

ACCESS for ELLs  
Year in Review  
October 2017-October 2018

December 14, 2018

# ACCESS for ELLs Year in Review

October 2017–October 2018

## Introduction

This report provides a review of the ACCESS for ELLs (ACCESS) test development and operational activities conducted over the course of the thirteen months from October 2017–October 2018. Written for stakeholders in the WIDA consortium states, the report includes overviews of test development, operational testing, psychometric analyses, and validation research.

## Annual Timeline

Figure 1 shows an overview of major test-development activities during October 2017—October 2018. Each row represents a test series, that is, an operational testing program for a given school year. Test Series 402 was operational in school year 2017–18, while field testing and final item selection were ongoing for Series 403, to be administered in school year 2018–19. As you read this document, it may help to refer to this timeline to see how the activities fit together.

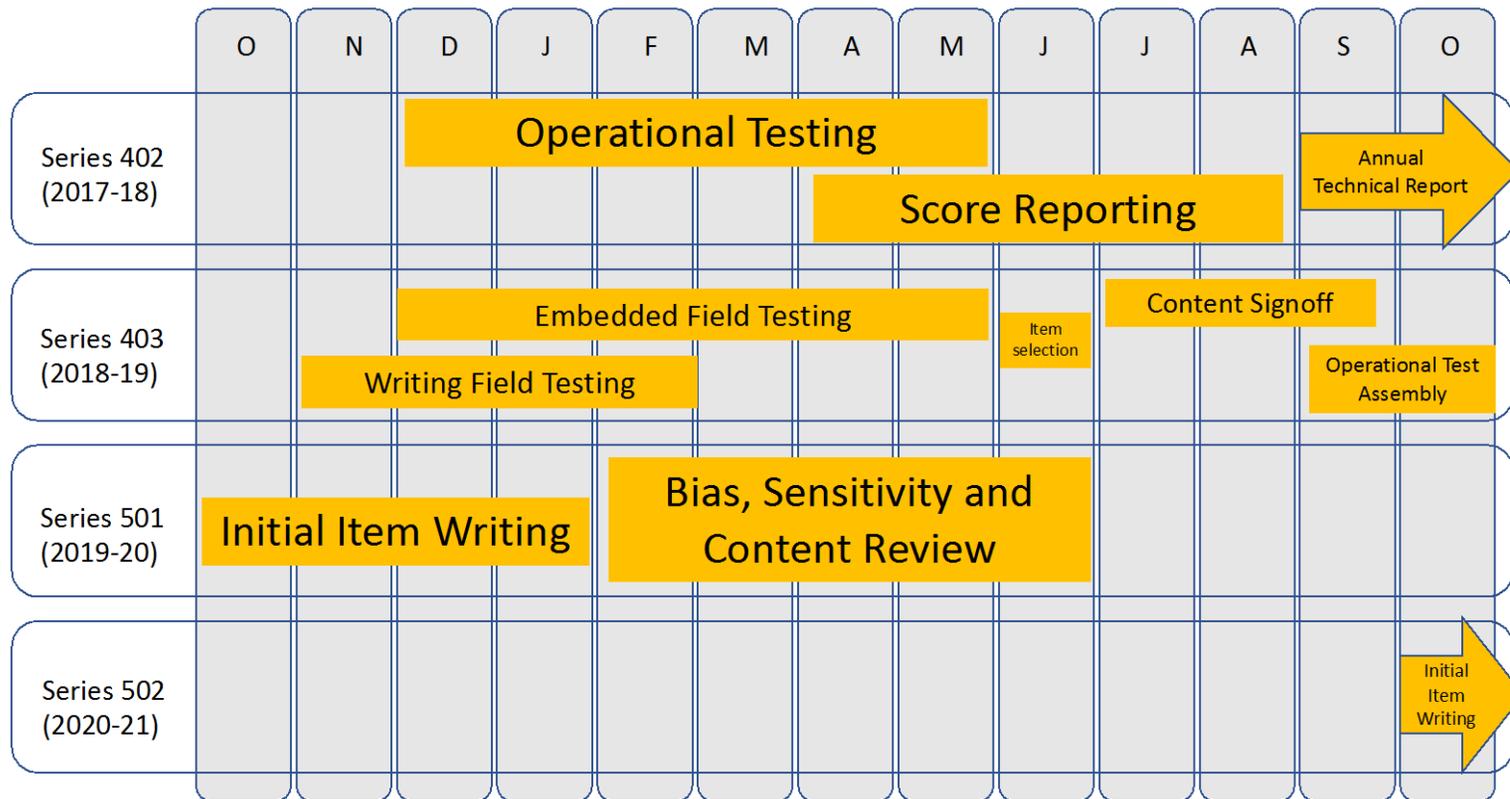


Figure 1. Major test development activities, October 2017–October 2018

## Development Overview

### How Many New Items were Developed?

Table 1 shows the number of new items developed for Series 403, the number of items that were chosen for operational testing based on field test results, and the number of items developed for Series 501 that passed bias, sensitivity and content reviews.

Table 1. *Field Test Item Development Breakdown by Domain and Series*

Cluster	Domain	Items Field Tested (on Series 402)	Series 403		Items Developed (for Series 401)	Items Evaluated in Bias & Sensitivity Review (Series 501)	
			New Operational Items	Total Operational Items		Submitted	Approved
Grade 1	Listening	24	9	54	18	18	18
	Reading	24	12	72	36	36	36
	Speaking	16	10	12	16	16	16
	Writing	2	4	10	6	6	6
Grade 2–3	Listening	24	3	54	18	18	17
	Reading	24	9	72	36	36	36
	Speaking	16	10	12	8	8	8
	Writing	4	4	9	6	6	6
Grade 4–5	Listening	24	9	54	18	18	18
	Reading	24	12	72	42	42	42
	Speaking	16	10	12	16	16	16
	Writing	4	4	9	6	6	6
Grade 6–8	Listening	24	15	54	24	24	24
	Reading	24	6	72	42	42	42
	Speaking	16	10	12	16	16	16
	Writing	4	4	9	6	6	4
Grade 9–12	Listening	24	9	54	30	30	29
	Reading	24	12	72	36	36	36
	Speaking	16	15	12	16	16	16
	Writing	4	3	9	6	6	6
Total	Listening	120	45	270	108	108	106
	Reading	120	51	360	192	192	192
	Speaking	80	55	60	72	72	72
	Writing	18	19	46	30	30	28

## What New Developments are in Store?

In 2018–19 ACCESS for ELLs will see the following changes:

- Tier B and Tier C test forms on ACCESS for ELLs Paper (hereafter referred to as *ACCESS Paper*) will be consolidated to a single Tier B/C test form for the Listening and Reading domains.
- WIDA will begin the development of content for new and additional ACCESS Paper test forms at Grades 6–8 and 9–12 for the Listening and Reading domains.

## Operational Overview

### How Does the Test Implement the Standards?

Appendix A shows how the WIDA ELD Standards are implemented across domains. The five Standards are Social and Instructional Language (SI), the Language of Language Arts (LA), the Language of Math (MA), the Language of Science (SC), and the Language of Social Studies (SS). Note that the numbers in Appendix A show the number of items by standard for the entire grade-level; because of the adaptivity of the online test and the tiered structure of the paper-based test they do not reflect the testing experience of the student. Table 2 shows the distribution of items for Listening and Reading on ACCESS for ELLs Online (hereafter referred to as *ACCESS Online*), while Table 3 shows the distribution for ACCESS Paper.

Table 2. *Distribution of Items on ACCESS Online*

<b>Listening</b>	<b>SI</b>	<b>LA</b>	<b>MA</b>	<b>SC</b>	<b>SS</b>	<b>Total</b>
Low proficiency	6	3	3	3	3	18
Mid/High proficiency	6	6	6	3	3	24
<b>Reading</b>						
Low proficiency	6	6	6	3	3	24
Mid/High proficiency	6	6	6	6	6	30

Table 3. *Distribution of items on ACCESS Paper*

<b>Listening</b>	<b>SI</b>	<b>LA</b>	<b>MA</b>	<b>SC</b>	<b>SS</b>	<b>Total</b>
Low proficiency	6	3	3	3	3	18
Mid/High proficiency	3	6	6	3	3	21
<b>Reading</b>						
Low proficiency	6	6	6	3	3	24
Mid/High proficiency	3	6	6	6	6	27

For the Writing domain, the online and paper-based tests have the same distribution of items, as shown in Table 4. (Note that Standards are combined for some items on the Writing test, as follows: MA and SC (MS) and an Integrated Task (IT) combining SI, LA, and SS.)

Table 4. *Distribution of Writing Items*

Writing	SI	MS	LA	IT	Total
Grade 1, Tier A	4				4
All other Tier A	1	1	1		3
Tiers B/C	1	1		1	3

For the Speaking domain, all students see two SI items, two combined LA/SS (LS) items, and two MS items.

### How Many Tests were Administered?

Table 5 shows the number of students tested by grade and mode of administration (online, paper, or mixed).

Table 5. *Number of Students Tested by Grade and Mode of Administration*<sup>1</sup>

Grade	Online	Paper	Mixed	Total
K	30	239,252	-	239,282
1	185,107	60,981	273	246,361
2	190,168	56,828	260	247,256
3	195,969	48,346	338	244,653
4	186,126	37,467	149	223,742
5	122,694	27,493	72	150,259
6	92,598	23,719	18	116,335
7	88,515	20,057	9	108,581
8	86,461	18,939	6	105,406
9	100,865	19,794	21	120,680
10	92,193	18,214	14	110,421
11	67,318	15,379	15	82,712
12	47,649	10,635	3	58,287
Total	1,455,693	597,104	1,178	2,053,975

### Operational Anomalies

In 2017–18 there no unplanned outages during the ACCESS for ELLs testing window. In addition, there were no scoring anomalies in the 2017–18 administration.

<sup>1</sup> These numbers are based on information provided to WIDA on October 6, 2018. Some minor changes may occur when finalizing the data (e.g., since the Kindergarten test is not online, it is likely that Kindergarten students listed as online or mixed will be reclassified).

## Psychometrics in Brief

### How were Scoring Tables Created for Speaking and Writing?

In spring of 2018, using data from early-testing states, data from scored Speaking and Writing test forms were used to create scoring tables for the operational ACCESS Series 402 Online assessment. This data was used to analyze the difficulty of each test item and to produce a scoring table by grade level that allowed students' performances to be interpreted using the ACCESS scale.

For the operational ACCESS Series 402 Paper assessment in Writing, the tasks that students took were the same sets of tasks taken by students in the Series 400 Paper Writing assessment; therefore, the scoring tables created for the Series 400 assessment were used for the Series 402 Paper Writing test. For the operational ACCESS Series 402 Paper assessment in Speaking, the Series 402 assessment was constructed by revising the Series 400 Speaking assessment to ensure that there was no overlap with the Screener assessment. Series 402 scoring tables were created by using difficulty measures of continuing tasks from the Series 400 assessments, and difficulty measures for new items drawn from the analyses of those items in the spring 2017 Speaking field test. These values were used to produce a scoring table by grade level that allowed students' performances to be interpreted using the ACCESS for ELLs scale.

### What Happened with Speaking Scores?

Some WIDA states reported having speaking scores (especially in Grade 1) that were lower than expected. WIDA checked the technical processes involved in scoring and reporting speaking test scores and found no anomalies. Nonetheless, WIDA is expanding its speaking training material to better support online scorers and examining its scoring methodology to see how it might be improved.

### How was the ACCESS Online Field Test Analyzed?

In the spring of 2018, CAL received results from the ACCESS Series 403 Field Test. New items were field tested in all four domains. (See Table 1.) Data collected from the field test were analyzed to provide input into the selection process for Series 403 operational items. Quantitative analyses are used to provide evidence that items are of appropriate difficulty for the test. For Listening and Reading assessments, Differential Item Functioning (DIF) analyses were conducted to ensure that there were no systematic differences between subgroups of students by

gender or ethnicity. The selection process for the Series 403 operational items uses a mixed method approach, i.e., combining quantitative results with qualitative reviews of items.

### What is WIDA Doing about the Issue of Comparability?

Yearly comparability analyses between ACCESS Paper and ACCESS Online are performed to examine domain and composite level score differences. To date, there have been observable scale score differences between modes. To correct for mode differences, an equipercentile linking procedure that adjusts paper domain scores to online scale scores is annually conducted (report forthcoming). Based upon WIDA's Technical Advisory Committee (TAC) recommendation, several studies have been undertaken to further understand scores difference. For more information, see the construct validity and Speaking rater study descriptions below.

### What Quality Control Reviews were Conducted?

Over the year, WIDA conducted several quality control reviews of partner organizations (DRC and CAL), including:

- User acceptance testing, to ensure the online adaptive Reading and Listening tests performed as expected from the test-taker's perspective
- Psychometric State Student Record (SSR) testing, to ensure that ACCESS scoring algorithms are calculated and recorded accurately
- DRC onsite handscoring review, to ensure that the Speaking and Writing tests are scored accurately<sup>2</sup>
- DRC onsite score reporting review, to ensure that the information provided in various score reports is correct<sup>2</sup>
- CAL test development quality control review, to ensure that high standards are consistently applied in the development of new test items<sup>2</sup>.

To better organize and coordinate reviews, WIDA has created and filled a Quality Control Officer position. This individual serves as an archivist for all QC documentation, organizes and leads the DRC onsite score reporting review, facilitates communications between QC teams, and produces an annual report on WIDA's QC reviews and their outcomes and recommendations.

---

<sup>2</sup> These reviews include member state participation.

## Validation Research Overview

WIDA is committed to providing the most valid ELD assessments possible. Accordingly, there is an ongoing series of studies supporting the validation ACCESS for ELLs. The following studies are this year's validation research activities.

### What Research has been Completed?

#### *Consequential Validity Study: EL Identification and Placement Decisions*

This study examined how K–12 educators identify and place English learners in language instruction educational programs. An online survey was completed by 476 educators (207 district-level and 269 school-level educators) across the WIDA Consortium. Findings indicate that decision makers were EL/Title III coordinators (district level) and EL teachers (school level). They heavily relied on home language survey data and WIDA screener scores (MODEL, W-APT, and the new WIDA Screener) for EL identification, and ACCESS for ELLs scores for EL placement. Educators generally perceived EL identification and placement decisions to be appropriate, and misidentification and misplacement to be rare; however, findings suggest some over-identification of ELs.

#### *Educator Perception of Online Accessibility Features*

This study examined how K–12 EL educators perceive and instruct online accessibility features embedded in ACCESS for ELLs. Ten features were examined: five available across the four language domains (Colored Overlays, Color Contrast, Highlighter, Line Guide, and Magnifier) and five that are specific to the Writing domain (Sticky Notes, Cut, Copy, Paste, and Underline). Findings from an online survey ( $n = 377$ ) and follow-up interview ( $n = 9$ ) show that educators perceived the Highlighter, Line Guide, Copy, and Paste to be more important than others. Their perception seemed to affect what features they chose to instruct ELs, both in classroom contexts and in preparation for assessments.

### What Research is Ongoing?

#### *ACCESS Screener Predictive Validity Study*

WIDA is conducting research into the predictive validity of the WIDA Screener Online. The aims of this study are the following:

- Understand the reliability of screener scores
- Understand how accurately screener scores identify ELs
- Understand the factor structure of screener scores

- Understand the predictive validity of screener score for ACCESS for ELLs 2.0 scores.

The findings of this study will answer the following research questions:

- To what extent do WIDA Screener scores predict ACCESS for ELLs scores?
- Which WIDA Screener domain/composite score best predicts ACCESS for ELLs overall scores?
- How reliable are WIDA Screener scores?
- What percentage of students are accurately identified as ELLs based on WIDA Screener scores?
- Is the factor structure of WIDA Screener similar to that of ACCESS for ELLs? Does Screener measure beyond a single factor?

### *Construct Validity Study*

There is an ongoing study to examine the construct validity of the internal structures of the paper and online versions of ACCESS for ELLs, Series 400 for Grades 1–12 and Series 303 for Kindergarten. Results will be used to support the underlying constructs of ACCESS for ELLs as they relate to the WIDA ELD standards.

Confirmatory factor analyses and structural equation modeling techniques, with orthogonal rotations, are being used to model the underlying structures of all items in the online and paper versions at each grade level for each tier. Initially, models were produced based on tiers within grade clusters. However, because it was anticipated comparisons would be needed between the two formats, separate grade level modelling is being undertaken.

To date, models for the online format for each Tier in Grades 1–7 provide evidence for the existence of four distinct language domains and a second order language domain factor (academic language). Model specifications are being kept as consistent as possible across the models.

### *Use of Accessibility Features Embedded in ACCESS for ELLs*

This study examined Grades 1–12 ELs' use of online accessibility features embedded in ACCESS for ELLs. Approximately 1.3 million ELs' ACCESS test data and telemetry information (i.e., records of test-takers' actions during the test) were analyzed. ELs with disabilities comprised 11% of the data. ELs' use of seven tools were examined: *Colored Overlays*, *Color Contrast*, *Help Tools*, *Line Guide*, *Highlighter*, *Magnifier*, and *Sticky Notes*. Preliminary findings show that ELs as a whole generally used the Line guide, Highlighter, and

Magnifier more frequently than other accessibility features, but used the Help Tools the least. Use of accessibility features was more common in the selected response Listening and Reading domains, which were administered prior to the constructed response Speaking and Writing sections. The comparison between ELs with and without disabilities revealed that higher percentages of ELs with disabilities activated the accessibility features across all domains than ELs without disabilities. Although the difference in the use of some features between the two groups was statistically significant, effect sizes were small.

### *Speaking Rater Study*

A mixed-methods study was conducted to explore scoring variations across different rating contexts in regard to ELL students' spoken responses. The study was partially motivated by the observed score gaps of spoken performance between the two administration modes of ACCESS Speaking in recent years; namely, face-to-face testing (ACCESS Paper; scoring done on site by local ELL educators/test administrators) and computer-delivered testing (ACCESS Online; scoring done post hoc by centrally trained raters). Preliminary results suggest that:

- According to local teachers/raters, student's familiarity with test administrators is important in eliciting extensive responses from students.
- For higher difficulty tasks, response times seemed short.
- Overall, local teacher ratings were higher than central scorer ratings.
- A small amount of variation was observed in ratings between central raters, while a substantial amount of variation was observed between local raters.
- The highest level of disagreement in the scoring rubric was observed between Adequate (2) and Strong (3), followed by Strong (3) and Exemplary (4).

## Appendix

### Standards Coverage

Table A-1. *Series 401 Listening and Reading Item Coverage by Standards: ACCESS Paper*

Cluster	Standard	Domain			
		Listening		Reading	
		#Items	Percent	#Items	Percent
K	SI	12	40%	12	40%
	LA	3	10%	0	0%
	SS	12	40%	3	10%
	MA	3	10%	0	0%
	SC	0	0%	15	50%
1	SI	12	20%	13	16%
	LA	15	25%	19	23%
	SS	9	15%	16	19%
	MA	15	25%	19	23%
	SC	9	15%	16	19%
2	SI	12	20%	13	16%
	LA	15	25%	19	23%
	SS	9	15%	16	19%
	MA	15	25%	19	23%
	SC	9	15%	16	19%
3	SI	12	20%	12	15%
	LA	15	25%	18	23%
	SS	9	15%	15	19%
	MA	15	25%	18	23%
	SC	9	15%	15	19%
4-5	SI	12	20%	12	15%
	LA	15	25%	18	23%
	SS	9	15%	15	19%
	MA	15	25%	18	23%
	SC	9	15%	15	19%
6-8	SI	12	20%	12	15%
	LA	15	25%	18	23%
	SS	9	15%	15	19%
	MA	15	25%	18	23%
	SC	9	15%	15	19%

9–12	SI	12	20%	12	15%
	LA	15	25%	18	23%
	SS	9	15%	15	19%
	MA	15	25%	18	23%
	SC	9	15%	15	19%

Table A-2. Series 401 Listening and Reading Item Coverage by Standards: ACCESS Online

Mode=Online	Standard	Domain			
		Listening		Reading	
		#Items	Percent	#Items	Percent
Cluster					
1	SI	6	11%	6	8%
	LA	15	28%	18	25%
	MA	15	28%	18	25%
	SC	9	17%	15	21%
	SS	9	17%	15	21%
2–3	SI	6	11%	6	8%
	LA	15	28%	18	25%
	MA	15	28%	18	25%
	SC	9	17%	15	21%
	SS	9	17%	15	21%
4–5	SI	6	11%	6	8%
	LA	15	28%	18	25%
	MA	15	28%	18	25%
	SC	9	17%	15	21%
	SS	9	17%	15	21%
6–8	SI	6	11%	6	8%
	LA	15	28%	18	25%
	MA	15	28%	18	25%
	SC	9	17%	15	21%
	SS	9	17%	15	21%
9–12	SI	6	11%	6	8%
	LA	15	28%	18	25%
	MA	15	28%	18	25%
	SC	9	17%	15	21%
	SS	9	17%	15	21%

Table A-3. Series 401 Writing Item Coverage by Standards

Writing			
Cluster	Standard	#Items	Percent
K <sup>3</sup>	SI	2	18%
	MA	3	27%
	SS	4	36%
	IT	2	18%
1	IT	2	20%
	LA	0	0%
	MS	2	20%
	SI	6	60%
2	IT	2	22%
	LA	1	11%
	MS	3	33%
	SI	3	33%
3	IT	2	22%
	LA	1	11%
	MS	3	33%
	SI	3	33%
4-5	IT	2	22%
	LA	1	11%
	MS	3	33%
	SI	3	33%
6-8	IT	2	22%
	LA	1	11%
	MS	3	33%
	SI	3	33%
9-12	IT	2	22%
	LA	1	11%
	MS	3	33%
	SI	3	33%

<sup>3</sup> The first Writing item for Kindergarten is intended to ensure that students have a basic ability to write and is not based on the WIDA Standards.

Table A-4. Series 401 Speaking Item Coverage by Standards

Speaking			
Cluster	Standard	#Items	Percent
K <sup>4</sup>	SI	2	20%
	LA	1	10%
	MA	3	30%
	SC	1	10%
	SS	3	30%
1	LS	4	33%
	MS	4	33%
	SI	4	33%
2-3	LS	4	33%
	MS	4	33%
	SI	4	33%
4-5	LS	4	33%
	MS	4	33%
	SI	4	33%
6-8	LS	4	33%
	MS	4	33%
	SI	4	33%
9-12	LS	4	33%
	MS	4	33%
	SI	4	33%