



Examining the relationship between the WIDA Screener and ACCESS for ELLs assessments

Prepared by
David MacGregor
Narek Sahakyan



Wisconsin Center for
Education Research
SCHOOL OF EDUCATION
UNIVERSITY OF WISCONSIN-MADISON

Contents

Executive Summary	4
Introduction	4
Screener Overview	5
Purpose of Screener	5
Screener Design	5
Online Administration	6
Screener Scoring	6
Purpose of this study	6
Methods	7
Data	7
Results and Discussion	9
Conclusion	13

List of Figures

Figure 1: Overview of the Test Design of Screener Online	5
Figure 2: Scatterplot of Screener and ACCESS Overall Composite Scale Scores.	10

List of Tables

Table 1: Calculation of Composite Scores	6
Table 2: Descriptive Statistics: Participation	8
Table 3: Descriptive Statistics: Distribution of Students by Grade	8
Table 4: Distribution of School by Type	8
Table 5: Distribution of States, Districts, and Schools by Relative Size	9
Table 6: Correlations between Screener and ACCESS Overall Composite Scale Scores.	9
Table 7: Mixed-Effects Model: Parameter Estimates	11

Executive Summary

This report describes a comparability study between WIDA Screener Online (hereafter, Screener) and ACCESS for ELLs Online (hereafter, ACCESS). The purpose of the study is to understand to what extent scores on Screener predict scores on ACCESS and how that relationship is affected by variables like grade level, time between tests, and special education needs.

Data was collected from more than 35,000 Grade 1-12 English learner (EL) students enrolled in over 10,000 schools and 2,500 districts across 35 WIDA states. These students took both Screener and ACCESS within a single academic year.

The data showed a strong correlation between a student's Overall composite scale score on Screener and that same student's score on ACCESS. Every individual factor analyzed, including grade, Individualized Education Program (IEP) status, gender, ethnicity, location, and time between test administrations, had a statistically significant effect on students' ACCESS performance. However, the predictive relationship between the two tests remains even when these factors are accounted for.

Results show that Screener scores are strongly predictive of ACCESS scores; a one-point increase in the Screener Overall composite scale score is associated with about a 0.70 point increase in the ACCESS Overall composite scale score.

The results of this study provide evidence for the predictive validity of Screener, and suggest that schools can continue to confidently use Screener as an identification tool for students across the WIDA Consortium.

Introduction

This report describes a comparability study between WIDA Screener Online (hereafter Screener) and ACCESS for ELLs Online (hereafter ACCESS). The purpose of the study is to understand to what extent scores on Screener predict scores on ACCESS, and whether and how that relationship is affected by several covariates, such as grade level, time between tests, and Individualized Education Program (IEP) status. The results of this study are intended to provide evidence of the predictive validity of Screener for schools that use ACCESS as a summative assessment.

Screener Overview

Purpose of Screener

Screener is an assessment designed to provide an initial measure of a student’s academic English language proficiency. The purpose of this assessment is to help educators identify students who are candidates for English language support services and to help parents and education teams make decisions about students’ need for language support.

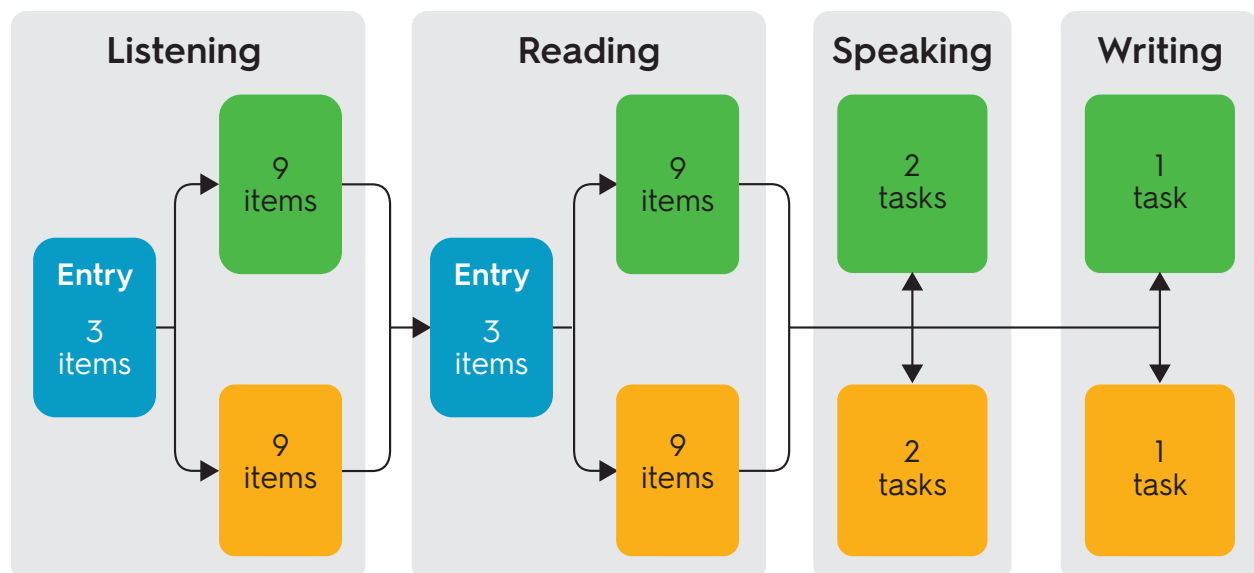
Screener Design

Screener is divided into five grade-level clusters: 1, 2-3, 4-5, 6-8, and 9-12. Each grade-level cluster test comprises four language domain tests: Listening, Reading, Speaking, and Writing.

The Screener Listening and Reading tests consist of items organized into groups of three, designed to test the language of progressively higher proficiency levels. Each of these item groups, or folders, is built around a common theme and tests the language of one of the five WIDA English Language Development Standards (Social Instructional Language, Language of Language Arts, Language of Math, Language of Science, and Language of Social Studies). Both the Listening and Reading tests include one entry folder that all students see. Depending on how students perform on the three initial items, they continue the test and see either the folders of Tier A (easier test content, appropriate for students at beginning levels of language proficiency) or Tier C (more challenging test content, appropriate for students at language proficiency levels that demonstrate they might not need English language services).

The Speaking and Writing tests are made up of performance tasks scored according to specific criteria. A student’s scale score on the Listening and Reading tests determines whether the student sees a Tier A or Tier C folder in both Speaking and Writing. Students in Grades 1-3 respond to Writing prompts in a paper-and-pencil format; for all other grade levels, students can type or handwrite their responses.

Figure 1: Overview of the Test Design of Screener Online



Online Administration

Each Screener domain test can be administered individually or in small groups. Listening and Reading are administered first and are scored automatically by the online test platform. Speaking and Writing tasks are scored locally by trained raters using WIDA-developed scoring scales.

Screener Scoring

Screener scores are calculated as scale scores on the ACCESS scale and reported as proficiency level scores, which are interpretations of the scale scores in terms of the six WIDA English language proficiency levels.

Proficiency level scores are provided for the following:

- Four language domains: Listening, Reading, Writing, and Speaking. Language domain scores are reported as integers from 1 to 6.
- Three composite areas: Oral Language (the average of Listening and Speaking), Literacy (the average of Reading and Writing), and Overall (a weighted average of all four domain scores). Composite scores are reported in 0.5 increments between integers (e.g., 1, 1.5, 2).

Table 1 shows the composite score weighting used for Screener.

Table 1: Calculation of Composite Scores

Type of Composite Score	Contribution of Language Domains (by %)			
	Listening	Reading	Writing	Speaking
Oral Language	50%	--	--	50%
Literacy	--	50%	50%	--
Overall	15%	35%	35%	15%

Purpose of this study

The purpose of the study described here was to estimate the predictive relationship between Screener and ACCESS. More specifically, the study addressed the following question: To what extent do Overall composite scale scores on Screener predict Overall composite scale scores on ACCESS?

Methods

To address this question we used data from Grades 1-12 English learner (EL) students who took both Screener and ACCESS within the same academic year. First, we estimated grade-level correlations between Screener and Access Overall composite scale scores. Scale scores are used for this estimation as they are interval variables, while proficiency level scores are ordinal. Next, to account for the possible effect of multiple student-, school-, district-, and state-level variables, multiple linear models with increasing complexity and more flexible covariance structures were fitted to the data to more reliably estimate the predictive relationship between Screener and ACCESS.

The model selected on the basis of best fit and most relevant covariates included both random and fixed state-, district-, school-, and individual-level effects. More specifically, the model included:

- student-level demographic covariates (IEP status, gender, ethnicity [Hispanic or not])
- a variable *interval*, measuring the time (in days) between Screener and ACCESS administrations¹
- grade-level fixed effects
- state-, district-, and school-level random effects

Data

As shown in Table 2, the analysis is based on test scores from 35,017 Grade 1-12 EL students enrolled in 10,300 schools, 2,658 districts, and 35 states who took both Screener and ACCESS within the same academic year. The average interval between the administration of the two tests for this sample was 144 days, with a standard deviation of about 35. Just over half of the sample was identified as Hispanic, 48% of the students were female, and 4% of the students had an IEP designation.

Table 3 presents the distribution of students by grade. Tables 4-5 provide information on the distribution of schools, districts, and states in the study. Table 4 shows the breakdown of schools by type (elementary, secondary, and high). Table 5 shows the distribution of states, districts, and schools by relative size (small, medium, and large), based on the number of students. For example, nine of the states had fewer than 100 students in the sample and are therefore categorized as small.

¹ In addition to the requirement that the tests be taken within the same academic year, the variable Interval was further restricted to a maximum of 300 days.

Table 2: Descriptive Statistics: Participation

Participants	N
States	35
Districts	2,657
Schools	10,297
Students	35,017

Table 3: Descriptive Statistics: Distribution of Students by Grade

Grade	Frequency	%
1	1,655	5
2	4,242	12
3	3,901	11
4	3,512	10
5	3,282	9
6	3,113	9
7	2,868	8
8	2,505	7
9	4,239	12
10	2,436	7
11	1,846	5
12	1,418	4
Total	35,017	100

Table 4: Distribution of School by Type

School Type	N	%
Elementary (K-5)	5,223	51
Secondary (6-8)	2,704	26
High (9-12)	2,370	23
Total	10,297	100

Table 5: Distribution of States, Districts, and Schools by Relative Size

States				
	Small (<100)	Medium (100–1,000)	Large (>1,000)	Total
N	9	12	14	35
%	25.7	34.3	40.0	100%

Districts				
	Small (<10)	Medium (10–100)	Large (>100)	Total
N	1,940	679	38	2,657
%	73.0	25.6	1.4	100%

Schools				
	Small (<10)	Medium (10–50)	Large (>50)	Total
N	9,697	582	18	10,297
%	94.2	5.7	0.2	100%

Results and Discussion

Table 6 shows the estimated correlation between the Overall composite scale scores on Screener and on ACCESS for Grades 1–12 and across all grade levels. The correlations for all grades were high, ranging from a low of 0.67 for Grade 1 to a high of 0.86 for Grades 7–9. The correlation across all grade levels was estimated at $\rho = 0.85$.

Table 6: Correlations between Screener and ACCESS Overall Composite Scale Scores

Grade	Pearson Correlation Coefficient ²
1	0.67
2	0.76
3	0.81
4	0.81
5	0.84
6	0.83
7	0.86
8	0.86
9	0.86
10	0.82
11	0.78
12	0.77
Overall	0.85

² Pearson product-moment correlation coefficient

Figure 2 shows a scatterplot of the relationship between Overall composite scale scores on Screener and on ACCESS across all grades, with each dot representing an individual student’s result. There is a clear positive relationship between the students’ performance on the two tests, with a narrow dispersion of scores around the red line, which represents the best linear fit with an estimated intercept of 114.9 and a slope of 0.74.³ This further illustrates the strong relationship between students’ performance on the two tests.

Figure 2: Scatterplot of Screener and ACCESS Overall Composite Scale Scores

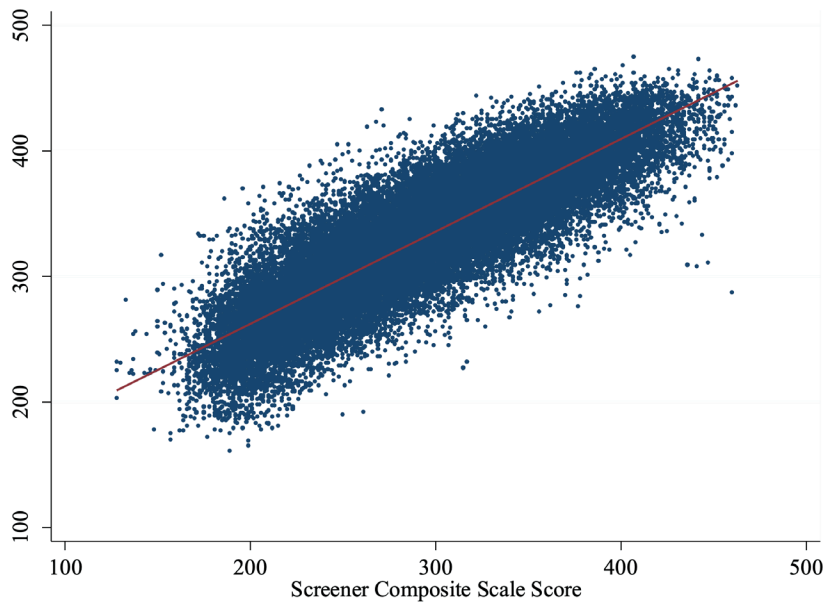


Table 7 presents the results of the linear regression model. The relationship between Screener and ACCESS is estimated controlling for:

- the interval (the number of days between the administration of Screener and the administration of ACCESS)
- individual-level factors (IEP status, female gender, Hispanic ethnicity)
- grade
- institutional-level effects (state, district and school)

Note that IEP status, female gender, and Hispanic ethnicity are indicator variables. The initial Screener score, the interval, and female gender all have statistically significant and positive parameter estimates. This indicates that higher scale scores on Screener, a longer time between the two test administrations, and female gender test performance are associated with higher scores on ACCESS. Conversely, students with a reported IEP status or Hispanic ethnicity have statistically significant and negative parameter estimates, indicating that these students, controlling for their initial Screener performance, other demographic variables, and school-, district-, and state-specific variation, scored lower than their no-IEP and not-Hispanic peers.

³ These estimates are obtained from a simple OLS regression of ACCESS scale scores on Screener scale scores. The adjusted R-squared for this model is estimated at 0.73.

Table 7: Mixed-Effects Model: Parameter Estimates

		Number of Observations = 35,013		
		Observations per Group		
Group Variable	Number of Groups	Minimum	Average	Maximum
State	35	21	1,000.40	3,837
District	2,657	1	13.2	1,064
School	10,297	1	3.4	154

Log Likelihood = -162,355.1

Wald χ^2 (16) = 90,289.35

Prob > χ^2 = 0.0000

Dependent variable: ACCESS Overall Composite Scale Score	Coefficient	Standard Error	z	P > z	[95% Confidence Interval]	
Screeners Scale Score	0.69	0.00	240.29	0.00	0.69	0.70
Interval	0.14	0.00	32.10	0.00	0.13	0.15
IEP Status	-7.93	0.68	-11.70	0.00	-9.26	-6.61
Female	2.28	0.27	8.60	0.00	1.76	2.80
Hispanic	-5.29	0.33	-16.17	0.00	-5.93	-4.65

Grade	Coefficient	Standard Error	z	P > z	[95% Confidence Interval]	
1 (baseline)						
2	0.31	0.77	0.40	0.69	-1.20	1.82
3	-2.08	0.78	-2.65	0.01	-3.61	-0.54
4	17.80	0.80	22.15	0.00	16.22	19.37
5	18.05	0.82	21.97	0.00	16.44	19.66
6	4.01	0.85	4.71	0.00	2.34	5.58
7	1.69	0.87	1.94	0.05	-0.02	3.40
8	-0.31	0.90	-0.35	0.73	-2.07	1.44
9	11.58	0.86	13.47	0.00	9.89	13.26
10	10.68	0.93	11.44	0.00	8.85	12.51
11	10.34	1.00	10.32	0.00	8.37	12.30
12	8.48	1.08	7.84	0.00	6.36	10.60

Random-Effects Parameters	Estimate	Standard Error	[95% Confidence Interval]	
State: Identity SD(constant)	2.72	0.50	1.90	3.89
District: Identity SD(constant)	6.18	0.32	5.58	6.84
School: Identity SD(constant)	9.00	0.24	8.54	9.49
SD(Residual)	23.45	0.10	23.25	23.65

Likelihood Ratio test versus linear model: χ^2 (3) = 2,358.63

Prob > χ^2 = 0.0000

The estimated grade fixed effects shown in Table 7, whenever statistically significant (all but Grade 2 and Grade 8), represent the average predicted differences in ACCESS Overall composite scale scores with respect to Grade 1 students. For example, the coefficient of -2.08 for Grade 3 students indicates that, on average, Grade 3 students score about two scale score points lower on ACCESS than do Grade 1 students, controlling for their Screener scores, the testing interval, individual-level factors, and state-, district-, and school-level random effects.

The estimated state, district, and school effects are all statistically significant, indicating that there is non-trivial variation in students' ACCESS scores due to these institutional factors. That is, the results suggest that in addition to the examined individual factors (grade, IEP status, female gender, and Hispanic ethnicity) and the time elapsed between the administration of Screener and the administration of ACCESS (testing interval), the state, district, and school students attend has an effect on their eventual performance on ACCESS.⁴

The final prediction equation is summarized as:

$$\text{ACCESS Scale Score} = 102.60 + 0.69 * \text{Screener Scale Score} + 0.14 * \text{Interval} - 7.93 * \text{IEP_Status} + 2.28 * \text{Female} - 5.29 * \text{Hispanic} + \text{Grade_Fixed-Effects} + \{ \text{State, District, and School} \} _ \text{Random-Effects} + \text{Random Error}$$

The parameter estimates provided in Table 3 indicate that, for example, 153 days (the number of days between September 1 and February 1) after her initial Screener administration, a randomly-selected, third-grade, Hispanic, female EL student with an IEP and a Screener Overall composite scale score of 285 (corresponding to a Screener proficiency level score of 2.5) would be expected, on average, to receive an ACCESS Overall composite scale score of approximately 308.

$$102.6 + 0.69 * 285 + 0.14 * 153 - 2.1 - 7.93 + 2.28 - 5.29 = 307.63 \text{ (with a 95\% confidence interval of 305.7-309.5)}^5$$

Similarly, another randomly-selected, Grade 4, non-Hispanic, male student without an IEP and with a Screener score of 285, would, 153 days after taking Screener, likely earn an ACCESS composite scale score of approximately 338.

$$102.6 + 0.69 * 285 + 0.14 * 153 + 17.80 = 338.47 \text{ (with a 95\% confidence interval of 337-340)}$$

Finally, if a randomly-selected student who is male, Hispanic, without an IEP, and in Grade 10 had a Screener score of about 300 in early September, 200 days later he would be expected to receive an ACCESS score of about 343.

$$102.6 + 0.69 * 300 + 0.14 * 200 - 5.28 + 10.68 = 343 \text{ (with a 95\% confidence interval of 341.3-344.7)}$$

4 The likelihood ratio test between the linear and random-effects specifications confirms the much better fit of the data when these effects are included in the model ($\chi^2(3) = 2,358.53$).

5 These results are descriptive and do not reflect a generalizable, causal relationship between Screener and ACCESS. The estimated parameters, while highly statistically significant for most variables, are only relevant for the sample of EL students used in this study.

Conclusion

The results of this analysis show that Screener scores are highly correlated with and strongly predictive of ACCESS scores, controlling for individual-level and institutional factors. Controlling for these factors, a one-point increase in the Screener Overall composite scale score is associated with about a 0.70 point increase in the predicted ACCESS Overall composite scale score. Students with special needs score about eight scale score points lower on ACCESS than students without IEPs. Hispanic ELs score about five points below non-Hispanic ELs, while female ELs score about 2.3 points higher than their male counterparts.

Parameter estimates indicate that in addition to individual-level covariates, there are statistically significant grade, school, district, and state effects; however, the predictive relationship between the two tests is not altered when these factors are accounted for.

These results provide evidence for the predictive validity of Screener for schools in the WIDA Consortium. The stable and strong predictive relationship between Screener and ACCESS Overall composite scale scores suggests that schools can confidently use the Screener Overall score specifically when making identification decisions for ELs.



Technical Report

No. TR-2020-1
February 2020

Wisconsin Center for Education Research
University of Wisconsin–Madison
1025 West Johnson St., MD #23
Madison, WI 53706

Client Services Center toll free:
(866) 276-7735

help@wida.us
wida.wisc.edu