REPORT

Alignment Analysis of Mathematics Standards and Assessments

Wisconsin

Grades 3-8 and 10

Norman L. Webb

July 31, 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.

REPORT

Alignment Analysis of Mathematics Standards and Assessments

Wisconsin
Grades 3-8 and 10

Norman L. Webb

July 31, 2006

Acknowledgements

Reviewers:

Rob Ely (Group Leader)	WI
Michael Kestner	DC
Lynn Raith	PA
Kristen Bieda	MO
Rosann Hollinger	Milwaukee
Faye Hilgart	Madison
Tony Pickar	Weston
Jodean Grunow	Dodgeville

CTB/McGraw Hill LLC funded this analysis as part of its contract from the Wisconsin Department of Instruction. Dennis Allion was the main contact person for the CTB-McGraw Hill and oversaw the coordination of the study. Dr. Lynette Russell, Director of Educational Accountability, was the main contact person for the Wisconsin Department of Public Instruction.

Table of Contents

Executive Summary

Eight reviewers analyzed the alignment of the Wisconsin grade level standards as specified in the assessment framework with the assessment for seven grades, 3-8 and 10. Five of the reviewers were from Wisconsin and three were from other states. The reviewers included mathematics content experts, district mathematics supervisors, mathematics teachers, and mathematics education doctoral graduate students.

The alignment between the standards and the assessments was found to be reasonable for four of the seven grades and needs slight improvement for the other three grades (grades 5, 6, and 7). For all seven grades, the assessments had a sufficient number of items for each of the six mathematics standards that were adequately distributed among the objectives. The main alignment issue was that not a high enough proportion of items had a DOK level that was the same or higher than the DOK level of the matching objective. This was the case primarily for one standard (Standard E, Statistics and Probability). Reviewers judged that items corresponding to Standard E had DOK levels of 1 or 2 whereas the DOK levels for the objectives under Standard E were judged to have DOK levels 2 and 3. About nine or ten items would need to be replaced on each of the assessments for grades 5, 6, and 7 to attain full alignment. The alignment for the other grades was found to be reasonable. Overall, the alignment is reasonable with the exception of one standard across the grades. By replacing a few items with those at a DOK level 2 or 3 full alignment would be attained.

Alignment Analysis of Mathematics 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 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 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 mathematics content experts, district mathematics supervisors, mathematics teachers, and mathematics education doctoral graduate students analyzed the agreement between the state's mathematics standards and 2006 assessments for grades 3–8 and 10. Five of the reviewers were from Wisconsin and three were from other states.

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 mathematics 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 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 both objectives within standards and assessment items is an essential requirement of alignment analysis. These descriptions help to clarify what the different levels represent in mathematics:

Level 1 (Recall) includes the recall of information such as a fact, definition, term, or a simple procedure, as well as performing a simple algorithm or applying a formula. That is, in mathematics, a one-step, well defined, and straight algorithmic procedure should be included at this lowest level. Other key words that signify a Level 1 include "identify," "recall," "recognize," "use," and "measure." Verbs such as "describe" and "explain" could be classified at different levels, depending on what is to be described and explained.

Level 2 (Skill/Concept) includes the engagement of some mental processing beyond an habitual response. A Level 2 assessment item requires students to make some decisions as to how to approach the problem or activity, whereas Level 1 requires students to demonstrate a rote response, perform a well-known algorithm, follow a set procedure (like a recipe), or perform a clearly defined series of steps. Keywords that generally distinguish a Level 2 item include "classify," "organize," "estimate," "make observations," "collect and display data," and "compare data." These actions imply more than one step. For example, to compare data requires first identifying characteristics of the objects or phenomenon and then grouping or ordering the objects. Some action verbs, such as "explain," "describe," or "interpret," could be classified at different levels depending on the object of the action. For example, interpreting information from a simple graph, or requiring mathematics information from the graph, also is at Level 2. Interpreting information from a complex graph that requires some decisions on what features of the graph need to be considered and how information from the graph can be aggregated is at Level 3. Level 2 activities are not limited solely to number skills, but can involve visualization skills and probability skills. Other Level 2 activities include noticing and describing non-trivial patterns; explaining the purpose and use of experimental procedures; carrying out experimental procedures; making observations and collecting data; classifying, organizing, and comparing data; and organizing and displaying data in tables, graphs, and charts.

Level 3 (Strategic Thinking) requires reasoning, planning, using evidence, and a higher level of thinking than the previous two levels. In most instances, requiring students to explain their thinking is at Level 3. Activities that require students to make conjectures are also at this level. The cognitive demands at Level 3 are complex and

abstract. The complexity does not result from the fact that there are multiple answers, a possibility for both Levels 1 and 2, but because the task requires more demanding reasoning. An activity, however, that has more than one possible answer and requires students to justify the response they give would most likely be at Level 3. Other Level 3 activities include drawing conclusions from observations; citing evidence and developing a logical argument for concepts; explaining phenomena in terms of concepts; and using concepts to solve problems.

Level 4 (Extended Thinking) requires complex reasoning, planning, developing, and thinking most likely over an extended period of time. The extended time period is not a distinguishing factor if the required work is only repetitive and does not require applying significant conceptual understanding and higher-order thinking. For example, if a student has to take the water temperature from a river each day for a month and then construct a graph, this would be classified as Level 2. However, if the student is to conduct a river study that requires taking into consideration a number of variables, this would be at Level 4. At Level 4, the cognitive demands of the task should be high and the work should be very complex. Students should be required to make several connections—relate ideas *within* the content area or *among* content areas—and to select one approach among many alternatives on how the situation should be solved, in order to be at this highest level. Level 4 activities include developing and proving conjectures; designing and conducting experiments; making connections between a finding and related concepts and phenomena; combining and synthesizing ideas into new concepts; and critiquing experimental designs.

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 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 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 standard 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 mathematics 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

Eight reviewers participated in the depth-of-knowledge (DOK) level consensus process for the standards and standards for the Wisconsin Mathematics standards. A summary of their deliberations is presented in Table 1. The complete group consensus values for each competency and standard can be found in Appendix A. The reviewers judged that the objectives were primarily at the skill and concept level of complexity. The proportion of objectives with a DOK level 2 does increase over the grades along with a slight increase in the proportion of objectives judged to have a DOK level 3. Thus, the objectives do increase some in sophistication and complexity over the grades.

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

Grade	Total number of objectives	DOK Level	# of objs by Level	% within std by Level
3	14	1 2 3	4 8 2	28 57 14
4	14	1 2 3	4 8 2	28 57 14
5	14	1 2 3	3 8 3	21 57 21
6	14	1 2 3	3 8 3	21 57 21
7	14	1 2 3	1 10 3	7 71 21
8	14	1 2 3	1 10 3	7 71 21
10	14	1 2 3	1 9 4	7 64 28

The reviewers were told that within each standard (e.g., *Number Operations and Relationships*), the standards were intended to fully span the content of that standard. For this reason, the reviewers only coded items to a standard if there were no objective that the item appeared to target. Such items are considered to target a generic objective. A large number of items coded to generic objectives may indicate ways in which a

standard's content is not fully spanned or described by its objectives. This may also simply indicate that these items are not as precise as the objectives. Table 2 shows the items on each of the seven assessments that were coded to a generic objective by more than one reviewer.

Only a total of seven items over the seven assessments were coded to generic standards by at least two reviewers. Item 43 on the grade 6 assessment was the only item coded by a majority of the reviewers to a generic objective. Reviewers strongly agreed that this item did not match any of the objectives under Standard M6.D (Measurement). This item should be reviewed and considered for replacement. The other items assigned to a generic objective were only done so by two or three reviewers. In general, reviewers found that the statement of expectations were clearly stated and for nearly all of the items could find at least one objective that was related.

Table 2
Items Coded to Generic Objectives by More Than One Reviewer, Wisconsin Alignment Analysis for Mathematics, Grades 3-8 and 10

Grade	Assessment Item	Generic Objective (Number of Reviewers)
5	33	M5.E (2)
6	6	M6.B (2)
	43	M6.D(6)
8	22	M8.F (2)
	48	M8.F (2)
10	31	M10.B(3)
	34	M10.D (2)

Alignment of Curriculum Standards and Assessments

The assessments for grades 3–8 and 10 were comprised of from 60 to 69 items (Table 3). Most items were 1-point multiple-choice items; the others were constructed response items worth 2 or 4 points each. The total point value ranged from 65 (grade 3) to 76 (grades 5, 6, and 7) points. No field test items were included on the test or in the analysis.

The results of the analysis for each of the four alignment criteria are summarized in Table 4. More detailed data on each of the criteria are given in the Appendix in the first three tables. In Table 4, "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.

Table 3
Number of items and point value by grade for Wisconsin assessments, grades 3-8 and 10

Grade Level	Number of Items	Number of Multi-Point Items	Total Point Value
3	60	5 2-points	65
4	62	6 2-points	68
5	69	7 2-points	76
6	69	7 2-points	76
7	69	7 2-points	76
8	66	8 2-points	74
10	61	5 2-points, 1 4-points	69

For four of the seven grades the alignment between the grade level standards as stated in the assessment framework and the assessments was reasonable. The alignment for grades 5, 6 and 7 need only slight improvement. The main alignment issue was an insufficient number of items corresponding to a standard with an appropriate DOK level. For all seven grades and all standards, the assessments were judged has having an adequate number of items with a sufficient coverage of objectives to fully meet the Categorical Concurrence and Range-of-Knowledge Correspondence criteria. The Balance of Representation criterion was also met for all but a few standards across the seven grades. The alignment for each grade is discussed below in more detail.

Grade 3

The alignment of the standards and assessment for grade 3 was found to be reasonably aligned (Table 4.1). Full alignment could be achieved by replacing three items with items that have a higher DOK level. One item corresponding to Standard M3.A and two items corresponding to Standard M3.E would need to be replaced by items with at least a DOK level 2.

Grade 4

The main alignment issue for grade 4 is the lack of attainment of the DOK Consistency criterion for Standard E (Statistics and Probability). At least one item corresponding to Standard E would need to be replaced by a item with a DOK level 2 or 3 to attain an acceptable level on DOK. The two balance weakness is not of a concerned because the other alignment criteria have been met.

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

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

Grade 3	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
M3.A - Mathematical Processes	YES	WEAK	YES	YES
M3.B - Number Operations and Relationships	YES	YES	YES	YES
M3.C – Geometry	YES	YES	YES	YES
M3.D – Measurement	YES	YES	YES	YES
M3.E - Statistics and Probability	YES	NO	YES	YES
M3.F - Algebraic Relationships	YES	YES	YES	YES

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

Grade 4	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
A - Mathematical Processes	YES	YES	YES	YES
B - Number Operations and Relationships	YES	YES	YES	YES
C – Geometry	YES	YES	YES	YES
D – Measurement	YES	YES	YES	WEAK
E - Statistics and Probability	YES	NO	YES	YES
F - Algebraic Relationships	YES	YES	YES	WEAK

Grade 5

About ten items need to be replaced on the grade 5 assessment by items with a higher DOK level to have full alignment. Two items that correspond to Standard A need to be replaced with items that have a DOK level of 3. Seven of the 13 items that were found to correspond to Standard E need to be replaced with items that have at least a

DOK level 2, but more preferably with items with a DOK level 3. All of the items were judged to have a DOK level that was lower than the DOK level of the match objective. The DOK weakness for Standard F could be removed by replacing one item, such as item 1 or 4, with an item that has a DOK level 2.

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

Grade 5	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
A - Mathematical Processes	YES	NO	YES	YES
B - Number Operations and Relationships	YES	YES	YES	YES
C - Geometry	YES	YES	YES	YES
D - Measurement	YES	YES	YES	YES
E - Statistics and Probability	YES	NO	YES	YES
F - Algebraic Relationships	YES	WEAK	YES	YES

Grade 6

As for grade 5, at least 10 grade 6 items need to be replaced by items that have a higher DOK level, levels 2 or 3. Two items corresponding to Standard A need to be replaced with items that have a DOK level 3. One item that corresponds to Standard B needs to be replaced by an item with a DOK level 2, such as items 5 or 16. None of the 13 items corresponding to Standard E, on the average, was judged to have an appropriate DOK level. At least seven of these items need to be replaced by items with a DOK level 2 or 3.

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

Grade 6	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
A - Mathematical Processes	YES	NO	YES	YES
B - Number Operations and Relationships	YES	WEAK	YES	YES
C - Geometry	YES	YES	YES	YES
D - Measurement	YES	YES	YES	YES
E - Statistics and Probability	YES	NO	YES	YES
F - Algebraic Relationships	YES	YES	YES	YES

Grade 7

Similar to the two previous grades, the main alignment issue for grade 7 is with depth-of-knowledge. About nine items on the grade 7 assessment would need to be replaced to attain full alignment—one item for Standard B, two items for Standard C, five items for Standard E, and one item for Standard F. All of these items need to be replaced by items that have at least a DOK level 2. For Standard E, some of the replacement items should have a DOK level 3.

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

Grade 7	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
A - Mathematical Processes	YES	YES	YES	YES
B - Number Operations and Relationships	YES	WEAK	YES	YES
C - Geometry	YES	WEAK	YES	YES
D - Measurement	YES	YES	YES	YES
E - Statistics and Probability	YES	NO	YES	YES
F - Algebraic Relationships	YES	WEAK	YES	YES

Grade 8

Full alignment could be attained for grade 8 by replacing five items of the 11 that were judged as corresponding to Standard E (Statistics and Probability). These items should be replaced by items with a DOK level 3. The balance weakness for Standard D is not considered of a concerned because the other three alignment criteria have been met.

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

Grade 8	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
A - Mathematical Processes	YES	YES	YES	YES
B - Number Operations and Relationships	YES	YES	YES	YES
C - Geometry	YES	YES	YES	YES
D - Measurement	YES	YES	YES	WEAK

E - Statistics and Probability	YES	NO	YES	YES
F - Algebraic Relationships	YES	YES	YES	YES

Grade 10

Only three items would need to be replaced on the grade 10 assessment to attain full alignment. Two items that corresponds to Standard A (Mathematical Processes) need to be replaced by items that have DOK level 3. One item that targets Standard F needs to be replaced by an item that has a DOK level 2 or 3. The weakness on the Balance of Representation for Standard D is not considered a major concerned.

Table 4.7
Summary of Acceptable Levels on Alignment Criteria for Mathematics 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*
A - Mathematical Processes	YES	NO	YES	YES
B - Number Operations and Relationships	YES	YES	YES	YES
C - Geometry	YES	YES	YES	YES
D - Measurement	YES	YES	YES	WEAK
E - Statistics and Probability	YES	YES	YES	YES
F - Algebraic Relationships	YES	WEAK	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 Appendix B. 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 Appendix B. 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 mathematics reviewers' assignment of DOK levels to items was good (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 standard level. The standard pairwise comparison values are very good (six of seven above .8), while the objective values are reasonable. Some lack

of agreement is explained by ambiguity in the standards, as reflected in the reviewer comments.

Table 5
Intraclass and Pairwise Comparisons, Wisconsin Alignment Analysis for Mathematics, Grades 3–8 and 10

Grade	Intraclass Correlation	Pairwise Comparison:	Pairwise: Objective	Pairwise: Standard
3	.88	.67	.76	.86
4	.92	.68	.77	.86
5	.81	.62	.75	.85
6	.87	.68	.74	.88
7	.91	.69	.74	.87
8	.90	.68	.69	.82
10	.80	.66	.59	.70

Summary

Eight reviewers analyzed the alignment of the Wisconsin grade level standards as specified in the assessment framework with the assessment for seven grades, 3-8 and 10. Five of the reviewers were from Wisconsin and three were from other states. The reviewers included mathematics content experts, district mathematics supervisors, mathematics teachers, and mathematics education doctoral graduate students.

The alignment between the standards and the assessments was found to be reasonable for four of the seven grades and needs slight improvement for the other three grades (grades 5, 6, and 7). For all seven grades, the assessments had a sufficient number of items for each of the six mathematics standards that were adequately distributed among the objectives. The main alignment issue was that not a high enough proportion of items had a DOK level that was the same or higher than the DOK level of the matching objective. This was the case primarily for one standard (Standard E, Statistics and Probability). Reviewers judged that items corresponding to Standard E had DOK levels of 1 or 2 whereas the DOK levels for the objectives under Standard E were judged to have DOK levels 2 and 3. About nine or ten items would need to be replaced on each of the assessments for grades 5, 6, and 7 to attain full alignment. The alignment for the other grades was found to be reasonable. Overall, the alignment is reasonable with the exception of one standard across the grades. By replacing a few items with those at a DOK level 2 or 3 full alignment would be attained.

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.

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