REPORT

Alignment Analysis of Mathematics Extended Grade Band Standards and Assessments

Wisconsin
Grades 3-8 and 10
2008

Norman L. Webb

June 25, 2008

REPORT

Alignment Analysis of Mathematics Extended Grade Band Standards and Assessments

Wisconsin
Grades 3-8 and 10
2008

Norman L. Webb

June 25. 2008

Acknowledgements

Reviewers

Linda McQuillen (Group Leader)   WI
Melvin Diggs   NC
Deena Young   WV
Amber Eckes   WI
Toyree Madison   WI
Stacy Tremaine   WI

Table of Contents

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

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

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

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

Executive Summary

A three-day alignment institute was held in Madison, Wisconsin, on April 18-20, 2008, to analyze the alignment between the Wisconsin Extended Grade Band Mathematics Standards and the Wisconsin Alternate Assessment for Students with Disabilities (WAA-SwD) for grades 3-8 and 10 administered in 2008. Six reviewers conducted the analysis including special education experts in mathematics, district special education coordinators, and special education teachers. Four of the reviewers were from Wisconsin and two were from other states.

The extended grade band standards and the WAA-SwD were found to be fully aligned for three grades (3, 4, and 7) and acceptably aligned for the other four grades (5, 6, 8, and 10). The alignment for grades 5, 6, 8, and 10 were judged to be acceptably aligned because fewer than six items would need to be replaced to attain full alignment. Only grade 8 had one standard that failed to meet an acceptable level for the Categorical Concurrence criterion of six or more items per standard. For grade 8, the majority of reviewers could only agree that the assessment had five items that mapped to objectives under Standard F (Algebraic Relationships). The most frequent alignment issue was with items having an appropriate DOK stage. For grades 5 and 6, fewer than half of the assessment items targeting objectives under Standard F (Algebraic Relationships) were at least the same as the DOK level of the assigned objective. For grade 10, this was true for items targeting Standard D (measurement). The Range-of-Knowledge Correspondence was acceptable for all standards and for all grades. The balance weakness for standards for grades 5 and 10 were because one objective under a standard was over emphasized. When the other three alignment criteria are acceptably met, an over emphasis of one objective is not considered as a crucial issue, but one of preference, because the balance weakness has not affected the overall number of items, the DOK levels of the items, or the number of objectives targeted.

The alignment results are summarized in the table below. The results improved from the analysis conducted in August, 2007. Although the alignment between the mathematics extended standards and the assessments could be improved by carefully selecting some items, the alignment was judged to be at least acceptable for all grades and fully aligned for three grades.

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 Form to be Replaced for Full Alignment*
3	100%	100%	100%	100%	0
4	100%	100%	100%	100%	0
5	100%	80%	100%	80%	4
6	100%	80%	100%	100%	3
7	100%	100%	100%	100%	0
8	80%	100%	100%	100%	2
10	100%	80%	100%	40%	3

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 Mathematics 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 Mathematics and Science Education* (Webb, 1997).

A three-day alignment analysis institute was conducted August 20-22, 2007, in Madison, Wisconsin. Six reviewers, including mathematics content experts, district special education consultants, and special education mathematics teachers analyzed the agreement between the Wisconsin extended grade band standards for mathematics drafted in May 2007 and the Wisconsin Alternate Assessment for Students with Disabilities (WAA-SwD) for grades 3–8 and 10 administered in 2008. 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 mathematics 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. 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 of the extended objectives and assessment items. This training included reviewing the definitions of the six extended depth-of-knowledge (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 extended 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 extended standards 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 mathematics 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.

3

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 mathematics:

Stage 1–Respond

Requires the ability to respond to, or indicate, or acknowledge mathematical features. Example:
♦ Points to a number
♦ Attends to someone counting
♦ Indicates a measuring device, i.e., ruler, measuring cup, scale
♦ Points to common shapes.

Stage 2–Reproduce

Requires the ability to copy, replicate, repeat, re-enact, mirror, or match mathematical features.
Example:
♦ Copies numbers from 1 to 10

- Counts from 1 to 10 with prompting
- Reproduces a pattern, shape, or figure, e.g., circle, square, triangle, rectangle
- Matches even numbers/even numbers
- Matches odd numbers/ and odd numbers
- Matches similar shapes together

Stage 3–Recall and Recognition

Requires students to recall or observe facts, definitions, terms. Involves simple one-step procedures. Involves computing simple algorithms (e.g., sum, quotient). Examples:
- Recalls or recognizes a fact, term, or property
- Conducts simple computation using manipulatives
- Represents in words, pictures, or symbols a math object or relationship
- Performs routine procedure like measuring
- Points to a number and the physical representation of that number (e.g., 2, two balls)
- Identifies common shapes and figures
- Identifies measuring devices, i.e., ruler, measuring cup, scale

Stage 4 (DOK Level 2 Basic Reasoning)

Requires students to make decisions on how to approach a problem. Requires students to compare, classify, organize, estimate, or order data. Typically involves two-step procedures. Examples:
- Specifies and explains relationships between facts, terms, properties, or operations
- Selects procedure according to criteria and performs it
- Solves routine multiple-step problems
- Recognizes the pattern in a series of objects/numbers
- Organizes fractional items from big to small
- Compares different attributes of data
  (Tells which has more/less/same)

Stage 5 (DOK Level 3 Complex Reasoning)

Requires reasoning, planning, or use of evidence to solve problem or algorithm. May involve activity with more than one possible answer. Requires conjecture or restructuring of problems. Involves drawing conclusions from observations, citing evidence and developing logical arguments for concepts. Uses concepts to solve non-routine problems. Examples:
- Analyzes similarities and differences between procedures
- Formulates original problem given situation
- Formulates mathematical model for complex situation

Stage 6 (DOK Level 4 Extended Reasoning)

Requires complex reasoning, planning, developing, and thinking. Typically requires extended time to complete problem, but time spent not on repetitive tasks. Requires students to make several connections and apply one approach among many to solve the problem. Involves complex restructuring of data, establishing and evaluating criteria to solve problems. Examples:

- Applies mathematical model to illuminate a problem, situation
- Conducts a project that specifies a problem, identifies solution paths, solves the problem, and reports results
- Designs a mathematical model to inform and solve a practical or abstract situation

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 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

The consensus DOK value for each mathematics extended objectives can be found in Appendix A. Table 1 shows the percentages of objectives at each DOK level. The DOK levels assigned by the mathematics review group in the August 20-22, 2007 study were used in this study conducted in April 18=20, 2008. Reviewers were asked to review the DOK level for each objective. If for any reason, the group found that the DOK level assigned in August was not appropriate they could change the value. However, reviewers accepted all of the assigned values from the previous study. The majority of objectives for all seven grades were assigned an extended DOK (EDOK) stage of 3 (recall and recognition) or 4 (skills and concepts). The level of complexity of the objective did have some increase in sophistication across the grades. The highest proportion of objectives assigned an EDOK stage 4 were for grade 10.

Table 1

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

Grade	Total Number of Objectives	DOK Level	# of objs by Level	% within std by Level
3	10	2	2	20
		3	7	70
		4	1	10
4	10	2	2	20
		3	7	70
		4	1	10
5	11	3	8	72
		4	3	27
6	11	3	8	72
		4	3	27
7	14	3	8	57
		4	6	42
8	14	3	8	57
		4	6	42
10	9	3	3	33
		4	6	66

If no particular extended objective is targeted by a given assessment item, reviewers were 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 topic not expressly or precisely described in the objectives. These items may highlight areas in the extended objectives that should be changed, or made more precise. Table 2 displays the assessment items coded to generic objectives by more than one reviewer.

Reviewers found a matching objective for all of the grade 4 items. The content targeted by all of the grade 4 items was contained within the objectives underlying the extended objectives. For the other six grades, the majority of the reviewers assigned from one to five items to generic objectives. Reviewers judged that these items did not precisely match the content expectations as described in the extended objectives, but did address content related to the general topic incorporated in a standard. For example, three grade 4 items (Items 15, 21, and 28) assessed students' knowledge of measurement tools other than calendars and clocks, but the objective only stated that students were to use calendars and clocks. Item 9 on the grade 7 assessment required students to identify a graph whereas the corresponding objective expected students to interpret data from tables and graphs. Reviewers indicated that item 9 match expectations in a lower grade and not the expectation as stated for grades 7 and 8.

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

Table 2
Items Coded to Generic Objectives by More Than One Reviewer, Wisconsin Alignment Analysis for Mathematics, Grades 3-8 and 10 2008

Grade	Generic Objective	Assessment Item (Number of Reviewers)
3	Ea	21 (5)
5	Da	15 (6), 28 (6), 31(6)
5	Ea	14 (2),19 (2)
6	Da	20 (3), 23 (3)
6	Ea	25 (4)
7	Ba	10 (2)
7	Cc	27 (4)
7	Da	16 (6), 23 (6)
7	Ea	9 (6)
8	Ba	10 (2), 23 (3)
8	Da	16 (4), 24 (6)
8	Ea	9 (6)
10	Ba	4 (6)
10	Ca	31 (6)
10	Da	26 (6)
10	Fa	3 (6)

Alignment of Curriculum Standards and Assessments

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. Each assessment had a total of 31 items with three of the items assigned two points each. The total point value for each assessment and grade was 34 points.

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.

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 Multi-Point Items (2 points each)	Total Point Value
3	31	3	34
4	31	3	34
5	31	3	34
6	31	3	34
7	31	3	34
8	31	3	34
10	31	3	34

Grade 3

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

Table 4.1
Summary of Acceptable Levels on Alignment Criteria for Mathematics 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*
B - Number Operations & Relationships	YES	YES	YES	YES
C – Geometry able to use	YES	YES	YES	YES
D – Measurement	YES	YES	YES	YES
E - Statistics and Probability	YES	YES	YES	YES
F - Algebraic Relationships	YES	YES	YES	YES

The grade 3 extended mathematics standards and assessment were found to be fully aligned. The assessment and the five mathematics standards had an acceptable level for each of the four alignment criteria—Categorical Concurrence, Depth-of-Knowledge Consistency, Range-of-Knowledge Correspondence, and Balance of Representation. The assessment had from six to nine items for each of the five standards; over 60% of the

items for each standard had a DOK level that was the same or higher than the DOK level of the assigned objective; all but one objective under Standard B had at least one assigned item; and the items were evenly distributed among the objectives. Reviewers' comments noted that the assessment and the standards were aligned. Some reviewers did indicate that the assessment should have more items related to money (e.g. sorting coins), but otherwise, reviewers were complementary of the relationship between the assessment and the standards.

Grade 4

As for grade 3, the grade 4 mathematics extended standards and assessment were found to be fully aligned. The assessment had from six to eight items for each of the five standards; over 80% of the items for each standard had a DOK level that was the same or higher than the DOK level of the assigned objective; all but one objective under Standard B had at least one assigned item; and the items were evenly distributed among the objectives. Reviewers, as for grade 3, would have liked to see more items related to students' use of money, an important life skill. One reviewer also noted that having more than one subtraction problem for grade 4 students would strengthen the assessment. Reviewers did think that the grade 4 assessment was not very different from the grade 3 assessment and had little progression. Even though the two assessments are designed to target the same set of standards, reviewers felt that the grade 4 assessment should targeted more advanced mathematics skills for the grade range (e.g. subtraction).

Table 4.2
Summary of Acceptable Levels on Alignment Criteria for Mathematics 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 Represent ation*
B - Number Operations & Relationships	YES	YES	YES	YES
C - Geometry	YES	YES	YES	YES
D - Measurement	YES	YES	YES	YES
E - Statistics and Probability	YES	YES	YES	YES
F - Algebraic Relationships	YES	YES	YES	YES

Grade 5

The alignment between the grade 5 extended standards and assessment for mathematics was acceptable. The assessment had from six to eight items for each of the five standards, thus meeting the acceptable level for the Categorical Concurrence criterion. The DOK Consistency criterion had an acceptable level for all of the standards except for Standard F (Algebra Relationships). Reviewers found that each of the six items that targeted objectives under Standard F had a DOK level that was lower than the DOK

level of the assigned objective. Other than for items that targeted objective under Standard F, reviewers felt the level of complexity of the items was good. Range was acceptable for all five standards. The Majority of reviewers found items that targeted all of the objectives. There was a balance weakness for Standard E because five of the six items that targeted objectives under this standard matched the same objective (Ea1). This was considered to be more of a preference than an alignment issue since the other three alignment criteria were successfully met for Standard E. Overall, four items would need to be replaced to attain full alignment—three items under Standard F with items that have a higher DOK level and at least one item that currently targets Ea1 with an item that targets Eb1.

Table 4.3
Summary of Acceptable Levels on Alignment Criteria for Mathematics 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*
B - Number Operations & Relationships	YES	YES	YES	YES
C - Geometry	YES	YES	YES	YES
D - Measurement	YES	YES	YES	YES
E - Statistics and Probability	YES	YES	YES	WEAK
F - Algebraic Relationships	YES	NO	YES	YES

Grade 6

The alignment between the grade 6 extended standards and assessment was acceptable. All four alignment criteria were acceptably met for each of the five standards except for DOK levels for items that corresponded to objectives under Standard F (Algebraic Relationships). The assessment had six to eight items for each of the five standards. Nearly all of the items had a DOK level that was the same or higher DOK level as the assigned objective. However, all six items that targeted objectives under Standard F had a DOK stage that was lower than the DOK stage of the assigned objective. The range criterion was fully met for all five standards with all of the objectives having at least one assigned item. The items were also sufficiently distributed among the objectives to have an acceptable level for the balance criterion.

Overall, three items currently targeting objectives under Standard F would need to be replaced to meet the minimum requirement for full alignment. Reviewers did make other comments on how the alignment could be improved. Reviewers felt that too much emphasis was given to pattern recognition rather than to computational skills. They also felt that not enough items had DOK stages 1 or 2 in order to provide some items for students with the lower functionality.

Table 4.4
Summary of Acceptable Levels on Alignment Criteria for Mathematics 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 Represent ation*
B - Number Operations & Relationships	YES	YES	YES	YES
C - Geometry	YES	YES	YES	YES
D - Measurement	YES	YES	YES	YES
E - Statistics and Probability	YES	YES	YES	YES
F - Algebraic Relationships	YES	NO	YES	YES

Grade 7

The grade 7 extended standards and the assessment were found to be fully aligned. All four alignment criteria were acceptably met by all five standards. Even though the assessment and standards were fully aligned, reviewers noted that the assessment did not include any items requiring computing. Reviewers felt computational items (basic fractions, addition, subtraction, etc.) should be included. These computational skills could be used in life situations such as reading price tags. Two or more reviewers did match five of the 31 assessment items to generic objectives. For these items reviewers did not think that these precisely assessed the expectation as stated in the objective. These items should be reviewed. At least one reviewer noted that the level of complexity of the items did not increase greatly from the lower grades. Even though the grade 7 extended standards and the assessment met the minimum requirements to be considered fully aligned, there were a number of ways the assessment could be improved to attain even better alignment.

Table 4.5
Summary of Acceptable Levels on Alignment Criteria for Mathematics 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*
B - Number Operations & Relationships	YES	YES	YES	YES
C – Geometry	YES	YES	YES	YES
D – Measurement	YES	YES	YES	YES
E - Statistics and Probability	YES	YES	YES	YES
F - Algebraic Relationships	YES	YES	YES	YES

Grade 8

The alignment between the grade 8 extended standards and the assessment is acceptable. Reviewers only found five items, on the average, that corresponded to objectives under Standard F (Algebraic Relationships), one fewer than the six needed to attain an acceptable level for the Categorical Concurrence criterion. The standards and assessment had an acceptable level for all other alignment criteria and standards. For Standards B through E, the reviewers found from six to nine corresponding items. The DOK levels of 70% or more of the items for each of the standards had a comparable DOK level. The assessment also had items that targeted at least 80% of the objectives under any one standard with the items sufficiently distributed among the objectives to have balance. Reviewers indicated that some items were too low in content for the grade level (e.g. Item 6). The items that targeted objectives under Standard E did not require students to interpret data from graphs as expected for grade 8. Two reviewers expected to see some items that related to estimation, but found none. As for grade 7, the assessment could have more items on numerical computations (fractions, addition, etc.). Overall, the alignment for grade 8 was acceptable with only two items needed to be replaced by items that clearly target objectives under Standard F. The alignment could be improved beyond this minimum level by including items related to estimation and computations.

Table 4.6
Summary of Acceptable Levels on Alignment Criteria for Mathematics 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*
B - Number Operations & Relationships	YES	YES	YES	YES
C – Geometry	YES	YES	YES	YES
D – Measurement	YES	YES	YES	YES
E - Statistics and Probability	YES	YES	YES	YES
F - Algebraic Relationships	NO (4.83)	YES	YES	YES

Grade 10

The alignment between the grade 10 mathematics extended standards and assessment was acceptable. Reviewers found from six to nine items that targeted objectives under each of the five standards. For four of the five standards, the Depth-of-Knowledge Consistency was acceptable. For Standard D, four of the six items that mapped to that standard targeted Objective Da1. All of these items were assigned a DOK stage 3 whereas Da1 was judged to have a DOK stage 4. On the average, only 20% of the items targeting objectives under Standard D had a DOK level that was the same or higher than the DOK level of the assigned objective. This was too few of items to have an acceptable level for the Depth-of-Knowledge Consistency criterion. Reviewers did find

14

items that mapped to all of the objectives satisfying the range criterion. However, the items were not evenly distributed for three of the five standards resulting in a weak balance for three of the standards. Reviewers, in general, made positive comments about the alignment for grade 10. One reviewer felt the items did not assess the most important parts of the standards. Reviewers did make some of the same comments as for other grades including a concern for the lack of DOK stage 1 and 2 items for the lowest functioning students, the use of the same type of graphs, and not as much progression from the grades 7-8 assessment as could be.

Overall, a minimum of three items would need to be replaced to attain full alignment for grade 10. Three of the items that target objectives under Standard D need to be replaced by those with a higher DOK stage. If one or two of these items target Objective Dc1, then this would remove the balance weakness for Standard D. The balance weakness for Standards C and F are not considered a major issue since the other three alignment criteria have acceptable levels.

Table 4.7
Summary of Acceptable Levels on Alignment Criteria for Mathematics Grade 10 Standards and Assessments for Wisconsin Alignment Analysis 2008

Grade 10	*Alignment Criteria*			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Representation*
B - Number Operations & Relationships	YES	YES	YES	YES
C - Geometry	YES	YES	YES	WEAK
D - Measurement	YES	NO	YES	WEAK
E - Statistics and Probability	YES	YES	YES	YES
F - Algebraic Relationships	YES	YES	YES	WEAK

Source of Challenge Issue and 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 C. Only one reviewer identified issues with one or two items for grades 3 and 5. 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 mathematics reviewers' assignment of DOK levels 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. The standard pairwise comparison values were also very high along with the agreement in assigning items to objectives. After coding the items independently with high agreement, the reviewers adjudicated their results. At this time reviewers could change their codings if they agreed with the argument given by other reviewers. The values presented in Table 5 are after adjudication.

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

Grade	Intraclass Correlation	Pairwise Comparison:	Pairwise: Objective	Pairwise: Standard
3	.94	.74	.90	.96
4	.96	.82	.93	1.00
5	.96	.84	.94	1.00
6	.96	.84	.94	1.00
7	.91	.80	.90	1.00
8	.88	.80	.91	.98
10	.94	.86	1.00	1.00

Summary

A three-day alignment institute was held in Madison, Wisconsin, on April 18-20, 2008, to analyze the alignment between the Wisconsin Extended Grade Band Mathematics Standards and the Wisconsin Alternate Assessment for Students with Disabilities (WAA-SwD) for grades 3-8 and 10 administered in 2008. Six reviewers conducted the analysis including special education experts in mathematics, district special education coordinators, and special education teachers. Four of the reviewers were from Wisconsin and two were from other states.

The extended grade band standards and the WAA-SwD were found to be fully aligned for three grades (3, 4, and 7) and acceptably aligned for the other four grades (5, 6, 8, and 10). The alignment for grades 5, 6, 8, and 10 were judged to be acceptably aligned because fewer than six items would need to be replaced to attain full alignment. Only grade 8 had one standard that failed to meet an acceptable level for the Categorical Concurrence criterion of six or more items per standard. For grade 8, the majority of reviewers could only agree that the assessment had five items that mapped to objectives under Standard F (Algebraic Relationships). The most frequent alignment issue was with items having an appropriate DOK stage. For grades 5 and 6, fewer than half of the assessment items targeting objectives under Standard F (Algebraic Relationships) were at least the same as the DOK level of the assigned objective. For grade 10, this was true for items targeting Standard D (measurement). The Range-of-Knowledge Correspondence was acceptable for all standards and for all grades. The balance weakness for standards

for grades 5 and 10 were because one objective under a standard was over emphasized. When the other three alignment criteria are acceptably met, an over emphasis of one objective is not considered as a crucial issue, but one of preference, because the balance weakness has not affected the overall number of items, the DOK levels of the items, or the number of objectives targeted.

The alignment results are summarized in the table below. The results improved from the analysis conducted in August, 2007. Although the alignment between the mathematics extended standards and the assessments could be improved by carefully selecting some items, the alignment was judged to be at least acceptable for all grades and fully aligned for three grades.

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 Form to be Replaced for Full Alignment
3	100%	100%	100%	100%	0
4	100%	100%	100%	100%	0
5	100%	80%	100%	80%	4
6	100%	80%	100%	100%	3
7	100%	100%	100%	100%	0
8	80%	100%	100%	100%	2
10	100%	80%	100%	40%	3

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, mathematics, 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.