	557	15,674
	C	782	5,422	15,892	10,847	2,578	3,545	39,066
4	A	2,782	1,541	753	429	196	0	5,701
	B	464	1,474	2,798	2,786	874	412	8,808
	C	174	1,429	4,766	8,331	3,338	2,037	20,075
5	A	2,823	849	619	506	109	0	4,906
	B	545	1,345	2,147	2,127	717	248	7,129
	C	238	1,515	5,137	8,315	3,901	1,449	20,555
6	A	2,581	948	849	439	180	52	5,049
	B	297	1,422	1,661	1,674	431	174	5,659
	C	116	1,309	3,386	5,327	2,207	1,217	13,562
7	A	2,474	951	703	579	67	66	4,840
	B	550	1,145	1,661	1,981	301	256	5,894
	C	175	730	2,441	5,123	1,278	1,278	11,025
8	A	2,825	457	1,016	310	139	0	4,747
	B	769	1,230	1,503	1,465	467	165	5,599
	C	183	842	2,368	4,276	2,208	1,035	10,912
9	A	5,123	672	1,045	228	127	0	7,195
	B	1,013	1,254	1,683	1,028	351	346	5,675
	C	254	1,049	3,018	3,128	1,419	1,231	10,099
10	A	3,110	465	790	305	0	0	4,670
	B	1,531	938	1,308	895	104	295	5,071
	C	488	875	2,605	3,097	615	1,047	8,727
11	A	2,070	395	692	263	0	0	3,420
	B	1,144	835	1,391	495	99	259	4,223
	C	406	729	2,831	1,834	499	961	7,260
12	A	743	294	267	94	0	0	1,398
	B	837	574	915	260	72	202	2,860
	C	452	767	2,295	1,096	355	687	5,652

Table 4.3.4.2B

Proficiency Level by Grade By Tier (Percent): Speaking S401 Paper

Grade	Tier	Speaking Proficiency Range						Total
		1	2	3	4	5	6	
K	-	22.1%	21.7%	7.8%	8.7%	11.3%	28.4%	100.0%
1	A	25.9%	34.2%	22.2%	13.0%	4.7%	0.0%	100.0%
	B	2.9%	24.1%	35.5%	28.8%	6.9%	1.8%	100.0%
	C	0.6%	11.4%	29.8%	38.2%	14.6%	5.3%	100.0%
2	A	43.5%	22.4%	25.6%	6.0%	2.5%	0.0%	100.0%
	B	8.4%	26.0%	40.5%	18.0%	4.9%	2.3%	100.0%
	C	2.1%	13.0%	37.2%	29.4%	11.4%	6.9%	100.0%
3	A	54.8%	22.4%	13.9%	8.9%	0.0%	0.0%	100.0%
	B	8.8%	25.7%	40.7%	18.2%	3.1%	3.6%	100.0%
	C	2.0%	13.9%	40.7%	27.8%	6.6%	9.1%	100.0%
4	A	48.8%	27.0%	13.2%	7.5%	3.4%	0.0%	100.0%
	B	5.3%	16.7%	31.8%	31.6%	9.9%	4.7%	100.0%
	C	0.9%	7.1%	23.7%	41.5%	16.6%	10.1%	100.0%
5	A	57.5%	17.3%	12.6%	10.3%	2.2%	0.0%	100.0%
	B	7.6%	18.9%	30.1%	29.8%	10.1%	3.5%	100.0%
	C	1.2%	7.4%	25.0%	40.5%	19.0%	7.0%	100.0%
6	A	51.1%	18.8%	16.8%	8.7%	3.6%	1.0%	100.0%
	B	5.2%	25.1%	29.4%	29.6%	7.6%	3.1%	100.0%
	C	0.9%	9.7%	25.0%	39.3%	16.3%	9.0%	100.0%
7	A	51.1%	19.6%	14.5%	12.0%	1.4%	1.4%	100.0%
	B	9.3%	19.4%	28.2%	33.6%	5.1%	4.3%	100.0%
	C	1.6%	6.6%	22.1%	46.5%	11.6%	11.6%	100.0%
8	A	59.5%	9.6%	21.4%	6.5%	2.9%	0.0%	100.0%
	B	13.7%	22.0%	26.8%	26.2%	8.3%	2.9%	100.0%
	C	1.7%	7.7%	21.7%	39.2%	20.2%	9.5%	100.0%
9	A	71.2%	9.3%	14.5%	3.2%	1.8%	0.0%	100.0%
	B	17.9%	22.1%	29.7%	18.1%	6.2%	6.1%	100.0%
	C	2.5%	10.4%	29.9%	31.0%	14.1%	12.2%	100.0%
10	A	66.6%	10.0%	16.9%	6.5%	0.0%	0.0%	100.0%
	B	30.2%	18.5%	25.8%	17.6%	2.1%	5.8%	100.0%
	C	5.6%	10.0%	29.8%	35.5%	7.0%	12.0%	100.0%
11	A	60.5%	11.5%	20.2%	7.7%	0.0%	0.0%	100.0%
	B	27.1%	19.8%	32.9%	11.7%	2.3%	6.1%	100.0%
	C	5.6%	10.0%	39.0%	25.3%	6.9%	13.2%	100.0%
12	A	53.1%	21.0%	19.1%	6.7%	0.0%	0.0%	100.0%
	B	29.3%	20.1%	32.0%	9.1%	2.5%	7.1%	100.0%
	C	8.0%	13.6%	40.6%	19.4%	6.3%	12.2%	100.0%

4.3.4.3 By Grade

Table 4.3.4.3A

Proficiency Level by Grade (Count): Speaking S401 Paper

	Speaking Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	53,899	52,866	18,905	21,100	27,612	69,280	243,662
1	8,167	17,899	20,720	17,860	5,695	1,400	71,741
2	7,530	14,063	27,183	16,248	5,761	3,140	73,925
3	5,917	10,980	23,231	14,306	3,065	4,102	61,601
4	3,420	4,444	8,317	11,546	4,408	2,449	34,584
5	3,606	3,709	7,903	10,948	4,727	1,697	32,590
6	2,994	3,679	5,896	7,440	2,818	1,443	24,270
7	3,199	2,826	4,805	7,683	1,646	1,600	21,759
8	3,777	2,529	4,887	6,051	2,814	1,200	21,258
9	6,390	2,975	5,746	4,384	1,897	1,577	22,969
10	5,129	2,278	4,703	4,297	719	1,342	18,468
11	3,620	1,959	4,914	2,592	598	1,220	14,903
12	2,032	1,635	3,477	1,450	427	889	9,910

Table 4.3.4.3B

Proficiency Level by Grade (Percent): Speaking S401 Paper

	Speaking Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	22.1%	21.7%	7.8%	8.7%	11.3%	28.4%	100.0%
1	11.4%	24.9%	28.9%	24.9%	7.9%	2.0%	100.0%
2	10.2%	19.0%	36.8%	22.0%	7.8%	4.2%	100.0%
3	9.6%	17.8%	37.7%	23.2%	5.0%	6.7%	100.0%
4	9.9%	12.8%	24.0%	33.4%	12.7%	7.1%	100.0%
5	11.1%	11.4%	24.2%	33.6%	14.5%	5.2%	100.0%
6	12.3%	15.2%	24.3%	30.7%	11.6%	5.9%	100.0%
7	14.7%	13.0%	22.1%	35.3%	7.6%	7.4%	100.0%
8	17.8%	11.9%	23.0%	28.5%	13.2%	5.6%	100.0%
9	27.8%	13.0%	25.0%	19.1%	8.3%	6.9%	100.0%
10	27.8%	12.3%	25.5%	23.3%	3.9%	7.3%	100.0%
11	24.3%	13.1%	33.0%	17.4%	4.0%	8.2%	100.0%
12	20.5%	16.5%	35.1%	14.6%	4.3%	9.0%	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 S401 Paper

Cluster	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	60,062	36,539	30,781	23,296	41,505	51,472	243,655
1	A	4,015	6,990	12,442	4,228	561	0	28,236
	B	296	2,311	11,677	9,289	2,236	0	25,809
	C	45	743	4,297	6,237	4,671	1,646	17,639
2	A	3,413	2,541	3,572	835	0	0	10,361
	B	260	2,744	13,080	8,693	1,873	0	26,650
	C	64	1,781	9,800	13,823	9,320	2,099	36,887
3	A	2,624	2,210	1,480	544	0	0	6,858
	B	184	2,272	7,377	4,870	969	0	15,672
	C	26	1,151	9,120	15,596	9,278	3,894	39,065
4-5	A	4,093	2,964	2,396	1,075	78	0	10,606
	B	274	1,605	5,849	6,173	2,030	0	15,931
	C	40	817	6,420	15,848	12,304	5,198	40,627
6-8	A	6,319	4,005	2,775	1,323	210	0	14,632
	B	671	3,338	6,024	5,750	1,363	0	17,146
	C	52	667	5,701	13,796	10,250	5,028	35,494
9-12	A	10,237	3,422	2,388	632	0	0	16,679
	B	1,868	4,312	6,459	4,002	1,183	0	17,824
	C	419	2,207	9,512	12,336	5,544	1,705	31,723

Table 4.3.5.1B

Proficiency Level by Cluster By Tier (Percent): Oral S401 Paper

Cluster	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	24.7%	15.0%	12.6%	9.6%	17.0%	21.1%	100.0%
1	A	14.2%	24.8%	44.1%	15.0%	2.0%	0.0%	100.0%
	B	1.1%	9.0%	45.2%	36.0%	8.7%	0.0%	100.0%
	C	0.3%	4.2%	24.4%	35.4%	26.5%	9.3%	100.0%
2	A	32.9%	24.5%	34.5%	8.1%	0.0%	0.0%	100.0%
	B	1.0%	10.3%	49.1%	32.6%	7.0%	0.0%	100.0%
	C	0.2%	4.8%	26.6%	37.5%	25.3%	5.7%	100.0%
3	A	38.3%	32.2%	21.6%	7.9%	0.0%	0.0%	100.0%
	B	1.2%	14.5%	47.1%	31.1%	6.2%	0.0%	100.0%
	C	0.1%	2.9%	23.3%	39.9%	23.8%	10.0%	100.0%
4-5	A	38.6%	27.9%	22.6%	10.1%	0.7%	0.0%	100.0%
	B	1.7%	10.1%	36.7%	38.7%	12.7%	0.0%	100.0%
	C	0.1%	2.0%	15.8%	39.0%	30.3%	12.8%	100.0%
6-8	A	43.2%	27.4%	19.0%	9.0%	1.4%	0.0%	100.0%
	B	3.9%	19.5%	35.1%	33.5%	7.9%	0.0%	100.0%
	C	0.1%	1.9%	16.1%	38.9%	28.9%	14.2%	100.0%
9-12	A	61.4%	20.5%	14.3%	3.8%	0.0%	0.0%	100.0%
	B	10.5%	24.2%	36.2%	22.5%	6.6%	0.0%	100.0%
	C	1.3%	7.0%	30.0%	38.9%	17.5%	5.4%	100.0%

4.3.5.2 By Grade by Tier

Table 4.3.5.2A

Proficiency Level by Grade By Tier (Count): Oral S401 Paper

Grade	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	60,062	36,539	30,781	23,296	41,505	51,472	243,655
1	A	4,015	6,990	12,442	4,228	561	0	28,236
	B	296	2,311	11,677	9,289	2,236	0	25,809
	C	45	743	4,297	6,237	4,671	1,646	17,639
2	A	3,413	2,541	3,572	835	0	0	10,361
	B	260	2,744	13,080	8,693	1,873	0	26,650
	C	64	1,781	9,800	13,823	9,320	2,099	36,887
3	A	2,624	2,210	1,480	544	0	0	6,858
	B	184	2,272	7,377	4,870	969	0	15,672
	C	26	1,151	9,120	15,596	9,278	3,894	39,065
4	A	2,042	1,628	1,437	515	78	0	5,700
	B	116	794	3,424	3,312	1,157	0	8,803
	C	19	363	3,244	7,605	6,241	2,600	20,072
5	A	2,051	1,336	959	560	0	0	4,906
	B	158	811	2,425	2,861	873	0	7,128
	C	21	454	3,176	8,243	6,063	2,598	20,555
6	A	1,904	1,462	1,126	456	100	0	5,048
	B	110	933	2,130	2,017	468	0	5,658
	C	14	213	2,247	5,364	3,745	1,978	13,561
7	A	2,137	1,361	821	466	55	0	4,840
	B	206	1,165	2,016	2,044	459	0	5,890
	C	20	244	1,784	4,354	3,219	1,401	11,022
8	A	2,278	1,182	828	401	55	0	4,744
	B	355	1,240	1,878	1,689	436	0	5,598
	C	18	210	1,670	4,078	3,286	1,649	10,911
9	A	4,491	1,559	906	238	0	0	7,194
	B	317	1,163	2,123	1,579	491	0	5,673
	C	46	415	2,314	4,426	2,236	657	10,094
10	A	2,912	891	668	198	0	0	4,669
	B	540	1,336	1,814	1,080	300	0	5,070
	C	121	602	2,507	3,457	1,524	513	8,724
11	A	2,034	708	533	144	0	0	3,419
	B	573	1,113	1,469	836	231	0	4,222
	C	112	630	2,480	2,604	1,087	343	7,256
12	A	800	264	281	52	0	0	1,397
	B	438	700	1,053	507	161	0	2,859
	C	140	560	2,211	1,849	697	192	5,649

Table 4.3.5.2B

Proficiency Level by Grade By Tier (Percent): Oral S401 Paper

Grade	Tier	Oral Language Proficiency Range						Total
		1	2	3	4	5	6	
K	-	24.7%	15.0%	12.6%	9.6%	17.0%	21.1%	100.0%
1	A	14.2%	24.8%	44.1%	15.0%	2.0%	0.0%	100.0%
	B	1.1%	9.0%	45.2%	36.0%	8.7%	0.0%	100.0%
	C	0.3%	4.2%	24.4%	35.4%	26.5%	9.3%	100.0%
2	A	32.9%	24.5%	34.5%	8.1%	0.0%	0.0%	100.0%
	B	1.0%	10.3%	49.1%	32.6%	7.0%	0.0%	100.0%
	C	0.2%	4.8%	26.6%	37.5%	25.3%	5.7%	100.0%
3	A	38.3%	32.2%	21.6%	7.9%	0.0%	0.0%	100.0%
	B	1.2%	14.5%	47.1%	31.1%	6.2%	0.0%	100.0%
	C	0.1%	2.9%	23.3%	39.9%	23.8%	10.0%	100.0%
4	A	35.8%	28.6%	25.2%	9.0%	1.4%	0.0%	100.0%
	B	1.3%	9.0%	38.9%	37.6%	13.1%	0.0%	100.0%
	C	0.1%	1.8%	16.2%	37.9%	31.1%	13.0%	100.0%
5	A	41.8%	27.2%	19.5%	11.4%	0.0%	0.0%	100.0%
	B	2.2%	11.4%	34.0%	40.1%	12.2%	0.0%	100.0%
	C	0.1%	2.2%	15.5%	40.1%	29.5%	12.6%	100.0%
6	A	37.7%	29.0%	22.3%	9.0%	2.0%	0.0%	100.0%
	B	1.9%	16.5%	37.6%	35.6%	8.3%	0.0%	100.0%
	C	0.1%	1.6%	16.6%	39.6%	27.6%	14.6%	100.0%
7	A	44.2%	28.1%	17.0%	9.6%	1.1%	0.0%	100.0%
	B	3.5%	19.8%	34.2%	34.7%	7.8%	0.0%	100.0%
	C	0.2%	2.2%	16.2%	39.5%	29.2%	12.7%	100.0%
8	A	48.0%	24.9%	17.5%	8.5%	1.2%	0.0%	100.0%
	B	6.3%	22.2%	33.5%	30.2%	7.8%	0.0%	100.0%
	C	0.2%	1.9%	15.3%	37.4%	30.1%	15.1%	100.0%
9	A	62.4%	21.7%	12.6%	3.3%	0.0%	0.0%	100.0%
	B	5.6%	20.5%	37.4%	27.8%	8.7%	0.0%	100.0%
	C	0.5%	4.1%	22.9%	43.8%	22.2%	6.5%	100.0%
10	A	62.4%	19.1%	14.3%	4.2%	0.0%	0.0%	100.0%
	B	10.7%	26.4%	35.8%	21.3%	5.9%	0.0%	100.0%
	C	1.4%	6.9%	28.7%	39.6%	17.5%	5.9%	100.0%
11	A	59.5%	20.7%	15.6%	4.2%	0.0%	0.0%	100.0%
	B	13.6%	26.4%	34.8%	19.8%	5.5%	0.0%	100.0%
	C	1.5%	8.7%	34.2%	35.9%	15.0%	4.7%	100.0%
12	A	57.3%	18.9%	20.1%	3.7%	0.0%	0.0%	100.0%
	B	15.3%	24.5%	36.8%	17.7%	5.6%	0.0%	100.0%
	C	2.5%	9.9%	39.1%	32.7%	12.3%	3.4%	100.0%

4.3.5.3 By Grade

Table 4.3.5.3A

Proficiency Level by Grade (Count): Oral S401 Paper

	Oral Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	60,062	36,539	30,781	23,296	41,505	51,472	243,655
1	4,356	10,044	28,416	19,754	7,468	1,646	71,684
2	3,737	7,066	26,452	23,351	11,193	2,099	73,898
3	2,834	5,633	17,977	21,010	10,247	3,894	61,595
4	2,177	2,785	8,105	11,432	7,476	2,600	34,575
5	2,230	2,601	6,560	11,664	6,936	2,598	32,589
6	2,028	2,608	5,503	7,837	4,313	1,978	24,267
7	2,363	2,770	4,621	6,864	3,733	1,401	21,752
8	2,651	2,632	4,376	6,168	3,777	1,649	21,253
9	4,854	3,137	5,343	6,243	2,727	657	22,961
10	3,573	2,829	4,989	4,735	1,824	513	18,463
11	2,719	2,451	4,482	3,584	1,318	343	14,897
12	1,378	1,524	3,545	2,408	858	192	9,905

Table 4.3.5.3B

Proficiency Level by Grade (Percent): Oral S401 Paper

	Oral Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	24.7%	15.0%	12.6%	9.6%	17.0%	21.1%	100.0%
1	6.1%	14.0%	39.6%	27.6%	10.4%	2.3%	100.0%
2	5.1%	9.6%	35.8%	31.6%	15.1%	2.8%	100.0%
3	4.6%	9.1%	29.2%	34.1%	16.6%	6.3%	100.0%
4	6.3%	8.1%	23.4%	33.1%	21.6%	7.5%	100.0%
5	6.8%	8.0%	20.1%	35.8%	21.3%	8.0%	100.0%
6	8.4%	10.7%	22.7%	32.3%	17.8%	8.2%	100.0%
7	10.9%	12.7%	21.2%	31.6%	17.2%	6.4%	100.0%
8	12.5%	12.4%	20.6%	29.0%	17.8%	7.8%	100.0%
9	21.1%	13.7%	23.3%	27.2%	11.9%	2.9%	100.0%
10	19.4%	15.3%	27.0%	25.6%	9.9%	2.8%	100.0%
11	18.3%	16.5%	30.1%	24.1%	8.8%	2.3%	100.0%
12	13.9%	15.4%	35.8%	24.3%	8.7%	1.9%	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 S401 Paper

Cluster	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	166,988	32,880	30,537	13,254	0	0	243,659
1	A	9,837	14,851	3,565	0	0	0	28,253
	B	1,704	8,225	15,129	765	2	0	25,825
	C	307	3,120	11,490	2,420	289	22	17,648
2	A	4,462	3,371	2,528	1	0	0	10,362
	B	1,473	6,147	16,152	2,874	10	0	26,656
	C	334	3,629	21,594	9,908	1,356	73	36,894
3	A	2,692	2,558	1,609	2	0	0	6,861
	B	533	2,705	9,901	2,523	4	0	15,666
	C	75	1,007	22,539	13,979	1,319	154	39,073
4-5	A	3,610	3,526	3,367	103	0	0	10,606
	B	444	1,800	9,557	3,987	124	1	15,913
	C	52	684	18,721	18,106	2,668	362	40,593
6-8	A	6,733	5,551	2,309	23	0	0	14,616
	B	1,024	3,992	9,315	2,789	13	0	17,133
	C	263	3,238	21,550	9,648	762	32	35,493
9-12	A	5,348	6,690	4,416	217	0	0	16,671
	B	1,415	4,718	8,846	2,776	62	0	17,817
	C	284	2,056	14,446	12,415	2,448	71	31,720

Table 4.3.6.1B

Proficiency Level by Cluster By Tier (Percent): Literacy S401 Paper

Cluster	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	68.5%	13.5%	12.5%	5.4%	0.0%	0.0%	100.0%
1	A	34.8%	52.6%	12.6%	0.0%	0.0%	0.0%	100.0%
	B	6.6%	31.8%	58.6%	3.0%	0.0%	0.0%	100.0%
	C	1.7%	17.7%	65.1%	13.7%	1.6%	0.1%	100.0%
2	A	43.1%	32.5%	24.4%	0.0%	0.0%	0.0%	100.0%
	B	5.5%	23.1%	60.6%	10.8%	0.0%	0.0%	100.0%
	C	0.9%	9.8%	58.5%	26.9%	3.7%	0.2%	100.0%
3	A	39.2%	37.3%	23.5%	0.0%	0.0%	0.0%	100.0%
	B	3.4%	17.3%	63.2%	16.1%	0.0%	0.0%	100.0%
	C	0.2%	2.6%	57.7%	35.8%	3.4%	0.4%	100.0%
4-5	A	34.0%	33.2%	31.7%	1.0%	0.0%	0.0%	100.0%
	B	2.8%	11.3%	60.1%	25.1%	0.8%	0.0%	100.0%
	C	0.1%	1.7%	46.1%	44.6%	6.6%	0.9%	100.0%
6-8	A	46.1%	38.0%	15.8%	0.2%	0.0%	0.0%	100.0%
	B	6.0%	23.3%	54.4%	16.3%	0.1%	0.0%	100.0%
	C	0.7%	9.1%	60.7%	27.2%	2.1%	0.1%	100.0%
9-12	A	32.1%	40.1%	26.5%	1.3%	0.0%	0.0%	100.0%
	B	7.9%	26.5%	49.6%	15.6%	0.3%	0.0%	100.0%
	C	0.9%	6.5%	45.5%	39.1%	7.7%	0.2%	100.0%

4.3.6.2 By Grade by Tier

Table 4.3.6.2A

Proficiency Level by Grade By Tier (Count): Literacy S401 Paper

Grade	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	166,988	32,880	30,537	13,254	0	0	243,659
1	A	9,837	14,851	3,565	0	0	0	28,253
	B	1,704	8,225	15,129	765	2	0	25,825
	C	307	3,120	11,490	2,420	289	22	17,648
2	A	4,462	3,371	2,528	1	0	0	10,362
	B	1,473	6,147	16,152	2,874	10	0	26,656
	C	334	3,629	21,594	9,908	1,356	73	36,894
3	A	2,692	2,558	1,609	2	0	0	6,861
	B	533	2,705	9,901	2,523	4	0	15,666
	C	75	1,007	22,539	13,979	1,319	154	39,073
4	A	1,924	1,826	1,906	45	0	0	5,701
	B	244	982	5,453	2,050	60	1	8,790
	C	27	320	9,928	8,459	1,149	178	20,061
5	A	1,686	1,700	1,461	58	0	0	4,905
	B	200	818	4,104	1,937	64	0	7,123
	C	25	364	8,793	9,647	1,519	184	20,532
6	A	2,099	1,976	962	6	0	0	5,043
	B	282	1,337	3,076	958	2	0	5,655
	C	93	1,257	8,656	3,363	183	9	13,561
7	A	2,162	1,976	686	9	0	0	4,833
	B	333	1,375	3,262	898	10	0	5,878
	C	79	1,063	6,709	2,939	217	15	11,022
8	A	2,472	1,599	661	8	0	0	4,740
	B	409	1,280	2,977	933	1	0	5,600
	C	91	918	6,185	3,346	362	8	10,910
9	A	2,399	2,914	1,751	125	0	0	7,189
	B	295	1,175	2,974	1,192	38	0	5,674
	C	37	365	3,939	4,739	963	49	10,092
10	A	1,418	1,887	1,301	60	0	0	4,666
	B	413	1,365	2,515	761	11	0	5,065
	C	62	474	3,974	3,501	697	15	8,723
11	A	1,048	1,357	995	18	0	0	3,418
	B	385	1,259	2,001	568	8	0	4,221
	C	67	578	3,372	2,679	551	7	7,254
12	A	483	532	369	14	0	0	1,398
	B	322	919	1,356	255	5	0	2,857
	C	118	639	3,161	1,496	237	0	5,651

Table 4.3.6.2B

Proficiency Level by Grade By Tier (Percent): Literacy S401 Paper

Grade	Tier	Literacy Proficiency Range						Total
		1	2	3	4	5	6	
K	-	68.5%	13.5%	12.5%	5.4%	0.0%	0.0%	100.0%
1	A	34.8%	52.6%	12.6%	0.0%	0.0%	0.0%	100.0%
	B	6.6%	31.8%	58.6%	3.0%	0.0%	0.0%	100.0%
	C	1.7%	17.7%	65.1%	13.7%	1.6%	0.1%	100.0%
2	A	43.1%	32.5%	24.4%	0.0%	0.0%	0.0%	100.0%
	B	5.5%	23.1%	60.6%	10.8%	0.0%	0.0%	100.0%
	C	0.9%	9.8%	58.5%	26.9%	3.7%	0.2%	100.0%
3	A	39.2%	37.3%	23.5%	0.0%	0.0%	0.0%	100.0%
	B	3.4%	17.3%	63.2%	16.1%	0.0%	0.0%	100.0%
	C	0.2%	2.6%	57.7%	35.8%	3.4%	0.4%	100.0%
4	A	33.7%	32.0%	33.4%	0.8%	0.0%	0.0%	100.0%
	B	2.8%	11.2%	62.0%	23.3%	0.7%	0.0%	100.0%
	C	0.1%	1.6%	49.5%	42.2%	5.7%	0.9%	100.0%
5	A	34.4%	34.7%	29.8%	1.2%	0.0%	0.0%	100.0%
	B	2.8%	11.5%	57.6%	27.2%	0.9%	0.0%	100.0%
	C	0.1%	1.8%	42.8%	47.0%	7.4%	0.9%	100.0%
6	A	41.6%	39.2%	19.1%	0.1%	0.0%	0.0%	100.0%
	B	5.0%	23.6%	54.4%	16.9%	0.0%	0.0%	100.0%
	C	0.7%	9.3%	63.8%	24.8%	1.3%	0.1%	100.0%
7	A	44.7%	40.9%	14.2%	0.2%	0.0%	0.0%	100.0%
	B	5.7%	23.4%	55.5%	15.3%	0.2%	0.0%	100.0%
	C	0.7%	9.6%	60.9%	26.7%	2.0%	0.1%	100.0%
8	A	52.2%	33.7%	13.9%	0.2%	0.0%	0.0%	100.0%
	B	7.3%	22.9%	53.2%	16.7%	0.0%	0.0%	100.0%
	C	0.8%	8.4%	56.7%	30.7%	3.3%	0.1%	100.0%
9	A	33.4%	40.5%	24.4%	1.7%	0.0%	0.0%	100.0%
	B	5.2%	20.7%	52.4%	21.0%	0.7%	0.0%	100.0%
	C	0.4%	3.6%	39.0%	47.0%	9.5%	0.5%	100.0%
10	A	30.4%	40.4%	27.9%	1.3%	0.0%	0.0%	100.0%
	B	8.2%	26.9%	49.7%	15.0%	0.2%	0.0%	100.0%
	C	0.7%	5.4%	45.6%	40.1%	8.0%	0.2%	100.0%
11	A	30.7%	39.7%	29.1%	0.5%	0.0%	0.0%	100.0%
	B	9.1%	29.8%	47.4%	13.5%	0.2%	0.0%	100.0%
	C	0.9%	8.0%	46.5%	36.9%	7.6%	0.1%	100.0%
12	A	34.5%	38.1%	26.4%	1.0%	0.0%	0.0%	100.0%
	B	11.3%	32.2%	47.5%	8.9%	0.2%	0.0%	100.0%
	C	2.1%	11.3%	55.9%	26.5%	4.2%	0.0%	100.0%

4.3.6.3 By Grade

Table 4.3.6.3A

Proficiency Level by Grade (Count): Literacy S401 Paper

	Literacy Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	166,988	32,880	30,537	13,254	0	0	243,659
1	11,848	26,196	30,184	3,185	291	22	71,726
2	6,269	13,147	40,274	12,783	1,366	73	73,912
3	3,300	6,270	34,049	16,504	1,323	154	61,600
4	2,195	3,128	17,287	10,554	1,209	179	34,552
5	1,911	2,882	14,358	11,642	1,583	184	32,560
6	2,474	4,570	12,694	4,327	185	9	24,259
7	2,574	4,414	10,657	3,846	227	15	21,733
8	2,972	3,797	9,823	4,287	363	8	21,250
9	2,731	4,454	8,664	6,056	1,001	49	22,955
10	1,893	3,726	7,790	4,322	708	15	18,454
11	1,500	3,194	6,368	3,265	559	7	14,893
12	923	2,090	4,886	1,765	242	0	9,906

Table 4.3.6.3B

Proficiency Level by Grade (Percent): Literacy S401 Paper

	Literacy Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	68.5%	13.5%	12.5%	5.4%	0.0%	0.0%	100.0%
1	16.5%	36.5%	42.1%	4.4%	0.4%	0.0%	100.0%
2	8.5%	17.8%	54.5%	17.3%	1.8%	0.1%	100.0%
3	5.4%	10.2%	55.3%	26.8%	2.1%	0.3%	100.0%
4	6.4%	9.1%	50.0%	30.5%	3.5%	0.5%	100.0%
5	5.9%	8.9%	44.1%	35.8%	4.9%	0.6%	100.0%
6	10.2%	18.8%	52.3%	17.8%	0.8%	0.0%	100.0%
7	11.8%	20.3%	49.0%	17.7%	1.0%	0.1%	100.0%
8	14.0%	17.9%	46.2%	20.2%	1.7%	0.0%	100.0%
9	11.9%	19.4%	37.7%	26.4%	4.4%	0.2%	100.0%
10	10.3%	20.2%	42.2%	23.4%	3.8%	0.1%	100.0%
11	10.1%	21.4%	42.8%	21.9%	3.8%	0.0%	100.0%
12	9.3%	21.1%	49.3%	17.8%	2.4%	0.0%	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 S401 Paper

Cluster	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	147,775	21,020	25,198	12,975	29,817	6,872	243,657
1	A	4,766	10,870	9,734	2,871	n/a	n/a	28,241
	B	191	1,045	10,858	7,511	6,214	n/a	25,819
	C	64	683	4,814	3,965	5,044	3,068	17,638
2	A	3,325	3,890	2,163	984	n/a	n/a	10,362
	B	147	3,356	9,707	5,931	7,507	n/a	26,648
	C	120	3,032	9,562	7,077	9,813	7,285	36,889
3	A	1,966	2,998	1,385	515	n/a	n/a	6,864
	B	180	2,494	6,352	3,914	2,729	n/a	15,669
	C	17	394	6,062	8,152	14,799	9,652	39,076
4-5	A	3,794	3,844	1,986	981	n/a	n/a	10,605
	B	424	3,287	5,661	3,779	2,763	n/a	15,914
	C	27	1,347	8,324	9,542	13,063	8,296	40,599
6-8	A	6,317	5,591	2,010	711	n/a	n/a	14,629
	B	746	5,931	5,661	3,027	1,783	n/a	17,148
	C	94	3,656	10,929	8,166	8,140	4,513	35,498
9-12	A	7,164	6,600	2,335	579	n/a	n/a	16,678
	B	1,306	7,202	5,443	2,344	1,532	n/a	17,827
	C	150	3,917	8,981	6,524	7,156	4,996	31,724

Table 4.3.7.1B

Proficiency Level by Cluster By Tier (Percent): Comprehension S401 Paper

Cluster	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	60.6%	8.6%	10.3%	5.3%	12.2%	2.8%	100.0%
1	A	16.9%	38.5%	34.5%	10.2%	n/a	n/a	100.0%
	B	0.7%	4.0%	42.1%	29.1%	24.1%	n/a	100.0%
	C	0.4%	3.9%	27.3%	22.5%	28.6%	17.4%	100.0%
2	A	32.1%	37.5%	20.9%	9.5%	n/a	n/a	100.0%
	B	0.6%	12.6%	36.4%	22.3%	28.2%	n/a	100.0%
	C	0.3%	8.2%	25.9%	19.2%	26.6%	19.7%	100.0%
3	A	28.6%	43.7%	20.2%	7.5%	n/a	n/a	100.0%
	B	1.1%	15.9%	40.5%	25.0%	17.4%	n/a	100.0%
	C	0.0%	1.0%	15.5%	20.9%	37.9%	24.7%	100.0%
4-5	A	35.8%	36.2%	18.7%	9.3%	n/a	n/a	100.0%
	B	2.7%	20.7%	35.6%	23.7%	17.4%	n/a	100.0%
	C	0.1%	3.3%	20.5%	23.5%	32.2%	20.4%	100.0%
6-8	A	43.2%	38.2%	13.7%	4.9%	n/a	n/a	100.0%
	B	4.4%	34.6%	33.0%	17.7%	10.4%	n/a	100.0%
	C	0.3%	10.3%	30.8%	23.0%	22.9%	12.7%	100.0%
9-12	A	43.0%	39.6%	14.0%	3.5%	n/a	n/a	100.0%
	B	7.3%	40.4%	30.5%	13.1%	8.6%	n/a	100.0%
	C	0.5%	12.3%	28.3%	20.6%	22.6%	15.7%	100.0%

4.3.7.2 By Grade by Tier

Table 4.3.7.2A
Proficiency Level by Grade By Tier (Count): Comprehension S401 Paper

Grade	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	147,775	21,020	25,198	12,975	29,817	6,872	243,657
1	A	4,766	10,870	9,734	2,871	n/a	n/a	28,241
	B	191	1,045	10,858	7,511	6,214	n/a	25,819
	C	64	683	4,814	3,965	5,044	3,068	17,638
2	A	3,325	3,890	2,163	984	n/a	n/a	10,362
	B	147	3,356	9,707	5,931	7,507	n/a	26,648
	C	120	3,032	9,562	7,077	9,813	7,285	36,889
3	A	1,966	2,998	1,385	515	n/a	n/a	6,864
	B	180	2,494	6,352	3,914	2,729	n/a	15,669
	C	17	394	6,062	8,152	14,799	9,652	39,076
4	A	1,852	2,125	1,161	562	n/a	n/a	5,700
	B	151	1,743	3,142	2,080	1,674	n/a	8,790
	C	13	475	4,101	4,596	6,895	3,984	20,064
5	A	1,942	1,719	825	419	n/a	n/a	4,905
	B	273	1,544	2,519	1,699	1,089	n/a	7,124
	C	14	872	4,223	4,946	6,168	4,312	20,535
6	A	1,831	2,133	803	280	n/a	n/a	5,047
	B	156	1,922	2,026	989	564	n/a	5,657
	C	28	1,288	4,721	3,134	3,100	1,293	13,564
7	A	2,125	1,867	615	229	n/a	n/a	4,836
	B	228	2,021	1,960	1,085	596	n/a	5,890
	C	29	1,248	3,283	2,725	2,350	1,389	11,024
8	A	2,361	1,591	592	202	n/a	n/a	4,746
	B	362	1,988	1,675	953	623	n/a	5,601
	C	37	1,120	2,925	2,307	2,690	1,831	10,910
9	A	3,135	2,960	907	193	n/a	n/a	7,195
	B	216	1,930	2,009	875	644	n/a	5,674
	C	14	726	2,451	2,291	2,741	1,870	10,093
10	A	1,918	1,895	668	187	n/a	n/a	4,668
	B	353	2,081	1,594	649	393	n/a	5,070
	C	27	984	2,518	1,875	1,836	1,484	8,724
11	A	1,483	1,248	559	128	n/a	n/a	3,418
	B	387	1,889	1,140	524	282	n/a	4,222
	C	38	1,083	2,179	1,230	1,617	1,110	7,257
12	A	628	497	201	71	n/a	n/a	1,397
	B	350	1,302	700	296	213	n/a	2,861
	C	71	1,124	1,833	1,128	962	532	5,650

Table 4.3.7.2B

Proficiency Level by Grade By Tier (Percent): Comprehension S401 Paper

Grade	Tier	Comprehension Proficiency Range						Total
		1	2	3	4	5	6	
K	-	60.6%	8.6%	10.3%	5.3%	12.2%	2.8%	100.0%
1	A	16.9%	38.5%	34.5%	10.2%	n/a	n/a	100.0%
	B	0.7%	4.0%	42.1%	29.1%	24.1%	n/a	100.0%
	C	0.4%	3.9%	27.3%	22.5%	28.6%	17.4%	100.0%
2	A	32.1%	37.5%	20.9%	9.5%	n/a	n/a	100.0%
	B	0.6%	12.6%	36.4%	22.3%	28.2%	n/a	100.0%
	C	0.3%	8.2%	25.9%	19.2%	26.6%	19.7%	100.0%
3	A	28.6%	43.7%	20.2%	7.5%	n/a	n/a	100.0%
	B	1.1%	15.9%	40.5%	25.0%	17.4%	n/a	100.0%
	C	0.0%	1.0%	15.5%	20.9%	37.9%	24.7%	100.0%
4	A	32.5%	37.3%	20.4%	9.9%	n/a	n/a	100.0%
	B	1.7%	19.8%	35.7%	23.7%	19.0%	n/a	100.0%
	C	0.1%	2.4%	20.4%	22.9%	34.4%	19.9%	100.0%
5	A	39.6%	35.0%	16.8%	8.5%	n/a	n/a	100.0%
	B	3.8%	21.7%	35.4%	23.8%	15.3%	n/a	100.0%
	C	0.1%	4.2%	20.6%	24.1%	30.0%	21.0%	100.0%
6	A	36.3%	42.3%	15.9%	5.5%	n/a	n/a	100.0%
	B	2.8%	34.0%	35.8%	17.5%	10.0%	n/a	100.0%
	C	0.2%	9.5%	34.8%	23.1%	22.9%	9.5%	100.0%
7	A	43.9%	38.6%	12.7%	4.7%	n/a	n/a	100.0%
	B	3.9%	34.3%	33.3%	18.4%	10.1%	n/a	100.0%
	C	0.3%	11.3%	29.8%	24.7%	21.3%	12.6%	100.0%
8	A	49.7%	33.5%	12.5%	4.3%	n/a	n/a	100.0%
	B	6.5%	35.5%	29.9%	17.0%	11.1%	n/a	100.0%
	C	0.3%	10.3%	26.8%	21.1%	24.7%	16.8%	100.0%
9	A	43.6%	41.1%	12.6%	2.7%	n/a	n/a	100.0%
	B	3.8%	34.0%	35.4%	15.4%	11.4%	n/a	100.0%
	C	0.1%	7.2%	24.3%	22.7%	27.2%	18.5%	100.0%
10	A	41.1%	40.6%	14.3%	4.0%	n/a	n/a	100.0%
	B	7.0%	41.0%	31.4%	12.8%	7.8%	n/a	100.0%
	C	0.3%	11.3%	28.9%	21.5%	21.0%	17.0%	100.0%
11	A	43.4%	36.5%	16.4%	3.7%	n/a	n/a	100.0%
	B	9.2%	44.7%	27.0%	12.4%	6.7%	n/a	100.0%
	C	0.5%	14.9%	30.0%	16.9%	22.3%	15.3%	100.0%
12	A	45.0%	35.6%	14.4%	5.1%	n/a	n/a	100.0%
	B	12.2%	45.5%	24.5%	10.3%	7.4%	n/a	100.0%
	C	1.3%	19.9%	32.4%	20.0%	17.0%	9.4%	100.0%

4.3.7.3 By Grade

Table 4.3.7.3A

Proficiency Level by Grade (Count): Comprehension S401 Paper

	Comprehension Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	147,775	21,020	25,198	12,975	29,817	6,872	243,657
1	5,021	12,598	25,406	14,347	11,258	3,068	71,698
2	3,592	10,278	21,432	13,992	17,320	7,285	73,899
3	2,163	5,886	13,799	12,581	17,528	9,652	61,609
4	2,016	4,343	8,404	7,238	8,569	3,984	34,554
5	2,229	4,135	7,567	7,064	7,257	4,312	32,564
6	2,015	5,343	7,550	4,403	3,664	1,293	24,268
7	2,382	5,136	5,858	4,039	2,946	1,389	21,750
8	2,760	4,699	5,192	3,462	3,313	1,831	21,257
9	3,365	5,616	5,367	3,359	3,385	1,870	22,962
10	2,298	4,960	4,780	2,711	2,229	1,484	18,462
11	1,908	4,220	3,878	1,882	1,899	1,110	14,897
12	1,049	2,923	2,734	1,495	1,175	532	9,908

Table 4.3.7.3B

Proficiency Level by Grade (Percent): Comprehension S401 Paper

	Comprehension Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	60.6%	8.6%	10.3%	5.3%	12.2%	2.8%	100.0%
1	7.0%	17.6%	35.4%	20.0%	15.7%	4.3%	100.0%
2	4.9%	13.9%	29.0%	18.9%	23.4%	9.9%	100.0%
3	3.5%	9.6%	22.4%	20.4%	28.5%	15.7%	100.0%
4	5.8%	12.6%	24.3%	20.9%	24.8%	11.5%	100.0%
5	6.8%	12.7%	23.2%	21.7%	22.3%	13.2%	100.0%
6	8.3%	22.0%	31.1%	18.1%	15.1%	5.3%	100.0%
7	11.0%	23.6%	26.9%	18.6%	13.5%	6.4%	100.0%
8	13.0%	22.1%	24.4%	16.3%	15.6%	8.6%	100.0%
9	14.7%	24.5%	23.4%	14.6%	14.7%	8.1%	100.0%
10	12.4%	26.9%	25.9%	14.7%	12.1%	8.0%	100.0%
11	12.8%	28.3%	26.0%	12.6%	12.7%	7.5%	100.0%
12	10.6%	29.5%	27.6%	15.1%	11.9%	5.4%	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 S401 Paper

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	126,385	44,166	39,861	28,732	4,499	0	243,643
1	A	6,307	13,378	8,532	6	0	0	28,223
	B	1,011	4,364	18,723	1,688	5	0	25,791
	C	139	1,130	10,976	4,656	692	44	17,637
2	A	3,860	3,600	2,877	20	0	0	10,357
	B	577	4,728	17,154	4,166	9	0	26,634
	C	106	2,013	18,547	13,662	2,469	82	36,879
3	A	2,649	2,461	1,728	16	0	0	6,854
	B	275	2,310	10,032	3,021	18	0	15,656
	C	36	549	16,409	18,877	2,951	232	39,054
4-5	A	3,800	3,296	3,342	166	0	0	10,604
	B	301	1,554	8,671	5,238	141	0	15,905
	C	33	354	12,048	22,686	4,980	487	40,588
6-8	A	6,639	4,869	3,000	104	0	0	14,612
	B	666	3,525	8,916	3,991	27	0	17,125
	C	94	1,207	14,595	17,381	2,099	104	35,480
9-12	A	7,563	5,342	3,579	184	0	0	16,668
	B	1,256	4,658	8,608	3,230	56	0	17,808
	C	171	1,533	12,890	14,358	2,678	77	31,707

Table 4.3.8.1B

Proficiency Level by Cluster By Tier (Percent): Overall S401 Paper

Cluster	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	51.9%	18.1%	16.4%	11.8%	1.8%	0.0%	100.0%
1	A	22.3%	47.4%	30.2%	0.0%	0.0%	0.0%	100.0%
	B	3.9%	16.9%	72.6%	6.5%	0.0%	0.0%	100.0%
	C	0.8%	6.4%	62.2%	26.4%	3.9%	0.2%	100.0%
2	A	37.3%	34.8%	27.8%	0.2%	0.0%	0.0%	100.0%
	B	2.2%	17.8%	64.4%	15.6%	0.0%	0.0%	100.0%
	C	0.3%	5.5%	50.3%	37.0%	6.7%	0.2%	100.0%
3	A	38.6%	35.9%	25.2%	0.2%	0.0%	0.0%	100.0%
	B	1.8%	14.8%	64.1%	19.3%	0.1%	0.0%	100.0%
	C	0.1%	1.4%	42.0%	48.3%	7.6%	0.6%	100.0%
4-5	A	35.8%	31.1%	31.5%	1.6%	0.0%	0.0%	100.0%
	B	1.9%	9.8%	54.5%	32.9%	0.9%	0.0%	100.0%
	C	0.1%	0.9%	29.7%	55.9%	12.3%	1.2%	100.0%
6-8	A	45.4%	33.3%	20.5%	0.7%	0.0%	0.0%	100.0%
	B	3.9%	20.6%	52.1%	23.3%	0.2%	0.0%	100.0%
	C	0.3%	3.4%	41.1%	49.0%	5.9%	0.3%	100.0%
9-12	A	45.4%	32.0%	21.5%	1.1%	0.0%	0.0%	100.0%
	B	7.1%	26.2%	48.3%	18.1%	0.3%	0.0%	100.0%
	C	0.5%	4.8%	40.7%	45.3%	8.4%	0.2%	100.0%

4.3.8.2 By Grade by Tier

Table 4.3.8.2A

Proficiency Level by Grade By Tier (Count): Overall S401 Paper

Grade	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	126,385	44,166	39,861	28,732	4,499	0	243,643
1	A	6,307	13,378	8,532	6	0	0	28,223
	B	1,011	4,364	18,723	1,688	5	0	25,791
	C	139	1,130	10,976	4,656	692	44	17,637
2	A	3,860	3,600	2,877	20	0	0	10,357
	B	577	4,728	17,154	4,166	9	0	26,634
	C	106	2,013	18,547	13,662	2,469	82	36,879
3	A	2,649	2,461	1,728	16	0	0	6,854
	B	275	2,310	10,032	3,021	18	0	15,656
	C	36	549	16,409	18,877	2,951	232	39,054
4	A	1,958	1,759	1,893	90	0	0	5,700
	B	155	774	4,918	2,873	65	0	8,785
	C	16	158	6,209	11,212	2,195	267	20,057
5	A	1,842	1,537	1,449	76	0	0	4,904
	B	146	780	3,753	2,365	76	0	7,120
	C	17	196	5,839	11,474	2,785	220	20,531
6	A	1,993	1,808	1,198	44	0	0	5,043
	B	157	1,068	3,038	1,384	7	0	5,654
	C	26	439	5,923	6,504	629	36	13,557
7	A	2,256	1,597	949	31	0	0	4,833
	B	214	1,213	3,114	1,320	14	0	5,875
	C	32	416	4,566	5,345	623	34	11,016
8	A	2,390	1,464	853	29	0	0	4,736
	B	295	1,244	2,764	1,287	6	0	5,596
	C	36	352	4,106	5,532	847	34	10,907
9	A	3,343	2,327	1,417	100	0	0	7,187
	B	231	1,101	2,950	1,348	39	0	5,669
	C	22	227	3,264	5,357	1,170	48	10,088
10	A	2,109	1,467	1,042	48	0	0	4,666
	B	386	1,384	2,397	888	9	0	5,064
	C	42	377	3,482	4,046	756	19	8,722
11	A	1,496	1,095	804	23	0	0	3,418
	B	363	1,238	1,977	639	4	0	4,221
	C	42	404	3,244	3,009	542	9	7,250
12	A	615	453	316	13	0	0	1,397
	B	276	935	1,284	355	4	0	2,854
	C	65	525	2,900	1,946	210	1	5,647

Table 4.3.8.2B

Proficiency Level by Grade By Tier (Percent): Overall S401 Paper

Grade	Tier	Overall Proficiency Range						Total
		1	2	3	4	5	6	
K	-	51.9%	18.1%	16.4%	11.8%	1.8%	0.0%	100.0%
1	A	22.3%	47.4%	30.2%	0.0%	0.0%	0.0%	100.0%
	B	3.9%	16.9%	72.6%	6.5%	0.0%	0.0%	100.0%
	C	0.8%	6.4%	62.2%	26.4%	3.9%	0.2%	100.0%
2	A	37.3%	34.8%	27.8%	0.2%	0.0%	0.0%	100.0%
	B	2.2%	17.8%	64.4%	15.6%	0.0%	0.0%	100.0%
	C	0.3%	5.5%	50.3%	37.0%	6.7%	0.2%	100.0%
3	A	38.6%	35.9%	25.2%	0.2%	0.0%	0.0%	100.0%
	B	1.8%	14.8%	64.1%	19.3%	0.1%	0.0%	100.0%
	C	0.1%	1.4%	42.0%	48.3%	7.6%	0.6%	100.0%
4	A	34.4%	30.9%	33.2%	1.6%	0.0%	0.0%	100.0%
	B	1.8%	8.8%	56.0%	32.7%	0.7%	0.0%	100.0%
	C	0.1%	0.8%	31.0%	55.9%	10.9%	1.3%	100.0%
5	A	37.6%	31.3%	29.5%	1.5%	0.0%	0.0%	100.0%
	B	2.1%	11.0%	52.7%	33.2%	1.1%	0.0%	100.0%
	C	0.1%	1.0%	28.4%	55.9%	13.6%	1.1%	100.0%
6	A	39.5%	35.9%	23.8%	0.9%	0.0%	0.0%	100.0%
	B	2.8%	18.9%	53.7%	24.5%	0.1%	0.0%	100.0%
	C	0.2%	3.2%	43.7%	48.0%	4.6%	0.3%	100.0%
7	A	46.7%	33.0%	19.6%	0.6%	0.0%	0.0%	100.0%
	B	3.6%	20.6%	53.0%	22.5%	0.2%	0.0%	100.0%
	C	0.3%	3.8%	41.4%	48.5%	5.7%	0.3%	100.0%
8	A	50.5%	30.9%	18.0%	0.6%	0.0%	0.0%	100.0%
	B	5.3%	22.2%	49.4%	23.0%	0.1%	0.0%	100.0%
	C	0.3%	3.2%	37.6%	50.7%	7.8%	0.3%	100.0%
9	A	46.5%	32.4%	19.7%	1.4%	0.0%	0.0%	100.0%
	B	4.1%	19.4%	52.0%	23.8%	0.7%	0.0%	100.0%
	C	0.2%	2.3%	32.4%	53.1%	11.6%	0.5%	100.0%
10	A	45.2%	31.4%	22.3%	1.0%	0.0%	0.0%	100.0%
	B	7.6%	27.3%	47.3%	17.5%	0.2%	0.0%	100.0%
	C	0.5%	4.3%	39.9%	46.4%	8.7%	0.2%	100.0%
11	A	43.8%	32.0%	23.5%	0.7%	0.0%	0.0%	100.0%
	B	8.6%	29.3%	46.8%	15.1%	0.1%	0.0%	100.0%
	C	0.6%	5.6%	44.7%	41.5%	7.5%	0.1%	100.0%
12	A	44.0%	32.4%	22.6%	0.9%	0.0%	0.0%	100.0%
	B	9.7%	32.8%	45.0%	12.4%	0.1%	0.0%	100.0%
	C	1.2%	9.3%	51.4%	34.5%	3.7%	0.0%	100.0%

4.3.8.3 By Grade

Table 4.3.8.3A

Proficiency Level by Grade (Count): Overall S401 Paper

	Overall Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	126,385	44,166	39,861	28,732	4,499	0	243,643
1	7,457	18,872	38,231	6,350	697	44	71,651
2	4,543	10,341	38,578	17,848	2,478	82	73,870
3	2,960	5,320	28,169	21,914	2,969	232	61,564
4	2,129	2,691	13,020	14,175	2,260	267	34,542
5	2,005	2,513	11,041	13,915	2,861	220	32,555
6	2,176	3,315	10,159	7,932	636	36	24,254
7	2,502	3,226	8,629	6,696	637	34	21,724
8	2,721	3,060	7,723	6,848	853	34	21,239
9	3,596	3,655	7,631	6,805	1,209	48	22,944
10	2,537	3,228	6,921	4,982	765	19	18,452
11	1,901	2,737	6,025	3,671	546	9	14,889
12	956	1,913	4,500	2,314	214	1	9,898

Table 4.3.8.3B

Proficiency Level by Grade (Percent): Overall S401 Paper

	Overall Proficiency Range						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
K	51.9%	18.1%	16.4%	11.8%	1.8%	0.0%	100.0%
1	10.4%	26.3%	53.4%	8.9%	1.0%	0.1%	100.0%
2	6.1%	14.0%	52.2%	24.2%	3.4%	0.1%	100.0%
3	4.8%	8.6%	45.8%	35.6%	4.8%	0.4%	100.0%
4	6.2%	7.8%	37.7%	41.0%	6.5%	0.8%	100.0%
5	6.2%	7.7%	33.9%	42.7%	8.8%	0.7%	100.0%
6	9.0%	13.7%	41.9%	32.7%	2.6%	0.1%	100.0%
7	11.5%	14.8%	39.7%	30.8%	2.9%	0.2%	100.0%
8	12.8%	14.4%	36.4%	32.2%	4.0%	0.2%	100.0%
9	15.7%	15.9%	33.3%	29.7%	5.3%	0.2%	100.0%
10	13.7%	17.5%	37.5%	27.0%	4.1%	0.1%	100.0%
11	12.8%	18.4%	40.5%	24.7%	3.7%	0.1%	100.0%
12	9.7%	19.3%	45.5%	23.4%	2.2%	0.0%	100.0%

World-Class Instructional Design and Assessment

Annual Technical Report for ACCESS for ELLs 2.0 Paper English Language Proficiency Test, Series 401, 2016–2017 Administration

Annual Technical Report No. 13B
Volume 2 of 3: Analyses of Test Forms

Prepared by:

Center for Applied Linguistics

Language Assessment Division
Psychometrics and Quantitative Research Team

April 25, 2018

Volume 2

5. Analyses of Test Forms: Overview 1

5.1 Background 1

5.1.1 Measurement Models Used 1

5.1.2 Data 4

5.1.3 Equating and Scaling 4

5.1.4 DIF Analyses 5

5.1.5 Analysis of Writing and Speaking Forms 7

5.2 Descriptions of Tables and Figures 7

5.2.1 Raw Score Information (Figure A and Table A) 7

5.2.2 Scale Score Information (Figure B and Table B) 8

5.2.3 Proficiency Level Information (Figure C and Table C) 9

5.2.4 Scaling Equation Table (Table D) 9

5.2.5 Equating Summary (Table E) 10

5.2.6 Test Characteristic Curve (Figure D) 12

5.2.7 Test Information Function (Figure E) 12

5.2.8 Reliability (Table F) 12

5.2.9 Complete Item or Task Analysis and Summary (Table G) 14

5.2.10 DIF Analysis and Summary (Table H) 15

5.2.11 Raw Score to Scale Score Conversion Chart (Table I) 15

5.2.12 Raw Score to Proficiency Level Score Conversion Table (Table J) 16

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-grouped rating scale model is used. Mathematically, this can be represented as

$$\log\left(\frac{P_{ngik}}{1 - P_{ngi(k-1)}}\right) = \beta_n - D_{gi} - F_{gk}$$

where

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

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

β_n = ability of person "n"

D_{gi} = difficulty of task "i" specific to rating scale "g"

F_{gk} = calibration of step "k" on rating scale "g"

The subscript "g" is a group index specifying the group of tasks to which task "i" belongs. It also identifies the scoring scale that was used for the group of tasks.

As described in section 1.6.2, ratings on the ACCESS 2.0 Writing Scoring Scale range from 0, 1, 1+,…, 6 and the possible raw scores range from 0-9. All Writing tasks are scored using this scoring scale except for Grade 1 Tier A Task 1 and 2. The profiles of the responses to these two tasks do not fit the generic scoring scale well, so additional task-specific instructions are provided to raters. These instructions guide raters in applying a limited number of score points on the scoring scale to responses elicited by these two tasks. The possible ratings for Grade 1 Tier A Task 1 are 0 or 1 and the possible ratings for Grade 1 Tier A Task 2 are 0, 1,1+, or 2. To simplify the year-to-year linking process, the Grade 1 Writing Tier A Task 1 is treated as a dichotomously-scored task. The Grade 1 Writing Tier B Task 2 is modeled using a rating scale with possible raw score of 0 to 3. All other Writing tasks are modeled using a rating scale with possible raw scores of 0 to 9. Thus there are total of two rating scales being modeled for ACCESS Writing. One rating scale is associated with the Grade 1 Writing Tier A Task 2, and the other rating scale is associated with all Writing tasks that are scored using the rating scale with raw score values 0 to 9.

For Speaking, PL 1 tasks are modeled as a group on a 0–2 scale and PL 3 and PL 5 tasks are modeled as a group on a 0–4 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 for ELLs, the vast majority of dichotomous items on the test forms have mean square fit statistics in the range of 0.5 and 1.5; thus, they fit the range that is "productive for measurement" according to the guidelines above.

Since performance tasks are constructed and scored very differently from dichotomous items, it is not as straightforward to apply this same guidance to interpret these fit statistics to performance tasks that are scored polytomously. Some performance tasks that were designed to elicit a restricted range of performances (for example, very easy tasks, for which it is expected that most students will get the highest score) can cause the model to predict the data too well (overfitting). Conversely, when performance tasks are scored using a very wide rubric scale, such as the case with ACCESS for ELL Writing tasks, sometimes un-modeled noise or other sources of variance in the data will cause the model to under predict the data (under-fitting). Overall, for ACCESS for ELL performance tasks, overfitting is more common than under-fitting. Under-fitting indicates that the task is less productive for measurement, but it is not degrading to the measurement of student performance.

5.1.2 Data

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 401 Paper in the academic year 2015–2016, with the exception of some students as described in Section 3.1. Also excluded from the data used to produce the results in Chapter 6 are students from the state of Michigan.

5.1.3 Equating and Scaling

No equating summaries are provided for ACCESS 2.0 Series 401 Paper Listening and Reading domains since the items were drawn from the ACCESS 1.0 Series 303 assessment (see Section 1.3.3). 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. For equating summaries for ACCESS 1.0 Series 303 assessment, see the Annual Technical Report for Series 303 (Center for Applied Linguistics, 2016).

Equating summaries are provided for ACCESS 2.0 Series 401 Paper Writing and Speaking. Series 401 Paper Writing and Speaking tasks are exactly the same as Series 401 Online Writing and Speaking. Some Writing and Speaking items were refreshed where a certain number of items on each test form is refreshed.

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. In the domains of Listening and Reading, the scaling equations from the original ACCESS scaling study continue to be used.

In the domains of Writing and Speaking, a scale reconstruction study was conducted in summer 2016, and new scaling equations were derived (see Center for Applied Linguistics, 2016).

Table D in Section 6 provides the equation for converting Rasch ability measures in logits to ACCESS scale scores.

5.1.4 DIF Analyses

Differential item function 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.)

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.

5.1.5 Analysis of Writing and Speaking Forms

As noted in Chapter 1, Section 1.2.5, ACCESS 2.0 Series 401 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 0–40. For all other grade-level clusters, the range on the Tier A test is 0–27. For all Tier B or C tests, the range is 0–54 (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, increases by grade within a grade-level cluster. For example, for Reading 6–8 Tier A, the scale score is capped at 373 for Grade 6, 380 for Grade 7, and 386 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.)

5.2.4 Scaling Equation Table (Table D)

Table D provides the scaling equation for each 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.

5.2.5 Equating Summary (Table E)

For Series 401 Paper, equating summaries are provided for Writing and Speaking only. In the domains of Listening and Reading, items were drawn from the Series 303 assessment (see Section 1.3.3), therefore no equating summaries are provided. For equating summaries for Series 303, please refer to ACCESS 1.0 Series 303 Annual Technical Report (Center for Applied Linguistics, 2016) Series 401 Paper Writing and Speaking tests are the same as Series 401 Online Writing and Speaking tests and the equating analyses were conducted using the Online students' responses. The equating summaries provided in this section contain the same information as the equating summaries for Series 401 Online Writing and Speaking tests (provided in the technical report for ACCESS Series 401 Online).

Series 401 Paper Writing and Speaking each include tasks which were included on Series 400 Writing and Speaking, and tasks which are new to Series 401. An equating procedure known as common-item equating is used to equate the results on the new forms to the older forms. In this procedure, the difficulty measures for items that appear on both the new and the old forms are kept constant across both forms. Thus, performances on the newer form may be interpreted with the same frame of reference. Many tasks appearing on ACCESS 2.0 Paper Series 401 also appeared on Series 400. All tasks common to both forms were anchored to their 400 values in the first equating run. In addition, for the Speaking domain, difficulty measures for the new tasks were anchored to their initial calibrated values from the Speaking field test analysis.[1] After the first equating run, some items that were originally anchored proved to have changed in their difficulty measure. This change is measured by the "Displacement" statistic. This statistic shows the difference between the difficulty value of the anchored item and what the difficulty value would have been had it not been anchored. Typically, random displacements of less than 0.5 logits are unlikely to have much impact on measurement in a test instrument (Linacre, n.d.). For Writing tasks, if this value was large (i.e., usually above .30 or below -.30), that item was unanchored in the final equating run (i.e., it was treated as if it were a new item). For Speaking tasks, a slightly different displacement criterion (above .50 or below -.50) was used since anchored tasks from the Speaking domain have been shown to be less stable than items and tasks from other domains.

For the Writing domain, the equating analysis was conducted using Series 401 Online operational data collected during the early testing window. The Writing equating study was conducting a random sample drawn with a target sample size of 3,000 students per grade-level cluster. The Writing equating sample was drawn proportionate to the Series 400 population for the Writing domain, by grade and tiered form.

For the Speaking domain, student data from the Series 401 Online appended Speaking field test administration were used to conduct the initial equating. These initial item parameters were then verified using Series 401 Online operational data collected during the early testing window. The

[1] The Speaking field test was administered only to those students who take ACCESS Online. For detail on the administration of the appended Speaking field test, please see the companion report for Series 401 Online.

verification study was conducted using a random sample drawn with a target of 3,000 Online students per grade-level cluster, proportionate to the Series 400 Online Speaking population by grades and tiers.

Table E presents a summary of the equating procedures. The first section of the table compares the current test (i.e., the Series 401 version of that test form) to the previous year's test (i.e., the Series 400 version of that test form). The number of items, the average item difficulty, the standard deviation of the item difficulty values, and the difficulty value of the easiest and hardest item on each test form are shown. These values are in terms of logits used in the Rasch measurement model.

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

For Writing and Speaking tasks, the table next provides the anchored step measures for the total score on each task. For the ACCESS Writing and Speaking tasks, a Rasch-group rating scale model is used (see 5.1.1 above). For Writing, the step difficulty values are the same for all the tasks that are scored on the 0-9 raw score scale. These constant step difficulty values help to provide anchors in the calibration of new Writing tasks onto the common WIDA score scale each year. For Speaking, the step difficulty values for all P1 tasks are the same and the step difficulty values for all P3 and P5 tasks are the same. As with Writing, these constant step difficulty values help to provide anchors in the calibration of new Speaking tasks onto the common WIDA score scale each year.

The final section of Table E shows the location of the anchor items or tasks, both by order of displacement statistics and by order of item difficulty. The displacement statistics provide information on the difference between the difficulty value of the anchored item and what that difficulty value would have been had the item not been anchored. Smaller displacement statistics indicate more consistency between the item's difficulty value on the Series 401 test form and on the Series 400 test form. It is desirable that the anchor items represent a wide range of difficulties across the entire spectrum of the item difficulty values on a test form. The greater the representation across the difficulty range, the more trustworthy the equating results will be.

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

5.2.6 Test Characteristic Curve (Figure D)

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.

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.

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

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 or task type (selected response or constructed response), the average item or task difficulty (in logits), the number of items or tasks, 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, the 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. The original scale scores subject to adjustment were computed from logit scores. Implementing the end-of-scale adjustment reduces the range of the original scale scores, however the lower and upper bounds of the adjusted scale scores always fall within the range of the original unadjusted scale scores.

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, Ryu, & MacGregor, 2013.) Hence, Table J includes a separate column for each grade.

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). This capped score is dependent on the grade level (not 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.

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 401, 2016–2017 Administration

Annual Technical Report No. 13B
Volume 3 of 3: Analyses Across Tiers

Prepared by:

Center for Applied Linguistics

Language Assessment Division
Psychometrics and Quantitative Research Team

April 25, 2018

Volume 3

7. Analysis Across Tiers: Overview 1

7.1 Background 1

7.1.1 Reliability of Composites 1

7.1.2 Accuracy and Consistency of Classification 1

7.2 Descriptions 3

7.2.1 Scale Score Information (Figure A and Table A) 3

7.2.2 Proficiency Level Information (Figure B and Table B) 3

7.2.3 Conditional Standard Error of Measurement at Cut Scores (Table C) 4

7.2.4 Test Characteristic Curve (Figure C) 4

7.2.5 Test Information Function (Figure D) 4

7.2.6 Reliability Information (Table D) 5

7.2.7 Accuracy and Consistency of Classification Tables (Table E) 5

8 Analyses Across Tiers: Results 6

8.1 Grade: K 6

8.1.1 Listening K 6

8.1.2 Reading K 8

8.1.3 Writing K 10

8.1.4 Speaking K 12

8.1.5 Oral Language Composite K 14

8.1.6 Literacy Composite K 16

8.1.7 Comprehension Composite K 18

8.1.8 Overall Composite K 20

8.2 Grade: 1 22

8.2.1 Listening 1 22

8.2.2 Reading 1 24

8.2.3 Writing 1 26

8.2.4 Speaking 1 28

8.2.5 Oral Language Composite 1 30

8.2.6 Literacy Composite 1 32
8.2.7 Comprehension Composite 1 34
8.2.8 Overall Composite 1 36
8.3 Grade: 2 38
8.3.1 Listening 2 38
8.3.2 Reading 2 40
8.3.3 Writing 2 42
8.3.4 Speaking 2 44
8.3.5 Oral Language Composite 2 46
8.3.6 Literacy Composite 2 48
8.3.7 Comprehension Composite 2 50
8.3.8 Overall Composite 2 52
8.4 Grade: 3 54
8.4.1 Listening 3 54
8.4.2 Reading 3 56
8.4.3 Writing 3 58
8.4.4 Speaking 3 60
8.4.5 Oral Language Composite 3 62
8.4.6 Literacy Composite 3 64
8.4.7 Comprehension Composite 3 66
8.4.8 Overall Composite 3 68
8.5 Grades: 4–5 70
8.5.1 Listening 4–5 70
8.5.2 Reading 4–5 73
8.5.3 Writing 4–5 76
8.5.4 Speaking 4–5 79
8.5.5 Oral Language Composite 4–5 82
8.5.6 Literacy Composite 4–5 85
8.5.7 Comprehension Composite 4–5 88
8.5.8 Overall Composite 4–5 91
8.6 Grades: 6–8 94
8.6.1 Listening 6–8 94
8.6.2 Reading 6–8 98

8.6.3 Writing 6–8 102
8.6.4 Speaking 6–8 106
8.6.5 Oral Language Composite 6–8 110
8.6.6 Literacy Composite 6–8 113
8.6.7 Comprehension Composite 6–8 116
8.6.8 Overall Composite 6–8 119
8.7 Grades: 9–12 122
8.7.1 Listening 9–12 122
8.7.2 Reading 9–12 126
8.7.3 Writing 9–12 130
8.7.4 Speaking 9–12 134
8.7.5 Oral Language Composite 9–12 138
8.7.6 Literacy Composite 9–12 142
8.7.7 Comprehension Composite 9–12 146
8.7.8 Overall Composite 9–12 150

References 154

Acknowledgements 156

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 stratified Cronbach's alpha is computed for the composites, it is computed by grade-level cluster, as there is not a one-to-one relationship between test forms and grade-level clusters. 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). However, 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.

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 is presented only by domain.

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 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.)

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 is provided only by domains.) 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, therefore direct comparison between tiers has limitations. 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 is provided only by domains.) 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. This information is presented by cluster. 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 tiered 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). These tables are presented by grade. 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 scores 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.

8 Analyses Across Tiers: Results

8.1 Grade: K

8.1.1 Listening K

Figure 8.1.1A
Scale Scores: List K S401 Paper

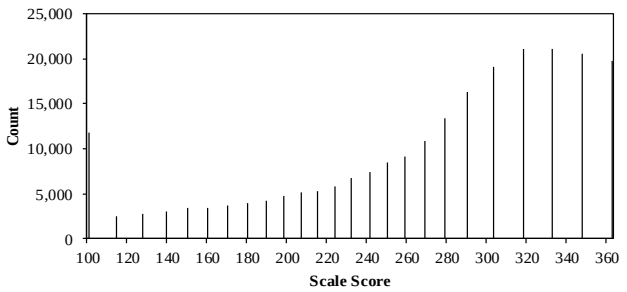


Table 8.1.1A
Scale Score Descriptive Statistics: List K S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	233,593	100	363	269.31	73.06

Figure 8.1.1B
Proficiency Level: List K S401 Paper

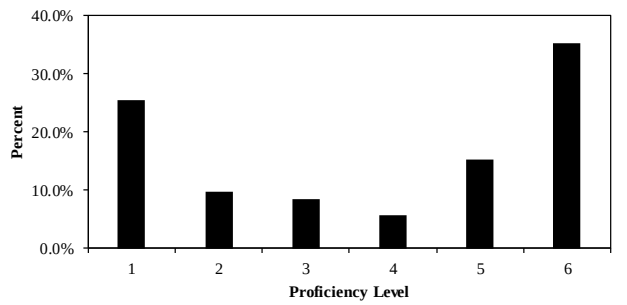
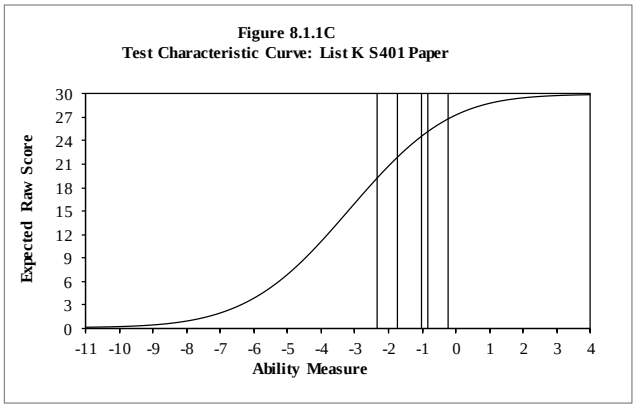


Table 8.1.1B
Proficiency Level Distribution:
List K S401 Paper

Level	Count	Percent
1	59,724	25.6%
2	22,676	9.7%
3	19,934	8.5%
4	13,397	5.7%
5	35,434	15.2%
6	82,428	35.3%
Total	233,593	100.0%

Table 8.1.1C
Conditional Standard Error of Measurement
at Cut Scores: List K S401 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



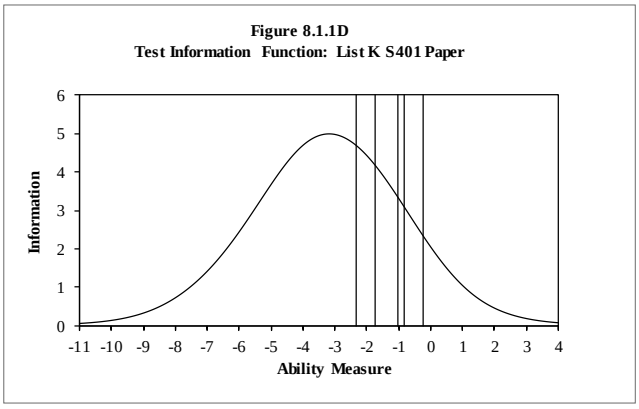


Table 8.1.1D

Reliability: List K S401 Paper

Tiers	No. of Students	Reliability
-	233,593	0.939

Table 8.1.1E

Accuracy and Consistency of Classification Indices: List (Grade K) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.688	0.624	0.504		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.878	0.826		
	2	0.462	0.344		
	3	0.328	0.249		
	4	0.215	0.158		
	5	0.464	0.351		
	6	0.825	0.774		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.943	0.032	0.024	0.921
	2/3	0.931	0.027	0.042	0.905
	3/4	0.919	0.046	0.034	0.888
	4/5	0.910	0.042	0.048	0.879
	5/6	0.899	0.033	0.068	0.860

8.1.2 Reading K

Figure 8.1.2A
Scale Scores: Read K S401 Paper

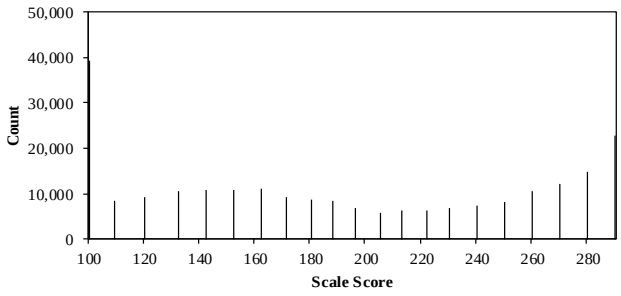


Table 8.1.2A
Scale Score Descriptive Statistics: Read K S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	233,592	100	290	189.73	67.16

Figure 8.1.2B
Proficiency Level: Read K S401 Paper

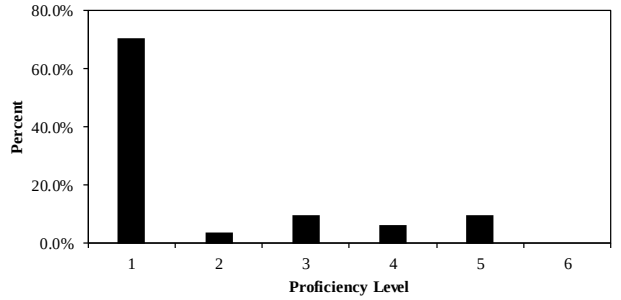
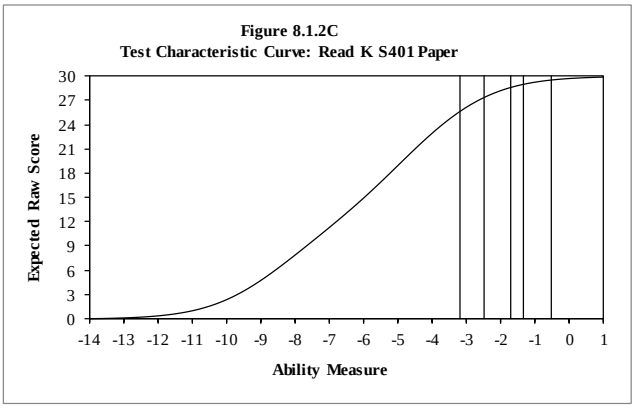


Table 8.1.2B
Proficiency Level Distribution:
Read K S401 Paper

Level	Count	Percent
1	165,131	70.7%
2	8,267	3.5%
3	22,717	9.7%
4	14,680	6.3%
5	22,797	9.8%
6	0	0.0%
Total	233,592	100.0%

Table 8.1.2C
Conditional Standard Error of Measurement
at Cut Scores: Read K S401 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



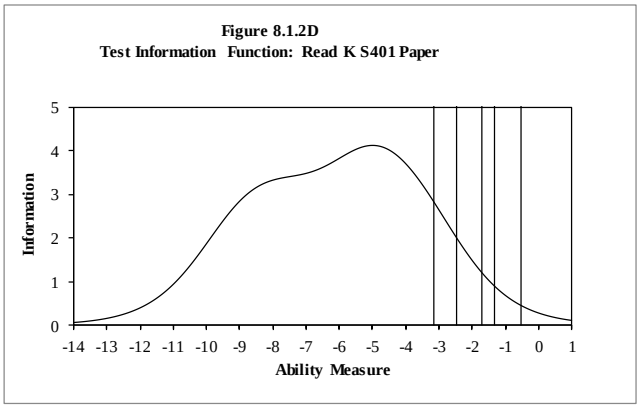


Table 8.1.2D

Reliability: Read K S401 Paper

Tiers	No. of Students	Reliability
-	233,592	0.949

Table 8.1.2E

Accuracy and Consistency of Classification Indices: Read (Grade K) S401 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)
	0.805	0.776	0.541

Conditional on Level	Level	Accuracy	Consistency
	1	0.974	0.963
	2	0.236	0.164
	3	0.537	0.410
	4	0.302	0.265
	5	0.740	0.573
	6	N/A	N/A

Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.956	0.018	0.026	0.935
	2/3	0.957	0.030	0.013	0.937
	3/4	0.946	0.034	0.019	0.925
	4/5	0.918	0.073	0.009	0.912
	5/6	N/A	N/A	N/A	N/A

8.1.3 Writing K

Figure 8.1.3A
Scale Scores: Writ K S401 Paper

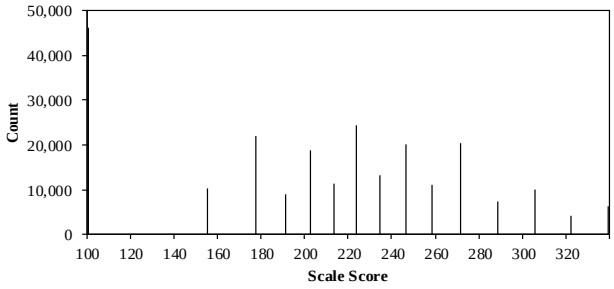


Table 8.1.3A
Scale Score Descriptive Statistics: Writ K S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	233,593	100	339	206.74	66.82

Figure 8.1.3B
Proficiency Level: Writ K S401 Paper

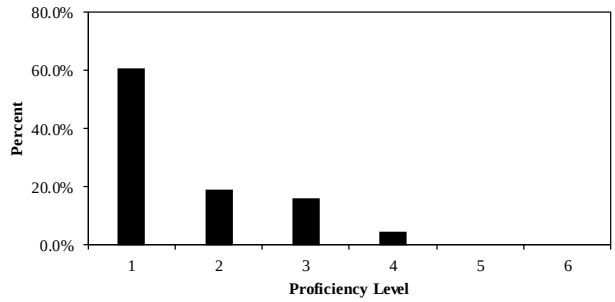
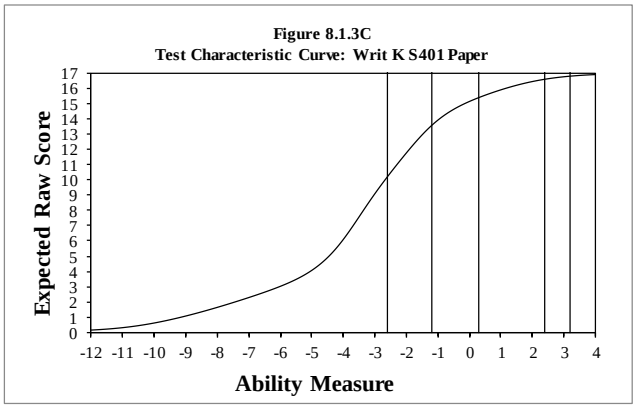


Table 8.1.3B
Proficiency Level Distribution:
Writ K S401 Paper

Level	Count	Percent
1	141,607	60.6%
2	44,143	18.9%
3	37,550	16.1%
4	10,293	4.4%
5	0	0.0%
6	0	0.0%
Total	233,593	100.0%

Table 8.1.3C
Conditional Standard Error of Measurement
at Cut Scores: Writ K S401 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



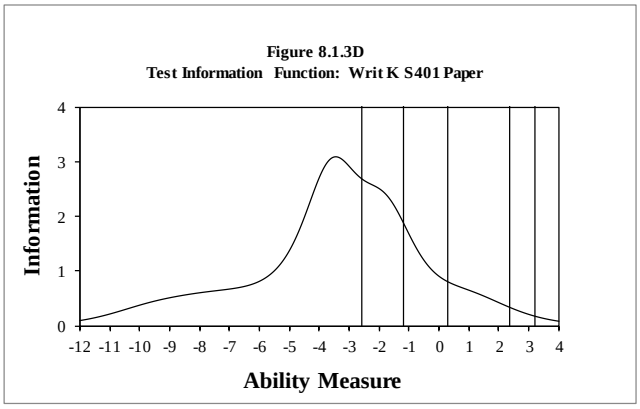


Table 8.1.3D

Reliability: Writ K S401 Paper

Tiers	No. of Students	Reliability
-	233,593	0.923

Table 8.1.3E

Accuracy and Consistency of Classification Indices: Writ (Grade K) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.775	0.720	0.509		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.952	0.921		
	2	0.528	0.404		
	3	0.519	0.463		
	4	-	0.177		
	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	0.926	0.028	0.046	0.895
	2/3	0.885	0.045	0.070	0.840
	3/4	0.956	0.044	0.000	0.949
	4/5	N/A	N/A	N/A	N/A
	5/6	N/A	N/A	N/A	N/A

8.1.4 Speaking K

Figure 8.1.4A
Scale Scores: Spek K S401 Paper

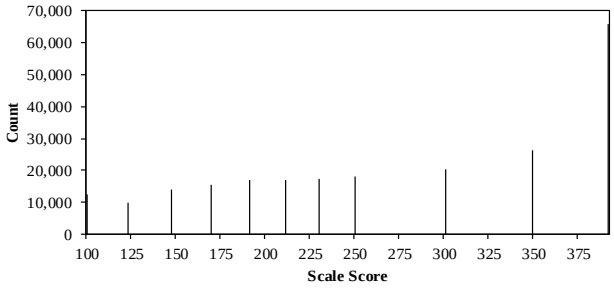


Table 8.1.4A
Scale Score Descriptive Statistics: Spek K S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	233,589	100	392	271.97	99.40

Figure 8.1.4B
Proficiency Level: Spek K S401 Paper

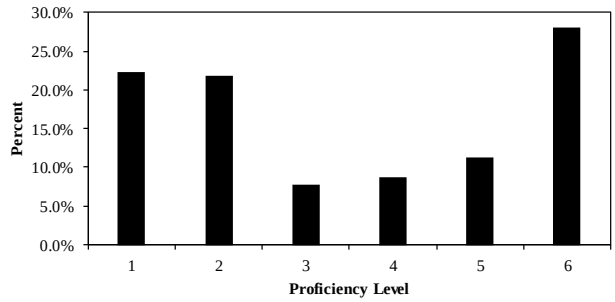
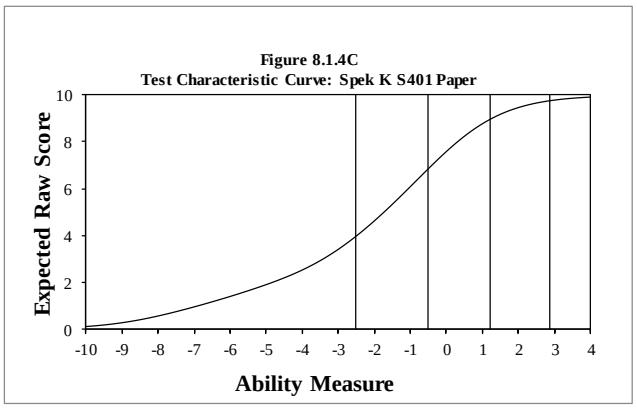


Table 8.1.4B
Proficiency Level Distribution:
Spek K S401 Paper

Level	Count	Percent
1	52,049	22.3%
2	51,007	21.8%
3	18,160	7.8%
4	20,238	8.7%
5	26,446	11.3%
6	65,689	28.1%
Total	233,589	100.0%

Table 8.1.4C
Conditional Standard Error of Measurement
at Cut Scores: Spek K S401 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



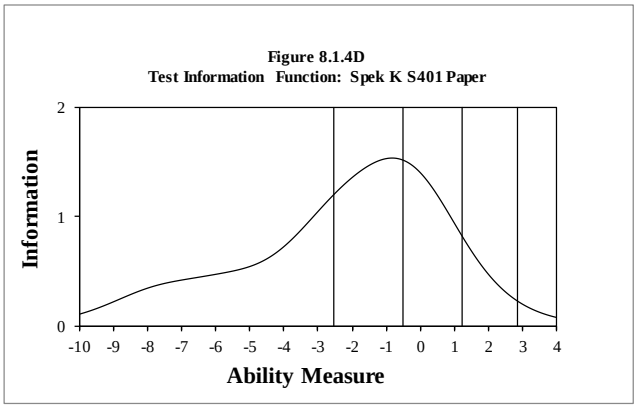


Table 8.1.4D

Reliability: Spek K S401 Paper

Tiers	No. of Students	Reliability
-	233,589	0.899

Table 8.1.4E

Accuracy and Consistency of Classification Indices: Spek (Grade K) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.453	0.464	0.349		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.683	0.623		
	2	0.632	0.492		
	3	0.349	0.254		
	4	0.468	0.324		
	5	0.244	0.240		
	6	-	0.740		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.880	0.090	0.030	0.843
	2/3	0.927	0.029	0.044	0.898
	3/4	0.948	0.027	0.025	0.923
	4/5	0.951	0.025	0.024	0.924
	5/6	0.719	0.281	0.000	0.792

8.1.5 Oral Language Composite K

Figure 8.1.5A
Scale Scores: Oral K S401 Paper

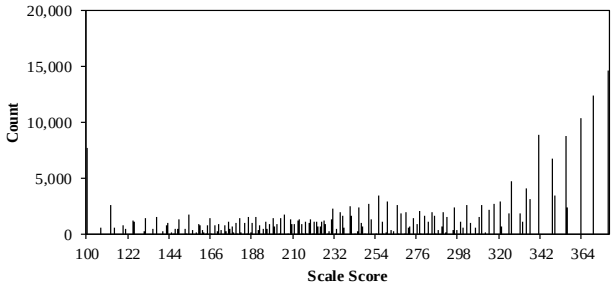


Table 8.1.5A
Scale Score Descriptive Statistics: Oral K S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	233,582	100	378	270.90	81.29

Figure 8.1.5B
Proficiency Level: Oral K S401 Paper

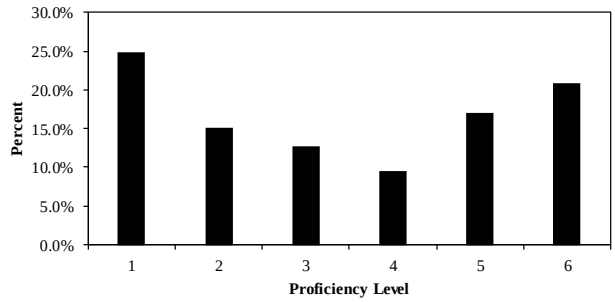


Table 8.1.5B
Proficiency Level Distribution:
Oral K S401 Paper

Level	Count	Percent
1	58,046	24.9%
2	35,290	15.1%
3	29,582	12.7%
4	22,331	9.6%
5	39,669	17.0%
6	48,664	20.8%
Total	233,582	100.0%

Table 8.1.5C
n/a

Figure 8.1.5C
n/a

Figure 8.1.5D
n/a

Table 8.1.5D
Oral Composite Reliability: Oral K S401 Paper

Component	Weight	Variance	Reliability
Listening	0.50	5337.016	0.939
Speaking	0.50	9880.902	0.899
Oral		6607.790	0.950

*Variances from students who had results in all four domains

Table 8.1.5E

Accuracy and Consistency of Classification Indices: Oral (Grade K) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.702	0.614	0.529
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.885	0.831
	2	0.627	0.511
	3	0.530	0.412
	4	0.426	0.317
	5	0.628	0.512
	6	0.851	0.771

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.945	0.029	0.027	0.922
	2/3	0.936	0.027	0.037	0.911
	3/4	0.941	0.029	0.030	0.914
	4/5	0.934	0.039	0.027	0.908
	5/6	0.934	0.036	0.030	0.907

8.1.6 Literacy Composite K

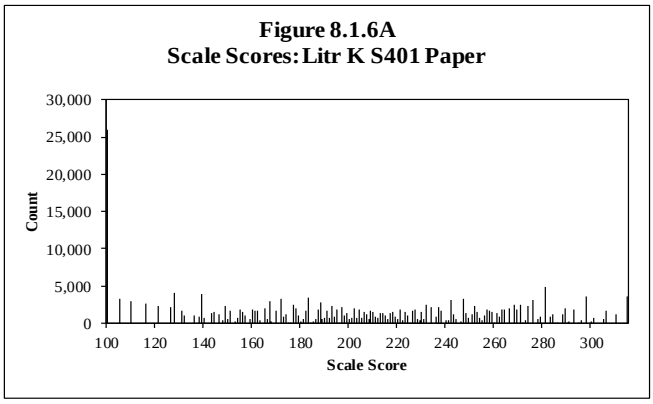


Table 8.1.6A

Scale Score Descriptive Statistics: Litr K S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	233,587	100	315	198.48	62.23

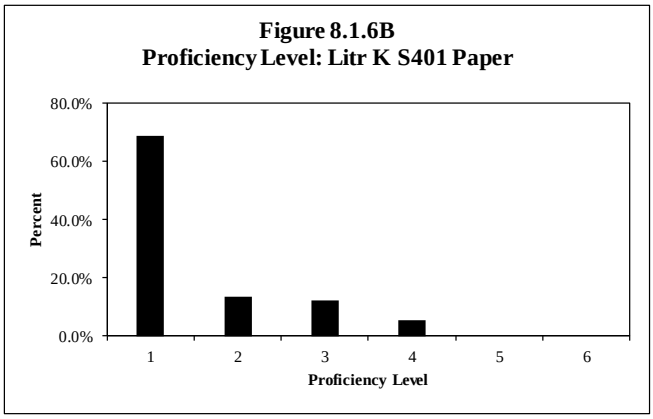


Table 8.1.6B

Proficiency Level Distribution: Litr K S401 Paper

Level	Count	Percent
1	160,754	68.8%
2	31,435	13.5%
3	28,913	12.4%
4	12,485	5.3%
5	0	0.0%
6	0	0.0%
Total	233,587	100.0%

Table 8.1.6C

n/a

Figure 8.1.6C

n/a

Figure 8.1.6D

n/a

Table 8.1.6D

Literacy Composite Reliability: Litr K S401 Paper

Component	Weight	Variance	Reliability
Reading	0.50	4510.975	0.949
Writing	0.50	4464.667	0.923
Literacy		3872.029	0.963

*Variances from students who had results in all four domains

Table 8.1.6E

Accuracy and Consistency of Classification Indices: Litr (Grade K) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.848	0.810	0.613

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.963	0.948
	2	0.643	0.514
	3	0.563	0.517
	4	-	0.439
	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	0.948	0.026	0.027	0.927
	2/3	0.953	0.020	0.027	0.931
	3/4	0.947	0.053	0.000	0.944
	4/5	N/A	N/A	N/A	N/A
	5/6	N/A	N/A	N/A	N/A

8.1.7 Comprehension Composite K

Figure 8.1.7A
Scale Scores: Cphn K S401 Paper

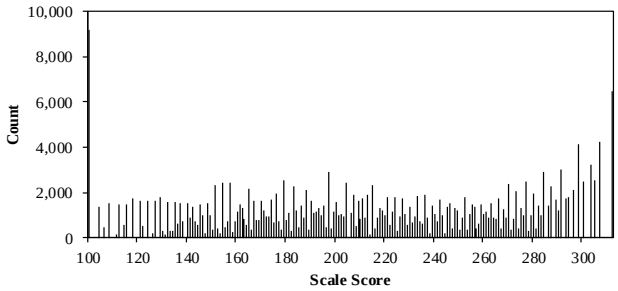


Table 8.1.7A
Scale Score Descriptive Statistics: Cphn K S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	233,585	100	312	213.60	61.52

Figure 8.1.7B
Proficiency Level: Cphn K S401 Paper

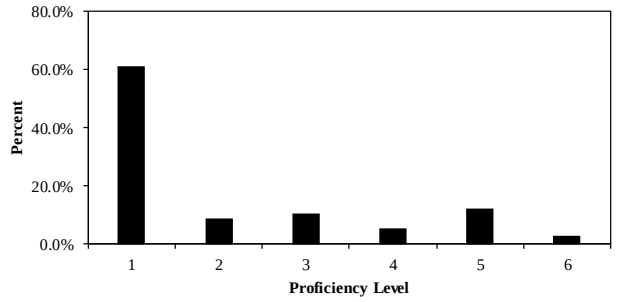


Table 8.1.7B
Proficiency Level Distribution:
Cphn K S401 Paper

Level	Count	Percent
1	142,366	60.9%
2	20,134	8.6%
3	24,096	10.3%
4	12,344	5.3%
5	28,183	12.1%
6	6,462	2.8%
Total	233,585	100.0%

Table 8.1.7C
n/a

Figure 8.1.7C
n/a

Figure 8.1.7D
n/a

Table 8.1.7D
Comprehension Composite Reliability: Cphn K S401 Paper

Component	Weight	Variance	Reliability
Listening	0.30	5337.016	0.939
Reading	0.70	4510.975	0.949
Comprehension		3784.041	0.963

*Variances from students who had results in all four domains

Table 8.1.7E

Accuracy and Consistency of Classification Indices: Cphn (Grade K) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.810	0.755	0.589

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.959	0.943
	2	0.474	0.358
	3	0.561	0.443
	4	0.361	0.271
	5	0.741	0.649
	6	0.744	0.503

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.948	0.025	0.027	0.926
	2/3	0.953	0.026	0.021	0.933
	3/4	0.954	0.028	0.018	0.936
	4/5	0.957	0.025	0.018	0.940
	5/6	0.981	0.014	0.005	0.973

8.1.8 Overall Composite K

Figure 8.1.8A
Scale Scores: Over K S401 Paper

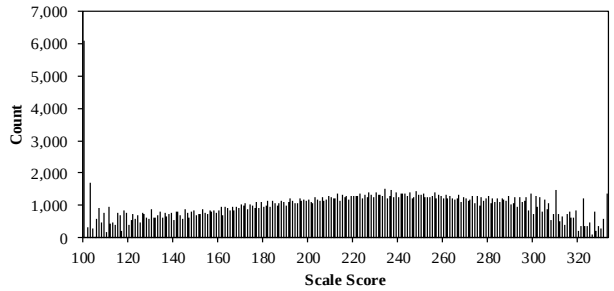


Table 8.1.8A
Scale Score Descriptive Statistics: Over K S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
K	233,571	100	333	219.98	61.35

Figure 8.1.8B
Proficiency Level: Over K S401 Paper

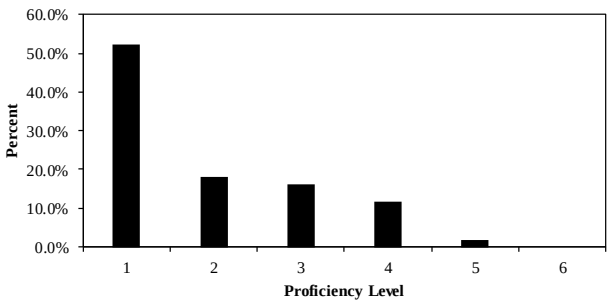


Table 8.1.8B
Proficiency Level Distribution:
Over K S401 Paper

Level	Count	Percent
1	121,933	52.2%
2	42,292	18.1%
3	38,029	16.3%
4	27,101	11.6%
5	4,216	1.8%
6	0	0.0%
Total	233,571	100.0%

Table 8.1.8C
n/a

Figure 8.1.8C
n/a

Figure 8.1.8D
n/a

Table 8.1.8D
Overall Composite Reliability: Over K S401 Paper

Component	Weight	Variance	Reliability
Listening	0.15	5337.016	0.939
Reading	0.35	4510.975	0.949
Speaking	0.15	9880.902	0.899
Writing	0.35	4464.667	0.923
Overall Composite		3763.784	0.973

*Variances from students who had results in all four domains

Table 8.1.8E

Accuracy and Consistency of Classification Indices: Over (Grade K) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.850	0.797	0.689

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.952	0.932
	2	0.748	0.648
	3	0.736	0.635
	4	0.728	0.666
	5	-	0.333
	6	N/A	N/A

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.952	0.025	0.023	0.932
	2/3	0.954	0.021	0.025	0.934
	3/4	0.963	0.019	0.018	0.947
	4/5	0.982	0.018	0.000	0.982
	5/6	N/A	N/A	N/A	N/A

8.2 Grade: 1

8.2.1 Listening 1

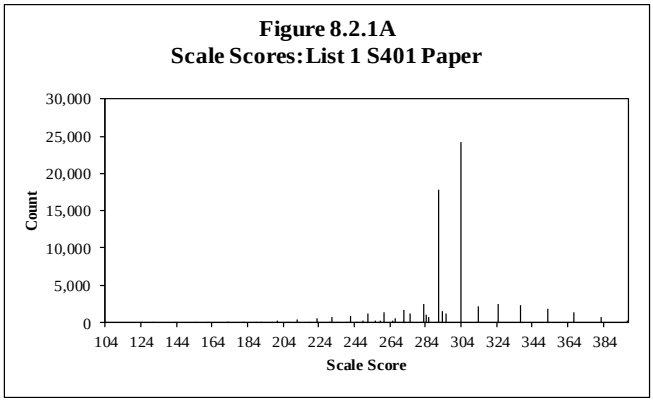


Table 8.2.1A

Scale Score Descriptive Statistics: List 1 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	71,276	104	397	295.82	29.08

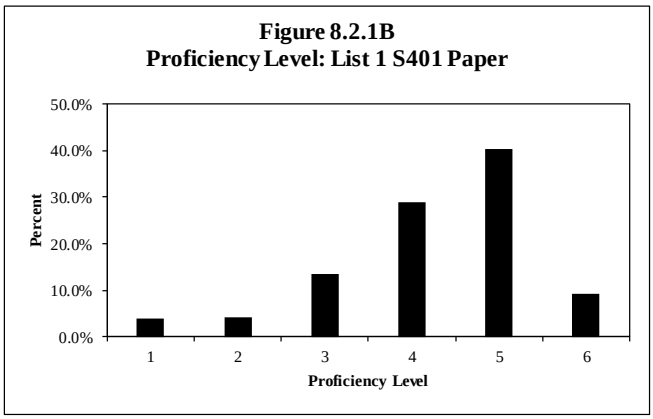


Table 8.2.1B

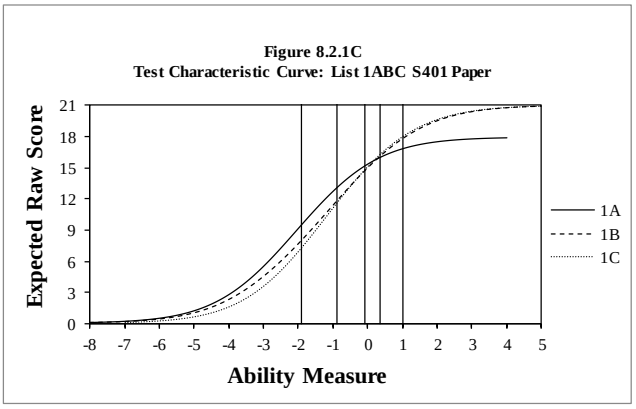
Proficiency Level Distribution:
List 1 S401 Paper

Level	Count	Percent
1	2,735	3.8%
2	3,042	4.3%
3	9,597	13.5%
4	20,639	29.0%
5	28,775	40.4%
6	6,488	9.1%
Total	71,276	100.0%

Table 8.2.1C

Conditional Standard Error of Measurement at Cut Scores: List 1 S401 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	1	236	19.16	20.29	19.16
2/3	1	259	19.16	19.16	18.41
3/4	1	291	22.17	19.54	18.79
4/5	1	303	n/a	19.91	19.24
5/6	1	327	n/a	n/a	21.42



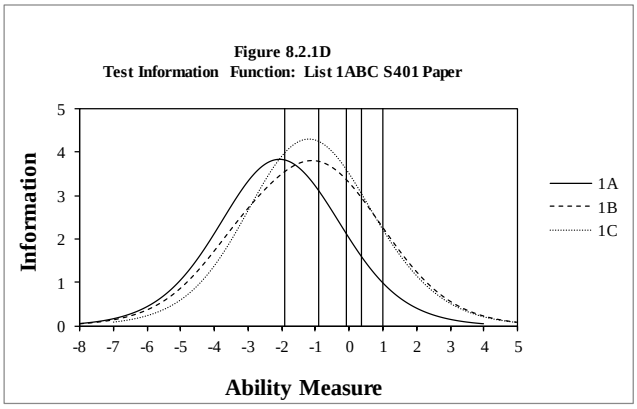


Table 8.2.1D

Weighted Reliability: List 1 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	28,097	0.778	0.696
B	25,565	0.664	
C	17,614	0.609	

Table 8.2.1E

Accuracy and Consistency of Classification Indices: List (Grade 1) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.488	0.371	0.177		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.857	0.614		
	2	0.349	0.225		
	3	0.344	0.243		
	4	0.440	0.368		
	5	0.600	0.525		
	6	0.498	0.280		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.973	0.002	0.025	0.967
	2/3	0.958	0.010	0.032	0.930
	3/4	0.827	0.129	0.044	0.756
	4/5	0.752	0.131	0.117	0.691
	5/6	0.909	0.053	0.039	0.844

8.2.2 Reading 1

Figure 8.2.2A
Scale Scores: Read 1 S401 Paper

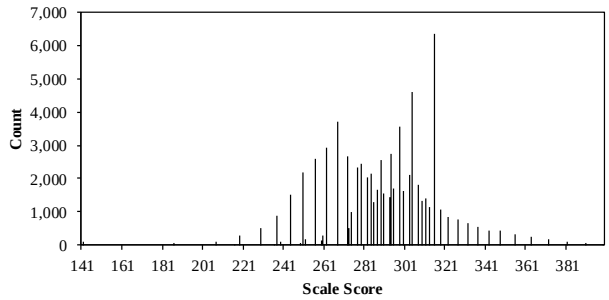


Table 8.2.2A
Scale Score Descriptive Statistics: Read 1 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	71,311	141	399	289.20	25.85

Figure 8.2.2B
Proficiency Level: Read 1 S401 Paper

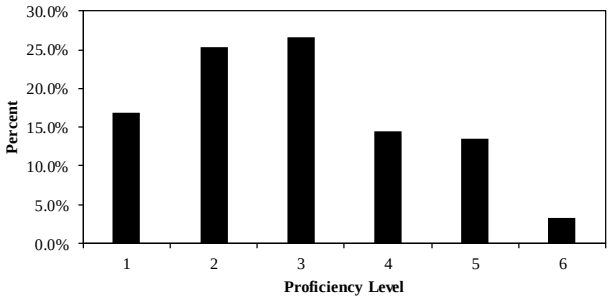
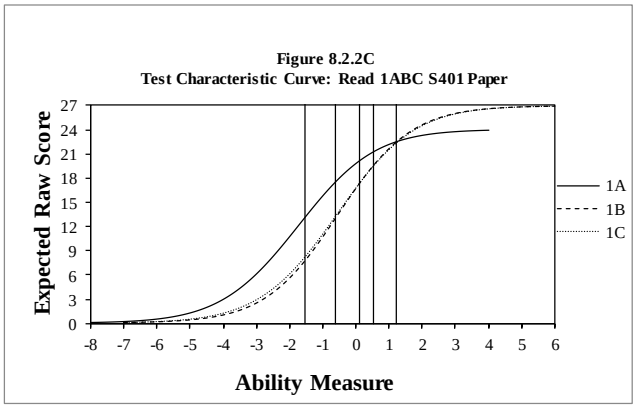


Table 8.2.2B
Proficiency Level Distribution:
Read 1 S401 Paper

Level	Count	Percent
1	11,979	16.8%
2	18,079	25.4%
3	18,954	26.6%
4	10,303	14.4%
5	9,682	13.6%
6	2,314	3.2%
Total	71,311	100.0%

Table 8.2.2C
Conditional Standard Error of Measurement at Cut Scores: Read 1 S401 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	1	264	11.70	13.78	14.04
2/3	1	286	11.44	11.18	11.70
3/4	1	304	12.37	10.40	10.87
4/5	1	315	n/a	10.53	10.92
5/6	1	334	n/a	n/a	11.44



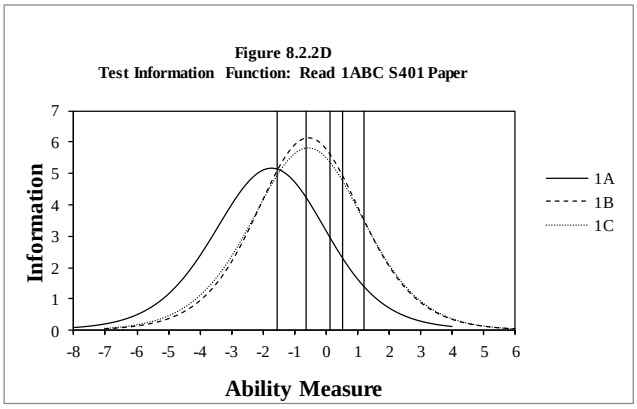


Table 8.2.2D

Weighted Reliability: Read 1 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	28,115	0.717	0.733
B	25,571	0.741	
C	17,625	0.748	

Table 8.2.2E

Accuracy and Consistency of Classification Indices: Read (Grade 1) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.506	0.401	0.248

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.775	0.603
	2	0.522	0.415
	3	0.464	0.377
	4	0.325	0.251
	5	0.506	0.365
	6	0.623	0.269

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.906	0.030	0.064	0.866
	2/3	0.838	0.075	0.087	0.775
	3/4	0.841	0.096	0.063	0.782
	4/5	0.884	0.080	0.035	0.839
	5/6	0.969	0.028	0.003	0.953

8.2.3 Writing 1

Figure 8.2.3A
Scale Scores: Writ 1 S401 Paper

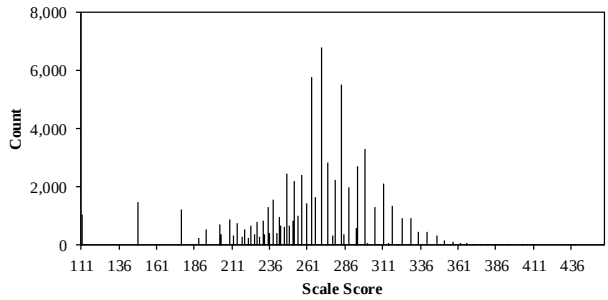


Table 8.2.3A
Scale Score Descriptive Statistics: Writ 1 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	71,305	111	457	262.12	40.91

Figure 8.2.3B
Proficiency Level: Writ 1 S401 Paper

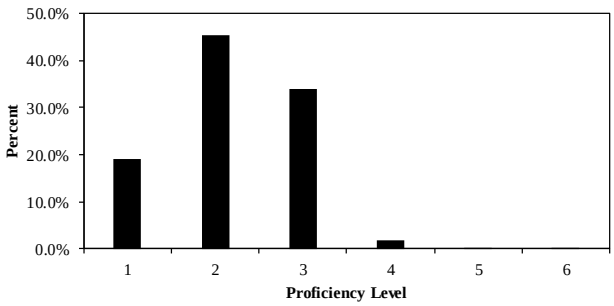
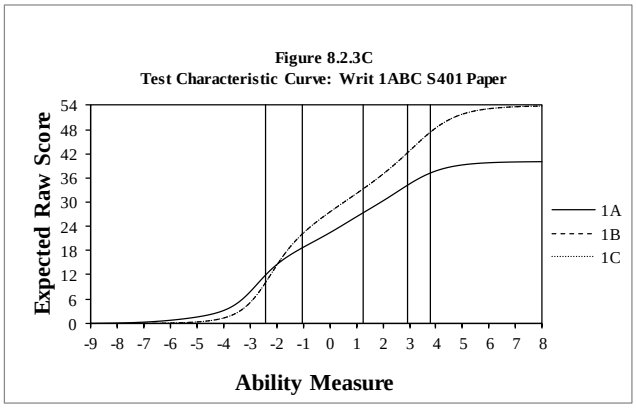


Table 8.2.3B
Proficiency Level Distribution:
Writ 1 S401 Paper

Level	Count	Percent
1	13,514	19.0%
2	32,296	45.3%
3	24,238	34.0%
4	1,242	1.7%
5	13	0.0%
6	2	0.0%
Total	71,305	100.0%

Table 8.2.3C
Conditional Standard Error of Measurement at Cut Scores: Writ 1 S401 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	1	238	10.31	8.54
2/3	1	275	13.96	10.74
3/4	1	337	13.69	12.35
4/5	1	382	13.16	10.74
5/6	1	405	16.38	11.81



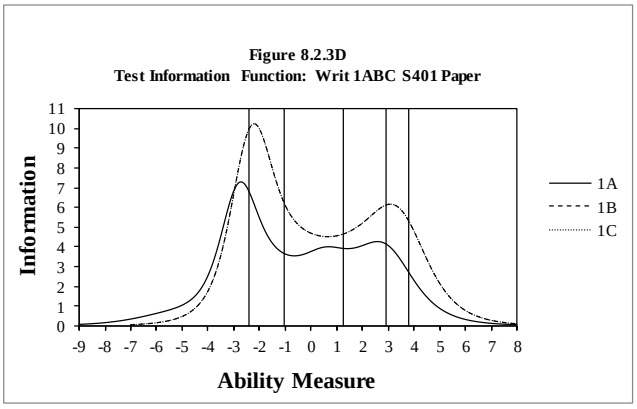


Table 8.2.3D

Weighted Reliability: Writ 1 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	28,106	0.874	0.912
B/C	43,199	0.937	

Table 8.2.3E

Accuracy and Consistency of Classification Indices: Writ (Grade 1) S401 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.786	0.715		0.564	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.834		0.747	
	2	0.857		0.765	
	3	0.705		0.659	
	4	-		0.118	
	5	-		-	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.945	0.033	0.022	0.918
	2/3	0.858	0.029	0.112	0.813
	3/4	0.982	0.018	0.000	0.982
	4/5	1.000	0.000	0.000	1.000
	5/6	1.000	0.000	0.000	1.000

8.2.4 Speaking 1

Figure 8.2.4A
Scale Scores: Spek 1 S401 Paper

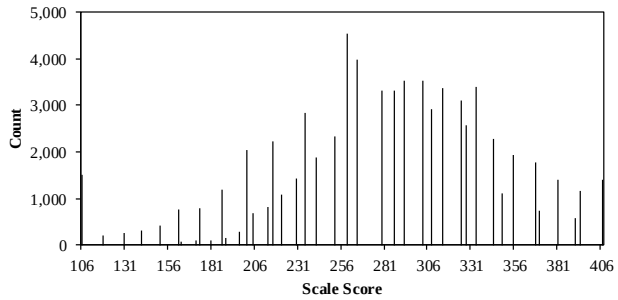


Table 8.2.4A
Scale Score Descriptive Statistics: Spek 1 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	71,303	106	407	282.05	63.73

Figure 8.2.4B
Proficiency Level: Spek 1 S401 Paper

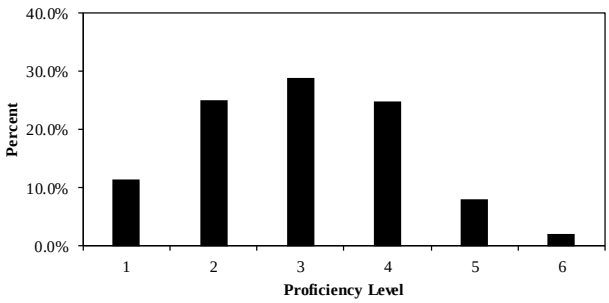
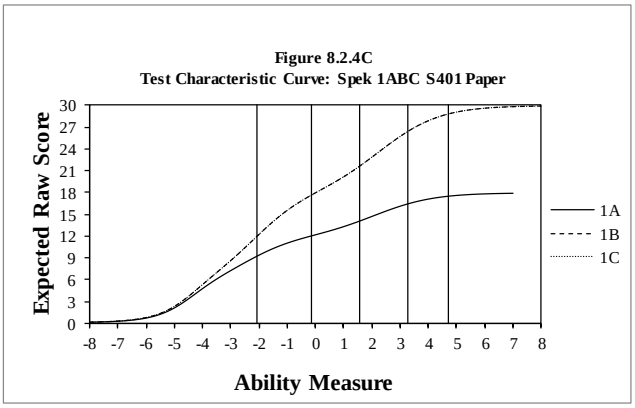


Table 8.2.4B
Proficiency Level Distribution:
Spek 1 S401 Paper

Level	Count	Percent
1	8,123	11.4%
2	17,808	25.0%
3	20,599	28.9%
4	17,720	24.9%
5	5,660	7.9%
6	1,393	2.0%
Total	71,303	100.0%

Table 8.2.4C
Conditional Standard Error of Measurement at Cut Scores: Spek 1 S401 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	1	205	20.77	15.21
2/3	1	261	28.37	19.89
3/4	1	311	24.28	17.55
4/5	1	361	28.08	19.01
5/6	1	403	45.63	29.25



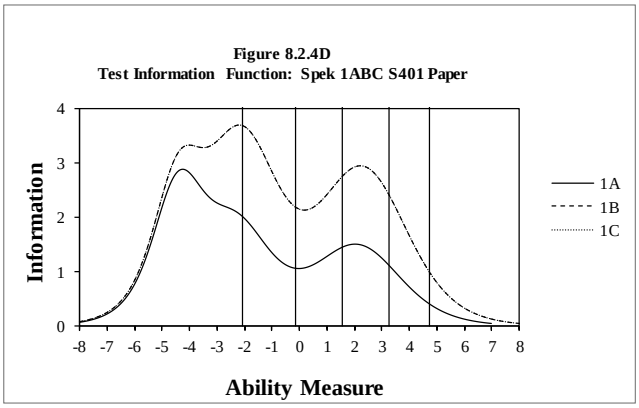


Table 8.2.4D

Weighted Reliability: Spek 1 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	28,111	0.880	0.899
B/C	43,192	0.911	

Table 8.2.4E

Accuracy and Consistency of Classification Indices: Spek (Grade 1) S401 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)	
	0.684	0.575	0.450	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.795	0.674	
	2	0.722	0.615	
	3	0.669	0.564	
	4	0.666	0.562	
	5	0.523	0.415	
	6	-	0.395	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.952	0.023	0.024	0.930
	2/3	0.904	0.043	0.054	0.867
	3/4	0.904	0.044	0.052	0.864
	4/5	0.939	0.041	0.020	0.912
	5/6	0.980	0.020	0.000	0.980

8.2.5 Oral Language Composite 1

Figure 8.2.5A
Scale Scores: Oral 1 S401 Paper

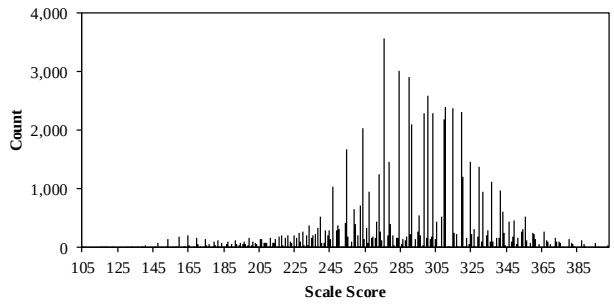


Table 8.2.5A
Scale Score Descriptive Statistics: Oral 1 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	71,247	105	402	289.19	41.39

Figure 8.2.5B
Proficiency Level: Oral 1 S401 Paper

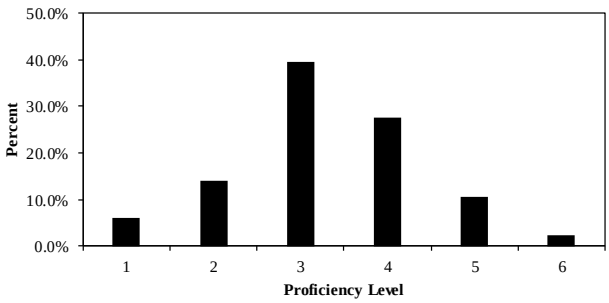


Table 8.2.5B
Proficiency Level Distribution:
Oral 1 S401 Paper

Level	Count	Percent
1	4,336	6.1%
2	9,981	14.0%
3	28,255	39.7%
4	19,610	27.5%
5	7,421	10.4%
6	1,644	2.3%
Total	71,247	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.50	845.329	0.696
Speaking	0.50	4058.802	0.899
Oral		1712.303	0.902

*Variances from students who had results in all four domains

Table 8.2.5E

Accuracy and Consistency of Classification Indices: Oral (Grade 1) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.712	0.607	0.469

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.838	0.719
	2	0.663	0.535
	3	0.789	0.708
	4	0.674	0.568
	5	0.570	0.464
	6	-	0.411

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.976	0.009	0.015	0.966
	2/3	0.932	0.035	0.033	0.902
	3/4	0.891	0.048	0.062	0.847
	4/5	0.935	0.034	0.030	0.904
	5/6	0.977	0.023	0.000	0.976

8.2.6 Literacy Composite 1

Figure 8.2.6A
Scale Scores: Litr 1 S401 Paper

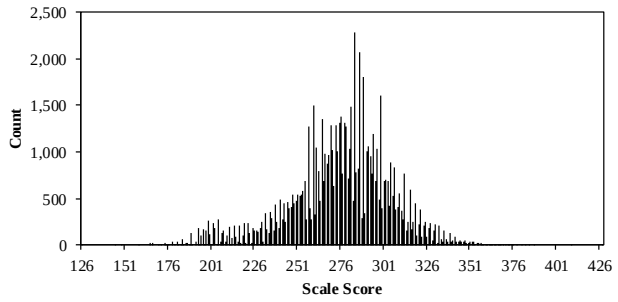


Table 8.2.6A
Scale Score Descriptive Statistics: Litr 1 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	71,288	126	428	275.91	29.52

Figure 8.2.6B
Proficiency Level: Litr 1 S401 Paper

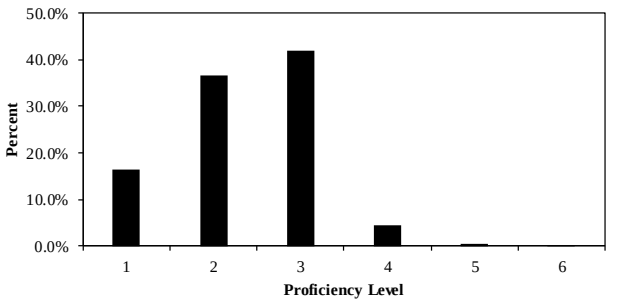


Table 8.2.6B
Proficiency Level Distribution: Litr 1 S401 Paper

Level	Count	Percent
1	11,780	16.5%
2	26,050	36.5%
3	29,988	42.1%
4	3,157	4.4%
5	291	0.4%
6	22	0.0%
Total	71,288	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 S401 Paper

Component	Weight	Variance	Reliability
Reading	0.50	668.053	0.733
Writing	0.50	1673.385	0.912
Literacy		871.209	0.907

*Variances from students who had results in all four domains

Table 8.2.6E

Accuracy and Consistency of Classification Indices: Litr (Grade 1) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.801	0.724	0.582

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.837	0.745
	2	0.796	0.714
	3	0.802	0.755
	4	0.568	0.382
	5	-	0.200
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.946	0.027	0.027	0.922
	2/3	0.898	0.045	0.057	0.857
	3/4	0.960	0.034	0.006	0.947
	4/5	0.996	0.004	0.000	0.996
	5/6	1.000	0.000	0.000	1.000

8.2.7 Comprehension Composite 1

Figure 8.2.7A
Scale Scores: Cphn 1 S401 Paper

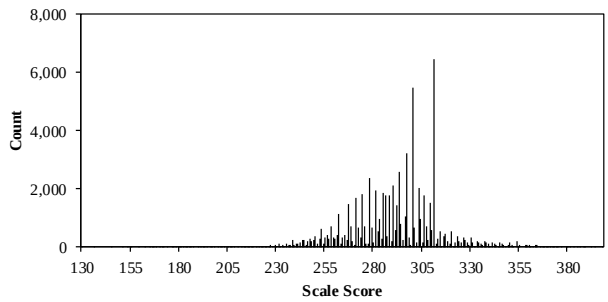


Table 8.2.7A
Scale Score Descriptive Statistics: Cphn 1 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	71,261	130	398	291.22	24.12

Figure 8.2.7B
Proficiency Level: Cphn 1 S401 Paper

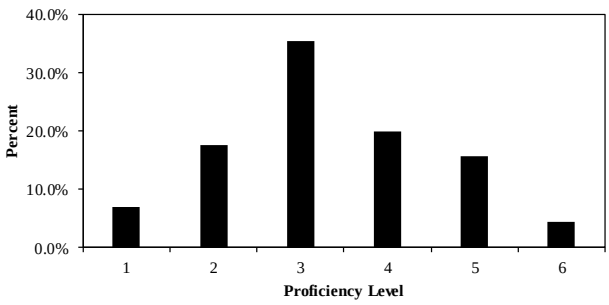


Table 8.2.7B
Proficiency Level Distribution:
Cphn 1 S401 Paper

Level	Count	Percent
1	4,993	7.0%
2	12,534	17.6%
3	25,256	35.4%
4	14,247	20.0%
5	11,174	15.7%
6	3,057	4.3%
Total	71,261	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.30	845.329	0.696
Reading	0.70	668.053	0.733
Comprehension		581.528	0.810

*Variances from students who had results in all four domains

Table 8.2.7E

Accuracy and Consistency of Classification Indices: Cphn (Grade 1) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.585	0.468	0.313

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.772	0.566
	2	0.569	0.439
	3	0.624	0.534
	4	0.454	0.356
	5	0.572	0.434
	6	0.757	0.496

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.958	0.012	0.031	0.939
	2/3	0.890	0.052	0.058	0.841
	3/4	0.848	0.091	0.062	0.795
	4/5	0.894	0.059	0.047	0.848
	5/6	0.970	0.024	0.006	0.957

8.2.8 Overall Composite 1

Figure 8.2.8A
Scale Scores: Over 1 S401 Paper

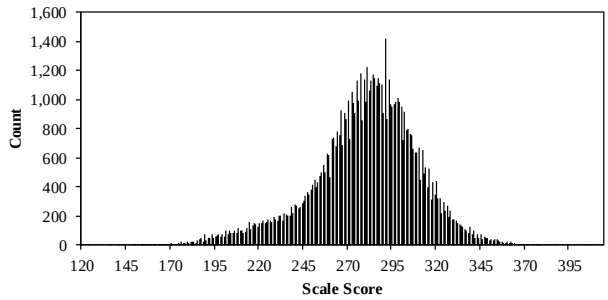


Table 8.2.8A
Scale Score Descriptive Statistics: Over 1 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	71,214	120	414	279.68	30.10

Figure 8.2.8B
Proficiency Level: Over 1 S401 Paper

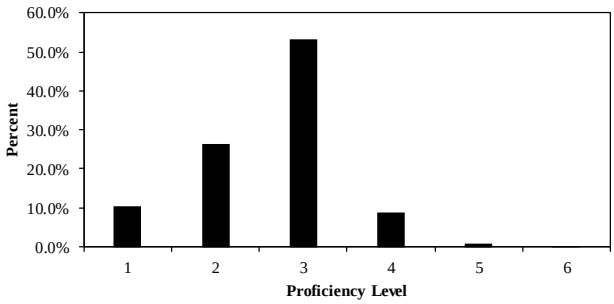


Table 8.2.8B
Proficiency Level Distribution:
Over 1 S401 Paper

Level	Count	Percent
1	7,420	10.4%
2	18,760	26.3%
3	37,995	53.4%
4	6,300	8.8%
5	695	1.0%
6	44	0.1%
Total	71,214	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.15	845.329	0.696
Reading	0.35	668.053	0.733
Speaking	0.15	4058.802	0.899
Writing	0.35	1673.385	0.912
Overall Composite		906.043	0.939

*Variances from students who had results in all four domains

Table 8.2.8E

Accuracy and Consistency of Classification Indices: Over (Grade 1) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.840	0.776	0.647

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.866	0.793
	2	0.810	0.730
	3	0.882	0.847
	4	0.654	0.534
	5	-	0.460
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.970	0.014	0.016	0.958
	2/3	0.929	0.035	0.036	0.900
	3/4	0.949	0.028	0.023	0.928
	4/5	0.990	0.010	0.000	0.990
	5/6	0.999	0.001	0.000	1.000

8.3 Grade: 2

8.3.1 Listening 2

Figure 8.3.1A
Scale Scores: List 2 S401 Paper

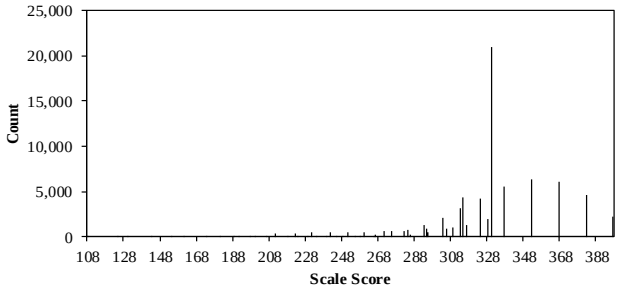


Table 8.3.1A
Scale Score Descriptive Statistics: List 2 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	73,483	108	397	329.45	33.05

Figure 8.3.1B
Proficiency Level: List 2 S401 Paper

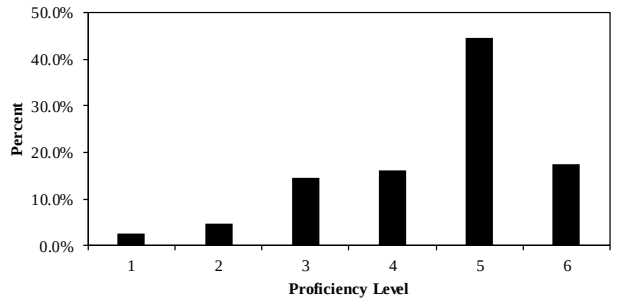
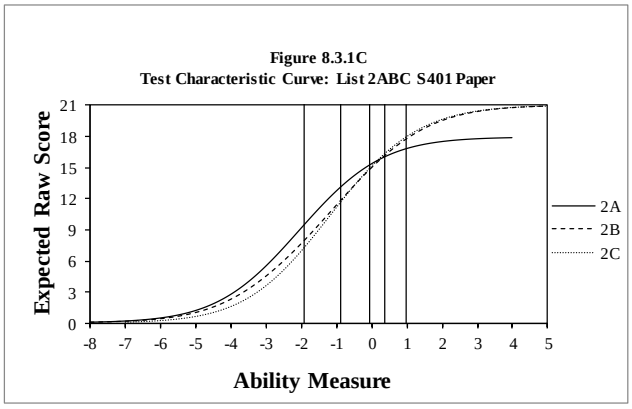


Table 8.3.1B
Proficiency Level Distribution:
List 2 S401 Paper

Level	Count	Percent
1	1,832	2.5%
2	3,539	4.8%
3	10,603	14.4%
4	11,892	16.2%
5	32,771	44.6%
6	12,846	17.5%
Total	73,483	100.0%

Table 8.3.1C
Conditional Standard Error of Measurement at Cut Scores: List 2 S401 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	2	245	19.16	19.75	18.79
2/3	2	283	21.04	19.16	18.41
3/4	2	314	25.92	20.66	19.91
4/5	2	330	n/a	21.79	21.79
5/6	2	354	n/a	n/a	25.55



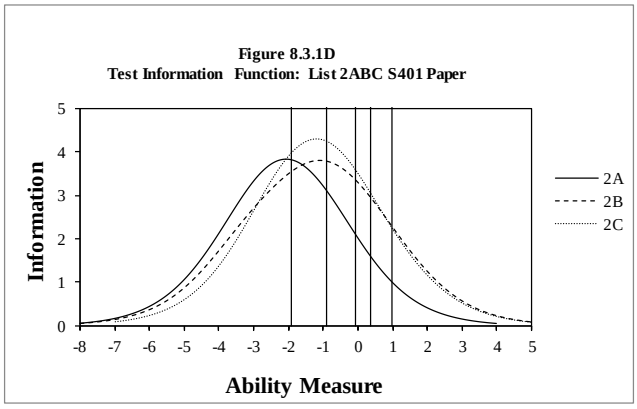


Table 8.3.1D

Weighted Reliability: List 2 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,265	0.826	0.642
B	26,342	0.647	
C	36,876	0.588	

Table 8.3.1E

Accuracy and Consistency of Classification Indices: List (Grade 2) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**	
	0.480	0.362	0.160	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.836	0.524	
	2	0.384	0.237	
	3	0.412	0.277	
	4	0.270	0.210	
	5	0.596	0.529	
	6	0.503	0.341	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.980	0.001	0.018	0.976
	2/3	0.953	0.014	0.033	0.922
	3/4	0.859	0.068	0.073	0.790
	4/5	0.774	0.127	0.099	0.703
	5/6	0.826	0.091	0.083	0.746

8.3.2 Reading 2

Figure 8.3.2A
Scale Scores: Read 2 S401 Paper

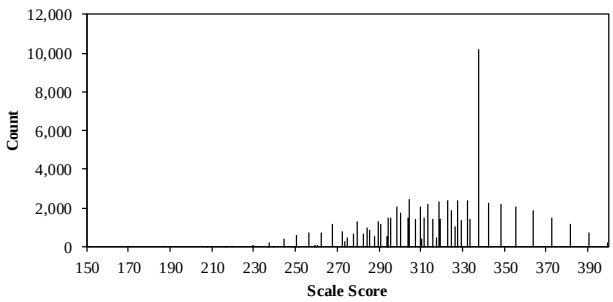


Table 8.3.2A
Scale Score Descriptive Statistics: Read 2 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	73,486	150	399	317.23	29.59

Figure 8.3.2B
Proficiency Level: Read 2 S401 Paper

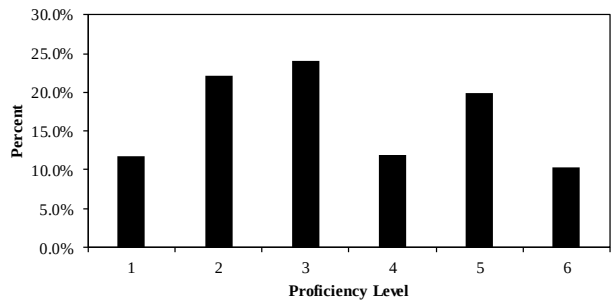
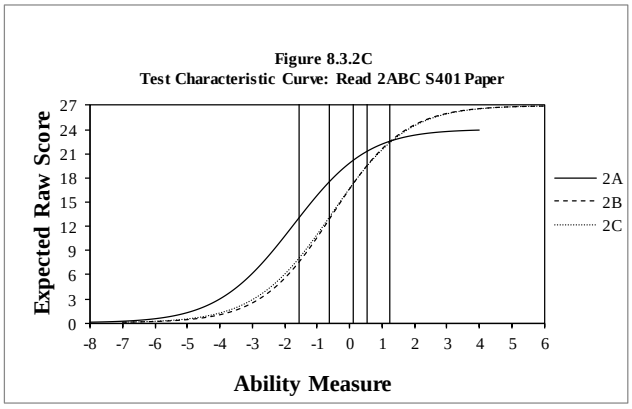


Table 8.3.2B
Proficiency Level Distribution:
Read 2 S401 Paper

Level	Count	Percent
1	8,631	11.7%
2	16,214	22.1%
3	17,663	24.0%
4	8,726	11.9%
5	14,653	19.9%
6	7,599	10.3%
Total	73,486	100.0%

Table 8.3.2C
Conditional Standard Error of Measurement at Cut Scores: Read 2 S401 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	2	283	11.44	11.44	11.96
2/3	2	307	12.74	10.47	10.92
3/4	2	326	15.08	10.92	11.18
4/5	2	337	n/a	11.70	11.67
5/6	2	355	n/a	n/a	13.54



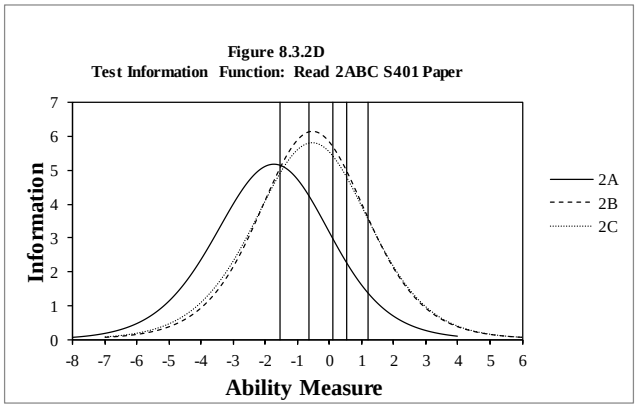


Table 8.3.2D

Weighted Reliability: Read 2 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,265	0.828	0.820
B	26,340	0.830	
C	36,881	0.812	

Table 8.3.2E

Accuracy and Consistency of Classification Indices: Read (Grade 2) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.548	0.439	0.315		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.741	0.571		
	2	0.568	0.457		
	3	0.509	0.405		
	4	0.292	0.221		
	5	0.584	0.462		
	6	0.752	0.568		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.930	0.026	0.044	0.900
	2/3	0.874	0.061	0.066	0.824
	3/4	0.871	0.070	0.058	0.818
	4/5	0.882	0.078	0.040	0.838
	5/6	0.939	0.040	0.021	0.911

8.3.3 Writing 2

Figure 8.3.3A
Scale Scores: Writ 2 S401 Paper

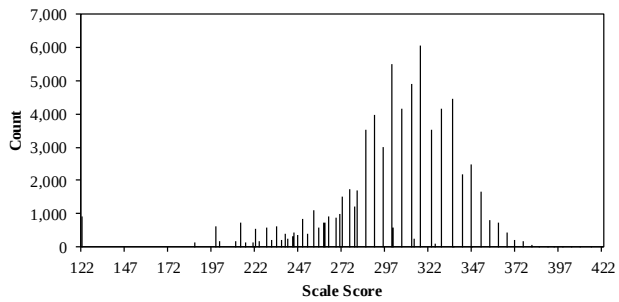


Table 8.3.3A
Scale Score Descriptive Statistics: Writ 2 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	73,500	122	422	297.93	40.19

Figure 8.3.3B
Proficiency Level: Writ 2 S401 Paper

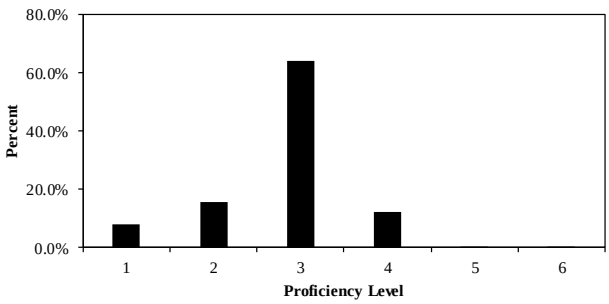
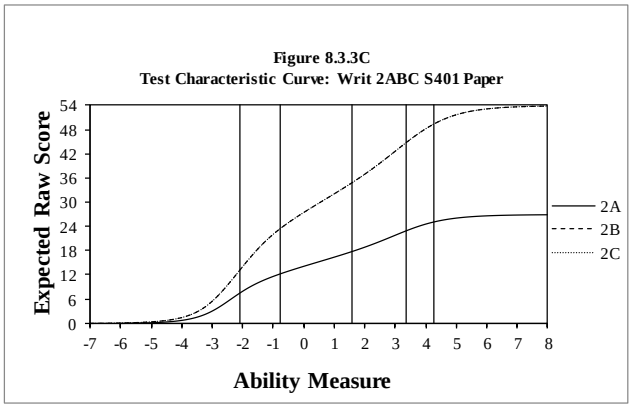


Table 8.3.3B
Proficiency Level Distribution:
Writ 2 S401 Paper

Level	Count	Percent
1	5,963	8.1%
2	11,571	15.7%
3	47,069	64.0%
4	8,842	12.0%
5	50	0.1%
6	5	0.0%
Total	73,500	100.0%

Table 8.3.3C
Conditional Standard Error of Measurement at Cut Scores: Writ 2 S401 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	2	242	11.81	8.86
2/3	2	279	16.27	11.01
3/4	2	341	17.18	12.30
4/5	2	388	15.31	11.01
5/6	2	411	18.26	12.35



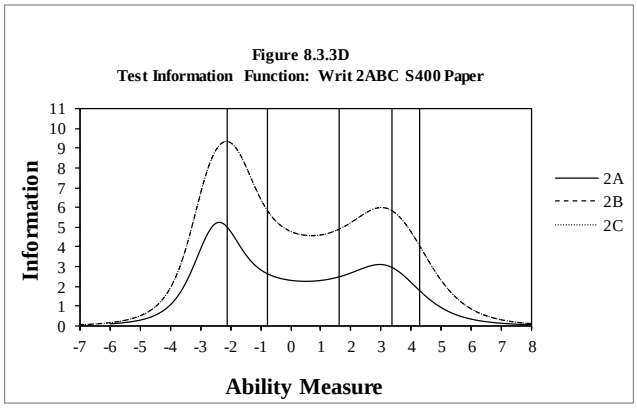


Table 8.3.3D

Weighted Reliability: Writ 2 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,268	0.910	0.930
B/C	63,232	0.934	

Table 8.3.3E

Accuracy and Consistency of Classification Indices: Writ (Grade 2) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.810	0.759	0.528		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.865	0.791		
	2	0.806	0.710		
	3	0.805	0.803		
	4	-	0.308		
	5	-	-		
	6	-	1.000		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.979	0.011	0.010	0.970
	2/3	0.952	0.018	0.029	0.933
	3/4	0.879	0.121	0.000	0.857
	4/5	0.999	0.001	0.000	0.999
	5/6	1.000	0.000	0.000	1.000

8.3.4 Speaking 2

Figure 8.3.4A
Scale Scores: Spek 2 S401 Paper

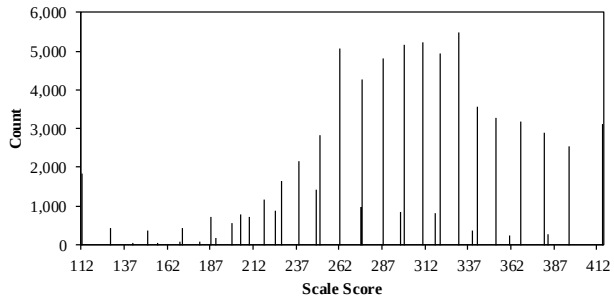


Table 8.3.4A
Scale Score Descriptive Statistics: Spek 2 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	73,488	112	415	298.21	64.46

Figure 8.3.4B
Proficiency Level: Spek 2 S401 Paper

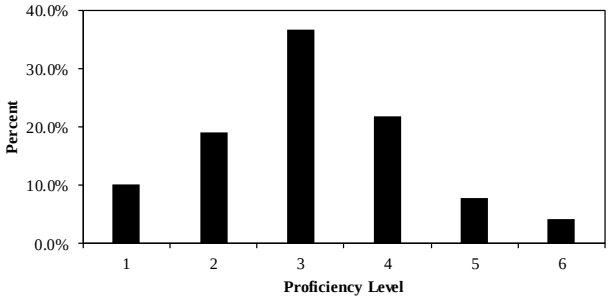
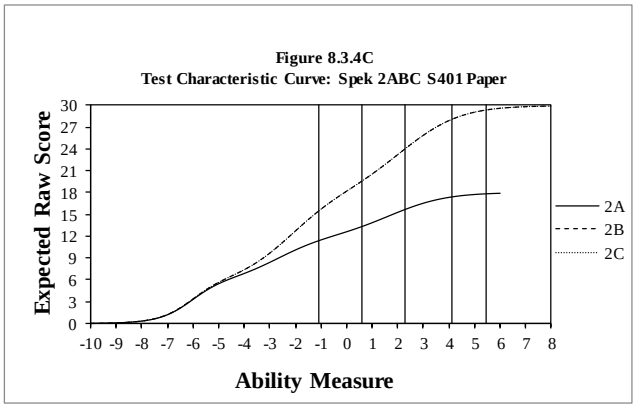


Table 8.3.4B
Proficiency Level Distribution:
Spek 2 S401 Paper

Level	Count	Percent
1	7,487	10.2%
2	13,993	19.0%
3	27,028	36.8%
4	16,129	21.9%
5	5,718	7.8%
6	3,133	4.3%
Total	73,488	100.0%

Table 8.3.4C
Conditional Standard Error of Measurement at Cut Scores: Spek 2 S401 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	2	220	24.86	16.67
2/3	2	273	26.91	19.30
3/4	2	322	24.57	17.55
4/5	2	374	35.39	21.64
5/6	2	415	62.30	34.80



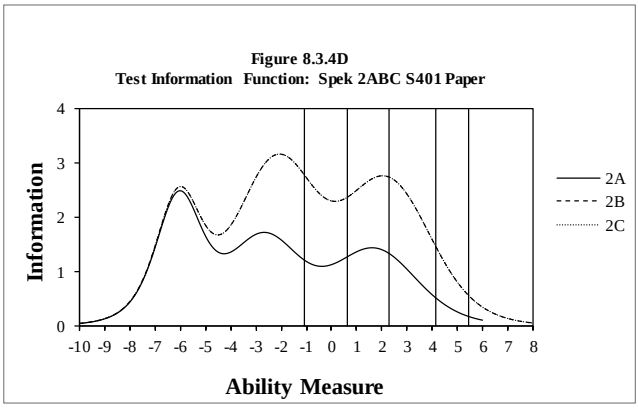


Table 8.3.4D

Weighted Reliability: Spek 2 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,265	0.921	0.913
B/C	63,223	0.912	

Table 8.3.4E

Accuracy and Consistency of Classification Indices: Spek (Grade 2) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.673	0.569	0.441		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.814	0.701		
	2	0.655	0.534		
	3	0.777	0.689		
	4	0.603	0.492		
	5	0.418	0.371		
	6	-	0.731		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.962	0.019	0.019	0.943
	2/3	0.909	0.050	0.041	0.872
	3/4	0.901	0.031	0.068	0.861
	4/5	0.938	0.044	0.018	0.909
	5/6	0.957	0.043	0.000	0.958

8.3.5 Oral Language Composite 2

Figure 8.3.5A
Scale Scores: Oral 2 S401 Paper

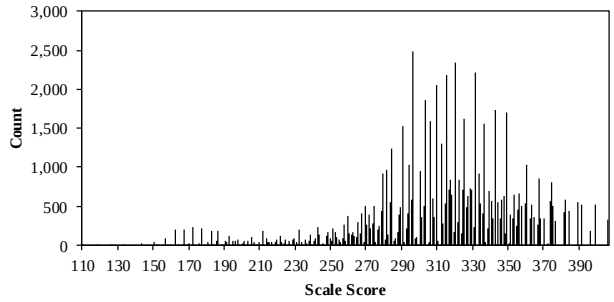


Table 8.3.5A
Scale Score Descriptive Statistics: Oral 2 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	73,461	110	406	314.08	43.50

Figure 8.3.5B
Proficiency Level: Oral 2 S401 Paper

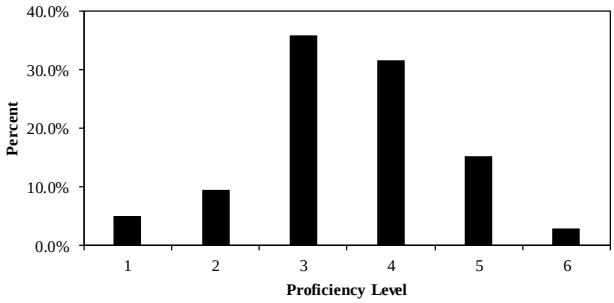


Table 8.3.5B
Proficiency Level Distribution:
Oral 2 S401 Paper

Level	Count	Percent
1	3,710	5.1%
2	7,031	9.6%
3	26,299	35.8%
4	23,184	31.6%
5	11,140	15.2%
6	2,097	2.9%
Total	73,461	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.50	1092.531	0.642
Speaking	0.50	4155.545	0.913
Oral		1892.177	0.901

*Variances from students who had results in all four domains

Table 8.3.5E

Accuracy and Consistency of Classification Indices: Oral (Grade 2) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.696	0.589	0.448

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.869	0.765
	2	0.595	0.458
	3	0.796	0.706
	4	0.665	0.560
	5	0.597	0.502
	6	-	0.305

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.982	0.006	0.013	0.975
	2/3	0.944	0.034	0.023	0.917
	3/4	0.887	0.043	0.070	0.844
	4/5	0.910	0.046	0.044	0.870
	5/6	0.971	0.029	0.000	0.970

8.3.6 Literacy Composite 2

Figure 8.3.6A
Scale Scores: Litr 2 S401 Paper

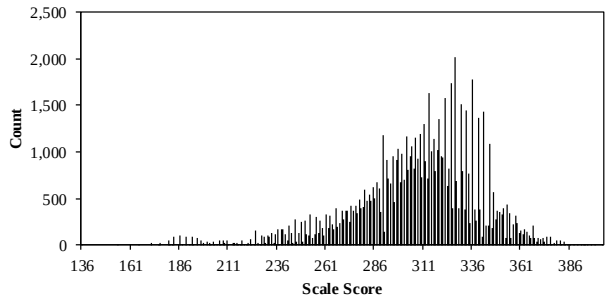


Table 8.3.6A
Scale Score Descriptive Statistics: Litr 2 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	73,475	136	403	307.84	31.65

Figure 8.3.6B
Proficiency Level: Litr 2 S401 Paper

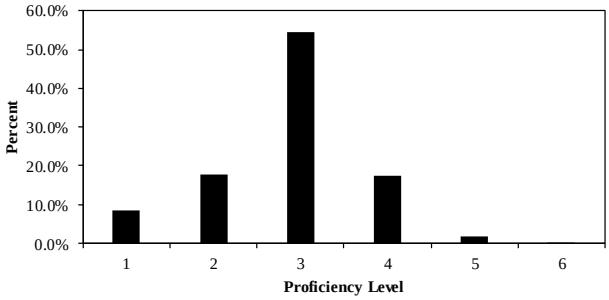


Table 8.3.6B
Proficiency Level Distribution: Litr 2 S401 Paper

Level	Count	Percent
1	6,228	8.5%
2	13,061	17.8%
3	40,017	54.5%
4	12,732	17.3%
5	1,365	1.9%
6	72	0.1%
Total	73,475	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 S401 Paper

Component	Weight	Variance	Reliability
Reading	0.50	875.493	0.820
Writing	0.50	1615.100	0.930
Literacy		1001.836	0.933

*Variances from students who had results in all four domains

Table 8.3.6E

Accuracy and Consistency of Classification Indices: Litr (Grade 2) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.811	0.740	0.596

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.855	0.781
	2	0.758	0.655
	3	0.872	0.826
	4	0.678	0.589
	5	-	0.282
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.974	0.012	0.014	0.964
	2/3	0.938	0.030	0.032	0.913
	3/4	0.916	0.038	0.046	0.881
	4/5	0.980	0.020	0.000	0.981
	5/6	0.999	0.001	0.000	1.000

8.3.7 Comprehension Composite 2

Figure 8.3.7A
Scale Scores: Cphn 2 S401 Paper

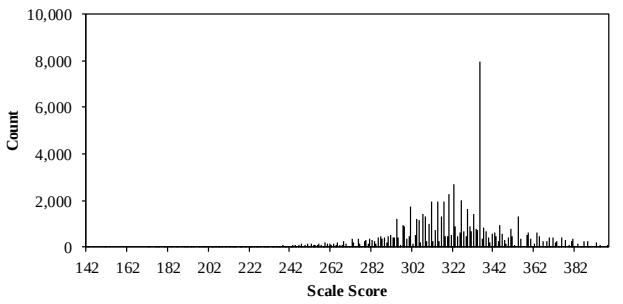


Table 8.3.7A
Scale Score Descriptive Statistics: Cphn 2 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	73,462	142	398	320.97	27.95

Figure 8.3.7B
Proficiency Level: Cphn 2 S401 Paper

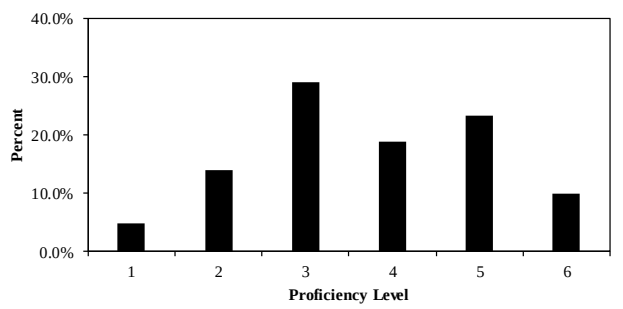


Table 8.3.7B
Proficiency Level Distribution:
Cphn 2 S401 Paper

Level	Count	Percent
1	3,565	4.9%
2	10,223	13.9%
3	21,303	29.0%
4	13,905	18.9%
5	17,189	23.4%
6	7,277	9.9%
Total	73,462	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.30	1092.531	0.642
Reading	0.70	875.493	0.820
Comprehension		781.491	0.856

*Variances from students who had results in all four domains

Table 8.3.7E

Accuracy and Consistency of Classification Indices: Cphn (Grade 2) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.607	0.491	0.362

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.799	0.609
	2	0.598	0.465
	3	0.639	0.533
	4	0.435	0.341
	5	0.637	0.518
	6	0.765	0.588

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.972	0.007	0.020	0.961
	2/3	0.917	0.041	0.042	0.880
	3/4	0.869	0.065	0.065	0.820
	4/5	0.882	0.067	0.051	0.835
	5/6	0.945	0.035	0.020	0.919

8.3.8 Overall Composite 2

Figure 8.3.8A
Scale Scores: Over 2 S401 Paper

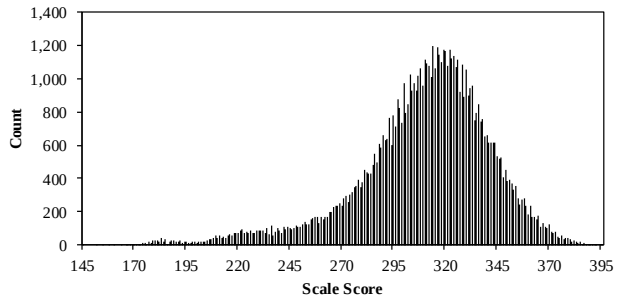


Table 8.3.8A
Scale Score Descriptive Statistics: Over 2 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	73,433	145	396	309.50	32.39

Figure 8.3.8B
Proficiency Level: Over 2 S401 Paper

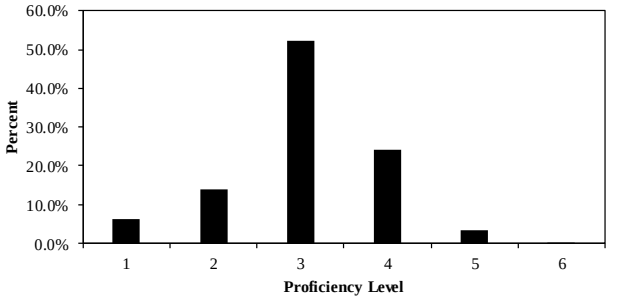


Table 8.3.8B
Proficiency Level Distribution:
Over 2 S401 Paper

Level	Count	Percent
1	4,514	6.1%
2	10,276	14.0%
3	38,338	52.2%
4	17,748	24.2%
5	2,476	3.4%
6	81	0.1%
Total	73,433	100.0%

Table 8.3.8C
n/a

Figure 8.3.8C
n/a

Figure 8.3.8D
n/a

Table 8.3.8D
Overall Composite Reliability: Over 2 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.15	1092.531	0.642
Reading	0.35	875.493	0.820
Speaking	0.15	4155.545	0.913
Writing	0.35	1615.100	0.930
Overall Composite		1049.191	0.952

*Variances from students who had results in all four domains

Table 8.3.8E

Accuracy and Consistency of Classification Indices: Over (Grade 2) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.828	0.771	0.648

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.880	0.832
	2	0.780	0.687
	3	0.902	0.861
	4	0.717	0.661
	5	-	0.359
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.984	0.007	0.009	0.979
	2/3	0.956	0.025	0.020	0.937
	3/4	0.920	0.030	0.050	0.889
	4/5	0.965	0.035	0.000	0.966
	5/6	0.999	0.001	0.000	1.000

8.4 Grade: 3

8.4.1 Listening 3

Figure 8.4.1A
Scale Scores: List 3 S401 Paper

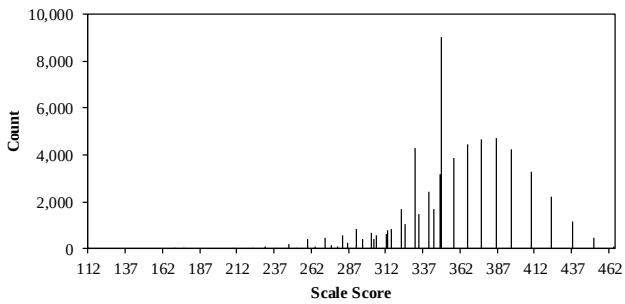


Table 8.4.1A
Scale Score Descriptive Statistics: List 3 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	61,360	112	465	357.28	37.45

Figure 8.4.1B
Proficiency Level: List 3 S401 Paper

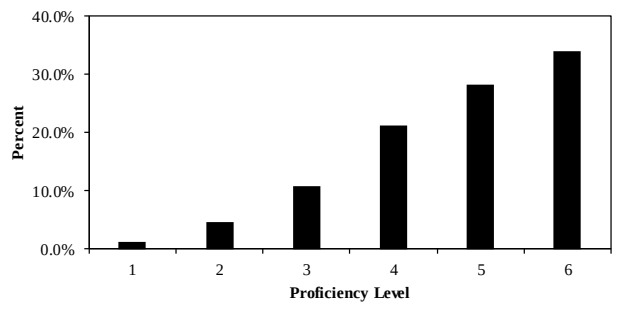
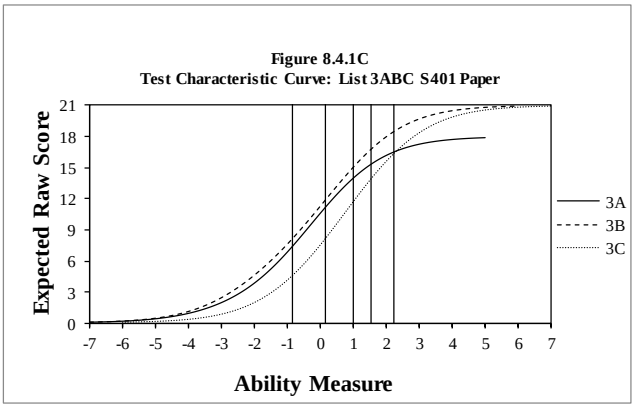


Table 8.4.1B
Proficiency Level Distribution:
List 3 S401 Paper

Level	Count	Percent
1	782	1.3%
2	2,823	4.6%
3	6,627	10.8%
4	12,979	21.2%
5	17,324	28.2%
6	20,825	33.9%
Total	61,360	100.0%

Table 8.4.1C
Conditional Standard Error of Measurement at Cut Scores: List 3 S401 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	3	262	21.42	21.04	25.55
2/3	3	300	19.16	19.16	19.91
3/4	3	331	19.91	18.79	18.25
4/5	3	349	n/a	19.54	18.03
5/6	3	374	n/a	n/a	18.79



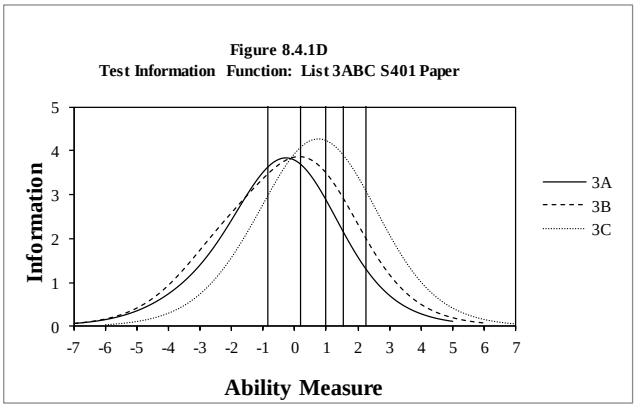


Table 8.4.1D

Weighted Reliability: List 3 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	6,832	0.749	0.641
B	15,488	0.644	
C	39,040	0.622	

Table 8.4.1E

Accuracy and Consistency of Classification Indices: List (Grade 3) S401 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)
	0.497	0.407	0.213

Conditional on Level	Level	Accuracy	Consistency
	1	-	0.199
	2	0.399	0.214
	3	0.313	0.218
	4	0.376	0.298
	5	0.422	0.354
	6	0.756	0.621

Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.987	0.000	0.013	0.984
	2/3	0.949	0.006	0.045	0.919
	3/4	0.862	0.074	0.064	0.796
	4/5	0.786	0.106	0.108	0.720
	5/6	0.817	0.109	0.074	0.746

8.4.2 Reading 3

Figure 8.4.2A
Scale Scores: Read 3 S401 Paper

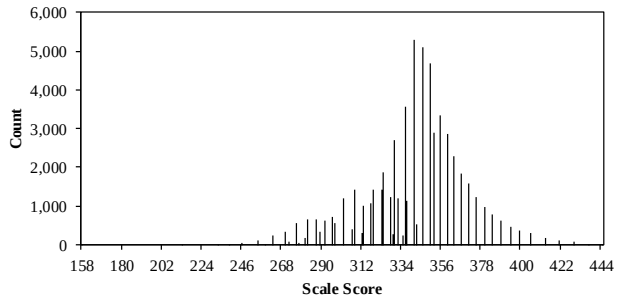


Table 8.4.2A
Scale Score Descriptive Statistics: Read 3 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	61,340	158	445	339.98	27.21

Figure 8.4.2B
Proficiency Level: Read 3 S401 Paper

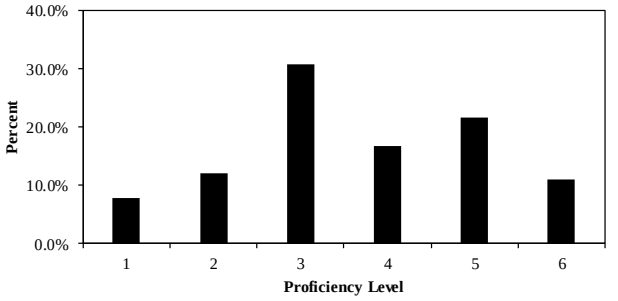
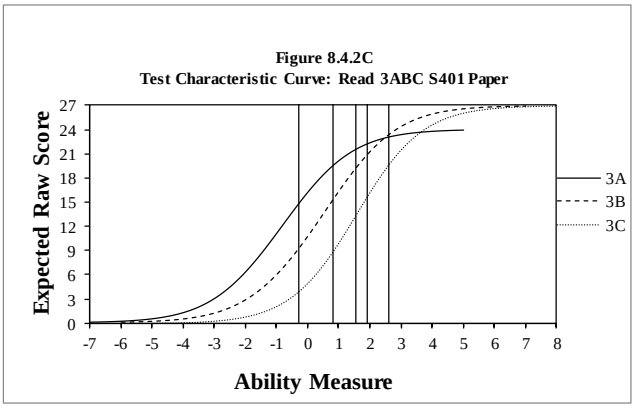


Table 8.4.2B
Proficiency Level Distribution:
Read 3 S401 Paper

Level	Count	Percent
1	4,722	7.7%
2	7,398	12.1%
3	18,930	30.9%
4	10,304	16.8%
5	13,218	21.5%
6	6,768	11.0%
Total	61,340	100.0%

Table 8.4.2C
Conditional Standard Error of Measurement at Cut Scores: Read 3 S401 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	3	297	11.41	13.00	19.24
2/3	3	323	11.93	11.10	13.26
3/4	3	342	13.78	10.92	11.18
4/5	3	352	n/a	11.18	10.66
5/6	3	370	n/a	n/a	10.40



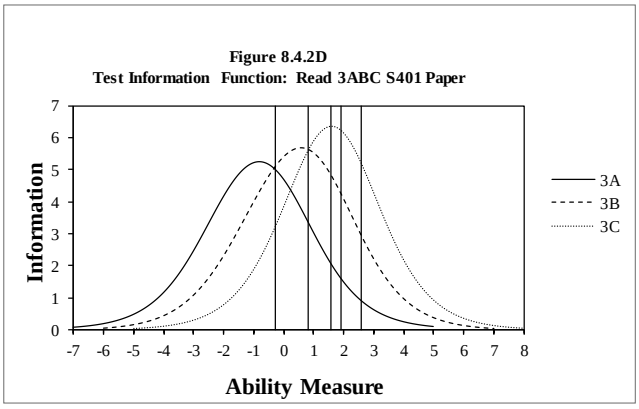


Table 8.4.2D

Weighted Reliability: Read 3 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	6,832	0.785	0.723
B	15,475	0.765	
C	39,033	0.695	

Table 8.4.2E

Accuracy and Consistency of Classification Indices: Read (Grade 3) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.466	0.363	0.211		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.824	0.599		
	2	0.409	0.288		
	3	0.545	0.442		
	4	0.283	0.228		
	5	0.440	0.348		
	6	0.625	0.381		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.954	0.008	0.037	0.936
	2/3	0.888	0.058	0.053	0.832
	3/4	0.801	0.089	0.111	0.737
	4/5	0.818	0.089	0.092	0.751
	5/6	0.906	0.068	0.025	0.860

8.4.3 Writing 3

Figure 8.4.3A
Scale Scores: Writ 3 S401 Paper

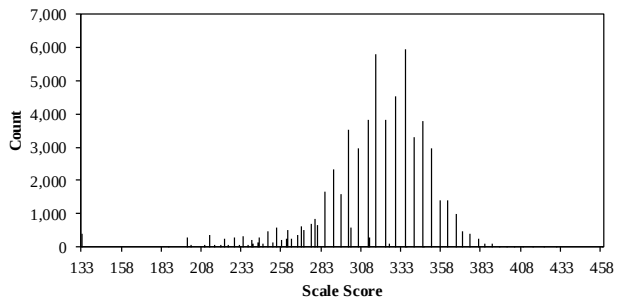


Table 8.4.3A
Scale Score Descriptive Statistics: Writ 3 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	61,358	133	459	314.29	36.53

Figure 8.4.3B
Proficiency Level: Writ 3 S401 Paper

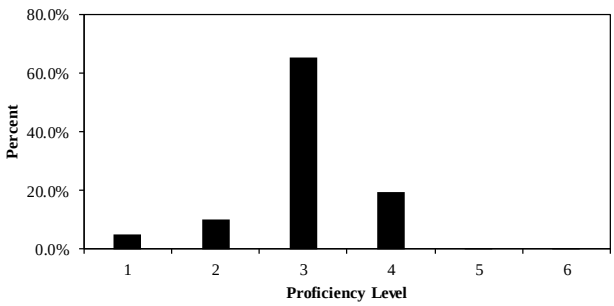
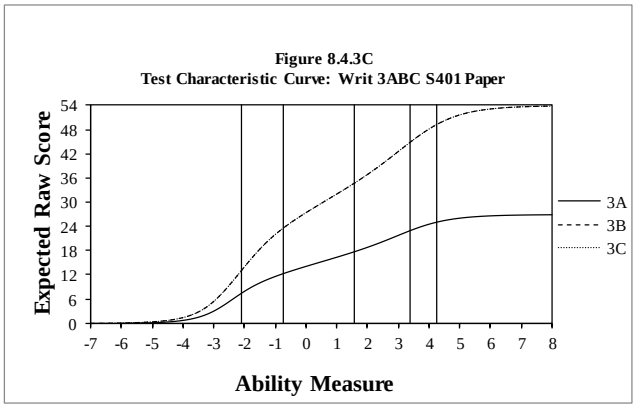


Table 8.4.3B
Proficiency Level Distribution:
Writ 3 S401 Paper

Level	Count	Percent
1	3,155	5.1%
2	6,087	9.9%
3	40,228	65.6%
4	11,799	19.2%
5	85	0.1%
6	4	0.0%
Total	61,358	100.0%

Table 8.4.3C
Conditional Standard Error of Measurement at Cut Scores: Writ 3 S401 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	3	247	12.08	8.86
2/3	3	283	16.65	11.28
3/4	3	346	17.18	12.14
4/5	3	394	15.65	11.14
5/6	3	418	20.41	13.43



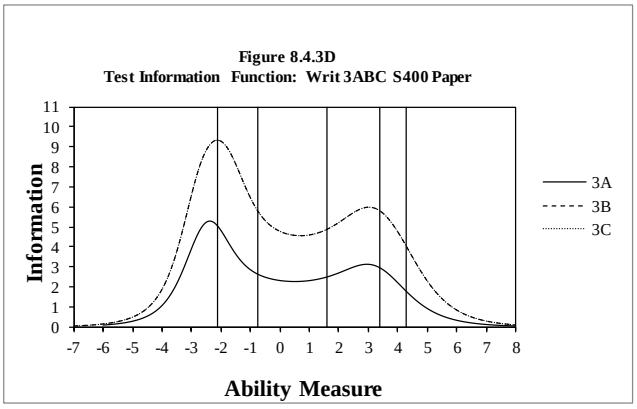


Table 8.4.3D

Weighted Reliability: Writ 3 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	6,831	0.909	0.917
B/C	54,527	0.918	

Table 8.4.3E

Accuracy and Consistency of Classification Indices: Writ (Grade 3) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.758	0.713	0.434		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.877	0.801		
	2	0.767	0.656		
	3	0.750	0.783		
	4	-	0.424		
	5	-	-		
	6	-	1.000		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.987	0.006	0.007	0.981
	2/3	0.965	0.015	0.020	0.951
	3/4	0.806	0.194	0.000	0.783
	4/5	0.999	0.001	0.000	0.999
	5/6	1.000	0.000	0.000	1.000

8.4.4 Speaking 3

Figure 8.4.4A
Scale Scores: Spek 3 S401 Paper

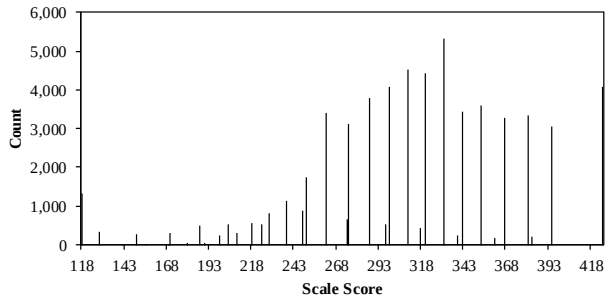


Table 8.4.4A
Scale Score Descriptive Statistics: Spek 3 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	61,329	118	425	312.71	64.89

Figure 8.4.4B
Proficiency Level: Spek 3 S401 Paper

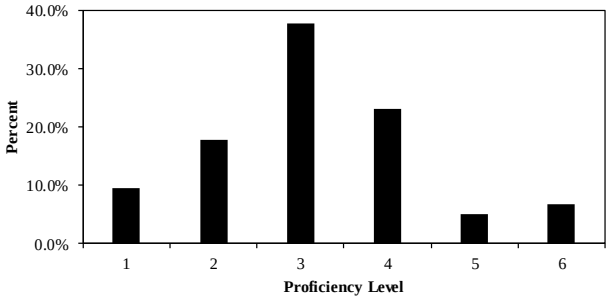
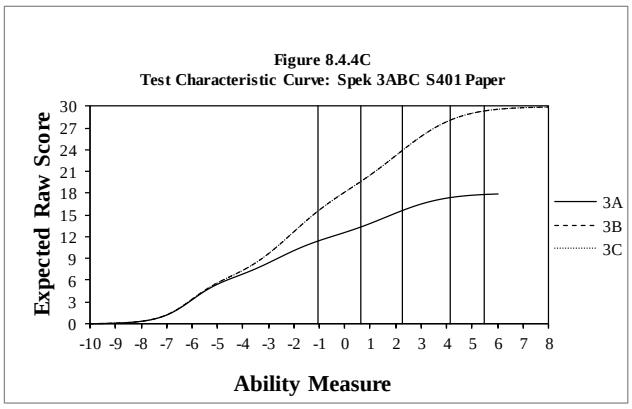


Table 8.4.4B
Proficiency Level Distribution:
Spek 3 S401 Paper

Level	Count	Percent
1	5,882	9.6%
2	10,944	17.8%
3	23,129	37.7%
4	14,235	23.2%
5	3,054	5.0%
6	4,085	6.7%
Total	61,329	100.0%

Table 8.4.4C
Conditional Standard Error of Measurement at Cut Scores: Spek 3 S401 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	3	234	26.62	17.55
2/3	3	283	26.03	19.01
3/4	3	332	25.45	17.55
4/5	3	386	40.95	24.28
5/6	3	425	72.83	40.07



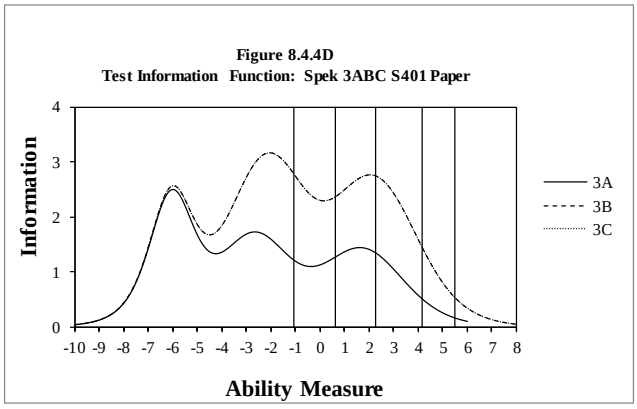


Table 8.4.4D

Weighted Reliability: Spek 3 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	6,828	0.921	0.911
B/C	54,501	0.910	

Table 8.4.4E

Accuracy and Consistency of Classification Indices: Spek (Grade 3) S401 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)	
	0.653	0.547	0.412	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.791	0.669	
	2	0.624	0.498	
	3	0.774	0.688	
	4	0.631	0.509	
	5	0.253	0.230	
	6	-	0.774	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.961	0.020	0.018	0.942
	2/3	0.905	0.052	0.043	0.867
	3/4	0.895	0.031	0.074	0.855
	4/5	0.947	0.033	0.020	0.914
	5/6	0.933	0.067	0.000	0.934

8.4.5 Oral Language Composite 3

Figure 8.4.5A
Scale Scores: Oral 3 S401 Paper

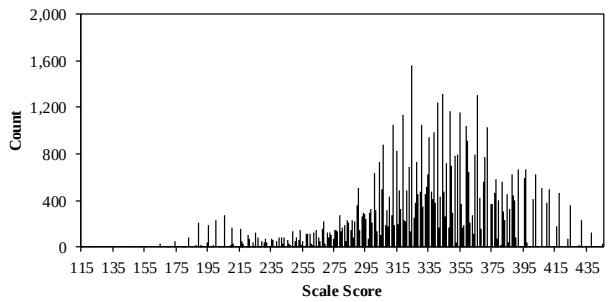


Table 8.4.5A
Scale Score Descriptive Statistics: Oral 3 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	61,323	115	445	335.25	45.28

Figure 8.4.5B
Proficiency Level: Oral 3 S401 Paper

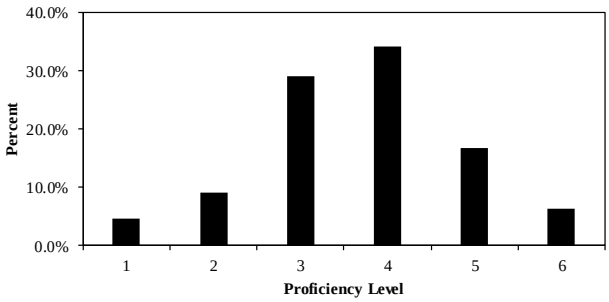


Table 8.4.5B
Proficiency Level Distribution:
Oral 3 S401 Paper

Level	Count	Percent
1	2,817	4.6%
2	5,604	9.1%
3	17,897	29.2%
4	20,907	34.1%
5	10,212	16.7%
6	3,886	6.3%
Total	61,323	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.50	1402.148	0.641
Speaking	0.50	4210.077	0.911
Oral		2050.509	0.893

*Variances from students who had results in all four domains

Table 8.4.5E

Accuracy and Consistency of Classification Indices: Oral (Grade 3) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.667	0.563	0.426

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.862	0.740
	2	0.601	0.462
	3	0.735	0.630
	4	0.701	0.602
	5	0.533	0.457
	6	0.759	0.482

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.982	0.005	0.013	0.974
	2/3	0.948	0.028	0.024	0.922
	3/4	0.889	0.049	0.062	0.846
	4/5	0.908	0.041	0.051	0.866
	5/6	0.940	0.059	0.001	0.935

8.4.6 Literacy Composite 3

Figure 8.4.6A
Scale Scores: Litr 3 S401 Paper

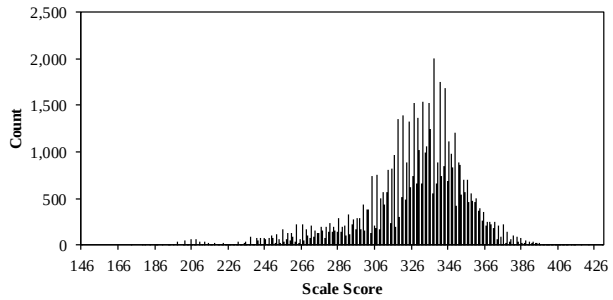


Table 8.4.6A
Scale Score Descriptive Statistics: Litr 3 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	61,328	146	430	327.39	28.67

Figure 8.4.6B
Proficiency Level: Litr 3 S401 Paper

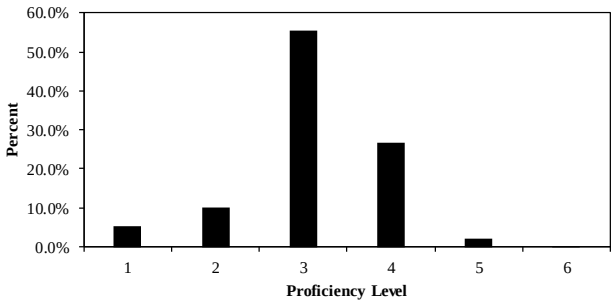


Table 8.4.6B
Proficiency Level Distribution: Litr 3 S401 Paper

Level	Count	Percent
1	3,279	5.3%
2	6,234	10.2%
3	33,930	55.3%
4	16,419	26.8%
5	1,313	2.1%
6	153	0.2%
Total	61,328	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 S401 Paper

Component	Weight	Variance	Reliability
Reading	0.50	739.959	0.723
Writing	0.50	1333.541	0.917
Literacy		821.927	0.904

*Variances from students who had results in all four domains

Table 8.4.6E

Accuracy and Consistency of Classification Indices: Litr (Grade 3) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.787	0.712	0.534

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.847	0.806
	2	0.688	0.560
	3	0.865	0.801
	4	0.685	0.611
	5	-	0.140
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.982	0.008	0.011	0.978
	2/3	0.952	0.025	0.022	0.934
	3/4	0.870	0.051	0.079	0.822
	4/5	0.976	0.024	0.000	0.976
	5/6	0.998	0.002	0.000	1.000

8.4.7 Comprehension Composite 3

Figure 8.4.7A
Scale Scores: Cphn 3 S401 Paper

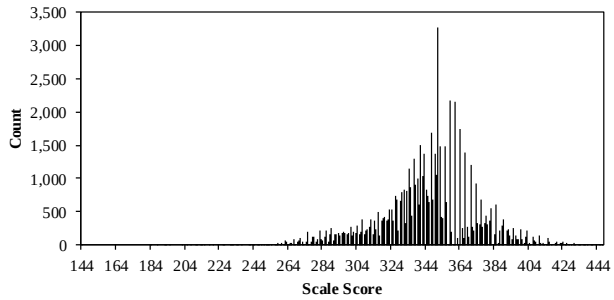


Table 8.4.7A
Scale Score Descriptive Statistics: Cphn 3 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	61,337	144	447	345.24	28.00

Figure 8.4.7B
Proficiency Level: Cphn 3 S401 Paper

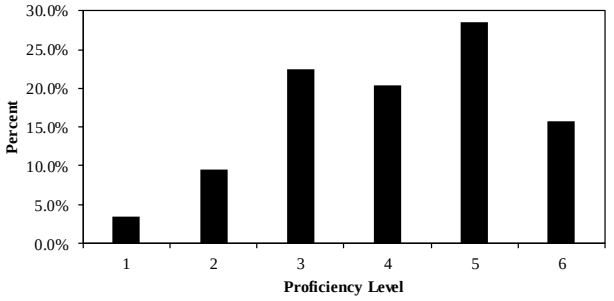


Table 8.4.7B
Proficiency Level Distribution:
Cphn 3 S401 Paper

Level	Count	Percent
1	2,144	3.5%
2	5,853	9.5%
3	13,726	22.4%
4	12,529	20.4%
5	17,452	28.5%
6	9,633	15.7%
Total	61,337	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.30	1402.148	0.641
Reading	0.70	739.959	0.723
Comprehension		783.831	0.814

*Variances from students who had results in all four domains

Table 8.4.7E

Accuracy and Consistency of Classification Indices: Cphn (Grade 3) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.562	0.450	0.308

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.831	0.620
	2	0.560	0.411
	3	0.555	0.437
	4	0.410	0.325
	5	0.572	0.472
	6	0.726	0.558

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.979	0.003	0.018	0.972
	2/3	0.937	0.026	0.038	0.905
	3/4	0.860	0.074	0.067	0.806
	4/5	0.841	0.080	0.079	0.785
	5/6	0.907	0.054	0.039	0.863

8.4.8 Overall Composite 3

Figure 8.4.8A
Scale Scores: Over 3 S401 Paper

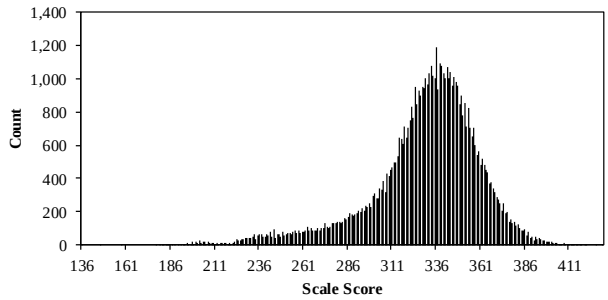


Table 8.4.8A
Scale Score Descriptive Statistics: Over 3 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
3	61,292	136	430	329.54	31.16

Figure 8.4.8B
Proficiency Level: Over 3 S401 Paper

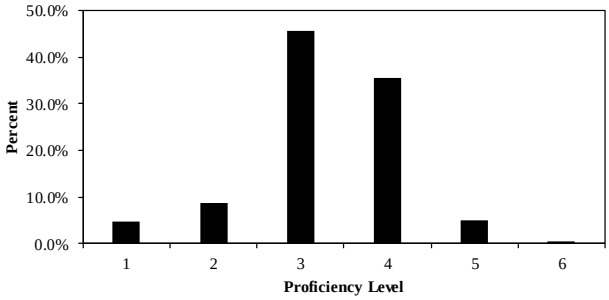


Table 8.4.8B
Proficiency Level Distribution:
Over 3 S401 Paper

Level	Count	Percent
1	2,938	4.8%
2	5,296	8.6%
3	28,054	45.8%
4	21,818	35.6%
5	2,958	4.8%
6	228	0.4%
Total	61,292	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.15	1402.148	0.641
Reading	0.35	739.959	0.723
Speaking	0.15	4210.077	0.911
Writing	0.35	1333.541	0.917
Overall Composite		971.188	0.940

*Variances from students who had results in all four domains

Table 8.4.8E

Accuracy and Consistency of Classification Indices: Over (Grade 3) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.814	0.756	0.625

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.848	0.861
	2	0.719	0.606
	3	0.891	0.839
	4	0.755	0.716
	5	-	0.311
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.985	0.007	0.008	0.984
	2/3	0.963	0.023	0.014	0.951
	3/4	0.908	0.036	0.056	0.876
	4/5	0.948	0.052	0.000	0.945
	5/6	0.996	0.004	0.000	1.000

8.5 Grades: 4–5

8.5.1 Listening 4–5

Figure 8.5.1A
Scale Scores: List 4-5 S401 Paper

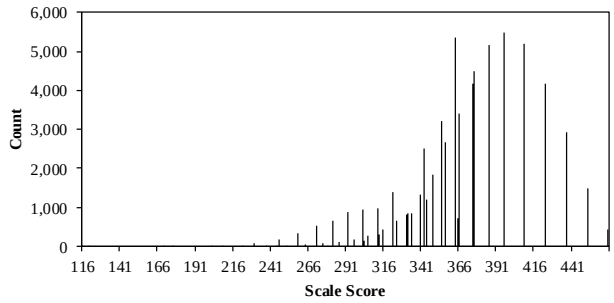


Table 8.5.1A
Scale Score Descriptive Statistics: List 4-5 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	34,323	116	465	365.03	39.21
5	32,341	120	465	376.72	40.54
Total	66,664	116	465	370.70	40.29

Figure 8.5.1B
Proficiency Level: List 4-5 S401 Paper

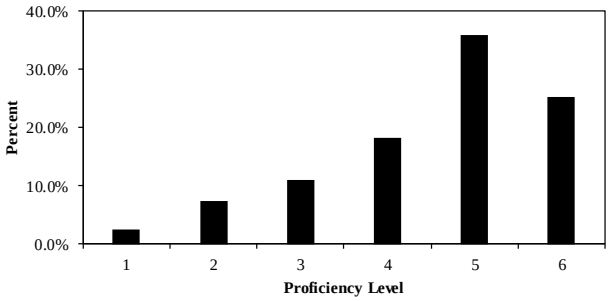


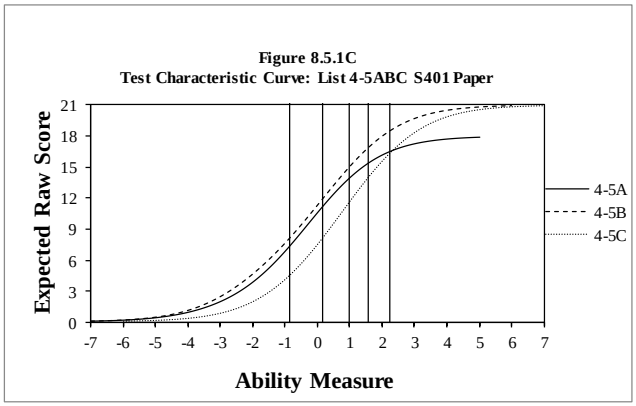
Table 8.5.1B
Proficiency Level Distribution: List 4-5 S401 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	770	2.2%	865	2.7%	1,635	2.5%
2	2,404	7.0%	2,444	7.6%	4,848	7.3%
3	3,975	11.6%	3,316	10.3%	7,291	10.9%
4	6,437	18.8%	5,663	17.5%	12,100	18.2%
5	12,317	35.9%	11,585	35.8%	23,902	35.9%
6	8,420	24.5%	8,468	26.2%	16,888	25.3%
Total	34,323	100.0%	32,341	100.0%	66,664	100.0%

Table 8.5.1C

Conditional Standard Error of Measurement at Cut Scores: List 4-5 S401 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	4	275	20.29	20.22	23.29
	5	285	19.54	19.54	21.79
2/3	4	313	19.16	18.79	19.05
	5	323	19.54	18.79	18.41
3/4	4	343	20.75	19.54	18.03
	5	354	22.17	19.97	18.03
4/5	4	363	n/a	20.66	18.41
	5	375	n/a	22.17	19.16
5/6	4	388	n/a	n/a	19.91
	5	401	n/a	n/a	21.42



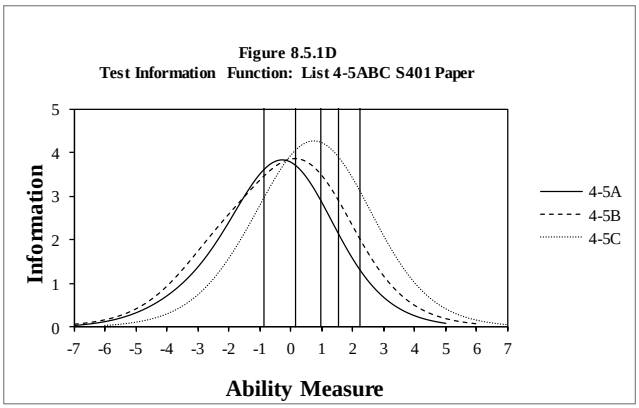


Table 8.5.1D

Weighted Reliability: List 4-5 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,541	0.771	0.667
B	15,653	0.695	
C	40,470	0.629	

Table 8.5.1Ei

Accuracy and Consistency of Classification Indices: List (Grade 4) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**	
	0.479	0.381	0.202	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.745	0.338	
	2	0.450	0.287	
	3	0.317	0.221	
	4	0.322	0.256	
	5	0.535	0.456	
	6	0.662	0.513	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.979	0.001	0.020	0.973
	2/3	0.934	0.014	0.052	0.900
	3/4	0.861	0.075	0.064	0.793
	4/5	0.781	0.131	0.088	0.716
	5/6	0.833	0.085	0.082	0.763

Table 8.5.1Eii

Accuracy and Consistency of Classification Indices: List (Grade 5) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**	
	0.479	0.382	0.201	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.735	0.381	
	2	0.462	0.300	
	3	0.290	0.200	
	4	0.310	0.242	
	5	0.527	0.451	
	6	0.670	0.521	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.976	0.002	0.022	0.969
	2/3	0.931	0.014	0.056	0.898
	3/4	0.870	0.068	0.062	0.802
	4/5	0.786	0.128	0.086	0.720
	5/6	0.822	0.096	0.081	0.751

8.5.2 Reading 4–5

Figure 8.5.2A
Scale Scores: Read 4-5 S401 Paper

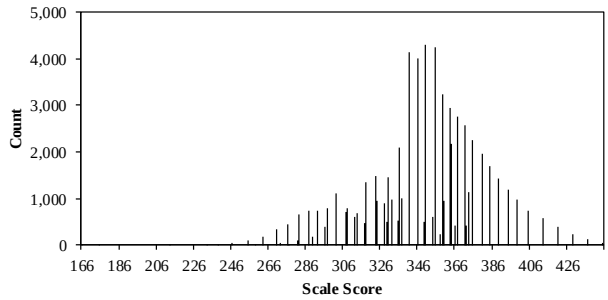


Table 8.5.2A

Scale Score Descriptive Statistics: Read 4-5 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	34,305	166	445	344.98	28.97
5	32,317	175	445	354.01	30.71
Total	66,622	166	445	349.36	30.16

Figure 8.5.2B
Proficiency Level: Read 4-5 S401 Paper

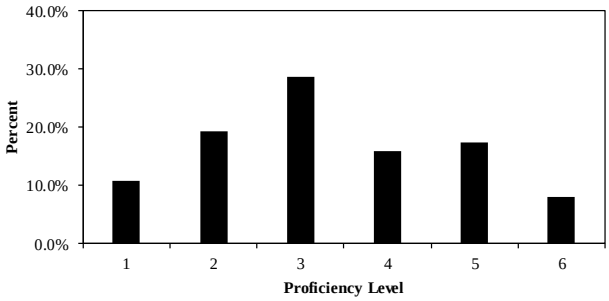


Table 8.5.2B

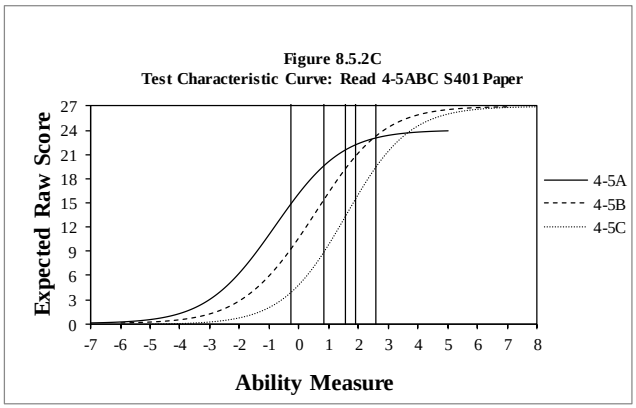
Proficiency Level Distribution: Read 4-5 S401 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,492	10.2%	3,645	11.3%	7,137	10.7%
2	6,343	18.5%	6,453	20.0%	12,796	19.2%
3	9,535	27.8%	9,559	29.6%	19,094	28.7%
4	6,498	18.9%	4,139	12.8%	10,637	16.0%
5	5,936	17.3%	5,655	17.5%	11,591	17.4%
6	2,501	7.3%	2,866	8.9%	5,367	8.1%
Total	34,305	100.0%	32,317	100.0%	66,622	100.0%

Table 8.5.2C

Conditional Standard Error of Measurement at Cut Scores: Read 4-5 S401 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	4	307	11.39	11.96	16.38
	5	316	11.70	11.44	14.56
2/3	4	335	12.86	10.92	11.70
	5	345	14.30	10.92	10.92
3/4	4	354	15.86	11.18	10.66
	5	364	17.94	11.96	10.40
4/5	4	364	n/a	11.96	10.40
	5	373	n/a	12.74	10.40
5/6	4	382	n/a	n/a	10.92
	5	391	n/a	n/a	11.44



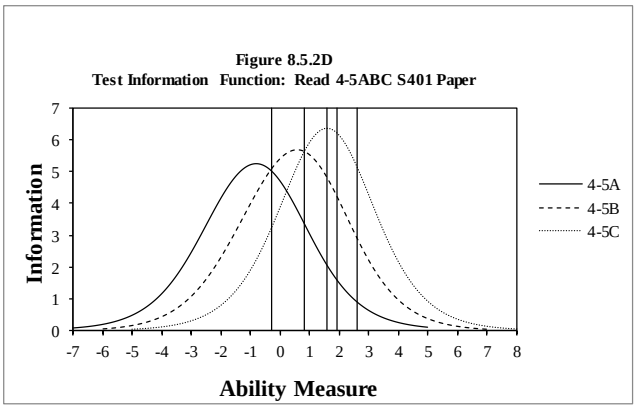


Table 8.5.2D

Weighted Reliability: Read 4-5 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,540	0.826	0.781
B	15,638	0.813	
C	40,444	0.756	

Table 8.5.2Ei

Accuracy and Consistency of Classification Indices: Read (Grade 4) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.501	0.393	0.252
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.808	0.631
	2	0.523	0.396
	3	0.499	0.409
	4	0.362	0.289
	5	0.448	0.342
	6	0.633	0.367

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.946	0.015	0.039	0.923
	2/3	0.874	0.066	0.060	0.818
	3/4	0.819	0.101	0.080	0.761
	4/5	0.850	0.075	0.075	0.792
	5/6	0.939	0.045	0.016	0.905

Table 8.5.2Eii

Accuracy and Consistency of Classification Indices: Read (Grade 5) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.503	0.397	0.257
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.793	0.611
	2	0.509	0.393
	3	0.526	0.433
	4	0.258	0.201
	5	0.454	0.344
	6	0.668	0.423

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.939	0.018	0.043	0.912
	2/3	0.860	0.076	0.064	0.803
	3/4	0.828	0.086	0.086	0.769
	4/5	0.855	0.078	0.067	0.797
	5/6	0.932	0.048	0.020	0.896

8.5.3 Writing 4–5

Figure 8.5.3A
Scale Scores: Writ 4-5 S401 Paper

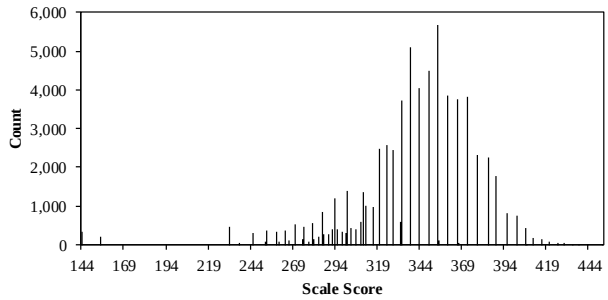


Table 8.5.3A
Scale Score Descriptive Statistics: Writ 4-5 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	34,322	144	443	333.20	37.64
5	32,336	155	452	343.56	37.69
Total	66,658	144	452	338.22	38.02

Figure 8.5.3B
Proficiency Level: Writ 4-5 S401 Paper

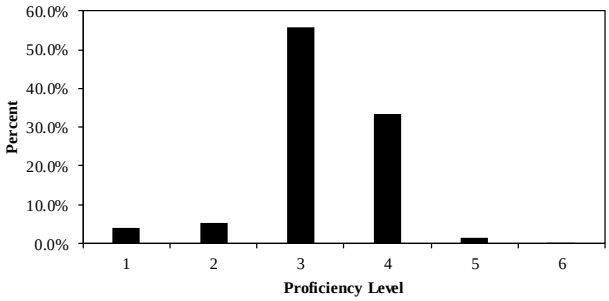


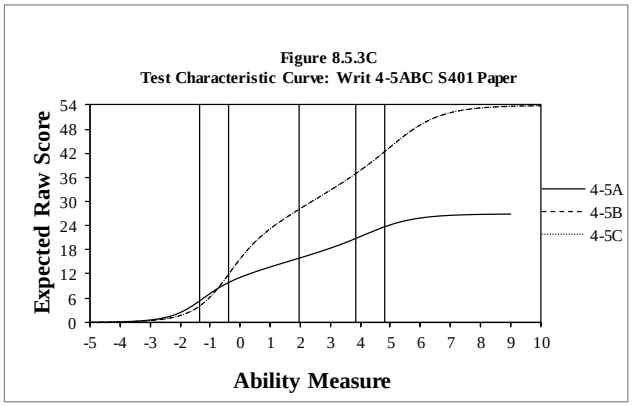
Table 8.5.3B
Proficiency Level Distribution: Writ 4-5 S401 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,521	4.4%	1,106	3.4%	2,627	3.9%
2	1,937	5.6%	1,667	5.2%	3,604	5.4%
3	19,923	58.0%	17,347	53.6%	37,270	55.9%
4	10,401	30.3%	11,827	36.6%	22,228	33.3%
5	507	1.5%	369	1.1%	876	1.3%
6	33	0.1%	20	0.1%	53	0.1%
Total	34,322	100.0%	32,336	100.0%	66,658	100.0%

Table 8.5.3C

Conditional Standard Error of Measurement at Cut Scores: Writ 4-5 S401 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	4	266	11.81	11.79
	5	267	11.55	11.55
2/3	4	288	13.43	8.59
	5	293	14.04	8.32
3/4	4	351	17.99	12.35
	5	356	17.72	12.35
4/5	4	401	15.57	11.98
	5	407	15.31	11.81
5/6	4	425	15.57	11.01
	5	433	16.65	10.74



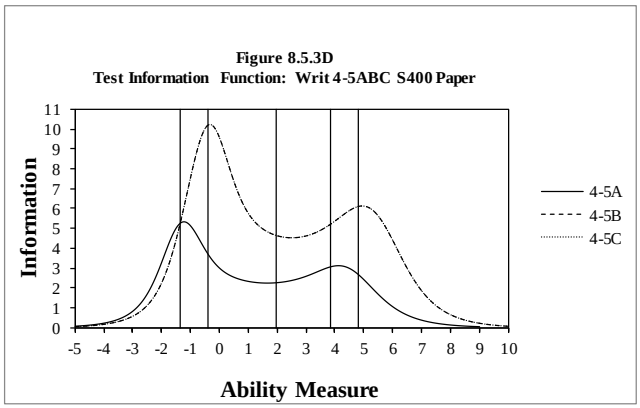


Table 8.5.3D

Weighted Reliability: Writ 4-5 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,542	0.879	0.894
B/C	56,116	0.897	

Table 8.5.3Ei

Accuracy and Consistency of Classification Indices: Writ (Grade 4) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.744	0.669	0.429
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.798	0.703
	2	0.582	0.435
	3	0.867	0.768
	4	0.615	0.559
	5	-	-
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.983	0.009	0.007	0.976
	2/3	0.965	0.016	0.019	0.951
	3/4	0.810	0.046	0.144	0.749
	4/5	0.984	0.016	0.000	0.985
	5/6	0.999	0.001	0.000	1.000

Table 8.5.3Eii

Accuracy and Consistency of Classification Indices: Writ (Grade 5) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.802	0.732	0.542
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.813	0.706
	2	0.573	0.422
	3	0.883	0.809
	4	0.742	0.691
	5	-	0.045
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.987	0.006	0.007	0.981
	2/3	0.967	0.017	0.016	0.953
	3/4	0.858	0.040	0.102	0.807
	4/5	0.988	0.012	0.000	0.988
	5/6	0.999	0.001	0.000	1.000

8.5.4 Speaking 4–5

Figure 8.5.4A
Scale Scores: Spek 4-5 S401 Paper

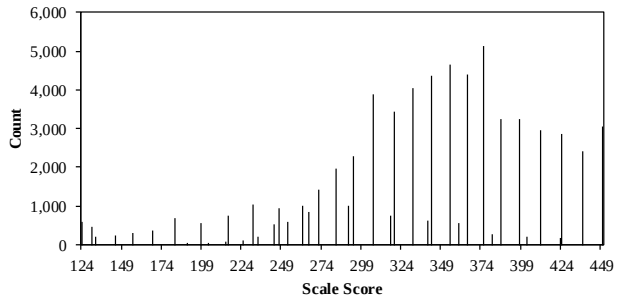


Table 8.5.4A
Scale Score Descriptive Statistics: Spek 4-5 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	34,328	124	450	334.34	70.01
5	32,340	130	450	343.69	70.04
Total	66,668	124	450	338.88	70.18

Figure 8.5.4B
Proficiency Level: Spek 4-5 S401 Paper

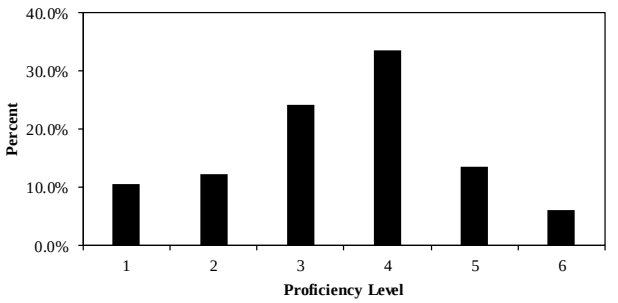


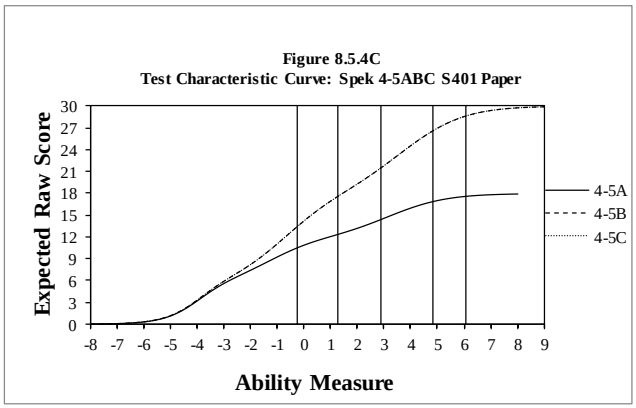
Table 8.5.4B
Proficiency Level Distribution: Spek 4-5 S401 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,400	9.9%	3,576	11.1%	6,976	10.5%
2	4,414	12.9%	3,696	11.4%	8,110	12.2%
3	8,279	24.1%	7,867	24.3%	16,146	24.2%
4	11,468	33.4%	10,879	33.6%	22,347	33.5%
5	4,341	12.6%	4,650	14.4%	8,991	13.5%
6	2,426	7.1%	1,672	5.2%	4,098	6.1%
Total	34,328	100.0%	32,340	100.0%	66,668	100.0%

Table 8.5.4C

Conditional Standard Error of Measurement at Cut Scores: Spek 4-5 S401 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	4	246	22.52	16.38
	5	258	23.98	16.67
2/3	4	293	28.08	18.72
	5	302	28.08	19.01
3/4	4	342	24.57	18.13
	5	350	24.28	17.84
4/5	4	397	29.25	19.01
	5	407	31.88	20.18
5/6	4	435	45.04	25.74
	5	443	50.60	28.08



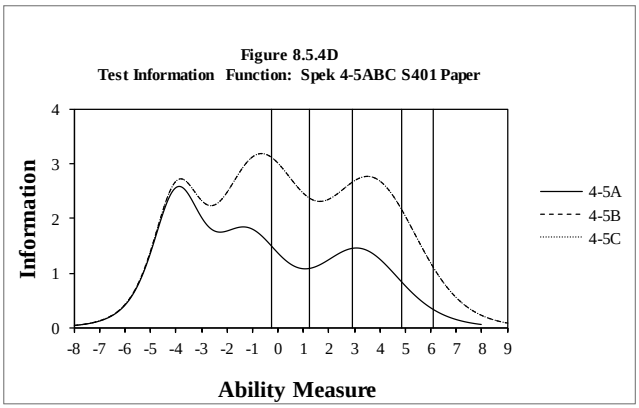


Table 8.5.4D

Weighted Reliability: Spek 4-5 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	10,541	0.897	0.906
B/C	56,127	0.908	

Table 8.5.4Ei

Accuracy and Consistency of Classification Indices: Spek (Grade 4) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.627	0.523	0.391
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.820	0.709
	2	0.589	0.460
	3	0.658	0.546
	4	0.694	0.578
	5	0.416	0.358
	6	-	0.350

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.964	0.018	0.019	0.946
	2/3	0.928	0.036	0.036	0.899
	3/4	0.899	0.042	0.059	0.860
	4/5	0.898	0.046	0.056	0.848
	5/6	0.929	0.071	0.000	0.926

Table 8.5.4Eii

Accuracy and Consistency of Classification Indices: Spek (Grade 5) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.640	0.536	0.402
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.835	0.729
	2	0.540	0.413
	3	0.676	0.563
	4	0.676	0.569
	5	0.484	0.413
	6	-	0.243

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.962	0.018	0.020	0.944
	2/3	0.929	0.037	0.034	0.900
	3/4	0.901	0.038	0.060	0.864
	4/5	0.889	0.057	0.053	0.841
	5/6	0.948	0.052	0.000	0.945

8.5.5 Oral Language Composite 4–5

Figure 8.5.5A
Scale Scores: Oral 4-5 S401 Paper

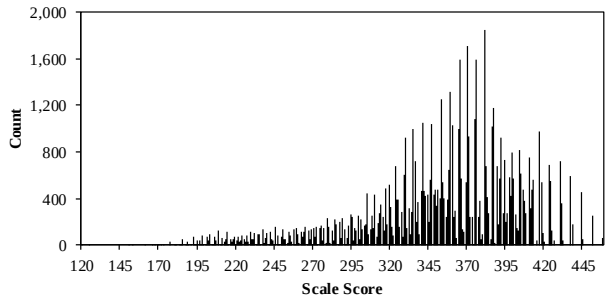


Table 8.5.5A
Scale Score Descriptive Statistics: Oral 4-5 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	34,319	120	458	349.94	49.33
5	32,339	125	458	360.46	49.94
Total	66,658	120	458	355.05	49.90

Figure 8.5.5B
Proficiency Level: Oral 4-5 S401 Paper

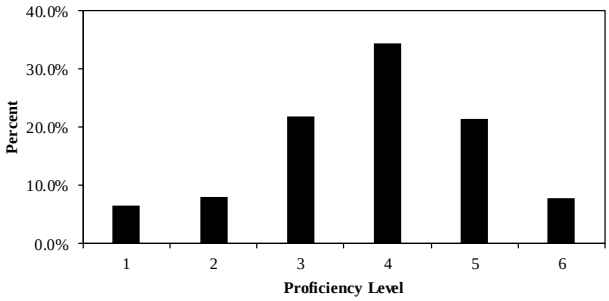


Table 8.5.5B
Proficiency Level Distribution: Oral 4-5 S401 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,167	6.3%	2,212	6.8%	4,379	6.6%
2	2,760	8.0%	2,582	8.0%	5,342	8.0%
3	8,059	23.5%	6,532	20.2%	14,591	21.9%
4	11,351	33.1%	11,581	35.8%	22,932	34.4%
5	7,407	21.6%	6,867	21.2%	14,274	21.4%
6	2,575	7.5%	2,565	7.9%	5,140	7.7%
Total	34,319	100.0%	32,339	100.0%	66,658	100.0%

Table 8.5.5C
n/a

Figure 8.5.5C
n/a

Figure 8.5.5D
n/a

Table 8.5.5D

Oral Composite Reliability: Oral 4-5 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.50	1622.783	0.667
Speaking	0.50	4926.133	0.906
Oral		2490.914	0.899

*Variances from students who had results in all four domains

Table 8.5.5Ei

Accuracy and Consistency of Classification Indices: Oral (Grade 4) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.656	0.550	0.419		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.883	0.783		
	2	0.564	0.429		
	3	0.700	0.583		
	4	0.695	0.591		
	5	0.567	0.497		
	6	0.659	0.413		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.978	0.006	0.016	0.970
	2/3	0.953	0.025	0.022	0.930
	3/4	0.901	0.045	0.054	0.862
	4/5	0.894	0.046	0.060	0.849
	5/6	0.928	0.069	0.003	0.915

Table 8.5.5Eii

Accuracy and Consistency of Classification Indices: Oral (Grade 5) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.649	0.546	0.412

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.881	0.782
	2	0.562	0.428
	3	0.659	0.536
	4	0.725	0.623
	5	0.540	0.476
	6	-	0.397

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.977	0.007	0.016	0.968
	2/3	0.954	0.023	0.023	0.932
	3/4	0.905	0.045	0.049	0.868
	4/5	0.890	0.045	0.065	0.842
	5/6	0.921	0.079	0.000	0.910

8.5.6 Literacy Composite 4–5

Figure 8.5.6A
Scale Scores: Litr 4-5 S401 Paper

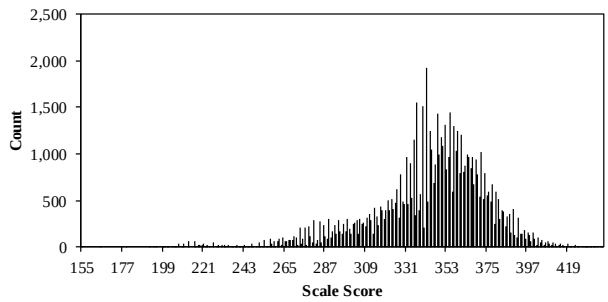


Table 8.5.6A
Scale Score Descriptive Statistics: Litr 4-5 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	34,296	155	433	339.35	30.35
5	32,310	165	438	349.05	31.26
Total	66,606	155	438	344.05	31.18

Figure 8.5.6B
Proficiency Level: Litr 4-5 S401 Paper

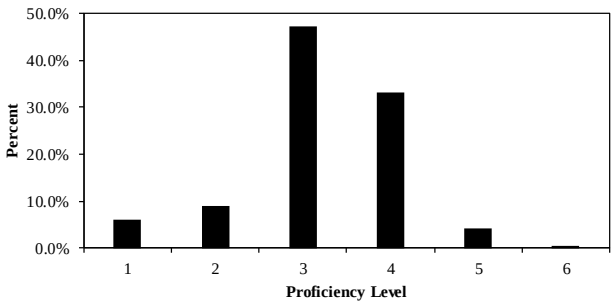


Table 8.5.6B
Proficiency Level Distribution: Litr 4-5 S401 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,180	6.4%	1,894	5.9%	4,074	6.1%
2	3,106	9.1%	2,858	8.8%	5,964	9.0%
3	17,149	50.0%	14,265	44.2%	31,414	47.2%
4	10,482	30.6%	11,550	35.7%	22,032	33.1%
5	1,202	3.5%	1,565	4.8%	2,767	4.2%
6	177	0.5%	178	0.6%	355	0.5%
Total	34,296	100.0%	32,310	100.0%	66,606	100.0%

Table 8.5.6C
n/a

Figure 8.5.6C
n/a

Figure 8.5.6D
n/a

Table 8.5.6D

Literacy Composite Reliability: Litr 4-5 S401 Paper

Component	Weight	Variance	Reliability
Reading	0.50	909.877	0.781
Writing	0.50	1445.912	0.894
Literacy		972.080	0.909

*Variances from students who had results in all four domains

Table 8.5.6Ei

Accuracy and Consistency of Classification Indices: Litr (Grade 4) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.781	0.708	0.549

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.810	0.783
	2	0.632	0.494
	3	0.862	0.799
	4	0.712	0.652
	5	-	0.222
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.976	0.012	0.012	0.972
	2/3	0.949	0.028	0.023	0.933
	3/4	0.883	0.046	0.071	0.842
	4/5	0.960	0.040	0.000	0.958
	5/6	0.995	0.005	0.000	1.000

Table 8.5.6Eii

Accuracy and Consistency of Classification Indices: Litr (Grade 5) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.777	0.697	0.544

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.827	0.784
	2	0.639	0.505
	3	0.845	0.776
	4	0.733	0.685
	5	-	0.269
	6	-	0.994

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.978	0.010	0.012	0.974
	2/3	0.953	0.025	0.022	0.936
	3/4	0.890	0.047	0.063	0.850
	4/5	0.946	0.054	0.000	0.935
	5/6	0.994	0.006	0.000	0.998

8.5.7 Comprehension Composite 4–5

Figure 8.5.7A
Scale Scores: Cphn 4-5 S401 Paper

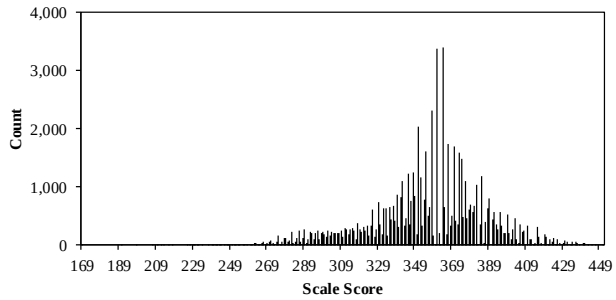


Table 8.5.7A
Scale Score Descriptive Statistics: Cphn 4-5 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	34,298	169	451	351.08	29.84
5	32,314	213	451	360.92	31.51
Total	66,612	169	451	355.85	31.05

Figure 8.5.7B
Proficiency Level: Cphn 4-5 S401 Paper

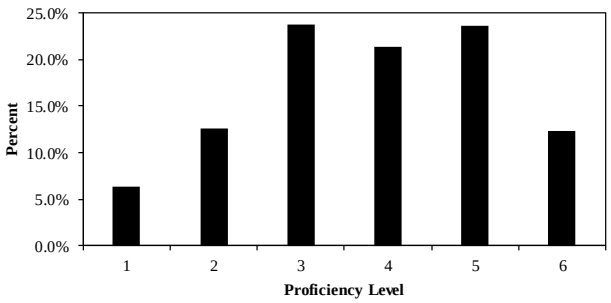


Table 8.5.7B
Proficiency Level Distribution: Cphn 4-5 S401 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,000	5.8%	2,205	6.8%	4,205	6.3%
2	4,302	12.5%	4,101	12.7%	8,403	12.6%
3	8,340	24.3%	7,514	23.3%	15,854	23.8%
4	7,190	21.0%	7,018	21.7%	14,208	21.3%
5	8,503	24.8%	7,198	22.3%	15,701	23.6%
6	3,963	11.6%	4,278	13.2%	8,241	12.4%
Total	34,298	100.0%	32,314	100.0%	66,612	100.0%

Table 8.5.7C
n/a

Figure 8.5.7C
n/a

Figure 8.5.7D
n/a

Table 8.5.7D

Comprehension Composite Reliability: Cphn 4-5 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.30	1622.783	0.667
Reading	0.70	909.877	0.781
Comprehension		964.340	0.848

*Variances from students who had results in all four domains

Table 8.5.7Ei

Accuracy and Consistency of Classification Indices: Cphn (Grade 4) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.582	0.468	0.340

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.855	0.706
	2	0.596	0.461
	3	0.584	0.470
	4	0.449	0.357
	5	0.583	0.476
	6	0.714	0.531

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.971	0.006	0.023	0.962
	2/3	0.929	0.033	0.038	0.895
	3/4	0.866	0.072	0.062	0.815
	4/5	0.861	0.071	0.067	0.811
	5/6	0.927	0.045	0.028	0.892

Table 8.5.7Eii

Accuracy and Consistency of Classification Indices: Cphn (Grade 5) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.566	0.454	0.329

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.854	0.712
	2	0.583	0.447
	3	0.561	0.449
	4	0.462	0.368
	5	0.516	0.416
	6	0.717	0.538

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.968	0.008	0.024	0.957
	2/3	0.927	0.035	0.038	0.892
	3/4	0.863	0.074	0.064	0.813
	4/5	0.859	0.063	0.078	0.807
	5/6	0.916	0.053	0.031	0.878

8.5.8 Overall Composite 4–5

Figure 8.5.8A
Scale Scores: Over 4-5 S401 Paper

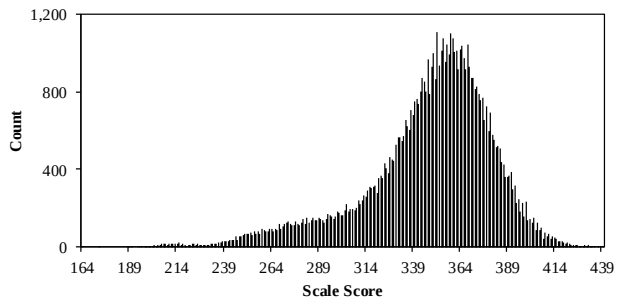


Table 8.5.8A
Scale Score Descriptive Statistics: Over 4-5 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	34,286	164	440	342.31	33.76
5	32,305	188	440	352.23	34.66
Total	66,591	164	440	347.12	34.56

Figure 8.5.8B
Proficiency Level: Over 4-5 S401 Paper

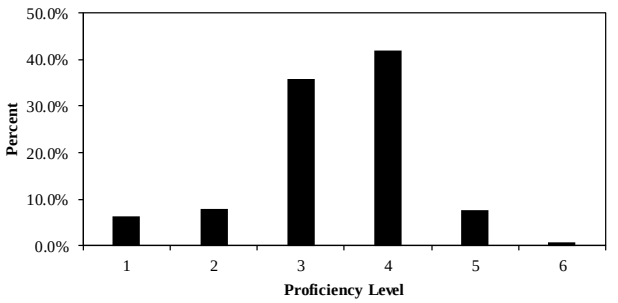


Table 8.5.8B
Proficiency Level Distribution: Over 4-5 S401 Paper

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,116	6.2%	1,988	6.2%	4,104	6.2%
2	2,670	7.8%	2,493	7.7%	5,163	7.8%
3	12,928	37.7%	10,978	34.0%	23,906	35.9%
4	14,065	41.0%	13,806	42.7%	27,871	41.9%
5	2,243	6.5%	2,825	8.7%	5,068	7.6%
6	264	0.8%	215	0.7%	479	0.7%
Total	34,286	100.0%	32,305	100.0%	66,591	100.0%

Table 8.5.8C
n/a

Figure 8.5.8C
n/a

Figure 8.5.8D
n/a

Table 8.5.8D

Overall Composite Reliability: Over 4-5 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.15	1622.783	0.667
Reading	0.35	909.877	0.781
Speaking	0.15	4926.133	0.906
Writing	0.35	1445.912	0.894
Overall Composite		1194.079	0.945

*Variances from students who had results in all four domains

Table 8.5.8Ei

Accuracy and Consistency of Classification Indices: Over (Grade 4) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.804	0.743	0.618

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.827	0.863
	2	0.697	0.578
	3	0.872	0.813
	4	0.769	0.747
	5	-	0.334
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.980	0.011	0.009	0.981
	2/3	0.963	0.023	0.014	0.955
	3/4	0.915	0.038	0.047	0.890
	4/5	0.927	0.073	0.000	0.918
	5/6	0.992	0.008	0.000	0.999

Table 8.5.8Eii

Accuracy and Consistency of Classification Indices: Over (Grade 5) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.787	0.725	0.596

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.844	0.862
	2	0.699	0.584
	3	0.865	0.799
	4	0.747	0.737
	5	-	0.367
	6	-	0.998

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.981	0.010	0.009	0.981
	2/3	0.965	0.021	0.014	0.956
	3/4	0.921	0.034	0.044	0.895
	4/5	0.906	0.094	0.000	0.893
	5/6	0.993	0.007	0.000	0.998

8.6 Grades: 6–8

8.6.1 Listening 6–8

Figure 8.6.1A
Scale Scores: List 6-8 S401 Paper

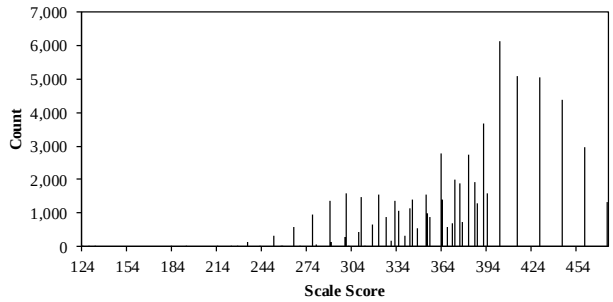


Table 8.6.1A
Scale Score Descriptive Statistics: List 6-8 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	23,942	124	474	375.18	45.81
7	21,442	128	474	379.26	48.94
8	20,969	132	474	384.85	51.99
Total	66,353	124	474	379.55	49.00

Figure 8.6.1B
Proficiency Level: List 6-8 S401 Paper

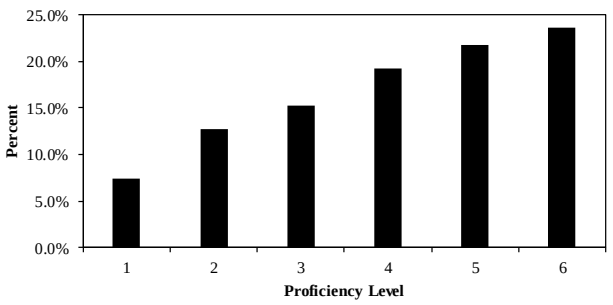


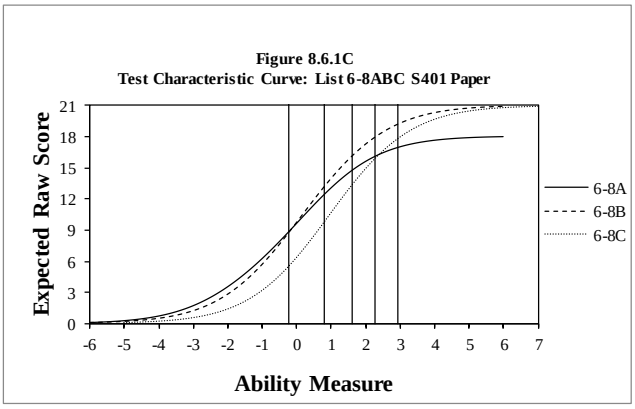
Table 8.6.1B
Proficiency Level Distribution: List 6-8 S401 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,352	5.6%	1,787	8.3%	1,806	8.6%	4,945	7.5%
2	2,459	10.3%	2,660	12.4%	3,280	15.6%	8,399	12.7%
3	3,611	15.1%	3,385	15.8%	3,138	15.0%	10,134	15.3%
4	5,388	22.5%	4,536	21.2%	2,826	13.5%	12,750	19.2%
5	5,412	22.6%	4,659	21.7%	4,392	20.9%	14,463	21.8%
6	5,720	23.9%	4,415	20.6%	5,527	26.4%	15,662	23.6%
Total	23,942	100.0%	21,442	100.0%	20,969	100.0%	66,353	100.0%

Table 8.6.1C

Conditional Standard Error of Measurement at Cut Scores: List 6-8 S401 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	6	294	20.29	18.79	21.42
	7	302	19.91	18.41	20.29
	8	308	19.91	18.24	19.54
2/3	6	332	20.03	18.41	18.03
	7	340	20.29	18.41	17.66
	8	347	20.66	18.79	17.66
3/4	6	363	22.17	20.00	17.66
	7	370	23.29	20.66	18.03
	8	377	24.05	21.42	18.41
4/5	6	385	n/a	22.54	19.16
	7	394	n/a	23.67	19.91
	8	402	n/a	25.18	20.63
5/6	6	411	n/a	n/a	21.79
	7	420	n/a	n/a	23.29
	8	427	n/a	n/a	24.80



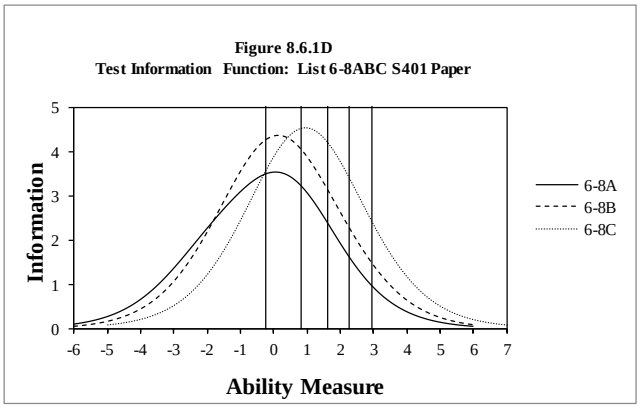


Table 8.6.1D

Weighted Reliability: List 6-8 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	14,480	0.734	0.661
B	16,617	0.682	
C	35,256	0.621	

Table 8.6.1Ei

Accuracy and Consistency of Classification Indices: List (Grade 6) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.437	0.351	0.198
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.726	0.406
	2	0.388	0.267
	3	0.317	0.239
	4	0.370	0.299
	5	0.380	0.309
	6	0.730	0.547

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.953	0.006	0.042	0.935
	2/3	0.897	0.033	0.069	0.846
	3/4	0.824	0.099	0.076	0.757
	4/5	0.793	0.114	0.093	0.726
	5/6	0.846	0.104	0.050	0.784

Table 8.6.1Eii

Accuracy and Consistency of Classification Indices: List (Grade 7) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.419	0.337	0.192
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.708	0.418
	2	0.364	0.266
	3	0.301	0.232
	4	0.345	0.278
	5	0.389	0.309
	6	0.710	0.517

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.930	0.009	0.060	0.905
	2/3	0.872	0.047	0.081	0.815
	3/4	0.813	0.105	0.082	0.744
	4/5	0.799	0.118	0.083	0.733
	5/6	0.862	0.091	0.047	0.802

Table 8.6.1Eiii

Accuracy and Consistency of Classification Indices: List (Grade 8) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.435	0.358	0.215

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.612	0.365
	2	0.421	0.327
	3	0.301	0.231
	4	0.247	0.189
	5	0.388	0.306
	6	0.762	0.604

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.922	0.015	0.063	0.891
	2/3	0.859	0.051	0.090	0.805
	3/4	0.839	0.084	0.077	0.770
	4/5	0.818	0.124	0.058	0.755
	5/6	0.845	0.106	0.049	0.790

8.6.2 Reading 6–8

Figure 8.6.2A
Scale Scores: Read 6-8 S401 Paper

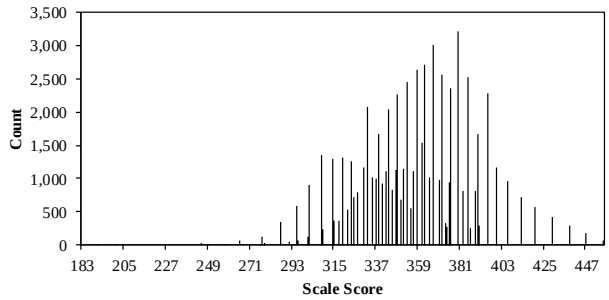


Table 8.6.2A
Scale Score Descriptive Statistics: Read 6-8 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	23,938	183	456	351.63	26.15
7	21,441	191	456	357.09	28.58
8	20,969	200	456	364.18	31.38
Total	66,348	183	456	357.36	29.13

Figure 8.6.2B
Proficiency Level: Read 6-8 S401 Paper

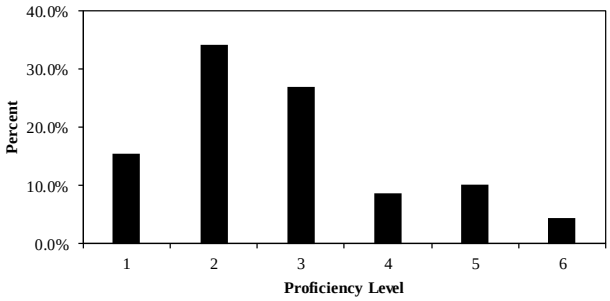


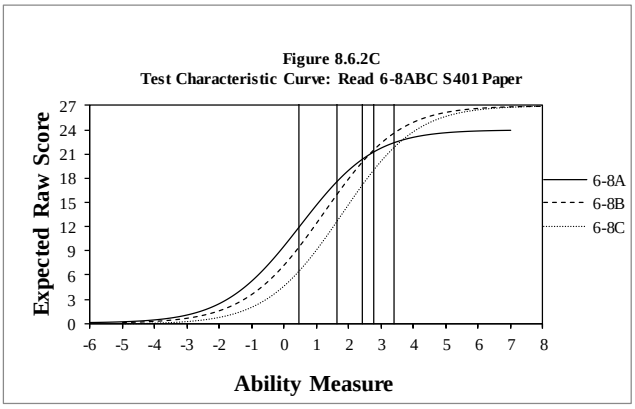
Table 8.6.2B
Proficiency Level Distribution: Read 6-8 S401 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,213	13.4%	3,415	15.9%	3,656	17.4%	10,284	15.5%
2	8,307	34.7%	7,664	35.7%	6,737	32.1%	22,708	34.2%
3	7,379	30.8%	5,117	23.9%	5,458	26.0%	17,954	27.1%
4	2,347	9.8%	2,153	10.0%	1,238	5.9%	5,738	8.6%
5	1,978	8.3%	2,201	10.3%	2,555	12.2%	6,734	10.1%
6	714	3.0%	891	4.2%	1,325	6.3%	2,930	4.4%
Total	23,938	100.0%	21,441	100.0%	20,969	100.0%	66,348	100.0%

Table 8.6.2C

Conditional Standard Error of Measurement at Cut Scores: Read 6-8 S401 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	6	323	11.70	12.22	13.78
	7	329	11.44	11.70	13.00
	8	335	11.45	11.44	12.48
2/3	6	353	11.96	10.92	11.13
	7	360	12.22	10.92	10.92
	8	366	12.74	10.92	10.92
3/4	6	373	13.52	11.44	10.66
	7	380	14.56	11.71	10.91
	8	386	15.60	12.22	11.18
4/5	6	382	n/a	11.96	10.92
	7	389	n/a	12.48	11.18
	8	395	n/a	13.26	11.61
5/6	6	399	n/a	n/a	11.96
	7	406	n/a	n/a	12.61
	8	412	n/a	n/a	13.26



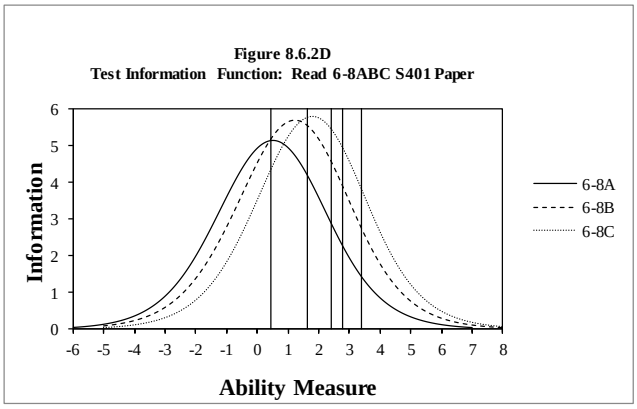


Table 8.6.2D

Weighted Reliability: Read 6-8 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	14,477	0.777	0.787
B	16,616	0.808	
C	35,255	0.780	

Table 8.6.2Ei

Accuracy and Consistency of Classification Indices: Read (Grade 6) S401 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.595	0.481		0.309	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.769		0.598	
	2	0.654		0.558	
	3	0.575		0.480	
	4	0.312		0.225	
	5	0.460		0.314	
	6	0.740		0.411	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.925	0.025	0.050	0.892
	2/3	0.844	0.083	0.073	0.785
	3/4	0.884	0.071	0.045	0.837
	4/5	0.927	0.048	0.025	0.891
	5/6	0.976	0.021	0.003	0.966

Table 8.6.2Eii

Accuracy and Consistency of Classification Indices: Read (Grade 7) S401 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.581	0.470		0.309	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.752		0.591	
	2	0.656		0.564	
	3	0.491		0.395	
	4	0.312		0.231	
	5	0.492		0.354	
	6	0.726		0.455	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.908	0.033	0.058	0.870
	2/3	0.854	0.077	0.069	0.796
	3/4	0.883	0.079	0.038	0.840
	4/5	0.916	0.056	0.027	0.882
	5/6	0.970	0.024	0.007	0.955

Table 8.6.2Eiii

Accuracy and Consistency of Classification Indices: Read (Grade 8) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.566	0.455	0.303

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.747	0.591
	2	0.599	0.506
	3	0.534	0.427
	4	0.179	0.129
	5	0.510	0.370
	6	0.738	0.508

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.900	0.038	0.062	0.858
	2/3	0.846	0.087	0.067	0.789
	3/4	0.890	0.061	0.049	0.840
	4/5	0.908	0.065	0.027	0.870
	5/6	0.957	0.031	0.011	0.938

8.6.3 Writing 6–8

Figure 8.6.3A
Scale Scores: Writ 6-8 S401 Paper

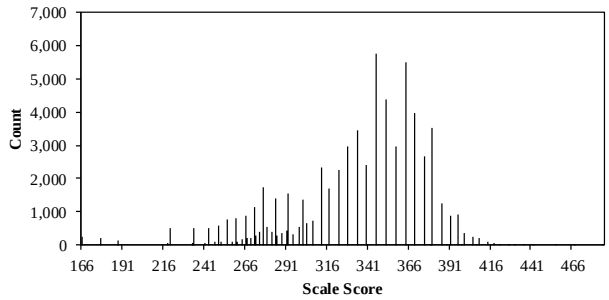


Table 8.6.3A
Scale Score Descriptive Statistics: Writ 6-8 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	23,934	166	456	326.64	42.30
7	21,426	177	485	330.61	43.71
8	20,962	188	467	336.53	43.82
Total	66,322	166	485	331.05	43.43

Figure 8.6.3B
Proficiency Level: Writ 6-8 S401 Paper

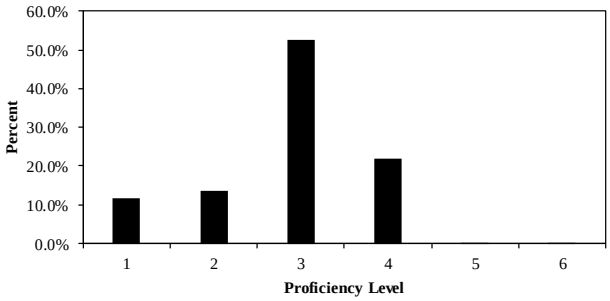


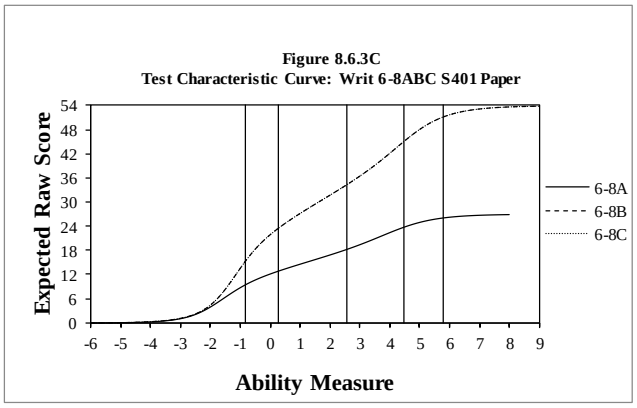
Table 8.6.3B
Proficiency Level Distribution: Writ 6-8 S401 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,276	9.5%	2,590	12.1%	2,992	14.3%	7,858	11.8%
2	3,455	14.4%	3,269	15.3%	2,277	10.9%	9,001	13.6%
3	12,582	52.6%	10,996	51.3%	11,200	53.4%	34,778	52.4%
4	5,601	23.4%	4,540	21.2%	4,463	21.3%	14,604	22.0%
5	19	0.1%	29	0.1%	28	0.1%	76	0.1%
6	1	0.0%	2	0.0%	2	0.0%	5	0.0%
Total	23,934	100.0%	21,426	100.0%	20,962	100.0%	66,322	100.0%

Table 8.6.3C

Conditional Standard Error of Measurement at Cut Scores: Writ 6-8 S401 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	6	268	12.35	8.32
	7	273	12.62	8.32
	8	281	13.69	8.59
2/3	6	298	15.84	10.20
	7	305	16.65	11.01
	8	311	16.92	11.55
3/4	6	361	17.45	12.62
	7	367	17.18	12.35
	8	372	16.92	12.35
4/5	6	413	15.57	10.74
	7	419	16.11	10.74
	8	424	16.92	11.01
5/6	6	441	20.68	12.35
	7	450	23.90	14.23
	8	459	27.93	16.38



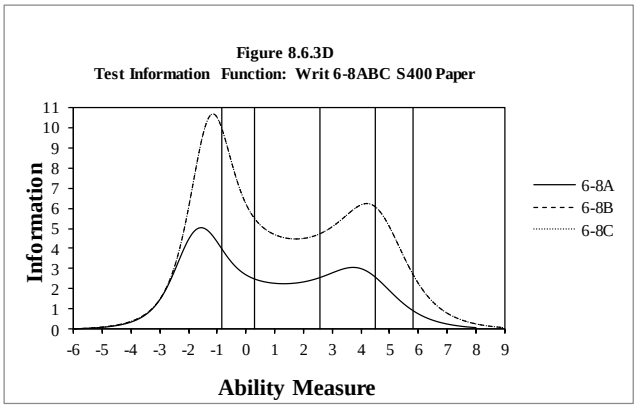


Table 8.6.3D

Weighted Reliability: Writ 6-8 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	14,468	0.873	0.903
B/C	51,854	0.911	

Table 8.6.3Ei

Accuracy and Consistency of Classification Indices: Writ (Grade 6) S401 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.765	0.681		0.497	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.836		0.739	
	2	0.704		0.582	
	3	0.797		0.730	
	4	0.696		0.594	
	5	-		-	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.968	0.016	0.016	0.954
	2/3	0.939	0.023	0.038	0.915
	3/4	0.858	0.072	0.070	0.809
	4/5	0.999	0.001	0.000	0.999
	5/6	1.000	0.000	0.000	1.000

Table 8.6.3Eii

Accuracy and Consistency of Classification Indices: Writ (Grade 7) S401 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.770	0.686		0.520	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.867		0.783	
	2	0.688		0.567	
	3	0.806		0.740	
	4	0.687		0.583	
	5	-		-	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.964	0.015	0.020	0.949
	2/3	0.936	0.025	0.038	0.911
	3/4	0.871	0.063	0.067	0.824
	4/5	0.999	0.001	0.000	0.999
	5/6	1.000	0.000	0.000	1.000

Table 8.6.3Eiii

Accuracy and Consistency of Classification Indices: Writ (Grade 8) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.778	0.696	0.529

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.904	0.836
	2	0.568	0.440
	3	0.836	0.771
	4	0.685	0.584
	5	-	-
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.963	0.013	0.024	0.949
	2/3	0.943	0.028	0.029	0.917
	3/4	0.873	0.057	0.070	0.826
	4/5	0.999	0.001	0.000	0.999
	5/6	1.000	0.000	0.000	1.000

8.6.4 Speaking 6–8

Figure 8.6.4A
Scale Scores: Spek 6-8 S401 Paper

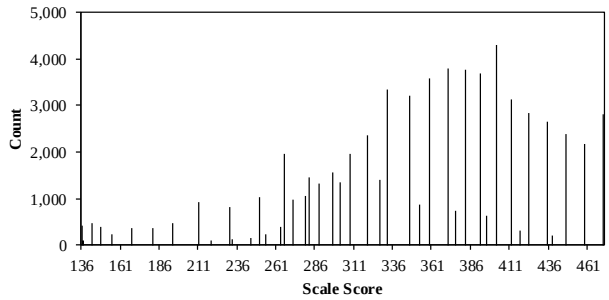


Table 8.6.4A
Scale Score Descriptive Statistics: Spek 6-8 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	23,940	136	471	351.12	72.44
7	21,445	142	471	352.99	74.59
8	20,966	148	471	358.61	76.08
Total	66,351	136	471	354.09	74.36

Figure 8.6.4B
Proficiency Level: Spek 6-8 S401 Paper

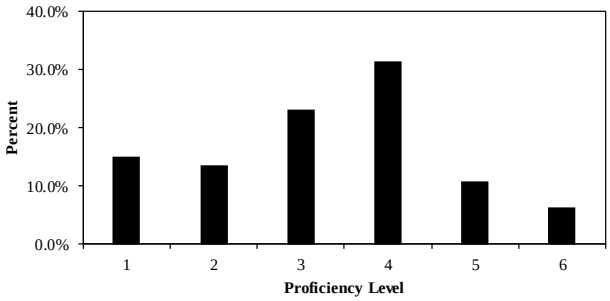


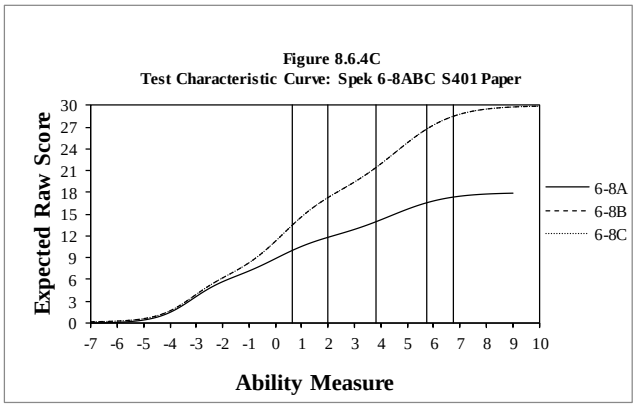
Table 8.6.4B
Proficiency Level Distribution: Spek 6-8 S401 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,974	12.4%	3,181	14.8%	3,755	17.9%	9,910	14.9%
2	3,650	15.2%	2,794	13.0%	2,506	12.0%	8,950	13.5%
3	5,831	24.4%	4,742	22.1%	4,836	23.1%	15,409	23.2%
4	7,314	30.6%	7,543	35.2%	5,948	28.4%	20,805	31.4%
5	2,765	11.5%	1,614	7.5%	2,743	13.1%	7,122	10.7%
6	1,406	5.9%	1,571	7.3%	1,178	5.6%	4,155	6.3%
Total	23,940	100.0%	21,445	100.0%	20,966	100.0%	66,351	100.0%

Table 8.6.4C

Conditional Standard Error of Measurement at Cut Scores: Spek 6-8 S401 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	6	268	21.64	15.79
	7	277	22.23	15.50
	8	284	22.81	15.79
2/3	6	310	26.62	18.13
	7	317	27.49	19.01
	8	323	28.08	19.60
3/4	6	360	26.03	19.01
	7	369	25.15	18.43
	8	377	24.57	17.84
4/5	6	417	25.74	17.84
	7	425	27.20	18.43
	8	433	28.96	19.60
5/6	6	451	35.10	23.11
	7	457	37.73	24.86
	8	463	40.95	26.62



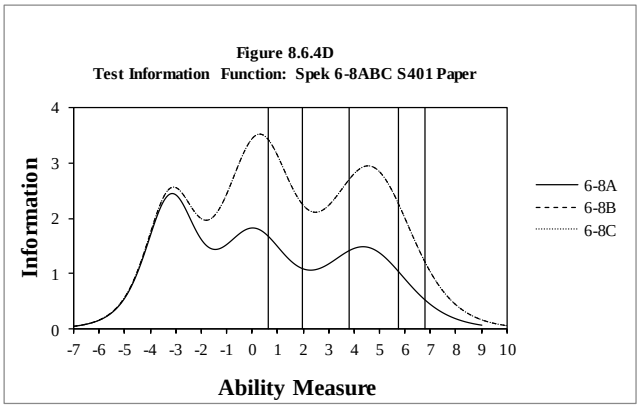


Table 8.6.4D

Weighted Reliability: Spek 6-8 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	14,482	0.889	0.909
B/C	51,869	0.915	

Table 8.6.4Ei

Accuracy and Consistency of Classification Indices: Spek (Grade 6) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.633	0.529	0.405
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.804	0.698
	2	0.599	0.475
	3	0.643	0.535
	4	0.686	0.572
	5	0.433	0.369
	6	-	0.328

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.953	0.025	0.022	0.931
	2/3	0.916	0.037	0.047	0.885
	3/4	0.904	0.039	0.058	0.865
	4/5	0.909	0.042	0.049	0.866
	5/6	0.941	0.059	0.000	0.936

Table 8.6.4Eii

Accuracy and Consistency of Classification Indices: Spek (Grade 7) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.628	0.528	0.399
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.833	0.740
	2	0.548	0.425
	3	0.606	0.494
	4	0.743	0.633
	5	0.294	0.243
	6	-	0.403

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.950	0.025	0.025	0.928
	2/3	0.923	0.034	0.043	0.893
	3/4	0.900	0.049	0.051	0.861
	4/5	0.912	0.036	0.052	0.866
	5/6	0.927	0.073	0.000	0.924

Table 8.6.4Eiii

Accuracy and Consistency of Classification Indices: Spek (Grade 8) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.639	0.538	0.420
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.857	0.774
	2	0.494	0.379
	3	0.642	0.530
	4	0.657	0.546
	5	0.483	0.412
	6	-	0.293

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.945	0.025	0.030	0.921
	2/3	0.923	0.036	0.041	0.892
	3/4	0.908	0.041	0.051	0.870
	4/5	0.905	0.055	0.041	0.863
	5/6	0.944	0.056	0.000	0.936

8.6.5 Oral Language Composite 6–8

Figure 8.6.5A
Scale Scores: Oral 6-8 S401 Paper

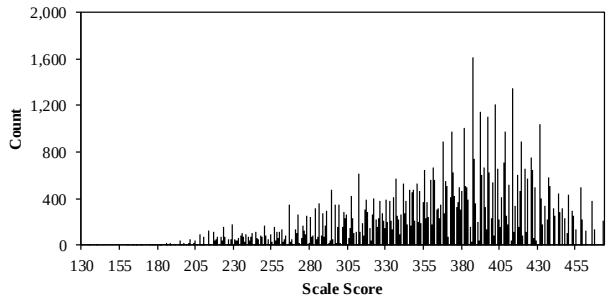


Table 8.6.5A
Scale Score Descriptive Statistics: Oral 6-8 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	23,937	130	473	363.41	53.58
7	21,438	135	473	366.37	56.52
8	20,961	140	473	371.99	58.98
Total	66,336	130	473	367.08	56.39

Figure 8.6.5B
Proficiency Level: Oral 6-8 S401 Paper

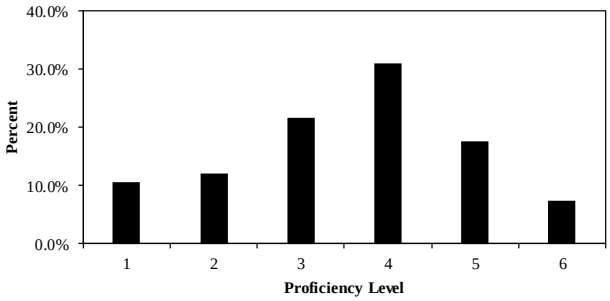


Table 8.6.5B
Proficiency Level Distribution: Oral 6-8 S401 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,020	8.4%	2,348	11.0%	2,636	12.6%	7,004	10.6%
2	2,584	10.8%	2,747	12.8%	2,610	12.5%	7,941	12.0%
3	5,439	22.7%	4,554	21.2%	4,327	20.6%	14,320	21.6%
4	7,705	32.2%	6,730	31.4%	6,064	28.9%	20,499	30.9%
5	4,244	17.7%	3,672	17.1%	3,712	17.7%	11,628	17.5%
6	1,945	8.1%	1,387	6.5%	1,612	7.7%	4,944	7.5%
Total	23,937	100.0%	21,438	100.0%	20,961	100.0%	66,336	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.50	2401.347	0.661
Speaking	0.50	5525.711	0.909
Oral		3179.188	0.897

*Variances from students who had results in all four domains

Table 8.6.5Ei

Accuracy and Consistency of Classification Indices: Oral (Grade 6) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.641	0.529	0.405		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.852	0.740		
	2	0.577	0.446		
	3	0.643	0.526		
	4	0.688	0.584		
	5	0.527	0.439		
	6	0.702	0.471		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.968	0.011	0.021	0.955
	2/3	0.939	0.028	0.032	0.913
	3/4	0.897	0.050	0.053	0.856
	4/5	0.899	0.047	0.054	0.856
	5/6	0.932	0.058	0.010	0.915

Table 8.6.5Eii

Accuracy and Consistency of Classification Indices: Oral (Grade 7) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.643	0.528	0.409
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.843	0.735
	2	0.579	0.453
	3	0.594	0.479
	4	0.686	0.585
	5	0.568	0.460
	6	0.674	0.443

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.959	0.016	0.026	0.942
	2/3	0.930	0.030	0.039	0.901
	3/4	0.897	0.054	0.049	0.856
	4/5	0.903	0.050	0.047	0.863
	5/6	0.948	0.040	0.012	0.928

Table 8.6.5Eiii

Accuracy and Consistency of Classification Indices: Oral (Grade 8) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.633	0.518	0.403
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.846	0.743
	2	0.545	0.424
	3	0.593	0.477
	4	0.661	0.556
	5	0.561	0.457
	6	0.661	0.448

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.953	0.018	0.029	0.934
	2/3	0.929	0.032	0.039	0.899
	3/4	0.901	0.052	0.047	0.861
	4/5	0.902	0.053	0.045	0.862
	5/6	0.939	0.045	0.016	0.915

8.6.6 Literacy Composite 6–8

Figure 8.6.6A
Scale Scores: Litr 6-8 S401 Paper

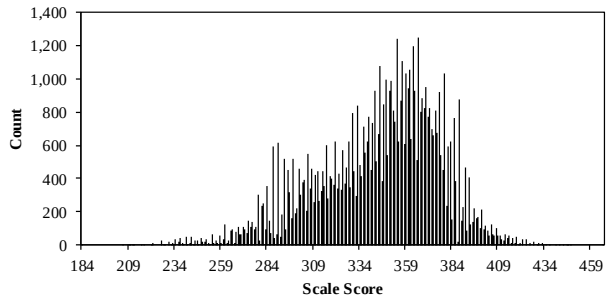


Table 8.6.6A
Scale Score Descriptive Statistics: Litr 6-8 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	23,929	206	435	339.38	31.09
7	21,419	184	466	344.11	33.14
8	20,958	210	448	350.62	34.71
Total	66,306	184	466	344.46	33.25

Figure 8.6.6B
Proficiency Level: Litr 6-8 S401 Paper

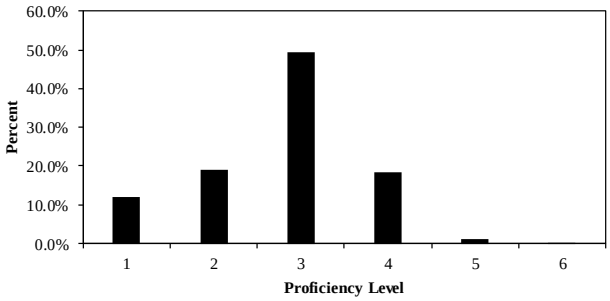


Table 8.6.6B
Proficiency Level Distribution: Litr 6-8 S401 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,460	10.3%	2,549	11.9%	2,954	14.1%	7,963	12.0%
2	4,525	18.9%	4,371	20.4%	3,753	17.9%	12,649	19.1%
3	12,528	52.4%	10,494	49.0%	9,680	46.2%	32,702	49.3%
4	4,223	17.6%	3,766	17.6%	4,206	20.1%	12,195	18.4%
5	184	0.8%	224	1.0%	357	1.7%	765	1.2%
6	9	0.0%	15	0.1%	8	0.0%	32	0.0%
Total	23,929	100.0%	21,419	100.0%	20,958	100.0%	66,306	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 S401 Paper

Component	Weight	Variance	Reliability
Reading	0.50	848.591	0.787
Writing	0.50	1885.940	0.903
Literacy		1105.560	0.918

*Variances from students who had results in all four domains

Table 8.6.6Ei

Accuracy and Consistency of Classification Indices: Litr (Grade 6) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.802	0.724	0.578
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.870	0.789
	2	0.736	0.629
	3	0.846	0.794
	4	0.711	0.606
	5	-	0.114
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.969	0.013	0.019	0.956
	2/3	0.932	0.032	0.036	0.903
	3/4	0.909	0.045	0.046	0.872
	4/5	0.992	0.008	0.000	0.992
	5/6	1.000	0.000	0.000	1.000

Table 8.6.6Eii

Accuracy and Consistency of Classification Indices: Litr (Grade 7) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.799	0.718	0.584
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.877	0.801
	2	0.726	0.621
	3	0.837	0.785
	4	0.731	0.628
	5	-	0.198
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.964	0.014	0.022	0.950
	2/3	0.926	0.036	0.038	0.896
	3/4	0.918	0.044	0.038	0.885
	4/5	0.989	0.011	0.000	0.986
	5/6	0.999	0.001	0.000	1.000

Table 8.6.6Eiii

Accuracy and Consistency of Classification Indices: Litr (Grade 8) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.787	0.703	0.576
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.891	0.820
	2	0.674	0.561
	3	0.833	0.774
	4	0.728	0.638
	5	-	0.257
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.960	0.014	0.026	0.945
	2/3	0.928	0.038	0.035	0.897
	3/4	0.915	0.043	0.042	0.880
	4/5	0.983	0.017	0.000	0.979
	5/6	1.000	0.000	0.000	1.000

8.6.7 Comprehension Composite 6–8

Figure 8.6.7A
Scale Scores: Cphn 6-8 S401 Paper

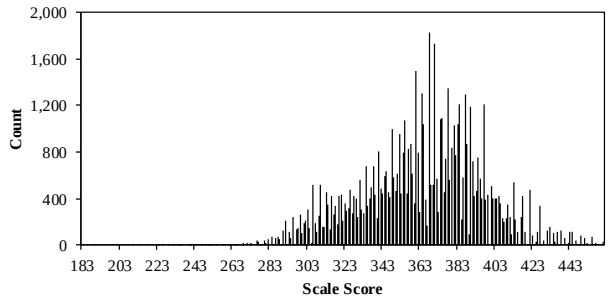


Table 8.6.7A
Scale Score Descriptive Statistics: Cphn 6-8 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	23,938	183	461	358.73	29.26
7	21,436	189	461	363.80	32.03
8	20,965	210	461	370.42	34.99
Total	66,339	183	461	364.06	32.41

Figure 8.6.7B
Proficiency Level: Cphn 6-8 S401 Paper

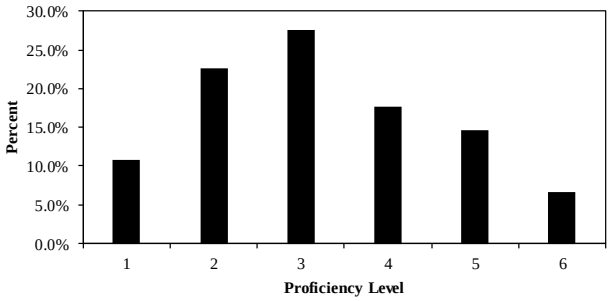


Table 8.6.7B
Proficiency Level Distribution: Cphn 6-8 S401 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,005	8.4%	2,365	11.0%	2,739	13.1%	7,109	10.7%
2	5,286	22.1%	5,080	23.7%	4,646	22.2%	15,012	22.6%
3	7,444	31.1%	5,758	26.9%	5,131	24.5%	18,333	27.6%
4	4,328	18.1%	3,971	18.5%	3,400	16.2%	11,699	17.6%
5	3,600	15.0%	2,893	13.5%	3,244	15.5%	9,737	14.7%
6	1,275	5.3%	1,369	6.4%	1,805	8.6%	4,449	6.7%
Total	23,938	100.0%	21,436	100.0%	20,965	100.0%	66,339	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.30	2401.347	0.661
Reading	0.70	848.591	0.787
Comprehension		1050.592	0.846

*Variances from students who had results in all four domains

Table 8.6.7Ei

Accuracy and Consistency of Classification Indices: Cphn (Grade 6) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**	
	0.606	0.493	0.358	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.829	0.673	
	2	0.653	0.538	
	3	0.619	0.520	
	4	0.450	0.352	
	5	0.563	0.432	
	6	0.739	0.500	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.958	0.011	0.031	0.942
	2/3	0.894	0.052	0.053	0.849
	3/4	0.867	0.074	0.059	0.818
	4/5	0.902	0.056	0.043	0.860
	5/6	0.963	0.027	0.009	0.947

Table 8.6.7Eii

Accuracy and Consistency of Classification Indices: Cphn (Grade 7) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.592	0.480	0.352
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.820	0.671
	2	0.639	0.528
	3	0.560	0.461
	4	0.471	0.370
	5	0.521	0.393
	6	0.757	0.538

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.945	0.016	0.039	0.924
	2/3	0.889	0.055	0.056	0.842
	3/4	0.869	0.079	0.052	0.823
	4/5	0.904	0.055	0.041	0.864
	5/6	0.960	0.028	0.011	0.941

Table 8.6.7Eiii

Accuracy and Consistency of Classification Indices: Cphn (Grade 8) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.578	0.469	0.349
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.823	0.683
	2	0.600	0.489
	3	0.536	0.434
	4	0.415	0.322
	5	0.531	0.409
	6	0.768	0.573

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.937	0.019	0.044	0.914
	2/3	0.889	0.056	0.055	0.842
	3/4	0.872	0.074	0.054	0.825
	4/5	0.896	0.060	0.043	0.855
	5/6	0.950	0.035	0.016	0.927

8.6.8 Overall Composite 6–8

Figure 8.6.8A
Scale Scores: Over 6-8 S401 Paper

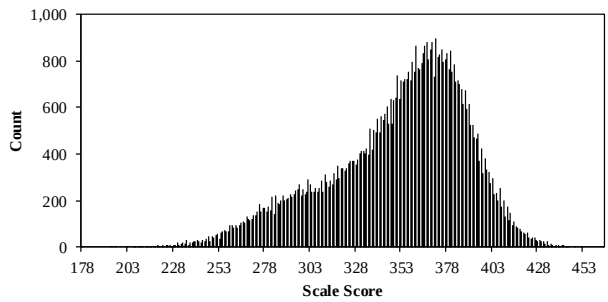


Table 8.6.8A
Scale Score Descriptive Statistics: Over 6-8 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	23,924	194	443	346.38	35.50
7	21,410	178	464	350.56	37.99
8	20,947	207	452	356.83	39.93
Total	66,281	178	464	351.03	37.99

Figure 8.6.8B
Proficiency Level: Over 6-8 S401 Paper

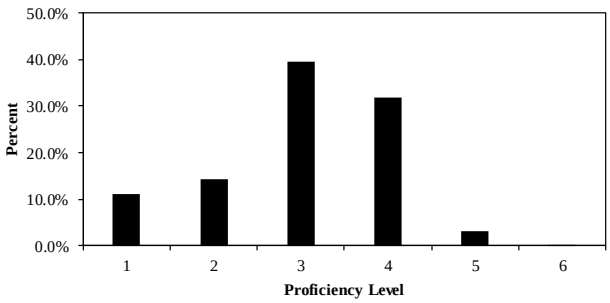


Table 8.6.8B
Proficiency Level Distribution: Over 6-8 S401 Paper

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,169	9.1%	2,484	11.6%	2,706	12.9%	7,359	11.1%
2	3,280	13.7%	3,194	14.9%	3,026	14.4%	9,500	14.3%
3	10,040	42.0%	8,502	39.7%	7,631	36.4%	26,173	39.5%
4	7,774	32.5%	6,567	30.7%	6,712	32.0%	21,053	31.8%
5	625	2.6%	629	2.9%	838	4.0%	2,092	3.2%
6	36	0.2%	34	0.2%	34	0.2%	104	0.2%
Total	23,924	100.0%	21,410	100.0%	20,947	100.0%	66,281	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.15	2401.347	0.661
Reading	0.35	848.591	0.787
Speaking	0.15	5525.711	0.909
Writing	0.35	1885.940	0.903
Overall Composite		1443.427	0.949

*Variances from students who had results in all four domains

Table 8.6.8Ei

Accuracy and Consistency of Classification Indices: Over (Grade 6) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.830	0.766	0.662

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.895	0.851
	2	0.752	0.653
	3	0.862	0.806
	4	0.807	0.763
	5	-	0.253
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.978	0.009	0.013	0.971
	2/3	0.954	0.024	0.022	0.937
	3/4	0.921	0.036	0.043	0.890
	4/5	0.972	0.028	0.000	0.969
	5/6	0.999	0.002	0.000	1.000

Table 8.6.8Eii

Accuracy and Consistency of Classification Indices: Over (Grade 7) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.819	0.750	0.649
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.902	0.857
	2	0.737	0.638
	3	0.847	0.788
	4	0.797	0.755
	5	-	0.311
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.973	0.011	0.016	0.964
	2/3	0.950	0.026	0.024	0.931
	3/4	0.923	0.037	0.039	0.893
	4/5	0.969	0.031	0.000	0.962
	5/6	0.998	0.002	0.000	1.000

Table 8.6.8Eiii

Accuracy and Consistency of Classification Indices: Over (Grade 8) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.807	0.735	0.635
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.904	0.860
	2	0.719	0.614
	3	0.833	0.768
	4	0.781	0.749
	5	2.200	0.346
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.970	0.012	0.018	0.960
	2/3	0.950	0.026	0.024	0.930
	3/4	0.924	0.038	0.039	0.895
	4/5	0.956	0.042	0.002	0.950
	5/6	0.998	0.002	0.000	1.000

8.7 Grades: 9–12

8.7.1 Listening 9–12

Figure 8.7.1A
Scale Scores: List 9-12 S401 Paper

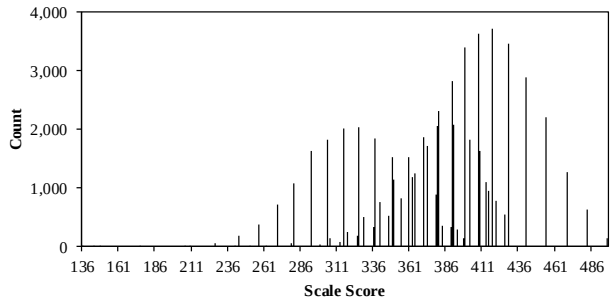


Table 8.7.1A
Scale Score Descriptive Statistics: List 9-12 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	22,491	136	497	375.02	51.44
10	18,210	175	497	380.92	47.85
11	14,660	144	497	384.22	46.36
12	9,741	148	497	390.45	42.06
Total	65,102	136	497	381.05	48.28

Figure 8.7.1B
Proficiency Level: List 9-12 S401 Paper

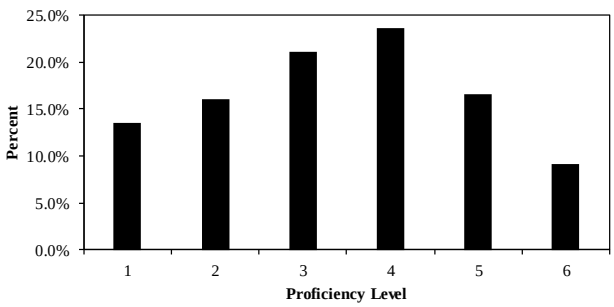


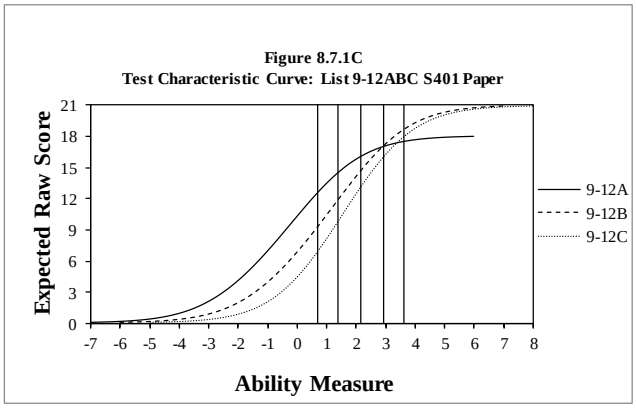
Table 8.7.1B
Proficiency Level Distribution: List 9-12 S401 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,036	13.5%	2,344	12.9%	2,139	14.6%	1,305	13.4%	8,824	13.6%
2	4,032	17.9%	2,801	15.4%	2,287	15.6%	1,296	13.3%	10,416	16.0%
3	4,250	18.9%	3,623	19.9%	3,524	24.0%	2,322	23.8%	13,719	21.1%
4	4,908	21.8%	4,548	25.0%	3,443	23.5%	2,491	25.6%	15,390	23.6%
5	4,051	18.0%	2,901	15.9%	2,212	15.1%	1,632	16.8%	10,796	16.6%
6	2,214	9.8%	1,993	10.9%	1,055	7.2%	695	7.1%	5,957	9.2%
Total	22,491	100.0%	18,210	100.0%	14,660	100.0%	9,741	100.0%	65,102	100.0%

Table 8.7.1C

Conditional Standard Error of Measurement at Cut Scores: List 9-12 S401 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	9	314	20.29	21.04	22.17
	10	325	20.66	20.29	20.60
	11	335	21.04	19.91	19.54
	12	342	21.42	19.91	19.16
2/3	9	353	22.17	19.54	18.41
	10	358	22.92	19.54	18.41
	11	364	23.29	19.54	18.20
	12	368	24.05	19.54	18.03
3/4	9	383	26.30	19.54	18.03
	10	389	27.43	19.91	18.03
	11	394	28.55	19.91	18.41
	12	398	29.68	20.29	18.41
4/5	9	409	n/a	21.04	19.16
	10	415	n/a	21.42	19.54
	11	420	n/a	22.17	19.91
	12	426	n/a	22.75	20.29
5/6	9	434	n/a	n/a	21.04
	10	441	n/a	n/a	22.10
	11	447	n/a	n/a	22.92
	12	452	n/a	n/a	24.05



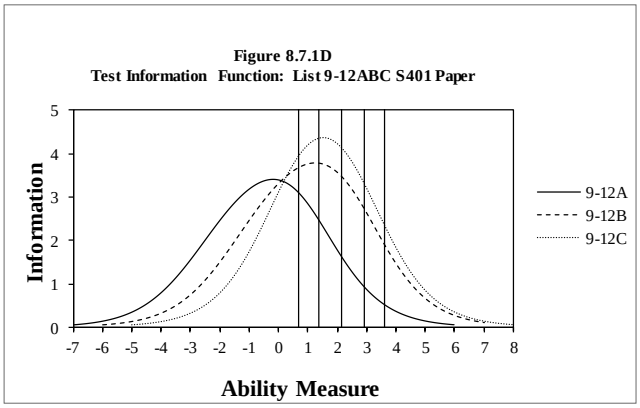


Table 8.7.1D

Weighted Reliability: List 9-12 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	16,436	0.660	0.657
B	17,255	0.667	
C	31,411	0.649	

Table 8.7.1Ei

Accuracy and Consistency of Classification Indices: List (Grade 9) S401 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.404	0.315		0.171	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.715		0.513	
	2	0.414		0.311	
	3	0.328		0.253	
	4	0.355		0.289	
	5	0.378		0.300	
	6	0.482		0.269	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.907	0.028	0.065	0.868
	2/3	0.845	0.064	0.091	0.781
	3/4	0.807	0.101	0.092	0.735
	4/5	0.808	0.111	0.080	0.744
	5/6	0.901	0.085	0.014	0.846

Table 8.7.1Eii

Accuracy and Consistency of Classification Indices: List (Grade 10) S401 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.406	0.319		0.173	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.739		0.513	
	2	0.376		0.274	
	3	0.337		0.266	
	4	0.400		0.328	
	5	0.343		0.266	
	6	0.640		0.347	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.912	0.022	0.065	0.874
	2/3	0.849	0.058	0.092	0.784
	3/4	0.797	0.118	0.085	0.725
	4/5	0.818	0.115	0.066	0.753
	5/6	0.900	0.088	0.012	0.859

Table 8.7.1Eiii

Accuracy and Consistency of Classification Indices: List (Grade 11) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.424	0.330	0.179
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.722	0.505
	2	0.347	0.259
	3	0.386	0.313
	4	0.395	0.320
	5	0.399	0.288
	6	0.639	0.313

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.900	0.029	0.071	0.856
	2/3	0.834	0.070	0.096	0.768
	3/4	0.794	0.114	0.092	0.723
	4/5	0.846	0.101	0.053	0.783
	5/6	0.935	0.056	0.009	0.900

Table 8.7.1Eiv

Accuracy and Consistency of Classification Indices: List (Grade 12) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.440	0.344	0.189
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.743	0.522
	2	0.333	0.244
	3	0.407	0.325
	4	0.424	0.348
	5	0.436	0.328
	6	0.666	0.323

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.909	0.023	0.068	0.872
	2/3	0.854	0.057	0.089	0.793
	3/4	0.806	0.107	0.087	0.734
	4/5	0.841	0.115	0.044	0.785
	5/6	0.934	0.061	0.005	0.907

8.7.2 Reading 9–12

Figure 8.7.2A
Scale Scores: Read 9-12 S401 Paper

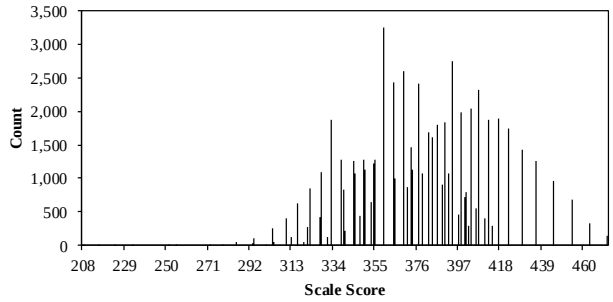


Table 8.7.2A
Scale Score Descriptive Statistics: Read 9-12 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	22,488	208	472	373.52	33.00
10	18,210	216	472	378.14	32.21
11	14,660	224	472	382.91	32.13
12	9,745	233	472	386.22	30.33
Total	65,103	208	472	378.83	32.53

Figure 8.7.2B
Proficiency Level: Read 9-12 S401 Paper

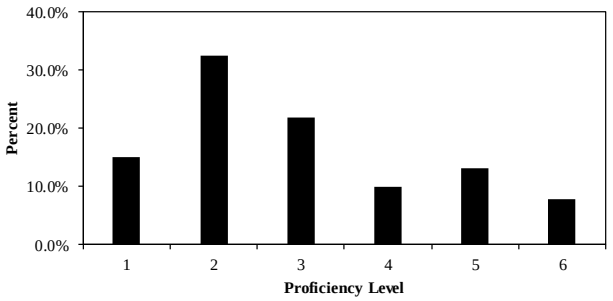


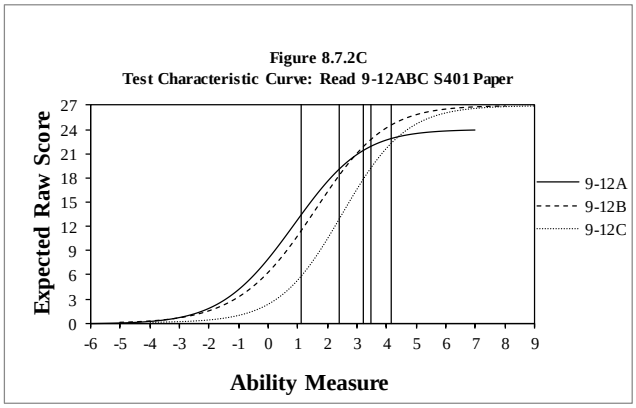
Table 8.7.2B
Proficiency Level Distribution: Read 9-12 S401 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	4,004	17.8%	2,507	13.8%	2,035	13.9%	1,229	12.6%	9,775	15.0%
2	6,733	29.9%	5,984	32.9%	4,884	33.3%	3,501	35.9%	21,102	32.4%
3	4,515	20.1%	4,132	22.7%	3,311	22.6%	2,243	23.0%	14,201	21.8%
4	2,389	10.6%	1,771	9.7%	1,346	9.2%	910	9.3%	6,416	9.9%
5	3,089	13.7%	2,533	13.9%	1,744	11.9%	1,218	12.5%	8,584	13.2%
6	1,758	7.8%	1,283	7.0%	1,340	9.1%	644	6.6%	5,025	7.7%
Total	22,488	100.0%	18,210	100.0%	14,660	100.0%	9,745	100.0%	65,103	100.0%

Table 8.7.2C

Conditional Standard Error of Measurement at Cut Scores: Read 9-12 S401 Paper

Proficiency Level	Grade	Cut Score	SEM		
			Tier A	Tier B	Tier C
1/2	9	340	11.70	11.96	14.33
	10	344	11.58	11.70	13.52
	11	348	11.70	11.44	13.00
	12	352	11.70	11.44	12.48
2/3	9	372	12.74	11.18	10.66
	10	377	13.01	11.18	10.52
	11	382	13.52	11.44	10.42
	12	386	14.04	11.70	10.37
3/4	9	392	14.88	11.96	10.40
	10	397	15.86	12.48	10.40
	11	402	16.90	13.00	10.66
	12	407	17.94	13.59	10.96
4/5	9	401	n/a	13.00	10.66
	10	406	n/a	13.52	10.92
	11	410	n/a	14.04	11.18
	12	414	n/a	14.56	11.44
5/6	9	418	n/a	n/a	11.96
	10	423	n/a	n/a	12.48
	11	427	n/a	n/a	13.00
	12	432	n/a	n/a	13.78



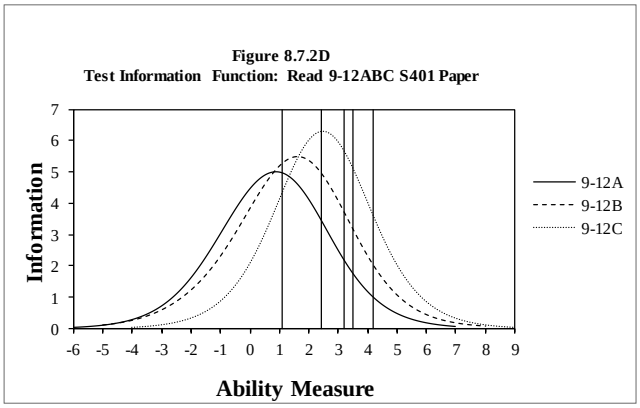


Table 8.7.2D

Weighted Reliability: Read 9-12 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	16,434	0.797	0.806
B	17,257	0.800	
C	31,412	0.814	

Table 8.7.2Ei

Accuracy and Consistency of Classification Indices: Read (Grade 9) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.554	0.449	0.312
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.770	0.622
	2	0.605	0.506
	3	0.436	0.343
	4	0.292	0.222
	5	0.493	0.372
	6	0.715	0.515

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.900	0.033	0.066	0.863
	2/3	0.868	0.070	0.062	0.812
	3/4	0.873	0.083	0.043	0.829
	4/5	0.894	0.066	0.040	0.856
	5/6	0.951	0.030	0.019	0.925

Table 8.7.2Eii

Accuracy and Consistency of Classification Indices: Read (Grade 10) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.563	0.453	0.305
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.702	0.532
	2	0.638	0.544
	3	0.478	0.379
	4	0.272	0.204
	5	0.524	0.391
	6	0.707	0.506

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.909	0.035	0.056	0.872
	2/3	0.861	0.078	0.062	0.805
	3/4	0.877	0.078	0.046	0.831
	4/5	0.898	0.065	0.037	0.860
	5/6	0.955	0.026	0.018	0.931

Table 8.7.2Eiii

Accuracy and Consistency of Classification Indices: Read (Grade 11) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.570	0.462	0.316
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.711	0.542
	2	0.643	0.551
	3	0.484	0.385
	4	0.268	0.198
	5	0.452	0.333
	6	0.782	0.593

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.910	0.033	0.057	0.873
	2/3	0.862	0.077	0.061	0.807
	3/4	0.880	0.075	0.045	0.836
	4/5	0.904	0.057	0.039	0.865
	5/6	0.950	0.034	0.016	0.926

Table 8.7.2Eiv

Accuracy and Consistency of Classification Indices: Read (Grade 12) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.591	0.479	0.326
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.669	0.494
	2	0.672	0.587
	3	0.503	0.401
	4	0.299	0.221
	5	0.532	0.400
	6	0.781	0.586

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.909	0.034	0.057	0.872
	2/3	0.867	0.074	0.060	0.812
	3/4	0.889	0.075	0.036	0.850
	4/5	0.913	0.052	0.035	0.878
	5/6	0.964	0.025	0.012	0.946

8.7.3 Writing 9–12

Figure 8.7.3A
Scale Scores: Writ 9-12 S401 Paper

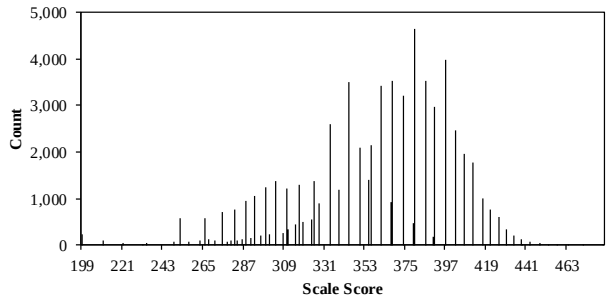


Table 8.7.3A
Scale Score Descriptive Statistics: Writ 9-12 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	22,486	199	472	351.11	44.65
10	18,203	210	472	356.41	41.32
11	14,658	221	483	362.32	39.77
12	9,739	232	472	368.25	36.64
Total	65,086	199	483	357.68	41.95

Figure 8.7.3B
Proficiency Level: Writ 9-12 S401 Paper

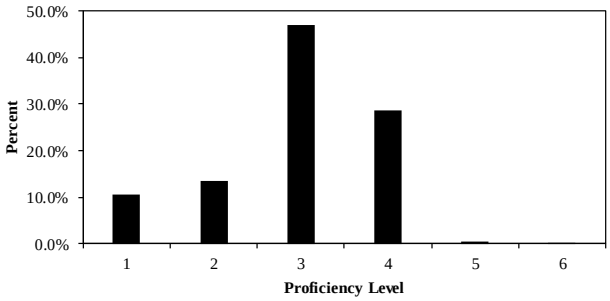


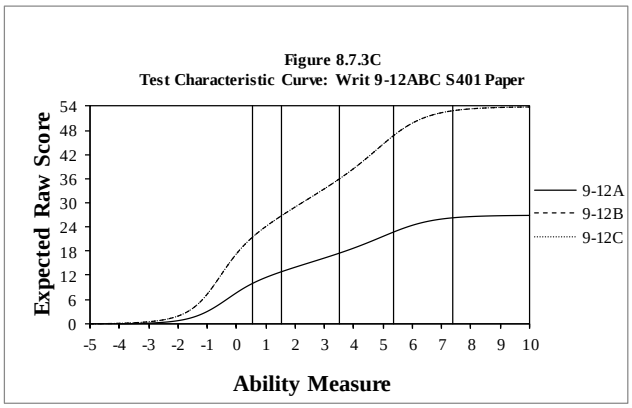
Table 8.7.3B
Proficiency Level Distribution: Writ 9-12 S401 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,479	11.0%	1,738	9.5%	1,580	10.8%	1,110	11.4%	6,907	10.6%
2	3,038	13.5%	2,637	14.5%	2,156	14.7%	910	9.3%	8,741	13.4%
3	9,359	41.6%	8,512	46.8%	6,801	46.4%	5,849	60.1%	30,521	46.9%
4	7,416	33.0%	5,246	28.8%	4,067	27.7%	1,850	19.0%	18,579	28.5%
5	192	0.9%	70	0.4%	54	0.4%	20	0.2%	336	0.5%
6	2	0.0%	0	0.0%	0	0.0%	0	0.0%	2	0.0%
Total	22,486	100.0%	18,203	100.0%	14,658	100.0%	9,739	100.0%	65,086	100.0%

Table 8.7.3C

Conditional Standard Error of Measurement at Cut Scores: Writ 9-12 S401 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	9	289	12.35	8.32
	10	298	12.08	8.59
	11	308	12.89	9.53
	12	318	14.23	10.74
2/3	9	319	14.23	10.77
	10	326	15.31	11.28
	11	335	16.38	11.81
	12	344	17.02	12.32
3/4	9	378	17.72	12.62
	10	385	17.72	12.35
	11	391	17.45	12.22
	12	398	17.18	12.08
4/5	9	430	15.47	10.74
	10	436	15.31	10.74
	11	441	15.57	11.01
	12	447	15.84	11.28
5/6	9	469	19.33	14.77
	10	479	22.29	17.45
	11	490	27.12	21.48
	12	501	33.03	26.42



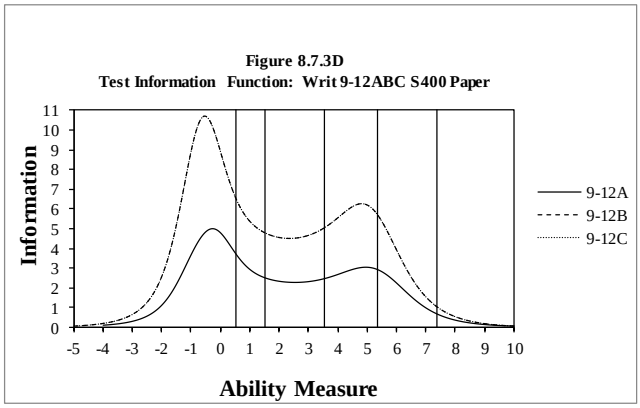


Table 8.7.3D

Weighted Reliability: Writ 9-12 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	16,430	0.850	0.898
B/C	48,656	0.914	

Table 8.7.3Ei

Accuracy and Consistency of Classification Indices: Writ (Grade 9) S401 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.780	0.697		0.557	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.851		0.754	
	2	0.641		0.517	
	3	0.781		0.701	
	4	0.810		0.746	
	5	-		0.065	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.963	0.015	0.022	0.947
	2/3	0.937	0.025	0.039	0.911
	3/4	0.889	0.055	0.056	0.843
	4/5	0.991	0.009	0.000	0.989
	5/6	1.000	0.000	0.000	1.000

Table 8.7.3Eii

Accuracy and Consistency of Classification Indices: Writ (Grade 10) S401 Paper

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.786	0.704		0.556	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.827		0.716	
	2	0.660		0.535	
	3	0.800		0.734	
	4	0.809		0.732	
	5	-		0.065	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.964	0.016	0.020	0.948
	2/3	0.932	0.026	0.042	0.906
	3/4	0.893	0.056	0.051	0.850
	4/5	0.996	0.004	0.000	0.995
	5/6	N/A	N/A	N/A	N/A

Table 8.7.3Eiii

Accuracy and Consistency of Classification Indices: Writ (Grade 11) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.784	0.701	0.555
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.842	0.739
	2	0.655	0.532
	3	0.798	0.731
	4	0.802	0.723
	5	-	0.076
	6	N/A	N/A

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.962	0.016	0.022	0.945
	2/3	0.932	0.026	0.041	0.905
	3/4	0.892	0.057	0.051	0.849
	4/5	0.996	0.004	0.000	0.996
	5/6	N/A	N/A	N/A	N/A

Table 8.7.3Eiv

Accuracy and Consistency of Classification Indices: Writ (Grade 12) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.796	0.716	0.529
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.897	0.819
	2	0.521	0.391
	3	0.872	0.824
	4	0.684	0.573
	5	-	0.949
	6	N/A	N/A

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.967	0.010	0.023	0.954
	2/3	0.942	0.032	0.026	0.915
	3/4	0.888	0.049	0.064	0.843
	4/5	0.998	0.002	0.000	0.998
	5/6	N/A	N/A	N/A	N/A

8.7.4 Speaking 9–12

Figure 8.7.4A
Scale Scores: Spek 9-12 S401 Paper

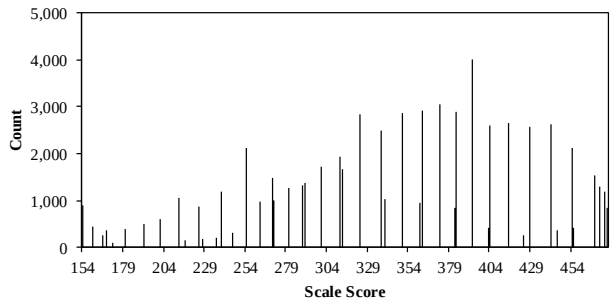


Table 8.7.4A
Scale Score Descriptive Statistics: Spek 9-12 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	22,491	154	468	341.64	84.90
10	18,214	160	471	346.15	79.82
11	14,663	166	474	352.32	76.59
12	9,742	172	476	364.56	69.54
Total	65,110	154	476	348.73	79.85

Figure 8.7.4B
Proficiency Level: Spek 9-12 S401 Paper

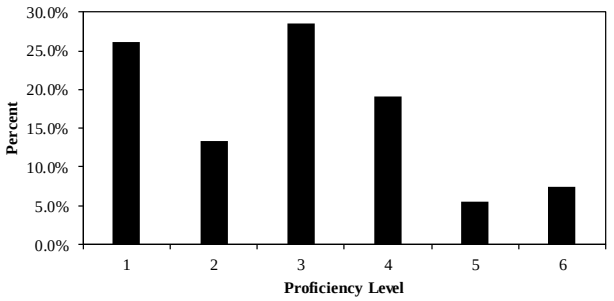


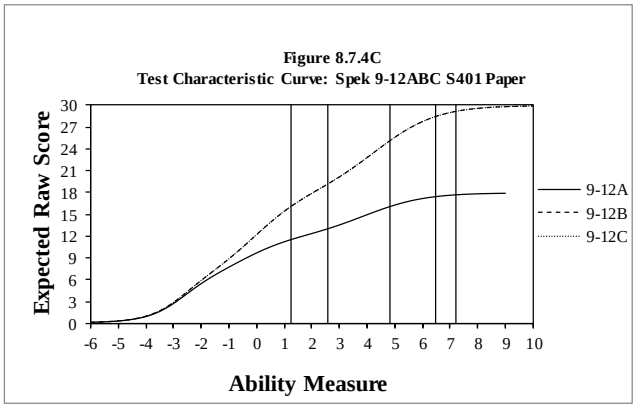
Table 8.7.4B
Proficiency Level Distribution: Spek 9-12 S401 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	6,246	27.8%	5,102	28.0%	3,598	24.5%	2,023	20.8%	16,969	26.1%
2	2,910	12.9%	2,263	12.4%	1,940	13.2%	1,623	16.7%	8,736	13.4%
3	5,654	25.1%	4,649	25.5%	4,849	33.1%	3,416	35.1%	18,568	28.5%
4	4,280	19.0%	4,205	23.1%	2,526	17.2%	1,422	14.6%	12,433	19.1%
5	1,871	8.3%	701	3.8%	573	3.9%	414	4.2%	3,559	5.5%
6	1,530	6.8%	1,294	7.1%	1,177	8.0%	844	8.7%	4,845	7.4%
Total	22,491	100.0%	18,214	100.0%	14,663	100.0%	9,742	100.0%	65,110	100.0%

Table 8.7.4C

Conditional Standard Error of Measurement at Cut Scores: Spek 9-12 S401 Paper

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	9	290	25.15	16.96
	10	295	26.03	17.55
	11	299	26.62	17.84
	12	302	26.91	18.13
2/3	9	328	27.79	19.60
	10	333	27.20	19.60
	11	337	26.91	19.60
	12	340	26.62	19.30
3/4	9	385	24.28	17.26
	10	393	24.86	17.26
	11	400	25.74	17.55
	12	406	26.32	17.84
4/5	9	440	36.27	22.52
	10	446	38.90	23.98
	11	451	41.82	25.15
	12	455	43.87	26.62
5/6	9	468	52.94	31.29
	10	471	55.57	32.46
	11	474	57.91	33.93
	12	476	59.67	35.10



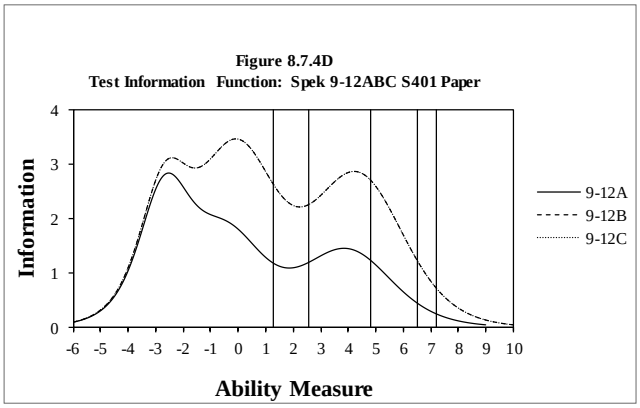


Table 8.7.4D

Weighted Reliability: Spek 9-12 S401 Paper

Tiers	No. of Students	Reliability	Weighted Reliability
A	16,438	0.900	0.927
B/C	48,672	0.937	

Table 8.7.4Ei

Accuracy and Consistency of Classification Indices: Spek (Grade 9) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.653	0.566	0.453
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.883	0.822
	2	0.478	0.369
	3	0.698	0.593
	4	0.577	0.463
	5	0.377	0.338
	6	-	0.410

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.934	0.032	0.033	0.906
	2/3	0.916	0.042	0.042	0.884
	3/4	0.923	0.029	0.048	0.890
	4/5	0.923	0.055	0.022	0.892
	5/6	0.932	0.068	0.000	0.931

Table 8.7.4Eii

Accuracy and Consistency of Classification Indices: Spek (Grade 10) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.682	0.598	0.489
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.895	0.837
	2	0.477	0.365
	3	0.695	0.593
	4	0.747	0.630
	5	0.227	0.198
	6	-	0.685

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.935	0.029	0.037	0.907
	2/3	0.924	0.039	0.037	0.892
	3/4	0.921	0.042	0.038	0.889
	4/5	0.961	0.021	0.018	0.935
	5/6	0.929	0.071	0.000	0.934

Table 8.7.4Eiii

Accuracy and Consistency of Classification Indices: Spek (Grade 11) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.679	0.603	0.494
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.881	0.812
	2	0.483	0.370
	3	0.776	0.692
	4	0.681	0.554
	5	0.225	0.214
	6	-	0.796

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.937	0.029	0.034	0.910
	2/3	0.916	0.046	0.038	0.882
	3/4	0.926	0.032	0.042	0.894
	4/5	0.970	0.019	0.012	0.951
	5/6	0.920	0.080	0.000	0.933

Table 8.7.4Eiv

Accuracy and Consistency of Classification Indices: Spek (Grade 12) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.682	0.610	0.503
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.851	0.773
	2	0.574	0.456
	3	0.799	0.722
	4	0.667	0.540
	5	0.237	0.237
	6	-	0.823

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.936	0.031	0.033	0.909
	2/3	0.916	0.041	0.043	0.882
	3/4	0.937	0.025	0.038	0.908
	4/5	0.973	0.018	0.009	0.959
	5/6	0.913	0.087	0.000	0.928

8.7.5 Oral Language Composite 9–12

Figure 8.7.5A
Scale Scores: Oral 9-12 S401 Paper

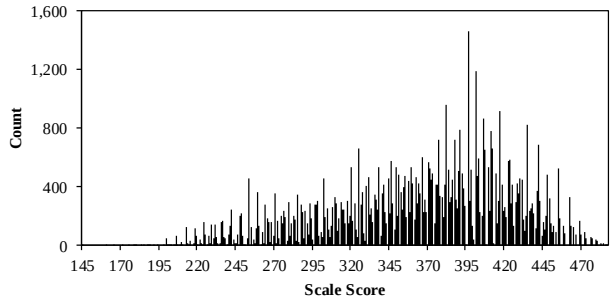


Table 8.7.5A
Scale Score Descriptive Statistics: Oral 9-12 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	22,483	145	483	358.58	63.41
10	18,209	183	484	363.76	58.56
11	14,657	171	486	368.52	55.94
12	9,737	160	487	377.75	49.55
Total	65,086	145	487	365.13	58.84

Figure 8.7.5B
Proficiency Level: Oral 9-12 S401 Paper

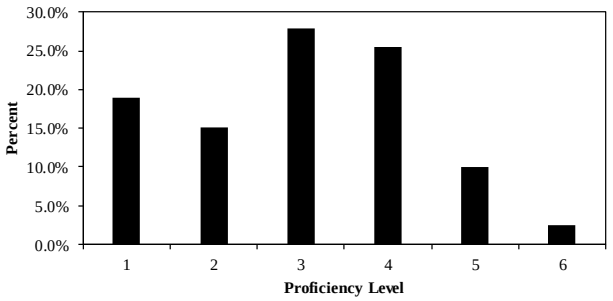


Table 8.7.5B
Proficiency Level Distribution: Oral 9-12 S401 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	4,744	21.1%	3,558	19.5%	2,706	18.5%	1,371	14.1%	12,379	19.0%
2	3,050	13.6%	2,806	15.4%	2,430	16.6%	1,514	15.5%	9,800	15.1%
3	5,271	23.4%	4,934	27.1%	4,430	30.2%	3,503	36.0%	18,138	27.9%
4	6,106	27.2%	4,641	25.5%	3,488	23.8%	2,357	24.2%	16,592	25.5%
5	2,664	11.8%	1,774	9.7%	1,270	8.7%	813	8.3%	6,521	10.0%
6	648	2.9%	496	2.7%	333	2.3%	179	1.8%	1,656	2.5%
Total	22,483	100.0%	18,209	100.0%	14,657	100.0%	9,737	100.0%	65,086	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.50	2329.153	0.657
Speaking	0.50	6375.280	0.927
Oral		3460.810	0.909

*Variances from students who had results in all four domains

Table 8.7.5Ei

Accuracy and Consistency of Classification Indices: Oral (Grade 9) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.663	0.556	0.442		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.884	0.817		
	2	0.533	0.413		
	3	0.647	0.530		
	4	0.657	0.552		
	5	0.522	0.422		
	6	-	0.189		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.944	0.023	0.033	0.920
	2/3	0.926	0.039	0.036	0.894
	3/4	0.905	0.046	0.049	0.867
	4/5	0.909	0.048	0.043	0.873
	5/6	0.971	0.029	0.000	0.958

Table 8.7.5Eii

Accuracy and Consistency of Classification Indices: Oral (Grade 10) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.676	0.570	0.457
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.871	0.792
	2	0.559	0.440
	3	0.671	0.565
	4	0.677	0.573
	5	0.534	0.427
	6	-	0.304

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.942	0.024	0.035	0.918
	2/3	0.919	0.039	0.042	0.885
	3/4	0.904	0.050	0.046	0.866
	4/5	0.933	0.040	0.027	0.904
	5/6	0.973	0.027	0.000	0.964

Table 8.7.5Eiii

Accuracy and Consistency of Classification Indices: Oral (Grade 11) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.691	0.585	0.470
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.855	0.768
	2	0.573	0.455
	3	0.698	0.600
	4	0.685	0.580
	5	0.584	0.456
	6	0.691	0.392

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.940	0.025	0.034	0.915
	2/3	0.914	0.041	0.045	0.879
	3/4	0.907	0.050	0.044	0.869
	4/5	0.946	0.034	0.020	0.922
	5/6	0.980	0.018	0.002	0.973

Table 8.7.5Eiv

Accuracy and Consistency of Classification Indices: Oral (Grade 12) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.719	0.616	0.497

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.846	0.747
	2	0.585	0.464
	3	0.756	0.674
	4	0.704	0.605
	5	0.654	0.518
	6	0.728	0.443

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.950	0.020	0.030	0.929
	2/3	0.919	0.038	0.043	0.885
	3/4	0.910	0.047	0.043	0.873
	4/5	0.953	0.033	0.014	0.934
	5/6	0.986	0.012	0.002	0.980

8.7.6 Literacy Composite 9–12

Figure 8.7.6A
Scale Scores: Litr 9-12 S401 Paper

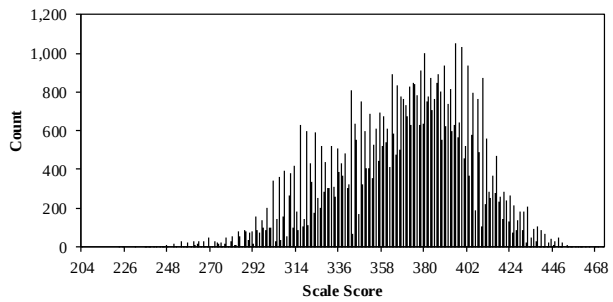


Table 8.7.6A
Scale Score Descriptive Statistics: Litr 9-12 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	22,477	204	472	362.58	36.06
10	18,200	231	468	367.53	33.80
11	14,653	223	469	372.88	32.78
12	9,738	248	465	377.49	30.01
Total	65,068	204	472	368.51	34.27

Figure 8.7.6B
Proficiency Level: Litr 9-12 S401 Paper

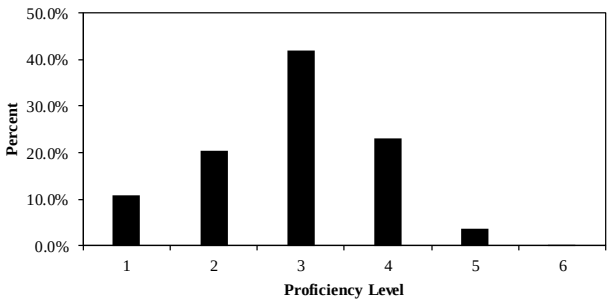


Table 8.7.6B
Proficiency Level Distribution: Litr 9-12 S401 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,669	11.9%	1,883	10.3%	1,490	10.2%	915	9.4%	6,957	10.7%
2	4,323	19.2%	3,695	20.3%	3,158	21.6%	2,062	21.2%	13,238	20.3%
3	8,529	37.9%	7,700	42.3%	6,272	42.8%	4,812	49.4%	27,313	42.0%
4	5,937	26.4%	4,219	23.2%	3,178	21.7%	1,715	17.6%	15,049	23.1%
5	970	4.3%	688	3.8%	548	3.7%	234	2.4%	2,440	3.7%
6	49	0.2%	15	0.1%	7	0.0%	0	0.0%	71	0.1%
Total	22,477	100.0%	18,200	100.0%	14,653	100.0%	9,738	100.0%	65,068	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 S401 Paper

Component	Weight	Variance	Reliability
Reading	0.50	1057.967	0.806
Writing	0.50	1759.854	0.898
Literacy		1174.139	0.918

*Variances from students who had results in all four domains

Table 8.7.6Ei

Accuracy and Consistency of Classification Indices: Litr (Grade 9) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.769	0.676	0.559		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.864	0.774		
	2	0.708	0.602		
	3	0.783	0.709		
	4	0.771	0.688		
	5	0.625	0.432		
	6	-	0.630		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.960	0.015	0.025	0.945
	2/3	0.929	0.033	0.038	0.900
	3/4	0.914	0.047	0.039	0.880
	4/5	0.967	0.023	0.010	0.952
	5/6	0.998	0.002	0.000	0.998

Table 8.7.6Eii

Accuracy and Consistency of Classification Indices: Litr (Grade 10) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.781	0.692	0.570
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.856	0.759
	2	0.721	0.617
	3	0.807	0.744
	4	0.763	0.672
	5	0.702	0.492
	6	-	0.935

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.963	0.013	0.023	0.949
	2/3	0.925	0.035	0.039	0.895
	3/4	0.918	0.045	0.037	0.885
	4/5	0.973	0.019	0.008	0.962
	5/6	0.999	0.001	0.000	0.999

Table 8.7.6Eiii

Accuracy and Consistency of Classification Indices: Litr (Grade 11) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.784	0.697	0.576
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.850	0.750
	2	0.729	0.628
	3	0.809	0.747
	4	0.767	0.673
	5	0.756	0.555
	6	-	0.974

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.963	0.014	0.023	0.948
	2/3	0.922	0.037	0.041	0.891
	3/4	0.922	0.044	0.034	0.891
	4/5	0.976	0.017	0.007	0.967
	5/6	1.000	0.000	0.000	1.000

Table 8.7.6Eiv

Accuracy and Consistency of Classification Indices: Litr (Grade 12) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.799	0.719	0.584

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.857	0.756
	2	0.733	0.631
	3	0.848	0.798
	4	0.722	0.628
	5	0.777	0.502
	6	N/A	N/A

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.967	0.012	0.021	0.953
	2/3	0.924	0.038	0.038	0.892
	3/4	0.928	0.037	0.035	0.897
	4/5	0.980	0.018	0.002	0.976
	5/6	N/A	N/A	N/A	N/A

8.7.7 Comprehension Composite 9–12

Figure 8.7.7A
Scale Scores: Cphn 9-12 S401 Paper

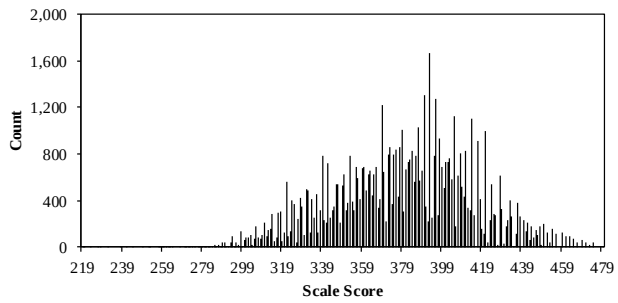


Table 8.7.7A
Scale Score Descriptive Statistics: Cphn 9-12 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	22,484	223	480	374.03	36.14
10	18,208	249	480	379.05	34.40
11	14,657	219	475	383.37	33.65
12	9,740	220	480	387.57	30.99
Total	65,089	219	480	379.56	34.71

Figure 8.7.7B
Proficiency Level: Cphn 9-12 S401 Paper

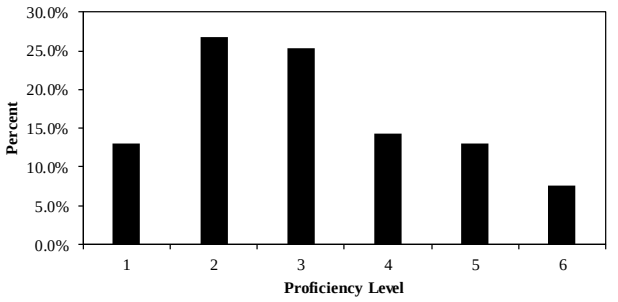


Table 8.7.7B
Proficiency Level Distribution: Cphn 9-12 S401 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,276	14.6%	2,292	12.6%	1,900	13.0%	1,040	10.7%	8,508	13.1%
2	5,494	24.4%	4,915	27.0%	4,163	28.4%	2,889	29.7%	17,461	26.8%
3	5,256	23.4%	4,709	25.9%	3,807	26.0%	2,699	27.7%	16,471	25.3%
4	3,311	14.7%	2,659	14.6%	1,851	12.6%	1,462	15.0%	9,283	14.3%
5	3,304	14.7%	2,174	11.9%	1,847	12.6%	1,133	11.6%	8,458	13.0%
6	1,843	8.2%	1,459	8.0%	1,089	7.4%	517	5.3%	4,908	7.5%
Total	22,484	100.0%	18,208	100.0%	14,657	100.0%	9,740	100.0%	65,089	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.30	2329.153	0.657
Reading	0.70	1057.967	0.806
Comprehension		1203.924	0.857

*Variances from students who had results in all four domains

Table 8.7.7Ei

Accuracy and Consistency of Classification Indices: Cphn (Grade 9) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.591	0.484	0.365
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.827	0.697
	2	0.630	0.522
	3	0.536	0.431
	4	0.403	0.311
	5	0.534	0.412
	6	0.753	0.566

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.934	0.021	0.045	0.908
	2/3	0.891	0.055	0.054	0.845
	3/4	0.881	0.069	0.050	0.837
	4/5	0.902	0.058	0.040	0.864
	5/6	0.952	0.031	0.017	0.929

Table 8.7.7Eii

Accuracy and Consistency of Classification Indices: Cphn (Grade 10) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.602	0.492	0.366
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.796	0.647
	2	0.660	0.555
	3	0.565	0.462
	4	0.423	0.323
	5	0.493	0.370
	6	0.790	0.605

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.937	0.022	0.042	0.911
	2/3	0.884	0.057	0.059	0.838
	3/4	0.884	0.067	0.049	0.840
	4/5	0.914	0.050	0.036	0.877
	5/6	0.956	0.031	0.013	0.937

Table 8.7.7Eiii

Accuracy and Consistency of Classification Indices: Cphn (Grade 11) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.612	0.501	0.375
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.774	0.625
	2	0.659	0.560
	3	0.581	0.475
	4	0.398	0.299
	5	0.550	0.419
	6	0.806	0.631

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.931	0.025	0.044	0.903
	2/3	0.882	0.061	0.057	0.835
	3/4	0.894	0.060	0.046	0.852
	4/5	0.921	0.047	0.032	0.887
	5/6	0.961	0.028	0.011	0.945

Table 8.7.7Eiv

Accuracy and Consistency of Classification Indices: Cphn (Grade 12) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.627	0.514	0.381

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.762	0.602
	2	0.694	0.595
	3	0.596	0.493
	4	0.469	0.358
	5	0.569	0.434
	6	0.803	0.605

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.941	0.022	0.037	0.916
	2/3	0.882	0.058	0.060	0.835
	3/4	0.892	0.063	0.045	0.850
	4/5	0.927	0.043	0.029	0.896
	5/6	0.970	0.023	0.007	0.957

8.7.8 Overall Composite 9–12

Figure 8.7.8A
Scale Scores: Over 9-12 S401 Paper

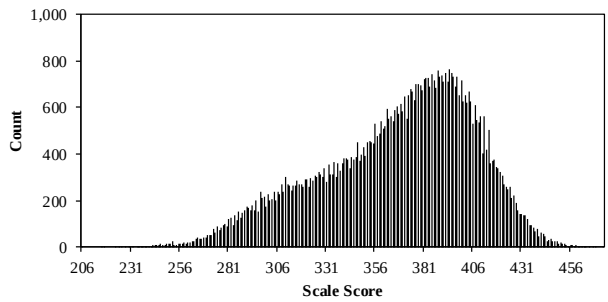


Table 8.7.8A
Scale Score Descriptive Statistics: Over 9-12 S401 Paper

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	22,466	206	473	361.16	42.24
10	18,198	225	466	366.19	39.02
11	14,649	216	469	371.35	37.37
12	9,730	235	468	377.37	32.99
Total	65,043	206	473	367.29	39.39

Figure 8.7.8B
Proficiency Level: Over 9-12 S401 Paper

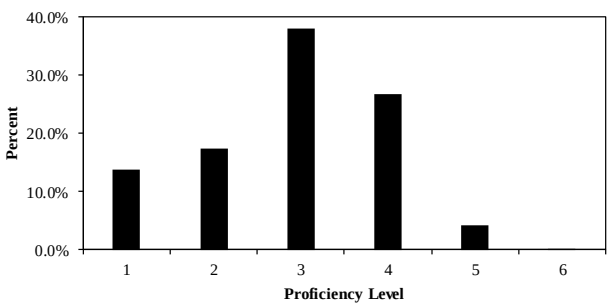


Table 8.7.8B
Proficiency Level Distribution: Over 9-12 S401 Paper

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,517	15.7%	2,527	13.9%	1,892	12.9%	951	9.8%	8,887	13.7%
2	3,541	15.8%	3,203	17.6%	2,708	18.5%	1,893	19.5%	11,345	17.4%
3	7,517	33.5%	6,834	37.6%	5,942	40.6%	4,444	45.7%	24,737	38.0%
4	6,667	29.7%	4,872	26.8%	3,564	24.3%	2,239	23.0%	17,342	26.7%
5	1,176	5.2%	743	4.1%	534	3.6%	202	2.1%	2,655	4.1%
6	48	0.2%	19	0.1%	9	0.1%	1	0.0%	77	0.1%
Total	22,466	100.0%	18,198	100.0%	14,649	100.0%	9,730	100.0%	65,043	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 S401 Paper

Component	Weight	Variance	Reliability
Listening	0.15	2329.153	0.657
Reading	0.35	1057.967	0.806
Speaking	0.15	6375.280	0.927
Writing	0.35	1759.854	0.898
Overall Composite		1551.732	0.951

*Variances from students who had results in all four domains

Table 8.7.8Ei

Accuracy and Consistency of Classification Indices: Over (Grade 9) S401 Paper

Overall Indices	Accuracy	Consistency	Kappa (k)
	0.800	0.722	0.629

Conditional on Level	Level	Accuracy	Consistency
	1	0.914	0.863
	2	0.724	0.622
	3	0.823	0.754
	4	0.788	0.730
	5	0.575	0.404
	6	-	0.750

Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.967	0.013	0.020	0.953
	2/3	0.950	0.026	0.024	0.929
	3/4	0.931	0.034	0.035	0.902
	4/5	0.954	0.032	0.014	0.938
	5/6	0.998	0.002	0.000	0.998

Table 8.7.8Eii

Accuracy and Consistency of Classification Indices: Over (Grade 10) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.816	0.742	0.651
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.898	0.845
	2	0.744	0.648
	3	0.838	0.779
	4	0.805	0.744
	5	0.699	0.484
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.966	0.014	0.020	0.954
	2/3	0.945	0.028	0.027	0.924
	3/4	0.933	0.035	0.032	0.907
	4/5	0.968	0.023	0.009	0.958
	5/6	0.999	0.001	0.000	1.000

Table 8.7.8Eiii

Accuracy and Consistency of Classification Indices: Over (Grade 11) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.824	0.752	0.659
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.891	0.831
	2	0.751	0.657
	3	0.848	0.794
	4	0.814	0.746
	5	0.747	0.560
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.967	0.013	0.019	0.955
	2/3	0.943	0.029	0.028	0.920
	3/4	0.935	0.034	0.030	0.910
	4/5	0.976	0.017	0.007	0.968
	5/6	0.999	0.001	0.000	1.000

Table 8.7.8Eiv

Accuracy and Consistency of Classification Indices: Over (Grade 12) S401 Paper

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**
	0.838	0.775	0.675

Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.886	0.816
	2	0.786	0.700
	3	0.869	0.823
	4	0.807	0.748
	5	0.765	0.492
	6	-	1.000

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.974	0.011	0.015	0.964
	2/3	0.944	0.027	0.028	0.921
	3/4	0.937	0.032	0.031	0.911
	4/5	0.982	0.017	0.001	0.979
	5/6	1.000	0.000	0.000	1.000

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.

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

Center for Applied Linguistics (2017). ACCESS for ELLs® 2.0 Speaking and Writing Score Scale Reconstruction: Technical Brief. Washington, D.C.: Author.

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. and MacGregor, D. (2017). *The ACCESS for ELLs 2.0 2016 Standard Setting Study* (Technical Report). Madison, WI: Board of Regents of the University of Wisconsin System.

Cronbach, L.J., Schonemann, P., & McKie, D. (1965). Alpha coefficients for stratified-parallel tests. *Educational and Psychological Measurement, 25,* 291–312.

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.

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

Lee, W., Hanson, B. A., & Brennan, R. L. (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement, 26,* 412–432.

Linacre, J. M. (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.

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

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).

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., 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.

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.

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement.* Chicago, IL: MESA Press.

Young, M. J., & Yoon, B. (1998, April). *Estimating the consistency and accuracy of classifications in a standards-referenced assessment* (CSE Tech. Rep. 475). Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education and Information Studies.

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.
Tanya Bitterman, M.A.
Claire Gellner, B.A.
Justin Kelly, Ph.D.
Dorry M. Kenyon, Ph.D.
Tristan Kirkman, M.A.
Daniel Lee, M.S.
Chen Li., M.S. Ed
Ashley Lipps, M.A., M.P.A.
Mohammed Louguit, Ph.D.
Michele Kawood, M.S.Ed.
Meg Montee, Ph.D.
Jennifer Norton, Ed.D.
Lin Pan, M.S.
Mary Spanarkel, M.A.
Jing Wei, Ph.D.
Yu-Chia Wu, M.Ed.
Shu Jing Yen, Ph.D.
Xin Yu, M.A.
Jiaqi Zheng, M.S.