REPORT

Alignment Analysis of Extended Reading Standards and Assessments

Wisconsin
Grades 3-8 and 10
2008

Norman L. Webb

June 25, 2008

REPORT

Alignment Analysis of Extended Reading Standards and Assessments

Wisconsin
Grades 3-8 and 10
2008

Norman L. Webb

June 25, 2008

Acknowledgements

Reviewers

Cindy Jacobson	(Group Leader)	WI
KarenVan Buren		NC
Rosanne Van Dyke		MI
Bart Wepking		WI
Patricia Schwenke		WI
Shelley Lehman		WI

Table of Contents

Appendix A
Wisconsin Grades 3-8 and 10 Extended Reading Standards and Group Consensus DOK Values

Appendix B
Data Analysis Tables Wisconsin Grades 3-8 and 10 Extended Reading 2008

Appendix C
Reviewers' Notes and Source-of-Challenge Comments Wisconsin Grades 3-8 and 10 Extended Reading 2008

Appendix D
Debriefing Summary Notes Wisconsin Grades 3-8 and 10 Extended Reading 2008

Executive Summary

A three-day alignment institute was held in Madison, Wisconsin, April 18-20, 2008, to analyze the alignment between the Wisconsin Alternate Assessment for Students with Disabilities (WAA-SwD) and standards for grades 3-8 and 10. Six reviewers participated including four from Wisconsin and two from other states. The reviewers included reading experts, special education classroom teachers, and those who had participated in assessment development.

The assessments and the standards for grades 3-8 were judged to have acceptable alignment with five or fewer items needed to be replaced to attain full alignment. The alignment for grade 10 needed improvement. The assessments and the standards had two alignment issues, one with Categorical Concurrence and one with Depth-of-Knowledge Consistency. Even though the assessments were designed to have six or more items for each of the three or four standards for a grade, reviewers found fewer than that number for four of the grades--grades 3, 4, 5, and 7. For grades 3 and 4, the assessments only had one and two items, respectively, that matched Objective 4A. This objective expected students to connect the text to self, a difficult expectation to assess with an on demand assessment. For these two grades reviewers only found one or two items that related in some way the text with the self. For grade 5, the assessment was short one item for Standard 3. For grade 7, the assessment was short two items for Standard 1.

The Depth-of-Knowledge Consistency criterion was not met for at least one standard for six of the seven grades, all but for grade 4. For at least one standard for each of these six grades, less than one-half of the items had a DOK stage that was the same or higher than the DOK stage for the assigned objective. The lack of match in the DOK stage for grades 3, 5, 6, 7, and 8 is considered only minor. However, for grade 10 the lack in agreement between the DOK stages expected by the standards and required by the assessment items is considered more problematic. Ten grade 10 items (one-third of the items on the assessment) would need to be replaced by items with higher DOK stages to attain full alignment. For the other grades four or fewer items would need to be replaced by items with higher DOK stages to fully satisfying the DOK criterion. Range and balance were fully attained in part because of the very small number of objectives underlying each standard.

Reviewers did make a number of comments on how they thought assessment items and the alignment could be improved. For some grades, reviewers suggested rewording some of the standards and objectives to improve the standards, such as eliminating "three sentences" for grades 4-3 Objective 2A and including something on main idea in the higher grades. Two or more reviewers also identified a number of items (one to 12) with source-of-challenge issue (e.g. unnecessary visual clue and confusing directions). These items should be reviewed.

Overall, the alignment for grades 3-8 was acceptable with only minor changes need to attain full alignment. The alignment for grade 10 needed improvement. The table below summarizes the results for each grade.

Summary Table
Percent of Wisconsin Extended Grade Band Mathematics Standards with Acceptable Level on Each Alignment Criteria for Grade 3-8 and 10 for WAA-SwD Analysis

Grade	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*	*Estimated Average Number of Items per Grade to be Replaced for Full Alignment*
3	75%	75%	75%	75%	5
4	75%	100%	100%	100%	4
5	75%	75%	100%	100%	3
6	100%	75%	100%	100%	1
7	67%	33%	100%	100%	5
8	100%	33%	100%	100%	4
10	100%	0%	100%	100%	10

Categorical Concurrence	>6 items
Depth-of-Knowledge	>50% with EDOK stage the same or higher than level of corresponding objective
Range-of-Knowledge	>50% of objective under a standard
Balance of Representation	>.70 index value

Alignment Analysis of Reading Standards and Assessments

Wisconsin
Grades 3-8 and 10
2008

Norman L. Webb

Introduction

The alignment of expectations for student learning with assessments for measuring students' attainment of these expectations is an essential attribute for an effective standards-based education system. Alignment is defined as the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide an education system toward students learning what they are expected to know and do. As such, alignment is a quality of the relationship between expectations and assessments and not an attribute of any one of these two system components. Alignment describes the match between expectations and an assessment that can be legitimately improved by changing either student expectations or the assessments. As a relationship between two or more system components, alignment is determined by using the multiple criteria described in detail in a National Institute for Science Education (NISE) research monograph, *Criteria for Alignment of Expectations and Assessments in Reading and Science Education* (Webb, 1997).

A three-day alignment analysis institute was conducted April 18-20, 2008, in Madison, Wisconsin. Six reviewers, including reading content experts and special education reading teachers analyzed the agreement between the Wisconsin extended grade band standards for reading drafted in May 2007 and the Wisconsin Alternate Assessment for Students with Disabilities (WAA-SwD) for grades 3–8 and 10. Four of the reviewers were from Wisconsin and two were from other states.

The State of Wisconsin uses the terminology of *model standards, extended grade band objectives (grades 3-4, 5-6, 7-8, and 10)*, and *achievement descriptors* in its reading content expectations for students with significant cognitive disabilities. For each extended grade band objective, the achievement descriptors were given for four performance levels—advanced, proficient, basic, and minimal. The proficient level descriptors were used in this analysis to further describe what students were expected to do to satisfy the extended grade band objectives. The model standards were the broad content requirements across all grades. The extended grade band objectives (referred to in this report as objectives) specified what students with significant cognitive disabilities were to know and do within a grade band. The standards and descriptors were "designed to allow students with significant cognitive disabilities to progress toward state standards linked to grade level expectations while beginning at each student's present level of performance." (Edvantia, Inc, draft, May 2007). The standards and extended objectives were designed to increase access by special education students to the general curriculum.

1

Data for this analysis were entered at the extended grade band objective level and reported out at the standards level.

As part of the alignment institute, reviewers were trained to identify the extended depth-of-knowledge (EDOK) of the extended objectives and assessment items. This training included reviewing the definitions of the six EDOK stages and reviewing examples of each. Then the reviewers participated in 1) a consensus process to determine the EDOK stages of the objectives and 2) individual analyses of the assessment items. Following individual analyses of the items, reviewers participated in a debriefing discussion in which they evaluated the degree to which they had coded particular items or types of content to the objectives.

To derive the results from the analysis, the reviewers' responses were averaged. Any variance among reviewers is considered legitimate, with the true EDOK stage for the item falling somewhere between the two or more assigned values. Such variation could signify a lack of clarity in how the standards and objectives were written, the robustness of an item that can legitimately correspond to more than one objective, and/or an EDOK that falls in between two of the six defined stages. Reviewers were allowed to identify one assessment item as corresponding to up to three objectives—one primary hit (objective) and up to two secondary hits. However, reviewers could only code one EDOK stage for each assessment item, even if the item corresponded to more than one objective.

Reviewers were instructed to focus primarily on the alignment between the state standards, extended objectives, and assessments. However, reviewers were encouraged to offer their opinions on the quality of the standards, or of the assessment activities/items, by writing a note about the item. Reviewers could also indicate whether there was a source-of-challenge issue with the item—i.e., a problem with the item that might cause the student who knows the material to give a wrong answer, or enable someone who does not have the knowledge being tested to answer the item correctly.

The results produced from the institute pertain only to the issue of alignment between the Wisconsin state standards, extended objectives, and the state alternate assessment instruments. Note that this alignment analysis does not serve as external verification of the general quality of the state's standards or assessments. Rather, only the degree of alignment is discussed in the results. For these results, the means of the reviewers' coding were used to determine whether the alignment criteria were met. When reviewers did vary in their judgments, the means lessened the error that might result from any one reviewer's finding. Standard deviations are reported in the tables provided in the Appendix B, which give one indication of the variance among reviewers.

The present report describes the results of an alignment study of extended objectives and the January 2008 tests in reading for grades 3–8 and 10 in Wisconsin. The study addressed specific criteria related to the content agreement between the state extended grade band standards and grade-level assessments. Four criteria received major attention: categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance of representation.

Alignment Criteria Used for This Analysis

This analysis judged the alignment between the standards and the assessments on the basis of four criteria. Information is also reported on the quality of items by identifying items with sources-of-challenge and other issues. For each alignment criterion, an acceptable level was defined by what would be required to assure that a student had met the standards.

Categorical Concurrence

An important aspect of alignment between standards and assessments is whether both address the same content categories. The categorical-concurrence criterion provides a very general indication of alignment if both documents incorporate the same content. *The criterion of categorical concurrence between standards and assessments is met if the same or consistent categories of content appear in both documents.* This criterion was judged by determining whether the assessment included items measuring content from each standard. The analysis assumed that the assessment had to have at least six items for measuring content from a standard in order for an acceptable level of categorical concurrence to exist between the standard and the assessment. The number of items, six, is based on estimating the number of items that could produce a reasonably reliable subscale for estimating students' mastery of content on that subscale. Of course, many factors have to be considered in determining what a reasonable number is, including the reliability of the subscale, the mean score, and cutoff score for determining mastery. Using a procedure developed by Subkoviak (1988) and assuming that the cutoff score is the mean and that the reliability of one item is .1, it was estimated that six items would produce an agreement coefficient of at least .63. This indicates that about 63% of the group would be consistently classified as masters or nonmasters if two equivalent test administrations were employed. The agreement coefficient would increase if the cutoff score is increased to one standard deviation from the mean to .77 and, with a cutoff score of 1.5 standard deviations from the mean, to .88. Usually states do not report student results by Standards or require students to achieve a specified cutoff score on subscales related to a standard. If a state did do this, then the state would seek a higher agreement coefficient than .63. Six items were assumed as a minimum for an assessment measuring content knowledge related to a standard, and as a basis for making some decisions about students' knowledge of that standard. If the mean for six items is 3 and one standard deviation is one item, then a cutoff score set at 4 would produce an agreement coefficient of .77. Any fewer items with a mean of one-half of the items would require a cutoff that would only allow a student to miss one item. This would be a very stringent requirement, considering a reasonable standard error of measurement on the subscale.

Depth-of-Knowledge Consistency

Standards and assessments can be aligned not only on the category of content covered by each, but also on the basis of the complexity of knowledge required by each. *Depth-of-knowledge consistency between standards and assessment indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what*

students are expected to know and do as stated in the standards. For consistency to exist between the assessment and the standard, as judged in this analysis, at least 50% of the items corresponding to a standard had to be at or above the level of knowledge of the standard: 50%, a conservative cutoff point, is based on the assumption that a minimal passing score for any one standard of 50% or higher would require the student to successfully answer at least some items at or above the depth-of-knowledge level of the corresponding standard. For example, assume an assessment included six items related to one standard and students were required to answer correctly four of those items to be judged proficient—i.e., 67% of the items. If three, 50%, of the six items were at or above the depth-of-knowledge level of the corresponding objectives, then for a student to achieve a proficient score would require the student to answer correctly at least one item at or above the depth-of-knowledge level of one objective. Some leeway was used in this analysis on this criterion. If a standard had between 40% and 50% of items at or above the depth-of-knowledge levels of the objectives, then it was reported that the criterion was "weakly" met.

Interpreting and assigning depth-of-knowledge levels to both objectives within standards and assessment items is an essential requirement of alignment analysis. For alternate assessments, six stages are used to judge complexity, rather than the traditional four depth-of-knowledge levels. The Extended Depth of Knowledge Stages for Special Education (EDOK) partitions the first DOK level (Recall and Recognition) into three stages—respond, reproduce, and recall. Stages 4, 5, and 6 are the same as DOK Levels 2, 3, and 4. The EDOKs were developed by Gary Cook and others. These descriptions help to clarify what the different levels represent in reading:

Stage 1-Respond
Requires the ability to respond to or indicate, or acknowledge text or discourse related features. Example:
♦ Points to the letters/words/pictures on a page
♦ Points to letters or words
♦ Acknowledges a discourse interaction with an interlocutor
♦ Responds to a conversation
♦ Acknowledges someone signing
♦ Attends to text

Stage 2-Reproduce
Requires the ability to copy, replicate, repeat, re-enact, mirror, or match text or discourse related features. Example:
♦ Copies letters
♦ Reproduces letters, text or words either verbally or through writing/signing
♦ Matches sound/sound
♦ Matches letter/letter
♦ Matches words
♦ Re-enacts a story or interaction either verbally or through text (also drawing)
♦ Matches picture/picture
♦ Matches symbol/symbol

Stage 3-Recall

Requires the ability to recite or recall facts or information. Involves the ability to distinguish between text-based or discourse features. Example:

♦ Identifies pictures of objects (animate or inanimate) though verbal cues or text-based cues.
♦ Identify details in text.
♦ Identifies correct spelling or meaning of words
♦ Identifies letters
♦ Identifies sounds
♦ Identifies figurative language
♦ Uses dictionary

Stage 4-Basic Reasoning

Requires processing beyond recall and observation. Requires both comprehension and subsequent processing of text. Involves ordering, classifying text as well as identifying patterns, relationships and main points. Examples:

♦ Uses context to identify unfamiliar words
♦ Predicts logical outcome
♦ Identifies and summarizes main points
♦ Associates/identifies letters with sounds
♦ Indicate what comes next in a story

Stage 5-Complex Reasoning

Requires students to go beyond the text. Requires students to explain, generalize and connect ideas. Involves inferencing, prediction, elaboration and summary. Requires students to support positions using prior knowledge and to manipulate themes across passages. Examples:

♦ Determines effect of author's purpose on text elements
♦ Summarizes information from multiple sources
♦ Critically analyzes literature/text
♦ Expresses an opinion about text citing evidence to support reasoning

Stage 6-Extended Reasoning

Requires extended higher order processing. Typically requires extended time to complete task, but time spent not on repetitive tasks. Involves taking information from one text/passage and applying this information to a new task. May require generating hypotheses and performing complex analyses and connections among texts. Examples:

♦ Analyzes and synthesizes information from multiple sources
♦ Examines and explains alternative perspectives across sources
♦ Describes and illustrates common themes across a variety of texts
♦ Creates compositions that synthesize, analyze, and evaluate

Range-of-Knowledge Correspondence

For standards and assessments to be aligned, the breadth of knowledge required on both should be comparable. *The range-of-knowledge criterion is used to judge whether 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.* The criterion for correspondence between span of knowledge for a standard and an assessment considers the number of objectives within the standard with one related assessment item/activity. Fifty percent of the objectives for a standard had to have at least one related assessment item in order for the alignment on this criterion to be judged acceptable. This level is based on the assumption that students' knowledge should be tested on content from over half of the domain of knowledge for a standard. This assumes that each objective for a standard should be given equal weight. Depending on the balance in the distribution of items and the need to have a low number of items related to any one objective, the requirement that assessment items need to be related to more than 50% of the objectives for an standard increases the likelihood that students will have to demonstrate knowledge on more than one objective per standard to achieve a minimal passing score. As with the other criteria, a state may choose to make the acceptable level on this criterion more rigorous by requiring an assessment to include items related to a greater number of the objectives. However, any restriction on the number of items included on the test will place an upper limit on the number of objectives that can be assessed. Range-of-knowledge correspondence is more difficult to attain if the content expectations are partitioned among a greater number of standards and a large number of objectives. If 50% or more of the objectives for a standard had a corresponding assessment item, then the range-of-knowledge correspondence criterion was met. If between 40% and 50% of the objectives for a standard had a corresponding assessment item, the criterion was "weakly" met.

Balance of Representation

In addition to comparable depth and breadth of knowledge, aligned standards and assessments require that knowledge be distributed equally in both. The range-of-knowledge criterion only considers the number of objectives within a standard hit (an objective with a corresponding item); it does not take into consideration how the hits (or assessment items/activities) are distributed among these objectives. *The balance-of-representation criterion is used to indicate the degree to which one objective is given more emphasis on the assessment than another.* An index is used to judge the distribution of assessment items. This index only considers the objectives for a standard that have at least one hit—i.e., one related assessment item per objective. The index is computed by considering the difference in the proportion of objectives and the proportion of hits assigned to the objective. An index value of 1 signifies perfect balance and is obtained if the hits (corresponding items) related to a standard are equally distributed among the objectives for the given standard. Index values that approach 0 signify that a large proportion of the hits are on only one or two of all of the objectives hit. Depending on the number of objectives and the number of hits, a unimodal distribution (most items related to one objective and only one item related to each of the remaining objectives) has an

index value of less than .5. A bimodal distribution has an index value of around .55 or .6. Index values of .7 or higher indicate that items/activities are distributed among all of the objectives at least to some degree (e.g., every objective has at least two items) and is used as the acceptable level on this criterion. Index values between .6 and .7 indicate the balance-of-representation criterion has only been "weakly" met.

Source-of-Challenge Criterion

The source-of-challenge criterion is only used to identify items on which the major cognitive demand is inadvertently placed and is other than the targeted reading objective, concept, or application. Cultural bias or specialized knowledge could be reasons for an item to have a source-of-challenge problem. Such item characteristics may result in some students not answering an assessment item, or answering an assessment item incorrectly, or at a lower level, even though they possess the understanding and skills being assessed.
.

Findings

Standards

The consensus EDOK value for each objective under the extended standards for reading can be found in Appendix A. Table 1 shows the percentages of extended objectives at each EDOK stage. The EDOK values used in the August 2007 analysis were used in this study. The complexity of the reading objectives does increase some in sophistication with grade. The percentage of objectives with an EDOK 5 increased from 20% (grade 3-4) to 66% (grade 10). The percentage of objectives with an EDOK 3 (recall and recognition) for grades 3 through 8 ranged from 40% to 60% while reviewers judged that none of the grade 10 objectives had an EDOK stage 3.

If no particular objective is targeted by a given assessment item, reviewers are instructed to code the item at the level of a standard. This coding to a generic objective sometimes indicates that the item is inappropriate for the grade level. However, if the item is grade-appropriate, then this situation may instead indicate that there is a part of the content not expressly or precisely described in the objectives. These items may highlight areas in the objectives that should be changed, or made more precise. Table 2 displays the assessment items coded to generic objectives by more than one reviewer.

At least two reviewers coded an item to a generic objective for five of the seven grades (3, 4, 7, 8, and 11) (Table 2). Most of the items were assigned by a majority of the reviewers to generic objectives. A majority of reviewers coded four grade 3 items to a generic objective because the items did not include three sentences as explicitly stated in Objective 2A (Recall basic facts and/or main ideas from a short paragraph of 3 simple sentences in length) as needed to reach proficiency. The same reason was given for assigning grade 4 items to generic objective Standard 2. Reviewers assigned grade 7 items to the generic objective Standard 2 because these items required students to identify the main idea, but this was not stated in the objective under Standard 2. Reviewers

assigned other items to generic objectives because the item did not have student use a context (e.g. grade 10 Item 22).

The statement of objectives included what was required for a student to be judged as proficient for the grade level. Reviewers used the proficient requirements in determining if an item matched the targeted objective. It is possible that some of the items were targeted to an objective, but at a level of attainment below proficiency. The alignment analysis considered items assigned to a generic objective as targeting a standard. However, if items were assigned to a generic objective, this increased the total number of objectives under a standard and the number of objectives needed to have an acceptable value for range.

Reviewers' debriefing comments also highlight some ambiguities in the objectives. These comments can be found in Appendix D.

Table 1
Percent of Grade-level Expectations by Depth-of-Knowledge (DOK) Levels for Grades 3-8 and 10 Wisconsin Alignment Analysis for Reading 2008 Study

Grade	Total Number of Objectives	DOK Stage	Number of Objectives by Stage	Percent within standard by Stage
3	5	3	3	60
		4	1	20
		5	1	20
4	5	3	3	60
		4	1	20
		5	1	20
5	5	3	3	60
		4	1	20
		5	1	20
6	5	3	3	60
		4	1	20
		5	1	20
7	4	3	2	50
		4	1	25
		5	1	25
8	4	3	2	50
		4	1	25
		5	1	25
10	3	4	1	33
		5	2	66

Table 2

Items Coded to Generic Objectives by More Than One Reviewer, Wisconsin Alignment Analysis for Reading, Grades 3-8 and 10 2008

Grade	Generic Objective	Assessment Item (Number of Reviewers)
3	1	22 (2)
3	2	7 (6), 9 (5), 14 (6), 22 (3)
3	3	26 (5)
4	2	6 (5), 11 (6), 23 (4), 26 (5)
4	3	27 (2)
7	2	6 (4), 15 (6), 23 (5), 26 (6)
7	3	2 (6), 16 (6)
8	2	6 (4)
8	3	2 (6)
10	1	22 (6), 23 (6)
10	2	4 (5), 6 (6), 9 (6), 17 (2), 24 (2)

Alignment of Curriculum Standards and Assessments

The results of the analysis for each of the four alignment criteria are summarized in Tables 4.1-4.7. More detailed data on each of the criteria are given in Appendix B, in the first three tables. With each table and for each grade, a description of the satisfaction of the alignment criteria for the given grade is provided. The reviewers' debriefing comments provide further detail about the individual reviewers' impressions of the alignment.

Table 3 displays the number of items and points for each assessment form. In the analysis that follows, multiple-point items are given additional weight for alignment purposes. For example, a 2-point item is counted towards the alignment as 2 identically coded 1-point items. The reading assessments had a total of 30 or 32 points for 28 items.

Table 3

Number of Items and Point Value by Grade for Wisconsin Assessments, Grades 3-8 and 10 2008

Grade Level	Number of Items	Number of Two-Point Items	Total Point Value
3	28	2	30
4	28	2	30
5	28	2	30
6	28	2	30
7	28	3	31
8	28	2	30
10	28	2	30

The results of the analysis for each of the four alignment criteria are summarized in Tables 4.1-4.7. More detailed data on each of the criteria are given in Appendix B, in the first three tables. With each table and for each grade, a description of the satisfaction of the alignment criteria for the given grade is provided. The reviewers' debriefing comments provide further detail about the individual reviewers' impressions of the alignment.

In Tables 4.1-4.7, "YES" indicates that an acceptable level was attained between the assessment and the learning goal on the criterion. "WEAK" indicates that the criterion was nearly met, within a margin that could simply be due to error in the system. "NO" indicates that the criterion was not met by a noticeable margin—10% over an acceptable level for Depth-of-Knowledge Consistency, 10% over an acceptable level for Range-of-Knowledge Correspondence, and .1 under an index value of .7 for Balance of Representation. If a standard has fewer than two items, it is considered not tested (NT).

Grade 3

The alignment between the Wisconsin alternate assessment and standards for grade 3 was found to be acceptable with minor improvement needed to be considered fully aligned. The main alignment issue for grade 3 was that reviewers only coded one item as corresponding to Standard 4 (evaluate and extend text). Having only one item that targets a standard is too few to even consider that the standard is measured in any way by the assessment. Reviewers only found one item (Item 19) that corresponded to connecting the text to the student. One reviewer commented, "(For) Standard 4 I felt that the example and the extended grade band did not necessarily match up. The example of the item could very well stand as an item on its own and not be a "connect" to self question. I do understand for this age group it is difficult to have a "connect to self" item." Reviewers found most items targeted expectations under Standards 1 and 2.

Other than for Standard 4, the assessment had a sufficient number of items for each standard (from seven to 11 items), with an appropriate level of complexity in relationship to the objectives, and with sufficient range (at least one item for each objective) and balance. Reviewers' comments primarily were directed towards formatting of items. A few reviewers did note that Objective 2B and Standard 4 needed more items. Overall, five items would need to be replaced on the grade 3 assessment with items that more explicitly targeted Objective 4A, similar to Item 19, to attain full alignment.

Table 4
Summary of Acceptable Levels on Alignment Criteria for Reading Grades 3-8 and 10, Standards and Assessments for Wisconsin Alignment Analysis 2008

Table 4.1
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 3, Standards and Assessments for Wisconsin Alignment Analysis 2008

Grade 3	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
1 - Determine the meaning of words and phrases in context	YES	YES	YES	YES
2 - Understand Text.	YES	YES	YES	YES
3 - Analyze text.	YES	YES	YES	YES
4 - Evaluate and extend text.	NO (1.0)	NT	NT	NT

Grade 4

The alignment between the grade 4 alternate assessment and the reading standards was considered acceptable. As for grade 3, the assessment and the standards were fully aligned for three of the four standards. Reviewers only found two items that corresponded to Objective 4A, a sufficient number to consider the standard as being at least assessed, but not a sufficient number of items to make a reliable judgment about students' proficiency on the standard. It is likely that other items were intended to target Objective 4A, but reviewers did not think the fit was close enough. One reviewer noted, "There seemed to be trouble connecting the text to self. Some of the questions tried to do so, but they were oddly worded or the questions could stand alone and not need text to support them." The two items that were mapped to Objective 4A had an appropriate DOK stage and satisfied the conditions for range and balance. Some reviewers also noted that the assessment did not have items that targeted Objective 2B (sequencing). In general, the DOK stages of the items were appropriate. One reviewer, however, observed that the there were no items with a DOK stage 1 or 2, "I would think with this population at this grade level you will very likely have students at a level 1 or 2 so therefore would be beneficial to have some questions that cover a wider range of the DOKs."

Overall, the alignment for grade 4 was acceptable. Full alignment could be attained by replacing four items with items that more clearly target Objective 4A. Reviewers made a number of comments about how the alignment for grade 4 could be improved. Many of their comments were on how the wording of the standards could be improved. For example, specifying three sentences in Standard 2A was problematic. One reviewer suggested for Standard 2A, "…possible rewording would be to include reference to pictures as well as simple sentences; stating a certain number of sentences may be problematic. [Suggested rewording:] 'Recall basic facts and/or main idea from

written, spoken, or verbal text'."

Table 4.2
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 4, Standards and Assessments for Wisconsin Alignment Analysis 2008

Grade 4	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
1 - Determine the meaning of words and phrases in context	YES	YES	YES	YES
2 - Understand Text.	YES	YES	YES	YES
3 - Analyze text.	YES	YES	YES	YES
4 - Evaluate and extend text.	NO (2.0)	YES	YES	YES

Grade 5

As for grades 3 and 4, the alignment between the grade 5 alternate assessment and standards was found to be acceptable. There were only two alignment issues for grade 5. The majority of reviewers only found five items that targeted Standard 3, one fewer than the six needed to have an acceptable level for the Categorical Concurrence criterion. The majority of reviewers found from seven to nine items that targeted content related to each of the other three standards. The set of items that targeted Standards 1, 2, and 3 had a sufficient DOK stage compared to what was expected by the underlying objectives to have an acceptable level for the Depth-of-Knowledge Consistency criterion for these standards. Only the seven items that reviewers found that targeted Objective 4A did not have a sufficiently high enough DOK stage. Only two of these items (27% on the average) had a DOK stage that was the same or higher than the DOK stage assigned to Objective 4A (Stage 5). Range and balance was acceptable for all four standards.

Overall, only three items would need to be replaced to attain full alignment. One item currently judged to target objectives under Standards 1 and 2 would need to be replaced by an item with an appropriate DOK stage that clearly targets Objective 3A. At least two of the items that correspond to Objective 4A would need to be replaced by items that have a DOK Stage 5. Reviewers noted that there were no grade 5 items that they found targeting Objective 2B. As one reviewer commented, "There were no 2B, following steps in a process questions. There were a couple main idea questions, but having the pictures underneath the question made the question seem more like matching then a main idea question." Reviewers also expressed concerned with the use and placement of pictures. For example, one reviewer suggested, "Use of pictures below the story in several early items and throughout the test is confusing; some later items without the pictures are better items and a higher level." Another reviewer noted, "I am uncomfortable with the pictures that go with the items, especially the ones that had a picture with the text and then the same picture in the student answer book. I am unsure as

to the intent of this."

Table 4.3
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 5 Standards and Assessments for Wisconsin Alignment Analysis 2008

Grade 5	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
1 - Determine the meaning of words and phrases in context	YES	YES	YES	YES
2 - Understand Text.	YES	YES	YES	YES
3 - Analyze text.	NO (5.33)	YES	YES	YES
4 - Evaluate and extend text.	YES	NO	YES	YES

Grade 6

The alignment between the grade 6 alternate assessment and the standards were acceptable. All four alignment criteria had an acceptable level for Standards 1 through 3. Standard 4 had a sufficient number of items to satisfy the Categorical Concurrence criterion, but the DOK stage for the eight items that mapped to this standard were a little low. Only three of the items had a DOK stage that was the same or higher than the DOK stage of the assigned objective. Range and balance were acceptable for all standards. The assessment had items for every objective except for Objective 2B. Reviewers thought the grade 6 items progressed some in complexity from the prior grades, but not as much as they had expected. One reviewer noted, "…progression of difficulty does not seem to be that large from grades 3-6. The DOK's were weak for some questions because they had pictures under the questions that would make the question more of a matching than a comprehension question." One reviewer indicated that that detail was not addressed under Objective 2A and that items that attempted to get at topic (main idea) was not as strong as they could have been.

Table 4.4
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 6 Standards and Assessments for Wisconsin Alignment Analysis 2008

Grade 6	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
1 - Determine the meaning of words and phrases in context	YES	YES	YES	YES
2 – Understand Text.	YES	YES	YES	YES
3 - Analyze text.	YES	YES	YES	YES
4 - Evaluate and extend text.	YES	NO	YES	YES

Overall the alignment at grade 6 was acceptable. Only one item would need to be replaced to attain full alignment. One of the items that target Objective 4A would need to be replaced by one with a DOK stage 5 such as Items 18 and 27.

Grade 7

The alignment between the grade 7 alternate assessment and the reading standards was also acceptable. Different from the previous grades, grade 7 had three standards rather than four. The assessment and only Standard 2 were found to be fully aligned. The alignment with the other two standards needed some improvement. The majority of reviewers found five items that corresponded to Standard 1, one fewer than needed to have an acceptable level for the Categorical Concurrence criterion. All of the items that reviewers mapped to Objective 1A had a DOK stage 3, one less than the DOK stage 4 assigned to the objective. Thus, the Depth-of-Knowledge Consistency criterion was not met for Standard 1. Reviewers assigned 11 items to Standard 3, but only 39% of the items had a DOK stage that was the same or higher than the DOK stage of Objective 3A (DOK stage 5). Range and balance was acceptable for all three standards.

Table 4.5
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 7 Standards and Assessments for Wisconsin Alignment Analysis 2008

Grade 7	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
1 - Determine the meaning of words and phrases in context	NO (4.83)	NO	YES	YES
2 - Analyze & Understand Text	YES	YES	YES	YES
3 - Evaluate and extend text.	YES	NO	YES	YES

Overall, the alignment for grade 7 was barely acceptable. Five items would need to be replaced to attain full alignment. Two items targeting objectives under Standards 2 and 3 would need to be replaced by items that clearly target Objective 1A. These items, as well as one item that currently maps to Objective 1A, would need to be replaced by items that have at least a DOK stage 4. At least two items that target Objective 3A need to be replaced by items that have a DOK stage 5 in order to have an acceptable level on the Depth-of-Knowledge Consistency criterion for Standard 3. Reviewers found that the standards missed some common topics that they expected to find at grade 7. They found items that they thought mapped to these common topics that were missing and had to assign six items (20% of the items) to generic objectives. One reviewer explained her coding, "No specific objective for main idea so those items had to be coded to generic; no specific objective for reality/fantasy at this grade level. No specific objective for fact/opinion so coded to generic. Inference coded under 3A." Even though the alignment at grade 7 was found to be acceptable using the minimum acceptable levels for alignment

and the format of items had improved, there were a number of issues that need to be resolved in order to have full alignment between the grade 7 standards and assessment.

Grade 8

As for most other grades, the alignment between the grade 8 alternate assessment and the standards was acceptable. Reviewers found from six to 14 items for each of the three grade 8 standards, sufficient to satisfy the Categorical Concurrence criterion. The main alignment issue was with the DOK stages of items for two of the three standards, 1 and 3. For both of these standards, only about one-third of the items had a DOK stage that was the same or higher than the DOK stage of the assigned objective. Range and balance was acceptable for all three standards. Reviewers felt that the alignment at grade 8 was better than for previous grades. They found at least some items that corresponded to Objective 2B and coded only two items to generic objectives. They repeated their comments for grade 7 about the need for the standards to expect students to identify the main idea.

Overall, the alignment at grade 8 was found to be acceptable. Four items would need to be replaced to attain full alignment. Two items that target Standard1 would need to be replaced by items that have at least a DOK stage 4 and two items that target Standard 3 would need to be replaced by items that have at least a DOK stage 5.

Table 4.6
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 8 Standards and Assessments for Wisconsin Alignment Analysis 2008

Grade 8	Alignment Criteria			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
1 - Determine the meaning of words and phrases in context	YES	NO	YES	YES
2 - Analyze & Understand Text	YES	YES	YES	YES
3 - Evaluate and extend text.	YES	NO	YES	YES

Grade 10

The alignment between the grade 10 alternate assessment and the reading standards needed improvement. The assessment had a sufficient number of items for each of the three standards, from eight to 14 items, but too few of the items had a DOK stage that matched the DOK stage of the assigned objective. For Standard 1, only about one-third of the eight items had a DOK stage that was comparable to the DOK stage of the objectives. For Standard 2, only about one of the 14 items had a DOK stage 5, the stage required to match the DOK stage of Objective 2A. For Standard 3, 18% of the items, on the average, had a DOK stage 5, the DOK stage assigned to Objective 3A. As for the

other grades and because of the low number of objectives, range and balance was acceptably met.

Reviewers noted, as for previous grades, missing topics they had expected to find such as main idea, fact and opinion, and reality or fantasy questions. Some reviewers indicated they did not see as much progression in items at the grade 10 level as they had expected. Some of the items were more basic than for grade 3. Overall, the alignment needed improvement because of the low DOK stages of items. A total of 10 items would need to be replaced with more complex items to have full alignment. At least one item that corresponded to Objective 1A would need to be replaced by an item with a DOK stage 4 or higher. For Standard 2, at least six items would need to be replaced by items with a DOK stage 5. For Standard 3, at least three of the eight items would need to be replaced by items with a DOK stage 5.

Table 4.7
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 10 Standards and Assessments for Wisconsin Alignment Analysis

Grade 10	Alignment Criteria			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
1 - Determine the meaning of words and phrases in context	YES	NO	YES	YES
2 - Understand Text. Analyze Text.	YES	NO	YES	YES
3 - Evaluate and extend text.	YES	NO	YES	YES

Source of Challenge Issues and Reviewers' Comments

Reviewers were instructed to document any source-of-challenge issue and to provide other comments they may have. These comments can be found in Tables (grade).5 and (grade).7 in Appendix C. Two or more reviewers identified a source-of-challenge issue with from one (grades 8 and 10) to 12 (grade 5) items. Some of the issues that reviewers identified included a faulty picture, confusing directions, information in the picture that could preempt necessity for reading, and unnecessary visual clue. Items noted as having a source-of-challenge issue should be reviewed to determine if the reviewers found a valid problem. It is possible that a comment by one reviewer can identify an important issue that was missed by the other reviewers. After coding each grade-level assessment, reviewers also were asked to respond to five debriefing questions. All of the comments made by the reviewers are given in Appendices D. The notes in general offer an opinion on the item or give an explanation of the reviewers' coding.

Reliability Among Reviewers

The overall intraclass correlation among the reading reviewers' assignment of DOK stages to items was high for six reviewers for Grades 3-8 and 10 (Table 5). An intraclass correlation value greater than 0.8 generally indicates a high level of agreement among the reviewers. A pairwise comparison is used to determine the degree of reliability of reviewer coding at the objective level and at the learning goal level. Both the standard and objective pairwise comparisons values were high. The values for the standard and objective pairwise comparisons were computed after the reviewers engaged in an adjudication of their coding to detect any miscodings.

Table 5
Intraclass and Pairwise Comparisons, Wisconsin Alignment Analysis for Reading Grades 3-8 and 10 Assessments

Grade	Intraclass Correlation	Pairwise Comparison:	Pairwise: Objective	Pairwise: Standard
3	.96	.86	.90	.92
4	.98	.93	.84	.91
5	.97	.96	.92	.91
6	.97	.86	.91	.90
7	.96	.78	.95	.96
8	.96	.76	.88	.92
10	.97	.87	.80	.88

Summary

A three-day alignment institute was held in Madison, Wisconsin, April 18-20, 2008, to analyze the alignment between the Wisconsin Alternate Assessment for Students with Disabilities (WAA-SwD) and standards for grades 3-8 and 10. Six reviewers participated including four from Wisconsin and two from other states. The reviewers included reading experts, special education classroom teachers, and those who had participated in assessment development.

The assessments and the standards for grades 3-8 were judged to have acceptable alignment with five or fewer items needed to be replaced to attain full alignment. The alignment for grade 10 needed improvement. The assessments and the standards had two alignment issues, one with Categorical Concurrence and one with Depth-of-Knowledge Consistency. Even though the assessments were designed to have six or more items for each of the three or four standards for a grade, reviewers found fewer than that number for four of the grades--grades 3, 4, 5, and 7. For grades 3 and 4, the assessments only had one and two items, respectively, that matched Objective 4A. This objective expected students to connect the text to self, a difficult expectation to assess with an on demand assessment. For these two grades reviewers only found one or two items that related in some way the text with the self. For grade 5, the assessment was short one item for Standard 3. For grade 7, the assessment was short two items for Standard 1.

17

The Depth-of-Knowledge Consistency criterion was not met for at least one standard for six of the seven grades, all but for grade 4. For at least one standard for each of these six grades, less than one-half of the items had a DOK stage that was the same or higher than the DOK stage for the assigned objective. The lack of match in the DOK stage for grades 3, 5, 6, 7, and 8 is considered only minor. However, for grade 10 the lack of agreement between the DOK stages expected by the standards and required by the assessment items is considered more problematic. Ten grade 10 items (one-third of the items on the assessment) would need to be replaced by items with higher DOK stages to attain full alignment. For the other grades four or fewer items would need to be replaced by items with higher DOK stages to fully satisfying the DOK criterion. Range and balance were fully attained in part because of the very small number of objectives underlying each standard.

Reviewers did make a number of comments on how they thought assessment items and the alignment could be improved. For some grades, reviewers suggested rewording some of the standards and objectives to improve the standards, such as eliminating "three sentences" for grades 4-3 Objective 2A and including something on main idea in the higher grades. Two or more reviewers also identified a number of items (one to 12) with source-of-challenge issue (e.g. unnecessary visual clue and confusing directions). These items should be reviewed.

Overall, the alignment for grades 3-8 was acceptable with only minor changes need to attain full alignment. The alignment for grade 10 needed improvement. The table below summarizes the results for each grade.

Summary Table
Percent of Wisconsin Extended Grade Band Mathematics Standards with Acceptable Level on Each Alignment Criteria for Grade 3-8 and 10 for WAA-SwD Analysis

Grade	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation	Estimated Average Number of Items per Grade to be Replaced for Full Alignment
3	75%	75%	75%	75%	5
4	75%	100%	100%	100%	4
5	75%	75%	100%	100%	3
6	100%	75%	100%	100%	1
7	67%	33%	100%	100%	5
8	100%	33%	100%	100%	4
10	100%	0%	100%	100%	10

Categorical Concurrence	>6 items
Depth-of-Knowledge	>50% with EDOK stage the same or higher than level of corresponding objective
Range-of-Knowledge	>50% of objective under a standard
Balance of Representation	>.70 index value

References

Edvantia, Inc. (May, 2007). *Wisconsin extended grade band standards: reading, reading, science*. A draft document submitted to the Wisconsin Department of Public Instruction. Charleston, West Virginia: Author.

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 Mathematics Education Research Monograph No. 6. Madison: University of Wisconsin, Wisconsin Center for Education Research.