

REPORT

Alignment Analysis of Extended Science Grade Band Standards and Alternate Assessments

**Wisconsin
Grades 4, 8 and 10
2008**

Norman L. Webb

June 25, 2008

REPORT

Alignment Analysis of Extended Science Grade Band Standards and Alternate Assessments

**Wisconsin
Grades 4, 8, and 10
2008**

Norman L. Webb

June 25, 2008

Acknowledgements

| | | |
|-------------------|----------------|----|
| Robert Stanley | (Group Leader) | NC |
| David Kessler | | NC |
| Larry Schoenemann | | WI |
| Mary Derginer | | WI |
| Jennifer Johnson | | WI |
| Linda Diaz | | WI |

CTB/McGraw Hill LLC funded this analysis as part of its contract from the Wisconsin Department of Public Instruction. Sandra Snell was the main contact person for CTB/McGraw Hill and oversaw the coordination of the study. Philip Olsen (Assistant Director) and Brian Johnson (Education Consultant), Office of Educational Accountability, Division for Reading and Student Achievement, were the main contact persons for the Wisconsin Department of Public Instruction.

Table of Contents

| | |
|---|----|
| Executive Summary | v |
| Introduction..... | 1 |
| Alignment Criteria Used for This Analysis | 3 |
| Categorical Concurrence..... | 3 |
| Depth-of-Knowledge Consistency..... | 4 |
| Range-of-Knowledge Correspondence..... | 6 |
| Balance of Representation | 7 |
| Source of Challenge..... | 7 |
| Findings..... | 7 |
| Extended Standards..... | 7 |
| Alignment of Curriculum Standards and Assessments..... | 9 |
| Source-of-Challenge Issues and Reviewers’ Comments | 13 |
| Reliability Among Reviewers..... | 13 |
| Summary | 14 |
| References..... | 16 |
| Appendix A Wisconsin Grades 4, 8, and 10 Science Extended Standards and Group Consensus DOK Values | |
| Appendix B Data Analysis Tables Wisconsin Grades 4, 8, and 10 Science 2008 | |
| Appendix C Reviewers’ Notes and Source-of-Challenge Comments Wisconsin Grades 4, 8, and 10 Science 2008 | |
| Appendix D Debriefing Summary Notes Wisconsin Grades 4, 8, and 10 Science 2008 | |

Executive Summary

A two-day alignment institute to analyze the agreement between the Wisconsin alternate assessments and extended standards in science was held April 18 and 19, 2008, in Madison, Wisconsin. Six reviewers including special education experts and special education science teachers participated in the analysis. Four of the reviewers were from Wisconsin and two were from another state. The Wisconsin Alternate Assessment for Students with Disabilities (WAA-SwD) in science for grades 4, 8 and 10 for 2008 was compared to the extended standards drafted in May 2007.

The alignment between the science extended standards and the alternate assessment varied by grade. The alignment for grades 4 and 8 needed slight improvement while the alignment for grade 10 was acceptable. At each grade the Categorical Concurrence criterion had an acceptable level (six or more items) for four or five of the six science standards. Reviewers did not find a sufficient number of items on any of the three assessments for Standard A-B (science connections/nature of science); on the grade 8 assessment for Standard C (inquiry); and on the grade 10 assessment for Standard G-H (science applications/science in social and personal perspectives). Even though the test specifications assigned at least six items to each standard for each grade, the reviewers did not agree with the mapping to the specific standard for at least one item (grade 10 Standards A-B and G-H) to five items (grade 4 Standard A-B).

The other alignment issue was with EDOK stages of the assessment items. Too small of a proportion of the items had an EDOK stage that was lower than the EDOK stage of the assigned objective for one grade 4 standard (D) and two grade 8 standards (C and F). For all grades the majority of items were assigned an EDOK stage 3 (recall). However, one or two objectives for each grade level was judged to require basic reasoning (an EDOK stage 4), such as sorting or classifying. As a result the Depth-of-Knowledge Consistency criterion was not met for grade 4 Standard D and grade 8 Standards C and F. The range and balance was acceptable for all standards for all three grades.

Overall, seven items for grade 4, six items for grade 8, and only two items for grade 10 would need to be replaced or added to attain full alignment. Thus, the alignment for grades 4 and 8 needed slight improvement while the alignment for grade 10 was judged as acceptable. The findings for science are summarized in the table below.

Summary Table

Percent of Wisconsin Extended Grade Band Science Standards with Acceptable Level on Each Alignment Criteria for Grades 4, 8, and 10 for WAA-SwD Analysis

| Grade | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> | <i>Estimated Average Number of Items per Form to be Replaced for Full Alignment</i> |
|-------|--------------------------------|---------------------------------------|---------------------------|----------------------------------|---|
| 3 | 84% | 80% | 100% | 80% | 7 |
| 4 | 67% | 67% | 100% | 100% | 6 |
| 5 | 67% | 100% | 100% | 100% | 2 |

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 Extended Grade Band Science Standards and Alternate Assessments

Wisconsin Grades 4, 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 Science and Science Education* (Webb, 1997).

A two-day alignment analysis institute for science was conducted April 18-19, 2008, in Madison, Wisconsin. Six reviewers, including special education experts and special education science teachers analyzed the agreement between the Wisconsin extended grade band standards for science drafted in May 2007 and the Wisconsin Alternate Assessment for Students with Disabilities (WAA-SwD) for grades 4, 8 and 10 for 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, 7-8, and 10), and *achievement descriptors* in its science 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 reviewed the consensus EDOK stages assigned to the objectives in the August 2007 study. The values from the August 2007 study were used in this analysis. Next the reviewers coded the items by assigning an EDOK stage to an item and the most appropriate objective. 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 science for grades 4, 8 and 10 in Wisconsin. The study addressed specific criteria related to the content agreement between the state

extended grade band standards and grade-level assessments. Four criteria received major attention: categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance of representation.

Alignment Criteria Used for This Analysis

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

Categorical Concurrence

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

Depth-of-Knowledge Consistency

Standards and assessments can be aligned not only on the category of content covered by each, but also on the basis of the complexity of knowledge required by each. *Depth-of-knowledge consistency between standards and assessment indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards.* For consistency to exist between the assessment and the standard, as judged in this analysis, at least 50% of the items corresponding to a standard had to be at or above the level of knowledge of the standard: 50%, a conservative cutoff point, is based on the assumption that a minimal passing score for any one standard of 50% or higher would require the student to successfully answer at least some items at or above the depth-of-knowledge level of the corresponding standard. For example, assume an assessment included six items related to one standard and students were required to answer correctly four of those items to be judged proficient—i.e., 67% of the items. If three, 50%, of the six items were at or above the depth-of-knowledge level of the corresponding objectives, then for a student to achieve a proficient score would require the student to answer correctly at least one item at or above the depth-of-knowledge level of one objective. Some leeway was used in this analysis on this criterion. If a standard had between 40% and 50% of items at or above the depth-of-knowledge levels of the objectives, then it was reported that the criterion was “weakly” met.

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

Stage 1 (Respond)

Requires the ability to respond to or indicate or acknowledge scientific features.

Example:

- ◆ Points to a rock.
- ◆ Attends to someone measuring.
- ◆ Indicates a measuring device, i.e., ruler, measuring cup, scale.
- ◆ Points to common animals, physical objectives, insects, etc.

Stage 2 (Reproduce)

Requires the ability to copy, replicate, repeat, re-enact, mirror, or match scientific ideas.

Example:

- ◆ Copies figure of insect, bird, or animal with distinguishing features.

- ◆ Traces movement of sun.
- ◆ Repeats indication of a plant growing with sun.
- ◆ Reproduces indication that fish live in water.
- ◆ Matches a number on a scale.
- ◆ Matches similar shapes together.

Stage 3 (Recall)

Requires students to recall or observe facts, definitions, terms. Involves simple one-step procedures. Requires a demonstration of a rote response, use of a well-known formula, or follow a set procedure (like a recipe), or perform a clearly defined series of steps.). Either knows the answer or not. Examples:

- ◆ Recall or recognize a fact, term, or property.
- ◆ Represent in words or diagrams a scientific concept or relationship.
- ◆ Provide or recognize a standard scientific representation for simple phenomenon.
- ◆ Perform a routine procedure, such as measuring length.
- ◆ Identifies common shapes and figures.
- ◆ Identifies measuring devices, i.e., ruler, measuring cup, scale.

Stage 4 (Basic Reasoning)

Requires students to make some decisions as to how to approach the question or problem. Keywords that generally distinguish a Stage 4 item include “classify,” “organize,” “estimate,” “make observations,” “collect and display data,” and “compare data.” These actions imply more than one step. Examples:

- ◆ Make observations and collect data.
- ◆ Classify, organize, and compare data.
- ◆ Organize and display data in tables, graphs, and charts.
- ◆ Select a procedure according to specified criteria and perform it.
- ◆ Formulate a routine problem, given data and conditions.
- ◆ Organize, represent, and interpret data.
- ◆ Specify and explain the relationship between facts, terms, properties, or variables.
- ◆ Describe and explain examples and non-examples of science concepts.

Stage 5 (Complex Reasoning)

Requires reasoning, planning or use of evidence than previous stages. May involve activity with more than one possible answer. An activity that has more than one possible answer and requires students to justify the response they give would most likely be a Stage 5. Experimental designs at Stage 5 typically involve more than one dependent variable. Requires drawing conclusions from observations; citing evidence and developing a logical argument for concepts; and explaining phenomena in terms of concepts.

Examples:

- ◆ Form conclusions from experimental data.
- ◆ Solve non-routine problems.
- ◆ Develop a scientific model for a complex situation.
- ◆ Identify research questions and design investigations for a scientific problem.

Level 6 (Extended Reasoning)

Involves high cognitive demands and complexity. Students are required to make several connections—relate ideas within the content area or among content areas—and have to select or devise one approach among many alternatives to solve the problem. Requires complex reasoning, experimental design and planning, and probably will require an extended period of time either for the science investigation required by an objective, or for carrying out the multiple steps of an assessment item. Examples:

- ◆ Based on data provided from a complex experiment that is novel to the student, deduct the fundamental relationship between several controlled variables.
- ◆ Conduct an investigation, from specifying a problem to designing and carrying out an experiment, to analyzing its data and forming conclusions.

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 science 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

Extended Standards

The consensus EDOK value for each extended objective under the model standards for science can be found in Appendix A. Table 1 shows the percentages of objectives at each EDOK stage. The complexity of the science objectives remained the same across the three grades. Reviewers judged that around 80% of the objectives had an

EDOK stage 3 and 20% of the objectives had an EDOK 4. In science, there was no increase in complexity across the grades.

Table 1
Percent of Objectives by Depth-of-Knowledge (DOK) Levels for Grades 4, 8 and 10 Wisconsin Alignment Analysis for Extended Standards for Science

| Grade | Total Number of Objectives | EDOK Stage | Number of Objectives by Stage | Percent within Standard by Stage |
|-------|----------------------------|------------|-------------------------------|----------------------------------|
| 4 | 7 | 3 | 6 | 85 |
| | | 4 | 1 | 14 |
| 8 | 8 | 3 | 6 | 75 |
| | | 4 | 2 | 25 |
| 10 | 10 | 3 | 8 | 80 |
| | | 4 | 2 | 20 |

If no particular objective was 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 not expressly or precisely described in the objectives. These items may highlight areas in the objectives that should be changed, or made more precise.

Table 2 displays the assessment items coded to generic objectives by more than one reviewer. The majority of reviewers assigned one grade 4 item to the generic objective F1. Grade 4 Item 4 required students to recognize a plant, but the item did not require students to recognize what was needed for a plant to live and grow as specified in Objective F1a. The reviewers were able to find an objective that matched each of the items on the grade 8 assessment. For grade 10, the majority of reviewers assigned four items to the generic objective C. For three of these items (17, 20, and 29), the reviewers noted that the question asked students to identify a tool. These items matched a grade 4 objective, but the grade 10 item required students to use tools. One reviewer explained, “The 4th grade standard for tools was used as scientific inquiry as written. The question could have been made a C1 (grade 10) standard by asking what would they use this tool for/what did they see/what happened . . .” The majority of reviewers also coded grade 10 Item 33 to the generic objective C. This item required students to read a graph. Reviewers found this item to be more of a mathematics item rather than science item.

Table 2
Items Coded to Generic Objectives by More Than One Reviewer, Wisconsin Alignment Analysis for Science, Grades 4, 8, and 10 2008

| Grade | Generic Objective | Assessment Item (Number of Reviewers) |
|-------|-------------------|---------------------------------------|
| 4 | F1 | 10 (5) |
| 10 | C | 17 (6), 20 (6), 29 (6), 33 (6) |

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

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 3-point item is counted towards the alignment as three identically coded 1-point items. Each science assessment had 36 items with one to three multiple point items. The grade 4 assessment had a total of 37 points. The grades 8 and 10 assessments had a total of 39 points.

Table 3
Number of Items and Point Value by Grade for Wisconsin Assessments, Grades 4, 8, and 10 2008

| Grade Level | Number of Items | Number of Multi-Point Items | Total Point Value |
|-------------|-----------------|-----------------------------|-------------------|
| 4 | 36 | One 2-point | 37 |
| 8 | 36 | Three 2-point | 39 |
| 10 | 36 | One 2-point One 3-point | 39 |

The results of the analysis for each of the four alignment criteria are summarized in Tables 4.1-4.3. 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 4.1-4.3, “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.

Grade 4

The alignment between the grade 4 alternate assessment and extended standards for science was found to need slight improvement. The majority of reviewers only coded one item as corresponding to Objective A-B1 (use science resources to gather information), Item 27. One reviewer assigned three other items to Objective A-B1 (Items 5, 7, and 10), whereas other reviewers assigned these items to Objectives G-H1 (Items 5), C1 (Item 7), and F1 (Item 10). It appears that one reviewer found these items to be

related to using science resources to gather information (Objective A-B1), but the majority of reviewers thought these items related to science inquiry or one of the science content areas rather than science processes. Even with the four items assigned to Objective A-B1 by at least one reviewer, the number of items found to relate to this objective still fell short of the six items that is used as an acceptable level for the Categorical Concurrence criterion. Reviewers found from six to 10 items that corresponded to the other five standards.

The other main alignment issue was with the EDOK stages for Standard D (physical science). Reviewers, on the average, only found one of six items that corresponded to Objective D1a (conceptual understanding including comparing and contrasting). The objective was assigned an EDOK stage 4. However, nearly all of the reviewers assigned items mapped to Objective D1a an EDOK stage 3 (recall). Otherwise, all of the other standards with enough items to be considered tested met an acceptable level on the Depth-of-Knowledge Consistency criterion.

The Range-of-Knowledge Correspondence criterion had an acceptable level for all of the grade 4 science standards. This would be expected with only one or two objectives for each standard. Balance of Representation had a weakness only for Standard E. Reviewers coded five of six items assigned to Standard E to Objective E1b and only one item to Objective E1a. Because all of the other alignment criteria had an acceptable level for Standard E, the balance weakness for Standard E is considered a matter of preference rather than a major alignment issue. It is worthy to consider if Objective E1b should be given more weight on the assessment than Objective E1a.

Table 4

Summary of Acceptable Levels on Alignment Criteria for Science Grades 4, 8, and 10, Standards and Assessments for Wisconsin Alignment Analysis 2008

Table 4.1

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

| Grade 4 | Alignment Criteria | | | |
|---|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Standards | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| A-B - A. Science Connections B. Nature of Science | NO (1.17) | NT | NT | NT |
| C - Science Inquiry | YES | YES | YES | YES |
| D - Physical Science | YES | NO | YES | YES |
| E - Earth and Space Science | YES | YES | YES | WEAK |
| F - Life and Environmental Science | YES | YES | YES | YES |
| G-H - G. Science Applications H. Science in Social and ... | YES | YES | YES | YES |

Overall, the alignment for grade 4 science needed slight improvement. A total of seven items would need to be replaced or added to attain full alignment. Five items would need to be replaced that currently map to standards other than Standards A-B by items that clearly target Objective A-B1. Reviewers struggled with items that possibly were designed to map Objective A-B1, but did not think that these items required students to gather information or matched the examples given under Objective A-B1. It also appears that reviewers had difficulty distinguishing between Objective A-B1 and Objective C1 and between the use of resources and the use of tools. In addition to the five items needed for Objective A-B1, two items targeting Objective D1a would need to be replaced by items with at least an EDOK stage 4. Reviewers coded nearly all of the items with an EDOK stage 2 or 3. At least three of the six items that targeted Objective D1a needed to have an EDOK stage 4.

Grade 8

The alignment between the grade 8 extended standards and assessment needed slight improvement. The assessment and four of the six standards satisfied an acceptable level of six or more items for the Categorical Concurrence criterion. The acceptable level was not met for Standards A-B and C. For these two standards reviewers did not find six items/points that required students to use specific materials to represent science concepts (Standard A-B) or identify simple cause and effect relationships (Standard C). Reviewers assigned nearly all of the assessment items an EDOK stage 3 (recall). However, two of the grade 8 objectives (C1 and F1a) were judged to have an EDOK stage 4 (basic reasoning). Objective C1 expected students to identify a cause and effect relationship. Objective F1a expected students to sort or classify. However, nearly all of the items asked students just to identify or match a relationship. As a result, the grade 8 standards and assessment did not have an acceptable level for the Depth-of-Knowledge Consistency criterion for Standards C and F. The depth-of-knowledge level was good for the other four standards. Both range and balance were acceptable for all six standards.

Overall, the alignment for grade 8 science needed slight improvement. Full alignment could be attained by replacing or adding six items to the assessment that would more clearly meet the expectations as stated in the standards. One item (or point) would need to be added to target Objective A-B1 and two items (or points) would need to be added to target Objective C1. If the additional two items that target Objective C1 had at least an EDOK stage 4, then the Depth-of-Knowledge Consistency criterion would be acceptable for Standard C. At least three of the items that currently target Objective F1a would need to be replaced by items that have an EDOK stage 4 (basic reasoning) to meet the DOK consistency criterion for Standard F. One reviewer noted that nearly all of the items that mapped to Objective F1a were matching a mother to an off spring. This reviewer felt that more items under Standard F were needed on choosing between living and non-living things. This reviewer wrote, "...the standard is focusing on sorting and classifying characteristics of a living thing. This assessment does not focus on sorting or classifying characteristics; more with matching off springs which is not at an EDOK level 4." Reviewers did note that the alignment for grade 8 was an improvement over the alignment for grade 4. The assessment items more clearly mapped to the proficiency

examples provided under the objectives.

Table 4.2

Summary of Acceptable Levels on Alignment Criteria for Science Grade 8, Standards and Assessments for Wisconsin Alignment Analysis 2008

| Grade 8 | Alignment Criteria | | | |
|---|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Standards | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| A-B - A. Science Connections B. Nature of Science | NO (5.17) | YES | YES | YES |
| C - C. Science Inquiry | NO (4.0) | NO | YES | YES |
| D - D. Physical Science | YES | YES | YES | YES |
| E - E. Earth and Space Science | YES | YES | YES | YES |
| F - F. Life and Environmental Science | YES | NO | YES | YES |
| G-H - G. Science Applications H. Science in Social and ... | YES | YES | YES | YES |

Grade 10

The alignment between the grade 10 science extended standards and assessment was found to be acceptable. Three of the four alignment criteria were acceptable for all six standards—Depth-of-Knowledge Consistency, range, and balance. The only issues were with the number of items reviewers found that mapped onto two of the standards—A-B and G-H. For each of these standards, the majority of reviewers only found five items rather than the six needed to have an acceptable level for Categorical Concurrence. The other four standards all had from six to nine corresponding items, a sufficient number to satisfy the Categorical Concurrence criterion. It should be noted, however, that four items that mapped to Standard C were coded by the reviewers to the generic objective rather than Objective C1. For most of these items, reviewers found that these items related to scientific tools, a topic included under Standard C for grade 4, but not mentioned in the grade 10 standards. Reviewers felt that the alignment for grade 10 was better than for either grade 4 or grade 8, but did think there were more items at grade 10 that only were somewhat related to the stated expectation rather than covering all that was sought under the objectives.

Overall, the alignment for grade 10 was acceptable. Only two items would need to be added or replaced to attain full alignment—one that mapped to Standard A-B and one that mapped to Standard G-H. The grade 10 assessment included three items that reviewers clearly felt were an EDOK stage 4, more than for the prior grades. There were items that could be improved because they only addressed a small part of an objective rather than the full intent of the expectation.

Table 4.3

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

| <i>Grade 10 Standards</i> | <i>Alignment Criteria</i> | | | |
|---|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| A-B - A. Science Connections B. Nature of Science | NO (5.33) | YES | YES | YES |
| C - C. Science Inquiry | YES | YES | YES | YES |
| D - D. Physical Science | YES | YES | YES | YES |
| E - E. Earth and Space Science | YES | YES | YES | YES |
| F - F. Life and Environmental Science | YES | YES | YES | YES |
| G-H - G. Science Applications H. Science in Social and ... | NO (5.0) | YES | YES | YES |

Source-of-Challenge Issues 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. Two reviewers identified two grade 4 items with a source-of-challenge issue—Items 5 and 27. These reviewers questioned if Item 5 was really a science question rather than only a reading question. Two reviewers felt that Item 27 could have more than one answer or the picture may imply the answer. One reviewer identified a source of challenge issue for a few items for grades 8 and 10. All of source-of-challenge notes should be reviewed, even if not verified by a second reviewer. It is possible that one reviewer found a valid issue that other reviewers missed.

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 science reviewers' assignment of DOK levels to items was high for six reviewers for grades 4, 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 standard level. Both the standard and objective pairwise comparison values were high. Reviewers adjudicated their codings for each grade after they independently assigned objectives and EDOK stages to each item.

Reviewers could change their coding after the discussion, but were not required to do so. The values in Table 5 are those after adjudication.

Table 5
Intraclass and Pairwise Comparisons, Wisconsin Alignment Analysis for Science Grades 4, 8, and 10 Assessments

| Grade | Intraclass Correlation | Pairwise Comparison: | Pairwise: Objective | Pairwise: Standard |
|-------|------------------------|----------------------|---------------------|--------------------|
| 4 | .90 | .90 | .92 | .92 |
| 8 | .89 | .86 | .96 | .97 |
| 10 | .90 | .88 | .96 | .98 |

Summary

A two-day alignment institute to analyze the agreement between the Wisconsin alternate assessments and extended standards in science was held April 18 and 19, 2008, in Madison, Wisconsin. Six reviewers including special education experts and special education science teachers participated in the analysis. Four of the reviewers were from Wisconsin and two were from another state. The Wisconsin Alternate Assessment for Students with Disabilities (WAA-SwD) in science for grades 4, 8 and 10 for 2008 was compared to the extended standards drafted in May 2007.

The alignment between the science extended standards and the alternate assessment varied by grade. The alignment for grades 4 and 8 needed slight improvement while the alignment for grade 10 was acceptable. At each grade the Categorical Concurrence criterion had an acceptable level (six or more items) for four or five of the six science standards. Reviewers did not find a sufficient number of items on any of the three assessments for Standard A-B (science connections/nature of science); on the grade 8 assessment for Standard C (inquiry); and on the grade 10 assessment for Standard G-H (science applications/science in social and personal perspectives). Even though the test specifications assigned at least six items to each standard for each grade, the reviewers did not agree with the mapping to the specific standard for at least one item (grade 10 Standards A-B and G-H) to five items (grade 4 Standard A-B).

The other alignment issue was with EDOK stages of the assessment items. Too small of a proportion of the items had an EDOK stage that was lower than the EDOK stage of the assigned objective for one grade 4 standard (D) and two grade 8 standards (C and F). For all grades the majority of items were assigned an EDOK stage 3 (recall). However, one or two objectives for each grade level was judged to require basic reasoning (an EDOK stage 4), such as sorting or classifying. As a result the Depth-of-Knowledge Consistency criterion was not met for grade 4 Standard D and grade 8 Standards C and F. The range and balance was acceptable for all standards for all three grades.

Overall, seven items for grade 4, six items for grade 8, and only two items for grade 10 would need to be replaced or added to attain full alignment. Thus, the alignment for grades 4 and 8 needed slight improvement while the alignment for grade 10 was judged as acceptable. The findings for science are summarized in the table below.

Summary Table

Percent of Wisconsin Extended Grade Band Science Standards with Acceptable Level on Each Alignment Criteria for Grades 4, 8, and 10 for WAA-SwD Analysis

| Grade | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> | <i>Estimated Average Number of Items per Form to be Replaced for Full Alignment</i> |
|-------|--------------------------------|---------------------------------------|---------------------------|----------------------------------|---|
| 3 | 84% | 80% | 100% | 80% | 7 |
| 4 | 67% | 67% | 100% | 100% | 6 |
| 5 | 67% | 100% | 100% | 100% | 2 |

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

References

- Edvantia, Inc. (May, 2007). *Wisconsin extended grade band standards: reading, reading, science*. A draft document submitted to the Wisconsin Department of Public Instruction. Charleston, West Virginia: Author.
- Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement*, 25(1), 47-55.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Council of Chief State School Officers and National Institute for Mathematics Education Research Monograph No. 6. Madison: University of Wisconsin, Wisconsin Center for Education Research.