World-Class Instructional Design and Assessment

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

Annual Technical Report No. 13A

Prepared by:

Center for Applied Linguistics

Language Assessment Division
Psychometrics and Quantitative Research Team

April 20, 2018

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 as ACCESS for ELLs, but the assessment is now 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 the 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 13A) provides technical information pertaining to ACCESS for ELLs 2.0 Series 401 Online. A second report (WIDA ACCESS Technical Report 13B) provides technical information for the ACCESS for ELLs 2.0 Series 401 Paper assessment, including the Kindergarten 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

Summary Highlights

This report presents a wealth of data documenting the technical properties of ACCESS for ELLs 2.0 Series 401 Online, 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. The annual analyses of the technical properties of ACCESS for ELLs test forms are used in the continual refinement and improvement of ACCESS for ELLs.

Argument-based validation framework for ACCESS for ELLs

The purpose of this report is to provide evidence for the validity of the online version of ACCESS for ELLs 2.0 (hereafter ACCESS 2.0 Online), when used for its intended purposes. This report is structured using a validation approach developed at the Center for Applied Linguistics (CAL), and based on Bachman and Palmer's (2010) Assessment Use Argument, integrated with the Evidence Centered Design principles outlined in Mislevy, Almond, and Lukas (2004). CAL's validation framework, shown in Figure 2 of Part I of this report, consists of 7 steps, leading the line of argumentation from *Plan* (Step 7) through *Consequences* (Step 1).

Part I of this report, *Foundations*, covers Steps 7–5 of CAL's Validation Framework (*Plan*, *Design*, and *Assessment Performance*).

Part II of the report, *Assessment Records* covers Step 4 in the Validation Framework. Part II has three subsections:

> II:1 Assessment Records for ACCESS 2.0 Online
>
> II:2 Background and Descriptions for the Presentation of Results
>
> II:3 Results by Grade Cluster

The first subsection provides the Assessment Use Argument (AUA), a set of claims which allow stakeholders to better interpret and use ACCESS for ELLs. These claims are each supported by evidence, much of which is drawn from the tables and figures presented in this report. The second subsection provides technical detail on the analyses conducted to provide evidence for the AUA claims, and the third subsection contains the tables and figures with the results of that analysis.

Demographic data

The Series 401 Online data set included the results of 1,231,359 students. The largest grade was Grade 3 with 182,698 students, while the smallest was Grade 12 with 33,370 students. Of the participating WIDA states, the largest was Illinois with 130,134 students, while the smallest was the United States Virgin Islands with 1,023 students.

Reliability and accuracy data

For most test users, the Overall Composite proficiency score, based on performances in Listening, Reading, Writing and Speaking, 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 Crohnbach's alpha) of the Overall Composite score for Series 401 Online, presented in Table C of Section 3.4 (see also section 2.3.3. of Part II) is very high across all grade-level clusters. For Grade 1 it was .937; for Grades 2–3, .947; for Grades 4–5, .950; for Grades 6–8, .959; and for Grades 9–12, .950.

Likewise, as Table 0.2 shows, the accuracy of classification for 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)
1	0.970	0.923	0.952	0.988	0.998
2	0.977	0.936	0.939	0.979	0.999
3	0.980	0.949	0.920	0.959	0.999
4	0.982	0.961	0.917	0.947	0.994
5	0.978	0.956	0.921	0.930	0.994
6	0.970	0.942	0.944	0.982	0.999
7	0.966	0.941	0.941	0.974	0.998
8	0.965	0.944	0.938	0.968	0.998
9	0.958	0.938	0.942	0.981	0.998
10	0.959	0.935	0.949	0.987	0.999
11	0.956	0.932	0.955	0.988	0.999
12	0.956	0.929	0.961	0.991	1.000

Series 401 Online: 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. Data from the state of Michigan were received 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 section 3.3 and 3.4) 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 84,524 of 1,231,539 total students in the ACCESS Online population, a proportion of 6.86%. 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 section 3.2).

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.

Center for Applied Linguistics (2016). ACCESS for ELLs® Series 400 Listening and Reading Scale Maintenance: Technical Brief. (WIDA Consortium).

> This brief summarizes the results of two scaling studies which were conducted to ensure that scores on ACCESS 2.0 in the domains of Listening and Reading remained on the original ACCESS score scale as the test transitioned from paper-based format to online delivery.

Center for Applied Linguistics (2017). ACCESS for ELLs® 2.0 Speaking and Writing Score Scale Reconstruction: Technical Brief. (WIDA Consortium).

> The purpose of this report is to summarize the results of two scaling studies that were conducted to the reconstruct ACCESS Speaking and Writing score scales as the transition was made to ACCESS 2.0 in the domains of Speaking and Writing.

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.

> This report summarizes the processes and procedures employed to set the new ACCESS 2.0 cut scores in the summer of 2016.

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).

Center for Applied Linguistics (2017a). *Annual Technical Report for ACCESS for ELLs® 2.0 Online English Language Proficiency Test, Series 400, 2015–2016 Administration* (WIDA Consortium Annual Technical Report No. 12A).

Center for Applied Linguistics (2017b). *Annual Technical Report for ACCESS for ELLs® 2.0 Paper English Language Proficiency Test, Series 400, 2015–2016 Administration* (WIDA Consortium Annual Technical Report No. 12B).

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.

Table of Contents

Part I

4 Assessment Performance: The Implementation of ACCESS 2.0 Online 24

Part II

Part I: Foundations

ACCESS for ELLs 2.0 Online is a secure, large-scale English language proficiency assessment administered to students in grades 1–12 who have been identified as English language learners (ELLs). It is administered annually in WIDA Consortium member states to monitor students' progress in acquiring academic English. ACCESS 2.0 Online is a standards-based English language proficiency test designed to measure both social and academic language proficiency of ELLs in English in a school context. It assesses social and instructional language, as well as the academic language associated with language arts, mathematics, science, and social studies, across the four language domains (Listening, Reading, Writing, and Speaking).

1 The Validation Framework for ACCESS 2.0 Online

1.1 Development of the Validation Framework for ACCESS 2.0 Online

As with any assessment, an important consideration during the development of ACCESS 2.0 was determining how to establish 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 consider the evidence that supports the interpretations and decisions made about students 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 that connect test design and administration to intended and actual score interpretation and consequences. The present section 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 Online.

An argument-based approach to the ACCESS 2.0 Online 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 Online). Specifically, tables and figures from this report explicitly address questions related to assessment data. Chapelle, Enright, and Jamieson (2010) support using such a structure for presenting 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 validity argument for ACCESS 2.0 Online shows the path from test design to student 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. Evidence includes results from analyses of test data, outside documentation, and other resources. In the validation argument, this process of identifying evidence to support claims encompasses the entire testing process, from the commencement of test design to the consequences of test use (Bachman & Palmer, 2010; Llosa, 2008); Figure 1 shows the process by which evidence supports validation actions, which are used to establish larger claims about ACCESS 2.0 Online. The figure shows the generic structure of the line of argumentation for validity.
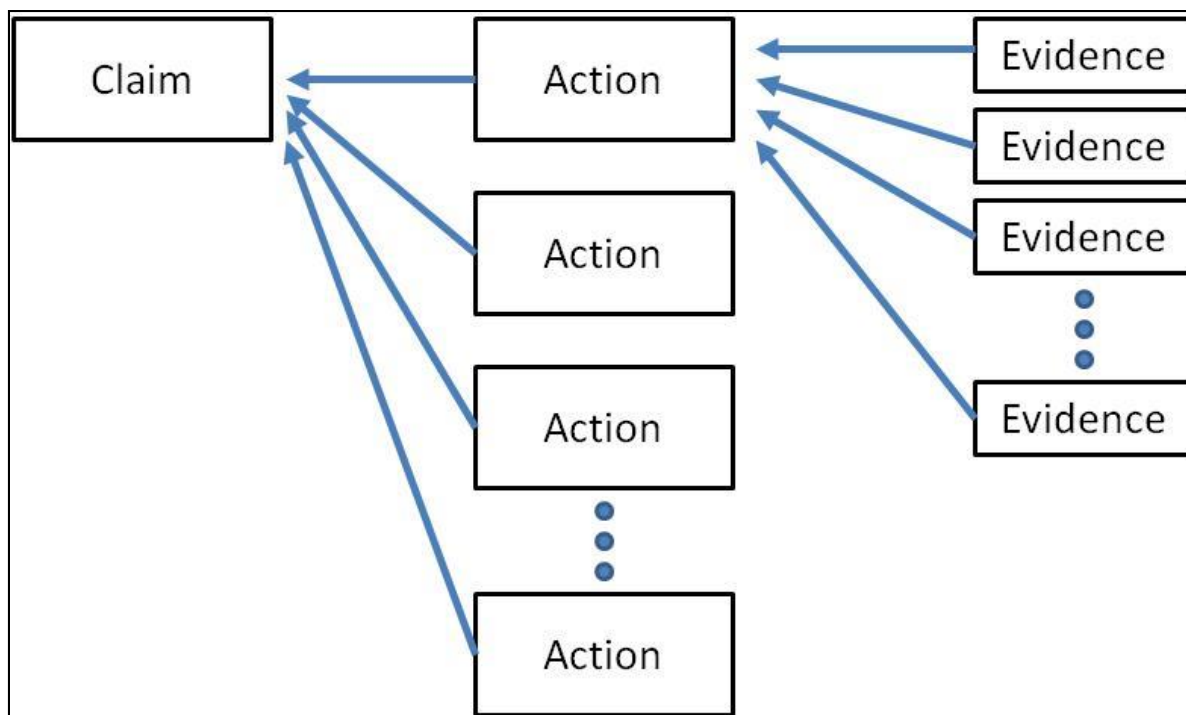
Figure 1. General Argument Structure for Assessment Validation (simplified from Toulmin, 2003).

1.2 Description of the Framework

The generic validation framework applied to the entire ACCESS 2.0 Online 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, combines models for both test development (i.e., Evidence-Centered Design (ECD) [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 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 going from bottom to top are numbered from seven to one. For example, during the initial *Plan* step (Step 7), test developers state the anticipated decisions and consequences of implementing the assessment program, which then drive the development and implementation of the assessment (Steps 6 through 4). Assessment results are then used to formulate Interpretations (Step 3) and to make Decisions (Step 2). Consequences (Step 1) represents the culmination of all previous steps. This structure highlights the fact that any weakness in a lower step affects the steps above it.

In CAL's Validation Framework, *Plan* involves an examination of possible decisions that state educational agencies might make and consequences that might result from use of 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 students' 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*).

Part I of this report presents evidence regarding the Planning, Designing, and Operationalization of the test, while information related to *Assessment Records* is found in Part II.
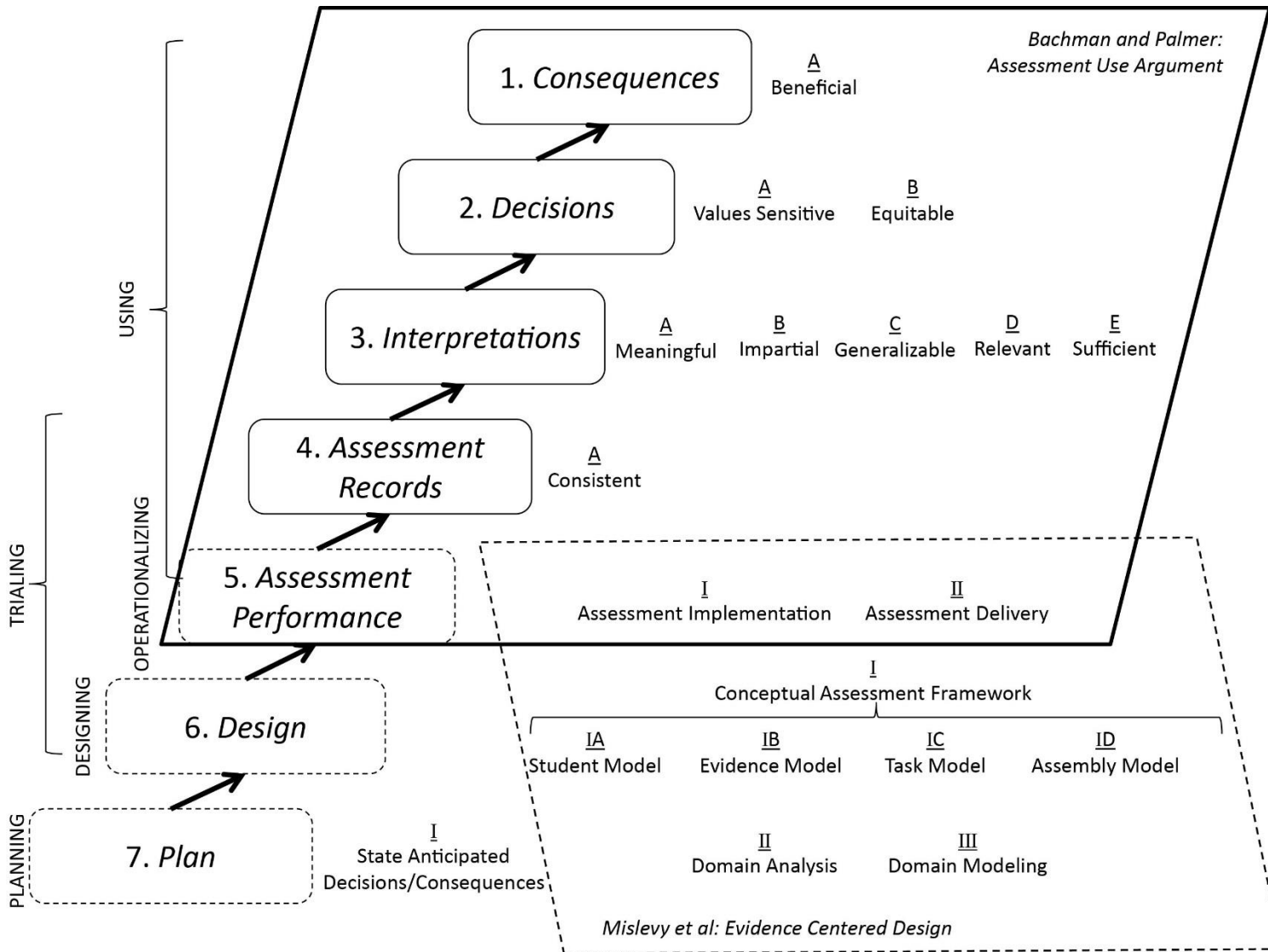
Figure 2. CAL's Validation Framework (based on Bachman & Palmer, 2010; Mislevy, Almond, & Lukas, 2004).

2   The Plan for ACCESS 2.0 Online

This section focuses on *Plan* (Step 7) of CAL's Validation Framework. This section details the decisions that the test is intended to inform, along with the consequences of those decisions. It then describes the domain analysis and modeling that undergirds WIDA's conceptualization of academic English language proficiency.

2.1   Purpose Statement: What are the intended decisions and consequences of using ACCESS?

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 Online, described in this report, and ACCESS 2.0 Paper, described in a companion report.

2.2 Domain Analysis: What is WIDA's conceptualization of the development of academic English language proficiency?

The Domain Analysis aspect of the Plan step in CAL's Validation Framework defines what ACCESS 2.0 is assessing as a measure of English language proficiency. In ECD (Mislevy Almond, & Lukas, 2004), Domain Analysis involves compiling and synthesizing all of the relevant information about what will be assessed, namely, academic language proficiency. WIDA's conceptualization of academic English language proficiency is encapsulated in the 2012 Amplification of the ELD Standards (WIDA, 2012), which built upon previous editions of the WIDA ELD Standards (Gottlieb, 2004; WIDA, 2007). The five WIDA ELD Standards form the basis of this conceptualization. In order to capture the language development of ELLs, the Standards include the following layers of organization: Grade-level clusters, Language Domains, and Language Proficiency Levels. Domain Analysis therefore also incorporates more granular information about the characteristics of a task and/or response for these various organizational levels.

2.2.1 The WIDA Standards

The five foundational WIDA ELD Standards, which inform the design, structure, and content of ACCESS 2.0 Online, are:

- *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 2.0 Online targets at least one of these five Standards. In the cases of some test items and tasks, the Standards are combined as follows:

- Integrated Language of Science (LoSC), Language of Language Arts (LoLA), and Language of Social Studies (LoSS): IT

- Language of Math (LoMA) and Language of Science (LoSC): MS

- Language of Language Arts (LoLA) and Language of Social Studies (LoSS): LS

2.2.2 Grade-Level Clusters

The WIDA ELD Standards describe developing English language proficiency within six grade-level clusters. These are K, 1, 2–3, 4–5, 6–8, and 9–12. Test forms follow this grade-level clustering.

2.2.3 Language Domains

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

2.2.4 Language Proficiency Levels

The WIDA ELD Standards describe the continuum of language development via five language proficiency levels (PLs) that are fully delineated in the WIDA ELD Standards document (WIDA 2012), with scores indicating progression through each level. These levels are *Entering*, *Emerging*, *Developing*, *Expanding*, and *Bridging*. There is also a final stage known as *Reaching* which is used to describe students who have progressed across the entire WIDA English language proficiency continuum; as such, scores do not indicate progression through this level. The proficiency levels are shown graphically in Figure 3.
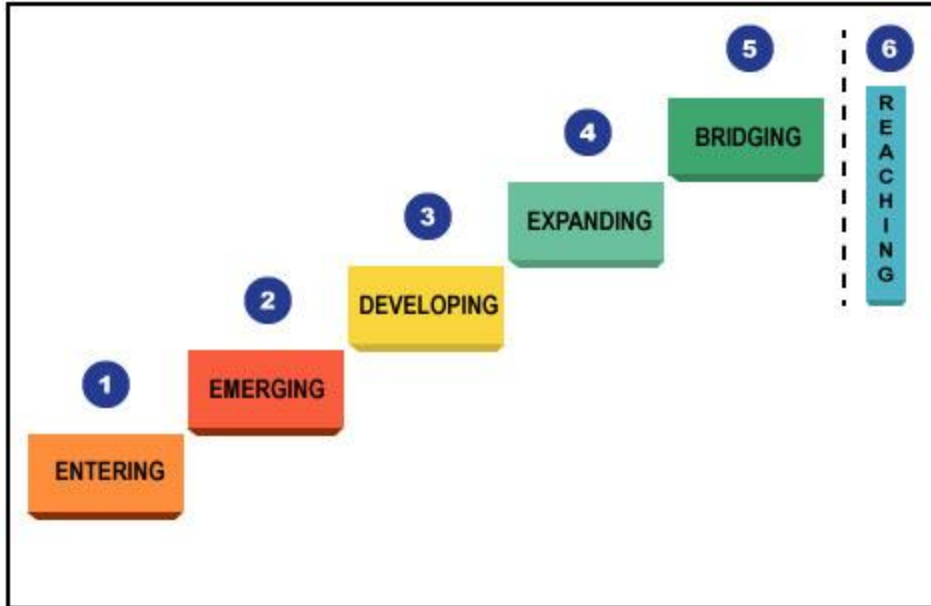
Figure 3. The Language Proficiency Levels of the WIDA ELD Standards.

2.3 Domain Modeling: How are the components of the ACCESS assessment program interrelated?

The Domain Model aspect of the *Plan* step in CAL's Validation Framework formulates the argument between the evidence collected about the test taker and the intended inferences to be made about them. In other words, in the Domain Model, we ask what evidence is necessary and sufficient to make the target inferences. In the case of ACCESS 2.0 Online, evidence is collected in order to make inferences about the test takers' language proficiency. This argument has been operationalized within ACCESS 2.0 Online in terms of the Model Performance Indicator (MPI). The MPIs convey what the test taker should be able to do with language. Each MPI is mapped to a PL, providing examples of how a test taker at each level would accomplish the task. This Domain Model serves as the basis for the Task Model in the *Design* step (Step 6) of CAL's Validation Framework, where the characteristics of individual items or tasks are defined. In ACCESS 2.0 Online, therefore, the overall enterprise of mapping evidence to inferences is mediated through the theoretical notion of the MPI in the Domain Model, while specific MPIs for actual test items are developed at a later stage.

3 The Design of ACCESS 2.0 Online

Step 6 in CAL's Validation Framework is the *Design* step, which has four components, derived from ECD (Mislevy, Almond, & Lukas, 2004): The Student Model; the Evidence Model; the Task Model; and the Assembly Model. For the benefit of the reader, the Task Model is presented prior to the Evidence Model in this section, as our description of the evidence derived from scoring is dependent upon a clear understanding of the nature of the tasks.

3.1 The Student Model: What knowledge, skills, and abilities does a student possess?

The Student Model defines the knowledge, skills, and abilities that a student possesses and that are going to be assessed. The Student Model for ACCESS 2.0 is operationalized in terms of the WIDA ELD Standards; the Standards define what ELLs process (in the Reading and Listening domains) or produce (in the Writing and Speaking domains) for a given grade-level cluster and proficiency level.

3.2 The Task Model: What do assessment tasks for ACCESS 2.0 Online look like?

This section describes how items and tasks are designed to reflect the elements of the domain analysis described in Section 2.2 in order to collect the necessary evidence required for later decision-making. Data Recognition Corporation (DRC), the vendor responsible for the online implementation of the assessment, administers the tasks in the online environment and carries out the automated scoring of the Listening and Reading tasks and the hand scoring of the Writing and Speaking tasks. Items and tasks are discussed by language domain.

3.2.1 Listening Items

All Listening items include a pre-recorded stimulus passage and question stem. Listening items are selected-response items, with one key and two distractors as answer choices. Answer choices are primarily illustrations; for Grades 2–12, items that test listening proficiency at PLs 3–5 may consist of short written text response options that are written to be about two PLs lower than the targeted PL of the Listening item. All operational Listening items are traditional multiple-choice items, though some items embedded for field-testing purposes involved enhanced item-presentations (see Section 4.1.1.).

Each item on the Listening test is written to reflect the language of one of the five WIDA ELD Standards and to test a student's ability to process language at one of the five fully delineated proficiency levels. *Folders* group together three test items that are written around a common theme, with each item targeting a progressively higher proficiency level. Thematic folders are targeted as A, B, or C, with A folders encompassing PLs 1–3, B folders encompassing PLs 2–4, and C folders encompassing PLs 3–5.

3.2.2 Reading Items

Reading items are similar in format to Listening items. The stimulus for Reading items is written text and answer choices primarily are also written text, though for Grades 1–12 graphic response options may be used for items targeting PLs 1 and 2. As with Listening items, Reading items are grouped into thematic folders of three test items each. Most items on the operational Reading assessment are traditional multiple choice, though some operational items and some items embedded for field-testing purposes involved enhanced item-presentations (see Section 4.1.1.).

3.2.3 Writing Tasks

Writing tasks are designed to elicit language from one or more of the WIDA ELD Standards. Tasks appearing on the Tier A test form (see Section 3.4.3) are designed to give students the opportunity to produce writing samples that fulfill linguistic expectations up to PL 3, while those appearing on the Tier B/C form are designed to give students the opportunity to produce writing samples that fulfill linguistic expectations up to PL 6.

With the exception of students in Grades 1–3 and those taking the paper-based accommodation, writing prompts appear on the computer screen. In the spirit of providing maximal support and making every provision to ensure that students are given the opportunity to demonstrate the full extent of their English language proficiency, modeling is sometimes used to make task expectations as clear as possible to students. For example, the first of a series of questions may already be partially completed, or a sentence starter may be provided.

Students in Grades 4–5 provide either handwritten or keyboarded responses, with the default response mode determined in advance at the state or district level. For students in Grades 6–12, keyboarding is the default response mode, with a handwriting option offered as an accommodation. For students in Grades 1–3, the test is not administered via computer. Rather, the familiar format from ACCESS 1.0 is utilized, where the test administrator reads from a script and the students respond in a printed test booklet.

3.2.4 Speaking Tasks

Stimuli on the Speaking test include graphics, audio and text. All stimuli are presented by a Virtual Test Administrator (VTA). The VTA serves as a narrator who guides students through the test and as a virtual interlocutor. The VTA is introduced to students during the test directions in order to establish the testing context.

Task modeling is an essential component of the Speaking test design. In addition to the VTA, students are introduced to a virtual model student during the test directions. Prior to responding to each task, test takers first listen to the model student respond to a parallel task. The purpose of the model is to demonstrate task expectations to both test takers and to DRC raters who score all Speaking task responses.

Students navigate through the Speaking test independently and at their own pace. They must listen to all audio on a screen before the test allows them to advance to the next screen. The

amount and complexity of task input varies by grade-level cluster and task level. The purpose of the input is to provide academic content for students to draw on in their responses.

Figure 4 shows the generic screen layout of the Speaking test.
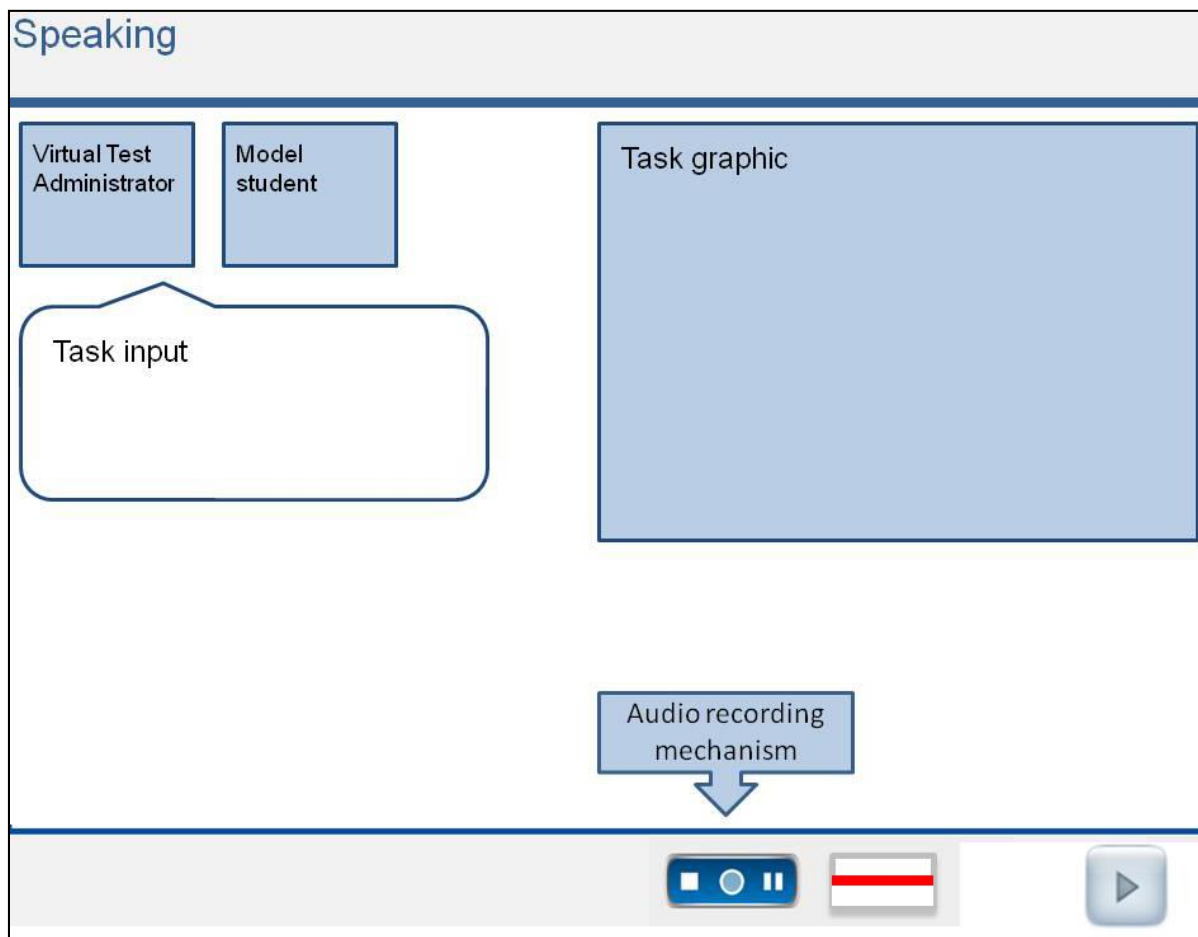
Figure 4. Visualization of the Speaking test screen layout.

Both the VTA and the model student are represented within the testing interface by static images. They are portrayed wearing computer headsets with microphones to reflect the actual testing scenario. Test input and stimuli are presented both aurally and in speech bubbles on the screen. Students respond orally to the tasks, with their responses recorded and transmitted to DRC for later scoring.

3.3 The Evidence Model

In determining what evidence should be sought at the *Design* phase of ACCESS 2.0 Online, two questions were articulated: (a) *How are student performances on ACCESS 2.0 Online scored?* and (b) *How are measures of student performance on ACCESS 2.0 Online calculated?* This section describes the scoring procedures and the methodologies used to score student performances in each domain.

3.3.1 How are student performances on ACCESS 2.0 Online scored?

3.3.1.1 Multiple Choice Scoring: Listening and Reading
Listening and Reading items are scored dichotomously, as correct or incorrect. Scale scores for each domain are calculated based on the items that are administered to the test taker and the number of those items that the student answers correctly. For details on how scale scores for Listening and Reading are calculated, see Section 3.3.2.1 below.

3.3.1.2 Performance-based Tasks: Writing and Speaking
Performance-based tasks in the domains of Writing and Speaking are scored by trained raters. 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 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 or Speaking 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 per 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 raters and as Team Leaders on previous projects. Scoring Directors are responsible for a specific set of tasks within a single domain. The Scoring Directors train and oversee the teams of raters 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.

Calculating Score Agreement for Scoring Monitoring

- For writing, 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,

scores 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.

- For speaking, agreement is defined as two scores being exactly the same. (See section 3.3.1.4 for a description of the speaking scoring scale.)

Routing Responses to Ensure "Blind" Second Ratings

- The DRC scoring system ensures that responses are routed to qualified raters until the prescribed number of ratings is performed for all responses.

- Raters do not know if they are the first or second rater.

Monitoring Scoring (Quality Control)

- Ongoing quality control checks and procedures help monitor and maintain the quality of the scoring sessions. DRC monitors rater reliability with a 20% read-behind protocol. Read-behind data are monitored daily.

- Responses can be retrieved on-demand (e.g., specific grade-level clusters, specific students) should the need arise during or subsequent to the scoring process.

- If needed, responses can be rescored based on task- or response-level information, such as task number, date, score value assigned, or rater ID.

- For Writing, DRC uses both recalibration and validation sets. For each of the first five days that raters score a task, they take one recalibration set of five responses per task. After the raters take the recalibration sets, the Scoring Director or Team Leader reviews them using descriptors from the Writing Scoring Scale and the anchor responses to confirm 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.

- For Speaking, DRC uses recalibration sets. At the start of the scoring window, raters take these sets every day to ensure that they are calibrated, and raters' performances on recalibration sets are used for monitoring purposes. Later in the scoring window, recalibration sets are used on a weekly basis to monitor scoring.

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.

3.3.1.3 The ACCESS 2.0 Writing Scoring Scale

The 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 (e.g., a response that falls between 3 and 4 is scored as 3+). The scale descriptors include three different yet interrelated dimensions: discourse, sentence and word/phrase. These scale descriptors guide raters as they consider all three dimensions in order to make holistic judgments about which score point best suits a response. The dimensions are distinguished as follows:

- The descriptors for the discourse dimension 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 descriptors for the sentence dimension evaluate the complexity and grammatical accuracy of sentence structures used in the response.

- The descriptors for the word/phrase dimension 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 makes 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 that response. If all three descriptors suit the response, a whole score point is awarded. If there is clear evidence that one or two descriptors from an adjacent score point are a better fit, a plus score point between the two applicable whole score points 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, while partially off-task responses are scored by ignoring the off-task portion of the response 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. 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. The 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 Tier B/C tests for all grade-level clusters, results from the different tasks are given different weights. 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/C tests can range from 0–54.

Table 1.
Rating to raw score conversion (Writing).

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

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 Online. The Writing Scoring Scale was designed specifically as a scoring tool and is not appropriate for any other purposes.

3.3.1.4 The ACCESS 2.0 Speaking Scoring Scale

The Speaking Scoring Scale defines five score points: *Exemplary*, *Strong*, *Adequate*, *Attempted*, and *No Response* (the final score point only applies if the rater uses one of three non-scorable codes: B= Blank response; F= Foreign language response; I = Indecipherable response). These score points are applied based on the proficiency level expectations of each task; that is, the level of language proficiency that each task is designed to elicit. These expectations are exemplified by the model student response (See Section 3.2.4). In this way, the model response serves as a scoring benchmark. Raters listen to the model response and score test taker responses relative to the model. A score of *Exemplary* means that the student response demonstrates English language use that is equal to or beyond the English language use illustrated by the model student's response.

The Speaking Scoring Scale includes descriptors for overall language use, response sophistication, language delivery, and word choice. As stated above, the scale is applied relative to the proficiency level demands of the task. For tasks targeting language elicitation at PL 1, there are only three possible score points: *No Response*, *Attempted*, and *Adequate and Above*.

This is the case because appropriate responses to PL 1 tasks are single words and short chunks of language, so it is not possible to reliably distinguish between *Adequate*, *Strong*, and *Exemplary* performances.

To calculate a raw score for the Speaking test, the five score points are converted to whole numbers, as shown in Table 2. To calculate a total raw score, the raw scores for each task are added together; additionally, in Tier B/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/C student would not be 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 pre-A test, scores can range from 0–6; on the A test, from 0–18; and on the B/C test, from 6–30.

Table 2.
Rating to raw score conversion (Speaking).

Rating	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

Speaking tasks are scored using the ACCESS 2.0 Speaking Scoring Scale. The Speaking Scoring Scale is distinct from the WIDA Speaking Rubric, which is a tool for classroom use and score interpretation. The Speaking Scoring Scale was designed specifically for test scoring use and is not intended for classroom purposes.

3.3.2 How are measures of student performances on ACCESS 2.0 Online calculated?

The measurement model that forms the basis of the analysis for the development of ACCESS 2.0 Online is the Rasch measurement model (Wright & Stone, 1979). Additional information on its use in the development of the ACCESS assessment program is available in the WIDA Consortium Technical Report No. 1, *Development and Field Test of ACCESS for ELLs* (Kenyon, 2006). The original ACCESS 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 guides the inclusion, revision, or deletion of items during the development and field testing of the test forms, and will continue to guide the refinement and further development

of the test. All Rasch analyses are conducted using the Rasch measurement software program *Winsteps* (Linacre, 2006).

3.3.2.1 Rasch Model for Dichotomous Scoring

For Listening and Reading, the dichotomous Rasch model is 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.

3.3.2.2. Rasch Model for Polytomous Scoring

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 3.3.1.3, 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 A 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 (see section 3.3.1.4).

3.3.2.2 Scale Scores and Proficiency Level Scores
Scale scores are calculated by transforming the person ability estimate via a scaling equation. The scaling equations for each domain are provided in II.1.2, under Claim 4.3 in the CAL Validation Framework. 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 logit scale (see Center for Applied Linguistics [2017]).

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 scores 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.

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.

3.3.2.3 Composite Scores

Four composite scores are calculated for ACCESS 2.0 Online: Oral language, Literacy, Comprehension, and Overall. Composite scores are calculated as weighted averages of domain scale scores, as follows:

- 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

3.4 The Assembly Model: How are the assessment components for ACCESS put together?

This section describes how ACCESS 2.0 Online is assembled to ensure that the evidence collected is (a) sufficient to make the intended decisions, and (b) appropriate for the student's level of proficiency. In order to tailor the test closely to student ability levels while still including items and tasks that assess all of the Standards, adaptivity has been built into the test. The Listening and Reading tests both use a multistage adaptive test design. The Writing and Speaking tests are tiered, and placement into the tiers depends on performance on the Listening and Reading tests. Details are presented below.

3.4.1 Listening

The Listening test uses a multistage adaptive design, as illustrated in Figure 5. All students begin the Listening test with two entry folders (with three items each) at Stage 1 and Stage 2, both targeting SIL (See Section 2.2.1 for the WIDA ELD Standards and their abbreviations). At that point, the student's ability is estimated based on performance on those six items, and that ability estimate is used to determine which of the three leveled LoLA folders in Stage 3 is administered next. Students whose ability estimate predicts a PL score of 5.0 or higher are routed into the folder at the highest level (C in Figure 5); students whose ability estimate predicts a PL score of 2.5 or lower are routed into the folder at the lowest level (A in Figure 5); all others are routed into the B folder.[1] Throughout the test, a student's underlying measure of ability is re-estimated with the completion of each folder, and the level of the next folder to be administered is chosen accordingly, following the decision rules above. Thus, each student will trace a tailor-made path through the test according to ability level, but the order of the stages is invariant across students. In total, there are eight possible stages, but students whose ability estimate falls below PL 2.5 after the sixth stage end the test at this point. The intent of this design is to ensure coverage of the Standards while delivering a test that closely matches the student's PL, thus minimizing measurement error.

Figure 5. Format of the Listening test.

---

[1] Note that although students' final scores for Series 401 were interpreted in terms of the ACCESS 2.0 proficiency level cuts set at the 2016 standard setting, the settings to route students through the adaptive test engine are keyed to ACCESS 1.0 proficiency level cuts.

3.4.2 Reading

Figure 6 shows the format of the Reading test. The format and adaptivity are similar to the Listening test, but the Reading test consists of ten stages rather than eight. This reflects the greater weight given to Reading in calculating the composite scores, as well as the view that literacy skills are paramount in developing academic language proficiency. The greater weight afforded to Reading and Writing resulted from a policy decision by the WIDA Board before the first operational administration of ACCESS.
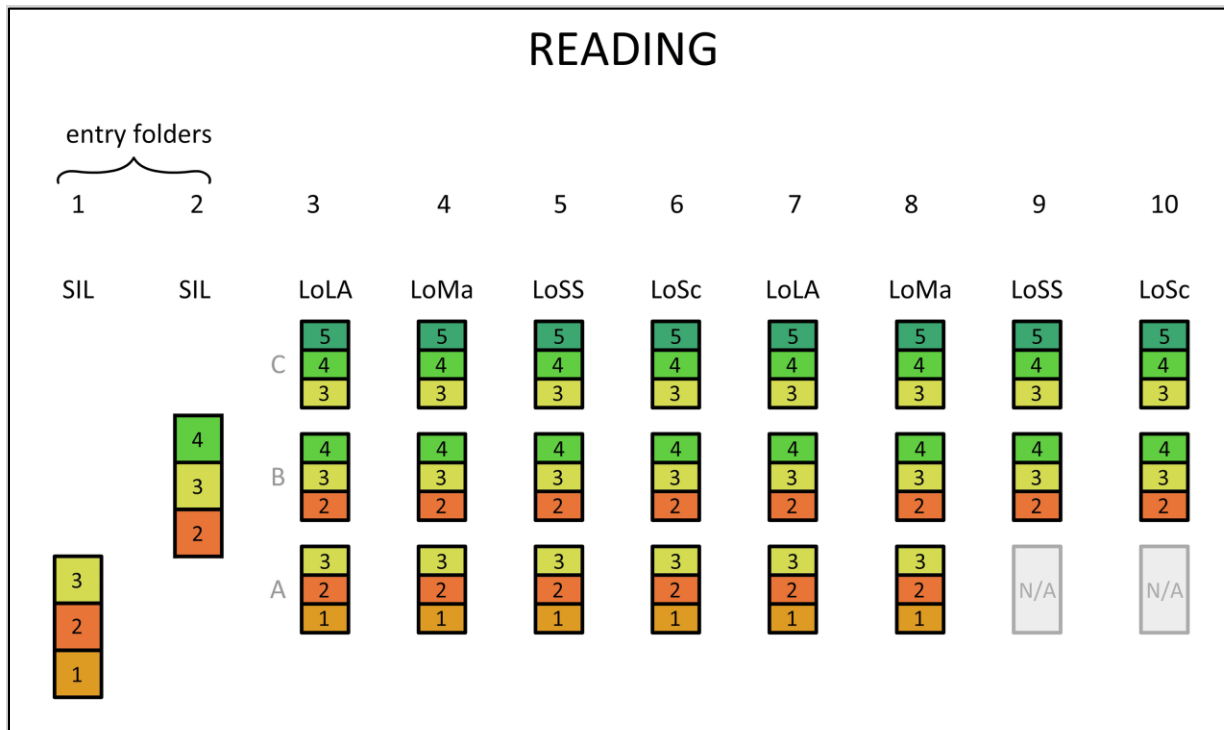
Figure 6. Format of the Reading test.

3.4.3 Writing

Figure 7 shows the format of the Writing test. As can be seen from the figure, Writing is tiered. Tier A consists of tasks written to elicit language at PLs 1–3, while Tier B/C is designed to elicit language at PLs 4–6. With the exception of Grade 1 Tier A, both tiers consist of three tasks. Both tiers include tasks that target a single standard and tasks that integrate more than one WIDA Standard. For example, in the Tier A forms (except for Grade 1), one task integrates the Language of Math and the Language of Science. On the Tier B/C forms, one task integrates the Language of Math and the Language of Science, while another extended task integrates Social Instructional Language, the Language of Language Arts, and the Language of Social Studies. The ways in which the Standards are targeted by these tasks vary across grade levels and are spelled out in the generative item specifications.
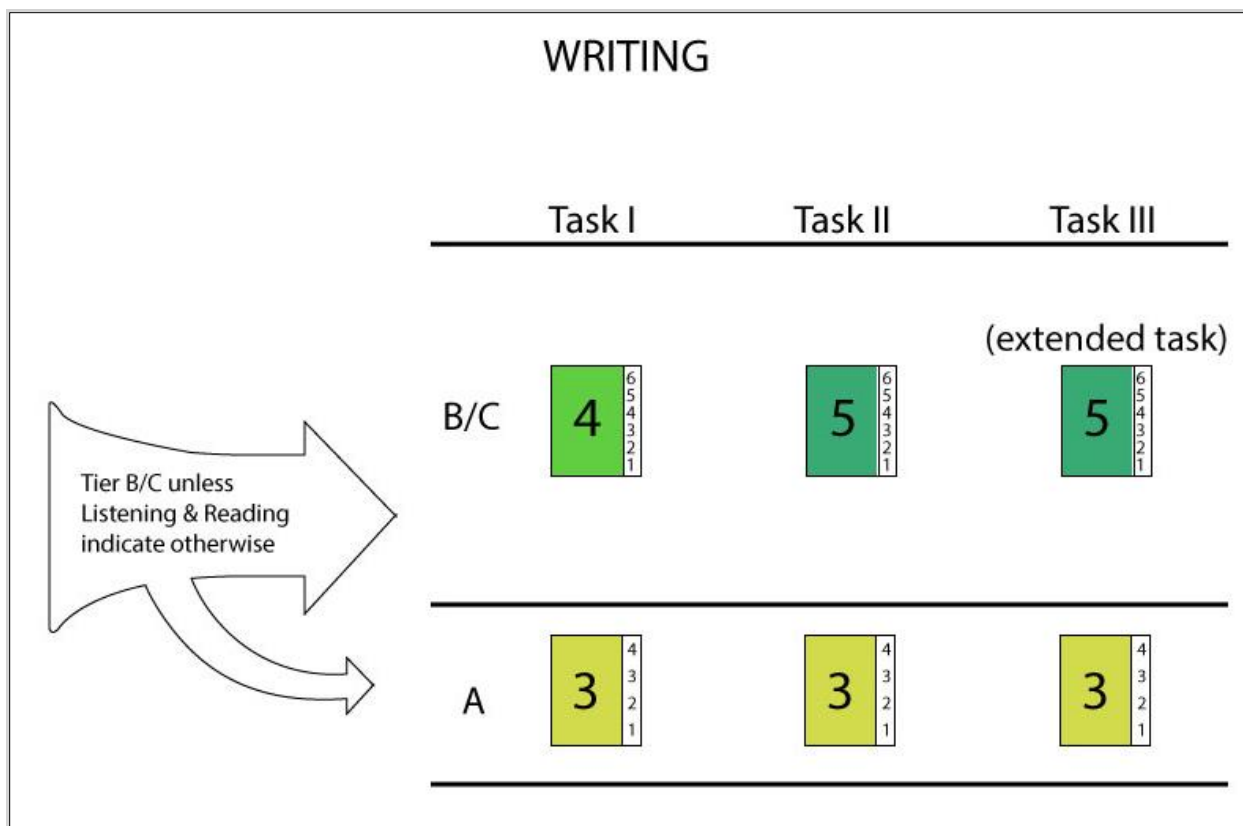
Figure 7. Format of the Writing test.
Note: Grade 1 Tier A follows a different model, and has four tasks targeting PLs 1, 1, 2, and 3. Numbers inside the boxes represent the targeted proficiency level of the task; the smaller numbers on the right edge of each box represent the range of proficiency levels that a task may elicit.

Placement into tiers on the Writing test depends on how students perform on the Listening and Reading tests, which receive computerized scores. To determine how to best place students into a tier, the previous year's test data for all students who were administered the assessment are analyzed to examine the relationship between how students perform on Listening and Reading and how they perform on Writing using logistic regression analyses. This information is used to program an algorithm into the ACCESS 2.0 Online test that will be used by the computer to determine which tier of the Writing test will be administered to each student. The purpose of the algorithm is to place students who are predicted to score above PL 3.0, based on their performances in Listening and Reading, into Tier B/C for Writing and Speaking, and all other students into Tier A.

3.4.4 Speaking

Figure 8 shows the format of the Speaking test. The Speaking test includes tasks that target language elicitation at three PLs: 1, 3, or 5. The tasks are grouped into thematic folders, which are aligned to one or two of the WIDA Standards.

As shown in Figure 8, the Speaking test includes three tiers: Tier Pre-A, Tier A, and Tier B/C. Tier Pre-A includes tasks that target language elicitation at PL 1. Tier A includes tasks that target language elicitation at PLs 1 and 3. Tier B/C includes tasks that target language elicitation at PLs 3 and 5.

A thematic panel refers to the folders across all tiers within a grade-level cluster that relate to a particular WIDA ELD Standard. For example, the Tier B/C, Tier A, and Tier Pre-A folders that address SIL make up a single thematic panel. Ideally, within a thematic panel, tasks at PL 1 and PL 3 are the same across tiers.[2] For example, within a SIL panel, the same PL 3 task appears on both the Tier A and the Tier B/C forms of the test.
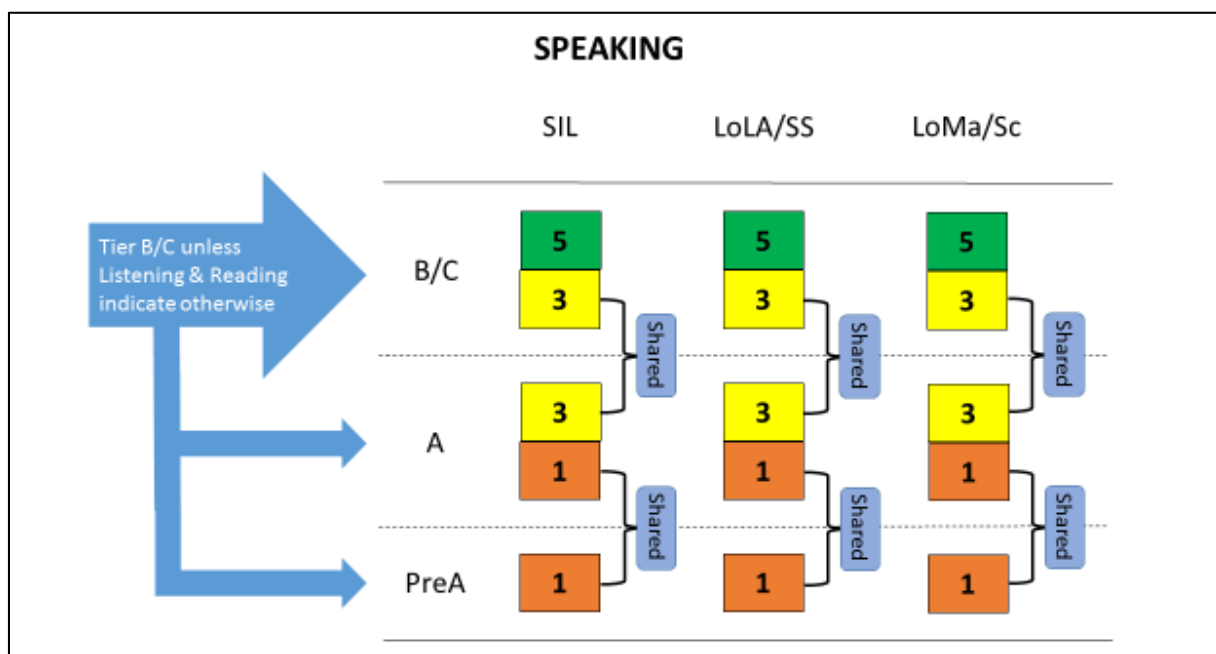
Figure 8. Format of the Speaking test.

As with Writing, placement into the three tiers on the Speaking test shown in Figure 8 depends on performance on the Listening and Reading tests. An algorithm is applied to the results of the Listening and Reading test to determine which tier is optimal for the student. Unlike Writing, the Speaking test has one additional tier, Tier pre-A. Students are placed into Tier pre-A when their scores on Listening and Reading fall below chance performance. The Speaking pre-A tier is designed to meet the needs of students in the very early stages of English language development. As noted above, tasks are targeted to the P1 level and are scored using a modified version of the full Speaking rating scale.

---

[2] Note, however, that on the Series 401 test there are a few exceptions where the PL3 task differs on Tiers A and B/C.

4 Assessment Performance: The Implementation of ACCESS 2.0 Online

This section focuses on *Assessment Performance* (Step 5) in CAL's Validation Framework. This section reviews how items and tasks for ACCESS 2.0 Online are developed, reviewed, revised, and chosen for inclusion in the operational test. It also describes the interaction between students and the test.

The development process for the first implementation of the ACCESS 2.0 Online represents a break from the regular cycle of item development and refreshment, as all of the items and tasks were developed specifically for the ACCESS 2.0 online environment.

4.1 How is ACCESS implemented?

4.1.1 Listening and Reading

Series 401 represents the second implementation of ACCESS 2.0. Prior to the transition to ACCESS 2.0, a refreshment cycle was in place in which roughly one-third of all Listening and one-third of all Reading items were targeted for refreshment annually. During the transition to ACCESS 2.0, item development in Listening and Reading followed a different trajectory.

Listening items are developed so that each item appears on its own screen, with associated graphic support. Audio recording scripts containing the item orientation, stimulus, and question stem are audio recorded with professional voice actors and produced by a professional recording studio. Audio playback of test item content is automatic when students advance to the next screen. Listening test content is played one time for students unless the student has a predetermined accommodation allowing for a single repetition of the item stimulus and question stem.

The Listening items on Series 400 were developed entirely for ACCESS 2.0 Online. No refreshment took place between Series 400 and Series 401 in the domain of Listening. The Listening items used on Series 401 were field tested prior to the operational launch of Series 400. For detail on the field testing of these items, see Center for Applied Linguistics (2016).

During the operational administration of Series 401, students experienced operational Listening items plus embedded Series 402 field test items. The embedded field test items included innovative item formats, including hot spot and drag-and-drop items, where the student either clicked on an area of the screen or dragged an image/text to a specified screen area to respond.

In the domain of Reading, items on Series 400 were developed based on operational test items from previous test series, adapted for implementation in the online environment. Item format was adjusted to optimize the items for presentation on computer screens. For example, on the paper test, students might have read a single "theme passage" with multiple items related to the single passage. The test booklet was laid out so that the student could see the passage and all three

items simultaneously. In the online format, the student sees only one item per screen. Therefore, the format was adjusted so that each item has its own passage.

For Series 401, approximately one-sixth of Reading items were targeted for refreshment. An enhancement in development for Series 401 was the inclusion of innovative item formats, including hot spot and drag-and-drop items, where the student either clicked on an area of the screen or dragged an image/text to a specified screen area to respond.

In both Listening and Reading, the item refreshment process spans approximately three years, beginning with the development of the refreshment plan and the updating of item specifications. Trained item writers work from these specifications to draft Listening and Reading items within a thematic folder. After initial development, folders are screened at CAL, and those that are approved for further development undergo a rigorous process of internal development and review, including reviews by standards experts and extensive fact checking. During this phase, images and other ancillary materials, such as scripts and directions, are produced.

At this point, items undergo external bias, sensitivity, and content reviews, after which they undergo further refinement. Items that reach this point are then administered as embedded field test items on students' operational assessments.

For Series 401, a total of 126 Reading items (42 folders) were field tested, across all five grade-level clusters, embedded into the Series 400 operational assessment. Each student receives one Listening and one Reading field test folder embedded into their operational test. Field test folders are targeted to refresh a specific operational folder on the test, and field test folder specifications include the stage, standard, and tier pool target (A, B, or C) of the folder. Students are administered the embedded field test folder at the stage targeted for refreshment, with administration randomized so that half of the students see the field test folder before the corresponding operational folder, and half see the operational folder before the field test folder. Field test folders are administered to those students who are routed to take the Operational folder that is either at the same tier or adjacent to the tier that the field test folder targets. When field test samples are drawn, the sample includes 50% of students at the tier targeted by the field test folder, and 50% at adjacent tiers. Field test sample targets in Listening and Reading are set at a minimum of 3,000 responses per folder.

After field test data are drawn, folders of items are analyzed for their psychometric properties, and those that meet established psychometric standards are eligible for selection in the next year's operational test.

As is the case for Listening, students taking the operational Series 401 Reading test receive one embedded Series 402 field test folder during their operational assessment.

4.1.2 Writing

The development of Writing tasks is similar to that of Listening and Reading items. Writing tasks, however, do not currently undergo large-scale field testing. Instead, after external bias,

sensitivity and content reviews, they are subject to small-scale tryouts, led by CAL staff. In these tryouts, candidate folders for Grades 4–12 are administered to students using a mocked-up computer delivery; as noted above, students in Grades 1–3 complete the Writing assessment with a traditional paper-and-pencil administration. Student responses, as well as observations and interviews, inform further revisions to the folders.

Then, a small-scale stand-alone field test of Writing folders is conducted. For Series 401, a total of 5 Writing tasks were field tested. A sample target of 75 students per task was established. The field test uses the online interface where applicable, and the field test is administered under standard testing conditions, with responses captured online where applicable. For the writing field test, responses are double-scored and adjudicated by CAL experts, and qualitative analysis of the collected responses is conducted. The main purposes of this small-scale field testing are (a) to confirm that the tasks are working as intended, (b) to identify anchor samples for rater training, and (c) to inform the rating of the tasks when they become operational.

The Writing items on Series 400 were primarily adapted to the computer from operational items from previous test series prior to the launch of Online ACCESS (Series 203, 301, and 302). Series 401 incorporates continuing items from Series 400 as well as items newly developed and field tested for Series 401. Major differences between Series 400 and Series 401 are that the extended task was refreshed on the Tier B/C form, with the introduction of tabbed browsing and click-to-enlarge functionalities in the input for Grades 4–12 tasks.

4.1.3 Speaking

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

Many of the 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 to the computer 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 Online.

All students are administered a Speaking field test folder appended to their operational Speaking assessment. A total of 84 tasks (42 folders) were field tested for Series 401, with a target sample size of 300 students per folder. Responses were double-scored by DRC trained raters, and adjudicated by CAL raters.

The Speaking test underwent a major overhaul between ACCESS 1.0 and ACCESS 2.0. The Speaking test was previously administered one-on-one with a live test administrator, who scored the test as it was administered. Each folder had tasks at each proficiency level (1–5). The test administrator used "stopping rules" to determine when the test taker could no longer answer

appropriately and when to move on to the next folder. The test administrator also was permitted to ask follow-up questions to elicit additional responses from the student.

For ACCESS 2.0 Online, folders were designed to target one or two proficiency levels: Tier Pre-A folders include one task which targets only PL 1. Tier A folders include tasks that target PLs 1 and 3. Tier B/C folders include tasks that target PLs 3 and 5. Students are routed into a tier based on their performance in the Reading and Listening sections of the test. The content is presented entirely on the computer, and the responses are recorded by the test engine and are transmitted to DRC for scoring.

4.2 What is the assessment delivery experience for students taking ACCESS 2.0 Online?

4.2.1 Listening and Reading

Listening and Reading are the first domains assessed. Students may take these in either order. Students sit at individual computer monitors and are administered the Listening and Reading tests online. They are issued headsets which are used to listen to directions for the Listening and Reading tests, as well as to the Listening items. Students use a computer mouse to select or record their answers.

4.2.2 Writing

Writing tasks are delivered on paper to students in Grades 1–3. All students in Grades 1–3 handwrite a response.

Writing tasks are delivered online to students in Grades 4-12. A student may provide handwritten or keyboarded responses, with the choice depending on a combination of local, state, and consortium-wide policies, as follows:

- Grades 4–5: A decision is made at the local or state level as to whether handwriting or keyboarding is the default response mode. In districts where keyboarding is the default, the option exists to use handwriting as an accommodation.

- Grades 6–12: Keyboarding is the default, with the option to use handwriting as an accommodation.

4.2.3 Speaking

Speaking tasks are delivered online. Students listen to prompts via headsets that are equipped with microphones to capture their responses. Extensive support is provided to the student through illustrations and written input designed to provide sufficient content for the response, as well as a model student response that is intended to provide guidance regarding the level of linguistic complexity required to respond adequately (see Section 3.2.4).

4.3 Assessment performance—interaction between test and student

Administration of ACCESS 2.0 Online takes place between December and April of the academic year, with testing windows determined at the state level. The Reading and Listening tests are administered first (in either order), followed by Writing and Speaking (in either order). The test may be administered in several sessions within one day or over a series of days. Student performance on the test forms the basis for developing *Assessment Records*, which are addressed in detail in Part II of this report.

Part II: Assessment Records

In Part II of the Annual Technical Report, the focus is on the *Assessment Records* step in the CAL Validation Framework (see Part I.1.2, for a full description of the framework). Section 1 details the claims made regarding assessment records and provide references to evidence that supports those claims. Section 2 provides descriptions of the data and analyses presented in Section 3. In Section 3, detailed data and analyses are presented regarding the most recent operational administration of ACCESS 2.0 Online.

1 Assessment Records for ACCESS 2.0 Online

The complete validation framework for ACCESS for ELLs assessment program, as described in Part I of this report contains seven steps. Part I of this report focuses on the initial three steps (*Plan, Design, and Assessment Performance)*. The argumentation and the data presented in this part (Part II) address *Assessment Records*, and present evidence specific to ACCESS 2.0 Online. By focusing on Assessment Records (i.e., test scores and proficiency level descriptions), the information here will be used to support claims related to the quality and consistency of the assessment data gathered and analyzed using ACCESS 2.0 Online. The claims in this step of the Assessment Use Argument (AUA) all pertain to the general question: *How do we know that the reported language domain scores and composite scores on ACCESS 2.0 Online are consistent and dependable?*

The diagram in Figure 1 shows a visual representation of an argument-based approach for supporting claims related to *Assessment Records* (Step 4). The figure shows how claims related to *Assessment Records* fit into the complete 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 Online. Section 1.2 provides an overview of the sources of evidence which support the argument.
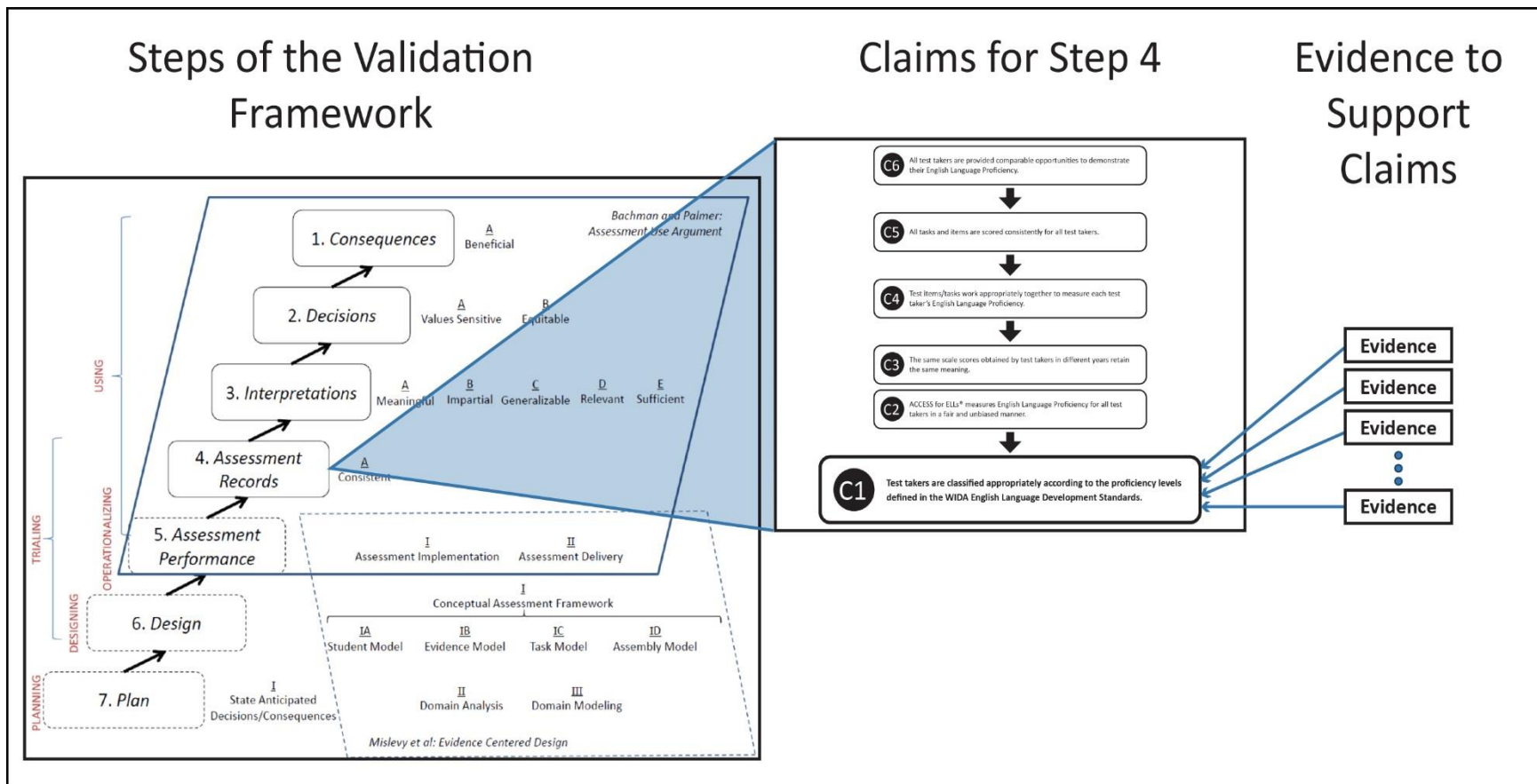
Figure 1. Structure of the argument-based approach supporting *Assessment Records* (Step 4).

1.1 Claims for the Assessment Records for ACCESS

Assessment Records (Step 4) of the CAL Validation Framework is broken down into the following six claims:

C4.6. All test takers are provided comparable opportunities to demonstrate their English Language Proficiency.

C4.5. All tasks and items are scored consistently for all test takers.

C4.4. Test items/tasks work appropriately together to measure each test taker's English Language Proficiency.

C4.3. The same scale scores obtained by test takers in different years retain the same meaning.

C4.2. ACCESS for ELLs measures English Language Proficiency for all test takers in a fair and unbiased manner.

C4.1. Test takers are classified appropriately according to the proficiency levels defined in the WIDA English Language Development (ELD) Standards.

As shown in Figure 2, these claims depend upon each other, moving from (C4.6) down to (C4.1). Within this organizational structure, each successive claim requires that the previous claim be met in order for it to support the validation argument.

The claim that tasks and items are scored consistently (C4.5) does not support the overall validity argument unless the claim that all test takers are provided with comparable opportunities (C4.6) is also met. In other words, tasks and items may be scored consistently for all test takers, but if all test takers are not provided with comparable opportunities, then consistent scoring in and of itself does not support the validity argument. Likewise, support for the claim that test items or tasks work appropriately together to measure English language proficiency (C4.4) requires that those items or tasks be consistently scored (C4.5), otherwise C4.4 cannot support the validity of the assessment. C4.3 asserts that scale score interpretation remains consistent over time—one requirement for this to be true is that the assessment must be able to measure students across a broad range of English language proficiency abilities (as claimed at C4.4). While comparability of opportunity is evinced by the steps taken to ensure that the implementation of the ACCESS test is equitable, C4.2 looks at measurement, or how student performance is translated into a quantifiable outcome. In order for this to be done in a fair and unbiased manner across time, C4.3 must be met. Finally, the appropriate classification of test takers (C4.1) cannot be accomplished unless the performance of all test takers is measured in a fair and unbiased manner (C4.2).

Each prior claim alone does not constitute the entirety of the evidence for the successive claims, however; while each claim requires the evidence from its predecessor, it also requires additional evidence to be supported fully. Section 1.2 below provides a fully fleshed out structure of the line of argumentation for *Assessment Records*, including actions that are taken to ensure the

consistency and reliability of the assessment records, and evidence to demonstrate that those actions are taken.
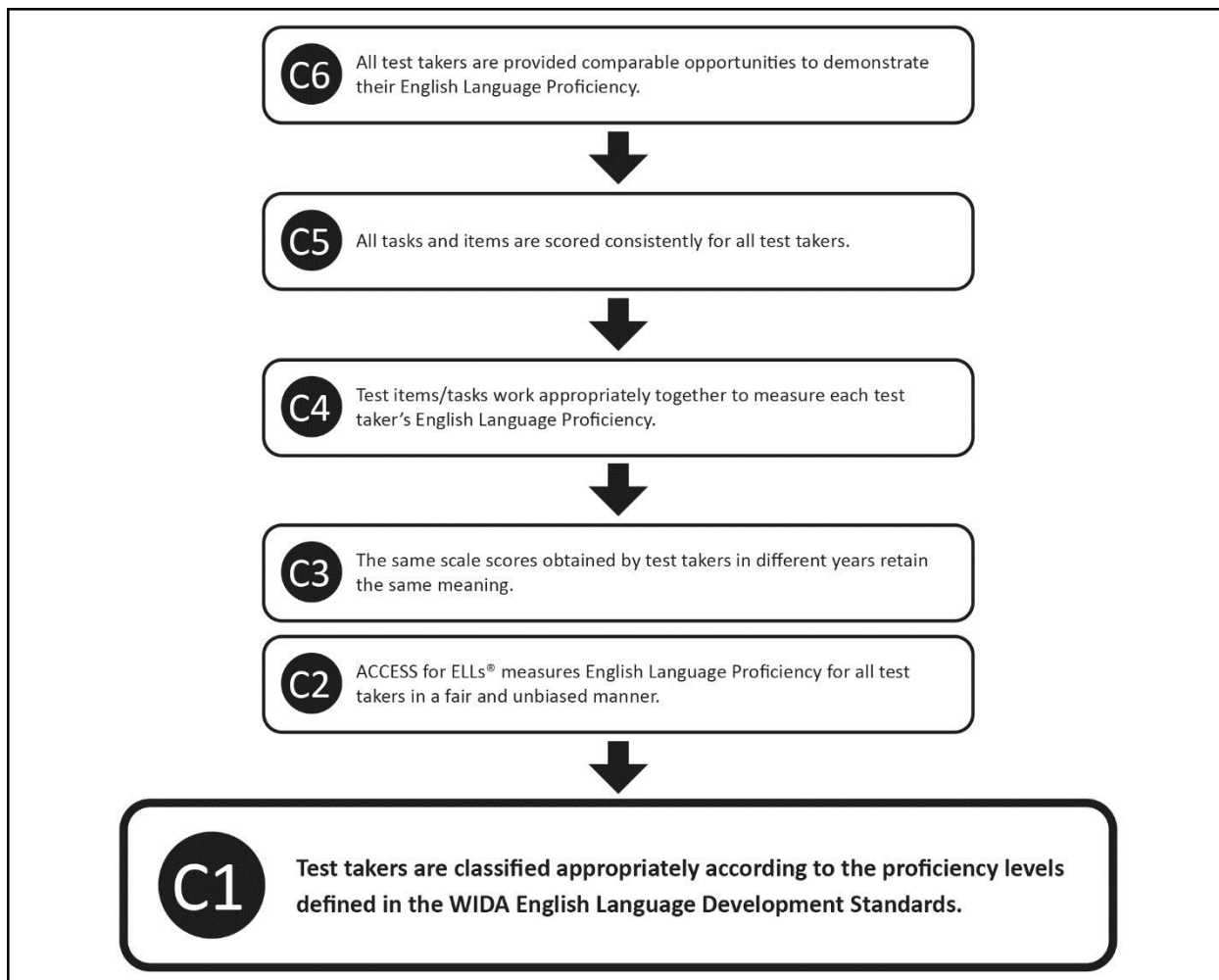
Figure 2. Progression of claims for *Assessment Records* (Step 4).

1.2 *Evidence for* Assessment Records *Claims of ACCESS 2.0 Online*

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. In what follows, we outline the location within this Annual Technical Report or the external sources that provide evidence related to each action. A summary table of this information is presented in Section 1.3, below.

Because these claims relate to *Assessment Records*, which is Step 4 of the overall validation framework, their numbering begins with 4. The second number (after the decimal) denotes the

level of the claim within Step 4. Individual actions to ensure each claim are denoted by the corresponding letter (a, b, c, and so on).

Note that the *Assessment Records* claims are claims for the ACCESS assessment program. The evidence provided for these claims in this report is evidence specific to ACCESS 2.0 Online Series 401.

C4.6. All test takers are provided comparable opportunities to demonstrate their English Language Proficiency.

Action 4.6a: Test design and student training procedures ensure that all students are able to interact with the technology of the test.

Evidence: CAL conducted extensive cognitive laboratories to ensure that students at all grade levels and at the lowest proficiency levels could successfully manipulate the student test interface.

A Test Demo video is available for all students to view prior to testing. This video walks the students through all aspects of testing.

The Test Practice items, which appear in the operational test prior to the operational items, are also available as stand-alone packages for students to familiarize themselves with the computer interface prior to testing.

Procedures for administering the test are documented in the Test Administration Manual.

All test domains contain an audio check prior to administration to ensure that students can hear the audio stimulus. In addition, the Speaking test, which requires that students speak into a microphone to capture their oral responses, contains a check to ensure that the students are speaking loudly enough for the interface to successfully record the response. This check occurs at the beginning of the Speaking section of the test. In addition, as the students record their responses, the interface detects the volume level as students respond, and prompts them to try again if they speak too softly. A further measure ensures that if the student does not speak loudly enough a second time, the test pauses and prompts the students to raise their hands for assistance.

Action 4.6b: Procedures are in place to address technical issues and interruptions.

Evidence: Procedures on handling technical issues and interruptions are detailed in the *DRC INSIGHT Technology User Guide*, Volume V (Data Recognition Corporation, 2016a), as well as in the *WIDA AMS User Guide* (Data Recognition Corporation, 2016b). The *ACCESS for ELLs 2.0 Test Administrator Manual* (WIDA Consortium, 2016) details the steps that test administrators can take during testing, and includes a troubleshooting chart, as well as information on how to contact DRC Customer Support. WIDA also offers a series of webinars

which focus on issue resolution during testing, including some with a specific focus on technological issues. The WIDA website also has a compilation of technology FAQs available.[1]

WIDA and DRC also collaborate to create documents and memos to address issues in the field. For example, on the ACCESS 2.0 Technology webpage, an iPad Troubleshooting Guide and a Whitelisting Memo have been added in response to common questions and concerns.

WIDA maintains a number of tools on their website in case of technical issues or interruptions.

WIDA has a system status page on their website to monitor and track system outages. A troubleshooting chart accompanies this page.

DRC maintains a customer service email account and phone number for technical issues. In the event of a systemic issue or outage, educators have access to the WIDA System Status Dashboard.[2]

Should an outage or technical issue occur, DRC notifies State Education Agencies (SEAs) via email as to when the systemic issue occurs as well as when the issue is resolved, noting which aspects of testing or testing devices were impacted. Additionally, for extended technical issues, WIDA posts general information pertaining to the outage on the main page of the WIDA website. In the event of extended technical issues, WIDA and DRC provide updates to SEAs via email as follows: (1) broadcast message/announcement of incident; (2) update(s) on the incident (if not resolved after two hours); (3) restoration of service message; (4) root cause analysis message; (5) solution confirmation message. In the event that DRC needs to schedule maintenance to fix the underlying issue, a final message is sent out once this maintenance occurs and a solution is implemented.

Action 4.6c: Administration procedures are in place to ensure consistency in test administration.

Evidence: Procedures for administering the test are documented in the Test Administrator Manual.

The Test Demo and Test Practice items (see Action 4.6a) are also available for teachers to familiarize themselves with the test prior to administration.

WIDA provides webinars and other training courses on their website to orient new test administrators to test administration procedures. The training courses include certification quizzes to ensure that test administrators properly understand the processes prior to administration.

---

[1] For WIDA webinars, see: https://www.wida.us/assessment/ACCESS%202.0/WebinarRecordings.aspx. For webinars focusing on technological issues, see: https://www.wida.us/assessment/video/DuringTestingTechnologyTroubleshooting.aspx. For technology FAQs, see: https://www.wida.us/assessment/ACCESS%202.0/technology.aspx#1.

[2] The WIDA System Status Dashboard is located at: http://status.drcedirect.com/WIDA.

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

Evidence: As detailed in Part I.4.1.1–I.4.1.3 of this report, 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 socioeconomic status, geographic area, or educational background, or introduce other systematic biases.

Action 4.6e: 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 conditions, testing environment, or other unusual occurrences can be found in the *District and School Test Coordinator Manual*. Specific testing situations, including where to start and stop the test, when breaks can be taken, material management protocols in the case of damaged testing materials, 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 for ELLs 2.0 State Checklists, found on the state pages[3] of the WIDA website and within the training course. Frequently asked questions regarding interruptions can be found in the ACCESS for ELLs 2.0 FAQ section of the WIDA website.[4] Additionally, the ACCESS for ELLs 2.0 Training Course highlights common testing irregularities and the resources to use in such circumstances.

Should the Test Administrator have additional questions about how to proceed in the event of a testing interruption or irregularity, the WIDA Client Services Center can be contacted via email or phone at help@wida.us or toll free at 1-866-276-7735.

C4.5. All tasks and items are scored consistently for all test takers.

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

Evidence: Part I.3.3.1.2 specifies the scoring procedure for performance-based tasks in ACCESS 2.0 Online. Raters of performance-based tasks are trained by DRC to appropriately use the Writing and Speaking scoring scales (detailed in Sections 3.3.1.3 and 3.3.1.4, respectively) to score performance-based tasks.

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

---

[3] WIDA state pages can be found at: https://www.wida.us/membership/states/index.aspx
[4] ACCESS for ELLs 2.0 FAQs can be found at:
https://www.wida.us/assessment/ACCESS%202.0/administration.aspx#8

Evidence: Part I.3.3.1 specifies the scoring procedure for ACCESS 2.0 Online. Listening and Reading items are dichotomous and are scored electronically by DRC (see Part I.3.3.1.1).

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

Evidence: Part I.3.3.1.2 specifies the scoring procedure for ACCESS 2.0 Online. Writing and Speaking 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, they rate calibration sets to ensure that they are properly calibrated to the grade-level cluster and task(s).

Action 4.5d: Raters of performance-based 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 Part I.3.3.1.2). 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 that any drift is identified and promptly corrected. For the Speaking test, pre-rated and vetted recalibration sets are administered to raters. Raters take these sets every day to ensure that they are calibrated. Due to the nature of the Speaking test structure, validation sets cannot be seeded into the Speaking scoring queues, so the recalibration sets are needed.

Action 4.5e: Scoring data for performance-based 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 and Speaking tasks (see Section 2.2.10). During operational scoring, these data are monitored daily for quality control purposes.

Action 4.5f: Raters of performance-based tasks are monitored over time to ensure that they apply the scales in a consistent way (internal consistency).

Evidence: Part I.3.3.1.2 details the procedures used by DRC to monitor raters. This includes ongoing quality control checks and procedures, and investigation of any irregularities.

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 that any drift is identified and promptly corrected.

For the Speaking test, pre-rated and vetted recalibration sets are administered to raters. Raters take these sets every day to ensure that they are calibrated. Due to the nature of the Speaking test structure, validation sets cannot be seeded into the Speaking scoring queues, so the recalibration sets are needed.

C4.4. Test items/tasks work appropriately together to measure each test taker's English Language Proficiency.

Action 4.4a: For each domain and grade-level cluster (e.g., Reading 6–8), item and task analyses are performed and psychometric properties of the items and tasks are evaluated to confirm that scores are internally consistent.

Evidence: Listening and Reading reliability are computed using the reliability coefficient described in Thissen (2000). For the Writing and Speaking domains, Cronbach's alpha is computed for each tier and also for each grade-level cluster, across tiers. Section 2.2.10 describes the ways in which test reliability is computed for the domains.

Action 4.4b: For each composite score, psychometric properties are evaluated to confirm that scores are internally consistent.

Evidence: To compute reliability for the composites, a stratified Cronbach's alpha is used. Section 2.3.3 describes the ways in which test reliability is computed for the composites.

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

Evidence: Section 2.2.1.1 describes the Rasch fit statistics that are computed for each item; the statistics are detailed in Table A, *Complete Item/Task Analysis and Summary,* in Section 3.3.

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

Evidence: The Complete Item or Task Analysis and Summary tables provide information on the difficulty of each item or task. Section 2.2.1 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.

Action 4.4e: Items in folders aimed at higher proficiency levels within a stage of the multistage adaptive tests (Listening and Reading) are more difficult than items in folders aimed at lower proficiency levels within the same stage.

Evidence: The Complete Item or Task Analysis and Summary tables include information on item difficulty (see Section 2.2.1.2).

Action 4.4f: Routing and placement procedures are in place to ensure that students are administered a test appropriate to their proficiency level.

Evidence: Part I of this report describes routing rules for Listening (I.3.4.1) and Reading (I.3.4.2), and placement rules for Writing (I.3.4.3) and Speaking (I.3.4.4).

Quality control procedures are in place to ensure that routing rules are implemented with fidelity in the computerized assessment.

Placement rules place students into tiers for Writing (A or B/C) and Speaking (pre-A, A, or B/C) tests. Evidence of the effects of these rules can be found in figures and tables which present raw

score and scale score distributions by tier and across tiers. Descriptions of the raw score distribution and scale score distribution tables can be found in Section 2.2.3 and Section 2.2.4, respectively.

C4.3. The same scale scores obtained by test takers in different years retain the same meaning.

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

Evidence: Each year, while a certain percentage of items on each ACCESS 2.0 Online test form is refreshed, a number of items and tasks are retained from the previous year's assessment for the purpose of scale maintenance. Section 2.2.7 of this report describes the equating procedures used, and Table G presents item-by-item information, including information on which items or tasks were used as anchor items or 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:

  i.   Section 2.2.7 describes the equating summary included in this report.

  ii. Previously used items and tasks (i.e., anchor items) are included on each test form along with new items and 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).

C4.2. ACCESS for ELLs measures English Language Proficiency for all test takers in a fair and unbiased manner.

Action 4.2a: Differential item functioning (DIF) analyses are conducted to determine whether any items or tasks may be biased against certain subgroups.

Evidence: The DIF analysis and summary table provides a summary of the findings of the DIF analyses, which look for measurement bias in test items (see Section 2.2.2). Ethnicity (Hispanic vs. non-Hispanic) and gender DIF analyses are conducted. In the domains of Listening and Reading, DIF analyses are conducted prior to operational testing, using data from the previous year's operational and embedded field test items. In the domains of Writing and Speaking, DIF analyses are conducted using population data, after the conclusion of operational testing.

Action 4.2b: Items that show evidence of DIF are carefully reviewed so that any that indicate bias are not included in the pool of items selected for Listening and Reading, or are removed from future test forms for Speaking and Writing.

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 decision on whether or not the item content is likely to favor or disfavor specific subgroups of students.

C4.1. Test takers are classified appropriately according to the proficiency levels defined in the WIDA English Language Development (ELD) Standards.

Action 4.1a: Distributions of scale scores and proficiency levels for each domain are analyzed to confirm that ACCESS 2.0 Online measures the performance of test takers across the range of English Language Proficiency levels defined by the WIDA ELD Standards. Distributions of raw scores are analyzed where appropriate.

Evidence:

    i.   The distribution of test takers' *raw scores* on ACCESS 2.0 Online for the Writing and Speaking tests, organized by individual test form (e.g., Writing 4–5B/C), shows the extent to which ACCESS 2.0 Online measures the performance of test takers across the range of ELD abilities that each form was designed to assess (see Section 2.2.3).

    ii.   The distribution of test takers' *scale scores* on ACCESS 2.0 Online for each domain, organized by test form, shows that ACCESS 2.0 Online measures the

performance of test takers across the range of ELD abilities that each form was designed to assess (see Section 2.2.4).

iii. The *proficiency level* distribution of test takers' scores on ACCESS 2.0 Online, for each domain, organized by individual test form, shows that ACCESS 2.0 Online measures the performance of test takers across the range of proficiency levels that each form was designed to assess (see Section 2.2.5).

iv. The Test Characteristic Curve graphically shows the relationship between test takers' ability measures (calculated based on test performance using Rasch modeling) on the horizontal axis and expected raw scores on the vertical axis. Test Characteristic Curves are provided for each tier for Writing and Speaking (see Section 2.2.8). (Note that there is no Test Characteristic Curve for Listening and Reading, as the notion of "expected raw score" is meaningless on the adaptive assessment.)

Action 4.1b: Distributions of scale scores and proficiency levels, for each domain and each composite, organized by grade-level cluster, are analyzed to confirm that ACCESS 2.0 Online measures the performance of test takers across the range of English Language Proficiency levels as defined by the WIDA ELD Standards.

Evidence:

i. The distribution of test takers' scale scores on ACCESS 2.0 Online, for each domain and each composite, organized by grade-level cluster, shows that ACCESS 2.0 Online measures the performance of test takers across the range of ELD abilities as described by the WIDA ELD Standards (see Section 2.2.4 and Section 2.3.1).

ii. The proficiency level distribution of test takers' scores on ACCESS 2.0 Online, for each domain and each composite, organized by grade-level cluster, shows that ACCESS 2.0 Online measures the performance of test takers across the range of proficiency levels as defined by the WIDA ELD Standards (see Section 2.2.5 and Section 2.3.2).

iii. The Test Characteristic Curve graphically shows the relationship between test takers' ability measures (calculated based on test performance using Rasch modeling) on the horizontal axis and expected raw scores on the vertical axis. Test Characteristic Curves are provided across each grade-level cluster for Writing and Speaking (see Section 2.2.8). (Note that there is no Test Characteristic Curve for Listening and Reading, as the notion of "expected raw score" is meaningless on the adaptive assessment.)

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

Evidence:

i.   The Test Information Function graphically shows the relationship between ability measure and the accuracy of test scores (see Section 2.2.9). Cut points are marked on the Test Information Function figures.

ii.   Tables provide information on the conditional standard error of measurement (CSEM) at the cut scores for Writing and Speaking (Section 2.2.11).

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

Evidence: Accuracy and consistency statistics are calculated for each domain for the grade-level cluster (see Section 2.2.12).

Action 4.1e: Students are placed into the appropriate proficiency level based on their test scores.

Evidence: A 2016 standard setting study established new cut scores for ACCESS 2.0 (Cook & MacGregor, 2017). A short history of ACCESS cut score setting can be found in Part I.3.3.2.3 of this report.

Action 4.1f: Items and tasks are aligned to the WIDA Standards.

Evidence: See Cook (2007) for evidence of alignment between the WIDA Standards and the ACCESS assessment program. Part I.3.2 details the continuing development of items and tasks for ACCESS 2.0 Online to maintain alignment.

1.3 Summary of Assessment Records Claims, Actions, and Evidence

Table 1

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. Test design and student training procedures ensure that all students are able to interact with the technology of the test.	a. Evidence summarized with claim at 4.6a.
	b. Procedures are in place to address technical issues and interruptions.	b. Evidence summarized with claim at 4.6b.
	c. Administration procedures are in place to ensure consistency in test administration.	c. Test Administration Manual, plus additional evidence summarized with claim at 4.6c.
	d. Procedures are in place to ensure that items and tasks do not have issues with bias or sensitivity.	d. Part I.4.1.1–I.4.1.3
	e. Test administrators document and report any irregularities that may occur so that appropriate action may be taken.	e. Evidence summarized with claim at 4.6e.
5. All items and tasks are scored consistently for all test takers.	a. Raters of performance-based tasks undergo thorough training so that they know how to score appropriately.	a. Part I.3.3.1.2
	b. Listening and Reading items are scored electronically using a carefully checked key.	b. Part I.3.3.1.1
	c. Raters of performance-based tasks are certified, demonstrating that they can score appropriately.	c. Part I.3.3.1.2
	d. Raters of Writing tasks are monitored daily to ensure that they are scoring appropriately.	d. Part I.3.3.1.2
	e. Scoring data for performance-based tasks are analyzed for rater agreement to understand how closely raters agree.	e. Section 2.2.10
	f. Raters of performance-based tasks are monitored over time to ensure that they apply the scales in a consistent way (internal consistency).	f. Part I.3.3.1.2

	a. For each domain and grade-level cluster (e.g., Reading 6–8), item and task analyses are performed and psychometric properties of the items and tasks are evaluated to confirm that scores are internally consistent.	a. Section 2.2.10
	b. For each composite score, psychometric properties are evaluated to confirm that scores are internally consistent.	b. Section 2.3.3
4. Test items/tasks work appropriately together to measure each test taker's English Language Proficiency.	c. Analyses of Rasch model fit statistics are conducted to show that individual tasks perform appropriately.	c. Section 2.2.1.1
	d. Items and task of appropriate difficulty are chosen for each domain.	d. Section 2.2.1
	e. Items in folders aimed at higher proficiency level within a stage of the multistage adaptive tests (Listening and Reading) are more difficult than items in folders aimed at lower proficiency levels within the same stage.	e. Section 2.2.1.2
	f. Routing and placement procedures are in place to ensure that students are administered a test appropriate to their proficiency level.	f. Sections I.3.4.1, I.3.4.2., I.3.4.3, I.3.4.4.
3. The same scale scores obtained by test takers in different years retain the same meaning.	a. A sufficient number of items and tasks are used as anchor items across adjacent years to maintain a consistent scale from year to year.	a. Section 2.2.7
	b. New items and tasks are calibrated with anchor items to ensure that their difficulty measures are on the same consistent scale that is used from year to year.	b. Section 2.2.7
	c. The same scaling equation is applied from year to year to ensure that scale scores are obtained consistently over time	c. Evidence summarized with claim at 4.3c.
2. ACCESS for ELLs measures English Language Proficiency for all test takers in a fair and unbiased manner.	a. Differential item functioning (DIF) analyses are conducted to determine whether any items or tasks are biased against certain subgroups.	a. Section 2.2.2
	b. Items that show evidence of DIF are carefully reviewed so that any that indicate bias are not included in the pool of items selected for Listening and Reading, or are removed from future test forms for Speaking and Writing.	b. Evidence summarized with claim at 4.2b.

1. Test takers are classified appropriately according to the proficiency levels defined in the WIDA English Language Development (ELD) Standards.	a. Distributions of scale scores and proficiency levels for each domain are analyzed to confirm that ACCESS 2.0 Online measures the performance of test takers across the range of English Language Proficiency levels defined by the WIDA ELD Standards. Distributions of raw scores are analyzed where appropriate.	a. Sections 2.2.3, 2.2.4, 2.2.5, 2.2.8
	b. Distributions of scale scores and proficiency levels, for each domain and each composite, organized by grade-level cluster, are analyzed to confirm that ACCESS 2.0 Online measures the performance of test takers across the range of English Language Proficiency levels as defined by the WIDA ELD Standards.	b. Sections 2.2.4, 2.2.5, 2.3.1, 2.3.2, 2.2.9
	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 grade.	c. Sections 2.2.9, 2.2.11
	d. Classification and accuracy analyses are conducted by grade level to confirm that proficiency level classifications are reliable for all domain and composite scores.	d. Section 2.2.12
	e. Students are placed into the appropriate proficiency level based on their test scores	e. Cook & MacGregor (2017) and Part I.3.3.2.3
	f. Items and tasks are aligned to the WIDA Standards.	f. Cook (2007) and Part I.3.2

2 Background and Descriptions for the Presentation of Results

This section describes the tables and figures included in Section 3 of this report.

2.1 Student Participation and Performance

Student participation and performance is detailed in Section 3.2, which has three subsections: *Participation* (3.2.1); *Scale score results* (3.2.2); and *Proficiency level results* (3.2.3).

There are two subsets of students who are included in the descriptions of student participation and performance but are excluded from subsequent analyses. There are a subset of students who were flagged as potentially having experienced test interruptions. Using telemetry data, three variables were selected which might potentially indicate interruption (that is, testing experiences that are outside of regular testing experiences). The interruption indicators WIDA used are 1) longer than expected testing time, 2) number of appearances (i.e. more than 1) of test items, and 3) number of log-ins. Records are flagged if they fall outside of established criteria for any of these three indicators. Students whose records are flagged as interrupted are included in the tables which describe participation in the assessment (these tables are described in Section 2.1.1) but are excluded from all subsequent analyses. Table 2.2 summarizes the numbers of students who are excluded from these analyses.

Table 2.2

Students Excluded from Analysis due to Test Interruptions by Domain and Cluster

Domain	Cluster	No. of Students	Percent
List	1	12,544	9.12%
	2-3	32,446	23.60%
	4-5	18,583	13.52%
	6-8	35,991	26.18%
	9-12	37,916	27.58%
	Total	137,480	100.00%
Read	1	12,298	5.89%
	2-3	39,689	19.01%
	4-5	39,707	19.02%
	6-8	44,875	21.50%
	9-12	72,170	34.57%
	Total	208,739	100.00%
Writ	1	0	0.00%
	2-3	0	0.00%
	4-5	15,581	25.21%
	6-8	19,846	32.11%
	9-12	26,379	42.68%
	Total	61,806	100.00%
Spek	1	16,908	13.86%
	2-3	37,582	30.80%
	4-5	18,276	14.98%
	6-8	25,507	20.91%
	9-12	23,737	19.45%
	Total	122,010	100.00%

Students from the state of Michigan are included in the analyses of participation and performance, but these students are excluded from subsequent analyses, as data from the state of Michigan arrived later than expected, and it was not possible to include these data in all analyses for this report.

2.1.1 Participation

Participation in ACCESS 2.0 Online is shown in three ways: by grade-level cluster, by grade, and, for Writing and Speaking only, by tier. This is the first subsection of *Student Participation and Performance*.

2.1.1.1 *Grade-Level Cluster*

Section 3.2.1.1 gives information on participation by grade-level cluster.

Table 3.2.1.1.1 shows participation across the 37 WIDA states that participated in the ACCESS 2.0 Online operational testing program in 2016–2017. The first row shows the grade-level

cluster, the next 37 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 37states.

Table 3.2.1.1.2 shows participation by grade-level cluster by gender across all 37 states combined, while Table 3.2.1.1.3 shows participation by grade-level cluster by ethnicity across all 36 states.

Table 3.2.1.1.4 shows participation by grade-level cluster and tier for all Writing and Speaking forms.

2.1.1.2 **Grade**
Section 3.2.1.2 gives similar data as in the previous section, but broken out by grade rather than by grade-level cluster.

2.1.2 Scale Score Results

The second subsection of *Student Participation and Performance* provides information on students' scale score results.

2.1.2.1 **Mean Scale Scores Across Domain and Composite Scores Section**
Section 3.2.2.1 shows mean (average) scale scores by grade-level cluster across the eight scores awarded, first for the four domains (Listening, Reading, Writing, and Speaking) and then for the four composites (Oral Language, Literacy, Comprehension, and Overall Composite). In this section, under each average, the number of students in each group is also given.

Table 3.2.2.1.1 shows mean scale scores by grade-level cluster, while Table 3.2.2.1.2 shows the same information broken down by gender, and Table 3.2.2.1.3 shows the same information broken down by ethnicity and race. Following the approach of the Census Bureau, ethnicity is 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. 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 in one racial category (e.g., Asian) who 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, they are labeled Non-Hispanic Multi-racial.

Section 3.2.2.2 shows the mean scale scores broken down by grade rather than by grade-level cluster. Table 3.2.2.2.1 shows mean scale scores by grade, while Table 3.2.2.2.2 shows the same information broken down by gender, and Table 3.2.2.2.3 shows the same information broken down by ethnicity and race.

2.1.2.2 ***Correlations***

Tables 3.2.2.3A through 3.2.2.3E show correlations among the four domain scale scores by grade-level cluster across all tiers, as well as the number of students included in each correlation. Table 3.2.2.3A shows the results for Grade 1, Table 3.2.2.3B shows the results for Cluster 2–3, Table 3.2.2.3C shows the results for Cluster 4–5, Table 3.2.2.3D shows the results for Cluster 6–8, and Table 3.2.2.3E shows the results for Cluster 9–12. Note that all correlations in Tables 3.2.2.3A through 3.2.2.3E are significant at the 0.01 level (2-tailed).

2.1.3 Proficiency Level Results

The third subsection of *Student Participation and Performance* covers results by proficiency level, and shows the distribution of students falling into the six language proficiency levels defined in the WIDA ELD Standards. Section 3.2.3.1 provides the results for Domains, while Section 3.2.3.2 provides the results for Composite scores.

Within each section, results are first presented by grade-level cluster, then by grade. For both, the first table shows the number of students classified into each language proficiency level (count), while the second table shows the results in terms of percentages within each row.

2.2 Analyses of Domain Scores

Section 3.3 presents a series of tables and figures pertaining to scores in the four domains. The tables and figures are organized by grade-level cluster, then by domain, then, where relevant, by tier. Tables and figures are numbered through the text according to their grade-level cluster and domain (and tier, where relevant); each table or figure is then labeled by a letter designation which indicates the table or figure type. Thus in Section 3.3, Table 3.3.1.1.A indicates that the table refers to the first grade-level cluster covered in the section (Grade 1) and the first domain covered (Listening). The letter designation, in this case, A, indicates that the table is a Complete Item Analysis and Summary table—so Table A appears for each relevant grade-level cluster, domain, and tier.

2.2.1 Complete Item or Task Analysis and Summary

Table A provides a summary of the analyses of the items (for Listening and Reading) or tasks (for Writing and Speaking), along with analyses of each item or task. Table A has either two parts (in the case of Listening and Reading) or three parts (in the case of Writing and Speaking). The first part of the table gives a summary of the total set of items or tasks on the form. The second part provides statistics pertaining to the individual items or tasks, and the third part (for Writing and Speaking only) expresses raw score distributions by task.

Statistics included across these three parts include item or task difficulties in logits, the number of items or tasks on the form, the average p-value (for forms with selected-response items), and the Rasch model fit statistics.

For Listening and Reading, Table A provides information on every item in the grade-level cluster. For Writing, Table Ai provides information on Tier A for the grade-level cluster, and Table Aii provides information on Tier B/C for the grade-level cluster. For Speaking, Table A provides information on every task in the grade-level cluster.

2.2.1.1   *Fit Statistics*

All Rasch analyses were conducted using the Rasch measurement software program *Winsteps* (Linacre, 2006). When speaking of the measure of person 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 *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.

Fit statistics for the Rasch model are calculated by comparing the observed empirical data with the data that would be expected to be produced by the Rasch model. Outfit mean square statistics are influenced by outliers. For example, a difficult item that some low-ability examinees get correct—for reasons unknown—will have a high outfit mean square statistic. Infit mean square statistics are influenced by unexpected patterns of observations by persons on items that are roughly targeted on them 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 of greater concern.

Linacre (2002) provides more guidance on how to interpret these statistics for dichotomous 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 are "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 vary differently from dichotomous items, it is not as straight forward to apply this same guidance to interpret these fit statistics to performance tasks that were scored polytomously. Some performance tasks that were designed to elicit a restricted range of performances (for example, it is expected that most students will get the highest score on a very easy task) 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.

2.2.1.2 *Structure of Complete Item Analysis and Summary Table*
The first section of Table A, the *Complete Item/Task Analysis and Summary* provides information about the total set of items or tasks, and includes the item type (selected response or constructed response), the average item difficulty (in logits), the number of items, the average p-value (for Listening and Reading only), the average infit mean square, and the average outfit mean square.

The second section of Table A presents results of the analyses of all of the items or tasks on the test form. The first column provides the unique item name. The second column in this section presents the item difficulty in logits. The third column provides information of whether the item or task served as an anchor item or task. 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.

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

2.2.2 DIF Analysis and Summary

Differential item analysis (DIF) attempts 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, DIF attempts to find items that may be functioning differently for different groups based on criteria irrelevant to what is being tested. The performance of students on ACCESS for ELLs items and tasks is compared by dividing students into two different groupings: first, males versus females; second, students of Hispanic ethnic background versus students of all other backgrounds. Students for whom gender or ethnicity[5] was unknown were excluded from both analyses. Two commonly used procedures for detecting

---

[5] In the dataset, Hispanic ethnicity, as well as each of the race categories, are coded as a binary variable (Y/N). Ethnicity information is counted as "Unknown" in cases where the student is recorded as N for Hispanic ethnicity and also N for every race category.

DIF were used: one for dichotomously scored items (Listening and Reading), conducted prior to operational testing, and one for polytomously scored items (Writing and Speaking), conducted on population data subsequent to the close of operational testing.

2.2.2.1 *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, a similar percentage of students in each group should get the item correct at any ability level (based on performance on the total test). The Mantel-Haenszel Chi-square statistic is used to check the probability that the two groups performed comparably 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.

The existing Mantel-Haenszel procedure was designed for fixed forms, where all test takers took exactly the same set of items, therefore, the test takers can be matched on the number-correct score when computing the M-H statistic. In the multistage computerized adaptive test (CAT) condition, however, not all students took exactly the same set of items, thus it is not possible to match students on the number-correct score. Instead, a CAT M-H DIF procedure (Zwick, Thayer, Wingersky, 1993) was used to examine DIF for the Listening and Reading domains. First, the examinee's expected true score for the entire item pool is derived. To derive the expected true score, each examinee's Rasch ability estimate is transformed into the expected true score metric by calculating the sum of the item response functions in the operational item pool, which is evaluated at the estimated ability level of the test taker. The expected true score of the examinees are used as the matching variable for the M-H DIF procedure. Once examinees are matched on the expected true score, the ordinary M-H DIF procedure and the ETS evaluation criterion for severity of M-H DIF can be applied. In CAL's implementation of this method, examinees are matched for M-H DIF analysis on the basis of this expected true score using two-unit intervals, as recommended by Zwick and Bridgeman (2014). A two-step purification process was used in conducting the DIF analysis; 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.

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 requires further review of the item. We follow the guidance provided by ETS to classify items into DIF levels as follows:

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

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

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

2.2.2.2 *Polytomous Items*

For polytomous items (i.e., Writing and Speaking tasks), a similar approach is used. It is based on the Mantel-Haenszel Chi-square statistic and the standardized mean difference following procedures again developed by ETS (Zwick, Donoghue, & Grima, 1993; Allen, Carlson, & Zalanak, 1999). The DIF procedures developed by ETS for polytomous items were used to identify 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 groups 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 interpretation challenges. 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 for ELLs 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, *DIF Analysis and Summary*, 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 non-Hispanic other ethnicities. Even though DIF may be negligible (category A or AA), this table shows the number of items that favored one group or the other at all levels of DIF. Optimally, even when items are all in category A or AA, there

should be roughly an even number of items favoring each of the two groups to ensure that there is no systematic biasing test effect across items.

Items and tasks which show C-level (or CC-level) DIF are investigated by a team of content experts to determine if any construct-irrelevant factors can be identified that may contribute to DIF. For dichotomous items, DIF analysis was conducted prior to operational testing, and items which show C-level DIF can be flagged before students' testing begins. For polytomous items, if content experts identify concerning construct-irrelevant factors, the task will be removed from the test for the next operational year.

2.2.3 Raw score distribution for Speaking and Writing

Figure C and Table C provide raw score information for Speaking and Writing only. Raw score distribution is presented by grade-level cluster and also by grade-level cluster and tier. For each test form, Figure C 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 received each raw score.

Table C shows, by 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.

2.2.4 Scale Score Distribution

Figure D and Table D relate to the ACCESS for ELLs scale scores on each test form. For each test form, raw scores were converted to vertically-equated scale scores. Scale score distribution is presented by grade-level cluster. For Writing and Speaking, it is also presented by grade-level cluster and tier.

Thus, for each test form, Figure D shows the distribution of the scale scores. The horizontal axis shows the scale scores based on performances on the test form. To provide a 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 received each scale score.

Table D shows, by grade and by total for the grade-level cluster:

- the number of students in the analyses (count),

- 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.

2.2.5 Proficiency Level distribution

Figure E and Table E provide information on the proficiency level distribution of the students who took the test form based on their performance. Proficiency level distribution is presented by grade-level cluster. For Writing and Speaking, it is also presented by grade-level cluster and tier. In Figure E, 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 E shows, by grade and by total for the grade-level cluster:

- the WIDA proficiency level designation (1 to 6),

- the number of students (count) whose performance on the test form placed them into that proficiency level in the domain being tested, and

- the percentage of students, out of the total number of students taking the form who were placed into that proficiency level in the domain being tested.

2.2.6 Raw to Scale Score Conversion for Speaking and Writing

The next table in this section, Table F, presents the raw score to scale score conversion table for the test form for Speaking and Writing only.

The first column shows all possible raw scores. The following column(s) show the corresponding scale score for the highest grade in the grade-level cluster. The next column shows the conditional standard error of measurement (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 fell below 100. In such cases, the lower bound has been set at 100, which has been determined to be the lowest score possible on the scale.

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. The standard error and the lower and upper bounds reported in Table F reflect the truncated score.

2.2.7 Equating Summary

Each year a certain percentage of items on each ACCESS for ELLs test form is refreshed, as determined by the refreshment plan for that Series. For Series 401, Reading, Writing, and Speaking were refreshed while Listening domain was not refreshed. An equating procedure known as common-item equating is used to equate the results on 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 items appearing on ACCESS 2.0 Online Series 401 also appeared on Series 400. All items 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. After the first equating run, some items that were originally anchored proved to have changed in their difficulty measure. This change is measured by the "Displacement" statistic. This statistic shows the difference between the difficulty value of the anchored item and what its difficulty value would have been had it not been anchored. 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 Listening and Reading items, and 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.

A pre equating design was used to conduct the annual equating for Listening and Reading domains. This design allows for Listening and Reading item parameters to be available for setting up the computer adaptive engine prior to operational administration. For the Listening domain, although the test was not refreshed in Series 401, student data from the Series 400 operational administration were used to refine the parameters. For the Reading domain, student data collected from the Series 401 Reading embedded field test were used to conduct the equating analyses. All available student data at the time the equating analyses were conducted were included in the analyses.

For the Writing domain, the equating analysis was conducted using Series 401 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 so that it was proportional to the Series 400 population for the Writing domain, by grade and tiered form.

For the Speaking domain, student data from the Series 401 appended Speaking field test administration were used to conduct the initial common-item equating. These initial item parameters were then verified using Series 401 operational data collected during the early testing window. The Speaking verification study was conducted using a random sample drawn with a target sample size of 3,000 students per grade-level cluster. The Speaking verification sample was drawn such that it was proportional to the Series 400 population for the Speaking domain, by grade and tier.

Table G presents a summary of the equating and verification 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 items) is shown, as well as the standard deviation of those items. For the Speaking domain, all item parameters were initially anchored to the field test measures. 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.

The third section of Table G gives information about 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.

In general, and for longer tests such as Listening and Reading, the greater the representation across the difficulty range of anchor items, the more trustworthy the equating results will be. For the Writing and Speaking domains, which are shorter and performance-based, and which have additional content and exposure considerations in terms of item refreshment, this rule-of-thumb may not always apply. In addition, the number of anchors is also a function of the targeted refreshment plan which can differ by Series and by domains.

For the Listening and Reading domains, the Series 400 parameters were derived from the scale maintenance studies in preparation for ACCESS transitioning from ACCESS 1.0 to 2.0 (see Center for Applied Linguistics, 2016). The scale maintenance studies used data from stand-alone field tests with small samples and the field tests were administered under a slightly different condition; therefore, there was a plan to refine these parameters using Series 400 operational data. However, due to technical issues related to interrupted test sessions which occurred during the Series 400 operational administration year, WIDA ultimately made the decision to retain the field test parameters in order to meet the reporting timeline (see the Annual Technical Report for Series 400 Online ACCESS for ELLs for further detail). Thus, it is not surprising that many of the repeated Listening and Reading items that were initially anchored needed to be re-estimated using the Series 401 Reading embedded field test data. In other words, had the refined parameters been used in the Series 400 Listening and Reading reporting, the final number of items being anchored in the Series 401 equating analyses would have been greater.

Note that, for the Writing and Speaking tasks, this table has a fourth section, which provides the anchored step measures for the score on each task. For the ACCESS Writing and Speaking tasks, a Rasch-grouped rating scale model is used (see Part I Section 3.3.2.2.). For Writing, the step difficulties 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.

2.2.8 Test Characteristic Curve

Test characteristic curves graphically show the relationship between the ability measure (in logits) on the horizontal axis and the expected raw score on the vertical axis. Five vertical lines indicate the five cut scores for the highest grade in the cluster for the test form, dividing the figure into six sections for each of the WIDA proficiency levels (Levels 1–6) for the domain being tested. (Note that for some domains for Tier A tests, it was not possible to place into all six language proficiency levels. As would be expected, higher raw scores are required to be placed into higher language proficiency levels. The relative width of each section between the cut score lines, however, gives an indication of how many items on that form must be answered correctly (or points on the Writing section must be earned) to be placed into a WIDA language Proficiency Level.

As the Listening and Reading assessments are multistage adaptive tests, raw scores are not a meaningful aspect of these tests, so no test characteristic curve is presented for these domains.

2.2.9 Test Information Function

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. Figure I shows graphically how well the test is measuring across the ability measure spectrum. High test information values indicate more accuracy in measurement. Figure I 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 errors of measurement.

Five vertical lines in Figure I indicate the five ACCESS cut scores for the highest grade in the grade-level cluster for the test form, dividing the figure into six sections for each of the WIDA proficiency levels (1–6) for the domain being tested. The ACCESS cut scores lines are presented

along with the test information function to facilitate the interpretation of the test information curves. The test information curve and the corresponding ACCESS cut score lines are both expressed on the ACCESS logit scale. Note that for Speaking, in Tier pre-A, all scores fall in the PL 1.0 range, so there are no vertical lines expressing the cuts between proficiency levels.

2.2.10 Reliability of Domain Scores

2.2.10.1 *Listening and Reading Domains*

In the Listening and Reading domains, Table J presents reliability information based on Item Response Theory. The table shows:

- the number of students (count),

- the number of items,

- Rasch Reliability (as a measure of internal consistency)

For tests administered using a multistage adaptive method, a reliability coefficient based on classical test theory such as Cronbach's coefficient alpha cannot be applied because not all students take the same set of items. Reliability for Listening and Reading was estimated using a method by Thissen (2000) by grade-level cluster:

$$\bar{\rho} = \frac{\sigma^2_{\theta-average(CSEM^2_{observed})}}{\sigma^2_\theta}$$

where

$\bar{\rho}$ is the average reliability,

σ^2_θ is the variance of the distribution of student measure,

$CSEM^2_{observed}$ is the squared observed conditional standard errors of measurement for each student

This estimate is equivalent to the Rasch separation reliability coefficient (Linacre, 1999). Like Cronbach's alpha, the Rasch reliability coefficient is an estimate of the ratio of "true measure variance" to "observed measure variance." To obtain these values, item parameters and population student data were used as inputs in the Winsteps program. The Rasch separation reliability coefficient can be interpreted like Cronbach's coefficient alpha. It expresses how well the items on a test appear to measure the same construct.

2.2.10.2 *Speaking and Writing Domains*
In the Speaking and Writing domains, Table J presents reliability and accuracy information based on classical test theory. Table J is provided for each tier, and it is also provided, in a different format, to express weighted reliability for each grade-level cluster.

For each tier, the table shows:

- the number of students (count),

- the number of tasks,

- for Writing, the response mode (keyboarded or handwritten)

- Cronbach's coefficient alpha (as a measure of internal consistency), and

- the classical standard error of measurement (SEM) in terms of raw scores.

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

Cronbach's alpha is also affected by the distribution of ability within the group of students tested. All things being equal, the greater the heterogeneity of abilities within the group of

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

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

For the Writing test, a slight modification was made in the estimation of the Cronbach's alpha for tiered forms that have differential weighting across tasks. This modification is an attempt to take into account that some tasks are weighted more than others when deriving student's ability measure for these tiered forms. For writing tasks with weight greater than one, student's response to the tasks are replicated as a function of their weights. For example, the fourth task is weighted three in Writing G1A, therefore, student's response to this task was repeated three times when computing the Cronbach's alpha. This modification means that the number of pieces of information or Writing tasks that contribute to the estimation of the Cronbach's alpha for G1A is actually six, not four.

Table J also presents the SEM based on classical test theory for Speaking and Writing. 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 (SD) of the test scores. It is calculated as

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

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

For the Writing and Speaking tests, information on interrater reliability for a sample of 20% of task raters is also provided in Table J. This portion of the table shows, for each of the tasks, the percent 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 Speaking, when the two raters agreed on the rating, an exact agreement was counted. If the two raters were different by one point, an adjacent agreement was counted. Otherwise, the raters are 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 3+, then a score of 3, 3+ or 4 from Rater 2 is categorized as agreement). Scores that are one whole score point apart are categorized as adjacent (for example, if Rater 1 assigns a 3+, then a score of 2+ or 4+ from Rater 2 is categorized as adjacent). Note that for Writing, interrater reliability is computed independently between ratings of keyboarded and handwritten responses.

For each grade-level cluster in Writing and Speaking, Table J is presents a single reliability value for the grade-level cluster. To produce this single value, values for Cronbach's alpha for each of the tiers in the grade-level cluster are weighted by the number of students who were administered the tier form, and a weighted average is expressed in Table J.

2.2.11   Conditional Standard Errors of Measurement at Cut Score

Table K presents information on the conditional standard errors of measurement (CSEM) at the most important points at which decisions are made about students based on performance on ACCESS—the cut points between language proficiency levels. Because the cut points depend on the grade level, information is provided for each grade level within a grade-level cluster. The leftmost column shows the cut (e.g., 1/2, which is the cut score between Proficiency Level 1 and Proficiency Level 2).

The second column shows the grade level. The third 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 Writing and Speaking.

For Writing and Speaking, the values are presented by tier. From Table K, it is possible to identify how well the different Writing and Speaking 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 B/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. Information from Table K provides comparable information on how well the two tier forms are targeted to provide the most accurate

measure in order to place their intended examinees into the language proficiency levels that they target.

Since the Listening and Reading tests are multistage adaptive tests, the CSEM will vary for the same scale score since students were routed to take different items; the mean, standard deviation, minimum, and maximum of the CSEM of all students at the cut scores are presented instead. Note that there are some rare cases where there are no observed scale scores corresponding to the cut score values, therefore these descriptive statistics cannot be provided.

2.2.12   Accuracy and consistency

Table L presents three sections of information related to the accuracy and consistency of placement into the WIDA language proficiency levels for each domain. A separate table is provided for each grade in a grade-level cluster. The first section provides overall indices related to the accuracy and consistency of classification, as well as Cohen's kappa. The second section of information shows accuracy and consistency information conditional on proficiency level. The third provides indices of classification accuracy, including the false positives and the false negatives, 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 (e.g., determining which students have reached Proficiency Level 5). Note that the consistency is generally higher at the cut points than over the proficiency levels.

There are several cases where there were no test takers who were placed into the proficiency level and accuracy of classification conditional on that level cannot be computed. In these cases, 'N/A' 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 the software program that is used (BB-CLASS, see below). In such cases, a hyphen (-) has been placed in the table.

For each domain, tables are provided that indicate estimates of the accuracy and consistency of classification of examinees into the WIDA language proficiency levels based on their performance on the test. It is important to know the reliability of any student's test score and the degree of precision with which it has been measured (i.e., the estimate of the invariant standard error of measure [SEM] of classical test theory and the estimate of the variable conditional standard errors of the Rasch measurement model). However, because decisions about students are ultimately made on the basis of their classification into language proficiency levels according to their performance on ACCESS, it is important to know how well these classifications are made. The analyses that were used utilize the methods outlined and implemented in Livingston and Lewis (1995) and Young and Yoon (1998) as implemented in the software program BB-CLASS (Brennan, 2004) (cf. also Lee, Hanson, & Brennan, 2002).

In the approach of Livingston and Lewis (1995), the *accuracy* of a decision is the extent to which decisions made on the basis of the administered test (i.e., the observed scores) would agree with those made if each student could somehow be tested with all possible parallel forms of the

assessments; that is, the examinee's "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 percentage of all students who would be classified into the same language proficiency level by both the administered test and either the true score distribution (accuracy) or a parallel test (consistency). (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 language proficiency level. These indices examine the percentage of students classified by both tests into a proficiency level divided by all students classified into that proficiency level according to either 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 Proficiency Level 5 ("Bridging"), one can look at the accuracy index provided in Table L for the cut score 4/5.

The Livingston and Lewis procedure requires that the reliability estimate of the test form be provided in estimating the classification consistency and accuracy statistics. For Listening and Reading, the Rasch reliability estimates by grade-level clusters were used in the procedure. Since the Writing and Speaking tests were tiered, it was necessary to produce a single reliability estimate across tiers for the Livingston and Lewis procedure. This is a weighted reliability

estimate across tiers. In other words, it is the average reliability weighted by the number of students who were administered that tier form. Thus, Table L, based on the information from Table J, provides the number of students and the reliability estimate for each tier. The final column presents the weighted reliability, an estimate of the reliability of the scale scores across the tiers.

2.3 Analyses of Composite Scores

In Section 3.4, analyses of the four composites—Oral Language, Literacy, Comprehension, and Overall Composite—are presented. Tables and figures pertaining to the composite scores are presented by grade-level cluster.

2.3.1 Scale Score distribution for Composites

Figure A and Table A provide scale score distributions for each of the composites, for each grade-level cluster.

Figure A 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 received each scale score.

Table A shows, by grade and by total for the grade-level cluster:

- the number of students in the analyses (count),
- 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.

2.3.2 Proficiency Level distribution for Composites

Figure B and Table B provide information on the proficiency level distribution for each of the composites for each grade-level cluster.

In Figure B, 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 B shows, by grade and by total for the grade-level cluster:

- the WIDA proficiency level designation (1 to 6),
- the number of students (count) whose performance on the test form placed them into that proficiency level in the domain being tested, and

- the percentage of students, out of the total number of students taking the form who were placed into that proficiency level in the domain being tested.

2.3.3 Reliability of Composites

To estimate the reliability of the composite scores, a stratified Cronbach's alpha coefficient (e.g., Rudner, 2001; Kamata, Turhan, & Darandari, 2003; Kane & Case, 2004;) is computed, weighted by the contribution of each domain score into the composite. Specifically, the formula is

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

where

k = number of components j

w_j = weight of component j

σ_j^2 = variance of component j

σ_c^2 = variance of composite

ρ_j = reliability coefficient of component j

The data used to compute the stratified Cronbach's alpha is provided in Table C. 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 domains forming the composite (note that these are the weighted reliabilities across the tiers for Speaking and Writing) and the reliability of the composite. Unlike the weighted composite, which is an average, the stratified alpha reflects the fact that there are two to four measures being combined into one single measure. Thus, the reliability of the composite score will be higher than the reliability of any single sub-score within the composite.

The stratified Cronbach's alpha, presented in Table C, was also used to produce the *Accuracy and Consistency* classification tables of the composites (Table D).

2.3.4 Accuracy and Consistency of Composites

Table D presents three sections of information related to the accuracy and consistency of placement into the WIDA language proficiency levels for each composite score. The first section provides overall indices related to the accuracy and consistency of classification, as well as Cohen's kappa. The second section shows accuracy and consistency information conditional per proficiency level. The third section provides indices of classification accuracy, including the false positives and false negatives, 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 (e.g., determining which students have reached Proficiency Level 5). Note that the consistency is generally higher

at the cut points than over the proficiency levels. For practical purposes, the primary score used for such decisions is the Overall Composite score.

As noted above in 2.2.12, there may be cases where there are no test takers placed into the proficiency level and accuracy of classification conditional on that level cannot be computed. In this case 'N/A' has been placed in the table. In addition, there may be 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 the software program that is used. In such cases, a hyphen (-) has been placed in the table.

3 Results By Grade Cluster

3.1 Guide to Tables and Figures

The remainder of the subsections of this report (3.2, 3.3, and 3.4) present tables and figures describing, respectively, students' participation and performance, analyses of the scores in the four language domains (Listening, Reading, Writing, and Speaking), and analyses of the scores in the four composites (Oral Language, Literacy, Comprehension, and Overall).

For ease of navigation through these subsequent sections, this section provides a visual overview of the numbered tables and figures. 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.

3.1.1 Guide to 3.2, *Student Participation and Performance*

Tables 3.1A-C provide a visual overview of the tables included in Section 3.2. There are three subsections:

3.2.1 *Participation* presents distributions of students' participation by grade and by grade-level cluster. Student participation by grade and grade-level cluster is further broken down by state, by gender, by ethnicity, and finally by tier and domain combined. Table 3.1A presents the tables included in this subsection.

Table 3.1A

Table Numbering System for Section 3.2.1, *Participation*

	3.2.1.1. By Grade-level Cluster	3.2.1.2. By Grade
By State	Table 3.2.1.1.1	Table 3.2.1.2.1
By Gender	Table 3.2.1.1.2	Table 3.2.1.2.2
By Ethnicity	Table 3.2.1.1.3	Table 3.2.1.2.3
By Tier by Domain	Table 3.2.1.1.4	Table 3.2.1.2.4

3.2.2 *Scale Score Results* presents distributions of students' scale score results. These are again presented by grade and grade-level cluster. Student scale score results by grade and grade-level cluster are further broken down by gender and by ethnicity, and correlations among scale score results are presented. Table 3.1B presents the section and table numbering system for this section.

Table 3.1B

Section and Table Numbering System for Section 3.2.2, *Scale Score Results*

Mean Scale Scores Across Domain and Composite		
	Section 3.2.2.1. **By Grade-level Cluster**	**Section 3.2.2.2.** **By Grade**
Alone	Table 3.2.2.1.1	Table 3.2.2.2.1
And by Gender	Table 3.2.2.1.2	Table 3.2.2.2.2
And by Ethnicity	Table 3.2.2.1.3	Table 3.2.2.2.3
Section 3.2.2.3 Correlations Among Scale Scores by Grade-level Cluster		

3.2.3 *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 3.1C lists the numbers of subsections. Each subsection contains a table expressing descriptive statistics as counts (Table A) and as percentages (Table B).

Table 3.1C

Section Numbering System for Section 3.2.3, *Proficiency Level Results*

		By Grade-Level Cluster	**By Grade**
		For each, distributions by count and by percent	
3.2.3.1 Domains			
3.2.3.1.1	**Listening**	3.2.3.1.1.1	3.2.3.1.1.2
3.2.3.1.2	**Reading**	3.2.3.1.2.1	3.2.3.1.2.2
3.2.3.1.3	**Writing**	3.2.3.1.3.1	3.2.3.1.3.2
3.2.3.1.4	**Speaking**	3.2.3.1.4.1	3.2.3.1.4.2
3.2.3.2 Composites			
3.2.3.2.1	**Oral Composite**	3.2.3.2.1.1	3.2.3.2.1.2
3.2.3.2.2	**Literacy Composite**	3.2.3.2.2.1	3.2.3.2.2.2
3.2.3.2.3	**Comprehension Composite**	3.2.3.2.3.1	3.2.3.2.3.2
3.2.3.2.4	**Overall Composite**	3.2.3.2.4.1	3.2.3.2.4.2

3.1.2 Guide to 3.3, *Analysis of Domain Scores*

An overview of the tables and figures in *Section 3.3 Analysis of Domain Scores* is provided in Table 3.1D. This section is organized by grade-level cluster, and the figure provides an overview of the detail in any given grade-level cluster. Note that the headers within the figure include an "X" to denote the grade-level cluster—for example, "Reading 3.3.X.2" would read "Reading 3.3.1.2" for grade 1, "Reading 3.3.2.2." for grades 2–3, and so on.

Table 3.1D

Naming conventions for tables and figures in Section 3.3. *Analysis of Domain Scores*

	Table	Figure	3.3.X.1. Listening	3.3.X.2. Reading	3.3.X.3. Writing		3.3.X.4. Speaking	
					By tier	*Across tiers*	*By tier*	*Across tiers*
Complete Item/Task Analysis and Summary	Table A		√	√	√	n/a	n/a	√
DIF Analysis and Summary	Table B		√	√	√	n/a	√	n/a
Raw Score Distribution (Spek and Writ only)	Table C	Figure C	n/a	n/a	√	√	√	√
Scale Score Distribution	Table D	Figure D	√	√	√	√	√	√
Proficiency Level Distribution	Table E	Figure E	√	√	√	√	√	√
Raw Score to Scale Score Conversion with CSEM (Spek and Writ only)	Table F		n/a	n/a	√	n/a	√	n/a
Equating Summary	Table G		√	√	√	n/a	n/a	√
Test Characteristic Curve		Figure H	n/a	n/a	√	√	√	√
Test Information Function		Figure I	√	√	√	√	√	√
Reliability	Table J		√	√	√	√	√	√
CSEM at Cut Score Points	Table K		√	√	n/a	√	n/a	√
Accuracy and Consistency (by grade)	Table L		√	√	n/a	√	n/a	√

Note: *By tier* means that a table is presented for each tier of the grade-level cluster. *Across tiers* means that one table is presented for the grade-level cluster, including information from all tiers.

The left column of Table 3.1D lists the content of the table of figure. The next two columns list the letter designations for the tables or figures as applicable. Check marks under a domain subheading indicate that the table and/or figure is included in the report or not applicable (n/a).

For the adaptive domains (Listening and Reading), tables or figures related to raw scores are not provided.

For the tiered domains (Writing and Speaking), differing subsets of tables are provided either by tier or across tiers in a grade-level cluster.

If a table or figure is provided multiple times with the same grade-level cluster and domain, it is denoted with a roman numeral—e.g. Table 3.3.1.4.Di provides scale score distribution information for Speaking Grade 1 pre-A; Table 3.3.1.4.Dii provides the same information for Speaking Grade 1 Tier A, and so on. For Writing, the two tables describing the individual Writing tasks (Table A, *Complete Task Analysis and Summary*, and Table B, *DIF Analysis and Summary*) are provided once for Tier A and once for Tier B/C. For Speaking, due to the design of the tiered assessment, Table A, *Complete Task Analysis and Summary*, is provided once across the three tiers. Table B, *DIF Analysis and Summary*, is provided separately for each tier. For both Writing and Speaking,. Figures and Tables C-E (*Raw Score Distribution*, *Scale Score Distribution*, and *Proficiency Level Distribution*) are provided first for each tier and then for the entire grade-level cluster. Table F, *Raw Score to Scale Score Conversion with SEM* is provided by Tier only. Table G, *Equating Summary*, is provided by tier for Writing and across tiers for Speaking. Figures H and I (*Test Characteristic Curve*; *Test Information Function*), and Table J (*Reliability*) are provided by each tier and also for the entire grade-level cluster. Finally, Tables K and L (*CSEM at Cut Score Points*, *Accuracy and Consistency*) are provided for the grade-level cluster.

3.1.3 Guide to 3.4, Analysis of Composite Scores

As with Section 3.3, Section 3.4. is first organized by grade-level cluster, and then by each of the four composites (Oral Language, Literacy, Comprehension, Overall). For each grade-level cluster/composite combination (e.g. Grade 4-5, Comprehension), the figures and tables presented in Table 3.1E below are provided.

Table 3.1E
Naming conventions for tables and figures in Section 3.4. *Analysis of Composite Scores*

Scale Score Distribution	Figure A	Table A
Proficiency Level Distribution	Figure B	Table B
Reliability		Table C
Accuracy and Consistency		Table D

3.2 Student Participation and Performance

3.2.1 Participation

3.2.1.1 Participation by Grade-level Cluster

3.2.1.1.1 By State

Table 3.2.1.1.1

Participation by Cluster by State S401 Online

State	Cluster					Total
	1	2-3	4-5	6-8	9-12	
AK	1,015	2,118	1,401	1,604	1,599	7,737
AL	2,138	3,747	1,799	1,847	2,118	11,649
CO	7,819	16,357	11,831	13,238	12,523	61,768
DC	897	1,495	796	1,003	1,419	5,610
DE	1,920	3,517	1,439	1,167	1,452	9,495
GA	12,795	23,783	12,432	12,924	12,058	73,992
HI	1,245	2,134	1,070	1,814	1,544	7,807
ID	2,290	3,428	2,596	2,661	2,392	13,367
IL	15,958	45,157	25,092	23,025	20,902	130,134
IN	7,016	12,765	5,748	7,620	8,821	41,970
KY	3,321	5,680	3,035	3,207	4,147	19,390
MA	5,734	11,696	8,582	9,493	10,769	46,274
MD	9,779	17,899	9,007	9,510	13,820	60,015
ME	458	1,022	892	1,066	1,273	4,711
MI	9,948	20,976	15,429	18,828	19,343	84,524
MN	7,948	16,601	11,001	12,108	12,248	59,906
MO	4,075	7,841	4,905	4,885	4,416	26,122
MP	123	240	272	443	224	1,302
MT	242	781	570	648	340	2,581
NC	12,082	26,978	11,127	13,644	15,637	79,468
ND	370	651	408	581	715	2,725
NH	477	978	615	659	865	3,594
NJ	10,563	16,209	8,538	10,053	14,703	60,066
NM	4,751	10,358	7,270	8,287	7,583	38,249
NV	7,834	18,000	12,208	14,595	11,743	64,380
OK	3,840	7,368	3,780	4,586	4,856	24,430
PA	4,176	8,565	6,839	9,208	12,286	41,074
RI	1,051	2,223	1,354	1,685	2,431	8,744
SC	3,825	8,671	7,823	10,425	8,630	39,374
SD	561	1,120	491	628	853	3,653
TN	5,448	12,164	6,843	7,468	6,949	38,872
UT	5,309	10,905	6,583	6,443	5,705	34,945
VA	9,626	22,681	12,510	13,565	18,465	76,847
VI	132	231	192	245	223	1,023
VT	197	390	199	208	301	1,295
WI	5,668	11,841	8,750	8,410	7,411	42,080
WY	350	712	314	377	433	2,186
Total	170,981	357,282	213,741	238,158	251,197	1,231,359

3.2.1.1.2 By Gender

Table 3.2.1.1.2

Participation by Cluster by Gender S401 Online

Cluster		Gender			Total
		F	M	Missing	
1	Count	79,220	88,591	3,170	170,981
	% within Cluster	46.3%	51.8%	1.9%	100.0%
2-3	Count	165,166	186,343	5,773	357,282
	% within Cluster	46.2%	52.2%	1.6%	100.0%
4-5	Count	92,320	118,209	3,212	213,741
	% within Cluster	43.2%	55.3%	1.5%	100.0%
6-8	Count	102,951	131,834	3,373	238,158
	% within Cluster	43.2%	55.4%	1.4%	100.0%
9-12	Count	107,652	138,931	4,614	251,197
	% within Cluster	42.9%	55.3%	1.8%	100.0%
Total	Count	547,309	663,908	20,142	1,231,359
	% within Cluster	44.4%	53.9%	1.6%	100.0%

3.2.1.1.3 By Ethnicity

Table 3.2.1.1.3

Participation by Cluster by Ethnicity S401 Online

Cluster		Hispanic/Non-Hispanic			Total
		Hispanic	Other	Unknown	
1	Count	110,766	52,683	7,532	170,981
	% within Cluster	64.8%	30.8%	4.4%	100.0%
2-3	Count	240,919	102,640	13,723	357,282
	% within Cluster	67.4%	28.7%	3.8%	100.0%
4-5	Count	145,104	59,671	8,966	213,741
	% within Cluster	67.9%	27.9%	4.2%	100.0%
6-8	Count	158,434	69,241	10,483	238,158
	% within Cluster	66.5%	29.1%	4.4%	100.0%
9-12	Count	161,945	76,987	12,265	251,197
	% within Cluster	64.5%	30.6%	4.9%	100.0%
Total	Count	817,168	361,222	52,969	1,231,359
	% within Cluster	66.4%	29.3%	4.3%	100.0%

3.2.1.1.4 By Tier by Domain

Table 3.2.1.1.4

Participation by Cluster by Tier by Domain S401 Online

Cluster			Domain	
			Writing	Speaking
1	Tier	Pre-A	-	5,726
		A	133,261	57,484
		BC	37,698	107,764
	Total		170,959	170,974
2-3	Tier	Pre-A	-	17,191
		A	108,599	89,259
		BC	248,613	250,817
	Total		357,212	357,267
4-5	Tier	Pre-A	-	6,315
		A	45,598	29,848
		BC	168,132	177,570
	Total		213,730	213,733
6-8	Tier	Pre-A	-	13,101
		A	96,875	55,371
		BC	141,270	169,669
	Total		238,145	238,141
9-12	Tier	Pre-A	-	22,978
		A	105,343	112,487
		BC	145,823	115,703
	Total		251,166	251,168

3.2.1.2 *Participation by Grade*

3.2.1.2.1 By State

Table 3.2.1.2.1

Participation by Grade by State S401 Online

State	\multicolumn{12}{c}{Grade}	Total
	1	2	3	4	5	6	7	8	9	10	11	12	
AK	1,015	1,034	1,084	791	610	536	536	532	544	433	330	292	7,737
AL	2,138	1,986	1,761	1,143	656	586	661	600	817	585	449	267	11,649
CO	7,819	8,276	8,081	6,431	5,400	4,432	4,459	4,347	4,632	3,289	2,585	2,017	61,768
DC	897	764	731	488	308	318	324	361	679	346	240	154	5,610
DE	1,920	1,824	1,693	943	496	424	378	365	661	400	229	162	9,495
GA	12,795	12,205	11,578	7,506	4,926	4,345	4,451	4,128	6,036	3,253	1,675	1,094	73,992
HI	1,245	985	1,149	599	471	600	619	595	652	376	300	216	7,807
ID	2,290	1,723	1,705	1,428	1,168	1,126	810	725	781	664	523	424	13,367
IL	15,958	18,111	27,046	15,470	9,622	8,332	7,572	7,121	8,144	5,958	3,951	2,849	130,134
IN	7,016	6,751	6,014	3,438	2,310	2,182	2,703	2,735	3,374	2,257	1,888	1,302	41,970
KY	3,321	3,009	2,671	1,752	1,283	1,072	1,073	1,062	1,829	1,119	714	485	19,390
MA	5,734	5,941	5,755	4,686	3,896	3,100	3,196	3,197	4,118	2,771	2,367	1,513	46,274
MD	9,779	9,596	8,303	5,347	3,660	3,154	3,092	3,264	6,988	3,655	1,992	1,185	60,015
ME	458	513	509	474	418	391	347	328	380	341	282	270	4,711
MI	9,948	10,252	10,724	8,126	7,303	6,280	6,303	6,245	6,852	5,259	3,782	3,450	84,524
MN	7,948	8,162	8,439	6,203	4,798	4,001	4,130	3,977	4,652	3,233	2,425	1,938	59,906
MO	4,075	4,035	3,806	2,705	2,200	1,752	1,630	1,503	1,741	1,216	875	584	26,122
MP	123	104	136	134	138	136	139	168	98	51	38	37	1,302
MT	242	379	402	328	242	212	232	204	140	106	55	39	2,581
NC	12,082	12,969	14,009	6,630	4,497	4,018	4,610	5,016	7,108	4,270	2,596	1,663	79,468
ND	370	337	314	202	206	168	191	222	272	187	131	125	2,725
NH	477	492	486	348	267	218	214	227	355	223	166	121	3,594
NJ	10,563	9,018	7,191	4,951	3,587	3,161	3,352	3,540	5,109	4,177	3,297	2,120	60,066
NM	4,751	5,227	5,131	4,080	3,190	2,626	2,890	2,771	3,087	2,144	1,401	951	38,249
NV	7,834	8,767	9,233	6,611	5,597	4,844	4,821	4,930	4,539	3,228	2,425	1,551	64,380
OK	3,840	3,810	3,558	2,314	1,466	1,183	1,706	1,697	2,150	1,316	874	516	24,430
PA	4,176	4,372	4,193	3,500	3,339	3,055	3,051	3,102	3,871	3,365	2,729	2,321	41,074
RI	1,051	1,055	1,168	766	588	541	549	595	896	671	511	353	8,744
SC	3,825	3,789	4,882	4,332	3,491	3,581	3,416	3,428	3,918	2,284	1,378	1,050	39,374
SD	561	547	573	299	192	197	187	244	385	201	153	114	3,653
TN	5,448	5,832	6,332	4,154	2,689	2,597	2,631	2,240	3,061	1,938	1,218	732	38,872
UT	5,309	5,687	5,218	3,847	2,736	1,999	2,229	2,215	1,985	1,595	1,247	878	34,945
VA	9,626	10,369	12,312	7,531	4,979	4,230	4,597	4,738	7,933	5,080	4,087	1,365	76,847
VI	132	113	118	102	90	100	88	57	84	54	43	42	1,023
VT	197	193	197	116	83	73	66	69	116	60	77	48	1,295
WI	5,668	5,989	5,852	5,012	3,738	2,950	2,753	2,707	2,846	2,019	1,471	1,075	42,080
WY	350	368	344	185	129	132	122	123	175	118	73	67	2,186
Total	170,981	174,584	182,698	122,972	90,769	78,652	80,128	79,378	101,008	68,242	48,577	33,370	1,231,359

3.2.1.2.2 By Gender

Table 3.2.1.2.2

Participation by Grade by Gender S401 Online

Grade		Gender			Total
		F	M	Missing	
1	Count	79,220	88,591	3,170	170,981
	% within Grade	46.3%	51.8%	1.9%	100.0%
2	Count	81,059	90,528	2,997	174,584
	% within Grade	46.4%	51.9%	1.7%	100.0%
3	Count	84,107	95,815	2,776	182,698
	% within Grade	46.0%	52.4%	1.5%	100.0%
4	Count	53,106	67,887	1,979	122,972
	% within Grade	43.2%	55.2%	1.6%	100.0%
5	Count	39,214	50,322	1,233	90,769
	% within Grade	43.2%	55.4%	1.4%	100.0%
6	Count	33,860	43,678	1,114	78,652
	% within Grade	43.1%	55.5%	1.4%	100.0%
7	Count	34,707	44,313	1,108	80,128
	% within Grade	43.3%	55.3%	1.4%	100.0%
8	Count	34,384	43,843	1,151	79,378
	% within Grade	43.3%	55.2%	1.5%	100.0%
9	Count	41,969	56,955	2,084	101,008
	% within Grade	41.6%	56.4%	2.1%	100.0%
10	Count	29,130	37,920	1,192	68,242
	% within Grade	42.7%	55.6%	1.7%	100.0%
11	Count	21,399	26,362	816	48,577
	% within Grade	44.1%	54.3%	1.7%	100.0%
12	Count	15,154	17,694	522	33,370
	% within Grade	45.4%	53.0%	1.6%	100.0%
Total	Count	547,309	663,908	20,142	1,231,359
	% within Grade	44.4%	53.9%	1.6%	100.0%

3.2.1.2.3 By Ethnicity

Table 3.2.1.2.3

Participation by Grade by Ethnicity S401 Online

Grade		Hispanic/Non-Hispanic			Total
		Hispanic	**Other**	**Unknown**	
1	Count	110,766	52,683	7,532	170,981
	% within Grade	64.8%	30.8%	4.4%	100.0%
2	Count	116,261	51,200	7,123	174,584
	% within Grade	66.6%	29.3%	4.1%	100.0%
3	Count	124,658	51,440	6,600	182,698
	% within Grade	68.2%	28.2%	3.6%	100.0%
4	Count	84,293	33,526	5,153	122,972
	% within Grade	68.5%	27.3%	4.2%	100.0%
5	Count	60,811	26,145	3,813	90,769
	% within Grade	67.0%	28.8%	4.2%	100.0%
6	Count	52,484	22,680	3,488	78,652
	% within Grade	66.7%	28.8%	4.4%	100.0%
7	Count	53,426	23,239	3,463	80,128
	% within Grade	66.7%	29.0%	4.3%	100.0%
8	Count	52,524	23,322	3,532	79,378
	% within Grade	66.2%	29.4%	4.4%	100.0%
9	Count	66,886	28,181	5,941	101,008
	% within Grade	66.2%	27.9%	5.9%	100.0%
10	Count	45,317	19,877	3,048	68,242
	% within Grade	66.4%	29.1%	4.5%	100.0%
11	Count	30,369	16,192	2,016	48,577
	% within Grade	62.5%	33.3%	4.2%	100.0%
12	Count	19,373	12,737	1,260	33,370
	% within Grade	58.1%	38.2%	3.8%	100.0%
Total	Count	817,168	361,222	52,969	1,231,359
	% within Grade	66.4%	29.3%	4.3%	100.0%

3.2.1.2.4 By Tier by Domain

Table 3.2.1.2.4

Participation by Grade by Tier by Domain S401 Online

Grade			Domain	
			Writing	Speaking
1	Tier	Pre-A	-	5,726
		A	133,261	57,484
		BC	37,698	107,764
	Total		170,959	170,974
2	Tier	Pre-A	-	6,671
		A	64,582	46,859
		BC	109,956	121,049
	Total		174,538	174,579
3	Tier	Pre-A	-	10,520
		A	44,017	42,400
		BC	138,657	129,768
	Total		182,674	182,688
4	Tier	Pre-A	-	2,407
		A	22,948	16,803
		BC	100,019	103,760
	Total		122,967	122,970
5	Tier	Pre-A	-	3,908
		A	22,650	13,045
		BC	68,113	73,810
	Total		90,763	90,763
6	Tier	Pre-A	-	3,069
		A	29,332	18,465
		BC	49,315	57,114
	Total		78,647	78,648
7	Tier	Pre-A	-	4,342
		A	33,601	13,664
		BC	46,522	62,115
	Total		80,123	80,121
8	Tier	Pre-A	-	5,690
		A	33,942	23,242
		BC	45,433	50,440
	Total		79,375	79,372

Grade			Domain	
			Writing	Speaking
9	Tier	Pre-A	-	8,458
		A	46,095	54,086
		BC	54,900	38,452
	Total		100,995	100,996
10	Tier	Pre-A	-	6,694
		A	28,532	29,789
		BC	39,701	31,750
	Total		68,233	68,233
11	Tier	Pre-A	-	4,750
		A	19,397	11,620
		BC	29,176	32,202
	Total		48,573	48,572
12	Tier	Pre-A	-	3,076
		A	11,319	16,992
		BC	22,046	13,299
	Total		33,365	33,367

3.2.2 Scale Score Results

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

3.2.2.1.1 By Cluster

Table 3.2.2.1.1

Mean Scale Scores by Cluster S401 Online

Cluster		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
1	Mean	328.49	290.64	260.96	271.80	300.96	275.63	301.76	283.21
	N	155,365	155,144	170,761	135,531	125,541	155,093	144,040	116,652
2-3	Mean	340.95	320.65	307.18	269.94	306.49	313.83	326.70	312.04
	N	318,299	309,909	356,941	293,645	266,501	309,803	282,857	237,158
4-5	Mean	403.89	345.40	328.57	311.83	358.84	335.79	362.28	342.49
	N	191,571	168,465	140,492	175,273	159,860	113,502	154,642	87,235
6-8	Mean	393.84	344.61	327.01	321.92	358.48	334.55	358.99	342.05
	N	195,700	186,282	210,933	180,383	153,304	171,107	160,625	118,380
9-12	Mean	386.17	368.94	351.35	314.68	351.01	358.44	373.54	356.52
	N	208,374	170,634	217,998	193,155	165,040	155,139	147,903	108,773

3.2.2.1.2 By Cluster by Gender

Table 3.2.2.1.2

Mean Scale Scores by Cluster by Gender S401 Online

Cluster	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
1	F	Mean	331.85	292.39	264.82	276.16	304.78	278.35	303.91	286.14
		N	72,356	71,329	79,119	63,893	59,372	71,307	66,458	54,696
	M	Mean	325.15	288.81	257.21	267.47	297.12	272.94	299.55	280.25
		N	80,376	81,212	88,474	69,470	64,179	81,185	75,183	60,123
	Missing	Mean	338.48	299.97	269.43	281.81	311.11	284.96	311.50	293.02
		N	2,633	2,603	3,168	2,168	1,990	2,601	2,399	1,833
2-3	F	Mean	342.68	322.84	313.05	272.74	308.65	317.87	328.72	315.34
		N	147,678	141,278	165,018	137,238	124,828	141,243	129,344	109,532
	M	Mean	339.28	318.61	301.76	267.23	304.35	310.18	324.81	308.97
		N	165,809	164,093	186,152	152,119	137,809	164,023	149,386	124,288
	Missing	Mean	345.52	326.45	313.80	276.58	312.70	320.28	332.15	318.53
		N	4,812	4,538	5,771	4,288	3,864	4,537	4,127	3,338
4-5	F	Mean	403.86	346.25	333.30	315.17	360.49	338.49	362.71	344.65
		N	82,963	70,943	61,000	76,024	69,520	48,060	65,374	37,236
	M	Mean	404.09	344.82	325.08	309.28	357.65	333.94	362.05	341.03
		N	105,916	95,146	78,128	96,855	88,193	64,324	87,110	49,202
	Missing	Mean	397.34	343.39	316.50	308.78	354.16	326.31	358.76	331.58
		N	2,692	2,376	1,364	2,394	2,147	1,118	2,158	797
6-8	F	Mean	394.03	347.71	332.67	324.27	359.68	338.90	361.11	345.24
		N	85,939	79,055	91,255	78,072	67,257	72,711	69,242	51,099
	M	Mean	394.09	342.53	322.91	320.38	357.87	331.55	357.65	339.86
		N	107,421	104,941	117,223	100,189	84,329	96,476	89,547	66,041
	Missing	Mean	375.05	332.71	312.31	307.86	341.56	320.47	344.64	326.89
		N	2,340	2,286	2,455	2,122	1,718	1,920	1,836	1,240
9-12	F	Mean	388.18	372.04	356.34	317.01	353.23	362.55	376.36	360.00
		N	90,670	71,047	92,873	82,653	71,612	64,445	62,413	45,722
	M	Mean	384.93	366.93	347.85	313.25	349.60	355.74	371.72	354.24
		N	114,293	96,921	121,598	107,479	90,989	88,420	83,355	61,676
	Missing	Mean	374.00	359.41	340.60	301.90	338.42	346.48	362.56	343.01
		N	3,411	2,666	3,527	3,023	2,439	2,274	2,135	1,375

3.2.2.1.3 By Cluster by Ethnicity

Table 3.2.2.1.3

Mean Scale Scores by Cluster by Ethnicity S401 Online

Cluster	Ethnicity			List	Read	Writ	Spek	Oral	Litr	Cphn	Over
1	Non-Hispanic Asian	Mean		338.38	305.96	273.71	278.38	309.17	289.68	315.43	295.57
		N		21,490	21,252	23,367	19,050	17,745	21,249	19,872	16,454
	Non-Hispanic Pacific Islander	Mean		308.48	284.09	259.62	263.94	286.42	271.98	291.41	276.15
		N		1,206	1,228	1,326	1,034	963	1,228	1,141	914
	Non-Hispanic Black	Mean		318.47	291.23	259.36	277.18	299.20	275.20	299.40	282.73
		N		7,395	7,524	8,353	6,657	6,021	7,522	6,807	5,571
	Hispanic (Of Any Race)	Mean		326.66	286.42	257.86	269.36	298.70	271.99	298.30	279.88
		N		101,226	101,400	110,678	87,788	81,434	101,363	94,323	76,004
	Non-Hispanic American Indian	Mean		319.82	282.46	254.27	262.72	292.81	268.42	293.70	276.16
		N		1,532	1,602	1,710	1,336	1,210	1,600	1,452	1,151
	Non-Hispanic Multi-racial	Mean		345.33	299.38	267.79	283.29	314.98	283.15	312.59	292.58
		N		786	778	887	711	645	778	709	580
	Non-Hispanic White	Mean		336.66	297.18	266.73	278.75	308.55	281.64	308.49	289.59
		N		15,375	15,050	16,951	13,645	12,596	15,045	13,879	11,412
	Unknown	Mean		320.01	292.80	256.77	266.05	294.66	274.68	300.59	280.86
		N		6,355	6,310	7,489	5,310	4,927	6,308	5,857	4,566
2-3	Non-Hispanic Asian	Mean		352.26	334.43	316.83	275.23	314.71	325.84	339.78	322.91
		N		38,534	38,276	42,587	35,911	32,929	38,273	35,268	30,252
	Non-Hispanic Pacific Islander	Mean		324.90	312.42	305.72	260.04	292.47	309.12	316.00	303.99
		N		2,262	2,219	2,524	2,024	1,861	2,219	2,032	1,671
	Non-Hispanic Black	Mean		332.85	316.32	299.73	269.91	302.64	307.64	321.17	306.51
		N		14,996	14,962	17,477	14,299	12,551	14,954	13,178	11,063
	Hispanic (Of Any Race)	Mean		338.81	317.66	305.92	268.72	304.74	311.64	323.97	309.89
		N		215,630	209,113	240,755	198,058	180,021	209,038	191,199	159,730
	Non-Hispanic American Indian	Mean		335.84	311.20	301.75	266.47	302.04	306.31	318.61	305.72
		N		3,768	3,754	4,332	3,474	3,056	3,750	3,338	2,717
	Non-Hispanic Multi-racial	Mean		356.28	329.19	310.87	278.01	317.93	320.26	337.12	319.86
		N		1,506	1,445	1,670	1,401	1,286	1,444	1,335	1,151
	Non-Hispanic White	Mean		351.74	328.78	311.88	276.20	314.98	320.21	335.57	319.12
		N		30,110	29,021	33,919	28,203	25,427	29,011	26,316	22,238
	Unknown	Mean		328.24	317.82	298.69	259.95	295.92	308.05	320.47	305.03
		N		11,493	11,119	13,677	10,275	9,370	11,114	10,191	8,336

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
4-5	Non-Hispanic Asian	Mean	410.38	355.66	337.03	315.38	363.87	346.34	371.61	352.00
		N	20,661	18,642	13,558	18,828	17,368	11,403	17,341	8,920
	Non-Hispanic Pacific Islander	Mean	390.23	338.51	323.47	304.44	347.59	330.06	353.79	334.83
		N	1,436	1,339	1,209	1,289	1,176	1,052	1,219	814
	Non-Hispanic Black	Mean	392.78	338.70	317.68	311.41	353.21	325.91	353.86	333.82
		N	10,781	9,671	6,394	10,246	9,113	5,026	8,636	3,798
	Hispanic (Of Any Race)	Mean	404.65	344.18	329.05	311.81	359.11	335.39	361.70	342.20
		N	130,639	114,056	97,228	119,270	108,993	78,192	104,852	60,189
	Non-Hispanic American Indian	Mean	403.20	341.66	324.95	309.64	357.68	332.89	359.90	340.67
		N	2,794	2,524	2,697	2,508	2,246	2,192	2,286	1,592
	Non-Hispanic Multi-racial	Mean	418.91	353.43	335.93	320.53	370.96	343.30	372.37	351.53
		N	729	648	456	663	619	381	603	299
	Non-Hispanic White	Mean	409.15	350.77	331.66	317.13	364.20	340.13	367.56	347.12
		N	16,832	14,527	13,653	15,692	14,216	10,768	13,242	8,339
	Unknown	Mean	379.10	337.99	305.51	291.92	337.12	318.65	348.62	322.24
		N	7,699	7,058	5,297	6,777	6,129	4,488	6,463	3,284
6-8	Non-Hispanic Asian	Mean	407.51	357.99	339.24	331.07	369.95	347.43	372.26	354.28
		N	21,984	19,932	23,201	20,402	17,668	18,317	17,474	13,236
	Non-Hispanic Pacific Islander	Mean	384.50	336.60	324.03	316.93	350.02	329.90	350.77	335.80
		N	1,581	1,687	1,841	1,525	1,177	1,535	1,341	961
	Non-Hispanic Black	Mean	388.56	337.99	319.04	321.25	354.68	326.93	352.30	334.43
		N	12,155	11,711	13,434	12,021	9,674	10,454	9,581	7,029
	Hispanic (Of Any Race)	Mean	392.28	343.53	326.74	320.86	357.19	333.97	357.88	341.32
		N	131,525	125,542	141,608	120,031	102,642	115,855	108,899	80,157
	Non-Hispanic American Indian	Mean	391.08	339.52	322.61	321.59	357.23	331.08	355.60	340.13
		N	2,910	3,016	3,242	2,619	2,231	2,835	2,575	1,885
	Non-Hispanic Multi-racial	Mean	414.76	352.83	334.91	332.83	374.68	342.86	371.54	352.81
		N	590	547	628	536	472	508	480	367
	Non-Hispanic White	Mean	405.76	351.17	332.55	330.56	368.92	340.56	367.08	349.51
		N	16,765	15,655	18,277	16,034	13,294	14,237	13,299	9,844
	Unknown	Mean	366.81	328.54	301.15	295.92	331.33	312.62	338.80	317.97
		N	8,190	8,192	8,702	7,215	6,146	7,366	6,976	4,901

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
9-12	Non-Hispanic Asian	Mean	398.35	383.22	364.07	326.63	363.10	372.00	387.35	369.46
		N	25,419	18,209	25,511	24,412	21,276	16,293	16,102	12,422
	Non-Hispanic Pacific Islander	Mean	385.59	368.19	353.76	316.07	351.66	360.28	373.55	357.73
		N	1,599	1,404	1,730	1,492	1,246	1,285	1,198	888
	Non-Hispanic Black	Mean	380.20	367.32	348.51	318.31	348.75	355.39	370.02	352.13
		N	15,913	11,670	16,762	15,650	12,900	10,300	9,744	7,196
	Hispanic (Of Any Race)	Mean	383.78	366.64	349.77	311.96	348.41	356.74	371.30	354.55
		N	134,578	114,173	142,388	123,687	105,423	104,536	98,887	72,390
	Non-Hispanic American Indian	Mean	394.57	371.79	357.73	323.33	359.60	364.08	378.28	362.91
		N	2,580	2,325	2,799	2,452	2,103	2,187	2,011	1,547
	Non-Hispanic Multi-racial	Mean	396.93	376.91	357.42	322.62	359.72	365.20	382.24	362.30
		N	560	437	581	518	457	402	393	300
	Non-Hispanic White	Mean	401.28	378.33	358.78	323.47	363.07	367.30	384.83	367.15
		N	17,995	14,145	18,158	16,367	14,368	12,812	12,526	9,268
	Unknown	Mean	366.48	354.39	330.34	293.36	330.46	338.87	356.37	335.99
		N	9,730	8,271	10,069	8,577	7,267	7,324	7,042	4,762

3.2.2.2 Mean Scale Scores by Grade Across Domain and Composite Scores

3.2.2.2.1 By Grade

Table 3.2.2.2.1

Mean Scale Scores by Grade S401 Online

Grade		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
1	Mean	328.49	290.64	260.96	271.80	300.96	275.63	301.76	283.21
	N	155,365	155,144	170,761	135,531	125,541	155,093	144,040	116,652
2	Mean	328.84	311.66	297.30	264.26	297.64	304.22	316.59	302.49
	N	154,676	151,219	174,393	141,170	127,608	151,153	137,371	113,409
3	Mean	352.39	329.23	316.61	275.21	314.62	323.00	336.25	320.80
	N	163,623	158,690	182,548	152,475	138,893	158,650	145,486	123,749
4	Mean	401.72	343.69	324.97	311.23	357.50	333.11	360.44	340.21
	N	110,019	97,309	80,224	100,539	91,410	64,908	89,088	49,613
5	Mean	406.82	347.74	333.35	312.63	360.63	339.36	364.79	345.50
	N	81,552	71,156	60,268	74,734	68,450	48,594	65,554	37,622
6	Mean	384.07	337.19	319.93	318.48	351.88	327.56	351.05	335.22
	N	64,820	62,850	69,202	60,103	51,264	57,241	54,115	39,883
7	Mean	395.25	344.69	327.17	321.61	359.34	334.64	359.43	342.55
	N	65,713	62,539	71,312	59,887	50,778	57,760	53,919	39,356
8	Mean	402.12	352.18	333.80	325.65	364.24	341.58	366.71	348.51
	N	65,167	60,893	70,419	60,393	51,262	56,106	52,591	39,141
9	Mean	384.38	364.10	345.20	311.46	349.00	352.58	369.34	352.28
	N	83,661	70,625	87,734	78,155	66,533	63,997	61,056	44,742
10	Mean	385.51	368.92	351.52	314.14	350.18	358.75	373.37	356.18
	N	56,588	46,083	59,166	52,350	44,777	41,909	39,938	29,366
11	Mean	387.49	373.75	357.17	317.14	352.61	364.04	377.54	360.44
	N	40,338	31,489	42,079	36,469	31,294	28,668	27,298	19,798
12	Mean	390.99	377.45	361.17	321.97	356.41	368.21	381.42	364.73
	N	27,787	22,437	29,019	26,181	22,436	20,565	19,611	14,867

3.2.2.2.2 By Grade by Gender

Table 3.2.2.2.2

Mean Scale Scores by Grade by Gender S401 Online

Grade	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
1	F	Mean	331.85	292.39	264.82	276.16	304.78	278.35	303.91	286.14
		N	72,356	71,329	79,119	63,893	59,372	71,307	66,458	54,696
	M	Mean	325.15	288.81	257.21	267.47	297.12	272.94	299.55	280.25
		N	80,376	81,212	88,474	69,470	64,179	81,185	75,183	60,123
	Missing	Mean	338.48	299.97	269.43	281.81	311.11	284.96	311.50	293.02
		N	2,633	2,603	3,168	2,168	1,990	2,601	2,399	1,833
2	F	Mean	330.73	313.35	302.80	266.89	299.80	307.76	318.26	305.45
		N	72,098	69,126	80,986	66,475	60,203	69,105	62,981	52,659
	M	Mean	326.93	309.96	292.07	261.59	295.39	300.88	314.89	299.58
		N	80,046	79,705	90,411	72,477	65,387	79,660	72,205	58,996
	Missing	Mean	335.45	319.44	306.54	272.69	305.90	312.93	324.31	311.47
		N	2,532	2,388	2,996	2,218	2,018	2,388	2,185	1,754
3	F	Mean	354.07	331.93	322.94	278.25	316.90	327.54	338.64	324.50
		N	75,580	72,152	84,032	70,763	64,625	72,138	66,363	56,873
	M	Mean	350.79	326.78	310.91	272.37	312.44	318.97	334.08	317.45
		N	85,763	84,388	95,741	79,642	72,422	84,363	77,181	65,292
	Missing	Mean	356.71	334.23	321.64	280.76	320.13	328.43	340.96	326.36
		N	2,280	2,150	2,775	2,070	1,846	2,149	1,942	1,584
4	F	Mean	401.50	344.22	329.54	314.82	359.15	335.59	360.52	342.07
		N	47,563	40,860	34,935	43,714	39,797	27,481	37,502	21,188
	M	Mean	401.96	343.29	321.58	308.40	356.20	331.42	360.40	338.93
		N	60,830	54,996	44,556	55,370	50,320	36,817	50,278	27,998
	Missing	Mean	399.20	343.96	312.47	311.45	357.13	324.18	359.78	331.15
		N	1,626	1,453	733	1,455	1,293	610	1,308	427
5	F	Mean	407.02	349.00	338.33	315.65	362.28	342.36	365.65	348.06
		N	35,400	30,083	26,065	32,310	29,723	20,579	27,872	16,048
	M	Mean	406.96	346.92	329.72	310.45	359.58	337.31	364.31	343.79
		N	45,086	40,150	33,572	41,485	37,873	27,507	36,832	21,204
	Missing	Mean	394.51	342.50	321.18	304.66	349.67	328.87	357.19	332.08
		N	1,066	923	631	939	854	508	850	370

Grade	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
6	F	Mean	383.96	340.35	326.25	320.70	352.84	332.30	353.11	338.64
		N	28,304	26,503	29,812	25,888	22,329	24,159	23,195	17,094
	M	Mean	384.40	335.04	315.31	317.05	351.39	324.26	349.67	332.82
		N	35,748	35,571	38,602	33,498	28,346	32,454	30,312	22,377
	Missing	Mean	373.21	327.88	307.47	305.00	339.64	316.02	341.58	323.32
		N	768	776	788	717	589	628	608	412
7	F	Mean	395.66	347.97	332.87	324.21	360.86	339.07	361.72	345.92
		N	28,898	26,674	30,901	25,960	22,284	24,671	23,333	17,066
	M	Mean	395.40	342.47	323.06	319.87	358.49	331.55	357.98	340.21
		N	36,036	35,119	39,610	33,244	27,935	32,449	29,978	21,884
	Missing	Mean	373.38	332.18	310.75	307.39	341.25	320.48	343.13	326.93
		N	779	746	801	683	559	640	608	406
8	F	Mean	402.32	354.99	338.73	327.86	365.28	345.41	368.65	351.22
		N	28,737	25,878	30,542	26,224	22,644	23,881	22,714	16,939
	M	Mean	402.49	350.38	330.28	324.23	363.82	339.03	365.58	346.75
		N	35,637	34,251	39,011	33,447	28,048	31,573	29,257	21,780
	Missing	Mean	378.46	338.12	318.15	311.15	343.84	324.74	349.14	330.35
		N	793	764	866	722	570	652	620	422
9	F	Mean	387.46	367.58	351.21	314.72	352.16	357.31	372.58	356.40
		N	35,295	28,578	36,271	32,496	28,069	25,876	25,036	18,363
	M	Mean	382.74	362.14	341.45	309.74	347.23	349.83	367.54	349.85
		N	46,798	40,743	49,824	44,260	37,335	37,007	34,987	25,713
	Missing	Mean	363.71	349.00	326.33	290.10	328.50	333.93	351.91	332.42
		N	1,568	1,304	1,639	1,399	1,129	1,114	1,033	666
10	F	Mean	387.01	371.43	355.79	315.82	351.85	362.16	375.57	358.92
		N	24,545	19,098	25,125	22,274	19,295	17,308	16,796	12,193
	M	Mean	384.58	367.25	348.42	313.05	349.11	356.46	371.90	354.38
		N	31,171	26,315	33,159	29,311	24,847	24,027	22,594	16,815
	Missing	Mean	376.22	363.21	346.80	307.18	341.69	352.25	366.30	347.11
		N	872	670	882	765	635	574	548	358
11	F	Mean	388.30	375.97	360.69	318.43	353.70	366.97	379.52	362.87
		N	18,012	13,456	18,405	16,058	13,948	12,196	11,764	8,512
	M	Mean	386.79	372.04	354.37	316.15	351.75	361.83	376.00	358.63
		N	21,703	17,607	23,053	19,891	16,923	16,115	15,188	11,072
	Missing	Mean	388.84	374.22	357.01	314.98	351.32	363.64	377.64	357.76
		N	623	426	621	520	423	357	346	214
12	F	Mean	392.23	380.73	365.54	323.66	358.11	372.34	384.38	368.25
		N	12,818	9,915	13,072	11,825	10,300	9,065	8,817	6,654
	M	Mean	389.97	374.81	357.51	320.63	355.01	364.92	378.98	361.90
		N	14,621	12,256	15,562	14,017	11,884	11,271	10,586	8,076
	Missing	Mean	388.25	377.17	360.67	318.56	352.95	366.34	380.42	360.72
		N	348	266	385	339	252	229	208	137

3.2.2.2.3 By Grade by Ethnicity

Table 3.2.2.2.3

Mean Scale Scores by Grade by Ethnicity S401 Online

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
1	Non-Hispanic Asian	Mean	338.38	305.96	273.71	278.38	309.17	289.68	315.43	295.57
		N	21,490	21,252	23,367	19,050	17,745	21,249	19,872	16,454
	Non-Hispanic Pacific Islander	Mean	308.48	284.09	259.62	263.94	286.42	271.98	291.41	276.15
		N	1,206	1,228	1,326	1,034	963	1,228	1,141	914
	Non-Hispanic Black	Mean	318.47	291.23	259.36	277.18	299.20	275.20	299.40	282.73
		N	7,395	7,524	8,353	6,657	6,021	7,522	6,807	5,571
	Hispanic (Of Any Race)	Mean	326.66	286.42	257.86	269.36	298.70	271.99	298.30	279.88
		N	101,226	101,400	110,678	87,788	81,434	101,363	94,323	76,004
	Non-Hispanic American Indian	Mean	319.82	282.46	254.27	262.72	292.81	268.42	293.70	276.16
		N	1,532	1,602	1,710	1,336	1,210	1,600	1,452	1,151
	Non-Hispanic Multi-racial	Mean	345.33	299.38	267.79	283.29	314.98	283.15	312.59	292.58
		N	786	778	887	711	645	778	709	580
	Non-Hispanic White	Mean	336.66	297.18	266.73	278.75	308.55	281.64	308.49	289.59
		N	15,375	15,050	16,951	13,645	12,596	15,045	13,879	11,412
	Unknown	Mean	320.01	292.80	256.77	266.05	294.66	274.68	300.59	280.86
		N	6,355	6,310	7,489	5,310	4,927	6,308	5,857	4,566
2	Non-Hispanic Asian	Mean	341.57	325.25	308.38	269.58	306.58	316.90	329.99	314.12
		N	19,577	19,400	21,740	18,034	16,474	19,400	17,813	15,055
	Non-Hispanic Pacific Islander	Mean	313.28	304.37	295.21	253.90	283.63	299.70	306.59	294.81
		N	1,086	1,054	1,191	950	887	1,054	980	801
	Non-Hispanic Black	Mean	322.56	308.37	290.45	266.48	296.09	299.12	312.41	298.67
		N	7,232	7,267	8,462	6,820	5,968	7,262	6,363	5,275
	Hispanic (Of Any Race)	Mean	325.85	308.39	295.36	262.58	295.20	301.56	313.40	299.74
		N	103,460	100,979	116,165	93,988	85,072	100,929	91,856	75,487
	Non-Hispanic American Indian	Mean	321.95	302.79	291.13	260.60	292.00	296.70	308.56	296.11
		N	1,756	1,780	2,034	1,607	1,412	1,778	1,575	1,276
	Non-Hispanic Multi-racial	Mean	346.84	320.88	304.05	273.98	311.00	312.43	328.02	311.76
		N	765	726	857	706	641	725	666	563
	Non-Hispanic White	Mean	340.58	318.97	302.59	270.54	306.61	310.31	325.16	309.51
		N	14,830	14,265	16,852	13,831	12,385	14,260	12,844	10,732
	Unknown	Mean	319.98	312.02	292.00	258.32	291.36	301.56	313.85	299.26
		N	5,970	5,748	7,092	5,234	4,769	5,745	5,274	4,220

Grade	Ethnicity			List	Read	Writ	Spek	Oral	Litr	Cphn	Over
3	Non-Hispanic Asian	Mean		363.29	343.86	325.64	280.92	322.86	335.02	349.77	331.62
		N		18,957	18,876	20,847	17,877	16,455	18,873	17,455	15,197
	Non-Hispanic Pacific Islander	Mean		335.63	319.70	315.11	265.48	300.52	317.64	324.76	312.44
		N		1,176	1,165	1,333	1,074	974	1,165	1,052	870
	Non-Hispanic Black	Mean		342.45	323.83	308.43	273.05	308.58	315.68	329.34	313.64
		N		7,764	7,695	9,015	7,479	6,583	7,692	6,815	5,788
	Hispanic (Of Any Race)	Mean		350.76	326.32	315.76	274.27	313.29	321.06	333.74	318.98
		N		112,170	108,134	124,590	104,070	94,949	108,109	99,343	84,243
	Non-Hispanic American Indian	Mean		347.97	318.79	311.15	271.52	310.66	314.99	327.59	314.23
		N		2,012	1,974	2,298	1,867	1,644	1,972	1,763	1,441
	Non-Hispanic Multi-racial	Mean		366.02	337.59	318.05	282.11	324.82	328.15	346.18	327.62
		N		741	719	813	695	645	719	669	588
	Non-Hispanic White	Mean		362.57	338.27	321.06	281.64	322.93	329.78	345.49	328.08
		N		15,280	14,756	17,067	14,372	13,042	14,751	13,472	11,506
	Unknown	Mean		337.17	324.03	305.88	261.64	300.66	315.00	327.57	310.93
		N		5,523	5,371	6,585	5,041	4,601	5,369	4,917	4,116
4	Non-Hispanic Asian	Mean		409.39	353.79	333.97	314.85	363.27	344.15	370.09	350.39
		N		11,822	10,782	7,738	10,763	9,872	6,538	9,974	5,030
	Non-Hispanic Pacific Islander	Mean		389.14	336.74	318.19	302.68	346.60	326.54	352.20	332.14
		N		776	717	659	701	642	567	650	439
	Non-Hispanic Black	Mean		391.10	336.90	313.20	310.42	351.80	322.73	352.11	330.83
		N		5,810	5,297	3,416	5,538	4,890	2,727	4,706	2,042
	Hispanic (Of Any Race)	Mean		402.02	342.48	325.47	311.17	357.47	332.72	359.70	339.83
		N		75,854	66,551	56,194	69,175	63,092	45,250	61,091	34,744
	Non-Hispanic American Indian	Mean		398.91	339.23	320.08	307.17	354.63	329.19	356.70	337.36
		N		1,580	1,457	1,526	1,445	1,279	1,247	1,294	892
	Non-Hispanic Multi-racial	Mean		416.19	352.08	333.15	319.04	368.56	342.59	371.23	351.45
		N		438	386	277	397	374	232	364	189
	Non-Hispanic White	Mean		407.34	348.39	327.53	316.30	362.92	336.60	365.37	344.14
		N		9,379	8,104	7,572	8,686	7,827	5,952	7,353	4,551
	Unknown	Mean		379.70	338.12	301.31	294.12	339.02	316.28	349.05	321.09
		N		4,360	4,015	2,842	3,834	3,434	2,395	3,656	1,726

Grade	Ethnicity			List	Read	Writ	Spek	Oral	Litr	Cphn	Over
5	Non-Hispanic Asian	Mean		411.71	358.23	341.09	316.08	364.65	349.28	373.65	354.07
		N		8,839	7,860	5,820	8,065	7,496	4,865	7,367	3,890
	Non-Hispanic Pacific Islander	Mean		391.51	340.55	329.78	306.55	348.77	334.17	355.61	337.98
		N		660	622	550	588	534	485	569	375
	Non-Hispanic Black	Mean		394.74	340.88	322.82	312.58	354.85	329.68	355.96	337.30
		N		4,971	4,374	2,978	4,708	4,223	2,299	3,930	1,756
	Hispanic (Of Any Race)	Mean		408.30	346.57	333.96	312.70	361.35	339.06	364.51	345.43
		N		54,785	47,505	41,034	50,095	45,901	32,942	43,761	25,445
	Non-Hispanic American Indian	Mean		408.80	344.99	331.30	312.99	361.70	337.78	364.08	344.89
		N		1,214	1,067	1,171	1,063	967	945	992	700
	Non-Hispanic Multi-racial	Mean		423.01	355.41	340.23	322.75	374.64	344.42	374.12	351.66
		N		291	262	179	266	245	149	239	110
	Non-Hispanic White	Mean		411.43	353.77	336.80	318.16	365.76	344.48	370.29	350.71
		N		7,453	6,423	6,081	7,006	6,389	4,816	5,889	3,788
	Unknown	Mean		378.31	337.81	310.38	289.05	334.69	321.35	348.06	323.50
		N		3,339	3,043	2,455	2,943	2,695	2,093	2,807	1,558
6	Non-Hispanic Asian	Mean		395.32	348.38	330.00	324.93	360.88	338.30	362.15	345.46
		N		7,007	6,559	7,358	6,530	5,678	5,985	5,777	4,375
	Non-Hispanic Pacific Islander	Mean		372.87	328.99	316.27	311.88	341.12	322.21	341.23	326.98
		N		561	575	636	520	417	523	476	345
	Non-Hispanic Black	Mean		377.50	330.73	310.33	317.69	347.62	319.20	343.98	327.00
		N		3,930	4,011	4,394	3,976	3,187	3,560	3,221	2,373
	Hispanic (Of Any Race)	Mean		382.65	336.33	320.12	317.89	350.88	327.34	350.16	334.83
		N		43,623	42,311	46,511	40,183	34,398	38,681	36,617	26,980
	Non-Hispanic American Indian	Mean		379.58	332.10	314.91	316.70	349.08	323.78	346.55	332.35
		N		962	993	1,045	870	745	910	841	602
	Non-Hispanic Multi-racial	Mean		400.04	340.44	323.63	323.33	363.26	330.64	357.33	341.37
		N		224	205	236	198	180	191	183	139
	Non-Hispanic White	Mean		395.78	342.91	325.64	326.71	361.88	332.85	358.21	341.76
		N		5,795	5,411	6,152	5,414	4,586	4,901	4,653	3,403
	Unknown	Mean		365.06	325.39	295.96	295.24	330.51	308.97	336.27	315.09
		N		2,718	2,785	2,870	2,412	2,073	2,490	2,347	1,666

Grade	Ethnicity			List	Read	Writ	Spek	Oral	Litr	Cphn	Over
7	Non-Hispanic Asian	Mean		408.91	358.77	339.87	331.14	370.72	348.23	372.93	355.07
		N		7,430	6,692	7,834	6,782	5,888	6,186	5,893	4,414
	Non-Hispanic Pacific Islander	Mean		386.29	336.14	322.75	315.59	351.19	329.02	351.05	335.88
		N		540	580	618	534	407	529	462	337
	Non-Hispanic Black	Mean		391.49	337.94	319.58	321.58	356.89	327.38	353.32	335.53
		N		4,129	3,943	4,550	4,011	3,236	3,529	3,237	2,357
	Hispanic (Of Any Race)	Mean		393.55	343.52	326.80	320.39	357.86	333.90	358.20	341.61
		N		44,310	42,247	48,050	39,917	34,122	39,214	36,649	26,722
	Non-Hispanic American Indian	Mean		394.47	340.37	324.40	322.12	359.38	331.91	357.35	341.54
		N		989	1,026	1,131	896	757	980	872	650
	Non-Hispanic Multi-racial	Mean		417.42	355.39	337.00	332.01	375.03	345.86	373.79	354.75
		N		190	178	201	168	150	166	157	120
	Non-Hispanic White	Mean		408.23	351.90	332.87	330.41	370.64	341.24	368.62	351.08
		N		5,400	5,176	6,069	5,253	4,228	4,722	4,330	3,183
	Unknown	Mean		366.21	326.97	299.88	295.25	331.40	311.46	337.47	317.68
		N		2,725	2,697	2,859	2,326	1,990	2,434	2,319	1,573
8	Non-Hispanic Asian	Mean		417.44	366.65	347.11	336.66	377.64	355.51	381.66	362.16
		N		7,547	6,681	8,009	7,090	6,102	6,146	5,804	4,447
	Non-Hispanic Pacific Islander	Mean		396.08	345.32	333.77	324.03	359.19	339.18	361.69	346.62
		N		480	532	587	471	353	483	403	279
	Non-Hispanic Black	Mean		396.23	345.79	327.01	324.42	359.40	334.64	359.83	340.99
		N		4,096	3,757	4,490	4,034	3,251	3,365	3,123	2,299
	Hispanic (Of Any Race)	Mean		400.64	350.97	333.22	324.31	362.89	340.79	365.49	347.66
		N		43,592	40,984	47,047	39,931	34,122	37,960	35,633	26,455
	Non-Hispanic American Indian	Mean		399.13	346.04	328.28	326.02	363.34	337.26	362.66	346.09
		N		959	997	1,066	853	729	945	862	633
	Non-Hispanic Multi-racial	Mean		430.64	365.52	346.64	344.69	388.80	355.01	387.59	365.37
		N		176	164	191	170	142	151	140	108
	Non-Hispanic White	Mean		413.75	359.23	339.25	334.58	374.50	348.05	375.11	356.06
		N		5,570	5,068	6,056	5,367	4,480	4,614	4,316	3,258
	Unknown	Mean		369.13	333.32	307.38	297.22	332.07	317.49	342.71	321.15
		N		2,747	2,710	2,973	2,477	2,083	2,442	2,310	1,662

Grade	Ethnicity			List	Read	Writ	Spek	Oral	Litr	Cphn	Over
9	Non-Hispanic Asian	Mean		402.52	382.51	361.56	327.20	365.75	370.18	387.87	369.29
		N		9,017	6,668	9,047	8,746	7,621	5,953	5,904	4,555
	Non-Hispanic Pacific Islander	Mean		386.07	364.62	350.67	314.87	350.22	356.96	371.58	355.90
		N		596	555	649	585	467	496	462	348
	Non-Hispanic Black	Mean		379.02	362.00	341.24	314.68	346.82	348.88	365.95	347.79
		N		5,793	4,519	6,103	5,723	4,705	3,970	3,762	2,753
	Hispanic (Of Any Race)	Mean		382.65	362.47	344.45	309.56	347.09	351.61	367.73	350.89
		N		55,520	48,169	58,936	51,454	43,733	44,037	41,698	30,519
	Non-Hispanic American Indian	Mean		392.16	367.51	353.75	321.63	358.59	359.64	374.17	359.41
		N		999	920	1,085	985	825	846	783	585
	Non-Hispanic Multi-racial	Mean		400.36	375.34	358.59	324.06	362.06	366.23	382.60	364.98
		N		239	185	242	230	201	171	168	132
	Non-Hispanic White	Mean		397.84	373.06	353.51	320.30	359.95	361.83	380.16	363.12
		N		6,780	5,355	6,751	6,306	5,516	4,774	4,703	3,501
	Unknown	Mean		354.50	343.31	314.30	280.09	318.52	325.36	344.93	323.48
		N		4,717	4,254	4,921	4,126	3,465	3,750	3,576	2,349
10	Non-Hispanic Asian	Mean		396.11	381.58	362.39	325.47	361.30	370.28	385.18	367.27
		N		6,524	4,616	6,559	6,215	5,449	4,156	4,075	3,150
	Non-Hispanic Pacific Islander	Mean		382.56	368.15	350.70	314.63	350.55	359.16	372.32	355.50
		N		414	358	442	385	328	325	311	231
	Non-Hispanic Black	Mean		379.48	367.08	348.49	318.26	348.20	355.25	369.44	351.45
		N		4,174	2,930	4,375	4,110	3,424	2,577	2,463	1,832
	Hispanic (Of Any Race)	Mean		383.22	366.71	350.00	311.33	347.61	357.14	371.32	354.45
		N		37,562	31,877	39,728	34,554	29,386	29,151	27,546	20,152
	Non-Hispanic American Indian	Mean		395.56	372.37	358.44	324.50	360.77	364.88	379.75	364.48
		N		683	615	755	638	551	588	530	419
	Non-Hispanic Multi-racial	Mean		394.21	379.29	355.79	321.53	357.75	365.84	383.58	360.38
		N		131	103	140	120	104	95	89	66
	Non-Hispanic White	Mean		400.06	377.74	358.08	322.26	361.75	366.88	383.73	365.69
		N		4,668	3,589	4,703	4,209	3,700	3,255	3,193	2,335
	Unknown	Mean		371.98	360.37	337.81	298.93	335.18	345.94	362.35	341.55
		N		2,432	1,995	2,464	2,119	1,835	1,762	1,731	1,181

Grade	Ethnicity			List	Read	Writ	Spek	Oral	Litr	Cphn	Over
11	Non-Hispanic Asian	Mean		397.21	385.43	367.27	327.07	362.86	374.86	389.09	371.94
		N		5,638	3,881	5,631	5,248	4,591	3,465	3,433	2,590
	Non-Hispanic Pacific Islander	Mean		387.40	370.69	356.63	315.47	352.18	362.33	374.17	357.55
		N		321	277	367	288	245	267	232	169
	Non-Hispanic Black	Mean		381.59	372.72	353.70	320.13	350.12	361.04	374.56	356.00
		N		3,277	2,333	3,485	3,175	2,594	2,073	1,938	1,418
	Hispanic (Of Any Race)	Mean		383.98	370.67	355.01	313.52	349.03	361.68	374.42	357.74
		N		25,267	20,435	26,598	22,593	19,362	18,690	17,677	12,712
	Non-Hispanic American Indian	Mean		395.86	376.16	359.97	321.32	358.28	367.67	381.34	364.24
		N		545	460	569	483	429	436	408	305
	Non-Hispanic Multi-racial	Mean		393.05	377.02	353.67	323.94	358.27	362.13	380.38	360.07
		N		110	89	117	100	90	82	81	61
	Non-Hispanic White	Mean		403.39	382.47	363.43	326.68	365.47	371.39	388.26	369.87
		N		3,550	2,750	3,638	3,144	2,753	2,519	2,433	1,767
	Unknown	Mean		382.43	370.21	350.50	308.61	344.99	358.07	372.13	352.18
		N		1,630	1,264	1,674	1,438	1,230	1,136	1,096	776
12	Non-Hispanic Asian	Mean		394.42	384.42	367.76	326.61	360.52	374.95	387.24	370.05
		N		4,240	3,044	4,274	4,203	3,615	2,719	2,690	2,127
	Non-Hispanic Pacific Islander	Mean		387.02	374.31	362.22	322.18	356.05	367.74	379.48	366.19
		N		268	214	272	234	206	197	193	140
	Non-Hispanic Black	Mean		382.17	373.78	357.94	324.03	352.17	364.04	375.03	358.55
		N		2,669	1,888	2,799	2,642	2,177	1,680	1,581	1,193
	Hispanic (Of Any Race)	Mean		388.59	375.11	359.36	319.20	353.73	366.37	379.13	362.69
		N		16,229	13,692	17,126	15,086	12,942	12,658	11,966	9,007
	Non-Hispanic American Indian	Mean		397.51	376.54	364.14	328.85	362.10	369.52	382.37	367.04
		N		353	330	390	346	298	317	290	238
	Non-Hispanic Multi-racial	Mean		396.48	377.48	362.11	317.74	357.52	365.48	381.73	360.12
		N		80	60	82	68	62	54	55	41
	Non-Hispanic White	Mean		408.47	386.04	365.96	329.03	369.57	374.89	392.66	374.76
		N		2,997	2,451	3,066	2,708	2,399	2,264	2,197	1,665
	Unknown	Mean		384.54	374.53	356.88	316.91	350.58	363.07	377.13	358.47
		N		951	758	1,010	894	737	676	639	456

3.2.2.3 Correlations Among Scale Scores by Cluster

Table 3.2.2.3A

Correlations Among Scale Scores: 1 S401 Online

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.405	.520	.558
	N	155,365	144,040	155,320	125,541
Reading	Pearson Correlation		1	.540	.367
	N		155,144	155,093	125,283
Writing	Pearson Correlation			1	.456
	N			170,761	135,492
Speaking	Pearson Correlation				1
	N				135,531

Table 3.2.2.3B

Correlations Among Scale Scores: 2-3 S401 Online

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.581	.598	.602
	N	318,299	282,857	318,192	266,501
Reading	Pearson Correlation		1	.637	.487
	N		309,909	309,803	258,985
Writing	Pearson Correlation			1	.538
	N			356,941	293,549
Speaking	Pearson Correlation				1
	N				293,645

Table 3.2.2.3C

Correlations Among Scale Scores: 4-5 S401 Online

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.629	.658	.641
	N	191,571	154,642	127,567	159,860
Reading	Pearson Correlation		1	.624	.531
	N		168,465	113,502	139,947
Writing	Pearson Correlation			1	.625
	N			140,492	117,092
Speaking	Pearson Correlation				1
	N				175,273

Table 3.2.2.3D

Correlations Among Scale Scores: 6-8 S401 Online

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.716	.693	.672
	N	195,700	160,625	178,705	153,304
Reading	Pearson Correlation		1	.756	.595
	N		186,282	171,107	145,221
Writing	Pearson Correlation			1	.664
	N			210,933	164,267
Speaking	Pearson Correlation				1
	N				180,383

Table 3.2.2.3E

Correlations Among Scale Scores: 9-12 S401 Online

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.711	.599	.638
	N	208,374	147,903	186,112	165,040
Reading	Pearson Correlation		1	.680	.635
	N		170,634	155,139	134,311
Writing	Pearson Correlation			1	.660
	N			217,998	171,946
Speaking	Pearson Correlation				1
	N				193,155

3.2.3 Proficiency Level Results

3.2.3.1 Domains

3.2.3.1.1 Listening

3.2.3.1.1.1 By Cluster

Table 3.2.3.1.1.1A

Proficiency Level by Cluster (Count): Listening S401 Online

Cluster	Listening Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	14,848	6,303	18,398	7,375	15,320	93,121	155,365
2-3	24,222	35,278	54,872	19,669	45,956	138,302	318,299
4-5	9,110	4,980	17,995	12,174	16,090	131,222	191,571
6-8	12,371	16,012	43,028	27,074	23,773	73,442	195,700
9-12	29,600	31,734	56,018	32,949	18,683	39,390	208,374

Table 3.2.3.1.1.1B

Proficiency Level by Cluster (Percent): Listening S401 Online

Cluster	Listening Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	9.6%	4.1%	11.8%	4.7%	9.9%	59.9%	100.0%
2-3	7.6%	11.1%	17.2%	6.2%	14.4%	43.5%	100.0%
4-5	4.8%	2.6%	9.4%	6.4%	8.4%	68.5%	100.0%
6-8	6.3%	8.2%	22.0%	13.8%	12.1%	37.5%	100.0%
9-12	14.2%	15.2%	26.9%	15.8%	9.0%	18.9%	100.0%

3.2.3.1.1.2 By Grade

Table 3.2.3.1.1.2A

Proficiency Level by Grade (Count): Listening S401 Online

Grade	Listening Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	14,848	6,303	18,398	7,375	15,320	93,121	155,365
2	11,676	18,369	31,542	9,072	19,280	64,737	154,676
3	12,546	16,909	23,330	10,597	26,676	73,565	163,623
4	4,283	2,745	8,242	6,807	10,564	77,378	110,019
5	4,827	2,235	9,753	5,367	5,526	53,844	81,552
6	2,970	4,956	16,003	9,655	10,069	21,167	64,820
7	4,072	5,190	14,059	9,647	8,383	24,362	65,713
8	5,329	5,866	12,966	7,772	5,321	27,913	65,167
9	9,317	14,482	20,420	14,120	7,263	18,059	83,661
10	8,310	7,585	16,594	8,533	5,365	10,201	56,588
11	6,775	5,486	11,268	6,669	3,238	6,902	40,338
12	5,198	4,181	7,736	3,627	2,817	4,228	27,787

Table 3.2.3.1.1.2B

Proficiency Level by Grade (Percent): Listening S401 Online

Grade	Listening Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	9.6%	4.1%	11.8%	4.7%	9.9%	59.9%	100.0%
2	7.5%	11.9%	20.4%	5.9%	12.5%	41.9%	100.0%
3	7.7%	10.3%	14.3%	6.5%	16.3%	45.0%	100.0%
4	3.9%	2.5%	7.5%	6.2%	9.6%	70.3%	100.0%
5	5.9%	2.7%	12.0%	6.6%	6.8%	66.0%	100.0%
6	4.6%	7.6%	24.7%	14.9%	15.5%	32.7%	100.0%
7	6.2%	7.9%	21.4%	14.7%	12.8%	37.1%	100.0%
8	8.2%	9.0%	19.9%	11.9%	8.2%	42.8%	100.0%
9	11.1%	17.3%	24.4%	16.9%	8.7%	21.6%	100.0%
10	14.7%	13.4%	29.3%	15.1%	9.5%	18.0%	100.0%
11	16.8%	13.6%	27.9%	16.5%	8.0%	17.1%	100.0%
12	18.7%	15.0%	27.8%	13.1%	10.1%	15.2%	100.0%

3.2.3.1.2 Reading

3.2.3.1.2.1   By Cluster

Table 3.2.3.1.2.1A

Proficiency Level by Cluster (Count): Reading S401 Online

Cluster	Reading Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	30,402	41,364	30,896	13,856	23,014	15,612	155,144
2-3	59,494	89,846	53,280	23,423	37,396	46,470	309,909
4-5	23,409	46,554	36,178	18,918	27,982	15,424	168,465
6-8	66,900	49,195	33,131	9,636	15,085	12,335	186,282
9-12	49,277	54,319	25,905	7,720	17,565	15,848	170,634

Table 3.2.3.1.2.1B

Proficiency Level by Cluster (Percent): Reading S401 Online

Cluster	Reading Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	19.6%	26.7%	19.9%	8.9%	14.8%	10.1%	100.0%
2-3	19.2%	29.0%	17.2%	7.6%	12.1%	15.0%	100.0%
4-5	13.9%	27.6%	21.5%	11.2%	16.6%	9.2%	100.0%
6-8	35.9%	26.4%	17.8%	5.2%	8.1%	6.6%	100.0%
9-12	28.9%	31.8%	15.2%	4.5%	10.3%	9.3%	100.0%

3.2.3.1.2.2 By Grade

Table 3.2.3.1.2.2A

Proficiency Level by Grade (Count): Reading S401 Online

Grade	Reading Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	30,402	41,364	30,896	13,856	23,014	15,612	155,144
2	26,869	46,113	30,183	12,658	17,874	17,522	151,219
3	32,625	43,733	23,097	10,765	19,522	28,948	158,690
4	10,541	24,448	24,091	11,395	18,532	8,302	97,309
5	12,868	22,106	12,087	7,523	9,450	7,122	71,156
6	22,705	17,132	12,098	3,259	5,099	2,557	62,850
7	22,501	16,676	11,042	3,457	4,627	4,236	62,539
8	21,694	15,387	9,991	2,920	5,359	5,542	60,893
9	22,281	20,143	10,934	2,869	7,259	7,139	70,625
10	13,198	14,476	7,615	2,076	4,700	4,018	46,083
11	8,409	11,073	4,246	1,620	3,287	2,854	31,489
12	5,389	8,627	3,110	1,155	2,319	1,837	22,437

Table 3.2.3.1.2.2B

Proficiency Level by Grade (Percent): Reading S401 Online

Grade	Reading Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	19.6%	26.7%	19.9%	8.9%	14.8%	10.1%	100.0%
2	17.8%	30.5%	20.0%	8.4%	11.8%	11.6%	100.0%
3	20.6%	27.6%	14.6%	6.8%	12.3%	18.2%	100.0%
4	10.8%	25.1%	24.8%	11.7%	19.0%	8.5%	100.0%
5	18.1%	31.1%	17.0%	10.6%	13.3%	10.0%	100.0%
6	36.1%	27.3%	19.2%	5.2%	8.1%	4.1%	100.0%
7	36.0%	26.7%	17.7%	5.5%	7.4%	6.8%	100.0%
8	35.6%	25.3%	16.4%	4.8%	8.8%	9.1%	100.0%
9	31.5%	28.5%	15.5%	4.1%	10.3%	10.1%	100.0%
10	28.6%	31.4%	16.5%	4.5%	10.2%	8.7%	100.0%
11	26.7%	35.2%	13.5%	5.1%	10.4%	9.1%	100.0%
12	24.0%	38.4%	13.9%	5.1%	10.3%	8.2%	100.0%

3.2.3.1.3 Writing

3.2.3.1.3.1   By Cluster

Table 3.2.3.1.3.1A

Proficiency Level by Cluster (Count): Writing S401 Online

Cluster	Writing Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	29,580	100,575	37,390	3,166	48	2	170,761
2-3	16,999	56,087	220,527	62,678	626	24	356,941
4-5	8,617	10,669	85,241	34,560	1,320	85	140,492
6-8	25,201	37,563	106,224	41,488	437	20	210,933
9-12	23,399	38,285	107,066	47,910	1,322	16	217,998

Table 3.2.3.1.3.1B

Proficiency Level by Cluster (Percent): Writing S401 Online

Cluster	Writing Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	17.3%	58.9%	21.9%	1.9%	0.0%	0.0%	100.0%
2-3	4.8%	15.7%	61.8%	17.6%	0.2%	0.0%	100.0%
4-5	6.1%	7.6%	60.7%	24.6%	0.9%	0.1%	100.0%
6-8	11.9%	17.8%	50.4%	19.7%	0.2%	0.0%	100.0%
9-12	10.7%	17.6%	49.1%	22.0%	0.6%	0.0%	100.0%

3.2.3.1.3.2 By Grade

Table 3.2.3.1.3.2A

Proficiency Level by Grade (Count): Writing S401 Online

Grade	Writing Proficiency Range						Total
	1	2	3	4	5	6	
1	29,580	100,575	37,390	3,166	48	2	170,761
2	10,144	31,959	112,559	19,619	110	2	174,393
3	6,855	24,128	107,968	43,059	516	22	182,548
4	5,016	6,077	50,373	17,947	773	38	80,224
5	3,601	4,592	34,868	16,613	547	47	60,268
6	7,507	12,290	36,910	12,396	97	2	69,202
7	8,310	14,410	34,535	13,917	133	7	71,312
8	9,384	10,863	34,779	15,175	207	11	70,419
9	10,502	14,309	39,237	22,847	833	6	87,734
10	4,839	11,581	29,853	12,654	232	7	59,166
11	4,098	8,392	20,999	8,423	165	2	42,079
12	3,960	4,003	16,977	3,986	92	1	29,019

Table 3.2.3.1.3.2B

Proficiency Level by Grade (Percent): Writing S401 Online

Grade	Writing Proficiency Range						Total
	1	2	3	4	5	6	
1	17.3%	58.9%	21.9%	1.9%	0.0%	0.0%	100.0%
2	5.8%	18.3%	64.5%	11.2%	0.1%	0.0%	100.0%
3	3.8%	13.2%	59.1%	23.6%	0.3%	0.0%	100.0%
4	6.3%	7.6%	62.8%	22.4%	1.0%	0.0%	100.0%
5	6.0%	7.6%	57.9%	27.6%	0.9%	0.1%	100.0%
6	10.8%	17.8%	53.3%	17.9%	0.1%	0.0%	100.0%
7	11.7%	20.2%	48.4%	19.5%	0.2%	0.0%	100.0%
8	13.3%	15.4%	49.4%	21.5%	0.3%	0.0%	100.0%
9	12.0%	16.3%	44.7%	26.0%	0.9%	0.0%	100.0%
10	8.2%	19.6%	50.5%	21.4%	0.4%	0.0%	100.0%
11	9.7%	19.9%	49.9%	20.0%	0.4%	0.0%	100.0%
12	13.6%	13.8%	58.5%	13.7%	0.3%	0.0%	100.0%

3.2.3.1.4 Speaking

3.2.3.1.4.1 By Cluster

Table 3.2.3.1.4.1A

Proficiency Level by Cluster (Count): Speaking S401 Online

Cluster	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
1	10,249	40,120	62,464	22,404	291	3	135,531
2-3	35,624	103,841	138,660	15,104	374	42	293,645
4-5	13,729	35,789	82,428	42,134	1,158	35	175,273
6-8	20,894	45,753	94,748	18,621	353	14	180,383
9-12	53,615	62,630	74,024	2,825	55	6	193,155

Table 3.2.3.1.4.1B

Proficiency Level by Cluster (Percent): Speaking S401 Online

Cluster	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
1	7.6%	29.6%	46.1%	16.5%	0.2%	0.0%	100.0%
2-3	12.1%	35.4%	47.2%	5.1%	0.1%	0.0%	100.0%
4-5	7.8%	20.4%	47.0%	24.0%	0.7%	0.0%	100.0%
6-8	11.6%	25.4%	52.5%	10.3%	0.2%	0.0%	100.0%
9-12	27.8%	32.4%	38.3%	1.5%	0.0%	0.0%	100.0%

3.2.3.1.4.2 By Grade

Table 3.2.3.1.4.2A

Proficiency Level by Grade (Count): Speaking S401 Online

Grade	Speaking Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	10,249	40,120	62,464	22,404	291	3	135,531
2	15,225	51,119	68,014	6,606	200	6	141,170
3	20,399	52,722	70,646	8,498	174	36	152,475
4	5,736	20,247	46,894	26,862	776	24	100,539
5	7,993	15,542	35,534	15,272	382	11	74,734
6	5,346	17,855	31,603	5,199	100	0	60,103
7	6,095	14,481	31,805	7,424	76	6	59,887
8	9,453	13,417	31,340	5,998	177	8	60,393
9	24,507	22,236	29,971	1,401	37	3	78,155
10	14,581	15,892	20,960	911	6	0	52,350
11	8,448	12,224	15,422	363	10	2	36,469
12	6,079	12,278	7,671	150	2	1	26,181

Table 3.2.3.1.4.2B

Proficiency Level by Grade (Percent): Speaking S401 Online

Grade	Speaking Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	7.6%	29.6%	46.1%	16.5%	0.2%	0.0%	100.0%
2	10.8%	36.2%	48.2%	4.7%	0.1%	0.0%	100.0%
3	13.4%	34.6%	46.3%	5.6%	0.1%	0.0%	100.0%
4	5.7%	20.1%	46.6%	26.7%	0.8%	0.0%	100.0%
5	10.7%	20.8%	47.5%	20.4%	0.5%	0.0%	100.0%
6	8.9%	29.7%	52.6%	8.7%	0.2%	0.0%	100.0%
7	10.2%	24.2%	53.1%	12.4%	0.1%	0.0%	100.0%
8	15.7%	22.2%	51.9%	9.9%	0.3%	0.0%	100.0%
9	31.4%	28.5%	38.3%	1.8%	0.0%	0.0%	100.0%
10	27.9%	30.4%	40.0%	1.7%	0.0%	0.0%	100.0%
11	23.2%	33.5%	42.3%	1.0%	0.0%	0.0%	100.0%
12	23.2%	46.9%	29.3%	0.6%	0.0%	0.0%	100.0%

3.2.3.2 Composites

3.2.3.2.1 Oral Composite

3.2.3.2.1.1 By Cluster

Table 3.2.3.2.1.1A

Proficiency Level by Cluster (Count): Oral S401 Online

Cluster	Oral Language Proficiency Range						Total
	1	2	3	4	5	6	
1	8,309	11,441	33,020	38,477	29,593	4,701	125,541
2-3	19,974	46,583	93,975	85,814	19,228	927	266,501
4-5	7,940	10,210	28,173	66,802	38,828	7,907	159,860
6-8	11,689	19,318	53,017	51,824	14,889	2,567	153,304
9-12	27,676	38,863	65,777	28,598	3,833	293	165,040

Table 3.2.3.2.1.1B

Proficiency Level by Cluster (Percent): Oral S401 Online

Cluster	Oral Language Proficiency Range						Total
	1	2	3	4	5	6	
1	6.6%	9.1%	26.3%	30.6%	23.6%	3.7%	100.0%
2-3	7.5%	17.5%	35.3%	32.2%	7.2%	0.3%	100.0%
4-5	5.0%	6.4%	17.6%	41.8%	24.3%	4.9%	100.0%
6-8	7.6%	12.6%	34.6%	33.8%	9.7%	1.7%	100.0%
9-12	16.8%	23.5%	39.9%	17.3%	2.3%	0.2%	100.0%

3.2.3.2.1.2 By Grade

Table 3.2.3.2.1.2A

Proficiency Level by Grade (Count): Oral S401 Online

Grade	Oral Language Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	8,309	11,441	33,020	38,477	29,593	4,701	125,541
2	9,129	24,958	48,031	35,823	9,220	447	127,608
3	10,845	21,625	45,944	49,991	10,008	480	138,893
4	3,587	5,271	15,479	40,131	22,355	4,587	91,410
5	4,353	4,939	12,694	26,671	16,473	3,320	68,450
6	2,796	6,186	19,901	17,276	4,446	659	51,264
7	3,887	6,447	17,402	16,817	5,372	853	50,778
8	5,006	6,685	15,714	17,731	5,071	1,055	51,262
9	10,582	14,826	24,873	14,030	2,043	179	66,533
10	7,511	10,501	18,379	7,429	901	56	44,777
11	5,628	7,895	12,849	4,314	575	33	31,294
12	3,955	5,641	9,676	2,825	314	25	22,436

Table 3.2.3.2.1.2B

Proficiency Level by Grade (Percent): Oral S401 Online

Grade	Oral Language Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	6.6%	9.1%	26.3%	30.6%	23.6%	3.7%	100.0%
2	7.2%	19.6%	37.6%	28.1%	7.2%	0.4%	100.0%
3	7.8%	15.6%	33.1%	36.0%	7.2%	0.3%	100.0%
4	3.9%	5.8%	16.9%	43.9%	24.5%	5.0%	100.0%
5	6.4%	7.2%	18.5%	39.0%	24.1%	4.9%	100.0%
6	5.5%	12.1%	38.8%	33.7%	8.7%	1.3%	100.0%
7	7.7%	12.7%	34.3%	33.1%	10.6%	1.7%	100.0%
8	9.8%	13.0%	30.7%	34.6%	9.9%	2.1%	100.0%
9	15.9%	22.3%	37.4%	21.1%	3.1%	0.3%	100.0%
10	16.8%	23.5%	41.0%	16.6%	2.0%	0.1%	100.0%
11	18.0%	25.2%	41.1%	13.8%	1.8%	0.1%	100.0%
12	17.6%	25.1%	43.1%	12.6%	1.4%	0.1%	100.0%

3.2.3.2.2 Literacy Composite

3.2.3.2.2.1 By Cluster

Table 3.2.3.2.2.1A

Proficiency Level by Cluster (Count): Literacy S401 Online

Cluster	Literacy Proficiency Range						Total
	1	2	3	4	5	6	
1	27,564	70,291	43,314	11,308	2,186	430	155,093
2-3	24,497	74,423	135,122	63,658	11,191	912	309,803
4-5	10,452	15,532	55,071	27,576	4,069	802	113,502
6-8	37,992	39,931	63,037	26,839	3,057	251	171,107
9-12	26,494	40,000	57,945	24,760	5,522	418	155,139

Table 3.2.3.2.2.1B

Proficiency Level by Cluster (Percent): Literacy S401 Online

Cluster	Literacy Proficiency Range						Total
	1	2	3	4	5	6	
1	17.8%	45.3%	27.9%	7.3%	1.4%	0.3%	100.0%
2-3	7.9%	24.0%	43.6%	20.5%	3.6%	0.3%	100.0%
4-5	9.2%	13.7%	48.5%	24.3%	3.6%	0.7%	100.0%
6-8	22.2%	23.3%	36.8%	15.7%	1.8%	0.1%	100.0%
9-12	17.1%	25.8%	37.4%	16.0%	3.6%	0.3%	100.0%

3.2.3.2.2.2   By Grade

Table 3.2.3.2.2.2A

Proficiency Level by Grade (Count): Literacy S401 Online

Grade	Literacy Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	27,564	70,291	43,314	11,308	2,186	430	155,093
2	12,286	43,248	67,839	24,235	3,255	290	151,153
3	12,211	31,175	67,283	39,423	7,936	622	158,650
4	5,322	8,333	33,546	15,455	1,851	401	64,908
5	5,130	7,199	21,525	12,121	2,218	401	48,594
6	12,373	13,345	23,379	7,478	616	50	57,241
7	12,836	13,694	21,205	8,863	1,063	99	57,760
8	12,783	12,892	18,453	10,498	1,378	102	56,106
9	12,601	14,283	22,700	11,486	2,669	258	63,997
10	6,465	11,008	16,451	6,565	1,327	93	41,909
11	4,280	8,371	10,835	4,233	897	52	28,668
12	3,148	6,338	7,959	2,476	629	15	20,565

Table 3.2.3.2.2.2B

Proficiency Level by Grade (Percent): Literacy S401 Online

Grade	Literacy Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	17.8%	45.3%	27.9%	7.3%	1.4%	0.3%	100.0%
2	8.1%	28.6%	44.9%	16.0%	2.2%	0.2%	100.0%
3	7.7%	19.7%	42.4%	24.8%	5.0%	0.4%	100.0%
4	8.2%	12.8%	51.7%	23.8%	2.9%	0.6%	100.0%
5	10.6%	14.8%	44.3%	24.9%	4.6%	0.8%	100.0%
6	21.6%	23.3%	40.8%	13.1%	1.1%	0.1%	100.0%
7	22.2%	23.7%	36.7%	15.3%	1.8%	0.2%	100.0%
8	22.8%	23.0%	32.9%	18.7%	2.5%	0.2%	100.0%
9	19.7%	22.3%	35.5%	17.9%	4.2%	0.4%	100.0%
10	15.4%	26.3%	39.3%	15.7%	3.2%	0.2%	100.0%
11	14.9%	29.2%	37.8%	14.8%	3.1%	0.2%	100.0%
12	15.3%	30.8%	38.7%	12.0%	3.1%	0.1%	100.0%

3.2.3.2.3 Comprehension Composite

3.2.3.2.3.1 By Cluster

Table 3.2.3.2.3.1A

Proficiency Level by Cluster (Count): Comprehension S401 Online

Cluster	Comprehension Proficiency Range						Total
	1	2	3	4	5	6	
1	14,004	22,852	30,988	19,230	27,932	29,034	144,040
2-3	29,538	58,801	65,092	33,997	44,225	51,204	282,857
4-5	9,916	19,473	29,182	23,798	34,903	37,370	154,642
6-8	32,204	37,239	33,116	19,486	19,582	18,998	160,625
9-12	31,162	42,685	30,206	13,955	14,402	15,493	147,903

Table 3.2.3.2.3.1B

Proficiency Level by Cluster (Percent): Comprehension S401 Online

Cluster	Comprehension Proficiency Range						Total
	1	2	3	4	5	6	
1	9.7%	15.9%	21.5%	13.4%	19.4%	20.2%	100.0%
2-3	10.4%	20.8%	23.0%	12.0%	15.6%	18.1%	100.0%
4-5	6.4%	12.6%	18.9%	15.4%	22.6%	24.2%	100.0%
6-8	20.0%	23.2%	20.6%	12.1%	12.2%	11.8%	100.0%
9-12	21.1%	28.9%	20.4%	9.4%	9.7%	10.5%	100.0%

3.2.3.2.3.2 By Grade

Table 3.2.3.2.3.2A

Proficiency Level by Grade (Count): Comprehension S401 Online

Grade	Comprehension Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	14,004	22,852	30,988	19,230	27,932	29,034	144,040
2	13,013	31,566	33,782	17,960	20,608	20,442	137,371
3	16,525	27,235	31,310	16,037	23,617	30,762	145,486
4	3,981	10,076	17,143	14,387	21,942	21,559	89,088
5	5,935	9,397	12,039	9,411	12,961	15,811	65,554
6	9,904	13,817	12,648	6,844	6,187	4,715	54,115
7	11,049	12,248	11,044	6,647	6,399	6,532	53,919
8	11,251	11,174	9,424	5,995	6,996	7,751	52,591
9	13,174	16,502	12,023	5,774	6,365	7,218	61,056
10	8,195	11,497	8,649	3,952	3,732	3,913	39,938
11	5,764	8,372	5,547	2,399	2,658	2,558	27,298
12	4,029	6,314	3,987	1,830	1,647	1,804	19,611

Table 3.2.3.2.3.2B

Proficiency Level by Grade (Percent): Comprehension S401 Online

Grade	Comprehension Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	9.7%	15.9%	21.5%	13.4%	19.4%	20.2%	100.0%
2	9.5%	23.0%	24.6%	13.1%	15.0%	14.9%	100.0%
3	11.4%	18.7%	21.5%	11.0%	16.2%	21.1%	100.0%
4	4.5%	11.3%	19.2%	16.1%	24.6%	24.2%	100.0%
5	9.1%	14.3%	18.4%	14.4%	19.8%	24.1%	100.0%
6	18.3%	25.5%	23.4%	12.6%	11.4%	8.7%	100.0%
7	20.5%	22.7%	20.5%	12.3%	11.9%	12.1%	100.0%
8	21.4%	21.2%	17.9%	11.4%	13.3%	14.7%	100.0%
9	21.6%	27.0%	19.7%	9.5%	10.4%	11.8%	100.0%
10	20.5%	28.8%	21.7%	9.9%	9.3%	9.8%	100.0%
11	21.1%	30.7%	20.3%	8.8%	9.7%	9.4%	100.0%
12	20.5%	32.2%	20.3%	9.3%	8.4%	9.2%	100.0%

3.2.3.2.4 Overall Composite

3.2.3.2.4.1   By Cluster

Table 3.2.3.2.4.1A

Proficiency Level by Cluster (Count): Overall S401 Online

Cluster	Overall Proficiency Range						Total
	1	2	3	4	5	6	
1	9,503	31,831	57,026	14,756	3,160	376	116,652
2-3	16,730	49,774	105,962	56,572	7,932	188	237,158
4-5	5,748	9,254	33,951	32,224	5,442	616	87,235
6-8	16,582	27,000	44,390	26,854	3,321	233	118,380
9-12	19,413	24,359	44,603	17,334	2,890	174	108,773

Table 3.2.3.2.4.1B

Proficiency Level by Cluster (Percent): Overall S401 Online

Cluster	Overall Proficiency Range						Total
	1	2	3	4	5	6	
1	8.1%	27.3%	48.9%	12.6%	2.7%	0.3%	100.0%
2-3	7.1%	21.0%	44.7%	23.9%	3.3%	0.1%	100.0%
4-5	6.6%	10.6%	38.9%	36.9%	6.2%	0.7%	100.0%
6-8	14.0%	22.8%	37.5%	22.7%	2.8%	0.2%	100.0%
9-12	17.8%	22.4%	41.0%	15.9%	2.7%	0.2%	100.0%

3.2.3.2.4.2 By Grade

Table 3.2.3.2.4.2A

Proficiency Level by Grade (Count): Overall S401 Online

Grade	Overall Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	9,503	31,831	57,026	14,756	3,160	376	116,652
2	7,449	29,005	52,867	21,413	2,594	81	113,409
3	9,281	20,769	53,095	35,159	5,338	107	123,749
4	2,682	4,845	20,216	18,854	2,672	344	49,613
5	3,066	4,409	13,735	13,370	2,770	272	37,622
6	4,672	9,621	16,932	7,868	739	51	39,883
7	5,549	9,009	14,579	8,991	1,132	96	39,356
8	6,361	8,370	12,879	9,995	1,450	86	39,141
9	8,354	8,692	17,684	8,396	1,499	117	44,742
10	5,110	6,570	12,512	4,518	626	30	29,366
11	3,489	4,924	8,241	2,668	456	20	19,798
12	2,460	4,173	6,166	1,752	309	7	14,867

Table 3.2.3.2.4.2B

Proficiency Level by Grade (Percent): Overall S401 Online

Grade	Overall Proficiency Range						Total
	1	**2**	**3**	**4**	**5**	**6**	
1	8.1%	27.3%	48.9%	12.6%	2.7%	0.3%	100.0%
2	6.6%	25.6%	46.6%	18.9%	2.3%	0.1%	100.0%
3	7.5%	16.8%	42.9%	28.4%	4.3%	0.1%	100.0%
4	5.4%	9.8%	40.7%	38.0%	5.4%	0.7%	100.0%
5	8.1%	11.7%	36.5%	35.5%	7.4%	0.7%	100.0%
6	11.7%	24.1%	42.5%	19.7%	1.9%	0.1%	100.0%
7	14.1%	22.9%	37.0%	22.8%	2.9%	0.2%	100.0%
8	16.3%	21.4%	32.9%	25.5%	3.7%	0.2%	100.0%
9	18.7%	19.4%	39.5%	18.8%	3.4%	0.3%	100.0%
10	17.4%	22.4%	42.6%	15.4%	2.1%	0.1%	100.0%
11	17.6%	24.9%	41.6%	13.5%	2.3%	0.1%	100.0%
12	16.5%	28.1%	41.5%	11.8%	2.1%	0.0%	100.0%

3.3.   Analyses of Domain Scores: Results

3.3.1 Grade: 1

3.3.1.1   Listening 1

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

Figure 3.3.1.1D
Scale Scores: List 1 S401 Online

Figure 3.3.1.1E
Proficiency Level: List 1 S401 Online
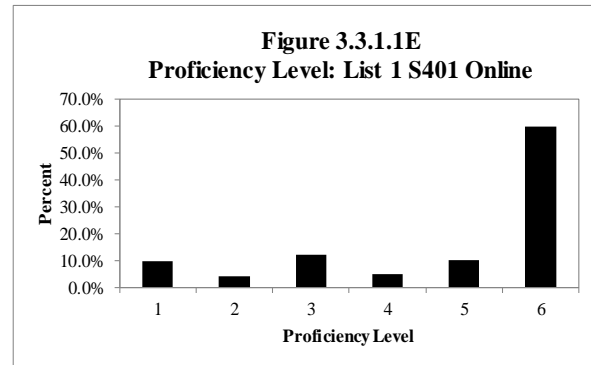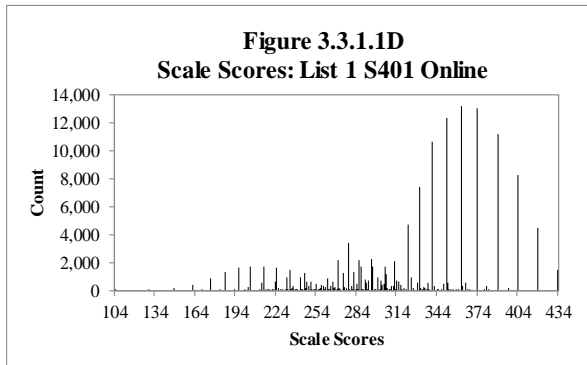
Table 3.3.1.1D

Scale Score Descriptive Statistics: List 1 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	146,126	104	434	328.03	58.63
Total	146,126	104	434	328.03	58.63

Table 3.3.1.1E

Proficiency Level Distribution: List 1 S401 Online

Level	Grade 1 Count	Grade 1 Percent	Total Count	Total Percent
1	14,016	9.59%	14,016	9.59%
2	5,963	4.08%	5,963	4.08%
3	17,462	11.95%	17,462	11.95%
4	6,992	4.78%	6,992	4.78%
5	14,551	9.96%	14,551	9.96%
6	87,142	59.63%	87,142	59.63%
Total	146,126	100.00%	146,126	100.00%

Table 3.3.1.1F
Raw Score to Scale Score Conversion: List 1 S401 Online
n/a

Figure 3.3.1.1H
Test Characteristic Curve: List 1 S401 Online
n/a

Figure 3.3.1.1I
Test Information Function: List 1 S401 Online
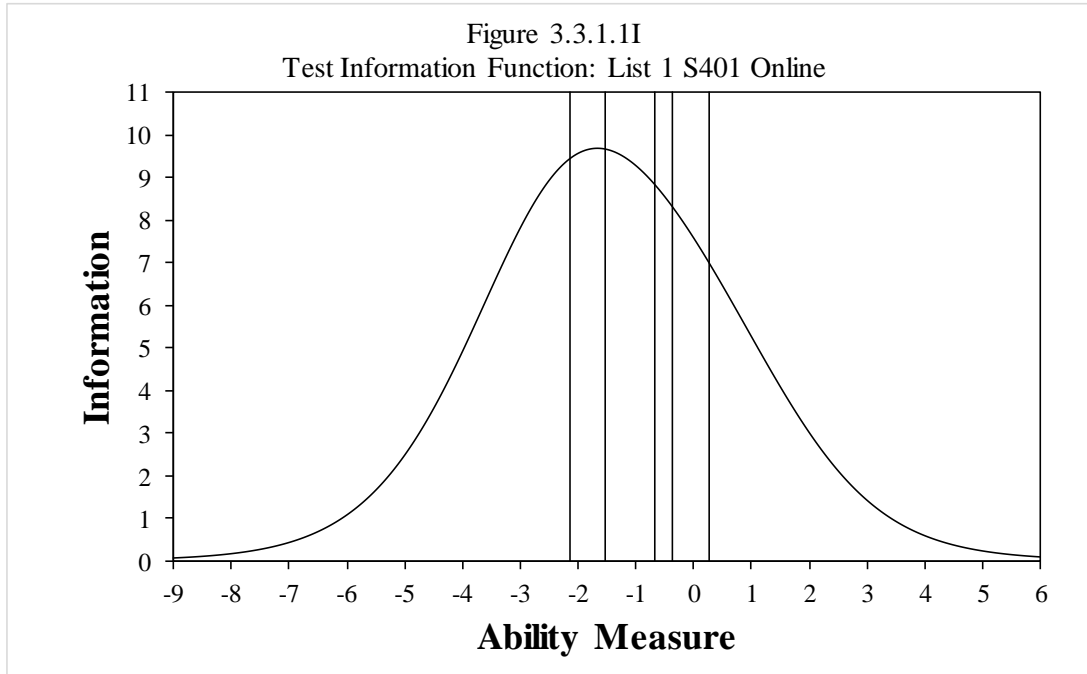
Table 3.3.1.1J

Reliability: List 1 S401 Online

No. of Students	No. of Items	Rasch Reliability Estimate
146,126	54	.86

Table 3.3.1.1K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: List 1 S401 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	1	236	184	16.84	19.39	19.21	0.66
2/3	1	259	284	16.84	17.86	17.45	0.43
3/4	1	291	576	17.86	20.41	18.61	0.79
4/5	1	303	358	17.86	21.94	19.57	0.98
5/6	1	327	N/A	N/A	N/A	N/A	N/A

Table 3.3.1.1L

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.719	0.659		0.447	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.853		0.717	
	2	0.254		0.183	
	3	0.500		0.375	
	4	0.179		0.128	
	5	0.306		0.223	
	6	0.919		0.886	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.958	0.011	0.030	0.942
	2/3	0.947	0.023	0.030	0.923
	3/4	0.919	0.031	0.050	0.887
	4/5	0.913	0.036	0.051	0.876
	5/6	0.895	0.057	0.048	0.850

3.3.1.2　Reading 1

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

Figure 3.3.1.2D
Scale Scores: Read 1 S401 Online

Figure 3.3.1.2E
Proficiency Level: Read 1 S401 Online
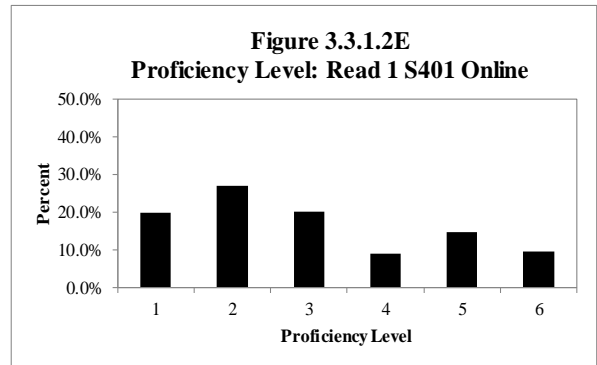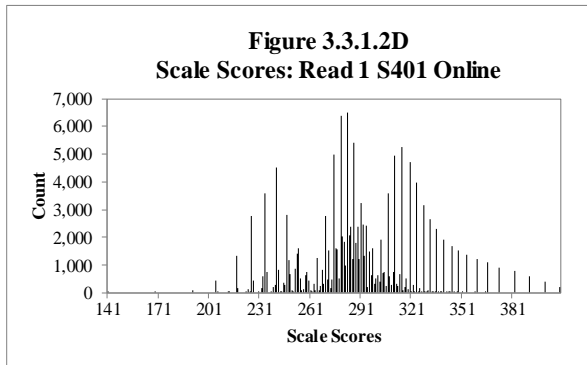
Table 3.3.1.2D

Scale Score Descriptive Statistics: Read 1 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	146,074	141	409	290.12	34.64
Total	146,074	141	409	290.12	34.64

Table 3.3.1.2E

Proficiency Level Distribution: Read 1 S401 Online

Level	Grade 1 Count	Grade 1 Percent	Total Count	Total Percent
1	28,938	19.81%	28,938	19.81%
2	39,222	26.85%	39,222	26.85%
3	29,296	20.06%	29,296	20.06%
4	13,132	8.99%	13,132	8.99%
5	21,419	14.66%	21,419	14.66%
6	14,067	9.63%	14,067	9.63%
Total	146,074	100.00%	146,074	100.00%

Table 3.3.1.2F
Raw Score to Scale Score Conversion: Read 1 S401 Online
n/a

Figure 3.3.1.2H
Test Characteristic Curve: Read 1 S401 Online
n/a

Figure 3.3.1.2I
Test Information Function: Read 1 S401 Online
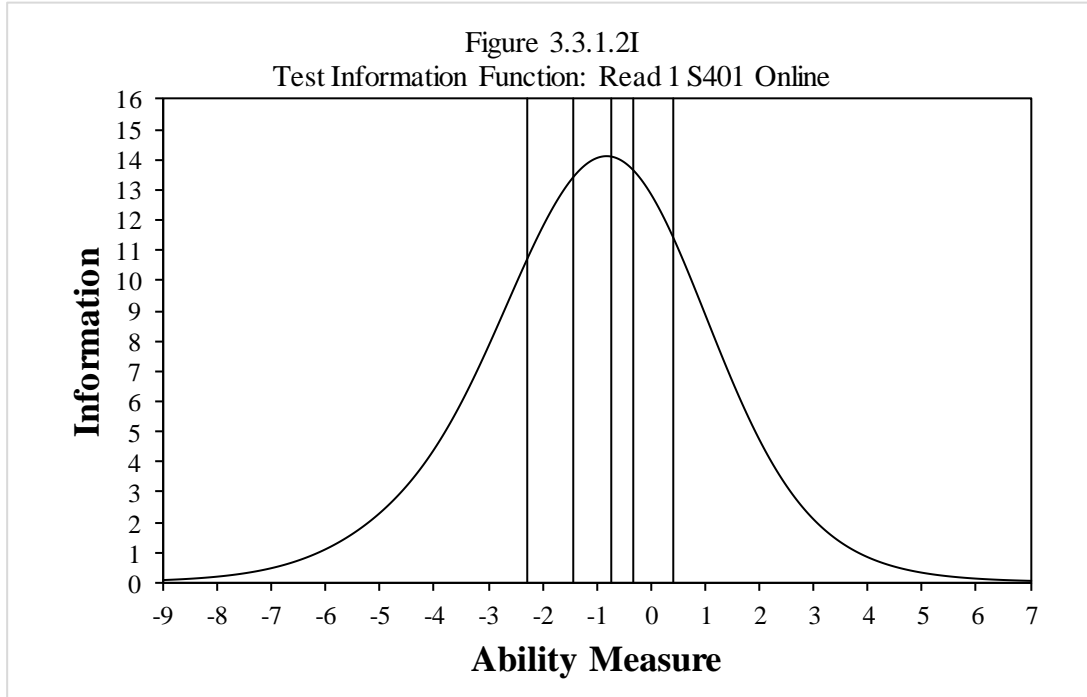
Table 3.3.1.2J

Reliability: Read 1 S401 Online

No. of Students	No. of Items	Rasch Reliability Estimate
146,074	72	.89

Table 3.3.1.2K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: Read 1 S401 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	1	264	102	10.71	11.22	10.88	0.24
2/3	1	286	1,237	9.69	10.71	9.97	0.26
3/4	1	304	709	9.69	10.71	10.25	0.24
4/5	1	315	5,235	10.20	10.71	10.20	0.01
5/6	1	334	11	10.20	10.20	10.20	0.00

Table 3.3.1.2L

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.614	0.508		0.396	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.763		0.658	
	2	0.632		0.525	
	3	0.517		0.405	
	4	0.312		0.229	
	5	0.613		0.480	
	6	0.791		0.661	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.910	0.049	0.041	0.872
	2/3	0.884	0.058	0.058	0.840
	3/4	0.908	0.049	0.043	0.868
	4/5	0.921	0.049	0.030	0.889
	5/6	0.958	0.023	0.019	0.938

3.3.1.3 Writing 1

3.3.1.3i *Writing 1 A*

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

Figure 3.3.1.3Ci
Raw Scores: Writ 1 A S401 Online
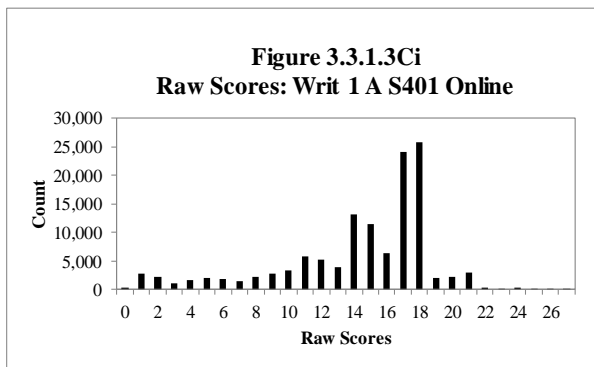
Table 3.3.1.3Ci

Raw Score Descriptive Statistics: Writ 1 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	126,315	0	27	14.31	4.73
Total	126,315	0	27	14.31	4.73

Figure 3.3.1.3Di
Scale Scores: Writ 1 A S401 Online
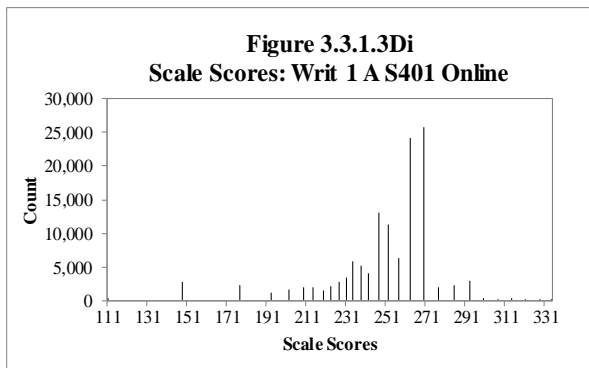
Table 3.3.1.3Di

Scale Score Descriptive Statistics: Writ 1 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	126,315	111	334	249.70	27.69
Total	126,315	111	334	249.70	27.69

Figure 3.3.1.3Ei
Proficiency Level: Writ 1 A S401 Online
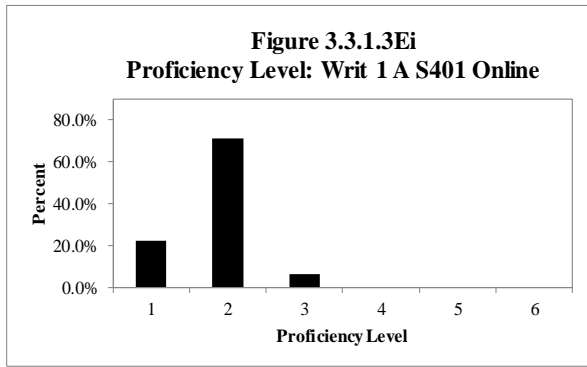
Table 3.3.1.3Ei

Proficiency Level Distribution: Writ 1 A S401 Online

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	28,030	22.19%	28,030	22.19%
2	89,943	71.21%	89,943	71.21%
3	8,342	6.60%	8,342	6.60%
4	0	0.00%	0	0.00%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	126,315	100.00%	126,315	100.00%

Figure 3.3.1.3Hi
Test Characteristic Curve: Writ 1 A S401 Online

Figure 3.3.1.3Ii
Test Information Function: Writ 1 A S401 Online
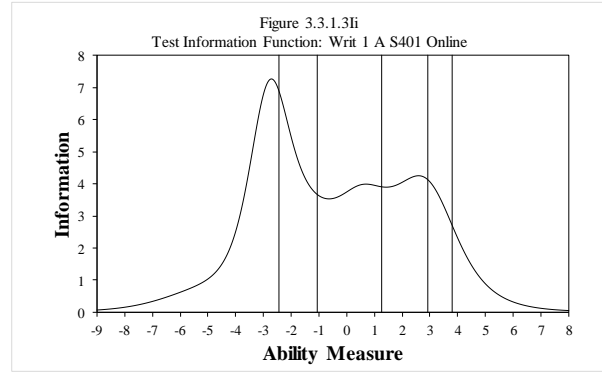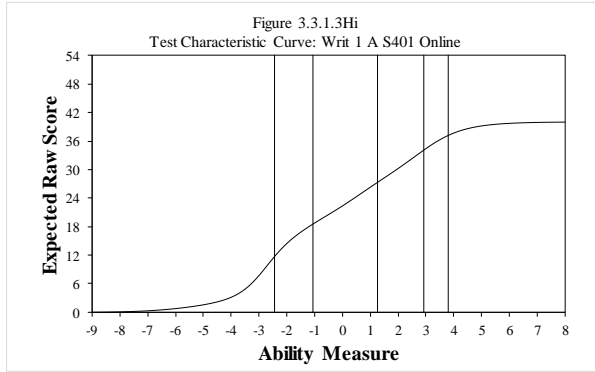
Table 3.3.1.3Ji

Reliability: Writ 1 A S401 Online

Reliability	No. of Students	No. of Tasks	Response Mode	Cronbach's Alpha	SEM
	126,315	4	Hand-written (HW)	.860	1.771
Interrater Reliability	**Task**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	60,726	100	0	0
	2	67,496	100	0	0
	3	71,834	98	2	0
	4	67,132	96	4	0

Table 3.3.1.3Ki
Conditional Standard Error of Measurement at Cut Scores: Writ 1 A S401 Online
n/a

Table 3.3.1.3Li
Accuracy and Consistency of Classification Indices: Writ 1 A S401 Online
n/a

3.3.1.3ii   *Writing 1 B/C*

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

Figure 3.3.1.3Cii
Raw Scores: Writ 1 B/C S401 Online
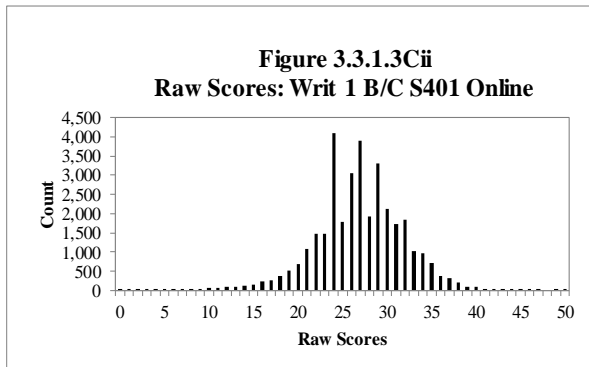
Table 3.3.1.3Cii

Raw Score Descriptive Statistics: Writ 1 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	34,590	0	50	26.77	5.18
Total	34,590	0	50	26.77	5.18

Figure 3.3.1.3Dii
Scale Scores: Writ 1 B/C S401 Online
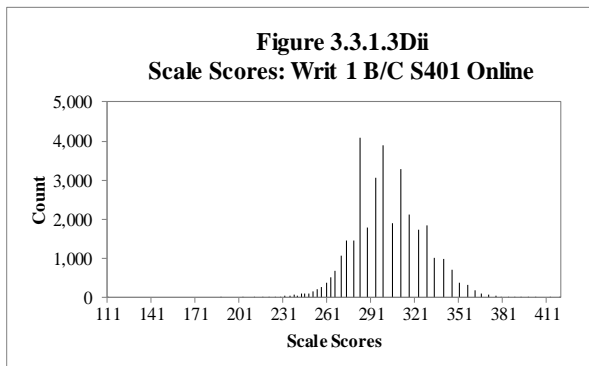
Table 3.3.1.3Dii

Scale Score Descriptive Statistics: Writ 1 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	34,590	111	420	300.04	26.27
Total	34,590	111	420	300.04	26.27

Figure 3.3.1.3Eii
Proficiency Level: Writ 1 B/C S401 Online
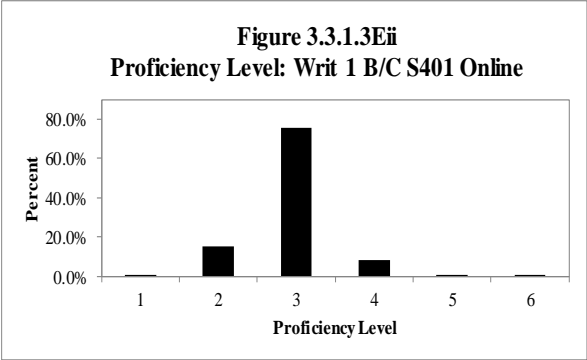
Table 3.3.1.3Eii

Proficiency Level Distribution: Writ 1 B/C S401 Online

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	269	0.78%	269	0.78%
2	5,232	15.13%	5,232	15.13%
3	26,214	75.78%	26,214	75.78%
4	2,829	8.18%	2,829	8.18%
5	44	0.13%	44	0.13%
6	2	0.01%	2	0.01%
Total	34,590	100.00%	34,590	100.00%

Figure 3.3.1.3Hii
Test Characteristic Curve: Writ 1 B/C S401 Online

Figure 3.3.1.3Iii
Test Information Function: Writ 1 B/C S401 Online
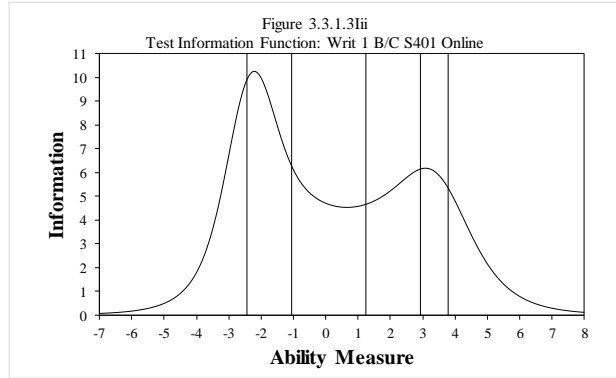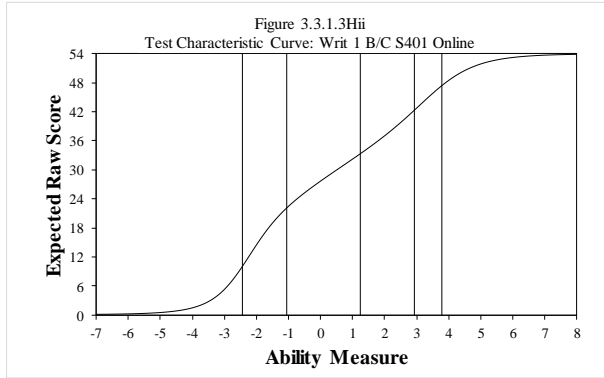
Table 3.3.1.3Jii

Reliability: Writ 1 B/C S401 Online

Reliability	No. of Students	No. of Tasks	Response Mode	Cronbach's Alpha	SEM
	34,590	3	Hand-written (HW)	.885	1.758
Interrater Reliability	Task	No. in Sample	% AG	% AD	% NA
	1	15,816	96	4	0
	2	16,302	94	5	0
	3	16,114	94	5	0

Table 3.3.1.3Kii
Conditional Standard Error of Measurement at Cut Scores: Writ 1 B/C S401 Online
n/a

Table 3.3.1.3Lii
Accuracy and Consistency of Classification Indices: Writ 1 B/C S401 Online
n/a

3.3.1.3iii   Writing 1 Across Tiers

Table 3.3.1.3Aiii
Complete Task Analysis and Summary: Writ 1 S401 Online
n/a

Table 3.3.1.3Biii
DIF Analysis and Summary: Writ 1 S401 Online
n/a

Figure 3.3.1.3Ciii
Raw Scores: Writ 1 S401 Online
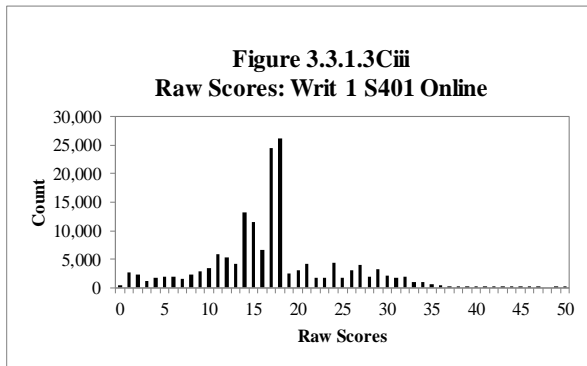
Table 3.3.1.3Ciii
Raw Score Descriptive Statistics: Writ 1 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	160,905	0	50	16.99	7.04
Total	160,905	0	50	16.99	7.04

Figure 3.3.1.3Diii
Scale Scores: Writ 1 S401 Online
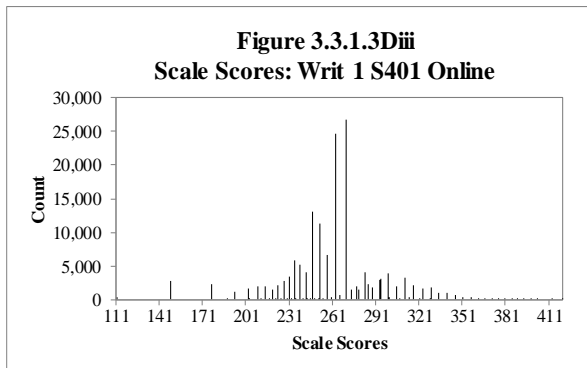
Table 3.3.1.3Diii
Scale Score Descriptive Statistics: Writ 1 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	160,905	111	420	260.52	34.32
Total	160,905	111	420	260.52	34.32

Figure 3.3.1.3Eiii
Proficiency Level: Writ 1 S401 Online
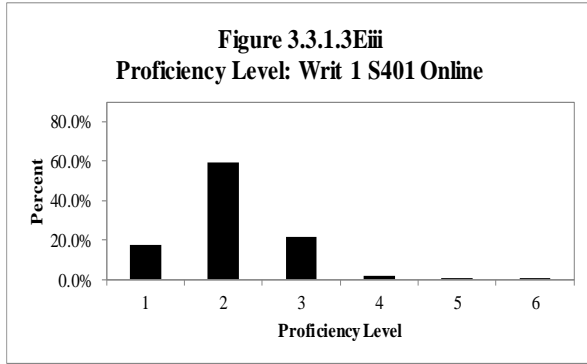
Table 3.3.1.3Eiii

Proficiency Level Distribution: Writ 1 S401 Online

Level	Grade 1 Count	Grade 1 Percent	Total Count	Total Percent
1	28,299	17.59%	28,299	17.59%
2	95,175	59.15%	95,175	59.15%
3	34,556	21.48%	34,556	21.48%
4	2,829	1.76%	2,829	1.76%
5	44	0.03%	44	0.03%
6	2	0.00%	2	0.00%
Total	160,905	100.00%	160,905	100.00%

Table 3.3.1.3Fiii
Raw Score to Scale Score Conversion: Writ 1 S401 Online
n/a

Table 3.3.1.3Giii
Equating Summary: Writ 1 S401 Online
n/a

Figure 3.3.1.3Hiii
Test Characteristic Curve: Writ 1 S401 Online

Figure 3.3.1.3Iiii
Test Information Function: Writ 1 S401 Online
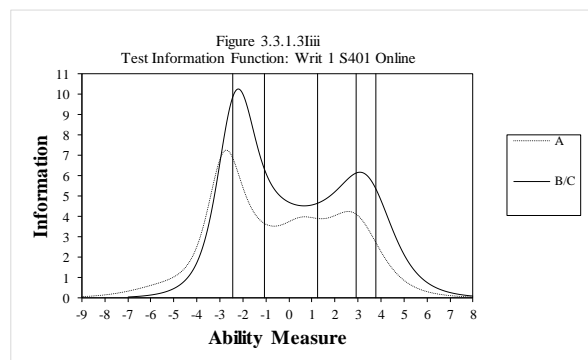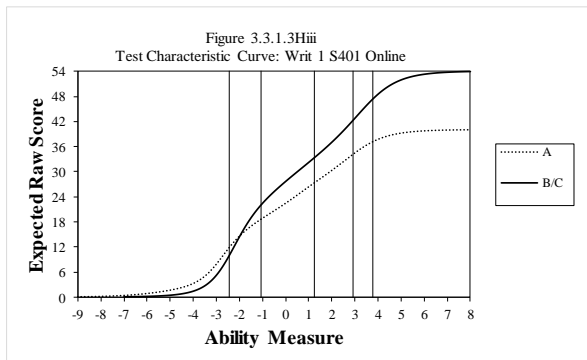
Table 3.3.1.3Jiii

Reliability: Writ 1 Weighted Reliability S401 Online

Tiers	No. of Students	Reliability	Weighted Reliability
A	126,315	0.860	0.865
B/C	34,590	0.885	

Table 3.3.1.3Kiii

Conditional Standard Error of Measurement at Cut Scores: Writ 1 S401 Online

| 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 3.3.1.3L

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.768	0.676		0.479	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.734		0.609	
	2	0.868		0.813	
	3	0.616		0.527	
	4	-		0.396	
	5	-		-	
	6	-		1.000	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.915	0.052	0.033	0.873
	2/3	0.871	0.035	0.095	0.820
	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

3.3.1.4 Speaking 1

3.3.1.4i   Speaking 1 Pre-A

Table 3.3.1.4Ai
Complete Task Analysis and Summary: Spek 1 Pre-A S401 Online
n/a

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

Figure 3.3.1.4Ci
Raw Scores: Spek 1 Pre-A S401 Online
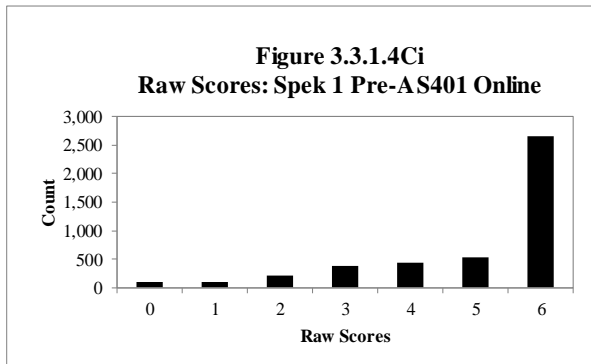
Table 3.3.1.4Ci
Raw Score Descriptive Statistics: Spek 1 Pre-A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	4,432	0	6	4.96	1.57
Total	4,432	0	6	4.96	1.57

Figure 3.3.1.4Di
Scale Scores: Spek 1 Pre-A S401 Online

Table 3.3.1.4Di

Scale Score Descriptive Statistics: Spek 1 Pre-A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	4,432	106	174	160.70	19.86
Total	4,432	106	174	160.70	19.86

Figure 3.3.1.4Ei
Proficiency Level: Spek 1 Pre-A S401 Online
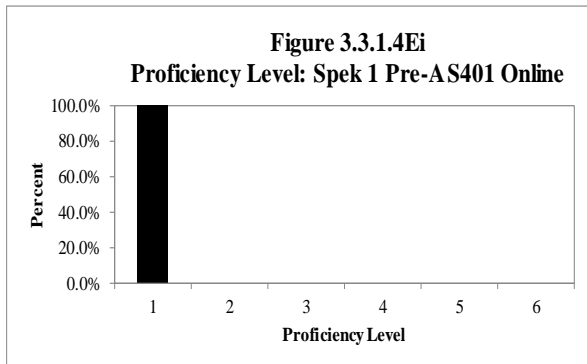
Table 3.3.1.4Ei

Proficiency Level Distribution: Spek 1 Pre-A S401 Online

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	4,432	100.00%	4,432	100.00%
Total	4,432	100.00%	4,432	100.00%

Table 3.3.1.4Gi
Equating Summary: Spek 1 Pre-A S401 Online
n/a

Figure 3.3.1.4Hi
Test Characteristic Curve: Spek 1 Pre-A S401 Online

Figure 3.3.1.4Ii
Test Information Function: Spek 1 Pre-A S401 Online
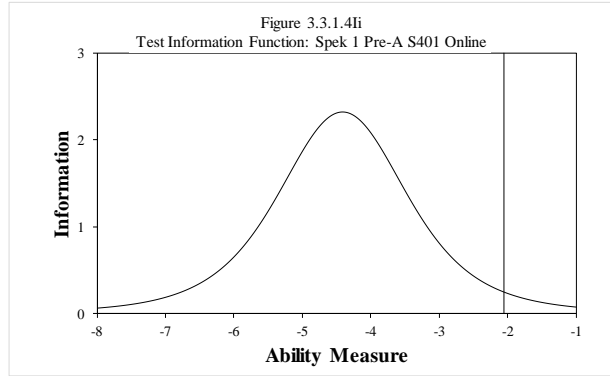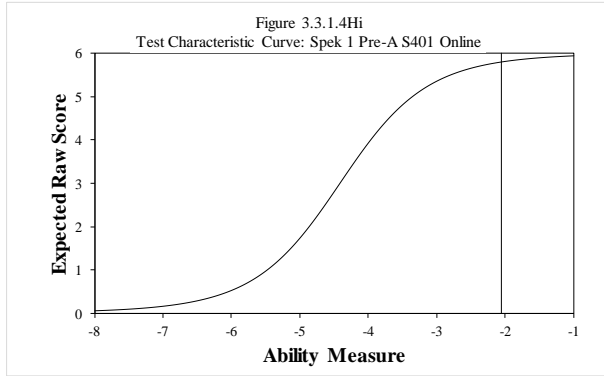
Table 3.3.1.4Ji

Reliability: Spek 1 Pre-A S401 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	4,432	3	.770		0.751
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	3,380	98	2	0
	2	3,468	97	3	0
	3	3,754	98	2	0

Table 3.3.1.4Ki
Conditional Standard Error of Measurement at Cut Scores: Spek 1 Pre-A S401 Online
**n/a**

Table 3.3.1.4Li
Accuracy and Consistency of Classification Indices: Spek 1 Pre-A S401 Online
n/a

3.3.1.4ii   Speaking 1 A

Table 3.3.1.4Aii
Complete Task Analysis and Summary: Spek 1 A S401 Online
n/a

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


Figure 3.3.1.4Cii
Raw Scores: Spek 1 A S401 Online
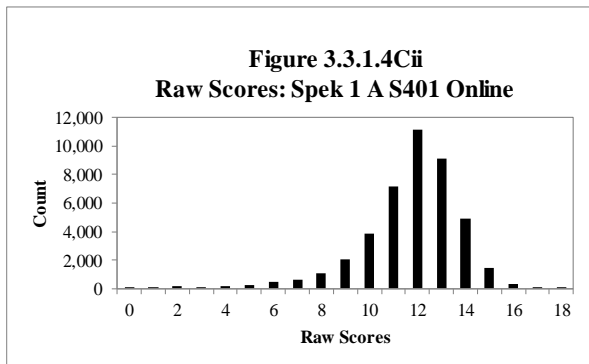
Table 3.3.1.4Cii
Raw Score Descriptive Statistics: Spek 1 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	43,064	0	18	11.68	2.19
Total	43,064	0	18	11.68	2.19

Figure 3.3.1.4Dii
Scale Scores: Spek 1 A S401 Online
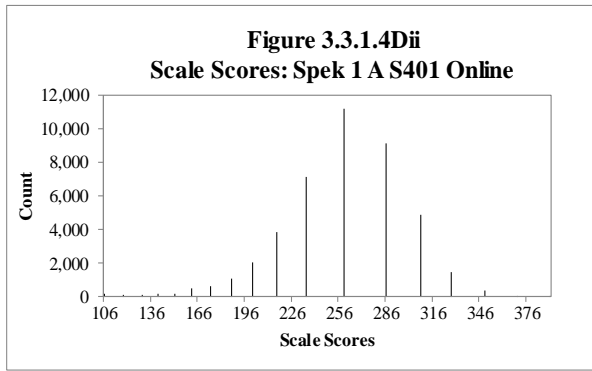
Table 3.3.1.4Dii

Scale Score Descriptive Statistics: Spek 1 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	43,064	106	391	256.34	41.88
Total	43,064	106	391	256.34	41.88

Figure 3.3.1.4Eii
Proficiency Level: Spek 1 A S401 Online
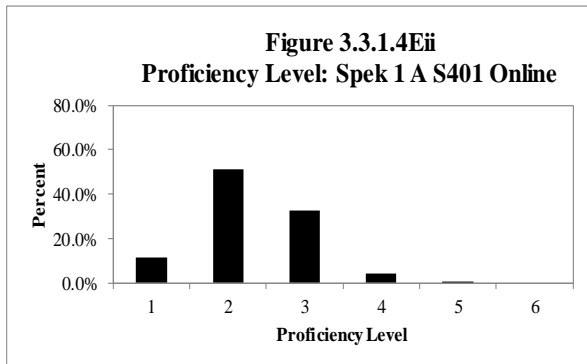
Table 3.3.1.4Eii

Proficiency Level Distribution: Spek 1 A S401 Online

Level	Grade 1 Count	Grade 1 Percent	Total Count	Total Percent
1	5,067	11.77%	5,067	11.77%
2	22,126	51.38%	22,126	51.38%
3	13,976	32.45%	13,976	32.45%
4	1,840	4.27%	1,840	4.27%
5	55	0.13%	55	0.13%
6	0	0.00%	0	0.00%
Total	43,064	100.00%	43,064	100.00%

Table 3.3.1.4Gii
Equating Summary: Spek 1 A S401 Online
n/a

Figure 3.3.1.4Hii
Test Characteristic Curve: Spek 1 A S401 Online

Figure 3.3.1.4Iii
Test Information Function: Spek 1 A S401 Online
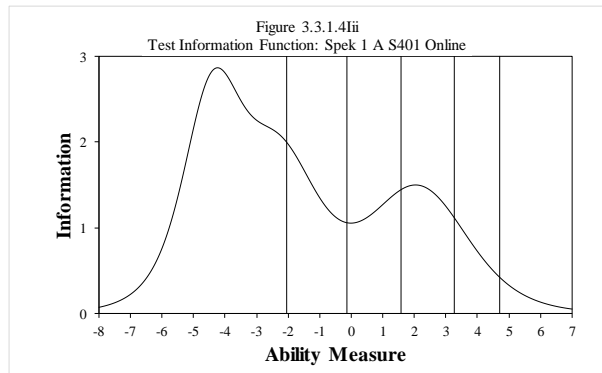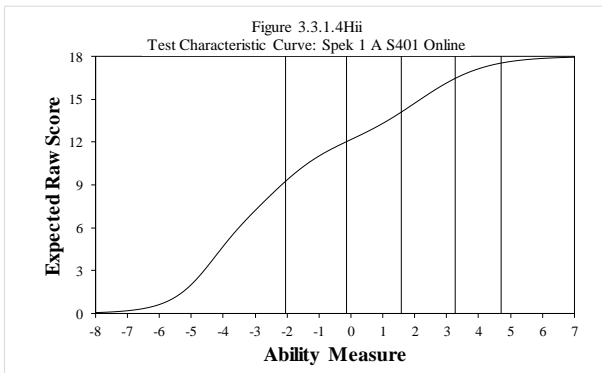
Table 3.3.1.4Jii

Reliability: Spek 1 A S401 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	43,064	6	.688		1.223
Interrater Reliability	Task	No. in Sample	% EX	% AD	% NA
	1	26,276	99	1	0
	2	26,286	88	12	0
	3	26,348	98	2	0
	4	26,346	88	12	0
	5	26,936	98	2	0
	6	26,936	88	12	1

Table 3.3.1.4Kii
Conditional Standard Error of Measurement at Cut Scores: Spek 1 A S401 Online
n/a

Table 3.3.1.4Lii
Accuracy and Consistency of Classification Indices: Spek 1 A S401 Online
n/a

3.3.1.4iii   *Speaking 1 B/C*

Table 3.3.1.4Aiii
Complete Task Analysis and Summary: Spek 1 B/C S401 Online
n/a

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

Figure 3.3.1.4Ciii
Raw Scores: Spek 1 B/C S401 Online
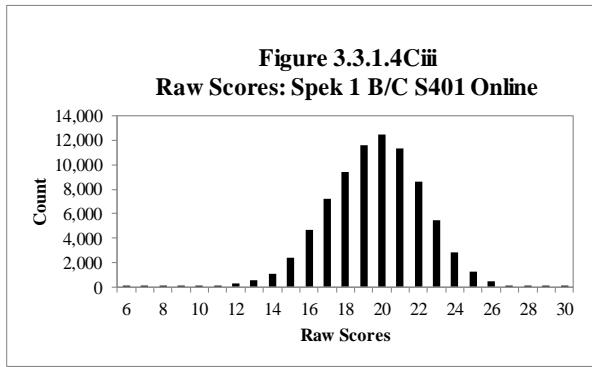
Table 3.3.1.4Ciii

Raw Score Descriptive Statistics: Spek 1 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	80,186	6	30	19.62	2.62
Total	80,186	6	30	19.62	2.62

Figure 3.3.1.4Diii
Scale Scores: Spek 1 B/C S401 Online
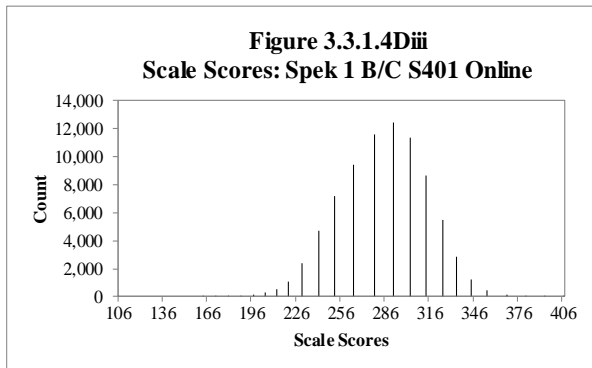
Table 3.3.1.4Diii

Scale Score Descriptive Statistics: Spek 1 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	80,186	106	407	285.48	30.78
Total	80,186	106	407	285.48	30.78
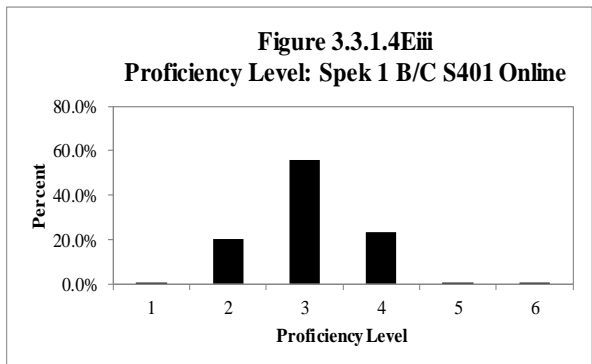
Table 3.3.1.4Eiii

Proficiency Level Distribution: Spek 1 B/C S401 Online

Figure 3.3.1.4Eiii
Proficiency Level: Spek 1 B/C S401 Online

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	272	0.34%	272	0.34%
2	16,149	20.14%	16,149	20.14%
3	44,825	55.90%	44,825	55.90%
4	18,725	23.35%	18,725	23.35%
5	212	0.26%	212	0.26%
6	3	0.00%	3	0.00%
Total	80,186	100.00%	80,186	100.00%

Table 3.3.1.4Giii
Equating Summary: Spek 1 B/C S401 Online
n/a

Figure 3.3.1.4Hiii
Test Characteristic Curve: Spek 1 B/C S401 Online

Figure 3.3.1.4Iiii
Test Information Function: Spek 1 B/C S401 Online
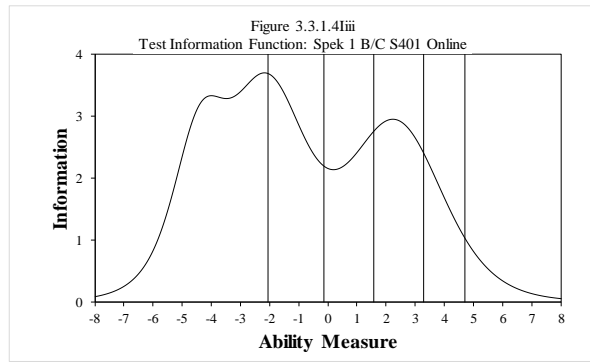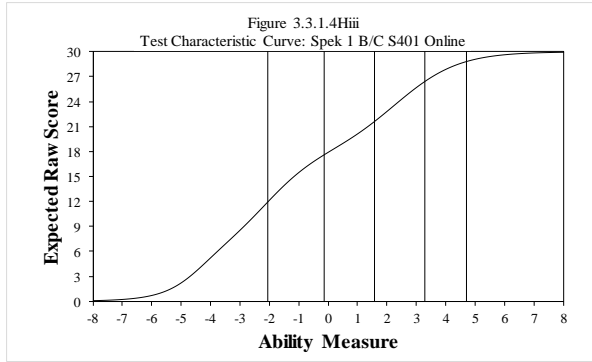
Table 3.3.1.4Jiii

Reliability: Spek 1 B/C S401 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	80,186	6	.651		1.549
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	45,470	84	16	0
	2	45,468	86	14	0
	3	46,086	89	11	0
	4	46,082	86	14	0
	5	43,866	85	15	0
	6	43,868	83	17	0

Table 3.3.1.4Kiii
Conditional Standard Error of Measurement at Cut Scores: Spek 1 B/C S401 Online
n/a

Table 3.3.1.4Liii
Accuracy and Consistency of Classification Indices: Spek 1 B/C S401 Online
n/a

3.3.1.4iv   Speaking 1 Across Tiers

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

Table 3.3.1.4Biv
DIF Analysis and Summary: Spek 1 S401 Online
n/a

Figure 3.3.1.4Civ
Raw Scores: Spek 1 S401 Online
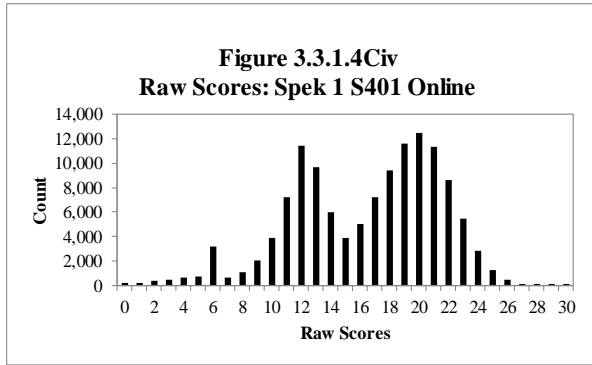
Table 3.3.1.4Civ

Raw Score Descriptive Statistics: Spek 1 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	127,682	0	30	16.43	4.96
Total	127,682	0	30	16.43	4.96

Figure 3.3.1.4Div
Scale Scores: Spek 1 S401 Online
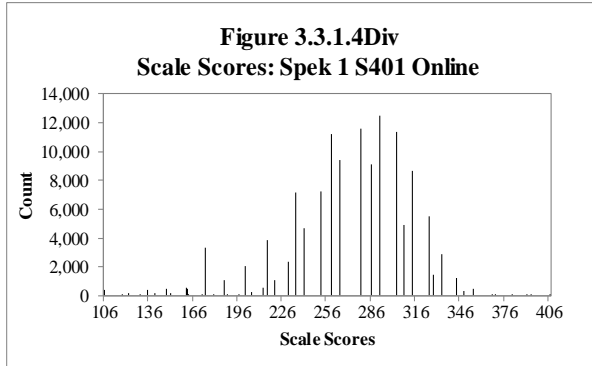
Table 3.3.1.4Div

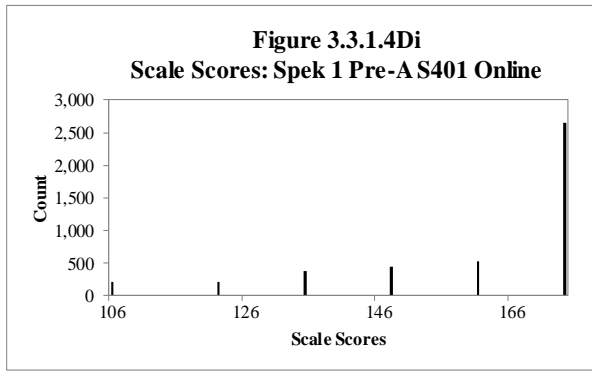Scale Score Descriptive Statistics: Spek 1 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	127,682	106	407	271.32	42.74
Total	127,682	106	407	271.32	42.74
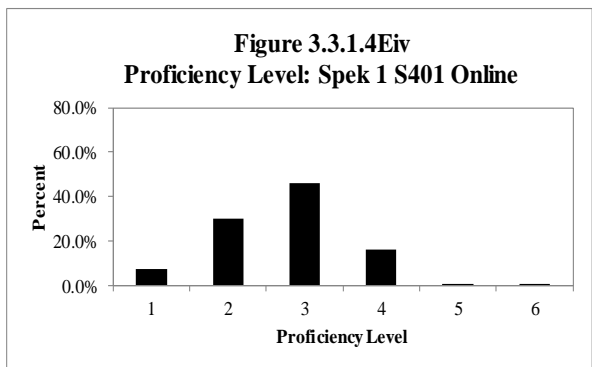
Table 3.3.1.4Eiv

Proficiency Level Distribution: Spek 1 S401 Online

Figure 3.3.1.4Eiv
Proficiency Level: Spek 1 S401 Online

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	9,771	7.65%	9,771	7.65%
2	38,275	29.98%	38,275	29.98%
3	58,801	46.05%	58,801	46.05%
4	20,565	16.11%	20,565	16.11%
5	267	0.21%	267	0.21%
6	3	0.00%	3	0.00%
Total	127,682	100.00%	127,682	100.00%

Table 3.3.1.4Fiv
Raw Score to Scale Score Conversion: Spek 1 S401 Online
n/a

Figure 3.3.1.4Hiv
Test Characteristic Curve: Spek 1 S401 Online

Figure 3.3.1.4Iiv
Test Information Function: Spek 1 S401 Online
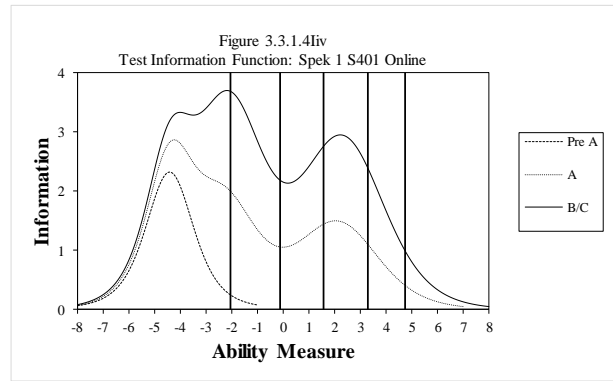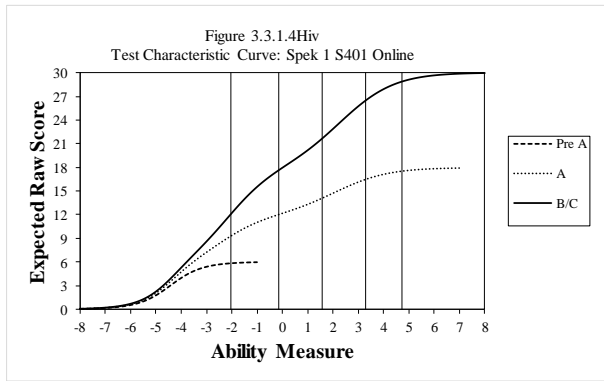
Table 3.3.1.4Jiv

Reliability: Spek 1 Weighted Reliability S401 Online

Tiers	No. of Students	Reliability	Weighted Reliability
Pre-A	4,432	0.770	
A	43,064	0.688	0.668
B/C	80,186	0.651	

Table 3.3.1.4Kiv

Conditional Standard Error of Measurement at Cut Scores: Spek 1 S401 Online

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

Note: Tier Pre-A is not presented as it is not possible for Tier Pre-A students to receive a proficeny level higher than 2.

Table 3.3.1.4L

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.623	0.504		0.247	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.774		0.602	
	2	0.704		0.509	
	3	0.585		0.558	
	4	-		0.262	
	5	-		-	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.961	0.015	0.024	0.941
	2/3	0.824	0.039	0.137	0.753
	3/4	0.837	0.163	0.000	0.766
	4/5	0.998	0.002	0.000	0.997
	5/6	1.000	0.000	0.000	1.000

3.3.2 Grades: 2-3

3.3.2.1   Listening 2-3

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

Figure 3.3.2.1D
Scale Scores: List 2-3 S401 Online

Figure 3.3.2.1E
Proficiency Level: List 2-3 S401 Online
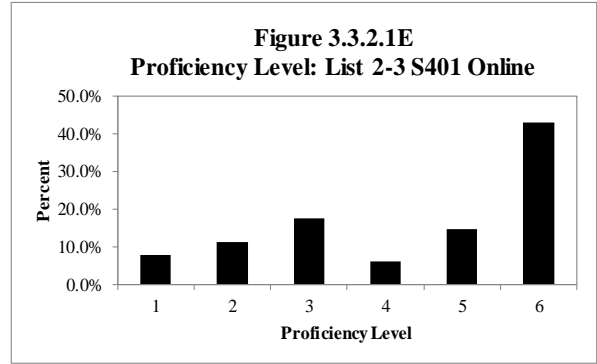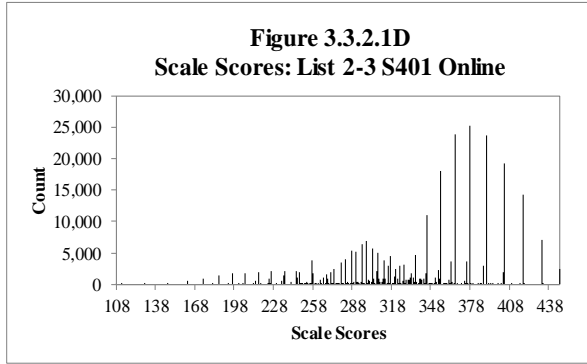
Table 3.3.2.1D

Scale Score Descriptive Statistics: List 2-3 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	145,452	108	446	328.06	54.92
3	153,970	112	446	351.63	55.54
Total	299,422	108	446	340.18	56.48

Table 3.3.2.1E

Proficiency Level Distribution: List 2-3 S401 Online

Level	Grade 2 Count	Grade 2 Percent	Grade 3 Count	Grade 3 Percent	Total Count	Total Percent
1	11,107	7.64%	11,937	7.75%	23,044	7.70%
2	17,509	12.04%	16,179	10.51%	33,688	11.25%
3	30,031	20.65%	22,228	14.44%	52,259	17.45%
4	8,594	5.91%	10,069	6.54%	18,663	6.23%
5	18,255	12.55%	25,313	16.44%	43,568	14.55%
6	59,956	41.22%	68,244	44.32%	128,200	42.82%
Total	145,452	100.00%	153,970	100.00%	299,422	100.00%

Table 3.3.2.1F
Raw Score to Scale Score Conversion: List 2-3 S401 Online
n/a

Figure 3.3.2.1H
Test Characteristic Curve: List 2-3 S401 Online
n/a

Figure 3.3.2.1I
Test Information Function: List 2-3 S401 Online
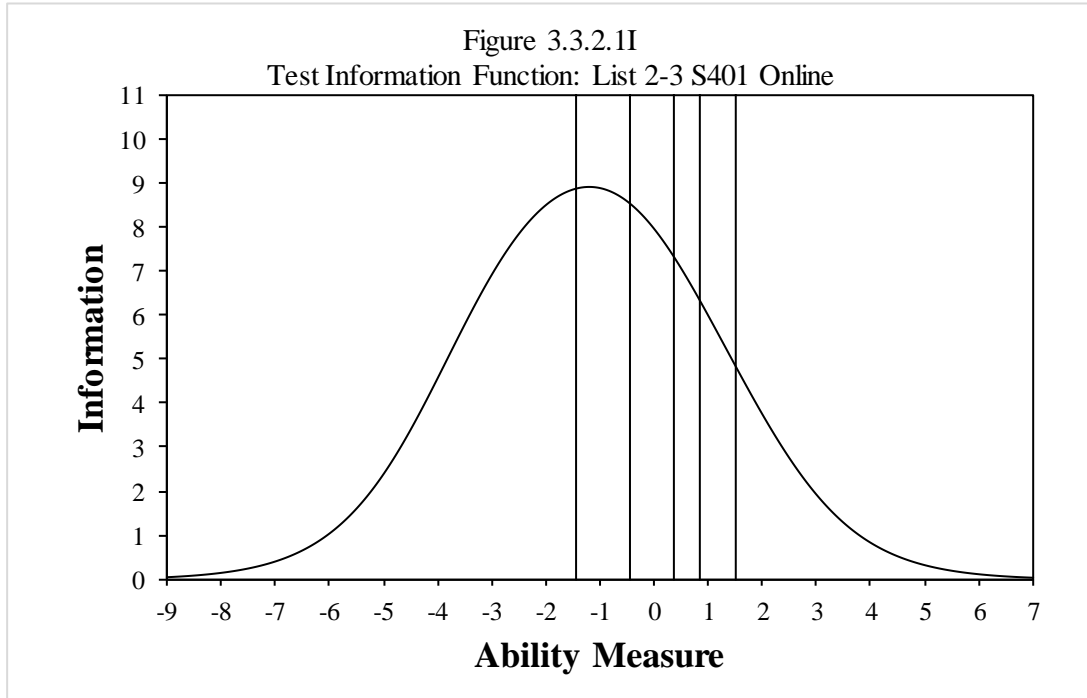
Table 3.3.2.1J

Reliability: List 2-3 S401 Online

No. of Students	No. of Items	Rasch Reliability Estimate
299,422	54	.86

Table 3.3.2.1K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: List 2-3 S401 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	2	245	860	19.90	20.41	19.96	0.17
	3	262	N/A	N/A	N/A	N/A	N/A
2/3	2	283	90	17.86	17.86	17.86	0.00
	3	300	428	17.35	17.35	17.35	0.00
3/4	2	314	185	17.86	18.88	18.29	0.39
	3	331	216	19.90	19.90	19.90	0.00
4/5	2	330	336	18.37	19.90	19.06	0.67
	3	349	82	20.41	22.45	22.03	0.44
5/6	2	354	924	20.41	21.94	20.57	0.46
	3	374	107	23.47	23.47	23.47	0.00

Table 3.3.2.1Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.629	0.545		0.404	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.789		0.622	
	2	0.514		0.396	
	3	0.591		0.474	
	4	0.177		0.127	
	5	0.346		0.252	
	6	0.894		0.832	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.961	0.014	0.025	0.943
	2/3	0.916	0.039	0.045	0.884
	3/4	0.894	0.033	0.073	0.853
	4/5	0.903	0.048	0.049	0.854
	5/6	0.887	0.073	0.040	0.842

Table 3.3.2.1Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.639	0.557		0.402	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.800		0.651	
	2	0.518		0.399	
	3	0.499		0.381	
	4	0.208		0.148	
	5	0.422		0.317	
	6	0.871		0.814	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.963	0.013	0.024	0.946
	2/3	0.929	0.029	0.043	0.902
	3/4	0.909	0.033	0.058	0.870
	4/5	0.902	0.052	0.046	0.856
	5/6	0.875	0.070	0.055	0.827

3.3.2.2   Reading 2-3

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

Figure 3.3.2.2C
Raw Scores: Read 2-3 S401 Online
n/a

Table 3.3.2.2C
Raw Score Descriptive Statistics: Read 2-3 S401 Online
n/a

Figure 3.3.2.2D
Scale Scores: Read 2-3 S401 Online

Figure 3.3.2.2E
Proficiency Level: Read 2-3 S401 Online
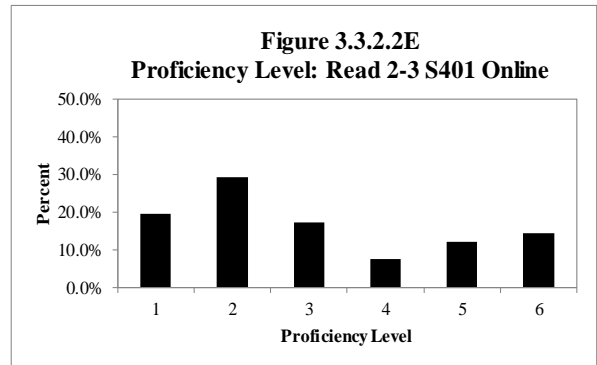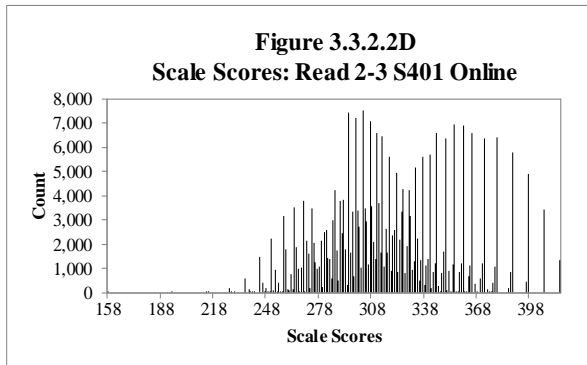
Table 3.3.2.2D

Scale Score Descriptive Statistics: Read 2-3 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	142,362	194	415	311.00	32.51
3	149,280	158	415	328.49	38.39
Total	291,642	158	415	319.95	36.70

Table 3.3.2.2E

Proficiency Level Distribution: Read 2-3 S401 Online

Level	Grade 2 Count	Grade 2 Percent	Grade 3 Count	Grade 3 Percent	Total Count	Total Percent
1	25,704	18.06%	31,198	20.90%	56,902	19.51%
2	43,882	30.82%	41,585	27.86%	85,467	29.31%
3	28,595	20.09%	21,938	14.70%	50,533	17.33%
4	11,938	8.39%	10,179	6.82%	22,117	7.58%
5	16,671	11.71%	18,250	12.23%	34,921	11.97%
6	15,572	10.94%	26,130	17.50%	41,702	14.30%
Total	142,362	100.00%	149,280	100.00%	291,642	100.00%

Table 3.3.2.2F
Raw Score to Scale Score Conversion: Read 2-3 S401 Online
n/a

Figure 3.3.2.2H
Test Characteristic Curve: Read 2-3 S401 Online
n/a

Figure 3.3.2.2I
Test Information Function: Read 2-3 S401 Online
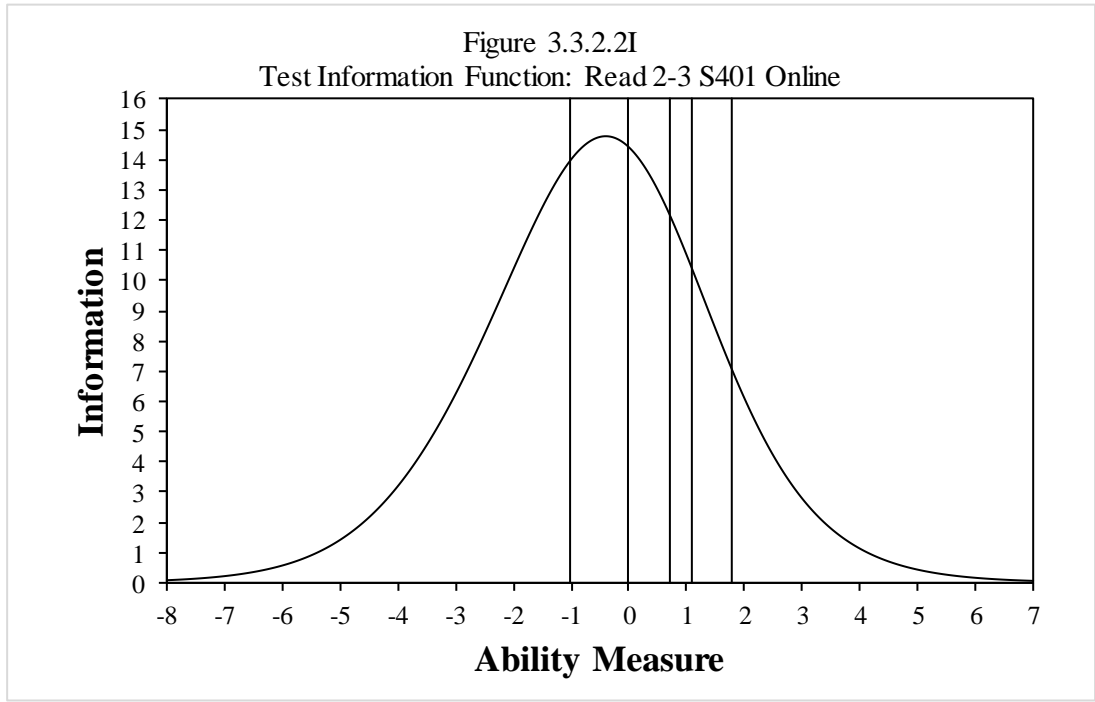
Table 3.3.2.2J

Reliability: Read 2-3 S401 Online

No. of Students	No. of Items	Rasch Reliability Estimate
291,642	72	.90

Table 3.3.2.2K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: Read 2-3 S401 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	2	283	906	10.20	12.24	11.00	0.46
	3	297	1,267	10.20	11.22	10.68	0.51
2/3	2	307	3,171	9.69	10.20	9.73	0.13
	3	323	464	9.69	11.22	9.96	0.29
3/4	2	326	578	9.69	10.20	10.01	0.25
	3	342	182	10.20	11.22	10.85	0.25
4/5	2	337	3,730	10.20	10.71	10.20	0.01
	3	352	861	10.71	12.24	11.30	0.21
5/6	2	355	2,937	11.22	11.73	11.23	0.03
	3	370	593	13.27	13.27	13.27	0.00

Table 3.3.2.2Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.651	0.544		0.433	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.754		0.645	
	2	0.698		0.599	
	3	0.564		0.446	
	4	0.353		0.256	
	5	0.583		0.455	
	6	0.851		0.747	
Indices at Cut Points		Accuracy			
			False Positives	False Negatives	
	Cut Point	Accuracy			Consistency
	1/2	0.914	0.046	0.041	0.877
	2/3	0.895	0.051	0.054	0.854
	3/4	0.924	0.040	0.036	0.890
	4/5	0.938	0.038	0.025	0.911
	5/6	0.960	0.025	0.015	0.944

Table 3.3.2.2Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.648	0.550		0.446	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.756		0.665	
	2	0.675		0.568	
	3	0.470		0.356	
	4	0.293		0.208	
	5	0.542		0.420	
	6	0.877		0.796	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.904	0.054	0.042	0.867
	2/3	0.904	0.044	0.052	0.865
	3/4	0.927	0.040	0.033	0.894
	4/5	0.933	0.040	0.027	0.905
	5/6	0.948	0.032	0.020	0.927

3.3.2.3   Writing 2-3

3.3.2.3i   *Writing 2-3 A*

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

Figure 3.3.2.3Ci
Raw Scores: Writ 2-3 A S401 Online
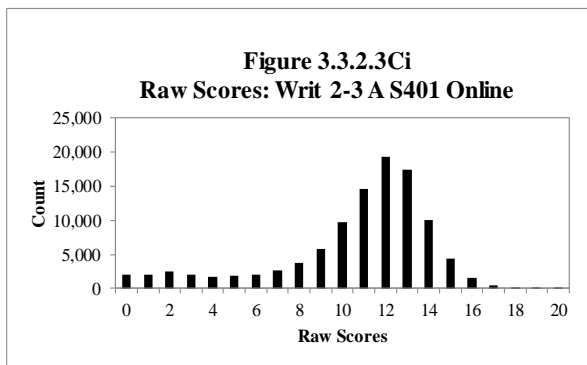
Table 3.3.2.3Ci
Raw Score Descriptive Statistics: Writ 2-3 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	61,602	0	19	10.33	3.62
3	42,063	0	20	10.92	3.64
Total	103,665	0	20	10.57	3.64

Figure 3.3.2.3Di
Scale Scores: Writ 2-3 A S401 Online
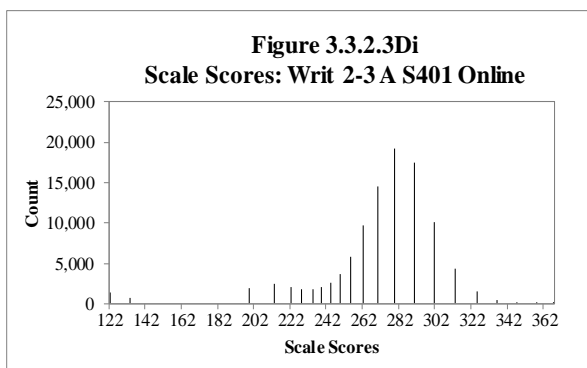
Table 3.3.2.3Di
Scale Score Descriptive Statistics: Writ 2-3 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	61,602	122	358	267.14	32.90
3	42,063	133	367	273.01	32.36
Total	103,665	122	367	269.52	32.81

Figure 3.3.2.3Ei
Proficiency Level: Writ 2-3 A S401 Online
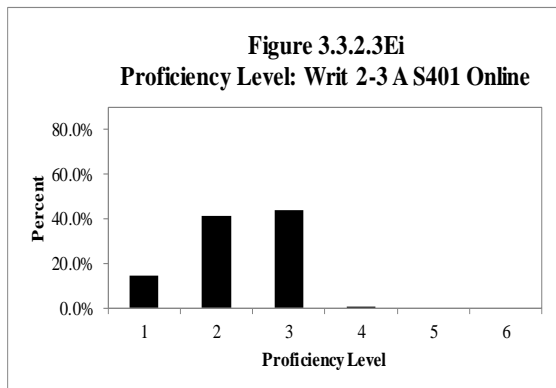
Table 3.3.2.3Ei

Proficiency Level Distribution: Writ 2-3 A S401 Online

Level	Grade 2		Grade 3		Total	
	Count	Percent	Count	Percent	Count	Percent
1	8,943	14.52%	6,192	14.72%	15,135	14.60%
2	23,378	37.95%	19,478	46.31%	42,856	41.34%
3	29,254	47.49%	16,327	38.82%	45,581	43.97%
4	27	0.04%	66	0.16%	93	0.09%
5	0	0.00%	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%	0	0.00%
Total	61,602	100.00%	42,063	100.00%	103,665	100.00%

Figure 3.3.2.3Hi
Test Characteristic Curve: Writ 2-3 A S401 Online

Figure 3.3.2.3Ii
Test Information Function: Writ 2-3 A S401 Online
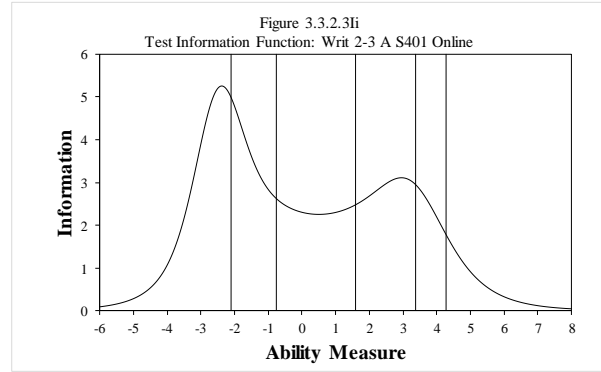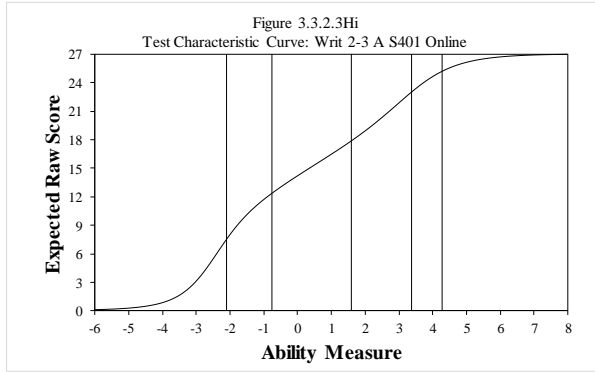
Table 3.3.2.3Ji

Reliability: Writ 2-3 A S401 Online

Reliability	No. of Students	No. of Tasks	Response Mode	Cronbach's Alpha	SEM
	103,665	3	Hand-written (HW)	.863	1.347
Interrater Reliability	Task	No. in Sample	% AG	% AD	% NA
	1	52,575	94	5	1
	2	52,446	96	3	0
	3	57,144	96	3	0

Table 3.3.2.3Ki
Conditional Standard Error of Measurement at Cut Scores: Writ 2-3 A S401 Online
n/a

Table 3.3.2.3Li
Accuracy and Consistency of Classification Indices: Writ 2-3 A S401 Online
n/a

3.3.2.3ii *Writing 2-3 B/C*

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

Figure 3.3.2.3Cii
Raw Scores: Writ 2-3 B/C S401 Online
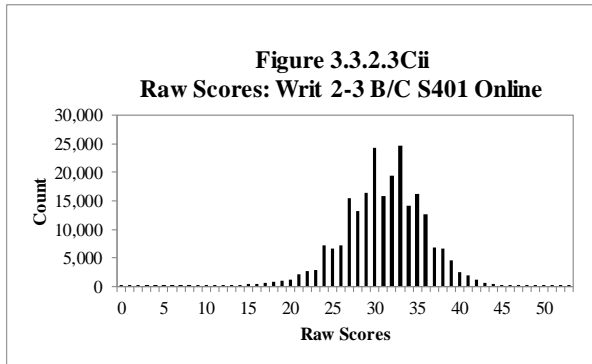
Table 3.3.2.3Cii

Raw Score Descriptive Statistics: Writ 2-3 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	102,601	0	52	29.36	5.10
3	129,812	0	53	32.25	4.68
Total	232,413	0	53	30.97	5.08

Figure 3.3.2.3Dii
Scale Scores: Writ 2-3 B/C S401 Online
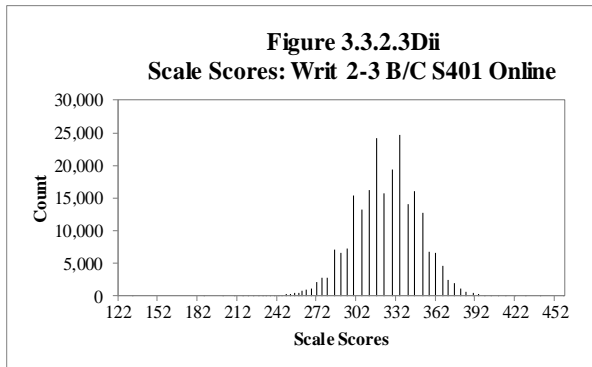
Table 3.3.2.3Dii

Scale Score Descriptive Statistics: Writ 2-3 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	102,601	122	441	314.83	26.31
3	129,812	133	459	330.37	25.12
Total	232,413	122	459	323.51	26.78

Figure 3.3.2.3Eii
Proficiency Level: Writ 2-3 B/C S401 Online
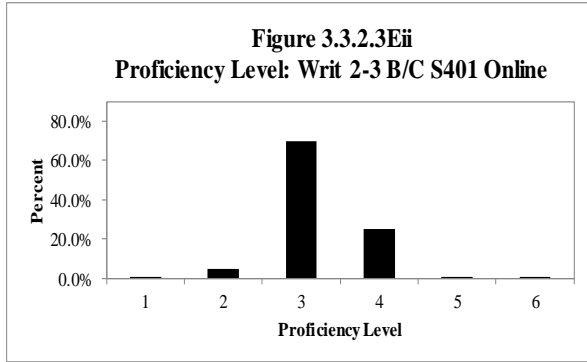
Table 3.3.2.3Eii

Proficiency Level Distribution: Writ 2-3 B/C S401 Online

Level	Grade 2		Grade 3		Total	
	Count	Percent	Count	Percent	Count	Percent
1	803	0.78%	379	0.29%	1,182	0.51%
2	7,130	6.95%	3,560	2.74%	10,690	4.60%
3	76,364	74.43%	85,198	65.63%	161,562	69.52%
4	18,198	17.74%	40,170	30.94%	58,368	25.11%
5	104	0.10%	484	0.37%	588	0.25%
6	2	0.00%	21	0.02%	23	0.01%
Total	102,601	100.00%	129,812	100.00%	232,413	100.00%

Figure 3.3.2.3Hii
Test Characteristic Curve: Writ 2-3 B/C S401 Online

Figure 3.3.2.3Iii
Test Information Function: Writ 2-3 B/C S401 Online
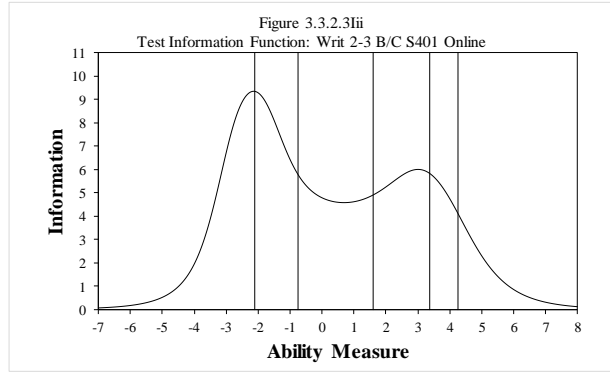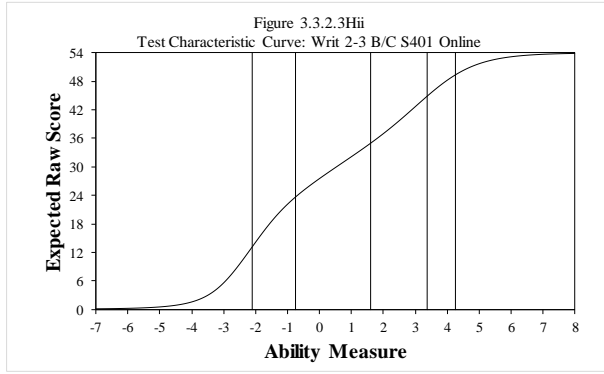
Table 3.3.2.3Jii

Reliability: Writ 2-3 B/C S401 Online

Reliability	No. of Students	No. of Tasks	Response Mode	Cronbach's Alpha	SEM
	232,413	3	Hand-written (HW)	.896	1.636
Interrater Reliability	**Task**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	101,738	95	5	0
	2	101,640	93	6	0
	3	102,426	95	5	0

Table 3.3.2.3Kii
Conditional Standard Error of Measurement at Cut Scores: Writ 2-3 B/C S401 Online
n/a

Table 3.3.2.3Lii
Accuracy and Consistency of Classification Indices: Writ 2-3 B/C S401 Online
n/a

3.3.2.3iii   *Writing 2-3 Across Tiers*

Table 3.3.2.3Aiii
Complete Task Analysis and Summary: Writ 2-3 S401 Online
n/a

Table 3.3.2.3Biii
DIF Analysis and Summary: Writ 2-3 S401 Online
n/a

Figure 3.3.2.3Ciii
Raw Scores: Writ 2-3 S401 Online
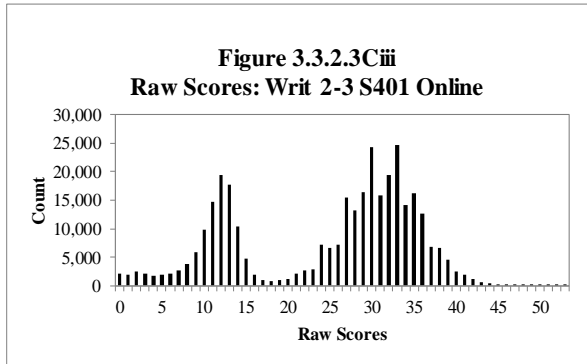
Table 3.3.2.3Ciii
Raw Score Descriptive Statistics: Writ 2-3 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	164,203	0	52	22.22	10.30
3	171,875	0	53	27.03	10.19
Total	336,078	0	53	24.68	10.52

Figure 3.3.2.3Diii
Scale Scores: Writ 2-3 S401 Online
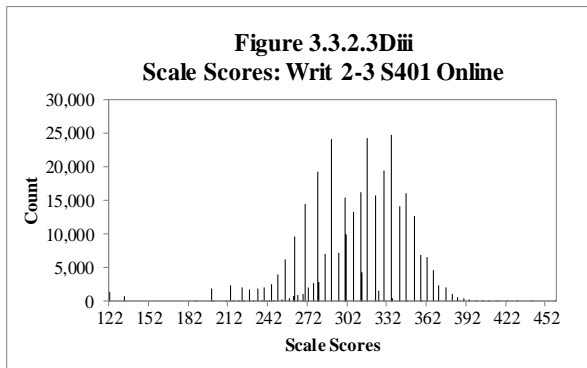
Table 3.3.2.3Diii
Scale Score Descriptive Statistics: Writ 2-3 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	164,203	122	441	296.94	37.03
3	171,875	133	459	316.33	36.62
Total	336,078	122	459	306.85	38.08

Figure 3.3.2.3Eiii
Proficiency Level: Writ 2-3 S401 Online
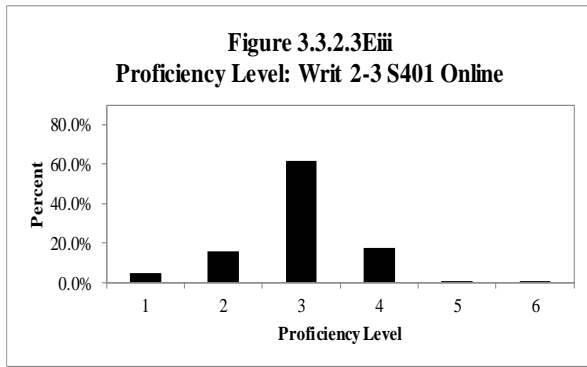
Table 3.3.2.3Eiii

Proficiency Level Distribution: Writ 2-3 S401 Online

Level	Grade 2 Count	Grade 2 Percent	Grade 3 Count	Grade 3 Percent	Total Count	Total Percent
1	9,746	5.94%	6,571	3.82%	16,317	4.86%
2	30,508	18.58%	23,038	13.40%	53,546	15.93%
3	105,618	64.32%	101,525	59.07%	207,143	61.64%
4	18,225	11.10%	40,236	23.41%	58,461	17.40%
5	104	0.06%	484	0.28%	588	0.17%
6	2	0.00%	21	0.01%	23	0.01%
Total	164,203	100.00%	171,875	100.00%	336,078	100.00%

Table 3.3.2.3Fiii
Raw Score to Scale Score Conversion: Writ 2-3 S401 Online
n/a

Table 3.3.2.3Giii
Equating Summary: Writ 2-3 S401 Online
n/a

Figure 3.3.2.3Hiii
Test Characteristic Curve: Writ 2-3 S401 Online

Figure 3.3.2.3Iiii
Test Information Function: Writ 2-3 S401 Online
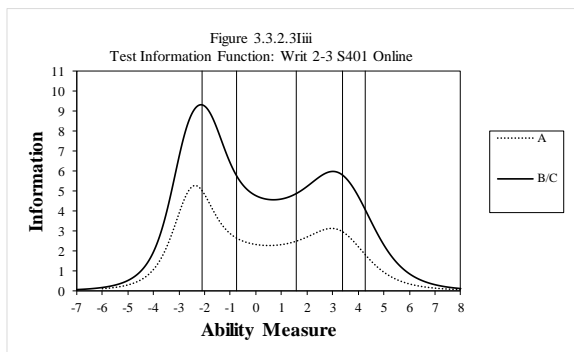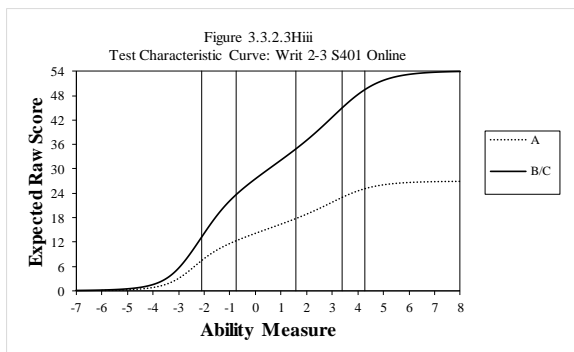
Table 3.3.2.3Jiii

Reliability: Writ 2-3 Weighted Reliability S401 Online

Tiers	No. of Students	Reliability	Weighted Reliability
A	103,665	0.863	0.886
B/C	232,413	0.896	

Table 3.3.2.3Kiii

Conditional Standard Error of Measurement at Cut Scores: Writ 2-3 S401 Online

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	2	242	11.81	8.86
	3	247	12.08	8.86
2/3	2	279	16.27	11.01
	3	283	16.65	11.28
3/4	2	341	17.18	12.30
	3	346	17.18	12.14
4/5	2	388	15.31	11.01
	3	394	15.65	11.14
5/6	2	411	18.26	12.35
	3	418	20.41	13.43

Table 3.3.2.3Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.795	0.732		0.476	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.794		0.673	
	2	0.776		0.661	
	3	0.799		0.795	
	4	-		0.324	
	5	-		-	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.976	0.013	0.011	0.965
	2/3	0.930	0.027	0.043	0.902
	3/4	0.888	0.112	0.000	0.865
	4/5	0.999	0.001	0.000	0.999
	5/6	1.000	0.000	0.000	1.000

Table 3.3.2.3Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.775	0.697		0.468	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.781		0.650	
	2	0.772		0.653	
	3	0.811		0.758	
	4	0.678		0.566	
	5	-		-	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.983	0.008	0.008	0.975
	2/3	0.948	0.019	0.033	0.927
	3/4	0.847	0.085	0.068	0.796
	4/5	0.997	0.003	0.000	0.997
	5/6	1.000	0.000	0.000	1.000

3.3.2.4 Speaking 2-3

3.3.2.4i Speaking 2-3 Pre-A

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

Figure 3.3.2.4Ci
Raw Scores: Spek 2-3 Pre-A S401 Online
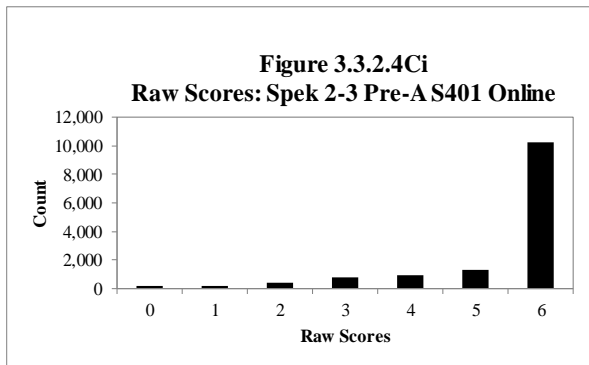
Table 3.3.2.4Ci
Raw Score Descriptive Statistics: Spek 2-3 Pre-A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	5,414	0	6	5.27	1.36
3	8,618	0	6	5.39	1.26
Total	14,032	0	6	5.35	1.30

Figure 3.3.2.4Di
Scale Scores: Spek 2-3 Pre-A S401 Online
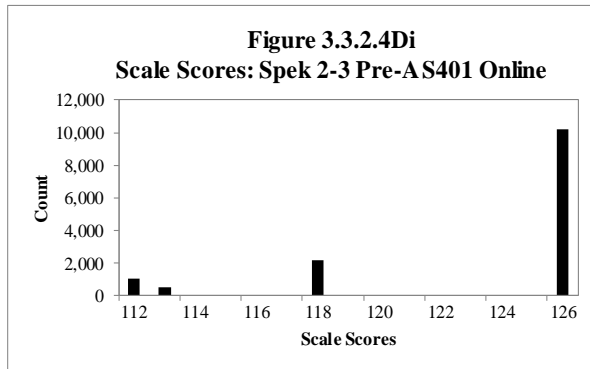
Table 3.3.2.4Di

Scale Score Descriptive Statistics: Spek 2-3 Pre-A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	5,414	112	126	121.95	6.25
3	8,618	118	126	123.95	3.49
Total	14,032	112	126	123.18	4.85

Figure 3.3.2.4Ei
Proficiency Level: Spek 2-3 Pre-A S401 Online
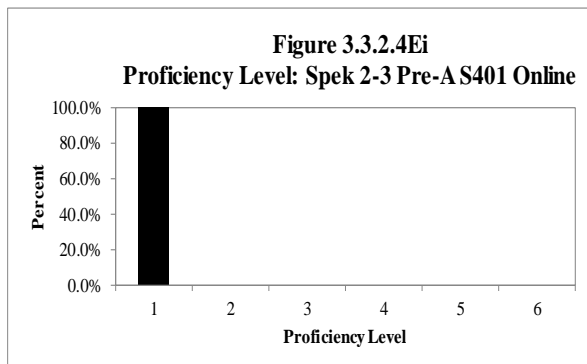
Table 3.3.2.4Ei

Proficiency Level Distribution: Spek 2-3 Pre-A S401 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	5,414	100.00%	8,618	100.00%	14,032	100.00%
Total	5,414	100.00%	8,618	100.00%	14,032	100.00%

Table 3.3.2.4Fi

Raw Score to Scale Score Conversion: Spek 2-3 Pre-A S401 Online

Raw Score	Scale Score	CSEM	Low Bound	High Bound
0	118^	26.62	100.00^	100.00^
1	118^	26.62	100.00^	100.00^
2	118^	26.62	100.00^	100.62
3	118^	26.62	100.00^	113.62
4	118^	26.62	100.00^	126.62
5	118*	26.62	100.00^	144.62
6	126*	31.00	116.00	178.00

^ Truncated

* Adjusted for end of scale effect

Table 3.3.2.4Gi
Equating Summary: Spek 2-3 Pre-A S401 Online
n/a

Figure 3.3.2.4Hi
Test Characteristic Curve: Spek 2-3 Pre-A S401 Online

Figure 3.3.2.4Ii
Test Information Function: Spek 2-3 Pre-A S401 Online
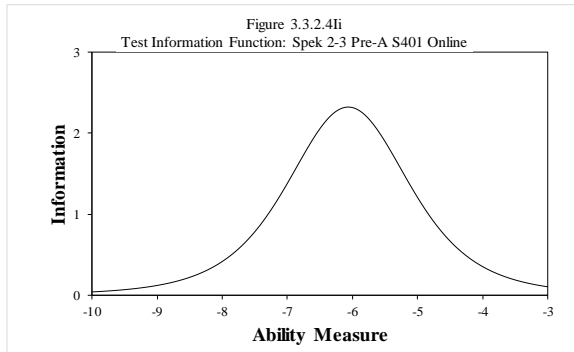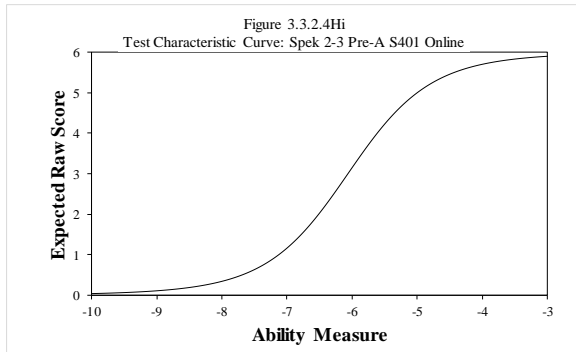
Table 3.3.2.4Ji
Reliability: Spek 2-3 Pre-A S401 Online

Reliability					
	No. of Students	**No. of Tasks**	**Cronbach's Alpha**		**SEM**
	14,032	3	.786		0.600
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	9,290	98	2	0
	2	9,956	98	2	0
	3	9,410	98	2	0

Table 3.3.2.4Ki
Conditional Standard Error of Measurement at Cut Scores: Spek 2-3 Pre-A S401 Online
n/a

Table 3.3.2.4Li
Accuracy and Consistency of Classification Indices: Spek 2-3 Pre-A S401 Online
n/a

3.3.2.4ii   *Speaking 2-3 A*

Table 3.3.2.4Aii
Complete Task Analysis and Summary: Spek 2-3 A S401 Online **n/a**

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

Figure 3.3.2.4Cii
Raw Scores: Spek 2-3 A S401 Online
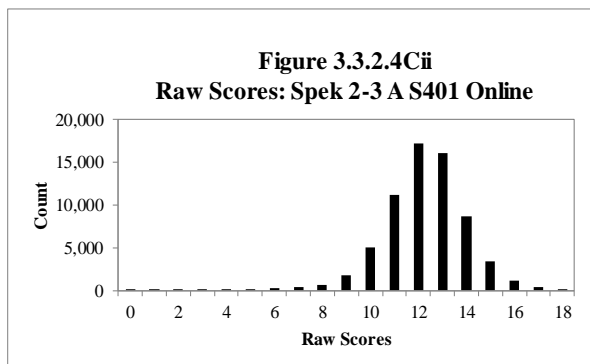
Table 3.3.2.4Cii
Raw Score Descriptive Statistics: Spek 2-3 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	34,384	0	18	11.96	1.90
3	32,253	0	18	12.49	1.67
Total	66,637	0	18	12.21	1.81

Figure 3.3.2.4Dii
Scale Scores: Spek 2-3 A S401 Online
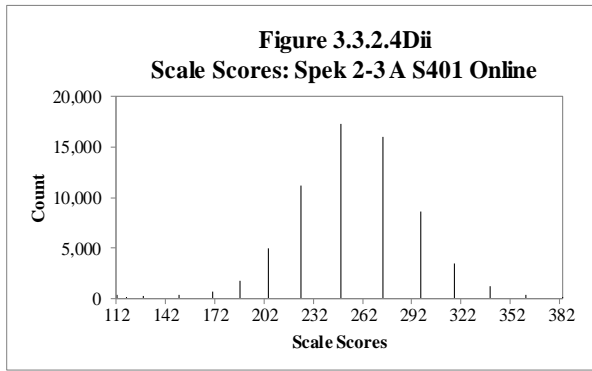
Table 3.3.2.4Dii

Scale Score Descriptive Statistics: Spek 2-3 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	34,384	112	383	249.37	40.12
3	32,253	118	383	260.92	37.04
Total	66,637	112	383	254.96	39.09

Figure 3.3.2.4Eii
Proficiency Level: Spek 2-3 A S401 Online
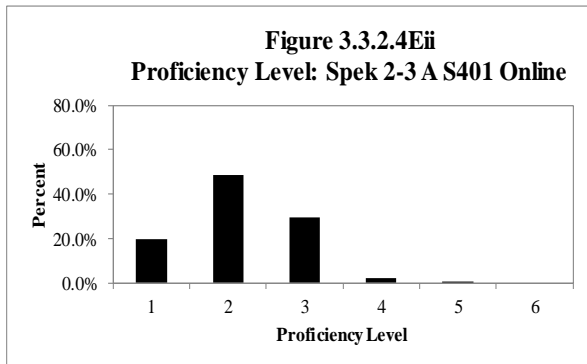
Table 3.3.2.4Eii

Proficiency Level Distribution: Spek 2-3 A S401 Online

Level	Grade 2		Grade 3		Total	
	Count	Percent	Count	Percent	Count	Percent
1	5,663	16.47%	7,542	23.38%	13,205	19.82%
2	15,345	44.63%	16,926	52.48%	32,271	48.43%
3	12,729	37.02%	6,822	21.15%	19,551	29.34%
4	618	1.80%	963	2.99%	1,581	2.37%
5	29	0.08%	0	0.00%	29	0.04%
6	0	0.00%	0	0.00%	0	0.00%
Total	34,384	100.00%	32,253	100.00%	66,637	100.00%

Table 3.3.2.4Gii
Equating Summary: Spek 2-3 A S401 Online
n/a

Figure 3.3.2.4Hii
Test Characteristic Curve: Spek 2-3 A S401 Online

Figure 3.3.2.4Iii
Test Information Function: Spek 2-3 A S401 Online
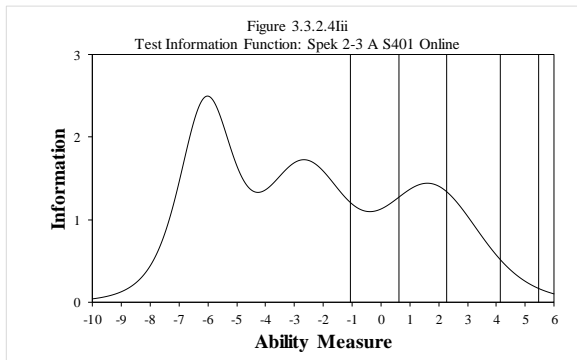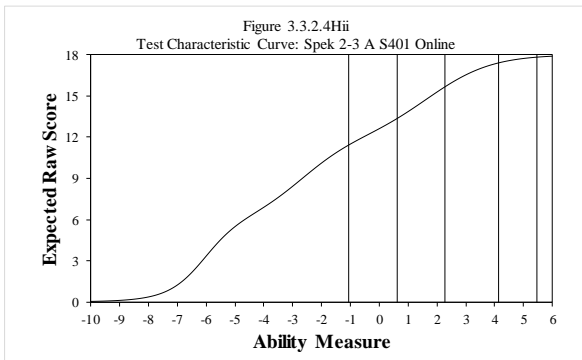
Table 3.3.2.4Jii

Reliability: Spek 2-3 A S401 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	66,637	6	.585		1.169
Interrater Reliability	Task	No. in Sample	% EX	% AD	% NA
	1	38,804	99	1	0
	2	38,805	83	16	0
	3	39,156	99	1	0
	4	39,152	88	11	0
	5	38,506	99	1	0
	6	38,506	77	22	1

Table 3.3.2.4Kii
Conditional Standard Error of Measurement at Cut Scores: Spek 2-3 A S401 Online
n/a

Table 3.3.2.4Lii
Accuracy and Consistency of Classification Indices: Spek 2-3 A S401 Online
n/a

3.3.2.4iii   Speaking 2-3 B/C

Table 3.3.2.4Aiii
Complete Task Analysis and Summary: Spek 2-3 B/C S401 Online **n/a**

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

Figure 3.3.2.4Ciii
Raw Scores: Spek 2-3 B/C S401 Online
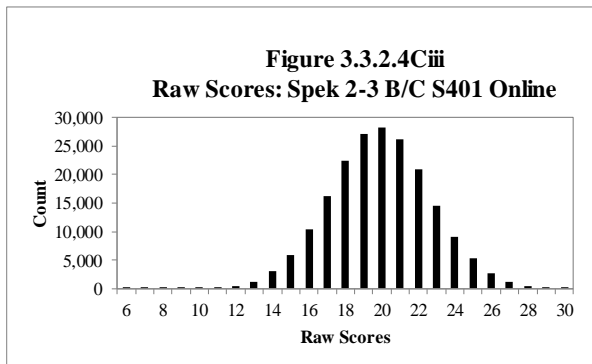
Table 3.3.2.4Ciii
Raw Score Descriptive Statistics: Spek 2-3 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	93,004	6	30	19.26	2.73
3	102,739	6	30	20.50	2.73
Total	195,743	6	30	19.91	2.80

Figure 3.3.2.4Diii
Scale Scores: Spek 2-3 B/C S401 Online
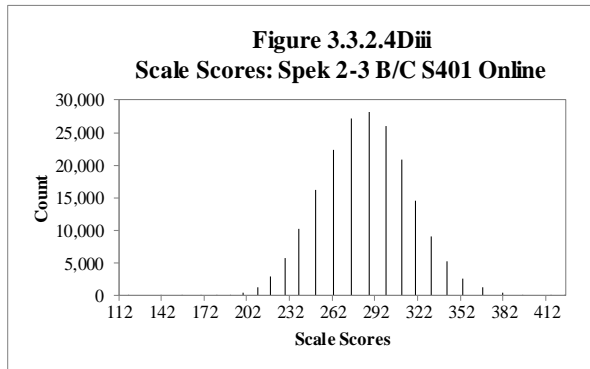
Table 3.3.2.4Diii

Scale Score Descriptive Statistics: Spek 2-3 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	93,004	112	415	277.46	31.63
3	102,739	118	425	291.88	31.49
Total	195,743	112	425	285.03	32.37

Figure 3.3.2.4Eiii
Proficiency Level: Spek 2-3 B/C S401 Online
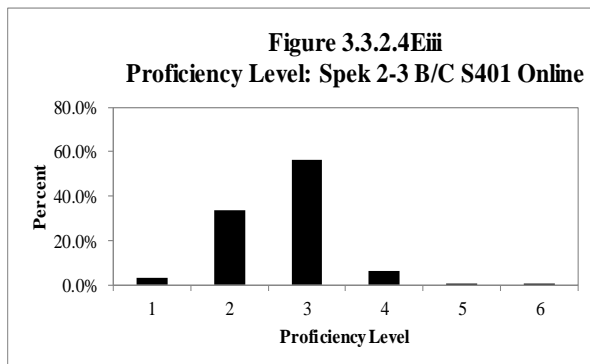
Table 3.3.2.4Eiii

Proficiency Level Distribution: Spek 2-3 B/C S401 Online

Level	Grade 2 Count	Grade 2 Percent	Grade 3 Count	Grade 3 Percent	Total Count	Total Percent
1	3,494	3.76%	3,367	3.28%	6,861	3.51%
2	33,072	35.56%	32,941	32.06%	66,013	33.72%
3	50,743	54.56%	59,221	57.64%	109,964	56.18%
4	5,525	5.94%	7,006	6.82%	12,531	6.40%
5	164	0.18%	168	0.16%	332	0.17%
6	6	0.01%	36	0.04%	42	0.02%
Total	93,004	100.00%	102,739	100.00%	195,743	100.00%

Table 3.3.2.4Giii
Equating Summary: Spek 2-3 B/C S401 Online
n/a

Figure 3.3.2.4Hiii
Test Characteristic Curve: Spek 2-3 B/C S401 Online

Figure 3.3.2.4Iiii
Test Information Function: Spek 2-3 B/C S401 Online
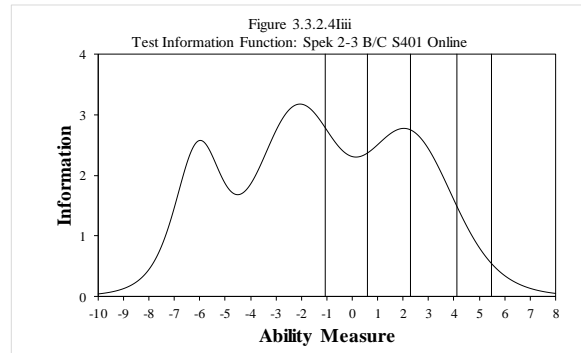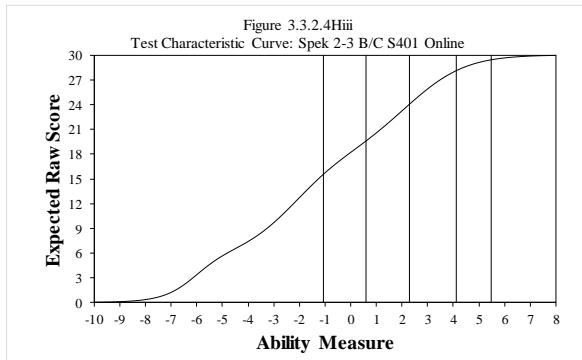
Table 3.3.2.4Jiii

Reliability: Spek 2-3 B/C S401 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	195,743	6	.688		1.561
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	102,167	74	25	1
	2	102,167	71	28	1
	3	105,128	78	21	1
	4	105,129	74	26	1
	5	105,894	78	22	1
	6	105,896	74	25	1

Table 3.3.2.4Kiii
Conditional Standard Error of Measurement at Cut Scores: Spek 2-3 B/C S401 Online
n/a

Table 3.3.2.4Liii
Accuracy and Consistency of Classification Indices: Spek 2-3 B/C S401 Online
n/a

3.3.2.4iv *Speaking 2-3 Across Tiers*

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

Table 3.3.2.4Biv
DIF Analysis and Summary: Spek 2-3 S401 Online
n/a

Figure 3.3.2.4Civ
Raw Scores: Spek 2-3 S401 Online
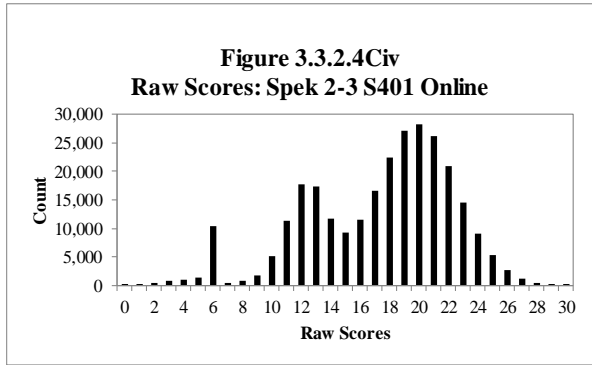
Table 3.3.2.4Civ
Raw Score Descriptive Statistics: Spek 2-3 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	132,802	0	30	16.80	4.68
3	143,610	0	30	17.79	5.18
Total	276,412	0	30	17.31	4.97

Figure 3.3.2.4Div
Scale Scores: Spek 2-3 S401 Online
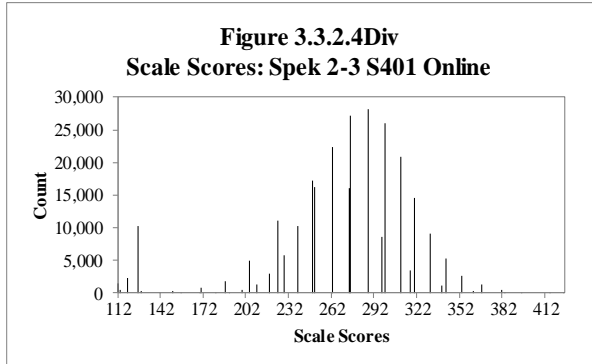
Table 3.3.2.4Div
Scale Score Descriptive Statistics: Spek 2-3 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	132,802	112	415	263.85	46.09
3	143,610	118	425	274.85	51.34
Total	276,412	112	425	269.56	49.19

Figure 3.3.2.4Eiv
Proficiency Level: Spek 2-3 S401 Online
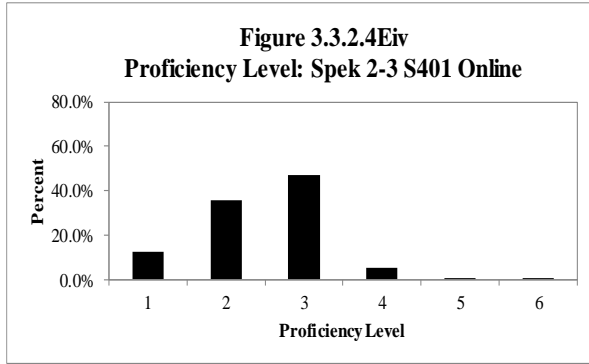
Table 3.3.2.4Eiv

Proficiency Level Distribution: Spek 2-3 S401 Online

Level	Grade 2		Grade 3		Total	
	Count	Percent	Count	Percent	Count	Percent
1	14,571	10.97%	19,527	13.60%	34,098	12.34%
2	48,417	36.46%	49,867	34.72%	98,284	35.56%
3	63,472	47.79%	66,043	45.99%	129,515	46.86%
4	6,143	4.63%	7,969	5.55%	14,112	5.11%
5	193	0.15%	168	0.12%	361	0.13%
6	6	0.00%	36	0.03%	42	0.02%
Total	132,802	100.00%	143,610	100.00%	276,412	100.00%

Table 3.3.2.4Fiv
Raw Score to Scale Score Conversion: Spek 2-3 S401 Online
n/a

Figure 3.3.2.4Hiv
Test Characteristic Curve: Spek 2-3 S401 Online

Figure 3.3.2.4Iiv
Test Information Function: Spek 2-3 S401 Online
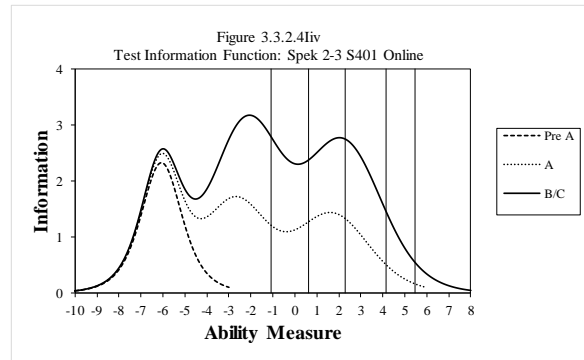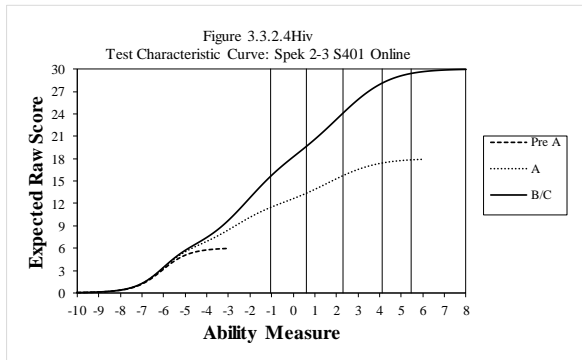
Table 3.3.2.4Jiv

Reliability: Spek 2-3 Weighted Reliability S401 Online

Tiers	No. of Students	Reliability	Weighted Reliability
Pre-A	14,032	0.786	
A	66,637	0.585	0.668
B/C	195,743	0.688	

Table 3.3.2.4Kiv

Conditional Standard Error of Measurement at Cut Scores: Spek 2-3 S401 Online

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	2	220	24.86	16.67
	3	234	26.62	17.55
2/3	2	273	26.91	19.30
	3	283	26.03	19.01
3/4	2	322	24.57	17.55
	3	332	25.45	17.55
4/5	2	374	35.39	21.64
	3	386	40.95	24.28
5/6	2	415	62.30	34.80
	3	425	72.83	40.07

Note: Tier Pre-A is not presented as it is not possible for Tier Pre-A students to receive a proficeny level higher than 2.

Table 3.3.2.4Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.617	0.497		0.203	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.865		0.675	
	2	0.806		0.415	
	3	0.554		0.568	
	4	-		0.061	
	5	-		-	
	6	-		-	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.960	0.013	0.027	0.932
	2/3	0.746	0.003	0.251	0.606
	3/4	0.952	0.048	0.000	0.906
	4/5	0.999	0.002	0.000	0.998
	5/6	1.000	0.000	0.000	1.000

Table 3.3.2.4Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.572	0.465		0.182	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.931		0.701	
	2	0.568		0.383	
	3	0.528		0.523	
	4	-		0.065	
	5	-		-	
	6	-		-	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.953	0.007	0.039	0.919
	2/3	0.654	0.016	0.330	0.576
	3/4	0.943	0.057	0.000	0.881
	4/5	0.999	0.001	0.000	0.998
	5/6	1.000	0.000	0.000	1.000

3.3.3 Grades: 4-5

3.3.3.1   Listening 4-5

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

Figure 3.3.3.1C
Raw Scores: List 4-5 S401 Online
n/a

Table 3.3.3.1C
Raw Score Descriptive Statistics: List 4-5 S401 Online
n/a

Figure 3.3.3.1D
Scale Scores: List 4-5 S401 Online

Figure 3.3.3.1E
Proficiency Level: List 4-5 S401 Online
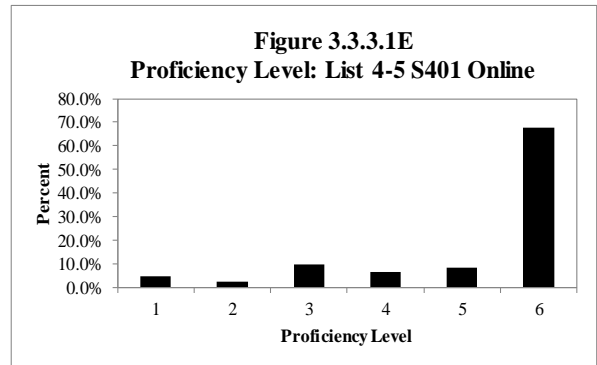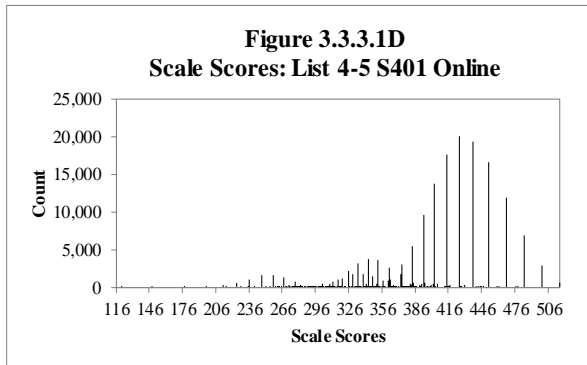
Table 3.3.3.1D

Scale Score Descriptive Statistics: List 4-5 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	102,781	116	516	400.80	53.57
5	74,998	120	516	405.35	60.33
Total	177,779	116	516	402.72	56.56

Table 3.3.3.1E

Proficiency Level Distribution: List 4-5 S401 Online

Level	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	4,070	3.96%	4,534	6.05%	8,604	4.84%
2	2,587	2.52%	2,114	2.82%	4,701	2.64%
3	7,858	7.65%	9,281	12.37%	17,139	9.64%
4	6,510	6.33%	5,124	6.83%	11,634	6.54%
5	10,085	9.81%	5,225	6.97%	15,310	8.61%
6	71,671	69.73%	48,720	64.96%	120,391	67.72%
Total	102,781	100.00%	74,998	100.00%	177,779	100.00%

Table 3.3.3.1F
Raw Score to Scale Score Conversion: List 4-5 S401 Online
n/a

Figure 3.3.3.1H
Test Characteristic Curve: List 4-5 S401 Online
n/a

Figure 3.3.3.1I
Test Information Function: List 4-5 S401 Online
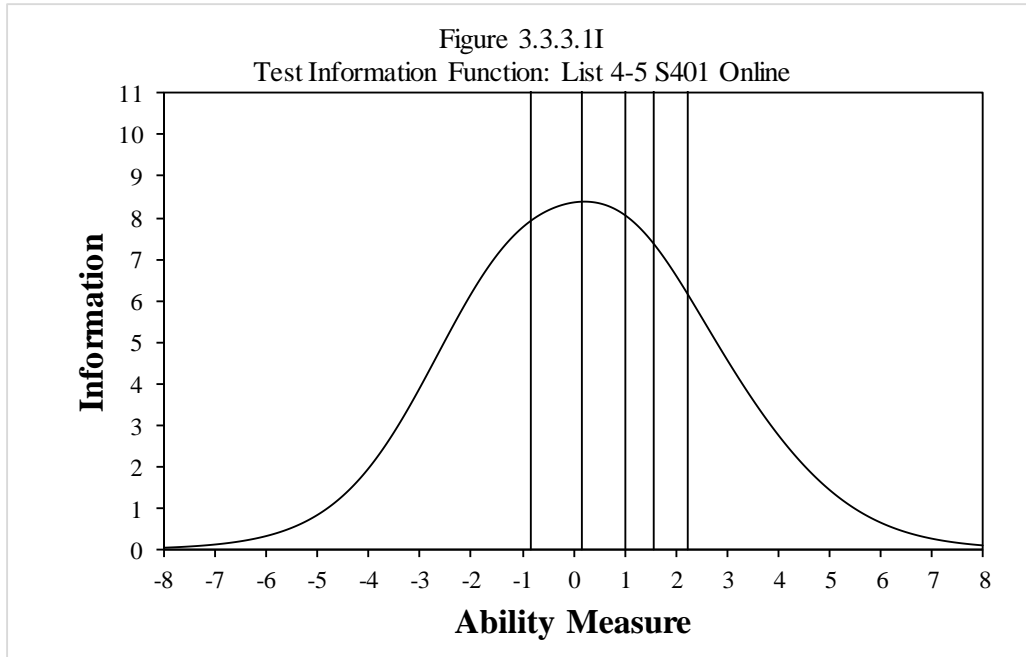
Table 3.3.3.1J

Reliability: List 4-5 S401 Online

No. of Students	No. of Items	Rasch Reliability Estimate
177,779	54	.85

Table 3.3.3.1K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: List 4-5 S401 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	4	275	9	17.35	18.88	18.71	0.51
	5	285	N/A	N/A	N/A	N/A	N/A
2/3	4	313	67	17.86	18.37	17.86	0.06
	5	323	N/A	N/A	N/A	N/A	N/A
3/4	4	343	1,684	18.37	19.39	18.37	0.04
	5	354	35	17.86	19.39	19.08	0.62
4/5	4	363	879	17.86	19.39	19.33	0.27
	5	375	6	19.39	20.92	19.90	0.79
5/6	4	388	N/A	N/A	N/A	N/A	N/A
	5	401	N/A	N/A	N/A	N/A	N/A

Table 3.3.3.1Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.780	0.724		0.451	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.907		0.781	
	2	0.331		0.230	
	3	0.540		0.393	
	4	0.314		0.226	
	5	0.347		0.249	
	6	0.932		0.911	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.983	0.003	0.014	0.978
	2/3	0.975	0.013	0.012	0.961
	3/4	0.946	0.019	0.035	0.926
	4/5	0.932	0.026	0.042	0.902
	5/6	0.906	0.047	0.048	0.863

Table 3.3.3.1Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.741	0.681		0.442	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.900		0.753	
	2	0.241		0.167	
	3	0.602		0.463	
	4	0.272		0.200	
	5	0.231		0.161	
	6	0.939		0.910	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.975	0.004	0.021	0.965
	2/3	0.961	0.023	0.016	0.940
	3/4	0.921	0.023	0.057	0.897
	4/5	0.922	0.026	0.053	0.886
	5/6	0.905	0.057	0.038	0.857

3.3.3.2   Reading 4-5

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

Figure 3.3.3.2C
Raw Scores: Read 4-5 S401 Online
n/a

Table 3.3.3.2C
Raw Score Descriptive Statistics: Read 4-5 S401 Online
n/a

Figure 3.3.3.2D
Scale Scores: Read 4-5 S401 Online

Figure 3.3.3.2E
Proficiency Level: Read 4-5 S401 Online
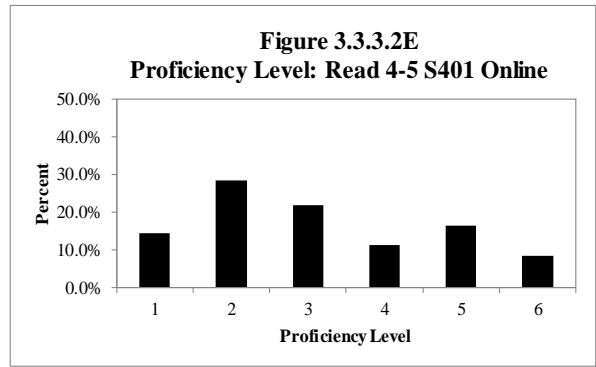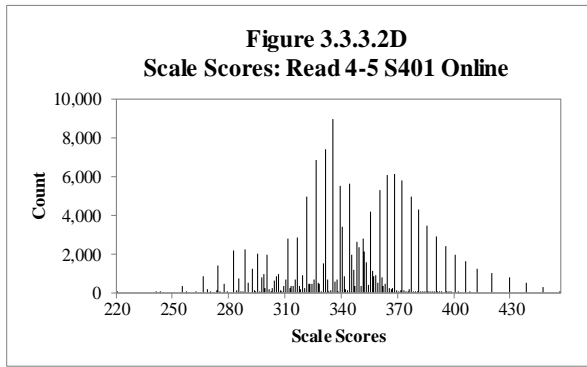
Table 3.3.3.2D

Scale Score Descriptive Statistics: Read 4-5 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	91,056	220	456	342.98	30.34
5	65,391	220	456	346.52	35.02
Total	156,447	220	456	344.46	32.43

Table 3.3.3.2E

Proficiency Level Distribution: Read 4-5 S401 Online

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	10,044	11.03%	12,235	18.71%	22,279	14.24%
2	23,320	25.61%	20,799	31.81%	44,119	28.20%
3	22,740	24.97%	11,189	17.11%	33,929	21.69%
4	10,713	11.77%	6,850	10.48%	17,563	11.23%
5	17,089	18.77%	8,353	12.77%	25,442	16.26%
6	7,150	7.85%	5,965	9.12%	13,115	8.38%
Total	91,056	100.00%	65,391	100.00%	156,447	100.00%

Table 3.3.3.2F
Raw Score to Scale Score Conversion: Read 4-5 S401 Online
n/a

Figure 3.3.3.2H
Test Characteristic Curve: Read 4-5 S401 Online
n/a

Figure 3.3.3.2I
Test Information Function: Read 4-5 S401 Online
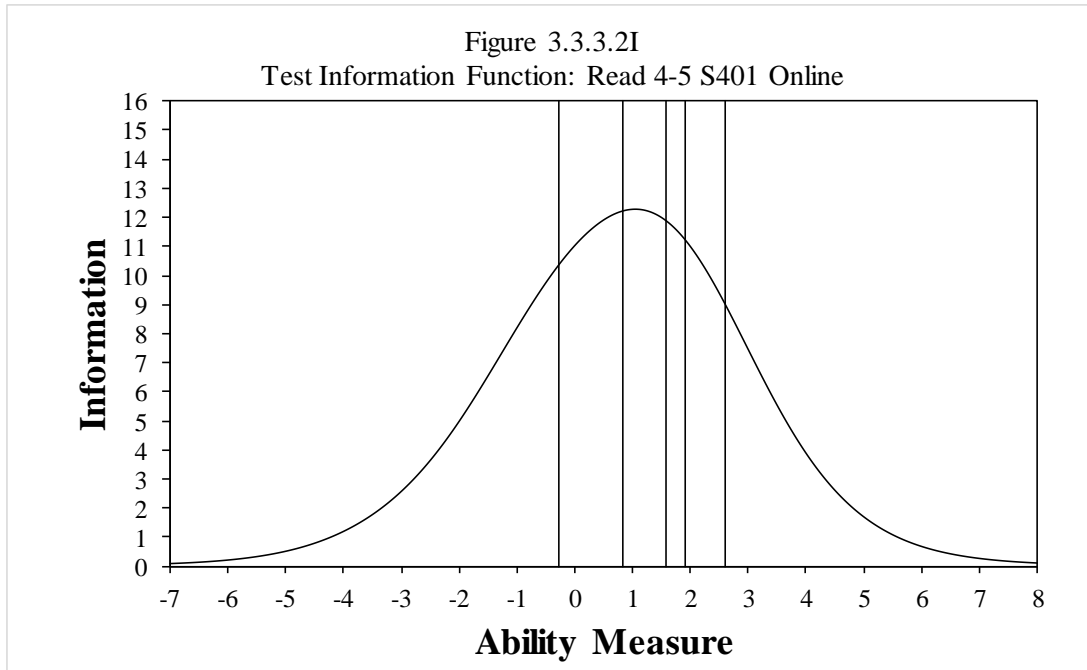
Table 3.3.3.2J

Reliability: Read 4-5 S401 Online

No. of Students	No. of Items	Rasch Reliability Estimate
156,447	66	.88

Table 3.3.3.2K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: Read 4-5 S401 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	4	307	111	11.22	13.27	11.82	0.48
	5	316	413	11.22	11.22	11.22	0.00
2/3	4	335	5,439	10.20	11.22	10.54	0.39
	5	345	304	10.20	10.71	10.56	0.23
3/4	4	354	284	10.20	10.71	10.30	0.20
	5	364	1,448	10.20	10.71	10.22	0.09
4/5	4	364	4,630	10.20	10.71	10.21	0.04
	5	373	99	10.20	10.71	10.56	0.23
5/6	4	382	17	10.71	11.22	10.92	0.26
	5	391	26	11.22	11.73	11.30	0.19

Table 3.3.3.2Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.614	0.504		0.386	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.804		0.666	
	2	0.686		0.578	
	3	0.574		0.465	
	4	0.328		0.250	
	5	0.629		0.502	
	6	0.714		0.543	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.949	0.019	0.032	0.927
	2/3	0.894	0.053	0.053	0.850
	3/4	0.887	0.062	0.051	0.841
	4/5	0.897	0.063	0.040	0.859
	5/6	0.953	0.025	0.021	0.930

Table 3.3.3.2Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.616	0.513		0.393	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.790		0.677	
	2	0.702		0.601	
	3	0.442		0.344	
	4	0.347		0.260	
	5	0.518		0.398	
	6	0.779		0.624	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.918	0.038	0.044	0.883
	2/3	0.889	0.050	0.061	0.843
	3/4	0.898	0.064	0.038	0.860
	4/5	0.913	0.048	0.039	0.880
	5/6	0.955	0.026	0.018	0.934

3.3.3.3   Writing 4-5

3.3.3.3i   *Writing 4-5 A*

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

Figure 3.3.3.3Ci
Raw Scores: Writ 4-5 A S401 Online
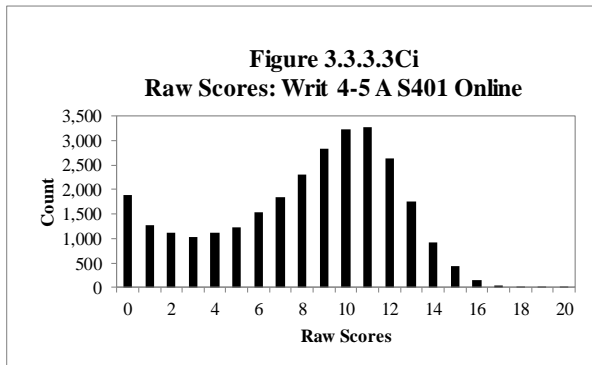
Table 3.3.3.3Ci

Raw Score Descriptive Statistics: Writ 4-5 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	14,443	0	19	7.47	4.10
5	14,205	0	20	8.53	4.11
Total	28,648	0	20	8.00	4.14

Figure 3.3.3.3Di
Scale Scores: Writ 4-5 A S401 Online
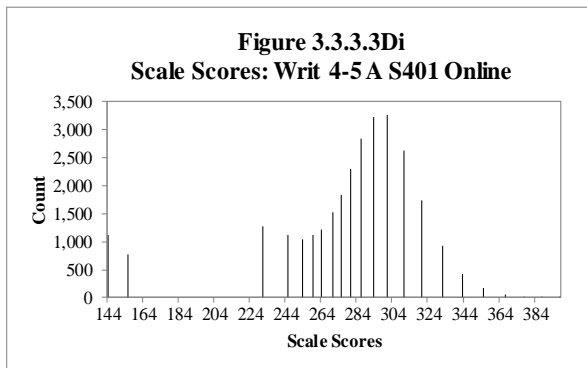
Table 3.3.3.3Di

Scale Score Descriptive Statistics: Writ 4-5 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	14,443	144	388	272.23	44.40
5	14,205	155	398	282.60	39.98
Total	28,648	144	398	277.37	42.58

Figure 3.3.3.3Ei
Proficiency Level: Writ 4-5 A S401 Online
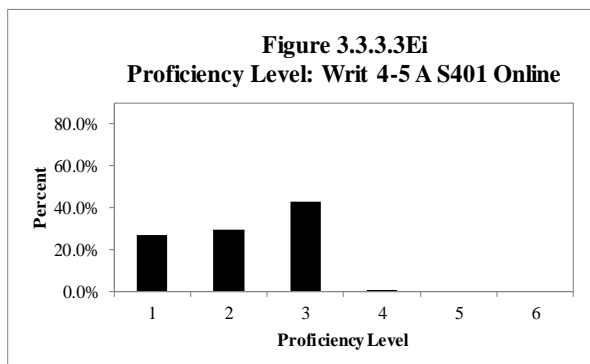
Table 3.3.3.3Ei

Proficiency Level Distribution: Writ 4-5 A S401 Online

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	4,389	30.39%	3,287	23.14%	7,676	26.79%
2	4,624	32.02%	3,885	27.35%	8,509	29.70%
3	5,359	37.10%	6,970	49.07%	12,329	43.04%
4	71	0.49%	63	0.44%	134	0.47%
5	0	0.00%	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%	0	0.00%
Total	14,443	100.00%	14,205	100.00%	28,648	100.00%

Figure 3.3.3.3Hi
Test Characteristic Curve: Writ 4-5 A S401 Online

Figure 3.3.3.3Ii
Test Information Function: Writ 4-5 A S401 Online
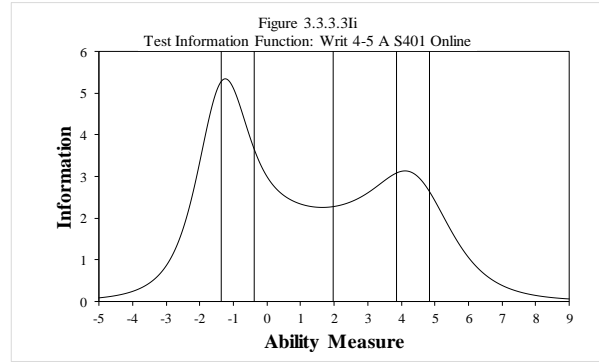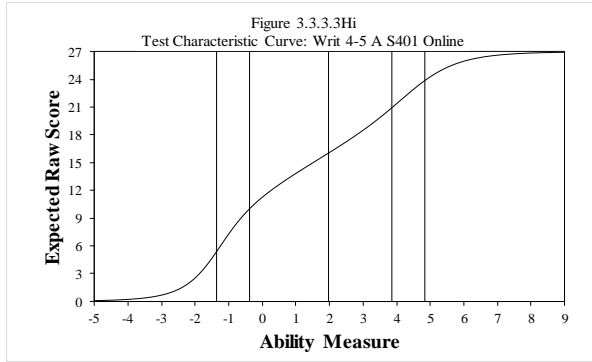
Table 3.3.3.3Ji

Reliability: Writ 4-5 A S401 Online

Reliability	No. of Students	No. of Tasks	Response Modes		Cronbach's Alpha	SEM
			Hand-written (HW)	Keyboarded (KB)		
	28,648	3			.881	1.429
Interrater Reliability	Task	Mode of Response	No. in Sample	% AG	% AD	% NA
	1	HW	9,116	98	2	0
		KB	14,122	95	5	0
	2	HW	7,186	97	3	0
		KB	13,974	96	4	0
	3	HW	8,108	98	2	0
		KB	14,000	96	4	0

Table 3.3.3.3Ki
Conditional Standard Error of Measurement at Cut Scores: Writ 4-5 A S401 Online
n/a

Table 3.3.3.3Li
Accuracy and Consistency of Classification Indices: Writ 4-5 A S401 Online
n/a

3.3.3.3ii   Writing 4-5 B/C

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

Figure 3.3.3.3Cii
Raw Scores: Writ 4-5 B/C S401 Online

Table 3.3.3.3Cii

Raw Score Descriptive Statistics: Writ 4-5 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	58,798	0	47	24.15	5.72
5	39,842	0	49	26.79	5.19
Total	98,640	0	49	25.22	5.66

Figure 3.3.3.3Dii
Scale Scores: Writ 4-5 B/C S401 Online
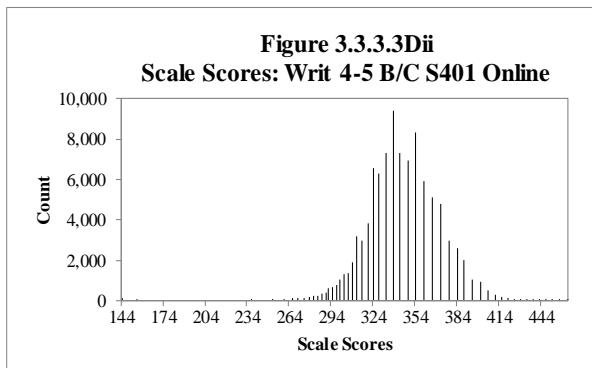
Table 3.3.3.3Dii

Scale Score Descriptive Statistics: Writ 4-5 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	58,798	144	452	336.84	27.03
5	39,842	155	463	349.67	26.19
Total	98,640	144	463	342.02	27.43

Figure 3.3.3.3Eii
Proficiency Level: Writ 4-5 B/C S401 Online
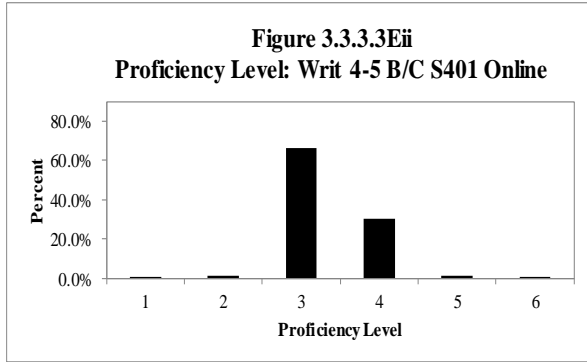
Table 3.3.3.3Eii

Proficiency Level Distribution: Writ 4-5 B/C S401 Online

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	344	0.59%	80	0.20%	424	0.43%
2	1,137	1.93%	431	1.08%	1,568	1.59%
3	40,885	69.53%	24,675	61.93%	65,560	66.46%
4	15,721	26.74%	14,157	35.53%	29,878	30.29%
5	676	1.15%	463	1.16%	1,139	1.15%
6	35	0.06%	36	0.09%	71	0.07%
Total	58,798	100.00%	39,842	100.00%	98,640	100.00%

Figure 3.3.3.3Hii
Test Characteristic Curve: Writ 4-5 B/C S401 Online

Figure 3.3.3.3Iii
Test Information Function: Writ 4-5 B/C S401 Online
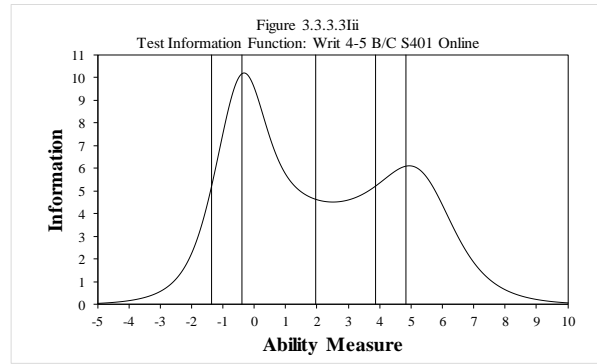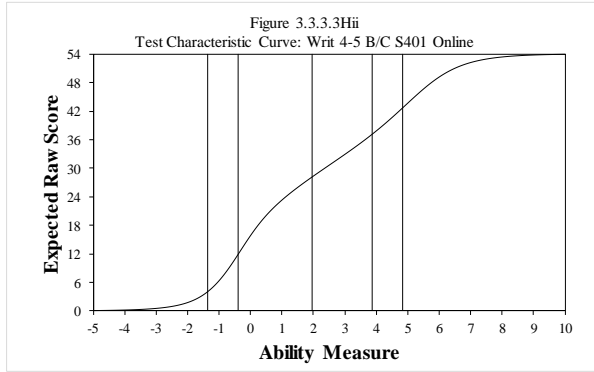
Table 3.3.3.3Jii

Reliability: Writ 4-5 B/C S401 Online

Reliability	No. of Students	No. of Tasks	Response Modes		Cronbach's Alpha	SEM
			Hand-written (HW)	Keyboarded (KB)		
	98,640	3			.901	1.778
Interrater Reliability	**Task**	**Mode of Response**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	HW	17,654	94	6	0
		KB	52,566	94	6	0
	2	HW	17,510	95	5	0
		KB	52,858	95	5	0
	3	HW	18,104	95	5	0
		KB	54,126	96	4	0

Table 3.3.3.3Kii
Conditional Standard Error of Measurement at Cut Scores: Writ 4-5 B/C S401 Online
n/a

Table 3.3.3.3Lii
Accuracy and Consistency of Classification Indices: Writ 4-5 B/C S401 Online
n/a

3.3.3.3iii   Writing 4-5 Across Tiers

Table 3.3.3.3Aiii
Complete Task Analysis and Summary: Writ 4-5 S401 Online
n/a

Table 3.3.3.3Biii
DIF Analysis and Summary: Writ 4-5 S401 Online
n/a

Figure 3.3.3.3Ciii
Raw Scores: Writ 4-5 S401 Online
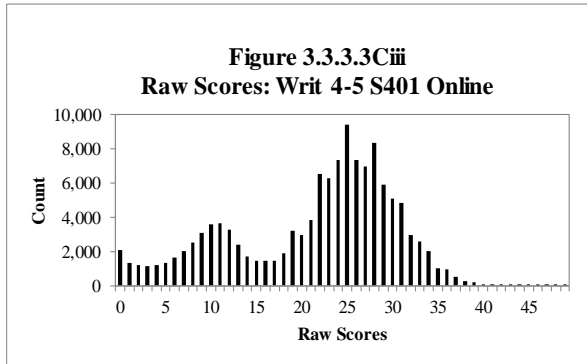
Table 3.3.3.3Ciii
Raw Score Descriptive Statistics: Writ 4-5 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	73,241	0	47	20.86	8.58
5	54,047	0	49	22.00	9.43
Total	127,288	0	49	21.34	8.97

Figure 3.3.3.3Diii
Scale Scores: Writ 4-5 S401 Online
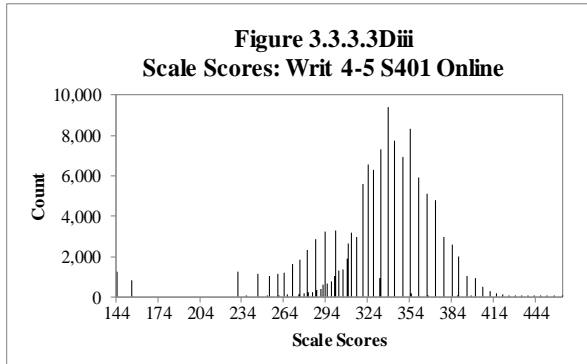
Table 3.3.3.3Diii
Scale Score Descriptive Statistics: Writ 4-5 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	73,241	144	452	324.10	40.45
5	54,047	155	463	332.05	42.39
Total	127,288	144	463	327.47	41.47

Figure 3.3.3.3Eiii
Proficiency Level: Writ 4-5 S401 Online
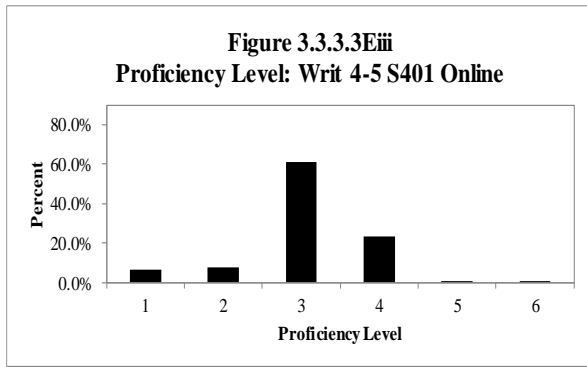
Table 3.3.3.3Eiii

Proficiency Level Distribution: Writ 4-5 S401 Online

Level	Grade 4		Grade 5		Total	
	Count	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	4,733	6.46%	3,367	6.23%	8,100	6.36%
2	5,761	7.87%	4,316	7.99%	10,077	7.92%
3	46,244	63.14%	31,645	58.55%	77,889	61.19%
4	15,792	21.56%	14,220	26.31%	30,012	23.58%
5	676	0.92%	463	0.86%	1,139	0.89%
6	35	0.05%	36	0.07%	71	0.06%
Total	73,241	100.00%	54,047	100.00%	127,288	100.00%

Table 3.3.3.3Fiii
Raw Score to Scale Score Conversion: Writ 4-5 S401 Online
n/a

Table 3.3.3.3Giii
Equating Summary: Writ 4-5 S401 Online
n/a

Figure 3.3.3.3Hiii
Test Characteristic Curve: Writ 4-5 S401 Online

Figure 3.3.3.3Iiii
Test Information Function: Writ 4-5 S401 Online
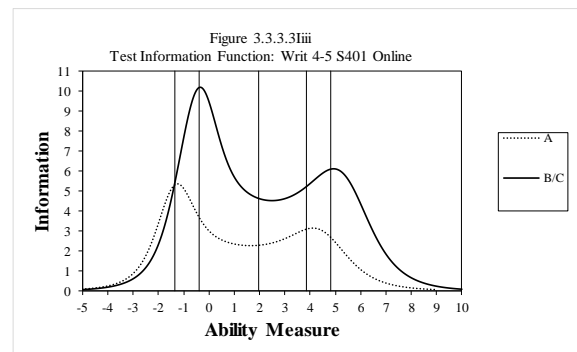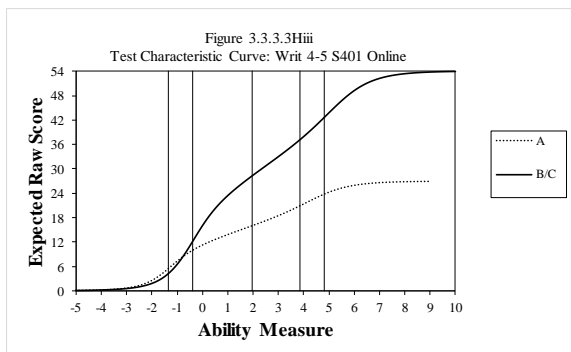
Table 3.3.3.3Jiii

Reliability: Writ 4-5 Weighted Reliability S401 Online

Tiers	No. of Students	Reliability	Weighted Reliability
A	28,648	0.881	0.897
B/C	98,640	0.901	

Table 3.3.3.3Kiii

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

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 3.3.3.3Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.714	0.635		0.334	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.811		0.722	
	2	0.658		0.516	
	3	0.710		0.740	
	4	-		0.372	
	5	-		-	
	6	-		1.000	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.979	0.013	0.008	0.969
	2/3	0.959	0.015	0.026	0.944
	3/4	0.775	0.225	0.000	0.721
	4/5	0.990	0.010	0.000	0.991
	5/6	1.000	0.000	0.000	1.000

Table 3.3.3.3Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.748	0.669		0.444	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.823		0.726	
	2	0.604		0.462	
	3	0.861		0.772	
	4	0.604		0.539	
	5	-		0.056	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.979	0.011	0.010	0.969
	2/3	0.954	0.021	0.025	0.936
	3/4	0.824	0.046	0.129	0.767
	4/5	0.991	0.009	0.000	0.991
	5/6	0.999	0.001	0.000	1.000

3.3.3.4   Speaking 4-5

3.3.3.4i   Speaking 4-5 Pre-A

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

Figure 3.3.3.4Ci
Raw Scores: Spek 4-5 Pre-A S401 Online
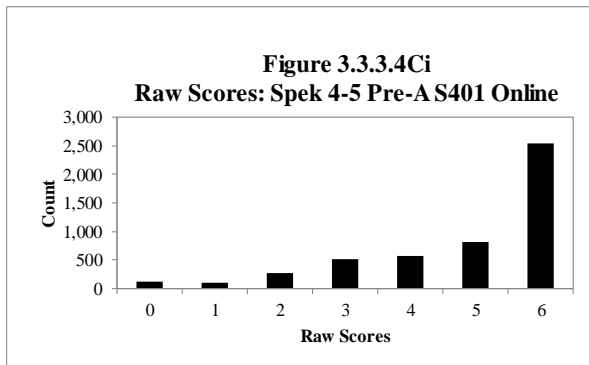
Table 3.3.3.4Ci
Raw Score Descriptive Statistics: Spek 4-5 Pre-A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	1,859	0	6	4.69	1.60
5	3,065	0	6	4.91	1.51
Total	4,924	0	6	4.83	1.55

Figure 3.3.3.4Di
Scale Scores: Spek 4-5 Pre-A S401 Online
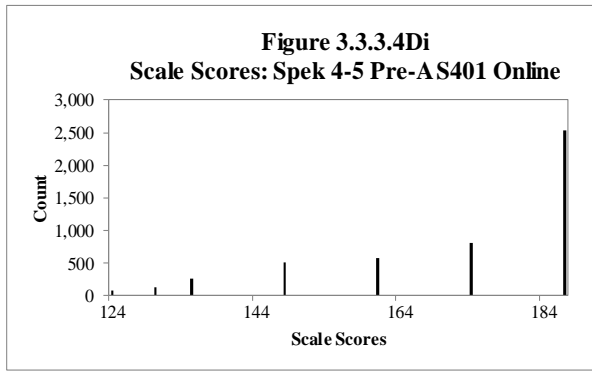
Table 3.3.3.4Di

Scale Score Descriptive Statistics: Spek 4-5 Pre-A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	1,859	124	187	170.42	19.79
5	3,065	130	187	173.49	17.95
Total	4,924	124	187	172.33	18.72

Figure 3.3.3.4Ei
Proficiency Level: Spek 4-5 Pre-A S401 Online
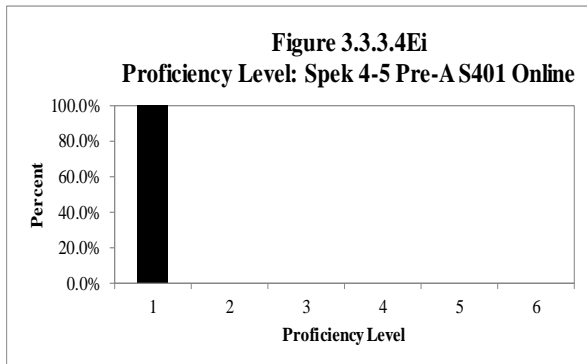
Table 3.3.3.4Ei

Proficiency Level Distribution: Spek 4-5 Pre-A S401 Online

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	1,859	100.00%	3,065	100.00%	4,924	100.00%
Total	1,859	100.00%	3,065	100.00%	4,924	100.00%

Table 3.3.3.4Gi
Equating Summary: Spek 4-5 Pre-A S401 Online
n/a

Figure 3.3.3.4Hi
Test Characteristic Curve: Spek 4-5 Pre-A S401 Online

Figure 3.3.3.4Ii
Test Information Function: Spek 4-5 Pre-A S401 Online
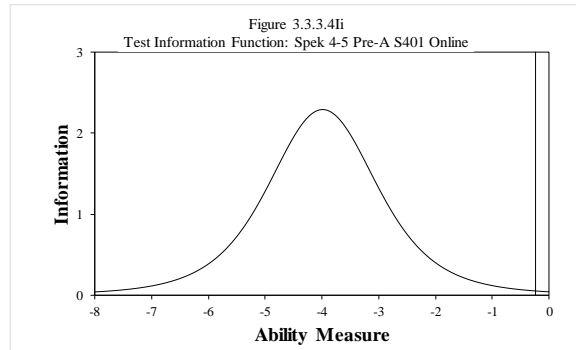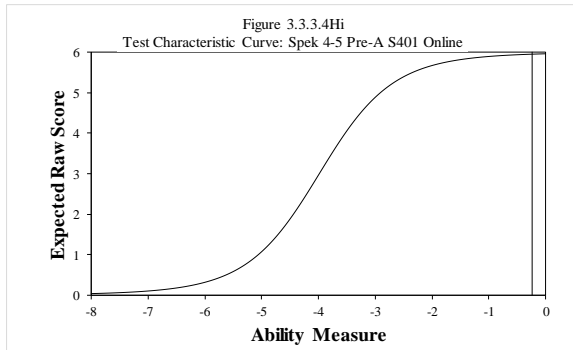
Table 3.3.3.4Ji

Reliability: Spek 4-5 Pre-A S401 Online

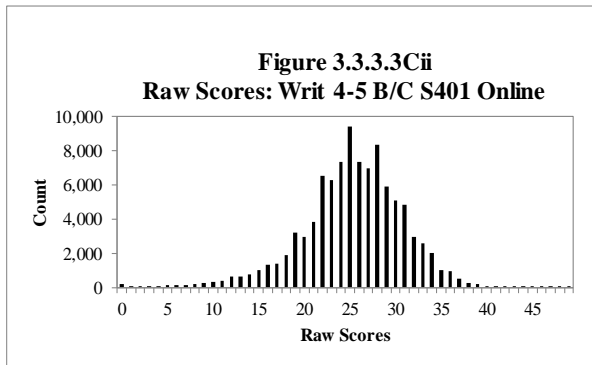| Reliability					
	No. of Students	**No. of Tasks**	**Cronbach's Alpha**		**SEM**
	4,924	3	.745		0.782
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	3,608	96	4	0
	2	4,116	96	4	0
	3	3,870	95	5	0

Table 3.3.3.4Ki
Conditional Standard Error of Measurement at Cut Scores: Spek 4-5 Pre-A S401 Online
n/a

Table 3.3.3.4Li
Accuracy and Consistency of Classification Indices: Spek 4-5 Pre-A S401 Online
n/a

3.3.3.4ii   *Speaking 4-5 A*

Table 3.3.3.4Aii
Complete Task Analysis and Summary: Spek 4-5 A S401 Online **n/a**

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

Figure 3.3.3.4Cii
Raw Scores: Spek 4-5 A S401 Online
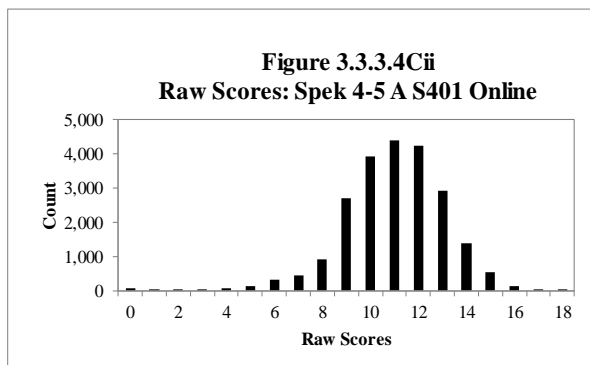
Table 3.3.3.4Cii
Raw Score Descriptive Statistics: Spek 4-5 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	12,552	0	18	10.79	2.23
5	9,791	0	18	11.13	2.08
Total	22,343	0	18	10.94	2.17

Figure 3.3.3.4Dii
Scale Scores: Spek 4-5 A S401 Online

Table 3.3.3.4Dii

Scale Score Descriptive Statistics: Spek 4-5 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	12,552	124	423	267.82	43.33
5	9,791	130	423	274.63	42.27
Total	22,343	124	423	270.80	43.00

Figure 3.3.3.4Eii
Proficiency Level: Spek 4-5 A S401 Online
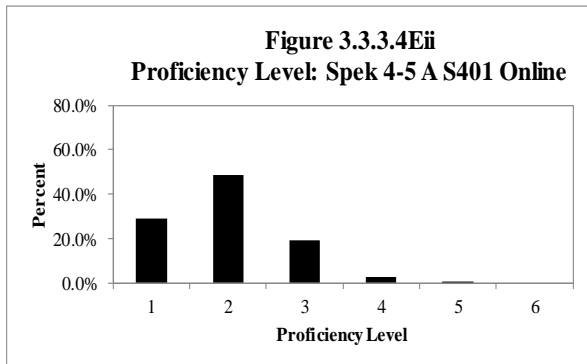
Table 3.3.3.4Eii

Proficiency Level Distribution: Spek 4-5 A S401 Online

Level	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	2,960	23.58%	3,510	35.85%	6,470	28.96%
2	7,029	56.00%	3,839	39.21%	10,868	48.64%
3	2,235	17.81%	2,080	21.24%	4,315	19.31%
4	318	2.53%	359	3.67%	677	3.03%
5	10	0.08%	3	0.03%	13	0.06%
6	0	0.00%	0	0.00%	0	0.00%
Total	12,552	100.00%	9,791	100.00%	22,343	100.00%

Table 3.3.3.4Gii
Equating Summary: Spek 4-5 A S401 Online
n/a

Figure 3.3.3.4Hii
Test Characteristic Curve: Spek 4-5 A S401 Online

Figure 3.3.3.4Iii
Test Information Function: Spek 4-5 A S401 Online
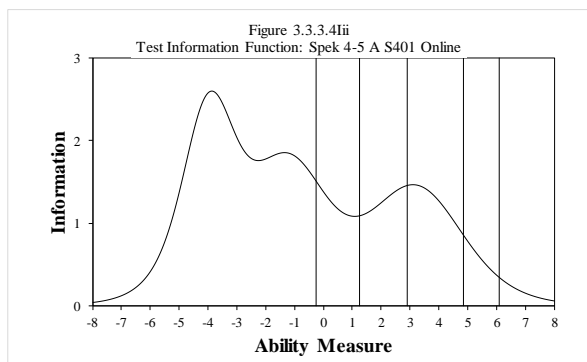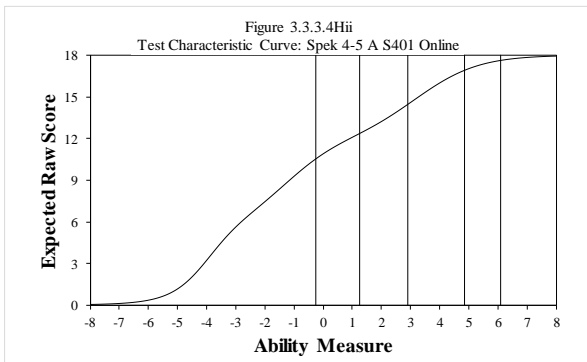
Table 3.3.3.4Jii

Reliability: Spek 4-5 A S401 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	22,343	6	.700		1.187
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	13,062	98	2	0
	2	13,062	84	16	0
	3	13,600	98	2	0
	4	13,600	84	16	0
	5	13,444	97	3	0
	6	13,444	81	18	0

Table 3.3.3.4Kii
Conditional Standard Error of Measurement at Cut Scores: Spek 4-5 A S401 Online
n/a

Table 3.3.3.4Lii
Accuracy and Consistency of Classification Indices: Spek 4-5 A S401 Online
n/a

3.3.3.4iii   *Speaking 4-5 B/C*

Table 3.3.3.4Aiii
Complete Task Analysis and Summary: Spek 4-5 B/C S401 Online **n/a**

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

Figure 3.3.3.4Ciii
Raw Scores: Spek 4-5 B/C S401 Online
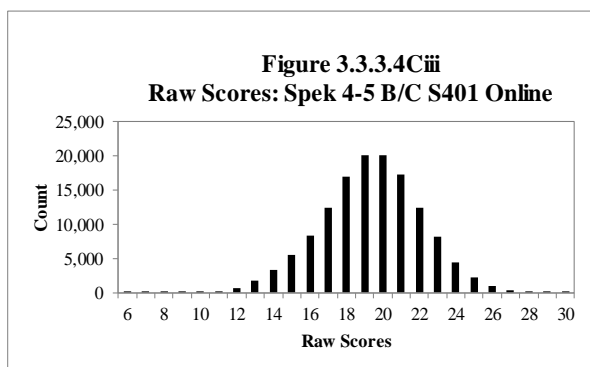
Table 3.3.3.4Ciii
Raw Score Descriptive Statistics: Spek 4-5 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	79,554	6	30	19.16	2.74
5	55,891	6	30	19.61	2.79
Total	135,445	6	30	19.35	2.77

Figure 3.3.3.4Diii
Scale Scores: Spek 4-5 B/C S401 Online
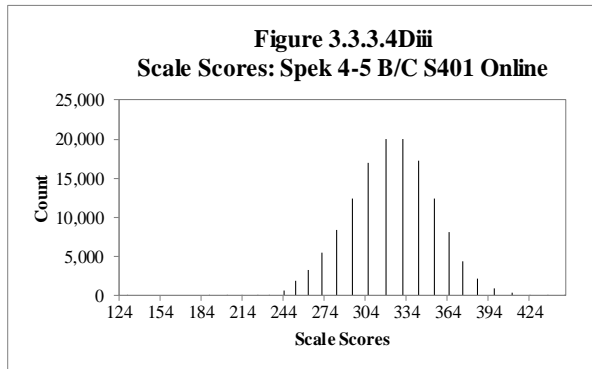
Table 3.3.3.4Diii

Scale Score Descriptive Statistics: Spek 4-5 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	79,554	124	450	320.62	31.68
5	55,891	130	450	325.75	32.28
Total	135,445	124	450	322.74	32.03

Figure 3.3.3.4Eiii
Proficiency Level: Spek 4-5 B/C S401 Online
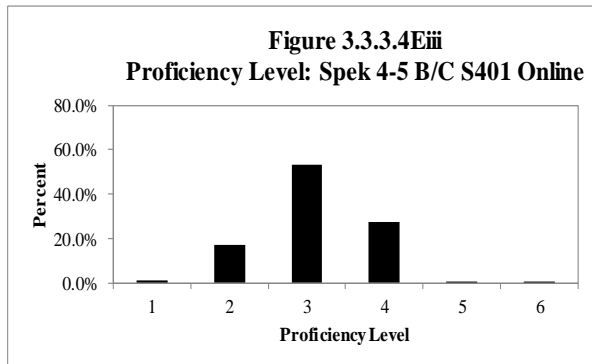
Table 3.3.3.4Eiii

Proficiency Level Distribution: Spek 4-5 B/C S401 Online

Level	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	640	0.80%	989	1.77%	1,629	1.20%
2	12,268	15.42%	10,832	19.38%	23,100	17.05%
3	41,699	52.42%	30,465	54.51%	72,164	53.28%
4	24,245	30.48%	13,260	23.72%	37,505	27.69%
5	679	0.85%	337	0.60%	1,016	0.75%
6	23	0.03%	8	0.01%	31	0.02%
Total	79,554	100.00%	55,891	100.00%	135,445	100.00%

Table 3.3.3.4Giii
Equating Summary: Spek 4-5 B/C S401 Online
n/a

Figure 3.3.3.4Hiii
Test Characteristic Curve: Spek 4-5 B/C S401 Online

Figure 3.3.3.4Iiii
Test Information Function: Spek 4-5 B/C S401 Online
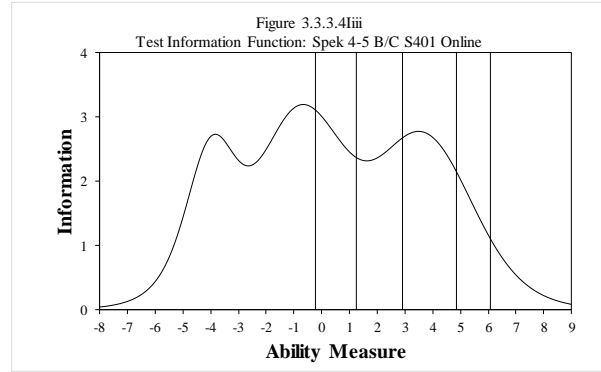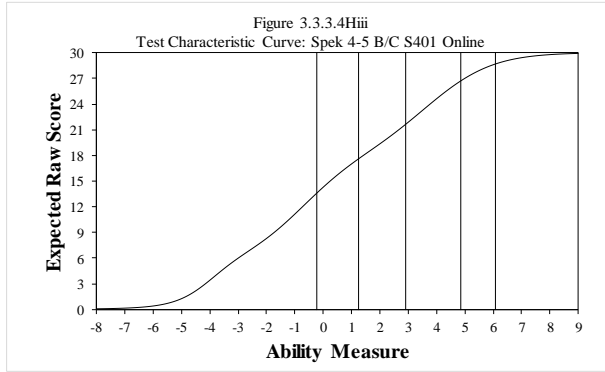
Table 3.3.3.4Jiii

Reliability: Spek 4-5 B/C S401 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	135,445	6	.726		1.448
Interrater Reliability	Task	No. in Sample	% EX	% AD	% NA
	1	74,322	82	17	0
	2	74,322	77	23	1
	3	74,390	72	27	1
	4	74,393	74	25	1
	5	73,178	72	27	1
	6	73,180	76	24	1

Table 3.3.3.4Kiii
Conditional Standard Error of Measurement at Cut Scores: Spek 4-5 B/C S401 Online
n/a

Table 3.3.3.4Liii
Accuracy and Consistency of Classification Indices: Spek 4-5 B/C S401 Online
n/a

3.3.3.4iv   *Speaking 4-5 Across Tiers*

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

Table 3.3.3.4Biv
DIF Analysis and Summary: Spek 4-5 S401 Online
n/a

Figure 3.3.3.4Civ
Raw Scores: Spek 4-5 S401 Online
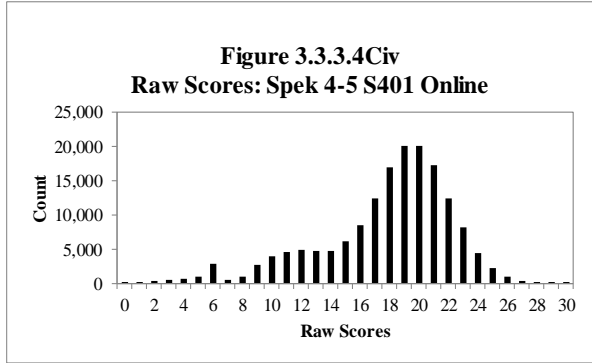
Table 3.3.3.4Civ
Raw Score Descriptive Statistics: Spek 4-5 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	93,965	0	30	17.76	4.31
5	68,747	0	30	17.75	4.84
Total	162,712	0	30	17.75	4.54

Figure 3.3.3.4Div
Scale Scores: Spek 4-5 S401 Online
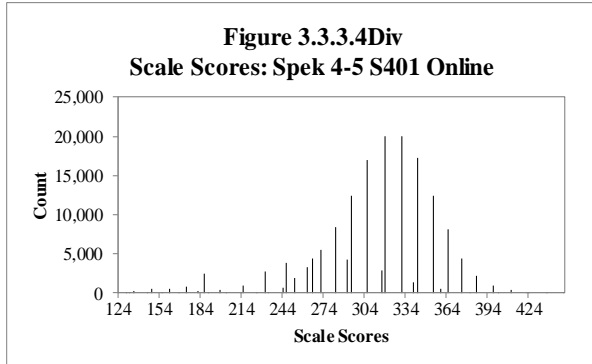
Table 3.3.3.4Div
Scale Score Descriptive Statistics: Spek 4-5 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	93,965	124	450	310.59	42.74
5	68,747	130	450	311.68	48.21
Total	162,712	124	450	311.05	45.13

Figure 3.3.3.4Eiv
Proficiency Level: Spek 4-5 S401 Online
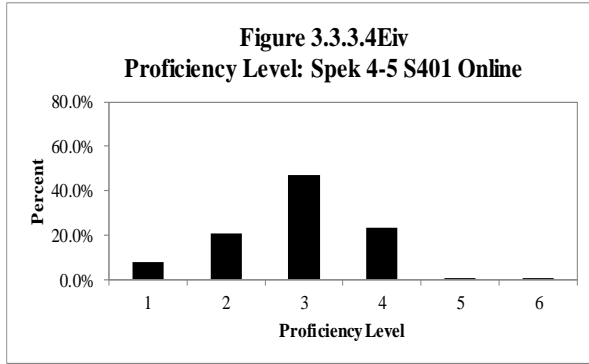
Table 3.3.3.4Eiv

Proficiency Level Distribution: Spek 4-5 S401 Online

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	5,459	5.81%	7,564	11.00%	13,023	8.00%
2	19,297	20.54%	14,671	21.34%	33,968	20.88%
3	43,934	46.76%	32,545	47.34%	76,479	47.00%
4	24,563	26.14%	13,619	19.81%	38,182	23.47%
5	689	0.73%	340	0.49%	1,029	0.63%
6	23	0.02%	8	0.01%	31	0.02%
Total	93,965	100.00%	68,747	100.00%	162,712	100.00%

Table 3.3.3.4Fiv
Raw Score to Scale Score Conversion: Spek 4-5 S401 Online
n/a

Figure 3.3.3.4Hiv
Test Characteristic Curve: Spek 4-5 S401 Online

Figure 3.3.3.4Iiv
Test Information Function: Spek 4-5 S401 Online
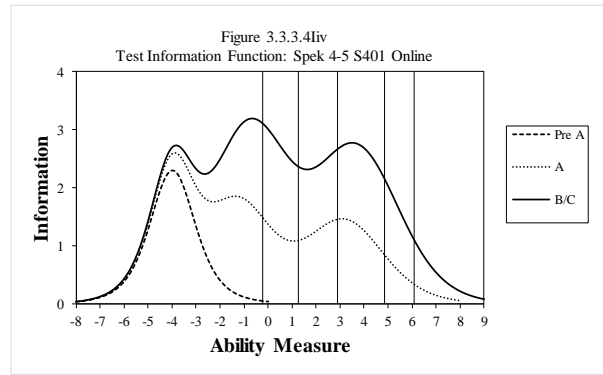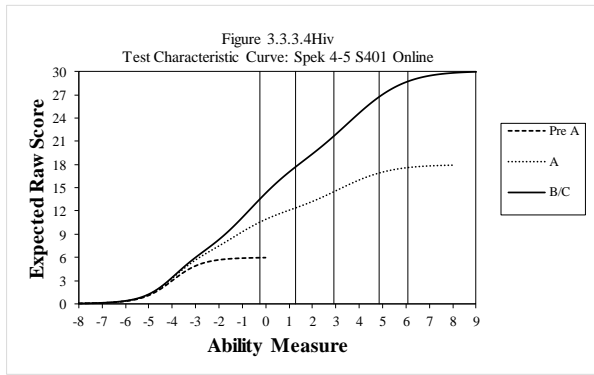
Table 3.3.3.4Jiv

Reliability: Spek 4-5 Weighted Reliability S401 Online

Tiers	No. of Students	Reliability	Weighted Reliability
Pre-A	4,924	0.745	
A	22,343	0.700	0.723
B/C	135,445	0.726	

Table 3.3.3.4Kiv

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

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

Note: Tier Pre-A is not presented as it is not possible for Tier Pre-A students to receive a proficeny level higher than 2.

Table 3.3.3.4Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.582	0.503		0.247	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.733		0.561	
	2	0.679		0.493	
	3	0.552		0.542	
	4	-		0.416	
	5	-		0.015	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.968	0.015	0.017	0.950
	2/3	0.881	0.035	0.084	0.829
	3/4	0.731	0.269	0.000	0.697
	4/5	0.992	0.008	0.000	0.991
	5/6	1.000	0.000	0.000	1.000

Table 3.3.3.4Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.605	0.484		0.236	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.816		0.678	
	2	0.586		0.373	
	3	0.579		0.560	
	4	-		0.292	
	5	-		0.008	
	6	-		-	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.955	0.019	0.026	0.932
	2/3	0.848	0.036	0.117	0.775
	3/4	0.797	0.203	0.000	0.717
	4/5	0.995	0.005	0.000	0.994
	5/6	1.000	0.000	0.000	1.000

3.3.4 Grades: 6-8

3.3.4.1   Listening 6-8

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

Figure 3.3.4.1C
Raw Scores: List 6-8 S401 Online
n/a

Table 3.3.4.1C
Raw Score Descriptive Statistics: List 6-8 S401 Online
n/a

Figure 3.3.4.1D
Scale Scores: List 6-8 S401 Online

Figure 3.3.4.1E
Proficiency Level: List 6-8 S401 Online
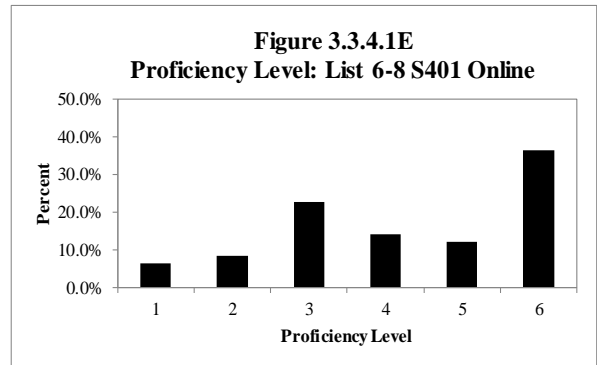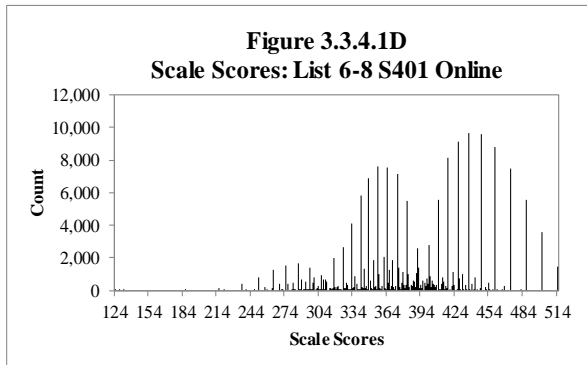
Table 3.3.4.1D

Scale Score Descriptive Statistics: List 6-8 S401 Online
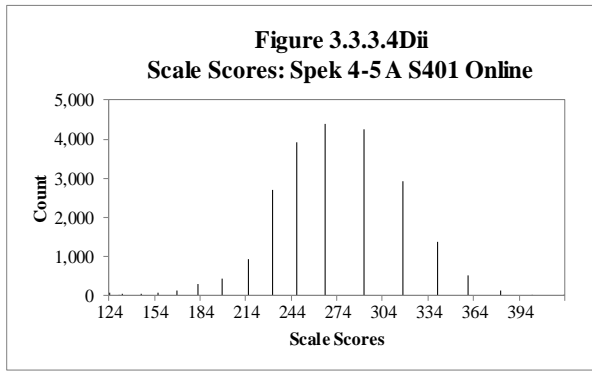
Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	59,759	124	515	382.49	50.81
7	60,913	128	515	393.56	56.64
8	60,312	132	515	400.37	61.06
Total	180,984	124	515	392.17	56.82

Table 3.3.4.1E

Proficiency Level Distribution: List 6-8 S401 Online

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	2,823	4.72%	3,905	6.41%	5,043	8.36%	11,771	6.50%
2	4,700	7.86%	4,968	8.16%	5,623	9.32%	15,291	8.45%
3	15,178	25.40%	13,354	21.92%	12,325	20.44%	40,857	22.57%
4	9,057	15.16%	9,080	14.91%	7,310	12.12%	25,447	14.06%
5	9,310	15.58%	7,829	12.85%	4,953	8.21%	22,092	12.21%
6	18,691	31.28%	21,777	35.75%	25,058	41.55%	65,526	36.21%
Total	59,759	100.00%	60,913	100.00%	60,312	100.00%	180,984	100.00%

Table 3.3.4.1F
Raw Score to Scale Score Conversion: List 6-8 S401 Online
n/a

Figure 3.3.4.1H
Test Characteristic Curve: List 6-8 S401 Online
n/a

Figure 3.3.4.1I
Test Information Function: List 6-8 S401 Online
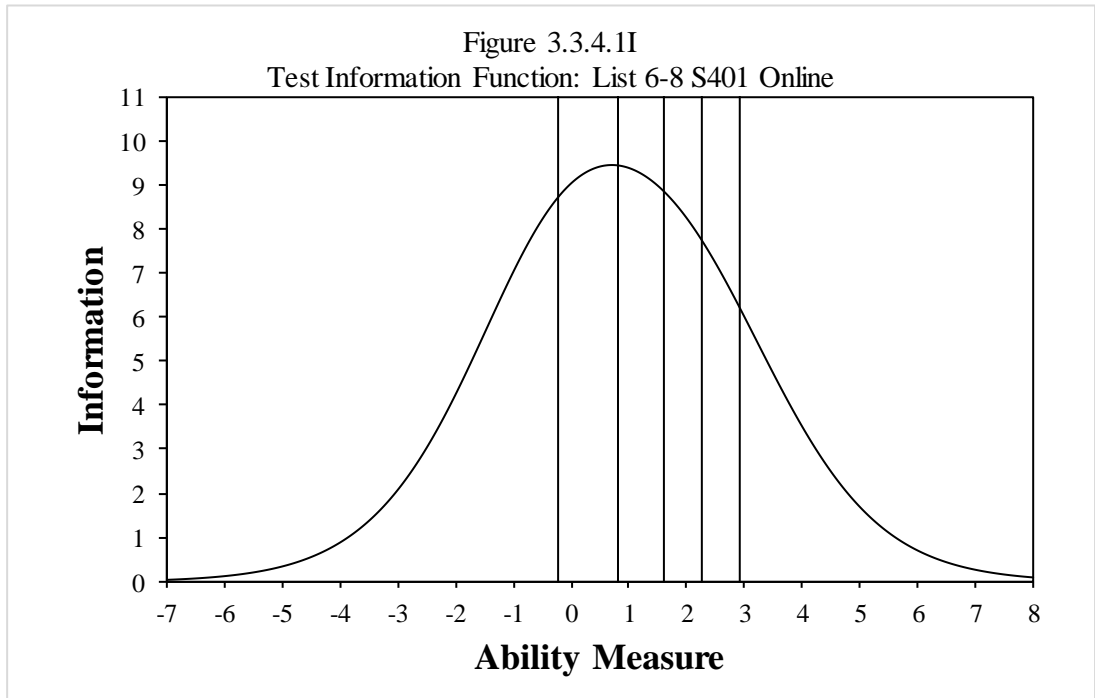
Table 3.3.4.1J

Reliability: List 6-8 S401 Online

No. of Students	No. of Items	Rasch Reliability Estimate
180,984	54	.87

Table 3.3.4.1K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: List 6-8 S401 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	6	294	2	17.35	17.86	17.60	0.36
	7	302	12	16.84	21.43	18.15	1.99
	8	308	24	17.86	17.86	17.86	0.00
2/3	6	332	13	17.35	17.35	17.35	0.00
	7	340	10	17.86	17.86	17.86	0.00
	8	347	1	17.35	17.35	17.35	0.00
3/4	6	363	3	17.86	17.86	17.86	0.00
	7	370	52	17.86	17.86	17.86	0.00
	8	377	74	17.35	17.86	17.37	0.12
4/5	6	385	145	17.35	18.88	17.87	0.19
	7	394	43	17.86	18.88	18.20	0.33
	8	402	653	17.86	19.90	19.55	0.61
5/6	6	411	N/A	N/A	N/A	N/A	N/A
	7	420	5	19.90	19.90	19.90	0.00
	8	427	2,681	18.37	19.39	18.37	0.06

Table 3.3.4.1Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.611	0.514		0.387	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.809		0.581	
	2	0.413		0.298	
	3	0.655		0.543	
	4	0.382		0.296	
	5	0.437		0.328	
	6	0.879		0.801	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.970	0.005	0.024	0.959
	2/3	0.928	0.043	0.029	0.893
	3/4	0.875	0.049	0.076	0.833
	4/5	0.894	0.050	0.056	0.846
	5/6	0.908	0.056	0.035	0.870

Table 3.3.4.1Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.609	0.521		0.389	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.793		0.578	
	2	0.376		0.275	
	3	0.596		0.484	
	4	0.388		0.301	
	5	0.374		0.276	
	6	0.895		0.826	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.961	0.009	0.031	0.945
	2/3	0.921	0.048	0.031	0.885
	3/4	0.879	0.049	0.072	0.839
	4/5	0.897	0.047	0.057	0.850
	5/6	0.908	0.057	0.035	0.869

Table 3.3.4.1Liii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.627	0.546		0.408	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.784		0.594	
	2	0.380		0.287	
	3	0.580		0.471	
	4	0.355		0.268	
	5	0.267		0.189	
	6	0.927		0.870	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.951	0.013	0.036	0.931
	2/3	0.913	0.051	0.036	0.878
	3/4	0.888	0.042	0.071	0.851
	4/5	0.912	0.038	0.050	0.867
	5/6	0.909	0.063	0.028	0.871

3.3.4.2   Reading 6-8

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

Figure 3.3.4.2C
Raw Scores: Read 6-8 S401 Online
n/a

Table 3.3.4.2C
Raw Score Descriptive Statistics: Read 6-8 S401 Online
n/a

Figure 3.3.4.2D
Scale Scores: Read 6-8 S401 Online

Figure 3.3.4.2E
Proficiency Level: Read 6-8 S401 Online
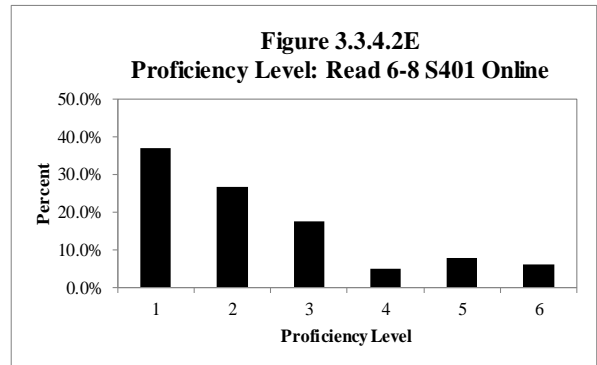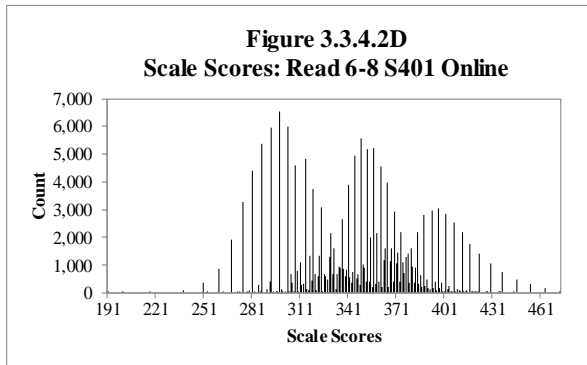
Table 3.3.4.2D

Scale Score Descriptive Statistics: Read 6-8 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	58,039	217	473	336.25	36.60
7	57,821	191	473	343.59	39.79
8	56,091	200	473	350.97	42.11
Total	171,951	191	473	343.52	39.99

Table 3.3.4.2E

Proficiency Level Distribution: Read 6-8 S401 Online

| Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	21,462	36.98%	21,347	36.92%	20,506	36.56%	63,315	36.82%
2	16,000	27.57%	15,531	26.86%	14,347	25.58%	45,878	26.68%
3	11,014	18.98%	10,127	17.51%	9,103	16.23%	30,244	17.59%
4	2,928	5.04%	3,127	5.41%	2,684	4.79%	8,739	5.08%
5	4,469	7.70%	4,096	7.08%	4,754	8.48%	13,319	7.75%
6	2,166	3.73%	3,593	6.21%	4,697	8.37%	10,456	6.08%
Total	58,039	100.00%	57,821	100.00%	56,091	100.00%	171,951	100.00%

Table 3.3.4.2F
Raw Score to Scale Score Conversion: Read 6-8 S401 Online
n/a

Figure 3.3.4.2H
Test Characteristic Curve: Read 6-8 S401 Online
n/a

Figure 3.3.4.2I
Test Information Function: Read 6-8 S401 Online
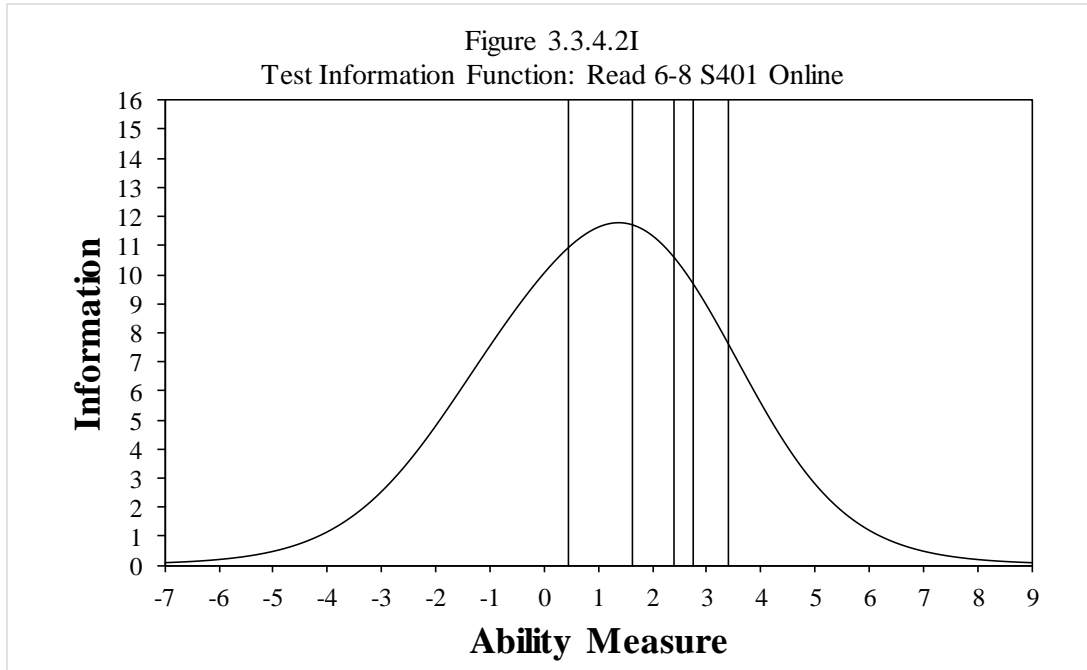
Table 3.3.4.2J

Reliability: Read 6-8 S401 Online

No. of Students	No. of Items	Rasch Reliability Estimate
171,951	69	.92

Table 3.3.4.2K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: Read 6-8 S401 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	6	323	471	10.71	12.24	12.01	0.51
	7	329	629	10.71	12.24	12.02	0.42
	8	335	477	10.71	12.24	12.09	0.46
2/3	6	353	2,514	10.20	10.71	10.20	0.02
	7	360	26	10.20	12.24	11.24	0.55
	8	366	79	10.20	11.73	10.49	0.53
3/4	6	373	107	10.20	10.71	10.32	0.21
	7	380	410	10.20	12.76	10.69	0.15
	8	386	327	10.71	12.24	11.09	0.24
4/5	6	382	101	10.71	10.71	10.71	0.00
	7	389	76	10.71	11.22	10.81	0.20
	8	395	221	10.71	11.73	10.92	0.38
5/6	6	399	86	10.71	11.22	10.72	0.06
	7	406	16	11.73	12.24	11.89	0.24
	8	412	1,108	11.22	12.24	11.23	0.04

Table 3.3.4.2Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.726	0.639		0.514	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.899		0.850	
	2	0.687		0.589	
	3	0.636		0.520	
	4	0.281		0.204	
	5	0.586		0.443	
	6	0.748		0.573	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.920	0.037	0.044	0.887
	2/3	0.915	0.050	0.035	0.882
	3/4	0.940	0.032	0.028	0.914
	4/5	0.953	0.029	0.017	0.933
	5/6	0.979	0.013	0.008	0.969

Table 3.3.4.2Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.721	0.634		0.512	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.893		0.844	
	2	0.682		0.581	
	3	0.608		0.491	
	4	0.301		0.217	
	5	0.512		0.381	
	6	0.828		0.694	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.918	0.039	0.043	0.885
	2/3	0.917	0.048	0.035	0.884
	3/4	0.939	0.033	0.028	0.914
	4/5	0.953	0.026	0.020	0.932
	5/6	0.973	0.018	0.009	0.962

Table 3.3.4.2Liii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.713	0.627		0.509	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.890		0.841	
	2	0.671		0.567	
	3	0.579		0.458	
	4	0.253		0.182	
	5	0.527		0.400	
	6	0.833		0.714	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.918	0.040	0.043	0.884
	2/3	0.917	0.046	0.037	0.885
	3/4	0.938	0.032	0.029	0.912
	4/5	0.948	0.032	0.020	0.926
	5/6	0.966	0.022	0.012	0.952

3.3.4.3 Writing 6-8

3.3.4.3i Writing 6-8 A

Table 3.3.4.3Ai

Complete Task Analysis and Summary: Writ 6-8 A S401 Online

Task Type	Average Task Difficulty (in logits)	No. of Tasks	Average Infit Mean Square	Average Outfit Mean Square
Constructed Response	0.92	3	0.46	0.49

Name	Task Difficulty (in logits)	Anchored?	Fit Statistics	
			Infit Mnsq	Outfit Mnsq
1.W68A_SI_Clubs_P100_A301_HW_14284	0.85	Yes	0.45	0.47
2.W68A_LA_Cat_P100_A301_HW_14285	0.60	Yes	0.45	0.47
3.W68A_MS_Ecosystems_MaMaNiVi_P100_A203_HW_14287	1.32	Yes	0.48	0.54

Raw Score Distribution by Task	Raw Score	Task 1	Task 2	Task 3	
	0	6.53%	4.18%	7.99%	
	1	5.97%	7.68%	10.24%	
	2	9.30%	8.04%	15.07%	
	3	17.57%	17.37%	39.18%	
	4	29.76%	29.66%	21.36%	
	5	24.07%	23.94%	5.34%	
	6	5.92%	7.71%	0.73%	
	7	0.81%	1.25%	0.07%	
	8	0.07%	0.16%	0.01%	
	9	0.01%	0.02%	0.00%	

Figure 3.3.4.3Ci
Raw Scores: Writ 6-8 A S401 Online

Figure 3.3.4.3Di
Scale Scores: Writ 6-8 A S401 Online
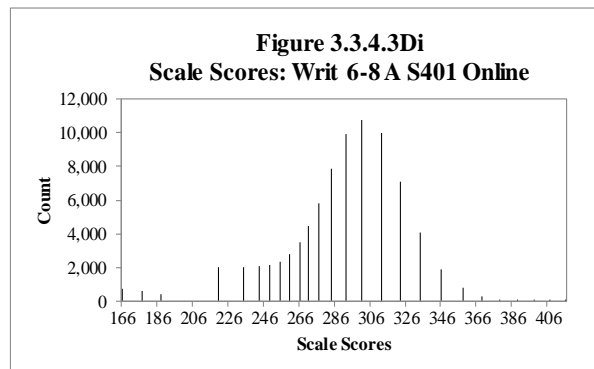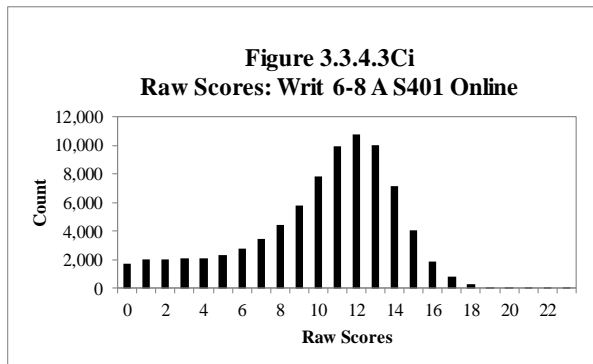
Table 3.3.4.3Ci

Raw Score Descriptive Statistics: Writ 6-8 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	24,619	0	21	9.33	4.01
7	28,520	0	23	10.09	3.95
8	28,652	0	23	10.60	3.94
Total	81,791	0	23	10.04	4.00

Table 3.3.4.3Di

Scale Score Descriptive Statistics: Writ 6-8 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	24,619	166	398	281.95	34.25
7	28,520	177	416	288.75	33.16
8	28,652	188	416	293.61	32.66
Total	81,791	166	416	288.41	33.65

Figure 3.3.4.3Ei
Proficiency Level: Writ 6-8 A S401 Online
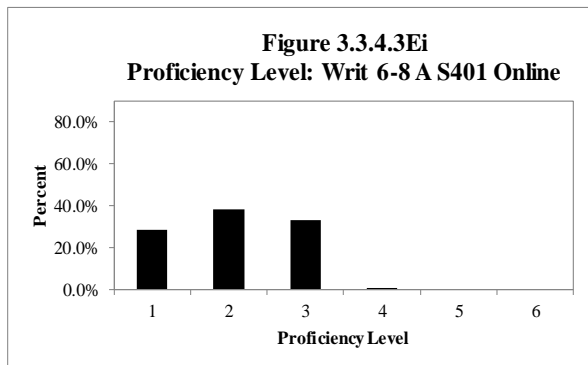
Table 3.3.4.3Ei

Proficiency Level Distribution: Writ 6-8 A S401 Online

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	6,769	27.50%	7,776	27.27%	8,738	30.50%	23,283	28.47%
2	9,399	38.18%	12,299	43.12%	9,488	33.11%	31,186	38.13%
3	8,409	34.16%	8,299	29.10%	10,356	36.14%	27,064	33.09%
4	42	0.17%	146	0.51%	70	0.24%	258	0.32%
5	0	0.00%	0	0.00%	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	24,619	100.00%	28,520	100.00%	28,652	100.00%	81,791	100.00%

Figure 3.3.4.3Hi
Test Characteristic Curve: Writ 6-8 A S401 Online

Figure 3.3.4.3Ii
Test Information Function: Writ 6-8 A S401 Online
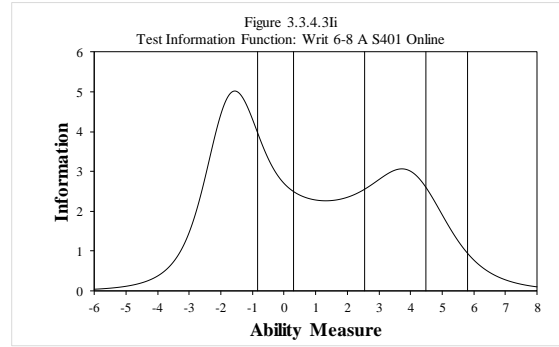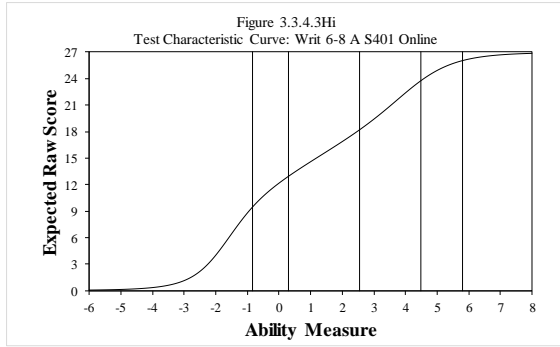
Table 3.3.4.3Ji

Reliability: Writ 6-8 A S401 Online

Reliability	No. of Students	No. of Tasks	Response Modes		Cronbach's Alpha	SEM
			Hand-written (HW)	Keyboarded (KB)		
	81,791	3			.870	1.439
Interrater Reliability	**Task**	**Mode of Response**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	HW	636	98	2	0
		KB	39,638	95	5	0
	2	HW	590	97	3	0
		KB	39,956	94	5	0
	3	HW	628	97	2	0
		KB	39,790	96	4	0

Table 3.3.4.3Ki
Conditional Standard Error of Measurement at Cut Scores: Writ 6-8 A S401 Online
n/a

Table 3.3.4.3Li
Accuracy and Consistency of Classification Indices: Writ 6-8 A S401 Online
n/a

3.3.4.3ii   Writing 6-8 B/C

Please note that this section 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.

Figure 3.3.4.3Cii
Raw Scores: Writ 6-8 B/C S401 Online

Figure 3.3.4.3Dii
Scale Scores: Writ 6-8 B/C S401 Online
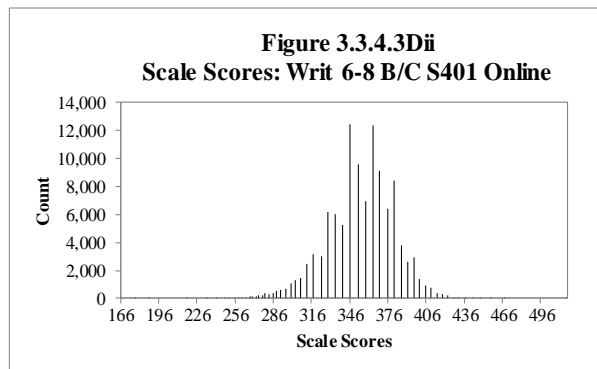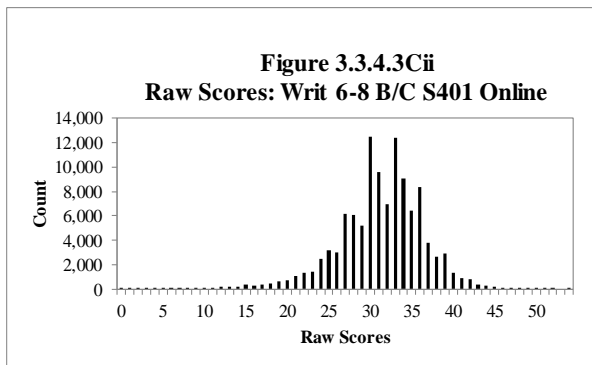
Table 3.3.4.3Cii

Raw Score Descriptive Statistics: Writ 6-8 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	39,209	0	50	29.05	5.63
7	37,403	0	52	31.46	4.97
8	36,268	0	54	33.17	4.81
Total	112,880	0	54	31.18	5.43

Table 3.3.4.3Dii

Scale Score Descriptive Statistics: Writ 6-8 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	39,209	166	448	341.89	28.70
7	37,403	177	467	354.75	26.40
8	36,268	188	516	364.06	25.65
Total	112,880	166	516	353.28	28.49

Figure 3.3.4.3Eii
Proficiency Level: Writ 6-8 B/C S401 Online
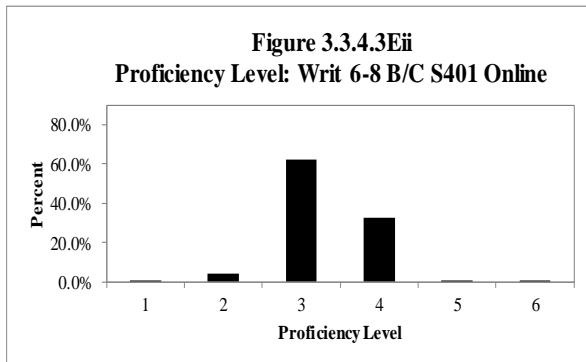
Table 3.3.4.3Eii

Proficiency Level Distribution: Writ 6-8 B/C S401 Online

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	408	1.04%	190	0.51%	203	0.56%	801	0.71%
2	2,321	5.92%	1,446	3.87%	790	2.18%	4,557	4.04%
3	25,530	65.11%	23,194	62.01%	21,364	58.91%	70,088	62.09%
4	10,862	27.70%	12,444	33.27%	13,710	37.80%	37,016	32.79%
5	86	0.22%	124	0.33%	192	0.53%	402	0.36%
6	2	0.01%	5	0.01%	9	0.02%	16	0.01%
Total	39,209	100.00%	37,403	100.00%	36,268	100.00%	112,880	100.00%

Figure 3.3.4.3Hii
Test Characteristic Curve: Writ 6-8 B/C S401 Online

Figure 3.3.4.3Iii
Test Information Function: Writ 6-8 B/C S401 Online
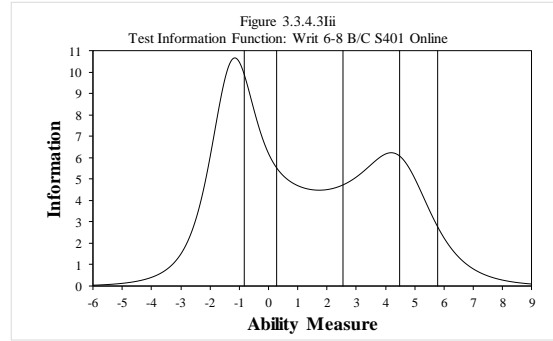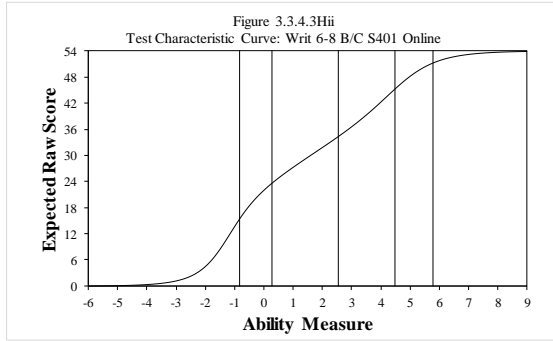
Table 3.3.4.3Jii

Reliability: Writ 6-8 B/C S401 Online

Reliability	No. of Students	No. of Tasks	Response Modes		Cronbach's Alpha	SEM
			Hand-written (HW)	Keyboarded (KB)		
	112,880	3			.904	1.686
Interrater Reliability	**Task**	**Mode of Response**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	HW	378	97	3	0
		KB	57,860	96	3	0
	2	HW	382	96	4	0
		KB	58,514	96	4	0
	3	HW	402	94	6	0
		KB	59,824	95	5	0

Table 3.3.4.3Kii
Conditional Standard Error of Measurement at Cut Scores: Writ 6-8 B/C S401 Online
n/a

Table 3.3.4.3Lii
Accuracy and Consistency of Classification Indices: Writ 6-8 B/C S401 Online
n/a

3.3.4.3iii   *Writing 6-8 Across Tiers*

Table 3.3.4.3Aiii
Complete Task Analysis and Summary: Writ 6-8 S401 Online
n/a

Table 3.3.4.3Biii
DIF Analysis and Summary: Writ 6-8 S401 Online
n/a

Figure 3.3.4.3Ciii
Raw Scores: Writ 6-8 S401 Online

Figure 3.3.4.3Diii
Scale Scores: Writ 6-8 S401 Online
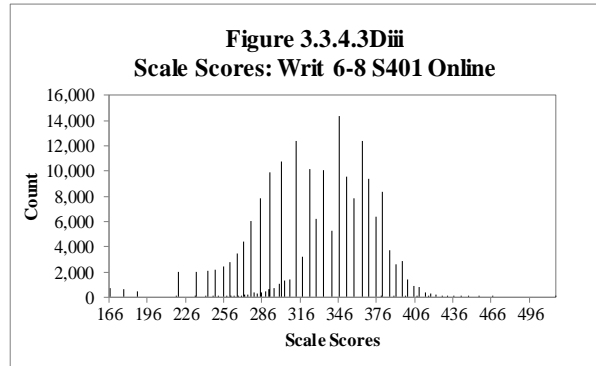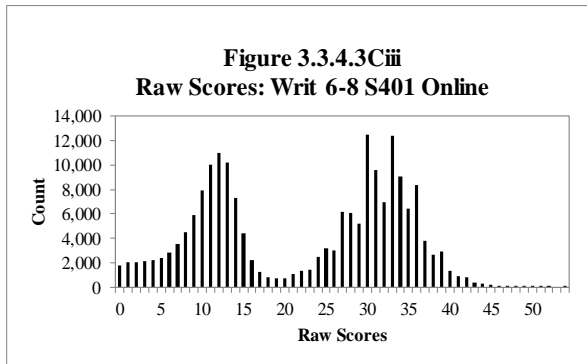
Table 3.3.4.3Ciii

Raw Score Descriptive Statistics: Writ 6-8 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	63,828	0	50	21.45	10.85
7	65,923	0	52	22.22	11.53
8	64,920	0	54	23.21	12.06
Total	194,671	0	54	22.30	11.52

Table 3.3.4.3Diii

Scale Score Descriptive Statistics: Writ 6-8 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	63,828	166	448	318.77	42.54
7	65,923	177	467	326.20	44.05
8	64,920	188	516	332.97	45.41
Total	194,671	166	516	326.02	44.40

Figure 3.3.4.3Eiii
Proficiency Level: Writ 6-8 S401 Online
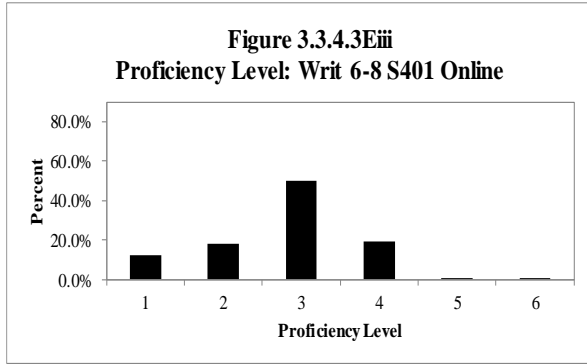
Table 3.3.4.3Eiii

Proficiency Level Distribution: Writ 6-8 S401 Online

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	7,177	11.24%	7,966	12.08%	8,941	13.77%	24,084	12.37%
2	11,720	18.36%	13,745	20.85%	10,278	15.83%	35,743	18.36%
3	33,939	53.17%	31,493	47.77%	31,720	48.86%	97,152	49.91%
4	10,904	17.08%	12,590	19.10%	13,780	21.23%	37,274	19.15%
5	86	0.13%	124	0.19%	192	0.30%	402	0.21%
6	2	0.00%	5	0.01%	9	0.01%	16	0.01%
Total	63,828	100.00%	65,923	100.00%	64,920	100.00%	194,671	100.00%

Table 3.3.4.3Fiii
Raw Score to Scale Score Conversion: Writ 6-8 S401 Online
n/a

Table 3.3.4.3Giii
Equating Summary: Writ 6-8 S401 Online
n/a

Figure 3.3.4.3Hiii
Test Characteristic Curve: Writ 6-8 S401 Online

Figure 3.3.4.3Iiii
Test Information Function: Writ 6-8 S401 Online
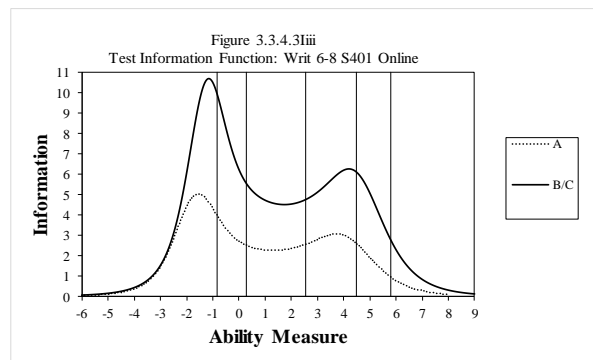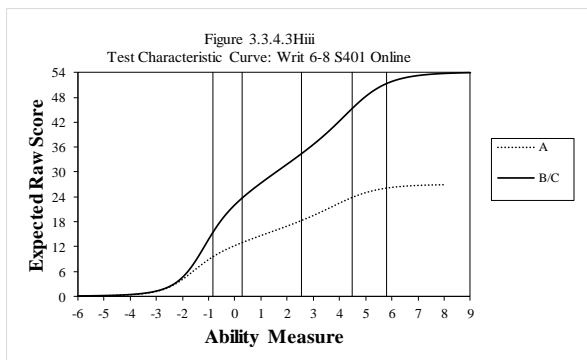
Table 3.3.4.3Jiii

Reliability: Writ 6-8 Weighted Reliability S401 Online

Tiers	No. of Students	Reliability	Weighted Reliability
A	81,791	0.870	0.890
B/C	112,880	0.904	

Table 3.3.4.3Kiii

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

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 3.3.4.3Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.761	0.676		0.490	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.811		0.701	
	2	0.688		0.565	
	3	0.790		0.740	
	4	0.687		0.550	
	5	-		-	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.957	0.021	0.022	0.937
	2/3	0.919	0.030	0.051	0.888
	3/4	0.885	0.070	0.045	0.846
	4/5	0.999	0.001	0.000	0.999
	5/6	1.000	0.000	0.000	1.000

Table 3.3.4.3Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.762	0.673		0.514	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.810		0.699	
	2	0.706		0.590	
	3	0.778		0.716	
	4	0.739		0.623	
	5	-		0.032	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.953	0.023	0.024	0.932
	2/3	0.913	0.031	0.056	0.881
	3/4	0.896	0.059	0.045	0.857
	4/5	0.998	0.002	0.000	0.998
	5/6	1.000	0.000	0.000	1.000

Table 3.3.4.3Liii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.767	0.678		0.522	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.855		0.760	
	2	0.623		0.497	
	3	0.800		0.736	
	4	0.738		0.633	
	5	-		0.034	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.955	0.019	0.026	0.935
	2/3	0.922	0.033	0.044	0.891
	3/4	0.891	0.056	0.052	0.849
	4/5	0.997	0.003	0.000	0.995
	5/6	1.000	0.000	0.000	1.000

3.3.4.4   Speaking 6-8

3.3.4.4i   Speaking 6-8 Pre-A

Table 3.3.4.4Ai
Complete Task Analysis and Summary: Spek 6-8 Pre-A S401 Online
n/a

Please note that this section 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.

Figure 3.3.4.4Ci
Raw Scores: Spek 6-8 Pre-A S401 Online

Figure 3.3.4.4Di
Scale Scores: Spek 6-8 Pre-A S401 Online
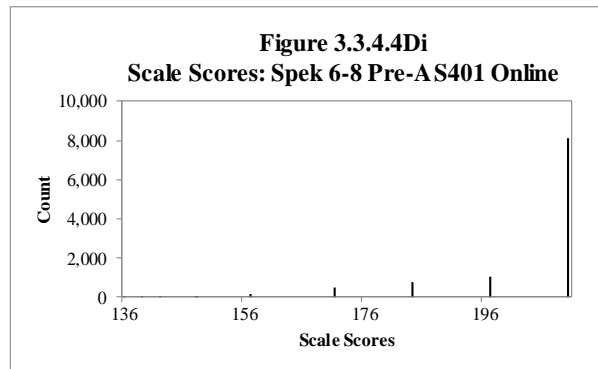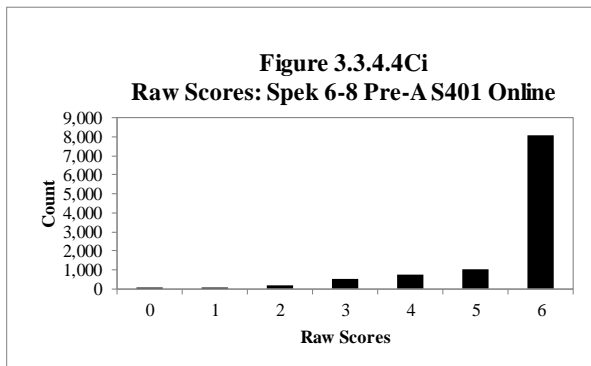
Table 3.3.4.4Ci

Raw Score Descriptive Statistics: Spek 6-8 Pre-A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,538	0	6	5.37	1.20
7	3,549	0	6	5.48	1.09
8	4,621	0	6	5.55	1.02
Total	10,708	0	6	5.48	1.09

Table 3.3.4.4Di

Scale Score Descriptive Statistics: Spek 6-8 Pre-A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	2,538	136	210	201.77	15.72
7	3,549	142	210	203.21	14.07
8	4,621	148	210	204.23	12.91
Total	10,708	136	210	203.31	14.03

Figure 3.3.4.4Ei
Proficiency Level: Spek 6-8 Pre-A S401 Online
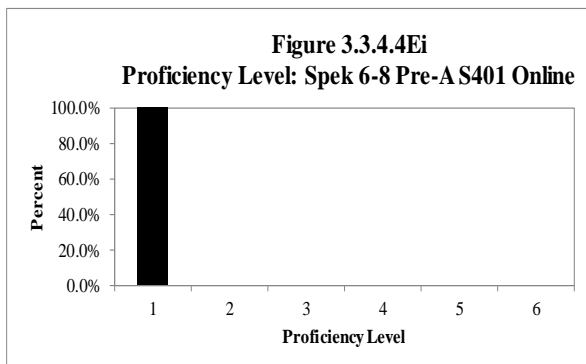
Table 3.3.4.4Ei

Proficiency Level Distribution: Spek 6-8 Pre-A S401 Online

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	2,538	100.00%	3,549	100.00%	4,621	100.00%	10,708	100.00%
Total	2,538	100.00%	3,549	100.00%	4,621	100.00%	10,708	100.00%

Table 3.3.4.4Gi
Equating Summary: Spek 6-8 Pre-A S401 Online
n/a

Figure 3.3.4.4Hi
Test Characteristic Curve: Spek 6-8 Pre-A S401 Online

Figure 3.3.4.4Ii
Test Information Function: Spek 6-8 Pre-A S401 Online
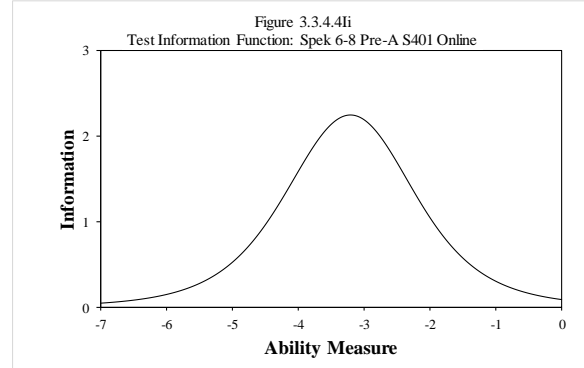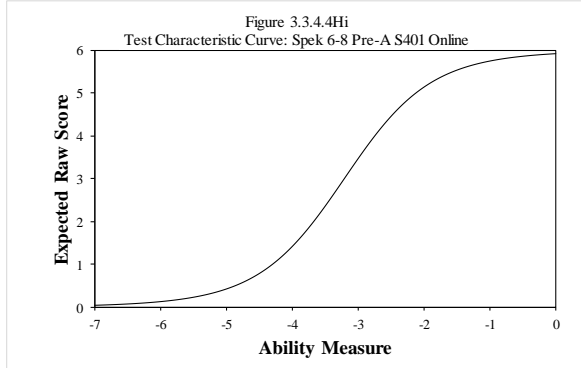
Table 3.3.4.4Ji

Reliability: Spek 6-8 Pre-A S401 Online

Reliability					
	No. of Students	**No. of Tasks**	**Cronbach's Alpha**		**SEM**
	10,708	3	.719		0.577
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	6,114	97	3	0
	2	6,100	97	3	0
	3	6,192	98	2	0

Table 3.3.4.4Ki
Conditional Standard Error of Measurement at Cut Scores: Spek 6-8 Pre-A S401 Online
n/a

Table 3.3.4.4Li
Accuracy and Consistency of Classification Indices: Spek 6-8 Pre-A S401 Online
n/a

3.3.4.4ii   Speaking 6-8 A

Table 3.3.4.4Aii
Complete Task Analysis and Summary: Spek 6-8 A S401 Online
n/a

Please note that this section 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.

Figure 3.3.4.4Cii
Raw Scores: Spek 6-8 A S401 Online

Figure 3.3.4.4Dii
Scale Scores: Spek 6-8 A S401 Online
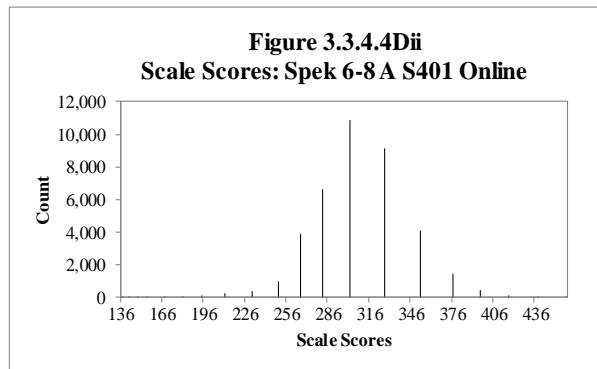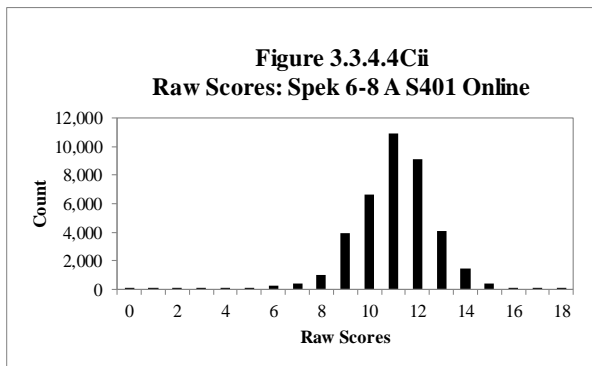
Table 3.3.4.4Cii

Raw Score Descriptive Statistics: Spek 6-8 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	13,195	0	17	10.79	1.70
7	9,301	0	18	10.73	1.66
8	16,134	0	18	11.41	1.62
Total	38,630	0	18	11.04	1.69

Table 3.3.4.4Dii

Scale Score Descriptive Statistics: Spek 6-8 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	13,195	136	438	301.53	34.46
7	9,301	142	459	300.17	33.91
8	16,134	148	459	314.85	35.27
Total	38,630	136	459	306.77	35.34

Figure 3.3.4.4Eii
Proficiency Level: Spek 6-8 A S401 Online

Table 3.3.4.4Eii

Proficiency Level Distribution: Spek 6-8 A S401 Online

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	2,402	18.20%	1,832	19.70%	3,938	24.41%	8,172	21.15%
2	6,375	48.31%	4,535	48.76%	4,320	26.78%	15,230	39.43%
3	4,006	30.36%	2,675	28.76%	7,486	46.40%	14,167	36.67%
4	394	2.99%	257	2.76%	371	2.30%	1,022	2.65%
5	18	0.14%	1	0.01%	19	0.12%	38	0.10%
6	0	0.00%	1	0.01%	0	0.00%	1	0.00%
Total	13,195	100.00%	9,301	100.00%	16,134	100.00%	38,630	100.00%

Table 3.3.4.4Gii
Equating Summary: Spek 6-8 A S401 Online
n/a

Figure 3.3.4.4Hii
Test Characteristic Curve: Spek 6-8 A S401 Online

Figure 3.3.4.4Iii
Test Information Function: Spek 6-8 A S401 Online
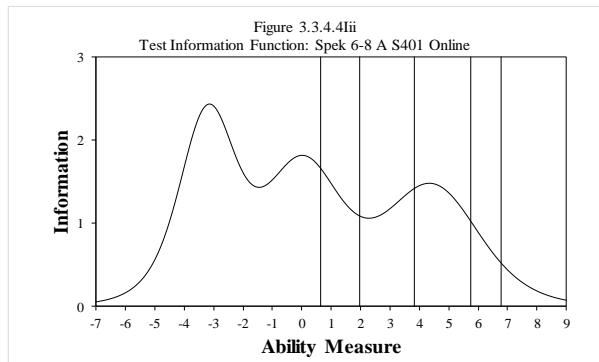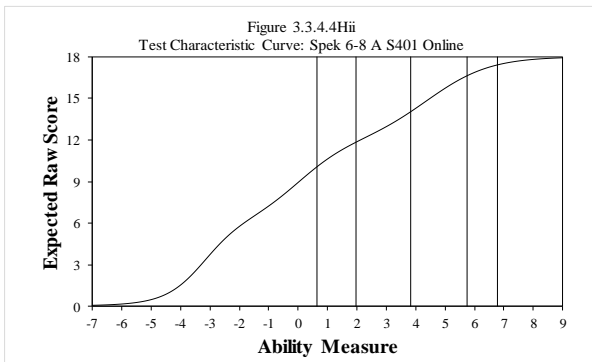
Table 3.3.4.4Jii

Reliability: Spek 6-8 A S401 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	38,630	6	.622		1.039
Interrater Reliability	Task	No. in Sample	% EX	% AD	% NA
	1	24,040	99	1	0
	2	24,045	83	17	0
	3	24,051	98	2	0
	4	24,044	83	17	0
	5	23,800	99	1	0
	6	23,798	83	17	1

Table 3.3.4.4Kii
Conditional Standard Error of Measurement at Cut Scores: Spek 6-8 A S401 Online
n/a

Table 3.3.4.4Lii
Accuracy and Consistency of Classification Indices: Spek 6-8 A S401 Online
n/a

3.3.4.4Iii   Speaking 6-8 B/C

Table 3.3.4.4Aiii
Complete Task Analysis and Summary: Spek 6-8 B/C S401 Online
n/a

Please note that this section 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.

Figure 3.3.4.4Ciii
Raw Scores: Spek 6-8 B/C S401 Online

Figure 3.3.4.4Diii
Scale Scores: Spek 6-8 B/C S401 Online
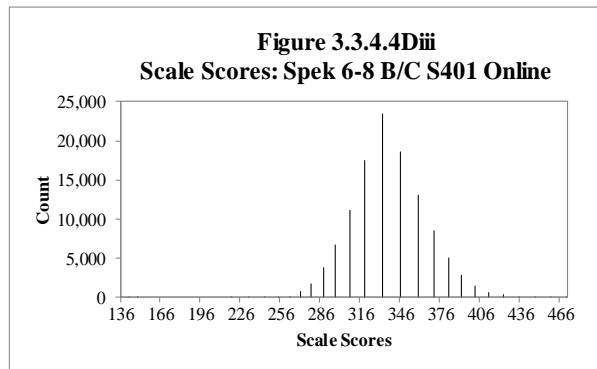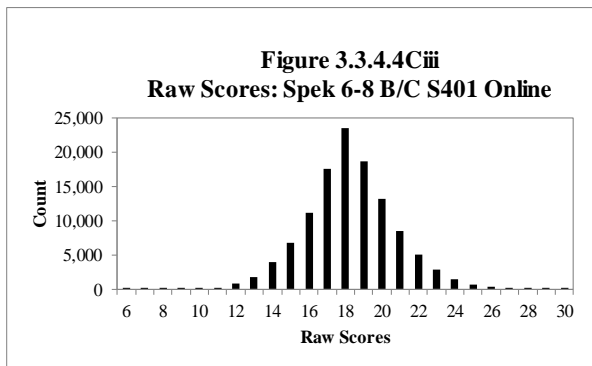
Table 3.3.4.4Ciii

Raw Score Descriptive Statistics: Spek 6-8 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	39,608	6	28	17.77	2.33
7	42,380	6	29	18.15	2.43
8	34,814	6	30	19.02	2.46
Total	116,802	6	30	18.28	2.46

Table 3.3.4.4Diii

Scale Score Descriptive Statistics: Spek 6-8 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	39,608	136	447	330.32	27.76
7	42,380	142	459	334.96	28.98
8	34,814	148	471	345.34	29.41
Total	116,802	136	471	336.48	29.34

Figure 3.3.4.4Eiii
Proficiency Level: Spek 6-8 B/C S401 Online
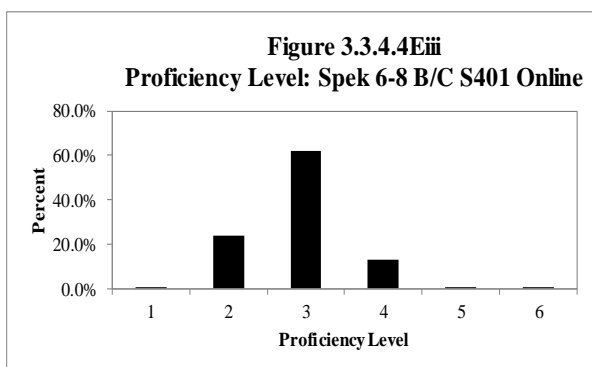
Table 3.3.4.4Eiii

Proficiency Level Distribution: Spek 6-8 B/C S401 Online

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	157	0.40%	433	1.02%	416	1.19%	1,006	0.86%
2	10,503	26.52%	9,158	21.61%	8,189	23.52%	27,850	23.84%
3	24,680	62.31%	26,375	62.23%	21,095	60.59%	72,150	61.77%
4	4,192	10.58%	6,342	14.96%	4,979	14.30%	15,513	13.28%
5	76	0.19%	68	0.16%	130	0.37%	274	0.23%
6	0	0.00%	4	0.01%	5	0.01%	9	0.01%
Total	39,608	100.00%	42,380	100.00%	34,814	100.00%	116,802	100.00%

Table 3.3.4.4Giii
Equating Summary: Spek 6-8 B/C S401 Online
n/a

Figure 3.3.4.4Hiii
Test Characteristic Curve: Spek 6-8 B/C S401 Online

Figure 3.3.4.4Iiii
Test Information Function: Spek 6-8 B/C S401 Online
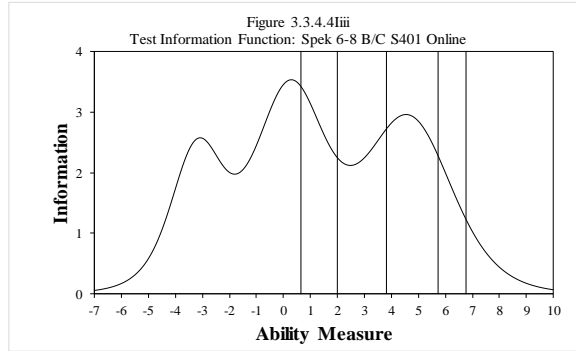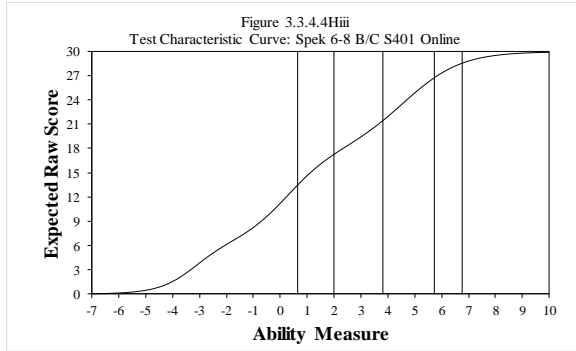
Table 3.3.4.4Jiii

Reliability: Spek 6-8 B/C S401 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	116,802	6	.731		1.277
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	70,442	78	22	1
	2	70,444	77	23	0
	3	70,708	80	20	0
	4	70,722	77	23	1
	5	70,984	79	20	1
	6	70,988	78	22	0

Table 3.3.4.4Kiii
Conditional Standard Error of Measurement at Cut Scores: Spek 6-8 B/C S401 Online
n/a

Table 3.3.4.4Liii
Accuracy and Consistency of Classification Indices: Spek 6-8 B/C S401 Online
n/a

3.3.4.4iv   *Speaking 6-8 Across Tiers*

Please note that this section 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.

Table 3.3.4.4Biv
DIF Analysis and Summary: Spek 6-8 S401 Online
n/a

Figure 3.3.4.4Civ
Raw Scores: Spek 6-8 S401 Online

Figure 3.3.4.4Div
Scale Scores: Spek 6-8 S401 Online
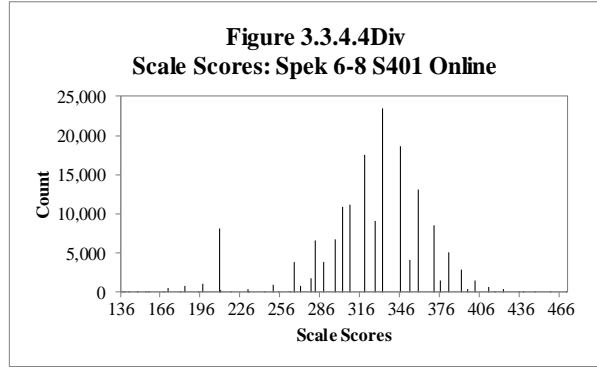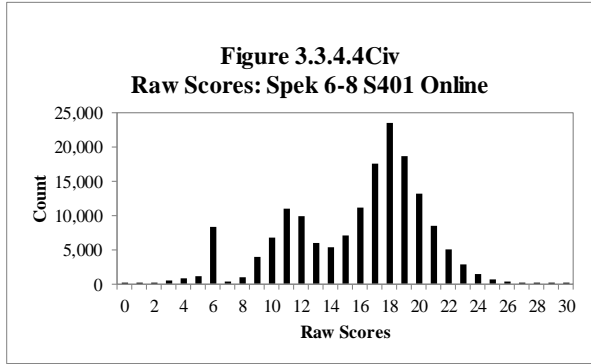
Table 3.3.4.4Civ

Raw Score Descriptive Statistics: Spek 6-8 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	55,341	0	28	15.54	4.28
7	55,230	0	29	16.09	4.52
8	55,569	0	30	15.69	5.04
Total	166,140	0	30	15.77	4.63

Table 3.3.4.4Div

Scale Score Descriptive Statistics: Spek 6-8 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	55,341	136	447	317.56	40.48
7	55,230	142	459	320.63	44.33
8	55,569	148	471	324.75	49.18
Total	166,140	136	471	320.99	44.91

Figure 3.3.4.4Eiv
Proficiency Level: Spek 6-8 S401 Online
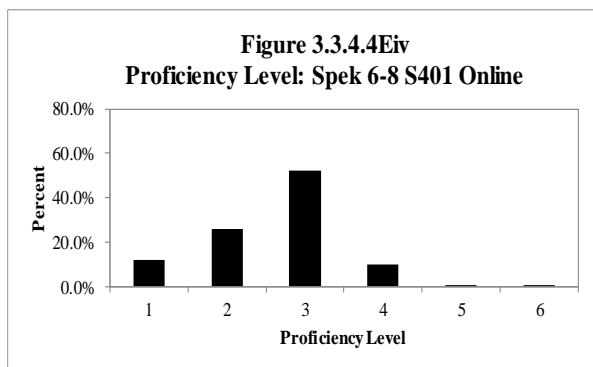
Table 3.3.4.4Eiv

Proficiency Level Distribution: Spek 6-8 S401 Online

Level	Grade 6		Grade 7		Grade 8		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	5,097	9.21%	5,814	10.53%	8,975	16.15%	19,886	11.97%
2	16,878	30.50%	13,693	24.79%	12,509	22.51%	43,080	25.93%
3	28,686	51.83%	29,050	52.60%	28,581	51.43%	86,317	51.95%
4	4,586	8.29%	6,599	11.95%	5,350	9.63%	16,535	9.95%
5	94	0.17%	69	0.12%	149	0.27%	312	0.19%
6	0	0.00%	5	0.01%	5	0.01%	10	0.01%
Total	55,341	100.00%	55,230	100.00%	55,569	100.00%	166,140	100.00%

Table 3.3.4.4Fiv
Raw Score to Scale Score Conversion: Spek 6-8 S401 Online
n/a

Figure 3.3.4.4Hiv
Test Characteristic Curve: Spek 6-8 S401 Online

Figure 3.3.4.4Iiv
Test Information Function: Spek 6-8 S401 Online
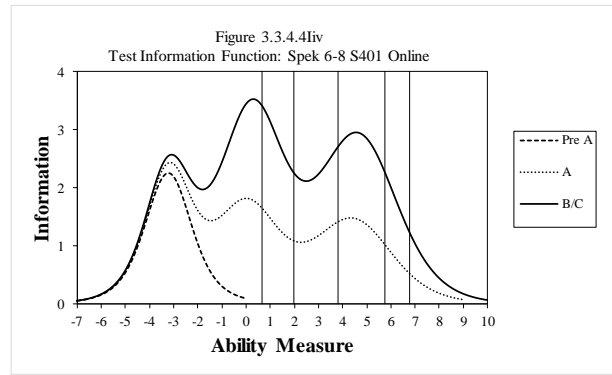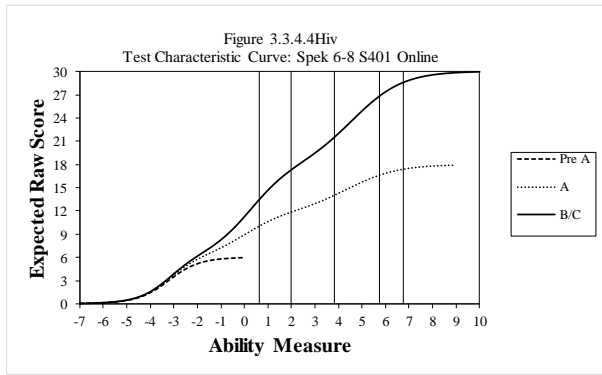
Table 3.3.4.4Jiv

Reliability: Spek 6-8 Weighted Reliability S401 Online

Tiers	No. of Students	Reliability	Weighted Reliability
Pre-A	10,708	0.719	
A	38,630	0.622	0.705
B/C	116,802	0.731	

Table 3.3.4.4Kiv

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

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

Note: Tier Pre-A is not presented as it is not possible for Tier Pre-A students to receive a proficeny level higher than 2.

Table 3.3.4.4Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.681	0.546		0.276	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.737		0.569	
	2	0.676		0.496	
	3	0.676		0.641	
	4	-		0.149	
	5	-		0.870	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.950	0.023	0.027	0.923
	2/3	0.813	0.043	0.144	0.746
	3/4	0.915	0.085	0.000	0.843
	4/5	0.998	0.002	0.000	0.998
	5/6	N/A	N/A	N/A	N/A

Table 3.3.4.4Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.652	0.523		0.243	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.771		0.602	
	2	0.576		0.376	
	3	0.654		0.629	
	4	-		0.189	
	5	-		-	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.948	0.023	0.029	0.920
	2/3	0.818	0.048	0.134	0.735
	3/4	0.879	0.121	0.000	0.817
	4/5	0.999	0.001	0.000	0.999
	5/6	1.000	0.000	0.000	1.000

Table 3.3.4.4Liii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.663	0.525		0.272	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.837		0.690	
	2	0.495		0.318	
	3	0.665		0.638	
	4	-		0.151	
	5	-		-	
	6	-		-	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.935	0.023	0.042	0.902
	2/3	0.815	0.047	0.138	0.729
	3/4	0.901	0.099	0.000	0.829
	4/5	0.997	0.003	0.000	0.997
	5/6	1.000	0.000	0.000	1.000

3.3.5 Grades: 9-12

3.3.5.1   Listening 9-12

Please note that this section 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.
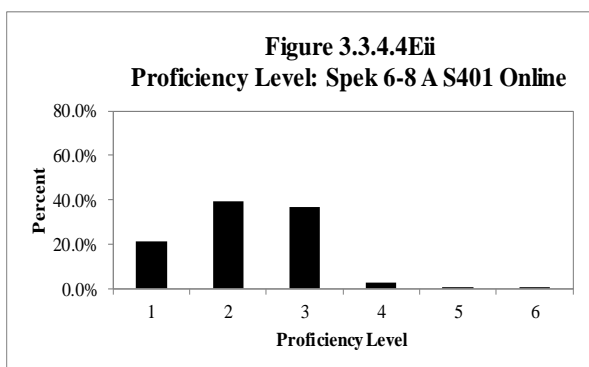
Figure 3.3.5.1C
Raw Scores: List 9-12 S401 Online
n/a

Table 3.3.5.1C
Raw Score Descriptive Statistics: List 9-12 S401 Online
n/a

Figure 3.3.5.1D
Scale Scores: List 9-12 S401 Online

Figure 3.3.5.1E
Proficiency Level: List 9-12 S401 Online
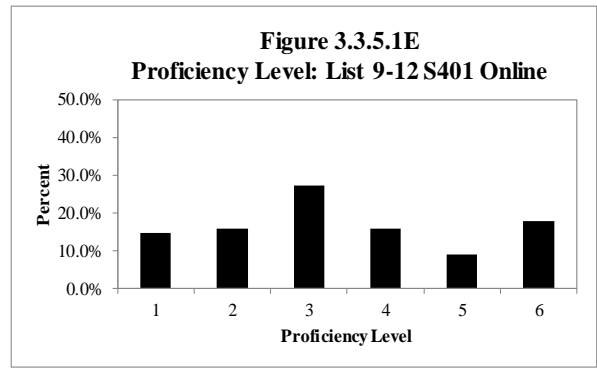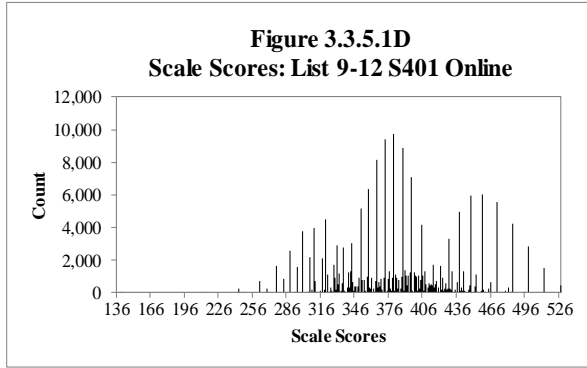
Table 3.3.5.1D

Scale Score Descriptive Statistics: List 9-12 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	77,726	136	527	382.96	54.82
10	52,056	140	527	383.51	53.02
11	37,047	214	527	385.78	52.25
12	24,783	214	527	388.18	51.19
Total	191,612	136	527	384.33	53.41

Table 3.3.5.1E

Proficiency Level Distribution: List 9-12 S401 Online

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	8,924	11.48%	7,916	15.21%	6,472	17.47%	4,834	19.51%	28,146	14.69%
2	13,792	17.74%	7,215	13.86%	5,155	13.91%	3,872	15.62%	30,034	15.67%
3	19,215	24.72%	15,500	29.78%	10,448	28.20%	7,051	28.45%	52,214	27.25%
4	13,037	16.77%	7,841	15.06%	6,065	16.37%	3,269	13.19%	30,212	15.77%
5	6,678	8.59%	4,887	9.39%	2,953	7.97%	2,461	9.93%	16,979	8.86%
6	16,080	20.69%	8,697	16.71%	5,954	16.07%	3,296	13.30%	34,027	17.76%
Total	77,726	100.00%	52,056	100.00%	37,047	100.00%	24,783	100.00%	191,612	100.00%

Table 3.3.5.1F
Raw Score to Scale Score Conversion: List 9-12 S401 Online
n/a

Figure 3.3.5.1H
Test Characteristic Curve: List 9-12 S401 Online
n/a

Figure 3.3.5.1I
Test Information Function: List 9-12 S401 Online
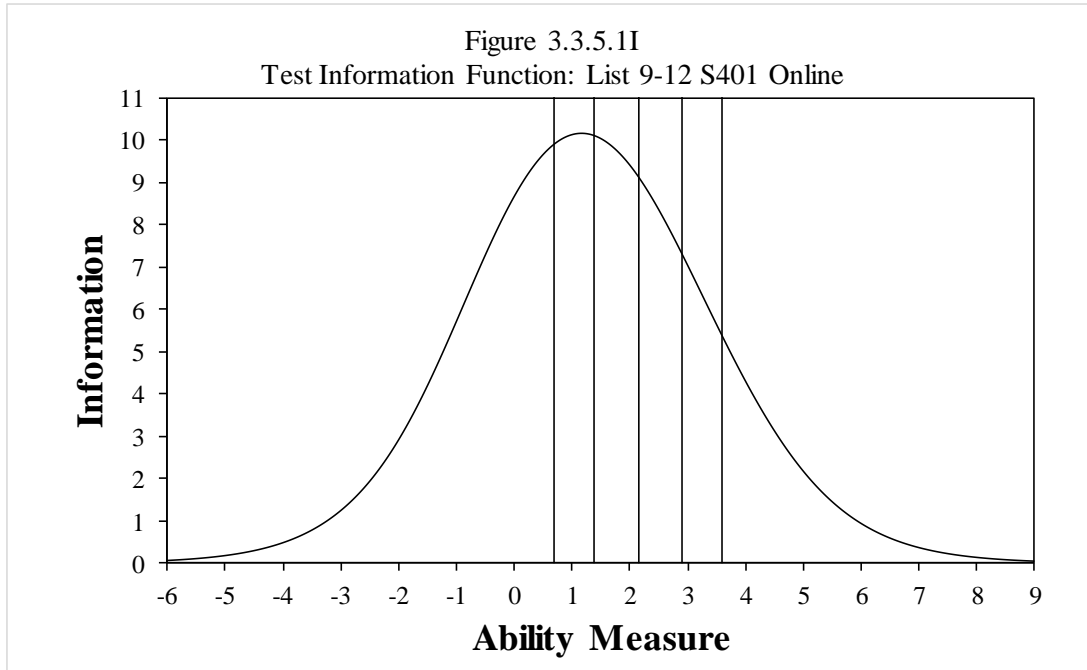
Table 3.3.5.1J

Reliability: List 9-12 S401 Online

No. of Students	No. of Items	Rasch Reliability Estimate
191,612	54	.87

Table 3.3.5.1K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: List 9-12 S401 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	9	314	N/A	N/A	N/A	N/A	N/A
	10	325	74	17.86	19.39	19.33	0.30
	11	335	248	16.33	19.90	19.44	0.26
	12	342	545	16.33	19.39	19.37	0.22
2/3	9	353	N/A	N/A	N/A	N/A	N/A
	10	358	1,532	16.33	16.33	16.33	0.00
	11	364	207	16.33	16.84	16.46	0.22
	12	368	N/A	N/A	N/A	N/A	N/A
3/4	9	383	352	16.33	16.84	16.83	0.04
	10	389	39	16.84	17.86	17.67	0.32
	11	394	2	18.37	18.37	18.37	0.00
	12	398	158	17.35	18.88	17.58	0.54
4/5	9	409	240	17.86	17.86	17.86	0.00
	10	415	285	18.37	19.39	19.22	0.38
	11	420	59	18.88	20.41	19.45	0.75
	12	426	171	18.37	20.92	19.50	1.07
5/6	9	434	108	18.37	19.39	18.41	0.19
	10	441	91	19.39	22.45	20.05	0.72
	11	447	111	20.92	20.92	20.92	0.00
	12	452	128	20.92	20.92	20.92	0.00

Table 3.3.5.1Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.585	0.483		0.372	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.742		0.567	
	2	0.495		0.396	
	3	0.569		0.464	
	4	0.471		0.360	
	5	0.345		0.245	
	6	0.897		0.807	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.927	0.022	0.050	0.899
	2/3	0.886	0.067	0.047	0.841
	3/4	0.880	0.058	0.062	0.837
	4/5	0.921	0.037	0.042	0.882
	5/6	0.938	0.043	0.019	0.912

Table 3.3.5.1Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.586	0.481		0.370	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.788		0.635	
	2	0.387		0.298	
	3	0.651		0.543	
	4	0.440		0.333	
	5	0.412		0.295	
	6	0.886		0.791	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.918	0.026	0.056	0.885
	2/3	0.883	0.078	0.039	0.837
	3/4	0.882	0.053	0.065	0.839
	4/5	0.928	0.037	0.035	0.891
	5/6	0.946	0.037	0.017	0.924

Table 3.3.5.1Liii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.586	0.482		0.372	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.786		0.648	
	2	0.380		0.289	
	3	0.621		0.515	
	4	0.479		0.366	
	5	0.365		0.258	
	6	0.899		0.801	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.913	0.033	0.054	0.875
	2/3	0.881	0.077	0.041	0.835
	3/4	0.881	0.057	0.062	0.839
	4/5	0.931	0.032	0.037	0.894
	5/6	0.948	0.038	0.014	0.928

Table 3.3.5.1Liv

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.587	0.481		0.368	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.758		0.629	
	2	0.391		0.303	
	3	0.637		0.523	
	4	0.424		0.316	
	5	0.481		0.353	
	6	0.864		0.760	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.900	0.044	0.056	0.856
	2/3	0.874	0.079	0.047	0.827
	3/4	0.895	0.045	0.060	0.853
	4/5	0.935	0.039	0.027	0.903
	5/6	0.952	0.032	0.016	0.934

3.3.5.2   Reading 9-12

Please note that this section 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.

Figure 3.3.5.2C
Raw Scores: Read 9-12 S401 Online
n/a

Table 3.3.5.2C
Raw Score Descriptive Statistics: Read 9-12 S401 Online
n/a

Figure 3.3.5.2D
Scale Scores: Read 9-12 S401 Online

Figure 3.3.5.2E
Proficiency Level: Read 9-12 S401 Online
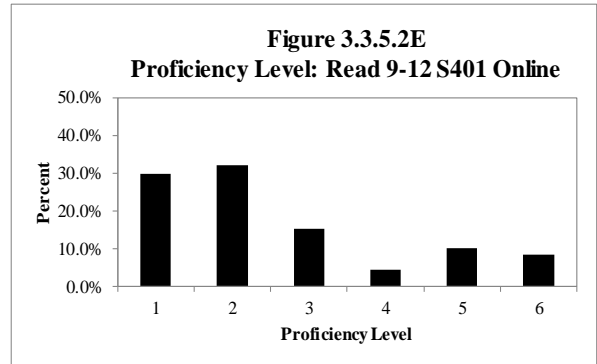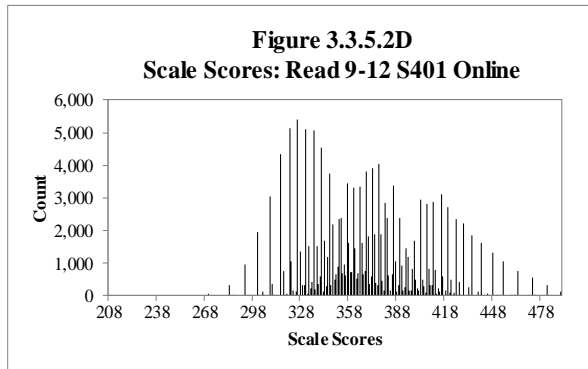
Table 3.3.5.2D

Scale Score Descriptive Statistics: Read 9-12 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	65,870	208	491	363.10	38.59
10	42,507	250	491	367.67	36.39
11	28,833	250	491	372.55	36.93
12	19,870	250	491	375.59	35.58
Total	157,080	208	491	367.65	37.61

Table 3.3.5.2E

Proficiency Level Distribution: Read 9-12 S401 Online

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	21,306	32.35%	12,531	29.48%	7,962	27.61%	4,925	24.79%	46,724	29.75%
2	18,917	28.72%	13,538	31.85%	10,221	35.45%	7,855	39.53%	50,531	32.17%
3	10,155	15.42%	7,001	16.47%	3,873	13.43%	2,794	14.06%	23,823	15.17%
4	2,627	3.99%	1,898	4.47%	1,474	5.11%	1,007	5.07%	7,006	4.46%
5	6,614	10.04%	4,219	9.93%	2,892	10.03%	1,949	9.81%	15,674	9.98%
6	6,251	9.49%	3,320	7.81%	2,411	8.36%	1,340	6.74%	13,322	8.48%
Total	65,870	100.00%	42,507	100.00%	28,833	100.00%	19,870	100.00%	157,080	100.00%

Table 3.3.5.2F
Raw Score to Scale Score Conversion: Read 9-12 S401 Online
n/a

Figure 3.3.5.2H
Test Characteristic Curve: Read 9-12 S401 Online
n/a

Figure 3.3.5.2I
Test Information Function: Read 9-12 S401 Online
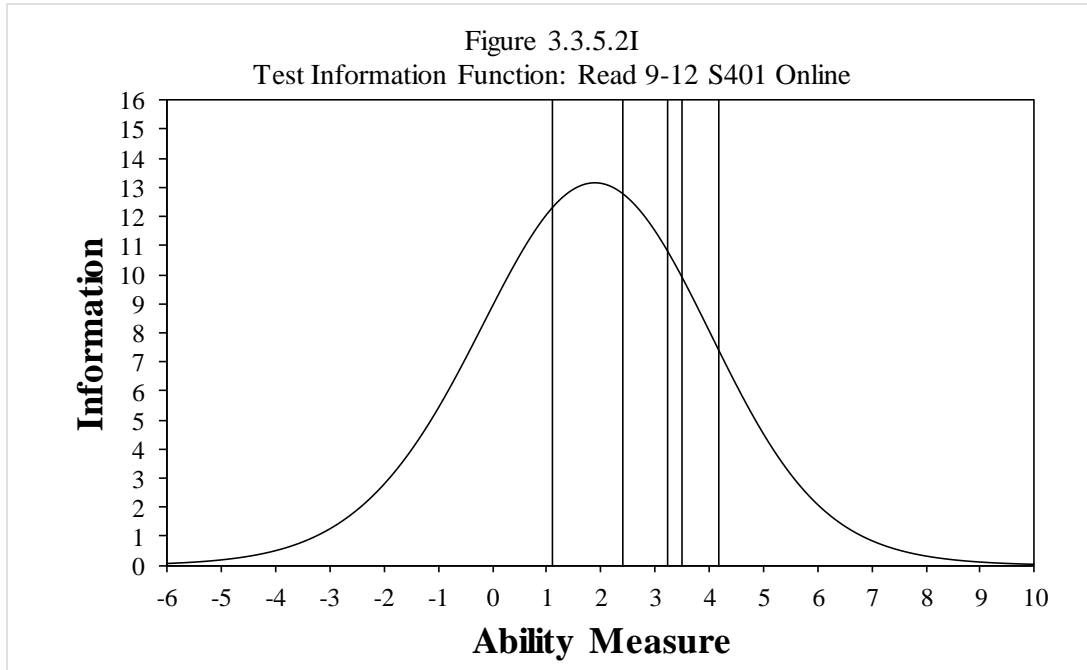
Table 3.3.5.2J

Reliability: Read 9-12 S401 Online

No. of Students	No. of Items	Rasch Reliability Estimate
157,080	72	.91

Table 3.3.5.2K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: Read 9-12 S401 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	9	340	455	11.22	12.24	11.24	0.12
	10	344	75	11.22	12.24	11.36	0.29
	11	348	475	11.22	12.24	11.23	0.07
	12	352	674	10.71	11.73	11.23	0.04
2/3	9	372	184	10.20	10.71	10.25	0.15
	10	377	1,168	10.20	11.22	10.21	0.07
	11	382	580	10.20	11.22	10.22	0.14
	12	386	591	10.20	11.73	10.24	0.16
3/4	9	392	49	10.20	11.73	10.94	0.38
	10	397	27	10.20	10.71	10.34	0.23
	11	402	71	10.20	11.73	10.91	0.35
	12	407	339	10.20	12.24	10.25	0.19
4/5	9	401	99	10.20	11.73	10.95	0.62
	10	406	26	11.22	12.24	11.75	0.51
	11	410	46	10.20	12.76	11.46	0.58
	12	414	73	11.22	12.24	11.32	0.21
5/6	9	418	2	11.22	11.73	11.48	0.36
	10	423	4	11.73	13.27	12.12	0.77
	11	427	81	11.22	12.76	11.79	0.22
	12	432	N/A	N/A	N/A	N/A	N/A

Table 3.3.5.2Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.706	0.614		0.498	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.858		0.798	
	2	0.687		0.585	
	3	0.578		0.450	
	4	0.218		0.157	
	5	0.602		0.472	
	6	0.831		0.730	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.905	0.045	0.050	0.866
	2/3	0.920	0.045	0.035	0.888
	3/4	0.943	0.031	0.026	0.917
	4/5	0.947	0.036	0.017	0.926
	5/6	0.965	0.019	0.015	0.951

Table 3.3.5.2Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.706	0.613		0.496	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.853		0.786	
	2	0.718		0.624	
	3	0.589		0.463	
	4	0.239		0.173	
	5	0.608		0.477	
	6	0.812		0.696	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.909	0.042	0.048	0.873
	2/3	0.917	0.046	0.036	0.884
	3/4	0.941	0.032	0.027	0.915
	4/5	0.947	0.036	0.018	0.926
	5/6	0.968	0.018	0.014	0.954

Table 3.3.5.2Liii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.705	0.611		0.491	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.824		0.750	
	2	0.759		0.669	
	3	0.509		0.389	
	4	0.273		0.200	
	5	0.592		0.468	
	6	0.816		0.702	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.905	0.049	0.046	0.866
	2/3	0.922	0.039	0.039	0.888
	3/4	0.940	0.038	0.022	0.915
	4/5	0.944	0.036	0.020	0.924
	5/6	0.967	0.018	0.015	0.953

Table 3.3.5.2Liv

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.713	0.618		0.492	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.797		0.716	
	2	0.792		0.711	
	3	0.523		0.403	
	4	0.285		0.209	
	5	0.614		0.489	
	6	0.811		0.688	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.906	0.052	0.041	0.868
	2/3	0.921	0.038	0.041	0.886
	3/4	0.942	0.038	0.020	0.919
	4/5	0.947	0.033	0.020	0.928
	5/6	0.973	0.015	0.012	0.960

3.3.5.3 Writing 9-12

3.3.5.3i Writing 9-12 A

Please note that this section 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.

Figure 3.3.5.3Ci
Raw Scores: Writ 9-12 A S401 Online

Figure 3.3.5.3Di
Scale Scores: Writ 9-12 A S401 Online
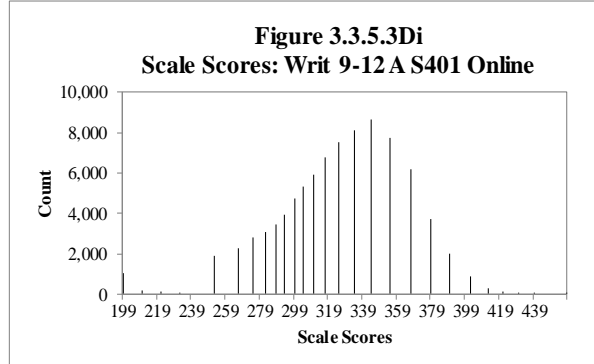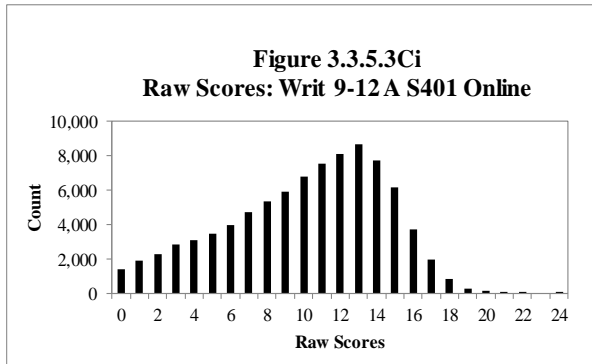
Table 3.3.5.3Ci

Raw Score Descriptive Statistics: Writ 9-12 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	38,220	0	24	9.01	4.62
10	23,565	0	21	10.39	4.02
11	15,956	0	22	11.23	3.74
12	9,142	0	21	11.70	3.63
Total	86,883	0	24	10.08	4.33

Table 3.3.5.3Di

Scale Score Descriptive Statistics: Writ 9-12 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	38,220	199	458	314.20	39.38
10	23,565	210	430	325.30	33.54
11	15,956	221	439	332.05	32.00
12	9,142	232	430	336.06	31.63
Total	86,883	199	458	322.79	36.73

Figure 3.3.5.3Ei
Proficiency Level: Writ 9-12 A S401 Online

Table 3.3.5.3Ei

Proficiency Level Distribution: Writ 9-12 A S401 Online

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	9,907	25.92%	4,212	17.87%	3,443	21.58%	2,915	31.89%	20,477	23.57%
2	12,043	31.51%	9,009	38.23%	5,833	36.56%	1,944	21.26%	28,829	33.18%
3	13,887	36.33%	9,464	40.16%	6,383	40.00%	4,061	44.42%	33,795	38.90%
4	2,372	6.21%	880	3.73%	297	1.86%	222	2.43%	3,771	4.34%
5	11	0.03%	0	0.00%	0	0.00%	0	0.00%	11	0.01%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	38,220	100.00%	23,565	100.00%	15,956	100.00%	9,142	100.00%	86,883	100.00%

Figure 3.3.5.3Hi
Test Characteristic Curve: Writ 9-12 A S401 Online

Figure 3.3.5.3Ii
Test Information Function: Writ 9-12 A S401 Online
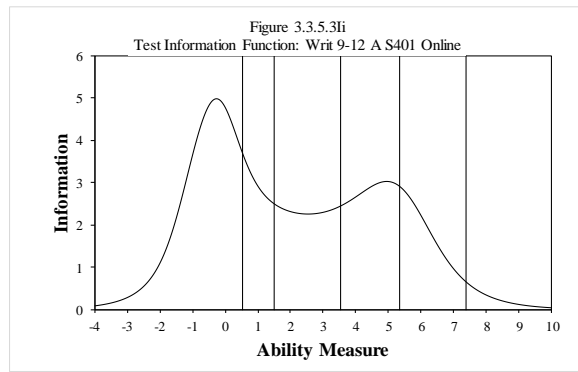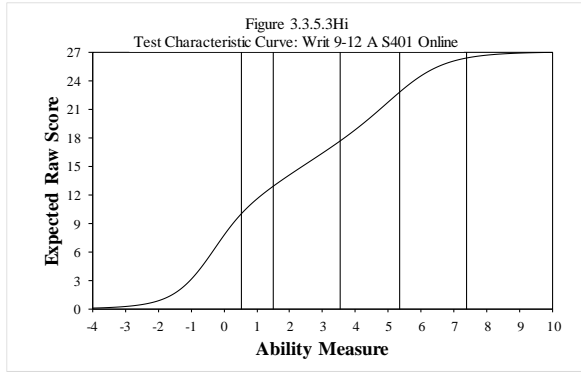
Table 3.3.5.3Ji

Reliability: Writ 9-12 A S401 Online

Reliability	No. of Students	No. of Tasks	Response Modes		Cronbach's Alpha	SEM
			Hand-written (HW)	Keyboarded (KB)		
	86,883	3			.848	1.688
Interrater Reliability	Task	Mode of Response	No. in Sample	% AG	% AD	% NA
	1	HW	802	98	2	0
		KB	43,556	93	7	0
	2	HW	630	100	0	0
		KB	43,410	95	5	0
	3	HW	826	99	1	0
		KB	43,321	94	6	0

Table 3.3.5.3Ki
Conditional Standard Error of Measurement at Cut Scores: Writ 9-12 A S401 Online
n/a

Table 3.3.5.3Li
Accuracy and Consistency of Classification Indices: Writ 9-12 A S401 Online
n/a

3.3.5.3ii *Writing 9-12 B/C*

Please note that this section 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.

Figure 3.3.5.3Cii
**Raw Scores: Writ 9-12 B/C S401 Online**


Figure 3.3.5.3Dii
Scale Scores: Writ 9-12 B/C S401 Online
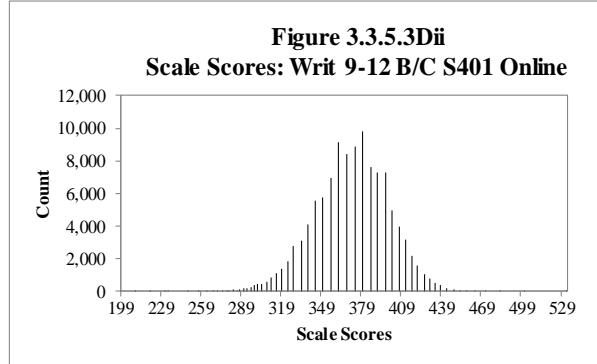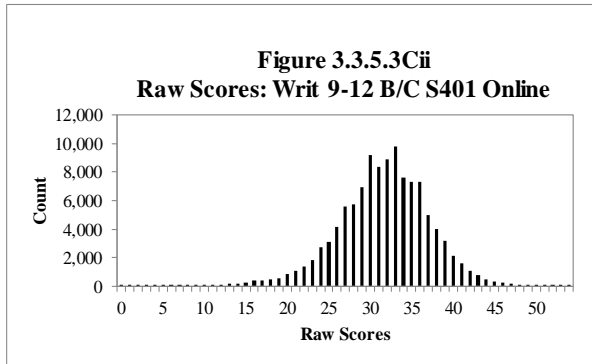
Table 3.3.5.3Cii

Raw Score Descriptive Statistics: Writ 9-12 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	43,685	0	53	31.36	5.61
10	30,962	0	53	31.25	5.67
11	22,723	0	54	31.93	5.70
12	16,772	0	52	31.87	5.84
Total	114,142	0	54	31.52	5.69

Table 3.3.5.3Dii

Scale Score Descriptive Statistics: Writ 9-12 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	43,685	199	501	370.30	29.74
10	30,962	210	501	369.75	29.92
11	22,723	221	533	373.44	30.11
12	16,772	232	483	373.20	30.60
Total	114,142	199	533	371.20	30.03

Figure 3.3.5.3Eii
Proficiency Level: Writ 9-12 B/C S401 Online

Table 3.3.5.3Eii

Proficiency Level Distribution: Writ 9-12 B/C S401 Online

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	239	0.55%	378	1.22%	454	2.00%	741	4.42%	1,812	1.59%
2	1,709	3.91%	2,023	6.53%	2,087	9.18%	1,740	10.37%	7,559	6.62%
3	22,749	52.08%	17,999	58.13%	12,797	56.32%	11,062	65.96%	64,607	56.60%
4	18,257	41.79%	10,349	33.42%	7,239	31.86%	3,153	18.80%	38,998	34.17%
5	725	1.66%	208	0.67%	144	0.63%	76	0.45%	1,153	1.01%
6	6	0.01%	5	0.02%	2	0.01%	0	0.00%	13	0.01%
Total	43,685	100.00%	30,962	100.00%	22,723	100.00%	16,772	100.00%	114,142	100.00%

Figure 3.3.5.3Hii
Test Characteristic Curve: Writ 9-12 B/C S401 Online

Figure 3.3.5.3Iii
Test Information Function: Writ 9-12 B/C S401 Online
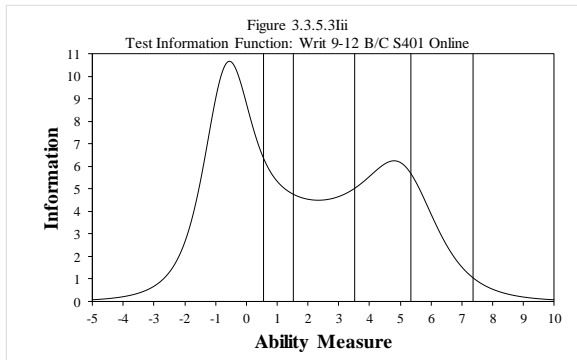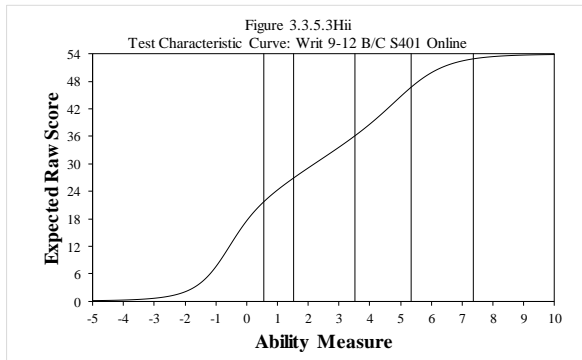
Table 3.3.5.3Jii

Reliability: Writ 9-12 B/C S401 Online

Reliability	No. of Students	No. of Tasks	Response Modes		Cronbach's Alpha	SEM
			Hand-written (HW)	Keyboarded (KB)		
	114,142	3			.881	1.959
Interrater Reliability	Task	Mode of Response	No. in Sample	% AG	% AD	% NA
	1	HW	178	99	1	0
		KB	59,618	95	5	0
	2	HW	198	98	2	0
		KB	60,552	95	5	0
	3	HW	208	99	1	0
		KB	60,474	95	5	0

Table 3.3.5.3Kii
Conditional Standard Error of Measurement at Cut Scores: Writ 9-12 B/C S401 Online
n/a

Table 3.3.5.3Lii
Accuracy and Consistency of Classification Indices: Writ 9-12 B/C S401 Online
n/a

3.3.5.3iii   *Writing 9-12 Across Tiers*

Table 3.3.5.3Aiii
Complete Task Analysis and Summary: Writ 9-12 S401 Online
n/a

Table 3.3.5.3Biii
DIF Analysis and Summary: Writ 9-12 S401 Online
n/a

Figure 3.3.5.3Ciii
Raw Scores: Writ 9-12 S401 Online

Figure 3.3.5.3Diii
Scale Scores: Writ 9-12 S401 Online
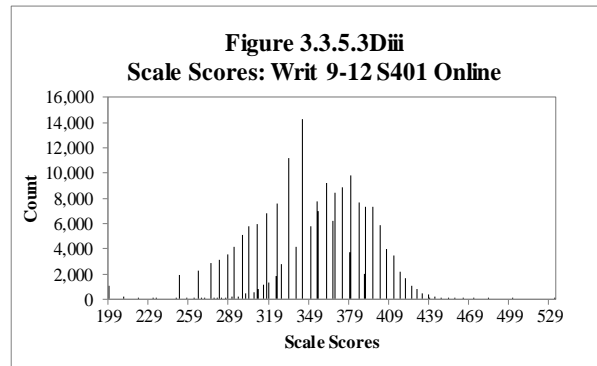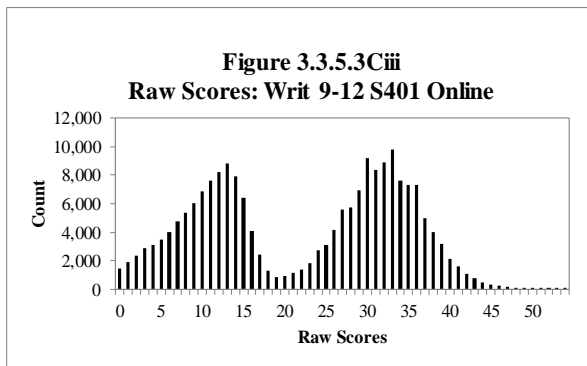
Table 3.3.5.3Ciii

Raw Score Descriptive Statistics: Writ 9-12 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	81,905	0	53	20.93	12.29
10	54,527	0	53	22.23	11.49
11	38,679	0	54	23.39	11.34
12	25,914	0	52	24.75	10.94
Total	201,025	0	54	22.25	11.80

Table 3.3.5.3Diii

Scale Score Descriptive Statistics: Writ 9-12 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	81,905	199	501	344.12	44.48
10	54,527	210	501	350.54	38.46
11	38,679	221	533	356.37	37.02
12	25,914	232	483	360.10	35.70
Total	201,025	199	533	350.28	40.87

Figure 3.3.5.3Eiii
Proficiency Level: Writ 9-12 S401 Online

Table 3.3.5.3Eiii

Proficiency Level Distribution: Writ 9-12 S401 Online

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	10,146	12.39%	4,590	8.42%	3,897	10.08%	3,656	14.11%	22,289	11.09%
2	13,752	16.79%	11,032	20.23%	7,920	20.48%	3,684	14.22%	36,388	18.10%
3	36,636	44.73%	27,463	50.37%	19,180	49.59%	15,123	58.36%	98,402	48.95%
4	20,629	25.19%	11,229	20.59%	7,536	19.48%	3,375	13.02%	42,769	21.28%
5	736	0.90%	208	0.38%	144	0.37%	76	0.29%	1,164	0.58%
6	6	0.01%	5	0.01%	2	0.01%	0	0.00%	13	0.01%
Total	81,905	100.00%	54,527	100.00%	38,679	100.00%	25,914	100.00%	201,025	100.00%

Table 3.3.5.3Fiii
Raw Score to Scale Score Conversion: Writ 9-12 S401 Online
n/a

Table 3.3.5.3Giii
Equating Summary: Writ 9-12 S401 Online
n/a

Figure 3.3.5.3Hiii
Test Characteristic Curve: Writ 9-12 S401 Online

Figure 3.3.5.3Iiii
Test Information Function: Writ 9-12 S401 Online
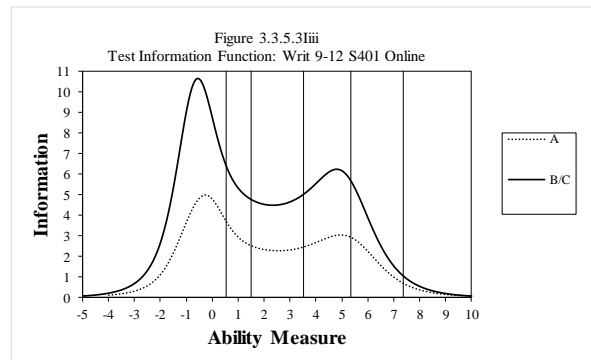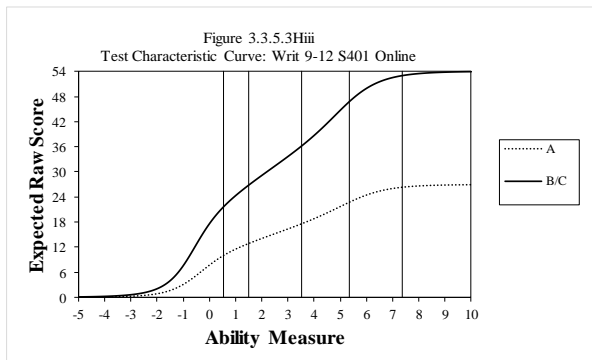
Table 3.3.5.3Jiii

Reliability: Writ 9-12 Weighted Reliability S401 Online

Tiers	No. of Students	Reliability	Weighted Reliability
A	86,883	0.848	0.867
B/C	114,142	0.881	

Table 3.3.5.3Kiii

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

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 3.3.5.3Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.739	0.644		0.483	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.825		0.711	
	2	0.627		0.501	
	3	0.754		0.676	
	4	0.740		0.646	
	5	-		0.073	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.952	0.020	0.028	0.931
	2/3	0.915	0.032	0.053	0.883
	3/4	0.878	0.064	0.057	0.831
	4/5	0.991	0.009	0.000	0.988
	5/6	1.000	0.000	0.000	1.000

Table 3.3.5.3Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.764	0.675		0.501	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.759		0.608	
	2	0.688		0.566	
	3	0.783		0.729	
	4	0.781		0.668	
	5	-		0.102	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.957	0.019	0.024	0.936
	2/3	0.905	0.033	0.061	0.870
	3/4	0.904	0.058	0.038	0.867
	4/5	0.996	0.004	0.000	0.995
	5/6	1.000	0.000	0.000	1.000

Table 3.3.5.3Liii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.761	0.673		0.503	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.789		0.652	
	2	0.679		0.559	
	3	0.781		0.726	
	4	0.771		0.654	
	5	-		0.097	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.954	0.020	0.027	0.932
	2/3	0.906	0.034	0.060	0.870
	3/4	0.904	0.059	0.037	0.868
	4/5	0.996	0.004	0.000	0.995
	5/6	1.000	0.000	0.000	1.000

Table 3.3.5.3Liv

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.779	0.692		0.502	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.868		0.760	
	2	0.519		0.396	
	3	0.852		0.811	
	4	0.691		0.540	
	5	-		0.371	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.948	0.016	0.036	0.927
	2/3	0.912	0.046	0.042	0.873
	3/4	0.920	0.045	0.035	0.884
	4/5	0.997	0.003	0.000	0.997
	5/6	N/A	N/A	N/A	N/A

3.3.5.4   Speaking 9-12

3.3.5.4i   Speaking 9-12 Pre-A

Please note that this section 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.
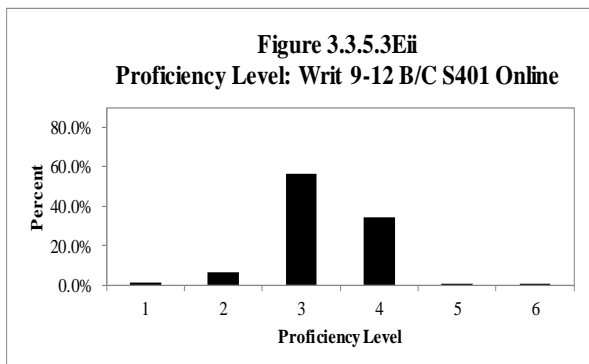

Figure 3.3.5.4Ci
Raw Scores: Spek 9-12 Pre-A S401 Online

Figure 3.3.5.4Di
Scale Scores: Spek 9-12 Pre-A S401 Online
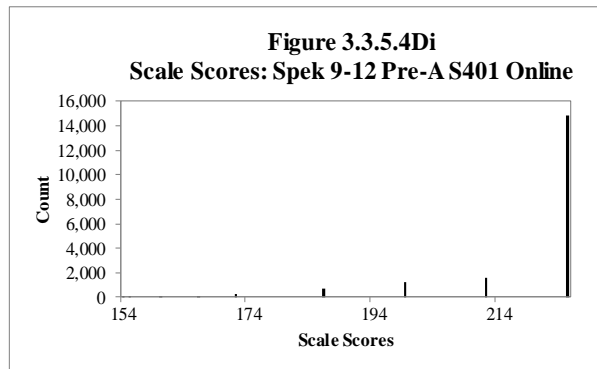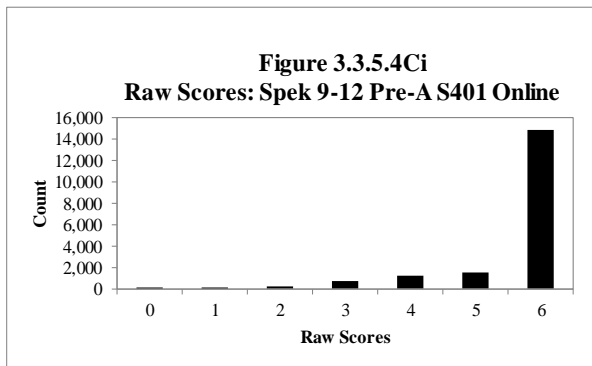
Table 3.3.5.4Ci

Raw Score Descriptive Statistics: Spek 9-12 Pre-A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	6,708	0	6	5.30	1.23
10	5,520	0	6	5.66	0.87
11	3,987	0	6	5.75	0.75
12	2,548	0	6	5.82	0.65
Total	18,763	0	6	5.57	0.99

Table 3.3.5.4Di

Scale Score Descriptive Statistics: Spek 9-12 Pre-A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	6,708	154	225	215.86	15.91
10	5,520	160	225	220.55	11.15
11	3,987	166	225	221.85	9.44
12	2,548	172	225	222.69	7.90
Total	18,763	154	225	219.44	12.73

Figure 3.3.5.4Ei
Proficiency Level: Spek 9-12 Pre-A S401 Online

Table 3.3.5.4Ei

Proficiency Level Distribution: Spek 9-12 Pre-A S401 Online

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	6,708	100.00%	5,520	100.00%	3,987	100.00%	2,548	100.00%	18,763	100.00%
Total	6,708	100.00%	5,520	100.00%	3,987	100.00%	2,548	100.00%	18,763	100.00%

Table 3.3.5.4Gi
Equating Summary: Spek 9-12 Pre-A S401 Online
n/a

Figure 3.3.5.4Hi
Test Characteristic Curve: Spek 9-12 Pre-A S401 Online

Figure 3.3.5.4Ii
Test Information Function: Spek 9-12 Pre-A S401 Online
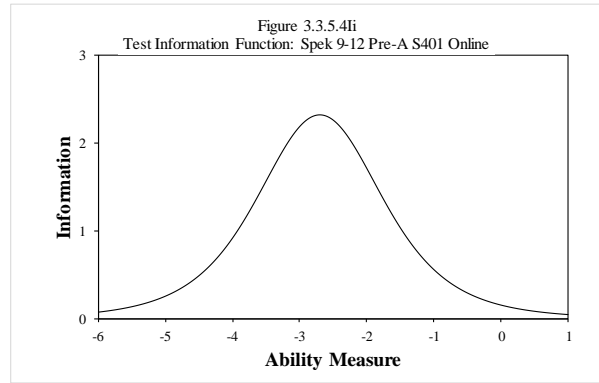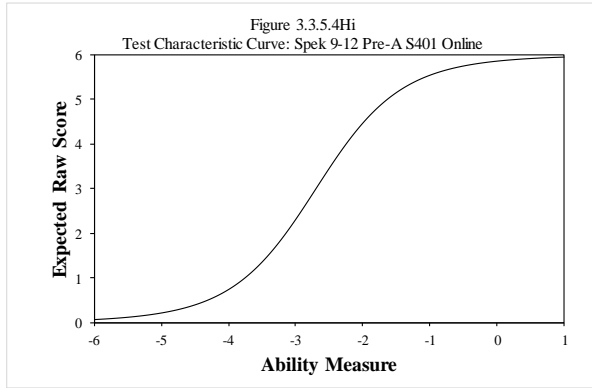
Table 3.3.5.4Ji

Reliability: Spek 9-12 Pre-A S401 Online

Reliability				
	No. of Students	**No. of Tasks**	**Cronbach's Alpha**	**SEM**
	18,763	3	.652	0.584
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	11,806	97	3	0
	2	11,474	98	2	0
	3	11,402	98	2	0

Table 3.3.5.4Ki
Conditional Standard Error of Measurement at Cut Scores: Spek 9-12 Pre-A S401 Online
n/a

Table 3.3.5.4Li
Accuracy and Consistency of Classification Indices: Spek 9-12 Pre-A S401 Online
n/a

3.3.5.4ii   *Speaking 9-12 A*

Please note that this section 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.

Figure 3.3.5.4Cii
Raw Scores: Spek 9-12 A S401 Online

Figure 3.3.5.4Dii
Scale Scores: Spek 9-12 A S401 Online
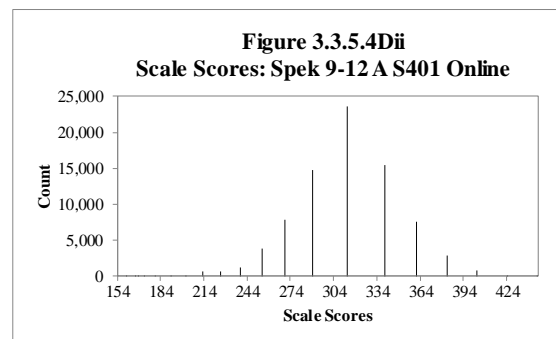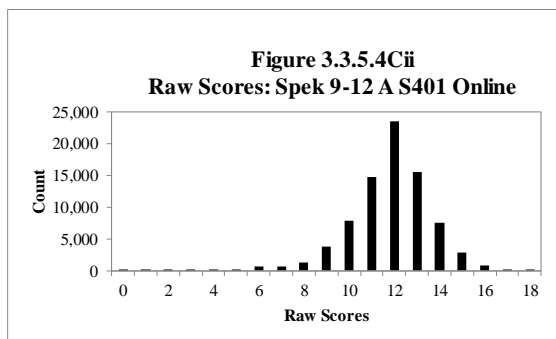
Table 3.3.5.4Cii

Raw Score Descriptive Statistics: Spek 9-12 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	38,370	0	18	11.55	1.98
10	21,147	0	18	11.89	1.66
11	8,122	0	17	11.80	1.55
12	12,300	0	18	12.49	1.50
Total	79,939	0	18	11.81	1.82

Table 3.3.5.4Dii

Scale Score Descriptive Statistics: Spek 9-12 A S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	38,370	154	445	306.16	39.67
10	21,147	160	445	312.46	35.28
11	8,122	166	424	310.26	33.16
12	12,300	172	445	325.62	33.43
Total	79,939	154	445	311.24	37.59

Figure 3.3.5.4Eii
Proficiency Level: Spek 9-12 A S401 Online

Table 3.3.5.4Eii

Proficiency Level Distribution: Spek 9-12 A S401 Online

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	16,247	42.34%	7,556	35.73%	3,070	37.80%	2,623	21.33%	29,496	36.90%
2	10,563	27.53%	6,592	31.17%	2,677	32.96%	6,919	56.25%	26,751	33.46%
3	11,151	29.06%	6,740	31.87%	2,302	28.34%	2,708	22.02%	22,901	28.65%
4	397	1.03%	259	1.22%	73	0.90%	50	0.41%	779	0.97%
5	12	0.03%	0	0.00%	0	0.00%	0	0.00%	12	0.02%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	38,370	100.00%	21,147	100.00%	8,122	100.00%	12,300	100.00%	79,939	100.00%

Table 3.3.5.4Gii
Equating Summary: Spek 9-12 A S401 Online
n/a

Figure 3.3.5.4Hii
Test Characteristic Curve: Spek 9-12 A S401 Online

Figure 3.3.5.4Iii
Test Information Function: Spek 9-12 A S401 Online
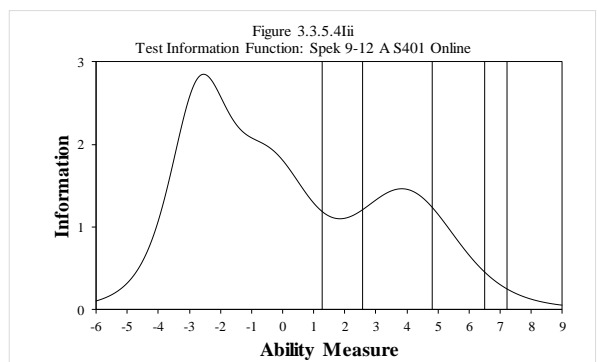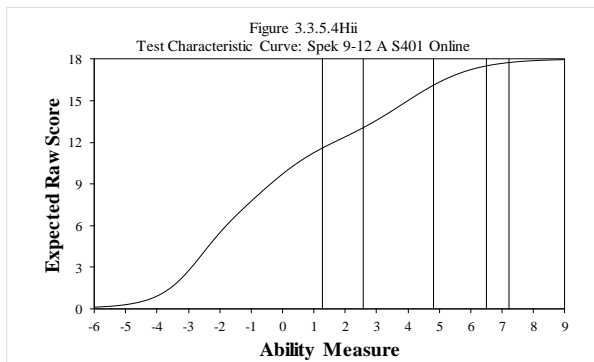
Table 3.3.5.4Jii

Reliability: Spek 9-12 A S401 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	79,939	6	.632		1.103
Interrater Reliability	Task	No. in Sample	% EX	% AD	% NA
	1	50,202	99	1	0
	2	50,202	82	17	0
	3	49,277	99	1	0
	4	49,276	82	17	1
	5	49,694	99	1	0
	6	49,694	79	20	1

Table 3.3.5.4Kii
Conditional Standard Error of Measurement at Cut Scores: Spek 9-12 A S401 Online
n/a

Table 3.3.5.4Lii
Accuracy and Consistency of Classification Indices: Spek 9-12 A S401 Online
n/a

3.3.5.4iii Speaking 9-12 B/C

Please note that this section 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.

Figure 3.3.5.4Ciii
Raw Scores: Spek 9-12 B/C S401 Online

Figure 3.3.5.4Diii
Scale Scores: Spek 9-12 B/C S401 Online
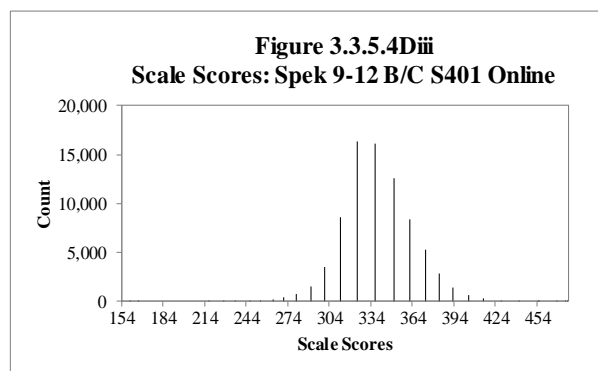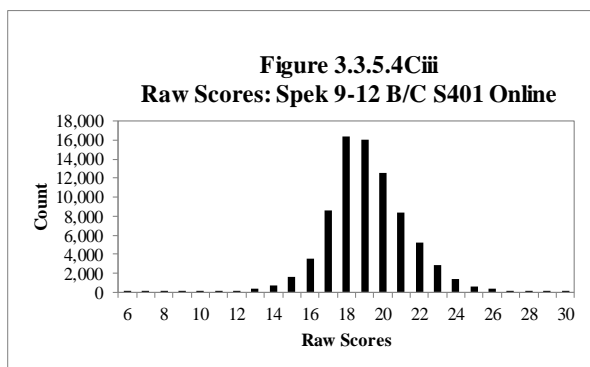
Table 3.3.5.4Ciii

Raw Score Descriptive Statistics: Spek 9-12 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	27,601	6	30	19.26	2.20
10	21,530	6	29	19.08	2.16
11	21,445	6	30	18.94	2.23
12	8,571	8	30	19.60	2.25
Total	79,147	6	30	19.16	2.21

Table 3.3.5.4Diii

Scale Score Descriptive Statistics: Spek 9-12 B/C S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	27,601	154	468	339.93	26.03
10	21,530	160	454	337.77	25.62
11	21,445	166	474	336.05	26.42
12	8,571	227	476	343.96	26.45
Total	79,147	154	476	338.73	26.18

Figure 3.3.5.4Eiii
Proficiency Level: Spek 9-12 B/C S401 Online

Table 3.3.5.4Eiii

Proficiency Level Distribution: Spek 9-12 B/C S401 Online

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	431	1.56%	811	3.77%	1,007	4.70%	465	5.43%	2,714	3.43%
2	9,966	36.11%	7,973	37.03%	8,528	39.77%	4,029	47.01%	30,496	38.53%
3	16,261	58.91%	12,144	56.41%	11,625	54.21%	3,982	46.46%	44,012	55.61%
4	919	3.33%	597	2.77%	273	1.27%	92	1.07%	1,881	2.38%
5	22	0.08%	5	0.02%	10	0.05%	2	0.02%	39	0.05%
6	2	0.01%	0	0.00%	2	0.01%	1	0.01%	5	0.01%
Total	27,601	100.00%	21,530	100.00%	21,445	100.00%	8,571	100.00%	79,147	100.00%

Table 3.3.5.4Giii
Equating Summary: Spek 9-12 B/C S401 Online
n/a

Figure 3.3.5.4Hiii
Test Characteristic Curve: Spek 9-12 B/C S401 Online

Figure 3.3.5.4Iiii
Test Information Function: Spek 9-12 B/C S401 Online
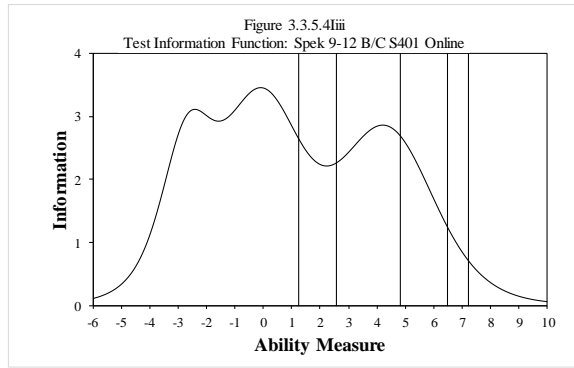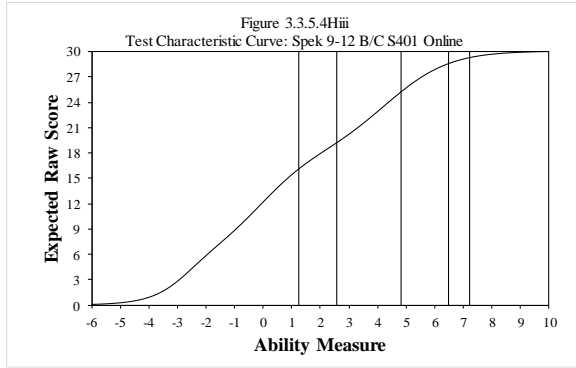
Table 3.3.5.4Jiii

Reliability: Spek 9-12 B/C S401 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	79,147	6	.686		1.240
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	49,278	82	17	0
	2	49,280	82	18	0
	3	49,070	77	22	1
	4	49,070	78	22	0
	5	49,012	75	25	0
	6	49,012	75	24	1

Table 3.3.5.4Kiii
Conditional Standard Error of Measurement at Cut Scores: Spek 9-12 B/C S401 Online
n/a

Table 3.3.5.4Liii
Accuracy and Consistency of Classification Indices: Spek 9-12 B/C S401 Online
n/a

3.3.5.4iv   *Speaking 9-12 Across Tiers*

Please note that this section 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.
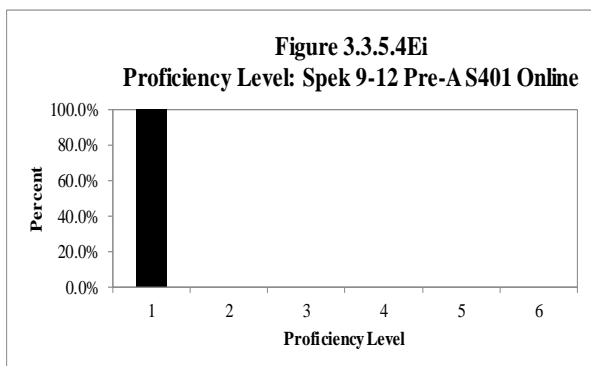
Table 3.3.5.4Biv
DIF Analysis and Summary: Spek 9-12 S401 Online
n/a

Figure 3.3.5.4Civ
Raw Scores: Spek 9-12 S401 Online

Figure 3.3.5.4Div
Scale Scores: Spek 9-12 S401 Online
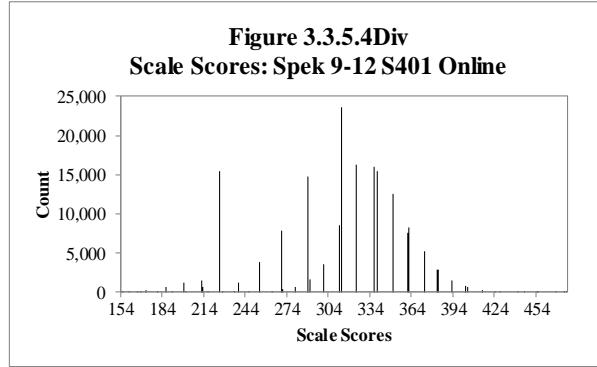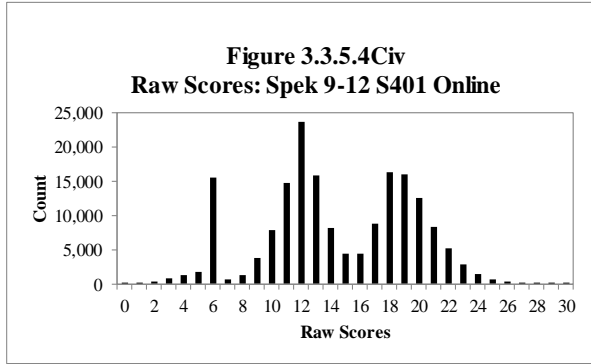
Table 3.3.5.4Civ

Raw Score Descriptive Statistics: Spek 9-12 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	72,679	0	30	13.90	4.97
10	48,197	0	29	14.39	4.97
11	33,554	0	30	15.64	5.10
12	23,419	0	30	14.36	4.79
Total	177,849	0	30	14.42	5.01

Table 3.3.5.4Div

Scale Score Descriptive Statistics: Spek 9-12 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	72,679	154	468	310.65	47.72
10	48,197	160	454	313.24	45.90
11	33,554	166	474	316.24	45.18
12	23,419	172	476	321.13	45.89
Total	177,849	154	476	313.79	46.65

Figure 3.3.5.4Eiv
Proficiency Level: Spek 9-12 S401 Online
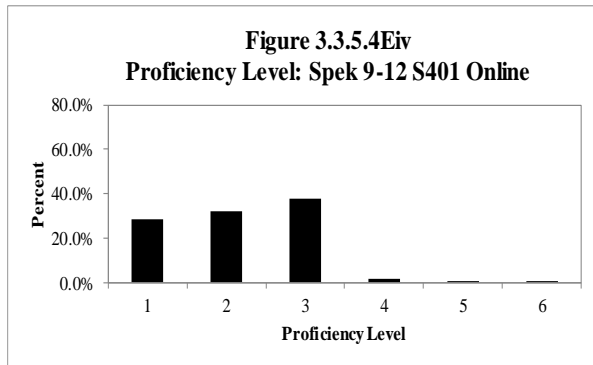
Table 3.3.5.4Eiv

Proficiency Level Distribution: Spek 9-12 S401 Online

Level	Grade 9		Grade 10		Grade 11		Grade 12		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
1	23,386	32.18%	13,887	28.81%	8,064	24.03%	5,636	24.07%	50,973	28.66%
2	20,529	28.25%	14,565	30.22%	11,205	33.39%	10,948	46.75%	57,247	32.19%
3	27,412	37.72%	18,884	39.18%	13,927	41.51%	6,690	28.57%	66,913	37.62%
4	1,316	1.81%	856	1.78%	346	1.03%	142	0.61%	2,660	1.50%
5	34	0.05%	5	0.01%	10	0.03%	2	0.01%	51	0.03%
6	2	0.00%	0	0.00%	2	0.01%	1	0.00%	5	0.00%
Total	72,679	100.00%	48,197	100.00%	33,554	100.00%	23,419	100.00%	177,849	100.00%

Table 3.3.5.4Fiv
Raw Score to Scale Score Conversion: Spek 9-12 S401 Online
n/a

Figure 3.3.5.4Hiv
Test Characteristic Curve: Spek 9-12 S401 Online

Figure 3.3.5.4Iiv
Test Information Function: Spek 9-12 S401 Online
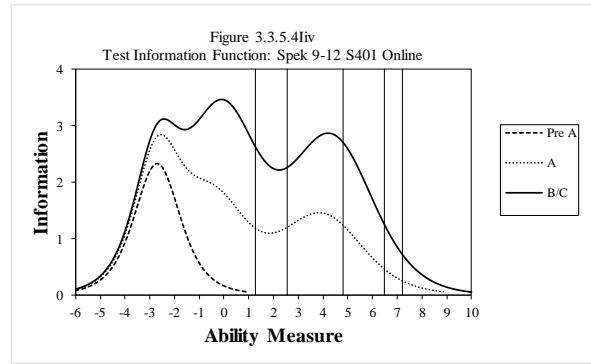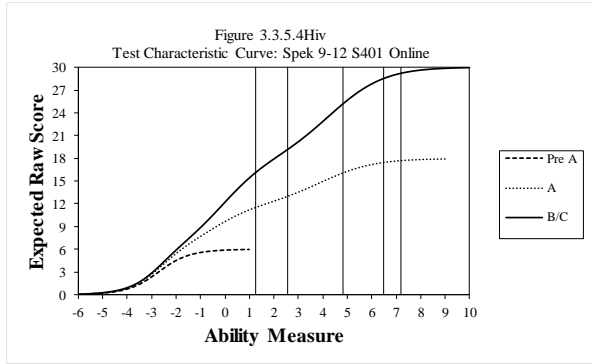
Table 3.3.5.4Jiv

Reliability: Spek 9-12 Weighted Reliability S401 Online

Tiers	No. of Students	Reliability	Weighted Reliability
Pre-A	18,763	0.652	
A	79,939	0.632	0.658
B/C	79,147	0.686	

Table 3.3.5.4Kiv

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

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

Note: Tier Pre-A is not presented as it is not possible for Tier Pre-A students to receive a proficeny level higher than 2.

Table 3.3.5.4Li

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.632	0.526		0.301	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.866		0.716	
	2	0.432		0.342	
	3	0.610		0.554	
	4	-		0.034	
	5	-		-	
	6	-		-	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.869	0.035	0.096	0.810
	2/3	0.752	0.078	0.171	0.674
	3/4	0.981	0.019	0.000	0.956
	4/5	1.000	0.001	0.000	0.999
	5/6	1.000	0.000	0.000	1.000

Table 3.3.5.4Lii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.616	0.516		0.281	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.863		0.685	
	2	0.437		0.356	
	3	0.593		0.545	
	4	-		0.030	
	5	-		-	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.881	0.032	0.087	0.818
	2/3	0.719	0.091	0.190	0.646
	3/4	0.982	0.018	0.000	0.971
	4/5	1.000	0.000	0.000	1.000
	5/6	N/A	N/A	N/A	N/A

Table 3.3.5.4Liii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.581	0.496		0.240	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.917		0.628	
	2	-		0.382	
	3	0.036		0.528	
	4	-		0.015	
	5	-		-	
	6	-		-	

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.906	0.015	0.079	0.825
	2/3	-	-	-	-
	3/4	0.989	0.011	0.000	0.983
	4/5	1.000	0.000	0.000	1.000
	5/6	1.000	0.000	0.000	1.000

Table 3.3.5.4Liv

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.548	0.491		0.236	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.820		0.609	
	2	0.964		0.533	
	3	0.352		0.383	
	4	-		0.010	
	5	-		-	
	6	-		-	

Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.887	0.036	0.077	0.817
	2/3	-	-	-	-
	3/4	0.994	0.006	0.000	0.987
	4/5	1.000	0.000	0.000	1.000
	5/6	1.000	0.000	0.000	1.000

3.4.   Analyses of Composite Scores: Results

3.4.1 Grade: 1

3.4.1.1   Oral Language Composite 1

Figure 3.4.1.1A
Scale Scores: Oral 1 S401 Online

Figure 3.4.1.1B
Proficiency Level: Oral 1 S401 Online
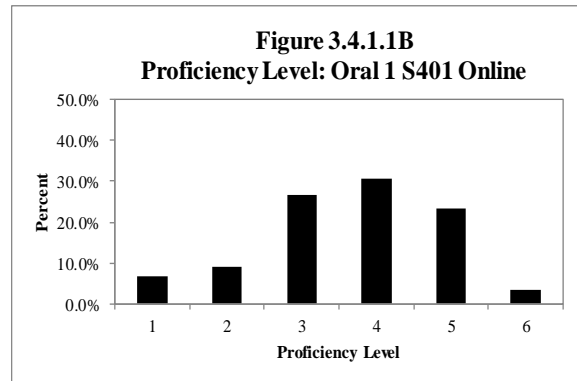
Table 3.4.1.1A

Scale Score Descriptive Statistics: Oral 1 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	118,149	118	414	300.49	44.56
Total	118,149	118	414	300.49	44.56

Table 3.4.1.1B

Proficiency Level Distribution: Oral 1 S401 Online

Level	Grade 1 Count	Grade 1 Percent	Total Count	Total Percent
1	7,882	6.67%	7,882	6.67%
2	10,862	9.19%	10,862	9.19%
3	31,456	26.62%	31,456	26.62%
4	36,343	30.76%	36,343	30.76%
5	27,382	23.18%	27,382	23.18%
6	4,224	3.58%	4,224	3.58%
Total	118,149	100.00%	118,149	100.00%

Table 3.4.1.1C

Reliability: Oral 1 S401 Online

Component	Weight	Variance	Reliability
Listening	0.50	3369.627	0.860
Speaking	0.50	1837.482	0.668
Oral		1994.085	0.864

*Variances from students who had results in all four domains

Table 3.4.1.1D

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.629	0.515		0.366	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.867		0.756	
	2	0.578		0.441	
	3	0.704		0.580	
	4	0.579		0.469	
	5	0.589		0.505	
	6	-		0.140	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.976	0.008	0.017	0.966
	2/3	0.947	0.026	0.027	0.922
	3/4	0.886	0.046	0.069	0.840
	4/5	0.849	0.070	0.081	0.790
	5/6	0.964	0.036	0.000	0.952

3.4.1.2 Literacy Language Composite 1

Figure 3.4.1.2A
Scale Scores: Litr 1 S401 Online

Figure 3.4.1.2B
Proficiency Level: Litr 1 S401 Online

Table 3.4.1.2A

Scale Score Descriptive Statistics: Litr 1 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	146,025	140	415	275.16	30.06
Total	146,025	140	415	275.16	30.06

Table 3.4.1.2B

Proficiency Level Distribution: Litr 1 S401 Online

Level	Grade 1 Count	Grade 1 Percent	Total Count	Total Percent
1	26,314	18.02%	26,314	18.02%
2	66,589	45.60%	66,589	45.60%
3	40,608	27.81%	40,608	27.81%
4	10,268	7.03%	10,268	7.03%
5	1,884	1.29%	1,884	1.29%
6	362	0.25%	362	0.25%
Total	146,025	100.00%	146,025	100.00%

Table 3.4.1.2C

Reliability: Litr 1 S401 Online

Component	Weight	Variance	Reliability
Reading	0.50	1226.883	0.890
Writing	0.50	1122.100	0.865
Literacy		909.410	0.921

*Variances from students who had results in all four domains

Table 3.4.1.2D

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.805	0.723		0.597	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.780		0.678	
	2	0.839		0.779	
	3	0.778		0.696	
	4	0.784		0.674	
	5	0.725		0.577	
	6	0.850		0.660	

Indices at Cut Points		Accuracy			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.925	0.041	0.033	0.892
	2/3	0.915	0.037	0.048	0.879
	3/4	0.973	0.018	0.009	0.962
	4/5	0.993	0.005	0.002	0.991
	5/6	0.999	0.001	0.000	0.998

3.4.1.3 Comprehension Language Composite 1

Figure 3.4.1.3A
Scale Scores: Cphn 1 S401 Online

Figure 3.4.1.3B
Proficiency Level: Cphn 1 S401 Online

Table 3.4.1.3A

Scale Score Descriptive Statistics: Cphn 1 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	135,505	130	417	301.27	35.20
Total	135,505	130	417	301.27	35.20

Table 3.4.1.3B

Proficiency Level Distribution: Cphn 1 S401 Online

Level	Grade 1 Count	Grade 1 Percent	Total Count	Total Percent
1	13,313	9.82%	13,313	9.82%
2	21,693	16.01%	21,693	16.01%
3	29,379	21.68%	29,379	21.68%
4	18,213	13.44%	18,213	13.44%
5	26,346	19.44%	26,346	19.44%
6	26,561	19.60%	26,561	19.60%
Total	135,505	100.00%	135,505	100.00%

Table 3.4.1.3C

Reliability: Cphn 1 S401 Online

Component	Weight	Variance	Reliability
Listening	0.30	3369.627	0.860
Reading	0.70	1226.883	0.890
Comprehension		1253.142	0.913

*Variances from students who had results in all four domains

Table 3.4.1.3D

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.646	0.543		0.445	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.790		0.654	
	2	0.606		0.489	
	3	0.596		0.486	
	4	0.420		0.321	
	5	0.614		0.499	
	6	0.861		0.777	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.952	0.018	0.030	0.932
	2/3	0.921	0.037	0.042	0.888
	3/4	0.903	0.054	0.043	0.866
	4/5	0.908	0.047	0.045	0.872
	5/6	0.940	0.034	0.026	0.915

3.4.1.4   Overall Language Composite 1

Figure 3.4.1.4A
Scale Scores: Over 1 S401 Online

Figure 3.4.1.4B
Proficiency Level: Over 1 S401 Online

Table 3.4.1.4A

Scale Score Descriptive Statistics: Over 1 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	109,820	153	405	282.73	30.64
Total	109,820	153	405	282.73	30.64

Table 3.4.1.4B

Proficiency Level Distribution: Over 1 S401 Online

Level	Grade 1 Count	Grade 1 Percent	Total Count	Total Percent
1	9,061	8.25%	9,061	8.25%
2	30,321	27.61%	30,321	27.61%
3	53,835	49.02%	53,835	49.02%
4	13,518	12.31%	13,518	12.31%
5	2,782	2.53%	2,782	2.53%
6	303	0.28%	303	0.28%
Total	109,820	100.00%	109,820	100.00%

Table 3.4.1.4C

Reliability: Over 1 S401 Online

Component	Weight	Variance	Reliability
Listening	0.15	3369.627	0.860
Reading	0.35	1226.883	0.890
Speaking	0.15	1837.482	0.668
Writing	0.35	1122.100	0.865
Overall Composite		939.072	0.937

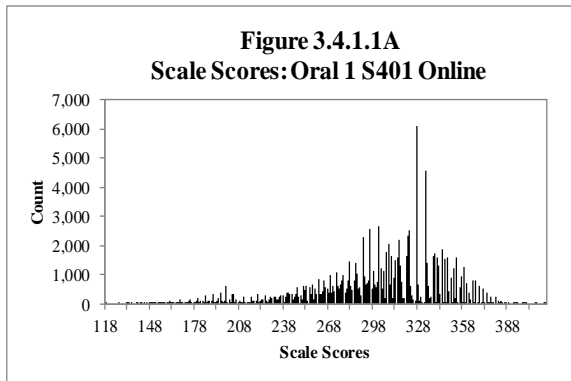*Variances from students who had results in all four domains

Table 3.4.1.4D

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.831	0.761		0.642	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.836		0.730	
	2	0.801		0.723	
	3	0.871		0.828	
	4	0.752		0.646	
	5	0.775		0.649	
	6	0.892		0.685	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.970	0.013	0.017	0.957
	2/3	0.923	0.039	0.038	0.892
	3/4	0.952	0.026	0.022	0.932
	4/5	0.988	0.010	0.003	0.984
	5/6	0.998	0.002	0.000	0.998

3.4.2 Grades: 2-3

3.4.2.1   Oral Language Composite 2-3

Figure 3.4.2.1A
Scale Scores: Oral 2-3 S401 Online

Figure 3.4.2.1B
Proficiency Level: Oral 2-3 S401 Online

Table 3.4.2.1A

Scale Score Descriptive Statistics: Oral 2-3 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	119,993	110	421	297.01	44.26
3	130,821	115	436	314.05	48.10
Total	250,814	110	436	305.90	47.08

Table 3.4.2.1B

Proficiency Level Distribution: Oral 2-3 S401 Online

Level	Grade 2 Count	Grade 2 Percent	Grade 3 Count	Grade 3 Percent	Total Count	Total Percent
1	8,713	7.26%	10,339	7.90%	19,052	7.60%
2	23,839	19.87%	20,719	15.84%	44,558	17.77%
3	45,482	37.90%	43,627	33.35%	89,109	35.53%
4	33,340	27.78%	46,689	35.69%	80,029	31.91%
5	8,217	6.85%	9,008	6.89%	17,225	6.87%
6	402	0.34%	439	0.34%	841	0.34%
Total	119,993	100.00%	130,821	100.00%	250,814	100.00%

Table 3.4.2.1C

Reliability: Oral 2-3 S401 Online

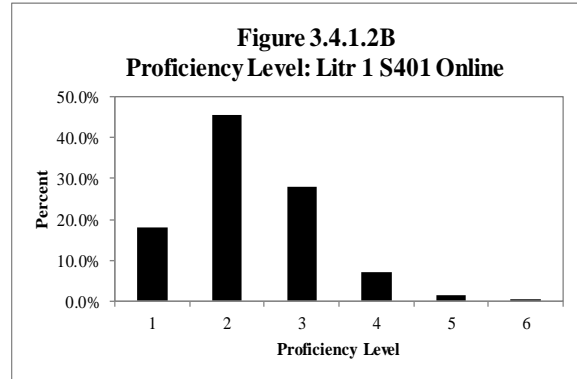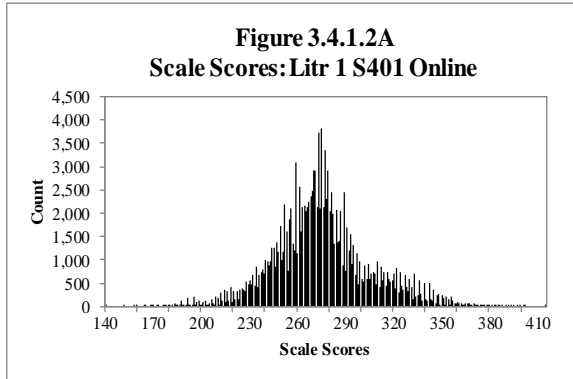| Component	Weight	Variance	Reliability
Listening	0.50	3230.825	0.860
Speaking	0.50	2449.825	0.668
Oral		2279.670	0.861

*Variances from students who had results in all four domains

Table 3.4.2.1Di

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

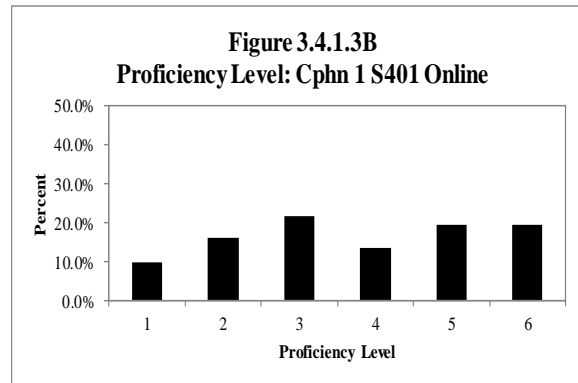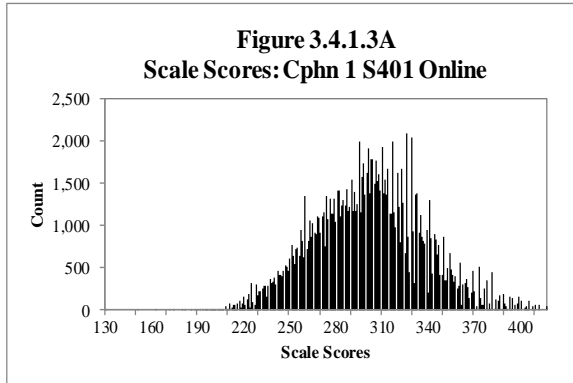| Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.652	0.543		0.372	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.791		0.665	
	2	0.726		0.597	
	3	0.694		0.567	
	4	0.558		0.500	
	5	-		0.205	
	6	-		0.667	
Indices at Cut Points		Accuracy			
			False	False	
	Cut Point	Accuracy	Positives	Negatives	Consistency
	1/2	0.970	0.015	0.015	0.955
	2/3	0.913	0.035	0.052	0.878
	3/4	0.837	0.054	0.109	0.774
	4/5	0.928	0.072	0.000	0.905
	5/6	0.997	0.003	0.000	0.997

Table 3.4.2.1Dii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.636	0.522		0.333	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.814		0.706	
	2	0.712		0.571	
	3	0.634		0.473	
	4	0.590		0.548	
	5	-		0.142	
	6	-		0.053	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.972	0.015	0.013	0.958
	2/3	0.927	0.026	0.047	0.897
	3/4	0.805	0.052	0.143	0.729
	4/5	0.928	0.072	0.000	0.895
	5/6	0.997	0.003	0.000	0.996

3.4.2.2　　　Literacy Language Composite 2-3

Figure 3.4.2.2A
Scale Scores: Litr 2-3 S401 Online

Figure 3.4.2.2B
Proficiency Level: Litr 2-3 S401 Online

Table 3.4.2.2A

Scale Score Descriptive Statistics: Litr 2-3 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	142,299	158	428	303.70	31.07
3	149,242	180	437	322.48	34.09
Total	291,541	158	437	313.32	33.97

Table 3.4.2.2B

Proficiency Level Distribution: Litr 2-3 S401 Online

Level	Grade 2		Grade 3		Total	
	Count	Percent	Count	Percent	Count	Percent
1	11,781	8.28%	11,676	7.82%	23,457	8.05%
2	41,196	28.95%	29,736	19.92%	70,932	24.33%
3	64,002	44.98%	63,691	42.68%	127,693	43.80%
4	22,277	15.66%	36,564	24.50%	58,841	20.18%
5	2,804	1.97%	7,024	4.71%	9,828	3.37%
6	239	0.17%	551	0.37%	790	0.27%
Total	142,299	100.00%	149,242	100.00%	291,541	100.00%

Table 3.4.2.2C

Reliability: Litr 2-3 S401 Online

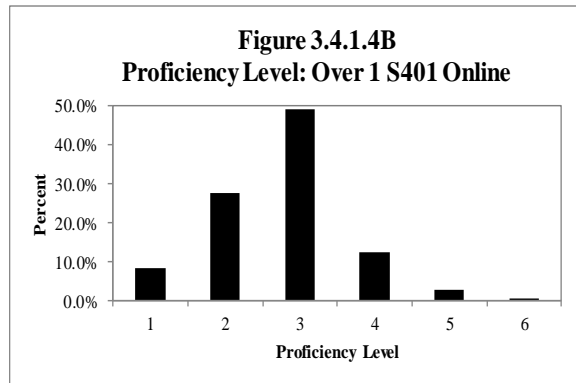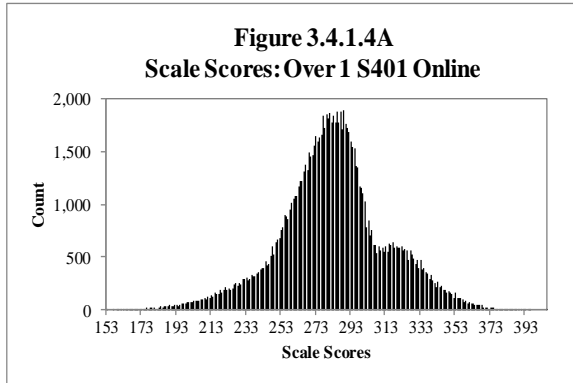| Component	Weight	Variance	Reliability
Reading	0.50	1359.713	0.900
Writing	0.50	1422.881	0.886
Literacy		1140.740	0.935

*Variances from students who had results in all four domains

Table 3.4.2.2Di

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.821	0.747		0.629	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.808		0.695	
	2	0.824		0.748	
	3	0.836		0.785	
	4	0.788		0.695	
	5	0.694		0.520	
	6	0.867		0.457	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.968	0.016	0.016	0.953
	2/3	0.923	0.033	0.045	0.892
	3/4	0.944	0.033	0.023	0.922
	4/5	0.987	0.010	0.003	0.983
	5/6	0.998	0.002	0.000	0.998

Table 3.4.2.2Dii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.799	0.718		0.603	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.847		0.749	
	2	0.783		0.690	
	3	0.829		0.771	
	4	0.767		0.687	
	5	0.650		0.485	
	6	-		0.233	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.974	0.011	0.015	0.962
	2/3	0.937	0.027	0.035	0.912
	3/4	0.924	0.040	0.036	0.893
	4/5	0.967	0.023	0.010	0.954
	5/6	0.996	0.004	0.000	0.996

3.4.2.3 Comprehension Language Composite 2-3

Figure 3.4.2.3A
Scale Scores: Cphn 2-3 S401 Online

Figure 3.4.2.3B
Proficiency Level: Cphn 2-3 S401 Online
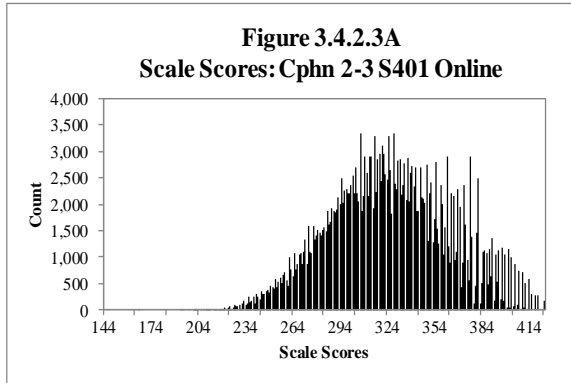
Table 3.4.2.3A

Scale Score Descriptive Statistics: Cphn 2-3 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	129,302	195	424	315.88	34.31
3	136,880	144	424	335.49	39.11
Total	266,182	144	424	325.97	38.14

Table 3.4.2.3B

Proficiency Level Distribution: Cphn 2-3 S401 Online

Level	Grade 2 Count	Grade 2 Percent	Grade 3 Count	Grade 3 Percent	Total Count	Total Percent
1	12,428	9.61%	15,801	11.54%	28,229	10.61%
2	30,102	23.28%	26,030	19.02%	56,132	21.09%
3	32,165	24.88%	29,790	21.76%	61,955	23.28%
4	17,019	13.16%	15,188	11.10%	32,207	12.10%
5	19,311	14.93%	22,268	16.27%	41,579	15.62%
6	18,277	14.14%	27,803	20.31%	46,080	17.31%
Total	129,302	100.00%	136,880	100.00%	266,182	100.00%

Table 3.4.2.3C

Reliability: Cphn 2-3 S401 Online

Component	Weight	Variance	Reliability
Listening	0.30	3230.825	0.860
Reading	0.70	1359.713	0.900
Comprehension		1468.833	0.927

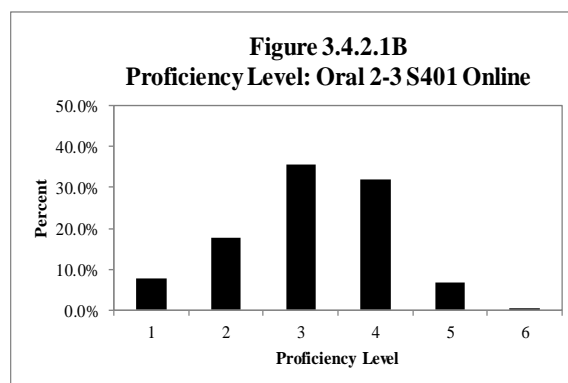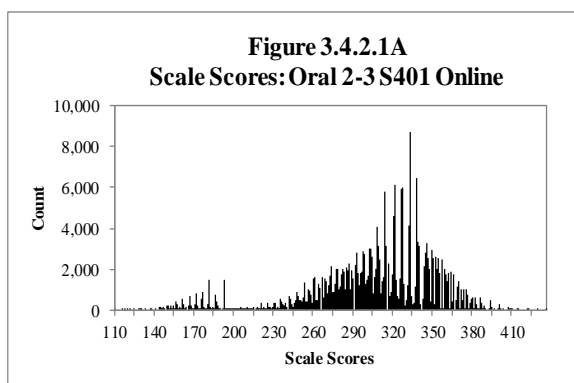*Variances from students who had results in all four domains

Table 3.4.2.3Di

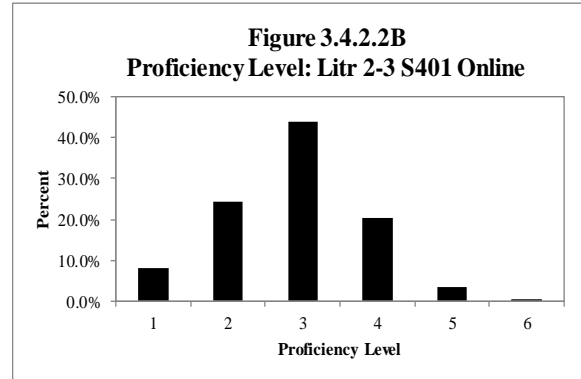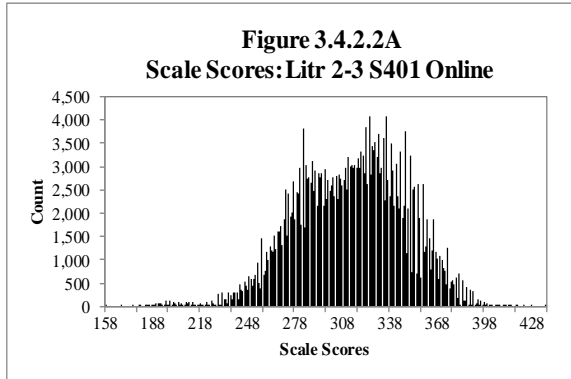Accuracy and Consistency of Classification Indices: Cphn (Grade 2) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.686	0.583		0.489	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.803		0.678	
	2	0.725		0.627	
	3	0.667		0.560	
	4	0.474		0.362	
	5	0.620		0.498	
	6	0.868		0.781	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.957	0.017	0.025	0.939
	2/3	0.918	0.041	0.041	0.885
	3/4	0.915	0.043	0.041	0.882
	4/5	0.930	0.037	0.033	0.901
	5/6	0.956	0.027	0.017	0.937

Table 3.4.2.3Dii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.677	0.577		0.487	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.819		0.712	
	2	0.670		0.559	
	3	0.635		0.524	
	4	0.403		0.303	
	5	0.601		0.480	
	6	0.881		0.806	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.954	0.020	0.027	0.934
	2/3	0.922	0.039	0.039	0.891
	3/4	0.916	0.042	0.042	0.882
	4/5	0.925	0.040	0.036	0.893
	5/6	0.944	0.033	0.023	0.921

3.4.2.4　Overall Language Composite 2-3

Figure 3.4.2.4A
Scale Scores: Over 2-3 S401 Online

Figure 3.4.2.4B
Proficiency Level: Over 2-3 S401 Online

Table 3.4.2.4A

Scale Score Descriptive Statistics: Over 2-3 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	106,741	157	398	301.91	31.90
3	116,535	171	422	320.25	35.44
Total	223,276	157	422	311.48	35.01

Table 3.4.2.4B

Proficiency Level Distribution: Over 2-3 S401 Online

Level	Grade 2 Count	Grade 2 Percent	Grade 3 Count	Grade 3 Percent	Total Count	Total Percent
1	7,127	6.68%	8,857	7.60%	15,984	7.16%
2	27,687	25.94%	19,909	17.08%	47,596	21.32%
3	50,067	46.91%	50,320	43.18%	100,387	44.96%
4	19,608	18.37%	32,718	28.08%	52,326	23.44%
5	2,190	2.05%	4,637	3.98%	6,827	3.06%
6	62	0.06%	94	0.08%	156	0.07%
Total	106,741	100.00%	116,535	100.00%	223,276	100.00%

Table 3.4.2.4C

Reliability: Over 2-3 S401 Online

Component	Weight	Variance	Reliability
Listening	0.15	3230.825	0.860
Reading	0.35	1359.713	0.900
Speaking	0.15	2449.825	0.668
Writing	0.35	1422.881	0.886
Overall Composite		1225.721	0.947

*Variances from students who had results in all four domains

Table 3.4.2.4Di

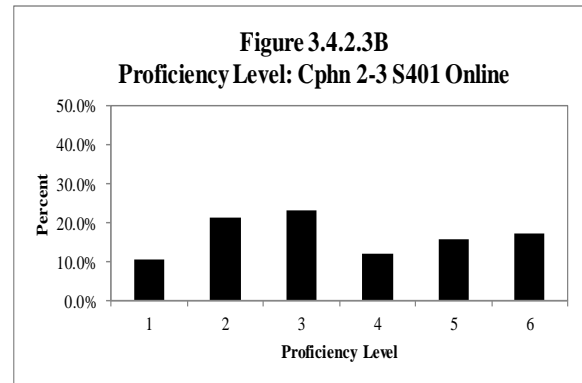Accuracy and Consistency of Classification Indices: Over (Grade 2) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.833	0.772		0.662	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.828		0.738	
	2	0.845		0.776	
	3	0.862		0.816	
	4	0.751		0.689	
	5	-		0.499	
	6	-		1.000	

Indices at Cut Points		Accuracy			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.977	0.012	0.011	0.968
	2/3	0.936	0.028	0.035	0.911
	3/4	0.939	0.031	0.030	0.914
	4/5	0.979	0.021	0.000	0.979
	5/6	0.999	0.001	0.000	1.000

Table 3.4.2.4Dii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.810	0.744		0.632	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.871		0.807	
	2	0.796		0.706	
	3	0.860		0.803	
	4	0.741		0.693	
	5	-		0.326	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.980	0.010	0.011	0.972
	2/3	0.949	0.025	0.026	0.929
	3/4	0.920	0.033	0.047	0.888
	4/5	0.959	0.041	0.000	0.955
	5/6	0.999	0.001	0.000	1.000

3.4.3 Grades: 4-5

3.4.3.1   Oral Language Composite 4-5

Figure 3.4.3.1A
Scale Scores: Oral 4-5 S401 Online

Figure 3.4.3.1B
Proficiency Level: Oral 4-5 S401 Online

Table 3.4.3.1A

Scale Score Descriptive Statistics: Oral 4-5 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	85,468	152	467	356.74	43.00
5	62,948	134	475	359.43	49.58
Total	148,416	134	475	357.89	45.93

Table 3.4.3.1B

Proficiency Level Distribution: Oral 4-5 S401 Online

Level	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	3,409	3.99%	4,097	6.51%	7,506	5.06%
2	5,026	5.88%	4,714	7.49%	9,740	6.56%
3	14,796	17.31%	12,041	19.13%	26,837	18.08%
4	37,822	44.25%	24,663	39.18%	62,485	42.10%
5	20,380	23.85%	14,607	23.20%	34,987	23.57%
6	4,035	4.72%	2,826	4.49%	6,861	4.62%
Total	85,468	100.00%	62,948	100.00%	148,416	100.00%

Table 3.4.3.1C

Reliability: Oral 4-5 S401 Online

Component	Weight	Variance	Reliability
Listening	0.50	3292.827	0.850
Speaking	0.50	2190.505	0.723
Oral		2252.993	0.878

*Variances from students who had results in all four domains

Table 3.4.3.1Di

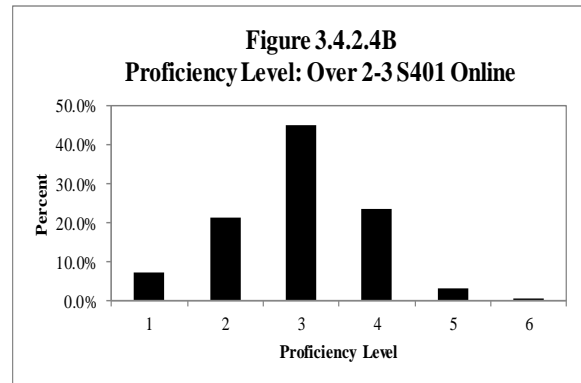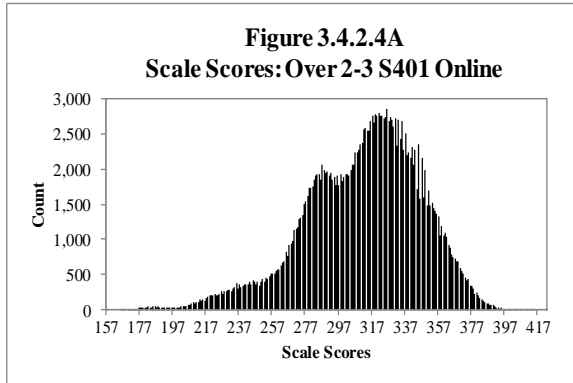Accuracy and Consistency of Classification Indices: Oral (Grade 4) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.640	0.537		0.356	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.881		0.787	
	2	0.623		0.482	
	3	0.685		0.548	
	4	0.735		0.614	
	5	0.507		0.445	
	6	-		0.138	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.987	0.004	0.009	0.982
	2/3	0.969	0.015	0.016	0.955
	3/4	0.918	0.036	0.046	0.883
	4/5	0.810	0.052	0.138	0.749
	5/6	0.953	0.047	0.000	0.940

Table 3.4.3.1Dii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.614	0.508		0.337	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.873		0.779	
	2	0.577		0.437	
	3	0.665		0.526	
	4	0.649		0.540	
	5	0.507		0.439	
	6	-		0.122	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.980	0.008	0.013	0.971
	2/3	0.957	0.021	0.022	0.937
	3/4	0.906	0.036	0.058	0.868
	4/5	0.808	0.073	0.120	0.747
	5/6	0.955	0.045	0.000	0.940

3.4.3.2 Literacy Language Composite 4-5

Figure 3.4.3.2A
Scale Scores: Litr 4-5 S401 Online

Figure 3.4.3.2B
Proficiency Level: Litr 4-5 S401 Online

Table 3.4.3.2A

Scale Score Descriptive Statistics: Litr 4-5 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	59,348	182	443	332.12	31.95
5	43,490	188	457	337.83	35.74
Total	102,838	182	457	334.54	33.72

Table 3.4.3.2B

Proficiency Level Distribution: Litr 4-5 S401 Online

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	5,016	8.45%	4,797	11.03%	9,813	9.54%
2	7,862	13.25%	6,757	15.54%	14,619	14.22%
3	31,137	52.47%	19,619	45.11%	50,756	49.36%
4	13,569	22.86%	10,281	23.64%	23,850	23.19%
5	1,463	2.47%	1,717	3.95%	3,180	3.09%
6	301	0.51%	319	0.73%	620	0.60%
Total	59,348	100.00%	43,490	100.00%	102,838	100.00%

Table 3.4.3.2C

Reliability: Litr 4-5 S401 Online

Component	Weight	Variance	Reliability
Reading	0.50	1025.399	0.880
Writing	0.50	1733.272	0.897
Literacy		1106.535	0.932

*Variances from students who had results in all four domains

Table 3.4.3.2Di

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

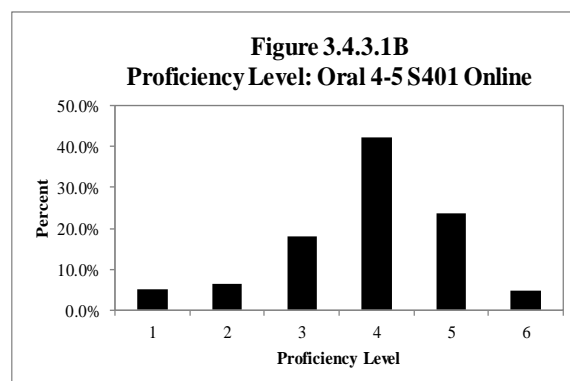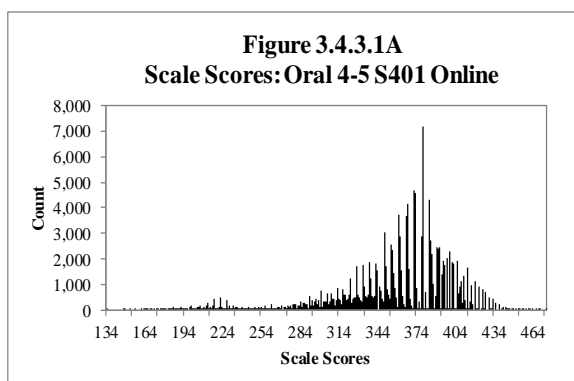| Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.799	0.731		0.589	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.825		0.792	
	2	0.711		0.592	
	3	0.874		0.821	
	4	0.695		0.626	
	5	-		0.231	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.972	0.015	0.013	0.966
	2/3	0.942	0.030	0.027	0.924
	3/4	0.900	0.041	0.059	0.866
	4/5	0.970	0.030	0.000	0.973
	5/6	0.995	0.005	0.000	1.000

Table 3.4.3.2Dii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.779	0.700		0.575	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.870		0.798	
	2	0.703		0.587	
	3	0.846		0.787	
	4	0.692		0.630	
	5	0.501		0.348	
	6	-		0.894	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.969	0.014	0.018	0.956
	2/3	0.938	0.030	0.032	0.913
	3/4	0.915	0.037	0.047	0.881
	4/5	0.957	0.041	0.002	0.950
	5/6	0.993	0.007	0.000	0.993

3.4.3.3   Comprehension Language Composite 4-5

Figure 3.4.3.3A
Scale Scores: Cphn 4-5 S401 Online

Figure 3.4.3.3B
Proficiency Level: Cphn 4-5 S401 Online

Table 3.4.3.3A

Scale Score Descriptive Statistics: Cphn 4-5 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	83,406	207	474	359.65	33.63
5	60,264	205	474	363.42	39.35
Total	143,670	205	474	361.23	36.18

Table 3.4.3.3B

Proficiency Level Distribution: Cphn 4-5 S401 Online

Level	Grade 4		Grade 5		Total	
	Count	Percent	Count	Percent	Count	Percent
1	3,775	4.53%	5,616	9.32%	9,391	6.54%
2	9,644	11.56%	8,965	14.88%	18,609	12.95%
3	16,391	19.65%	11,394	18.91%	27,785	19.34%
4	13,658	16.38%	8,791	14.59%	22,449	15.63%
5	20,586	24.68%	11,876	19.71%	32,462	22.59%
6	19,352	23.20%	13,622	22.60%	32,974	22.95%
Total	83,406	100.00%	60,264	100.00%	143,670	100.00%

Table 3.4.3.3C

Reliability: Cphn 4-5 S401 Online

Component	Weight	Variance	Reliability
Listening	0.30	3292.827	0.850
Reading	0.70	1025.399	0.880
Comprehension		1278.180	0.918

*Variances from students who had results in all four domains

Table 3.4.3.4Di

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

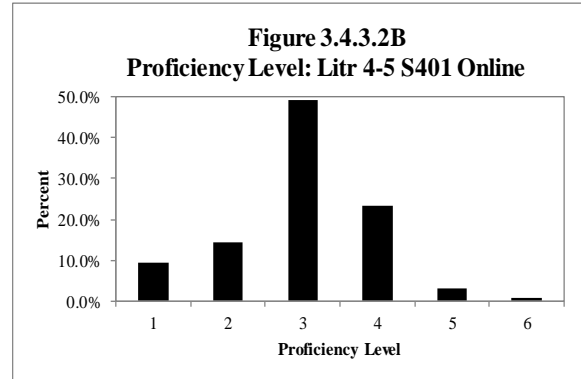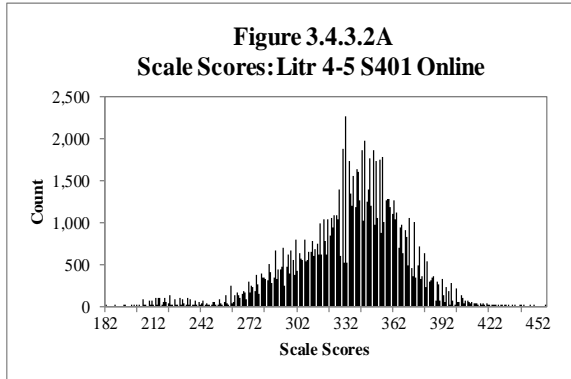| Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.676	0.574		0.472	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.861		0.744	
	2	0.703		0.583	
	3	0.650		0.532	
	4	0.482		0.376	
	5	0.639		0.531	
	6	0.830		0.744	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.982	0.005	0.013	0.975
	2/3	0.951	0.023	0.027	0.929
	3/4	0.912	0.045	0.043	0.876
	4/5	0.898	0.050	0.052	0.860
	5/6	0.921	0.040	0.039	0.887

Table 3.4.3.4Dii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.654	0.554		0.458	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.862		0.760	
	2	0.663		0.544	
	3	0.586		0.470	
	4	0.439		0.339	
	5	0.572		0.461	
	6	0.841		0.754	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.966	0.011	0.022	0.953
	2/3	0.935	0.031	0.034	0.907
	3/4	0.906	0.050	0.044	0.869
	4/5	0.901	0.049	0.050	0.863
	5/6	0.925	0.040	0.035	0.893

3.4.3.4　　Overall Language Composite 4-5

Figure 3.4.3.4A
Scale Scores: Over 4-5 S401 Online

Figure 3.4.3.4B
Proficiency Level: Over 4-5 S401 Online

Table 3.4.3.4A

Scale Score Descriptive Statistics: Over 4-5 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	45,375	176	445	339.17	33.02
5	33,626	189	450	343.86	37.97
Total	79,001	176	450	341.16	35.29

Table 3.4.3.4B

Proficiency Level Distribution: Over 4-5 S401 Online

Level	Grade 4 Count	Grade 4 Percent	Grade 5 Count	Grade 5 Percent	Total Count	Total Percent
1	2,504	5.52%	2,842	8.45%	5,346	6.77%
2	4,597	10.13%	4,163	12.38%	8,760	11.09%
3	18,992	41.86%	12,699	37.77%	31,691	40.11%
4	16,885	37.21%	11,581	34.44%	28,466	36.03%
5	2,142	4.72%	2,131	6.34%	4,273	5.41%
6	255	0.56%	210	0.62%	465	0.59%
Total	45,375	100.00%	33,626	100.00%	79,001	100.00%

Table 3.4.3.4C

Reliability: Over 4-5 S401 Online

Component	Weight	Variance	Reliability
Listening	0.15	3292.827	0.850
Reading	0.35	1025.399	0.880
Speaking	0.15	2190.505	0.723
Writing	0.35	1733.272	0.897
Overall Composite		1245.101	0.950

*Variances from students who had results in all four domains

Table 3.4.3.4Di
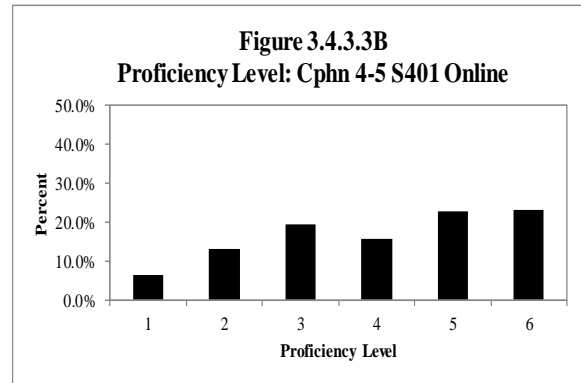
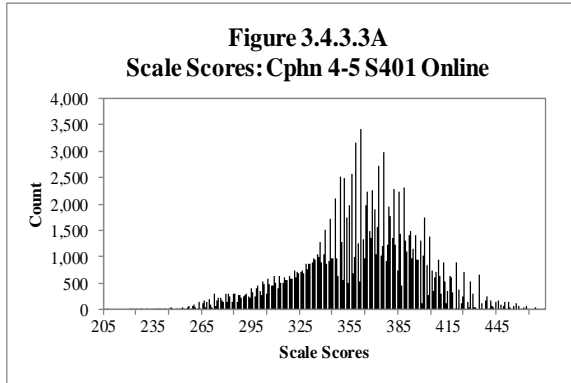Accuracy and Consistency of Classification Indices: Over (Grade 4) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.822	0.769		0.653	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.812		0.831	
	2	0.768		0.662	
	3	0.886		0.833	
	4	0.779		0.747	
	5	-		0.279	
	6	-		1.000	

Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.982	0.011	0.007	0.982
	2/3	0.961	0.022	0.017	0.952
	3/4	0.917	0.034	0.049	0.891
	4/5	0.947	0.053	0.000	0.945
	5/6	0.994	0.006	0.000	1.000

Table 3.4.3.4Dii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.794	0.728		0.616	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.864		0.836	
	2	0.754		0.648	
	3	0.861		0.797	
	4	0.735		0.708	
	5	-		0.340	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.978	0.012	0.011	0.973
	2/3	0.956	0.023	0.021	0.941
	3/4	0.921	0.032	0.046	0.893
	4/5	0.930	0.070	0.000	0.922
	5/6	0.994	0.006	0.000	0.997

3.4.4   Grades: 6-8

3.4.4.1   Oral Language Composite 6-8

Figure 3.4.4.1A
Scale Scores: Oral 6-8 S401 Online

Figure 3.4.4.1B
Proficiency Level: Oral 6-8 S401 Online

Table 3.4.4.1A

Scale Score Descriptive Statistics: Oral 6-8 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	47,283	161	475	350.63	41.68
7	47,065	164	487	357.98	46.97
8	47,355	171	493	362.92	51.56
Total	141,703	161	493	357.18	47.18

Table 3.4.4.1B

Proficiency Level Distribution: Oral 6-8 S401 Online

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	2,658	5.62%	3,733	7.93%	4,738	10.01%	11,129	7.85%
2	5,921	12.52%	6,193	13.16%	6,408	13.53%	18,522	13.07%
3	18,769	39.70%	16,407	34.86%	14,774	31.20%	49,950	35.25%
4	15,592	32.98%	15,346	32.61%	16,123	34.05%	47,061	33.21%
5	3,782	8.00%	4,674	9.93%	4,435	9.37%	12,891	9.10%
6	561	1.19%	712	1.51%	877	1.85%	2,150	1.52%
Total	47,283	100.00%	47,065	100.00%	47,355	100.00%	141,703	100.00%

Table 3.4.4.1C

Reliability: Oral 6-8 S401 Online

Component	Weight	Variance	Reliability
Listening	0.50	3391.687	0.870
Speaking	0.50	2114.953	0.705
Oral		2290.219	0.884

*Variances from students who had results in all four domains

Table 3.4.4.1Di

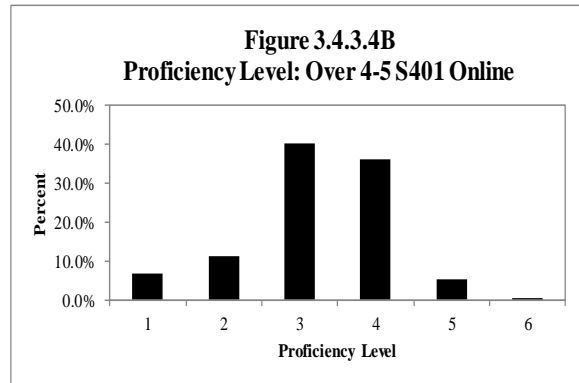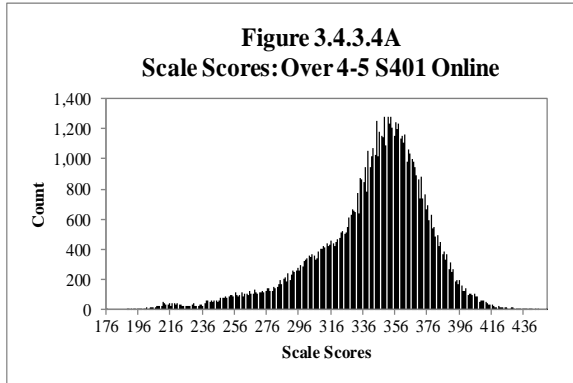Accuracy and Consistency of Classification Indices: Oral (Grade 6) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.704	0.601		0.440	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.833		0.712	
	2	0.637		0.499	
	3	0.787		0.696	
	4	0.654		0.588	
	5	0.506		0.356	
	6	-		0.315	
Indices at Cut Points		**Accuracy**			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.978	0.009	0.014	0.968
	2/3	0.933	0.035	0.031	0.903
	3/4	0.878	0.046	0.076	0.830
	4/5	0.920	0.068	0.012	0.897
	5/6	0.988	0.012	0.000	0.988

Table 3.4.4.1Dii

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

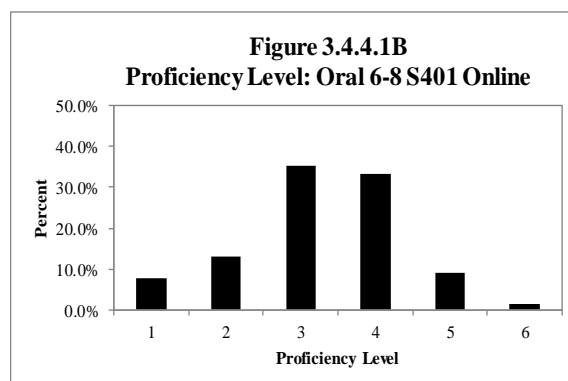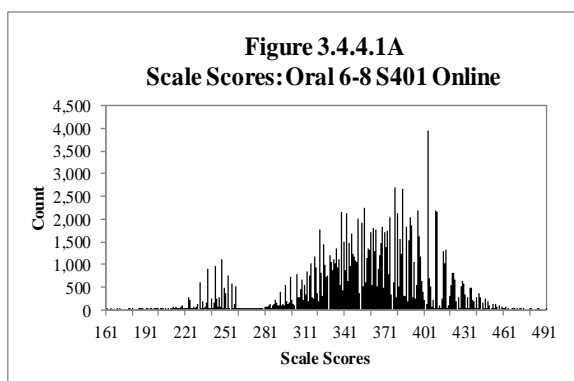| Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.673	0.565		0.413	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.837		0.721	
	2	0.611		0.475	
	3	0.745		0.641	
	4	0.633		0.562	
	5	0.505		0.368	
	6	-		0.117	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.971	0.012	0.017	0.957
	2/3	0.928	0.037	0.036	0.896
	3/4	0.878	0.046	0.076	0.831
	4/5	0.903	0.074	0.022	0.874
	5/6	0.985	0.015	0.000	0.983

Table 3.4.4.1Diii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.658	0.546		0.395	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.842		0.735	
	2	0.600		0.466	
	3	0.703		0.587	
	4	0.629		0.570	
	5	0.428		0.304	
	6	-		0.099	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.965	0.015	0.020	0.949
	2/3	0.925	0.036	0.038	0.894
	3/4	0.877	0.048	0.076	0.828
	4/5	0.894	0.087	0.019	0.861
	5/6	0.981	0.019	0.000	0.977

3.4.4.2   Literacy Language Composite 6-8

Figure 3.4.4.2A
Scale Scores: Litr 6-8 S401 Online

Figure 3.4.4.2B
Proficiency Level: Litr 6-8 S401 Online
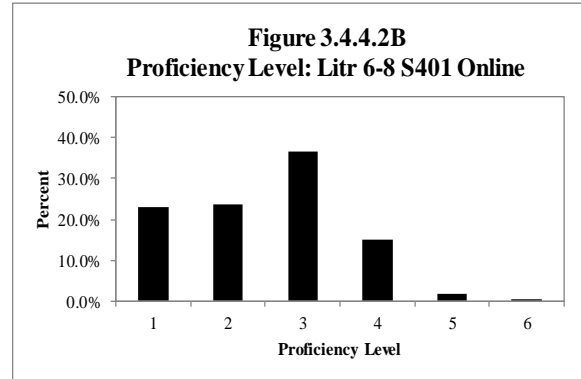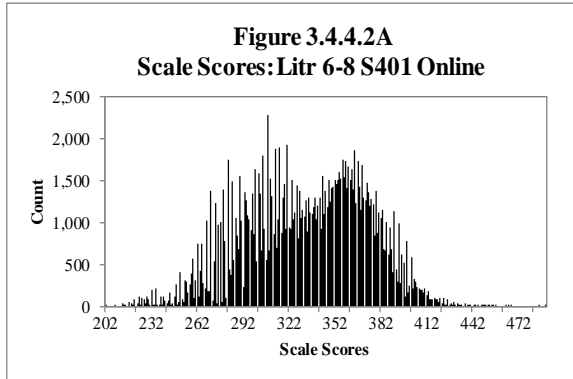
Table 3.4.4.2A

Scale Score Descriptive Statistics: Litr 6-8 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	52,892	202	452	326.49	36.69
7	53,499	214	453	333.57	39.63
8	51,693	213	490	340.48	41.66
Total	158,084	202	490	333.46	39.77

Table 3.4.4.2B

Proficiency Level Distribution: Litr 6-8 S401 Online

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	11,793	22.30%	12,305	23.00%	12,163	23.53%	36,261	22.94%
2	12,616	23.85%	12,900	24.11%	12,118	23.44%	37,634	23.81%
3	21,427	40.51%	19,436	36.33%	16,843	32.58%	57,706	36.50%
4	6,476	12.24%	7,853	14.68%	9,274	17.94%	23,603	14.93%
5	533	1.01%	922	1.72%	1,206	2.33%	2,661	1.68%
6	47	0.09%	83	0.16%	89	0.17%	219	0.14%
Total	52,892	100.00%	53,499	100.00%	51,693	100.00%	158,084	100.00%

Table 3.4.4.2C

Reliability: Litr 6-8 S401 Online

Component	Weight	Variance	Reliability
Reading	0.50	1625.254	0.920
Writing	0.50	2000.288	0.890
Literacy		1597.394	0.945

*Variances from students who had results in all four domains

Table 3.4.4.2Di

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.823	0.752		0.654	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.906		0.855	
	2	0.743		0.650	
	3	0.852		0.800	
	4	0.748		0.652	
	5	0.628		0.372	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.952	0.020	0.028	0.933
	2/3	0.930	0.036	0.034	0.902
	3/4	0.950	0.027	0.023	0.929
	4/5	0.990	0.009	0.001	0.988
	5/6	0.999	0.001	0.000	1.000

Table 3.4.4.2Dii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.811	0.734		0.640	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.905		0.852	
	2	0.745		0.652	
	3	0.826		0.765	
	4	0.761		0.666	
	5	0.585		0.389	
	6	-		0.864	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.950	0.021	0.029	0.930
	2/3	0.931	0.035	0.034	0.903
	3/4	0.945	0.030	0.025	0.923
	4/5	0.986	0.011	0.003	0.979
	5/6	0.998	0.002	0.000	0.999

Table 3.4.4.2Diii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.798	0.717		0.624	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.900		0.847	
	2	0.750		0.654	
	3	0.796		0.722	
	4	0.756		0.675	
	5	0.530		0.334	
	6	-		0.300	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.950	0.023	0.027	0.929
	2/3	0.933	0.032	0.035	0.906
	3/4	0.938	0.033	0.029	0.913
	4/5	0.978	0.017	0.005	0.968
	5/6	0.998	0.002	0.000	0.998

3.4.4.3 Comprehension Language Composite 6-8

Figure 3.4.4.3A
Scale Scores: Cphn 6-8 S401 Online

Figure 3.4.4.3B
Proficiency Level: Cphn 6-8 S401 Online
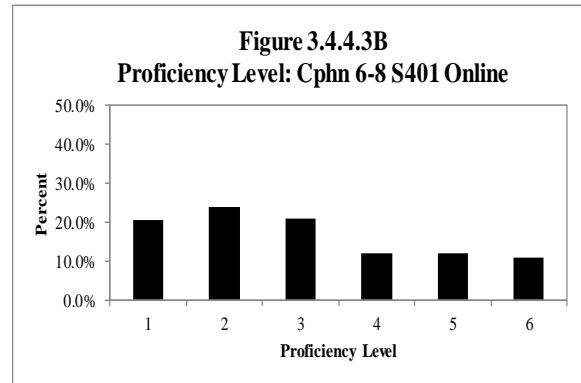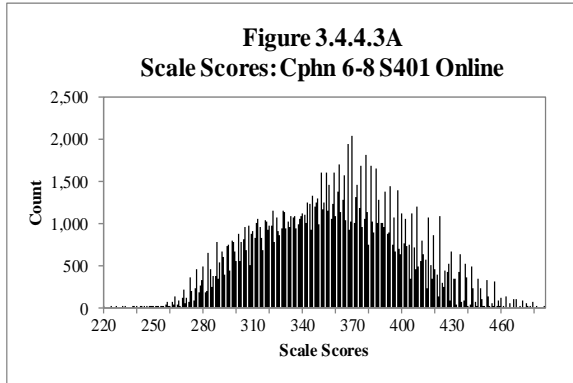
Table 3.4.4.3A

Scale Score Descriptive Statistics: Cphn 6-8 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	50,026	238	486	349.91	37.32
7	49,974	220	486	358.12	41.86
8	48,606	224	486	365.31	45.29
Total	148,606	220	486	357.71	42.05

Table 3.4.4.3B

Proficiency Level Distribution: Cphn 6-8 S401 Online

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	9,378	18.75%	10,570	21.15%	10,727	22.07%	30,675	20.64%
2	13,109	26.20%	11,593	23.20%	10,552	21.71%	35,254	23.72%
3	11,828	23.64%	10,338	20.69%	8,835	18.18%	31,001	20.86%
4	6,224	12.44%	6,094	12.19%	5,491	11.30%	17,809	11.98%
5	5,478	10.95%	5,741	11.49%	6,338	13.04%	17,557	11.81%
6	4,009	8.01%	5,638	11.28%	6,663	13.71%	16,310	10.98%
Total	50,026	100.00%	49,974	100.00%	48,606	100.00%	148,606	100.00%

Table 3.4.4.3C

Reliability: Cphn 6-8 S401 Online

Component	Weight	Variance	Reliability
Listening	0.30	3391.687	0.870
Reading	0.70	1625.254	0.920
Comprehension		1813.255	0.943

*Variances from students who had results in all four domains

Table 3.4.4.3Di

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.725	0.632		0.543	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.894		0.828	
	2	0.752		0.665	
	3	0.688		0.585	
	4	0.521		0.406	
	5	0.611		0.485	
	6	0.847		0.740	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.951	0.019	0.030	0.932
	2/3	0.927	0.040	0.034	0.897
	3/4	0.926	0.042	0.033	0.897
	4/5	0.944	0.029	0.028	0.920
	5/6	0.971	0.018	0.011	0.959

Table 3.4.4.3Dii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.713	0.619		0.534	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.894		0.833	
	2	0.717		0.620	
	3	0.652		0.542	
	4	0.501		0.388	
	5	0.585		0.461	
	6	0.856		0.762	

Indices at Cut Points		Accuracy			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.947	0.021	0.032	0.926
	2/3	0.929	0.038	0.033	0.899
	3/4	0.926	0.041	0.033	0.897
	4/5	0.940	0.032	0.029	0.915
	5/6	0.963	0.022	0.015	0.947

Table 3.4.4.3Diii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.705	0.612		0.529	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.892		0.832	
	2	0.704		0.602	
	3	0.612		0.497	
	4	0.462		0.354	
	5	0.592		0.472	
	6	0.854		0.765	

Indices at Cut Points		Accuracy			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.946	0.023	0.031	0.924
	2/3	0.930	0.036	0.034	0.901
	3/4	0.927	0.040	0.033	0.898
	4/5	0.935	0.035	0.030	0.909
	5/6	0.956	0.025	0.019	0.937

3.4.4.4 Overall Language Composite 6-8

Figure 3.4.4.4A
Scale Scores: Over 6-8 S401 Online

Figure 3.4.4.4B
Proficiency Level: Over 6-8 S401 Online
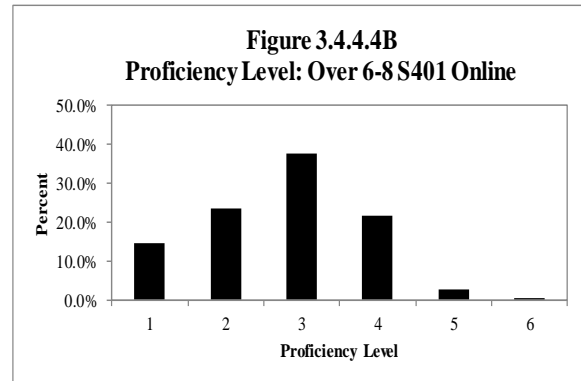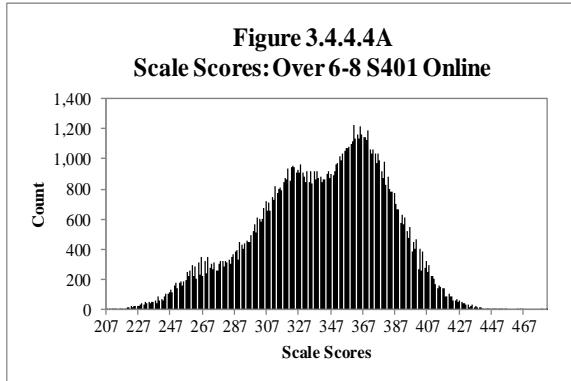
Table 3.4.4.4A

Scale Score Descriptive Statistics: Over 6-8 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	36,879	207	454	334.09	36.03
7	36,536	209	455	341.35	40.14
8	36,097	216	481	347.29	43.32
Total	109,512	207	481	340.86	40.28

Table 3.4.4.4B

Proficiency Level Distribution: Over 6-8 S401 Online

Level	Grade 6 Count	Grade 6 Percent	Grade 7 Count	Grade 7 Percent	Grade 8 Count	Grade 8 Percent	Total Count	Total Percent
1	4,452	12.07%	5,346	14.63%	6,063	16.80%	15,861	14.48%
2	9,175	24.88%	8,592	23.52%	7,943	22.00%	25,710	23.48%
3	15,730	42.65%	13,531	37.03%	11,892	32.94%	41,153	37.58%
4	6,847	18.57%	8,021	21.95%	8,879	24.60%	23,747	21.68%
5	627	1.70%	967	2.65%	1,247	3.45%	2,841	2.59%
6	48	0.13%	79	0.22%	73	0.20%	200	0.18%
Total	36,879	100.00%	36,536	100.00%	36,097	100.00%	109,512	100.00%

Table 3.4.4.4C

Reliability: Over 6-8 S401 Online

Component	Weight	Variance	Reliability
Listening	0.15	3391.687	0.870
Reading	0.35	1625.254	0.920
Speaking	0.15	2114.953	0.705
Writing	0.35	2000.288	0.890
Overall Composite		1622.385	0.959

*Variances from students who had results in all four domains

Table 3.4.4.4Di

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.841	0.782		0.693	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.877		0.821	
	2	0.832		0.761	
	3	0.866		0.819	
	4	0.779		0.722	
	5	-		0.435	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.970	0.015	0.015	0.959
	2/3	0.942	0.028	0.031	0.919
	3/4	0.944	0.028	0.028	0.922
	4/5	0.982	0.018	0.000	0.982
	5/6	0.999	0.001	0.000	1.000

Table 3.4.4.4Dii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.826	0.760		0.675	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.881		0.829	
	2	0.812		0.734	
	3	0.839		0.781	
	4	0.790		0.737	
	5	0.742		0.439	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.966	0.017	0.017	0.953
	2/3	0.941	0.028	0.031	0.919
	3/4	0.941	0.030	0.029	0.919
	4/5	0.974	0.023	0.003	0.971
	5/6	0.998	0.002	0.000	0.999

Table 3.4.4.4Diii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.816	0.742		0.658	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.899		0.845	
	2	0.802		0.719	
	3	0.816		0.747	
	4	0.789		0.735	
	5	0.582		0.385	
	6	-		0.889	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.965	0.017	0.018	0.951
	2/3	0.944	0.025	0.031	0.922
	3/4	0.938	0.031	0.030	0.913
	4/5	0.968	0.025	0.006	0.958
	5/6	0.998	0.002	0.000	0.998

3.4.5 Grades: 9-12

3.4.5.1   Oral Language Composite 9-12

Figure 3.4.5.1A
Scale Scores: Oral 9-12 S401 Online

Figure 3.4.5.1B
Proficiency Level: Oral 9-12 S401 Online

Table 3.4.5.1A

Scale Score Descriptive Statistics: Oral 9-12 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	61,666	154	491	347.85	47.25
10	41,084	207	484	348.68	45.25
11	28,692	215	501	351.22	44.70
12	19,971	217	477	354.46	43.09
Total	151,413	154	501	349.58	45.75

Table 3.4.5.1B

Proficiency Level Distribution: Oral 9-12 S401 Online

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	10,143	16.45%	7,150	17.40%	5,382	18.76%	3,674	18.40%	26,349	17.40%
2	13,979	22.67%	9,884	24.06%	7,372	25.69%	5,194	26.01%	36,429	24.06%
3	23,042	37.37%	16,894	41.12%	11,703	40.79%	8,645	43.29%	60,284	39.81%
4	12,577	20.40%	6,369	15.50%	3,708	12.92%	2,214	11.09%	24,868	16.42%
5	1,772	2.87%	738	1.80%	494	1.72%	224	1.12%	3,228	2.13%
6	153	0.25%	49	0.12%	33	0.12%	20	0.10%	255	0.17%
Total	61,666	100.00%	41,084	100.00%	28,692	100.00%	19,971	100.00%	151,413	100.00%

Table 3.4.5.1C

Reliability: Oral 9-12 S401 Online

Component	Weight	Variance	Reliability
Listening	0.50	2992.442	0.870
Speaking	0.50	2358.334	0.658
Oral		2210.830	0.865

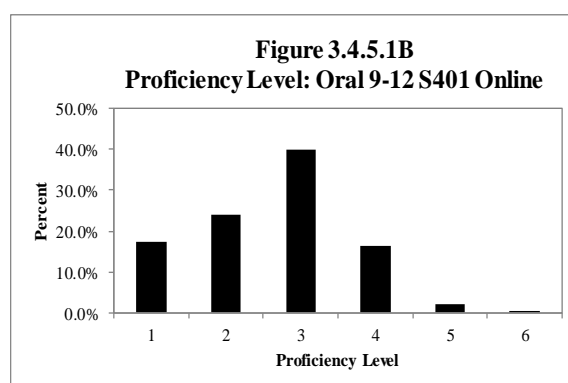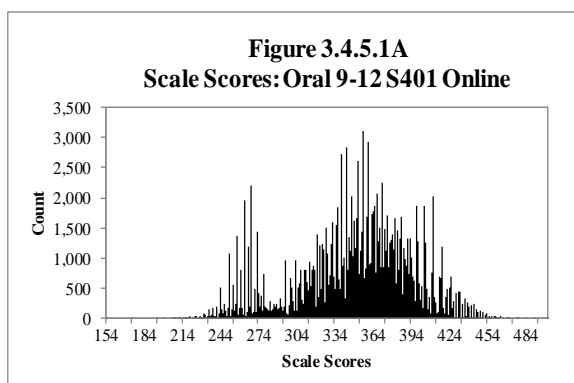*Variances from students who had results in all four domains

Table 3.4.5.1Di

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.692	0.582		0.438	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.833		0.724	
	2	0.619		0.502	
	3	0.709		0.617	
	4	0.645		0.545	
	5	0.515		0.263	
	6	-		0.109	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.937	0.026	0.037	0.910
	2/3	0.890	0.054	0.056	0.846
	3/4	0.893	0.055	0.052	0.849
	4/5	0.969	0.030	0.001	0.959
	5/6	0.998	0.002	0.000	0.997

Table 3.4.5.1Dii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.699	0.592		0.437	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.827		0.721	
	2	0.637		0.517	
	3	0.728		0.643	
	4	0.591		0.474	
	5	-		0.156	
	6	-		0.667	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.936	0.029	0.035	0.908
	2/3	0.886	0.054	0.061	0.840
	3/4	0.894	0.054	0.052	0.849
	4/5	0.981	0.019	0.000	0.976
	5/6	0.999	0.001	0.000	0.999

Table 3.4.5.1Diii

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

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.699	0.593		0.435	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.819		0.714	
	2	0.647		0.527	
	3	0.724		0.641	
	4	0.556		0.433	
	5	-		0.159	
	6	-		0.600	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.931	0.034	0.035	0.900
	2/3	0.881	0.053	0.066	0.834
	3/4	0.902	0.052	0.045	0.860
	4/5	0.982	0.018	0.000	0.978
	5/6	0.999	0.001	0.000	0.999

Table 3.4.5.1Div

Accuracy and Consistency of Classification Indices: Oral (Grade 12) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.713	0.610		0.445	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.802		0.696	
	2	0.645		0.527	
	3	0.746		0.673	
	4	0.573		0.431	
	5	-		0.145	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.928	0.037	0.036	0.896
	2/3	0.878	0.055	0.067	0.832
	3/4	0.914	0.050	0.036	0.877
	4/5	0.988	0.012	0.000	0.987
	5/6	0.999	0.001	0.000	1.000

3.4.5.2　Literacy Language Composite 9-12

Figure 3.4.5.2A
Scale Scores: Litr 9-12 S401 Online

Figure 3.4.5.2B
Proficiency Level: Litr 9-12 S401 Online

Table 3.4.5.2A

Scale Score Descriptive Statistics: Litr 9-12 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	59,767	235	477	351.47	38.83
10	38,625	230	474	357.56	34.27
11	26,179	252	487	362.91	33.67
12	18,165	261	482	366.57	31.97
Total	142,736	230	487	357.14	36.31

Table 3.4.5.2B

Proficiency Level Distribution: Litr 9-12 S401 Online

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	12,166	20.36%	6,178	15.99%	4,064	15.52%	2,882	15.87%	25,290	17.72%
2	13,606	22.77%	10,399	26.92%	7,803	29.81%	5,778	31.81%	37,586	26.33%
3	21,157	35.40%	15,177	39.29%	9,853	37.64%	7,090	39.03%	53,277	37.33%
4	10,324	17.27%	5,732	14.84%	3,676	14.04%	1,988	10.94%	21,720	15.22%
5	2,301	3.85%	1,069	2.77%	741	2.83%	421	2.32%	4,532	3.18%
6	213	0.36%	70	0.18%	42	0.16%	6	0.03%	331	0.23%
Total	59,767	100.00%	38,625	100.00%	26,179	100.00%	18,165	100.00%	142,736	100.00%

Table 3.4.5.2C

Reliability: Litr 9-12 S401 Online

Component	Weight	Variance	Reliability
Reading	0.50	1432.799	0.910
Writing	0.50	1677.917	0.867
Literacy		1309.167	0.933

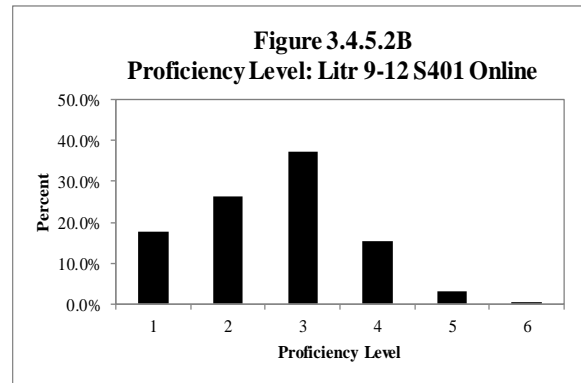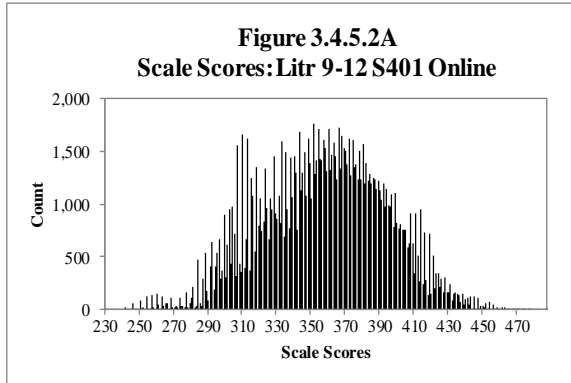*Variances from students who had results in all four domains

Table 3.4.5.2Di

Accuracy and Consistency of Classification Indices: Litr (Grade 9) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.782	0.695		0.594	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.893		0.826	
	2	0.701		0.601	
	3	0.798		0.728	
	4	0.757		0.654	
	5	0.692		0.537	
	6	0.778		0.449	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.946	0.020	0.034	0.925
	2/3	0.924	0.039	0.036	0.893
	3/4	0.936	0.036	0.028	0.910
	4/5	0.978	0.014	0.008	0.968
	5/6	0.997	0.003	0.000	0.996

Table 3.4.5.2Dii

Accuracy and Consistency of Classification Indices: Litr (Grade 10) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.802	0.721		0.616	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.870		0.788	
	2	0.748		0.661	
	3	0.825		0.765	
	4	0.780		0.678	
	5	0.757		0.603	
	6	0.870		0.658	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.950	0.019	0.031	0.930
	2/3	0.921	0.040	0.039	0.889
	3/4	0.946	0.031	0.023	0.924
	4/5	0.986	0.009	0.005	0.980
	5/6	0.999	0.001	0.000	0.999

Table 3.4.5.2Diii

Accuracy and Consistency of Classification Indices: Litr (Grade 11) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.801	0.720		0.614	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.860		0.773	
	2	0.774		0.693	
	3	0.813		0.749	
	4	0.771		0.665	
	5	0.759		0.604	
	6	0.871		0.663	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.951	0.021	0.029	0.930
	2/3	0.919	0.040	0.041	0.886
	3/4	0.947	0.031	0.022	0.926
	4/5	0.986	0.009	0.005	0.979
	5/6	0.999	0.001	0.000	0.999

Table 3.4.5.2Div

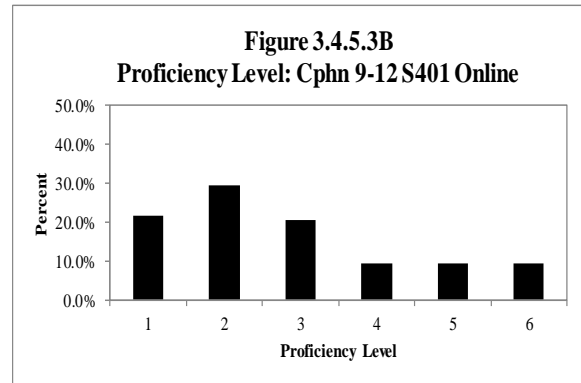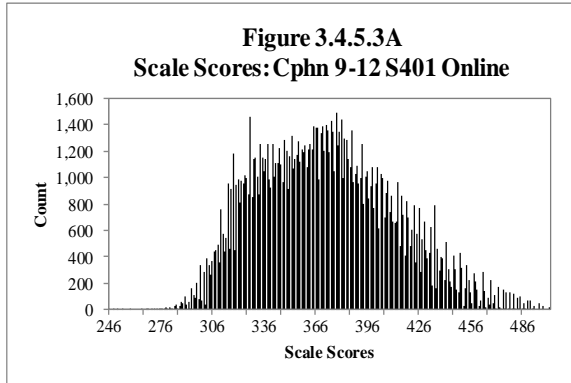Accuracy and Consistency of Classification Indices: Litr (Grade 12) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.809	0.730		0.621	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.853		0.765	
	2	0.783		0.704	
	3	0.830		0.770	
	4	0.750		0.635	
	5	0.819		0.665	
	6	1.000		0.737	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.949	0.022	0.029	0.927
	2/3	0.916	0.042	0.042	0.882
	3/4	0.956	0.026	0.019	0.937
	4/5	0.989	0.008	0.003	0.984
	5/6	1.000	0.000	0.000	1.000

3.4.5.3 Comprehension Language Composite 9-12

Figure 3.4.5.3A
Scale Scores: Cphn 9-12 S401 Online

Figure 3.4.5.3B
Proficiency Level: Cphn 9-12 S401 Online

Table 3.4.5.3A

Scale Score Descriptive Statistics: Cphn 9-12 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	56,813	246	502	368.17	40.84
10	36,742	269	502	371.84	38.44
11	24,906	268	502	376.07	38.52
12	17,298	277	502	379.12	36.99
Total	135,759	246	502	372.01	39.50

Table 3.4.5.3B

Proficiency Level Distribution: Cphn 9-12 S401 Online

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	12,623	22.22%	7,811	21.26%	5,487	22.03%	3,686	21.31%	29,607	21.81%
2	15,611	27.48%	10,783	29.35%	7,748	31.11%	5,782	33.43%	39,924	29.41%
3	11,163	19.65%	8,037	21.87%	5,069	20.35%	3,568	20.63%	27,837	20.50%
4	5,356	9.43%	3,589	9.77%	2,150	8.63%	1,596	9.23%	12,691	9.35%
5	5,744	10.11%	3,318	9.03%	2,333	9.37%	1,381	7.98%	12,776	9.41%
6	6,316	11.12%	3,204	8.72%	2,119	8.51%	1,285	7.43%	12,924	9.52%
Total	56,813	100.00%	36,742	100.00%	24,906	100.00%	17,298	100.00%	135,759	100.00%

Table 3.4.5.3C

Reliability: Cphn 9-12 S401 Online

Component	Weight	Variance	Reliability
Listening	0.30	2992.442	0.870
Reading	0.70	1432.799	0.910
Comprehension		1591.969	0.938

*Variances from students who had results in all four domains

Table 3.4.5.3Di

Accuracy and Consistency of Classification Indices: Cphn (Grade 9) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)		
	0.721	0.628		0.538		
Conditional on Level	Level	Accuracy		Consistency		
	1	0.868		0.799		
	2	0.729		0.639		
	3	0.658		0.544		
	4	0.461		0.349		
	5	0.600		0.474		
	6	0.875		0.794		
Indices at Cut Points		Accuracy				
	Cut Point	Accuracy	False Positives	False Negatives	Consistency	
	1/2	0.933	0.028	0.039	0.907	
	2/3	0.926	0.041	0.033	0.896	
	3/4	0.936	0.034	0.030	0.911	
	4/5	0.950	0.028	0.022	0.928	
	5/6	0.968	0.019	0.013	0.954	

Table 3.4.5.3Dii

Accuracy and Consistency of Classification Indices: Cphn (Grade 10) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.728	0.634		0.540	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.867		0.794	
	2	0.749		0.664	
	3	0.683		0.575	
	4	0.484		0.368	
	5	0.590		0.462	
	6	0.862		0.768	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.937	0.027	0.036	0.911
	2/3	0.924	0.041	0.035	0.892
	3/4	0.936	0.034	0.029	0.911
	4/5	0.953	0.026	0.021	0.932
	5/6	0.972	0.017	0.011	0.960

Table 3.4.5.3Diii

Accuracy and Consistency of Classification Indices: Cphn (Grade 11) S401 Online

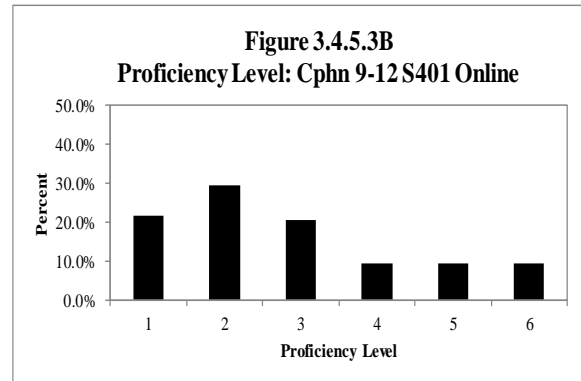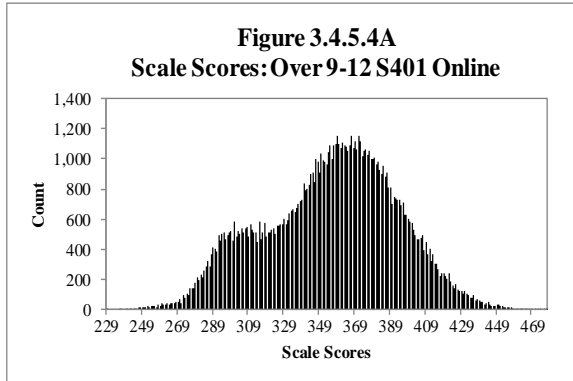| Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.729	0.635		0.539	
Conditional on Level	Level	Accuracy		Consistency	
	1	0.851		0.778	
	2	0.758		0.673	
	3	0.667		0.555	
	4	0.454		0.341	
	5	0.619		0.490	
	6	0.864		0.773	
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.932	0.032	0.036	0.904
	2/3	0.923	0.041	0.035	0.892
	3/4	0.939	0.032	0.028	0.915
	4/5	0.954	0.026	0.020	0.935
	5/6	0.973	0.016	0.011	0.962

Table 3.4.5.3Div

Accuracy and Consistency of Classification Indices: Cphn (Grade 12) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.737	0.643		0.543	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.838		0.761	
	2	0.778		0.697	
	3	0.673		0.561	
	4	0.509		0.387	
	5	0.606		0.476	
	6	0.877		0.788	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.931	0.035	0.034	0.902
	2/3	0.922	0.041	0.037	0.891
	3/4	0.942	0.031	0.027	0.918
	4/5	0.960	0.022	0.019	0.942
	5/6	0.977	0.015	0.008	0.968

3.4.5.4   Overall Language Composite 9-12

Figure 3.4.5.4A
Scale Scores: Over 9-12 S401 Online

Figure 3.4.5.3B
Proficiency Level: Cphn 9-12 S401 Online

Table 3.4.5.4A

Scale Score Descriptive Statistics: Over 9-12 S401 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	41,580	229	469	351.06	39.70
10	26,964	244	474	354.75	36.06
11	18,004	248	477	359.02	35.73
12	13,070	257	472	362.71	33.53
Total	99,618	229	477	355.03	37.48

Table 3.4.5.4B

Proficiency Level Distribution: Over 9-12 S401 Online

Level	Grade 9 Count	Grade 9 Percent	Grade 10 Count	Grade 10 Percent	Grade 11 Count	Grade 11 Percent	Grade 12 Count	Grade 12 Percent	Total Count	Total Percent
1	8,065	19.40%	4,877	18.09%	3,341	18.56%	2,272	17.38%	18,555	18.63%
2	8,236	19.81%	6,205	23.01%	4,562	25.34%	3,809	29.14%	22,812	22.90%
3	16,441	39.54%	11,527	42.75%	7,454	41.40%	5,442	41.64%	40,864	41.02%
4	7,463	17.95%	3,861	14.32%	2,269	12.60%	1,358	10.39%	14,951	15.01%
5	1,284	3.09%	470	1.74%	362	2.01%	185	1.42%	2,301	2.31%
6	91	0.22%	24	0.09%	16	0.09%	4	0.03%	135	0.14%
Total	41,580	100.00%	26,964	100.00%	18,004	100.00%	13,070	100.00%	99,618	100.00%

Table 3.4.5.4C

Reliability: Over 9-12 S401 Online

Component	Weight	Variance	Reliability
Listening	0.15	2992.442	0.870
Reading	0.35	1432.799	0.910
Speaking	0.15	2358.334	0.658
Writing	0.35	1677.917	0.867
Overall Composite		1405.064	0.950

*Variances from students who had results in all four domains

Table 3.4.5.4Di

Accuracy and Consistency of Classification Indices: Over (Grade 9) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.818	0.745		0.654	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.918		0.868	
	2	0.718		0.622	
	3	0.856		0.802	
	4	0.777		0.691	
	5	0.693		0.536	
	6	-		0.688	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.958	0.015	0.027	0.942
	2/3	0.938	0.035	0.027	0.913
	3/4	0.942	0.029	0.028	0.918
	4/5	0.981	0.013	0.006	0.973
	5/6	0.998	0.002	0.000	0.998

Table 3.4.5.4Dii

Accuracy and Consistency of Classification Indices: Over (Grade 10) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.832	0.765		0.672	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.908		0.856	
	2	0.754		0.667	
	3	0.870		0.824	
	4	0.771		0.682	
	5	0.766		0.554	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.959	0.016	0.025	0.944
	2/3	0.935	0.036	0.029	0.908
	3/4	0.949	0.026	0.025	0.929
	4/5	0.987	0.010	0.002	0.984
	5/6	0.999	0.001	0.000	1.000

Table 3.4.5.4Diii

Accuracy and Consistency of Classification Indices: Over (Grade 11) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.832	0.763		0.670	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.897		0.839	
	2	0.767		0.683	
	3	0.866		0.818	
	4	0.770		0.674	
	5	0.794		0.635	
	6	-		1.000	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.956	0.018	0.026	0.939
	2/3	0.932	0.036	0.032	0.904
	3/4	0.955	0.024	0.021	0.936
	4/5	0.988	0.009	0.003	0.985
	5/6	0.999	0.001	0.000	1.000

Table 3.4.5.4Div

Accuracy and Consistency of Classification Indices: Over (Grade 12) S401 Online

Overall Indices	Accuracy	Consistency		Kappa (k)	
	0.838	0.772		0.677	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.884		0.820	
	2	0.797		0.722	
	3	0.869		0.822	
	4	0.759		0.662	
	5	0.854		0.682	
	6	-		1.000	

Indices at Cut Points		Accuracy			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.956	0.020	0.024	0.939
	2/3	0.929	0.037	0.034	0.901
	3/4	0.961	0.021	0.019	0.944
	4/5	0.991	0.008	0.001	0.989
	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, D.C.: 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. (2004). Linking with equivalent group or single group design (LEGS) (version 2.0). [Computer software]. University of Iowa: Center for Advanced Studies in Measurement and Assessment (CASMA).

Center for Applied Linguistics (2016). *ACCESS for ELLs® Series 400 Listening and Reading Scale Maintenance: Technical Brief.* Washington, D.C.: Author.

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. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language.* London: Routledge.

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

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

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.

Data Recognition Corporation (2016a). *DRC INSIGHT Online Learning System Technology User Guide.* Maple Grove, MN: DRC/WIDA.

Data Recognition Corporation (2016b). *WIDA Assessment Management System (WIDA AMS).* Maple Grove, MN: DRC/WIDA.

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,* 221-240.

Kenyon, D. M. (2006). *Development and Field Test of ACCESS for ELLs® (WIDA Consortium Technical Report No. 1).* Center for Applied Linguistics: Washington, DC.

Kenyon, D. M., Ryu, J. R., & MacGregor, D. (2013). *Setting Grade Level Cut Scores for ACCESS for ELLs® (WIDA Consortium Technical Report No. 4).* Center for Applied Linguistics: Washington, DC.

Lee, W., Hanson, B.A., & Brennan, R.L. (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement, 26,* 412-432.

Linacre, J. M. (n.d). Displacement measures. Retrieved from http://www.winsteps.com/winman/displacement.htm.

Linacre, J. M. (1999, Autumn). Relating Cronbach and Rasch Reliabilities *Rasch Measurement Transactions, 1999, 13:2 p. 696.* Retrieved from http://www.rasch.org/rmt/rmt132i.htm

Linacre, J. M. (2002, Autumn). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878. Retrieved from http://www.rasch.org/rmt/rmt162f.htm

Linacre, J. M. (2006). Winsteps Rasch analysis (Version 3.60.1) [Computer software]. Retrieved from http://www.winsteps.com

Livingston, S.A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32,* 179-197.

Llosa, L. (2008). Building and supporting a validity argument for a standards-based classroom assessment of English proficiency based on teacher judgments. *Educational Measurement: Issues and Practice, 27*(3), 32–42.

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]. Retrieved from 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.

Thissen, D. (2000). Reliability and measurement precision. In H. Wainer, N. Dorans, D. Eignor, R. Flaugher, B. Green, R. Mislevy, L. Steinberg & D. Thissen (Eds.), *Computerized adaptive testing: A primer* (Second Edition). Hillsdale, NJ: Lawrence Erlbaum Associates, 159-184.

Toulmin, S. E. (2003). *The uses of argument* (Updated ed.). Cambridge: Cambridge University Press.

WIDA Consortium. (2007). *English Language Proficiency Standards and Resource Guide, 2007 Edition, PreKindergarten through Grade 12*. Madison, WI: Board of Regents of the University of Wisconsin System.

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

WIDA Consortium. (2016). *WIDA ACCESS for ELLs 2.0 2016–2017 Test Administrator Manual.* Madison, WI: Author.

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., & Bridgeman, B. (2014). Evaluating validity, fairness, and differential item functioning in multistage testing. In Y. Duanli, A. A. von Davier, & C. Lewis (Eds.), *Computer Multistage Testing: Theory and Applications* (pp. 271–284). Hoboken, NJ: CRC Press.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement, 30,* 233–251.

Zwick R., Thayer D. T., & Wingersky, M. (1993). A simulation study of methods for addressing differential item functioning in computer-adaptive tests. *ETS Research Report,* 93–11.

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.
Michele Kawood, M.S.Ed.
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.
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.