

REPORT

Alignment Analysis of Reading Standards and Assessments

**South Dakota
Grades 3-8 and 11
2009 and 2010**

Norman L. Webb

December 2, 2009

REPORT

Alignment Analysis of Reading Standards and Assessments

**South Dakota
Grades 3-8 and 11
2009 and 2010**

Norman L. Webb

December 2, 2009

Acknowledgements

Reviewers

| | | |
|----------------|--------------|--------------|
| John Fortier | Group Leader | Wisconsin |
| Ellen Last | | Wisconsin |
| Cindy Jacobson | | Wisconsin |
| Tammy Bauck | | South Dakota |
| Rich Meier | | South Dakota |
| Darci Love | | South Dakota |

The South Dakota Department of Education funded this analysis. Gay Pickner, Director of Assessment, Office of Curriculum, Technology, and Assessment, was the main contact person for the South Dakota Department of Education and oversaw the coordination of the study.

Table of Contents

| | |
|---|----|
| Executive Summary | v |
| Introduction..... | 1 |
| Alignment Criteria Used for This Analysis | 3 |
| Categorical Concurrence..... | 3 |
| Depth-of-Knowledge Consistency..... | 3 |
| Range-of-Knowledge Correspondence..... | 5 |
| Balance of Representation | 6 |
| Source of Challenge..... | 6 |
| Findings..... | 7 |
| Standards..... | 7 |
| Alignment of Curriculum Standards and Assessments..... | 8 |
| Source-of-Challenge Issues and Reviewers’ Comments | 22 |
| Reliability Among Reviewers..... | 22 |
| Summary | 23 |
| References..... | 25 |
| Appendix A | |
| South Dakota Grades 3-8 and 11 Reading Standards and Group Consensus DOK Values 2009 and 2010 | |
| Appendix B | |
| Data Analysis Tables South Dakota Grades 3-8 and 11 Reading 2009 and 2010 | |
| Appendix C | |
| Reviewers’ Notes and Source-of-Challenge Comments South Dakota Grades 3-8 and 11 Reading 2009 and 2010 | |
| Appendix D | |
| Debriefing Summary Notes South Dakota Grades 3-8 and 11 Reading 2009 and 2010 | |

Executive Summary

A four- day alignment institute was held October 19 through October 22, 2009 in Sioux Falls, South Dakota to analyze the South Dakota 2009 Reading Standards and the assessments for two years, 2009 and 2010. One group of six reviewers participated in the institute. Three of the reviewers were from South Dakota and three were from other states. The reviewers included reading education content experts, district reading supervisors, reading teachers, and reading assessment experts. The group analyzed assessments and standards for grades 3-8 and 11.

The alignment between the South Dakota 2009 reading indicators and the 2009 and 2010 assessments were found to be acceptable or in need of slight improvement with some variation by grade. The alignment for both years was acceptable for grades 3, 4, 7, and 11. The alignment needed some improvement for grades 5 and 6. The grade 8 2009 assessment and the standards needed slight improvement while the grade 8 2010 assessment was considered as acceptable.

At least four indicators for each assessment satisfied an acceptable level for the Categorical Concurrence criterion of six or more corresponding items. However, the majority of reviewers found fewer than six items that targeted one indicator for 10 of the 14 assessments. For most of these assessments the reviewers only found four or five items that corresponded to Indicator R.4 (interpret and respond to diverse, multicultural, and time period texts). Reviewers indicated that they had difficulty deciphering between items that responded to multicultural text and those that assessed general comprehension (Indicator R.2).

The other main alignment issue was with Depth-of-Knowledge Consistency. Reviewers judged that the majority of the items on each assessment had a DOK level 2 (e.g. contextual clues and comprehension of text). However, the proportion of standards judged to have a DOK level 3 (e.g. making inferences) increased over the grades from about 9% (grade 4) to 55% (grade 8). There was some increase in the proportion of items with a DOK level 3 in the higher grades (particularly for grade 11), but for most grades the proportion of items with a DOK level 3 was not sufficient to have at least 50% of the items with a DOK that was the same or higher than the DOK level of the assigned standard. The level of complexity for the items was low mainly for Indicators R.4 and R.5.

Range and balance was acceptable for nearly all of the indicators for all of the 14 assessments. The items were adequately distributed among the standards under an indicator without over emphasizing any one of the standards. An adequate range and balance was fairly easily attained because of the low number of standards that was under each indicator.

Overall, nine of the 14 assessments were judged to have acceptable alignment with the reading indicators. For these assessments fewer than five items would need to be replaced or added to attain full alignment. Five of the assessments and reading indicators

were judged to need slight improvement in alignment. For these assessments five or six items would need to be replaced or added to attain full alignment.

Summary Table

Percent of South Dakota Reading 2009 Standards with Acceptable Level on Each Alignment Criteria for Grade 3-8 and 112009 and 2010 Assessments

| Grade | <i>Categorical Concurrence</i> (six or more items) | <i>Depth-of-Knowledge Consistency</i> (50% at/above) | <i>Range of Knowledge</i> (50% of objectives) | <i>Balance of Representation</i> (without possible weakness) | <i>Estimated Range of Items per to be Added or Replaced for Full Alignment</i> |
|---------|---|---|--|---|--|
| 3 2009 | 80 | 80 | 100 | 100 | 3 |
| 3 2010 | 80 | 80 | 100 | 100 | 3 |
| 4 2009 | 80 | 80 | 80 | 80 | 4 |
| 4 2010 | 100 | 80 | 80 | 100 | 4 |
| 5 2009 | 80 | 40 | 100 | 100 | 6 |
| 5 2010 | 100 | 40 | 100 | 100 | 5 |
| 6 2009 | 80 | 60 | 100 | 80 | 6 |
| 6 2010 | 80 | 60 | 100 | 80 | 6 |
| 7 2009 | 80 | 100 | 100 | 100 | 1 |
| 7 2010 | 100 | 80 | 100 | 100 | 3 |
| 8 2009 | 100 | 40 | 100 | 100 | 5 |
| 8 2010 | 80 | 60 | 100 | 100 | 3 |
| 11 2009 | 80 | 100 | 100 | 100 | 1 |
| 11 2010 | 80 | 100 | 100 | 100 | 1 |

Categorical Concurrence >6 items
 Depth-of-Knowledge >50% with DOK level the same or higher than level of corresponding Objectives
 Range-of-Knowledge >70% of objectives under a standard
 Balance of Representation A possible weakness if one or more objectives with a relative large number of items (e.g. five or more than the objective with the next highest number of items)

Alignment Analysis of Reading Standards and Assessments

South Dakota Grades 3-8 and 11 2009 and 2010

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 four- day alignment institute was held October 19 through October 22, 2009 in Sioux Falls, South Dakota to analyze the South Dakota 2009 Reading Standards and the assessments for two years, 2009 and 2010. One group of six reviewers participated in the institute. Three of the reviewers were from South Dakota and three were from other states. External reviewers from other states are likely to provide a more objective view of the standards and the assessments because they have no invested interest in either. The three external reviewers who participated in this analysis had participated in similar alignment studies for over 10 states. Their experiences gave them a more global perspective. The South Dakota reviewers also provided important knowledge for the process. The instate reviewers were more aware of the interpretation of the standards by South Dakota teachers and a classroom perspective. The balance between external reviewers and South Dakota reviewers provided a balance between a more objective view and important knowledge of practices from the state. The reviewers included reading education content experts, district reading supervisors, reading teachers, and reading assessment experts. The group analyzed assessments and standards for grades 3-8 and 11.

South Dakota used the terminology of *goal/strand*, *indicator*, and *standard*. Reading only had one goal/strand. The Reading Strand was delineated into five indicators:

- R.1 Recognize and analyze words;
- R.2 Comprehend and fluently read text;

- R.3 Apply knowledge of text structures, literary devices, and literary elements to develop interpretations and form responses;
- R.4 Interpret and respond to diverse, multicultural, and time period texts; and
- R.5 Access, analyze, synthesize, and evaluate informational texts.

Indicators described key aspects of the goals/strands. The standards under the indicators specified what students were to know and be able to do related to the indicator at the specific grade level.

As part of the alignment institute, reviewers were trained to identify the depth of knowledge (DOK) of the standards and assessment items. This training included reviewing the definitions of the four DOK levels and reviewing examples of each. Then, the reviewers participated in a consensus process to determine the DOK levels of the standards and individual analyses of the assessment items. Following individual analyses of the items, reviewers participated in a debriefing discussion in which they assessed the degree to which they had coded particular items or types of content to the standards.

To derive the results from the analysis, the reviewers' responses were averaged. Any variance among reviewers was considered legitimate, with the true DOK level for the item falling somewhere between the two or more assigned values. Such variation could signify a lack of clarity in how the standards were written, the robustness of an item that could legitimately correspond to more than one standard and/or a DOK that falls in between two of the four defined levels. Reviewers adjudicated their results after each grade level analysis. The adjudication process included the discussion of any results in assigning items to standards without a majority of reviewers in agreement. Reviewers were not required to change their results after the discussion. Thus, large variations among reviewers represented true differences in opinion among the reviewers and not because of coding error. Reviewers were allowed to identify one assessment item as corresponding to up to three standards—one primary hit (standard) and up to two secondary hits. However, reviewers could only code one DOK level to each assessment item, even if the item corresponded to more than one standard.

Reviewers were instructed to focus primarily on the alignment between the state's 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 also could 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 South Dakota Reading Content Standards and the assessments. Note that an alignment analysis of this nature 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. Standard deviations are reported in

the tables provided in Appendix B, which give one indication of the variance among reviewers.

This report describes the results of an alignment study of the South Dakota assessment administered in the spring 2009 and 2010 and reading content standards for grades 3-8 and 11. The study addressed specific criteria related to the content agreement between the state's 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 strand. The analysis assumed that the assessment had to have at least six items for measuring content from a strand in order for an acceptable level of categorical concurrence to exist between the strand 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 strands or require students to achieve a specified cutoff score on indicators related to a strand. 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 strand, and as a basis for making some decisions about students' knowledge of that strand. 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 strands, as judged in this analysis, at least 50% of the items corresponding to a strand had to be at or above the depth-of-knowledge level of the corresponding standards: 50%, a conservative cutoff point, is based on the assumption that a minimal passing score for any one strand 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 standards. For example, assume an assessment included six items related to one strand and students were required to answer correctly four of those items to be judged proficient—i.e., 67% of the items. If three, 50%, of the six items were at or above the depth-of-knowledge level of the corresponding standards, then for a student to achieve a proficient score would require the student to answer correctly at least one item at or above the depth-of-knowledge level of one standard. Some leeway was used in this analysis on this criterion. If a strand had between 40% and 50% of items at or above the depth-of-knowledge levels of the standards, then it was reported that the criterion was “weakly” met.

Interpreting and assigning depth-of-knowledge levels to both standards within strands and assessment items are essential requirements of alignment analysis. The reading levels are based on Valencia and Wixson (2000, pp. 909-935). These descriptions help to clarify what the different levels represent in reading:

:
Reading

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

- Support ideas by reference to details in the text.
- Use a dictionary to find the meanings of words.
- Identify figurative language in a reading passage.

Reading Level 2. Level 2 includes the engagement of some mental processing beyond recalling or reproducing a response; it requires both comprehension and subsequent processing of text or portions of text. Inter-sentence analysis or inference is

required. Some important concepts are covered but not in a complex way. Standards and items at this level may include words such as summarize, interpret, infer, classify, organize, collect, display, compare, and determine whether fact or opinion. Literal main ideas are stressed. A Level 2 assessment item may require students to apply skills and concepts that are covered in Level 1. Some examples that represent, but do not constitute all of, Level 2 performance are:

- Use context cues to identify the meaning of unfamiliar words.
- Predict a logical outcome based on information in a reading selection.
- Identify and summarize the major events in a narrative.

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

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

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

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

Range-of-Knowledge Correspondence

For standards and assessments to be aligned, the breadth of knowledge required on both should be comparable. *The range-of-knowledge criterion is used to judge whether a comparable span of knowledge expected of students by a strand 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 strand and an assessment considers the number of standards within the strand with one related assessment item/activity. Fifty percent of the standards for a

strand 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 strand. This assumes that each standard for a strand 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 standard, the requirement that assessment items need to be related to more than 50% of the standards for an strand increases the likelihood that students will have to demonstrate knowledge on more than one standard per strand 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 standards. However, any restriction on the number of items included on the test will place an upper limit on the number of standards that can be assessed. Range-of-knowledge correspondence is more difficult to attain if the content expectations are partitioned among a greater number of strands and a large number of standards. If 50% or more of the standards for a strand had a corresponding assessment item, then the range-of-knowledge correspondence criterion was met. If between 40% and 50% of the standards for a strand 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 standards within a strand hit (a standard with a corresponding item); it does not take into consideration how the hits (or assessment items/activities) are distributed among these standards. *The balance-of-representation criterion is used to indicate the degree to which one standard 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 standards for a strand that have at least one hit—i.e., one related assessment item per standard. The index is computed by considering the difference in the proportion of standards and the proportion of hits assigned to the standard. An index value of 1 signifies perfect balance and is obtained if the hits (corresponding items) related to a strand are equally distributed among the standards for the given strand. Index values that approach 0 signify that a large proportion of the hits are on only one or two of all of the standards hit. Depending on the number of standards and the number of hits, a unimodal distribution (most items related to one standard and only one item related to each of the remaining standards) 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 standards at least to some degree (e.g., every standard has at least two items) and is used as the acceptable level on this criterion. Index values between .6 and .7 indicate the balance-of-representation criterion has only been "weakly" met.

Source-of-Challenge Criterion

The source-of-challenge criterion is only used to identify items on which the major cognitive demand is inadvertently placed and is other than the targeted reading standard or expectation. 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 2009 reading standard for each grade 3-8 and 11 can be found in Appendix A. Table 1 shows the percentages of standards at each DOK level by grade. Reviewers found the distribution of the different DOK levels to vary some across the grades with a general pattern towards a larger proportion of DOK 3 standards with the higher grades. For grade 3, reviewers judged that most of the standards had a DOK 2 with two of the nine standards with a DOK 1 and one of the standards with a DOK 3. Nearly all of the standards for grades 4-8 and 11 were judged to have a DOK level 2 or 3. Standard 6.R.1.1 was the only exception and was assigned a DOK level 1. For grades 7, 8, and 11 at least half of the standards had a DOK level 3. Generally, the standards increased in the proportion of those with a DOK level 3 with an increase in grade.

Table 1

Percent of Standards by Depth-of-Knowledge (DOK) Levels for Grades 3-8 and 11 South Dakota Alignment Analysis for Reading

| Grade | Total Number of Standards | DOK Level | Number of Standards by Level | Percent within Strand by Level |
|-------|---------------------------|-----------|------------------------------|--------------------------------|
| 3 | 9 | 1 | 2 | 22 |
| | | 2 | 6 | 66 |
| | | 3 | 1 | 11 |
| 4 | 11 | 1 | 0 | 0 |
| | | 2 | 10 | 90 |
| | | 3 | 1 | 9 |
| 5 | 11 | 1 | 0 | 0 |
| | | 2 | 6 | 54 |
| | | 3 | 5 | 45 |
| 6 | 11 | 1 | 1 | 9 |
| | | 2 | 6 | 54 |
| | | 3 | 4 | 36 |
| 7 | 12 | 1 | 0 | 0 |
| | | 2 | 6 | 50 |
| | | 3 | 6 | 50 |

Table 1 (continued)

| Grade | Total Number of Standards | DOK Level | Number of Standards by Level | Percent within Strand by Level |
|-------|---------------------------|-----------|------------------------------|--------------------------------|
| 8 | 9 | 1 | 0 | 0 |
| | | 2 | 4 | 44 |
| | | 3 | 5 | 55 |
| 11 | 6 | 1 | 0 | 0 |
| | | 2 | 3 | 50 |
| | | 3 | 3 | 50 |

If no particular standard is targeted by a given assessment item, reviewers are instructed to code the item at the level of an indicator or strand. This coding to a generic standard 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 standards. These items may highlight areas in the standards that should be changed, or made more precise. There were no items coded to generic standards by more than one reviewer in this study. Reviewers were able to match each item to one of the given standards at each grade level. Thus, no item targeted content outside of the reading expectations as defined by South Dakota. Reviewers' debriefing comments also highlight some ambiguities in the standards. 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. There were no multi-point items on the assessments. The total number of points was the same as the total number of items. The assessment at each grade level for both years, 2009 and 2010, had the same number of items.

Table 3

Number of Items and Point Value by Grade for South Dakota 2009 and 2010 Assessments, Grades 3-8 and 11

| Grade Level | Number of Items 2009 | Number of Items 2010 | Number of Multi-Point Items | Total Point Value |
|-------------|----------------------|----------------------|-----------------------------|-------------------|
| 3 | 56 | 56 | 0 | 56 |
| 4 | 56 | 56 | 0 | 56 |
| 5 | 56 | 56 | 0 | 56 |
| 6 | 56 | 56 | 0 | 56 |
| 7 | 56 | 56 | 0 | 56 |
| 8 | 56 | 56 | 0 | 56 |
| 11 | 50 | 50 | 0 | 50 |

The results of the analysis for each of the four alignment criteria are summarized in Tables 4.1-4.14. 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 4.1-4.14, "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 3

The alignment between the grade 3 assessments, both for 2009 and 2010, and the 2009 reading standards was acceptable. The only alignment issue reviewers found was too few items that mapped to Indicator 3.R.4 (respond to ideas and attitudes expressed in multicultural and historical texts by making connections).

Reviewers found from eight to 18 items that mapped to each of the indicators, 3.R.1 (9-10 items), 3.R.2 (17-18 items), 3.R.3 (14-16 items), and 3.R.5 (8-9 items). The number of items was sufficient to have an acceptable level for the Categorical Concurrence criteria for each of these four indicators. Reviewers only found from three or four items that mapped to Indicator 3.R.4. This would be too few items to reliably judge the proficiency of students on Indicator 3.R.4. Reviewers noted that the assessments included multicultural passages, but that items related to these passages were judged to more appropriately target general comprehension (3.R.2) rather than to respond to ideas and attitudes expressed in the texts. One reviewer commented, "Although some passages represented various cultures, the corresponding items were easily coded to a more general comprehension standard."

The DOK levels of the items appropriately matched the DOK levels of the standards. The Depth-of-Knowledge Consistency criterion was acceptable for four of the five indicators, all except for Indicator 3.R.4. Reviewers found that most of the items had a DOK level 2 with some DOK level 3 items. Of the three items that mapped to Indicator 3.R.4, two of these items, on the average, were judged to have a DOK level 2 which did not match the expectation that required students to make connection among ideas and attitudes, a DOK level 3.

The Range-of-Knowledge Correspondence and Balance of Representation was acceptable for all five reading indicators. Reviewers found items that targeted each of the nine grade 3 standards except for Standard 3.R.2.2 (fluently read aloud and silently to comprehend text). This standard was more a performance standard that would be difficult to assess on an on-demand assessment. The items also were fairly evenly distributed

among the standard without any one standard having a large number of standards compared to other standards under the indicator.

Overall, the alignment between the grade 3 assessments for 2009 and 2010 and the standards was acceptable. Only three items would need to be replaced or added to attain full alignment. These items would all need to target clearly Standard 3.R.4.1 and have a DOK level 3. Reviewers only made comments on a few items (see Appendix C) and about the assessment (Appendix D). Two reviewers noted that they thought the passages for grade 3 were a little short and that this age group could have longer passages. Five of the six reviewers indicated that the alignment was acceptable concurring with the results from the analyses.

Table 4.1
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 3 Indicators and Assessment for South Dakota Alignment Analysis 2009

| Grade 3 | Alignment Criteria | | | |
|---|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Indicators | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| 3.R.1 - Recognize and analyze words | YES | YES | YES | YES |
| 3.R.2 - Comprehend and fluently read text | YES | YES | YES | YES |
| 3.R.3 - Apply knowledge of text structures | YES | YES | YES | YES |
| 3.R.4 - Interpret and respond to diverse texts | NO (3.83) | NO | YES | YES |
| 3.R.5 - Access, analyze, synthesize, and evaluate informational texts | YES | YES | YES | YES |

Table 4.2
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 3 Indicators and Assessment for South Dakota Alignment Analysis 2010

| Grade 3 | Alignment Criteria | | | |
|---|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Indicators | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| 3.R.1 - Recognize and analyze words | YES | YES | YES | YES |
| 3.R.2 - Comprehend and fluently read text | YES | YES | YES | YES |
| 3.R.3 - Apply knowledge of text structures | YES | YES | YES | YES |
| 3.R.4 - Interpret and respond to diverse texts | NO (3.67) | NO | YES | YES |
| 3.R.5 - Access, analyze, synthesize, and evaluate informational texts | YES | YES | YES | YES |

Grade 4

The alignment between the grade 4 assessments for 2009 and 2010 and the grade 4 2009 reading standards was acceptable. However, reviewers found more alignment issues at grade 4 than they did for grade 3. The Categorical Concurrence alignment criterion was acceptable for four of the five indicators for the 2009 assessment and for all five indicators for the 2010 assessment. The majority of reviewers only found four items on the 2009 grade 4 assessment that mapped to the standard under Indicator 4.R.4. Half of them agreed that a fifth item mapped to this indicator. One reviewer found other items that he or she judged targeted Indicator 4.R.4, but none of the other five reviewers agreed with this judgment. A majority of reviewers did agree that six items on the 2010 grade 4 assessment targeted Indicator 4.R.4. Thus, the categorical concurrence was acceptable for the 2010 grade 4 assessment and all five indicators.

The Depth-of-Knowledge Consistency criterion was acceptable for both grade 4 assessment forms and four of the five indicators. The majority of the reviewers judged that the items on both assessments that mapped to the Indicator 4.R.4 had a DOK level 2 (comprehension or processing of text) but not a DOK level 3 (drawing inferences or making connections among multiple texts). Nearly all of the items on both the grade 4 2009 and 2010 assessments were judged to have a DOK level 2. The level of complexity of these items matched favorably the expected level of complexity of all but one of the grade 4 standards. Ten of the 11 grade 4 standards were judged to have a DOK level 2. Standard 4.R.4.1 was assigned a DOK level 3.

Range was acceptable for both grade 4 assessments and the grade 4 standards except for Indicator 4.R.2. This indicator had three underlying standards. However, reviewers only found items that mapped to one of the three, 4.R.2.1. Reviewers did not find any items on either of the grade 4 assessments that they judged required students to use fluency strategies. One reviewer commented, “References to ‘fluency’ in the standards don’t lend themselves to assessment of this type, but emphasizing ‘comprehension’ and strategic reading work well.” The Balance of Representation was acceptable except for the 2009 assessment and Indicator 4.R.3. The majority of the reviewers found seven items on the 2009 assessment that mapped to Standard 4.R.3.1, but only one item that mapped to each of the other two standards under Indicator 4.R.3. This was similar to how the items from the 2010 grade 4 assessment mapped to Indicator 4.R.3, but the majority of the reviewers found a second item that they felt mapped to each of the other two standards. This balance issue is not considered a major alignment issue, but more a matter of preference since the other three alignment criteria were acceptable for Indicator 4.R.3.

Overall, the alignment for grade 4 between the two assessments and the standards was acceptable. Only four items would need to be replaced or added to attain full alignment for each of the two assessment forms. Three items that currently mapped to Standard 4.R.4.1 would need to be replaced by items that clearly have a DOK level 3. For the 2009 items, one of these items would need to be an additional item to have at least six items that mapped to Indicator 4.R.4. Another item that currently maps to Standard

4.R.2.1 would need to be replaced by an item that maps to one of the other two standards under Indicator 4.R.2 in order to have an acceptable level for range.

Table 4.3

Summary of Acceptable Levels on Alignment Criteria for Reading Grade 4 Indicators and Assessment for South Dakota Alignment Analysis 2009

| Grade 4 | Alignment Criteria | | | |
|---|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Indicators | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| 4.R.1 - Recognize and analyze words | YES | YES | YES | YES |
| 4.R.2 - Comprehend and fluently read text | YES | YES | NO | YES |
| 4.R.3 - Apply knowledge of text structures | YES | YES | YES | WEAK |
| 4.R.4 - Interpret and respond to diverse texts | NO (5.67) | NO | YES | YES |
| 4.R.5 - Access, analyze, synthesize, and evaluate informational texts | YES | YES | YES | YES |

Table 4.4

Summary of Acceptable Levels on Alignment Criteria for Reading Grade 4 Indicators and Assessment for South Dakota Alignment Analysis 2010

| Grade 4 | Alignment Criteria | | | |
|---|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Indicators | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| 4.R.1 - Recognize and analyze words | YES | YES | YES | YES |
| 4.R.2 - Comprehend and fluently read text | YES | YES | NO | YES |
| 4.R.3 - Apply knowledge of text structures | YES | YES | YES | YES |
| 4.R.4 - Interpret and respond to diverse texts | YES | NO | YES | YES |
| 4.R.5 - Access, analyze, synthesize, and evaluate informational texts | YES | YES | YES | YES |

Four of the reviewers for the 2009 assessment and five for the reviewers for the grade 4 2010 assessment thought that the alignment was acceptable. Reviewers thought the passages were more challenging for grade 4 than for grade 3 and were more interesting. One reviewer noted regarding the 2010 assessment, “The questions surrounding the use of informational texts and time period texts were excellent. Many of

them made students think at higher levels.” When asked if the grade 4 standards had an appropriate level of specificity, one reviewer felt the standards could be improved:

No, [standards were] just the same as the previous grade level. No specific objective that lists specific literary elements and literary devices that students are taught at this grade level. Missing "making connections" as had been specified in grade 3. No specific objective for main idea but one item asked about that specifically; main idea is usually taught at this grade level. Good to see some passages about South Dakota.

Grade 5

For grade 5, the alignment between the two assessments, 2009 and 2010, and the grade 5 standards needed slight improvement. Both assessments failed to reach an acceptable level for the Depth-of-Knowledge Consistency for three of the five indicators—5.R.1, 5.R.4, and 5.R.5. As for the two previous grades reviewers assigned a DOK level 1 or 2 to most of the assessment items. However, nearly half of the grade 5 standards had a DOK level 3.

The Categorical Concurrence criterion was acceptable for four of the five indicators for the 2009 assessment and for all of the 2010 assessment. The average number of items on the 2009 assessment reviewers assigned to Indicator 5.R.4 was only five, one fewer than needed to make a reasonably reliable decision about a student’s proficiency on an indicator. Reviewers still had some disagreement on which items targeted the standards under Indicator 5.R.4. On the 2010 assessment, at least five of the six reviewers agreed that six of the items targeted Indicator 5.R.4. Thus, the average number of hits for the 2010 assessment was over six for Indicator 5.R.4. For the other four indicators, the average number of items ranged from nine (5.R.1) to 18 (5.R.3).

The main alignment issue for grade 5 was with the Depth-of-Knowledge Consistency. This criterion was acceptable for Indicators 5.R.2 and 5.R.3, but was not acceptable for each of the assessments for the other three indicators. About five of the nine items on each assessment that were judged to target Indicator 5.R.1, were judged to have a DOK level 1 (recall or verbatim skill). Most of these items corresponded to Standard 5.R.1.1 which had a DOK level 2. The objectives under Indicators 5.R.4 and 5.R.5 were all assigned a DOK level 3. Most of the reviewers agreed that maybe one of the items on each assessment that targeted the standard under Indicator 5.R.4 and three or four of the items on each assessment that targeted Indicator 5.R.5 was a DOK level 3. However, on the average, for both Indicators 5.R.4 and 5.R.5 and both assessments, 40% or fewer of the assigned items had a DOK level that was the same or higher than the DOK level of the corresponding objective.

Range and balance was acceptable for all five of the indicators for both of the assessments. The items were adequately distributed among the objectives underlying each of the indicators.

Table 4.5

Summary of Acceptable Levels on Alignment Criteria for Reading Grade 5 Indicators and Assessment for South Dakota Alignment Analysis 2009

| Grade 5 | Alignment Criteria | | | |
|---|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Indicators | <i>Categorical Concurrency</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| 5.R.1 - Recognize and analyze words | YES | WEAK | YES | YES |
| 5.R.2 - Comprehend and fluently read text | YES | YES | YES | YES |
| 5.R.3 - Apply knowledge of text structures | YES | YES | YES | YES |
| 5.R.4 - Interpret and respond to diverse texts | NO (5.33) | NO | YES | YES |
| 5.R.5 - Access, analyze, synthesize, and evaluate informational texts | YES | NO | YES | YES |

Table 4.6

Summary of Acceptable Levels on Alignment Criteria for Reading Grade 5 Indicators and Assessment for South Dakota Alignment Analysis 2010

| Grade 5 | Alignment Criteria | | | |
|---|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Indicators | <i>Categorical Concurrency</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| 5.R.1 - Recognize and analyze words | YES | WEAK | YES | YES |
| 5.R.2 - Comprehend and fluently read text | YES | YES | YES | YES |
| 5.R.3 - Apply knowledge of text structures | YES | YES | YES | YES |
| 5.R.4 - Interpret and respond to diverse texts | YES | NO | YES | YES |
| 5.R.5 - Access, analyze, synthesize, and evaluate informational texts | YES | WEAK | YES | YES |

Overall, the alignment between the grade 5 assessments and the grade 5 standards for reading needed slight improvement. Six items would need to be replaced or added to attain full alignment. One item that targeted Standard 5.R.1.1 on each assessment would need to be replaced by an item with at least a DOK level 2. On the 2009 assessment, one additional item is needed for Indicator 5.R.4. This item would need to have a DOK level 3 along with replacing two of the existing items with DOK level 3 items. Also three items that target Indicator 5.R.5 would need to be replaced by items that clearly have a DOK level 3. For the 2010 assessment, three items that currently target Indicator 5.R.4 and two items that currently target Indicator 5.R.5 need to be replaced by items that have a DOK

3. There was strong agreement among reviewers that 2009 items 18, 61, and 62 and 2010 items 18, 59, 60, and 61 had a level of complexity that reached a DOK level 3. Reviewers only made comments on one or two grade 5 items (Appendix C). In the judgment of the reviewers, five felt that alignment was acceptable. In general, reviewers complimented the grade 5 passages and the distribution of items.

Grade 6

The alignment between the 2009 and 2010 reading assessments and the reading standards needed slight improvement. The main alignment issue for grade 6 and for both assessment forms was an insufficient number of items that clearly targeted Indicator 6.R.4 (respond to ideas and attitudes expressed in multicultural and historical texts by making connections).

Both the 2009 and the 2010 grade 6 assessment forms had from nine to 21 items for each of four indicators—6.R.1 (9 items), 6.R.2 (10 or 12 items), 6.R.3 (21 or 20 items), and 6.R.5 (11 or 12 items). The number of items for each of these indicators was judged to be sufficient to make a reliable judgment on a student’s proficiency for each indicator. However, for Indicator 6.R.4 all reviewers could only agree that one item on each of the two assessment forms mapped to this indicator. Some of the reviewers (four or five) judged that a second item also mapped to Indicator 6.R.4. Thus, the grade 6 assessments had an unacceptable level for the Categorical Concurrence criterion for this one indicator.

The Depth-of-Knowledge Consistency criterion was acceptable for three of the five indicators (6.R.1, 6.R.2, and 6.R.3). On both assessment forms, only one of the two items for Indicator 6.R.4 and only about five of the 12 items that mapped to Indicator 6.R.5 had items that had a DOK level that was the at least the same as the DOK level of the assigned standard.

Range was acceptable for all five reading indicators for both assessment forms. Balance was acceptable for four of the five indicators. The analysis did reveal a balance weakness for Indicator 6.R.3. On each form, reviewers found about 15 items that targeted Standard 6.R.3.2 and only two to five items that targeted the other two standards under Indicator 6.R.3. Since the other alignment criteria were acceptable, an overemphasis of one standard over another is not considered a major fault of the assessments, but one of preference.

Overall, the alignment at grade 6 between the two assessment forms and the reading standards needed slight improvement. At least six items would need to be replaced or added to each assessment to attain full alignment. Four additional items on each assessment form are needed to improve the Categorical Concurrence for Indicator 6.R.4. If each of these items is at a DOK level 3, then the DOK consistency would be acceptable for Indicator 6.R.4. Two items that currently target Indicator 6.R.5 need to be replaced by items that are clearly at a DOK level 3. One reviewer wrote a note for two of the items on the 2009 grade 6 assessment form (Appendix C). Four of the reviewers felt

that the alignment for grade 6 was acceptable. In general, their comments complemented the passages and the assessments. One reviewer did note some issues with some of the standards when trying to distinguish among them:

These standards are better, especially 6.R.3, which specify text structure, literary elements, and devices such as discrete objectives. Great! Consider adding these to other grade levels. The difference between general comprehension strategies 2.1 and literary elements 3.2 continues to be problematic for coding. Objective statements do not help reviewers distinguish between these two objectives.

Table 4.7

Summary of Acceptable Levels on Alignment Criteria for Reading Grade 6 Indicators and Assessment for South Dakota Alignment Analysis 2009

| Grade 6 | Alignment Criteria | | | |
|---|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Indicators | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| 6.R.1 - Recognize and analyze words | YES | YES | YES | YES |
| 6.R.2 - Comprehend and fluently read text | YES | YES | YES | YES |
| 6.R.3 - Apply knowledge of text structures | YES | YES | YES | WEAK |
| 6.R.4 - Interpret and respond to diverse texts | NO (2.67) | NO | YES | YES |
| 6.R.5 - Access, analyze, synthesize, and evaluate informational texts | YES | WEAK | YES | YES |

Table 4.8

Summary of Acceptable Levels on Alignment Criteria for Reading Grade 6 Indicators and Assessment for South Dakota Alignment Analysis 2010

| Grade 6 | Alignment Criteria | | | |
|---|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Indicators | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| 6.R.1 - Recognize and analyze words | YES | YES | YES | YES |
| 6.R.2 - Comprehend and fluently read text | YES | YES | YES | YES |
| 6.R.3 - Apply knowledge of text structures | YES | YES | YES | WEAK |
| 6.R.4 - Interpret and respond to diverse texts | NO (2.0) | NO | YES | YES |
| 6.R.5 - Access, analyze, synthesize, and evaluate informational texts | YES | WEAK | YES | YES |

Grade 7

The alignment for grade 7 between the two reading assessment forms (2009 and 2010) and the reading standards was acceptable. The 2009 assessment form only had an issue with Categorical Concurrence and the 2010 assessment form only had an issue with the DOK levels of items for one indicator.

The 2009 assessment form had from seven to 14 items for four of the five indicators, all except for Indicator 7.R.2. The majority of the reviewers only could agree that four items on the 2009 assessment targeted standards under Indicator 7.R.2. For the 2010 assessment, reviewers found at least six items for each of the five indicators.

The Depth-of-Knowledge Consistency was acceptable for all five indicators and the 2009 assessment, but was only acceptable for four of the five indicators and the 2010 assessment. Only about one third of the 11 items on the 2010 assessment that mapped to Indicator 7.R.5 had a DOK that was comparable to the DOK of the assigned standard. Range and balance was acceptable for all five indicators and both assessment forms.

Table 4.9

Summary of Acceptable Levels on Alignment Criteria for Reading Grade 7 Indicators and Assessment for South Dakota Alignment Analysis 2009

| Grade 7 | Alignment Criteria | | | |
|---|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Indicators | <i>Categorical Concurrence</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| 7.R.1 - Recognize and analyze words | YES | YES | YES | YES |
| 7.R.2 - Comprehend and fluently read text | NO (1.29) | YES | YES | YES |
| 7.R.3 - Apply knowledge of text structures | YES | YES | YES | YES |
| 7.R.4 - Interpret and respond to diverse texts | YES | YES | YES | YES |
| 7.R.5 - Access, analyze, synthesize, and evaluate informational texts | YES | YES | YES | YES |

Overall, the alignment for both grade 7 assessment forms and the five grade 7 reading indicators was acceptable. Only one item on the 2009 assessment and three items on the 2010 assessment would need to be added or replaced to attain full alignment. At least one more item is needed for the 2009 assessment that clearly targets a standard under Indicator 7.R.2 and at least three items that currently map to Indicator 7.R.5 on the 2010 assessment needs to be replaced by items that have a DOK that matches the DOK level of the assigned standard such as Items 6 and 14. Six of the reviewers thought the alignment was acceptable with the 2009 assessment and five of the reviewers thought the alignment was acceptable with the 2010 assessment. Reviewers indicated that they had

some problem distinguishing between Standard 7.R.2.1 and Standard 7.R.5.2. One reviewer did note some confusion in coding items regarding author's purpose:

One challenge to the coding was that the reference to "author's purpose" appears only in the standard related to "informational texts." Some items relating to author's purpose in literary texts were confusing to code. Passages presented a variety of genres and historical settings, with considerable emphasis on informational text.

Table 4.10

Summary of Acceptable Levels on Alignment Criteria for Reading Grade 7 Indicators and Assessment for South Dakota Alignment Analysis 2010

| <i>Grade 7</i> <i>Indicators</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> |
| 7.R.1 - Recognize and analyze words | YES | YES | YES | YES |
| 7.R.2 - Comprehend and fluently read text | YES | YES | YES | YES |
| 7.R.3 - Apply knowledge of text structures | YES | YES | YES | YES |
| 7.R.4 - Interpret and respond to diverse texts | YES | YES | YES | YES |
| 7.R.5 - Access, analyze, synthesize, and evaluate informational texts | YES | NO | YES | YES |

Grade 8

The results from the analysis of the alignment between the grade 8 2009 assessment and the grade 8 2009 reading standards indicated that the alignment needed slight improvement. The alignment with the grade 8 2010 assessment, however, was acceptable. For both assessment forms, the main alignment issue was in attaining an acceptable level for Depth-of-Knowledge Consistency.

Both the 2009 and the 2010 had a sufficient number of items, six or more, for each of the indicators to have an acceptable level for the Categorical Concurrence criterion except for Indicator 8.R.4 and the 2010 assessment. Reviewers only found five items, on the average, on the 2010 assessment that mapped to the one standard under Indicator 8.R.4. On the 2009 assessment reviewers found, on the average, nearly eight items that mapped to this indicator. For the other four indicators, from nine to 17 items were mapped to each indicator.

The DOK level of the items that mapped to Indicators 8.R.1 and 2 was acceptable, but there was some consistency weakness for the other three indicators. This was mainly due to an insufficient number of items at a DOK level 3. For the 2009 assessment, the Depth-of-Knowledge consistency was only weakly met for Indicators 8.R.3 and 5, and

not acceptable for Indicator 8.R.4. The 2010 assessment had a DOK weakness for two indicators, 8.R.3 and 8.R.4. The range and balance was acceptable for all five indicators and for both assessments.

Overall, the alignment between the 2009 assessment and the grade 8 reading standards needed slight improvement and the alignment between the 2010 assessment and the grade 8 reading standards was acceptable. Five items for the 2009 assessment and three items for the 2010 assessment would need to be replaced or added to attain full alignment. For the 2009 assessment, the number of items that would need to be replaced was two for Indicator 8.R.3, two for Indicator 8.R.4, and one for Indicator 8.R.5. For the 2010 assessment one more item is needed that would target Indicator 8.R.4. If this item had an appropriate DOK level then this would also resolve the DOK weakness. Two items that currently map to Indicator 8.R.3 would need to be replaced by items with a DOK level 3.

Reviewers only made notes about one or two grade 8 items (Appendix C). In their debriefing comments (Appendix D) for grade 8, reviewers indicated that the items seemed to be evenly distributed among the standards. They noted that there could have been more items with a DOK level 3. This coincides with the results from the analyses. Five of the reviewers thought that the alignment was acceptable. One reviewer did comment on one perceived deficiency in the grade 8 standards:

As compared to previous standards, one noticeable lack in the Grade 8 standards is a reference to literary genres. No items referred specifically to this topic, so coding was not impaired. This lack may be a concern regarding curriculum and instruction on this grade level. It would seem logical that emphasis on genre in previous grades would be enhanced and extended for grade 8 as students become more able to think conceptually.

Table 4.11
Summary of Acceptable Levels on Alignment Criteria for Reading Grade 8 Indicators and Assessment for South Dakota Alignment Analysis 2009

| <i>Grade 8 Indicators</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> |
| 8.R.1 - Recognize and analyze words | YES | YES | YES | YES |
| 8.R.2 - Comprehend and fluently read text | YES | YES | YES | YES |
| 8.R.3 - Apply knowledge of text structures | YES | WEAK | YES | YES |
| 8.R.4 - Interpret and respond to diverse texts | YES | NO | YES | YES |
| 8.R.5 - Access, analyze, synthesize, and evaluate informational texts | YES | WEAK | YES | YES |

Table 4.12

Summary of Acceptable Levels on Alignment Criteria for Reading Grade 8 Indicators and Assessment for South Dakota Alignment Analysis 2010

| <i>Grade 8</i> <i>Indicators</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> |
| 8.R.1 - Recognize and analyze words | YES | YES | YES | YES |
| 8.R.2 - Comprehend and fluently read text | YES | YES | YES | YES |
| 8.R.3 - Apply knowledge of text structures | YES | WEAK | YES | YES |
| 8.R.4 - Interpret and respond to diverse texts | NO (5.0) | WEAK | YES | YES |
| 8.R.5 - Access, analyze, synthesize, and evaluate informational texts | YES | YES | YES | YES |

Grade 11

The alignment between the grade 11 reading assessments for 2009 and 2010 and the grade 11 standards was acceptable. At this grade level there was only a minor alignment issue of not enough items on each assessment for one indicator. The majority of reviewers only found five items on the 2009 assessment that mapped to Indicator 11.R.5 and only five items on the 2010 assessment that mapped to Indicator 11.R.1. On each of the assessments, reviewers found from seven to 14 items that targeted the other four indicators. Thus, each assessment was only one item short from meeting an acceptable level for the Categorical Concurrence criterion for all five standards.

The Depth-of-Knowledge Consistency criterion was acceptable for all five of the indicators for each of the two assessments. Over 55% of the items that targeted each of the indicators had a DOK level that was at least as high as the DOK level of the assigned standard. Range and balance also were acceptable for all five indicators and each of the two assessments. Range was not as difficult to satisfy because all but one of the indicators only had one underlying standard.

Overall, the alignment for grade 11 between the two assessments and the reading standards was acceptable. Only one item would need to be replaced or added for each assessment to attain full alignment. One additional item would be needed for Indicator 11.R.5 for the 2009 assessment and one additional item would be needed for Indicator 11.R.1 for the 2010 assessment. One or two reviewers made comments about five items on the 2009 grade 11 assessment (Appendix C).

Table 4.13

Summary of Acceptable Levels on Alignment Criteria for Reading Grade 11 Indicators and Assessment for South Dakota Alignment Analysis 2009

| Grade 11 | Alignment Criteria | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Indicators | <i>Categorical Concurrency</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| 11.R.1 - Recognize and analyze words | YES | YES | YES | YES |
| 11.R.2 - Comprehend and fluently read text | YES | YES | YES | YES |
| 11.R.3 - Apply knowledge of text structures | YES | YES | YES | YES |
| 11.R.4 - Interpret and respond to diverse texts | YES | YES | YES | YES |
| 11.R.5 - Access, analyze, synthesize, and evaluate informational texts | NO (5.67) | YES | YES | YES |

Table 4.14

Summary of Acceptable Levels on Alignment Criteria for Reading Grade 11 Indicators and Assessment for South Dakota Alignment Analysis 2010

| Grade 11 | Alignment Criteria | | | |
|--|--------------------------------|---------------------------------------|---------------------------|----------------------------------|
| Indicators | <i>Categorical Concurrency</i> | <i>Depth-of-Knowledge Consistency</i> | <i>Range of Knowledge</i> | <i>Balance of Representation</i> |
| 11.R.1 - Recognize and analyze words | NO (5.0) | YES | YES | YES |
| 11.R.2 - Comprehend and fluently read text | YES | YES | YES | YES |
| 11.R.3 - Apply knowledge of text structures | YES | YES | YES | YES |
| 11.R.4 - Interpret and respond to diverse texts | YES | YES | YES | YES |
| 11.R.5 - Access, analyze, synthesize, and evaluate informational texts | YES | YES | YES | YES |

Reviewers had some difficulty in assigning grade 11 items to standards because of the spread in content among the standards. One reviewer explained their frustration:

A number of topics have disappeared from previous grade-level standards, making coding challenging. There were no standards for literary elements, reading comprehension strategies, or discourse types such as persuasion, exposition, etc. The options for coding literary and informational items were severely limited. Because "fluency" can't be

tested in this format, that standard was used as the general "comprehension" standard for coding.

Another reviewer had a similar comment about the 2010 assessment:

[Standards were] quite clear, but with numerous omissions of material usually stressed at this level. Sometimes one has to stretch to make the item fit one of the available objectives when coding. There could be more exact matches if there were not so few objectives.

Five of the six reviewers indicated that the alignment for grade 11 was acceptable. They felt that the rigor of the items were higher than normally seen for grade 11 with most of the items judged to have a DOK level 3.

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. Two or more reviewers did not comment on a source-of-challenge issue for any one item. A reviewer did make a comment on an occasional item. After coding each grade-level assessment, reviewers also were asked to respond to five debriefing questions. All of the comments made by the reviewers are given in Appendices D. The notes in general offer an opinion on the item or give an explanation of the reviewers' coding.

Reliability Among Reviewers

The overall intraclass correlation among the reading reviewers' assignment of DOK levels to items was reasonable for six reviewers for Grades 3-8 and 11 (Table 5). An intraclass correlation value greater than 0.8 generally indicates a high level of agreement among the reviewers. The intraclass correlation for all 14 analyses were higher than 0.80. A pairwise comparison was used to determine the degree of reliability of reviewer coding at the standard level and at the indicator level. The pairwise comparison was computed by considering for each item the coding assigned by each reviewer compared to the coding by each of the other five reviewers. With six reviewers a total of 15 comparisons were computed for each item. For 56 items a total of 336 comparison were computed as agree or disagree. The reported pairwise comparison was the total of those with exact agreement divided by 336. The pairwise standard agreements were all above 0.70 which is higher than for most alignment studies. The pairwise indicator agreement ranged between 0.76 and 0.88. Normally the pairwise agreement at the indicator level is 0.90. The lower agreement among reviewers in assigning items to indicators reflects, in part, how the indicators were constructed. Reviewers' comments noted that some reviewers had a difficult time distinguishing between R.4 (multicultural) and the comprehension indicator R.2.

Table 5

Intraclass and Pairwise Comparisons, South Dakota Alignment Analysis for Reading Grades 3-8 and 11 Assessments 2009 and 2010

| Grade | Intraclass Correlation | Pairwise Comparison: | Pairwise: Standard | Pairwise: Indicator |
|---------|------------------------|----------------------|--------------------|---------------------|
| 3 2009 | 0.90 | 0.66 | 0.76 | 0.77 |
| 3 2010 | 0.90 | 0.74 | 0.76 | 0.76 |
| 4 2009 | 0.91 | 0.83 | 0.77 | 0.81 |
| 4 2010 | 0.90 | 0.80 | 0.72 | 0.78 |
| 5 2009 | 0.86 | 0.75 | 0.80 | 0.85 |
| 5 2010 | 0.88 | 0.73 | 0.79 | 0.82 |
| 6 2009 | 0.84 | 0.78 | 0.78 | 0.84 |
| 6 2010 | 0.84 | 0.75 | 0.74 | 0.79 |
| 7 2009 | 0.88 | 0.74 | 0.83 | 0.88 |
| 7 2010 | 0.86 | 0.73 | 0.78 | 0.86 |
| 8 2009 | 0.84 | 0.69 | 0.76 | 0.81 |
| 8 2010 | 0.90 | 0.75 | 0.78 | 0.84 |
| 11 2009 | 0.88 | 0.77 | 0.79 | 0.79 |
| 11 2010 | 0.86 | 0.73 | 0.77 | 0.77 |

Summary

A four- day alignment institute was held October 19 through October 22, 2009 in Sioux Falls, South Dakota to analyze the South Dakota 2009 Reading Standards and the assessments for two years, 2009 and 2010. One group of six reviewers participated in the institute. Three of the reviewers were from South Dakota and three were from other states. The reviewers included reading education content experts, district reading supervisors, reading teachers, and reading assessment experts. The group analyzed assessments and standards for grades 3-8 and 11.

The alignment between the South Dakota 2009 reading indicators and the 2009 and 2010 assessments were found to be acceptable or in need of slight improvement with some variation by grade. The alignment for both years was acceptable for grades 3, 4, 7, and 11. The alignment needed some improvement for grades 5 and 6. The grade 8 2009 assessment and the standards needed slight improvement while the grade 8 2010 assessment was considered as acceptable.

At least four indicators for each assessment satisfied an acceptable level for the Categorical Concurrence criterion of six or more corresponding items. However, the majority of reviewers found fewer than six items that targeted one indicator for 10 of the 14 assessments. For most of these assessments the reviewers only found four or five items that corresponded to Indicator R.4 (interpret and respond to diverse, multicultural, and time period texts). Reviewers indicated that they had difficulty deciphering between items that responded to multicultural text and those that assessed general comprehension (Indicator R.2).

The other main alignment issue was with Depth-of-Knowledge Consistency. Reviewers judged that the majority of the items on each assessment had a DOK level 2 (e.g. contextual clues and comprehension of text). However, the proportion of standards judged to have a DOK level 3 (e.g. making inferences) increased over the grades from about 9% (grade 4) to 55% (grade 8). There was some increase in the proportion of items with a DOK level 3 in the higher grades (particularly for grade 11), but for most grades the proportion of items with a DOK level 3 was not sufficient to have at least 50% of the items with a DOK that was the same or higher than the DOK level of the assigned standard. The level of complexity for the items was low mainly for Indicators R.4 and R.5.

Range and balance was acceptable for nearly all of the indicators for all of the 14 assessments. The items were adequately distributed among the standards under an indicator without over emphasizing any one of the standards. An adequate range and balance was fairly easily attained because of the low number of standards that was under each indicator.

Overall, nine of the 14 assessments were judged to have acceptable alignment with the reading indicators. For these assessments fewer than five items would need to be replaced or added to attain full alignment. Five of the assessments and reading indicators were judged to need slight improvement in alignment. For these assessments five or six items would need to be replaced or added to attain full alignment.

Summary Table

Percent of South Dakota Reading 2009 Standards with Acceptable Level on Each Alignment Criteria for Grade 3-8 and 11 for the 2009 and 2010 Assessments

| Grade | <i>Categorical Concurrence</i> (six or more items) | <i>Depth-of-Knowledge Consistency</i> (50% at/above) | <i>Range of Knowledge</i> (50% of objectives) | <i>Balance of Representation</i> (without possible weakness) | <i>Estimated Range of Items per to be Added or Replaced for Full Alignment</i> |
|---------|---|---|--|---|--|
| 3 2009 | 80 | 80 | 100 | 100 | 3 |
| 3 2010 | 80 | 80 | 100 | 100 | 3 |
| 4 2009 | 80 | 80 | 80 | 80 | 4 |
| 4 2010 | 100 | 80 | 80 | 100 | 4 |
| 5 2009 | 80 | 40 | 100 | 100 | 6 |
| 5 2010 | 100 | 40 | 100 | 100 | 5 |
| 6 2009 | 80 | 60 | 100 | 80 | 6 |
| 6 2010 | 80 | 60 | 100 | 80 | 6 |
| 7 2009 | 80 | 100 | 100 | 100 | 1 |
| 7 2010 | 100 | 80 | 100 | 100 | 3 |
| 8 2009 | 100 | 40 | 100 | 100 | 5 |
| 8 2010 | 80 | 60 | 100 | 100 | 3 |
| 11 2009 | 80 | 100 | 100 | 100 | 1 |
| 11 2010 | 80 | 100 | 100 | 100 | 1 |

| | |
|---------------------------|--|
| Categorical Concurrence | >6 items |
| Depth-of-Knowledge | >50% with DOK level the same or higher than level of corresponding Objectives |
| Range-of-Knowledge | >70% of objectives under a standard |
| Balance of Representation | A possible weakness if one or more objectives with a relative large number of items (e.g. five or more than the objective with the next highest number of items) |

References

- Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement*, 25(1), 47-55.
- Valencia, S. W., & Wixson, K. K. (2000). Policy-oriented research on literary standards and assessment. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Vol. III*. Mahwah, NJ: Lawrence Erlbaum.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and Mathematics 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.