WIDA™ RESEARCH REPORT

EXAMINING RELATIONSHIPS BETWEEN ALTERNATE ACCESS AND STATE ALTERNATE ASSESSMENTS: EXPLORING NOTIONS OF ENGLISH LANGUAGE PROFICIENCY

PREPARED BY
H. Gary Cook, Ph.D., Wisconsin Center for Education Research,
University of Wisconsin-Madison

DECEMBER 2014

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

WIDA advances academic language development and academic achievement for linguistically diverse students. WIDA was formed as the result of a federal grant to comply with the requirements of the No Child Left Behind Act. It is a consortium of states and districts working together to promote achievement of English language learners. The organization has created a comprehensive system that includes English Language Development Standards, Spanish Language Development Standards, English language proficiency assessments, professional development for educators of ELLs, and research on all aspects of English language learning.

RESEARCH

WIDA's Research Department seeks to provide timely, meaningful, and actionable research that promotes educational equity and academic achievement for linguistically and culturally diverse students. Its annual research agenda is developed under the guidance of the WIDA Consortium Board Research Subcommittee and includes topics in the areas of academic language, standards, professional learning, and policy.

The Alternate ACCESS for ELLs (Alt ACCESS) is a task-based assessment of English language proficiency (ELP) created by the WIDA Consortium (WIDA) for English learners (ELs) with significant cognitive disabilities. WIDA created the Alternate ACCESS for ELLs to meet federal accountability requirements and to provide educators with a measure sensitive to English language proficiency growth of ELs with significant cognitive disabilities (see www.wida.us/assessment/alternateaccess.aspx). This report examines how Alt ACCESS serves as a tool that identifies English proficiency attainment for English learners with significant cognitive disabilities.

When defining an English learner (Limited English Proficient student statutes), the No Child Left Behind Act states the following:

> LIMITED ENGLISH PROFICIENT– The term "limited English proficient," when used with respect to an individual, means an individual whose difficulties in speaking, reading, writing, or understanding the English language may be sufficient to deny the individual
>
> - ***the ability to meet the State's proficient level of achievement on State assessments described in section 1111(b)(3) [emphasis added];***
> - the ability to successfully achieve in classrooms where the language of instruction is English; or
> - the opportunity to participate fully in society (s.9101(25)(D)).

This statute highlights three characteristics that identify English learners: (a) their ability to perform proficiently on state content assessments, (b) their ability to be successful in the English-only classroom, and (c) their opportunity to participate in society. This report provides evidence examining Alt ACCESS's potential to meet the first characteristic (highlighted above).

About Alternate ACCESS for ELLs

The Alt ACCESS assessment has four cluster-based forms (Grades 1–2, 3–5, 6–8, and 9–12) in the four language domains of listening, speaking, reading and writing. Alt ACCESS was developed from and is aligned to WIDA's Alternate Modified Performance Indicators (AMPIs). For more information on WIDA's AMPIs, see www.wida.us/assessment/alternateaccess.aspx.

Alt ACCESS's receptive language (listening and reading) test items are selected response, and productive language (writing and speaking) test items are constructed response. The assessment is administered and scored locally. Once scored, test forms are then sent to an assessment vendor for processing and reporting.

Alt ACCESS provides six proficiency levels: A1-Initiating, A2-Exploring, A3-Engaging, P1-Entering, P2-Emerging, and P3- Developing (only in writing). Scores and proficiency levels are provided in four language domains (listening, speaking, reading and writing) and in four weighted composites (oral language, literacy, comprehension, and overall). Like ACCESS for ELLs, Alt ACCESS weights scale scores to create composite proficiency scores and proficiency levels. Composite weights are shown in the table below.

TABLE 1: ALT ACCESS SCORE WEIGHTING SCHEME

Composite	Percentage of Domain Weight			
	Listening	Speaking	Reading	Writing
Oral	50%	50%	--	--
Literacy	--	--	50%	50%
Comprehension	30%	--	70%	--
Overall	15%	15%	35%	35%

Research Questions

Cook, Linquanti, Chinen, and Jung (2012)[1] prepared a report for the U.S. Department of Education that examined several questions related to English learner accountability. One question, germane here, is "How do you determine a meaningful English language performance standard?" Cook et al. (2012) argued that one approach to answering this question was to explore the relationship between English language and academic content proficiency assessments. They provided three procedures for examining this relationship: a descriptive boxplot, decision consistency analyses, and logistic regression. Two of these procedures (descriptive boxplots and logistic regression) are used here to address two research questions:

1. What overall composite score(s) on Alt ACCESS represents a potential "English proficiency" criterion?
2. What is an optimal domain score weighting formula for creating an overall composite score on Alt ACCESS?

This research aims to provide guidance to states in AMAO determinations for ELs participating in Alt ACCESS and to provide feedback on the efficacy of the Alt ACCESS overall composite weighting formula.

Methodology

Subjects

Four WIDA states provided Alt ACCESS and state alternate assessment (in English language arts [ELA] or reading, and mathematics) data for all participating ELs for the 2012–2013 school year. Analyses reported here are "census" data for each state. Note that for each student record, only a state identifier, grade, Alt ACCESS, and ELA or reading, and mathematics alternate assessment score was provided. No other demographic variables were requested or provided by states. The tables below display the numbers of students provided by states and used for analyses. Note that one WIDA state had only 23 records across all grades. That state is not shown in tables displaying state information.[2]

[1] Cook, H. G., Linquanti, R., Chinen, M., & Jung, H. (2012). *National Evaluation of Title III Implementation Supplemental Report: Exploring approaches to setting English language proficiency and performance criteria and monitoring English learner progress.* Washington, DC: U.S. Department of Education. Retrieved from http://www2.ed.gov/rschstat/eval/title-iii/implementation-supplemental-report.html.

[2] The total number of ELs in this state is too small for public reporting, although student records from this state are used in aggregate reporting and analyses.

TABLE 2: ALT ACCESS DATASET PROVIDED BY WIDA STATES BY GRADE

Grade	State			Total
	A	B	C	
1	82	46	.	130
2	82	61	.	144
3	107	49	35	195
4	106	44	42	193
5	112	34	18	164
6	89	36	38	163
7	89	27	31	148
8	59	25	26	112
9	60	12	.	75
10	49	19	20	91
11	54	14	.	73
12	58	15	.	74
Total	947	382	210	1562

TABLE 3: ALT ACCESS AND ALTERNATE ASSESSMENT DATASET USED FOR ANALYSES[3]

Grade	State			Total
	A	B	C	
3-5	315	130	93	538
6-8	219	85	88	392
9-12	106	18	23	147
Total	640	233	204	1077

[3] Grades are clustered in this table to prevent suppression rule violations, i.e., individual cells with less than 10 student records.

Materials: Alternate Assessments

Each state provided Alt ACCESS scores (described above) and state ELA or reading, and mathematics alternate assessments scores. Note that state alternate assessments are provided for students with significant cognitive disabilities, whose educational goals are not the same as their noncognitively disabled student peers. By federal statute, alternate assessments must be provided in reading or ELA, and mathematics and in Grades 3–8, and once in high school (§.1111(b)(3)(v)(I)[4]). One state has alternate assessment records in 9th and 10th grade. The other states have student alternate assessment scores in 10th or 12th grade. Score ranges for these assessments differ (as might be expected) and scaling methodologies vary. To establish comparability across states, the alternate assessment scores were standardized (using t-score transformations[5]) within grade for each state.

Procedures for Handling Data

Alt ACCESS and alternate assessment data were securely uploaded and entered into the WIDA data warehouse. Student records had IDs and names redacted, and each record received an anonymized identifier. Researchers used the SAS statistical software program for analyses. This statistical software is housed at the Social Science Computing Collaborative (SSCC) Citrix web client at the University of Wisconsin. SSCC allows researchers to conduct statistical analyses on a secure server, which prevents any data from being exposed to external sources. Further, WIDA research team's procedures restrict the creation of secondary data files. All research files were created using a PROC SQL call and filed in SAS's "Work" library. Once analyses were complete and the SAS program closed, all secondary datasets were deleted, since they reside in cache memory.

Methods of Analyses

The first research question seeks to determine Alt ACCESS composite scores that represent an English proficiency criterion. We used two approaches to explore this question. First, we examined descriptive statistical analyses and graphic representations between Alt ACCESS scores and alternate assessment. The goal of these analyses was to identify any patterns that emerged in the relationship between ELP and content alternate assessment scores. The second approach employed logistic regression, as outlined in Cook, et al. (2012), to examine a potential English language proficiency criterion.

The second research question examines optimal composite score weighting for Alt ACCES. With logistic regression approach, models are developed using domain tests as predictors and proficiency on state alternate assessments as outcomes. It then compares each domain tests' odds ratio estimate to identify potential weights.

---

[4] No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

[5] For t-score transformation, see chapter 7 of Brown, J.D. (1988). *Understanding research in second language learning: a teacher's guide to statistics and research design.* NY: Cambridge University Press.

Research Question #1: What Overall Composite Score(s) on Alt ACCESS Represents a Potential "English proficiency" Criterion?

Results and Discussion

Descriptive analyses. Table 4 and Figure 1 show the number and percentage of EL students who met (proficient) or did not meet (not proficient) state alternate assessment performance criteria in ELA (for three WIDA states) or reading (one WIDA state), and mathematics (all four WIDA states). Note, State A does not have an ELA alternate assessment. It has separate reading and writing alternate assessments. For this analysis, only the reading alternate will be reported for this state, and it is presumed to provide similar information to that of state B's and C's ELA assessments.

TABLE 4: FREQUENCIES AND PERCENTAGES OF ELS MEETING OR NOT MEETING STATE ALTERNATE ASSESSMENT PROFICIENCY DESIGNATION BY ALT ACCESS COMPOSITE PROFICIENCY LEVELS IN 4 WIDA STATES[6]

Alt ACC Level	English Language Arts*				Mathematics			
	Not Met		Met		Not Met		Met	
	N	Percent	N	Percent	N	Percent	N	Percent
P2	6	10.0	54	90.0	51	23.5	166	76.5
P1	30	22.2	105	77.8	122	33.3	244	66.7
A3	62	56.4	48	43.6	151	67.1	74	32.9
A2	37	82.2	8	17.8	88	89.8	10	10.2
A1	70	94.6	4	5.4	134	95.0	7	5.0
Overall	205	48.3	219	51.7	546	52.1	501	47.9

Alt ACC Level	Reading**			
	Not Met		Met	
	N	Percent	N	Percent
P2	77	48.1	83	51.9
P1	157	66.2	80	33.8
A3	113	95.0	6	5.0
A2	52	94.5	3	5.5
A1	77	98.7	1	1.3
Overall	476	73.3	173	26.7

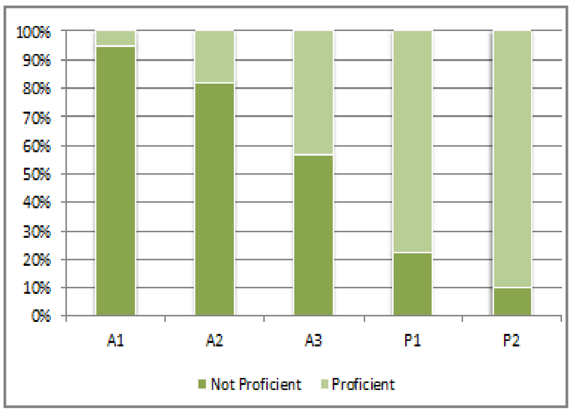
* Data from 3 states, ** Data from 1 state

[6] Note that totals in Table 3 and Table 4 differ slightly due to off-grade testing, i.e., some ELs were provided the alternate assessment in Grades 1 and 2, primarily in State A.

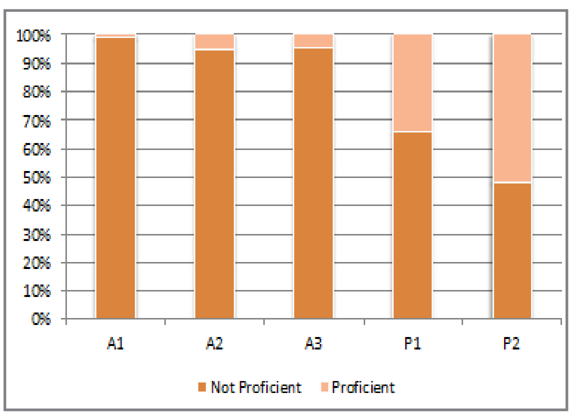
Table 4 and Figure 1 show that as Alt ACCESS proficiency levels increase, the percentage of students receiving proficient scores on the alternate assessments also increases. This trend is consistent with what Cook, et al. (2012) found in their research. The increase in the percentage of students attaining proficient scores on alternate assessments as English proficiency increases is an indicator that language is having less and less of an impact on student content performance.

FIGURE 1: STACKED BAR CHART OF ELS MEETING AND NOT MEETING STATE ALTERNATE ASSESSMENT PERFORMANCE LEVELS IN 4 WIDA STATES BY ALT ACCESS PL

English Language Arts



Reading



Mathematics

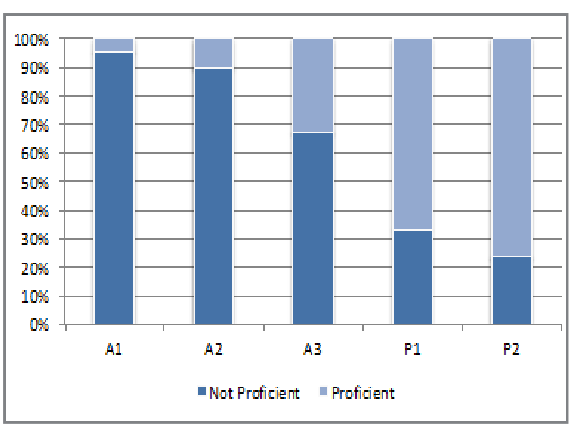
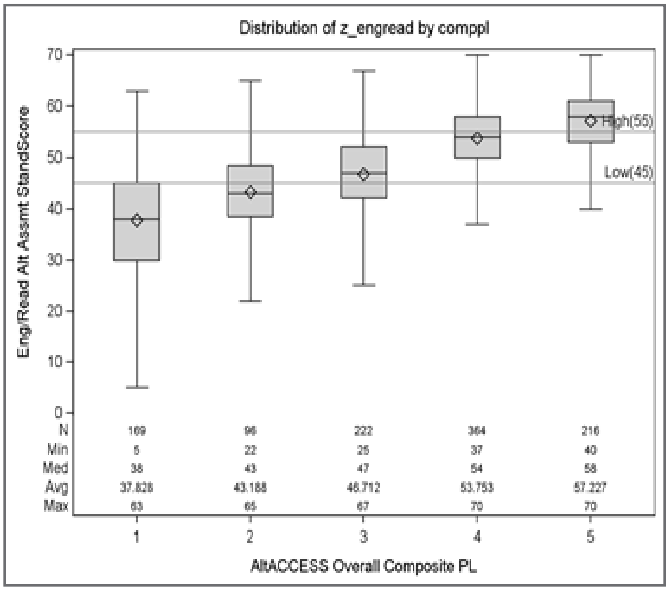
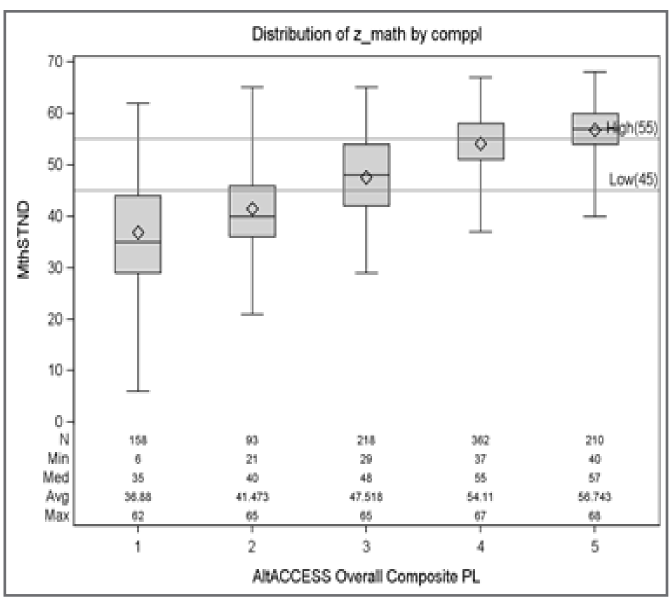


FIGURE 2: BOXPLOTS OF STATE ALTERNATE ASSESSMENT STANDARDIZED SCALE SCORE DISTRIBUTION BY ALT ACCESS OVERALL COMPOSITE PROFICIENCY LEVEL IN 4 WIDA STATES

English Language Arts and Reading



Mathematics



Note. On x-axis, number values are as follows: 1=A1, 2=A2, 3=A3, 4=P1, and 5=P2.

Figure 2 displays the relationship between Alt ACCESS and state alternate assessments in boxplot form. To create a common scale between states' alternate assessment scores, we conducted a t-score transformation.[7]

The horizontal lines on the two boxplots display the range of "proficient" across state alternate assessments (in t-score units). For both ELA and reading, and for mathematics, t-scores that represent proficiency across states ranged between 45 and 55. Two things can be immediately observed from these plots. First, as Alt ACCESS proficiency level increases (x-axis), the number of students attaining proficiency on alternate assessments also increases. Second, a majority of students taking Alt ACCESS attain proficiency on state alternate assessments between the A3 and P2 proficiency levels. That is, the median t-score ("Med" in the statistics section at the bottom of the graph) for the Alt ACCESS proficiency level goes above horizontal lines which identify alternate assessment proficiency rages, between A3 and P2 on both boxplots.

In sum, the descriptive analyses identify two major findings. As Alt ACCESS proficiency levels increase, the number of students attaining proficiency on state alternate assessments also increases. Second, a majority of students attain proficiency on ELA and reading, or on mathematics assessments between standardized t-scores of 45 to 55, depending on the state.

Logistic Regression. Tables 5 through 7 display univariate logistic regression analysis results between Alt ACCESS composite scale scores and proficiency categorizations on state alternate assessments. To interpret these results, we examined the model's fit and parameter estimates. Model fit explores whether a statistical model has any predictive power, and nonsignificant model fit tests (e.g., Likelihood Ratio test) indicate no predictive power. If evaluation tests are significant ($p<.05$), parameter estimates are then examined. All overall model fit tests for regressions in Tables 5–7 had probabilities less than 0.05. This suggests that parameters in the logistic regression models are exhibiting some predictability.

The next step was to examine the parameter estimates. The parameter of interest here is the Alt ACCESS overall composite scale score *(Composite).* In all three tables, the Alt ACCESS overall composite scale score is a significant predictor of proficiency on state alternate assessments. In other words, the Alt ACCESS overall composite scale score significantly predicts a student's likelihood of being proficient on state ELA, reading, or mathematics alternate assessments.

[7] T-score transformation is a standardization procedure where $t = 10z + 50$, and $z = (x-\bar{x})/SD$.

TABLE 5: LOGISTIC REGRESSION ANALYSIS OF PROFICIENCY ON ELA ALTERNATE ASSESSMENT AND ALT ACCESS PERFORMANCE BY OVERALL COMPOSITE IN 3 WIDA STATES

Parameter	*df*	Estimate	Standard Error	Wald Chi-Square	p	Odds ratio
Intercept	1	-171.3	18.3524	87.1402	<.0001	NA
Composite	1	0.1831	0.0196	87.4673	<.0001	1.201
Overall model fit						
Test				**Chi-Square**	***df***	**p**
Likelihood Ratio				175.2993	1	<.0001
Score				141.2633	1	<.0001
Wald				87.4673	1	<.0001

Note. SAS programming codes: [proc logistic; model sc_english_pl_proficient (event='1') = composite / pevent=0.5 rsquare lackfit;] *Cox & Snell* R^2 = 0.3386, *Nagelkerke* R^2 *(Max-rescaled* R^2*)* = 0.4517, *Kendall's Tau-a* = 0.349, *Goodman-Kruskal Gamma* = 0.715, *Somers' D* = 0.698, *c-statistic* = 0.849.

TABLE 6: LOGISTIC REGRESSION ANALYSIS OF PROFICIENCY ON READING ALTERNATE ASSESSMENT AND ALT ACCESS PERFORMANCE BY OVERALL COMPOSITE IN 1 WIDA STATE

Parameter	*df*	Estimate	Standard Error	Wald Chi-Square	p	Odds ratio
Intercept	1	-203.8	24.4705	69.3602	<.0001	NA
Composite	1	0.2155	0.0260	68.9423	<.0001	1.240
Overall model fit						
Test				**Chi-Square**	***df***	**p**
Likelihood Ratio				148.0090	1	<.0001
Score				90.5869	1	<.0001
Wald				68.9423	1	<.0001

Note. SAS programming codes: [proc logistic; model sc_reading_pl_proficient (event='1') = composite / pevent=0.5 rsquare lackfit;] *Cox & Snell* R^2 = 0.2039, *Nagelkerke* R^2 *(Max-rescaled* R^2*)* = 0.2971, *Kendall's Tau-a* = 0.233, *Goodman-Kruskal Gamma* = 0.630, *Somers' D* = 0.596, *c-statistic* = 0.798.

TABLE 7: LOGISTIC REGRESSION ANALYSIS OF PROFICIENCY ON MATHEMATICS ALTERNATE ASSESSMENT AND ALT ACCESS PERFORMANCE BY OVERALL COMPOSITE IN 4 WIDA STATES

Parameter	*df*	Estimate	Standard Error	Wald Chi-Square	p	Odds ratio
Intercept	1	-157.0	11.8134	176.5200	<.0001	NA
Composite	1	0.1672	0.0126	176.8901	<.0001	1.182
Overall model fit						
Test				**Chi-Square**	***df***	**p**
Likelihood Ratio				333.9284	1	<.0001
Score				261.8804	1	<.0001
Wald				176.8901	1	<.0001

Note. SAS programming codes: [proc logistic; model sc_math_pl_proficient (event='1') = composite / pevent=0.5 rsquare lackfit;] *Cox & Snell* R^2 = 0.2731, *Nagelkerke* R^2 *(Max-rescaled* R^2*)* = 0.3643, *Kendall's Tau-a* = 0.304, *Goodman-Kruskal Gamma* = 0.629, *Somers' D* = 0.608, *c-statistic* = 0.804.

It is important to know that Alt ACCESS scores are predictive of alternate assessments, as required by federal statute (NCLB: §.3113(b)(2) and §.9101(25)(D)(i)). But what can be done with this information? We can use logistic regression analyses to support establishing an English proficient performance criterion(ia) on Alt ACCESS.

Cook et al. (2012) wrote,

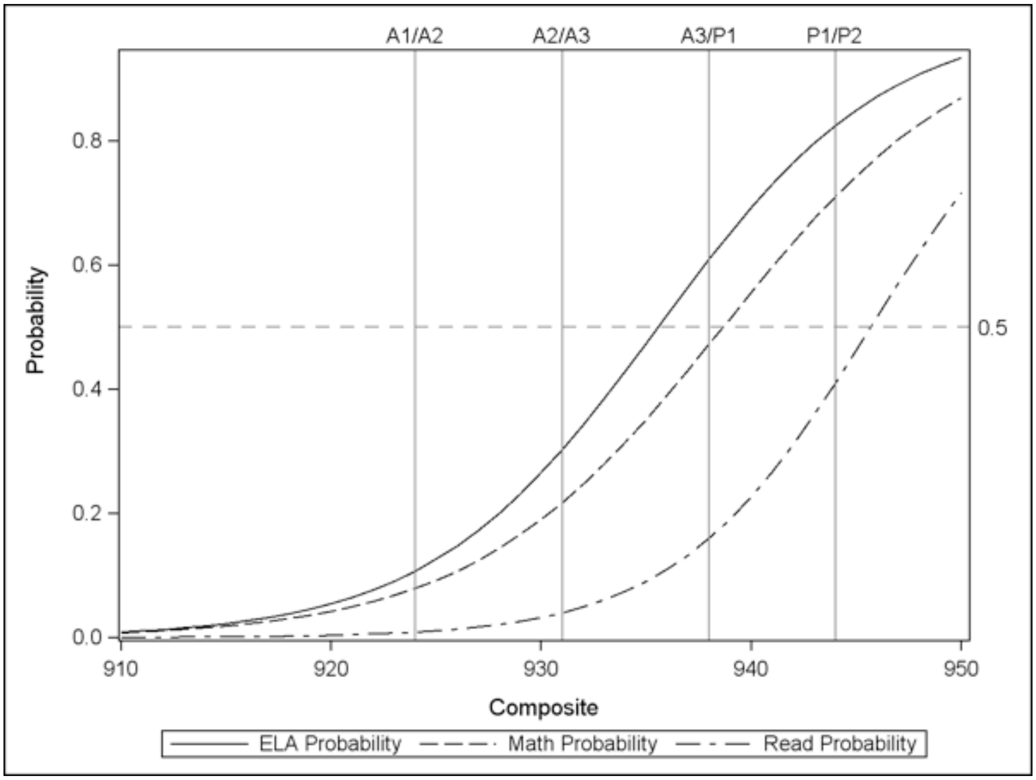
> Empirically, researchers can define "English language proficient" as the point at which EL students' academic content achievement assessed using English becomes *less* related to their ELP. That is, there is a point at which EL students have sufficient English language skills to adequately function in English on content assessments. (p.8)

With logistic regression, these researchers conceptualized the English proficient point as a 50-50 criterion or a probability of 50% (.50). They argued "students with ELP scores at this level have an equal likelihood of attaining content-area-proficient performance" (Cook et al. (2012), p. 8, footnote). Scores at or above the 50% probability indicate that something more than language proficiency is determining how well students perform on content tests, presumably content knowledge.

We can take results from Tables 5–7 and calculate where the 50% probability lies.[8] Figure 3 displays logistic probability curves from above regression analyses, and the 50% probability level is identified by a horizontal broken line. The vertical lines show current Alt ACCESS overall composite proficiency level cutpoint barriers (A1/A2 = 924, A2/A3 = 931, A3/P1 = 938, and P1/P2 = 944).

[8] The Composite parameter estimates in Tables 5-7 require transformation to identify probabilities. Univariate logistic regression takes the form p= $\frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}$. The intercept and Composite parameter estimates can be inserted into this formula to identify Composite scale score probabilities.

FIGURE 3: LOGISTIC REGRESSION CURVES IDENTIFYING THE RELATIONSHIP BETWEEN ALT ACCESS COMPOSITE SCALE SCORES AND 3 STATES' ALTERNATE ASSESSMENT PROFICIENCY CATEGORIZATIONS



The three curves in the figure above are from the four aforementioned WIDA states.

- The leftmost curve identifies the relationship between Alt ACCESS composite scale scores and state ELA alternate assessments for three of the WIDA states.
- The middle curve identifies the relationship between Alt ACCESS composite scale scores and state mathematics alternate assessments. All four analyzed states are included in this curve.
- The rightmost curve identifies the relationship between Alt ACCESS and State A's reading alternate assessment.

For all three curves, the point at which the curve goes above the horizontal dotted line represents a probability of 50% or greater. Interestingly, the subject area that crossed the 50% probability line at the lowest Alt ACCESS score was ELA. The ELA curve crossed over the 50% line in the middle of Alt ACCESS levels A3. The mathematics alternate assessment curve crossed over the 50% line just above the P1 Alt ACCESS level, Additionally, State A's reading alternate assessment crossed over the 50% line well into the P2 level.

What this suggests is proficiency on State A's reading alternate assessment requires the higher level of academic language, as measured by Alt ACCESS, when compared to the mathematics and ELA alternate assessments. Surprisingly, the ELA alternate assessment requires the lowest Alt ACCESS score to go above the 50% probability line.

We can think of several reasons for this. One might be that the cognitive demand of the mathematics and reading assessments are greater than that of the ELA alternate assessment. Another more plausible reason for the observed curves might be in aggregating across states. That is, different states' alternate assessments might have different relationships to Alt ACCESS, and depending on sample size, this may over influence the probability curves. To explore this, we examined an analysis of each state's alternate assessment relationship to Alt ACCESS. Results of this analysis are presented below in the form of logistic regression curves.

FIGURE 4: LOGISTIC REGRESSION CURVES IDENTIFYING THE RELATIONSHIP BETWEEN ALT ACCESS COMPOSITE SCALE SCORES AND STATES' ELA OR READING ALTERNATE ASSESSMENT PROFICIENCY CATEGORIZATIONS

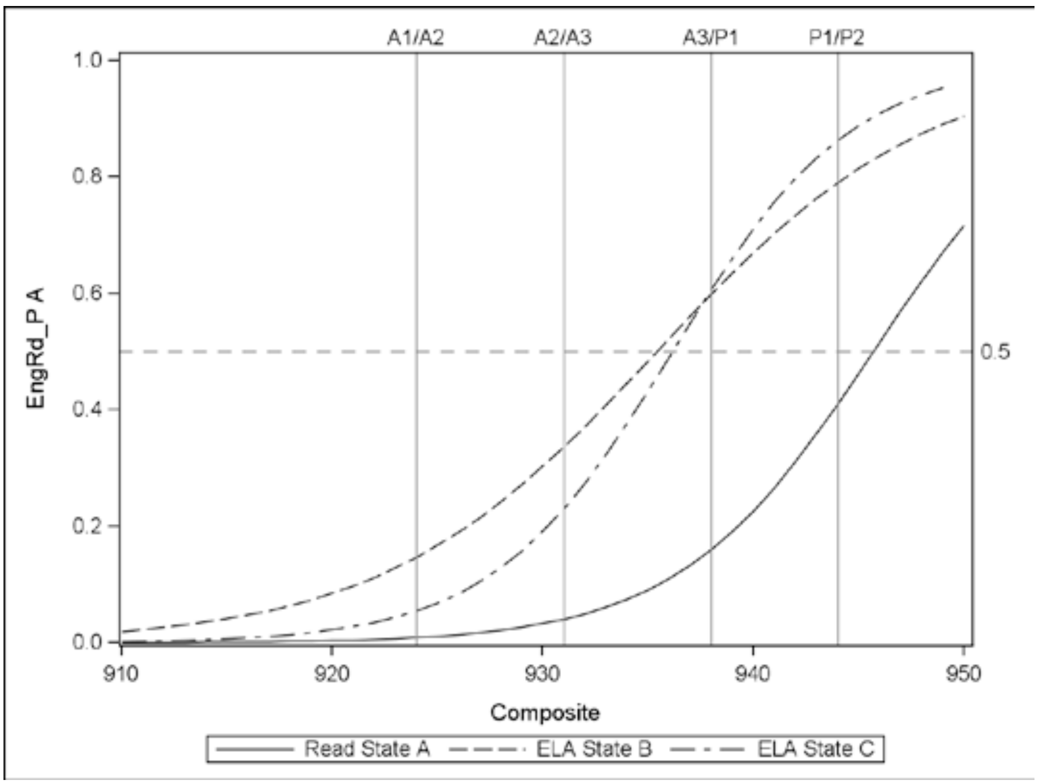
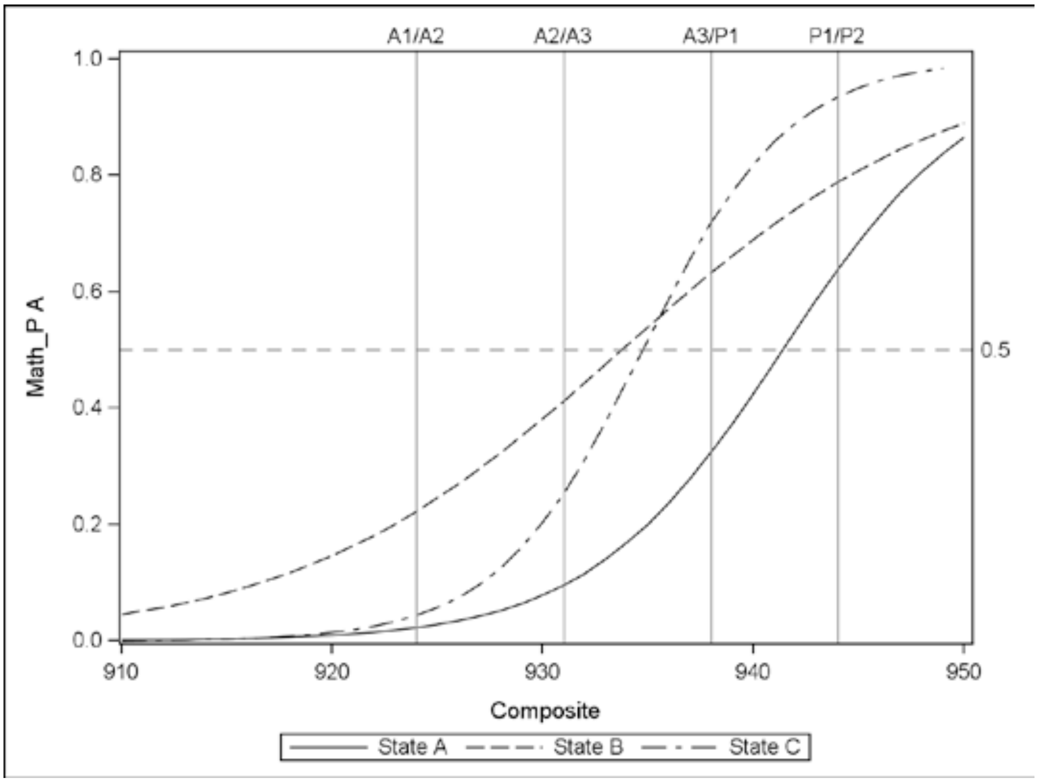


Figure 4 shows the relationship between state ELA or reading alternate assessments and Alt ACCESS composite scale scores. Like in Figure 3, vertical lines represent Alt ACCESS proficiency level cut points, and the horizontal dotted line is at the 50% probability level. Again we see that State A's reading alternate assessment crosses the 50% probability line well into Alt ACCESS's P2 proficiency level. For States B and C, the point where ELA curves across the 50% probability level is at Alt ACCESS level A3. States B and C cross this 50% line at very similar scale scores. Clearly, the academic English language required to be proficient on State A's reading alternate assessment is much higher than that required of either State B or State C's ELA alternate assessments.

Figure 5 below displays the relationship between states' mathematics alternate assessments and Alt ACCESS. Unlike the curves in Figure 4 that show two different types of content tests (ELA and reading), Figure 5 shows logistic regression curves for the same content area, mathematics.

FIGURE 5: LOGISTIC REGRESSION CURVES IDENTIFYING THE RELATIONSHIP BETWEEN ALT ACCESS COMPOSITE SCALE SCORES AND STATES' MATHEMATICS PROFICIENCY CATEGORIZATIONS



Here again we see that State A requires a higher Alt ACCESS score to cross the 50% probability line. Note, however, that the point where the mathematics curve crosses the 50% line is less in mathematics than it is for reading in State A. This is what we would expect to see, i.e., more academic language is needed for the reading alternate assessment than for mathematics. States B and C's mathematics assessments cross the 50% line at a lower Alt ACCESS score. As with State A, the mathematics alternate assessments for States B and C cross the 50% line at a lower Alt ACCESS score compared to the ELA alternate assessments. Also note that the slopes on the logistic regression curves reflect the accuracy in predicting relationship between Alt ACCESS and state content alternate assessments—the steeper the slope, the more accurate the prediction.

Summary

From both descriptive and box plot analyses, observable relationships between Alt ACCESS and states' ELA or reading and mathematics alternate assessments are clearly identified. As English language proficiency increases, the proportion of students attaining proficiency on state alternate assessments also increases. Cook et al. (2012) argued that English language proficiency is the language proficiency level where students are equally likely to be proficient on state content assessments. In descriptive and box plot analyses, we observe this somewhere between the A3 and P2 Alt ACCESS proficiency levels.

Findings from the logistic regression analysis further clarify this observation. Different states exhibit different relationships between Alt ACCESS and alternate assessments when analyzed via logistic regression. Sometimes this difference is substantial; hence, it is prudent for each individual state to establish English language proficiency on Alt ACCESS independently, taking their unique alternate assessments into account. The procedures forwarded by Cook et al. (2012) seem to fit well in supporting the development of English language proficiency levels on Alt ACCESS.

Research Question #2: What is an Optimal Domain Score Weighting Formula for Creating an Overall Composite Score on Alt ACCESS?

Results and Discussion

The analyses presented here outline an empirical method for identifying optimal domain weights on Alt ACCESS as it relates to state ELA and mathematics alternate assessments. It is important to note that ultimately, the selection of domain weights for Alt ACCESS is a policy rather than an empirical decision. Particular domain weights might differ for many different reasons. For example, policymakers may wish to emphasize one domain over another for instructional reasons. This analysis uses state alternate assessments as outcomes. Policy makers may want Alt ACCESS to predict other meaningful outcome variables. Analyses presented here are intended merely to inform potential domain weighting possibilities.

Identifying optimal domain weights. To identify optimal domain weights, we used a logistic regression approach, similar to that used for the first research question. The difference here is each domain scale score was used as a predictor. The goal was to determine the contribution each domain score had in predicting proficiency on state alternate assessments. Tables 8 and 9 display results from logistic regression analyses between Alt ACCESS domain scores and state ELA and mathematics alternate assessments.

The first step in interpreting results was to identify model fit. Model fit in both tables are observed to be significant (i.e., Likelihood Ratio, Score, Wald Tests $p<.05$). The next step examined each Alt ACCESS domain score's significance value. In both the ELA and mathematics logistic regression models, all domain scores are significant predictors of alternate assessment proficiency except listening. In the states, Alt ACCESS's listening domain score does not seem to significantly contribute to proficiency on ELA or mathematics alternate assessments. This does not mean that the listening domain test on Alt ACCESS is not important. It only indicates that listening does not serve as a significant predictor to proficiency on alternate assessments in these states. It does suggest that the listening domain score should be one of the least weighted components of an Alt ACCESS overall composite score if predicting to alternate assessment proficiency is important. A quick glance at Table 1 shows that the listening domain score is one of the least weighted components of the overall composite.

TABLE 8: LOGISTIC REGRESSION ANALYSIS OF PROFICIENCY ON ELA ALTERNATE ASSESSMENT AND ALT ACCESS PERFORMANCE BY DOMAINS IN 3 WIDA STATES BY SAS PROC LOGISTIC

Parameter	*df*	Estimate	Standard Error	Wald Chi-Square	p	Odds ratio
Intercept	1	-178.1	21.1665	70.7935	<.0001	NA
aa_listening_scale_s	1	0.0422	0.0295	2.0364	0.1536	1.043
aa_speaking_scale_sc	1	0.0477	0.0166	8.2234	0.0041	1.049
aa_reading_scale_sco	1	0.0613	0.0246	6.2392	0.0125	1.063
aa_writing_scale_sco	1	0.0390	0.0175	4.9836	0.0256	1.040
Overall model fit						
Test				**Chi-Square**	***df***	**p**
Likelihood Ratio				178.7962	4	<.0001
Score				143.3205	4	<.0001
Wald				84.1304	4	<.0001

Note. SAS programming codes: [proc logistic; model sc_english_pl_proficient (event='1') = aa_listening_scale_score aa_speaking_scale_score aa_reading_scale_score aa_writing_scale_score / pevent=0.5 rsquare;] *Cox & Snell* R^2 = 0.3441, *Nagelkerke* R^2 *(Max-rescaled* R^2*)* = 0.4589, *Kendall's Tau-a* = 0.357, *Goodman-Kruskal Gamma* = 0.714, *Somers' D* = 0.713, *c-statistic* = 0.856.

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
aa_listening_scale_s	1.043	0.984	1.105
aa_speaking_scale_sc	1.049	1.015	1.084
aa_reading_scale_sco	1.063	1.013	1.116
aa_writing_scale_sco	1.040	1.005	1.076

Weighting domain scores. How should other domain scores be weighted, given that they are all significant predictors? Examining the odds ratio statistics in Tables 8 and 9 can help answer this question. The odds ratio is a statistic that describes the relationship between a predictor and an outcome in logistic regression. It can range from zero to infinity. An odds ratio of 1.0 indicates no relationship between the predictor and the outcome. Odds ratios greater than 1.0 reflect stronger relationships.

Isn't the listening domain test odds ratio greater than 1.0 in Tables 8 and 9? Yes (1.043 in Table 8 and 1.081 in Table 9), but notice another panel at the bottom of each table. This panel displays odds ratios and their confidence intervals (limits). If the confidence interval goes through 1.0, that variable is considered nonsignificant. In both Table 8 and Table

9, the confidence band for the listening domain test goes through 1.0. The other domain tests do not, which signifies a significant relationship.

TABLE 9: LOGISTIC REGRESSION ANALYSIS OF PROFICIENCY ON MATHEMATICS ALTERNATE ASSESSMENT AND ALT ACCESS PERFORMANCE BY DOMAINS IN 4 WIDA STATES BY SAS PROC LOGISTIC

Parameter	*df*	Estimate	Standard Error	Wald Chi-Square	p	Odds ratio
Intercept	1	-175.0	14.9563	136.9297	<.0001	NA
aa_listening_scale_s	1	0.0363	0.0210	3.0020	0.0832	1.081
aa_speaking_scale_sc	1	0.0783	0.0134	33.9253	<.0001	1.037
aa_reading_scale_sco	1	0.0368	0.0173	4.4999	0.0339	1.035
aa_writing_scale_sco	1	0.0348	0.0101	11.8815	0.0006	1.037
Overall model fit						
Test				**Chi-Square**	***df***	**p**
Likelihood Ratio				357.1896	4	<.0001
Score				272.5220	4	<.0001
Wald				163.1275	4	<.0001

Note. SAS programming codes: [proc logistic; model sc_math_pl_proficient (event='1') = aa_listening_scale_score aa_speaking_scale_score aa_reading_scale_score aa_writing_scale_score / pevent=0.5 rsquare;] *Cox & Snell* R^2 = 0.2891, *Nagelkerke* R^2 *(Max-rescaled* R^2*)* = 0.3856, *Kendall's Tau-a* = 0.316, *Goodman-Kruskal Gamma* = 0.636, *Somers' D* = 0.633, *c-statistic* = 0.817.

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
aa_listening_scale_s	1.037	0.995	1.081
aa_speaking_scale_sc	1.081	1.053	1.110
aa_reading_scale_sco	1.037	1.003	1.073
aa_writing_scale_sco	1.035	1.015	1.056

Using the odds ratio values, we can identify which domain tests have higher associations with state ELA or mathematics alternate assessment proficiency. For the ELA alternate assessment, the reading domain test has the highest odds ratio. This isn't surprising. The next highest odds ratio is speaking, which is somewhat of a surprise. The lowest significant predictor, and the one with the lowest odds ratio, is writing. Were we to create a weighting scheme for the Alt ACCESS using just the ELA alternate assessment, the reading domain test would be waited the most, and a listening domain test would be weighted the least. The speaking domain test on Alt ACCESS has the highest odds ratio for mathematics. The weighting order in mathematics would be speaking, reading, writing, and listening.

The odds ratio identifies which predictors have higher associations with the outcome. They do not describe the magnitude of those relationships. A procedure that can assist in understanding the magnitude of these relationships is stepwise regression. In stepwise regression, each predictor variable (in this case domain scores) is examined independently. The predictor that has the greatest relationship to the outcome is placed in the regression model first. The procedure then identifies the variable that has the next highest relationship and adds it to the regression model. This continues until all variables are selected.

TABLE 10: RESULTS FROM A STEPWISE LOGISTIC REGRESSION BETWEEN ALT ACCESS AND STATE ELA AND MATHEMATICS ALTERNATE ASSESSMENTS

Stepwise Logistic Regression Results ELA					
Step	**Parameter**	***df***	**Wald Chi-Square**	**p**	**Max-rescaled R-Square**
1	aa_reading_scale_sco	1	122.0509	<.0001	0.3944
2	aa_speaking_scale_sc	1	21.9103	<.0001	0.4430
3	aa_writing_scale_sco	1	5.7663	0.0163	0.4546
4	aa_listening_scale_s	1	2.0406	0.1531	0.4589
Step	**Parameter**	***df***	**Wald Chi-Square**	**p**	**Max-rescaled R-Square**
1	aa_speaking_scale_sc	1	241.2500	<.0001	0.3365
2	aa_reading_scale_sco	1	36.0339	<.0001	0.3715
3	aa_writing_scale_sco	1	12.7193	0.0004	0.3829
4	aa_listening_scale_s	1	3.0143	0.0825	0.3856
Note. SAS PROC LOGISTIC stepwise code options: [selection=stepwise slentry=0.3 slstay=0.35 details lackfit]					

Table 10 shows results from a foreword selection stepwise regression analysis of Alt ACCESS and state alternate assessments. The table has six columns. The first column displays the step in the stepwise regression procedure. The next column identifies the variable (parameter) associated with that step. This is followed by the degrees of freedom. Columns 4 and 5 display the statistical test (Wald chi-square) and the p-value for each step. The last column identifies the R^2 value. The Max-rescaled R-Square is a statistic that identifies the strength or magnitude of the relationship between the predictor variable(s) and the outcome.[9]

[9] There is no generally accepted analog to R-Square for logistic regression models. A variety of procedures have been suggested in the literature. For more detail see Allison, P. D. (2014, March). *Measures of fit for logistic regression.* Paper 1485-2014 presented at SAS Global Forum, Washington, DC. Retrieved from http://support.sas.com/resources/papers/proceedings14/1485-2014.pdfHowever, Allison (2014) notes that the Max Rescaled R-Square method may overestimate predictive power of logistic models. That was not a major concern here since the intent was to identify the relative predictive relationship between domain tests not the overall predictive power of the logistical model.

Step 1 of the ELA regression analysis identifies the reading domain score as the strongest predictor. This is not a surprise. The reading domain score captures 39.4% (0.3944) of the variability in the likelihood of being proficient in ELA. Step 2 adds another variable. For the ELA analysis that variable is speaking. When the speaking domain score is added, the R-square value increases to 44.3%. Adding the speaking domain score to the regression model increased the predictive likelihood by 5%. When all four domain tests are added to the regression model, the predictive likelihood is 45.9%. In the ELA model, the lion's share of the predictive variability is from the reading domain test. In the mathematics regression model, the predictor with the highest R-Square is the speaking domain test, followed by the reading domain test. When all four domain tests are added to the mathematics logistic regression model, the R-square value is 38.6%.

Summary of Findings and Follow up Analysis with Alternate Composite Weighting Schemes

To establish optimal weighting procedures for Alt ACCESS as related to state alternate assessments, we conducted a logistic regression analysis. All four domain assessments were used to predict the likelihood that a student would be proficient on state ELA and mathematics alternate assessments. ELs' listening domain test scores were found to not be a significant predictor of proficiency in the state's alternate assessments. Using an odds ratio analysis, domain test scores were ordered in their relationship to the likelihood of being proficient on state alternate assessments. Stepwise regression analysis was conducted on each state alternate assessment using Alt ACCESS domain scores.

For the ELA alternate assessment, we found that Alt ACCESS's reading assessment was the strongest predictor for overall proficiency. For the mathematics alternate assessment, we found the speaking test to be the strongest protector. In both the ELA and mathematics regression analyses, we found writing to be the third strongest predictor. Thus, the reading and speaking domain scores are more strongly associated with proficiency in states' alternate assessments.

If predicting state alternate assessment proficiency is a value, weighting the reading and speaking domain tests higher than the writing and listening domain test seems appropriate. From the analyses provided here, the current Alt ACCESS overall composite weighting formula (Table 1) does not seem empirically justified.

What other overall composites weighing formulae might be suggested? Analyses reported here suggest that reading and speaking should be weighted more than writing and listening, but other models might be plausible. To examine possibilities, additional logistic regression analyses were conducted with alternate weightings. See the table below for alternate weighting models. *Note that the weights listed in this table are illustrative. We are not endorsing these values as the only or even most appropriate composite score weights. Instead we recommend that WIDA stakeholders and policymakers use this information in their discussions about composite score weighting for Alt ACCESS.*

TABLE 11: ALT ACCESS SCORE ALTERNATE WEIGHTING SCHEMES

Composite Weighting Models	Percentage of Domain Weight			
	Listening	Speaking	Reading	Writing
Current Overall Composite	15%	15%	35%	35%
Equally Weighted Composite	25%	25%	25%	25%
Reading & Speaking (R/S) Weighted Composite	10%	35%	35%	20%

For reference purposes, a logistic model was run with all Alt ACCESS domain scores and state alternate assessments proficiency scores were used as outcomes. As noted in a footnote above, the Max Rescaled R-Square statistic for model fit may introduce upward bias in results. Allison (2014) recommends that Tjur's Coefficient of Discrimination (Tjur R-Square)[10] be used as an alternative, and it will be used in comparisons here. For more on this see Allison (2014) and Tjur (2009). The table below displays the Tjur R-Square findings for 12 logistic regression models.

The highlighted row (All Domains) represents the upper end of predictability. That is, 30% of the variance in math alternate assessment proficiency categorizations, 23.6% of the variance in reading alternate assessment categorizations, and 36.8% of the variance in ELA alternate assessment proficiency categorizations can be attributed to Alt ACCESS's listening, speaking, reading, and writing domain scores. Alternate composite models closest to the All Domain reference Tjur R-Square values are preferred.

TABLE 12: ALT ACCESS SCORE ALTERNATE WEIGHTING SCHEMES TJUR R-SQUARE VALUES

Alt ACCESS Composite Model	Tjur R-Square Statistic for Alternate Assessments		
	Math (4 states)	Reading (3 States)	ELA (1 State)
Current Composite	28.2%	20.4%	36.0%
Equal Weighted Composite	29.5%	21.1%	36.4%
R/S Weighted Composite	29.9%	22.6%	36.8%
All Domains - Reference	30.0%	23.6%	36.8%

For all subject areas, the Reading and Speaking (R/S) weighted composite was closest to the All Domains Tjur R-Square reference values. This should not be surprising, given the findings with domain scores described above. Thus, predictability to state alternate assessment proficiency, at least in these four WIDA states, is greater with an Alt ACCESS composite that weighs reading and speaking higher than writing and listening. It should be clearly understood that the weights listed in Table 11 are arbitrary. A number of weighting models could have been used (e.g., R=40%, S=40%, W=15%, L=5%). Ultimately, weighting choices are policy decisions; ideally these decisions are informed by empirical analyses like those presented here.

This study has several limitations. First, a limited number of states were used in these analyses. Research Question #1 found that states had observable differences in the relationship between Alt ACCESS and their alternate assessments. This suggests that analyses using aggregations of alternate assessments across states may be misleading. More states are needed to better understand the variability in how Alt ACCESS interacts with state alternate assessments. The analyses used here to identify optimal composite score weighting may be similarly limited. Also, approaches other than the ones mentioned here might better identify the relationship between Alt ACCESS and alternate assessments. Ultimately, the goal is to provide empirical support to inform the development of meaningful composite scores for Alt ACCESS. The methods used here are but one approach.

[10] Tjur, T. (2009). Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination. *The American Statistician, 63*:4, pp. 366-372.

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

The Wisconsin Center for Education Research (WCER) is one of the nation's oldest university-based education research and development centers. WCER is based in the UW–Madison School of Education, which is consistently ranked one of the top schools of education in the country. With annual outside funding exceeding $47 million, WCER is home to centers for research on the improvement of mathematics and science education from kindergarten through postsecondary levels, the strategic management of human capital in public education, and value-added achievement, as well as the Minority Student Achievement Network and a multistate collaborative project to develop assessments for English language learners.

WIDA™ RESEARCH REPORT

Wisconsin Center for Education Research | University of Wisconsin-Madison
1025 West Johnson St. | Madison WI 53706 | www.wida.us