World-Class Instructional Design and Assessment

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

Annual Technical Report No. 12A

Prepared by:

Center for Applied Linguistics

Language Assessment Division
Psychometrics and Quantitative Research Team

May 2017

The WIDA ACCESS for ELLs Technical Advisory Committee

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

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

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

Executive Summary

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

This technical report is produced as a service to members and potential members of the WIDA Consortium. The technical information herein is intended for use by those who have technical knowledge of test construction and measurement procedures, as stated in *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2014).

ACCESS for ELLs is intended to assess reliably and validly the English language development (ELD) of English language learners (ELLs) in Grades K–12 according to WIDA 2012 Amplification of the English Language Development Standards Kindergarten–Grade 12 (WIDA Consortium, 2012). Results on ACCESS for ELLs are used by WIDA Consortium states for monitoring the progress of students, for making decisions about exiting students from language support services, and for accountability.

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

Table 0.1

Participation in ACCESS for ELLs Online and Paper, Series 400

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

Summary Highlights

This report presents a wealth of data documenting the technical properties of ACCESS for ELLs 2.0 Series 400 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.

Launch of ACCESS 2.0

Series 400 Online is the first series of the ACCESS 2.0 assessment. ACCESS 2.0 is now offered in two formats. The Paper format is available for grades K–12, and the Online format is available for grades 1–12. There are a number of key changes between ACCESS 2.0 Series 400 Online and the test offered in the prior year; further detail on these elements can be found in Part I of this report.

- ACCESS 2.0 Online is administered entirely online, with the exception of the Writing test in grades 1–3, which are administered on paper.
- The cluster structure of ACCESS 2.0 Online has been updated from the cluster structure of ACCESS 1.0. The cluster structure of ACCESS 2.0 Online has been modified to add an additional cluster in the early grades to account for the more rapid developmental changes that younger children undergo. The cluster structure of ACCESS 2.0 Online is: 1, 2–3, 4–5, 6–8, 9–12.
- The Listening and Reading tests of ACCESS 2.0 Online are multistage adaptive assessments. The three-tier structure of ACCESS 1.0 no longer applies to Listening and Reading Online.
- The Writing test is offered in both keyboarded and handwritten response formats. For grades 1–3, administered on paper, responses are handwritten. For grades 4–5, either response mode is permitted, and the choice of response mode may be made at a state or local level. For grades 6 and above, the handwritten response mode is offered as an accommodation. Writing responses are centrally scored.
- The Writing test now has two tiers: A and B/C.
- The Speaking test is administered online, with recorded responses scored by central raters. The Speaking test has three tiers: pre-A, A, and B/C.

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 400 Online data set included the results of 1,148,337 students. The largest grade was Grade 2 with 172,489 students, while the smallest was Grade 12 with 30,026 students. Of the participating WIDA states, the largest was Illinois with 119,961 students, while the smallest was the United States Virgin Islands with 743 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 400 Online, presented in Table C of Section 3.4, is very high across all grade-level clusters. For Grade 1 it was .929; for Grades 2–3, .936; for Grades 4–5, .935; for Grades 6–8, .944; and for Grades 9–12, .951.

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	0.976	0.925	0.927	0.959	0.991
2	0.992	0.963	0.916	0.937	0.979
3	0.991	0.977	0.94	0.911	0.905
4	0.993	0.98	0.937	0.906	0.908
5	0.989	0.972	0.927	0.895	0.928
6	0.986	0.957	0.918	0.923	0.979
7	0.98	0.949	0.915	0.934	0.983
8	0.977	0.945	0.915	0.923	0.985
9	0.969	0.948	0.93	0.931	0.96
10	0.973	0.943	0.927	0.943	0.972
11	0.975	0.945	0.924	0.942	0.964
12	0.974	0.942	0.921	0.947	0.954

Series 400 Online: Special Considerations

Test Interruption Issues

During the 2015–2016 online administration of ACCESS 2.0, a substantial number of interruptions occurred during students' test sessions. Interrupted test sessions were observed by states early in the ACCESS 2.0 test administration cycle (i.e., December 2015). States observing these early interruptions reported them WIDA; WIDA then reported this to Data Recognition Corporation (DRC), the vendor responsible for the online administration of the test. Interruptions occurred for a variety of reasons—some caused by test administrators, but many due to technical issues associated with DRC's online test administration engine. The frequency of interruptions was such that there was concern about the meaning of students' scores. In January of 2016, WIDA requested online telemetry data from DRC to examine the scope of technical or unexplained interruptions. With early ACCESS administration data, WIDA conducted preliminary interruption analyses and discovered small but noticeable differences between interrupted and non-interrupted student scores. WIDA decided not to correct for interruptions on score reports; however, WIDA directed the Center for Applied Linguistics to not include students with interrupted test sessions in their psychometric analyses. Hence, psychometric

analyses reported in this year's annual report do not included students with interrupted tests sessions. (Note that tables which summarize counts of students who participated in the assessment do include students who experienced test interruptions.) A formal report on the nature and effect of interruptions is forthcoming and provides the complete scope of interruptions and their effects during the 2015-2016 test administration.

Writing Grade 1, Tier A

In February 2017, during CAL's preparation of the final data for this ATR, an anomaly in the Grade 1, Tier A Writing scores was discovered; namely, there were no Proficiency Level (PL) 3 scores awarded for that test, when usually 3-5% of Tier A students would have achieved PL 3. Looking further into the matter, CAL and WIDA discovered that scores on the fourth and last task on the Grade 1, Tier A Writing test were not weighted by a factor of 3 as defined in the scoring algorithm specifications.

Once the scoring error was confirmed, WIDA Consortium member states were notified of the error in a March 14, 2017 memorandum to State Education Agency Assessment Directors.

The results reported herein are drawn from the data as CAL received it 2016. The tables listed below have been produced using the erroneously scored Grade 1 Tier A Writing data as it was received by CAL in October 2016. A report correcting these tables is forthcoming.

Scale Score Descriptive Statistics for Writing Grade 1 Tier A and for Writing Grade 1 Across Tiers

Proficiency Level Descriptive Statistics for Writing Grade 1 Tier A and for Writing Grade 1 Across Tiers

Scale Score Descriptive Statistics for the Literacy and Overall Composites, Grade 1

Annotated Bibliography

Technical Reports

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

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

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

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

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

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

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

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

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

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

This report describes the technical procedures and outcomes of the process to move from grade-level cluster cut scores to grade-level cut scores. Proposed cut scores were

determined mathematically and then reviewed and revised in a standard-setting process involving 75 teachers from 14 WIDA Consortium states.

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

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

Annual Technical Reports for ACCESS for ELLs

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

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

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

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

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

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

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

Yanosky, T., Chong, A., Louguit, M., Olson, E., Choi, Y., MacGregor, D., . . .Kenyon, D. M. (2012). *Annual Technical Report for ACCESS for ELLs® English Language*

Proficiency Test, Series 202, 2010–2011 Administration (WIDA Consortium Annual Technical Report No. 7).

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

Center for Applied Linguistics (2014). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test, Series 301, 2012–2013 Administration* (WIDA Consortium Annual Technical Report No. 9).

Center for Applied Linguistics (2015). *Annual Technical Report for ACCESS for ELLs® English Language Proficiency Test, Series 302, 2013–2014 Administration* (WIDA Consortium Annual Technical Report No. 10).

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

Other Documentation

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Fox, J. & Fairbairn, S. (2011). Test review: ACCESS for ELLs®. *Language Testing, 28* (3): 425–431.

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

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

These documents contain the WIDA Standards and describe the rationale behind and development of the frameworks for large-scale state and classroom assessments. These frameworks comprise English Language Development standards, language domains,

grade-level clusters, language proficiency levels and the model performance indicators upon which ACCESS for ELLs is based. They are meant to guide curriculum development, instruction, and assessment of English language learners.

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

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

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

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

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

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

National Research Council. (2011). *Allocating federal funds for state programs for English language learners.* Washington, DC: The National Academies Press.

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

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

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

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

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

WIDA Consortium. (2007). *English Language Proficiency Standards and Resource Guide, 2007 Edition, PreKindergarten through Grade 12*. Madison, Wisconsin: Board of Regents of the University of Wisconsin System.

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

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

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

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

This report provides an overview on how ACCESS for ELLs is scored and how those scores are reported. Part 1 gives a description of scores for 2014. Part 2 gives suggestions on how states can use scores, as well as examples of score reports to various stakeholders. Part 3 provides guidance on interpreting the reports.

Wolf, M., Kao, J., Griffin, N., Herman, J., Bachman, P., Chang, S., & Farnsworth, T. (2008). Issues in assessing English language learners: English language proficiency measures and accommodation uses—Practice review (Part 2 of 3) (CRESST Report 732). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing Web site: http://www.cse.ucla.edu/products/rsearch.asp.

This paper describes the English language proficiency tests in use in school year 2005–2006, including ACCESS for ELLs, and provides a summary of validity evidence for the tests.

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337-347)**.** Hillsdale, NJ: Lawrence Erlbaum Associates.

This book chapter describes procedures for conducting DIF analysis.

Table of Contents

Part I

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, and it 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.

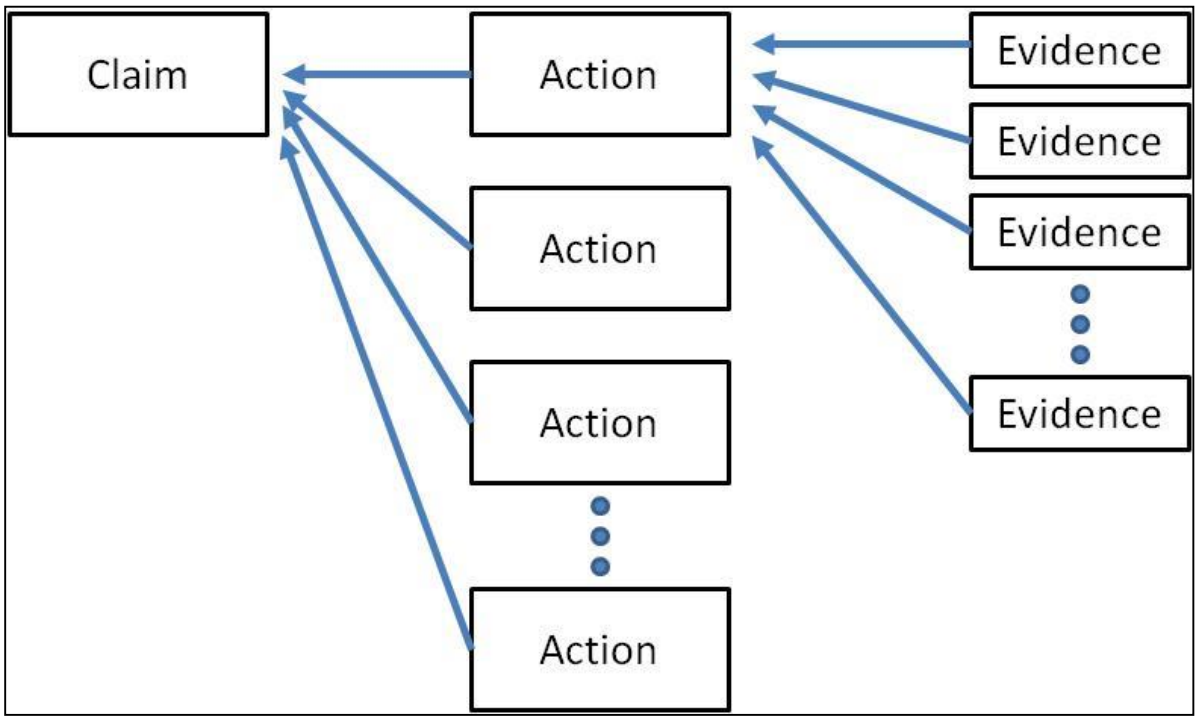

Figure 1. General Argument Structure for Assessment Validation (simplified from Toulmin, 2003).

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.

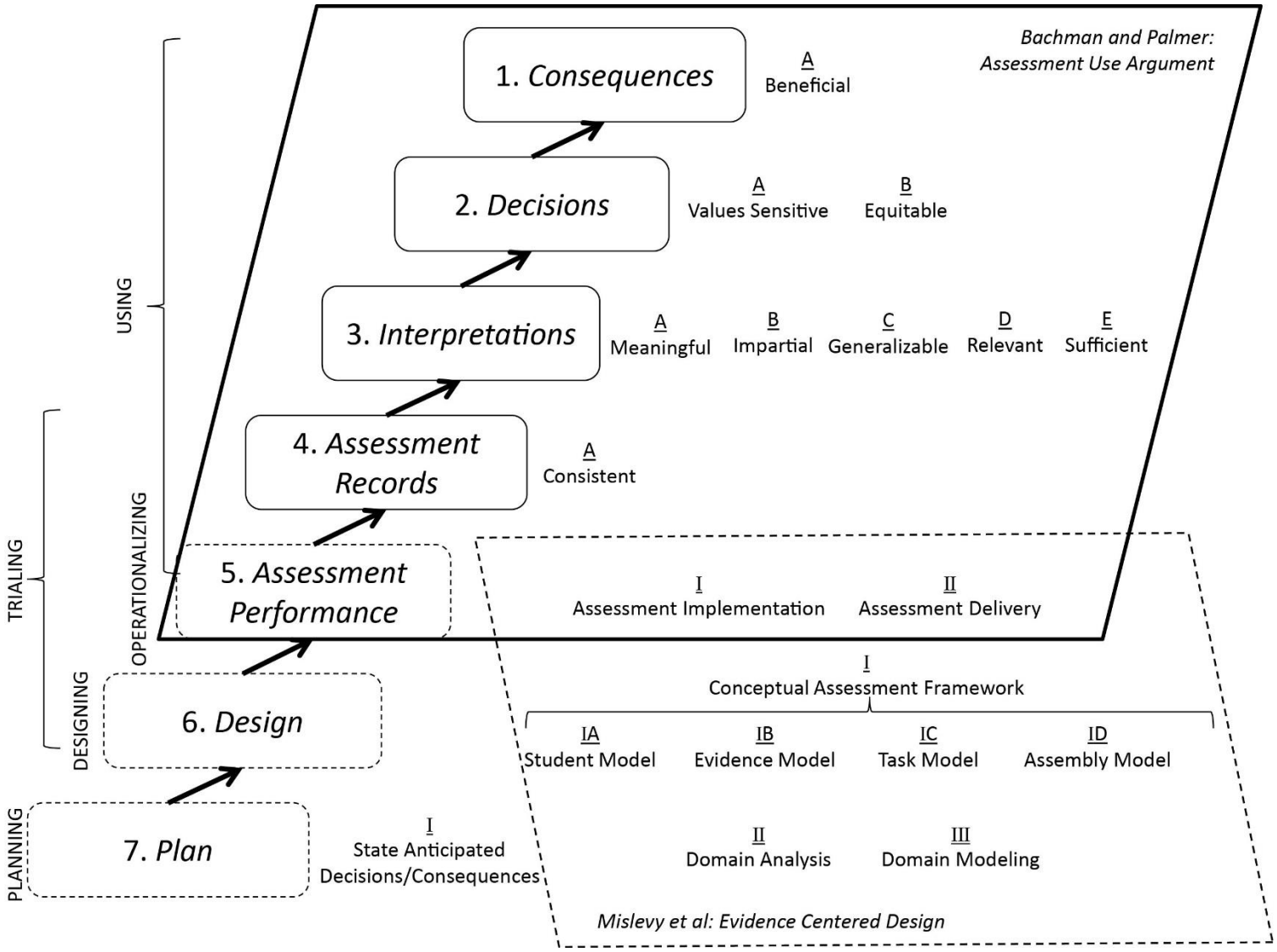

Figure 2. CAL's Validation Framework (based on Bachman & Palmer, 2010; Mislevy, Almond, & Lukas, 2004).

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 English Language Arts: LoLA
- Language of Math: LoMA
- Language of Science: LoSC
- Language of Social Studies: LoSS

Every selected response item and every performance-based task on ACCESS 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 English 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.

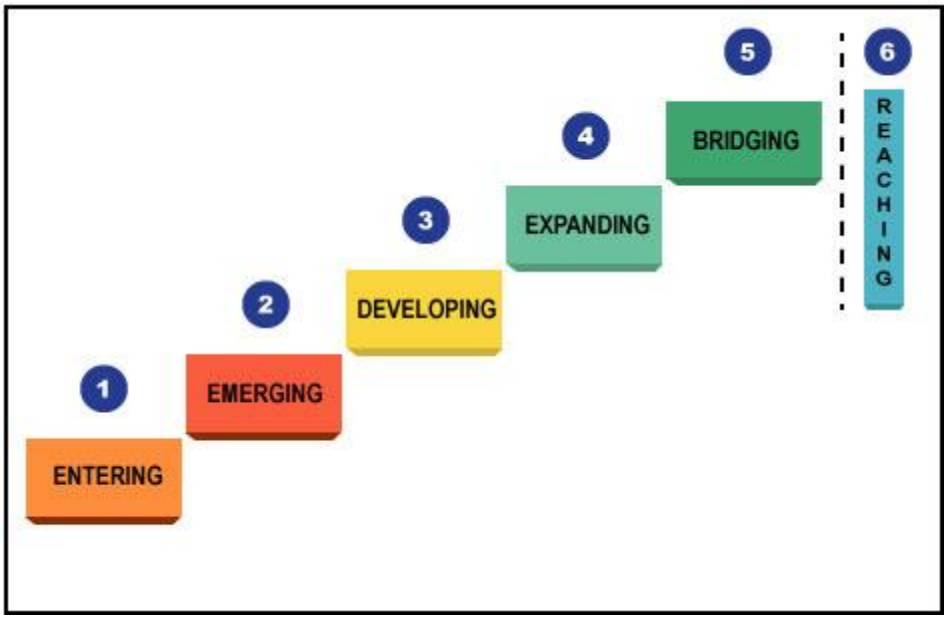

Figure 3. The Language Proficiency Levels of the WIDA ELD Standards.

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. So in ACCESS 2.0 Online, 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 question prompt and answer stem. Listening items are selected-response items, with one key and two distractors as answer choices. Answer choices are primarily illustrations, though for Grades 2–12, items that test listening at PLs 3–5 may consist of short written text that is at two PLs lower than the targeted PL of the item. All items are traditional multiple-choice items.

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 labeled, 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 1–8 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. All items on the operational Reading assessment are traditional multiple choice, though 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 used to make task expectations as clear as possible to students, particularly for tasks in Tier A. For example, the first of a series of questions may already be 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.

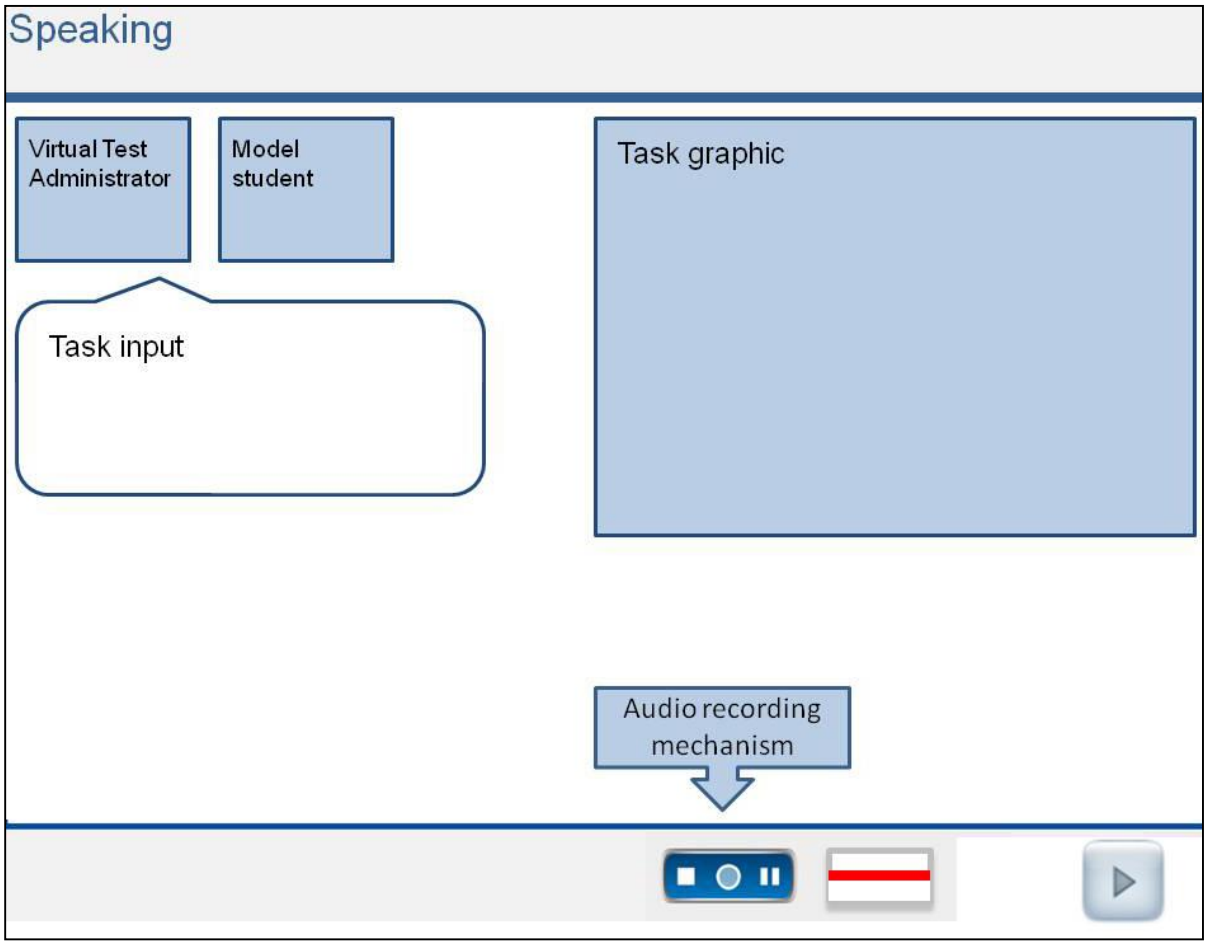

Figure 4. Visualization of the Speaking test screen layout.

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 scores 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 and grade-level cluster. 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 on qualifying sets before scoring live responses. Scoring proficiency is defined as 70% agreement on at least one of two qualifying sets for Speaking; and 70% agreement on two qualifying sets for Writing.
- Once raters are qualified, they are further trained for their grade-level cluster on the specific tasks for which they will rate responses. After this more specific training, they take calibration sets to ensure a consistent interrater understanding of how to apply the scoring scale (See Sections 3.3.1.3 and 0) to their particular tasks.
- DRC uses calibration sets to calibrate the raters to the actual tasks they will be scoring.

Routing Responses to Ensure "Blind" Second Ratings

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

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

Monitoring Scoring (Quality Control)

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

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

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

- For Writing, DRC uses validation sets. These are sets of items seeded into the operational sets that, on a daily basis, monitor how raters are doing when compared to the known ratings of the validity sets. The raters do not know which items are operational and which are from a validation set.

- For Speaking, DRC uses re-calibration sets. Raters take these every day to ensure that they are calibrated, and raters' performances on re-calibration sets are used for monitoring purposes.

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 score point (1 to 6) best describes a response and then determines whether the three descriptors for that 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-task and off-topic responses are scored in their entirety.

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

On 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 (5*1 + 2*6 + 3*7), while a student with raw scores of 7, 6, and 5 on the three tasks would have a total raw score of 34 (7*1 + 2*6 + 3*5). Raw scores on the Tier B/C tests can range from 0–66.

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

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

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 tasks, a Rasch rating scale model is used. Mathematically, this can be represented as

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

where

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

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

B_n = ability of person "n"

D_i = difficulty of task "i"

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

For the Speaking tasks, a Rasch-grouped rating scale model is used. Mathematically, this can be represented as

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

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)

B_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 (e.g., for Speaking, PL 1 tasks are scored as a group on a 0–2 scale and PL 3 and PL 5 tasks are scored as a group on a 0–4 scale).

3.3.2.3 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. Note that for Series 400, scaling equations were used for the Listening and Reading domains (evidence for scale maintenance from ACCES 1.0 to ACCESS 2.0 can be found in Center for Applied Linguistics [2016]). An equipercentile linking method (see Part II.2.2.7) was used to link scale scores between ACCESS 1.0 (Series 303) and ACCESS 2.0 (Series 400) for the Writing and Speaking domains, therefore scaling equations were not used for these domains.

Proficiency Level (PL) scores are interpretations of these scale scores in terms of the PLs described in the WIDA ELD Standards. These interpretations derive from a series of standard setting studies, in which educators reviewed evidence from the test, either in the form of items for the selected response sections (Listening and Reading) or student portfolios for the constructed response sections (Writing and Speaking), to establish cut scores between the PLs. The first standard setting study for ACCESS took place in 2005; it established cut scores for all four domains by grade-level cluster (Kenyon, 2006). The second cut score study took place in 2007; it established cut sores for all four domains by grade level (Kenyon, Ryu, & MacGregor, 2013). These cut scores were used to derive PL scores through Series 400 of ACCESS 2.0 Online. With the release of ACCESS 2.0 Online, another standard setting study was conducted in 2016; the cut scores resulting from this will be used to derive PL scores beginning with Series 401 of ACCESS 2.0 Online.

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 in effect for Series 400, if a Grade 2 student receives a 350 in Listening, it would be interpreted as a PL score of 6.0; if a Grade 5 student receives a 350 in Listening, it would be a 4.0; if a Grade 8 student receives a 350 in

Listening, it would be a 3.2; and if a Grade 12 student receives a 350 in Listening, it would be a 2.3.

Because the bands between cut scores 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.

3.3.2.4 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. 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.

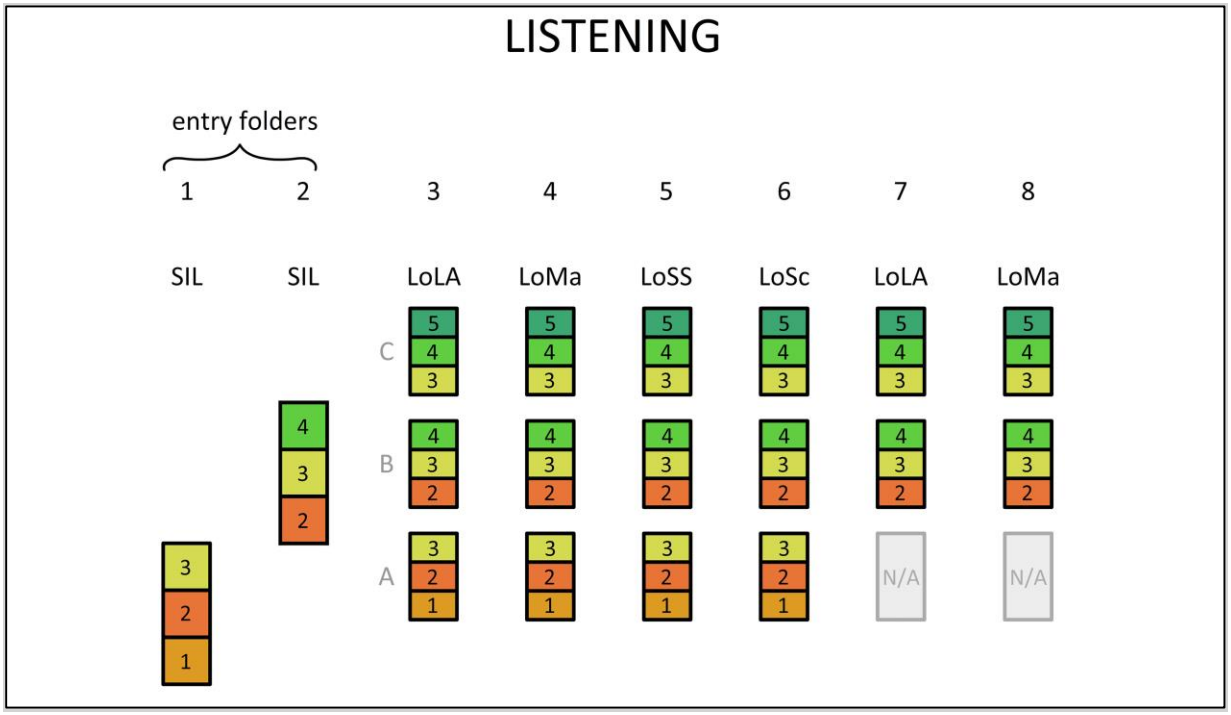

Figure 5. Format of the Listening test.

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.

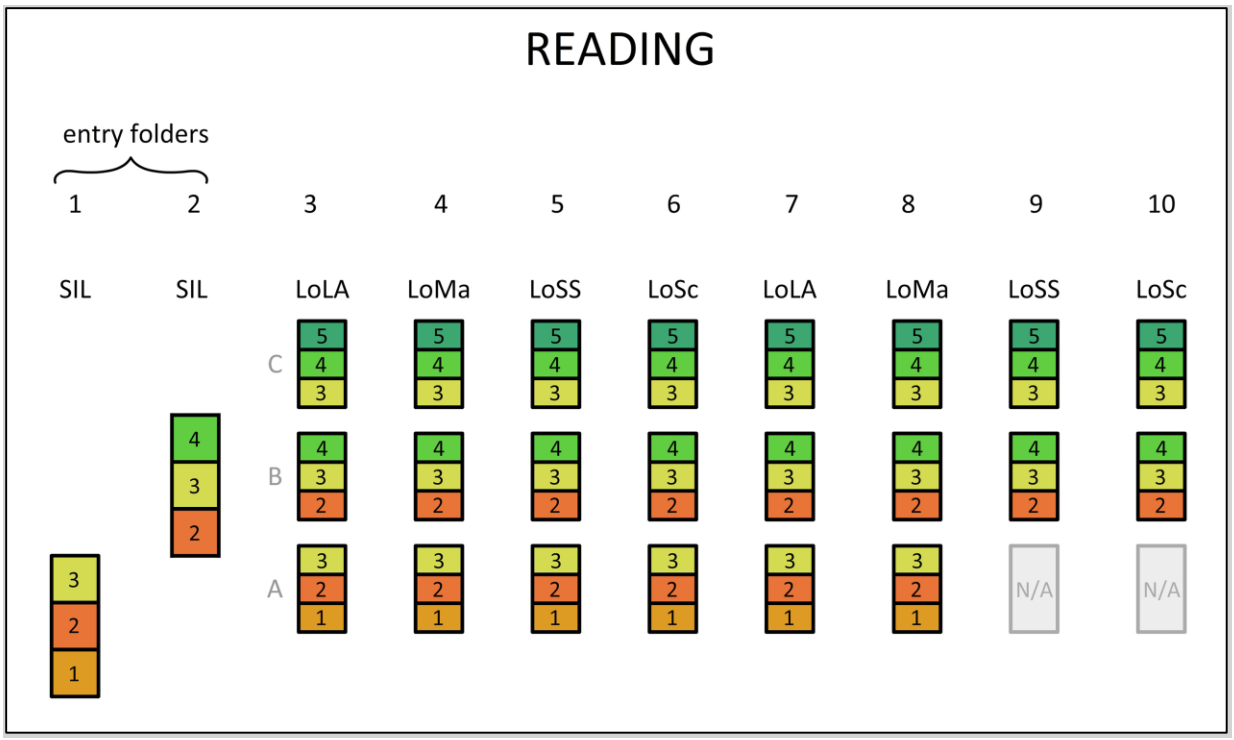

Figure 6. Format of the Reading test.

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 is designed to elicit language at PLs 4–6. Both tiers consist of three tasks. Both tiers include 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.

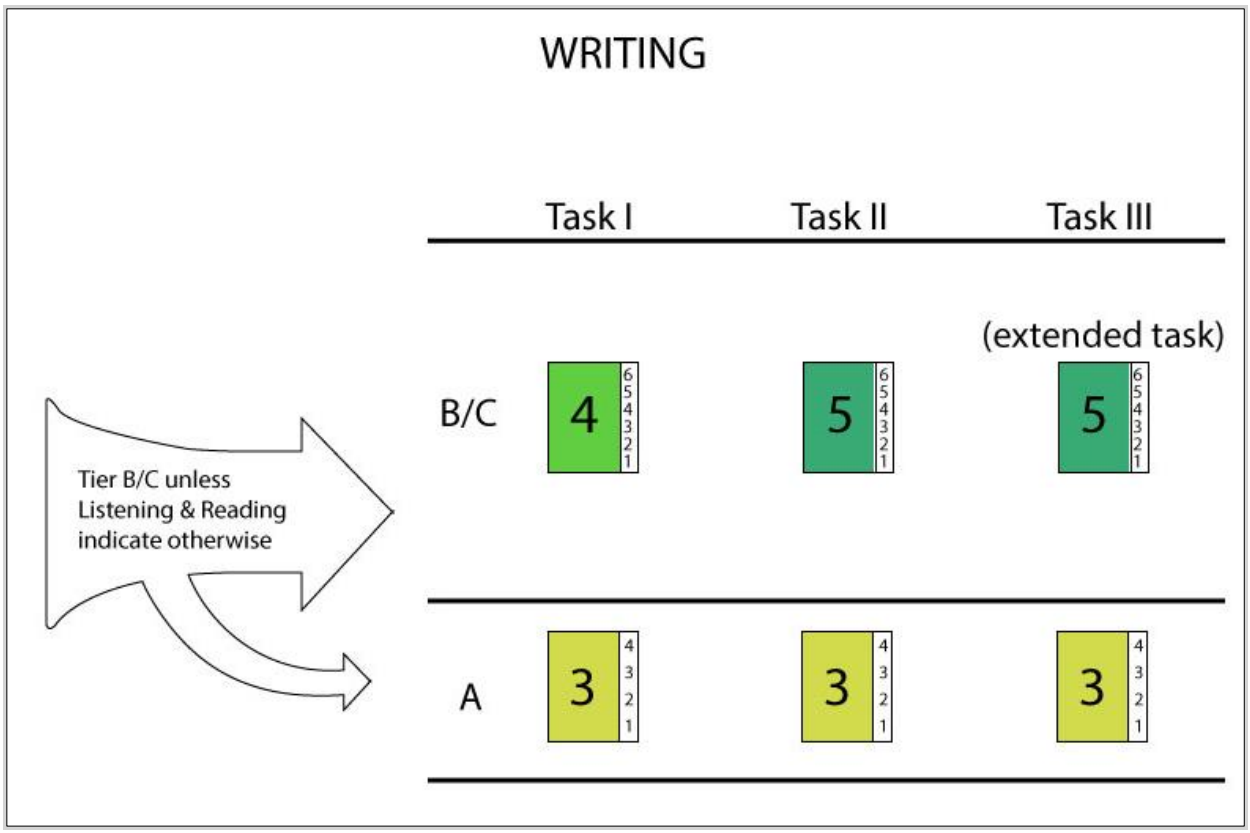

Figure 7. Format of the Writing test.

Note: Grade 1 Tier A follows a different model, and has four tasks, not three. Numbers inside the boxes represent the targeted proficiency level of the task; the smaller numbers on the right edge of each box represent the possible 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. 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.[1] 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.

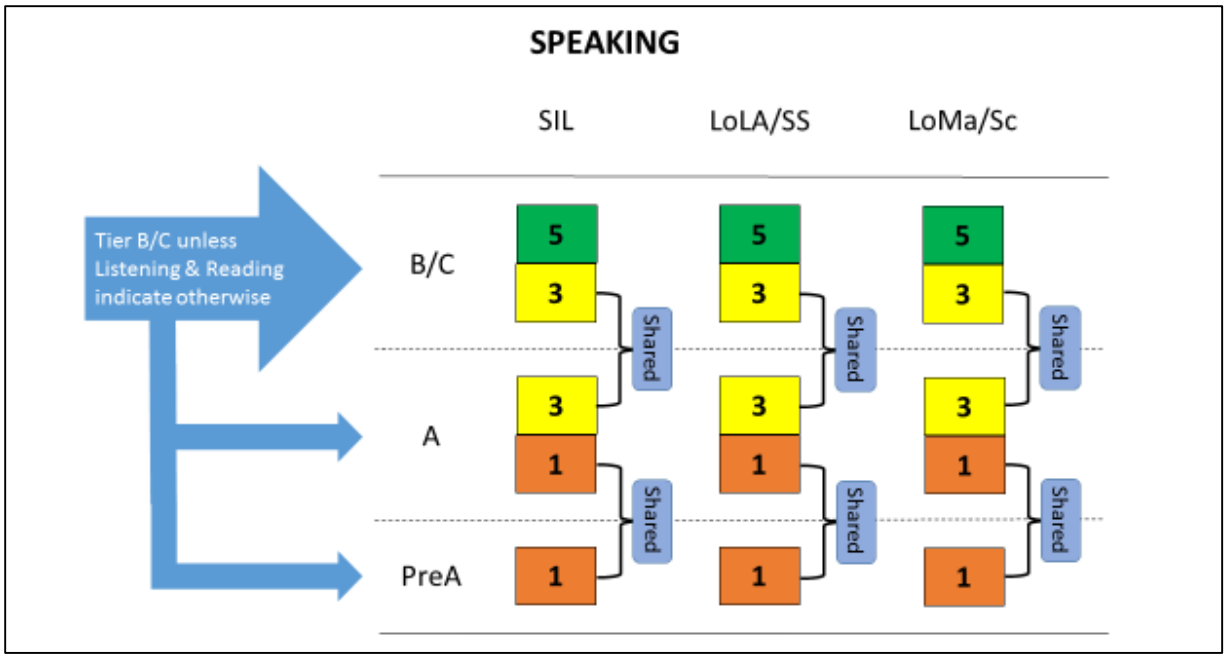

Figure 8. Format of the Speaking test.

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.

[1] Note, however, that on the Series 400 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

In order to minimize exposure of items and improve the quality of the test, in the regular cycle of item development, roughly one-third of all Listening and Reading items are refreshed annually. The item refreshment process spans approximately three years, beginning with the development of the refreshment plan and the updating of item and folder 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 the test series for the current operational year. After field testing, folders of items are analyzed for their psychometric properties, and those that meet established psychometric standards are eligible for inclusion in the next year's operational test.

For the first implementation of ACCESS 2.0 Online, item development in Listening and Reading followed a different trajectory.

The Listening items on Series 400 were developed entirely for ACCESS 2.0 Online. The folders were similar to the media-delivered, paper-and-pencil version of the Listening test in that each item had a stand-alone passage along with the stem and response options. The primary difference ACCESS 2.0 Online is related to the computer administration and its effect on test layout. The paper-based Listening test generally had all three items and associated graphics laid out in the test booklet so that all were visible to the student simultaneously. In the online test, each item appears on its own screen, with its own graphics. For Series 400, students experienced only the operational Listening items—no embedded field test items were included for Listening.

The Reading items on Series 400 were developed based on operational test items from previous test series (predominantly Series 301), 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 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. Additionally, while all of the operational Reading items in Series 400 were traditional multiple-choice items, students also experienced embedded field test items with 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.

For both the Listening and Reading domains, the assessment design calls for two entry folders of three items apiece. Every student is administered these entry folders; based on the student's performance on the entry folders, the student is then routed to the next item on the adaptive test. Entry folders were therefore developed (in the same manner as indicated above for Listening and Reading) to cover the needed proficiency levels (see Figure 5 and Figure 6).

For further information on the field testing of Listening and Reading items, see Center for Applied Linguistics (2016).

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 two rounds of small-scale tryouts, the first led by CAL and the second by teachers in the field. In these tryouts, candidate folders for Grades 4–12 are administered to students using the online interface; 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 field test of Writing folders is conducted. The field test also uses the online interface where applicable, and the field test is administered under standard testing conditions, with responses captured online and rated by DRC raters. For the writing field test, responses are adjudicated at CAL and qualitative analysis of the collected responses is conducted. The main purposes of this small-scale field testing is (a) to confirm that the tasks are working as intended, (b) to identify benchmark 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 (Series 203, 301, and 302). Major differences between the prior test series and Series 400 are that the content is presented entirely on the computer without a live test administrator (except Grades 1–3), and students may have keyboarded their responses (see Section 3.2.3).

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.

The Speaking items on Series 400 were primarily adapted to the computer from both operational items from previous test series and from materials that were not developed to finality for previous test series. A few folders were creates specifically for ACCESS 2.0 Online.

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 all domains, as well as to the Listening items and Speaking tasks. Students use a computer mouse to select or record their answers.

4.2.2 Writing

Writing tasks are delivered to students online. 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 1–3: All responses are handwritten.

- 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 and scaffolding are 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 June 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.

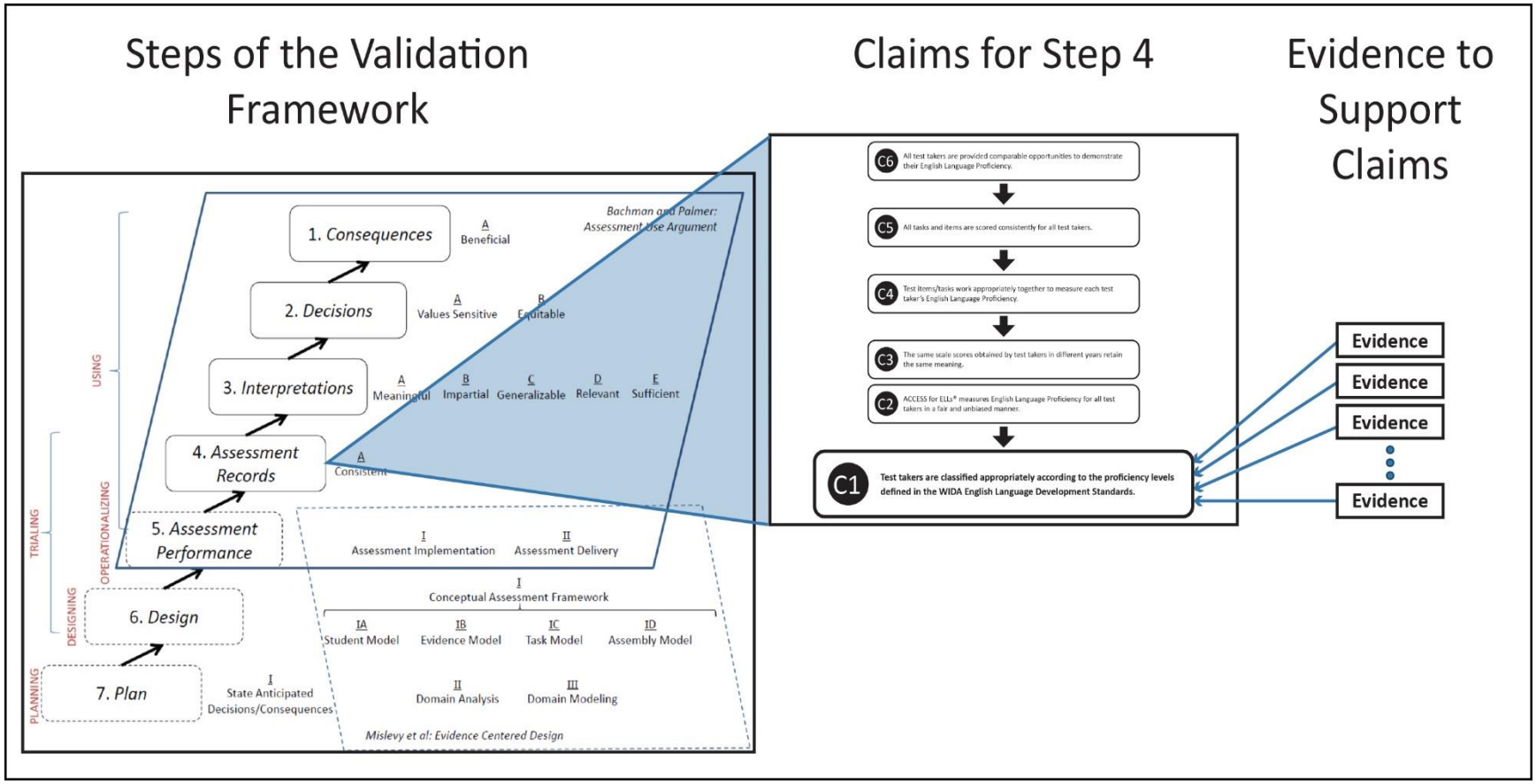

Figure 1. Structure of the argument-based approach supporting *Assessment Records* (Step 4).

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.

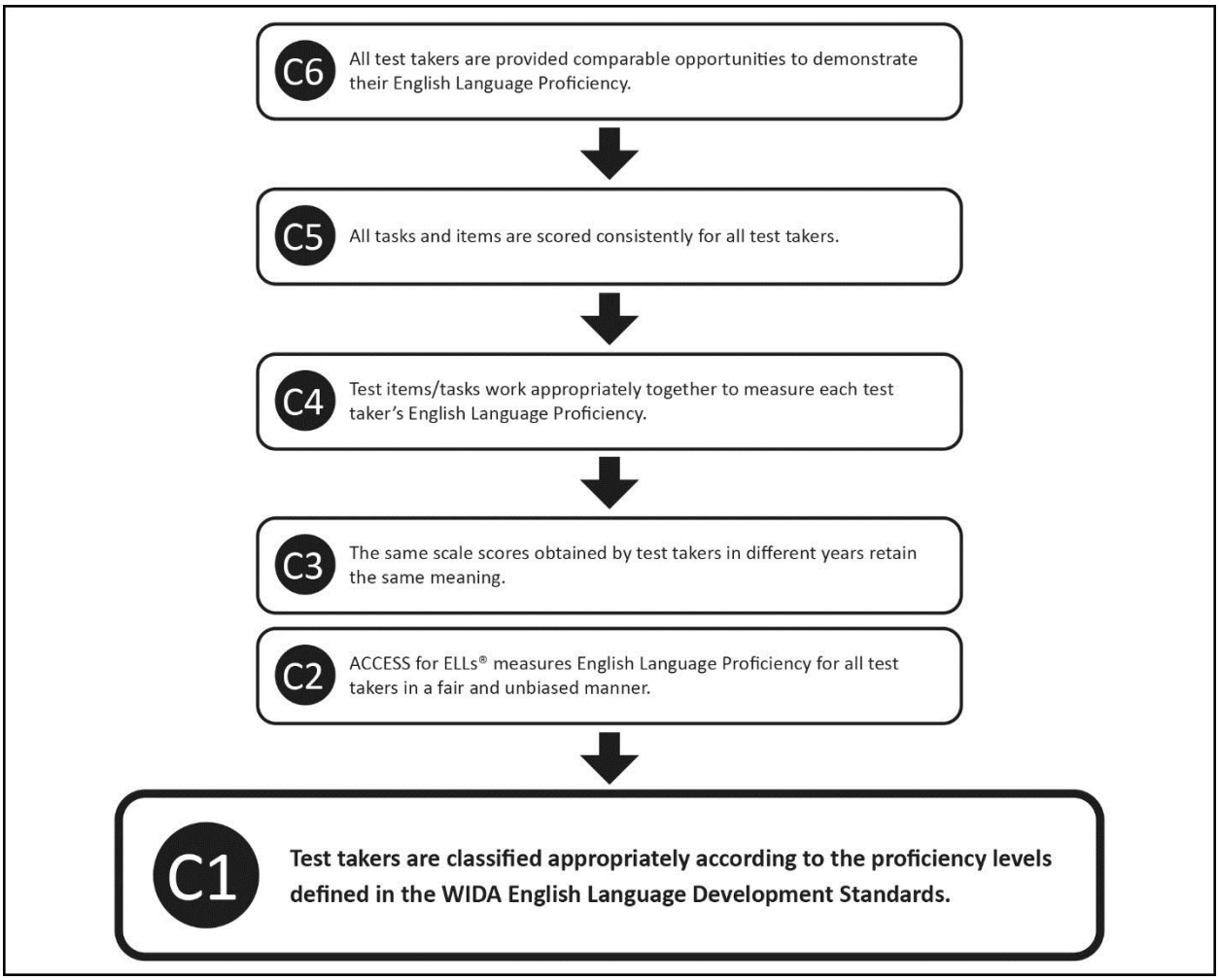

Figure 2. Progression of claims for *Assessment Records* (Step 4).

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 400. ACCESS 2.0 Online Series 400 represents a transitional year between the administration of ACCESS 1.0, which was entirely paper-based, and ACCESS 2.0, which has both online and paper formats. Certain claims, therefore, are relaxed for this transitional year. To maintain the structure of the validation argument, validity claims for the ACCESS assessment program are presented below. Those claims which do not apply to the Series 400 transitional year are presented in grey text.

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 standalone 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.

Note that for Series 400, there were a substantial number of interruptions to student test sessions. See Section 2.1 for further detail on interrupted Series 400 test sessions and actions taken.

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.

---

[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: https://sites.google.com/a/datarecognitioncorp.com/wida-systemstatus/.

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.

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 Test Administrator Manual*.[3] 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.[4] 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[5] 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.[6] 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.

---

[3] The District and School Test Coordinator Test Administration Manual can be found at: https://www.wida.us/assessment/access%202.0/documents/2016DistrictandSchoolTestCoordinatorManual.pdf

[4] The Test Administrator Manual can be found at: https://www.wida.us/assessment/access%202.0/documents/2016TestAdministratorManual.pdf

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

[6] ACCESS for ELLs 2.0 FAQs can be found at: https://www.wida.us/assessment/ACCESS%202.0/administration.aspx#8

Evidence: Part I.3.3.1.2 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.

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 re-calibration sets are administered to raters. Raters take these 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 re-calibration 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 re-calibration sets are administered to raters. Raters take these 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 re-calibration 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.

For the Series 400 transitional year, items administered in the Listening and Reading domains were field tested during the Series 302 and Series 303 operational testing seasons. Series 400 items were anchored using a common person design. This procedure is detailed in the *ACCESS for ELLs Series 400 Listening and Reading Scale Maintenance: Technical Brief* (Center for Applied Linguistics 2016). For the Writing and Speaking domains, equipercentile linking was conducted to link the distribution of scores on Series 400 to the distribution of scores on Series 303, and new scaling equations for these domains will be applied to Series 401.

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.

All Series 400 items and tasks are new.

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

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

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

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

For the transitional Series 400 year, 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, because an equipercentile approach was used for scaling results, scaling equations were not used for Series 400. Scaling was conducted during the Series 400 operational year, and new scaling equations for these domains will be applied to Series 401.

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 using population data.

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

Evidence: If an item shows C-level DIF, a content review panel is convened to examine the content of the item. The panel is composed of diverse members and is chosen carefully so that panelists include male and female members as well as bilingual individuals who speak either English and Spanish or English and another language. The panel then comes to a consensus decision on whether or not the item content is likely to favor or disfavor specific subgroups of students.

Note that for Series 400, new methods for DIF analysis were under development to accommodate analysis for the multistage adaptive assessment, thus DIF analyses were not conducted until after operational testing had been completed. While there were a small number of items which showed C-level DIF and which were reviewed by the panel, there were no items for which the panel review resulted in concern that the item content was likely to result in systematic bias.

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. (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.)

Note that for Series 400, the test forms for Writing and Speaking were linked to series 303 using an equipercentile linking methodology (described in Section 2.2.7). The Test Characteristic Curve (iv) is not appropriate for this year's 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. (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.)

Note that for Series 400, the test forms for Writing and Speaking were linked to Series 303 using an equipercentile linking methodology (described in Section 2.2.7). The Test Characteristic Curve (iv) is not appropriate for this year's assessment.

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

Note that for Series 400, Test Information Function figures (i) are provided for Listening and Reading. Test Information Function figures are not provided for Writing and Speaking, as the equipercentile linking methodology means that the Writing and Speaking task parameters are not on the ACCESS logit scale. Note also that for the CSEM at the cut score tables (ii), the values provided are estimated from the equipercentile linking (see Section 2.2.7 and 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: The standard setting study that established cut scores used through Series 400 is described in Part I.3.3.2.3 of this report and in greater detail in Kenyon, Ryu, and MacGregor (2013).[7]

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.

[7] A new standard setting was conducted in 2016, and the results of that study will be applied starting with Series 401 (see Cook and MacGregor, 2017).

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
	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

4. Test items/tasks work appropriately together to measure each test taker's English Language Proficiency.	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
	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
	f. Routing and placement procedures are in place to ensure that students are administered a test appropriate to their proficiency level.	f. Sections 3.4.1, 3.4.2., 3.4.3, 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. n/a for Series 400
	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. n/a for Series 400
	c. The same scaling equation is applied from year to year to ensure that scale scores are obtained consistently over time	c. For Series 400, applies to Writing and Speaking only. 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 used for scoring and are removed from future test forms.	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
	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
	c. For each test form, analyses are run to confirm that English Language Proficiency is measured with high precision at the cut points pertinent to each tier.	c. 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. Kenyon, Ryu and MacGregor (2013) 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.

Note that in some circumstances there was a mismatch between a student's reported grade and the reported grade-level cluster of the test the student took (e.g., a student who was reported to be in Grade 5 was administered a test for the 6–8 grade-level cluster). In all, 78 students were administered a test form from a grade-level cluster that did not match their reported grade level. Table 2.1 below shows the number of students in each grade who were administered out-of-grade-level tests and the test forms that they were administered. The data for these students was eliminated from all analyses in this report.

Table 2.1

Students Excluded from Analysis due to Grade/Cluster Mismatch

	Cluster					
Grade	**1**	**2-3**	**4-5**	**6-8**	**9-12**	**Total**
1		12	1	0	0	13
2	11		1	0	0	12
3	1		3	1	0	5
4	0	11		0	0	11
5	0	0		3	1	4
6	0	1	11		0	12
7	0	0	2		0	2
8	0	0	0		1	1
9	0	0	0	5		5
10	0	0	0	1		1
11	0	0	0	0		0
12	0	0	0	0		0
Missing	9	2	1	0	0	12
Total	21	26	19	10	2	78

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).

During the 2015–2016 administration of ACCESS 2.0 Online, a substantial number of interruptions occurred during test sessions. Interrupted test sessions were observed by states early in the ACCESS 2.0 test administration cycle (i.e., December 2015). States that observed these early interruptions reported them to WIDA; WIDA then reported them to DRC. Interruptions occurred for a variety of reasons—some caused by test administrators—but many were due to technical issues associated with DRC's online test administration engine. The frequency of interruptions was such that there was concern about the meaning of students'

scores. In January of 2016, WIDA requested online telemetry data from DRC to examine the scope of technical or unexplained interruptions. With early ACCESS 2.0 Online administration data, WIDA conducted preliminary interruption analyses and discovered small but noticeable differences between interrupted and non-interrupted student scores. WIDA decided not to correct for interruptions on score reports; however, WIDA directed CAL to exclude data for students with interrupted test sessions from their psychometric analyses. Hence, psychometric analyses reported in this Annual Technical Report do not included students with interrupted tests sessions. A formal report on the nature and effect of interruptions is forthcoming and will provide the complete scope of interruptions and their effects during the 2015–2016 test administration.

Students who experienced interrupted test sessions 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	14,504	11.30%
	2-3	34,556	26.91%
	4-5	19,663	15.31%
	6-8	30,590	23.82%
	9-12	29,088	22.65%
	Total	128,401	100.00%
Read	1	16,070	6.97%
	2-3	53,382	23.14%
	4-5	42,467	18.41%
	6-8	53,714	23.29%
	9-12	65,027	28.19%
	Total	230,660	100.00%
Writ	1	0	0.00%
	2-3	0	0.00%
	4-5	10,638	25.20%
	6-8	14,132	33.47%
	9-12	17,447	41.33%
	Total	42,217	100.00%
Spek	1	10,785	12.09%
	2-3	22,628	25.37%
	4-5	13,122	14.71%
	6-8	22,397	25.11%
	9-12	20,276	22.73%
	Total	89,208	100.00%

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 36 WIDA states that participated in the ACCESS 2.0 Online operational testing program in 2015–2016. The first row shows the grade-level cluster, the next 36 rows show the number of students in that grade-level cluster who took the test by state, and the final row shows the total number of participants across all 36 states.

Table 3.2.1.1.2 shows participation by grade-level cluster by gender across all 36 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, indicating that the item may not be measuring the same thing as other items on the test. Infit mean square statistics are influenced by more aberrant response patterns and generally indicate a more serious measurement problem. The expectation for both of these statistics is 1.00 and values near 1.00 are not of great concern. Values less than 1.00 indicate that the observations are too predictable and thus redundant, but are not of great concern. High values are of greater concern.

Linacre (2002) provides more guidance on how to interpret these statistics for test items. He writes:

- values greater than 2.0 "distort or degrade the measurement system;"
- values between 1.5 and 2.0 are "unproductive for construction of measurement, but not degrading;"
- values between 0.5 and 1.5 should be considered "productive for measurement;" and

- values below 0.5 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 items and tasks on the test forms have mean square fit statistics in the range of 0.75 and 1.25; thus, they fit the range that is "productive for measurement" according to the guidelines above.

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. 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[8] was missing were excluded from both analyses. Two commonly used procedures for detecting DIF were used: one for dichotomously scored items (Listening and Reading) and one for polytomously scored items (Writing and Speaking).

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

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. If such a factor is identified, that item or 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. As described in Section 2.2.7, the conditional standard errors of measurement from the equated Series 303 Speaking and Writing scale scores are used to approximate of the conditional standard errors of measurement for Series 400 scale scores. These conditional standard errors of measurement are approximations since they were obtained using the equipercentile relationship between Series 303 and 400 scale scores.

The first column shows all possible raw scores. The following column(s) show the corresponding scale score for each grade level in the grade-level cluster. 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

No equating summary is presented for ACCESS 2.0 Online Series 400 because the Series 400 test was not directed equated to Series 303. Listening and Reading scales were maintained using field test data collected from the Series 400 Listening and Reading field test. Scaling analyses were conducted to ensure that scores on the operational ACCESS 2.0 Listening and Reading tests remained on the original ACCESS score scale during the transition. The scaling was accomplished using a common-person linking design in which were administered ACCESS 2.0 Listening and Reading field test items contiguously with the operational ACCESS 1.0 test. The scaled field test items served as the item pool for creating Series 400. Verification studies were planned to refine the Listening and Reading field test parameters using Series 400 operational data. Due to the test interruption issues, however (see Section 2.1), field test parameters were used to score students in order to meet the score report timeline. For more information on the scaling, see the *ACCESS for ELLs Series 400 Listening and Reading Scale Maintenance: Technical Brief* (Center for Applied Linguistics, 2016).

The Series 400 Speaking and Writing tests were designed to measure the same constructs and had the same specifications as Series 303. However, several changes have been made to the Series 400 Speaking and Writing scoring scales and scoring procedures such that the reporting scales cannot be adequately maintained through traditional scaling procedures (Mislevy, 1992). An equipercentile linking study (Kolen & Brennan, 2004) was conducted to link the Series 400 and Series 303 Speaking and Writing scale scores in order to maintain the ACCESS Speaking and Writing score distribution. The linking study adapted a process for concordance that was proposed by Pommerich, Hanson, Harris, and Sconing (2004) and seen in Pommerich (2007). The main analysis involves linking the scale score distribution of Series 400 early return data to that of the Series 303 population data. The computer software program LEGS (*L*inking with *E*quivalent *G*roups or *S*ingle Group Design; Brennan, 2004) was used in conducting the linking.

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

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

Since the goal of the equipercentile procedure is to preserve the distribution of the ACCESS Series 303 Writing and Speaking scale scores, the proportion of students at each observable scale score and WIDA proficiency level is constrained to be more or less the same between series at the level where the linking was conducted. Such an approach provides stability for the ACCESS 2.0 Series 400 Writing and Speaking scores.

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. 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.

For the Writing and Speaking domains, no test characteristic curve is presented for Series 400. As described in Section 2.2.7, a temporary logit scale was created for the Writing and Speaking tests solely for the purpose of conducting the linking analyses. These Writing and Speaking temporary logit scales are not on the ACCESS scales. Therefore, it is not informative to present the test characteristic curve for the Series 400 Writing and Speaking 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 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.

For Series 400, Figure I is provided by grade-level cluster for Listening and Reading only, since Writing and Speaking task parameters are not on the ACCESS logit scale. As described in Section 2.2.7, a temporary logit scale was created for the Writing and Speaking tests solely for the purpose of conducting the linking analyses. These temporary logit scales and the ACCESS cut scores are not on the same scale. Therefore, it is not appropriate to present the test information function curves for the Series 400 Writing and Speaking domains.

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:

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

where

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

σ_θ^2 is the variance of the distribution of student measure,

$CSEM_{observed}^2$ 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 Speaking and Writing. As described in Section 2.2.7, the CSEM from the equated Series 303 Speaking and Writing scale scores are used to approximate the CSEM for Series 400 scale scores. These CSEMs are approximations since there were obtained through the equipercentile relationship between Series 303 and 400 scale scores. Since the Series 303 Speaking test was not tiered, the CSEM at the cut scores for the Speaking test are presented by grade-level cluster.

For Writing, the values are presented by tier. From Table K, it is possible to identify how well the different Writing 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 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 rows 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 row provides overall indices related to the accuracy and consistency of classification, as well as Cohen's kappa. The second row of information shows accuracy and consistency information conditional on proficiency level. The third provides indices of classification accuracy and consistency at the cut points. These indices are perhaps the most important of all when using any of these as an absolute cut-point (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 was no test takers get placed into the proficiency level and accuracy of classification conditional on that level can not be computed, a 'NA' has been placed in the table. In addition, there are a few cases where due to the small percentage of test takers placed into the proficiency level and the range of observed scale scores, accuracy of classification conditional on that level can not 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 rows of information related to the accuracy and consistency of placement into the WIDA language proficiency levels for each composite score. The first row provides overall indices related to the accuracy and consistency of classification, as well as Cohen's kappa. The second row shows accuracy and consistency information conditional per proficiency level. The third row provides indices of classification accuracy and consistency at the cut points. These indices are perhaps the most important of all when using any of these as an absolute cut-point (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 can not be computed. In this case 'NA' 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 can not 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 descriptives 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 Figure 1. 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.

Figure 1: Overview of tables and figures in Section 3.3 Analysis of Domain Scores

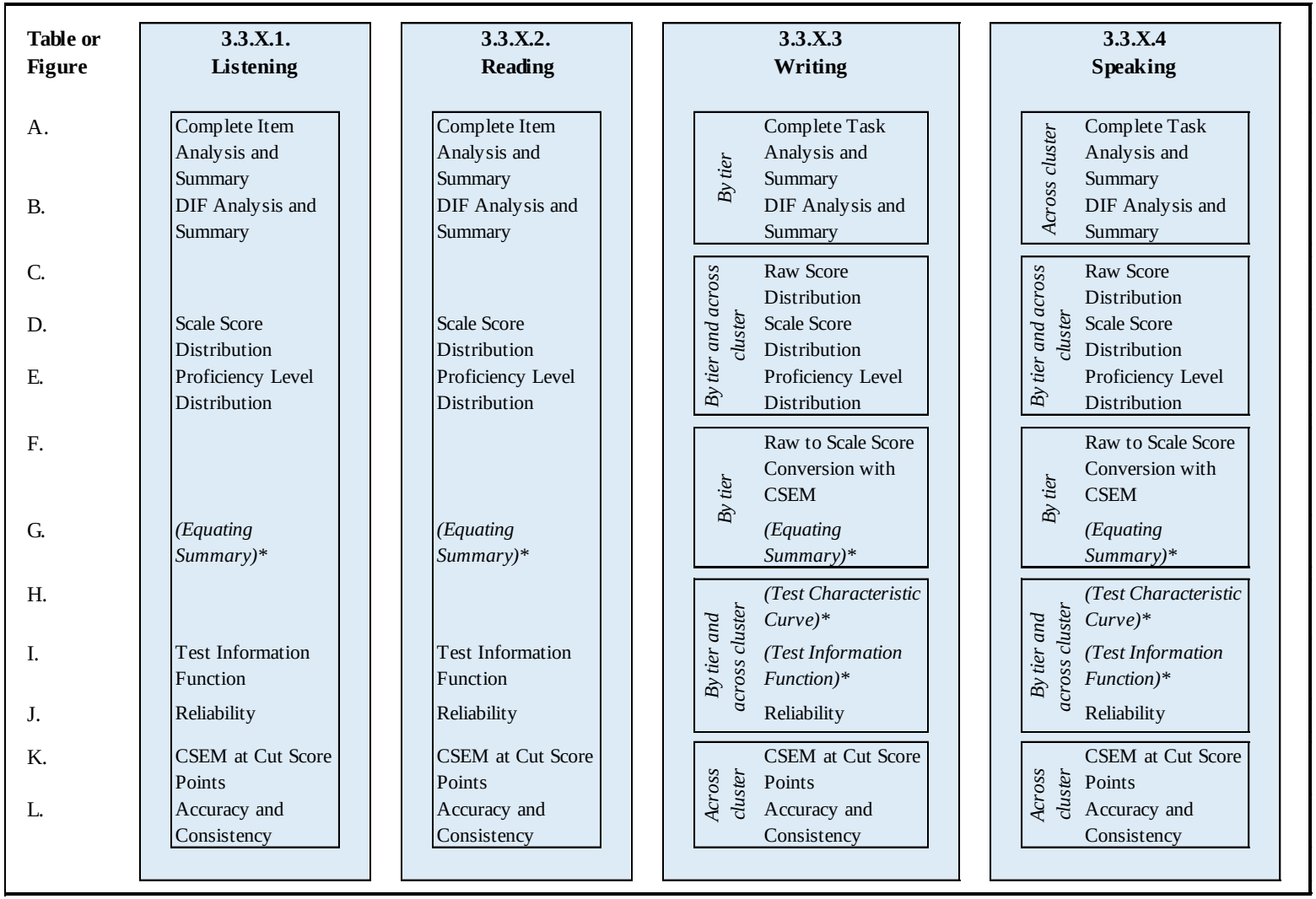

*Table is not produced for Series 400

NOTE: By tier means that a table is presented for each tier of the grade-level cluster. Across cluster means that one table is presented for the grade-level cluster. By tier and across cluster means that a table is produced for each tier and a table is also presented for the grade-level cluster.

Tables or figures within each domain subsection have letter designations; these are aligned on the left side of the figure. Table 3.1.D provides additional detail on the table and figure letter denotation conventions.

Table 3.1D
Naming conventions for tables and figures in Section 3.3. *Analysis of Domain Scores*

Complete Item/Task Analysis and Summary		Table A
DIF Analysis and Summary		Table B
Raw Score Distribution (Speaking and Writing only)	Figure C	Table C
Scale Score Distribution	Figure D	Table D
Proficiency Level Distribution	Figure E	Table E
Raw Score to Scale Score Conversion with CSEM (Speaking and Writing only)		Table F
Equating Summary		Table G
Test Characteristic Curve	Figure H	
Test Information Function	Figure I	
Reliability		Table J
CSEM at Cut Score Points		Table K
Accuracy and Consistency		Table L

Note that 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, this table is provided once, across the tiers in the cluster. 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 and Table G (*Raw Score to Scale Score Conversion with SEM, Equating Summary*) are provided by Tier only. 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.

Note additionally that there are specific tables or figures which do not apply to Series 400 Online (for full explanation, see Section 2.2). These are marked accordingly in Figure 1.

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 S400 Online

State	Cluster					Total
	1	2-3	4-5	6-8	9-12	
AK	1,217	2,598	1,776	2,018	2,087	9,696
AL	2,397	4,052	1,705	1,816	2,142	12,112
CO	8,334	17,133	12,043	14,131	11,672	63,313
DC	997	1,627	720	920	1,234	5,498
DE	1,917	2,999	1,150	1,051	1,224	8,341
GA	12,464	22,822	11,577	11,916	10,335	69,114
ID	1,760	3,117	2,281	2,205	2,135	11,498
IL	16,083	42,065	22,560	20,386	18,867	119,961
IN	7,424	13,366	6,497	9,004	8,690	44,981
KY	3,302	5,849	2,904	2,934	3,389	18,378
MA	4,111	8,009	5,837	7,404	7,860	33,221
MD	9,779	17,357	7,297	8,425	11,492	54,350
ME	462	986	742	793	799	3,782
MI	9,463	19,294	13,867	16,926	16,584	76,134
MN	8,161	16,427	11,247	12,257	11,357	59,449
MO	4,154	7,812	4,699	4,535	3,985	25,185
MP	55	160	304	361	214	1,094
MT	327	681	499	620	343	2,470
NC	13,208	27,636	11,904	14,510	14,437	81,695
ND	371	656	381	614	676	2,698
NH	400	838	472	547	750	3,007
NJ	10,080	16,472	7,294	8,767	12,784	55,397
NM	5,200	11,406	7,339	8,634	7,657	40,236
NV	8,980	19,819	14,458	14,561	10,687	68,505
OK	3,078	5,422	2,302	3,837	3,074	17,713
PA	3,769	8,367	6,075	7,775	11,050	37,036
RI	386	727	1,158	1,504	1,871	5,646
SC	2,894	7,918	5,957	8,559	6,536	31,864
SD	506	1,086	407	657	744	3,400
TN	5,864	12,270	6,185	6,300	5,316	35,935
UT	5,699	10,861	5,122	6,396	5,003	33,081
VA	9,031	19,757	10,373	12,618	16,208	67,987
VI	84	178	119	164	198	743
VT	194	367	163	207	312	1,243
WI	5,901	11,919	8,522	7,988	7,048	41,378
WY	365	699	351	382	399	2,196
Total	168,417	342,752	196,287	221,722	219,159	1,148,337

3.2.1.1.2 By Gender

Table 3.2.1.1.2

Participation by Cluster by Gender S400 Online

Cluster		Gender			Total
		F	M	Missing	
1	Count	79,187	87,414	1,816	168,417
	% within Cluster	47.0%	51.9%	1.1%	100.0%
2-3	Count	160,784	178,672	3,296	342,752
	% within Cluster	46.9%	52.1%	1.0%	100.0%
4-5	Count	87,518	106,305	2,464	196,287
	% within Cluster	44.6%	54.2%	1.3%	100.0%
6-8	Count	96,370	122,372	2,980	221,722
	% within Cluster	43.5%	55.2%	1.3%	100.0%
9-12	Count	94,233	121,487	3,439	219,159
	% within Cluster	43.0%	55.4%	1.6%	100.0%
Total	Count	518,092	616,250	13,995	1,148,337
	% within Cluster	45.1%	53.7%	1.2%	100.0%

3.2.1.1.3 By Ethnicity

Table 3.2.1.1.3

Participation by Cluster by Ethnicity S400 Online

Cluster		Hispanic/Non-Hispanic			Total
		Hispanic	Other	Unknown	
1	Count	111,872	50,336	6,209	168,417
	% within Cluster	66.4%	29.9%	3.7%	100.0%
2-3	Count	234,751	97,227	10,774	342,752
	% within Cluster	68.5%	28.4%	3.1%	100.0%
4-5	Count	134,118	54,867	7,302	196,287
	% within Cluster	68.3%	28.0%	3.7%	100.0%
6-8	Count	149,314	63,794	8,614	221,722
	% within Cluster	67.3%	28.8%	3.9%	100.0%
9-12	Count	139,788	70,206	9,165	219,159
	% within Cluster	63.8%	32.0%	4.2%	100.0%
Total	Count	769,843	336,430	42,064	1,148,337
	% within Cluster	67.0%	29.3%	3.7%	100.0%

3.2.1.1.4 By Tier by Domain

Table 3.2.1.1.4

Participation by Cluster by Tier by Domain S400 Online

Cluster			Domain	
			Writing	Speaking
1	Tier	Pre-A	-	1,191
		A	79,064	34,040
		BC	89,353	133,186
	Total		168,417	168,417
2-3	Tier	Pre-A	-	1,415
		A	77,980	29,178
		BC	264,772	312,159
	Total		342,752	342,752
4-5	Tier	Pre-A	-	1,748
		A	18,591	19,048
		BC	177,696	175,491
	Total		196,287	196,287
6-8	Tier	Pre-A	-	4,401
		A	47,052	40,442
		BC	174,670	176,879
	Total		221,722	221,722
9-12	Tier	Pre-A	-	9,465
		A	45,676	67,855
		BC	173,483	141,839
	Total		219,159	219,159

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 S400 Online

State	Grade												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
AK	1,217	1,221	1,377	977	799	685	657	676	729	527	445	386	9,696
AL	2,397	2,189	1,863	996	709	611	578	627	982	564	382	214	12,112
CO	8,334	8,701	8,432	6,677	5,366	4,680	4,731	4,720	4,263	3,083	2,244	2,082	63,313
DC	997	913	714	431	289	311	310	299	644	254	203	133	5,498
DE	1,917	1,698	1,301	685	465	341	340	370	592	314	165	153	8,341
GA	12,464	11,822	11,000	6,705	4,872	3,986	3,866	4,064	5,522	2,536	1,416	861	69,114
ID	1,760	1,695	1,422	1,149	1,132	806	690	709	702	591	473	369	11,498
IL	16,083	17,296	24,769	13,179	9,381	7,266	6,631	6,489	8,172	4,941	3,414	2,340	119,961
IN	7,424	7,534	5,832	3,519	2,978	2,974	2,977	3,053	3,442	1,749	2,211	1,288	44,981
KY	3,302	3,165	2,684	1,724	1,180	925	1,006	1,003	1,524	870	563	432	18,378
MA	4,111	4,128	3,881	3,192	2,645	2,456	2,526	2,422	3,031	2,152	1,507	1,170	33,221
MD	9,779	9,329	8,028	4,181	3,116	2,616	2,889	2,920	5,733	3,363	1,462	934	54,350
ME	462	477	509	396	346	332	220	241	273	205	161	160	3,782
MI	9,463	9,808	9,486	7,541	6,326	5,811	5,671	5,444	5,600	4,415	3,235	3,334	76,134
MN	8,161	8,461	7,966	6,295	4,952	4,229	3,946	4,082	4,334	2,967	2,210	1,846	59,449
MO	4,154	4,128	3,684	2,714	1,985	1,614	1,495	1,426	1,571	1,075	745	594	25,185
MP	55	48	112	163	141	143	106	112	75	67	44	28	1,094
MT	327	358	323	269	230	233	200	187	147	89	59	48	2,470
NC	13,208	13,813	13,823	6,752	5,152	4,483	4,731	5,296	6,938	3,767	2,183	1,549	81,695
ND	371	353	303	207	174	159	222	233	262	152	125	137	2,698
NH	400	423	415	287	185	178	180	189	294	193	155	108	3,007
NJ	10,080	9,072	7,400	4,228	3,066	2,726	2,914	3,127	4,290	3,707	2,840	1,947	55,397
NM	5,200	5,700	5,706	4,141	3,198	2,954	2,814	2,866	3,253	2,089	1,391	924	40,236
NV	8,980	9,889	9,930	8,651	5,807	4,819	4,874	4,868	4,396	2,995	1,944	1,352	68,505
OK	3,078	2,764	2,658	1,408	894	1,042	1,364	1,431	1,348	852	503	371	17,713
PA	3,769	4,327	4,040	3,277	2,798	2,532	2,583	2,660	3,402	2,980	2,570	2,098	37,036
RI	386	424	303	647	511	457	481	566	776	496	329	270	5,646
SC	2,894	3,913	4,005	2,991	2,966	2,848	2,878	2,833	2,952	1,511	1,197	876	31,864
SD	506	562	524	213	194	176	214	267	300	211	133	100	3,400
TN	5,864	6,630	5,640	3,136	3,049	2,397	1,975	1,928	2,428	1,439	859	590	35,935
UT	5,699	5,657	5,204	3,109	2,013	2,164	2,186	2,046	1,763	1,361	1,041	838	33,081
VA	9,031	9,372	10,385	5,822	4,551	4,009	4,115	4,494	7,162	4,410	3,316	1,320	67,987
VI	84	91	87	60	59	59	58	47	86	43	39	30	743
VT	194	188	179	85	78	70	66	71	102	93	62	55	1,243
WI	5,901	5,990	5,929	5,077	3,445	2,721	2,672	2,595	2,976	1,790	1,279	1,003	41,378
WY	365	350	349	201	150	111	127	144	156	93	64	86	2,196
Total	168,417	172,489	170,263	111,085	85,202	73,924	73,293	74,505	90,220	57,944	40,969	30,026	1,148,337

3.2.1.2.2 By Gender

Table 3.2.1.2.2

Participation by Grade by Gender S400 Online

Grade		Gender			Total
		F	M	Missing	
1	Count	79,187	87,414	1,816	168,417
	% within Grade	47.0%	51.9%	1.1%	100.0%
2	Count	81,278	89,445	1,766	172,489
	% within Grade	47.1%	51.9%	1.0%	100.0%
3	Count	79,506	89,227	1,530	170,263
	% within Grade	46.7%	52.4%	0.9%	100.0%
4	Count	49,967	59,807	1,311	111,085
	% within Grade	45.0%	53.8%	1.2%	100.0%
5	Count	37,551	46,498	1,153	85,202
	% within Grade	44.1%	54.6%	1.4%	100.0%
6	Count	32,300	40,601	1,023	73,924
	% within Grade	43.7%	54.9%	1.4%	100.0%
7	Count	31,744	40,531	1,018	73,293
	% within Grade	43.3%	55.3%	1.4%	100.0%
8	Count	32,326	41,240	939	74,505
	% within Grade	43.4%	55.4%	1.3%	100.0%
9	Count	37,910	50,944	1,366	90,220
	% within Grade	42.0%	56.5%	1.5%	100.0%
10	Count	24,754	32,359	831	57,944
	% within Grade	42.7%	55.8%	1.4%	100.0%
11	Count	18,003	22,145	821	40,969
	% within Grade	43.9%	54.1%	2.0%	100.0%
12	Count	13,566	16,039	421	30,026
	% within Grade	45.2%	53.4%	1.4%	100.0%
Total	Count	518,092	616,250	13,995	1,148,337
	% within Grade	45.1%	53.7%	1.2%	100.0%

3.2.1.2.3 By Ethnicity

Table 3.2.1.2.3

Participation by Grade by Ethnicity S400 Online

Grade		Hispanic/Non-Hispanic			Total
		Hispanic	Other	Unknown	
1	Count	111,872	50,336	6,209	168,417
	% within Grade	66.4%	29.9%	3.7%	100.0%
2	Count	116,049	50,604	5,836	172,489
	% within Grade	67.3%	29.3%	3.4%	100.0%
3	Count	118,702	46,623	4,938	170,263
	% within Grade	69.7%	27.4%	2.9%	100.0%
4	Count	76,453	30,644	3,988	111,085
	% within Grade	68.8%	27.6%	3.6%	100.0%
5	Count	57,665	24,223	3,314	85,202
	% within Grade	67.7%	28.4%	3.9%	100.0%
6	Count	49,773	21,212	2,939	73,924
	% within Grade	67.3%	28.7%	4.0%	100.0%
7	Count	49,328	21,110	2,855	73,293
	% within Grade	67.3%	28.8%	3.9%	100.0%
8	Count	50,213	21,472	2,820	74,505
	% within Grade	67.4%	28.8%	3.8%	100.0%
9	Count	60,949	25,040	4,231	90,220
	% within Grade	67.6%	27.8%	4.7%	100.0%
10	Count	37,634	18,087	2,223	57,944
	% within Grade	64.9%	31.2%	3.8%	100.0%
11	Count	24,491	14,671	1,807	40,969
	% within Grade	59.8%	35.8%	4.4%	100.0%
12	Count	16,714	12,408	904	30,026
	% within Grade	55.7%	41.3%	3.0%	100.0%
Total	Count	769,843	336,430	42,064	1,148,337
	% within Grade	67.0%	29.3%	3.7%	100.0%

3.2.1.2.4 By Tier by Domain

Table 3.2.1.2.4

Participation by Grade by Tier by Domain S400 Online

Grade			Domain	
			Writing	Speaking
1	Tier	Pre-A	-	1,191
		A	79,064	34,040
		BC	89,353	133,186
	Total		168,417	168,417
2	Tier	Pre-A	-	431
		A	65,426	8,273
		BC	107,063	163,785
	Total		172,489	172,489
3	Tier	Pre-A	-	984
		A	12,554	20,905
		BC	157,709	148,374
	Total		170,263	170,263
4	Tier	Pre-A	-	513
		A	7,593	8,204
		BC	103,492	102,368
	Total		111,085	111,085
5	Tier	Pre-A	-	1,235
		A	10,998	10,844
		BC	74,204	73,123
	Total		85,202	85,202
6	Tier	Pre-A	-	539
		A	12,563	10,496
		BC	61,361	62,889
	Total		73,924	73,924
7	Tier	Pre-A	-	1,528
		A	14,284	13,748
		BC	59,009	58,017
	Total		73,293	73,293
8	Tier	Pre-A	-	2,334
		A	20,205	16,198
		BC	54,300	55,973
	Total		74,505	74,505

Grade			Domain	
			Writing	Speaking
9	Tier	Pre-A	-	3,242
		A	23,114	34,805
		BC	67,106	52,173
	Total		90,220	90,220
10	Tier	Pre-A	-	2,423
		A	11,711	20,441
		BC	46,233	35,080
	Total		57,944	57,944
11	Tier	Pre-A	-	1,897
		A	6,072	10,461
		BC	34,897	28,611
	Total		40,969	40,969
12	Tier	Pre-A	-	1,903
		A	4,779	2,148
		BC	25,247	25,975
	Total		30,026	30,026

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 S400 Online

Cluster		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
1	Mean	313.61	289.90	264.29	335.82	325.04	277.02	296.90	291.06
	N	153,766	152,182	168,417	157,417	144,287	152,182	140,229	131,864
2-3	Mean	355.62	325.34	312.62	354.36	355.30	318.91	334.35	329.48
	N	307,921	289,079	342,752	319,709	288,703	289,079	262,758	247,141
4-5	Mean	374.34	348.68	347.58	361.78	368.41	347.62	356.10	353.13
	N	176,543	153,729	134,549	182,996	165,411	106,146	140,191	90,764
6-8	Mean	382.14	354.27	352.68	363.72	373.08	353.03	362.21	358.26
	N	191,073	167,943	204,601	199,203	173,525	157,264	147,878	128,009
9-12	Mean	382.11	371.36	389.41	373.47	377.88	379.41	374.13	377.80
	N	190,006	153,991	198,503	198,614	174,002	142,810	136,116	118,181

3.2.2.1.2 By Cluster by Gender

Table 3.2.2.1.2

Mean Scale Scores by Cluster by Gender S400 Online

Cluster	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
1	F	Mean	315.78	291.56	267.36	340.50	328.47	279.32	298.68	293.66
		N	72,460	70,968	79,187	74,203	68,153	70,968	65,556	61,780
	M	Mean	311.73	288.50	261.57	331.67	322.03	275.04	295.38	288.80
		N	79,610	79,541	87,414	81,526	74,551	79,541	73,105	68,617
	Missing	Mean	308.72	286.41	261.08	330.57	319.41	273.62	292.99	287.20
		N	1,696	1,673	1,816	1,688	1,583	1,673	1,568	1,467
2-3	F	Mean	356.77	327.39	316.93	358.97	358.17	322.12	336.11	332.56
		N	144,832	133,341	160,784	150,109	135,926	133,341	121,589	114,493
	M	Mean	354.67	323.69	308.85	350.37	352.83	316.23	332.93	326.90
		N	160,072	152,885	178,672	166,538	149,968	152,885	138,527	130,170
	Missing	Mean	351.12	317.98	306.59	345.06	347.97	312.09	327.69	322.26
		N	3,017	2,853	3,296	3,062	2,809	2,853	2,642	2,478
4-5	F	Mean	374.73	350.29	350.87	364.98	370.20	350.04	357.32	355.26
		N	79,012	66,993	59,618	81,681	74,091	46,015	61,306	39,547
	M	Mean	374.10	347.53	345.07	359.33	367.07	345.88	355.23	351.61
		N	95,283	84,737	72,885	99,038	89,233	58,450	77,032	49,743
	Missing	Mean	370.77	343.71	341.62	353.40	362.19	341.80	351.59	346.89
		N	2,248	1,999	2,046	2,277	2,087	1,681	1,853	1,474
6-8	F	Mean	382.24	356.47	356.02	365.31	373.88	355.81	363.70	360.34
		N	83,663	71,739	88,827	86,542	75,887	67,186	63,695	55,119
	M	Mean	382.22	352.72	350.20	362.75	372.66	351.04	361.20	356.83
		N	104,806	93,861	113,037	110,056	95,312	87,885	82,089	71,091
	Missing	Mean	376.03	348.94	346.57	351.60	363.98	346.96	356.49	351.39
		N	2,604	2,343	2,737	2,605	2,326	2,193	2,094	1,799
9-12	F	Mean	383.54	373.51	393.04	376.61	380.18	382.37	376.10	380.57
		N	82,231	64,453	84,867	85,260	75,151	59,630	57,345	49,650
	M	Mean	381.13	369.87	386.80	371.23	376.25	377.37	372.77	375.89
		N	104,829	87,164	110,455	110,225	96,151	80,945	76,718	66,729
	Missing	Mean	377.47	367.61	383.16	366.40	372.31	374.09	370.09	372.55
		N	2,946	2,374	3,181	3,129	2,700	2,235	2,053	1,802

3.2.2.1.3 By Cluster by Ethnicity

Table 3.2.2.1.3

Mean Scale Scores by Cluster by Ethnicity S400 Online

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
1	Non-Hispanic Asian	Mean	321.92	302.88	274.32	340.54	331.52	288.56	308.47	301.01
		N	20,101	19,837	21,930	20,574	18,949	19,837	18,374	17,362
	Non-Hispanic Pacific Islander	Mean	304.39	286.95	260.96	327.34	315.27	273.45	291.66	285.27
		N	714	687	768	725	674	687	644	611
	Non-Hispanic Black	Mean	307.65	291.22	263.75	344.78	326.96	277.45	296.21	292.25
		N	6,892	6,913	7,647	6,980	6,334	6,913	6,285	5,791
	Hispanic (Of Any Race)	Mean	311.89	286.64	261.87	333.88	323.21	274.19	294.10	288.53
		N	102,167	101,380	111,872	104,755	95,983	101,380	93,390	87,917
	Non-Hispanic American Indian	Mean	307.86	284.67	257.44	329.72	319.24	271.25	292.03	285.98
		N	1,785	1,691	1,964	1,811	1,648	1,691	1,551	1,438
	Non-Hispanic Multi-racial	Mean	327.55	296.55	269.70	347.83	338.37	283.02	305.82	299.64
		N	726	697	783	719	668	697	654	605
	Non-Hispanic White	Mean	320.18	295.47	269.84	343.07	331.96	282.49	302.61	296.83
		N	15,692	15,345	17,244	16,154	14,755	15,345	14,098	13,276
	Unknown	Mean	305.22	287.33	259.42	324.50	314.92	273.22	292.60	285.32
		N	5,689	5,632	6,209	5,699	5,276	5,632	5,233	4,864
2-3	Non-Hispanic Asian	Mean	364.38	338.12	317.32	359.55	362.25	327.85	346.03	337.86
		N	36,629	35,120	40,370	37,829	34,482	35,120	32,173	30,408
	Non-Hispanic Pacific Islander	Mean	349.31	321.90	313.20	347.61	348.69	317.40	329.61	326.04
		N	1,483	1,434	1,660	1,532	1,382	1,434	1,305	1,212
	Non-Hispanic Black	Mean	350.69	323.15	308.73	358.00	354.72	315.75	331.28	327.16
		N	14,039	13,428	15,915	14,462	12,899	13,428	11,993	11,080
	Hispanic (Of Any Race)	Mean	353.90	322.68	312.37	352.97	353.75	317.41	331.96	327.96
		N	211,255	197,605	234,751	219,403	198,363	197,605	179,851	169,343
	Non-Hispanic American Indian	Mean	351.00	316.39	307.11	346.69	349.22	312.10	327.18	323.29
		N	3,670	3,377	4,188	3,820	3,359	3,377	3,003	2,763
	Non-Hispanic Multi-racial	Mean	368.37	334.28	316.19	365.31	367.11	325.49	344.54	337.80
		N	1,495	1,386	1,657	1,521	1,385	1,386	1,266	1,178
	Non-Hispanic White	Mean	363.57	332.47	314.77	362.11	363.15	323.61	341.70	335.06
		N	29,687	27,594	33,437	31,198	27,821	27,594	24,832	23,340
	Unknown	Mean	343.71	317.85	300.96	337.95	340.79	308.95	325.06	317.69
		N	9,663	9,135	10,774	9,944	9,012	9,135	8,335	7,817

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
4-5	Non-Hispanic Asian	Mean	379.68	358.47	352.57	366.37	373.43	355.67	364.58	360.50
		N	19,714	17,581	12,719	20,388	18,614	10,343	16,192	9,017
	Non-Hispanic Pacific Islander	Mean	369.19	344.79	346.24	357.41	363.48	344.69	351.97	349.90
		N	1,012	937	775	1,018	929	672	864	580
	Non-Hispanic Black	Mean	369.64	344.49	342.87	364.35	367.52	342.73	351.60	348.96
		N	9,285	8,070	5,636	9,669	8,604	4,316	7,245	3,542
	Hispanic (Of Any Race)	Mean	374.06	347.19	347.53	361.16	367.94	346.77	354.98	352.32
		N	120,786	104,643	94,429	125,159	113,230	74,262	95,518	63,537
	Non-Hispanic American Indian	Mean	371.56	343.71	341.67	355.32	363.95	343.10	352.23	349.32
		N	2,632	2,255	2,594	2,771	2,432	1,950	2,020	1,643
	Non-Hispanic Multi-racial	Mean	382.03	354.22	349.16	371.17	377.10	351.53	361.75	358.74
		N	706	613	448	715	661	357	566	323
	Non-Hispanic White	Mean	379.16	353.93	350.43	368.98	374.35	352.15	361.31	358.40
		N	15,794	13,700	12,889	16,573	14,815	10,101	12,331	8,547
	Unknown	Mean	359.61	341.46	337.08	340.17	350.28	339.09	346.12	342.14
		N	6,614	5,930	5,059	6,703	6,126	4,145	5,455	3,575
6-8	Non-Hispanic Asian	Mean	392.51	364.12	358.46	373.32	382.86	360.85	372.19	366.58
		N	22,055	18,254	23,181	22,500	19,930	17,145	16,364	14,189
	Non-Hispanic Pacific Islander	Mean	379.70	351.77	353.52	360.39	369.73	352.75	359.33	356.84
		N	1,084	1,039	1,270	1,196	987	994	858	755
	Non-Hispanic Black	Mean	381.43	350.67	349.61	369.55	375.80	349.52	359.26	356.94
		N	11,224	9,415	11,770	11,757	9,980	8,409	8,095	6,602
	Hispanic (Of Any Race)	Mean	380.05	353.11	352.45	361.91	371.16	352.37	360.83	357.28
		N	128,994	114,855	138,670	134,888	117,627	108,081	101,234	88,231
	Non-Hispanic American Indian	Mean	377.16	349.12	349.14	361.47	369.72	349.28	357.33	355.17
		N	3,287	2,985	3,699	3,452	2,929	2,847	2,575	2,213
	Non-Hispanic Multi-racial	Mean	397.50	361.40	357.99	374.22	386.25	359.43	372.72	366.74
		N	564	494	630	601	510	469	425	375
	Non-Hispanic White	Mean	392.43	359.04	355.78	373.34	383.13	357.06	368.80	364.18
		N	16,488	14,080	17,510	17,174	14,902	12,999	12,333	10,514
	Unknown	Mean	367.19	344.78	338.65	337.37	352.01	340.57	350.45	342.64
		N	7,377	6,821	7,871	7,635	6,660	6,320	5,994	5,130

Cluster	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
9-12	Non-Hispanic Asian	Mean	390.40	382.16	398.57	389.62	390.08	389.45	384.28	388.67
		N	24,992	18,098	25,419	25,662	22,718	16,599	16,141	13,831
	Non-Hispanic Pacific Islander	Mean	381.24	371.06	392.88	376.23	378.36	381.00	373.29	378.72
		N	1,124	956	1,192	1,157	1,040	902	861	764
	Non-Hispanic Black	Mean	378.61	369.45	388.92	382.62	380.37	378.00	371.24	377.35
		N	14,206	9,910	14,071	14,852	12,768	8,658	8,620	6,979
	Hispanic (Of Any Race)	Mean	379.79	369.02	387.67	368.70	374.42	377.46	371.90	375.52
		N	121,388	101,633	128,197	127,163	111,495	95,034	89,890	78,799
	Non-Hispanic American Indian	Mean	385.66	372.95	390.63	373.65	380.26	381.78	376.95	380.69
		N	3,139	2,723	3,469	3,398	2,881	2,564	2,328	2,056
	Non-Hispanic Multi-racial	Mean	390.63	376.85	393.37	383.66	387.41	384.34	380.64	383.83
		N	507	417	545	528	456	390	363	309
	Non-Hispanic White	Mean	394.34	380.15	395.70	386.66	390.57	387.43	384.08	387.49
		N	16,670	13,545	17,306	17,512	15,306	12,493	11,978	10,361
	Unknown	Mean	370.29	361.66	374.74	351.42	360.38	366.67	363.15	362.87
		N	7,980	6,709	8,304	8,342	7,338	6,170	5,935	5,082

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 S400 Online

Grade		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
1	Mean	313.61	289.90	264.29	335.82	325.04	277.02	296.90	291.06
	N	153,766	152,182	168,417	157,417	144,287	152,182	140,229	131,864
2	Mean	347.79	316.22	289.04	355.16	351.79	302.57	325.51	316.91
	N	155,299	145,816	172,489	160,457	145,251	145,816	132,672	124,475
3	Mean	363.60	334.63	336.50	353.54	358.85	335.53	343.36	342.23
	N	152,622	143,263	170,263	159,252	143,452	143,263	130,086	122,666
4	Mean	371.56	345.97	345.90	360.09	366.21	345.54	353.35	351.10
	N	99,842	87,027	76,025	103,495	93,483	60,075	79,273	51,294
5	Mean	377.95	352.22	349.78	363.97	371.26	350.33	359.68	355.77
	N	76,701	66,702	58,524	79,501	71,928	46,071	60,918	39,470
6	Mean	374.84	348.25	348.96	360.92	368.12	348.24	355.95	353.63
	N	63,480	57,603	68,050	66,479	57,670	53,744	50,390	43,390
7	Mean	382.31	354.05	352.89	363.22	372.98	353.04	362.08	358.24
	N	63,119	55,527	67,683	65,871	57,397	52,017	48,883	42,329
8	Mean	389.17	360.82	356.15	366.99	378.07	358.01	368.82	363.04
	N	64,474	54,813	68,868	66,853	58,458	51,503	48,605	42,290
9	Mean	380.75	367.73	386.00	364.72	372.87	375.81	371.02	373.77
	N	78,210	65,310	82,018	82,327	72,047	60,498	57,682	50,085
10	Mean	379.85	370.14	388.27	374.53	377.21	378.33	372.65	376.90
	N	50,098	40,516	52,409	52,443	45,861	37,535	35,733	31,029
11	Mean	384.50	376.65	393.88	382.30	383.44	384.46	378.81	383.08
	N	35,474	27,381	36,896	36,790	32,221	25,437	24,212	21,026
12	Mean	387.27	378.15	395.82	386.02	386.83	386.12	380.60	385.25
	N	26,224	20,784	27,180	27,054	23,873	19,340	18,489	16,041

3.2.2.2.2 By Grade by Gender

Table 3.2.2.2.2

Mean Scale Scores by Grade by Gender S400 Online

Grade	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
1	F	Mean	315.78	291.56	267.36	340.50	328.47	279.32	298.68	293.66
		N	72,460	70,968	79,187	74,203	68,153	70,968	65,556	61,780
	M	Mean	311.73	288.50	261.57	331.67	322.03	275.04	295.38	288.80
		N	79,610	79,541	87,414	81,526	74,551	79,541	73,105	68,617
	Missing	Mean	308.72	286.41	261.08	330.57	319.41	273.62	292.99	287.20
		N	1,696	1,673	1,816	1,688	1,583	1,673	1,568	1,467
2	F	Mean	349.19	317.90	292.90	359.52	354.66	305.35	327.08	319.68
		N	73,353	67,482	81,278	75,699	68,680	67,482	61,575	57,851
	M	Mean	346.58	314.84	285.60	351.39	349.30	300.22	324.21	314.56
		N	80,338	76,789	89,445	83,118	75,071	76,789	69,675	65,291
	Missing	Mean	343.99	310.98	285.63	345.30	344.64	298.31	320.60	311.60
		N	1,608	1,545	1,766	1,640	1,500	1,545	1,422	1,333
3	F	Mean	364.55	337.12	341.50	358.41	361.75	339.31	345.38	345.71
		N	71,479	65,859	79,506	74,410	67,246	65,859	60,014	56,642
	M	Mean	362.83	332.62	332.15	349.35	356.37	332.38	341.74	339.32
		N	79,734	76,096	89,227	83,420	74,897	76,096	68,852	64,879
	Missing	Mean	359.26	326.25	330.79	344.79	351.79	328.37	335.95	334.66
		N	1,409	1,308	1,530	1,422	1,309	1,308	1,220	1,145
4	F	Mean	372.06	347.33	349.06	363.67	368.23	347.82	354.43	353.19
		N	45,035	38,163	33,976	46,641	42,232	26,196	34,871	22,504
	M	Mean	371.26	345.01	343.47	357.30	364.68	343.93	352.60	349.62
		N	53,610	47,779	40,948	55,651	50,145	32,958	43,395	27,985
	Missing	Mean	366.25	340.02	338.55	350.16	358.42	338.49	347.91	343.84
		N	1,197	1,085	1,101	1,203	1,106	921	1,007	805
5	F	Mean	378.26	354.21	353.27	366.71	372.81	352.97	361.14	358.00
		N	33,977	28,830	25,642	35,040	31,859	19,819	26,435	17,043
	M	Mean	377.75	350.77	347.11	361.93	370.13	348.41	358.63	354.18
		N	41,673	36,958	31,937	43,387	39,088	25,492	33,637	21,758
	Missing	Mean	375.92	348.08	345.18	357.04	366.44	345.81	355.98	350.55
		N	1,051	914	945	1,074	981	760	846	669

Grade	Gender		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
6	F	Mean	374.90	350.57	352.67	362.30	368.81	351.26	357.55	355.83
		N	27,911	24,756	29,701	29,018	25,315	23,119	21,824	18,807
	M	Mean	374.90	346.52	346.14	360.11	367.75	345.99	354.77	352.02
		N	34,694	32,013	37,417	36,564	31,566	29,846	27,835	23,952
	Missing	Mean	370.57	345.86	344.39	349.28	360.27	345.05	353.44	349.33
		N	875	834	932	897	789	779	731	631
7	F	Mean	382.33	356.23	356.18	364.50	373.57	355.79	363.44	360.20
		N	27,593	23,592	29,318	28,566	25,096	22,127	20,994	18,209
	M	Mean	382.47	352.54	350.47	362.49	372.73	351.11	361.21	356.91
		N	34,616	31,152	37,436	36,421	31,497	29,165	27,170	23,517
	Missing	Mean	375.72	348.06	346.81	351.81	364.26	346.15	355.30	350.71
		N	910	783	929	884	804	725	719	603
8	F	Mean	389.43	362.97	359.22	369.12	379.21	360.63	370.40	365.17
		N	28,159	23,391	29,808	28,958	25,476	21,940	20,877	18,103
	M	Mean	389.12	359.36	353.93	365.62	377.40	356.20	367.78	361.62
		N	35,496	30,696	38,184	37,071	32,249	28,874	27,084	23,622
	Missing	Mean	382.21	353.44	348.62	353.91	367.67	349.99	361.30	354.42
		N	819	726	876	824	733	689	644	565
9	F	Mean	382.98	370.49	390.37	368.43	375.84	379.43	373.68	377.28
		N	33,074	26,710	34,297	34,565	30,420	24,676	23,716	20,569
	M	Mean	379.35	366.01	383.06	362.35	370.96	373.52	369.37	371.55
		N	43,969	37,607	46,470	46,516	40,560	34,889	33,112	28,775
	Missing	Mean	370.49	358.40	375.44	350.04	360.65	365.69	360.87	362.51
		N	1,167	993	1,251	1,246	1,067	933	854	741
10	F	Mean	380.73	371.80	391.40	377.06	378.95	380.78	374.09	379.04
		N	21,527	16,751	22,234	22,345	19,647	15,480	14,896	12,860
	M	Mean	379.14	368.92	385.97	372.67	375.89	376.59	371.56	375.34
		N	27,854	23,196	29,403	29,345	25,553	21,519	20,339	17,728
	Missing	Mean	381.06	370.95	386.06	371.96	376.56	377.74	373.64	376.95
		N	717	569	772	753	661	536	498	441
11	F	Mean	385.01	377.89	396.57	384.29	384.70	386.43	379.74	384.74
		N	15,730	11,841	16,147	16,156	14,281	10,987	10,555	9,166
	M	Mean	384.19	375.73	391.94	380.83	382.50	383.06	378.11	381.87
		N	19,040	15,014	19,986	19,893	17,298	13,956	13,195	11,453
	Missing	Mean	381.43	375.42	387.62	378.46	380.77	380.24	377.70	379.91
		N	704	526	763	741	642	494	462	407
12	F	Mean	388.22	379.78	398.88	388.83	388.68	388.60	382.10	387.55
		N	11,900	9,151	12,189	12,194	10,803	8,487	8,178	7,055
	M	Mean	386.51	376.83	393.34	383.68	385.31	384.18	379.38	383.42
		N	13,966	11,347	14,596	14,471	12,740	10,581	10,072	8,773
	Missing	Mean	385.26	378.57	393.28	385.06	385.06	384.52	380.90	384.31
		N	358	286	395	389	330	272	239	213

3.2.2.2.3 By Grade by Ethnicity

Table 3.2.2.2.3

Mean Scale Scores by Grade by Ethnicity S400 Online

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
1	Non-Hispanic Asian	Mean	321.92	302.88	274.32	340.54	331.52	288.56	308.47	301.01
		N	20,101	19,837	21,930	20,574	18,949	19,837	18,374	17,362
	Non-Hispanic Pacific Islander	Mean	304.39	286.95	260.96	327.34	315.27	273.45	291.66	285.27
		N	714	687	768	725	674	687	644	611
	Non-Hispanic Black	Mean	307.65	291.22	263.75	344.78	326.96	277.45	296.21	292.25
		N	6,892	6,913	7,647	6,980	6,334	6,913	6,285	5,791
	Hispanic (Of Any Race)	Mean	311.89	286.64	261.87	333.88	323.21	274.19	294.10	288.53
		N	102,167	101,380	111,872	104,755	95,983	101,380	93,390	87,917
	Non-Hispanic American Indian	Mean	307.86	284.67	257.44	329.72	319.24	271.25	292.03	285.98
		N	1,785	1,691	1,964	1,811	1,648	1,691	1,551	1,438
	Non-Hispanic Multi-racial	Mean	327.55	296.55	269.70	347.83	338.37	283.02	305.82	299.64
		N	726	697	783	719	668	697	654	605
	Non-Hispanic White	Mean	320.18	295.47	269.84	343.07	331.96	282.49	302.61	296.83
		N	15,692	15,345	17,244	16,154	14,755	15,345	14,098	13,276
	Unknown	Mean	305.22	287.33	259.42	324.50	314.92	273.22	292.60	285.32
		N	5,689	5,632	6,209	5,699	5,276	5,632	5,233	4,864

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
2	Non-Hispanic Asian	Mean	357.55	329.70	295.55	360.57	359.33	312.75	338.00	326.37
		N	19,332	18,486	21,264	19,879	18,152	18,486	16,956	15,997
	Non-Hispanic Pacific Islander	Mean	342.09	314.19	290.28	348.73	345.40	302.27	322.15	314.37
		N	755	725	836	775	706	725	670	623
	Non-Hispanic Black	Mean	343.12	314.98	286.54	359.19	351.75	300.74	323.26	315.74
		N	7,161	6,875	8,082	7,330	6,569	6,875	6,147	5,653
	Hispanic (Of Any Race)	Mean	345.45	313.07	287.90	353.53	349.82	300.39	322.58	314.76
		N	104,703	98,017	116,049	108,144	98,057	98,017	89,309	83,855
	Non-Hispanic American Indian	Mean	343.70	309.08	284.35	350.63	347.82	296.97	319.78	312.22
		N	1,818	1,654	2,078	1,900	1,661	1,654	1,462	1,348
	Non-Hispanic Multi-racial	Mean	360.39	324.83	291.66	364.96	362.93	308.20	335.12	324.32
		N	742	689	829	765	693	689	626	590
	Non-Hispanic White	Mean	356.83	323.22	292.58	362.98	360.12	307.83	333.17	323.09
		N	15,592	14,414	17,515	16,304	14,582	14,414	13,004	12,203
	Unknown	Mean	338.20	310.91	282.09	339.87	338.92	296.20	318.51	308.24
		N	5,196	4,956	5,836	5,360	4,831	4,956	4,498	4,206
3	Non-Hispanic Asian	Mean	372.01	347.48	341.56	358.41	365.49	344.63	354.99	350.61
		N	17,297	16,634	19,106	17,950	16,330	16,634	15,217	14,411
	Non-Hispanic Pacific Islander	Mean	356.79	329.78	336.47	346.46	352.12	332.88	337.48	338.39
		N	728	709	824	757	676	709	635	589
	Non-Hispanic Black	Mean	358.57	331.72	331.62	356.78	357.81	331.50	339.71	339.06
		N	6,878	6,553	7,833	7,132	6,330	6,553	5,846	5,427
	Hispanic (Of Any Race)	Mean	362.19	332.15	336.29	352.42	357.59	334.17	341.22	340.91
		N	106,552	99,588	118,702	111,259	100,306	99,588	90,542	85,488
	Non-Hispanic American Indian	Mean	358.16	323.40	329.51	342.80	350.59	326.63	334.20	333.82
		N	1,852	1,723	2,110	1,920	1,698	1,723	1,541	1,415
	Non-Hispanic Multi-racial	Mean	376.23	343.62	340.75	365.67	371.30	342.58	353.76	351.33
		N	753	697	828	756	692	697	640	588
	Non-Hispanic White	Mean	371.04	342.59	339.19	361.17	366.50	340.86	351.07	348.19
		N	14,095	13,180	15,922	14,894	13,239	13,180	11,828	11,137
	Unknown	Mean	350.11	326.10	323.27	335.70	342.94	324.08	332.75	328.70
		N	4,467	4,179	4,938	4,584	4,181	4,179	3,837	3,611
4	Non-Hispanic Asian	Mean	377.08	355.08	350.71	364.41	371.20	352.93	361.32	357.79
		N	11,083	9,956	7,210	11,480	10,439	5,866	9,149	5,112
	Non-Hispanic Pacific Islander	Mean	364.90	341.69	346.05	354.71	359.71	343.47	348.44	347.65
		N	561	522	440	552	511	384	485	330
	Non-Hispanic Black	Mean	366.72	341.91	341.00	363.55	365.57	340.46	348.83	346.77
		N	4,999	4,410	2,989	5,208	4,615	2,309	3,954	1,883
	Hispanic (Of Any Race)	Mean	371.26	344.61	345.93	359.51	365.75	344.87	352.32	350.44
		N	68,821	59,564	53,733	71,311	64,505	42,333	54,295	36,135
	Non-Hispanic American Indian	Mean	367.46	341.19	338.88	352.79	360.50	340.53	349.25	346.51
		N	1,424	1,196	1,390	1,500	1,314	1,027	1,077	870
	Non-Hispanic Multi-racial	Mean	377.06	348.31	347.31	369.09	373.79	346.62	356.03	354.45
		N	399	340	251	396	368	196	316	175
	Non-Hispanic White	Mean	376.20	350.83	348.28	366.65	371.77	349.52	358.29	355.97
		N	8,948	7,772	7,255	9,395	8,394	5,683	7,000	4,823
	Unknown	Mean	357.58	338.99	335.02	338.82	348.77	336.67	344.00	340.12
		N	3,607	3,267	2,757	3,653	3,337	2,277	2,997	1,966

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
5	Non-Hispanic Asian	Mean	383.02	362.90	355.01	368.90	376.29	359.25	368.82	364.05
		N	8,631	7,625	5,509	8,908	8,175	4,477	7,043	3,905
	Non-Hispanic Pacific Islander	Mean	374.53	348.70	346.50	360.61	368.10	346.33	356.48	352.87
		N	451	415	335	466	418	288	379	250
	Non-Hispanic Black	Mean	373.05	347.59	344.97	365.28	369.78	345.34	354.93	351.45
		N	4,286	3,660	2,647	4,461	3,989	2,007	3,291	1,659
	Hispanic (Of Any Race)	Mean	377.77	350.60	349.65	363.34	370.83	349.28	358.49	354.79
		N	51,965	45,079	40,696	53,848	48,725	31,929	41,223	27,402
	Non-Hispanic American Indian	Mean	376.39	346.56	344.89	358.30	368.00	345.96	355.63	352.47
		N	1,208	1,059	1,204	1,271	1,118	923	943	773
	Non-Hispanic Multi-racial	Mean	388.49	361.58	351.51	373.74	381.25	357.52	368.98	363.80
		N	307	273	197	319	293	161	250	148
	Non-Hispanic White	Mean	383.04	358.00	353.20	372.02	377.73	355.52	365.28	361.54
		N	6,846	5,928	5,634	7,178	6,421	4,418	5,331	3,724
	Unknown	Mean	362.05	344.49	339.54	341.78	352.09	342.04	348.71	344.60
		N	3,007	2,663	2,302	3,050	2,789	1,868	2,458	1,609
6	Non-Hispanic Asian	Mean	384.08	357.29	354.32	367.91	376.12	355.48	364.96	361.00
		N	7,146	6,252	7,582	7,386	6,490	5,884	5,548	4,827
	Non-Hispanic Pacific Islander	Mean	368.79	343.30	348.21	352.69	360.88	345.62	350.02	349.21
		N	363	350	412	389	332	333	294	256
	Non-Hispanic Black	Mean	374.64	344.79	345.16	366.81	371.45	344.46	353.33	352.21
		N	3,794	3,265	3,935	3,927	3,355	2,915	2,813	2,267
	Hispanic (Of Any Race)	Mean	372.63	347.07	348.84	359.62	366.32	347.62	354.51	352.68
		N	42,847	39,363	46,070	44,945	39,049	36,872	34,497	29,877
	Non-Hispanic American Indian	Mean	370.74	343.29	345.25	357.64	364.76	344.36	351.65	350.73
		N	1,045	967	1,202	1,114	927	911	810	680
	Non-Hispanic Multi-racial	Mean	387.18	354.29	354.31	369.94	377.87	354.21	364.35	360.44
		N	188	169	212	205	175	163	143	131
	Non-Hispanic White	Mean	385.77	352.59	352.20	369.72	378.21	352.03	362.46	359.41
		N	5,590	4,844	5,961	5,904	5,081	4,456	4,204	3,582
	Unknown	Mean	364.01	342.21	335.68	336.65	350.22	338.26	348.17	340.92
		N	2,507	2,393	2,676	2,609	2,261	2,210	2,081	1,770
7	Non-Hispanic Asian	Mean	393.33	364.04	358.68	373.04	383.15	361.00	372.37	366.82
		N	7,375	6,139	7,722	7,565	6,713	5,744	5,510	4,785
	Non-Hispanic Pacific Islander	Mean	382.02	353.13	354.24	363.21	372.18	354.34	361.07	358.74
		N	363	345	419	395	327	329	287	254
	Non-Hispanic Black	Mean	380.63	350.56	349.66	368.75	374.86	349.52	359.04	356.77
		N	3,586	3,054	3,799	3,817	3,214	2,729	2,601	2,128
	Hispanic (Of Any Race)	Mean	380.19	352.86	352.65	361.18	370.93	352.37	360.68	357.19
		N	42,580	37,909	45,863	44,543	38,857	35,703	33,392	29,103
	Non-Hispanic American Indian	Mean	375.79	348.65	348.70	361.42	368.87	348.79	356.63	354.53
		N	1,092	999	1,211	1,135	979	957	872	758
	Non-Hispanic Multi-racial	Mean	404.91	361.80	359.41	376.62	392.58	359.53	375.54	369.17
		N	187	164	204	196	167	154	140	118
	Non-Hispanic White	Mean	392.56	358.47	356.02	373.80	383.60	356.98	368.32	364.18
		N	5,467	4,673	5,852	5,683	4,911	4,321	4,092	3,479
	Unknown	Mean	366.86	344.30	338.92	337.56	352.33	340.12	349.83	342.50
		N	2,469	2,244	2,613	2,537	2,229	2,080	1,989	1,704

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
8	Non-Hispanic Asian	Mean	399.70	371.49	362.23	378.90	389.09	366.42	379.57	372.22
		N	7,534	5,863	7,877	7,549	6,727	5,517	5,306	4,577
	Non-Hispanic Pacific Islander	Mean	388.41	359.03	357.82	364.97	376.24	358.31	367.40	362.83
		N	358	344	439	412	328	332	277	245
	Non-Hispanic Black	Mean	388.89	356.98	353.90	372.99	380.98	354.87	365.69	361.97
		N	3,844	3,096	4,036	4,013	3,411	2,765	2,681	2,207
	Hispanic (Of Any Race)	Mean	387.22	359.67	355.80	364.91	376.14	357.31	367.51	362.07
		N	43,567	37,583	46,737	45,400	39,721	35,506	33,345	29,251
	Non-Hispanic American Indian	Mean	384.31	355.10	353.21	365.07	375.03	354.33	363.16	359.70
		N	1,150	1,019	1,286	1,203	1,023	979	893	775
	Non-Hispanic Multi-racial	Mean	400.45	368.45	360.29	376.25	388.67	364.93	378.37	371.03
		N	189	161	214	200	168	152	142	126
	Non-Hispanic White	Mean	399.15	366.47	359.28	376.69	387.77	362.45	375.89	369.14
		N	5,431	4,563	5,697	5,587	4,910	4,222	4,037	3,453
	Unknown	Mean	370.84	348.08	341.46	337.93	353.56	343.53	353.55	344.64
		N	2,401	2,184	2,582	2,489	2,170	2,030	1,924	1,656
9	Non-Hispanic Asian	Mean	393.16	382.43	397.82	385.73	389.73	389.43	385.43	388.81
		N	8,715	6,441	8,911	8,928	7,935	5,947	5,786	4,968
	Non-Hispanic Pacific Islander	Mean	381.78	367.20	390.23	368.10	374.38	377.57	370.29	374.88
		N	377	324	401	390	350	306	287	262
	Non-Hispanic Black	Mean	378.23	365.85	385.17	374.65	376.50	374.24	368.62	374.01
		N	5,080	3,803	5,085	5,378	4,587	3,299	3,299	2,642
	Hispanic (Of Any Race)	Mean	379.01	366.00	384.90	360.95	370.14	374.45	369.33	372.06
		N	52,872	45,361	56,046	55,840	48,856	42,352	40,021	35,101
	Non-Hispanic American Indian	Mean	384.92	371.18	388.70	367.40	377.17	379.88	375.77	378.76
		N	1,212	1,025	1,323	1,311	1,117	963	889	785
	Non-Hispanic Multi-racial	Mean	394.45	377.46	393.16	381.82	389.14	384.89	382.50	385.22
		N	200	157	206	198	178	148	146	123
	Non-Hispanic White	Mean	391.47	375.77	392.50	380.27	386.05	383.71	380.15	383.74
		N	6,052	4,956	6,269	6,407	5,601	4,546	4,372	3,776
	Unknown	Mean	360.18	351.06	362.97	328.97	343.82	355.42	352.48	349.57
		N	3,702	3,243	3,777	3,875	3,423	2,937	2,882	2,428
10	Non-Hispanic Asian	Mean	388.05	380.53	397.28	389.13	388.35	387.82	382.29	386.94
		N	6,329	4,580	6,458	6,531	5,753	4,192	4,070	3,487
	Non-Hispanic Pacific Islander	Mean	380.75	368.72	390.86	378.34	379.44	378.79	371.16	377.10
		N	300	258	318	309	276	241	230	198
	Non-Hispanic Black	Mean	376.52	368.60	387.88	383.60	379.35	377.23	370.04	376.26
		N	3,633	2,520	3,607	3,831	3,275	2,197	2,186	1,778
	Hispanic (Of Any Race)	Mean	376.92	367.39	386.11	369.47	373.37	375.98	369.92	374.18
		N	32,620	27,239	34,423	34,183	29,947	25,419	24,050	21,065
	Non-Hispanic American Indian	Mean	385.75	371.95	389.79	374.35	380.23	380.98	376.66	380.45
		N	852	756	974	930	775	726	623	556
	Non-Hispanic Multi-racial	Mean	385.46	371.99	389.42	379.79	382.96	379.53	375.72	378.23
		N	129	120	146	137	115	111	97	80
	Non-Hispanic White	Mean	393.76	380.80	395.45	388.43	391.18	387.65	384.59	388.20
		N	4,283	3,427	4,455	4,499	3,922	3,149	3,032	2,610
	Unknown	Mean	374.78	366.07	379.96	363.89	369.15	372.15	368.02	370.31
		N	1,952	1,616	2,028	2,023	1,798	1,500	1,445	1,255

Grade	Ethnicity		List	Read	Writ	Spek	Oral	Litr	Cphn	Over
11	Non-Hispanic Asian	Mean	389.99	383.76	400.58	393.67	391.93	391.13	385.32	390.29
		N	5,351	3,759	5,432	5,483	4,837	3,417	3,327	2,816
	Non-Hispanic Pacific Islander	Mean	380.50	374.86	396.79	382.66	381.40	384.86	375.97	382.69
		N	261	207	272	268	244	196	192	171
	Non-Hispanic Black	Mean	381.39	375.05	393.41	388.96	384.72	383.42	376.14	382.44
		N	2,976	1,902	2,931	3,052	2,634	1,685	1,651	1,349
	Hispanic (Of Any Race)	Mean	381.84	374.34	392.08	377.67	379.86	382.59	376.62	380.93
		N	21,261	17,049	22,343	22,078	19,371	15,985	15,113	13,257
	Non-Hispanic American Indian	Mean	386.21	375.39	392.77	381.05	384.03	384.42	378.47	383.51
		N	587	534	653	654	544	493	451	396
	Non-Hispanic Multi-racial	Mean	392.51	381.71	398.48	386.40	389.69	389.82	384.11	389.30
		N	99	72	107	108	93	67	66	56
	Non-Hispanic White	Mean	396.31	383.50	398.32	391.11	393.58	390.02	386.75	389.60
		N	3,387	2,647	3,493	3,521	3,089	2,459	2,350	2,057
	Unknown	Mean	381.60	375.26	387.24	375.56	378.57	379.98	376.53	377.97
		N	1,552	1,211	1,665	1,626	1,409	1,135	1,062	924
12	Non-Hispanic Asian	Mean	388.91	382.09	399.46	392.97	390.97	389.82	383.60	388.98
		N	4,597	3,318	4,618	4,720	4,193	3,043	2,958	2,560
	Non-Hispanic Pacific Islander	Mean	381.95	377.46	396.05	380.42	380.44	386.20	378.78	383.58
		N	186	167	201	190	170	159	152	133
	Non-Hispanic Black	Mean	379.12	372.51	392.85	390.24	384.63	381.37	373.40	380.57
		N	2,517	1,685	2,448	2,591	2,272	1,477	1,484	1,210
	Hispanic (Of Any Race)	Mean	386.06	376.60	394.80	382.50	384.56	384.83	379.28	383.83
		N	14,635	11,984	15,385	15,062	13,321	11,278	10,706	9,376
	Non-Hispanic American Indian	Mean	386.66	376.08	394.41	379.01	383.49	384.65	378.44	382.36
		N	488	408	519	503	445	382	365	319
	Non-Hispanic Multi-racial	Mean	387.08	378.88	394.21	390.71	387.29	385.66	380.24	383.24
		N	79	68	86	85	70	64	54	50
	Non-Hispanic White	Mean	398.83	384.39	399.56	392.27	395.65	391.62	388.28	391.65
		N	2,948	2,515	3,089	3,085	2,694	2,339	2,224	1,918
	Unknown	Mean	384.64	378.55	390.38	378.93	381.99	382.90	380.57	381.86
		N	774	639	834	818	708	598	546	475

3.2.2.3 Correlations among Scale Scores by Cluster

Table 3.2.2.3A

Correlations Among Scale Scores: 1 S400 Online

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.426	.591	.435
	N	153,766	140,229	153,766	144,287
Reading	Pearson Correlation		1	.673	.274
	N		152,182	152,182	142,791
Writing	Pearson Correlation			1	.382
	N			168,417	157,417
Speaking	Pearson Correlation				1
	N				157,417

Table 3.2.2.3B

Correlations Among Scale Scores: 2-3 S400 Online

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.603	.511	.484
	N	307,921	262,758	307,921	288,703
Reading	Pearson Correlation		1	.593	.413
	N		289,079	289,079	270,978
Writing	Pearson Correlation			1	.340
	N			342,752	319,709
Speaking	Pearson Correlation				1
	N				319,709

Table 3.2.2.3C

Correlations Among Scale Scores: 4-5 S400 Online

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.625	.585	.575
	N	176,543	140,191	120,943	165,411
Reading	Pearson Correlation		1	.575	.498
	N		153,729	106,146	144,191
Writing	Pearson Correlation			1	.600
	N			134,549	125,531
Speaking	Pearson Correlation				1
	N				182,996

Table 3.2.2.3D

Correlations Among Scale Scores: 6-8 S400 Online

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.705	.593	.548
	N	191,073	147,878	177,631	173,525
Reading	Pearson Correlation		1	.637	.524
	N		167,943	157,264	153,296
Writing	Pearson Correlation			1	.600
	N			204,601	185,023
Speaking	Pearson Correlation				1
	N				199,203

Table 3.2.2.3E

Correlations Among Scale Scores: 9-12 S400 Online

		Listening	Reading	Writing	Speaking
Listening	Pearson Correlation	1	.726	.610	.520
	N	190,006	136,116	173,628	174,002
Reading	Pearson Correlation		1	.687	.553
	N		153,991	142,810	142,106
Writing	Pearson Correlation			1	.619
	N			198,503	181,292
Speaking	Pearson Correlation				1
	N				198,614

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 S400 Online

Cluster	Listening Proficiency Range						Total
	1	2	3	4	5	6	
1	7,966	11,480	28,739	13,639	33,030	58,912	153,766
2-3	5,363	7,619	28,764	28,648	90,027	147,500	307,921
4-5	4,118	5,522	16,282	29,272	65,570	55,779	176,543
6-8	5,374	18,042	45,974	41,020	38,436	42,227	191,073
9-12	13,039	36,791	53,645	42,745	21,728	22,058	190,006

Table 3.2.3.1.1.1B

Proficiency Level by Cluster (Percent): Listening S400 Online

Cluster	Listening Proficiency Range						Total
	1	2	3	4	5	6	
1	5.2%	7.5%	18.7%	8.9%	21.5%	38.3%	100.0%
2-3	1.7%	2.5%	9.3%	9.3%	29.2%	47.9%	100.0%
4-5	2.3%	3.1%	9.2%	16.6%	37.1%	31.6%	100.0%
6-8	2.8%	9.4%	24.1%	21.5%	20.1%	22.1%	100.0%
9-12	6.9%	19.4%	28.2%	22.5%	11.4%	11.6%	100.0%

3.2.3.1.1.2 By Grade

Table 3.2.3.1.1.2A

Proficiency Level by Grade (Count): Listening S400 Online

Grade	Listening Proficiency Range						Total
	1	2	3	4	5	6	
1	7,966	11,480	28,739	13,639	33,030	58,912	153,766
2	2,801	3,573	15,556	11,260	45,203	76,906	155,299
3	2,562	4,046	13,208	17,388	44,824	70,594	152,622
4	1,819	2,887	8,424	15,209	38,157	33,346	99,842
5	2,299	2,635	7,858	14,063	27,413	22,433	76,701
6	933	5,041	15,492	14,557	14,825	12,632	63,480
7	1,835	6,160	15,073	13,243	12,596	14,212	63,119
8	2,606	6,841	15,409	13,220	11,015	15,383	64,474
9	3,955	14,726	20,356	18,033	10,333	10,807	78,210
10	3,252	11,013	15,973	9,655	5,648	4,557	50,098
11	2,805	6,742	10,413	8,423	3,434	3,657	35,474
12	3,027	4,310	6,903	6,634	2,313	3,037	26,224

Table 3.2.3.1.1.2B

Proficiency Level by Grade (Percent): Listening S400 Online

Grade	Listening Proficiency Range						Total
	1	2	3	4	5	6	
1	5.2%	7.5%	18.7%	8.9%	21.5%	38.3%	100.0%
2	1.8%	2.3%	10.0%	7.3%	29.1%	49.5%	100.0%
3	1.7%	2.7%	8.7%	11.4%	29.4%	46.3%	100.0%
4	1.8%	2.9%	8.4%	15.2%	38.2%	33.4%	100.0%
5	3.0%	3.4%	10.2%	18.3%	35.7%	29.2%	100.0%
6	1.5%	7.9%	24.4%	22.9%	23.4%	19.9%	100.0%
7	2.9%	9.8%	23.9%	21.0%	20.0%	22.5%	100.0%
8	4.0%	10.6%	23.9%	20.5%	17.1%	23.9%	100.0%
9	5.1%	18.8%	26.0%	23.1%	13.2%	13.8%	100.0%
10	6.5%	22.0%	31.9%	19.3%	11.3%	9.1%	100.0%
11	7.9%	19.0%	29.4%	23.7%	9.7%	10.3%	100.0%
12	11.5%	16.4%	26.3%	25.3%	8.8%	11.6%	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 S400 Online

	Reading Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	10,746	13,283	33,992	39,388	30,760	24,013	152,182
2-3	8,355	32,016	62,260	28,532	70,725	87,191	289,079
4-5	7,105	14,734	39,398	14,850	34,731	42,911	153,729
6-8	22,444	41,546	48,126	14,107	24,956	16,764	167,943
9-12	31,841	42,268	23,064	10,056	20,784	25,978	153,991

Table 3.2.3.1.2.1B

Proficiency Level by Cluster (Percent): Reading S400 Online

	Reading Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	7.1%	8.7%	22.3%	25.9%	20.2%	15.8%	100.0%
2-3	2.9%	11.1%	21.5%	9.9%	24.5%	30.2%	100.0%
4-5	4.6%	9.6%	25.6%	9.7%	22.6%	27.9%	100.0%
6-8	13.4%	24.7%	28.7%	8.4%	14.9%	10.0%	100.0%
9-12	20.7%	27.4%	15.0%	6.5%	13.5%	16.9%	100.0%

3.2.3.1.2.2 By Grade

Table 3.2.3.1.2.2A

Proficiency Level by Grade (Count): Reading S400 Online

Grade	Reading Proficiency Range						Total
	1	2	3	4	5	6	
1	10,746	13,283	33,992	39,388	30,760	24,013	152,182
2	3,070	14,569	35,895	15,788	39,504	36,990	145,816
3	5,285	17,447	26,365	12,744	31,221	50,201	143,263
4	2,547	7,139	19,882	9,867	21,626	25,966	87,027
5	4,558	7,595	19,516	4,983	13,105	16,945	66,702
6	5,690	13,122	18,703	5,291	9,636	5,161	57,603
7	7,928	12,841	16,756	5,168	7,670	5,164	55,527
8	8,826	15,583	12,667	3,648	7,650	6,439	54,813
9	13,884	15,802	12,033	3,372	9,107	11,112	65,310
10	8,525	12,329	5,756	2,585	5,465	5,856	40,516
11	5,167	8,173	3,073	2,216	3,674	5,078	27,381
12	4,265	5,964	2,202	1,883	2,538	3,932	20,784

Table 3.2.3.1.2.2B

Proficiency Level by Grade (Percent): Reading S400 Online

Grade	Reading Proficiency Range						Total
	1	2	3	4	5	6	
1	7.1%	8.7%	22.3%	25.9%	20.2%	15.8%	100.0%
2	2.1%	10.0%	24.6%	10.8%	27.1%	25.4%	100.0%
3	3.7%	12.2%	18.4%	8.9%	21.8%	35.0%	100.0%
4	2.9%	8.2%	22.8%	11.3%	24.8%	29.8%	100.0%
5	6.8%	11.4%	29.3%	7.5%	19.6%	25.4%	100.0%
6	9.9%	22.8%	32.5%	9.2%	16.7%	9.0%	100.0%
7	14.3%	23.1%	30.2%	9.3%	13.8%	9.3%	100.0%
8	16.1%	28.4%	23.1%	6.7%	14.0%	11.7%	100.0%
9	21.3%	24.2%	18.4%	5.2%	13.9%	17.0%	100.0%
10	21.0%	30.4%	14.2%	6.4%	13.5%	14.5%	100.0%
11	18.9%	29.8%	11.2%	8.1%	13.4%	18.5%	100.0%
12	20.5%	28.7%	10.6%	9.1%	12.2%	18.9%	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 S400 Online

	Writing Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	17,708	86,187	56,383	8,139	0	0	168,417
2-3	13,854	65,869	128,373	107,979	26,414	263	342,752
4-5	3,422	7,058	31,760	83,954	8,312	43	134,549
6-8	12,812	32,217	112,934	46,308	330	0	204,601
9-12	12,858	22,943	63,057	74,405	24,081	1,159	198,503

Table 3.2.3.1.3.1B

Proficiency Level by Cluster (Percent): Writing S400 Online

	Writing Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	10.5%	51.2%	33.5%	4.8%	0.0%	0.0%	100.0%
2-3	4.0%	19.2%	37.5%	31.5%	7.7%	0.1%	100.0%
4-5	2.5%	5.2%	23.6%	62.4%	6.2%	0.0%	100.0%
6-8	6.3%	15.7%	55.2%	22.6%	0.2%	0.0%	100.0%
9-12	6.5%	11.6%	31.8%	37.5%	12.1%	0.6%	100.0%

3.2.3.1.3.2 By Grade

Table 3.2.3.1.3.2A

Proficiency Level by Grade (Count): Writing S400 Online

Grade	Writing Proficiency Range						Total
	1	2	3	4	5	6	
1	17,708	86,187	56,383	8,139	0	0	168,417
2	7,508	55,842	104,064	5,075	0	0	172,489
3	6,346	10,027	24,309	102,904	26,414	263	170,263
4	1,486	4,014	14,376	50,333	5,775	41	76,025
5	1,936	3,044	17,384	33,621	2,537	2	58,524
6	2,504	9,039	29,575	26,654	278	0	68,050
7	4,280	9,691	39,299	14,362	51	0	67,683
8	6,028	13,487	44,060	5,292	1	0	68,868
9	4,458	12,052	19,421	28,466	16,666	955	82,018
10	3,855	4,971	17,703	21,174	4,550	156	52,409
11	2,356	3,464	14,058	14,909	2,063	46	36,896
12	2,189	2,456	11,875	9,856	802	2	27,180

Table 3.2.3.1.3.2B

Proficiency Level by Grade (Percent): Writing S400 Online

Grade	Writing Proficiency Range						Total
	1	2	3	4	5	6	
1	10.5%	51.2%	33.5%	4.8%	0.0%	0.0%	100.0%
2	4.4%	32.4%	60.3%	2.9%	0.0%	0.0%	100.0%
3	3.7%	5.9%	14.3%	60.4%	15.5%	0.2%	100.0%
4	2.0%	5.3%	18.9%	66.2%	7.6%	0.1%	100.0%
5	3.3%	5.2%	29.7%	57.4%	4.3%	0.0%	100.0%
6	3.7%	13.3%	43.5%	39.2%	0.4%	0.0%	100.0%
7	6.3%	14.3%	58.1%	21.2%	0.1%	0.0%	100.0%
8	8.8%	19.6%	64.0%	7.7%	0.0%	0.0%	100.0%
9	5.4%	14.7%	23.7%	34.7%	20.3%	1.2%	100.0%
10	7.4%	9.5%	33.8%	40.4%	8.7%	0.3%	100.0%
11	6.4%	9.4%	38.1%	40.4%	5.6%	0.1%	100.0%
12	8.1%	9.0%	43.7%	36.3%	3.0%	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 S400 Online

	Speaking Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	14,896	38,369	24,842	21,204	18,941	39,165	157,417
2-3	23,873	57,279	38,059	41,559	45,061	113,878	319,709
4-5	18,064	21,284	16,215	20,551	27,816	79,066	182,996
6-8	25,789	21,785	30,884	35,843	23,800	61,102	199,203
9-12	39,030	19,820	22,703	34,722	9,851	72,488	198,614

Table 3.2.3.1.4.1B

Proficiency Level by Cluster (Percent): Speaking S400 Online

	Speaking Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	9.5%	24.4%	15.8%	13.5%	12.0%	24.9%	100.0%
2-3	7.5%	17.9%	11.9%	13.0%	14.1%	35.6%	100.0%
4-5	9.9%	11.6%	8.9%	11.2%	15.2%	43.2%	100.0%
6-8	12.9%	10.9%	15.5%	18.0%	11.9%	30.7%	100.0%
9-12	19.7%	10.0%	11.4%	17.5%	5.0%	36.5%	100.0%

3.2.3.1.4.2 By Grade

Table 3.2.3.1.4.2A

Proficiency Level by Grade (Count): Speaking S400 Online

Grade	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
1	14,896	38,369	24,842	21,204	18,941	39,165	157,417
2	10,717	24,233	16,898	20,979	23,608	64,022	160,457
3	13,156	33,046	21,161	20,580	21,453	49,856	159,252
4	9,961	12,924	9,799	12,749	16,084	41,978	103,495
5	8,103	8,360	6,416	7,802	11,732	37,088	79,501
6	8,002	5,216	10,664	17,422	8,299	16,876	66,479
7	8,868	7,441	12,545	9,430	7,961	19,626	65,871
8	8,919	9,128	7,675	8,991	7,540	24,600	66,853
9	23,568	4,604	4,809	17,183	4,537	27,626	82,327
10	8,931	7,638	7,632	7,049	3,194	17,999	52,443
11	3,928	4,479	6,109	5,501	1,786	14,987	36,790
12	2,603	3,099	4,153	4,989	334	11,876	27,054

Table 3.2.3.1.4.2B

Proficiency Level by Grade (Percent): Speaking S400 Online

Grade	Speaking Proficiency Range						Total
	1	2	3	4	5	6	
1	9.5%	24.4%	15.8%	13.5%	12.0%	24.9%	100.0%
2	6.7%	15.1%	10.5%	13.1%	14.7%	39.9%	100.0%
3	8.3%	20.8%	13.3%	12.9%	13.5%	31.3%	100.0%
4	9.6%	12.5%	9.5%	12.3%	15.5%	40.6%	100.0%
5	10.2%	10.5%	8.1%	9.8%	14.8%	46.7%	100.0%
6	12.0%	7.8%	16.0%	26.2%	12.5%	25.4%	100.0%
7	13.5%	11.3%	19.0%	14.3%	12.1%	29.8%	100.0%
8	13.3%	13.7%	11.5%	13.4%	11.3%	36.8%	100.0%
9	28.6%	5.6%	5.8%	20.9%	5.5%	33.6%	100.0%
10	17.0%	14.6%	14.6%	13.4%	6.1%	34.3%	100.0%
11	10.7%	12.2%	16.6%	15.0%	4.9%	40.7%	100.0%
12	9.6%	11.5%	15.4%	18.4%	1.2%	43.9%	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 S400 Online

	Oral Language Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	8,610	15,960	32,837	24,053	33,274	29,553	144,287
2-3	7,984	16,141	40,556	46,366	70,468	107,188	288,703
4-5	7,594	8,099	19,763	28,878	48,194	52,883	165,411
6-8	12,302	15,974	28,278	41,542	41,161	34,268	173,525
9-12	18,256	26,978	33,541	40,085	32,528	22,614	174,002

Table 3.2.3.2.1.1B

Proficiency Level by Cluster (Percent): Oral S400 Online

	Oral Language Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	6.0%	11.1%	22.8%	16.7%	23.1%	20.5%	100.0%
2-3	2.8%	5.6%	14.0%	16.1%	24.4%	37.1%	100.0%
4-5	4.6%	4.9%	11.9%	17.5%	29.1%	32.0%	100.0%
6-8	7.1%	9.2%	16.3%	23.9%	23.7%	19.7%	100.0%
9-12	10.5%	15.5%	19.3%	23.0%	18.7%	13.0%	100.0%

3.2.3.2.1.2 By Grade

Table 3.2.3.2.1.2A

Proficiency Level by Grade (Count): Oral S400 Online

Grade	Oral Language Proficiency Range						Total
	1	2	3	4	5	6	
1	8,610	15,960	32,837	24,053	33,274	29,553	144,287
2	3,713	7,735	20,586	21,628	33,727	57,862	145,251
3	4,271	8,406	19,970	24,738	36,741	49,326	143,452
4	3,783	4,267	11,624	16,668	25,458	31,683	93,483
5	3,811	3,832	8,139	12,210	22,736	21,200	71,928
6	3,229	4,965	9,254	15,435	14,395	10,392	57,670
7	4,182	5,478	9,787	12,854	14,033	11,063	57,397
8	4,891	5,531	9,237	13,253	12,733	12,813	58,458
9	8,761	11,875	11,421	13,857	14,975	11,158	72,047
10	4,535	7,562	9,142	10,932	8,538	5,152	45,861
11	2,689	4,731	7,015	8,554	5,609	3,623	32,221
12	2,271	2,810	5,963	6,742	3,406	2,681	23,873

Table 3.2.3.2.1.2B

Proficiency Level by Grade (Percent): Oral S400 Online

Grade	Oral Language Proficiency Range						Total
	1	2	3	4	5	6	
1	6.0%	11.1%	22.8%	16.7%	23.1%	20.5%	100.0%
2	2.6%	5.3%	14.2%	14.9%	23.2%	39.8%	100.0%
3	3.0%	5.9%	13.9%	17.2%	25.6%	34.4%	100.0%
4	4.0%	4.6%	12.4%	17.8%	27.2%	33.9%	100.0%
5	5.3%	5.3%	11.3%	17.0%	31.6%	29.5%	100.0%
6	5.6%	8.6%	16.0%	26.8%	25.0%	18.0%	100.0%
7	7.3%	9.5%	17.1%	22.4%	24.4%	19.3%	100.0%
8	8.4%	9.5%	15.8%	22.7%	21.8%	21.9%	100.0%
9	12.2%	16.5%	15.9%	19.2%	20.8%	15.5%	100.0%
10	9.9%	16.5%	19.9%	23.8%	18.6%	11.2%	100.0%
11	8.3%	14.7%	21.8%	26.5%	17.4%	11.2%	100.0%
12	9.5%	11.8%	25.0%	28.2%	14.3%	11.2%	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 S400 Online

	Literacy Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	9,556	66,417	39,717	25,045	9,635	1,812	152,182
2-3	8,351	39,547	86,870	78,834	55,677	19,800	289,079
4-5	2,849	7,921	26,635	39,490	23,148	6,103	106,146
6-8	13,047	32,866	69,653	33,719	7,153	826	157,264
9-12	15,424	29,611	37,912	29,560	21,983	8,320	142,810

Table 3.2.3.2.2.1B

Proficiency Level by Cluster (Percent): Literacy S400 Online

	Literacy Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	6.3%	43.6%	26.1%	16.5%	6.3%	1.2%	100.0%
2-3	2.9%	13.7%	30.1%	27.3%	19.3%	6.8%	100.0%
4-5	2.7%	7.5%	25.1%	37.2%	21.8%	5.7%	100.0%
6-8	8.3%	20.9%	44.3%	21.4%	4.5%	0.5%	100.0%
9-12	10.8%	20.7%	26.5%	20.7%	15.4%	5.8%	100.0%

3.2.3.2.2.2 By Grade

Table 3.2.3.2.2.2A

Proficiency Level by Grade (Count): Literacy S400 Online

Grade	Literacy Proficiency Range						Total
	1	2	3	4	5	6	
1	9,556	66,417	39,717	25,045	9,635	1,812	152,182
2	3,557	28,645	62,922	35,250	13,034	2,408	145,816
3	4,794	10,902	23,948	43,584	42,643	17,392	143,263
4	1,154	3,518	13,136	23,583	14,823	3,861	60,075
5	1,695	4,403	13,499	15,907	8,325	2,242	46,071
6	2,886	10,044	23,118	14,383	2,994	319	53,744
7	4,411	10,699	23,317	11,025	2,328	237	52,017
8	5,750	12,123	23,218	8,311	1,831	270	51,503
9	6,821	11,886	14,068	12,426	11,259	4,038	60,498
10	3,996	8,349	10,454	7,751	5,207	1,778	37,535
11	2,449	5,092	7,534	5,496	3,382	1,484	25,437
12	2,158	4,284	5,856	3,887	2,135	1,020	19,340

Table 3.2.3.2.2.2B

Proficiency Level by Grade (Percent): Literacy S400 Online

Grade	Literacy Proficiency Range						Total
	1	2	3	4	5	6	
1	6.3%	43.6%	26.1%	16.5%	6.3%	1.2%	100.0%
2	2.4%	19.6%	43.2%	24.2%	8.9%	1.7%	100.0%
3	3.3%	7.6%	16.7%	30.4%	29.8%	12.1%	100.0%
4	1.9%	5.9%	21.9%	39.3%	24.7%	6.4%	100.0%
5	3.7%	9.6%	29.3%	34.5%	18.1%	4.9%	100.0%
6	5.4%	18.7%	43.0%	26.8%	5.6%	0.6%	100.0%
7	8.5%	20.6%	44.8%	21.2%	4.5%	0.5%	100.0%
8	11.2%	23.5%	45.1%	16.1%	3.6%	0.5%	100.0%
9	11.3%	19.6%	23.3%	20.5%	18.6%	6.7%	100.0%
10	10.6%	22.2%	27.9%	20.7%	13.9%	4.7%	100.0%
11	9.6%	20.0%	29.6%	21.6%	13.3%	5.8%	100.0%
12	11.2%	22.2%	30.3%	20.1%	11.0%	5.3%	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 S400 Online

	Comprehension Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	3,597	14,382	33,031	23,185	39,056	26,978	140,229
2-3	2,723	15,463	43,633	35,326	73,755	91,858	262,758
4-5	2,959	10,357	26,736	19,947	41,261	38,931	140,191
6-8	9,476	29,682	43,644	20,745	26,690	17,641	147,878
9-12	18,990	36,302	27,144	17,510	18,381	17,789	136,116

Table 3.2.3.2.3.1B

Proficiency Level by Cluster (Percent): Comprehension S400 Online

	Comprehension Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	2.6%	10.3%	23.6%	16.5%	27.9%	19.2%	100.0%
2-3	1.0%	5.9%	16.6%	13.4%	28.1%	35.0%	100.0%
4-5	2.1%	7.4%	19.1%	14.2%	29.4%	27.8%	100.0%
6-8	6.4%	20.1%	29.5%	14.0%	18.0%	11.9%	100.0%
9-12	14.0%	26.7%	19.9%	12.9%	13.5%	13.1%	100.0%

3.2.3.2.3.2 By Grade

Table 3.2.3.2.3.2A

Proficiency Level by Grade (Count): Comprehension S400 Online

Grade	Comprehension Proficiency Range						Total
	1	2	3	4	5	6	
1	3,597	14,382	33,031	23,185	39,056	26,978	140,229
2	939	6,413	22,210	21,851	39,268	41,991	132,672
3	1,784	9,050	21,423	13,475	34,487	49,867	130,086
4	987	4,706	13,586	11,872	24,300	23,822	79,273
5	1,972	5,651	13,150	8,075	16,961	15,109	60,918
6	1,653	9,564	16,311	7,374	10,082	5,406	50,390
7	3,402	9,718	14,883	7,014	8,163	5,703	48,883
8	4,421	10,400	12,450	6,357	8,445	6,532	48,605
9	7,819	14,193	12,259	6,875	8,683	7,853	57,682
10	4,894	10,519	7,253	4,606	4,521	3,940	35,733
11	3,342	6,718	4,389	3,264	3,110	3,389	24,212
12	2,935	4,872	3,243	2,765	2,067	2,607	18,489

Table 3.2.3.2.3.2B

Proficiency Level by Grade (Percent): Comprehension S400 Online

Grade	Comprehension Proficiency Range						Total
	1	2	3	4	5	6	
1	2.6%	10.3%	23.6%	16.5%	27.9%	19.2%	100.0%
2	0.7%	4.8%	16.7%	16.5%	29.6%	31.7%	100.0%
3	1.4%	7.0%	16.5%	10.4%	26.5%	38.3%	100.0%
4	1.2%	5.9%	17.1%	15.0%	30.7%	30.1%	100.0%
5	3.2%	9.3%	21.6%	13.3%	27.8%	24.8%	100.0%
6	3.3%	19.0%	32.4%	14.6%	20.0%	10.7%	100.0%
7	7.0%	19.9%	30.4%	14.3%	16.7%	11.7%	100.0%
8	9.1%	21.4%	25.6%	13.1%	17.4%	13.4%	100.0%
9	13.6%	24.6%	21.3%	11.9%	15.1%	13.6%	100.0%
10	13.7%	29.4%	20.3%	12.9%	12.7%	11.0%	100.0%
11	13.8%	27.7%	18.1%	13.5%	12.8%	14.0%	100.0%
12	15.9%	26.4%	17.5%	15.0%	11.2%	14.1%	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 S400 Online

	Overall Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	5,123	35,285	46,061	28,754	14,060	2,581	131,864
2-3	5,893	19,624	60,177	73,423	63,046	24,978	247,141
4-5	2,997	5,640	17,093	29,900	27,193	7,941	90,764
6-8	9,446	19,977	42,379	38,030	15,925	2,252	128,009
9-12	12,775	21,961	30,347	26,742	18,770	7,586	118,181

Table 3.2.3.2.4.1B

Proficiency Level by Cluster (Percent): Overall S400 Online

	Overall Proficiency Range						
Cluster	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
1	3.9%	26.8%	34.9%	21.8%	10.7%	2.0%	100.0%
2-3	2.4%	7.9%	24.3%	29.7%	25.5%	10.1%	100.0%
4-5	3.3%	6.2%	18.8%	32.9%	30.0%	8.7%	100.0%
6-8	7.4%	15.6%	33.1%	29.7%	12.4%	1.8%	100.0%
9-12	10.8%	18.6%	25.7%	22.6%	15.9%	6.4%	100.0%

3.2.3.2.4.2 By Grade

Table 3.2.3.2.4.2A

Proficiency Level by Grade (Count): Overall S400 Online

Grade	Overall Proficiency Range						Total
	1	2	3	4	5	6	
1	5,123	35,285	46,061	28,754	14,060	2,581	131,864
2	2,526	11,571	42,530	39,934	22,881	5,033	124,475
3	3,367	8,053	17,647	33,489	40,165	19,945	122,666
4	1,305	2,741	9,032	16,668	16,461	5,087	51,294
5	1,692	2,899	8,061	13,232	10,732	2,854	39,470
6	2,196	6,164	13,805	14,806	5,522	897	43,390
7	3,192	6,552	14,669	11,610	5,605	701	42,329
8	4,058	7,261	13,905	11,614	4,798	654	42,290
9	6,021	9,024	10,789	11,022	9,409	3,820	50,085
10	3,140	6,362	8,433	6,892	4,570	1,632	31,029
11	1,993	3,735	6,040	5,017	2,985	1,256	21,026
12	1,621	2,840	5,085	3,811	1,806	878	16,041

Table 3.2.3.2.4.2B

Proficiency Level by Grade (Percent): Overall S400 Online

Grade	Overall Proficiency Range						Total
	1	2	3	4	5	6	
1	3.9%	26.8%	34.9%	21.8%	10.7%	2.0%	100.0%
2	2.0%	9.3%	34.2%	32.1%	18.4%	4.0%	100.0%
3	2.7%	6.6%	14.4%	27.3%	32.7%	16.3%	100.0%
4	2.5%	5.3%	17.6%	32.5%	32.1%	9.9%	100.0%
5	4.3%	7.3%	20.4%	33.5%	27.2%	7.2%	100.0%
6	5.1%	14.2%	31.8%	34.1%	12.7%	2.1%	100.0%
7	7.5%	15.5%	34.7%	27.4%	13.2%	1.7%	100.0%
8	9.6%	17.2%	32.9%	27.5%	11.3%	1.5%	100.0%
9	12.0%	18.0%	21.5%	22.0%	18.8%	7.6%	100.0%
10	10.1%	20.5%	27.2%	22.2%	14.7%	5.3%	100.0%
11	9.5%	17.8%	28.7%	23.9%	14.2%	6.0%	100.0%
12	10.1%	17.7%	31.7%	23.8%	11.3%	5.5%	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 3.3.1.1C
Raw Scores: List 1 S400 Online
n/a

Table 3.3.1.1C
Raw Score Descriptive Statistics: List 1 S400 Online
n/a

Figure 3.3.1.1D
Scale Scores: List 1 S400 Online

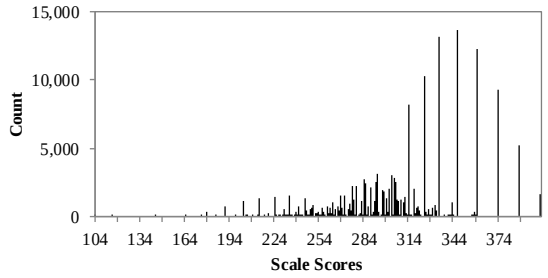

Figure 3.3.1.1E
Proficiency Level: List 1 S400 Online

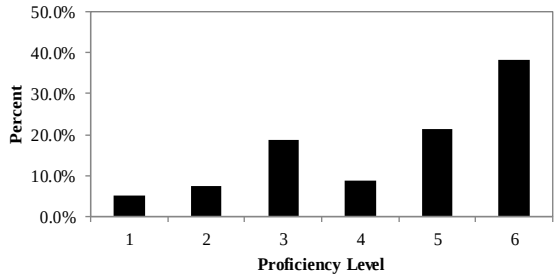

Table 3.3.1.1D

Scale Score Descriptive Statistics: List 1 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	153,766	104	402	313.61	42.85
Total	153,766	104	402	313.61	42.85

Table 3.3.1.1E

Proficiency Level Distribution: List 1 S400 Online

	Grade 1		Total	
Level	**Count**	**Percent**	**Count**	**Percent**
1	7,966	5.18%	7,966	5.18%
2	11,480	7.47%	11,480	7.47%
3	28,739	18.69%	28,739	18.69%
4	13,639	8.87%	13,639	8.87%
5	33,030	21.48%	33,030	21.48%
6	58,912	38.31%	58,912	38.31%
Total	153,766	100.00%	153,766	100.00%

Table 3.3.1.1F
Raw Score to Scale Score Conversion: List 1 S400 Online
n/a

Table 3.3.1.1G
Equating Summary: List 1 S400 Online
n/a for S400

Figure 3.3.1.1H
Test Characteristic Curve: List 1 S400 Online
n/a

Figure 3.3.1.1I
Test Information Function: List 1 S400 Online

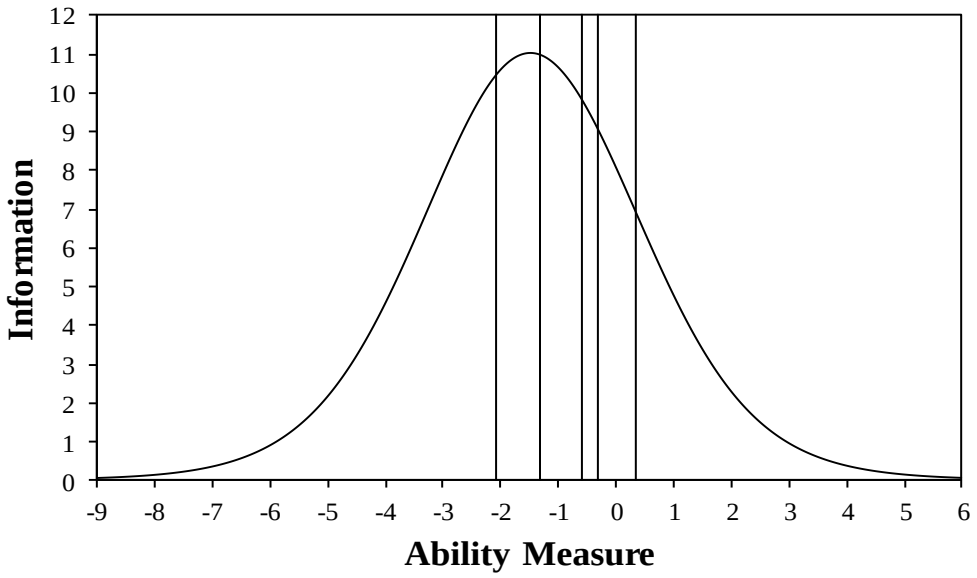

Table 3.3.1.1J
Reliability: List 1 S400 Online

No. of Students	No. of Items	Rasch Reliability Estimate
153,766	54	.76

Table 3.3.1.1K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: List 1 S400 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	1	238	307	16.79	19.76	19.71	0.18
2/3	1	267	403	17.13	17.51	17.27	0.16
3/4	1	295	60	17.88	18.67	17.93	0.18
4/5	1	305	2,522	18.07	18.86	18.14	0.15
5/6	1	330	599	20.63	20.81	20.80	0.05

Table 3.3.1.1L

Accuracy and Consistency of Classification Indices: List (Grade 1) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.550	0.467	0.295		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.750	0.479		
	2	0.356	0.243		
	3	0.464	0.357		
	4	0.189	0.145		
	5	0.400	0.319		
	6	0.818	0.719		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.963	0.007	0.030	0.946
	2/3	0.921	0.032	0.047	0.881
	3/4	0.851	0.064	0.085	0.798
	4/5	0.840	0.071	0.089	0.782
	5/6	0.847	0.087	0.066	0.786

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 3.3.1.2C
Raw Scores: Read 1 S400 Online
n/a

Table 3.3.1.2C
Raw Score Descriptive Statistics: Read 1 S400 Online
n/a

Figure 3.3.1.2D
Scale Scores: Read 1 S400 Online

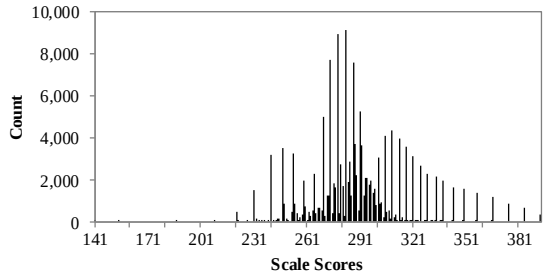

Figure 3.3.1.2E
Proficiency Level: Read 1 S400 Online

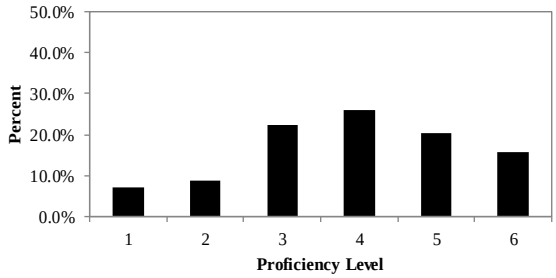

Table 3.3.1.2D

Scale Score Descriptive Statistics: Read 1 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	152,182	141	393	289.90	27.47
Total	152,182	141	393	289.90	27.47

Table 3.3.1.2E

Proficiency Level Distribution: Read 1 S400 Online

	Grade 1		Total	
Level	**Count**	**Percent**	**Count**	**Percent**
1	10,746	7.06%	10,746	7.06%
2	13,283	8.73%	13,283	8.73%
3	33,992	22.34%	33,992	22.34%
4	39,388	25.88%	39,388	25.88%
5	30,760	20.21%	30,760	20.21%
6	24,013	15.78%	24,013	15.78%
Total	152,182	100.00%	152,182	100.00%

Table 3.3.1.2F
Raw Score to Scale Score Conversion: Read 1 S400 Online
n/a

Table 3.3.1.2G
Equating Summary: Read 1 S400 Online
n/a for S400

Figure 3.3.1.2H
Test Characteristic Curve: Read 1 S400 Online
n/a

Figure 3.3.1.2I
Test Information Function: Read 1 S400 Online

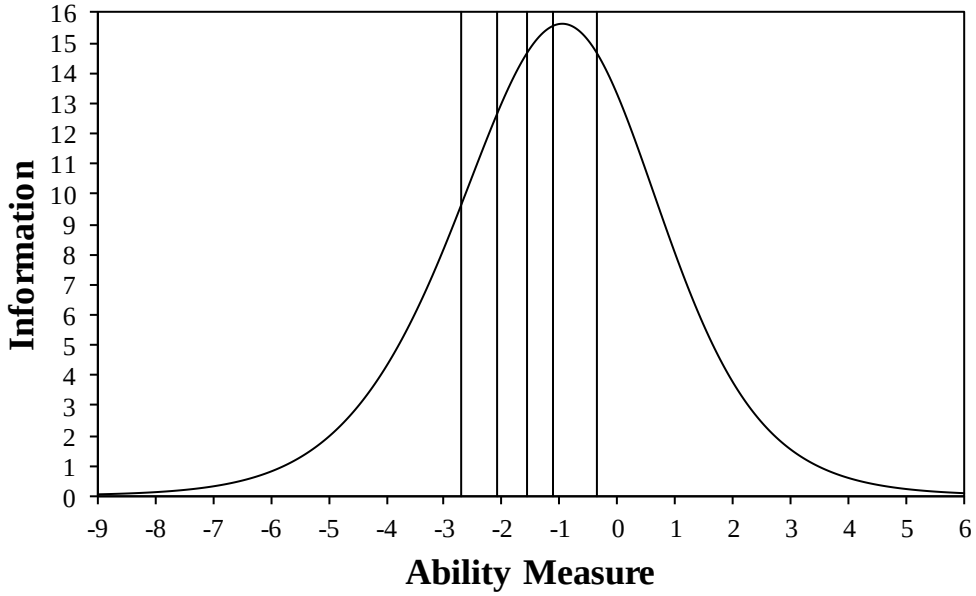

Table 3.3.1.2J
Reliability: Read 1 S400 Online

No. of Students	No. of Items	Rasch Reliability Estimate
152,182	72	.84

Table 3.3.1.2K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: Read 1 S400 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	1	253	3,269	11.73	13.55	12.71	0.27
2/3	1	269	516	10.53	11.36	10.78	0.13
3/4	1	283	9,159	10.14	10.61	10.27	0.06
4/5	1	294	2,081	9.88	10.24	10.04	0.04
5/6	1	314	73	10.06	10.40	10.19	0.06

Table 3.3.1.2L

Accuracy and Consistency of Classification Indices: Read (Grade 1) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.495	0.408	0.275		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	-	0.305		
	2	0.273	0.209		
	3	0.429	0.349		
	4	0.500	0.404		
	5	0.536	0.414		
	6	0.829	0.706		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.929	0.000	0.071	0.911
	2/3	0.873	0.085	0.042	0.816
	3/4	0.822	0.124	0.055	0.771
	4/5	0.865	0.066	0.070	0.815
	5/6	0.942	0.032	0.026	0.913

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 S400 Online

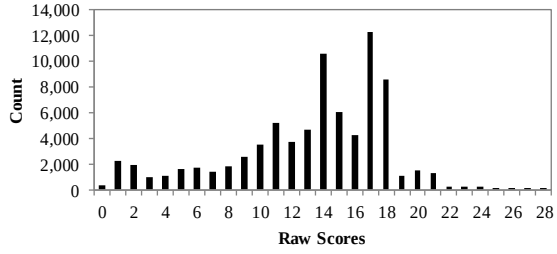

Table 3.3.1.3Ci

Raw Score Descriptive Statistics: Writ 1 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	79,064	0	28	13.15	4.94
Total	79,064	0	28	13.15	4.94

Figure 3.3.1.3Di
Scale Scores: Writ 1 A S400 Online

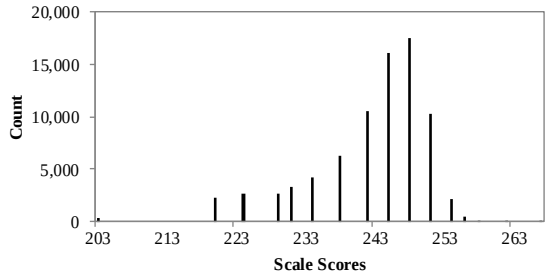

Table 3.3.1.3Di

Scale Score Descriptive Statistics: Writ 1 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	79,064	203	267	242.48	8.58
Total	79,064	203	267	242.48	8.58

Figure 3.3.1.3Ei
Proficiency Level: Writ 1 A S400 Online

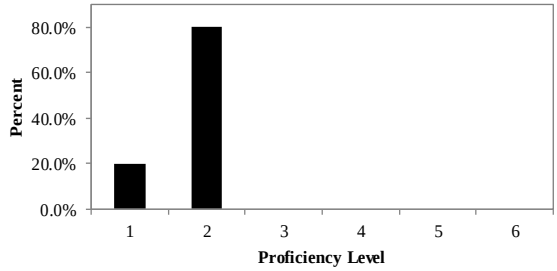

Table 3.3.1.3Ei

Proficiency Level Distribution: Writ 1 A S400

	Grade 1		Total	
Level	**Count**	**Percent**	**Count**	**Percent**
1	15,593	19.72%	15,593	19.72%
2	63,471	80.28%	63,471	80.28%
3	0	0.00%	0	0.00%
4	0	0.00%	0	0.00%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	79,064	100.00%	79,064	100.00%

Table 3.3.1.3Gi
Equating Summary: Writ 1 A S400 Online
n/a for S400

Figure 3.3.1.3Hi
Test Characteristic Curve: Writ 1 A S400 Online
n/a for S400

Figure 3.3.1.3Ii
Test Information Function: Writ 1 A S400 Online
n/a for S400

Table 3.3.1.3Ji
Reliability: Writ 1 A S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Response Mode**	**Cronbach's Alpha**	**SEM**
	79,064	4	Hand-written (HW)	.862	1.833
Interrater Reliability	**Task**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	32,682	100	0	0
	2	41,420	100	0	0
	3	44,994	97	3	0
	4	41,720	96	4	0

Table 3.3.1.3Ki
Conditional Standard Error of Measurement at Cut Scores: Writ 1 A S400 Online
n/a

Table 3.3.1.3Li
Accuracy and Consistency of Classification Indices: Writ 1 A S400 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 S400 Online

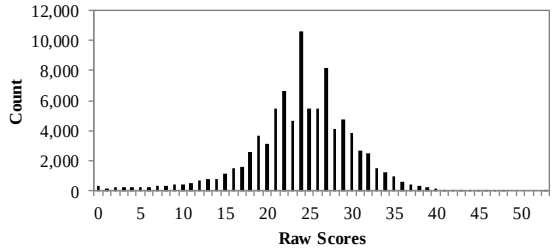

Table 3.3.1.3Cii

Raw Score Descriptive Statistics: Writ 1 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	89,353	0	53	24.08	6.18
Total	89,353	0	53	24.08	6.18

Figure 3.3.1.3Dii
Scale Scores: Writ 1 B/C S400 Online

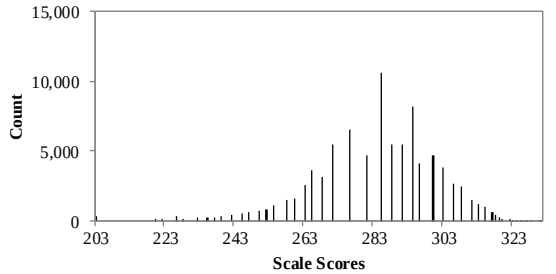

Table 3.3.1.3Dii

Scale Score Descriptive Statistics: Writ 1 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	89,353	203	331	283.58	19.13
Total	89,353	203	331	283.58	19.13

Figure 3.3.1.3Eii
Proficiency Level: Writ 1 B/C S400 Online

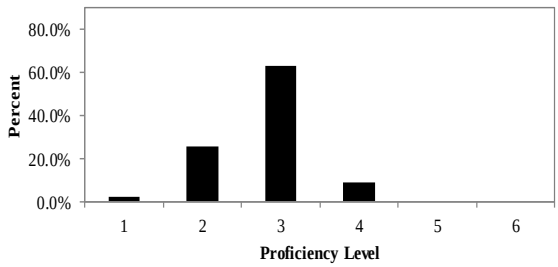

Table 3.3.1.3Eii

Proficiency Level Distribution: Writ 1 B/C S400 Online

	Grade 1		Total	
Level	Count	Percent	Count	Percent
1	2,115	2.37%	2,115	2.37%
2	22,716	25.42%	22,716	25.42%
3	56,383	63.10%	56,383	63.10%
4	8,139	9.11%	8,139	9.11%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	89,353	100.00%	89,353	100.00%

Table 3.3.1.3Gii
Equating Summary: Writ 1 B/C S400 Online
n/a for S400

Figure 3.3.1.3Hii
Test Characteristic Curve: Writ 1 B/C S400 Online
n/a for S400

Figure 3.3.1.3Iii
Test Information Function: Writ 1 B/C S400 Online
n/a for S400

Table 3.3.1.3Jii
Reliability: Writ 1 B/C S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Response Mode**	**Cronbach's Alpha**	**SEM**
	89,353	3	Hand-written (HW)	.914	1.809
Interrater Reliability	**Task**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	37,586	95	5	0
	2	40,070	93	6	0
	3	38,292	96	4	0

Table 3.3.1.3Kii
Conditional Standard Error of Measurement at Cut Scores: Writ 1 B/C S400 Online
n/a

Table 3.3.1.3Lii
Accuracy and Consistency of Classification Indices: Writ 1 B/C S400 Online
n/a

3.3.1.3iii Writing 1 Across Tiers

Table 3.3.1.3Aiii
Complete Task Analysis and Summary: Writ 1 S400 Online
n/a

Table 3.3.1.3Biii
DIF Analysis and Summary: Writ 1 S400 Online
n/a

Figure 3.3.1.3Ciii
Raw Scores: Writ 1 S400 Online

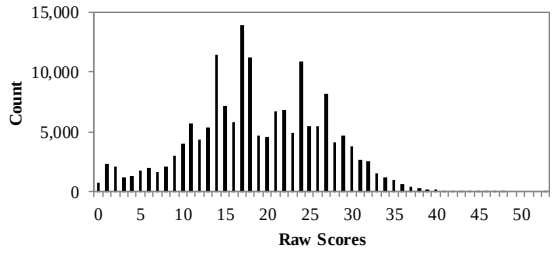

Table 3.3.1.3Ciii
Raw Score Descriptive Statistics: Writ 1 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	168,417	0	53	18.95	7.84
Total	168,417	0	53	18.95	7.84

Figure 3.3.1.3Diii
Scale Scores: Writ 1 S400 Online

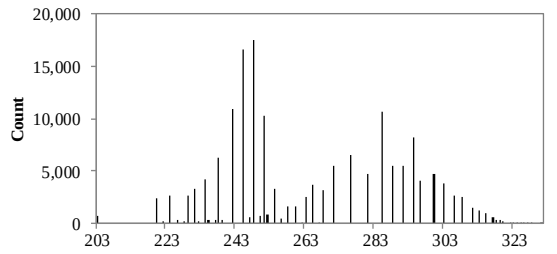

Table 3.3.1.3Diii
Scale Score Descriptive Statistics: Writ 1 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	168,417	203	331	264.29	25.48
Total	168,417	203	331	264.29	25.48

Figure 3.3.1.3Eiii
Proficiency Level: Writ 1 S400 Online

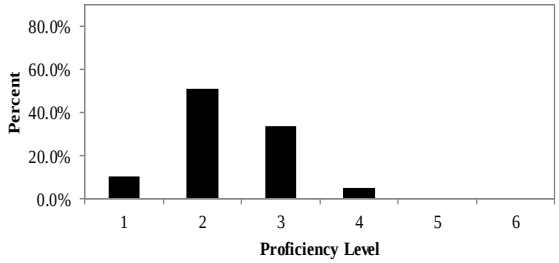

Table 3.3.1.3Eiii

Proficiency Level Distribution: Writ 1 S400 Online

	Grade 1		Total	
Level	**Count**	**Percent**	**Count**	**Percent**
1	17,708	10.51%	17,708	10.51%
2	86,187	51.17%	86,187	51.17%
3	56,383	33.48%	56,383	33.48%
4	8,139	4.83%	8,139	4.83%
5	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%
Total	168,417	100.00%	168,417	100.00%

Table 3.3.1.3Fiii
Raw Score to Scale Score Conversion: Writ 1 S400 Online
n/a

Table 3.3.1.3Giii
Equating Summary: Writ 1 S400 Online
n/a

Figure 3.3.1.3Hiii
Test Characteristic Curve: Writ 1 S400 Online
n/a for S400

Figure 3.3.1.3Iiii
Test Information Function: Writ 1 S400 Online
n/a for S400

Table 3.3.1.3Jiii

Reliability: Writ 1 Weighted Reliability S400 Online

Tiers	No. of Students	Reliability	Weighted Reliability
A	79,064	0.862	0.890
B/C	89,353	0.914	

Table 3.3.1.3Kiii

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

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

Table 3.3.1.3L

Accuracy and Consistency of Classification Indices: Writ (Grade 1) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.772	0.690	0.510		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.505	0.390		
	2	0.844	0.784		
	3	0.795	0.737		
	4	0.583	0.394		
	5	N/A	N/A		
	6	N/A	N/A		
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.896	0.069	0.035	0.862
	2/3	0.920	0.039	0.041	0.886
	3/4	0.956	0.033	0.011	0.941
	4/5	N/A	N/A	N/A	N/A
	5/6	N/A	N/A	N/A	N/A

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 S400 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 S400 Online

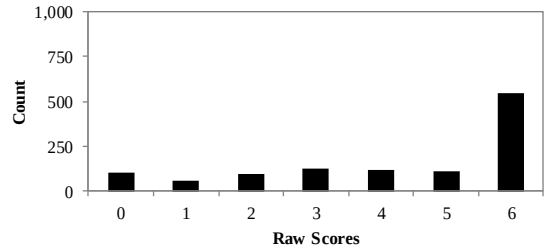

Table 3.3.1.4Ci
Raw Score Descriptive Statistics: Spek 1 Pre-A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	1,152	0	6	4.26	2.06
Total	1,152	0	6	4.26	2.06

Figure 3.3.1.4Di
Scale Scores: Spek 1 Pre-A S400 Online

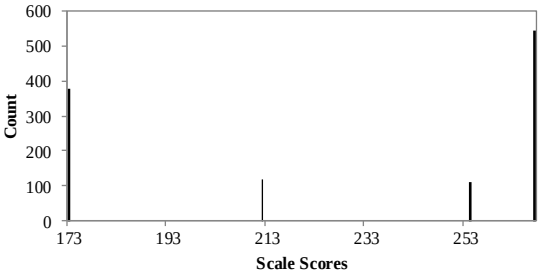

Table 3.3.1.4Di
Scale Score Descriptive Statistics: Spek 1 Pre-A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	1,152	173	267	229.17	42.44
Total	1,152	173	267	229.17	42.44

Figure 3.3.1.4Ei
Proficiency Level: Spek 1 Pre-A S400 Online

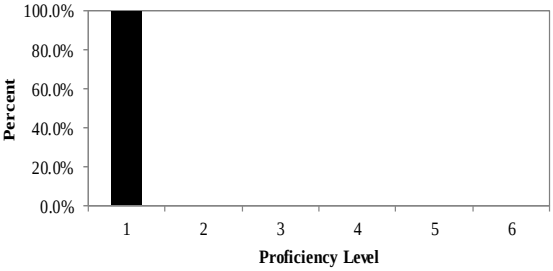

Table 3.3.1.4Ei

Proficiency Level Distribution: Spek 1 Pre-A S400 Online

	Grade 1		Total	
Level	**Count**	**Percent**	**Count**	**Percent**
1	1,152	100.00%	1,152	100.00%
Total	1,152	100.00%	1,152	100.00%

Table 3.3.1.4Gi
Equating Summary: Spek 1 Pre-A S400 Online
n/a for S400

Figure 3.3.1.4Hi
Test Characteristic Curve: Spek 1 Pre-A S400 Online
n/a for S400

Figure 3.3.1.4Ii
Test Information Function: Spek 1 Pre-A S400 Online
n/a for S400

Table 3.3.1.4Ji
Reliability: Spek 1 Pre-A S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Cronbach's Alpha**		**SEM**
	1,152	3	.814		0.887
Interrater	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	624	95	5	0
	2	634	96	4	0
	3	640	96	4	0

Table 3.3.1.4Ki
Conditional Standard Error of Measurement at Cut Scores: Spek 1 Pre-A S400 Online
n/a

Table 3.3.1.4Li
Accuracy and Consistency of Classification Indices: Spek 1 Pre-A S400 Online
n/a

3.3.1.4ii Speaking 1 A

Table 3.3.1.4Aii
Complete Task Analysis and Summary: Spek 1 A S400 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 S400 Online

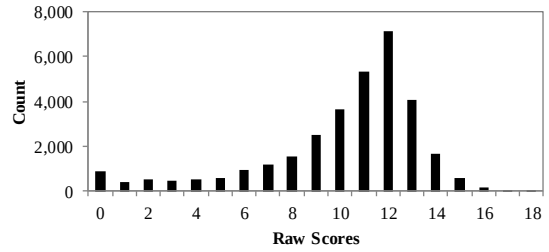

Table 3.3.1.4Cii
Raw Score Descriptive Statistics: Spek 1 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	32,047	0	18	10.13	3.36
Total	32,047	0	18	10.13	3.36

Figure 3.3.1.4Dii
Scale Scores: Spek 1 A S400 Online

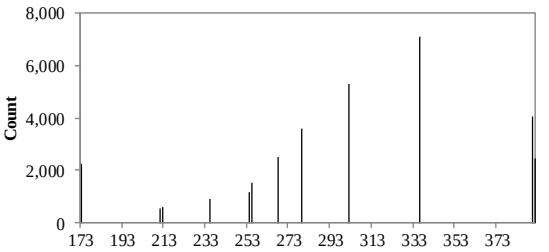

Table 3.3.1.4Dii
Scale Score Descriptive Statistics: Spek 1 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	32,047	173	391	304.21	61.34
Total	32,047	173	391	304.21	61.34

Figure 3.3.1.4Eii
Proficiency Level: Spek 1 A S400 Online

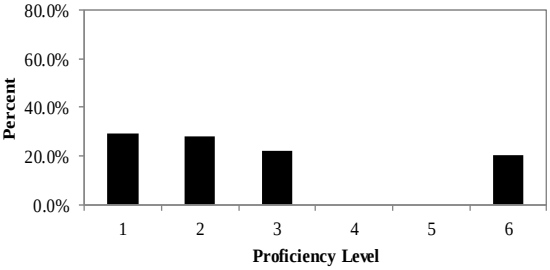

Table 3.3.1.4Eii
Proficiency Level Distribution: Spek 1 A S400 Online

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	9,477	29.57%	9,477	29.57%
2	8,926	27.85%	8,926	27.85%
3	7,107	22.18%	7,107	22.18%
4	0	0.00%	0	0.00%
5	0	0.00%	0	0.00%
6	6,537	20.40%	6,537	20.40%
Total	32,047	100.00%	32,047	100.00%

Table 3.3.1.4Gii
Equating Summary: Spek 1 A S400 Online
n/a for S400

Figure 3.3.1.4Hii
Test Characteristic Curve: Spek 1 A S400 Online
n/a for S400

Figure 3.3.1.4Iii
Test Information Function: Spek 1 A S400 Online
n/a for S400

Table 3.3.1.4Jii

Reliability: Spek 1 A S400 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	32,047	6	.850		1.303
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	13,546	97	3	0
	2	13,336	81	19	0
	3	14,091	98	2	0
	4	13,927	84	16	0
	5	14,644	98	2	0
	6	14,452	82	18	0

Table 3.3.1.4Kii

Conditional Standard Error of Measurement at Cut Scores: Spek 1 A S400 Online

n/a

Table 3.3.1.4Lii

Accuracy and Consistency of Classification Indices: Spek 1 A S400 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 S400 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 S400 Online

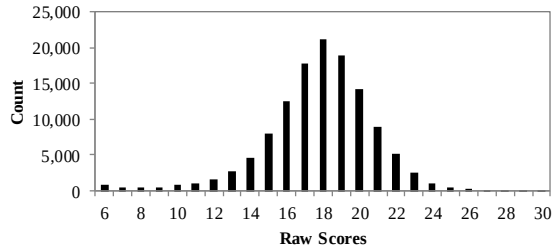

Table 3.3.1.4Ciii

Raw Score Descriptive Statistics: Spek 1 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	124,218	6	30	17.77	2.99
Total	124,218	6	30	17.77	2.99

Figure 3.3.1.4Diii
Scale Scores: Spek 1 B/C S400 Online

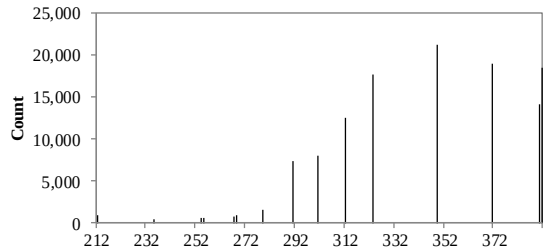

Table 3.3.1.4Diii

Scale Score Descriptive Statistics: Spek 1 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	124,218	212	391	344.96	38.98
Total	124,218	212	391	344.96	38.98

Figure 3.3.1.4Eiii
Proficiency Level: Spek 1 B/C S400 Online

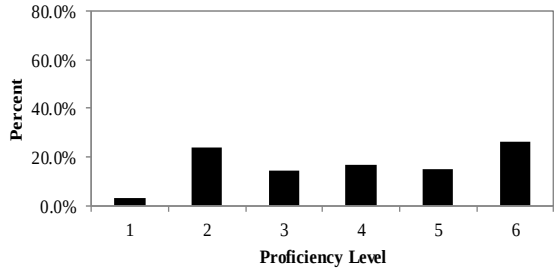

Table 3.3.1.4Eiii

Proficiency Level Distribution: Spek 1 B/C S400 Online

Level	Grade 1		Total	
	Count	Percent	Count	Percent
1	4,267	3.44%	4,267	3.44%
2	29,443	23.70%	29,443	23.70%
3	17,735	14.28%	17,735	14.28%
4	21,204	17.07%	21,204	17.07%
5	18,941	15.25%	18,941	15.25%
6	32,628	26.27%	32,628	26.27%
Total	124,218	100.00%	124,218	100.00%

Table 3.3.1.4Giii
Equating Summary: Spek 1 B/C S400 Online
n/a for S400

Figure 3.3.1.4Hiii
Test Characteristic Curve: Spek 1 B/C S400 Online
n/a for S400

Figure 3.3.1.4Iiii
Test Information Function: Spek 1 B/C S400 Online
n/a for S400

Table 3.3.1.4Jiii

Reliability: Spek 1 B/C S400 Online

Reliability	**No. of Students**	**No. of Items**	**Cronbach's Alpha**		**SEM**
	124,218	6	.772		1.425
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	55,482	82	18	0
	2	55,226	82	18	0
	3	54,485	78	22	0
	4	54,309	80	20	0
	5	52,027	84	16	0
	6	51,919	76	23	0

Table 3.3.1.4Kiii
Conditional Standard Error of Measurement at Cut Scores: Spek 1 B/C S400 Online
n/a

Table 3.3.1.4Liii
Accuracy and Consistency of Classification Indices: Spek 1 B/C S400 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 S400 Online
n/a

Figure 3.3.1.4Civ
Raw Scores: Spek 1 S400 Online

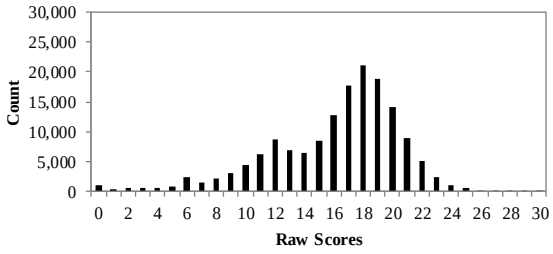

Table 3.3.1.4Civ
Raw Score Descriptive Statistics: Spek 1 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	157,417	0	30	16.11	4.46
Total	157,417	0	30	16.11	4.46

Figure 3.3.1.4Div
Scale Scores: Spek 1 S400 Online

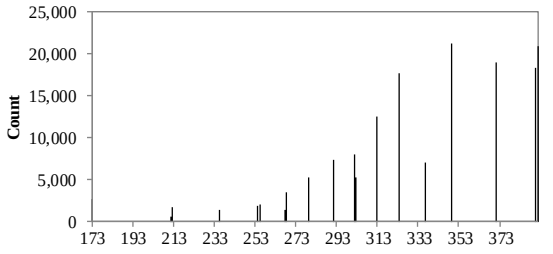

Table 3.3.1.4Div
Scale Score Descriptive Statistics: Spek 1 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	157,417	173	391	335.82	48.28
Total	157,417	173	391	335.82	48.28

Figure 3.3.1.4Eiv
Proficiency Level: Spek 1 S400 Online

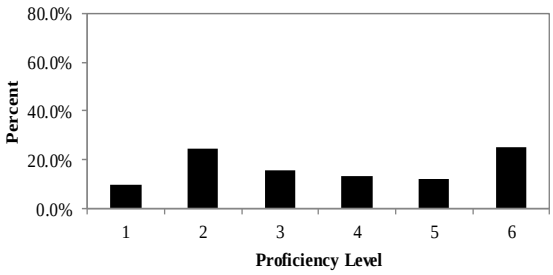

Table 3.3.1.4Eiv
Proficiency Level Distribution: Spek 1 S400 Online

	Grade 1		Total	
Level	Count	Percent	Count	Percent
1	14,896	9.46%	14,896	9.46%
2	38,369	24.37%	38,369	24.37%
3	24,842	15.78%	24,842	15.78%
4	21,204	13.47%	21,204	13.47%
5	18,941	12.03%	18,941	12.03%
6	39,165	24.88%	39,165	24.88%
Total	157,417	100.00%	157,417	100.00%

Table 3.3.1.4Fiv
Raw Score to Scale Score Conversion: Spek 1 S400 Online
n/a

Table 3.3.1.4Giv
Equating Summary: Spek 1 S400 Online
n/a

Figure 3.3.1.4Hiv
Test Characteristic Curve: Spek 1 S400 Online
n/a for S400

Figure 3.3.1.4Iiv
Test Information Function: Spek 1 S400 Online
n/a for S400

Table 3.3.1.4Jiv

Reliability: Spek 1 Weighted Reliability S400 Online

Tiers	No. of Students	Reliability	Weighted Reliability
Pre-A	1,152	0.814	0.788
A	32,047	0.850	
B/C	124,218	0.772	

Table 3.3.1.4Kiv

Conditional Standard Error of Measurement at Cut Scores: Spek 1 S400 Online

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

Table 3.3.1.4L

Accuracy and Consistency of Classification Indices: Spek (Grade 1) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.441	0.379	0.253		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.630	0.460		
	2	0.616	0.506		
	3	0.340	0.265		
	4	0.303	0.219		
	5	0.255	0.210		
	6	0.850	0.655		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.931	0.036	0.034	0.895
	2/3	0.857	0.049	0.094	0.813
	3/4	0.866	0.058	0.076	0.804
	4/5	0.874	0.084	0.042	0.815
	5/6	0.814	0.172	0.013	0.804

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 3.3.2.1C
Raw Scores: List 2-3 S400 Online
n/a

Table 3.3.2.1C
Raw Score Descriptive Statistics: List 2-3 S400 Online
n/a

Figure 3.3.2.1D
Scale Scores: List 2-3 S400 Online

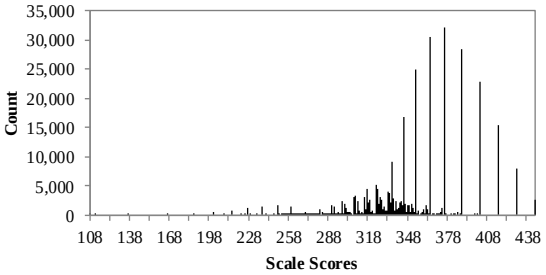

Figure 3.3.2.1E
Proficiency Level: List 2-3 S400 Online

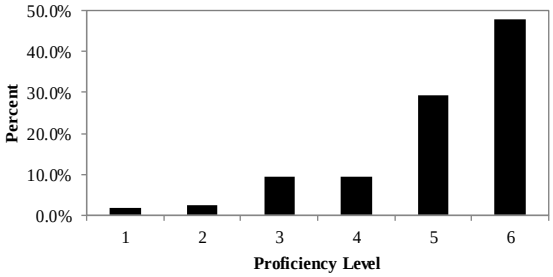

Table 3.3.2.1D
Scale Score Descriptive Statistics: List 2-3 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	155,299	108	444	347.79	38.58
3	152,622	112	444	363.60	39.52
Total	307,921	108	444	355.62	39.84

Table 3.3.2.1E
Proficiency Level Distribution: List 2-3 S400 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,801	1.80%	2,562	1.68%	5,363	1.74%
2	3,573	2.30%	4,046	2.65%	7,619	2.47%
3	15,556	10.02%	13,208	8.65%	28,764	9.34%
4	11,260	7.25%	17,388	11.39%	28,648	9.30%
5	45,203	29.11%	44,824	29.37%	90,027	29.24%
6	76,906	49.52%	70,594	46.25%	147,500	47.90%
Total	155,299	100.00%	152,622	100.00%	307,921	100.00%

Table 3.3.2.1F
Raw Score to Scale Score Conversion: List 2-3 S400 Online
n/a

Table 3.3.2.1G
Equating Summary: List 2-3 S400 Online
n/a for S400

Figure 3.3.2.1H
Test Characteristic Curve: List 2-3 S400 Online
n/a

Figure 3.3.2.1I
Test Information Function: List 2-3 S400 Online

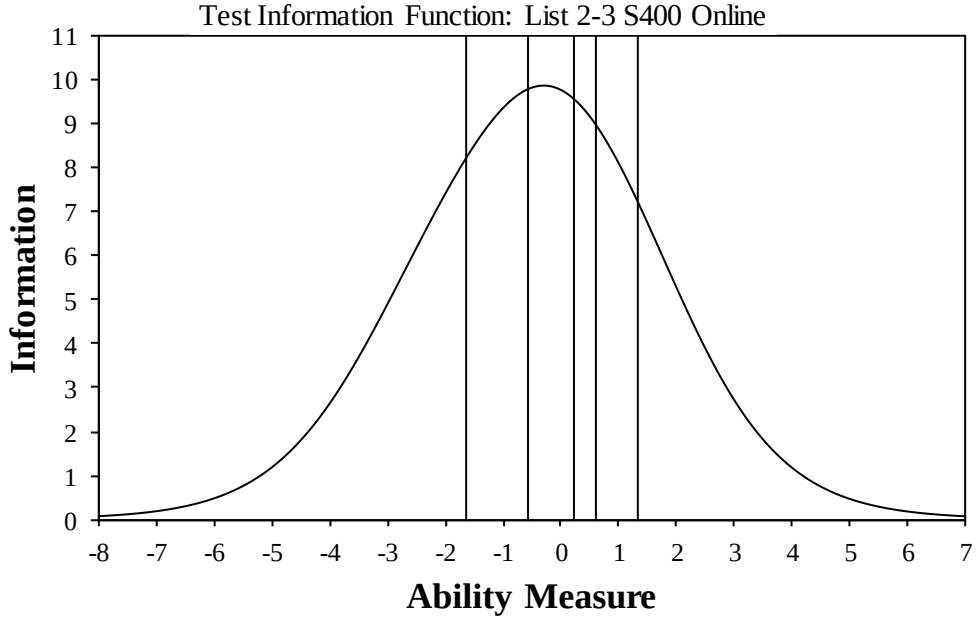

Table 3.3.2.1J
Reliability: List 2-3 S400 Online

No. of Students	No. of Items	Rasch Reliability Estimate
307,921	54	.73

Table 3.3.2.1K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: List 2-3 S400 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	2	247	N/A	N/A	N/A	N/A	N/A
	3	255	N/A	N/A	N/A	N/A	N/A
2/3	2	281	720	17.85	18.60	18.33	0.08
	3	295	322	18.00	18.64	18.29	0.24
3/4	2	311	730	17.81	18.94	17.83	0.10
	3	325	2,837	18.00	19.01	18.00	0.03
4/5	2	324	3,608	18.03	19.54	18.13	0.09
	3	340	604	18.41	18.94	18.57	0.16
5/6	2	350	265	18.97	19.35	19.14	0.12
	3	367	132	20.63	20.81	20.76	0.09

Table 3.3.2.1Li

Accuracy and Consistency of Classification Indices: List (Grade 2) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.629	0.542	0.300		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.867	0.577		
	2	0.307	0.186		
	3	0.469	0.307		
	4	0.202	0.145		
	5	0.519	0.423		
	6	0.794	0.733		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.987	0.001	0.013	0.984
	2/3	0.974	0.009	0.017	0.955
	3/4	0.911	0.035	0.054	0.869
	4/5	0.872	0.060	0.068	0.820
	5/6	0.813	0.080	0.107	0.748

Table 3.3.2.1Lii

Accuracy and Consistency of Classification Indices: List (Grade 3) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.606	0.519	0.286		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.861	0.573		
	2	0.345	0.204		
	3	0.393	0.255		
	4	0.294	0.217		
	5	0.501	0.414		
	6	0.774	0.705		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.988	0.001	0.012	0.985
	2/3	0.972	0.010	0.018	0.951
	3/4	0.910	0.041	0.049	0.866
	4/5	0.854	0.060	0.086	0.803
	5/6	0.812	0.075	0.113	0.745

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 S400 Online
n/a

Table 3.3.2.2C
Raw Score Descriptive Statistics: Read 2-3 S400 Online
n/a

Figure 3.3.2.2D
Scale Scores: Read 2-3 S400 Online

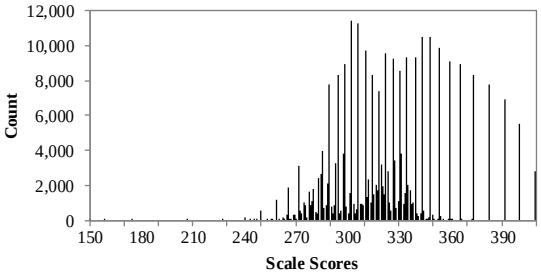

Figure 3.3.2.2E
Proficiency Level: Read 2-3 S400 Online

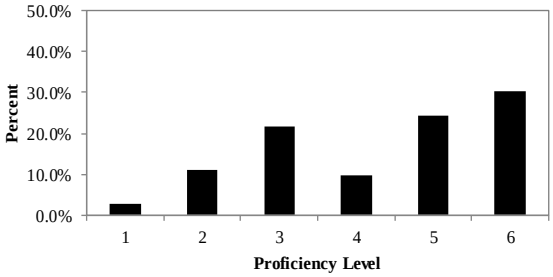

Table 3.3.2.2D
Scale Score Descriptive Statistics: Read 2-3 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	145,816	150	409	316.22	28.49
3	143,263	158	409	334.63	33.08
Total	289,079	150	409	325.34	32.19

Table 3.3.2.2E
Proficiency Level Distribution: Read 2-3 S400 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,070	2.11%	5,285	3.69%	8,355	2.89%
2	14,569	9.99%	17,447	12.18%	32,016	11.08%
3	35,895	24.62%	26,365	18.40%	62,260	21.54%
4	15,788	10.83%	12,744	8.90%	28,532	9.87%
5	39,504	27.09%	31,221	21.79%	70,725	24.47%
6	36,990	25.37%	50,201	35.04%	87,191	30.16%
Total	145,816	100.00%	143,263	100.00%	289,079	100.00%

Table 3.3.2.2F
Raw Score to Scale Score Conversion: Read 2-3 S400 Online
n/a

Table 3.3.2.2G
Equating Summary: Read 2-3 S400 Online
n/a for S400

Figure 3.3.2.2H
Test Characteristic Curve: Read 2-3 S400 Online
n/a

Figure 3.3.2.2I
Test Information Function: Read 2-3 S400 Online

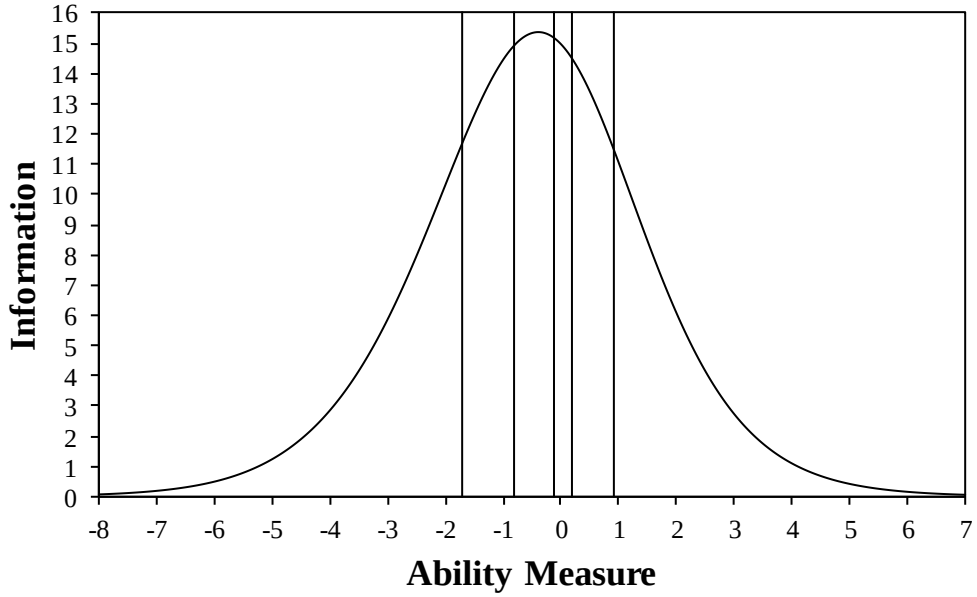

Table 3.3.2.2J
Reliability: Read 2-3 S400 Online

No. of Students	No. of Items	Rasch Reliability Estimate
289,079	72	.86

Table 3.3.2.2K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: Read 2-3 S400 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	2	267	37	12.22	13.83	13.11	0.45
	3	279	150	11.44	12.48	12.03	0.27
2/3	2	286	423	10.89	11.02	10.97	0.05
	3	302	3,845	10.30	10.61	10.34	0.03
3/4	2	303	597	10.32	10.63	10.48	0.14
	3	320	263	10.27	10.71	10.38	0.06
4/5	2	312	2,288	10.11	10.48	10.37	0.09
	3	328	613	10.37	10.79	10.50	0.09
5/6	2	331	930	10.35	10.71	10.39	0.03
	3	347	168	11.36	11.70	11.44	0.12

Table 3.3.2.2Li

Accuracy and Consistency of Classification Indices: Read (Grade 2) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.588	0.493	0.352		
Conditional on Level	Level	Accuracy	Consistency		
	1	-	0.101		
	2	0.408	0.296		
	3	0.518	0.446		
	4	0.249	0.184		
	5	0.651	0.522		
	6	0.846	0.759		
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.979	0.000	0.021	0.972
	2/3	0.886	0.032	0.082	0.848
	3/4	0.865	0.077	0.058	0.806
	4/5	0.871	0.086	0.043	0.822
	5/6	0.921	0.040	0.039	0.886

Table 3.3.2.2Lii

Accuracy and Consistency of Classification Indices: Read (Grade 3) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.600	0.516	0.372		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	-	0.225		
	2	0.469	0.375		
	3	0.462	0.364		
	4	0.232	0.173		
	5	0.545	0.430		
	6	0.873	0.802		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.963	0.000	0.037	0.948
	2/3	0.897	0.052	0.051	0.855
	3/4	0.879	0.069	0.052	0.832
	4/5	0.879	0.070	0.050	0.834
	5/6	0.906	0.050	0.044	0.867

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 S400 Online

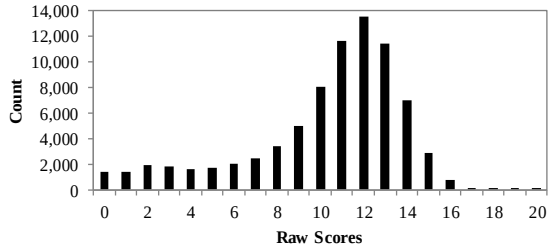

Table 3.3.2.3Ci
Raw Score Descriptive Statistics: Writ 2-3 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	65,426	0	20	10.54	3.43
3	12,554	0	19	8.82	4.23
Total	77,980	0	20	10.26	3.63

Figure 3.3.2.3Di
Scale Scores: Writ 2-3 A S400 Online

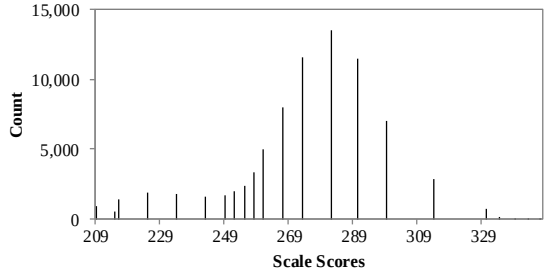

Table 3.3.2.3Di
Scale Score Descriptive Statistics: Writ 2-3 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	65,426	209	347	274.62	21.86
3	12,554	215	343	264.78	25.15
Total	77,980	209	347	273.03	22.72

Figure 3.3.2.3Ei
Proficiency Level: Writ 2-3 A S400 Online

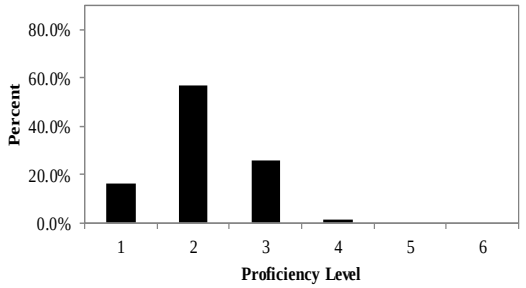

Table 3.3.2.3Ei

Proficiency Level Distribution: Writ 2-3 A S400 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	6,874	10.51%	5,660	45.09%	12,534	16.07%
2	38,792	59.29%	5,667	45.14%	44,459	57.01%
3	18,969	28.99%	1,098	8.75%	20,067	25.73%
4	791	1.21%	129	1.03%	920	1.18%
5	0	0.00%	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%	0	0.00%
Total	65,426	100.00%	12,554	100.00%	77,980	100.00%

Table 3.3.2.3Gi
Equating Summary: Writ 2-3 A S400 Online
n/a for S400

Figure 3.3.2.3Hi
Test Characteristic Curve: Writ 2-3 A S400 Online
n/a for S400

Figure 3.3.2.3Ii
Test Information Function: Writ 2-3 A S400 Online
n/a for S400

Table 3.3.2.3Ji
Reliability: Writ 2-3 A S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Response Mode**	**Cronbach's Alpha**	**SEM**
	77,980	3	Hand-written (HW)	.859	1.361
Interrater Reliability	**Task**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	36,218	94	5	1
	2	36,064	95	5	0
	3	39,069	96	4	0

Table 3.3.2.3Ki
Conditional Standard Error of Measurement at Cut Scores: Writ 2-3 A S400 Online
n/a

Table 3.3.2.3Li
Accuracy and Consistency of Classification Indices: Writ 2-3 A S400 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 S400 Online

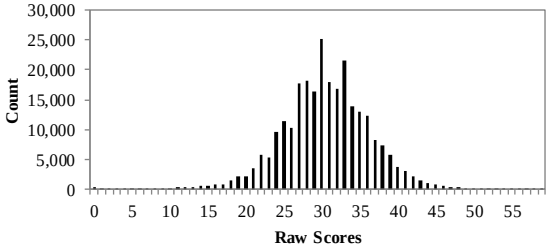

Table 3.3.2.3Cii
Raw Score Descriptive Statistics: Writ 2-3 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	107,063	0	54	29.25	5.50
3	157,709	0	59	31.19	6.13
Total	264,772	0	59	30.41	5.96

Figure 3.3.2.3Dii
Scale Scores: Writ 2-3 B/C S400 Online

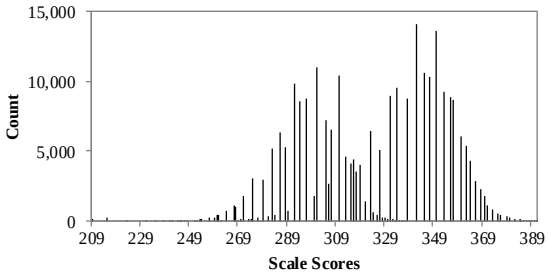

Table 3.3.2.3Dii
Scale Score Descriptive Statistics: Writ 2-3 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	107,063	209	337	297.86	15.38
3	157,709	215	391	342.21	20.08
Total	264,772	209	391	324.28	28.46

Figure 3.3.2.3Eii
Proficiency Level: Writ 2-3 B/C S400 Online

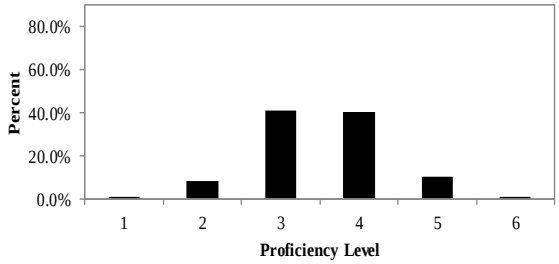

Table 3.3.2.3Eii

Proficiency Level Distribution: Writ 2-3 B/C S400 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	634	0.59%	686	0.43%	1,320	0.50%
2	17,050	15.93%	4,360	2.76%	21,410	8.09%
3	85,095	79.48%	23,211	14.72%	108,306	40.91%
4	4,284	4.00%	102,775	65.17%	107,059	40.43%
5	0	0.00%	26,414	16.75%	26,414	9.98%
6	0	0.00%	263	0.17%	263	0.10%
Total	107,063	100.00%	157,709	100.00%	264,772	100.00%

Table 3.3.2.3Gii
Equating Summary: Writ 2-3 B/C S400 Online
n/a for S400

Figure 3.3.2.3Hii
Test Characteristic Curve: Writ 2-3 B/C S400 Online
n/a for S400

Figure 3.3.2.3Iii
Test Information Function: Writ 2-3 B/C S400 Online
n/a for S400

Table 3.3.2.3Jii
Reliability: Writ 2-3 B/C S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Response Mode**	**Cronbach's Alpha**	**SEM**
	264,772	3	Hand-written (HW)	.907	1.822
Interrater Reliability	**Task**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	108,696	94	5	0
	2	108,604	91	9	1
	3	109,108	90	9	1

Table 3.3.2.3Kii
Conditional Standard Error of Measurement at Cut Scores: Writ 2-3 B/C S400 Online
n/a

Table 3.3.2.3Lii
Accuracy and Consistency of Classification Indices: Writ 2-3 B/C S400 Online
n/a for S400

3.3.2.3iii Writing 2-3 Across Tiers

Table 3.3.2.3Aiii
Complete Task Analysis and Summary: Writ 2-3 S400 Online
n/a

Table 3.3.2.3Biii
DIF Analysis and Summary: Writ 2-3 S400 Online
n/a

Figure 3.3.2.3Ciii
Raw Scores: Writ 2-3 S400 Online

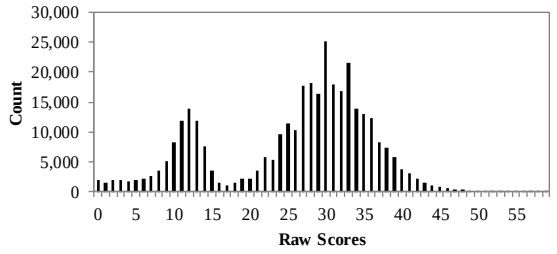

Table 3.3.2.3Ciii
Raw Score Descriptive Statistics: Writ 2-3 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	172,489	0	54	22.15	10.28
3	170,263	0	59	29.54	8.39
Total	342,752	0	59	25.83	10.09

Figure 3.3.2.3Diii
Scale Scores: Writ 2-3 S400 Online

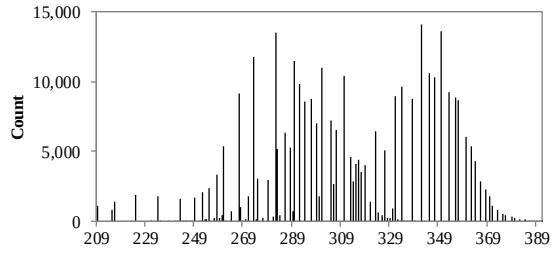

Table 3.3.2.3Diii
Scale Score Descriptive Statistics: Writ 2-3 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	172,489	209	347	289.04	21.34
3	170,263	215	391	336.50	28.81
Total	342,752	209	391	312.62	34.70

Figure 3.3.2.3Eiii
Proficiency Level: Writ 2-3 S400 Online

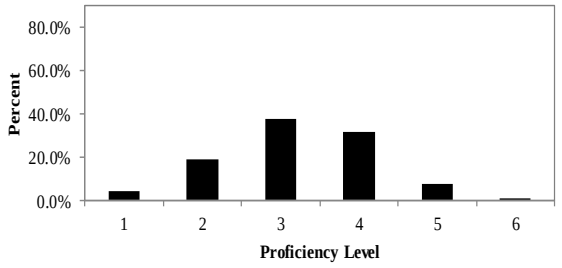

Table 3.3.2.3Eiii

Proficiency Level Distribution: Writ 2-3 S400 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	7,508	4.35%	6,346	3.73%	13,854	4.04%
2	55,842	32.37%	10,027	5.89%	65,869	19.22%
3	104,064	60.33%	24,309	14.28%	128,373	37.45%
4	5,075	2.94%	102,904	60.44%	107,979	31.50%
5	0	0.00%	26,414	15.51%	26,414	7.71%
6	0	0.00%	263	0.15%	263	0.08%
Total	172,489	100.00%	170,263	100.00%	342,752	100.00%

Table 3.3.2.3Fiii
Raw Score to Scale Score Conversion: Writ 2-3 S400 Online
n/a

Table 3.3.2.3Giii
Equating Summary: Writ 2-3 S400 Online
n/a

Figure 3.3.2.3Hiii
Test Characteristic Curve: Writ 2-3 S400 Online
n/a for S400

Figure 3.3.2.3Iiii
Test Information Function: Writ 2-3 S400 Online
n/a for S400

Table 3.3.2.3Jiii

Reliability: Writ 2-3 Weighted Reliability S400 Online

Tiers	No. of Students	Reliability	Weighted Reliability
A	77,980	0.859	0.896
B/C	264,772	0.907	

Table 3.3.2.3Kiii

Conditional Standard Error of Measurement at Cut Scores: Writ 2-3 S400 Online

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	2	251	6.84	7.46
	3	264	10.88	7.77
2/3	2	285	8.40	8.40
	3	297	11.19	7.77
3/4	2	320	7.77	7.46
	3	330	11.51	8.09
4/5	2	348	6.84	6.53
	3	360	10.57	7.46
5/6	2	373	6.53	7.15
	3	384	9.33	6.53

Table 3.3.2.3Li

Accuracy and Consistency of Classification Indices: Writ (Grade 2) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.867	0.813		0.639	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.775		0.639	
	2	0.861		0.788	
	3	0.876		0.853	
	4	-		0.125	
	5	N/A		N/A	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.981	0.010	0.009	0.971
	2/3	0.915	0.033	0.052	0.881
	3/4	0.971	0.029	0.000	0.961
	4/5	N/A	N/A	N/A	N/A
	5/6	N/A	N/A	N/A	N/A

Table 3.3.2.3Lii

Accuracy and Consistency of Classification Indices: Writ (Grade 3) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.755	0.657		0.397	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.887		0.812	
	2	0.736		0.622	
	3	0.749		0.601	
	4	0.751		0.741	
	5	-		0.248	
	6	-		-	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.990	0.004	0.006	0.986
	2/3	0.979	0.010	0.012	0.969
	3/4	0.943	0.021	0.037	0.915
	4/5	0.843	0.157	0.000	0.783
	5/6	0.998	0.002	0.000	0.998

3.3.2.4 Speaking 2-3

3.3.2.4i Speaking 2-3 Pre-A

Table 3.3.2.4Ai
Complete Task Analysis and Summary: Spek 2-3 Pre-A S400 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 3.3.2.4Ci
Raw Scores: Spek 2-3 Pre-A S400 Online

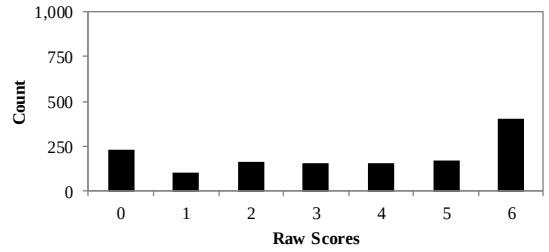

Table 3.3.2.4Ci
Raw Score Descriptive Statistics: Spek 2-3 Pre-A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	410	0	6	3.30	2.30
3	964	0	6	3.53	2.19
Total	1,374	0	6	3.46	2.23

Figure 3.3.2.4Di
Scale Scores: Spek 2-3 Pre-A S400 Online

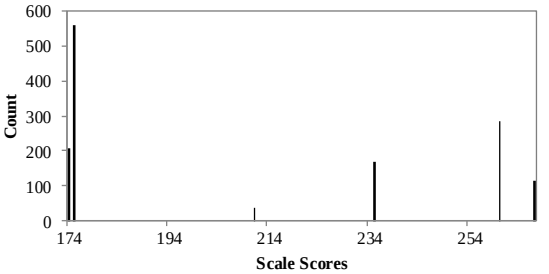

Table 3.3.2.4Di
Scale Score Descriptive Statistics: Spek 2-3 Pre-A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	410	174	267	211.01	40.58
3	964	175	260	207.50	39.06
Total	1,374	174	267	208.55	39.54

Figure 3.3.2.4Ei
Proficiency Level: Spek 2-3 Pre-A S400 Online

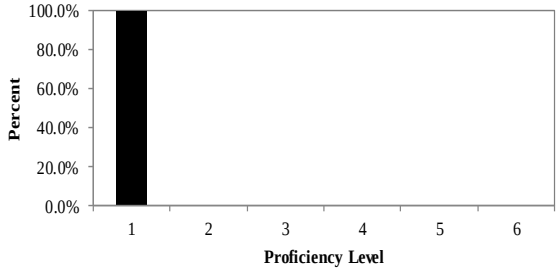

Table 3.3.2.4Ei

Proficiency Level Distribution: Spek 2-3 Pre-A S400 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	410	100.00%	964	100.00%	1,374	100.00%
Total	410	100.00%	964	100.00%	1,374	100.00%

3.3.2.4Fi

Raw Score to Scale Score Conversion: Spek 2-3 Pre-A S400 Online

Raw Score	Scale Score	CSEM	Low Bound	High Bound
0	175	24.49	150.51	199.49
1	175	24.49	150.51	199.49
2	175	24.49	150.51	199.49
3	175	24.49	150.51	199.49
4	175	24.49	150.51	199.49
5	235	22.45	212.55	257.45
6	260	19.39	240.61	279.39

Table 3.3.2.4Gi
Equating Summary: Spek 2-3 Pre-A S400 Online
n/a for S400

Figure 3.3.2.4Hi
Test Characteristic Curve: Spek 2-3 Pre-A S400 Online
n/a for S400

Figure 3.3.2.4Ii
Test Information Function: Spek 2-3 Pre-A S400 Online
n/a for S400

Table 3.3.2.4Ji
Reliability: Spek 2-3 Pre-A S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Cronbach's Alpha**		**SEM**
	1,374	3	.820		0.943
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	974	99	1	0
	2	1,054	96	4	0
	3	1,004	96	4	0

Table 3.3.2.4Ki
Conditional Standard Error of Measurement at Cut Scores: Spek 2-3 Pre-A S400 Online
n/a

Table 3.3.2.4Li
Accuracy and Consistency of Classification Indices: Spek 2-3 Pre-A S400 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 S400 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 S400 Online

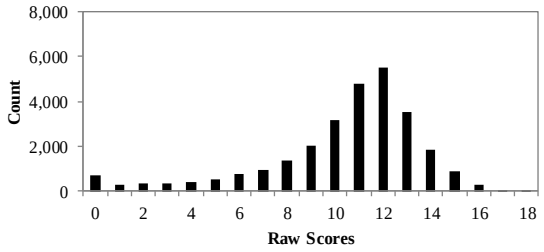

Table 3.3.2.4Cii
Raw Score Descriptive Statistics: Spek 2-3 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	7,782	0	17	8.86	3.77
3	19,788	0	18	10.94	2.95
Total	27,570	0	18	10.35	3.34

Figure 3.3.2.4Dii
Scale Scores: Spek 2-3 A S400 Online

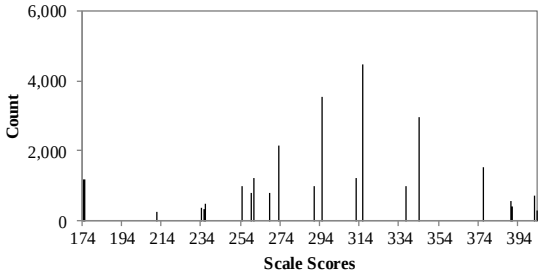

Table 3.3.2.4Dii
Scale Score Descriptive Statistics: Spek 2-3 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	7,782	174	391	282.74	64.61
3	19,788	175	403	303.13	51.86
Total	27,570	174	403	297.38	56.50

Figure 3.3.2.4Eii
Proficiency Level: Spek 2-3 A S400 Online

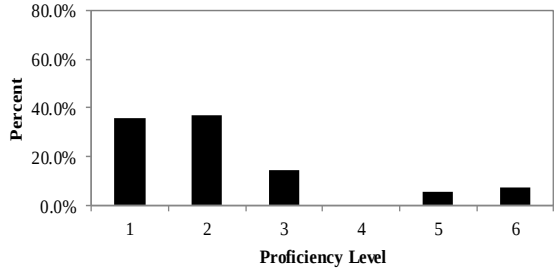

Table 3.3.2.4Eii

Proficiency Level Distribution: Spek 2-3 A S400 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,586	46.08%	6,214	31.40%	9,800	35.55%
2	2,206	28.35%	8,028	40.57%	10,234	37.12%
3	1,011	12.99%	2,957	14.94%	3,968	14.39%
4	0	0.00%	0	0.00%	0	0.00%
5	0	0.00%	1,544	7.80%	1,544	5.60%
6	979	12.58%	1,045	5.28%	2,024	7.34%
Total	7,782	100.00%	19,788	100.00%	27,570	100.00%

Table 3.3.2.4Gii
Equating Summary: Spek 2-3 A S400 Online
n/a for S400

Figure 3.3.2.4Hii
Test Characteristic Curve: Spek 2-3 A S400 Online
n/a for S400

Figure 3.3.2.4Iii
Test Information Function: Spek 2-3 A S400 Online
n/a for S400

Table 3.3.2.4Jii

Reliability: Spek 2-3 A S400 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	27,570	6	.838		1.343
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	12,952	97	3	0
	2	12,988	78	22	1
	3	12,732	92	8	0
	4	12,729	71	29	1
	5	12,702	97	3	0
	6	12,702	77	23	0

Table 3.3.2.4Kii

Conditional Standard Error of Measurement at Cut Scores: Spek 2-3 A S400 Online

n/a

Table 3.3.2.4Lii

Accuracy and Consistency of Classification Indices: Spek 2-3 A S400 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 S400 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 S400 Online

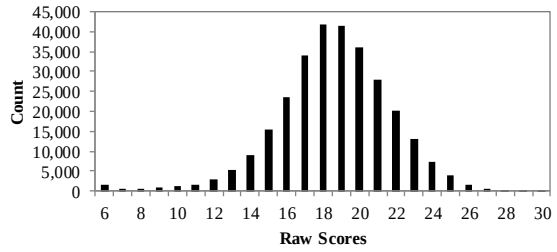

Table 3.3.2.4Ciii
Raw Score Descriptive Statistics: Spek 2-3 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	152,265	6	30	17.82	3.07
3	138,500	6	30	19.43	2.96
Total	290,765	6	30	18.58	3.12

Figure 3.3.2.4Diii
Scale Scores: Spek 2-3 B/C S400 Online

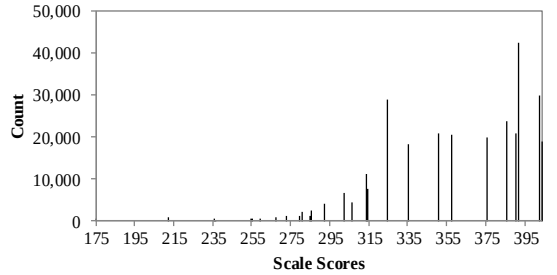

Table 3.3.2.4Diii
Scale Score Descriptive Statistics: Spek 2-3 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	152,265	212	391	359.25	39.00
3	138,500	175	403	361.76	38.54
Total	290,765	175	403	360.45	38.80

Figure 3.3.2.4Eiii
Proficiency Level: Spek 2-3 B/C S400 Online

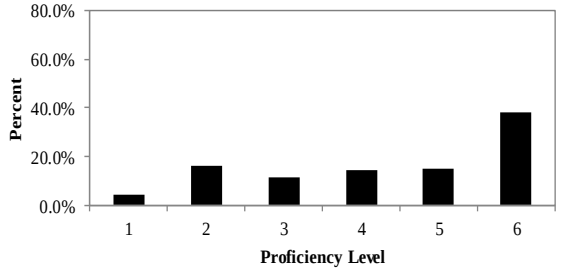

Table 3.3.2.4Eiii

Proficiency Level Distribution: Spek 2-3 B/C S400 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	6,721	4.41%	5,978	4.32%	12,699	4.37%
2	22,027	14.47%	25,018	18.06%	47,045	16.18%
3	15,887	10.43%	18,204	13.14%	34,091	11.72%
4	20,979	13.78%	20,580	14.86%	41,559	14.29%
5	23,608	15.50%	19,909	14.37%	43,517	14.97%
6	63,043	41.40%	48,811	35.24%	111,854	38.47%
Total	152,265	100.00%	138,500	100.00%	290,765	100.00%

Table 3.3.2.4Giii
Equating Summary: Spek 2-3 B/C S400 Online
n/a for S400

Figure 3.3.2.4Hiii
Test Characteristic Curve: Spek 2-3 B/C S400 Online
n/a for S400

Figure 3.3.2.4Iiii
Test Information Function: Spek 2-3 B/C S400 Online
n/a for S400

Table 3.3.2.4Jiii

Reliability: Spek 2-3 B/C S400 Online

Reliability	**No. of Students**	**No. of Items**	**Cronbach's Alpha**		**SEM**
	290,765	6	.752		1.555
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	120,912	72	28	0
	2	120,834	70	29	0
	3	123,890	76	24	0
	4	124,160	70	30	0
	5	125,336	73	26	1
	6	125,425	68	31	1

Table 3.3.2.4Kiii
Conditional Standard Error of Measurement at Cut Scores: Spek 2-3 B/C S400 Online
n/a

Table 3.3.2.4Liii
Accuracy and Consistency of Classification Indices: Spek 2-3 B/C S400 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 S400 Online
n/a

Figure 3.3.2.4Civ
Raw Scores: Spek 2-3 S400 Online

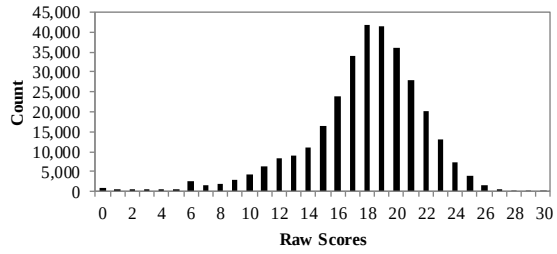

Table 3.3.2.4Civ
Raw Score Descriptive Statistics: Spek 2-3 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	160,457	0	30	17.34	3.72
3	159,252	0	30	18.28	4.23
Total	319,709	0	30	17.81	4.01

Figure 3.3.2.4Div
Scale Scores: Spek 2-3 S400 Online

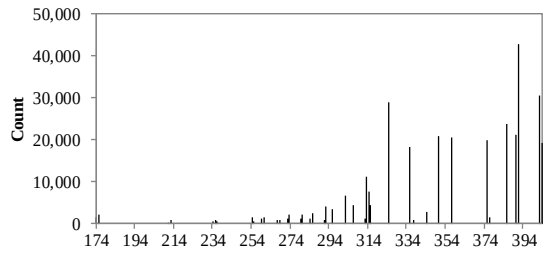

Table 3.3.2.4Div
Scale Score Descriptive Statistics: Spek 2-3 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	160,457	174	391	355.16	44.42
3	159,252	175	403	353.54	46.25
Total	319,709	174	403	354.36	45.35

Figure 3.3.2.4Eiv
Proficiency Level: Spek 2-3 S400 Online

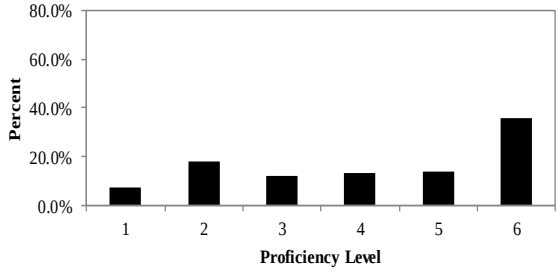

Table 3.3.2.4Eiv

Proficiency Level Distribution: Spek 2-3 S400 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	10,717	6.68%	13,156	8.26%	23,873	7.47%
2	24,233	15.10%	33,046	20.75%	57,279	17.92%
3	16,898	10.53%	21,161	13.29%	38,059	11.90%
4	20,979	13.07%	20,580	12.92%	41,559	13.00%
5	23,608	14.71%	21,453	13.47%	45,061	14.09%
6	64,022	39.90%	49,856	31.31%	113,878	35.62%
Total	160,457	100.00%	159,252	100.00%	319,709	100.00%

Table 3.3.2.4Fiv
Raw Score to Scale Score Conversion: Spek 2-3 S400 Online
n/a

Table 3.3.2.4Giv
Equating Summary: Spek 2-3 S400 Online
n/a

Figure 3.3.2.4Hiv
Test Characteristic Curve: Spek 2-3 S400 Online
n/a for S400

Figure 3.3.2.4Iiv
Test Information Function: Spek 2-3 S400 Online
n/a for S400

Table 3.3.2.4Jiv

Reliability: Spek 2-3 Weighted Reliability S400 Online

Tiers	No. of Students	Reliability	Weighted Reliability
Pre-A	1,374	0.820	0.760
A	27,570	0.838	
B/C	290,765	0.752	

Table 3.3.2.4Kiv

Conditional Standard Error of Measurement at Cut Scores: Spek 2-3 S400 Online

Proficiency Level	Grade	Cut Score	SEM
1/2	2	286	21.43
	3	293	18.88
2/3	2	322	20.41
	3	326	22.45
3/4	2	345	19.39
	3	346	23.98
4/5	2	368	19.39
	3	369	23.98
5/6	2	386	19.39
	3	389	24.49

Table 3.3.2.4Li

Accuracy and Consistency of Classification Indices: Spek (Grade 2) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.486	0.409		0.246	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.569		0.403	
	2	0.534		0.424	
	3	0.291		0.223	
	4	0.308		0.224	
	5	0.244		0.195	
	6	0.809		0.691	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.943	0.030	0.027	0.915
	2/3	0.895	0.039	0.066	0.861
	3/4	0.885	0.048	0.066	0.838
	4/5	0.869	0.069	0.062	0.806
	5/6	0.785	0.158	0.057	0.727

Table 3.3.2.4Lii

Accuracy and Consistency of Classification Indices: Spek (Grade 3) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.459	0.383		0.244	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.586		0.412	
	2	0.553		0.444	
	3	0.291		0.232	
	4	0.264		0.197	
	5	0.266		0.201	
	6	0.819		0.665	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.932	0.036	0.032	0.897
	2/3	0.850	0.048	0.101	0.811
	3/4	0.855	0.048	0.098	0.799
	4/5	0.860	0.074	0.066	0.787
	5/6	0.825	0.136	0.039	0.773

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 S400 Online
n/a

Table 3.3.3.1C
Raw Score Descriptive Statistics: List 4-5 S400 Online
n/a

Figure 3.3.3.1D
Scale Scores: List 4-5 S400 Online

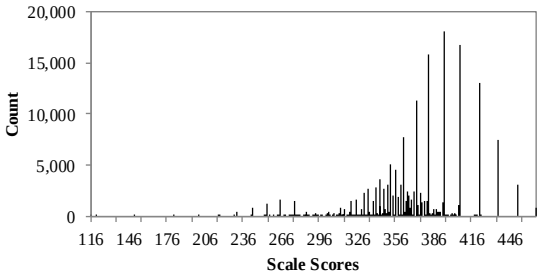

Figure 3.3.3.1E
Proficiency Level: List 4-5 S400 Online

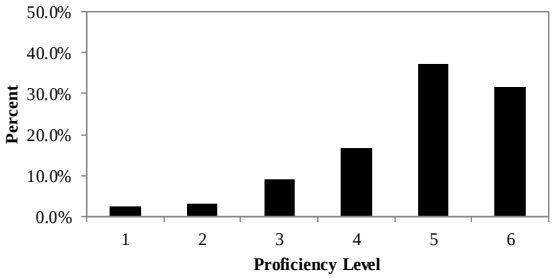

Table 3.3.3.1D
Scale Score Descriptive Statistics: List 4-5 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	99,842	116	467	371.56	38.38
5	76,701	120	467	377.95	41.07
Total	176,543	116	467	374.34	39.70

Table 3.3.3.1E
Proficiency Level Distribution: List 4-5 S400 Online

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,819	1.82%	2,299	3.00%	4,118	2.33%
2	2,887	2.89%	2,635	3.44%	5,522	3.13%
3	8,424	8.44%	7,858	10.24%	16,282	9.22%
4	15,209	15.23%	14,063	18.33%	29,272	16.58%
5	38,157	38.22%	27,413	35.74%	65,570	37.14%
6	33,346	33.40%	22,433	29.25%	55,779	31.60%
Total	99,842	100.00%	76,701	100.00%	176,543	100.00%

Table 3.3.3.1F
Raw Score to Scale Score Conversion: List 4-5 S400 Online
n/a

Table 3.3.3.1G
Equating Summary: List 4-5 S400 Online
n/a for S400

Figure 3.3.3.1H
Test Characteristic Curve: List 4-5 S400 Online
n/a

Figure 3.3.3.1I
Test Information Function: List 4-5 S400 Online

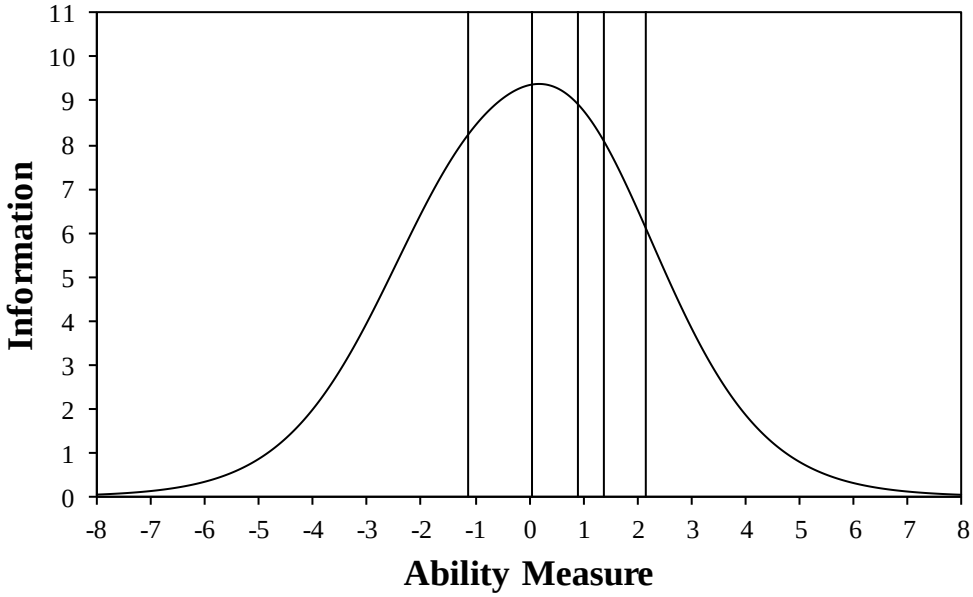

Table 3.3.3.1J
Reliability: List 4-5 S400 Online

No. of Students	No. of Items	Rasch Reliability Estimate
176,543	54	.72

Table 3.3.3.1K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: List 4-5 S400 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	4	264	26	20.29	20.70	20.59	0.19
	5	274	N/A	N/A	N/A	N/A	N/A
2/3	4	307	293	18.45	18.52	18.45	0.02
	5	318	2	18.07	18.07	18.07	0.00
3/4	4	338	1,034	18.30	18.97	18.36	0.15
	5	350	2,184	18.94	20.21	19.14	0.06
4/5	4	355	N/A	N/A	N/A	N/A	N/A
	5	368	90	19.57	20.29	19.72	0.22
5/6	4	383	18	21.04	21.04	21.04	0.00
	5	397	20	22.35	22.35	22.35	0.00

Table 3.3.3.1Li

Accuracy and Consistency of Classification Indices: List (Grade 4) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.558	0.456	0.245		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.885	0.661		
	2	0.395	0.240		
	3	0.383	0.243		
	4	0.353	0.262		
	5	0.566	0.484		
	6	0.664	0.572		
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.988	0.001	0.011	0.985
	2/3	0.974	0.009	0.017	0.953
	3/4	0.910	0.046	0.044	0.862
	4/5	0.833	0.071	0.097	0.777
	5/6	0.796	0.071	0.133	0.725

Table 3.3.3.1Lii

Accuracy and Consistency of Classification Indices: List (Grade 5) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.519	0.416	0.219		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.890	0.689		
	2	0.351	0.217		
	3	0.390	0.258		
	4	0.381	0.288		
	5	0.513	0.438		
	6	0.631	0.519		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.982	0.002	0.016	0.978
	2/3	0.966	0.014	0.019	0.940
	3/4	0.893	0.055	0.052	0.843
	4/5	0.814	0.072	0.114	0.756
	5/6	0.795	0.081	0.124	0.721

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 S400 Online
n/a

Table 3.3.3.2C
Raw Score Descriptive Statistics: Read 4-5 S400 Online
n/a

Figure 3.3.3.2D
Scale Scores: Read 4-5 S400 Online

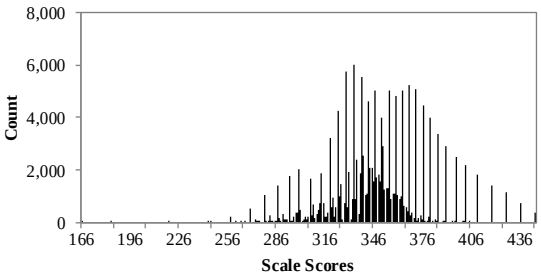

Figure 3.3.3.2E
Proficiency Level: Read 4-5 S400 Online

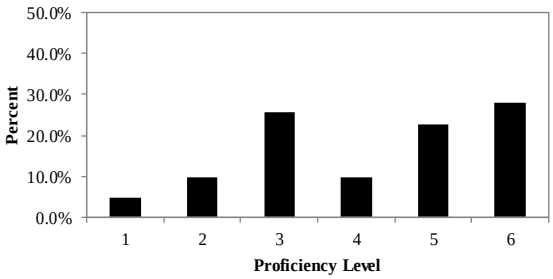

Table 3.3.3.2D
Scale Score Descriptive Statistics: Read 4-5 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	87,027	166	446	345.97	27.63
5	66,702	184	446	352.22	31.42
Total	153,729	166	446	348.68	29.49

Table 3.3.3.2E
Proficiency Level Distribution: Read 4-5 S400 Online

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,547	2.93%	4,558	6.83%	7,105	4.62%
2	7,139	8.20%	7,595	11.39%	14,734	9.58%
3	19,882	22.85%	19,516	29.26%	39,398	25.63%
4	9,867	11.34%	4,983	7.47%	14,850	9.66%
5	21,626	24.85%	13,105	19.65%	34,731	22.59%
6	25,966	29.84%	16,945	25.40%	42,911	27.91%
Total	87,027	100.00%	66,702	100.00%	153,729	100.00%

Table 3.3.3.2F
Raw Score to Scale Score Conversion: Read 4-5 S400 Online
n/a

Table 3.3.3.2G
Equating Summary: Read 4-5 S400 Online
n/a for S400

Figure 3.3.3.2H
Test Characteristic Curve: Read 4-5 S400 Online
n/a

Figure 3.3.3.2I
Test Information Function: Read 4-5 S400 Online

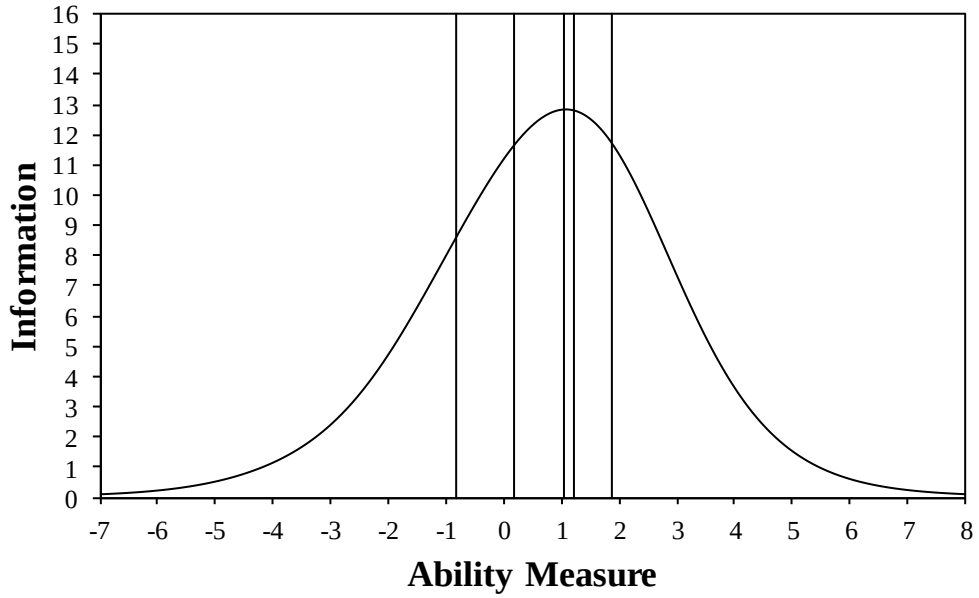

Table 3.3.3.2J
Reliability: Read 4-5 S400 Online

No. of Students	No. of Items	Rasch Reliability Estimate
153,729	66	.85

Table 3.3.3.2K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: Read 4-5 S400 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	4	291	120	15.37	15.37	15.37	0.00
	5	302	20	12.90	13.47	13.21	0.29
2/3	4	316	131	11.36	11.65	11.48	0.11
	5	328	107	11.00	11.47	11.07	0.14
3/4	4	336	1,834	10.79	11.10	10.93	0.03
	5	350	433	10.40	10.87	10.57	0.07
4/5	4	343	1,816	10.56	10.97	10.75	0.05
	5	355	867	10.35	10.71	10.48	0.08
5/6	4	360	3,914	10.37	10.74	10.38	0.03
	5	372	2,366	10.50	10.95	10.58	0.02

Table 3.3.3.2Li

Accuracy and Consistency of Classification Indices: Read (Grade 4) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.601	0.508	0.367		
Conditional on Level	Level	Accuracy	Consistency		
	1	0.811	0.462		
	2	0.461	0.349		
	3	0.583	0.469		
	4	0.271	0.208		
	5	0.547	0.442		
	6	0.838	0.746		
Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.972	0.000	0.027	0.970
	2/3	0.935	0.031	0.034	0.900
	3/4	0.870	0.070	0.060	0.822
	4/5	0.864	0.071	0.065	0.815
	5/6	0.897	0.056	0.047	0.854

Table 3.3.3.2Lii

Accuracy and Consistency of Classification Indices: Read (Grade 5) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.587	0.492	0.358		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.800	0.561		
	2	0.427	0.328		
	3	0.624	0.519		
	4	0.178	0.136		
	5	0.482	0.376		
	6	0.832	0.729		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.950	0.006	0.044	0.937
	2/3	0.908	0.055	0.037	0.864
	3/4	0.861	0.073	0.067	0.812
	4/5	0.867	0.067	0.066	0.818
	5/6	0.909	0.050	0.041	0.868

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 S400 Online

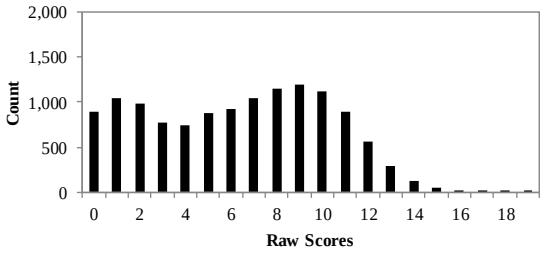

Table 3.3.3.3Ci
Raw Score Descriptive Statistics: Writ 4-5 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	5,144	0	17	5.38	3.71
5	7,551	0	19	6.96	3.85
Total	12,695	0	19	6.32	3.88

Figure 3.3.3.3Di
Scale Scores: Writ 4-5 A S400 Online

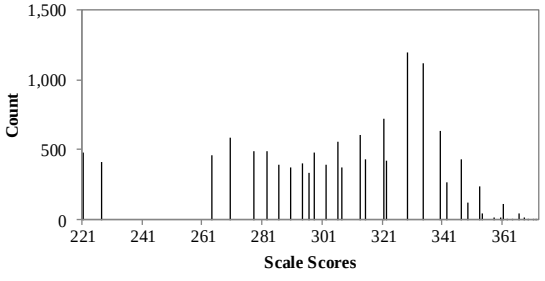

Table 3.3.3.3Di
Scale Score Descriptive Statistics: Writ 4-5 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	5,144	221	364	299.35	33.76
5	7,551	227	372	310.40	31.80
Total	12,695	221	372	305.92	33.05

Figure 3.3.3.3Ei
Proficiency Level: Writ 4-5 A S400 Online

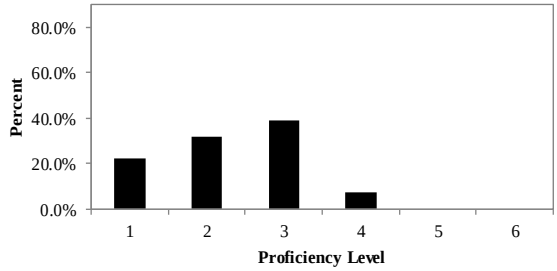

Table 3.3.3.3Ei

Proficiency Level Distribution: Writ 4-5 A S400 Online

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,063	20.66%	1,756	23.26%	2,819	22.21%
2	1,963	38.16%	2,048	27.12%	4,011	31.60%
3	1,638	31.84%	3,328	44.07%	4,966	39.12%
4	480	9.33%	419	5.55%	899	7.08%
5	0	0.00%	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%	0	0.00%
Total	5,144	100.00%	7,551	100.00%	12,695	100.00%

Table 3.3.3.3Gi
Equating Summary: Writ 4-5 A S400 Online
n/a for S400

Figure 3.3.3.3Hi
Test Characteristic Curve: Writ 4-5 A S400 Online
n/a for S400

Figure 3.3.3.3Ii
Test Information Function: Writ 4-5 A S400 Online
n/a for S400

Table 3.3.3.3Ji
Reliability: Writ 4-5 A S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Response Modes**		**Cronbach's Alpha**	**SEM**
	12,695	3	Hand-written (HW)	Keyboarded (KB)	.862	1.440
Interrater Reliability	**Task**	**Mode of Response**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	HW	4,766	98	2	0
		KB	3,887	94	6	0
	2	HW	2,736	97	3	0
		KB	4,909	95	5	0
	3	HW	4,006	99	1	0
		KB	4,516	95	5	0

Table 3.3.3.3Ki
Conditional Standard Error of Measurement at Cut Scores: Writ 4-5 A S400 Online
n/a

Table 3.3.3.3Li
Accuracy and Consistency of Classification Indices: Writ 4-5 A S400 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 S400 Online

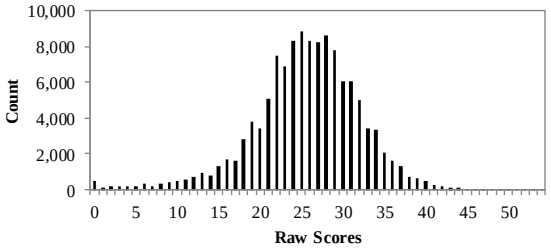

Table 3.3.3.3Cii
Raw Score Descriptive Statistics: Writ 4-5 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	70,881	0	52	24.45	6.53
5	50,973	0	54	27.16	6.04
Total	121,854	0	54	25.59	6.47

Figure 3.3.3.3Dii
Scale Scores: Writ 4-5 B/C S400 Online

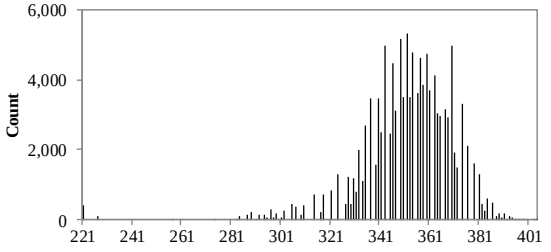

Table 3.3.3.3Dii
Scale Score Descriptive Statistics: Writ 4-5 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	70,881	221	400	349.27	19.71
5	50,973	227	404	355.61	16.85
Total	121,854	221	404	351.92	18.83

Figure 3.3.3.3Eii
Proficiency Level: Writ 4-5 B/C S400 Online

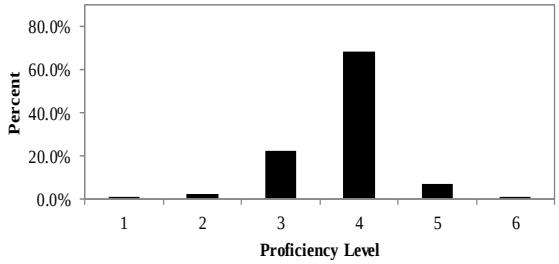

Table 3.3.3.3Eii

Proficiency Level Distribution: Writ 4-5 B/C S400 Online

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	423	0.60%	180	0.35%	603	0.49%
2	2,051	2.89%	996	1.95%	3,047	2.50%
3	12,738	17.97%	14,056	27.58%	26,794	21.99%
4	49,853	70.33%	33,202	65.14%	83,055	68.16%
5	5,775	8.15%	2,537	4.98%	8,312	6.82%
6	41	0.06%	2	0.00%	43	0.04%
Total	70,881	100.00%	50,973	100.00%	121,854	100.00%

Table 3.3.3.3Gii
Equating Summary: Writ 4-5 B/C S400 Online
n/a for S400

Figure 3.3.3.3Hii
Test Characteristic Curve: Writ 4-5 B/C S400 Online
n/a for S400

Figure 3.3.3.3Iii
Test Information Function: Writ 4-5 B/C S400 Online
n/a for S400

Table 3.3.3.3Jii
Reliability: Writ 4-5 B/C S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Response Modes**		**Cronbach's Alpha**	**SEM**
	121,854	3	Hand-written (HW)	Keyboarded (KB)	.916	1.878
Interrater Reliability	**Task**	**Mode of Response**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	HW	20,666	93	7	0
		KB	56,758	92	7	0
	2	HW	20,410	96	4	0
		KB	56,546	95	5	0
	3	HW	20,028	95	5	0
		KB	56,474	93	6	0

Table 3.3.3.3Kii
Conditional Standard Error of Measurement at Cut Scores: Writ 4-5 B/C S400 Online
n/a

Table 3.3.3.3Lii
Accuracy and Consistency of Classification Indices: Writ 4-5 B/C S400 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 S400 Online
n/a

Table 3.3.3.3Biii
DIF Analysis and Summary: Writ 4-5 S400 Online
n/a

Figure 3.3.3.3Ciii
Raw Scores: Writ 4-5 S400 Online

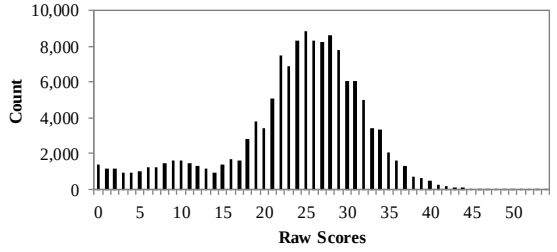

Table 3.3.3.3Ciii
Raw Score Descriptive Statistics: Writ 4-5 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	76,025	0	52	23.16	7.98
5	58,524	0	54	24.56	8.92
Total	134,549	0	54	23.77	8.43

Figure 3.3.3.3Diii
Scale Scores: Writ 4-5 S400 Online

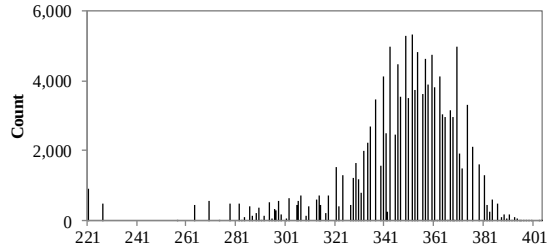

Table 3.3.3.3Diii
Scale Score Descriptive Statistics: Writ 4-5 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	76,025	221	400	345.90	24.42
5	58,524	227	404	349.78	24.65
Total	134,549	221	404	347.58	24.60

Figure 3.3.3.3Eiii
Proficiency Level: Writ 4-5 S400 Online

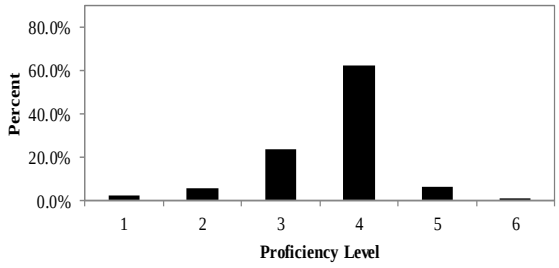

Table 3.3.3.3Eiii

Proficiency Level Distribution: Writ 4-5 S400 Online

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,486	1.95%	1,936	3.31%	3,422	2.54%
2	4,014	5.28%	3,044	5.20%	7,058	5.25%
3	14,376	18.91%	17,384	29.70%	31,760	23.60%
4	50,333	66.21%	33,621	57.45%	83,954	62.40%
5	5,775	7.60%	2,537	4.33%	8,312	6.18%
6	41	0.05%	2	0.00%	43	0.03%
Total	76,025	100.00%	58,524	100.00%	134,549	100.00%

Table 3.3.3.3Fiii
Raw Score to Scale Score Conversion: Writ 4-5 S400 Online
n/a

Table 3.3.3.3Giii
Equating Summary: Writ 4-5 S400 Online
n/a

Figure 3.3.3.3Hiii
Test Characteristic Curve: Writ 4-5 S400 Online
n/a for S400

Figure 3.3.3.3Iiii
Test Information Function: Writ 4-5 S400 Online
n/a for S400

Table 3.3.3.3Jiii

Reliability: Writ 4-5 Weighted Reliability S400 Online

Tiers	No. of Students	Reliability	Weighted Reliability
A	12,695	0.862	0.911
B/C	121,854	0.916	

Table 3.3.3.3Kiii

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

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

Table 3.3.3.3Li

Accuracy and Consistency of Classification Indices: Writ (Grade 4) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.841	0.793	0.567		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.843	0.783		
	2	0.774	0.667		
	3	0.849	0.738		
	4	0.844	0.834		
	5	-	0.131		
	6	-	1.000		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.994	0.003	0.003	0.992
	2/3	0.981	0.010	0.009	0.974
	3/4	0.941	0.016	0.043	0.917
	4/5	0.923	0.076	0.000	0.910
	5/6	0.999	0.001	0.000	1.000

Table 3.3.3.3Lii

Accuracy and Consistency of Classification Indices: Writ (Grade 5) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.825	0.770	0.580		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.912	0.855		
	2	0.681	0.555		
	3	0.864	0.742		
	4	0.820	0.801		
	5	-	0.090		
	6	-	1.000		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.993	0.003	0.005	0.990
	2/3	0.977	0.014	0.009	0.965
	3/4	0.899	0.023	0.078	0.859
	4/5	0.957	0.043	0.000	0.954
	5/6	1.000	0.000	0.000	1.000

3.3.3.4 Speaking 4-5

3.3.3.4i Speaking 4-5 Pre-A

Table 3.3.3.4Ai
Complete Task Analysis and Summary: Spek 4-5 Pre-A S400 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 3.3.3.4Ci
Raw Scores: Spek 4-5 Pre-A S400 Online

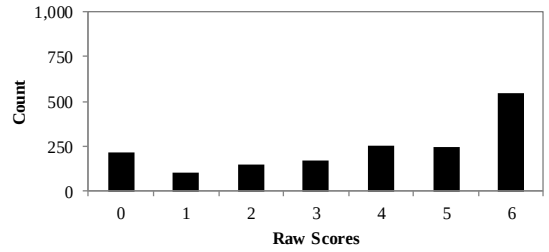

Table 3.3.3.4Ci
Raw Score Descriptive Statistics: Spek 4-5 Pre-A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	490	0	6	3.50	2.22
5	1,190	0	6	3.95	2.05
Total	1,680	0	6	3.82	2.11

Figure 3.3.3.4Di
Scale Scores: Spek 4-5 Pre-A S400 Online

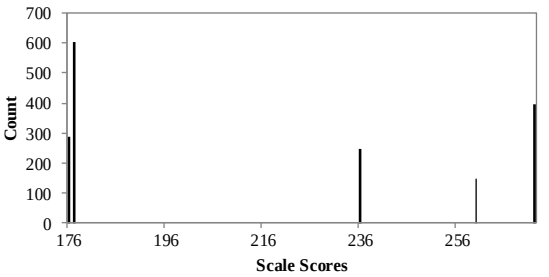

Table 3.3.3.4Di
Scale Score Descriptive Statistics: Spek 4-5 Pre-A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	490	176	260	207.98	38.83
5	1,190	177	272	218.07	43.25
Total	1,680	176	272	215.13	42.25

Figure 3.3.3.4Ei
Proficiency Level: Spek 4-5 Pre-A S400 Online

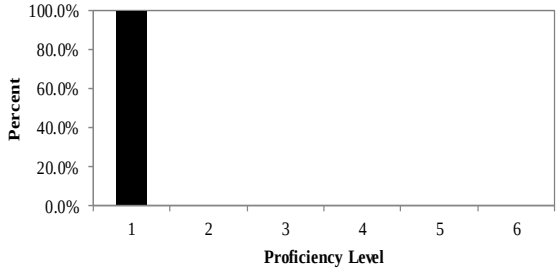

Table 3.3.3.4Ei
Proficiency Level Distribution: Spek 4-5 Pre-A S400 Online

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	490	100.00%	1,190	100.00%	1,680	100.00%
Total	490	100.00%	1,190	100.00%	1,680	100.00%

Table 3.3.3.4Gi
Equating Summary: Spek 4-5 Pre-A S400 Online
n/a for S400

Figure 3.3.3.4Hi
Test Characteristic Curve: Spek 4-5 Pre-A S400 Online
n/a for S400

Figure 3.3.3.4Ii
Test Information Function: Spek 4-5 Pre-A S400 Online
n/a for S400

Table 3.3.3.4Ji
Reliability: Spek 4-5 Pre-A S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Cronbach's Alpha**		**SEM**
	1,680	3	.775		1.002
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	1,116	95	5	0
	2	1,236	94	6	0
	3	1,014	97	3	0

Table 3.3.3.4Ki
Conditional Standard Error of Measurement at Cut Scores: Spek 4-5 Pre-A S400 Online
n/a

Table 3.3.3.4Li
Accuracy and Consistency of Classification Indices: Spek 4-5 Pre-A S400 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 S400 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 S400 Online

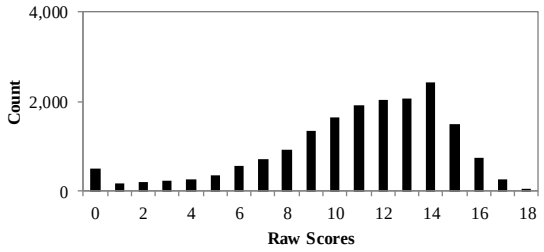

Table 3.3.3.4Cii
Raw Score Descriptive Statistics: Spek 4-5 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	7,638	0	18	10.09	3.95
5	10,148	0	18	11.49	3.58
Total	17,786	0	18	10.89	3.80

Figure 3.3.3.4Dii
Scale Scores: Spek 4-5 A S400 Online

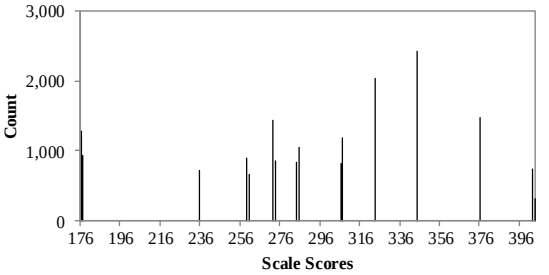

Table 3.3.3.4Dii
Scale Score Descriptive Statistics: Spek 4-5 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	7,638	176	403	281.11	61.53
5	10,148	177	403	305.51	59.09
Total	17,786	176	403	295.03	61.35

Figure 3.3.3.4Eii
Proficiency Level: Spek 4-5 A S400 Online

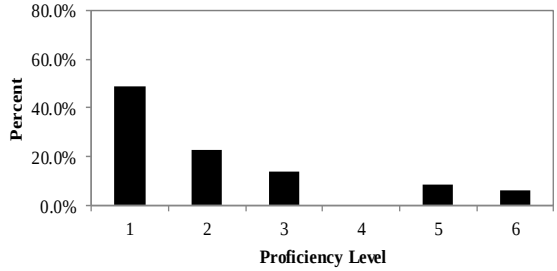

Table 3.3.3.4Eii

Proficiency Level Distribution: Spek 4-5 A S400 Online

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	4,451	58.27%	4,277	42.15%	8,728	49.07%
2	1,625	21.28%	2,448	24.12%	4,073	22.90%
3	832	10.89%	1,601	15.78%	2,433	13.68%
4	0	0.00%	0	0.00%	0	0.00%
5	462	6.05%	1,016	10.01%	1,478	8.31%
6	268	3.51%	806	7.94%	1,074	6.04%
Total	7,638	100.00%	10,148	100.00%	17,786	100.00%

Table 3.3.3.4Gii
Equating Summary: Spek 4-5 A S400 Online
n/a for S400

Figure 3.3.3.4Hii
Test Characteristic Curve: Spek 4-5 A S400 Online
n/a for S400

Figure 3.3.3.4Iii
Test Information Function: Spek 4-5 A S400 Online
n/a for S400

Table 3.3.3.4Jii

Reliability: Spek 4-5 A S400 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	17,786	6	.821		1.609
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	8,590	97	3	0
	2	8,474	85	13	1
	3	9,102	97	3	0
	4	8,872	77	22	1
	5	8,508	97	3	0
	6	8,180	81	19	0

Table 3.3.3.4Kii

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

n/a

Table 3.3.3.4Lii

Accuracy and Consistency of Classification Indices: Spek 4-5 A S400 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 S400 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 S400 Online

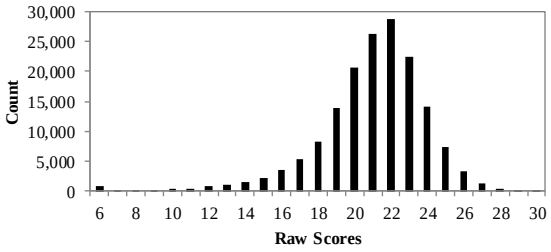

Table 3.3.3.4Ciii
Raw Score Descriptive Statistics: Spek 4-5 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	95,367	6	30	20.73	3.03
5	68,163	6	30	21.32	2.95
Total	163,530	6	30	20.98	3.01

Figure 3.3.3.4Diii
Scale Scores: Spek 4-5 B/C S400 Online

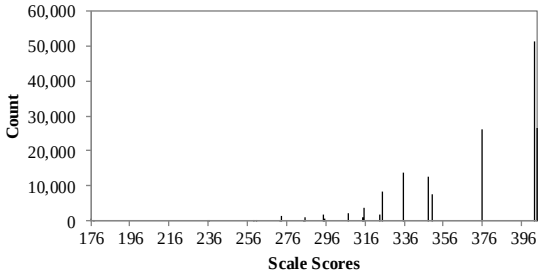

Table 3.3.3.4Diii
Scale Score Descriptive Statistics: Spek 4-5 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	95,367	176	403	367.20	39.52
5	68,163	177	403	375.22	36.16
Total	163,530	176	403	370.54	38.36

Figure 3.3.3.4Eiii
Proficiency Level: Spek 4-5 B/C S400 Online

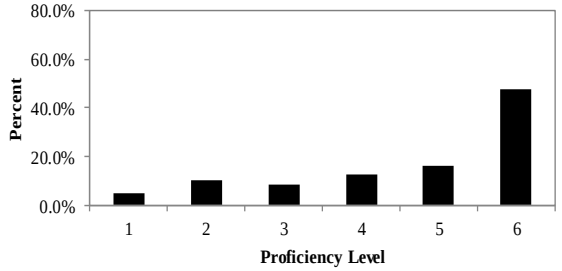

Table 3.3.3.4Eiii

Proficiency Level Distribution: Spek 4-5 B/C S400 Online

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	5,020	5.26%	2,636	3.87%	7,656	4.68%
2	11,299	11.85%	5,912	8.67%	17,211	10.52%
3	8,967	9.40%	4,815	7.06%	13,782	8.43%
4	12,749	13.37%	7,802	11.45%	20,551	12.57%
5	15,622	16.38%	10,716	15.72%	26,338	16.11%
6	41,710	43.74%	36,282	53.23%	77,992	47.69%
Total	95,367	100.00%	68,163	100.00%	163,530	100.00%

Table 3.3.3.4Giii
Equating Summary: Spek 4-5 B/C S400 Online
n/a for S400

Figure 3.3.3.4Hiii
Test Characteristic Curve: Spek 4-5 B/C S400 Online
n/a for S400

Figure 3.3.3.4Iiii
Test Information Function: Spek 4-5 B/C S400 Online
n/a for S400

Table 3.3.3.4Jiii
Reliability: Spek 4-5 B/C S400 Online

Reliability	**No. of Students**	**No. of Items**	**Cronbach's Alpha**		**SEM**
	163,530	6	.727		1.574
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	72,176	89	10	1
	2	71,814	78	21	1
	3	70,094	79	21	0
	4	69,436	70	29	0
	5	69,187	74	25	1
	6	69,151	80	20	0

Table 3.3.3.4Kiii
Conditional Standard Error of Measurement at Cut Scores: Spek 4-5 B/C S400 Online
n/a

Table 3.3.3.4Liii
Accuracy and Consistency of Classification Indices: Spek 4-5 B/C S400 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.

DIF Analysis and Summary: Spek 4-5 S400 Online
n/a

Figure 3.3.3.4Civ
Raw Scores: Spek 4-5 S400 Online

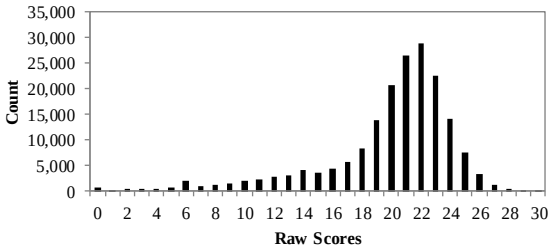

Table 3.3.3.4Civ
Raw Score Descriptive Statistics: Spek 4-5 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	103,495	0	30	19.86	4.32
5	79,501	0	30	19.81	4.87
Total	182,996	0	30	19.84	4.57

Figure 3.3.3.4Div
Scale Scores: Spek 4-5 S400 Online

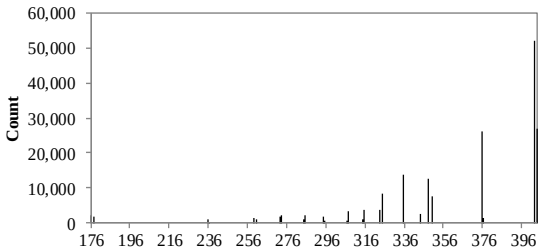

Table 3.3.3.4Div
Scale Score Descriptive Statistics: Spek 4-5 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	103,495	176	403	360.09	48.40
5	79,501	177	403	363.97	49.58
Total	182,996	176	403	361.78	48.95

Figure 3.3.3.4Eiv
Proficiency Level: Spek 4-5 S400 Online

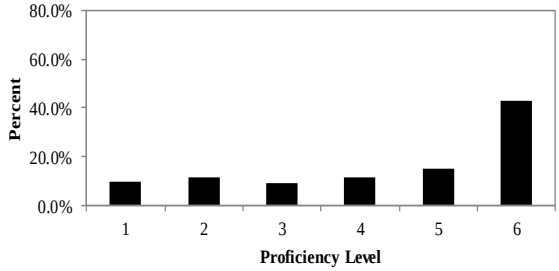

Table 3.3.3.4Eiv

Proficiency Level Distribution: Spek 4-5 S400 Online

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	9,961	9.62%	8,103	10.19%	18,064	9.87%
2	12,924	12.49%	8,360	10.52%	21,284	11.63%
3	9,799	9.47%	6,416	8.07%	16,215	8.86%
4	12,749	12.32%	7,802	9.81%	20,551	11.23%
5	16,084	15.54%	11,732	14.76%	27,816	15.20%
6	41,978	40.56%	37,088	46.65%	79,066	43.21%
Total	103,495	100.00%	79,501	100.00%	182,996	100.00%

Table 3.3.3.4Fiv
Raw Score to Scale Score Conversion: Spek 4-5 S400 Online
n/a

Table 3.3.3.4Giv
Equating Summary: Spek 4-5 S400 Online
n/a

Figure 3.3.3.4Hiv
Test Characteristic Curve: Spek 4-5 S400 Online
n/a for S400

Figure 3.3.3.4Iiv
Test Information Function: Spek 4-5 S400 Online
n/a for S400

Table 3.3.3.4Jiv

Reliability: Spek 4-5 Weighted Reliability S400 Online

Tiers	No. of Students	Reliability	Weighted Reliability
Pre-A	1,680	0.775	0.736
A	17,786	0.821	
B/C	163,530	0.727	

Table 3.3.3.4Kiv

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

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

Table 3.3.3.4Li

Accuracy and Consistency of Classification Indices: Spek (Grade 4) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.408	0.376		0.203	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.689		0.534	
	2	0.424		0.311	
	3	0.247		0.174	
	4	0.234		0.161	
	5	0.221		0.193	
	6	0.741		0.625	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.939	0.029	0.032	0.912
	2/3	0.886	0.037	0.077	0.849
	3/4	0.868	0.040	0.092	0.815
	4/5	0.833	0.070	0.098	0.743
	5/6	0.707	0.233	0.060	0.676

Table 3.3.3.4Lii

Accuracy and Consistency of Classification Indices: Spek (Grade 5) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.246	0.365		0.171	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.725		0.586	
	2	0.412		0.281	
	3	0.197		0.122	
	4	0.156		0.111	
	5	0.179		0.172	
	6	-		0.623	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.942	0.027	0.031	0.918
	2/3	0.902	0.026	0.072	0.867
	3/4	0.868	0.031	0.102	0.804
	4/5	0.813	0.072	0.115	0.694
	5/6	0.533	0.467	0.000	0.617

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 S400 Online
n/a

Table 3.3.4.1C
Raw Score Descriptive Statistics: List 6-8 S400 Online
n/a

Figure 3.3.4.1D
Scale Scores: List 6-8 S400 Online

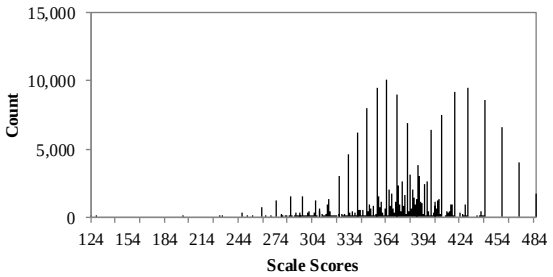

Figure 3.3.4.1E
Proficiency Level: List 6-8 S400 Online

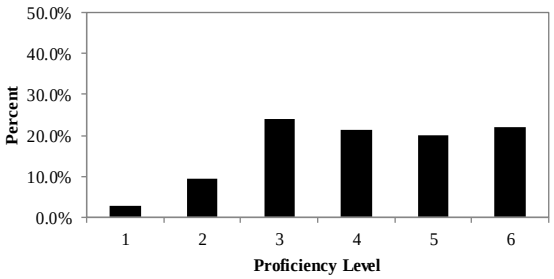

Table 3.3.4.1D

Scale Score Descriptive Statistics: List 6-8 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	63,480	124	486	374.84	38.71
7	63,119	128	486	382.31	43.57
8	64,474	198	486	389.17	46.01
Total	191,073	124	486	382.14	43.29

Table 3.3.4.1E

Proficiency Level Distribution: List 6-8 S400 Online

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	933	1.47%	1,835	2.91%	2,606	4.04%	5,374	2.81%
2	5,041	7.94%	6,160	9.76%	6,841	10.61%	18,042	9.44%
3	15,492	24.40%	15,073	23.88%	15,409	23.90%	45,974	24.06%
4	14,557	22.93%	13,243	20.98%	13,220	20.50%	41,020	21.47%
5	14,825	23.35%	12,596	19.96%	11,015	17.08%	38,436	20.12%
6	12,632	19.90%	14,212	22.52%	15,383	23.86%	42,227	22.10%
Total	63,480	100.00%	63,119	100.00%	64,474	100.00%	191,073	100.00%

Table 3.3.4.1F
Raw Score to Scale Score Conversion: List 6-8 S400 Online
n/a

Table 3.3.4.1G
Equating Summary: List 6-8 S400 Online
n/a for S400

Figure 3.3.4.1H
Test Characteristic Curve: List 6-8 S400 Online
n/a

Figure 3.3.4.1I
Test Information Function: List 6-8 S400 Online

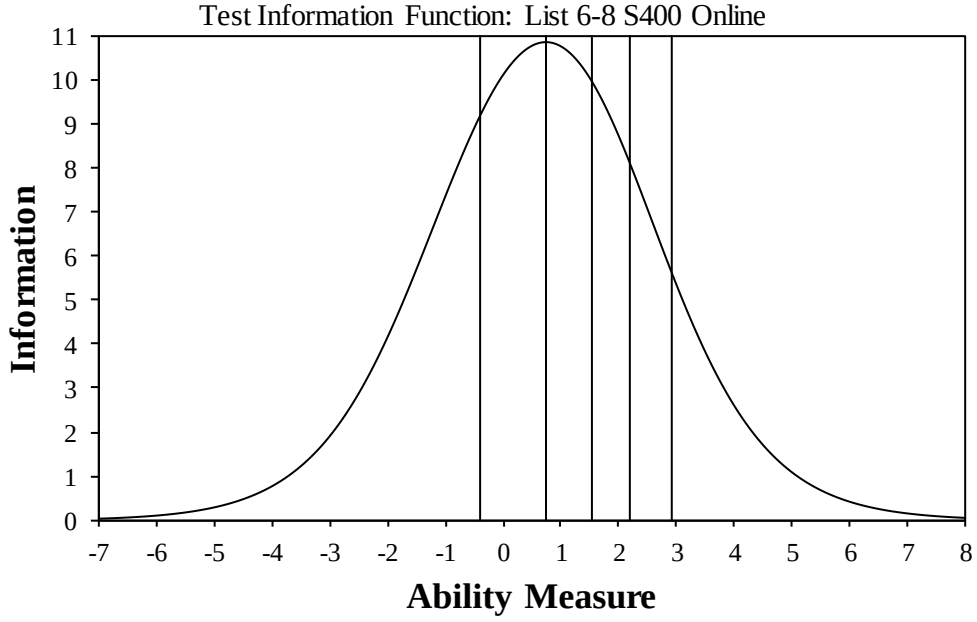

Table 3.3.4.1J
Reliability: List 6-8 S400 Online

No. of Students	No. of Items	Rasch Reliability Estimate
191,073	54	.79

Table 3.3.4.1K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: List 6-8 S400 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	6	283	96	19.16	21.38	21.31	0.39
	7	293	2	17.66	18.33	18.00	0.48
	8	302	3	17.32	17.92	17.52	0.35
2/3	6	328	19	16.76	17.17	16.90	0.08
	7	337	17	17.13	17.17	17.13	0.01
	8	345	106	17.06	17.32	17.10	0.04
3/4	6	359	169	17.09	17.28	17.12	0.06
	7	368	347	17.21	18.26	17.46	0.07
	8	375	329	17.70	19.16	17.81	0.23
4/5	6	380	230	17.51	17.81	17.72	0.13
	7	390	1,532	18.00	19.01	19.00	0.09
	8	399	25	18.11	18.11	18.11	0.00
5/6	6	409	2,691	18.86	18.86	18.86	0.00
	7	418	10	20.03	20.03	20.03	0.00
	8	426	94	21.64	21.83	21.72	0.09

Table 3.3.4.1Li

Accuracy and Consistency of Classification Indices: List (Grade 6) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.562	0.454	0.308		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	-	0.294		
	2	0.485	0.335		
	3	0.565	0.457		
	4	0.448	0.359		
	5	0.518	0.409		
	6	0.816	0.675		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.985	0.000	0.015	0.983
	2/3	0.930	0.029	0.041	0.893
	3/4	0.848	0.075	0.077	0.795
	4/5	0.855	0.073	0.072	0.798
	5/6	0.909	0.059	0.032	0.870

Table 3.3.4.1Lii

Accuracy and Consistency of Classification Indices: List (Grade 7) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.539	0.437	0.298		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.728	0.359		
	2	0.453	0.329		
	3	0.536	0.432		
	4	0.419	0.333		
	5	0.455	0.351		
	6	0.824	0.688		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.973	0.001	0.026	0.965
	2/3	0.914	0.041	0.044	0.871
	3/4	0.846	0.075	0.079	0.793
	4/5	0.858	0.071	0.071	0.800
	5/6	0.901	0.065	0.034	0.859

Table 3.3.4.1Liii

Accuracy and Consistency of Classification Indices: List (Grade 8) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.530	0.431	0.295		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.742	0.425		
	2	0.445	0.326		
	3	0.533	0.428		
	4	0.413	0.328		
	5	0.401	0.305		
	6	0.830	0.694		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.966	0.003	0.030	0.955
	2/3	0.908	0.047	0.046	0.863
	3/4	0.845	0.070	0.085	0.793
	4/5	0.863	0.066	0.071	0.805
	5/6	0.895	0.070	0.035	0.852

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 S400 Online
n/a

Table 3.3.4.2C
Raw Score Descriptive Statistics: Read 6-8 S400 Online
n/a

Figure 3.3.4.2D
Scale Scores: Read 6-8 S400 Online

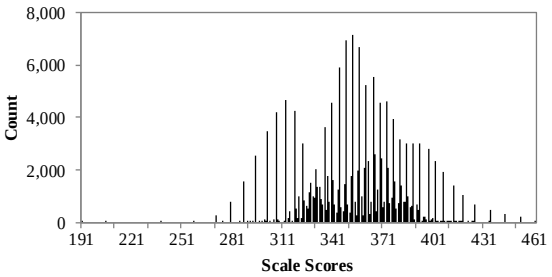

Figure 3.3.4.2E
Proficiency Level: Read 6-8 S400 Online

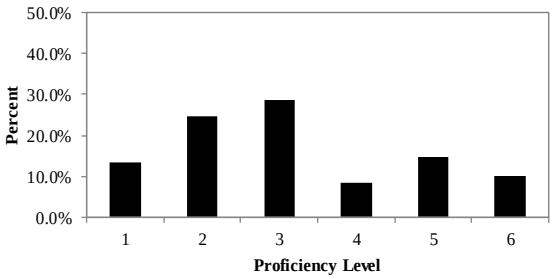

Table 3.3.4.2D

Scale Score Descriptive Statistics: Read 6-8 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	57,603	205	462	348.25	27.08
7	55,527	191	462	354.05	29.09
8	54,813	205	462	360.82	31.04
Total	167,943	191	462	354.27	29.53

Table 3.3.4.2E

Proficiency Level Distribution: Read 6-8 S400 Online

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	5,690	9.88%	7,928	14.28%	8,826	16.10%	22,444	13.36%
2	13,122	22.78%	12,841	23.13%	15,583	28.43%	41,546	24.74%
3	18,703	32.47%	16,756	30.18%	12,667	23.11%	48,126	28.66%
4	5,291	9.19%	5,168	9.31%	3,648	6.66%	14,107	8.40%
5	9,636	16.73%	7,670	13.81%	7,650	13.96%	24,956	14.86%
6	5,161	8.96%	5,164	9.30%	6,439	11.75%	16,764	9.98%
Total	57,603	100.00%	55,527	100.00%	54,813	100.00%	167,943	100.00%

Table 3.3.4.2F
Raw Score to Scale Score Conversion: Read 6-8 S400 Online
n/a

Table 3.3.4.2G
Equating Summary: Read 6-8 S400 Online
n/a for S400

Figure 3.3.4.2H
Test Characteristic Curve: Read 6-8 S400 Online
n/a

Figure 3.3.4.2I
Test Information Function: Read 6-8 S400 Online

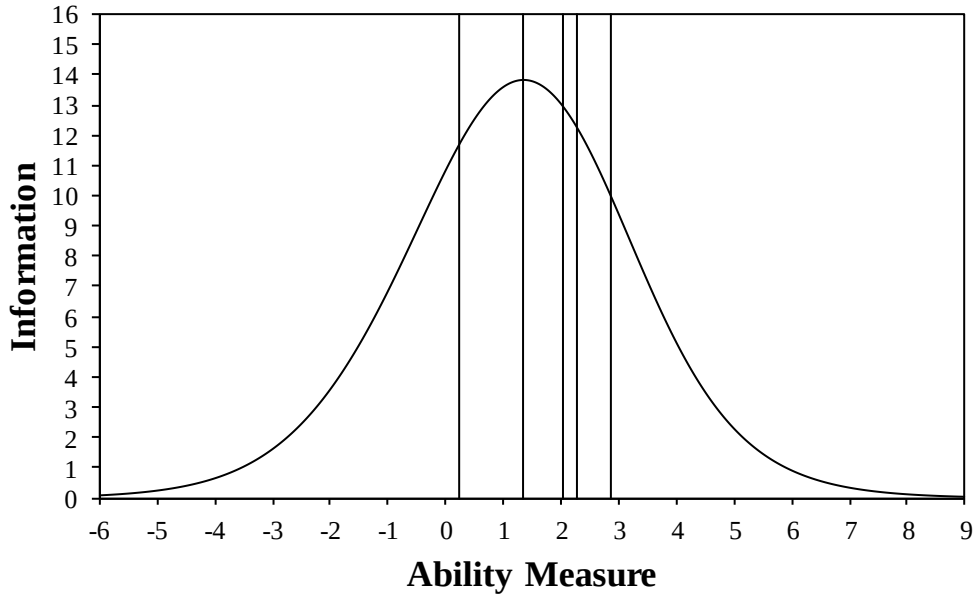

Table 3.3.4.2J
Reliability: Read 6-8 S400 Online

No. of Students	No. of Items	Rasch Reliability Estimate
167,943	69	.86

Table 3.3.4.2K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: Read 6-8 S400 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	6	312	2	12.06	12.06	12.06	0.00
	7	321	212	12.06	12.64	12.50	0.12
	8	329	335	11.62	11.65	11.62	0.00
2/3	6	340	2,489	10.35	10.87	10.60	0.04
	7	349	2,589	10.17	10.61	10.30	0.03
	8	358	103	10.27	10.84	10.41	0.20
3/4	6	360	1,102	10.19	10.48	10.26	0.04
	7	369	1,746	10.19	10.61	10.31	0.04
	8	376	352	10.35	11.57	10.40	0.12
4/5	6	366	1,061	10.22	10.37	10.26	0.05
	7	375	251	10.22	10.56	10.44	0.09
	8	382	1018	10.37	10.95	10.74	0.25
5/6	6	382	53	10.37	10.66	10.46	0.11
	7	391	88	10.66	11.44	10.87	0.24
	8	398	1261	10.92	11.39	10.94	0.09

Table 3.3.4.2Li

Accuracy and Consistency of Classification Indices: Read (Grade 6) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.610	0.502	0.371		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.829	0.684		
	2	0.637	0.525		
	3	0.642	0.544		
	4	0.244	0.184		
	5	0.543	0.418		
	6	0.744	0.564		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.953	0.013	0.033	0.935
	2/3	0.889	0.061	0.050	0.843
	3/4	0.873	0.070	0.058	0.826
	4/5	0.890	0.061	0.049	0.847
	5/6	0.949	0.031	0.020	0.923

Table 3.3.4.2Lii

Accuracy and Consistency of Classification Indices: Read (Grade 7) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.603	0.498	0.371		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.841	0.715		
	2	0.599	0.490		
	3	0.616	0.519		
	4	0.260	0.196		
	5	0.488	0.367		
	6	0.777	0.598		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.938	0.019	0.044	0.914
	2/3	0.885	0.067	0.049	0.837
	3/4	0.876	0.071	0.053	0.832
	4/5	0.898	0.054	0.048	0.856
	5/6	0.950	0.032	0.017	0.926

Table 3.3.4.2Liii

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

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.607	0.504	0.382		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.810	0.683		
	2	0.650	0.548		
	3	0.535	0.429		
	4	0.203	0.149		
	5	0.493	0.375		
	6	0.800	0.648		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.927	0.027	0.046	0.897
	2/3	0.880	0.063	0.057	0.833
	3/4	0.890	0.061	0.049	0.847
	4/5	0.903	0.056	0.041	0.864
	5/6	0.943	0.037	0.020	0.918

3.3.4.3 Writing 6-8

3.3.4.3i Writing 6-8 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 3.3.4.3Ci
Raw Scores: Writ 6-8 A S400 Online

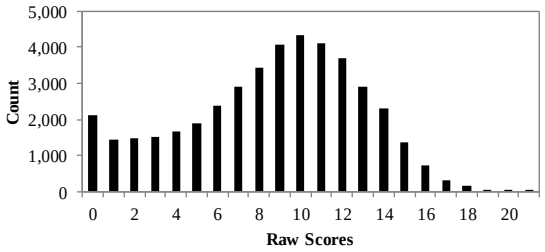

Figure 3.3.4.3Di
Scale Scores: Writ 6-8 A S400 Online

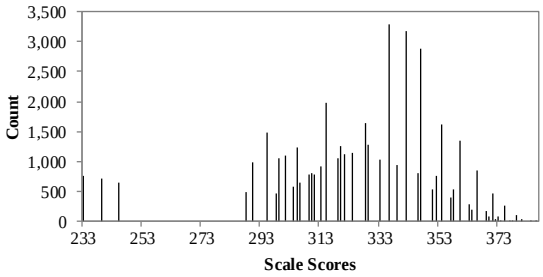

Table 3.3.4.3Ci

Raw Score Descriptive Statistics: Writ 6-8 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	11,375	0	19	7.56	4.14
7	13,019	0	21	8.12	4.17
8	18,481	0	21	9.53	4.19
Total	42,875	0	21	8.58	4.26

Table 3.3.4.3Di

Scale Score Descriptive Statistics: Writ 6-8 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	11,375	233	375	316.67	29.70
7	13,019	239	379	322.58	28.42
8	18,481	245	386	331.61	27.41
Total	42,875	233	386	324.91	29.02

Figure 3.3.4.3Ei
Proficiency Level: Writ 6-8 A S400 Online

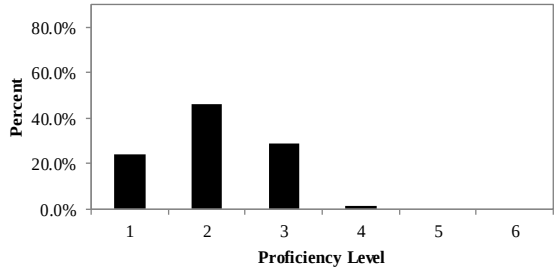

Table 3.3.4.3Ei

Proficiency Level Distribution: Writ 6-8 A S400 Online

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,783	15.67%	3,450	26.50%	5,090	27.54%	10,323	24.08%
2	5,554	48.83%	5,476	42.06%	8,696	47.05%	19,726	46.01%
3	3,710	32.62%	3,977	30.55%	4,637	25.09%	12,324	28.74%
4	328	2.88%	116	0.89%	58	0.31%	502	1.17%
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	11,375	100.00%	13,019	100.00%	18,481	100.00%	42,875	100.00%

Table 3.3.4.3Gi
Equating Summary: Writ 6-8 A S400 Online
n/a for S400

Figure 3.3.4.3Hi
Test Characteristic Curve: Writ 6-8 A S400 Online
n/a for S400

Figure 3.3.4.3Ii
Test Information Function: Writ 6-8 A S400 Online
n/a for S400

Table 3.3.4.3Ji
Reliability: Writ 6-8 A S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Response Modes**		**Cronbach's Alpha**	**SEM**
	42,875	3	Hand-written (HW)	Keyboarded (KB)	.856	1.614
Interrater Reliability	**Task**	**Mode of Response**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	HW	1,066	95	4	1
		KB	16,737	89	10	1
	2	HW	962	96	4	1
		KB	16,977	91	8	1
	3	HW	1,040	96	4	0
		KB	15,817	92	7	1

Table 3.3.4.3Ki
Conditional Standard Error of Measurement at Cut Scores: Writ 6-8 A S400 Online
n/a

Table 3.3.4.3Li
Accuracy and Consistency of Classification Indices: Writ 6-8 A S400 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 agencies (SEAs) may request this information; please contact us at help@wida.us.

Figure 3.3.4.3Cii
Raw Scores: Writ 6-8 B/C S400 Online

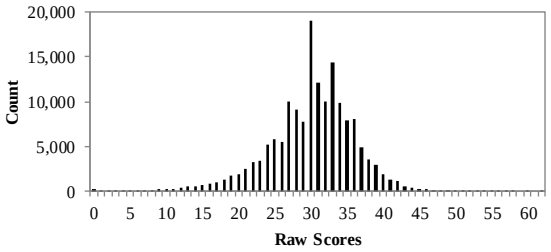

Figure 3.3.4.3Dii
Scale Scores: Writ 6-8 B/C S400 Online

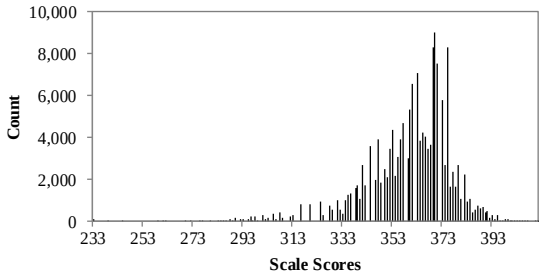

Table 3.3.4.3Cii

Raw Score Descriptive Statistics: Writ 6-8 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	56,675	0	60	27.89	6.28
7	54,664	0	54	30.05	5.88
8	50,387	0	62	32.29	5.38
Total	161,726	0	62	29.99	6.14

Table 3.3.4.3Dii

Scale Score Descriptive Statistics: Writ 6-8 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	56,675	233	410	355.44	18.68
7	54,664	239	404	360.11	17.11
8	50,387	245	411	365.15	15.43
Total	161,726	233	411	360.05	17.64

Figure 3.3.4.3Eii
Proficiency Level: Writ 6-8 B/C S400 Online

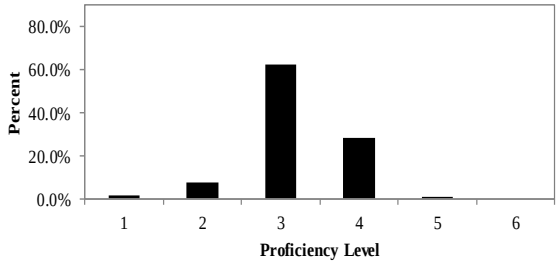

Table 3.3.4.3Eii

Proficiency Level Distribution: Writ 6-8 B/C S400 Online

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	721	1.27%	830	1.52%	938	1.86%	2,489	1.54%
2	3,485	6.15%	4,215	7.71%	4,791	9.51%	12,491	7.72%
3	25,865	45.64%	35,322	64.62%	39,423	78.24%	100,610	62.21%
4	26,326	46.45%	14,246	26.06%	5,234	10.39%	45,806	28.32%
5	278	0.49%	51	0.09%	1	0.00%	330	0.20%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	56,675	100.00%	54,664	100.00%	50,387	100.00%	161,726	100.00%

Table 3.3.4.3Gii
Equating Summary: Writ 6-8 B/C S400 Online
n/a for S400

Figure 3.3.4.3Hii
Test Characteristic Curve: Writ 6-8 B/C S400 Online
n/a for S400

Figure 3.3.4.3Iii
Test Information Function: Writ 6-8 B/C S400 Online
n/a for S400

Table 3.3.4.3Jii
Reliability: Writ 6-8 B/C S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Response Modes**		**Cronbach's Alpha**	**SEM**
	161,726	3	Hand-written (HW)	Keyboarded (KB)	.915	1.794
Interrater Reliability	**Task**	**Mode of Response**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	HW	1,130	96	4	0
		KB	70,600	94	5	0
	2	HW	1,108	93	7	1
		KB	71,284	92	7	1
	3	HW	1,146	90	9	1
		KB	71,628	93	7	1

Table 3.3.4.3Kii
Conditional Standard Error of Measurement at Cut Scores: Writ 6-8 B/C S400 Online
n/a

Table 3.3.4.3Lii
Accuracy and Consistency of Classification Indices: Writ 6-8 B/C S400 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 S400 Online
n/a

Table 3.3.4.3Biii
DIF Analysis and Summary: Writ 6-8 S400 Online
n/a

Figure 3.3.4.3Ciii
Raw Scores: Writ 6-8 S400 Online

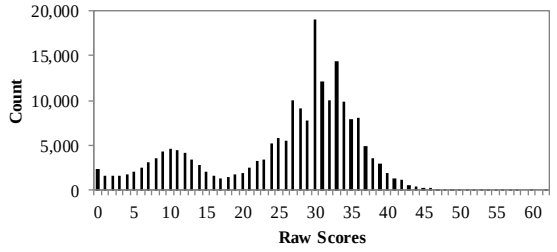

Figure 3.3.4.3Diii
Scale Scores: Writ 6-8 S400 Online

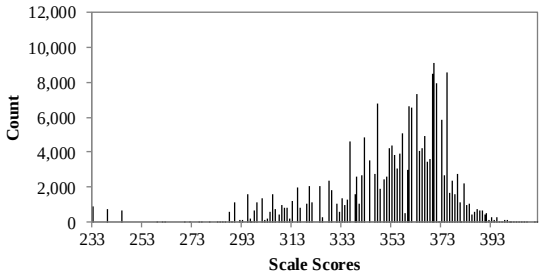

Table 3.3.4.3Ciii
Raw Score Descriptive Statistics: Writ 6-8 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	68,050	0	60	24.49	9.65
7	67,683	0	54	25.83	10.29
8	68,868	0	62	26.18	11.29
Total	204,601	0	62	25.50	10.46

Table 3.3.4.3Diii
Scale Score Descriptive Statistics: Writ 6-8 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	68,050	233	410	348.96	25.45
7	67,683	239	404	352.89	24.71
8	68,868	245	411	356.15	24.43
Total	204,601	233	411	352.68	25.04

Figure 3.3.4.3Eiii
Proficiency Level: Writ 6-8 S400 Online

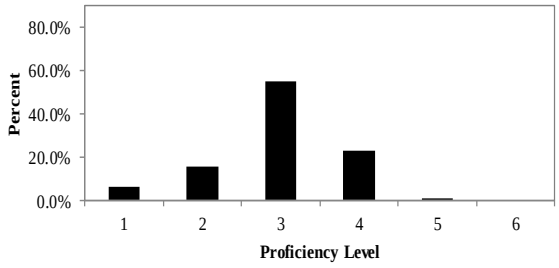

Table 3.3.4.3Eiii

Proficiency Level Distribution: Writ 6-8 S400 Online

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,504	3.68%	4,280	6.32%	6,028	8.75%	12,812	6.26%
2	9,039	13.28%	9,691	14.32%	13,487	19.58%	32,217	15.75%
3	29,575	43.46%	39,299	58.06%	44,060	63.98%	112,934	55.20%
4	26,654	39.17%	14,362	21.22%	5,292	7.68%	46,308	22.63%
5	278	0.41%	51	0.08%	1	0.00%	330	0.16%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	68,050	100.00%	67,683	100.00%	68,868	100.00%	204,601	100.00%

Table 3.3.4.3Fiii
Raw Score to Scale Score Conversion: Writ 6-8 S400 Online
n/a

Table 3.3.4.3Giii
Equating Summary: Writ 6-8 S400 Online
n/a

Figure 3.3.4.3Hiii
Test Characteristic Curve: Writ 6-8 S400 Online
n/a for S400

Figure 3.3.4.3Iiii
Test Information Function: Writ 6-8 S400 Online
n/a for S400

Table 3.3.4.3Jiii

Reliability: Writ 6-8 Weighted Reliability S400 Online

Tiers	No. of Students	Reliability	Weighted Reliability
A	42,875	0.856	0.902
B/C	161,726	0.915	

Table 3.3.4.3Kiii

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

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

Table 3.3.4.3Li

Accuracy and Consistency of Classification Indices: Writ (Grade 6) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.720	0.626		0.411	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.803		0.698	
	2	0.835		0.741	
	3	0.748		0.593	
	4	0.666		0.620	
	5	-		0.444	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.987	0.008	0.005	0.981
	2/3	0.960	0.015	0.026	0.944
	3/4	0.778	0.062	0.160	0.704
	4/5	0.996	0.004	0.000	0.996
	5/6	N/A	N/A	N/A	N/A

Table 3.3.4.3Lii

Accuracy and Consistency of Classification Indices: Writ (Grade 7) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.724	0.650		0.400	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.859		0.787	
	2	0.801		0.699	
	3	0.700		0.711	
	4	-		0.373	
	5	-		1.000	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.982	0.009	0.009	0.975
	2/3	0.954	0.018	0.028	0.935
	3/4	0.787	0.213	0.000	0.740
	4/5	0.999	0.001	0.000	1.000
	5/6	N/A	N/A	N/A	N/A

Table 3.3.4.3Liii

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

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.838	0.771		0.558	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.870		0.793	
	2	0.814		0.714	
	3	0.841		0.827	
	4	-		0.161	
	5	-		1.000	
	6	N/A		N/A	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.977	0.011	0.012	0.966
	2/3	0.939	0.021	0.041	0.914
	3/4	0.923	0.077	0.000	0.890
	4/5	1.000	0.000	0.000	1.000
	5/6	N/A	N/A	N/A	N/A

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 S400 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 3.3.4.4Ci
Raw Scores: Spek 6-8 Pre-A S400 Online

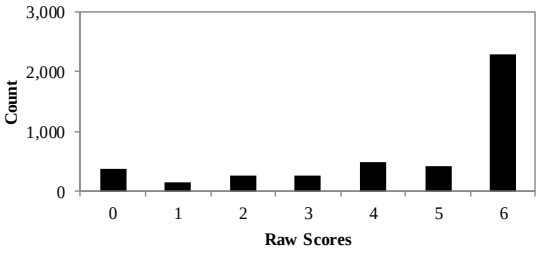

Figure 3.3.4.4Di
Scale Scores: Spek 6-8 Pre-A S400 Online

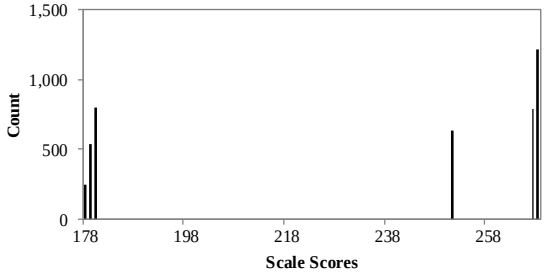

Table 3.3.4.4Ci

Raw Score Descriptive Statistics: Spek 6-8 Pre-A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	523	0	6	4.46	2.01
7	1,450	0	6	4.52	2.02
8	2,253	0	6	4.54	2.01
Total	4,226	0	6	4.53	2.01

Table 3.3.4.4Di

Scale Score Descriptive Statistics: Spek 6-8 Pre-A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	523	178	251	216.11	36.50
7	1,450	179	267	233.18	41.62
8	2,253	180	268	234.98	41.11
Total	4,226	178	268	232.03	41.18

Figure 3.3.4.4Ei
Proficiency Level: Spek 6-8 Pre-A S400 Online

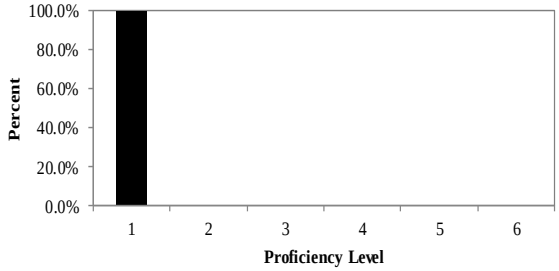

Table 3.3.4.4Ei
Proficiency Level Distribution: Spek 6-8 Pre-A S400 Online

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	523	100.00%	1,450	100.00%	2,253	100.00%	4,226	100.00%
Total	523	100.00%	1,450	100.00%	2,253	100.00%	4,226	100.00%

Table 3.3.4.4Gi
Equating Summary: Spek 6-8 Pre-A S400 Online
n/a for S400

Figure 3.3.4.4Hi
Test Characteristic Curve: Spek 6-8 Pre-A S400 Online
n/a for S400

Figure 3.3.4.4Ii
Test Information Function: Spek 6-8 Pre-A S400 Online
n/a for S400

Table 3.3.4.4Ji
Reliability: Spek 6-8 Pre-A S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Cronbach's Alpha**		**SEM**
	4,226	3	.814		0.868
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	2,292	96	4	0
	2	2,520	98	2	0
	3	2,364	98	2	0

Table 3.3.4.4Ki
Conditional Standard Error of Measurement at Cut Scores: Spek 6-8 Pre-A S400 Online
n/a

Table 3.3.4.4Li
Accuracy and Consistency of Classification Indices: Spek 6-8 Pre-A S400 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 S400 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 3.3.4.4Cii
Raw Scores: Spek 6-8 A S400 Online

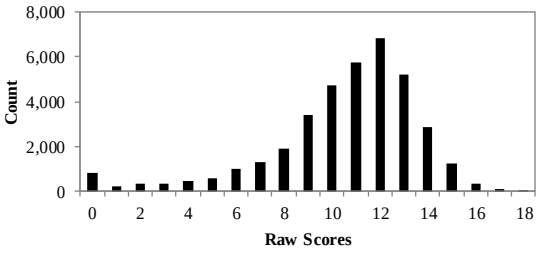

Figure 3.3.4.4Dii
Scale Scores: Spek 6-8 A S400 Online

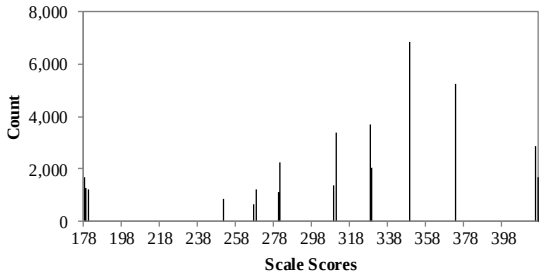

Table 3.3.4.4Cii

Raw Score Descriptive Statistics: Spek 6-8 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	9,702	0	18	9.92	3.23
7	12,706	0	18	10.47	3.11
8	14,959	0	18	10.97	3.04
Total	37,367	0	18	10.53	3.14

Table 3.3.4.4Dii

Scale Score Descriptive Statistics: Spek 6-8 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	9,702	178	416	305.06	69.25
7	12,706	179	416	320.53	63.25
8	14,959	180	416	331.41	62.37
Total	37,367	178	416	320.87	65.36

Figure 3.3.4.4Eii
Proficiency Level: Spek 6-8 A S400 Online

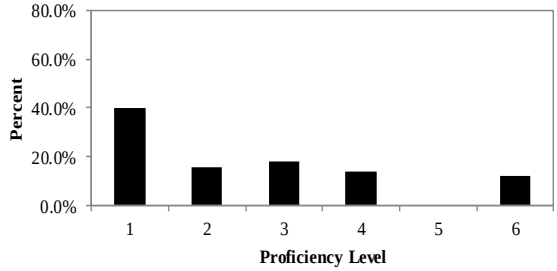

Table 3.3.4.4Eii

Proficiency Level Distribution: Spek 6-8 A S400 Online

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	4,800	49.47%	5,226	41.13%	4,995	33.39%	15,021	40.20%
2	1,504	15.50%	2,030	15.98%	2,212	14.79%	5,746	15.38%
3	1,592	16.41%	2,293	18.05%	2,934	19.61%	6,819	18.25%
4	1,042	10.74%	1,737	13.67%	2,447	16.36%	5,226	13.99%
5	0	0.00%	0	0.00%	0	0.00%	0	0.00%
6	764	7.87%	1,420	11.18%	2,371	15.85%	4,555	12.19%
Total	9,702	100.00%	12,706	100.00%	14,959	100.00%	37,367	100.00%

Table 3.3.4.4Gii
Equating Summary: Spek 6-8 A S400 Online
n/a for S400

Figure 3.3.4.4Hii
Test Characteristic Curve: Spek 6-8 A S400 Online
n/a for S400

Figure 3.3.4.4Iii
Test Information Function: Spek 6-8 A S400 Online
n/a for S400

Table 3.3.4.4Jii

Reliability: Spek 6-8 A S400 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	37,367	6	.799		1.407
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	15,742	99	1	0
	2	15,748	76	24	0
	3	16,230	99	1	0
	4	16,218	79	20	1
	5	15,696	99	1	0
	6	15,684	80	20	0

Table 3.3.4.4Kii

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

n/a

Table 3.3.4.4Lii

Accuracy and Consistency of Classification Indices: Spek 6-8 A S400 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 S400 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 3.3.4.4Ciii
Raw Scores: Spek 6-8 B/C S400 Online

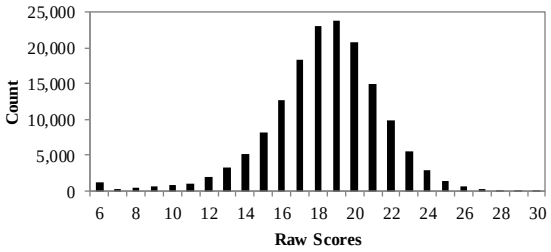

Figure 3.3.4.4Diii
Scale Scores: Spek 6-8 B/C S400 Online

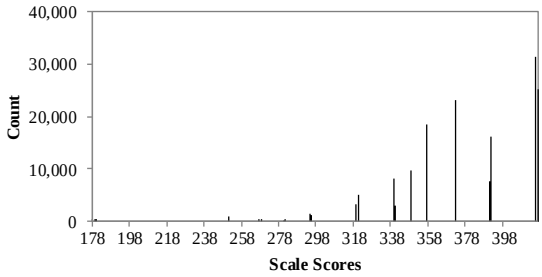

Table 3.3.4.4Ciii

Raw Score Descriptive Statistics: Spek 6-8 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	56,254	6	30	17.85	3.05
7	51,715	6	30	18.31	3.13
8	49,641	6	30	18.93	3.17
Total	157,610	6	30	18.34	3.15

Table 3.3.4.4Diii

Scale Score Descriptive Statistics: Spek 6-8 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	56,254	178	416	371.90	40.62
7	51,715	179	416	377.35	39.41
8	49,641	180	416	383.70	38.34
Total	157,610	178	416	377.41	39.81

Figure 3.3.4.4Eiii
Proficiency Level: Spek 6-8 B/C S400 Online

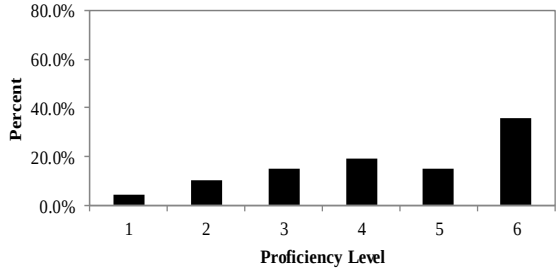

Table 3.3.4.4Eiii

Proficiency Level Distribution: Spek 6-8 B/C S400 Online

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,679	4.76%	2,192	4.24%	1,671	3.37%	6,542	4.15%
2	3,712	6.60%	5,411	10.46%	6,916	13.93%	16,039	10.18%
3	9,072	16.13%	10,252	19.82%	4,741	9.55%	24,065	15.27%
4	16,380	29.12%	7,693	14.88%	6,544	13.18%	30,617	19.43%
5	8,299	14.75%	7,961	15.39%	7,540	15.19%	23,800	15.10%
6	16,112	28.64%	18,206	35.20%	22,229	44.78%	56,547	35.88%
Total	56,254	100.00%	51,715	100.00%	49,641	100.00%	157,610	100.00%

Table 3.3.4.4Giii
Equating Summary: Spek 6-8 B/C S400 Online
n/a for S400

Figure 3.3.4.4Hiii
Test Characteristic Curve: Spek 6-8 B/C S400 Online
n/a for S400

Figure 3.3.4.4Iiii
Test Information Function: Spek 6-8 B/C S400 Online
n/a for S400

Table 3.3.4.4Jiii

Reliability: Spek 6-8 B/C S400 Online

Reliability	**No. of Students**	**No. of Items**	**Cronbach's Alpha**		**SEM**
	157,610	6	.778		1.482
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	71,252	74	26	0
	2	71,227	78	22	0
	3	69,332	71	28	0
	4	69,406	72	28	1
	5	71,644	69	31	0
	6	71,672	67	32	1

Table 3.3.4.4Kiii
Conditional Standard Error of Measurement at Cut Scores: Spek 6-8 B/C S400 Online
n/a

Table 3.3.4.4Liii
Accuracy and Consistency of Classification Indices: Spek 6-8 B/C S400 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 agencies (SEAs) may request this information; please contact us at help@wida.us.

Table 3.3.4.4Biv
DIF Analysis and Summary: Spek 6-8 S400 Online
n/a

Figure 3.3.4.4Civ
Raw Scores: Spek 6-8 S400 Online

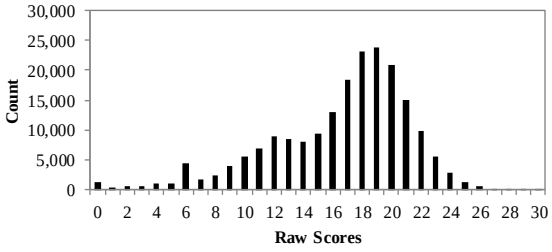

Figure 3.3.4.4Div
Scale Scores: Spek 6-8 S400 Online

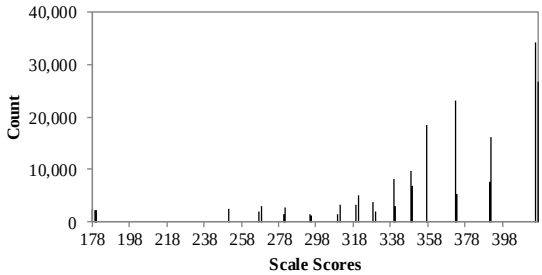

Table 3.3.4.4Civ
Raw Score Descriptive Statistics: Spek 6-8 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	66,479	0	30	16.59	4.29
7	65,871	0	30	16.49	4.73
8	66,853	0	30	16.67	5.07
Total	199,203	0	30	16.58	4.71

Table 3.3.4.4Div
Scale Score Descriptive Statistics: Spek 6-8 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	66,479	178	416	360.92	53.19
7	65,871	179	416	363.22	53.94
8	66,853	180	416	366.99	55.65
Total	199,203	178	416	363.72	54.33

Figure 3.3.4.4Eiv
Proficiency Level: Spek 6-8 S400 Online

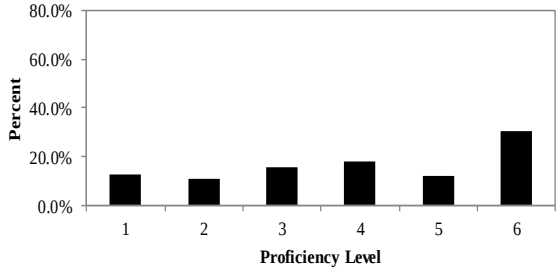

Table 3.3.4.4Eiv

Proficiency Level Distribution: Spek 6-8 S400 Online

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	8,002	12.04%	8,868	13.46%	8,919	13.34%	25,789	12.95%
2	5,216	7.85%	7,441	11.30%	9,128	13.65%	21,785	10.94%
3	10,664	16.04%	12,545	19.04%	7,675	11.48%	30,884	15.50%
4	17,422	26.21%	9,430	14.32%	8,991	13.45%	35,843	17.99%
5	8,299	12.48%	7,961	12.09%	7,540	11.28%	23,800	11.95%
6	16,876	25.39%	19,626	29.79%	24,600	36.80%	61,102	30.67%
Total	66,479	100.00%	65,871	100.00%	66,853	100.00%	199,203	100.00%

Table 3.3.4.4Fiv
Raw Score to Scale Score Conversion: Spek 6-8 S400 Online
n/a

Table 3.3.4.4Giv
Equating Summary: Spek 6-8 S400 Online
n/a

Figure 3.3.4.4Hiv
Test Characteristic Curve: Spek 6-8 S400 Online
n/a for S400

Figure 3.3.4.4Iiv
Test Information Function: Spek 6-8 S400 Online
n/a for S400

Table 3.3.4.4Jiv

Reliability: Spek 6-8 Weighted Reliability S400 Online

Tiers	No. of Students	Reliability	Weighted Reliability
Pre-A	4,226	0.814	0.783
A	37,367	0.799	
B/C	157,610	0.778	

Table 3.3.4.4Kiv

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

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

Table 3.3.4.4Li

Accuracy and Consistency of Classification Indices: Spek (Grade 6) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.291	0.316	0.161		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.792	0.662		
	2	0.283	0.182		
	3	0.357	0.231		
	4	0.348	0.302		
	5	0.172	0.158		
	6	-	0.361		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.950	0.025	0.025	0.924
	2/3	0.902	0.043	0.055	0.861
	3/4	0.818	0.034	0.147	0.764
	4/5	0.687	0.042	0.271	0.630
	5/6	0.746	0.254	0.000	0.679

Table 3.3.4.4Lii

Accuracy and Consistency of Classification Indices: Spek (Grade 7) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.290	0.325	0.181		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.743	0.619		
	2	0.352	0.248		
	3	0.376	0.265		
	4	0.208	0.169		
	5	0.177	0.160		
	6	-	0.457		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.934	0.036	0.030	0.904
	2/3	0.879	0.047	0.075	0.839
	3/4	0.798	0.031	0.172	0.743
	4/5	0.760	0.049	0.191	0.676
	5/6	0.702	0.298	0.000	0.681

Table 3.3.4.4Liii

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

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.269	0.342	0.188		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.732	0.616		
	2	0.420	0.290		
	3	0.250	0.164		
	4	0.212	0.159		
	5	0.155	0.144		
	6	-	0.545		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.932	0.038	0.031	0.903
	2/3	0.872	0.036	0.092	0.832
	3/4	0.833	0.030	0.137	0.773
	4/5	0.784	0.051	0.164	0.687
	5/6	0.632	0.368	0.000	0.656

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 agencies (SEAs) may request this information; please contact us at help@wida.us.

Figure 3.3.5.1C
Raw Scores: List 9-12 S400 Online
n/a

Table 3.3.5.1C
Raw Score Descriptive Statistics: List 9-12 S400 Online
n/a

Figure 3.3.5.1D
Scale Scores: List 9-12 S400 Online

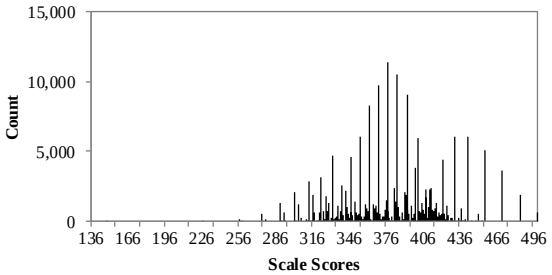

Figure 3.3.5.1E
Proficiency Level: List 9-12 S400 Online

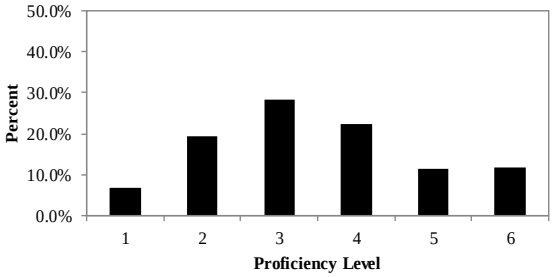

Table 3.3.5.1D
Scale Score Descriptive Statistics: List 9-12 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	78,210	136	499	380.75	42.59
10	50,098	227	499	379.85	40.65
11	35,474	227	499	384.50	40.17
12	26,224	148	499	387.27	40.23
Total	190,006	136	499	382.11	41.40

Table 3.3.5.1E
Proficiency Level Distribution: List 9-12 S400 Online

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,955	5.06%	3,252	6.49%	2,805	7.91%	3,027	11.54%	13,039	6.86%
2	14,726	18.83%	11,013	21.98%	6,742	19.01%	4,310	16.44%	36,791	19.36%
3	20,356	26.03%	15,973	31.88%	10,413	29.35%	6,903	26.32%	53,645	28.23%
4	18,033	23.06%	9,655	19.27%	8,423	23.74%	6,634	25.30%	42,745	22.50%
5	10,333	13.21%	5,648	11.27%	3,434	9.68%	2,313	8.82%	21,728	11.44%
6	10,807	13.82%	4,557	9.10%	3,657	10.31%	3,037	11.58%	22,058	11.61%
Total	78,210	100.00%	50,098	100.00%	35,474	100.00%	26,224	100.00%	190,006	100.00%

Table 3.3.5.1F
Raw Score to Scale Score Conversion: List 9-12 S400 Online
n/a

Table 3.3.5.1G
Equating Summary: List 9-12 S400 Online
n/a

Figure 3.3.5.1H
Test Characteristic Curve: List 9-12 S400 Online
n/a

Figure 3.3.5.1I
Test Information Function: List 9-12 S400 Online

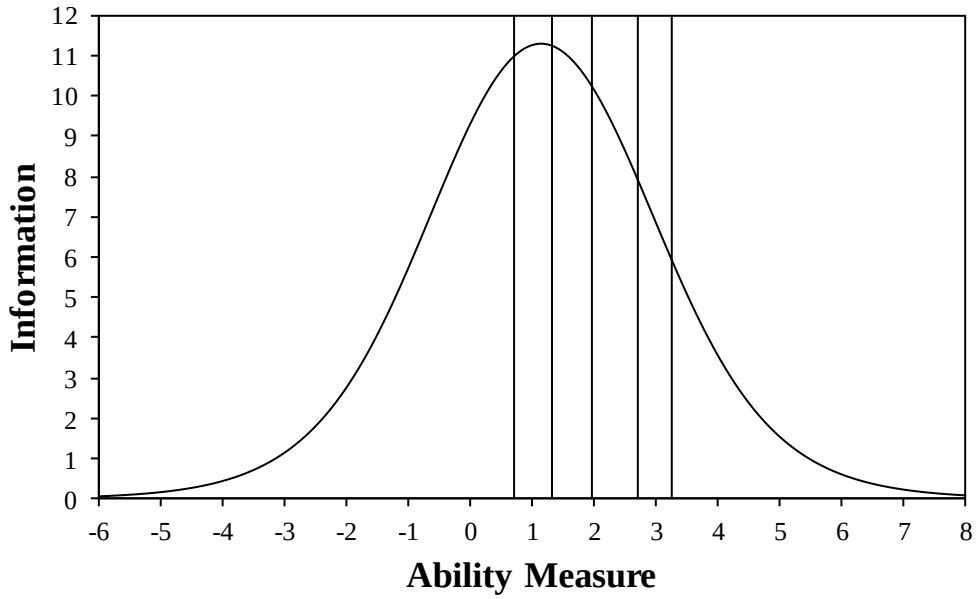

Table 3.3.5.1J
Reliability: List 9-12 S400 Online

No. of Students	No. of Items	Rasch Reliability Estimate
190,006	54	.79

Table 3.3.5.1K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: List 9-12 S400 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	9	312	13	17.96	18.03	18.03	0.02
	10	322	99	16.91	20.29	19.75	1.18
	11	332	17	16.76	17.02	16.77	0.06
	12	343	830	16.64	19.35	19.18	0.11
2/3	9	352	70	16.61	16.68	16.62	0.03
	10	358	90	16.61	16.64	16.61	0.01
	11	363	1,128	16.61	16.94	16.63	0.08
	12	366	324	16.76	17.17	17.04	0.18
3/4	9	381	130	16.87	17.51	17.02	0.14
	10	386	66	17.25	17.55	17.54	0.05
	11	389	67	17.21	18.26	17.76	0.44
	12	391	297	16.91	18.15	17.57	0.20
4/5	9	406	622	17.77	18.75	18.34	0.35
	10	412	604	17.81	19.76	19.34	0.66
	11	416	69	18.75	18.75	18.75	0.00
	12	418	48	20.03	20.03	20.03	0.00
5/6	9	432	3,026	19.76	21.34	19.80	0.22
	10	436	N/A	N/A	N/A	N/A	N/A
	11	438	26	21.34	21.38	21.36	0.02
	12	439	N/A	N/A	N/A	N/A	N/A

Table 3.3.5.1Li

Accuracy and Consistency of Classification Indices: List (Grade 9) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.536	0.425		0.288	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.630		0.348	
	2	0.547		0.435	
	3	0.506		0.413	
	4	0.493		0.387	
	5	0.418		0.308	
	6	0.836		0.673	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.955	0.007	0.038	0.934
	2/3	0.872	0.062	0.065	0.821
	3/4	0.847	0.083	0.070	0.792
	4/5	0.892	0.058	0.050	0.843
	5/6	0.933	0.050	0.017	0.906

Table 3.3.5.1Lii

Accuracy and Consistency of Classification Indices: List (Grade 10) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.539	0.423		0.274	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.609		0.344	
	2	0.534		0.434	
	3	0.574		0.476	
	4	0.451		0.346	
	5	0.451		0.323	
	6	0.796		0.611	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.941	0.011	0.047	0.914
	2/3	0.851	0.082	0.067	0.795
	3/4	0.853	0.072	0.075	0.797
	4/5	0.910	0.054	0.036	0.869
	5/6	0.951	0.035	0.014	0.929

Table 3.3.5.1Liii

Accuracy and Consistency of Classification Indices: List (Grade 11) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.521	0.409	0.264		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.603	0.358		
	2	0.467	0.373		
	3	0.531	0.436		
	4	0.519	0.408		
	5	0.381	0.267		
	6	0.814	0.633		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.929	0.016	0.055	0.897
	2/3	0.852	0.086	0.062	0.795
	3/4	0.843	0.085	0.072	0.787
	4/5	0.910	0.049	0.041	0.866
	5/6	0.948	0.037	0.015	0.924

Table 3.3.5.1Liv

Accuracy and Consistency of Classification Indices: List (Grade 12) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.509	0.404	0.267		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.583	0.391		
	2	0.380	0.304		
	3	0.484	0.392		
	4	0.565	0.444		
	5	0.364	0.254		
	6	0.843	0.685		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.899	0.035	0.067	0.857
	2/3	0.846	0.090	0.064	0.788
	3/4	0.846	0.091	0.062	0.793
	4/5	0.915	0.044	0.041	0.873
	5/6	0.949	0.036	0.015	0.926

3.3.5.2 Reading 9-12

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 3.3.5.2C
Raw Scores: Read 9-12 S400 Online
n/a

Table 3.3.5.2C
Raw Score Descriptive Statistics: Read 9-12 S400 Online
n/a

Figure 3.3.5.2D
Scale Scores: Read 9-12 S400 Online

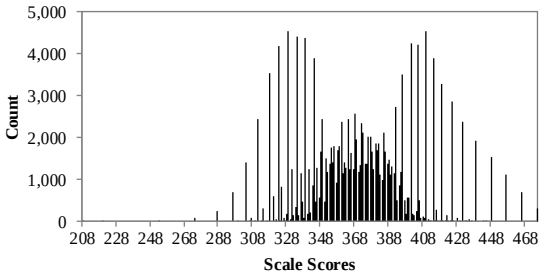

Figure 3.3.5.2E
Proficiency Level: Read 9-12 S400 Online

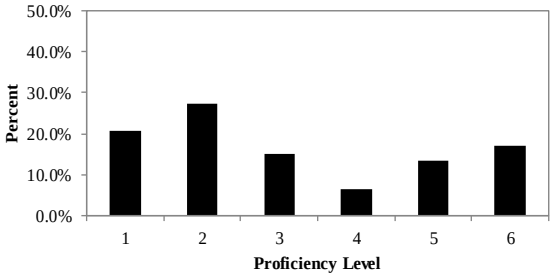

Table 3.3.5.2D
Scale Score Descriptive Statistics: Read 9-12 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	65,310	208	475	367.73	34.82
10	40,516	253	475	370.14	33.35
11	27,381	274	475	376.65	33.35
12	20,784	253	475	378.15	33.19
Total	153,991	208	475	371.36	34.21

Table 3.3.5.2E
Proficiency Level Distribution: Read 9-12 S400 Online

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	13,884	21.26%	8,525	21.04%	5,167	18.87%	4,265	20.52%	31,841	20.68%
2	15,802	24.20%	12,329	30.43%	8,173	29.85%	5,964	28.70%	42,268	27.45%
3	12,033	18.42%	5,756	14.21%	3,073	11.22%	2,202	10.59%	23,064	14.98%
4	3,372	5.16%	2,585	6.38%	2,216	8.09%	1,883	9.06%	10,056	6.53%
5	9,107	13.94%	5,465	13.49%	3,674	13.42%	2,538	12.21%	20,784	13.50%
6	11,112	17.01%	5,856	14.45%	5,078	18.55%	3,932	18.92%	25,978	16.87%
Total	65,310	100.00%	40,516	100.00%	27,381	100.00%	20,784	100.00%	153,991	100.00%

Table 3.3.5.2F
Raw Score to Scale Score Conversion: Read 9-12 S400 Online
n/a

Table 3.3.5.2G
Equating Summary: Read 9-12 S400 Online
n/a

Figure 3.3.5.2H
Test Characteristic Curve: Read 9-12 S400 Online
n/a

Figure 3.3.5.2I
Test Information Function: Read 9-12 S400 Online

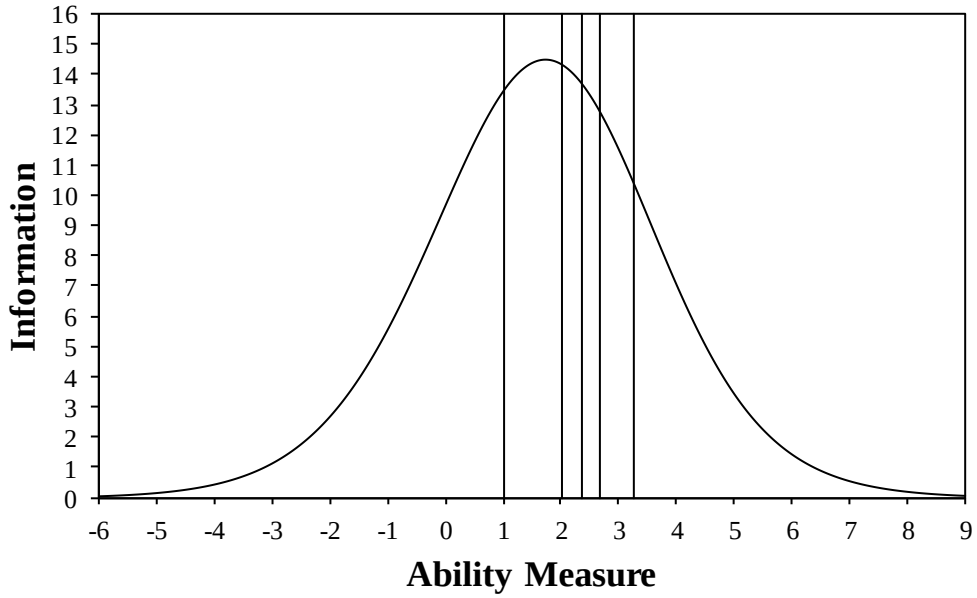

Table 3.3.5.2J
Reliability: Read 9-12 S400 Online

No. of Students	No. of Items	Rasch Reliability Estimate
153,991	72	.89

Table 3.3.5.2K

Descriptive Statistics of Conditional Standard Error of Measurement at Cut Scores: Read 9-12 S400 Online

Proficiency Level	Grade	Cut Score	No. of Students	Min.	Max.	Mean	Std. Dev.
1/2	9	336	583	11.44	11.96	11.45	0.05
	10	341	353	11.34	12.01	11.40	0.15
	11	346	208	11.28	11.47	11.29	0.04
	12	350	51	11.26	11.62	11.41	0.12
2/3	9	364	1,214	10.11	10.45	10.12	0.03
	10	370	200	10.17	10.71	10.35	0.15
	11	374	158	10.22	10.87	10.29	0.15
	12	376	257	10.22	10.82	10.30	0.11
3/4	9	381	327	10.17	11.08	10.36	0.10
	10	383	243	10.22	11.13	10.35	0.15
	11	384	160	10.24	11.08	10.36	0.16
	12	385	355	10.22	10.74	10.55	0.08
4/5	9	387	673	10.32	11.05	10.47	0.13
	10	390	454	10.37	10.92	10.52	0.13
	11	392	381	10.32	11.26	10.49	0.18
	12	393	87	10.40	11.31	11.00	0.25
5/6	9	402	69	10.74	11.78	11.19	0.24
	10	406	128	10.97	12.38	11.25	0.38
	11	407	11	11.44	11.78	11.66	0.17
	12	408	24	11.05	12.40	11.41	0.32

Table 3.3.5.2Li

Accuracy and Consistency of Classification Indices: Read (Grade 9) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.634	0.539		0.431	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.834		0.738	
	2	0.610		0.504	
	3	0.518		0.405	
	4	0.178		0.131	
	5	0.512		0.394	
	6	0.831		0.730	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.916	0.032	0.052	0.883
	2/3	0.901	0.056	0.043	0.860
	3/4	0.906	0.052	0.042	0.869
	4/5	0.911	0.054	0.035	0.877
	5/6	0.938	0.035	0.028	0.911

Table 3.3.5.2Lii

Accuracy and Consistency of Classification Indices: Read (Grade 10) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.636	0.541		0.425	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.801		0.699	
	2	0.673		0.574	
	3	0.423		0.322	
	4	0.226		0.166	
	5	0.530		0.408	
	6	0.827		0.720	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.910	0.040	0.051	0.873
	2/3	0.898	0.056	0.046	0.857
	3/4	0.908	0.051	0.041	0.871
	4/5	0.916	0.049	0.035	0.883
	5/6	0.946	0.030	0.024	0.922

Table 3.3.5.2Liii

Accuracy and Consistency of Classification Indices: Read (Grade 11) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.625	0.534	0.420		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.789	0.678		
	2	0.673	0.573		
	3	0.331	0.248		
	4	0.264	0.195		
	5	0.481	0.367		
	6	0.840	0.743		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.915	0.038	0.047	0.880
	2/3	0.894	0.056	0.049	0.852
	3/4	0.901	0.054	0.046	0.861
	4/5	0.909	0.051	0.040	0.874
	5/6	0.935	0.036	0.028	0.908

Table 3.3.5.2Liv

Accuracy and Consistency of Classification Indices: Read (Grade 12) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.622	0.531	0.417		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.772	0.668		
	2	0.643	0.540		
	3	0.313	0.235		
	4	0.298	0.221		
	5	0.461	0.348		
	6	0.859	0.767		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.905	0.046	0.049	0.866
	2/3	0.893	0.060	0.047	0.851
	3/4	0.899	0.056	0.045	0.861
	4/5	0.913	0.046	0.041	0.878
	5/6	0.941	0.034	0.025	0.914

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 agencies (SEAs) may request this information; please contact us at help@wida.us.

Figure 3.3.5.3Ci
Raw Scores: Writ 9-12 A S400 Online

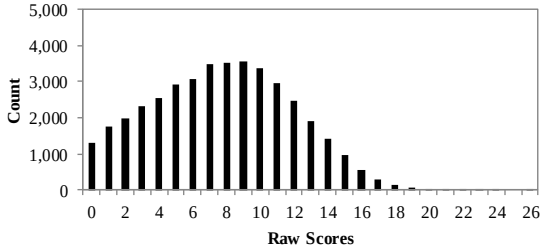

Figure 3.3.5.3Di
Scale Scores: Writ 9-12 A S400 Online

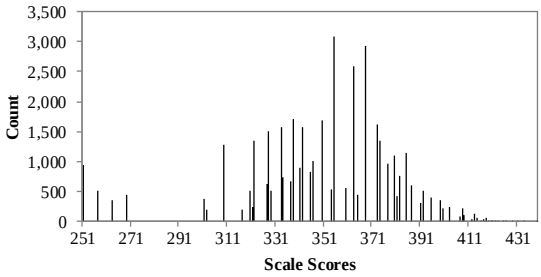

Table 3.3.5.3Ci

Raw Score Descriptive Statistics: Writ 9-12 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	20,472	0	24	6.90	4.20
10	10,412	0	22	7.95	3.77
11	5,486	0	21	8.73	3.82
12	4,293	0	26	9.82	3.93
Total	40,663	0	26	7.72	4.14

Table 3.3.5.3Di

Scale Score Descriptive Statistics: Writ 9-12 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	20,472	251	425	346.03	30.99
10	10,412	257	423	350.81	31.53
11	5,486	263	423	353.83	33.97
12	4,293	269	439	357.22	38.02
Total	40,663	251	439	349.49	32.58

Figure 3.3.5.3Ei
Proficiency Level: Writ 9-12 A S400 Online

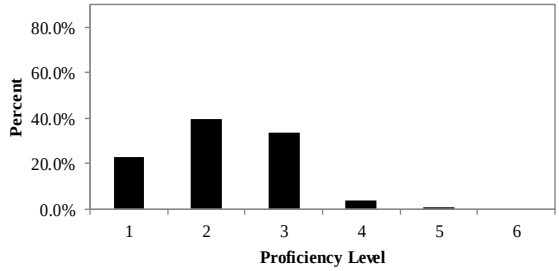

Table 3.3.5.3Ei

Proficiency Level Distribution: Writ 9-12 A S400 Online

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,573	17.45%	2,762	26.53%	1,480	26.98%	1,449	33.75%	9,264	22.78%
2	9,660	47.19%	2,952	28.35%	2,194	39.99%	1,299	30.26%	16,105	39.61%
3	6,398	31.25%	4,269	41.00%	1,630	29.71%	1,416	32.98%	13,713	33.72%
4	799	3.90%	427	4.10%	182	3.32%	127	2.96%	1,535	3.77%
5	42	0.21%	2	0.02%	0	0.00%	2	0.05%	46	0.11%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
Total	20,472	100.00%	10,412	100.00%	5,486	100.00%	4,293	100.00%	40,663	100.00%

Table 3.3.5.3Gi
Equating Summary: Writ 9-12 A S400 Online
n/a for S400

Figure 3.3.5.3Hi
Test Characteristic Curve: Writ 9-12 A S400 Online
n/a for S400

Figure 3.3.5.3Ii
Test Information Function: Writ 9-12 A S400 Online
n/a for S400

Table 3.3.5.3Ji
Reliability: Writ 9-12 A S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Response Modes**		**Cronbach's Alpha**	**SEM**
	40,663	3	Hand-written (HW)	Keyboarded (KB)	.835	1.681
Interrater Reliability	**Task**	**Mode of Response**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	HW	1,382	96	3	1
		KB	15,508	90	8	1
	2	HW	1,054	95	5	0
		KB	16,981	93	6	1
	3	HW	1,484	98	2	0
		KB	15,042	93	7	1

Table 3.3.5.3Ki
Conditional Standard Error of Measurement at Cut Scores: Writ 9-12 A S400 Online
n/a

Table 3.3.5.3Li
Accuracy and Consistency of Classification Indices: Writ 9-12 A S400 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 agencies (SEAs) may request this information; please contact us at help@wida.us.

Figure 3.3.5.3Cii
Raw Scores: Writ 9-12 B/C S400 Online

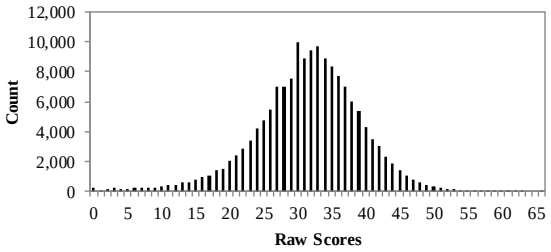

Figure 3.3.5.3Dii
Scale Scores: Writ 9-12 B/C S400 Online

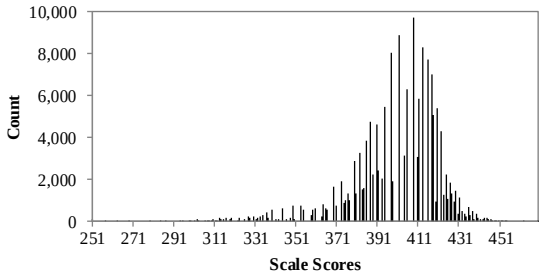

Table 3.3.5.3Cii

Raw Score Descriptive Statistics: Writ 9-12 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	61,546	0	62	31.35	7.59
10	41,997	0	59	30.78	7.75
11	31,410	0	63	31.80	7.64
12	22,887	0	66	32.58	7.54
Total	157,840	0	66	31.47	7.65

Table 3.3.5.3Dii

Scale Score Descriptive Statistics: Writ 9-12 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	61,546	251	454	399.29	23.32
10	41,997	257	453	397.56	23.72
11	31,410	263	462	400.87	22.41
12	22,887	269	469	403.06	21.78
Total	157,840	251	469	399.69	23.10

Figure 3.3.5.3Eii
Proficiency Level: Writ 9-12 B/C S400 Online

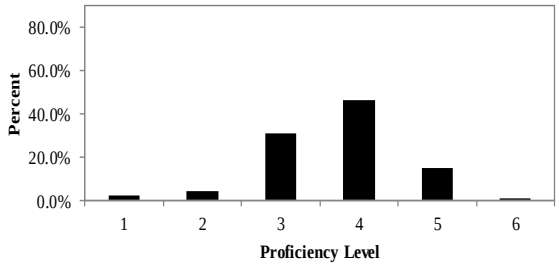

Table 3.3.5.3Eii

Proficiency Level Distribution: Writ 9-12 B/C S400 Online

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	885	1.44%	1,093	2.60%	876	2.79%	740	3.23%	3,594	2.28%
2	2,392	3.89%	2,019	4.81%	1,270	4.04%	1,157	5.06%	6,838	4.33%
3	13,023	21.16%	13,434	31.99%	12,428	39.57%	10,459	45.70%	49,344	31.26%
4	27,667	44.95%	20,747	49.40%	14,727	46.89%	9,729	42.51%	72,870	46.17%
5	16,624	27.01%	4,548	10.83%	2,063	6.57%	800	3.50%	24,035	15.23%
6	955	1.55%	156	0.37%	46	0.15%	2	0.01%	1,159	0.73%
Total	61,546	100.00%	41,997	100.00%	31,410	100.00%	22,887	100.00%	157,840	100.00%

Table 3.3.5.3Gii
Equating Summary: Writ 9-12 B/C S400 Online
n/a for S400

Figure 3.3.5.3Hii
Test Characteristic Curve: Writ 9-12 B/C S400 Online
n/a for S400

Figure 3.3.5.3Iii
Test Information Function: Writ 9-12 B/C S400 Online
n/a for S400

Table 3.3.5.3Jii
Reliability: Writ 9-12 B/C S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Response Modes**		**Cronbach's Alpha**	**SEM**
	157,840	3	Hand-written (HW)	Keyboarded (KB)	.914	2.248
Interrater Reliability	**Task**	**Mode of Response**	**No. in Sample**	**% AG**	**% AD**	**% NA**
	1	HW	952	93	7	0
		KB	70,866	93	7	0
	2	HW	984	93	7	0
		KB	72,686	90	10	1
	3	HW	988	91	9	0
		KB	73,606	90	9	1

Table 3.3.5.3Kii
Conditional Standard Error of Measurement at Cut Scores: Writ 9-12 B/C S400 Online
n/a

Table 3.3.5.3Lii
Accuracy and Consistency of Classification Indices: Writ 9-12 B/C S400 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 S400 Online
n/a

Table 3.3.5.3Biii
DIF Analysis and Summary: Writ 9-12 S400 Online
n/a

Figure 3.3.5.3Ciii
Raw Scores: Writ 9-12 S400 Online

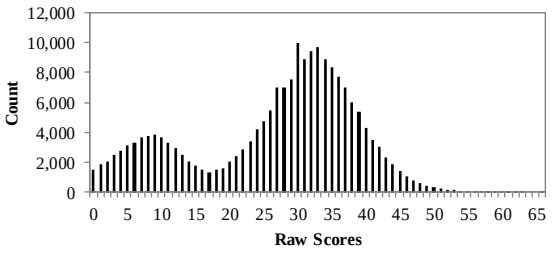

Figure 3.3.5.3Diii
Scale Scores: Writ 9-12 S400 Online

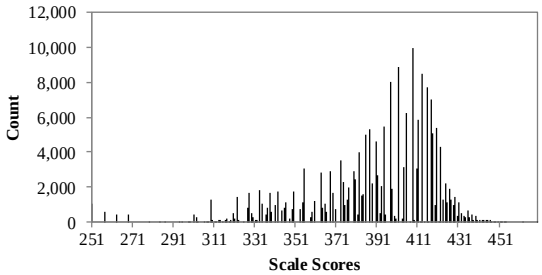

Table 3.3.5.3Ciii
Raw Score Descriptive Statistics: Writ 9-12 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	82,018	0	62	25.25	12.63
10	52,409	0	59	26.24	11.57
11	36,896	0	63	28.37	10.92
12	27,180	0	66	28.99	10.92
Total	198,503	0	66	26.60	11.91

Table 3.3.5.3Diii
Scale Score Descriptive Statistics: Writ 9-12 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	82,018	251	454	386.00	34.34
10	52,409	257	453	388.27	31.56
11	36,896	263	462	393.88	29.65
12	27,180	269	469	395.82	30.12
Total	198,503	251	469	389.41	32.44

Figure 3.3.5.3Eiii
Proficiency Level: Writ 9-12 S400 Online

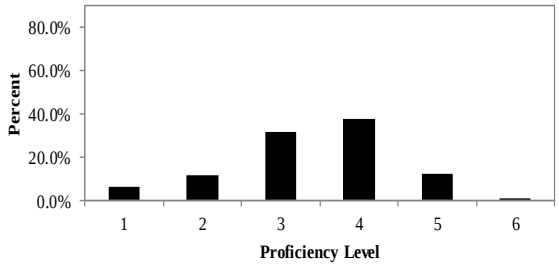

Table 3.3.5.3Eiii

Proficiency Level Distribution: Writ 9-12 S400 Online

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	4,458	5.44%	3,855	7.36%	2,356	6.39%	2,189	8.05%	12,858	6.48%
2	12,052	14.69%	4,971	9.49%	3,464	9.39%	2,456	9.04%	22,943	11.56%
3	19,421	23.68%	17,703	33.78%	14,058	38.10%	11,875	43.69%	63,057	31.77%
4	28,466	34.71%	21,174	40.40%	14,909	40.41%	9,856	36.26%	74,405	37.48%
5	16,666	20.32%	4,550	8.68%	2,063	5.59%	802	2.95%	24,081	12.13%
6	955	1.16%	156	0.30%	46	0.12%	2	0.01%	1,159	0.58%
Total	82,018	100.00%	52,409	100.00%	36,896	100.00%	27,180	100.00%	198,503	100.00%

Table 3.3.5.3Fiii
Raw Score to Scale Score Conversion: Writ 9-12 S400 Online
n/a

Table 3.3.5.3Giii
Equating Summary: Writ 9-12 S400 Online
n/a

Figure 3.3.5.3Hiii
Test Characteristic Curve: Writ 9-12 S400 Online
n/a for S400

Figure 3.3.5.3Iiii
Test Information Function: Writ 9-12 S400 Online
n/a for S400

Table 3.3.5.3Jiii

Reliability: Writ 9-12 Weighted Reliability S400 Online

Tiers	No. of Students	Reliability	Weighted Reliability
A	40,663	0.835	0.898
B/C	157,840	0.914	

Table 3.3.5.3Kiii

Conditional Standard Error of Measurement at Cut Scores: Writ 9-12 S400 Online

Proficiency Level	Grade	Cut Score	SEM	
			Tier A	Tier B/C
1/2	9	327	9.02	6.53
	10	336	10.26	7.15
	11	344	11.19	8.09
	12	352	11.82	8.40
2/3	9	356	12.13	8.40
	10	363	12.13	8.40
	11	370	11.82	8.40
	12	377	11.51	8.09
3/4	9	389	11.19	8.09
	10	397	11.19	7.77
	11	404	10.88	7.77
	12	410	10.57	7.46
4/5	9	415	10.26	7.15
	10	422	9.95	6.84
	11	428	9.33	6.84
	12	434	9.33	6.53
5/6	9	435	9.33	6.53
	10	441	9.02	6.53
	11	447	9.02	6.53
	12	452	9.33	6.53

Table 3.3.5.3Li

Accuracy and Consistency of Classification Indices: Writ (Grade 9) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.616	0.548		0.396	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.762		0.638	
	2	0.765		0.649	
	3	0.696		0.574	
	4	0.518		0.528	
	5	-		0.459	
	6	-		0.039	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.976	0.014	0.010	0.964
	2/3	0.942	0.020	0.038	0.921
	3/4	0.911	0.032	0.057	0.872
	4/5	0.764	0.234	0.003	0.782
	5/6	0.988	0.012	0.000	0.985

Table 3.3.5.3Lii

Accuracy and Consistency of Classification Indices: Writ (Grade 10) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.719	0.617		0.446	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.870		0.783	
	2	0.639		0.505	
	3	0.782		0.655	
	4	0.678		0.640	
	5	-		0.222	
	6	-		0.375	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.978	0.009	0.013	0.969
	2/3	0.953	0.022	0.024	0.933
	3/4	0.876	0.040	0.083	0.824
	4/5	0.910	0.090	0.000	0.878
	5/6	0.997	0.003	0.000	0.997

Table 3.3.5.3Liii

Accuracy and Consistency of Classification Indices: Writ (Grade 11) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.713	0.616	0.421		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.857	0.775		
	2	0.696	0.564		
	3	0.786	0.616		
	4	0.660	0.626		
	5	-	0.120		
	6	-	1.000		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.982	0.009	0.009	0.974
	2/3	0.959	0.018	0.023	0.941
	3/4	0.830	0.040	0.130	0.758
	4/5	0.943	0.057	0.000	0.928
	5/6	0.999	0.001	0.000	0.999

Table 3.3.5.3Liv

Accuracy and Consistency of Classification Indices: Writ (Grade 12) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.658	0.572	0.346		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.883	0.814		
	2	0.700	0.566		
	3	0.724	0.565		
	4	0.577	0.537		
	5	-	0.053		
	6	-	-		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.981	0.009	0.009	0.973
	2/3	0.961	0.015	0.023	0.945
	3/4	0.744	0.064	0.193	0.674
	4/5	0.970	0.030	0.000	0.965
	5/6	1.000	0.000	0.000	1.000

3.3.5.4 Speaking 9-12

3.3.5.4i Speaking 9-12 Pre-A

Table 3.3.5.4Ai
Complete Task Analysis and Summary: Spek 9-12 Pre-A S400 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 3.3.5.4Ci
Raw Scores: Spek 9-12 Pre-A S400 Online

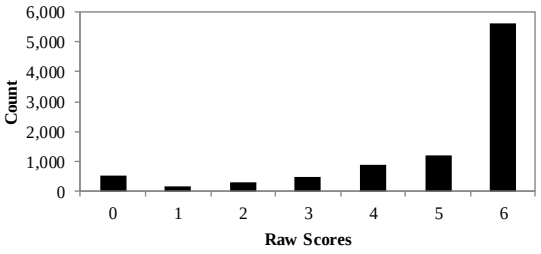

Figure 3.3.5.4Di
Scale Scores: Spek 9-12 Pre-A S400 Online

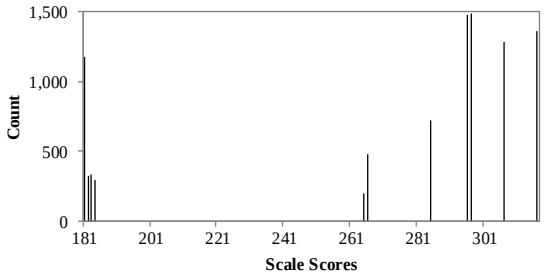

Table 3.3.5.4Ci

Raw Score Descriptive Statistics: Spek 9-12 Pre-A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	3,127	0	6	4.52	1.88
10	2,340	0	6	5.07	1.60
11	1,827	0	6	5.25	1.49
12	1,833	0	6	5.32	1.48
Total	9,127	0	6	4.97	1.69

Table 3.3.5.4Di

Scale Score Descriptive Statistics: Spek 9-12 Pre-A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	3,127	181	296	248.21	53.14
10	2,340	182	297	276.52	39.20
11	1,827	183	307	281.73	47.31
12	1,833	184	317	292.58	48.31
Total	9,127	181	317	271.09	50.79

Figure 3.3.5.4Ei
Proficiency Level: Spek 9-12 Pre-A S400 Online

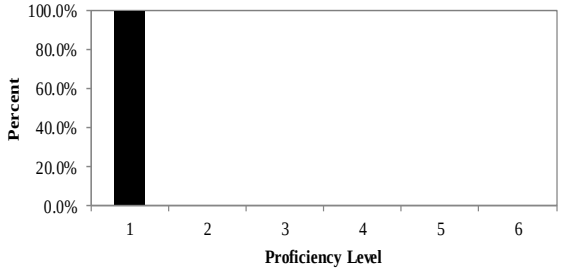

Table 3.3.5.4Ei
Proficiency Level Distribution: Spek 9-12 Pre-A S400 Online

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,127	100.00%	2,340	100.00%	1,827	100.00%	1,833	100.00%	9,127	100.00%
Total	3,127	100.00%	2,340	100.00%	1,827	100.00%	1,833	100.00%	9,127	100.00%

Table 3.3.5.4Gi
Equating Summary: Spek 9-12 Pre-A S400 Online
n/a for S400

Figure 3.3.5.4Hi
Test Characteristic Curve: Spek 9-12 Pre-A S400 Online
n/a for S400

Figure 3.3.5.4Ii
Test Information Function: Spek 9-12 Pre-A S400 Online
n/a for S400

Table 3.3.5.4Ji

Reliability: Spek 9-12 Pre-A S400 Online

Reliability	**No. of Students**	**No. of Tasks**	**Cronbach's Alpha**		**SEM**
	9,127	3	.803		0.751
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	4,426	97	3	0
	2	4,680	96	4	0
	3	5,024	96	4	0

Table 3.3.5.4Ki
Conditional Standard Error of Measurement at Cut Scores: Spek 9-12 Pre-A S400 Online
n/a

Table 3.3.5.4Li
Accuracy and Consistency of Classification Indices: Spek 9-12 Pre-A S400 Online
n/a

3.3.5.4ii Speaking 9-12 A

Table 3.3.5.4Aii
Complete Task Analysis and Summary: Spek 9-12 A S400 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 3.3.5.4Cii
Raw Scores: Spek 9-12 A S400 Online

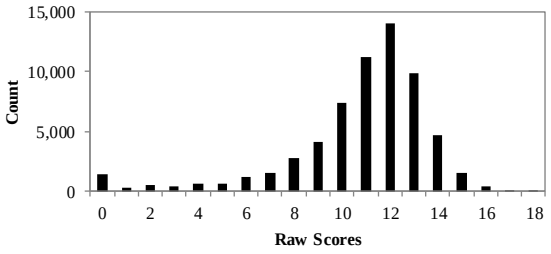

Figure 3.3.5.4Dii
Scale Scores: Spek 9-12 A S400 Online

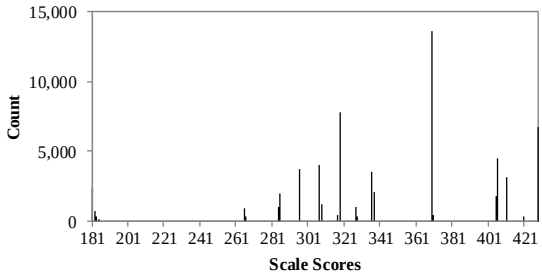

Table 3.3.5.4Cii

Raw Score Descriptive Statistics: Spek 9-12 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	32,196	0	18	10.49	3.13
10	18,899	0	18	10.99	2.79
11	9,689	0	18	11.29	2.70
12	1,983	0	17	11.03	2.80
Total	62,767	0	18	10.78	2.97

Table 3.3.5.4Dii

Scale Score Descriptive Statistics: Spek 9-12 A S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	32,196	181	428	334.44	64.48
10	18,899	182	428	352.12	56.09
11	9,689	183	428	359.22	52.69
12	1,983	184	428	355.87	59.85
Total	62,767	181	428	344.27	61.06

Figure 3.3.5.4Eii
Proficiency Level: Spek 9-12 A S400 Online

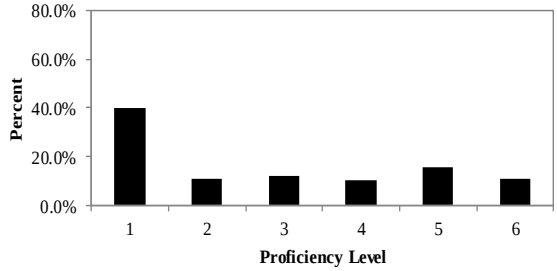

Table 3.3.5.4Eii

Proficiency Level Distribution: Spek 9-12 A S400 Online

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	17,892	55.57%	5,678	30.04%	1,419	14.65%	243	12.25%	25,232	40.20%
2	0	0.00%	3,502	18.53%	2,712	27.99%	704	35.50%	6,918	11.02%
3	0	0.00%	4,497	23.79%	2,479	25.59%	475	23.95%	7,451	11.87%
4	6,586	20.46%	0	0.00%	0	0.00%	0	0.00%	6,586	10.49%
5	4,537	14.09%	3,194	16.90%	1,786	18.43%	334	16.84%	9,851	15.69%
6	3,181	9.88%	2,028	10.73%	1,293	13.35%	227	11.45%	6,729	10.72%
Total	32,196	100.00%	18,899	100.00%	9,689	100.00%	1,983	100.00%	62,767	100.00%

Table 3.3.5.4Gii
Equating Summary: Spek 9-12 A S400 Online
n/a for S400

Figure 3.3.5.4Hii
Test Characteristic Curve: Spek 9-12 A S400 Online
n/a for S400

Figure 3.3.5.4Iii
Test Information Function: Spek 9-12 A S400 Online
n/a for S400

Table 3.3.5.4Jii

Reliability: Spek 9-12 A S400 Online

Reliability	No. of Students	No. of Tasks	Cronbach's Alpha		SEM
	62,767	6	.800		1.329
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	27,202	98	2	0
	2	27,198	77	22	0
	3	26,516	98	2	0
	4	26,435	72	27	1
	5	26,890	98	2	0
	6	26,905	77	22	1

Table 3.3.5.4Kii

Conditional Standard Error of Measurement at Cut Scores: Spek 9-12 A S400 Online

n/a

Table 3.3.5.4Lii

Accuracy and Consistency of Classification Indices: Spek 9-12 A S400 Online

n/a

3.3.5.4iii Speaking 9-12 B/C

Table 3.3.5.4Aii
Complete Task Analysis and Summary: Spek 9-12 B/C S400 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 3.3.5.4Ciii
Raw Scores: Spek 9-12 B/C S400 Online

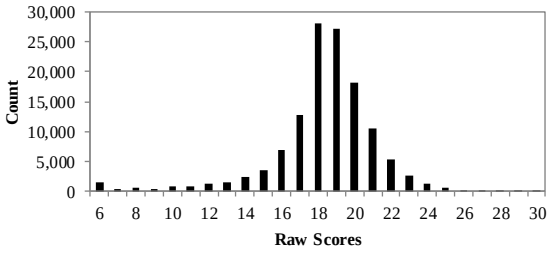

Figure 3.3.5.4Diii
Scale Scores: Spek 9-12 B/C S400 Online

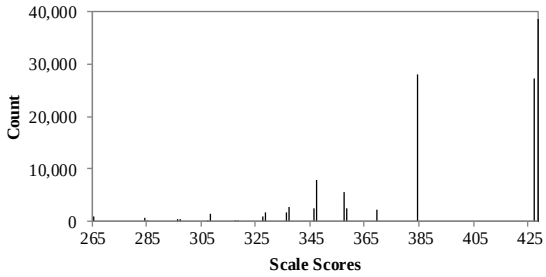

Table 3.3.5.4Ciii

Raw Score Descriptive Statistics: Spek 9-12 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	47,004	6	30	18.34	2.87
10	31,204	6	29	18.25	2.89
11	25,274	6	30	18.41	2.86
12	23,238	6	30	18.11	3.04
Total	126,720	6	30	18.29	2.91

Table 3.3.5.4Diii

Scale Score Descriptive Statistics: Spek 9-12 B/C S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	47,004	265	428	393.21	41.67
10	31,204	284	428	395.46	37.27
11	25,274	284	428	398.42	35.62
12	23,238	265	428	395.96	36.42
Total	126,720	265	428	395.31	38.55

Figure 3.3.5.4Eiii
Proficiency Level: Spek 9-12 B/C S400 Online

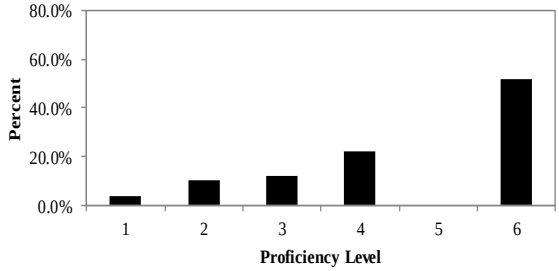

Table 3.3.5.4Eiii

Proficiency Level Distribution: Spek 9-12 B/C S400 Online

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,549	5.42%	913	2.93%	682	2.70%	527	2.27%	4,671	3.69%
2	4,604	9.79%	4,136	13.25%	1,767	6.99%	2,395	10.31%	12,902	10.18%
3	4,809	10.23%	3,135	10.05%	3,630	14.36%	3,678	15.83%	15,252	12.04%
4	10,597	22.54%	7,049	22.59%	5,501	21.77%	4,989	21.47%	28,136	22.20%
5	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
6	24,445	52.01%	15,971	51.18%	13,694	54.18%	11,649	50.13%	65,759	51.89%
Total	47,004	100.00%	31,204	100.00%	25,274	100.00%	23,238	100.00%	126,720	100.00%

Table 3.3.5.4Giii
Equating Summary: Spek 9-12 B/C S400 Online
n/a for S400

Figure 3.3.5.4Hiii
Test Characteristic Curve: Spek 9-12 B/C S400 Online
n/a for S400

Figure 3.3.5.4Iiii
Test Information Function: Spek 9-12 B/C S400 Online
n/a for S400

Table 3.3.5.4Jiii
Reliability: Spek 9-12 B/C S400 Online

Reliability	**No. of Students**	**No. of Items**	**Cronbach's Alpha**		**SEM**
	126,720	6	.791		1.329
Interrater Reliability	**Task**	**No. in Sample**	**% EX**	**% AD**	**% NA**
	1	55,542	80	19	0
	2	55,534	76	24	0
	3	55,932	82	17	1
	4	55,972	78	21	1
	5	56,138	80	20	0
	6	56,060	76	24	0

Table 3.3.5.4Kiii
Conditional Standard Error of Measurement at Cut Scores: Spek 9-12 B/C S400 Online
n/a

Table 3.3.5.4Liii
Accuracy and Consistency of Classification Indices: Spek 9-12 B/C S400 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 agencies (SEAs) may request this information; please contact us at help@wida.us.

Table 3.3.5.4Biv
DIF Analysis and Summary: Spek 9-12 S400 Online
n/a

Figure 3.3.5.4Civ
Raw Scores: Spek 9-12 S400 Online

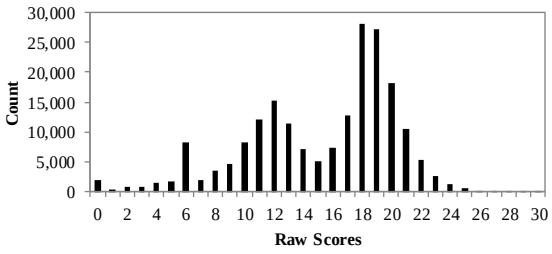

Figure 3.3.5.4Div
Scale Scores: Spek 9-12 S400 Online

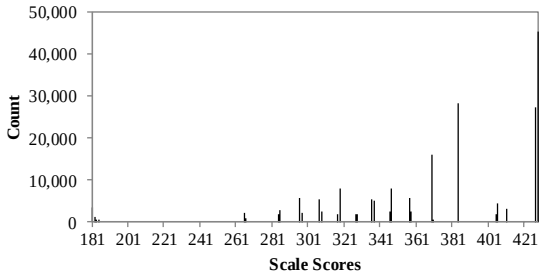

Table 3.3.5.4Civ
Raw Score Descriptive Statistics: Spek 9-12 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	82,327	0	30	14.75	5.20
10	52,443	0	29	15.04	4.94
11	36,790	0	30	15.88	4.82
12	27,054	0	30	16.73	4.63
Total	198,614	0	30	15.31	5.04

Table 3.3.5.4Div
Scale Score Descriptive Statistics: Spek 9-12 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	82,327	181	428	364.72	63.73
10	52,443	182	428	374.53	53.84
11	36,790	183	428	382.30	50.35
12	27,054	184	428	386.02	47.99
Total	198,614	181	428	373.47	57.45

Figure 3.3.5.4Eiv
Proficiency Level: Spek 9-12 S400 Online

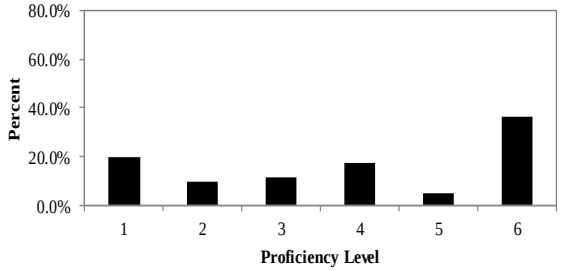

Table 3.3.5.4Eiv

Proficiency Level Distribution: Spek 9-12 S400 Online

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	23,568	28.63%	8,931	17.03%	3,928	10.68%	2,603	9.62%	39,030	19.65%
2	4,604	5.59%	7,638	14.56%	4,479	12.17%	3,099	11.45%	19,820	9.98%
3	4,809	5.84%	7,632	14.55%	6,109	16.61%	4,153	15.35%	22,703	11.43%
4	17,183	20.87%	7,049	13.44%	5,501	14.95%	4,989	18.44%	34,722	17.48%
5	4,537	5.51%	3,194	6.09%	1,786	4.85%	334	1.23%	9,851	4.96%
6	27,626	33.56%	17,999	34.32%	14,987	40.74%	11,876	43.90%	72,488	36.50%
Total	82,327	100.00%	52,443	100.00%	36,790	100.00%	27,054	100.00%	198,614	100.00%

Table 3.3.5.4Fiv
Raw Score to Scale Score Conversion: Spek 9-12 S400 Online
n/a

Table 3.3.5.4Giv
Equating Summary: Spek 9-12 S400 Online
n/a

Figure 3.3.5.4Hiv
Test Characteristic Curve: Spek 9-12 S400 Online
n/a for S400

Figure 3.3.5.4Iiv
Test Information Function: Spek 9-12 S400 Online
n/a for S400

Table 3.3.5.4Jiv

Reliability: Spek 9-12 Weighted Reliability S400 Online

Tiers	No. of Students	Reliability	Weighted Reliability
Pre-A	9,127	0.803	0.794
A	62,767	0.800	
B/C	126,720	0.791	

Table 3.3.5.4Kiv

Conditional Standard Error of Measurement at Cut Scores: Spek 9-12 S400 Online

Proficiency Level	Grade	Cut Score	SEM
1/2	9	319	20.92
	10	321	20.92
	11	322	20.92
	12	323	21.43
2/3	9	347	22.45
	10	351	22.96
	11	354	23.47
	12	357	23.47
3/4	9	366	24.49
	10	371	24.49
	11	377	25.00
	12	384	26.02
4/5	9	388	26.53
	10	393	27.04
	11	399	28.06
	12	405	29.08
5/6	9	407	29.59
	10	412	30.61
	11	416	31.12
	12	421	32.14

Table 3.3.5.4Li

Accuracy and Consistency of Classification Indices: Spek (Grade 9) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.551	0.481		0.334	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.885		0.791	
	2	0.162		0.116	
	3	0.155		0.112	
	4	0.458		0.340	
	5	0.105		0.083	
	6	0.797		0.663	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.903	0.028	0.069	0.863
	2/3	0.895	0.054	0.052	0.843
	3/4	0.874	0.077	0.049	0.825
	4/5	0.841	0.053	0.106	0.782
	5/6	0.836	0.106	0.059	0.765

Table 3.3.5.4Lii

Accuracy and Consistency of Classification Indices: Spek (Grade 10) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.499	0.431		0.295	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.745		0.610	
	2	0.374		0.294	
	3	0.319		0.252	
	4	0.285		0.214	
	5	0.131		0.095	
	6	0.873		0.726	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.908	0.041	0.051	0.870
	2/3	0.859	0.059	0.082	0.815
	3/4	0.847	0.052	0.101	0.795
	4/5	0.868	0.053	0.079	0.799
	5/6	0.849	0.119	0.033	0.789

Table 3.3.5.4Liii

Accuracy and Consistency of Classification Indices: Spek (Grade 11) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.490	0.416	0.268		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.710	0.554		
	2	0.399	0.305		
	3	0.401	0.315		
	4	0.316	0.229		
	5	0.090	0.068		
	6	0.878	0.746		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.937	0.031	0.032	0.907
	2/3	0.885	0.052	0.063	0.848
	3/4	0.855	0.047	0.098	0.808
	4/5	0.863	0.049	0.088	0.789
	5/6	0.808	0.157	0.035	0.744

Table 3.3.5.4Liv

Accuracy and Consistency of Classification Indices: Spek (Grade 12) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.394	0.396	0.247		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.728	0.570		
	2	0.420	0.317		
	3	0.406	0.312		
	4	0.381	0.263		
	5	0.020	0.017		
	6	0.904	0.768		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.945	0.024	0.031	0.919
	2/3	0.898	0.043	0.059	0.863
	3/4	0.874	0.048	0.078	0.826
	4/5	0.852	0.049	0.099	0.766
	5/6	0.702	0.281	0.017	0.710

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 S400 Online

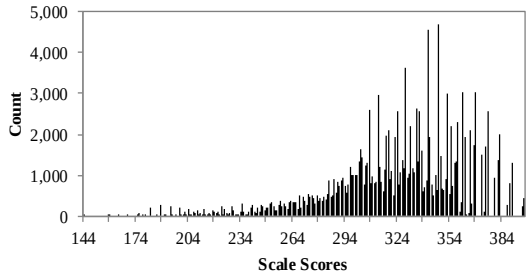

Figure 3.4.1.1B
Proficiency Level: Oral 1 S400 Online

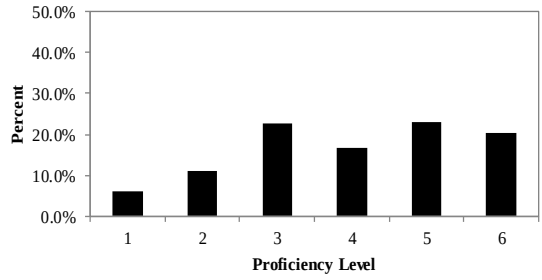

Table 3.4.1.1A

Scale Score Descriptive Statistics: Oral 1 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	144,287	144	397	325.04	38.63
Total	144,287	144	397	325.04	38.63

Table 3.4.1.1B

Proficiency Level Distribution: Oral 1 S400 Online

	Grade 1		Total	
Level	**Count**	**Percent**	**Count**	**Percent**
1	8,610	5.97%	8,610	5.97%
2	15,960	11.06%	15,960	11.06%
3	32,837	22.76%	32,837	22.76%
4	24,053	16.67%	24,053	16.67%
5	33,274	23.06%	33,274	23.06%
6	29,553	20.48%	29,553	20.48%
Total	144,287	100.00%	144,287	100.00%

Table 3.4.1.1C

Reliability: Oral 1 S400 Online

Component	Weight	Variance	Reliability
Listening	0.50	1834.451	0.760
Speaking	0.50	2337.055	0.788
Oral		1496.179	0.844

*Variances from students who had results in all four domains

Table 3.4.1.1D

Accuracy and Consistency of Classification Indices: Oral (Grade 1) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.553	0.444		0.315	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.808		0.646	
	2	0.542		0.404	
	3	0.587		0.471	
	4	0.378		0.290	
	5	0.482		0.385	
	6	0.727		0.581	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.971	0.010	0.019	0.958
	2/3	0.926	0.036	0.038	0.893
	3/4	0.873	0.053	0.074	0.827
	4/5	0.866	0.059	0.075	0.811
	5/6	0.875	0.076	0.048	0.828

3.4.1.2 Literacy Language Composite 1

Figure 3.4.1.2A
Scale Scores: Litr 1 S400 Online

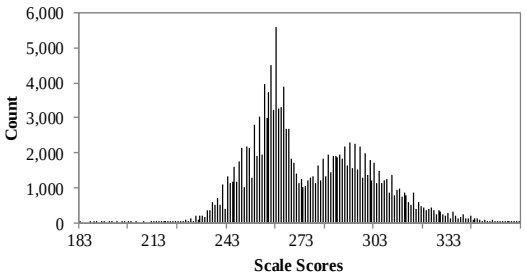

Figure 3.4.1.2B
Proficiency Level: Litr 1 S400 Online

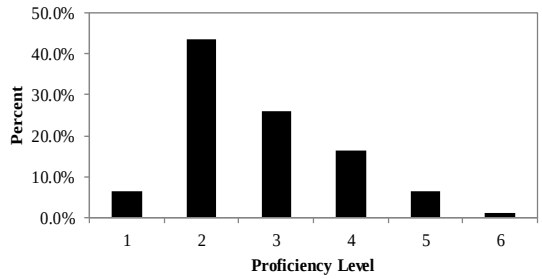

Table 3.4.1.2A

Scale Score Descriptive Statistics: Litr 1 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	152,182	183	362	277.02	24.18
Total	152,182	183	362	277.02	24.18

Table 3.4.1.2B

Proficiency Level Distribution: Litr 1 S400 Online

	Grade 1		Total	
Level	**Count**	**Percent**	**Count**	**Percent**
1	9,556	6.28%	9,556	6.28%
2	66,417	43.64%	66,417	43.64%
3	39,717	26.10%	39,717	26.10%
4	25,045	16.46%	25,045	16.46%
5	9,635	6.33%	9,635	6.33%
6	1,812	1.19%	1,812	1.19%
Total	152,182	100.00%	152,182	100.00%

Table 3.4.1.2C

Reliability: Litr 1 S400 Online

Component	Weight	Variance	Reliability
Reading	0.50	756.922	0.840
Writing	0.50	645.334	0.890
Literacy		586.432	0.918

*Variances from students who had results in all four domains

Table 3.4.1.2D

Accuracy and Consistency of Classification Indices: Litr (Grade 1) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.768	0.672	0.540		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.546	0.357		
	2	0.832	0.785		
	3	0.706	0.606		
	4	0.778	0.678		
	5	0.746	0.626		
	6	0.847	0.699		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.942	0.023	0.035	0.915
	2/3	0.913	0.038	0.049	0.876
	3/4	0.945	0.036	0.019	0.924
	4/5	0.974	0.014	0.012	0.962
	5/6	0.994	0.004	0.001	0.992

3.4.1.3 Comprehension Language Composite 1

Figure 3.4.1.3A
Scale Scores: Cphn 1 S400 Online

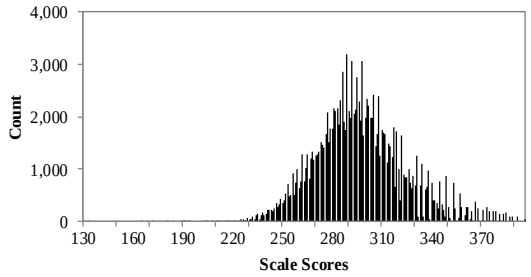

Figure 3.4.1.3B
Proficiency Level: Cphn 1 S400 Online

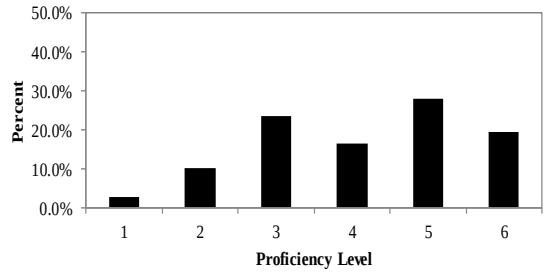

Table 3.4.1.3A

Scale Score Descriptive Statistics: Cphn 1 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	140,229	130	396	296.90	27.31
Total	140,229	130	396	296.90	27.31

Table 3.4.1.3B

Proficiency Level Distribution: Cphn 1 S400 Online

	Grade 1		Total	
Level	**Count**	**Percent**	**Count**	**Percent**
1	3,597	2.57%	3,597	2.57%
2	14,382	10.26%	14,382	10.26%
3	33,031	23.56%	33,031	23.56%
4	23,185	16.53%	23,185	16.53%
5	39,056	27.85%	39,056	27.85%
6	26,978	19.24%	26,978	19.24%
Total	140,229	100.00%	140,229	100.00%

Table 3.4.1.3C

Reliability: Cphn 1 S400 Online

Component	Weight	Variance	Reliability
Listening	0.30	1834.451	0.760
Reading	0.70	756.922	0.840
Comprehension		748.075	0.868

*Variances from students who had results in all four domains

Table 3.4.1.3D

Accuracy and Consistency of Classification Indices: Cphn (Grade 1) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.606	0.503	0.372		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	-	0.214		
	2	0.487	0.372		
	3	0.552	0.452		
	4	0.397	0.306		
	5	0.670	0.557		
	6	0.838	0.736		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.974	0.000	0.026	0.966
	2/3	0.913	0.036	0.051	0.875
	3/4	0.873	0.075	0.051	0.825
	4/5	0.877	0.068	0.054	0.832
	5/6	0.934	0.036	0.030	0.904

3.4.1.4 Overall Language Composite 1

Figure 3.4.1.4A
Scale Scores: Over 1 S400 Online

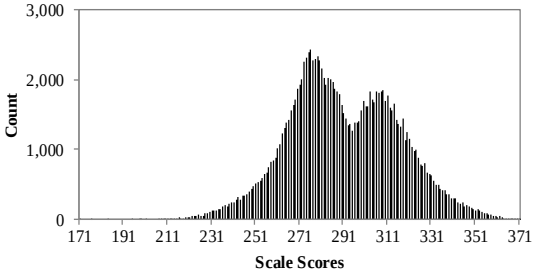

Figure 3.4.1.4B
Proficiency Level: Over 1 S400 Online

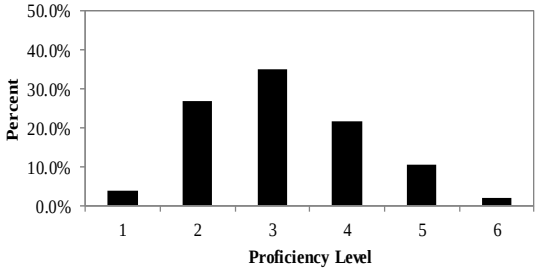

Table 3.4.1.4A

Scale Score Descriptive Statistics: Over 1 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
1	131,864	171	371	291.06	25.08
Total	131,864	171	371	291.06	25.08

Table 3.4.1.4B

Proficiency Level Distribution: Over 1 S400 Online

	Grade 1		Total	
Level	**Count**	**Percent**	**Count**	**Percent**
1	5,123	3.89%	5,123	3.89%
2	35,285	26.76%	35,285	26.76%
3	46,061	34.93%	46,061	34.93%
4	28,754	21.81%	28,754	21.81%
5	14,060	10.66%	14,060	10.66%
6	2,581	1.96%	2,581	1.96%
Total	131,864	100.00%	131,864	100.00%

Table 3.4.1.4C

Reliability: Over 1 S400 Online

Component	Weight	Variance	Reliability
Listening	0.15	1834.451	0.760
Reading	0.35	756.922	0.840
Speaking	0.15	2337.055	0.788
Writing	0.35	645.334	0.890
Overall Composite		629.161	0.929

*Variances from students who had results in all four domains

Table 3.4.1.4D

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

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.778	0.689	0.582		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.735	0.548		
	2	0.813	0.742		
	3	0.773	0.697		
	4	0.749	0.648		
	5	0.768	0.659		
	6	0.851	0.706		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.976	0.008	0.016	0.965
	2/3	0.925	0.034	0.041	0.894
	3/4	0.927	0.043	0.030	0.898
	4/5	0.959	0.023	0.018	0.942
	5/6	0.991	0.007	0.002	0.988

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 S400 Online

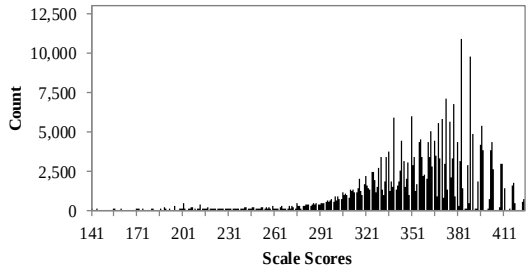

Figure 3.4.2.1B
Proficiency Level: Oral 2-3 S400 Online

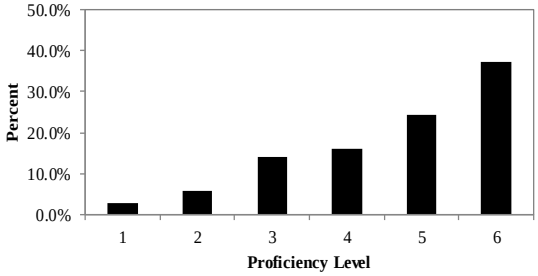

Table 3.4.2.1A

Scale Score Descriptive Statistics: Oral 2-3 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	145,251	141	418	351.79	35.73
3	143,452	144	424	358.85	37.41
Total	288,703	141	424	355.30	36.74

Table 3.4.2.1B

Proficiency Level Distribution: Oral 2-3 S400 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,713	2.56%	4,271	2.98%	7,984	2.77%
2	7,735	5.33%	8,406	5.86%	16,141	5.59%
3	20,586	14.17%	19,970	13.92%	40,556	14.05%
4	21,628	14.89%	24,738	17.24%	46,366	16.06%
5	33,727	23.22%	36,741	25.61%	70,468	24.41%
6	57,862	39.84%	49,326	34.39%	107,188	37.13%
Total	145,251	100.00%	143,452	100.00%	288,703	100.00%

Table 3.4.2.1C

Reliability: Oral 2-3 S400 Online

Component	Weight	Variance	Reliability
Listening	0.50	1626.759	0.730
Speaking	0.50	2102.180	0.760
Oral		1388.688	0.830

*Variances from students who had results in all four domains

Table 3.4.2.1Di

Accuracy and Consistency of Classification Indices: Oral (Grade 2) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.613	0.510		0.340	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.806		0.619	
	2	0.507		0.355	
	3	0.570		0.437	
	4	0.412		0.311	
	5	0.460		0.352	
	6	0.801		0.724	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.987	0.004	0.009	0.980
	2/3	0.959	0.020	0.021	0.939
	3/4	0.908	0.035	0.057	0.875
	4/5	0.881	0.045	0.074	0.833
	5/6	0.843	0.077	0.080	0.778

Table 3.4.2.1Dii

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

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.590	0.487	0.327		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.815	0.628		
	2	0.495	0.344		
	3	0.522	0.393		
	4	0.432	0.334		
	5	0.487	0.384		
	6	0.773	0.681		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.985	0.004	0.011	0.978
	2/3	0.955	0.022	0.023	0.931
	3/4	0.900	0.042	0.058	0.863
	4/5	0.868	0.053	0.080	0.817
	5/6	0.847	0.074	0.079	0.785

3.4.2.2 Literacy Language Composite 2-3

Figure 3.4.2.2A
Scale Scores: Litr 2-3 S400 Online

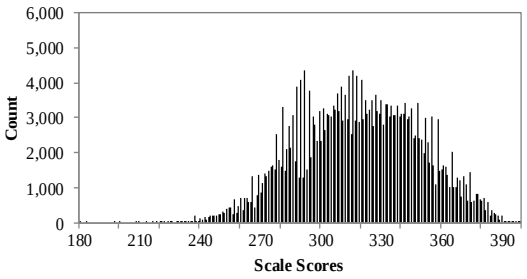

Figure 3.4.2.2B
Proficiency Level: Litr 2-3 S400 Online

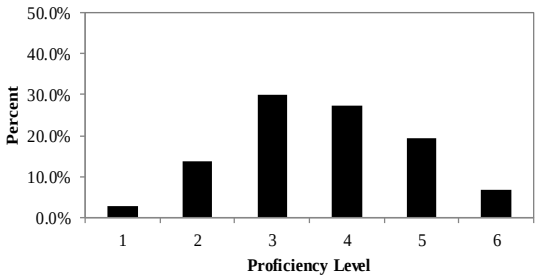

Table 3.4.2.2A

Scale Score Descriptive Statistics: Litr 2-3 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	145,816	180	373	302.57	22.08
3	143,263	209	399	335.53	27.86
Total	289,079	180	399	318.91	30.03

Table 3.4.2.2B

Proficiency Level Distribution: Litr 2-3 S400 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,557	2.44%	4,794	3.35%	8,351	2.89%
2	28,645	19.64%	10,902	7.61%	39,547	13.68%
3	62,922	43.15%	23,948	16.72%	86,870	30.05%
4	35,250	24.17%	43,584	30.42%	78,834	27.27%
5	13,034	8.94%	42,643	29.77%	55,677	19.26%
6	2,408	1.65%	17,392	12.14%	19,800	6.85%
Total	145,816	100.00%	143,263	100.00%	289,079	100.00%

Table 3.4.2.2C

Reliability: Litr 2-3 S400 Online

Component	Weight	Variance	Reliability
Reading	0.50	1038.162	0.860
Writing	0.50	1228.495	0.896
Literacy		902.637	0.924

*Variances from students who had results in all four domains

Table 3.4.2.2Di

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

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.792	0.708	0.588		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.780	0.575		
	2	0.777	0.690		
	3	0.821	0.762		
	4	0.764	0.667		
	5	0.747	0.634		
	6	0.901	0.773		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.984	0.003	0.012	0.979
	2/3	0.932	0.033	0.035	0.903
	3/4	0.920	0.044	0.036	0.888
	4/5	0.964	0.020	0.016	0.947
	5/6	0.992	0.007	0.001	0.990

Table 3.4.2.2Dii

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

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.713	0.607	0.490		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.883	0.784		
	2	0.721	0.607		
	3	0.681	0.563		
	4	0.744	0.643		
	5	0.682	0.595		
	6	0.737	0.586		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.988	0.003	0.009	0.984
	2/3	0.970	0.013	0.017	0.957
	3/4	0.927	0.040	0.033	0.897
	4/5	0.902	0.039	0.059	0.863
	5/6	0.925	0.050	0.026	0.898

3.4.2.3 Comprehension Language Composite 2-3

Figure 3.4.2.3A
Scale Scores: Cphn 2-3 S400 Online

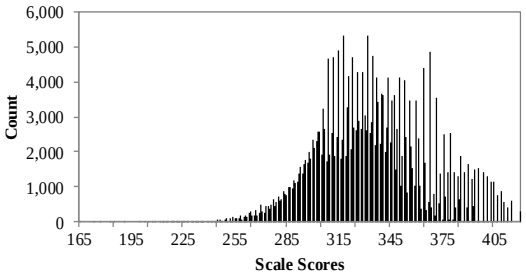

Figure 3.4.2.3B
Proficiency Level: Cphn 2-3 S400 Online

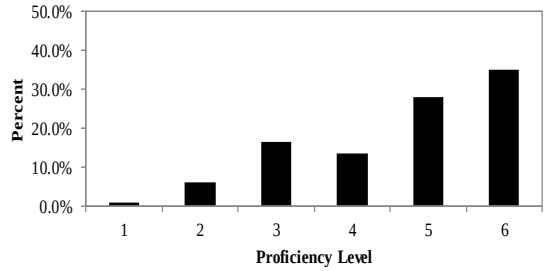

Table 3.4.2.3A

Scale Score Descriptive Statistics: Cphn 2-3 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	132,672	165	420	325.51	28.11
3	130,086	185	420	343.36	31.89
Total	262,758	165	420	334.35	31.34

Table 3.4.2.3B

Proficiency Level Distribution: Cphn 2-3 S400 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	939	0.71%	1,784	1.37%	2,723	1.04%
2	6,413	4.83%	9,050	6.96%	15,463	5.88%
3	22,210	16.74%	21,423	16.47%	43,633	16.61%
4	21,851	16.47%	13,475	10.36%	35,326	13.44%
5	39,268	29.60%	34,487	26.51%	73,755	28.07%
6	41,991	31.65%	49,867	38.33%	91,858	34.96%
Total	132,672	100.00%	130,086	100.00%	262,758	100.00%

Table 3.4.2.3C

Reliability: Cphn 2-3 S400 Online

Component	Weight	Variance	Reliability
Listening	0.30	1626.759	0.730
Reading	0.70	1038.162	0.860
Comprehension		986.002	0.888

*Variances from students who had results in all four domains

Table 3.4.2.3Di

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

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.664	0.571		0.431	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	-		0.140	
	2	0.493		0.326	
	3	0.550		0.450	
	4	0.437		0.338	
	5	0.663		0.558	
	6	0.871		0.799	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.993	0.000	0.007	0.992
	2/3	0.953	0.008	0.039	0.934
	3/4	0.898	0.057	0.045	0.856
	4/5	0.882	0.063	0.055	0.839
	5/6	0.915	0.045	0.040	0.878

Table 3.4.2.3Dii

Accuracy and Consistency of Classification Indices: Cphn (Grade 3) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.675	0.587		0.442	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.741		0.366	
	2	0.567		0.437	
	3	0.589		0.466	
	4	0.318		0.239	
	5	0.626		0.518	
	6	0.872		0.810	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.987	0.000	0.013	0.984
	2/3	0.953	0.019	0.028	0.931
	3/4	0.909	0.044	0.046	0.872
	4/5	0.894	0.056	0.050	0.853
	5/6	0.901	0.050	0.049	0.861

3.4.2.4 Overall Language Composite 2-3

Figure 3.4.2.4A
Scale Scores: Over 2-3 S400 Online

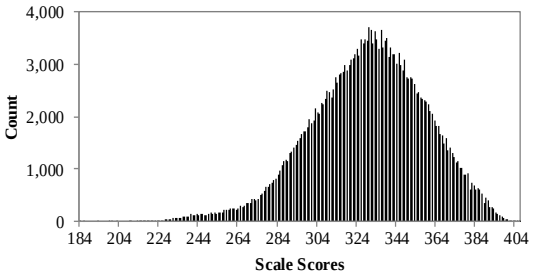

Figure 3.4.2.4B
Proficiency Level: Over 2-3 S400 Online

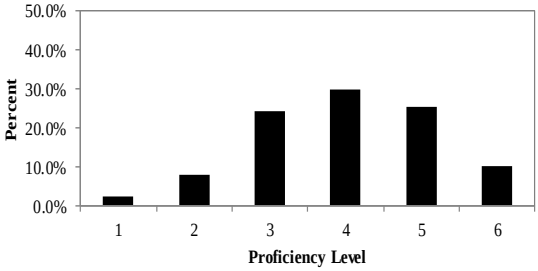

Table 3.4.2.4A

Scale Score Descriptive Statistics: Over 2-3 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
2	124,475	184	386	316.91	23.74
3	122,666	213	406	342.23	28.63
Total	247,141	184	406	329.48	29.17

Table 3.4.2.4B

Proficiency Level Distribution: Over 2-3 S400 Online

	Grade 2		Grade 3		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,526	2.03%	3,367	2.74%	5,893	2.38%
2	11,571	9.30%	8,053	6.56%	19,624	7.94%
3	42,530	34.17%	17,647	14.39%	60,177	24.35%
4	39,934	32.08%	33,489	27.30%	73,423	29.71%
5	22,881	18.38%	40,165	32.74%	63,046	25.51%
6	5,033	4.04%	19,945	16.26%	24,978	10.11%
Total	124,475	100.00%	122,666	100.00%	247,141	100.00%

Table 3.4.2.4C

Reliability: Over 2-3 S400 Online

Component	Weight	Variance	Reliability
Listening	0.15	1626.759	0.730
Reading	0.35	1038.162	0.860
Speaking	0.15	2102.180	0.760
Writing	0.35	1228.495	0.896
Overall Composite		850.925	0.936

*Variances from students who had results in all four domains

Table 3.4.2.4Di

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

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.786	0.701	0.595		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.881	0.772		
	2	0.746	0.635		
	3	0.825	0.756		
	4	0.768	0.684		
	5	0.752	0.665		
	6	0.859	0.708		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.992	0.002	0.006	0.989
	2/3	0.963	0.019	0.018	0.946
	3/4	0.916	0.041	0.043	0.882
	4/5	0.937	0.032	0.031	0.910
	5/6	0.979	0.018	0.004	0.973

Table 3.4.2.4Dii

Accuracy and Consistency of Classification Indices: Over (Grade 3) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.725	0.622	0.507		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.896	0.814		
	2	0.749	0.643		
	3	0.701	0.583		
	4	0.754	0.652		
	5	0.700	0.610		
	6	0.725	0.606		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.991	0.002	0.006	0.988
	2/3	0.977	0.011	0.013	0.966
	3/4	0.940	0.033	0.027	0.915
	4/5	0.911	0.034	0.055	0.876
	5/6	0.905	0.053	0.042	0.871

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 S400 Online

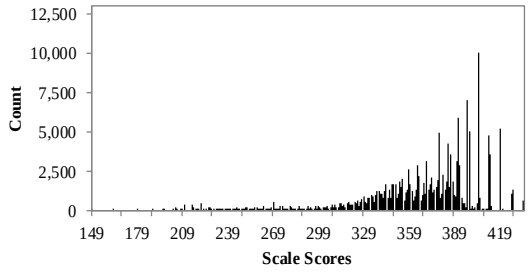

Figure 3.4.3.1B
Proficiency Level: Oral 4-5 S400 Online

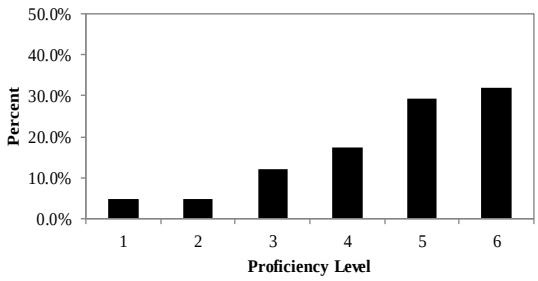

Table 3.4.3.1A

Scale Score Descriptive Statistics: Oral 4-5 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	93,483	163	435	366.21	38.38
5	71,928	149	435	371.26	40.62
Total	165,411	149	435	368.41	39.45

Table 3.4.3.1B

Proficiency Level Distribution: Oral 4-5 S400 Online

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,783	4.05%	3,811	5.30%	7,594	4.59%
2	4,267	4.56%	3,832	5.33%	8,099	4.90%
3	11,624	12.43%	8,139	11.32%	19,763	11.95%
4	16,668	17.83%	12,210	16.98%	28,878	17.46%
5	25,458	27.23%	22,736	31.61%	48,194	29.14%
6	31,683	33.89%	21,200	29.47%	52,883	31.97%
Total	93,483	100.00%	71,928	100.00%	165,411	100.00%

Table 3.4.3.1C

Reliability: Oral 4-5 S400 Online

Component	Weight	Variance	Reliability
Listening	0.50	1628.018	0.720
Speaking	0.50	2575.673	0.736
Oral		1649.175	0.828

*Variances from students who had results in all four domains

Table 3.4.3.1Di

Accuracy and Consistency of Classification Indices: Oral (Grade 4) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.560	0.452		0.277	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.866		0.726	
	2	0.438		0.300	
	3	0.531		0.395	
	4	0.461		0.345	
	5	0.449		0.359	
	6	0.689		0.599	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.984	0.004	0.012	0.977
	2/3	0.962	0.021	0.017	0.941
	3/4	0.910	0.036	0.054	0.878
	4/5	0.866	0.041	0.092	0.815
	5/6	0.799	0.087	0.114	0.728

Table 3.4.3.1Dii

Accuracy and Consistency of Classification Indices: Oral (Grade 5) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.525	0.426	0.248		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.859	0.728		
	2	0.439	0.307		
	3	0.488	0.356		
	4	0.431	0.308		
	5	0.474	0.404		
	6	0.606	0.509		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.979	0.006	0.014	0.970
	2/3	0.956	0.023	0.021	0.934
	3/4	0.913	0.034	0.053	0.881
	4/5	0.862	0.045	0.093	0.804
	5/6	0.770	0.110	0.120	0.711

3.4.3.2 Literacy Language Composite 4-5

Figure 3.4.3.2A
Scale Scores: Litr 4-5 S400 Online

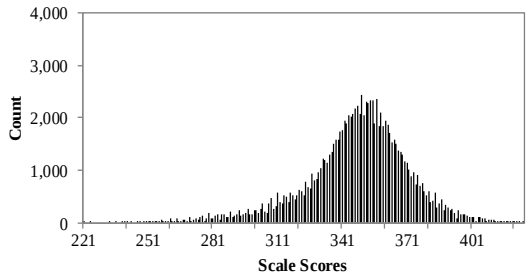

Figure 3.4.3.2B
Proficiency Level: Litr 4-5 S400 Online

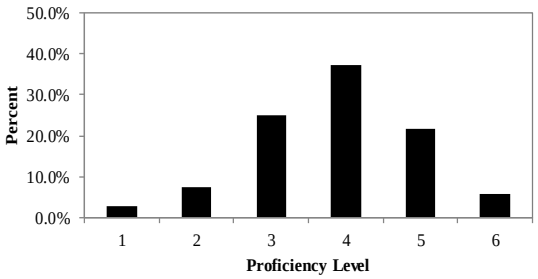

Table 3.4.3.2A

Scale Score Descriptive Statistics: Litr 4-5 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	60,075	221	423	345.54	23.18
5	46,071	236	425	350.33	25.37
Total	106,146	221	425	347.62	24.27

Table 3.4.3.2B

Proficiency Level Distribution: Litr 4-5 S400 Online

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,154	1.92%	1,695	3.68%	2,849	2.68%
2	3,518	5.86%	4,403	9.56%	7,921	7.46%
3	13,136	21.87%	13,499	29.30%	26,635	25.09%
4	23,583	39.26%	15,907	34.53%	39,490	37.20%
5	14,823	24.67%	8,325	18.07%	23,148	21.81%
6	3,861	6.43%	2,242	4.87%	6,103	5.75%
Total	60,075	100.00%	46,071	100.00%	106,146	100.00%

Table 3.4.3.2C

Reliability: Litr 4-5 S400 Online

Component	Weight	Variance	Reliability
Reading	0.50	847.164	0.850
Writing	0.50	649.891	0.911
Literacy		588.478	0.921

*Variances from students who had results in all four domains

Table 3.4.3.2Di

Accuracy and Consistency of Classification Indices: Litr (Grade 4) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.738	0.646		0.518	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.913		0.831	
	2	0.734		0.621	
	3	0.761		0.659	
	4	0.798		0.716	
	5	0.648		0.577	
	6	0.720		0.485	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.993	0.001	0.005	0.991
	2/3	0.977	0.011	0.012	0.966
	3/4	0.922	0.041	0.037	0.888
	4/5	0.902	0.037	0.061	0.862
	5/6	0.944	0.050	0.006	0.934

Table 3.4.3.2Dii

Accuracy and Consistency of Classification Indices: Litr (Grade 5) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.717	0.623		0.500	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.894		0.806	
	2	0.720		0.606	
	3	0.775		0.682	
	4	0.749		0.654	
	5	0.584		0.512	
	6	-		0.435	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.988	0.003	0.009	0.983
	2/3	0.962	0.020	0.019	0.944
	3/4	0.906	0.047	0.047	0.868
	4/5	0.910	0.034	0.056	0.871
	5/6	0.951	0.049	0.000	0.948

3.4.3.3 Comprehension Language Composite 4-5

Figure 3.4.3.4A
Scale Scores: Cphn 4-5 S400 Online

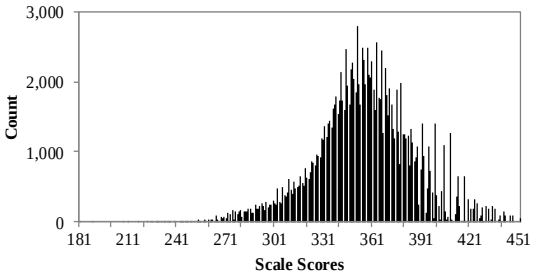

Figure 3.4.3.4B
Proficiency Level: Cphn 4-5 S400 Online

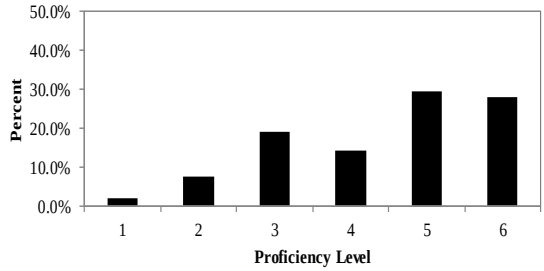

Table 3.4.3.3A

Scale Score Descriptive Statistics: Cphn 4-5 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	79,273	181	452	353.35	28.05
5	60,918	208	452	359.68	31.65
Total	140,191	181	452	356.10	29.83

Table 3.4.3.3B

Proficiency Level Distribution: Cphn 4-5 S400 Online

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	987	1.25%	1,972	3.24%	2,959	2.11%
2	4,706	5.94%	5,651	9.28%	10,357	7.39%
3	13,586	17.14%	13,150	21.59%	26,736	19.07%
4	11,872	14.98%	8,075	13.26%	19,947	14.23%
5	24,300	30.65%	16,961	27.84%	41,261	29.43%
6	23,822	30.05%	15,109	24.80%	38,931	27.77%
Total	79,273	100.00%	60,918	100.00%	140,191	100.00%

Table 3.4.3.3C

Reliability: Cphn 4-5 S400 Online

Component	Weight	Variance	Reliability
Listening	0.30	1628.018	0.720
Reading	0.70	847.164	0.850
Comprehension		867.536	0.881

*Variances from students who had results in all four domains

Table 3.4.3.3Di

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

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.663	0.566		0.429	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.868		0.622	
	2	0.628		0.486	
	3	0.612		0.487	
	4	0.410		0.315	
	5	0.644		0.545	
	6	0.837		0.753	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.990	0.001	0.009	0.989
	2/3	0.965	0.012	0.023	0.948
	3/4	0.905	0.051	0.044	0.865
	4/5	0.878	0.060	0.062	0.834
	5/6	0.901	0.050	0.049	0.859

Table 3.4.3.3Dii

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

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.636	0.536		0.411	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.853		0.678	
	2	0.617		0.480	
	3	0.617		0.500	
	4	0.343		0.263	
	5	0.614		0.512	
	6	0.817		0.718	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.982	0.003	0.015	0.977
	2/3	0.948	0.023	0.029	0.923
	3/4	0.888	0.060	0.053	0.844
	4/5	0.874	0.062	0.065	0.828
	5/6	0.907	0.048	0.045	0.867

3.4.3.4 Overall Language Composite 4-5

Figure 3.4.3.4A
Scale Scores: Over 4-5 S400 Online

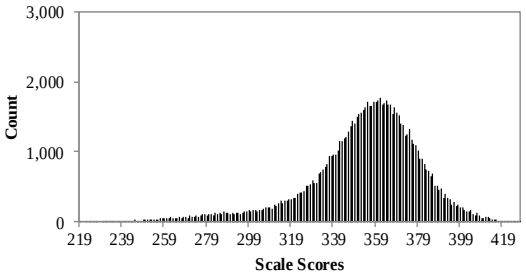

Figure 3.4.3.4B
Proficiency Level: Over 4-5 S400 Online

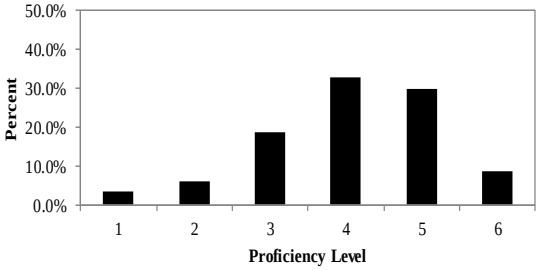

Table 3.4.3.4A

Scale Score Descriptive Statistics: Over 4-5 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
4	51,294	219	427	351.10	26.04
5	39,470	231	426	355.77	28.39
Total	90,764	219	427	353.13	27.18

Table 3.4.3.4B

Proficiency Level Distribution: Over 4-5 S400 Online

	Grade 4		Grade 5		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,305	2.54%	1,692	4.29%	2,997	3.30%
2	2,741	5.34%	2,899	7.34%	5,640	6.21%
3	9,032	17.61%	8,061	20.42%	17,093	18.83%
4	16,668	32.50%	13,232	33.52%	29,900	32.94%
5	16,461	32.09%	10,732	27.19%	27,193	29.96%
6	5,087	9.92%	2,854	7.23%	7,941	8.75%
Total	51,294	100.00%	39,470	100.00%	90,764	100.00%

Table 3.4.3.4C

Reliability: Over 4-5 S400 Online

Component	Weight	Variance	Reliability
Listening	0.15	1628.018	0.720
Reading	0.35	847.164	0.850
Speaking	0.15	2575.673	0.736
Writing	0.35	649.891	0.911
Overall Composite		738.918	0.935

*Variances from students who had results in all four domains

Table 3.4.3.4Di

Accuracy and Consistency of Classification Indices: Over (Grade 4) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.724	0.638		0.514	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.925		0.864	
	2	0.730		0.621	
	3	0.762		0.656	
	4	0.787		0.693	
	5	0.660		0.615	
	6	0.641		0.451	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.993	0.002	0.005	0.991
	2/3	0.980	0.010	0.010	0.971
	3/4	0.937	0.033	0.030	0.910
	4/5	0.906	0.033	0.062	0.869
	5/6	0.908	0.082	0.010	0.893

Table 3.4.3.4Dii

Accuracy and Consistency of Classification Indices: Over (Grade 5) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.712	0.622	0.499		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.916	0.854		
	2	0.719	0.608		
	3	0.756	0.650		
	4	0.770	0.665		
	5	0.623	0.572		
	6	-	0.328		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.989	0.003	0.008	0.985
	2/3	0.972	0.015	0.013	0.960
	3/4	0.927	0.037	0.036	0.897
	4/5	0.895	0.031	0.074	0.854
	5/6	0.928	0.072	0.000	0.917

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 S400 Online

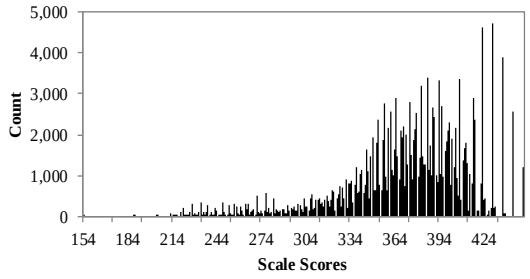

Figure 3.4.4.1B
Proficiency Level: Oral 6-8 S400 Online

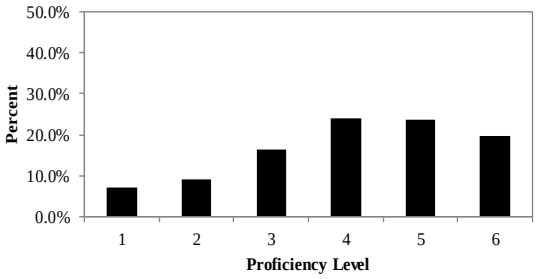

Table 3.4.4.1A

Scale Score Descriptive Statistics: Oral 6-8 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	57,670	188	451	368.12	40.21
7	57,397	154	451	372.98	43.16
8	58,458	189	451	378.07	45.28
Total	173,525	154	451	373.08	43.14

Table 3.4.4.1B

Proficiency Level Distribution: Oral 6-8 S400 Online

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	3,229	5.60%	4,182	7.29%	4,891	8.37%	12,302	7.09%
2	4,965	8.61%	5,478	9.54%	5,531	9.46%	15,974	9.21%
3	9,254	16.05%	9,787	17.05%	9,237	15.80%	28,278	16.30%
4	15,435	26.76%	12,854	22.39%	13,253	22.67%	41,542	23.94%
5	14,395	24.96%	14,033	24.45%	12,733	21.78%	41,161	23.72%
6	10,392	18.02%	11,063	19.27%	12,813	21.92%	34,268	19.75%
Total	57,670	100.00%	57,397	100.00%	58,458	100.00%	173,525	100.00%

Table 3.4.4.1C

Reliability: Oral 6-8 S400 Online

Component	Weight	Variance	Reliability
Listening	0.50	1938.414	0.790
Speaking	0.50	3117.260	0.783
Oral		1947.065	0.861

*Variances from students who had results in all four domains

Table 3.4.4.1Di

Accuracy and Consistency of Classification Indices: Oral (Grade 6) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.548	0.440	0.300		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.843	0.710		
	2	0.538	0.392		
	3	0.519	0.397		
	4	0.566	0.452		
	5	0.470	0.394		
	6	0.636	0.482		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.978	0.008	0.015	0.967
	2/3	0.942	0.031	0.027	0.914
	3/4	0.889	0.050	0.062	0.849
	4/5	0.856	0.048	0.096	0.801
	5/6	0.859	0.089	0.052	0.814

Table 3.4.4.1Dii

Accuracy and Consistency of Classification Indices: Oral (Grade 7) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)
	0.558	0.448	0.319
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.835	0.699
	2	0.503	0.368
	3	0.509	0.393
	4	0.494	0.392
	5	0.503	0.405
	6	0.701	0.556

Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.970	0.010	0.020	0.956
	2/3	0.932	0.035	0.033	0.901
	3/4	0.883	0.052	0.065	0.841
	4/5	0.864	0.056	0.079	0.811
	5/6	0.878	0.069	0.053	0.830

Table 3.4.4.1Diii

Accuracy and Consistency of Classification Indices: Oral (Grade 8) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)
	0.537	0.432	0.302
Conditional on Level	**Level**	**Accuracy**	**Consistency**
	1	0.839	0.708
	2	0.480	0.351
	3	0.485	0.372
	4	0.504	0.394
	5	0.428	0.346
	6	0.687	0.548

Indices at Cut Points		Accuracy			
	Cut Point	Accuracy	False Positives	False Negatives	Consistency
	1/2	0.967	0.012	0.021	0.951
	2/3	0.930	0.037	0.033	0.898
	3/4	0.885	0.049	0.066	0.845
	4/5	0.864	0.051	0.085	0.809
	5/6	0.852	0.089	0.059	0.801

3.4.4.2 Literacy Language Composite 6-8

Figure 3.4.4.2A
Scale Scores: Litr 6-8 S400 Online

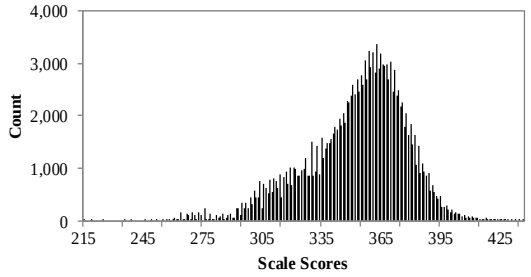

Figure 3.4.4.2B
Proficiency Level: Litr 6-8 S400 Online

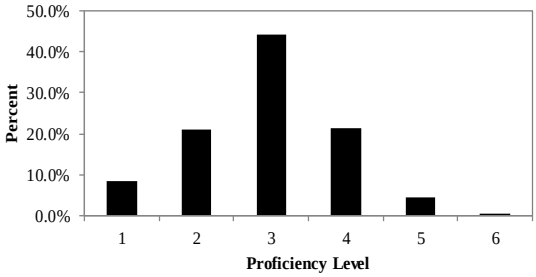

Table 3.4.4.2A

Scale Score Descriptive Statistics: Litr 6-8 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	53,744	219	431	348.24	23.82
7	52,017	215	427	353.04	24.76
8	51,503	225	437	358.01	25.72
Total	157,264	215	437	353.03	25.09

Table 3.4.4.2B

Proficiency Level Distribution: Litr 6-8 S400 Online

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,886	5.37%	4,411	8.48%	5,750	11.16%	13,047	8.30%
2	10,044	18.69%	10,699	20.57%	12,123	23.54%	32,866	20.90%
3	23,118	43.02%	23,317	44.83%	23,218	45.08%	69,653	44.29%
4	14,383	26.76%	11,025	21.19%	8,311	16.14%	33,719	21.44%
5	2,994	5.57%	2,328	4.48%	1,831	3.56%	7,153	4.55%
6	319	0.59%	237	0.46%	270	0.52%	826	0.53%
Total	53,744	100.00%	52,017	100.00%	51,503	100.00%	157,264	100.00%

Table 3.4.4.2C

Reliability: Litr 6-8 S400 Online

Component	Weight	Variance	Reliability
Reading	0.50	870.854	0.860
Writing	0.50	677.305	0.902
Literacy		633.842	0.926

*Variances from students who had results in all four domains

Table 3.4.4.2Di

Accuracy and Consistency of Classification Indices: Litr (Grade 6) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.775	0.689	0.559		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.853	0.776		
	2	0.790	0.698		
	3	0.826	0.759		
	4	0.707	0.626		
	5	0.559	0.385		
	6	-	0.924		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.982	0.007	0.011	0.976
	2/3	0.940	0.030	0.030	0.915
	3/4	0.904	0.044	0.051	0.867
	4/5	0.950	0.037	0.012	0.932
	5/6	0.994	0.006	0.000	0.995

Table 3.4.4.2Dii

Accuracy and Consistency of Classification Indices: Litr (Grade 7) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.774	0.690		0.562	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.853		0.801	
	2	0.763		0.667	
	3	0.830		0.767	
	4	0.668		0.581	
	5	0.637		0.393	
	6	-		0.993	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.972	0.012	0.016	0.965
	2/3	0.929	0.038	0.033	0.903
	3/4	0.908	0.044	0.048	0.874
	4/5	0.957	0.037	0.007	0.947
	5/6	0.995	0.005	0.000	0.998

Table 3.4.4.2Diii

Accuracy and Consistency of Classification Indices: Litr (Grade 8) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.771	0.688		0.559	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.864		0.809	
	2	0.762		0.668	
	3	0.830		0.768	
	4	0.608		0.515	
	5	0.592		0.348	
	6	-		0.992	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.966	0.015	0.019	0.956
	2/3	0.922	0.042	0.036	0.892
	3/4	0.914	0.041	0.045	0.881
	4/5	0.962	0.035	0.003	0.955
	5/6	0.995	0.005	0.000	0.997

3.4.4.3 Comprehension Language Composite 6-8

Figure 3.4.4.3A
Scale Scores: Cphn 6-8 S400 Online

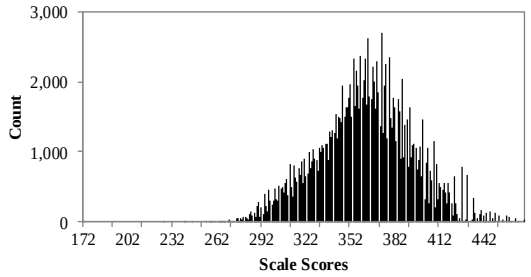

Figure 3.4.4.3B
Proficiency Level: Cphn 6-8 S400 Online

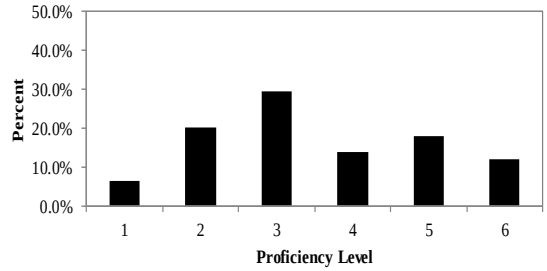

Table 3.4.4.3A

Scale Score Descriptive Statistics: Cphn 6-8 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	50,390	240	469	355.95	28.00
7	48,883	172	469	362.08	31.23
8	48,605	226	469	368.82	33.53
Total	147,878	172	469	362.21	31.41

Table 3.4.4.3B

Proficiency Level Distribution: Cphn 6-8 S400 Online

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	1,653	3.28%	3,402	6.96%	4,421	9.10%	9,476	6.41%
2	9,564	18.98%	9,718	19.88%	10,400	21.40%	29,682	20.07%
3	16,311	32.37%	14,883	30.45%	12,450	25.61%	43,644	29.51%
4	7,374	14.63%	7,014	14.35%	6,357	13.08%	20,745	14.03%
5	10,082	20.01%	8,163	16.70%	8,445	17.37%	26,690	18.05%
6	5,406	10.73%	5,703	11.67%	6,532	13.44%	17,641	11.93%
Total	50,390	100.00%	48,883	100.00%	48,605	100.00%	147,878	100.00%

Table 3.4.4.3C

Reliability: Cphn 6-8 S400 Online

Component	Weight	Variance	Reliability
Listening	0.30	1938.414	0.790
Reading	0.70	870.854	0.860
Comprehension		987.307	0.902

*Variances from students who had results in all four domains

Table 3.4.4.3Di

Accuracy and Consistency of Classification Indices: Cphn (Grade 6) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.678	0.575	0.459		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.826	0.655		
	2	0.752	0.648		
	3	0.714	0.621		
	4	0.424	0.325		
	5	0.648	0.531		
	6	0.811	0.680		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.983	0.004	0.013	0.976
	2/3	0.925	0.035	0.039	0.893
	3/4	0.893	0.056	0.050	0.852
	4/5	0.906	0.049	0.044	0.868
	5/6	0.953	0.028	0.018	0.933

Table 3.4.4.3Dii

Accuracy and Consistency of Classification Indices: Cphn (Grade 7) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.660	0.555	0.444		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.829	0.688		
	2	0.692	0.585		
	3	0.689	0.591		
	4	0.431	0.330		
	5	0.600	0.477		
	6	0.830	0.707		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.967	0.010	0.024	0.955
	2/3	0.917	0.043	0.041	0.881
	3/4	0.895	0.056	0.049	0.854
	4/5	0.913	0.046	0.041	0.876
	5/6	0.953	0.030	0.018	0.932

Table 3.4.4.3Diii

Accuracy and Consistency of Classification Indices: Cphn (Grade 8) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.645	0.541	0.437		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.841	0.716		
	2	0.693	0.587		
	3	0.632	0.525		
	4	0.391	0.298		
	5	0.585	0.465		
	6	0.823	0.701		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.961	0.012	0.027	0.946
	2/3	0.913	0.044	0.043	0.877
	3/4	0.895	0.055	0.050	0.855
	4/5	0.907	0.050	0.043	0.870
	5/6	0.945	0.033	0.022	0.921

3.4.4.4 Overall Language Composite 6-8

Figure 3.4.4.4A
Scale Scores: Over 6-8 S400 Online

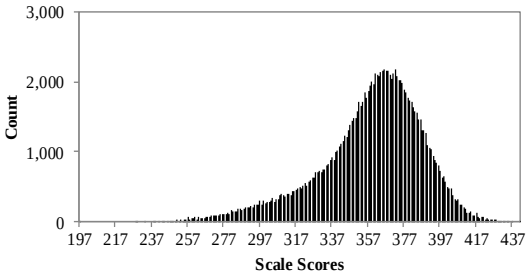

Figure 3.4.4.4B
Proficiency Level: Over 6-8 S400 Online

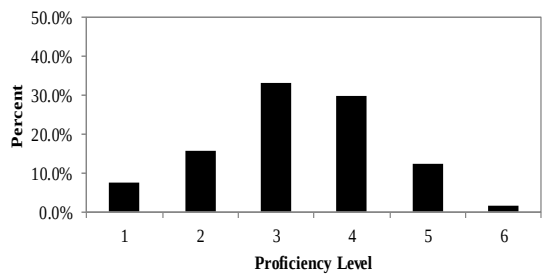

Table 3.4.4.4A

Scale Score Descriptive Statistics: Over 6-8 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
6	43,390	228	437	353.63	26.94
7	42,329	197	434	358.24	28.84
8	42,290	239	441	363.04	30.34
Total	128,009	197	441	358.26	28.98

Table 3.4.4.4B

Proficiency Level Distribution: Over 6-8 S400 Online

	Grade 6		Grade 7		Grade 8		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	2,196	5.06%	3,192	7.54%	4,058	9.60%	9,446	7.38%
2	6,164	14.21%	6,552	15.48%	7,261	17.17%	19,977	15.61%
3	13,805	31.82%	14,669	34.65%	13,905	32.88%	42,379	33.11%
4	14,806	34.12%	11,610	27.43%	11,614	27.46%	38,030	29.71%
5	5,522	12.73%	5,605	13.24%	4,798	11.35%	15,925	12.44%
6	897	2.07%	701	1.66%	654	1.55%	2,252	1.76%
Total	43,390	100.00%	42,329	100.00%	42,290	100.00%	128,009	100.00%

Table 3.4.4.4C

Reliability: Over 6-8 S400 Online

Component	Weight	Variance	Reliability
Listening	0.15	1938.414	0.790
Reading	0.35	870.854	0.860
Speaking	0.15	3117.260	0.783
Writing	0.35	677.305	0.902
Overall Composite		840.030	0.944

*Variances from students who had results in all four domains

Table 3.4.4.4Di

Accuracy and Consistency of Classification Indices: Over (Grade 6) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.765	0.679		0.569	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.890		0.823	
	2	0.793		0.703	
	3	0.815		0.735	
	4	0.768		0.679	
	5	0.599		0.507	
	6	-		0.391	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.986	0.005	0.008	0.981
	2/3	0.957	0.022	0.021	0.938
	3/4	0.918	0.037	0.045	0.885
	4/5	0.923	0.035	0.042	0.890
	5/6	0.979	0.021	0.000	0.979

Table 3.4.4.4Dii

Accuracy and Consistency of Classification Indices: Over (Grade 7) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.763	0.674	0.571		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.894	0.829		
	2	0.759	0.662		
	3	0.813	0.737		
	4	0.722	0.624		
	5	0.671	0.575		
	6	-	0.401		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.980	0.007	0.012	0.973
	2/3	0.949	0.027	0.024	0.927
	3/4	0.915	0.039	0.045	0.882
	4/5	0.934	0.033	0.033	0.906
	5/6	0.983	0.017	0.000	0.983

Table 3.4.4.4Diii

Accuracy and Consistency of Classification Indices: Over (Grade 8) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.746	0.654	0.548		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.897	0.836		
	2	0.764	0.668		
	3	0.803	0.719		
	4	0.693	0.592		
	5	0.590	0.487		
	6	-	0.378		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.977	0.009	0.014	0.968
	2/3	0.945	0.029	0.026	0.922
	3/4	0.915	0.036	0.049	0.881
	4/5	0.923	0.041	0.036	0.891
	5/6	0.985	0.015	0.000	0.984

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 S400 Online

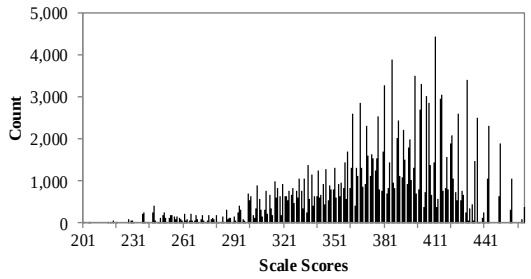

Figure 3.4.5.1B
Proficiency Level: Oral 9-12 S400 Online

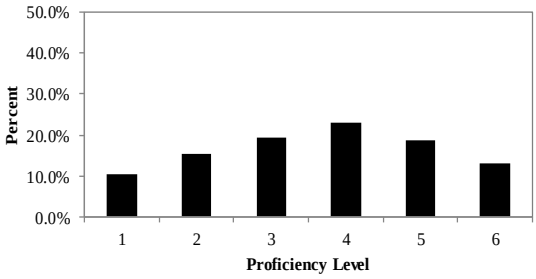

Table 3.4.5.1A

Scale Score Descriptive Statistics: Oral 9-12 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	72,047	201	464	372.87	47.33
10	45,861	220	464	377.21	41.17
11	32,221	205	464	383.44	39.23
12	23,873	217	464	386.83	37.87
Total	174,002	201	464	377.88	43.40

Table 3.4.5.1B

Proficiency Level Distribution: Oral 9-12 S400 Online

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	8,761	12.16%	4,535	9.89%	2,689	8.35%	2,271	9.51%	18,256	10.49%
2	11,875	16.48%	7,562	16.49%	4,731	14.68%	2,810	11.77%	26,978	15.50%
3	11,421	15.85%	9,142	19.93%	7,015	21.77%	5,963	24.98%	33,541	19.28%
4	13,857	19.23%	10,932	23.84%	8,554	26.55%	6,742	28.24%	40,085	23.04%
5	14,975	20.79%	8,538	18.62%	5,609	17.41%	3,406	14.27%	32,528	18.69%
6	11,158	15.49%	5,152	11.23%	3,623	11.24%	2,681	11.23%	22,614	13.00%
Total	72,047	100.00%	45,861	100.00%	32,221	100.00%	23,873	100.00%	174,002	100.00%

Table 3.4.5.1C

Reliability: Oral 9-12 S400 Online

Component	Weight	Variance	Reliability
Listening	0.50	1777.541	0.790
Speaking	0.50	3569.231	0.794
Oral		2009.393	0.862

*Variances from students who had results in all four domains

Table 3.4.5.1Di

Accuracy and Consistency of Classification Indices: Oral (Grade 9) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.542	0.437	0.321		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.805	0.681		
	2	0.580	0.455		
	3	0.431	0.329		
	4	0.446	0.343		
	5	0.479	0.388		
	6	0.673	0.509		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.948	0.022	0.030	0.925
	2/3	0.906	0.039	0.054	0.869
	3/4	0.885	0.054	0.062	0.838
	4/5	0.872	0.064	0.064	0.821
	5/6	0.889	0.070	0.041	0.848

Table 3.4.5.1Dii

Accuracy and Consistency of Classification Indices: Oral (Grade 10) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.562	0.450	0.330		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.804	0.668		
	2	0.598	0.471		
	3	0.508	0.399		
	4	0.524	0.417		
	5	0.489	0.393		
	6	0.675	0.478		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.956	0.018	0.027	0.936
	2/3	0.909	0.041	0.050	0.871
	3/4	0.879	0.056	0.065	0.831
	4/5	0.879	0.062	0.059	0.830
	5/6	0.915	0.060	0.025	0.883

Table 3.4.5.1Diii

Accuracy and Consistency of Classification Indices: Oral (Grade 11) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.564	0.451	0.324		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.799	0.655		
	2	0.582	0.450		
	3	0.547	0.435		
	4	0.555	0.449		
	5	0.462	0.369		
	6	0.675	0.471		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.962	0.015	0.023	0.944
	2/3	0.913	0.041	0.047	0.877
	3/4	0.875	0.055	0.070	0.828
	4/5	0.880	0.060	0.061	0.829
	5/6	0.913	0.064	0.023	0.881

Table 3.4.5.1Div

Accuracy and Consistency of Classification Indices: Oral (Grade 12) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.563	0.456	0.326		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.836	0.706		
	2	0.513	0.384		
	3	0.607	0.492		
	4	0.581	0.472		
	5	0.389	0.313		
	6	0.696	0.472		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.961	0.014	0.025	0.944
	2/3	0.920	0.040	0.040	0.885
	3/4	0.871	0.056	0.073	0.825
	4/5	0.883	0.052	0.065	0.830
	5/6	0.907	0.079	0.015	0.882

3.4.5.2 Literacy Language Composite 9-12

Figure 3.4.5.2A
Scale Scores: Litr 9-12 S400 Online

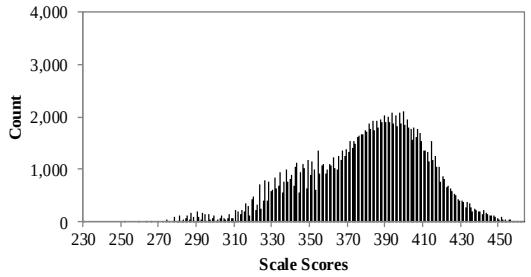

Figure 3.4.5.2B
Proficiency Level: Litr 9-12 S400 Online

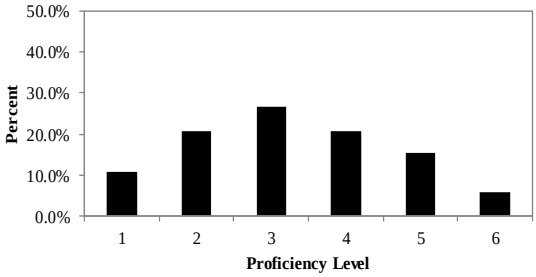

Table 3.4.5.2A

Scale Score Descriptive Statistics: Litr 9-12 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	60,498	230	462	375.81	32.62
10	37,535	266	462	378.33	30.17
11	25,437	269	463	384.46	29.23
12	19,340	272	463	386.12	29.25
Total	142,810	230	463	379.41	31.22

Table 3.4.5.2B

Proficiency Level Distribution: Litr 9-12 S400 Online

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	6,821	11.27%	3,996	10.65%	2,449	9.63%	2,158	11.16%	15,424	10.80%
2	11,886	19.65%	8,349	22.24%	5,092	20.02%	4,284	22.15%	29,611	20.73%
3	14,068	23.25%	10,454	27.85%	7,534	29.62%	5,856	30.28%	37,912	26.55%
4	12,426	20.54%	7,751	20.65%	5,496	21.61%	3,887	20.10%	29,560	20.70%
5	11,259	18.61%	5,207	13.87%	3,382	13.30%	2,135	11.04%	21,983	15.39%
6	4,038	6.67%	1,778	4.74%	1,484	5.83%	1,020	5.27%	8,320	5.83%
Total	60,498	100.00%	37,535	100.00%	25,437	100.00%	19,340	100.00%	142,810	100.00%

Table 3.4.5.2C

Reliability: Litr 9-12 S400 Online

Component	Weight	Variance	Reliability
Reading	0.50	1164.435	0.890
Writing	0.50	1153.896	0.898
Literacy		979.985	0.937

*Variances from students who had results in all four domains

Table 3.4.5.2Di

Accuracy and Consistency of Classification Indices: Litr (Grade 9) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.715	0.611	0.522		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.867	0.783		
	2	0.758	0.663		
	3	0.688	0.584		
	4	0.638	0.526		
	5	0.698	0.589		
	6	0.734	0.577		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.965	0.014	0.021	0.950
	2/3	0.939	0.026	0.034	0.914
	3/4	0.923	0.044	0.033	0.891
	4/5	0.924	0.041	0.035	0.894
	5/6	0.962	0.021	0.017	0.945

Table 3.4.5.2Dii

Accuracy and Consistency of Classification Indices: Litr (Grade 10) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.728	0.626	0.531		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.864	0.778		
	2	0.769	0.678		
	3	0.724	0.628		
	4	0.657	0.544		
	5	0.665	0.551		
	6	0.768	0.590		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.967	0.014	0.019	0.953
	2/3	0.932	0.033	0.036	0.904
	3/4	0.918	0.044	0.038	0.885
	4/5	0.937	0.033	0.030	0.911
	5/6	0.972	0.020	0.008	0.961

Table 3.4.5.2Diii

Accuracy and Consistency of Classification Indices: Litr (Grade 11) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.721	0.618	0.521		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.864	0.778		
	2	0.753	0.655		
	3	0.745	0.652		
	4	0.661	0.547		
	5	0.612	0.505		
	6	0.792	0.615		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.971	0.012	0.017	0.958
	2/3	0.933	0.034	0.033	0.905
	3/4	0.914	0.042	0.044	0.880
	4/5	0.937	0.030	0.032	0.910
	5/6	0.964	0.028	0.008	0.953

Table 3.4.5.2Div

Accuracy and Consistency of Classification Indices: Litr (Grade 12) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.711	0.612	0.511		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.856	0.771		
	2	0.756	0.658		
	3	0.746	0.651		
	4	0.646	0.529		
	5	0.543	0.456		
	6	0.794	0.579		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.966	0.016	0.018	0.952
	2/3	0.926	0.037	0.037	0.896
	3/4	0.915	0.038	0.046	0.881
	4/5	0.943	0.028	0.029	0.916
	5/6	0.958	0.038	0.004	0.952

3.4.5.3 Comprehension Language Composite 9-12

Figure 3.4.5.3A
Scale Scores: Cphn 9-12 S400 Online

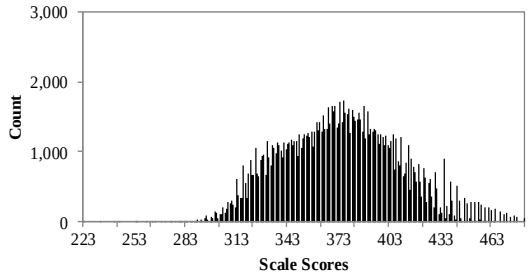

Figure 3.4.5.3B
Proficiency Level: Cphn 9-12 S400 Online

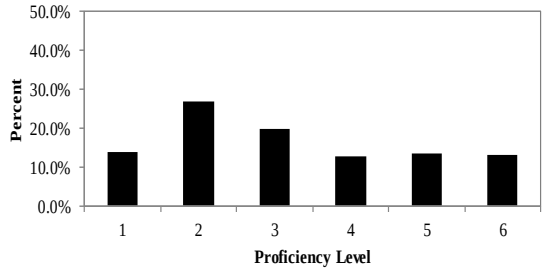

Table 3.4.5.3A

Scale Score Descriptive Statistics: Cphn 9-12 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	57,682	223	482	371.02	35.29
10	35,733	279	482	372.65	33.48
11	24,212	275	482	378.81	33.19
12	18,489	254	482	380.60	32.95
Total	136,116	223	482	374.13	34.35

Table 3.4.5.3B

Proficiency Level Distribution: Cphn 9-12 S400 Online

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	7,819	13.56%	4,894	13.70%	3,342	13.80%	2,935	15.87%	18,990	13.95%
2	14,193	24.61%	10,519	29.44%	6,718	27.75%	4,872	26.35%	36,302	26.67%
3	12,259	21.25%	7,253	20.30%	4,389	18.13%	3,243	17.54%	27,144	19.94%
4	6,875	11.92%	4,606	12.89%	3,264	13.48%	2,765	14.95%	17,510	12.86%
5	8,683	15.05%	4,521	12.65%	3,110	12.84%	2,067	11.18%	18,381	13.50%
6	7,853	13.61%	3,940	11.03%	3,389	14.00%	2,607	14.10%	17,789	13.07%
Total	57,682	100.00%	35,733	100.00%	24,212	100.00%	18,489	100.00%	136,116	100.00%

Table 3.4.5.3C

Reliability: Cphn 9-12 S400 Online

Component	Weight	Variance	Reliability
Listening	0.30	1777.541	0.790
Reading	0.70	1164.435	0.890
Comprehension		1167.610	0.917

*Variances from students who had results in all four domains

Table 3.4.5.3Di

Accuracy and Consistency of Classification Indices: Cphn (Grade 9) S400 Online

Overall Indices	**Accuracy**	**Consistency**		**Kappa (k)**	
	0.666	0.565		0.469	
Conditional on Level	**Level**	**Accuracy**		**Consistency**	
	1	0.832		0.723	
	2	0.700		0.604	
	3	0.608		0.494	
	4	0.423		0.323	
	5	0.604		0.482	
	6	0.840		0.740	
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.943	0.020	0.037	0.922
	2/3	0.917	0.043	0.040	0.882
	3/4	0.915	0.048	0.038	0.880
	4/5	0.923	0.044	0.033	0.893
	5/6	0.952	0.027	0.021	0.932

Table 3.4.5.3Dii

Accuracy and Consistency of Classification Indices: Cphn (Grade 10) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)	
	0.668	0.567	0.463	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.815	0.699	
	2	0.731	0.642	
	3	0.581	0.468	
	4	0.463	0.356	
	5	0.581	0.455	
	6	0.836	0.729	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.940	0.023	0.037	0.917
	2/3	0.910	0.048	0.042	0.874
	3/4	0.913	0.047	0.040	0.879
	4/5	0.930	0.038	0.031	0.902
	5/6	0.960	0.023	0.017	0.942

Table 3.4.5.3Diii

Accuracy and Consistency of Classification Indices: Cphn (Grade 11) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)	
	0.656	0.556	0.455	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.804	0.689	
	2	0.708	0.614	
	3	0.533	0.422	
	4	0.464	0.357	
	5	0.559	0.434	
	6	0.855	0.761	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.939	0.025	0.037	0.914
	2/3	0.908	0.050	0.042	0.871
	3/4	0.909	0.047	0.043	0.874
	4/5	0.928	0.039	0.033	0.898
	5/6	0.954	0.026	0.019	0.935

Table 3.4.5.3Div

Accuracy and Consistency of Classification Indices: Cphn (Grade 12) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.649	0.546	0.446		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.778	0.668		
	2	0.670	0.567		
	3	0.515	0.405		
	4	0.514	0.400		
	5	0.534	0.407		
	6	0.869	0.782		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.927	0.034	0.039	0.896
	2/3	0.905	0.054	0.042	0.867
	3/4	0.909	0.049	0.042	0.874
	4/5	0.933	0.034	0.032	0.904
	5/6	0.958	0.024	0.018	0.940

3.4.5.4 Overall Language Composite 9-12

Figure 3.4.5.4A
Scale Scores: Over 9-12 S400 Online

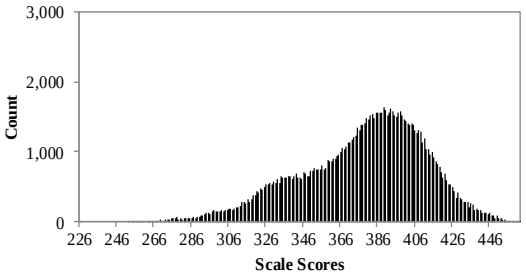

Figure 3.4.5.3B
Proficiency Level: Cphn 9-12 S400 Online

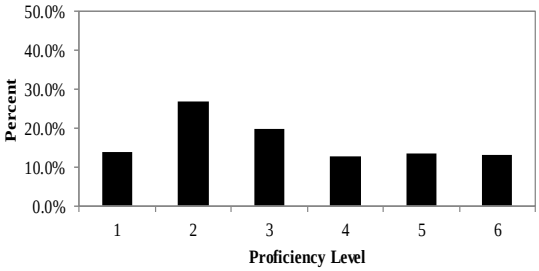

Table 3.4.5.4A

Scale Score Descriptive Statistics: Over 9-12 S400 Online

Grade	No. of Students	Min.	Max.	Mean	Std. Dev.
9	50,085	226	461	373.77	35.55
10	31,029	258	462	376.90	31.99
11	21,026	264	461	383.08	30.77
12	16,041	269	462	385.25	30.26
Total	118,181	226	462	377.80	33.42

Table 3.4.5.4B

Proficiency Level Distribution: Over 9-12 S400 Online

	Grade 9		Grade 10		Grade 11		Grade 12		Total	
Level	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**	**Count**	**Percent**
1	6,021	12.02%	3,140	10.12%	1,993	9.48%	1,621	10.11%	12,775	10.81%
2	9,024	18.02%	6,362	20.50%	3,735	17.76%	2,840	17.70%	21,961	18.58%
3	10,789	21.54%	8,433	27.18%	6,040	28.73%	5,085	31.70%	30,347	25.68%
4	11,022	22.01%	6,892	22.21%	5,017	23.86%	3,811	23.76%	26,742	22.63%
5	9,409	18.79%	4,570	14.73%	2,985	14.20%	1,806	11.26%	18,770	15.88%
6	3,820	7.63%	1,632	5.26%	1,256	5.97%	878	5.47%	7,586	6.42%
Total	50,085	100.00%	31,029	100.00%	21,026	100.00%	16,041	100.00%	118,181	100.00%

Table 3.4.5.4C

Reliability: Over 9-12 S400 Online

Component	Weight	Variance	Reliability
Listening	0.15	1777.541	0.790
Reading	0.35	1164.435	0.890
Speaking	0.15	3569.231	0.794
Writing	0.35	1153.896	0.898
Overall Composite		1116.783	0.951

*Variances from students who had results in all four domains

Table 3.4.5.4Di

Accuracy and Consistency of Classification Indices: Over (Grade 9) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)		
	0.739	0.640	0.560		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.891	0.825		
	2	0.768	0.678		
	3	0.708	0.605		
	4	0.694	0.586		
	5	0.704	0.599		
	6	0.755	0.616		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.969	0.013	0.018	0.957
	2/3	0.948	0.024	0.028	0.927
	3/4	0.930	0.039	0.031	0.902
	4/5	0.931	0.035	0.034	0.903
	5/6	0.960	0.023	0.017	0.942

Table 3.4.5.4Dii

Accuracy and Consistency of Classification Indices: Over (Grade 10) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)	
	0.758	0.663	0.579	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.884	0.811	
	2	0.787	0.703	
	3	0.757	0.667	
	4	0.710	0.605	
	5	0.699	0.594	
	6	0.793	0.640	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.973	0.011	0.016	0.961
	2/3	0.943	0.028	0.029	0.919
	3/4	0.927	0.038	0.035	0.898
	4/5	0.943	0.029	0.028	0.919
	5/6	0.972	0.019	0.009	0.961

Table 3.4.5.4Diii

Accuracy and Consistency of Classification Indices: Over (Grade 11) S400 Online

Overall Indices	Accuracy	Consistency	Kappa (k)	
	0.751	0.656	0.569	
Conditional on Level	**Level**	**Accuracy**	**Consistency**	
	1	0.887	0.816	
	2	0.765	0.671	
	3	0.776	0.689	
	4	0.722	0.619	
	5	0.648	0.549	
	6	0.802	0.642	
Indices at Cut Points		**Accuracy**		
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.975	0.010	0.015	0.965
	2/3	0.945	0.029	0.027	0.921
	3/4	0.924	0.037	0.039	0.893
	4/5	0.942	0.027	0.031	0.918
	5/6	0.964	0.028	0.008	0.954

Table 3.4.5.4Div

Accuracy and Consistency of Classification Indices: Over (Grade 12) S400 Online

Overall Indices	**Accuracy**	**Consistency**	**Kappa (k)**		
	0.739	0.650	0.557		
Conditional on Level	**Level**	**Accuracy**	**Consistency**		
	1	0.882	0.811		
	2	0.754	0.654		
	3	0.795	0.713		
	4	0.723	0.618		
	5	0.543	0.473		
	6	0.818	0.613		
Indices at Cut Points		**Accuracy**			
	Cut Point	**Accuracy**	**False Positives**	**False Negatives**	**Consistency**
	1/2	0.974	0.012	0.014	0.963
	2/3	0.942	0.031	0.027	0.917
	3/4	0.921	0.035	0.043	0.890
	4/5	0.947	0.023	0.030	0.923
	5/6	0.954	0.044	0.002	0.952

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.

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

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

Cook, H. G. and MacGregor, D. (2017). *The ACCESS for ELLs 2.0 2016 Standard Setting Study* (Technical Report). Madison, WI: Board of Regents of the University of Wisconsin System.

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.

Data Recognition Corporation (2016a). DRC INSIGHT Online Learning System Technology User Guide. Maple Grove, MN: DRC/WIDA. Retrieved on February 28, 2017 from https://www.wida.us/assessment/access%202.0/documents/2016InsightTechnologyUserGuide.pdf.

Data Recognition Corporation (2016b). WIDA Assessment Management System (WIDA AMS). Maple Grove, MN: DRC/WIDA. Retrieved on February 28, 2017 from https://www.wida.us/assessment/access%202.0/documents/WIDAAMSUserGuide.pdf.

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).

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

Kolen, M. J., & Brennan, R. L. (2004). *Test Equating: Methods and Practices* (2nd ed.). New York, NY: Springer-Verlag.

Lee, W., Hanson, B.A., & Brennan, R.L. (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement, 26, 412-432.*

Linacre, J. M. (1999, Autumn). Relating Cronbach and Rasch Reliabilities Clauser B., Linacre J.M. … Rasch Measurement Transactions, 1999, 13:2 p. 696. Retrieved on February 26, 2017 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 on April 10, 2006 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. (1992). *Linking Educational Assessments: Concepts, Issues, Methods, and Prospects*. Princeton, NJ: Educational Testing Service, Policy Information Center.

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

Pommerich, M. (2007). Concordance: The good, the bad, and the ugly. In N. J. Dorans, M. Pommerich, & P. W. Holland (Ed.), *Linking and Aligning Scores and Scales* (pp. 199–230). New York, NY: Springer-Verlag.

Pommerich, M. Hanson, B. A., Harris, D. J., & Sconing, J. A. (2004). Issues in conducting linkages between distinct tests. *Applied Psychological Measurement, 28(4),* 247–273.

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. (2016). WIDA ACCESS for ELLs 2.0 2016–2017 Test Administrator Manual. Madison, WI: Author. Retrieved on February 28, 2017 from https://www.wida.us/assessment/access%202.0/documents/2016TestAdministratorManual.pdf.

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.

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., 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.

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.

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.
Min Feng, B.S.
Justin Kelly, Ph.D.
Daniel Lee, M.S.
Chi-Kai (Cary) Lin, Ph.D.
Mohammed Louguit, Ph.D.
Dorry M. Kenyon, Ph.D.
David MacGregor, Ph.D.
Meg Montee, M.A.T
Jennifer Norton, Ed.D.
Jennifer Renn, Ph.D.
Shu Jing Yen, Ph.D.
Jing Wei, Ph.D.
Xin Yu, M.A.
Yu-Chia Wu, M.Ed.