

Alignment between the WIDA English Language Proficiency Standards, 2007 Edition, Pre-Kindergarten through Grade 12 and ACCESS for ELLs® Assessment

A report on research carried out by the University of Illinois on behalf of the WIDA Consortium

Submitted December 27, 2011



I L L I N O I S



Youngshin Chi

Independent consultant

Chih-Kai (Cary) Lin

Jinming Zhang

University of Illinois at Urbana-Champaign

Preamble

This document and the research activities that support it were prepared by an independent consultant and researchers at the University of Illinois at Urbana-Champaign (UIUC).

This research was supported by a contract between UIUC and the University of Wisconsin (Madison).

Members of the WIDA headquarters team at the Wisconsin Center for Educational Research (University of Wisconsin)¹ facilitated the logistics of the research and provided review and comment on this report.

Enquiries about this report should be directed to WIDA, on behalf of UIUC.²

Jinming Zhang, PhD

Associate Professor of Educational Psychology

Principal Investigator, UIUC-WIDA Research Contract

¹ Wisconsin Center for Education Research; 1025 W. Johnson Street; University of Wisconsin-Madison; Madison, WI 53706

² <http://www.wida.us> -- Contact the WIDA Help Desk: help@wida.us, or call toll free 1-866-276-7735; + 608-263-4216 from outside the USA.

Table of Contents

Executive Summary	1
Introduction.....	3
Background.....	3
WIDA English Language Proficiency Standards.....	3
ACCESS for ELLs® Assessment.....	5
ELP Standards Alignment Procedure.....	6
Match.....	6
Depth	6
Breadth.....	7
Alignment Criteria.....	7
WIDA ELP Standards to ACCESS Alignment Process	11
Alignment Results.....	13
Summary.....	33
Reliability among Committee Members.....	34
References	36
Appendix	37
Reviewers' comment.....	37

List of Tables

Table 1: ELL Categorical Concurrence Criterion Example.....	6
Table 2: ACCESS Test Forms by Grade Cluster and Domain	7
Table 3: Domain by Level Alignment Criteria	9
Table 4: Domain by Standard Alignment Criteria	10
Table 5: WIDA to ACCESS Alignment Reviewers.....	11
Table 6: Level Alignment of Kindergarten Listening ACCESS Items	13
Table 7: Standard Alignment of Kindergarten Listening ACCESS Items.....	13
Table 8: Level Alignment of Kindergarten Speaking ACCESS Items.....	14
Table 9: Standard Alignment of Kindergarten Speaking ACCESS Items.....	14
Table 10: Level Alignment of Kindergarten Reading ACCESS Items	15
Table 11: Standard Alignment of Kindergarten Reading ACCESS Items	15
Table 12: Level Alignment of Kindergarten Writing ACCESS Items	16
Table 13: Standard Alignment of Kindergarten Writing ACCESS Items	16
Table 14: Level Alignment of Grades 1-2 Listening ACCESS Items.....	17
Table 15: Standard Alignment of Grades 1-2 Listening ACCESS Items.....	17
Table 16: Level Alignment of Grades 1-2 Speaking ACCESS Items	18
Table 17: Standard Alignment of Grades 1-2 Speaking ACCESS Items	18
Table 18: Level Alignment of Grades 1-2 Reading ACCESS Items.....	19
Table 19: Standard Alignment of Grades 1-2 Reading ACCESS Items.....	19
Table 20: Level Alignment of Grades 1-2 Writing ACCESS Items	20
Table 21: Standard Alignment of Grades 1-2 Writing ACCESS Items.....	20
Table 22: Level Alignment of Grades 3-5 Listening ACCESS Items.....	21
Table 23: Standard Alignment of Grades 3-5 Listening ACCESS Items.....	21
Table 24: Level Alignment of Grades 3-5 Speaking ACCESS Items	22
Table 25: Standard Alignment of Grades 3-5 Speaking ACCESS Items	22

Table 26: Level Alignment of Grades 3-5 Reading ACCESS Items.....	23
Table 27: Standard Alignment of Grades 3-5 Reading ACCESS Items.....	23
Table 28: Level Alignment of Grades 3-5 Writing ACCESS Items	24
Table 29: Standard Alignment of Grades 3-5 Writing ACCESS Items.....	24
Table 30: Level Alignment of Grades 6-8 Listening ACCESS Items.....	25
Table 31: Standard Alignment of Grades 6-8 Listening ACCESS Items.....	25
Table 32: Level Alignment of Grades 6-8 Speaking ACCESS Items	26
Table 33: Standard Alignment of Grades 6-8 Speaking ACCESS Items	26
Table 34: Level Alignment of Grades 6-8 Reading ACCESS Items.....	27
Table 35: Standard Alignment of Grades 6-8 Reading ACCESS Items.....	27
Table 36: Level Alignment of Grades 6-8 Writing ACCESS Items	28
Table 37: Standard Alignment of Grades 6-8 Writing ACCESS Items.....	28
Table 38: Level Alignment of Grades 9-12 Listening ACCESS Items.....	29
Table 39: Standard Alignment of Grades 9-12 Listening ACCESS Items.....	29
Table 40: Level Alignment of Grades 9-12 Speaking ACCESS Items	30
Table 41: Standard Alignment of Grades 9-12 Speaking ACCESS Items	30
Table 42: Level Alignment of Grades 9-12 Reading ACCESS Items.....	31
Table 43: Standard Alignment of Grades 9-12 Reading ACCESS Items.....	31
Table 44: Level Alignment of Grades 9-12 Writing ACCESS Items	31
Table 45: Standard Alignment of Grades 9-12 Writing ACCESS Items.....	32
Table 46: Reliability among Committee Members	34

List of Figures

Figure 1 Example of the Model Performance Indicators for Language Arts Grades 3-5	4
Figure 2 ACCESS for ELLs Assessment Model	5

Executive Summary

This report presents the findings of an alignment study conducted between the WIDA English Language Proficiency (ELP) Standards, 2007 Edition, Pre-Kindergarten through Grade 12 and the ACCESS for ELLs[®] assessment. The study was conducted October 19-20, 2010 in Schaumburg, Illinois with the assistance of 55 educators from WIDA member states. The protocol is based on Dr. Gary Cook's (2006) adaptation of Dr. Norman Webb's (1997) alignment framework.

What is Alignment?

Alignment is a mechanism for contributing evidence to a test's validity argument. It is good test development practice to consider alignment during all phases of development. Post-development alignment studies serve as a means to validate the presumptions of test developers. When analyzing the relationship between standards and assessments, an alignment study answers the question, 'How valid is this assessment in measuring what students are expected to know?'

A closely aligned standards-based assessment system is important for student success. In addition, the 2001 reauthorization of the Elementary and Secondary Education Act (ESEA) requires that states align their standards, including ELP standards, and assessments. Over the past decade, alignment has come to refer not only to matching the content of test items but also to ascertaining the breadth and depth of the items in relation to the standards (Webb, 1997; Cook, 2006).

Procedures

The Cook ELP alignment procedure used in this study is adapted from Webb's alignment methodology and uses many of the same measures. An important difference is that in modifying the methodology to address the construct of language proficiency, an ELP study uses a metric to examine the Linguistic Difficulty Level (LDL) of the standards and assessment rather than the Depth of Knowledge (DOK) metric used in a Webb alignment study. The procedure examines the relationship between standards and assessments in three dimensions: match, depth and breadth. Match examines how well assessment items can be matched to the ELP standards. Depth examines the relationship between the linguistic difficulty of the standards and assessment. Breadth addresses the degree to which the assessment covers the standards. Each of the three areas has associated statistics.

For this study, two alignment analyses were conducted for each grade cluster (K, 1-2, 3-5, 6-8, and 9-12): a domain by standard alignment and a domain by level alignment. The domain by standard alignment examines the relationship between ACCESS test items and the listening, speaking, reading, and writing WIDA model performance indicators for each of the five WIDA ELP Standards: Social and Instructional Language, the Language of Language Arts, the Language of Mathematics, the Language of Science, and the Language of Social Studies.

In the domain by level alignment, ACCESS test items were aligned to listening, speaking, reading, and writing model performance indicators for five proficiency levels expressed in the WIDA ELP Standards.

Results

Domain by Level Alignment

The domain by level alignment showed that the match and depth criteria were met for almost all levels across the domains but the breadth criterion was not entirely met for most levels except those in 3-5 Reading, 9-12 Reading and 6-8 Listening.

Domain by Standard Alignment

The domain by standard alignment showed strong or moderate findings across the domains and grade clusters. The match and depth criteria were generally strongly or moderately met. Limited range, which is one of the breadth criteria, was found for most clusters in Language of Language Arts, Language of Mathematics, and Language of Social Studies in some domains. Of the domains, writing had the weakest range across grade clusters.

Overall Results

The results of this study indicate that the ACCESS for ELLs® met most of the alignment criteria for the domain by level analysis. In particular:

- Most levels had adequate alignment for match for all domains
- Most levels had adequate alignment for depth for all domains
- Some levels were weak or did not meet the Range criterion (primarily in writing)

The domain by standard analysis indicates:

- The match criterion was strongly or moderately met for most standards across grade clusters
- The depth criterion was strongly met for most standards across grade clusters
- Breadth revealed some issues in that one or more standards for most domains did not meet the Range and Balance criteria

Introduction

Background

The ACCESS for ELLs[®] assessment is designed to enable English language learners (ELLs) to demonstrate their English language proficiency (ELP) and ELP growth. This alignment study was conducted to analyze the relationship between the WIDA ELP Standards, Pre-Kindergarten through Grade 12, 2007 edition, and the ACCESS for ELLs assessment. Webb's (1997) alignment methodology, traditionally used to evaluate the alignment between academic content standards and academic content assessments, has been adapted to study the alignment between ELP standards and assessments (Cook, 2006). This report analyzes the alignment between the assessment and standards in three areas: match, depth, and breadth.

WIDA English Language Proficiency Standards

The WIDA English Language Proficiency Standards (WIDA Consortium, 2007) focus on the social and academic language needed by ELLs to succeed in school. In contrast to content standards, the WIDA ELP Standards include multiple dimensions – standards, domains, and levels – which, taken as a whole, are intended to assist students in developing language proficiency in specific academic contexts (WIDA English Language Proficiency Standards and Resource Guide, 2007). It is important to emphasize that although the ELP standards refer to content areas, they address the “language” of those content areas. The standards are structured to provide examples of the types of language activities that will support access to content knowledge.

The WIDA ELP Standards include the following five standards:

1. English language learners communicate in English for SOCIAL AND INSTRUCTIONAL purposes within the school setting.
2. English language learners communicate information, ideas, and concepts necessary for academic success in the content area of LANGUAGE ARTS.
3. English language learners communicate information, ideas, and concepts necessary for academic success in the content area of MATHEMATICS.
4. English language learners communicate information, ideas, and concepts necessary for academic success in the content area of SCIENCE.
5. English language learners communicate information, ideas, and concepts necessary for academic success in the content area of SOCIAL STUDIES.

The standards are organized by grade level clusters (Pre K-K, 1-2, 3-5, 6-8, 9-12) and for each cluster, the five general standards (Social and Instructional Language, the Language of Language Arts, the Language of Mathematics, the Language of Science, and the Language of Social Studies) are elaborated. Per grade cluster, each standard is divided into four domains: listening, speaking, reading, and writing.

A visual layout of the components of WIDA’s model performance indicators for [one of] the WIDA ELP Standards is displayed in Figure 1. The English language proficiency levels head each column and the language domains begin each row. The horizontal cells contain five model performance indicators, creating a strand or strands across proficiency levels within a language domain.

The five language proficiency levels covered in the WIDA ELP Standards and included in the standards matrix are:

- Level 1 – Entering;
- Level 2 – Beginning;
- Level 3 – Developing;
- Level 4 – Expanding; and
- Level 5 – Bridging

Grades 3-5							32
ELP Standard 2: The Language of Language Arts, Summative Framework							
		Level 1 Entering	Level 2 Beginning	Level 3 Developing	Level 4 Expanding	Level 5 Bridging	Level 6 - Reaching
LISTENING	Example Genre	Match pictures to individual clues based on oral statements	Identify pictures associated with solutions to short mysteries read aloud	Make predictions based on pictures of clues/ pieces of evidence from mysteries and oral descriptions	Sequence pictures of clues/pieces of evidence from mysteries read aloud	Apply analogies of events or characters in mysteries read aloud to students' lives	
	Mysteries						
LISTENING	Example Topic	Match oral statements from narrative or expository material to their illustrated representations	Determine literal meanings of oral passages from narrative or expository material and match to illustrations	Project next in a sequence from oral discourse on narrative or expository material supported by illustrations	Identify cause/effect in oral discourse from narrative or expository material supported by illustrations	Make connections and draw conclusions from oral discourse using grade-level materials	
	Explicit & inferential information						
SPEAKING	Example Genre	Answer WH- questions to distinguish between pictures of real and imaginary people, objects or situations	Describe pictures of imaginary people, objects or situations	Provide details of pictures of imaginary people, objects or situations	Complete scenarios from pictures of imaginary people, objects or situations	Make up fantasies about imaginary people, objects or situations	
	Fantasies						
SPEAKING	Example Topic	Name story elements of various genres (e.g., non-fiction works, fairy tales, myths, fables or legends) depicted visually	Describe story elements of various genres supported by illustrations	Summarize story lines, issues or conflicts in various genres supported by illustrations	Discuss relationships among ideas or offer opinions on issues in various genres supported by illustrations	Propose options or solutions to issues in various genres and support responses with details	
	Story elements & types of genres						

Figure 1. Example of the Model Performance Indicators for Language Arts Grades 3-5

WIDA’s model performance indicators are functional, measurable indices of the language domains (listening, speaking, reading, and writing) aimed at the targeted age/developmental levels of English language learners. Since they are examples drawn from a myriad of English language proficiency and state academic content standards, they are not intended to be exhaustive of all of the ways in which students use and produce language.

ACCESS for ELLs® Assessment

The alignment reported here is designed to express the relationship between the WIDA English Language Proficiency Standards (WIDA) and the ACCESS for ELLs® (ACCESS) assessment. ACCESS is a uniquely designed assessment in that there are three “Tiers” associated with each grade band of the test (with the exception of Kindergarten). Figure 2 (www.wida.us/ACCESSForELLs/02_the_tiers.html) displays this relationship:

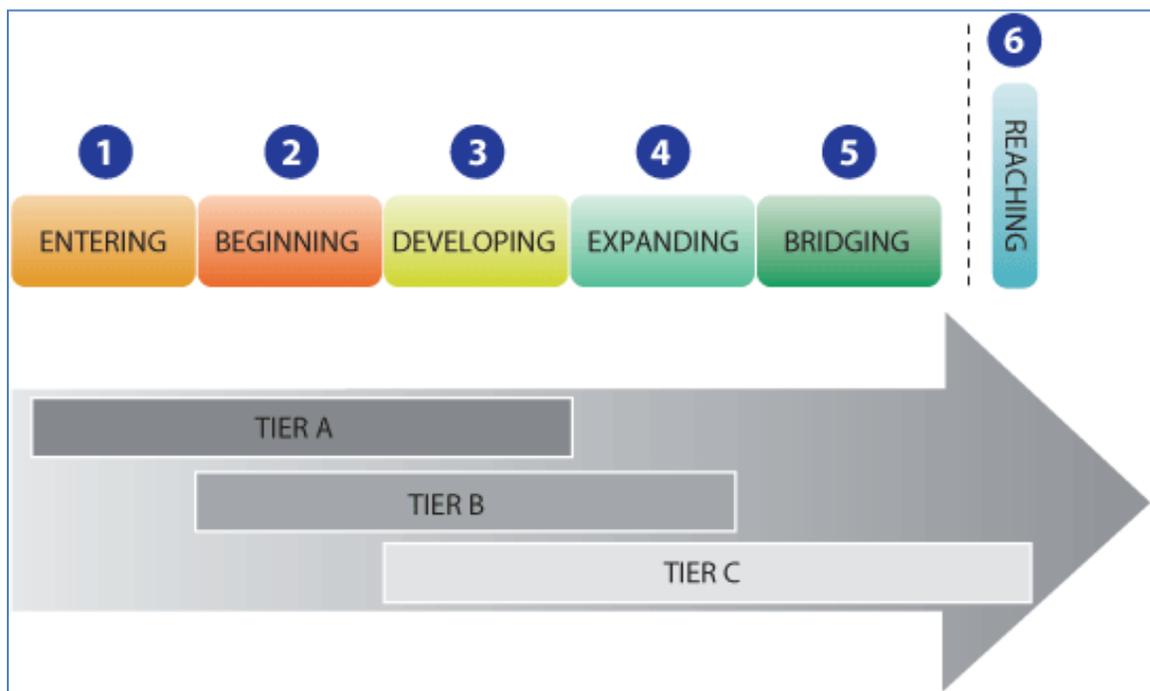


Figure 2 ACCESS for ELLs® Assessment Model

The ACCESS for ELLs® assessment was developed based on the WIDA ELP Standards and thus includes forms for each grade cluster: PreK-K, 1-2, 3-5, 6-8, and 9-12. Each grade cluster form is divided into the domains of reading, writing, listening, and speaking. Form 202, the operational test in the 2011-12 school year, was used in this study.

To appropriately align ACCESS to the WIDA Standards, all elements of the standards should be addressed. Based on consultation with WIDA staff, two alignment analyses were conducted. The first analysis focuses on level within domains. The question addressed is ‘How well do ACCESS items match WIDA’s linguistic complexity within each domain?’ The second analysis examines the standard by domain alignment. That is, ‘How well are the WIDA ELP Standards within domains expressed on the ACCESS test?’ Taken together, the two analyses capture the structural

interaction between standards, domains, and levels expressed in the WIDA ELP Standards. The following section of this report outlines the alignment methodology used for both analyses.

ELP Standards Alignment Procedure

Match

Match refers to how well the assessment matches, or has items associated with, the standards. To evaluate this aspect, the statistic, **Categorical Concurrence (CAT)**, is used. CAT is calculated by averaging the number of items reviewers assign to specific ELP standards. Reviewers, educators and ELL experts who participate in the alignment process, select specific model performance indicators unique to their grade cluster for each item on the tests being reviewed. The number of coded items are averaged across reviewers and reported as CAT. This statistic may be considered a proxy for the average number of items that reviewers determine to address specific standards. It is important to note that some items can address more than one standard, and reviewers are allowed to code accordingly. The criterion used for CAT for an ELL alignment differs depending on the alignment area examined. The underlying concept behind the criterion is there must be at least 6 items aligned at any decision point. The number 6 is taken from Webb’s research (1999). For example, if a set of ELP standards has five levels of proficiency (beginning, low intermediate, high intermediate, low advanced, high advanced), at least three items would need to be aligned at each level to be acceptable. The table below displays this concept.

Table 1: ELL Categorical Concurrence Criterion Example

Proficiency Level	Aligned Items	Items at Decision Points			
Beginning	3	6			
Low Intermediate	3		6		
High Intermediate	3			6	
Low Advanced	3				6
High Advanced	3				6

At each proficiency level, 3 items are aligned. At the decision point between Beginning and Low Intermediate there are 6 items as well as at all other decision points (e.g., Low Intermediate/High Intermediate, High Intermediate/Low Advanced, etc.). This concept of six items is applied to all relevant levels, skills and content. It is important to note that six items at each decision point is a minimum criterion.

Depth

To evaluate depth, a metric representing the percent of items at the **Linguistic Difficulty Level (LDL)** of the standards is used. This measure is somewhat akin to Webb’s Depth of Knowledge statistic, but in this case it refers to linguistic instead of cognitive complexity. Reviewers code each language proficiency standard (i.e, model performance indicator) with a linguistic difficulty level of 1, 2, or 3. As will be further outlined below, Level 1 stands for elementary linguistic features, Level 2 stands for standard constructions, and Level 3 refers to complex formulations. During the alignment process, LDLs are also assigned to each test item. The purpose is to identify the connection between the standards’ LDLs and the test items’ LDLs. The statistic reported here refers to the percent of test items coded at the LDL level of the standard. The main idea of

assessing ELL students is to help identify and monitor linguistic progress. If items are not reflective of the linguistic levels of the standards, the ability to properly evaluate students is limited. This is not to say that all items need to match the LDL of the standard. The acceptability criterion for LDL is 50% for both skill-based content and language proficiency levels.

Breadth

Two statistics are used to evaluate how well an assessment covers the breadth of the ELP standards. The first measure is range. **Range of Knowledge** (Range) refers to how well a test’s items cover a set of standards. Webb (2001) states that “[t]his criterion is met if a comparable span of knowledge expected of students by a standard is the same as, or corresponds to, the span of knowledge that students need in order to correctly answer the assessment items/activities.” This statistic is calculated by identifying the number of standards/levels that have at least one aligned item. If all standard/levels have at least one aligned item, the Range would be 100%. The purpose of this statistic is to help identify the breadth of coverage of standards. Again, using Webb’s criterion, the suggested minimum acceptable Range criterion is 50%, i.e., 50% of the standards or levels have at least one aligned item. The second statistic used to examine breadth is **Balance of Representation** (Balance). Balance is met if “the degree to which one objective is given emphasis on the assessment is comparable to the emphasis given to the other objectives within a standard.” An index is calculated to obtain Balance. The index identifies the degree to which item coverage is spread evenly across standards/level. According to Webb (2001), this index must be .70 or higher to show acceptable balance. Summarized briefly, the Range statistic examines the proportion of standards/levels covered by the assessment and Balance describes how equally dispersed those items are across standards/levels.

Alignment Criteria

One of the unique features of the ACCESS test is that grade clusters after the Kindergarten test have three tiers per domain. The speaking domain for those clusters has only one tier. Kindergarten has only one form that includes all domains. Table 2 displays the number of forms and tiers for each grade cluster.

Table 2: ACCESS Test Forms by Grade Cluster and Domain

Clusters	Domains and Forms			
	Listening Form	Speaking Form	Reading Form	Writing Form
Kindergarten	One combined form			
Grades 1-2	Tiers A, B, C	One form per grade cluster	Tiers A, B, C	Tiers A, B, C
Grades 3-5	Tiers A, B, C		Tiers A, B, C	Tiers A, B, C
Grades 6-8	Tiers A, B, C		Tiers A, B, C	Tiers A, B, C
Grades 9-12	Tiers A, B, C		Tiers A, B, C	Tiers A, B, C

In order to address all grade clusters and domains, the following decisions about the set-up of the alignment study were made:

1. Alignment studies would be set up based on domain (i.e. listening, speaking, reading, and writing).

2. Each study would include a combination of all tiers for each domain. Common items across tiers would be coded once, at the lowest tier.
3. Each alignment reviewer would be in a group of 6-7 people representing two domains (i.e. listening/speaking for grade cluster 1-2 or reading/writing for 1-2, etc.).
4. Two sets of alignment criteria were established that utilize Cook's 2006 nomenclature. The first addresses domain by level alignment; the second addresses domain by standard alignment. The domain by level alignment criteria are expressed as YES – for acceptable, WEAK – for weak alignment, and NO – for not aligned. For the domain by standard alignment, the nomenclature is expressed as STRONG, MODERATE, and LIMITED.

The WIDA Standards differ from traditional language proficiency expectations in that students are assessed on discourse features associated with academic content, not just English or language arts. Alignment criteria for the domain by standard alignment express the strength of that relationship. While criteria used here adopt Webb's alignment specifications, little guidance from the literature or federal or state guidelines is available as to what is or is not acceptable in an ELP alignment. We adopt the terms YES, WEAK, and NO (for the domain by level alignment) and STRONG, MODERATE, and LIMITED (for the domain by standard alignment) to describe the relationships. This nomenclature has been used for numerous WIDA alignment studies for both standards-to-standards correspondence studies and assessment-to-standards alignment studies

Table 3: Domain by Level Alignment Criteria

Alignment Areas		Alignment Criteria								
		Listening and Reading			Speaking			Writing		
		NO	WEAK	YES	NO	WEAK	YES	NO	WEAK	YES
Categorical Concurrence	Level 1	<2	≥2	≥3	<2	≥2	≥3	<0.5	≥0.5	≥1
	Level 2	<4	≥4	≥6				<1.5	≥1.5	≥2
	Level 3	<7	≥7	≥9				<2.5	≥2.5	≥3
	Level 4	<4	≥4	≥6				<1.5	≥1.5	≥2
	Level 5	<2	≥2	≥3				<0.5	≥0.5	≥1
LDL Consistency		<40%	≥40%	≥50%						
Range of Knowledge		<40%	≥40%	≥50%						
Balance of Representation		<0.6	≥0.6	≥0.7						

*Kindergarten is not tiered, so the Categorical Concurrence criterion is the same (3 items) for all levels for Listening, Reading, and Speaking. For writing, the criterion for all levels is 1.

Table 3 outlines the alignment criteria for the domain by level alignment. For CAT, it is similar to Table 1. However, the example given there presumed only one form of a test is aligned. The alignment criteria presented here are across three tiers (forms). Since tiers are designed to span three levels at a time, different numbers of items would be needed to meet Webb’s CAT criterion of 6 items. Recall that Tier A spans levels 1, 2 and 3, Tier B spans levels 2, 3, and 4, and Tier C spans levels 3, 4 and 5. Across proficiency levels, levels 1 and 5 are represented once (Tier A and Tier C respectively). Levels 2 and 4 are represented twice (Tiers A & B and Tiers B & C), and level 3 is represented three times (Tiers A, B and C). Using the example shown in Table 1, we would expect at a minimum 3 items at levels 1 and 5; 6 items at levels 2 and 4, and 9 items at level 3.

By design, the ACCESS speaking test has only one form per grade cluster. Therefore, the CAT criterion is a minimum of 3 items across all levels. The CAT criterion for writing differs from listening, speaking or reading. By design, the ACCESS writing test 1-12 forms have only writing prompts. There are limited numbers of items in the writing test at each tier because productive writing tasks take more time to complete than multiple-choice items. Each prompt is graded with a scoring rubric. If we consider each score point on the rubric as an item, we could posit that a CAT finding of 1 for writing would represent 3 items (based on the rubric). Upon consultation with WIDA staff, that is the criteria adopted here. For Levels 1 and 5, acceptable CAT is 1; for Levels 2 and 4, it is 2; and for Level 3 the criteria is 3.

The domain by level LDL Consistency criterion is consistent with the criterion described in the previous section, as are the Range of Knowledge and Balance of Representation criteria. Again, the expectation is that strong LDL consistency would be demonstrated by 50% of assessment items being at the LDL level of the standards, strong range would indicate that 50% or more of the standards for that level were targeted, and strong balance would be an index of 0.7 or above.

Table 4: Domain by Standard Alignment Criteria

Alignment Areas	Alignment Criteria								
	Listening and Reading			Speaking			Writing		
	Limited	Moderate	Strong	Limited	Moderate	Strong	Limited	Moderate	Strong
Categorical Concurrence	<4	≥4	≥6	<3	≥3	≥4	<0.5	≥0.5	≥1
LDL Consistency	<40%	≥40%	≥50%						
Range of Knowledge	<40%	≥40%	≥50%						
Balance of Representation	<0.6	≥0.6	≥0.7						

As shown in Table 4, the domain by standard alignment criteria is similar to that of a traditional Webb alignment study. The nomenclature is less definitive as the primary goal of this study is to ensure that proficiency levels are sufficiently addressed by the items. Standards coverage is secondary. For the receptive domains (Listening and Reading), strong Categorical Concurrence alignment indicates 6 or more aligned items. As noted above, the productive domains (Speaking and Writing) have fewer items. In addition, grades 1-12 Speaking forms have the Language of Language Arts and Language of Social Studies standards as a combined unit and have the Language of Mathematics and Language of Science standards as another combined unit on the test forms. Three units of standards are addressed in some test items rather than five standards. To ensure equal coverage of the three units of standards, the criterion of CAT for grades 1-12 Speaking will be 4 or more aligned items per unit standards. For Kindergarten, the criterion of CAT will require 4 or more aligned items per each standard. For writing test forms for all clusters, adequate coverage of the standards is determined as one or more item per standard.

WIDA ELP Standards to ACCESS Alignment Process

The alignment workshop was conducted at the Harper College Professional Center in Schaumburg, Illinois on October 19-20, 2010. WIDA staff invited all member states to nominate participants. Fifty-five ELL teachers and content experts with extensive experiences working with ELLs were selected for the study. Participants represented 23 WIDA member states.

The table below displays reviewers' names, the states they represent, and the grade clusters they represented. Grey highlighted names are of reviewers who acted as table leaders for each group:

Table 5: WIDA to ACCESS Alignment Reviewers

Grade Cluster	Name	State
Grades K-2 (Listening and Speaking)	Karen Botts	KY
	Lauretta Kloer	GA
	Lynne Duffy	IL
	Rebecca King	OK
	Patricia Hedden	VA
	Pat Lewno	WI
	Taffi Brown	WY
Grades K-2 (Reading and Writing)	Monica Coe	AL
	Mindi Teich	DC
	Diane Murakami	HI
	Jenifer Albright-Borts	MO
	Timothy Sims	NC
	Brittany Holten	ND
	Patricia Gardner	NH
Grades 3-5 (Listening and Speaking)	Noreen Segal	IL
	Karen Snyder	AL
	Anne Betbeze	AL
	Jennifer Corum	MS
	Elva Autry	NM
	Jennifer Paolozzi	RI
	Ruth Hicks	VT
Grades 3-5 (Reading and Writing)	Mae Lindenberg	ND
	Jacqueline Rose	GA
	Daniele Novak	KY
	Jane Cross	ME
	Tracy Klingbiel	NC
	Nancy Carnevale	RI
	Marcia Gaudet	SD
Grades 6-8	Jewel Kastrantas	MS

Grade Cluster	Name	State
(Listening and Speaking)	Meredith Vesoulis	MO
	Isela Rosales	NM
	Deborah Dobbs	OK
	Bethany Pennacchio	RI
	Kim Hanneman	WY
Grades 6-8 (Reading and Writing)	Deborah Blackledge	MS
	Beata Kruss	IL
	Betsy Burton	NC
	Kevin LaMastra	NJ
	Teresa Vignaroli	VA
	Ervina Kuckovic	VT
	Wanda Grasse	WI
Grades 9-12 (Listening and Speaking)	Elizabeth Allen	DE
	Sandra Byrd	KY
	Louise Moses	ME
	Pamela Croxton	PA
	Patricia Sinclair	VA
	Kristen Gundry	WI
	Paul Jimenez	WY
Grades 9-12 (Reading and Writing)	Patricia Morris	RI
	Kristie Segers	AL
	Joy Cowart	GA
	Kathryn Khoshaba	IL
	Karla Winslow	OK
	Kristin Grinager	SD
	Olga Benoit	VT

After an initial training for all reviewers, the alignment process was conducted in two parts.

In Part 1, reviewers assigned linguistic difficulty levels (LDLs) to each of the WIDA Model Performance Indicators (MPIs). The three levels of the LDL metric are:

- Level 1: Elementary — A limited to basic ability to process formulaic English linguistic features
- Level 2: Standard Constructions — A basic to moderate ability and facility to process English linguistic features
- Level 3: Complex Formulations — A moderate to sophisticated ability and facility to process English linguistic features

After coding the LDL of each MPI, reviewers met in groups to discuss the rationale for their codings and to come to a consensus on the LDL of each MPI. The consensus values for the MPIs are later used in the alignment analysis.

In Part 2, reviewers assigned LDLs to ACCESS test items and identified MPI(s) that best matched the test items. Table leaders were directed to discuss two or three practice test items as part of the training and then reviewers completed the coding independently. Once reviewers completed the coding process for a test form, they filled out debriefing questions describing their overall impressions of the alignment (see Appendix).

Upon completion of the alignment, the data were extracted from the Web Alignment Tool and analyzed. The next section of this report presents findings from the alignment analyses.

Alignment Results

Alignment results are presented for each grade cluster. Each grade cluster has two alignment tables: a domain by level alignment table and a domain by standard alignment table. Each table displays alignment statistics in the left four columns and alignment findings in the right four columns. CAT stands for Categorical Concurrence; LDL represents Linguistic Difficulty Level Correspondence; Range refers to Range of Knowledge, and Balance is Balance of Representation.

Kindergarten

Table 6: Level Alignment of Kindergarten Listening ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	11	76%	60%	0.70	YES	YES	YES	YES
Level 2	12.71	47%	80%	0.81	YES	WEAK	YES	YES
Level 3	10.14	58%	69%	0.85	YES	YES	YES	YES
Level 4	6.14	32%	49%	0.84	YES	NO	WEAK	YES
Level 5	4.86	82%	37%	0.77	YES	YES	NO	YES

Table 7: Standard Alignment of Kindergarten Listening ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	21	68%	84%	0.83	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	6.14	58%	48%	0.83	STRONG	STRONG	MODERATE	STRONG
Language of Mathematics	7	56%	53%	0.85	STRONG	STRONG	STRONG	STRONG
Language of Science	7	51%	60%	0.85	STRONG	STRONG	STRONG	STRONG
Language of Social Studies	9.14	36%	58%	0.85	STRONG	LIMITED	STRONG	STRONG

Kindergarten listening alignment results are presented in Table 6 and 7. In Table 6, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at all levels. The LDL level was sufficient for most levels; however, level 2 was weak and level 4 did not meet the LDL criterion. Levels 1-3 met the Range criterion but level 4 was weak and level 5 did not meet the criterion. Balance was met at all levels.

In Table 7, the domain by standard alignment showed that the CAT criterion was strongly met for all standards. The LDL criterion was strongly met for most standards except for limited LDL in Language of Social Studies. Range was strong or moderate for all standards and Balance was strong for all standards.

Table 8: Level Alignment of Kindergarten Speaking ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	2.86	98%	37%	0.90	WEAK	YES	NO	YES
Level 2	2.29	36%	40%	0.95	WEAK	NO	NO	YES
Level 3	1.57	89%	26%	0.70	NO	YES	NO	YES
Level 4	4.86	66%	63%	0.88	YES	YES	YES	YES
Level 5	2.86	88%	37%	0.89	WEAK	YES	NO	YES

Table 9: Standard Alignment of Kindergarten Speaking ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	4.43	69%	57%	0.73	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	3.86	85%	47%	0.72	MODERATE	STRONG	MODERATE	STRONG
Language of Mathematics	2.86	56%	45%	0.90	LIMITED	STRONG	MODERATE	STRONG
Language of Science	2.29	65%	30%	0.68	LIMITED	STRONG	LIMITED	MODERATE
Language of Social Studies	2.43	83%	32%	0.95	LIMITED	STRONG	LIMITED	STRONG

Kindergarten speaking alignment results are presented in Tables 8 and 9. In Table 8, the domain by level alignment showed that only level 4 met the CAT criterion. The LDL level was sufficient for all levels except at level 2. Range was sufficient only at level 4. Balance was met at all levels.

In Table 9, the domain by standard alignment showed that the CAT criterion was not met for the majority of standards. The LDL criterion was strongly met for all standards. Range was limited

for Language of Science and Language of Social Studies. Balance was strong or moderate for all standards.

Table 10: Level Alignment of Kindergarten Reading ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	6.57	89%	40%	0.95	YES	YES	WEAK	YES
Level 2	0.57	50%	6%	0.29	NO	YES	NO	NO
Level 3	5.71	75%	37%	0.80	YES	YES	NO	YES
Level 4	5.43	93%	26%	0.66	YES	YES	NO	WEAK
Level 5	21.43	90%	49%	0.84	YES	YES	WEAK	YES

Table 11: Standard Alignment of Kindergarten Reading ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	20.29	84%	60%	0.68	STRONG	STRONG	STRONG	MODERATE
Language of Language Arts	3.43	30%	9%	0.10	LIMITED	LIMITED	LIMITED	LIMITED
Language of Mathematics	3	86%	19%	0.86	LIMITED	STRONG	LIMITED	STRONG
Language of Science	11	84%	41%	0.87	STRONG	STRONG	MODERATE	STRONG
Language of Social Studies	6	81%	33%	0.93	STRONG	STRONG	LIMITED	STRONG

Kindergarten reading alignment results are presented in Table 10 and 11. In Table 10, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at levels 1, 3, 4 and 5. However, an insufficient number of aligned items was found at level 2. The LDL level was sufficient at all levels. Range was weak or not met at all levels. Balance was met for most levels although level 2 was not met and level 4 was weak.

In Table 11, the domain by standard alignment showed that the CAT criterion was strongly met for most standards whereas Language of Language Arts and Language of Mathematics did not meet the criterion. The LDL criterion was strongly met for most standards except for Language of Language Arts. Range was strong for Social and Instructional Language and moderate for Language of Science. Balance was strong for most standards but limited for Language of Language Arts.

Table 12: Level Alignment of Kindergarten Writing ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	2.86	100%	37%	1.00	YES	YES	NO	YES
Level 2	2.86	78%	23%	0.68	YES	YES	NO	WEAK
Level 3	1.43	80%	14%	0.57	YES	YES	NO	NO
Level 4	2.71	50%	29%	0.67	YES	YES	NO	WEAK
Level 5	2.71	85%	23%	0.82	YES	YES	NO	YES

Table 13: Standard Alignment of Kindergarten Writing ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	4.57	69%	42%	0.78	STRONG	STRONG	MODERATE	STRONG
Language of Language Arts	4.29	77%	36%	0.67	STRONG	STRONG	LIMITED	MODERATE
Language of Mathematics	1.43	100%	20%	0.57	STRONG	STRONG	LIMITED	LIMITED
Language of Science	0.86	100%	6%	0.29	MODERATE	STRONG	LIMITED	LIMITED
Language of Social Studies	3.43	66%	30%	0.80	STRONG	STRONG	LIMITED	STRONG

Kindergarten writing alignment results are presented in Tables 12 and 13. In Table 12, the domain by level alignment showed that reviewers identified a sufficient number of aligned items to meet the Categorical Concurrence criterion at all levels. The LDL level was sufficient at all levels. Range was insufficient at all levels. Balance was only met at levels 1 and 5.

In Table 13, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for most standards and moderately met for Language of Science. The LDL criterion was strongly met for all standards. Range was moderate for Social and Instructional Language and limited for the rest of the standards. Balance was strongly met for Social and Instructional Language and Language of Social Studies and moderately met for Language of Language Arts.

Grades 1-2

Table 14: Level Alignment of Grades 1-2 Listening ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	14.43	71%	86%	0.81	YES	YES	YES	YES
Level 2	18.43	55%	91%	0.77	YES	YES	YES	YES
Level 3	14.43	76%	77%	0.79	YES	YES	YES	YES
Level 4	13.86	54%	80%	0.78	YES	YES	YES	YES
Level 5	6	92%	46%	0.88	YES	YES	WEAK	YES

Table 15: Standard Alignment of Grades 1-2 Listening ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	13.86	70%	80%	0.79	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	23	71%	91%	0.75	STRONG	STRONG	STRONG	STRONG
Language of Mathematics	17.57	74%	81%	0.81	STRONG	STRONG	STRONG	STRONG
Language of Science	9.43	50%	66%	0.79	STRONG	STRONG	STRONG	STRONG
Language of Social Studies	10.43	60%	64%	0.76	STRONG	STRONG	STRONG	STRONG

Grades 1-2 listening alignment results are presented in Tables 14 and 15. In Table 14, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at all levels. The LDL level was sufficient at all levels. Range and Balance were met at most levels except for weak Range identified at level 5.

In Table 15, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for all standards. The LDL criterion was strongly met for all standards. Range and Balance were strong for all standards.

Table 16: Level Alignment of Grades 1-2 Speaking ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	4.57	92%	51%	0.80	YES	YES	YES	YES
Level 2	4.86	74%	60%	0.87	YES	YES	YES	YES
Level 3	4.57	71%	66%	0.84	YES	YES	YES	YES
Level 4	2.14	59%	31%	0.79	WEAK	YES	NO	YES
Level 5	3.29	69%	46%	0.90	YES	YES	WEAK	YES

Table 17: Standard Alignment of Grades 1-2 Speaking ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	5.71	86%	53%	0.82	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	8.15	78%	45%	0.70	STRONG	STRONG	LIMITED	STRONG
Language of Social Studies								
Language of Mathematics	7.43	66%	60%	0.95	STRONG	STRONG	STRONG	STRONG
Language of Science								

Grades 1-2 speaking alignment results are presented in Table 16 and 17. In Table 16, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at most levels except for level 4. The LDL level was sufficient at all levels. Range was met at most levels but was weak at level 5 and did not meet the criterion at level 4. Balance was met for all levels.

As noted earlier, grades 1-2 ACCESS speaking forms combine the Language of Language Arts and Language of Social Studies standards as a unit, and the Language of Mathematics and Language of Science standards as another unit. In the analyses of combined units of standards, the speaking alignment statistic for Categorical Concurrence is the sum of CAT for individual standards whereas the alignment statistics for LDL, Range, and Balance are the average of these statistics of individual standards. In Table 17, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for all standards. The LDL criterion was strongly met for all standards. Range and Balance were strong for most standards except for limited Range in the combined unit of Language of Language Arts and Language of Social Studies.

Table 18: Level Alignment of Grades 1-2 Reading ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	7.57	64%	57%	0.85	YES	YES	YES	YES
Level 2	11.29	32%	60%	0.77	YES	NO	YES	YES
Level 3	20.57	74%	66%	0.69	YES	YES	YES	WEAK
Level 4	21.71	64%	60%	0.69	YES	YES	YES	WEAK
Level 5	26	90%	74%	0.80	YES	YES	YES	YES

Table 19: Standard Alignment of Grades 1-2 Reading ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	25.86	70%	63%	0.67	STRONG	STRONG	STRONG	MODERATE
Language of Language Arts	19.29	74%	56%	0.77	STRONG	STRONG	STRONG	STRONG
Language of Mathematics	25.14	55%	83%	0.80	STRONG	STRONG	STRONG	STRONG
Language of Science	23.71	55%	80%	0.83	STRONG	STRONG	STRONG	STRONG
Language of Social Studies	17	71%	48%	0.64	STRONG	STRONG	MODERATE	MODERATE

Grades 1-2 reading alignment results are presented in Tables 18 and 19. In Table 18, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at all levels. The LDL level was sufficient at most levels but level 2 did not meet the LDL criterion. Range was met at all levels. Balance was met for most levels except for levels 3 and 4.

In Table 19, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for all standards. The LDL criterion was strongly met for all standards. Range and Balance were strong for most standards but were moderate for Social and Instructional Language (Balance) and Language of Social Studies (Range and Balance).

Table 20: Level Alignment of Grades 1-2 Writing ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	6	91%	29%	0.94	YES	YES	NO	YES
Level 2	4.14	46%	37%	0.92	YES	WEAK	NO	YES
Level 3	6.43	51%	37%	0.87	YES	YES	NO	YES
Level 4	2	56%	17%	0.54	YES	YES	NO	NO
Level 5	6	85%	54%	0.85	YES	YES	YES	YES

Table 21: Standard Alignment of Grades 1-2 Writing ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	16.43	63%	72%	0.75	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	4.43	59%	41%	0.78	STRONG	STRONG	MODERATE	STRONG
Language of Mathematics	1.43	75%	16%	0.56	STRONG	STRONG	LIMITED	LIMITED
Language of Science	4.43	65%	44%	0.94	STRONG	STRONG	MODERATE	STRONG
Language of Social Studies	0.86	67%	9%	0.43	MODERATE	STRONG	LIMITED	LIMITED

Grades 1-2 writing alignment results are presented in Tables 20 and 21. In Table 20, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at all levels. The LDL level was sufficient at most levels although level 2 was weak for the LDL criterion. Range was met only at level 5. Balance was met for most levels except for level 4.

In Table 21, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for most standards except for Language of Social Studies. The LDL criterion was strongly met for all standards. Range was either strong or moderate for most standards except for Language of Mathematics and Language of Social Studies. Balance was strongly met for most standards except for Language of Mathematics and Language of Social Studies.

Grades 3-5

Table 22: Level Alignment of Grades 3-5 Listening ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	12.71	87%	91%	0.76	YES	YES	YES	YES
Level 2	17.71	71%	91%	0.79	YES	YES	YES	YES
Level 3	15.71	73%	91%	0.73	YES	YES	YES	YES
Level 4	13.14	69%	74%	0.71	YES	YES	YES	YES
Level 5	15.86	91%	83%	0.64	YES	YES	YES	WEAK

Table 23: Standard Alignment of Grades 3-5 Listening ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	19.14	77%	94%	0.82	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	24.43	82%	91%	0.83	STRONG	STRONG	STRONG	STRONG
Language of Mathematics	15.57	81%	86%	0.82	STRONG	STRONG	STRONG	STRONG
Language of Science	8.86	71%	86%	0.77	STRONG	STRONG	STRONG	STRONG
Language of Social Studies	7.14	80%	74%	0.85	STRONG	STRONG	STRONG	STRONG

Grades 3-5 listening alignment results are presented in Tables 22 and 23. In Table 22, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at all levels. The LDL level was also sufficient at all levels. Range was sufficient at all levels, and the Balance criterion was met at most levels except for the weak Balance identified at level 5.

In Table 23, the domain by standard alignment results showed that the Categorical Concurrence criterion was strongly met for all standards. The LDL criterion was strongly met for all standards. Range and Balance were strong for all standards.

Table 24: Level Alignment of Grades 3-5 Speaking ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	5.29	96%	69%	0.79	YES	YES	YES	YES
Level 2	4.43	59%	66%	0.86	YES	YES	YES	YES
Level 3	4.14	67%	60%	0.92	YES	YES	YES	YES
Level 4	3.14	72%	43%	0.64	YES	YES	WEAK	WEAK
Level 5	3.71	98%	49%	0.75	YES	YES	WEAK	YES

Table 25: Standard Alignment of Grades 3-5 Speaking ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	8.57	74%	86%	0.85	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	6.29	81%	50%	0.73	STRONG	STRONG	STRONG	STRONG
Language of Social Studies								
Language of Mathematics	5.86	45%	50%	0.74	STRONG	MODERATE	STRONG	STRONG
Language of Science								

Grades 3-5 speaking alignment results are presented in Tables 24 and 25. In Table 24, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at all levels. The LDL level was also sufficient at all levels. Range and Balance were sufficient at most levels except for weak Range and Balance at level 4 and weak Range at level 5.

In Table 25, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for all standards. The LDL criterion was strongly met for most standards except for the combined unit of Language of Mathematics and Language of Science. Range and Balance were strong for all standards.

Table 26: Level Alignment of Grades 3-5 Reading ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	7.29	71%	80%	0.86	YES	YES	YES	YES
Level 2	13.29	55%	86%	0.75	YES	YES	YES	YES
Level 3	16.43	65%	91%	0.80	YES	YES	YES	YES
Level 4	22.57	56%	94%	0.72	YES	YES	YES	YES
Level 5	22.29	95%	97%	0.75	YES	YES	YES	YES

Table 27: Standard Alignment of Grades 3-5 Reading ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	15.86	69%	92%	0.72	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	24.71	56%	91%	0.74	STRONG	STRONG	STRONG	STRONG
Language of Mathematics	15.57	73%	80%	0.82	STRONG	STRONG	STRONG	STRONG
Language of Science	14.86	66%	94%	0.80	STRONG	STRONG	STRONG	STRONG
Language of Social Studies	11	79%	91%	0.76	STRONG	STRONG	STRONG	STRONG

Grades 3-5 reading alignment results are presented in Tables 26 and 27. In Table 26, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at all levels. The LDL level was also sufficient at all levels. Range and Balance were sufficient at all levels.

In Table 27, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for all standards. The LDL criterion was strongly met for all standards. Range and Balance were strong for all standards.

Table 28: Level Alignment of Grades 3-5 Writing ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	0.86	83%	17%	0.86	WEAK	YES	NO	YES
Level 2	1.57	80%	29%	0.83	WEAK	YES	NO	YES
Level 3	5.43	49%	54%	0.77	YES	WEAK	YES	YES
Level 4	3.29	100%	37%	0.91	YES	YES	NO	YES
Level 5	2	100%	31%	0.80	YES	YES	NO	YES

Table 29: Standard Alignment of Grades 3-5 Writing ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	3.43	80%	43%	0.85	STRONG	STRONG	MODERATE	STRONG
Language of Language Arts	1.29	83%	17%	0.70	STRONG	STRONG	LIMITED	STRONG
Language of Mathematics	2.57	65%	49%	0.98	STRONG	STRONG	MODERATE	STRONG
Language of Science	5.14	87%	49%	0.81	STRONG	STRONG	MODERATE	STRONG
Language of Social Studies	0.71	88%	11%	0.57	MODERATE	STRONG	LIMITED	LIMITED

Grades 3-5 writing alignment results are presented in Tables 28 and 29. In Table 28, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at most levels; however, weak Categorical Concurrence was found at levels 1 and 2. The LDL level was also sufficient at most levels except for level 3. Range was only met at level 3. Balance was sufficient for all levels.

In Table 29, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for most standards except for Language of Social Studies. The LDL criterion was strongly met for all standards. Range was moderate or limited for all standards. Balance was strongly met for most standards except for Language of Social Studies.

Grades 6-8

Table 30: Level Alignment of Grades 6-8 Listening ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	17.67	79%	93%	0.81	YES	YES	YES	YES
Level 2	12.67	53%	87%	0.70	YES	YES	YES	YES
Level 3	10.33	75%	87%	0.79	YES	YES	YES	YES
Level 4	10.83	44%	90%	0.81	YES	WEAK	YES	YES
Level 5	13.33	63%	90%	0.84	YES	YES	YES	YES

Table 31: Standard Alignment of Grades 6-8 Listening ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	18.5	67%	90%	0.79	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	9.5	63%	80%	0.78	STRONG	STRONG	STRONG	STRONG
Language of Mathematics	16.17	70%	97%	0.80	STRONG	STRONG	STRONG	STRONG
Language of Science	11.5	60%	90%	0.74	STRONG	STRONG	STRONG	STRONG
Language of Social Studies	9.17	54%	90%	0.80	STRONG	STRONG	STRONG	STRONG

Grades 6-8 listening alignment results are presented in Table 30 and 31. In Table 30, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at all levels. The LDL level was also sufficient at most levels except for level 4. Range and Balance were met at all levels.

In Table 31, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for all standards. The LDL criterion was strongly met for all standards. Range and Balance were strong for all standards.

Table 32: Level Alignment of Grades 6-8 Speaking ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	8.83	62%	63%	0.73	YES	YES	YES	YES
Level 2	3.17	47%	33%	0.93	YES	WEAK	NO	YES
Level 3	3.33	72%	53%	0.96	YES	YES	YES	YES
Level 4	2.5	79%	47%	0.97	WEAK	YES	WEAK	YES
Level 5	2.83	92%	40%	0.78	WEAK	YES	WEAK	YES

Table 33: Standard Alignment of Grades 6-8 Speaking ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	8.5	54%	63%	0.77	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	7.83	76%	52%	.084	STRONG	STRONG	STRONG	STRONG
Language of Social Studies								
Language of Mathematics	4.33	76%	35%	0.82	STRONG	STRONG	LIMITED	STRONG
Language of Science								

Grades 6-8 speaking alignment results are presented in Tables 32 and 33. In Table 32, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at most levels; however, weak Categorical Concurrence was found at levels 4 and 5. The LDL level was sufficient at most levels except for level 2. Range was sufficient at levels 1 and 3 and Balance was sufficient at all levels.

In Table 33, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for all standards. The LDL criterion was strongly met for all standards. Range was strong for most standards except for the combined unit of Language of Mathematics and Language of Science. Balance was strong for all standards.

Table 34: Level Alignment of Grades 6-8 Reading ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	8.29	63%	51%	0.59	YES	YES	YES	NO
Level 2	28.14	84%	91%	0.68	YES	YES	YES	WEAK
Level 3	23.57	54%	91%	0.72	YES	YES	YES	YES
Level 4	35.71	83%	94%	0.74	YES	YES	YES	YES
Level 5	23.29	91%	94%	0.67	YES	YES	YES	WEAK

Table 35: Standard Alignment of Grades 6-8 Reading ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	32.86	77%	83%	0.77	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	44.43	73%	90%	0.73	STRONG	STRONG	STRONG	STRONG
Language of Mathematics	17.29	78%	91%	0.80	STRONG	STRONG	STRONG	STRONG
Language of Science	12.86	77%	78%	0.73	STRONG	STRONG	STRONG	STRONG
Language of Social Studies	13	76%	83%	0.73	STRONG	STRONG	STRONG	STRONG

Grades 6-8 reading alignment results are presented in Tables 34 and 35. In Table 34, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at all levels. The LDL level was also sufficient at all levels. Range was sufficient at all levels. Balance was met at levels 3 and 4.

In Table 35, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for all standards. The LDL criterion was strongly met for all standards. Range and Balance were strong for all standards.

Table 36: Level Alignment of Grades 6-8 Writing ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	2.29	100%	40%	0.98	YES	YES	WEAK	YES
Level 2	5.86	81%	66%	0.83	YES	YES	YES	YES
Level 3	5.29	45%	63%	0.81	YES	WEAK	YES	YES
Level 4	5.14	94%	49%	0.85	YES	YES	WEAK	YES
Level 5	2	83%	34%	0.86	YES	YES	NO	YES

Table 37: Standard Alignment of Grades 6-8 Writing ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	6.71	77%	86%	0.86	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	6.43	83%	74%	0.71	STRONG	STRONG	STRONG	STRONG
Language of Mathematics	3.43	57%	39%	0.93	STRONG	STRONG	LIMITED	STRONG
Language of Science	3.86	80%	49%	0.85	STRONG	STRONG	MODERATE	STRONG
Language of Social Studies	0.43	100%	6%	0.29	LIMITED	STRONG	LIMITED	LIMITED

Grades 6-8 writing alignment results are presented in Tables 36 and 37. In Table 36, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at all levels. The LDL level was sufficient at most levels; however, weak LDL was found at level 3. Range criterion was met at levels 2 and 3. Balance was sufficient at all levels.

In Table 37, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for most standards except for Language of Social Studies. The LDL criterion was strongly met for all standards. Range was strongly or moderately met for most standards except for Language of Mathematics and Language of Social Studies. The Balance criterion was strongly met for most standards except for Language of Social Studies.

Grades 9-12

Table 38: Level Alignment of Grades 9-12 Listening ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	19	53%	91%	0.77	YES	YES	YES	YES
Level 2	19.43	48%	97%	0.63	YES	WEAK	YES	WEAK
Level 3	13.14	47%	83%	0.76	YES	WEAK	YES	YES
Level 4	9.86	76%	69%	0.78	YES	YES	YES	YES
Level 5	17.14	83%	71%	0.76	YES	YES	YES	YES

Table 39: Standard Alignment of Grades 9-12 Listening ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	34.43	54%	85%	0.71	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	12.71	76%	80%	0.78	STRONG	STRONG	STRONG	STRONG
Language of Mathematics	15.86	58%	90%	0.76	STRONG	STRONG	STRONG	STRONG
Language of Science	8.71	60%	77%	0.77	STRONG	STRONG	STRONG	STRONG
Language of Social Studies	10.29	42%	83%	0.78	STRONG	MODERATE	STRONG	STRONG

Grades 9-12 listening alignment results are presented in Tables 38 and 39. In Table 38, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at all levels. The LDL level was also sufficient at most levels except for levels 2 and 3. Range was met at all levels. Balance was met at most levels except for weak Balance at level 2.

In Table 39, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for all standards. The LDL criterion was strongly or moderately met for all standards. Range and Balance were strong for all standards.

Table 40: Level Alignment of Grades 9-12 Speaking ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	8.57	90%	63%	0.79	YES	YES	YES	YES
Level 2	5.86	45%	51%	0.82	YES	WEAK	YES	YES
Level 3	3.71	67%	43%	0.89	YES	YES	WEAK	YES
Level 4	4.86	71%	46%	0.70	YES	YES	WEAK	YES
Level 5	2.29	100%	34%	0.98	WEAK	YES	NO	YES

Table 41: Standard Alignment of Grades 9-12 Speaking ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	11.29	71%	74%	0.81	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	7.57	60%	40%	0.62	STRONG	STRONG	MODERATE	MODERATE
Language of Social Studies								
Language of Mathematics	6.58	92%	42%	0.49	STRONG	STRONG	MODERATE	LIMITED
Language of Science								

Grades 9-12 speaking alignment results are presented in Table 40 and 41. In Table 40, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at most levels. However, weak Categorical Concurrence was found at level 5. The LDL level was sufficient at most levels except for level 2. Range was only sufficient at levels 1 and 2. Balance was sufficient at all levels.

In Table 41, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for all standards. The LDL criterion was strongly met for all standards. Range was strong or moderate for all standards. Balance was strong or moderate for most standards except for the combined unit of Language of Mathematics and Language of Science.

Table 42: Level Alignment of Grades 9-12 Reading ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	11.71	77%	89%	0.73	YES	YES	YES	YES
Level 2	15.29	69%	94%	0.74	YES	YES	YES	YES
Level 3	19.29	60%	89%	0.72	YES	YES	YES	YES
Level 4	23.57	74%	91%	0.73	YES	YES	YES	YES
Level 5	14.86	88%	89%	0.78	YES	YES	YES	YES

Table 43: Standard Alignment of Grades 9-12 Reading ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	14.57	61%	89%	0.82	STRONG	STRONG	STRONG	STRONG
Language of Language Arts	32.29	74%	94%	0.81	STRONG	STRONG	STRONG	STRONG
Language of Mathematics	13.57	71%	91%	0.77	STRONG	STRONG	STRONG	STRONG
Language of Science	15.14	83%	94%	0.80	STRONG	STRONG	STRONG	STRONG
Language of Social Studies	10	72%	83%	0.83	STRONG	STRONG	STRONG	STRONG

Grades 9-12 reading alignment results are presented in Tables 42 and 43. In Table 42, the domain by level alignment showed that reviewers identified a sufficient number of aligned items for Categorical Concurrence at all levels. The LDL level was also sufficient at all levels. Range and Balance were sufficient at all levels.

In Table 43, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for all standards. The LDL criterion was strongly met for all standards. Range and Balance were strong for all standards.

Table 44: Level Alignment of Grades 9-12 Writing ACCESS Items

Levels	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Level 1	5.14	92%	66%	0.82	YES	YES	YES	YES
Level 2	4.14	60%	51%	0.89	YES	YES	YES	YES
Level 3	4.29	48%	46%	0.90	YES	WEAK	WEAK	YES
Level 4	3.86	71%	54%	0.89	YES	YES	YES	YES
Level 5	2.86	100%	37%	0.93	YES	YES	NO	YES

Table 45: Standard Alignment of Grades 9-12 Writing ACCESS Items

Standards	Alignment Statistics				Alignment Findings			
	CAT	LDL	Range	Balance	CAT	LDL	Range	Balance
Social & Instructional Language	2.86	68%	49%	0.98	STRONG	STRONG	MODERATE	STRONG
Language of Language Arts	8.57	70%	74%	0.81	STRONG	STRONG	STRONG	STRONG
Language of Mathematics	1.86	73%	31%	0.81	STRONG	STRONG	LIMITED	STRONG
Language of Science	5.43	73%	66%	0.87	STRONG	STRONG	STRONG	STRONG
Language of Social Studies	1.71	100%	34%	1.00	STRONG	STRONG	LIMITED	STRONG

Grades 9-12 writing alignment results are presented in Table 44 and 45. In Table 44, the domain by level alignment showed that reviewers identified sufficient number of aligned items for Categorical Concurrence at all levels. The LDL criterion was strongly met at most levels except for level 3. Range criterion was met at most levels except for weak Range at level 3 and insufficient Range at level 5. Balance was sufficient at all levels.

In Table 45, the domain by standard alignment showed that the Categorical Concurrence criterion was strongly met for all standards. The LDL criterion was strongly met for all standards. Range was strongly or moderately met for most standards except for Language of Mathematics and Language of Social Studies. The balance criterion was strongly met for all standards.

Summary

Summary of the Domain by Level Alignment

The level alignment showed that the match and depth criteria were met for almost all domains and levels, while weak Range and Balance findings at some levels caused moderate or weak alignment.

For pre-Kindergarten to Kindergarten, most language domains had moderate alignment for all levels. The alignment criteria were sufficiently met but the majority of levels had limited range in all language domains. For the grade cluster 1-2, language domains had moderate alignment due to weak findings for depth and breadth (particularly in the Range criterion). Categorical Concurrence and Balance were met for most levels, but level 2 showed weak depth in reading and writing. Moreover, levels 1-4 did not meet the Range criterion. For the grade cluster 3-5, listening and reading showed adequate alignment at all levels. Levels 4 and 5 in speaking showed weak breadth. The majority of the levels showed weak alignment in writing. For the grade cluster 6-8, all levels in listening showed adequate alignment. Speaking, reading and writing exhibited moderate alignment because two or more levels (e.g., level 2 and level 5) did not meet the breadth criteria. For the grade cluster 9-12, reading had adequate alignment that fulfilled all alignment criteria at all levels. Listening, speaking and writing had moderate alignment because two or more levels were either weak or did not meet the depth and breadth criteria.

Summary of the Domain by Standard Alignment

The domain by standard alignment showed strong or moderate findings across the domains and grade clusters. Though the match and depth criteria were generally strongly or moderately met, limited range, which is one of the breadth criteria, was found in the majority of the domains. Of the domains, writing showed the most limitations for range across grade clusters.

Findings from the standard alignment for Pre-Kindergarten to Kindergarten showed that listening and writing had moderate alignment because of limited depth and range in some standards. Speaking and reading had weak alignment due to limited match, depth and breath criteria in some standards such as the Language of Language Arts, the Language of Mathematics, or the Language of Social Studies. The 1-2 grade cluster showed that the alignment criteria for match and depth were strongly or moderately met for almost all standards and domains. Limited range was found for some standards: the combined unit of Language of Language Arts and Language of Social Studies standards in speaking and the Language of Mathematics and the Language of Social Studies in writing. For grade cluster 3-5, listening, speaking, and reading showed strong alignment for all standards. Alignment findings for writing indicated that the match and breadth criteria were limited for the Language of Social Studies. Moreover, limited range was found in the Language of Language Arts. For the grade cluster 6-8, the alignment criteria for match and depth were strongly met for all standards and domains except the Language of Social Studies (in Writing). The breadth findings were moderately met, with the exception of Range. Speaking and writing domains had one or more standard with limited alignment findings for Range or Balance.

For the grade cluster 9-12, the alignment criteria for match and depth were strongly or moderately met for all standards.

Reliability among Committee Members

The following table shows the intraclass correlation coefficients³ for each grade level, which indicate the degree of agreement among committee members in each group. Values larger than 0.7 indicate a good level of reliability among committee members (Subkoviak, 1988). This criterion has been met for the majority of groups in this alignment study. It should, however, be noted that these values are highly dependent on the number of committee members in each group. The pairwise comparison values represent the average agreement for pairs of committee members in each group. As a previous alignment report argued⁴, a result of 0.6 or higher reflects reasonable agreement, 0.7 or higher demonstrates good agreement, and a result of less than 0.5 reflects poor agreement among pairs of committee members.

Table 46: Reliability among Committee Members

Domain	Standards	Number of standards (MPIs)	Number of Committee members	Intraclass correlation	LDL Pairwise comparison
<i>Kindergarten</i>					
Listening	MPI Listening	25	7	0.91	0.62
Speaking	MPI Speaking	25	7	0.96	0.67
Reading	MPI Reading	25	7	0.97	0.73
Writing	MPI Writing	25	7	0.84	0.50
<i>Grades 1-2</i>					
Listening	MPI Listening	25	7	0.88	0.55
Speaking	MPI Speaking	25	7	0.94	0.71
Reading	MPI Reading	25	7	0.91	0.60
Writing	MPI Writing	25	7	0.89	0.49
<i>Grades 3-5</i>					
Listening	MPI Listening	25	7	0.89	0.58
Speaking	MPI Speaking	25	7	0.97	0.83
Reading	MPI Reading	25	7	0.90	0.61
Writing	MPI Writing	25	7	0.96	0.79
<i>Grades 6-8</i>					
Listening	MPI Listening	25	6	0.88	0.54
Speaking	MPI Speaking	25	6	0.98	0.81

³ An intraclass correlation coefficient (ICC) serves as an index of the reliability of more than two raters rating the same set of items (Shrout, P.E., & Fleiss, J.L., 1979).

⁴ Cook, H.G. and Wilmes, C. (October, 2007). Alignment Between the Kentucky Core Content for Assessment and the WIDA Consortium English Language Proficiency Standards. Madison, WI: University of Wisconsin, Wisconsin Center for Education Research.

Domain	Standards	Number of standards (MPIs)	Number of Committee members	Intraclass correlation	LDL Pairwise comparison
Reading	MPI Reading	25	7	0.79	0.51
Writing	MPI Writing	25	7	0.93	0.66
<i>Grades 9-12</i>					
Listening	MPI Listening	25	7	0.89	0.54
Speaking	MPI Speaking	25	7	0.95	0.65
Reading	MPI Reading	25	7	0.90	0.60
Writing	MPI Writing	25	7	0.95	0.70

References

Cook, H.G. (2006). Aligning English Language Proficiency Tests to English Language Learning Standards. In *Aligning Assessment to Guide the Learning of All Students*. Council of Chief State School Officers; Washington, D.C.

Cook, H.G. & Wilmes, C. (October, 2007). Alignment Between the Kentucky Core Content for Assessment and the WIDA Consortium English Language Proficiency Standards. Madison, WI: University of Wisconsin, Wisconsin Center for Education Research.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 2, 420-428.

Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement*, 25(1), 47-55.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6. Madison: University of Wisconsin, Wisconsin Center for Education Research.

Webb, N.L. (1999). Alignment of Science and Mathematics Standards and Assessments in Four States. Research Monograph No.18. Madison, WI: National Institute for Science Education University of Wisconsin-Madison.

Webb, N.L. (2001). Reviewer Background Information and Instructions, Mathematics Standards and Assessments Alignment Analysis, CCSSO TILSA Alignment Study. Unpublished Document.

WIDA Consortium (2007). WIDA English Language Proficiency Standards and Resource Guide. Madison, WI: State of Wisconsin.

Appendix

Reviewers' comment

WIDA ACCESS Kindergarten Listening

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- I did not see many items that covered science topics.
- It seemed that this assessment did not cover all the standards - either that, or I am having difficulty identifying the additional standards each item covered.
- All the standards were evenly distributed except for social and instructional standard. I believe it was more heavily represented, particularly in the Part A portion of the listening test. I did not consider "point to the bear" an LA standard, however, in looking at the standards, I think "identify pictures of make-believe animals as modeled orally" might fit that description. Perhaps I am wrong in not assigning more standards to the first half of the test. It seemed more real in the second half. Perhaps because the first half was in a home.
- Yes, the items cover the most important topics
- I was confused by the living/nonliving distinction in the science domain. It's actually irrelevant. Children identify and describe various living and nonliving things but never have to distinguish the two concepts. Patterns seemed underrepresented or at least if patterns were being assessed, it wasn't obvious.
- repeat, describe, complete, make, use

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- Yes
- I felt that all the levels were assessed equally.
- As we moved up the levels, I think there should be less isolated commands. Following a story line rather than independent multi-step directions would require more complex discourse listening skills. I think it would better assess the higher levels. A good example of higher-level discourse is D4 questions using rhyme that the students need to have a better command of language to infer a bit. A riddle format? Perhaps it is the way of a listening assessment.
- Yes the items cover the most important performance levels
- Yes

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Yes
- For the most part, the wording was very appropriate toward not only the grade level, but also getting you to the desired answer.

- Standards are directed towards expectations appropriate for grade level but need more specificity. For example, there seems to be debate as to what the MPI is asking the student to do at each level. We debated whether the standard was at a different level because we all had a different idea as to what the students were expected to do for that level. Another example would be the line "based on oral directions," only once did I see a distinction between a level four and five of "according to descriptive oral discourse," rather than just oral direction/description. Specificity isn't addressed, or maybe that is implied with the performance definitions.
- Yes
- The current standards are heavy on function, light on content words, but success on the assessment depends on knowing specific content words, it seems.
- I think some were too high....Linguistically difficult for their age and grade....Too much info and directions in one step..

D. What is your general opinion of the alignment between the standards and assessment:

- i. Perfect Alignment (1) : 12%
- ii. Acceptable Alignment (4) : 50%
- iii. Needs slight improvement (2) : 25%
- iv. Needs major improvement (1) : 12%

E. Comments

- I think that more items should address science standards.
- Based on the input summary report, there appears to be a need for major improvement, but I think this is because I had only the Summative Framework with example topics to use as a guide. I believe I could have responded to the alignment task better if I had in front of me the general standards and not the standards in the context of specific topics. My state, Virginia, just adopted the standards last year, and as a K-4 teacher with two sets of standards, state and WIDA, spanning five grades, I cannot remember wording off the top of my head.
- Students at Kinder level have selective listening, so results of performance are restrictive/limited.... Average vocabulary is limited at this age....Too much description and direction.....Are we testing for gifted and talented at this age in certain item questions.....Language acquisition compared to a native speaker at this age is what????

WIDA ACCESS Kindergarten Speaking

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- No, I felt like the items did not cover science well.
- Yes
- I did not find anything to do with rhyming. I am not sure if I just did not understand how to change the thread to adjust or what.
- Yes

- There did not appear to be sufficient coverage of math and social studies.
- adaptation of language to certain audience.....

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- I feel like the level 3 of linguistic difficulty is never really reached in the speaking test. We need more rigor in this test as well as tests across grade level clusters. The level of rigor in this test does not reflect the level of performance that students are asked to achieve in schools.
- Yes
- I think that it was fairly evenly spread over difficulty levels.
- I feel many items covered the standard's content area but did not stretch to a level five of the standards, so it did cover the DOK levels, maybe just not the five proficiency levels.
- Yes
- Yes

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- I think there is not enough rigor in this test.
- Yes
- I think the questions that were there did address the expectations.
- Could use more specific information as to the length of the utterance. Some had it specified, others did not. Vocabulary specificity is also not addressed, ie "describe different types of homes or habitats from illustrate scenes using phrases or short sentences, using specific location or ?? words.
- Yes
- Yes

D. What is your general opinion of the alignment between the standards and assessment:

- i. Perfect Alignment (1) : 10%
- ii. Acceptable Alignment (4) : 40%
- iii. Needs slight improvement (4) : 40%
- iv. Need major improvement (1) : 10%

E. Comments

- There needs to be more complex speaking items in the test.
- I appreciate the clarification this morning on applying examples of standards to different topics, but it is still hard to spot some cross-curricular topics like time and schedules as a math topic. I strongly recommend giving alignment participants copies of the general

standards and if possible, topic areas for each subject. Identifying science, social studies, and math topics for grades 1-4 is fairly easy, but kinder is difficult for those of us without experience as a kinder classroom teacher.

- I still think the compare and contrast with certain items should be very straight forward....small big, color...etc.

WIDA ACCESS Kindergarten Reading

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- I felt there weren't many questions in the mathematics standard. Also, there weren't many (if any) questions in levels 2-4.
- Yes. Perhaps more comprehension is needed?
- I didn't notice a questions for the Language of Language Arts standard addressing the standard's PreK-K benchmarks of print/non print materials. I think there could be more question related to the math and language arts standards included in the reading portion.
- Based on my ratings, most standards received sufficient coverage. However, many were making connections that a topic "could cover" a science or social studies topic. The math standard could use more representation today.
- Not all standards were represented.
- There were some topics that were not covered in the items, especially under Math and Science. But, it should be noted that we need to keep in mind what and how much information is expected of a Grade K student before creating items using academic language items for ELL Grade K students.
- The language of the 4 contents was not as explicit as needed. SI and LA assessed items that are specific to those standards are also secondary within content language. The primary alignment could have been more closely aligned to the ever-needed and sought assessment data on content language ability.

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- Again, not many LDL in 1-2.
- Yes
- Yes, the questions seemed to go from low DOK levels to higher ones as the test progressed.
- The lower and middle linguistic difficulty levels were not as well represented.
- Level 3 was assessed the most. There were not enough items at level 2.
- Some of the LDLs were less covered than others. There were also some items that were not scaffolded with varying LDLs. This might be something to look at so that ELLs would be able to demonstrate their current levels of language proficiency even if they could not answer an item with a high LDL. (This is for those items that they all need to try before stopping a section and moving on.)

- The linguistic difficulty levels were well represented within the ACCESS items, level 1,2 and 3 aligned with the consensus reached within the cohort. This was the most effective alignment I have seen in both alignment studies I have been involved in.

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Yes
- Yes the standards were written at an appropriate level for kindergarten students. I tend to have higher expectations and therefore identified the DOK levels at a 2 for the final section. In our district, we use the Scott Foresman reading program and this has higher expectations on comprehension and fluency.
- Yes. I think thought the questions were developmentally appropriate for the grade level.
- The standards were geared more towards the higher linguistic difficulty levels for kindergarten, especially considering the time frames for when the testing is conducted.
- There needs to be more specificity in the standards for this grade level.
- There were some MPIs that were not clear due to the varying interpretations of the MPIs and their expectations. Also - we need to take into consideration which of the "basic" grade K skills are actually expected of mainstream English-speaking grade K students at certain points in time as we make decisions as to which skills to assess our ELL grade K students on.
- Specificity of the MPIs is the most critical issue with the use of the WIDA Standards. The "picture" needed of "what we are asking an ELL to do within the MPI" is inconsistent per educator. The vague representation and interpretation has to be debated within any consensus building versus the actual alignment focus.

D. What is your general opinion of the alignment between the standards and assessment:

- i. Perfect Alignment (1) : 12%
- ii. Acceptable Alignment (2) : 25%
- iii. Needs slight improvement (5) : 62%

E. Comments

- Include more subject and LDL variety. I believe it would keep the student more focused.
- Thank you. Excellent work by WIDA.
- I had a difficult time distinguishing between the questions pertaining to science standards and social studies standards.
- There were some test items that seemed to go even beyond the expectation of the standards -- like matching the sentences to the pictures. Not all standards were represented.
- Some of the comments are listed above since they were tied into the responses. Thank you.
- Besides the need for more items with the primary assessment being within the content of Science, Social Studies, and Mathematics.

WIDA ACCESS Kindergarten Writing

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- I didn't see any items pertaining to environmental print.
- Difficulty attaching a standard for B1 and E1 (Tracing and copying name); Standard for initial letters; but not final letters.
- No, the assessment did not cover the standards of math and science. I think the writing task was weak in the areas of social and instructional and language arts.
- The assessment did not represent the standards equally based on the summative MPIs presented. This was especially the case at the mid proficiency levels. Clearly this is a function of this type of assessment in Kindergarten. Is it appropriate to force a the rating the linguistic difficulty level when it does not match the construct of the test?
- Some of the test items did not seem to correspond to the writing domain, but rather to the reading domain. (E3) The only way I see this as a writing question is if it relates to the correct formation of a letter, and I don't see that in the standards.
- Regarding topics - there seemed to be a lot of items on the Language of Social Studies. There did not seem to be many items on the Language of Math and the Language of Science. But, the question of how much "academic language" should a grade K student know at a given point in time should be researched. We seem to be requiring our ELLs to be familiar with some academic language that even our mainstream English-speaking students have still not been exposed to. This is why it is difficult to respond to the question of "what topics were not assessed that should have been". There seemed to at times be a question of the extent of correlation between the things included in the topic examples/test items and what our state standards require.
- The writing was weak in regard to the linguistic difficulty, the transition from tracing and copying to independent production of text does not assess the writing domain of the standards (or should be noted that the ability gap between the items is apparent without a transitional expectation). The reading standards that are assessed within the writing (beginning/ending sounds) does not align to nor assess the writing standard embedded within.

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- There was a good representation of all LDL.
- Yes
- It was difficult to assign a DOL to some of the questions because they were accepting a variety of answers.
- Most of questions for kindergarten were written so that students would be encouraged to write what they could. While appropriate, assigning the linguistic difficulty level is mainly dependent on rating what the students produced, rather than the design of the question. The lower linguistic difficulty levels were not well addressed.

- It was difficult to determine the linguistic difficulty level for some of the items since the test administrator had the option of providing some scaffolding to the student to get some writing down on paper. (B2-5, E4-5)
- Some items did not include level 3 LDLs. There were items that also did not include the lower level ones that could be used for those students who might not be at that higher level. Some of the items did include those "scaffolded" tasks that could assist in determining the student's appropriate language proficiency levels. The extent to how much or how well an ELL Kindergartener should be able to perform at a certain point in time needs to be taken into consideration when performance expectations are decided. They are a "special" group.
- The performance expectations had a large gap and some of the items required a reading standard expectation to reliably assess the expectation of the writing standard.

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- I like items B2/3/4/5 and E4/5 that allows the student to produce writing on their level.
- Yes, I do feel the writing expectations are appropriate for beginning, mid year and end of year student. I feel these expectations are as they should be and will prepare the student for grade 1.
- The standards and assessment appeared to be written for kindergarten age students. The fact that students can write what they are able to whether it is a letter, word, phrase, sentence, or story is appropriate for this age group. Tracing and copying letters is also an appropriate assessment for kindergarten age students.
- For the most part they were directed appropriately towards expectations of kindergarten.
- The items contained performances that were not clear at times - "Depict, express, relate, represent, etc." There were also some discrepancies with what "icons" and "models" were referring to. Some of the MPIs were vague so it was difficult to assign LDLs to them. Is this question asking about the Standards or the MPIs? At times, it was confusing was when the MPIs were referred to as the "standards" when the standards, according to my understanding, are actually the 5 statements, e.g. - "ELLs communicate for Social and Instructional purposes within the school setting.; etc." This also happened when we had a WIDA presenter come to our state to do a presentation. The MPIs were called "standards" and it got really confusing. Even the handout that was used incorrectly used the terms.
- Specificity has always been a concern within the performance indicators that many teachers use to access and use the standards. The MPI are crucial to mainstream teachers being able to support and assess their ELLs.

D. What is your general opinion of the alignment between the standards and assessment:

- i. Perfect Alignment (1) : 12%
- ii. Acceptable Alignment (2) : 25%
- iii. Needs slight improvement (3) : 38%
- iv. Needs major improvement (2) : 25%

E. Comments

- Thank You WIDA for all your efforts in creating such excellent assessments for the students, teachers and administrators.
- It was difficult to see the correlation between the standards and the writing assessment.
- I think that I wrote some of the comments that should have been in this box in the above boxes since they were tied into the questions. Thank you.

WIDA ACCESS Grades 1-2 Listening

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- Not really... it seems like some standards were represented over and over while others were not touched at all. I would expect to see more lower standards at tier a and more higher standards at tier c, but I think they could be distributed more evenly.
- Yes
- All standard areas were addressed to some part on this assessment.
- Topics related to the content stem is irrelevant. However, I believe the items related to landforms are that of science, not SS. These tests (A-C) seemed to need more science related items.
- For this grade level, Language Arts and Math are the key standards. The others are important but students must be able to read and comprehend. Basics in math computation is critical also for future success in science. The logic of maps proves difficult for student. Very often the compass rose is introduced to students during this test. Then I wonder if I am assessing their understanding of the directions/compass rose or are they able to follow directions.
- Yes
- Yes

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- I think that this test was fairly rigorous with the language used, across tiers and that is good. I think that in some items, there is cognitive complexity when we really need linguistic complexity, so that needs to be worked out.
- Yes
- The lower standards were addressed, but not the higher ones. (e.g. LA1, LA2, LA3, but not LA4 or LA5).
- It appeared the Tier A test covered DOK levels 1-3, which would make sense. Tier B seemed to be DOK levels 2 and 3. The Tier C test did not seem to have as many DOK level 3 as I would have expected.
- Yes
- Higher level science and social studies not covered in listening

- Yes

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Over all, yes, but the map questions are too complex for first and second graders. I see them year after year after year and see the same results for my students. I think those map questions should be replaced with something else. Following map directions is a complex skill for Native English speakers... ELL's need something different.
- I did not see as many Level 5 items in Tier C as I expected.
- Most questions were written appropriately. Some feel a little stilted - especially the directions using a compass rose.
- Standards are written at appropriate level, items on the test do not appear to be at a challenging level, particularly for tier C.
- Yes
- Yes
- Map and directions are higher level for this age group especially for a 1st grader.....Too much info.....

D. What is your general opinion of the alignment between the standards and assessment:

- i. Perfect Alignment (1) : 12%
- ii. Acceptable Alignment (2) : 25%
- iii. Needs slight improvement (4) : 50%
- iv. Needs major improvement (1) : 12%

E. Comments

- The map questions skew the test results to make it look like our students are lacking social studies language. Social studies encompasses so much more than just map skills... I would think that you could find a different skill to use on the test.

WIDA ACCESS Grades 1-2 Speaking

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- I think that the speaking test does a better job of covering all standards than the listening test.
- Yes
- It was probably just me, but I did not find many (if any) questions that addressed math, science, and social studies.
- Standards not fully addressed would be SS.
- Yes

- Yes
- pretty much, I would like to see/hear the students formulate questions....You had this in prior testing....

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- I think that more complex speaking items should be in the speaking test. There were few items that required level 3 linguistic difficulty.
- Yes
- I feel that Language Arts was covered sufficiently. Social Interactions was covered to a lower degree, the others were not covered sufficiently.
- DOK levels for social and instructional are 1-2, which is appropriate for the tasks. The other ELP standards are evenly distributed 1-3.
- Yes
- Need more LDL 3s for all subjects but math.
- Yes

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- I think that more complex speaking items should be in the speaking test. There were few items that required level 3 linguistic difficulty.
- Yes
- Most of the items were very appropriate as far as audience and expectations.
- The items need to have higher proficiency level tasks for SS. Social Studies seems to be misrepresented.
- Yes
- Yes
- Yes

D. What is your general opinion of the alignment between the standards and assessment:

- i. Perfect Alignment (1) : 12%
- ii. Acceptable Alignment (4) : 50%
- iii. Needs slight improvement (3) : 38%

WIDA ACCESS Grades 1-2 Reading

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- This was a very difficult task b/c the questions didn't match up with the standards.

- Should more phonemic awareness be added...more blending of nonsense words focusing on medial vowels; Should more word study exercises be tested...compound words, prefixes and suffixes
- There were not many questions assessing the language of Language Arts standard. I felt Science and Math were assessed very well.
- Most of the items only reflected the strand represented. Math was largely covered only in estimation. That is not representative of the broader language of math. This is representative of the MPIs but not helpful in the broader context of teaching ELLs. In most areas I found an MPI that relates, but generally it was by stretching what the MPI truly stated to reflect a much more difficult task.
- No. I think there was an over emphasis on living things and estimation. I am not sure what topics should be covered, but surely there are a greater variety of topics in science and math that could be covered.
- There seemed to be a lot of items that dealt with estimation (math). Is this the most important thing in math for grades 1-2? It was difficult to tell since some of the items seemed to be under more than one standard. This could be because I needed to look over several MPIs to find a match. More of the grade level expectations need to be researched to see if enough of the relevant topics are covered.
- The topics were covered well within all three tiers as a separate whole. If assessment contained the various topics used within all 3, coverage is almost perfect between, SI, Science, Math, LA, and Soc Studies (a little weak in Soc Studies).

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- I felt all three levels had mainly 2-3 LDL.
- My students have a difficult time with math/reading; perhaps there should be more "2" DOK levels;
- I didn't think there were an appropriate number of DOK levels 1 and 2 for each standard. I thought many of the questions were a DOK level 3, especially in the science and math standard. The reading required and linguistic ability did not match the DOK levels of the standards that we evaluated in our groups. I felt like it was hard to assign a DOK level based on the standards that we discussed as a group as many of the questions were level 3 DOK questions.
- No. Based on the MPIs, the higher level linguistic demand level was covered much more significantly. The lower and middle levels were not. A large part of that is due to the construction of the items. Having a picture that relates generically to the topic is not the same as have text supported by illustrations. This caused many items to rise to higher linguistic demand, based on what students in these grade levels can do and the representation by the MPIs.
- There seemed to be not enough of LDL2 items. Also - for the higher tier, shouldn't all items be LDL3 items? I think that there were some of the lower level ones. Likewise - shouldn't all tier A items be LDL1 items and all tier B be LDL2 items? Not sure if I am looking at the

intent of the levels correctly, but I understood that they were leveled so that the "appropriate" test could be administered to the students.

- The LDL levels were either a 1 or 3 with an Unexpected inconsistency as consensus on MPIs were reached by cohort. The 3's were much more difficult in regard to the expectation in the MPIs, the amount of reading and text that was not easily decodable with sight word matches were a level 3 example of an MPI or standard that was more simplified in interpretation of the framework.

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- I think the level is harder/higher than 1-2 grade
- PreK-K..Standard 5 Reading Level 2 (various sources- need more examples) Grades 1-2...Standard 2: Reading Level 2, Perhaps should say "segment sounds and blend/ Level 3 Instead of Cover Up....perhaps onset and rhyme;
- I thought the standards were written at an appropriate grade level, but the questions in the test were fairly difficult. In my opinion, many of my ELL students would struggle with the DOK level 2 and 3 questions.
- I think the items were appropriate for students who are closer towards grade level expectations, but not for students at lower levels of proficiency.
- I found that the standards were not directed towards expectations appropriate for the grade level. In some cases, the level of expectation was higher than that expected of a typical student at the grade level. There was a great deal of written language (that students are expected to read) in some items that I thought was too much for the grade level being tested.
- It was difficult to tell due to the MPI descriptions. Many of the items seemed to be under a certain standard (e.g. - social/instructional language; math; science; or social studies), yet the closest linguistic function was listed in an MPI under a different standard.
- Either too decodable or too complex for grade level 1-2.

D. What is your general opinion of the alignment between the standards and assessment:

- i. Perfect Alignment (1) : 12%
- ii. Acceptable Alignment (1) : 12%
- iii. Needs slight improvement (4) : 50%
- iv. Needs major improvement (2) : 25%

E. Comments

- I am not sure how to answer this b/c I'm uncertain of my answers. I do think the content is too hard.
- I often had a difficult time seeing the connection between the standards and the test questions. I think this could be improved by reviewing both the standards as well as the test questions.
- Overall, I'm disappointed in the quality of this assessment, especially in connection to the standards it is supposed to represent. It was extremely difficult to match tasks students were

expected to accomplish compared to our groups consensus on the linguistic demand levels. I think in the future, it would be helpful to develop more specific definitions of LDL for younger grades.

- In too many cases I could not find the connection between a test item and a standard. It seems that many of the items simply expected students to match a picture to a written statement or find an answer in the text.
- Some of the things being asked on the test need to be looked at more closely to ensure that it is indeed academic "language" and not academic "content" that is being tested. Some of the items seemed questionable. Also - some of the MPIs were confusing. We need to make sure that we are not expecting our ELLs to "know" things that our mainstream English-speaking students are not expected to know at a certain point in time. We need to remember that our ELLs should be assessed on what they have been exposed to and are expected to know based on their length of time in our school system. This is a concern since some school districts, such as mine, do annual reassessments in January/February. I am not sure if some of the topics on the items would be covered by that time. This would ensure that a more "valid" assessment will be done. Thank you.

WIDA ACCESS Grades 1-2 Writing

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- Yes
- Yes
- I didn't think there were questions that covered the social studies, math, and language arts standards.
- The writing assessment generally covered the topics represented by the MPIs. This can provide a representative sample of writing in context of each of the standards.
- There seemed to be more topics under social/instructional language and science. When looking at "academic language", more alignment with individual state standards might be needed. This would provide a more informative response to the question.
- The topics assessed were comparable to the reading items. SI, Science, SS, MA, and LA were represented well and equally, this being the best feature within these items.

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- Yes
- Yes
- I thought the writing section lacked questions that required a DOK level 2. There were many questions that were either a 1 or a 3, but few, if any, that were a 2.
- The LDL was representative of the MPI and the consensus of our group. Of course students performance on each question may reflect different LDLs.

- Once again (as in my reading section response) - the question of which LDLs should be covered in each level needs to be addressed. For Tier A, should it be LDL1 items; for Tier B, should it be LDL2 items; and for Tier C, should it be LDL3 items? Weren't the tiers devised so that ELLs at different levels in proficiency would be administered a test that was "appropriate" for them? There also seemed to be a limited number of LDL2 items.
- The level 3 LDL for many writing items is due to the language that is needed to perform the expectation that would be needed to meet the standard. In the standards and MPI framework there is not a correlation to the test items, the amount of writing is not represented in the standards.

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Yes
- Yes. Minor Notes: The term reproduce...is this meant the same as "copy";
- I thought the content was appropriate for the grade level. In my opinion, the expectations in the last part of the writing test are a bit high. I don't think that most first and second grade students should be expected to write 4-6 sentences about a topic. For the tier B and C tests, they are required to write even more than one writing passage with 4-6 sentences each. This is a big expectation for first and second grade students. I think both ELL and mainstream students would struggle.
- Yes
- Some of the MPIs were kind of confusing the way that they were written. Also - some of the expectations need to be reviewed to see if they match what the student is required to know at a certain point in time.
- Depending on your expectation, the writing assessment had a 3 LDL in most cases and would have been so for higher grade levels than 1-2. UNLESS we are assessing grade level expectations, which would have more correlation to the test items we have looked at.

D. What is your general opinion of the alignment between the standards and assessment:

- i. Perfect Alignment (1) : 12%
- ii. Acceptable Alignment (4) : 50%
- iii. Needs slight improvement (2) : 25%
- iv. Needs major improvement (1) : 12%

E. Comments

- I think there needs to be more writing assessment pieces with level 2 DOK questions. There is a big difference between the DOK level 1 questions and the level 3 questions.
- Thank you for the opportunity to participate in this alignment study. (By the way - any chance of an interim ACCESS reassessment tool? Currently, a student cannot be exited since WIDA does not have an interim reassessment tool. There is only the annual ACCESS

assessment tool. Please consider this so that our ELLs can have that option to exit if they show a lot of progress and meet the exit criteria prior to the annual reassessment. Thank you.)

WIDA ACCESS Grades 3-5 Listening

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- NO, I did feel as though there enough items at the higher levels. There were many between levels 1-3, but not as many between 4-5.
- The assessment was what I had expected because I have given the assessment multiple times.
- Yes, they did.
- I felt the items did cover pertinent topics for the grade levels. Force and motion and the body are in many curriculums for @ 4th grade. Trade routes & early travel is pertinent to Soc. Studies curriculum, Ghana-only in that they may do reports on different countries at some point.
- Yes
- The training we received at breakfast was very vague and confusing. Most items did cover important topics measured by the standards.

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- Yes, there was a fairly even distribution of LDL levels across the listening section.
- It did, I thought there could have been more DOK level 3.
- Yes, they did.
- The technical language of math, social studies, and science seemed to be at a much higher level than the technical language of language arts. The word prediction was used, but not too much other technical language of that field was used, so I rated the linguistic difficulty lower in these instances. The complexity seemed to be more in the what was being asked in some questions. The understanding of the language of language arts was not used as much.
- Yes

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Yes. The topics and questions were very grade level appropriate for the listening section. The questions and responses were very specific as was the visual support.
- I think they were appropriate and aligned well with the assessment, however, there was a lot of information in many questions that was not needed to answer the questions. That may be confusing for ELLs.
- I believe so.

- The questions did have very specific math language. Ex. Part F: 3-D Shapes- The students should be able to answer these questions given the drawings and the use of the words in the text. Some of these will be more difficult for the third graders, but as they proceed into the next two grades 4th and 5th, they should come in contact with this vocabulary and concepts. The charts & maps are another example of the specific language that should be grade appropriate.
- Yes
- Yes

D. What is your general opinion of the alignment between the standards and assessment:

- i. Perfect Alignment (2) : 25%
- ii. Acceptable Alignment (6) : 75%

E. Comments

- I felt as though the test had many questions at the social and instructional and language arts standards. I do not know what the breakdown needs to be for all 5 standards but not sure if enough items addressed science or social science.
- I struggled with aligning a few higher level questions.
- I liked the process. It really helps you to think and analyze what we are teaching/testing.
- This test will be difficult for my students to pass since 60% their instruction is in their native language. They will be familiar with the standards but not in English. I realize that more instruction needs to be in English but overall this test is well done.

WIDA ACCESS Grades 3-5 Speaking

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- The items covered many social and instructional language functions necessary to connect the content area functions
- Yes
- Yes, with the speaking section, expansion is able to occur as the student moves through the questions.
- These areas of testing seemed to fit the domain of speaking. There was no or limited math references, but that would be more difficult for extended discourse without technical language. So I think subject areas were good.
- Yes
- Yes, this test met the standards well.

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- Yes, there was a range of Linguistic Difficulty Levels present throughout the speaking test.
- Yes, I felt there were many level 3 LDLs.
- Yes
- Yes. There were identifying, describing, comparing and higher levels of discourse are possible with these examples.
- Yes
- The speaking test is much better prepared than the listening test.

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Yes
- Yes
- Yes with difficulty increases as several questions were asked.
- Yes. There were appropriate and had enough information from the administrator of the test and from the graphics to support the test taker in using appropriate language for answering.
- Yes
- Yes, but I feel the training that we had this morning clarified the process much better and made it much easier to continue the process. I highly recommend that you rethink your training the first day for any future groups.

D. What is your general opinion of the alignment between the standards and assessment:

- i. Perfect Alignment (2) : 25%
- ii. Acceptable Alignment (5) : 62%

E. Comments

- The speaking test aligns easily and well.
- Crisp, clean, relevant that would visually support the student in the content areas is evident.
- Thanks for included us in the process.

WIDA ACCESS Grades 3-5 Reading

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- Yes
- Yes, I think the items covered the most important topics, however I had difficulty matching the items with the most appropriate MPI for the primary one.
- Yes
- Yes

- I think that the choices made covered the topics in the best possible way without making the test too long or cumbersome.
- Yes

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- Yes
- Yes, I think that the items covered the most important topics.
- LDL was not assessed much, but this is probably due to time constraints and would make the test extremely long.
- Yes
- For this grade cluster the difference in reading level between 3rd and 5th graders is great. I think that the level here is high for 3rd graders and about right for 4th and 5th.
- Yes

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Yes
- Yes, they were written at an appropriate level of specificity.
- No, but this is due to the fact that the standards and tests assess a grade band of 3-5. Grade-level text for a 3rd grade student is much different from that of a 5th grade student.
- Yes
- I would have liked more math related examples.
- Yes

D. What is your general opinion of the alignment between the standards and assessment:

- ii. Acceptable Alignment (7) : 88%
- iii. Needs slight improvement (1) : 12%

E. Comments

- I am wondering if the "Can do" statements would have been helpful with this study.
- I had a great deal of difficulty with part two as did many of the members in the group. Many of us felt we really were unclear about how to apply the MPIs to the items and seemed to get different explanations of how to do it. I tried to take the information given to us and add what I know and I hope that I represented it effectively.
- The process was not very clear. I would have benefitted from completing a few examples of the alignment process with old (no longer used) items with the entire group.
- Some problems have incorrect information and need to be corrected. ie. Math

- On the reading standards examples I found items which fit at level 3 and 4 but were not supported by graphics or visuals.
- The thing that makes this task challenging is that we have so many older students coming into our program from refugee camps with no prior literacy in their first language. We have found it takes almost a year, with systematic literacy instruction for students to attain beginning literacy skills, which are light years below grade level the older they are. When I taught ELL in Wisconsin this test felt appropriate. Now that I am in a refugee dominate district with over 50 languages and hundreds of students who have experienced severe trauma, I am wishing there was a test to measure the lower levels of literacy we are moving mountains to achieve. Coming from trauma and nonliterate cultures we are finding it takes over two years to reach a level 2 on the WIDA for older students. We are working with young girls who were not allowed to attend school, who were raped in their refugee camps, who have seen relatives murdered and homes burned. I would be nice if there were a Level 1 A and a level A B to honor the hours and hours of instruction and acculturation these students require to learn academic English.

WIDA ACCESS Grades 3-5 Writing

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- Yes
- Yes, I think that the most important topics were covered.
- Yes
- Yes
- There was no social studies topic directly assessed but the topics chosen present a good chance for all students to show their writing ability.
- Yes

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- Most of the items were at a higher DOK level.
- I don't think there are enough items covered at the lower LDL levels and it doesn't give students at those levels the opportunity to show what they know and can produce in writing.
- Yes
- I really didn't see any level 1 writing according to the types of writing asked for in the example MPI's
- The general LDL of the writing test seems high and my experience is that my students score low, but I don't believe in lowering the bar so I guess that means we need better ways to teach ELLs to write well.
- Yes

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Yes
- Yes I think the standards were written at an appropriated level of specificity and directed towards expectations appropriate for the grade level.
- With the test being written for such a wide grade span (3-5), it would be much more difficult for a 3rd grade student who may not have had exposure to some of the content (such as circuits). This imbalance could be, and hopefully already is, addressed through scoring variance by grade level.
- Yes
- Yes
- Yes

D. What is your general opinion of the alignment between the standards and assessment:

- ii. Acceptable Alignment (4) : 50%
- iii. Needs slight improvement (4) : 50%

E. Comments

- I found the writing easier to score than the reading. Perhaps that is due to the fact that I feel more comfortable with the scoring today and the directions were clearer. As a team leader I felt I had a better handle on how to conduct the consensus and we had more discussions about each MPI today. Instead of saying who wasn't in consensus, I told the group that we didn't all agree and we came to consensus through discussion rather than one person feeling on the spot to defend her position. Everyone gave their opinion and it was much easier to agree when done that way. People who tried to push their opinion on others yesterday were much better "behaved" when done this way!
- Since the summative framework does not actually encompass all the content that is assessed, labeling it as such is misleading. Why have a "formative" framework and a "summative" framework when parts of both are in the test? Let's just call it a "framework" then.
- I think the first writing task should be a little lower and increase in difficulty.
- Tier A assessment should encompass more LDL 1s, since our "newcomers" are the targeted population.
- Once again, I would mention that the expectations seem extremely high for the bottom of the tier, but I know of no way to rectify that without reversing the situation.
- I do wish there was a test to measure students who come in from nonliterate cultures. It takes a year of hard work to understand abstract concepts like letters and sounds; maps; the entire concept of expressing language through a written format. Students come to us rich in relationships - if you need directions for going somewhere, you walk with your friend, you do not Google directions so you can do it independently. There is a huge acculturation piece that is a part of literacy. I am not sure how you would do it, but I wish there was a way to measure the growth of these amazing students who must transition their worldview from one of

community to our rugged individualism in American culture. Language reflects our culture, it is more than phonemes, morphemes, and syntax. How do you measure this remarkable transformation of the soul?

WIDA ACCESS Grades 6-8 Listening

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- Yes, I believe each standard covered the essential items and topics that were expected by the standard.
- The items really closely followed the topics in the MPIs.
- I feel that the majority of the topics were indeed covered by the standards. They seem to offer a variety of topics and content areas.
- Yes
- Yes

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- Yes, the majority of the LDL levels were covered by the expected standards.
- Yes
- I believe the items covered most of the performance levels (LDL) that would be expected as a part of the standard. A couple of them were a little confusing, but when focusing on the listening aspect only, I felt they were adequately represented.
- It seemed there were a lot of level 1 and 2 questions and could have been more 3 questions.
- Yes

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- During the listening test I feel many of the standards are written at an appropriate level.
- As far as specificity, I do feel that the standards are appropriate. This is where I would address some confusion in language operations perhaps. For example differences in the terms of retell/paraphrase or define/describe as each of these could have vast differences in interpretation. I could ask a student to paraphrase or retell instructions: If this was a regurgitation of the instructions given daily this I some thing, but if it is of a new topic being presented that could be completely different.
- There could have been more advanced questions on the C level test.
- Yes

D. What is your general opinion of the alignment between the standards and assessment:

- i. Perfect Alignment (1) : 14%
- ii. Acceptable Alignment (5) : 71%

WIDA ACCESS Grades 6-8 Speaking

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- Yes, for speaking I felt that the standard did cover the most important topics.
- Yes, the standards for both the Speaking/Listening tests did a good job in covering items/topics that were expected by the standards.
- These tasks seemed to bridge standards. If the goal of a particular Part is to assess the language of a single subject area, I don't believe that the T1 and T2 questions elicit the language of that subject area well enough. If the goal is to bridge, SI/LA/SS and SI/MA/SC, then I believe these topics were very well chosen.
- I believe so...may have been a bit light in the math department.
- Yes, I felt the topics assessed were appropriate.
- Yes

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- Yes, I felt the standard did cover the items for the LDL.
- For the most part the LDL's were aligned with the particular levels on the MPIs. Assigning the LDL to specific MPIs and working within our groups assisted when working individually. I felt the observations and insight that were discussed within our small groups aided in completing the second task.
- Yes
- I did feel there were a variety of levels addressed. I did not see a great deal of "2's"....
- I sometimes felt there were more 3 DOKs on the Social Studies area.
- Yes

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Yes, many standards are written at an appropriate level, but, there are some points of interest that were brought up within our group. Please see notes below. The majority of these standards are concise and written clearly. When working within my small group we did find some language of the standards that made it difficult to assign particular LDLs to the MPI. For example, Level 2 MPI for ELP Standard 1 in speaking asks to paraphrase or retell. If a student needs to retell simple homework directions then that could be placed at a 1 LDL, but, if they need to paraphrase that is going to make it a 2 LDL. Paraphrasing is a higher order thinking skill compared to simple retelling. Furthermore, ELP Standard 4 Level 4 of

speaking is asking the student to "imagine". There is no oral language needed for this particular MPI. More clarification is needed for this MPI because as of right now the task is not clear. If it stated imagine and DESCRIBE that would then incorporate an oral aspect into the MPI.

- My only concern is the "reading corner" topic. I'm not certain that 6-8th graders could identify with a reading corner in a classroom--maybe in a library.
- I do feel they were appropriate.
- I sometimes felt there were more 3 DOKs on the Social Studies area.
- Yes

D. What is your general opinion of the alignment between the standards and assessment:

ii. Acceptable Alignment (7) : 100%

E. Comments

- All in all, I feel the second day working on this went much better than the first. The first day there were a lot of "gray" areas and a little bit of confusion. For next time, I feel the conversation we had in the morning with Meredith and Carson is something that should be done on Day 1 before we go in to the computer lab. I felt the group work on the second day was more helpful and accommodating because all of us felt we had a better grasp and hold of the MPIs, standards, LDLs, and what was being expected. The study was interesting and insightful, met many experienced and knowledgeable people, and cannot wait to participate in another one!
- I wish that we would have received more explanation about LDLs and perhaps have gone through a couple of the examples in our binder as a group to get everyone on the same page and to tease out some of the questions groups would have before we were turned loose. Even though it was explained that we were rating the MPIs and items for LDL and not DOK, I think that just by having DOK on the WAT pages made people think of content or cognitive complexity instead of linguistic complexity.
- There are a few "gray areas" for me in the standards. Some of the areas are close, but could be completely different. For example: Soc./Instr. speaking Level 2 - Paraphrase or retell... retelling can be quite simple, however paraphrasing can be quite involved. Science-Speaking-Level 4, Imagine... then there is no language correlation for this action, it can be quite easy for them to imagine, but then what?
- I wish there were a category that said "good" alignment because I felt this alignment has gone through a very thorough process. I have been on pilots of other ESL assessments and the ELL ACCESS is the most professionally done and a better measure than any I have seen. Thank you for providing us with the opportunity to have this input.
- When assessing items during the first part of this section comments were made regarding telling and paraphrasing. The consensus was that telling and paraphrasing are two different tasks and that these were also at different levels of difficulty. therefore the consensus is that these two should be separated at some point.

WIDA ACCESS Grades 6-8 Reading

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- Yes
- Yes
- Yes. There seemed to be a representative coverage of the most important topics.
- I think once I realized that I can transform the standards, it was much easier to apply it to the items in the test.
- I thought that the reading level was very high and that level 1 and 2 students would become extremely disheartened, even by the Tier A items. It looks like I found many more examples of higher level performance indicators.
- Yes, I think that the reading part of the test covers wide range of topics and they seem to be evenly spread.
- I feel that all standards were covered, but in science there was too much focus on processes and cycles; in language arts nearly all questions dealt with figurative language; in math nearly all questions focused on perimeter, volume, area. I think the items should have been spread out among more topics in each standard.

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- I think that in Form A more level 1 difficulties should have been in the test items. There was not enough difference of Form A and Form B.
- Yes
- Yes. The items moved from least linguistically demanding to most challenging.
- I am not sure if LDLs were covered evenly in reading. It seemed like there were just a few of 1s; however, I am guessing that in order to assess student's proficiency, there should be more 2s and 3s...so then the selection was proper.
- Yes
- Yes, the items covered all the DOK levels.
- There were not many items that I felt addressed level 1 of the LDLs. Mostly, however, I think this area was covered.

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Somewhat
- Yes
- Considering that the items cover grade ranges, yes, they are appropriate for grade level.
- Look at the above (1) response; yes, they were.

- I believe that students at my school, even non-ELLS would have difficulty with the reading levels in these tests, especially Tier C. However, it should be noted that my low income school has been flagged for the 2nd year for not making AYP in the area of reading.
- As I said in my comments from yesterday, they are if you have a student that has not had interrupted schooling, trauma and is literate in his/her native language.
- Yes, I think so. All seemed to be grade-level appropriate. This test addressed many more things actually covered in our state's SCOS, but I realize that is not the main issue for the test - it is just a fortunate happenstance for me.

D. What is your general opinion of the alignment between the standards and assessment:

- ii. Acceptable Alignment (5) : 62%
- iii. Needs slight improvement (3) : 38%

E. Comments

- I believe Form A needs to have more test items addressing Level 1-2 difficulty.
- The reading part is time-consuming; I wish the distribution of time or tested areas was managed differently. I wanted to do a good job, but I felt like I was disturbed by people finishing other areas all the time.
- I think that the lower level skills could be broken down into an initial test for newcomers that could better inform instructions.
- I have to say that this part extremely difficult and time consuming. It was difficult because a lot of times I was not able to pick a standard for some items. I was able to generalize the content of the item, but not make a connection with a specific standard. I wish that we had done this one first.

WIDA ACCESS Grades 6-8 Writing

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- Social Studies was not assessed. If the test is supposed to cover all standards, then Social Studies should be included.
- Yes
- I expected to see more representation of the three following content areas: math, science, and history. While I saw some math and science representation - and of course - Social Instructional & Language Arts, I saw not Social Studies represented at all. This could be that there are other areas of the test in which these standards are reflected.
- I am not sure if I understand the question; however, I haven't seen enough of social studies(history) topics there. I also think some items should include topics that would be culturally closer to second language learners.
- I only used one Math standard and zero from Social Studies. I used the most from Language Arts and Social and Instructional language. This would be bad for students in Madison

schools because we use the Connected Mathematics Program (CMP). CMP requires students to explain how they achieve their answers and to justify why the answers are correct.

- I do think there were necessarily any items that discussed social studies topics. A lot of the items covered the social and instructional language.
- I think there is too much emphasis on soc and instruc language and a too heavy emphasis on school life within that standard.

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- The Math standard was only used in Form A. It seems that the Math standard should have been included in Forms B and C as well.
- Yes
- Yes, I believe that the LDLs were fairly well represented within the test items for each tier. Tasks ranged from low LDLs to higher LDLs with graduating decrease of support.
- I thought linguistic difficulty levels were spread evenly.
- Math was very poorly covered. In teir C, the tomatoe garden questions could have included dimensions and reasoning out how much could be planted and so on. I didn't see Social Studies covered in the writing domain at all. I would think that justifying and making inferences would be a key part, illustrated in the writing domain, to effective functioning in the language of Social Studies.
- I do think that the items covered all of the DOK levels.
- All of the language difficulty levels were addressed

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Yes
- Yes
- Yes, for the most part they were specific and directed towards expectations appropriate for the grade level, but some of the higher level writing tasks seemed more challenging than what is indicated by the levels 3-4 MPIs indicate.
- I think standards were written specifically enough.
- No, they were too specific and seemed kind of randomly selected. Why, for example, and synonyms and antonyms more important than, say summarizing or explaining the plot of a story.
- I believe that they are if you assume that your students have had previous schooling, are literate in their native language and have not been exposed to any trauma.
- Yes.

D. What is your general opinion of the alignment between the standards and assessment:

- ii. Acceptable Alignment (5) : 62%
- iii. Needs slight improvement (3) : 38%

E. Comments

- The writing section does not include all standards in every test form; consequently, it seems like it would be impossible to include all standards in the writing section of the test.
- I have only looked at the writing domain. It may be that more of the standards for Math, Science and Social Studies were more fully addressed in the other domains. However, at least in Madison, students are graded on being able to explain their reasoning as a key part of the math and science programs. Our Social Studies program doesn't require that much writing, but I believe that other schools do.

WIDA ACCESS Grades 9-12 Listening

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- Yes
- In my experience, many students who would be taking the Tier A test may have gaps in content knowledge that interfere with their rapid acquisition of academic language. While a teacher may adapt MPIs to suit his/her needs for students who are not working at or near grade level, the content-language expectations for Tier A test-takers seems out of their proximal zone of development in some Standards. This seems especially true for LoMa and somewhat true for LoSc.
- In the listening language arts section the MPI about comedies is very limited.
- Yes
- I would like to see a summary of my results and compare them to the standards to see what topics were not assessed.
- For the most part, it seemed each standard was represented. It seemed that Lang. Arts overall was the least represented, or was confusing that it in fact was represented. Since this is representative of content standards, it is difficult to distinguish what topics should have been included. Math, Science and Social studies were all well represented.

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- Yes
- Generally, each tier of ACCESS seems appropriate for linguistic complexity, with the exception noted about students with educational gaps on Tier A. I don't believe it is possible to entirely separate content and language. Students without strong L1 foundations are at a disadvantage for connecting to prior knowledge.
- Yes it did
- Yes.. they were covered

- I do think that DOK levels were covered by the tests.
- I think each standard was covered well.
- Seemed like it...

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Some were more difficult than I expected.
- Generally, the standards and expectations are appropriate. I was surprised that the Tier C test did not ask for demonstration of more complex language and thinking skills in LoLA. I understand that this is a LISTENING test; however, some of the longer questions on Tiers B and C seemed more about recalling a particular detail without actually attempting to invest meaning in the recall of those details.
- Yes, for the most part. I noted exceptions on particular items.
- Very wordy. I often wonder how our native speaking students would do on this assessment.
- Language Arts was a little difficult to cross reference
- Yes
- Yes

D. What is your general opinion of the alignment between the standards and assessment:

- ii. Acceptable Alignment (5) : 62%
- iii. Needs slight improvement (3) : 38%

E. Comments

- Too many to do in one sitting!
- Even with the suggestions and critique noted above, I still find the ACCESS test a superior testing instrument for gauging annual growth in students' language proficiency to tests used previously. This final suggestion might result in a prohibitive cost increase, but I have noticed that, like myself, students who have taken the ACCESS test for several years in a row do recall when a particular theme set of questions has been repeated from the previous year's test. Would it be possible to avoid repeating themes from year to year to make the test even more authentic?
- I think more frequent breaks would have been good during this part of alignment. Did not feel as fresh when working on Tier C. ALSO, starting on about question 10 to question 60 I incorrectly placed the SocIn in the primary box when I should have placed either LA, Science, Math or Social Studies.
- The notes I left were my justification for the answers that I chose.
- Will the test ever be given to native speakers to see how their data/results would compare to ELLs taking the Access?

WIDA ACCESS Grades 9-12 Speaking

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- Yes
- There were no opportunities to use LoMA on the speaking assessment and few opportunities to use LoSS. How about trying to integrate the math and science standards?
- The speaking test does a fine job of covering the various highlights of the summative speaking standards
- There was no math in the speaking portion.
- Yes I think the most important topics were covered
- Yes--the deliberate and well planned questions are appreciated.
- Math was not represented. Language Arts levels 3, 4, & 5 were also not represented. Social Studies standards levels 3, 4, & 5 were also not represented. The question's need to represent all levels of the standards.

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- Yes
- Perhaps it's because we use the same speaking test regardless of Tier, but there were significantly fewer questions in which there were opportunities to demonstrate complex linguistic abilities. I suppose that the argument could be made that higher linguistic ability is generated naturally from students who have that proficiency level, but as someone who gave 275 speaking tests last year and 243 speaking tests the previous year, the test administrator is supposed to stick to the script and its given options to ensure rater reliability. We're not supposed to generate these opportunities ourselves. I'm curious as to why we don't seem to be concerned with discovering whether students are functioning beyond "acceptable" levels.
- Yes
- Don't really think level 3 DOK was covered
- Yep
- Level 3 was underrepresented.

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Sometimes they seem too specific. However, if a teacher understands well how to SUBSTITUTE or TRANSFORM the content, it would be fine. Training is necessary in this area. I have only a burgeoning understanding of this, and believe that many people could benefit from examples.
- Compared to the other 3 domains, the speaking test seems significantly less challenging, and even juvenile, in places.
- Yes

- Yes
- Yes, I think all of the standards were written at an appropriate level
- Yes
- Yes the questions were specific enough, but some are culturally biased. Questions are also grade level appropriate, or under.

D. What is your general opinion of the alignment between the standards and assessment:

- ii. Acceptable Alignment (6) : 75%
- iii. Needs slight improvement (2) : 25%

E. Comments

- Next time, I recommend that the concept of TRANSFORMATIONS be discussed in the beginning. Understanding this concept made the process less stressful and more clear for me. Also, I recommend that the sample standards for LANGUAGE ARTS LISTENING be changed. It is very difficult to transform the "Multiple Meaning" examples. Overall, I think that Meredith and Carsten did an excellent job of facilitating the process. They were professional, courteous, informative, and stern. I know that teachers can be the worst students sometimes, and they dealt well with the frustration and the few smart-alecs and rebels. GOOD JOB!
- All 5 standards should be included in the speaking test. Is there a way to embed LoSI within the other standards so that the test is more comprehensive without be necessarily longer? I've found that testing takes 20-30 minutes per student as it is currently.
- Practically perfect- but not totally perfect
- Math should be represented.

WIDA ACCESS Grades 9-12 Reading

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- Yes
- I think Social Studies Level 5 was not covered by any of the items in the test. Other standards were more or less equally covered.
- I felt that the topics were covered adequately.
- Yes

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- Yes
- I think this was achieved.

- I felt the DOK levels were adequate for each tier level of the test.
- My impression is that I saw a lot of 1's and 3's but not so many 2's.
- Yes

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- Perhaps appropriate for expectations for the grade level of the general population of students but not the subset of ELLs.
- Very much so, especially in math.
- Yes
- Yes

D. What is your general opinion of the alignment between the standards and assessment:

ii. Acceptable Alignment (8) : 100%

E. Comments

- Have time constraints on administering these tests been considered? The testing season is more and more consuming for students and teachers as state testing and ELL testing demands have increased. As a classroom teacher traveling to four schools, administering the test to 3 tier groups and the speaking test to individual students (complicated by the numerous dates blocked by state testing) has become more and more demanding. Has consideration ever been given to reducing number of test items, resulting in a more realistic time allotment for the lengthy reading passages? Have researchers participated in the WIDA testing process onsite to witness student frustration level with the length of the test? Many students will quit or fill in answers indiscriminately due to fatigue. Just curious how the field testing has informed the construction of the test.
- For some of the items, I had a difficult time matching them to a standard - especially those items involving the Internet. I decided to use the Social and Instructional standards at times, because it seemed to make the most sense to me.
- I struggled with where to put situations that dealt with research.

WIDA ACCESS Grades 9-12 Writing

A. For each standard, did the items cover the most important topics you expected by the standard? If not, what topics were not assessed that should have been?

- For the most part, yes, important topics were covered by the items.
- I think that Standard 3 (Language of Math) and Standard 5 (Social Studies) could be represented more within the test items.
- There did not seem to be any items that assessed figurative language.

- Level A - OK Level B - I don't think figurative language is addressed Level C - I don't think figurative language is addressed Figurative language is part of our state assessment - Dakota Step
- Yes
- Yes

B. For each standard, did the items cover the most important performance (DOK levels) you expected by the standard? If not, what performance was not assessed?

- Yes
- I think that DOK levels are what I expected.
- Yes
- Yes
- Yes
- Yes

C. Were the standards written at an appropriate level of specificity and directed towards expectations appropriate for the grade level?

- The term "viscosity" (Writing Tier B Part B) would challenge many grade 9 or 10 students, whether or not the students were ELLs. Have such test items been field tested in the general student population? Although a definition of viscosity is provided, ELLs might be intimidated by the term and its later adjective usage, "viscous."
- I think that this was achieved.
- Yes
- Yes
- Yes
- Yes

D. What is your general opinion of the alignment between the standards and assessment:

ii. Acceptable Alignment (8) : 100%