REPORT

Alignment Analysis of Reading Standards and Assessments

Wisconsin

Grades 3-8 and 10

Norman L. Webb

August 13, 2006

REPORT

Alignment Analysis of Reading Standards and Assessments

Wisconsin
Grades 3-8 and 10

Norman L. Webb

August 13, 2006

This study is one of three alignment studies conducted for the State of Wisconsin. The Alignment Analysis Institute was held July 10-12, 2006, in Madison, Wisconsin. The report consists of a description of the four criteria used to judge the alignment between Wisconsin Assessment Framework and the Wisconsin assessments, with tables listing the results from the analysis of the coding by eight reviewers.

Acknowledgements

Reviewers:

John Fortier (Group Leader)	Madison, WI
Cynthia Jacobson	West Salem, WI
Bart Wepking	Wheatland, WI
Lisa Wiedman	Rhinelander, WI
Mark Heike	Green Bay, WI
Joseph Papenfuss	Racine, WI
Maria Cormier	New York
Elizabeth Vander Schaaf	Minnesota

Table of Contents

Appendix A
Group Consensus Values for Wisconsin Reading Standards and Objectives

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

Appendix C
Debriefing Summary Comments Wisconsin Grades 3-8 and 10 Reading

Executive Summary

Eight experienced language arts reviewers analyzed the alignment of the reading standards and the assessments for grades 3-8 and 10 at a three-day institute held in Madison, Wisconsin on July 10-12. Six of the reviewers were from Wisconsin and two were from other states. All eight reviewers had participated in previous alignment studies. The reviewers included reading and language arts content experts, district reading coordinators, and reading teachers.

The alignment between the reading standards from the assessment framework and the assessments was found to be reasonable for the seven grades. Reviewers judged that the complexity of the standards was high with 50% of the 12 objectives under the four standards with a depth-of-knowledge level 3 (drawing inferences using information beyond the text, drawing conclusions, and analyzing author's purpose, etc.). For five of the seven grades, the alignment between the standards and assessment was found to be reasonable. Three or fewer items would need to be replaced to achieve full alignment. The alignment for grades 3 and 4 was found to need slight improvement. This is largely due to the relatively high DOK levels expected for these grades, 50% at a DOK level 3. For grades 3 and 4, eight and ten items need to be replaced by items corresponding to Standards 3 (analyze text) and with a DOK level 3. Also, for grades 3 and 4, two or three items need to be added or replaced by items that correspond to Standard 4 (evaluate and extend text). A large proportion of the items for grades 3 and 4 were judged to have a DOK level 2, fairly demanding for these grade levels. Overall, the alignment is considered to be reasonable and the assessment appropriately increasing in complexity over the grades.

Alignment Analysis of Reading Standards and Assessments

Wisconsin
Grades 3-8 and 10

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 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 Language Arts and Science Education* (Webb, 1997).

A three-day Alignment Analysis Institute was conducted July 10-12, 2006, in Madison, Wisconsin. Eight reviewers, including language arts and reading content experts, retired district reading supervisors, and reading teachers analyzed the agreement between the state's reading assessment framework standards and 2006 assessments for grades 3–8 and 10. Six of the reviewers were from Wisconsin and two were from other states. All of the reviewers were experienced reviewers and had participated in over two other alignment analyses.

The State of Wisconsin uses the terminology of *standards* and *objectives* in its mathematics content expectations. Standards are the broad content requirements across all grades. Objectives specify in more detail under a standard what students are to know and do. Wisconsin Model Academic Standards specify what students are to know and do for three grade ranges—K-4, 5-8, and 9-12. An assessment framework was developed to specify the expectations for each grade. The grade level expectations developed for the assessment framework were used in this analysis. Data for this analysis were entered at the objective level and reported out at the standards level.

As part of the alignment institute, reviewers were trained to identify the depth-of-knowledge of the objectives and assessment items. This training included reviewing the definitions of the four depth-of-knowledge (DOK) levels and reviewing examples of each. Then the reviewers participated in 1) a consensus process to determine the depth-of-knowledge levels 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 assessed 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 are averaged. Any variance among reviewers is considered legitimate, with the true depth-of-knowledge level for the item falling somewhere in 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 a depth of knowledge that falls in between two of the four defined levels. 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 depth-of-knowledge level to 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 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 and the state 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 these 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, which give one indication of the variance among reviewers.

The present report describes the results of an alignment study of objectives and the 2005 operational tests in reading for grades 3–8 and 10 in Wisconsin. The study addressed specific criteria related to the content agreement between the state 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 assessment 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

3

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 standards, 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 standard. 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 standards, then it was reported that the criterion was "weakly" met.

Interpreting and assigning depth-of-knowledge levels to standards and assessment items is an essential requirement of alignment analysis. The reading levels are based on Valencia and Wixson (2000, pp. 909-935). The following definitions of depth-of-knowledge levels were used in this reading analysis:

Reading Level 1. Level 1 requires students to receive or recite facts or to use simple skills or abilities. Oral reading that does not include analysis of the text as well as basic comprehension of a text is included. Items require only a shallow understanding of the text presented and often consist of verbatim recall from text, slight paraphrasing of specific details from the text, or simple understanding of a single word or phrase. Some examples that represent, but do not constitute all of, Level 1 performance are:

- Support ideas by reference to verbatim or only slightly paraphrased details from the text.
- Use a dictionary to find the meanings of words.
- Recognize figurative language in a reading passage.

Reading Level 2. Level 2 includes the engagement of some mental processing beyond recalling or reproducing a response; it requires both comprehension and subsequent processing of text or portions of text. Inter-sentence analysis of inference is required. Some important concepts are covered but not in a complex way. Standards and items at this level may include words such as summarize, interpret, infer, classify, organize, collect, display, compare, and determine whether fact or opinion. Literal main ideas are stressed. A Level 2 assessment item may require students to apply skills and concepts that are covered in Level 1. However, items require closer understanding of text, possibly through the item's paraphrasing both of both the question and the answer. Some examples that represent, but do not constitute all of, Level 2 performance are:

- Use context cues to identify the meaning of unfamiliar words, phrases, and expressions that could otherwise have multiple meanings.
- Predict a logical outcome based on information in a reading selection.
- Identify and summarize the major events in a narrative.

Reading Level 3. Deep knowledge becomes a greater focus at Level 3. Students are encouraged to go beyond the text; however, they are still required to show

understanding of the ideas in the text. Students may be encouraged to explain, generalize, or connect ideas. Standards and items at Level 3 involve reasoning and planning. Students must be able to support their thinking. Items may involve abstract theme identification, inference across an entire passage, or students' application of prior knowledge. Items may also involve more superficial connections between texts. Some examples that represent, but do not constitute all of, Level 3 performance are:

- Explain or recognize how author's purpose affects the interpretation of a reading selection.
- Summarize information from multiple sources to address a specific topic.
- Analyze and describe the characteristics of various types of literature.

Reading Level 4. Higher-order thinking is central and knowledge is deep at Level 4. The standard or assessment item at this level will probably be an extended activity, with extended time provided for completing it. The extended time period is not a distinguishing factor if the required work is only repetitive and does not require the application of significant conceptual understanding and higher-order thinking. Students take information from at least one passage of a text and are asked to apply this information to a new task. They may also be asked to develop hypotheses and perform complex analyses of the connections among texts. Some examples that represent, but do not constitute all of, Level 4 performance are:

- Analyze and synthesize information from multiple sources.
- Examine and explain alternative perspectives across a variety of sources.
- Describe and illustrate how common themes are found across texts from different cultures.

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 benchmarks within the standard with one related assessment item/activity. Fifty percent of the benchmarks 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 benchmark 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 benchmark, the requirement that assessment items need to be related to more than 50% of the benchmarks for a standard increases the likelihood that students will have to demonstrate knowledge on more than one benchmark 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 benchmarks. However, any restriction on the number of items included on the test will place an upper limit on the number of benchmarks 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 benchmarks. If 50% or more of the benchmarks for a standard had a corresponding assessment item, then the range-of-knowledge criterion was met. If between 40% and 50% of the benchmarks 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 benchmarks within a standard hit (a standard with a corresponding item); it does not take into consideration how the hits (or assessment items/activities) are distributed among these benchmarks. *The balance-of-representation criterion is used to indicate the degree to which one benchmark 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 benchmarks for a standard that have at least one hit—i.e., one related assessment item per benchmark. The index is computed by considering the difference in the proportion of benchmarks and the proportion of hits assigned to the benchmark. 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 benchmarks 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 benchmarks hit. Depending on the number of benchmarks and the number of hits, a unimodal distribution (most items related to one benchmark and only one item related to each of the remaining benchmarks) 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 benchmarks at least to some degree (e.g., every benchmark 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 language arts skill, 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.

Standards

Eight reviewers participated in the depth-of-knowledge (DOK) level consensus process for the standards and objectives for the Wisconsin reading standards from the assessment framework. A summary of their deliberations is presented in Table 1. The complete group consensus values for each competency and objective can be found in Appendix A. The structure of the standards and objectives are very similar for each grade. For each grade, there are four standards and 12 objectives. The 12 objectives are very similar within grade ranges (3-6, 7-8 & 10) and parallel across grades. The reviewers found the same proportion of objectives at each of three DOK levels for each grade, five with a DOK level 1 (42%), one with a DOK level 2 (8%), and six with a DOK level 3 (50%) (Table 1). The level of complexity is high with 50% of the 12 objectives rated at the DOK level 3 indicating students are expected to analyze texts, draw information from beyond the text to make inferences, and make conclusions. With the level of complexity high across all grades, there is no increase in sophistication in the DOK level of the objectives. What increase there would be comes from having students read more complex text.

Table 1
Percent of Objectives by Depth-of-Knowledge (DOK) Levels for Grades 3-8 and 10
Wisconsin Alignment Analysis for Reading

Grade	Total number of objectives	DOK Level	# of objs by Level	% within std by Level
3	12	1	5	42
		2	1	8
		3	6	50
4	12	1	5	42
		2	1	8
		3	6	50
5	12	1	5	42
		2	1	8
		3	6	50
6	12	1	5	42
		2	1	8
		3	6	50
7	12	1	5	42
		2	1	8
		3	6	50
8	12	1	5	42
		2	1	8
		3	6	50
10	12	1	5	42
		2	1	8
		3	6	50

Reviewers were instructed to assign an item to a generic objective (the standard) if the item corresponded to the standard, but did not match any of the underlying objectives under the standard. The reading reviewers did not assign any items to a generic objective for each of the seven grades analyzed. This indicates that the statements of the objectives were clearly stated and reviewers were able to find at least one objective that matched each assessment item. Reviewers noted that they found the statement of the objectives to be transparent and clearly written.

Alignment of Curriculum Standards and Assessments

The reading assessments for each grade had from 59 to 63 items (Table 2). The grades 3-8 assessments had 60 multiple-choice items, each assigned one point. The grades 3 and 4 assessments had two constructed-response items, each assigned up to three points. The grades 5-8 had three constructed-response items, each assigned up to three points. The grade 10 items had four constructed-response items, each assigned up to three points.

Table 2
Number of Assessment Items and Item Point Values for Grades 3-8 and 10 Assessments for the Wisconsin Alignment Analysis for Reading

Grade	Number of Multiple-Choice Items	Number and Point Value of Constructed-Response Items	Total Number of Items	Total Number of Points
3	60	2 (3 points)	62	66
4	60	2 (3 points)	62	66
5	60	3 (3 points)	63	69
6	60	3 (3 points)	63	69
7	60	3 (3 points)	63	69
8	60	3 (3 points)	63	69
10	55	4 (3 points)	59	67

The results of the analysis for each of the four alignment criteria are summarized in Table 3. More detailed data on each of the criteria are given in the Appendix in the first three tables. In Table 3, "YES" indicates that an acceptable level was attained between the assessment and the standard 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.

Across the seven grades the alignment of the reading standards and assessments is considered reasonable. At the lower grades, grades 3 and 4, the alignment needs slight improvement primarily because too low of a percentage of the items corresponding to

Standard 3 (analyze text) had a DOK level that was below the DOK level of the corresponding objective. The complexity distribution over the objectives across the grades was constant and higher than for most states that we have reviewed. It is not surprising that the DOK levels at the two lowest grades were not as complex as for other grades. Reviewers' notes and judgments indicate that the proportion of items with a DOK level 3 increased from grade 3 through grade 10. Therefore, the set of assessments had a reasonable progression in complexity across the grades. The alignment for five of the seven grades analyzed were all judged to be at least reasonable. In general, the standards represented high expectations and the assessments increased in a reasonable progression to adequately measure these expectations. More detailed information on the alignment for each grade is given below.

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

Table 3.1
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 3 Standards and Assessment for Wisconsin Alignment Analysis

Grade 3	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
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	NO	YES	YES
4 - Evaluate and extend text.	NO	YES	WEAK	YES

Grade 3

The grade 3 assessment needs slight improvement to attain full alignment with the grade 3 standards. The assessment and Standards 1 and 2 were found to be fully aligned, but some improvement is needed for the assessment to be fully aligned with Standards 3 and 4. Too low of a proportion of the 25 items corresponding to Standard 3 (analyze text) was judged to have a DOK level that was lower than the DOK level of the corresponding objective. All three of the objectives under Standard 3 were judged to have a DOK 3. Items 16, 33, 37, 52, and 61 are example of ones judged to have a DOK level 3 and at an appropriate DOK level as expected for by the objective. However, most of the other items targeting Standard 3 objectives were judged to have a DOK level 2. The depth-of knowledge issue for Standard 3 could be fixed by replacing at least four items now with a DOK level 2 with those that have a DOK level 3 or higher. The other main alignment issue was that the assessment had too few items that corresponded to Standard 4, only three items. Reviewers' comments indicated that they found no items on making predictions, making connections to the text, extending themes to other situations, or

evaluating comparisons and contrasts. Three or four items need to be added (or replace items assigned to the other three standards) to achieve full alignment between the assessment and Standard 4 for grade 3. At least one of these three or four items need to measure students knowledge related to the Objectives 4.2 or 4.3 (informational text or author's use of language). Thus, at least seven items of the grade 3 items need to be replaced to achieve full alignment.

Grade 4

The alignment between the grade 4 standards and assessment needs slight improvement. As for grade 3, the alignment between the assessment and the Standards 1 and 2 is reasonable. The assessment does over emphasize Objective 1.1, but this is not considered a critical alignment issue because the other alignment criteria have been met for Standard 1. The assessment has too few of items that corresponded to Standard 4. A majority of the reviewers indicated that items 36 and 42 (3 points) measured students' knowledge and skills related to Standard 4. This would only give four point values related to this standard. At least two more items need to be added or replace existing items corresponding to other standards to achieve full alignment. The most significant alignment issue at grade 4 is that too many of the items measuring content related Standard 3 were judged to have a DOK level 2, below the DOK level of the objectives under this standard. Reviewers had strong agreement that item 16 had a DOK level 3. But reviewers rated nearly all of the other items corresponding to objective under Standard 3 as having a DOK level 2, even a higher proportion of items than for grade 3. At least eight of the nearly 30 items judged as targeting Standard 3 need to be replaced by items with a DOK level 3. This number could be reduced by replacing some of these items with items that have an appropriate DOK level and correspond to objectives under Standard 4. Overall, from eight to ten items on the grade 4 assessment need to be replaced to attain full alignment.

Table 3.2
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 4 Standards and Assessment for Wisconsin Alignment Analysis

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	WEAK
2 - Understand text.	YES	YES	YES	YES
3 - Analyze text.	YES	NO	YES	YES
4 - Evaluate and extend text.	NO	YES	YES	YES

Grade 5

The alignment between the grade 5 standards was reasonable. The assessment had an acceptable number of items for each of the four standards, six or greater, and had

sufficient coverage of the objectives under each standard. The reviewers noted that the items were more complex than on the assessments for the previous grades. However, the level of complexity of the items targeting Standard 3 still fell below the desired proportion. About three items corresponding to Standard 3 need to be replaced with items at a DOK level 3 to fully meet the acceptable level for the Depth-of-Knowledge Consistency criterion. The majority of reviewers indicated that items 6, 13, 15, 43, and 53 had a DOK level 3 and appropriately matched the desired level of complexity for Standard 3. Items 1, 8, and 11 would be possible items to be replaced with items at a DOK level 3 to achieve full alignment. The balance weakness for Standard 1 is not critical because the other three alignment criteria were met. Overall, only three items would need to be replaced by those with a DOK level 3 for the grade 5 assessment and standards to be fully aligned.

Table 3.3
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 5 Standards and Assessment for Wisconsin Alignment Analysis

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	NO
2 - Understand text.	YES	YES	YES	YES
3 - Analyze text.	YES	NO	YES	YES
4 - Evaluate and extend text.	YES	YES	YES	YES

Table 3.4
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 6 Standards and Assessment for Wisconsin Alignment Analysis

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	WEAK
2 - Understand text.	YES	YES	YES	YES
3 - Analyze text.	YES	WEAK	YES	YES
4 - Evaluate and extend text.	YES	YES	YES	YES

Grade 6

The Grade 6 standards and assessment for reading were found to be reasonably aligned. The reviewers found a higher proportion of items than for previous grades with a DOK level 3 indicating a positive progression in complexity. Only two items that

correspond to objectives under Standard 3 would need to be replaced by items that have a DOK level 3 to attain full alignment. The assessment does over emphasize Objective 1.1, but this is not considered a critical alignment issue because the other alignment criteria for Standard 1 were met. Overall, the alignment at grade 6 is reasonable.

Grade 7

As for grade 6, the alignment of the grade 7 standards and assessment was reasonable. Only three items that correspond to objectives under Standard 3 need to be replaced by items with a DOK level 3 to attain full alignment. Reviewers reported that they found an increasing number of items at a DOK level 3 from the previous grades. With only minor changes needed, the results indicate that the alignment at grade 7 is reasonable.

Table 3.5
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 7 Standards and Assessment for Wisconsin Alignment Analysis

Grade 7	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
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	WEAK	YES	YES
4 - Evaluate and extend text.	YES	YES	YES	YES

Grade 8

The grade 8 standards and assessment were found to be fully aligned. No modifications are needed.

Table 3.6
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 8 Standards and Assessment for Wisconsin Alignment Analysis

Grade 8	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
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	YES	YES	YES

Grade 10

The alignment between the Grade 10 standards and assessment is fully aligned. Even though the assessment over emphasizes Objective 1.1 when compared to number of items that correspond to the other two objectives under Standard 1, this is not considered a major alignment issue. Therefore, the grade10 standards and assessment are considered to be fully aligned.

Table 3.7
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 10 Standards and Assessment for Wisconsin Alignment Analysis

Grade 10	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent ation*
1 - Determine the meaning of words and phrases in context.	YES	YES	YES	WEAK
2 - Understand text.	YES	YES	YES	YES
3 - Analyze text.	YES	YES	YES	YES
4 - Evaluate and extend text.	YES	YES	YES	YES

Reviewers' Comments

Reviewers were instructed to document any Source-of-Challenge issue and to provide any other comments they may have. These comments can be found in Tables (grade).5 and (grade).7 in Appendices B. Only one or two reviewers noted a source-of-challenge issue for fewer than five items for all seven grades. This indicates that reviewers found the items to target the appropriate content knowledge. Reviewers need write a number of notes that mainly indicated how an item or the order of the items could be improved. After coding each grade-level assessment, reviewers also were asked to respond to five debriefing questions. All of the debriefing comments made by the reviewers are given in Appendix C. 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 levels to items was very high (Table 5). An intraclass correlation value greater than 0.8 generally indicates a high level of agreement among the reviewers. All of the intraclass correlations in assigning DOK levels to the items were above .90. A pairwise comparison is used to determine the degree of reliability in the reviewers' coding of items to objectives and standards. An adjudication process was incorporated into the process. The team of eight reviewers reviewed the coding of items to objectives for each item where a majority of reviewers had not coded the item to the same objective. This step in

the process helps to assure the consistency in coding to objectives and standards. After the discussion, the reviewers were not required to change the coding if they felt that their original coding was appropriate. The standard pairwise comparison values are fairly high (nearly all above .80), whereas the objective values are for most grades reasonable.

Table 4
Intraclass and Pairwise Comparisons, Wisconsin Alignment Analysis for Reading

Grade	Intraclass Correlation	Pairwise Comparison:	Pairwise: Objective	Pairwise: Standard
3	.94	.71	.72	.78
4	.92	.68	.76	.82
5	.93	.67	.78	.84
6	.96	.73	.82	.87
7	.96	.76	.78	.85
8	.97	.74	.72	.82
10	.97	.78	.77	.86

Summary

Eight experienced language arts reviewers analyzed the alignment of the reading standards and the assessments for grades 3-8 and 10 at a three-day institute held in Madison, Wisconsin on July 10-12. Six of the reviewers were from Wisconsin and two were from other states. All eight reviewers had participated in previous alignment studies. The reviewers included reading and language arts content experts, district reading coordinators, and reading teachers.

The alignment between the reading standards from the assessment framework and the assessments was found to be reasonable for the seven grades. Reviewers judged that the complexity of the standards was high with 50% of the 12 objectives under the four standards with a depth-of-knowledge level 3 (drawing inferences using information beyond the text, drawing conclusions, and analyzing author's purpose, etc.). For five of the seven grades, the alignment between the standards and assessment was found to be reasonable. Three or fewer items would need to be replaced to achieve full alignment. The alignment for grades 3 and 4 was found to need slight improvement. This is largely due to the relatively high DOK levels expected for these grades, 50% at a DOK level 3. For grades 3 and 4, eight and ten items need to be replaced by items corresponding to Standards 3 (analyze text) and with a DOK level 3. Also, for grades 3 and 4, two or three items need to be added or replaced by items that correspond to Standard 4 (evaluate and extend text). A large proportion of the items for grades 3 and 4 were judged to have a DOK level 2, fairly demanding for these grade levels. Overall, the alignment is considered to be reasonable and the assessment appropriately increasing in complexity over the grades.

References

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.

Valencia, S. W., & Wixson, K. K. (2000). Policy-oriented research on literary standards and assessment. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Vol III.* Mahwah, NJ: Lawrence Erlbaum.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in language arts 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.