

# ***Independent Alignment Review of the Reading Minnesota Comprehensive Assessment - Modified (MCA-MOD)***

***Yvette M. Nemeth  
Arthur Thacker  
Richard Deatz  
Wade Buckland  
Susan Fry  
Michele Hardoin  
Caroline Wiley***

Prepared for: Minnesota Department of Education  
1500 Highway 36 West  
Roseville, MN 55113

December 22, 2011

***HumRRO***  
Human Resources Research Organization

66 Canal Center Plaza, Suite 700 • Alexandria, Virginia 22314  
www.humrro.org • Phone: (703) 549-3611 - Fax: (703) 519-9661

# ***Independent Alignment Review of the Reading Minnesota Comprehensive Assessment - Modified (MCA-MOD)***

***Yvette M. Nemeth  
Arthur Thacker  
Richard Deatz  
Wade Buckland  
Susan Fry  
Michele Hardoin  
Caroline Wiley***

Prepared for: Minnesota Department of Education  
1500 Highway 36 West  
Roseville, MN 55113

December 22, 2011

## **Executive Summary**

### ***Scope of Work***

The Minnesota Department of Education (MDE) requested an external independent alignment study (review and analysis) of the Reading Minnesota Comprehensive Assessment - Modified (MCA-MOD) in grades 5 – 8 and 10. Specifically, MDE wanted an evaluation of the alignment of the MCA-MOD for grades 5 – 8 and 10 to the Minnesota Academic Standards<sup>1</sup>. Minnesota uses the Reading MCA-MOD in the federal and state accountability programs. The Human Resources Research Organization (HumRRO) was contracted to conduct this alignment study.

MDE requested the alignment study in order to meet both state and federal requirements. The federal requirements of the U.S. Department of Education (USDE) stem from the No Child Left Behind (NCLB) Act of 2001. NCLB challenges each state to establish a coherent assessment system based on solid academic standards. This law calls for states to provide independent evidence of the validity of their assessments used to calculate Adequate Yearly Progress (AYP). All states receiving Title I funds must present evidence of establishing a fair and consistent assessment system that is based on rigorous standards, sufficient alignment between standards and assessments, and high-quality educational results.

An alignment review can provide one form of evidence supporting the validity of the state assessment system. Alignment results should demonstrate that the assessments represent the full range of the content standards and that the assessments measure student knowledge in the same manner and at the same level of complexity as specified in the content standards. All aspects of the state assessment system must coincide, including the academic content standards, achievement standards (linked to cut scores), performance level descriptors, and each assessment.

### ***Methodology***

Two different types of alignment evaluations were performed for this Minnesota study. These evaluations involved a comparison of the 2011 Reading MCA-MOD to the Minnesota Academic Standards. The content alignment evaluation involved a review by current and recently retired Minnesota educators highly familiar with the content standards and the assessment.

### ***Review of Content Alignment and Accessibility***

For the content alignment review, HumRRO convened panels of Minnesota educators to review the grades 5 – 8 and 10 Reading MCA-MOD. The review involved two major tasks: (a) matching the Reading items to grade span Minnesota Academic Standards for Reading, and (b) evaluating test quality with respect to students who take the test.

---

<sup>1</sup> Minnesota Academic Standards can be found at  
<http://education.state.mn.us/MDE/EdExc/StanCurri/K-12AcademicStandards/>

HumRRO developed three review panels with the assistance of MDE and Pearson, Minnesota’s testing contractor during the administration of the assessment. Panelists were recruited by Pearson from their database of Minnesota educators. Every effort was made to produce panels consisting of teachers reflecting the population of students who take the assessments. Panels were convened in facilities procured by Pearson. HumRRO directed the actual reviews independently of MDE and Pearson. Each panel included 4–5 reviewers.

To conduct the content alignment review, HumRRO applied the Webb (2005) alignment method. Dr. Norman Webb developed a procedure to evaluate alignment of the assessment to the content standards using four statistics. These statistics indicate how well an assessment covers the content standards in terms of content breadth and depth. The alignment indicators include the following:

- Categorical concurrence – determines the degree of overall content coverage by the assessment for each content strand.
- Range-of-knowledge representation – indicates the specific content expectations (e.g., standard, benchmark) assessed within each strand.
- Balance-of-knowledge representation – provides a statistical index reflecting the distribution of assessed content within each strand (i.e., how evenly the content is assessed.)
- Depth-of-knowledge consistency – compares the cognitive complexity ratings of the items with the complexity ratings of each content standard.

The content reviews also involved a broad examination of test quality that went beyond content alignment. Other facets of test validity are critical as well, such as whether the assessment enables students to demonstrate what they know. For example, are test items free of biases, clear in language, or appropriate for the grade level? Evaluating these aspects of the assessments ensures that the test items are appropriate and accessible to “the widest possible range of students, including students with disabilities and students with limited English proficiency” (NCLB, 2001, Section 200.2(b)(2)). To examine test quality, panelists evaluated the Reading MCA-MOD on several dimensions at the item level and across each grade as a whole.

All assessments should “be designed from the beginning to be accessible and valid with respect to the widest possible range of students, including students with disabilities and students with limited English proficiency” (NCLB, 2001, Section 200.2(b)(2)). The Reading MCA-MOD underwent bias reviews as part of the item development process; however, review of quality and accessibility from an independent evaluator provides further confirmation of a fair process and assessment.

## **Summary of Results**

### ***Key Findings and Conclusions***

The results of the alignment and quality reviews provide support overall for the content validity of the Reading MCA-MOD for each grade (5 – 8 and 10) based on several outcomes. First, panelists found that the test items assessed the majority of the targeted content strands. Of those benchmarks within the strands, items were distributed rather evenly across these content

expectations. However, some issues were identified related to the level of cognitive complexity assessed by the items. Second, panelists considered whether the majority of test items provided reasonable access to the population of students who take the assessments. Most items were judged appropriate, clear in language, and free from bias.

***Alignment of Reading MCA-MOD to Minnesota Academic Standards***

Table 1 provides summary conclusions on the alignment of the Reading MCA-MOD to the Minnesota Academic Standards per grade tested. The conclusions are based on the following decision criteria (Webb, 2005):

- Fully aligned – assessments align to all content strands (91%–100%);
- Highly aligned – assessments align to the majority of strands (70%–90%);
- Partially aligned – assessments align well to some strands (50%–69%);
- Weakly aligned – assessments align to less than half the strands (below 50%).

Webb’s alignment method does not allow for a *single* judgment of overall alignment across the four alignment indicators. However, one can get a sense of overall alignment between the assessments and standards by looking at all of the alignment indicators together.

***Table 1. Summary Alignment Outcomes on Each Webb Criterion by Grade Level for Reading MCA-MOD***

| Grade Assessment | Percentage of Strands that Met Webb Criteria |                                |                                   |                                     |
|------------------|--|--------------------------------|-----------------------------------|-------------------------------------|
|                  | Categorical Concurrence                      | Depth-of-Knowledge Consistency | Range-of-Knowledge Correspondence | Balance-of-Knowledge Representation |
| 5                | Partially aligned (67%)                      | Partially aligned (67%)        | Partially aligned (67%)           | Partially aligned (67%)             |
| 6                | Fully aligned (100%)                         | Fully aligned (100%)           | Fully aligned (100%)              | Fully aligned (100%)                |
| 7                | Fully aligned (100%)                         | Partially aligned (67%)        | Fully aligned (100%)              | Fully aligned (100%)                |
| 8                | Partially aligned (67%)                      | Partially aligned (67%)        | Partially aligned (67%)           | Partially aligned (67%)             |
| 10               | Partially aligned (67%)                      | Partially aligned (67%)        | Partially aligned (67%)           | Partially aligned (67%)             |

Overall, panelists’ findings showed that the MCA-MOD Reading grade 6 assessment was aligned across all of Webb’s criteria. The Reading grade 7 assessment was also aligned to all of Webb’s criteria except for depth of knowledge. Across grades 5, 8, and 10, there were issues that appeared: (a) the number of items to assess one of the sub-strand content areas; (b) the level of cognitive complexity assessed by the items; (c) the number of benchmarks within a content area that are assessed by items; and (d) the distribution of items linked to each benchmark within a content area.

***Quality of Reading MCA-MOD Tests***

Table 2 presents the summary outcomes on the item quality ratings. The table includes conclusions regarding the quality of the items on each assessment, along with the percentage of items that received favorable ratings. The conclusions are based on the following decision criteria (adapted from Thompson, Johnstone, Anderson, & Miller, 2005):

- Excellent – all items are acceptable;
- Good – most items are acceptable (at least 90%);
- Acceptable – many items are acceptable (70%-89%);
- Questionable – few items are acceptable (less than 70%).

***Table 2. Item Quality Ratings for Reading MCA-MOD by Grade***

| Grade | Percentage of Items with Acceptable Ratings |                      |
|-------|---|----------------------|
|       | Written Content                             | Overall Item Quality |
| 5     | Excellent<br>(100%)                         | Acceptable<br>(84%)  |
| 6     | Excellent<br>(100%)                         | Good<br>(97%)        |
| 7     | Excellent<br>(100%)                         | Good<br>(98%)        |
| 8     | Acceptable<br>(89%)                         | Acceptable<br>(89%)  |
| 10    | Acceptable<br>(88%)                         | Good<br>(90%)        |

The independent item ratings, along with whole test reviews for each grade span group, suggest that the Reading MCA-MOD functions well for the majority of students who take the assessment. A few items on each grade’s assessment may require review to enhance clarity in wording and reduce potential bias against particular student groups.

***Recommendations***

HumRRO makes the following recommendations to strengthen the alignment between the components of the Minnesota assessment system.

- ***Review the cognitive complexity (depth of knowledge) for items on the Grade 5, 7, 8, and 10 assessments.*** The panelists reviewing these assessments rated a number of items as less demanding cognitively than the Minnesota Academic Standards. Thus, the assessments may not adequately reflect the rigor of the content expectations. Finding a disproportionate number of items assessing more basic cognitive skills is not uncommon among large-scale assessments. However, such a circumstance also is not an inevitable consequence of standardized testing, particularly for Reading assessments. Given the outcomes on depth-of-knowledge ratings along with the accessibility outcomes, it is likely that increasing the complexity of the assessment would involve modifications to current operational

items, rather than item replacement, because no items were rated as seriously flawed.

- ***Review the categorical concurrence (strand coverage) for items on the Grade 5, 8, and 10 assessments.*** Webb’s criterion of a minimum of six items per content strand was missed on three of the grade assessments based on panelists’ reviews. The assessment at grades 5, 8 and 10 needs to be reviewed to ensure that an adequate number of items are being assessed at each sub-strand level.
- ***Review the range of knowledge and balance of knowledge on the Grade 5, 8, and 10 assessments.*** These two Webb criteria examine the breadth of knowledge covered by the assessment and the distribution of items at the benchmark level within a content area. Missing these two criteria indicate that not all of the sub-strand content areas are being adequately assessed by the items and the items are not distributed evenly among the benchmarks being assessed within a content area. Increasing the number of items chosen to assess a content area and ensuring that items are adequately assessing all of the benchmarks instead of some and not others should provide a better range and balance of assessed content.
- ***Review those items that received the lowest ratings on test quality for possible revision.*** As noted in Recommendation 1, no items were rated as seriously flawed or requiring replacement. However, panelists found a small number of items for each grade test that could benefit from review to increase clarity in language. There were a couple of items in grade 10 that reviewers noted as having, potentially, more than one correct answer.

**INDEPENDENT ALIGNMENT REVIEW OF THE READING MINNESOTA COMPREHENSIVE  
ASSESSMENT - SERIES III (MCA-MOD)**

**Table of Contents**

|  |            |
|--|------------|
| <b>Chapter 1. Introduction.....</b>  | <b>1</b>   |
| Organization and Contents of the Report.....                                 | 1          |
| <b>Chapter 2. Alignment Study Design and Methodology.....</b>                | <b>2</b>   |
| Alignment of Assessments and Standards on Content .....                      | 2          |
| Content Alignment and Accessibility .....                                    | 2          |
| Webb Alignment Method .....  | 2          |
| Scope of Alignment Evaluations for Reading MCA-MOD.....                      | 3          |
| Review of Content Alignment and Accessibility.....                           | 4          |
| Panelists .....  | 4          |
| Materials .....  | 6          |
| Procedures.....  | 6          |
| <b>Chapter 3. Results: Content Alignment.....</b>                            | <b>8</b>   |
| Reliability Results.....   | 8          |
| Panelist-Test Developer Analyses .....                                       | 8          |
| Webb Alignment Results .....   | 9          |
| Categorical Concurrence.....   | 9          |
| Depth-of-Knowledge Consistency.....  | 10         |
| Range of Knowledge.....  | 17         |
| Balance-of-Knowledge Representation .....                                    | 18         |
| Summary and Discussion on Webb Alignment Indicators.....                     | 19         |
| <b>Chapter 4. Results: Test Quality of Reading MCA-MOD.....</b>              | <b>22</b>  |
| Written Content.....   | 22         |
| Overall Item Quality .....   | 23         |
| Whole Test Evaluation.....   | 24         |
| Summary and Discussion of Test Quality Results.....                          | 26         |
| <b>Chapter 5. Summary and Recommendations .....</b>                          | <b>28</b>  |
| <b>References.....</b>   | <b>30</b>  |
| <b>Appendix A. Content Alignment Results per Grade Level Assessment.....</b> | <b>A-1</b> |
| <b>Appendix B. Sample Alignment Review Materials.....</b>                    | <b>B-1</b> |

### *List of Tables*

|   |     |
|---|-----|
| Table 1. Summary Alignment Outcomes on Each Webb Criterion by Grade Level for Reading MCA-MOD.....  | iii |
| Table 2. Item Quality Ratings for Reading MCA-MOD by Grade .....  | iv  |
| Table 2.1. Professional and Demographic Characteristics of Reading MCA-MOD Panelists.....   | 5   |
| Table 2.2. Characteristics of 2011 Reading MCA-MOD Test Forms Reviewed .....  | 6   |
| Table 3.1. Percent Agreement between Panelists and Pearson on Target Content for Operational Items .....                                  | 8   |
| Table 3.2. Summary of Categorical Concurrence Results for Reading MCA-MOD .....   | 9   |
| Table 3.3. Panelists Ratings on Overall Item Alignment.....   | 10  |
| Table 3.4. Summary of Depth-of-Knowledge Results for Reading MCA-MOD Operational Items  | 11  |
| Table 3.5. Cognitive Level Minimum Distribution of Items in Reading from 2010 MCA-MOD Test Specifications.....                            | 12  |
| Table 3.6. Number of Content Sub-Strands and Benchmarks per Grade Level Reading MCA-MOD .....   | 18  |
| Table 3.7. Summary of Range-of-Knowledge Results for Reading MCA-MOD .....  | 18  |
| Table 3.8. Summary of Balance-of-Knowledge Representation Results for Reading MCA-MOD   | 19  |
| Table 3.9. Summary Alignment Outcomes on Each Webb Criterion per Grade Level for Reading MCA-MOD.....                                     | 21  |
| Table 4.1. Mean Number of Items Rated as Accessible in Content to Range of Students per Grade Assessment.....                             | 22  |
| Table 4.3. Panelist Ratings on Overall Item Quality .....   | 24  |
| Table 4.4. Grade 5 – 7 Reading MCA-MOD: Consensus Ratings on Whole Test Evaluation .....  | 25  |
| Table 4.5. Grade 8 and 10 Reading MCA-MOD: Consensus Ratings on Whole Test Evaluation   | 26  |
| Table 4.6. Item Quality Ratings for Reading MCA-MOD per Grade Assessment .....  | 27  |
| Table A-1. Categorical Concurrence for Reading MCA-MOD, Grade 5: Mean Number of Items per Sub-Strand .....                                | A-1 |
| Table A-2. Categorical Concurrence for Reading MCA-MOD, Grade 6: Mean Number of Items per Sub-Strand .....                                | A-1 |
| Table A-3. Categorical Concurrence for Reading MCA-MOD, Grade 7: Mean Number of Items per Sub-Strand .....                                | A-2 |
| Table A-4. Categorical Concurrence for Reading MCA-MOD, Grade 8: Mean Number of Items per Sub-Strand .....                                | A-2 |
| Table A-5. Categorical Concurrence for Reading MCA-MOD, Grade 10: Mean Number of Items per Sub-Strand .....                               | A-2 |
| Table A-6. DOK Consistency for Reading MCA-MOD, Grade 5: Mean Percent of Items with DOK Below, At, and Above DOK Level of Benchmarks..... | A-3 |
| Table A-7. DOK Consistency for Reading MCA-MOD, Grade 6: Mean Percent of Items with DOK Below, At, and Above DOK Level of Benchmarks..... | A-3 |
| Table A-8. DOK Consistency for Reading MCA-MOD, Grade 7: Mean Percent of Items with DOK Below, At, and Above DOK Level of Benchmarks..... | A-4 |

|   |      |
|---|------|
| Table A-9. DOK Consistency for Reading MCA-MOD, Grade 8: Mean Percent of Items with DOK Below, At, and Above DOK Level of Benchmarks.....   | A-4  |
| Table A-10. DOK Consistency for Reading MCA-MOD, Grade 10: Mean Percent of Items with DOK Below, At, and Above DOK Level of Benchmarks..... | A-5  |
| Table A-11. Range-of-Knowledge for Reading MCA-MOD, Grade 5: Mean Percent of Benchmarks per Sub-Strand Linked with Items.....               | A-5  |
| Table A-12. Range-of-Knowledge for Reading MCA-MOD, Grade 6: Mean Percent of Benchmarks per Sub-Strand Linked with Items.....               | A-6  |
| Table A-13. Range-of-Knowledge for Reading MCA-MOD, Grade 7: Mean Percent of Benchmarks per Sub-Strand Linked with Items.....               | A-6  |
| Table A-14. Range-of-Knowledge for Reading MCA-MOD, Grade 8: Mean Percent of Benchmarks per Sub-Strand Linked with Items.....               | A-7  |
| Table A-15. Range-of-Knowledge for Reading MCA-MOD, Grade 10: Mean Percent of Benchmarks per Sub-Strand Linked with Items.....              | A-7  |
| Table A-16. Balance-of-Knowledge Representation for Reading MCA-MOD, Grade 5: Mean Balance Index per Sub-Strand.....                        | A-8  |
| Table A-17. Balance-of-Knowledge Representation for Reading MCA-MOD, Grade 6: Mean Balance Index per Sub-Strand.....                        | A-8  |
| Table A-18. Balance-of-Knowledge Representation for Reading MCA-MOD, Grade 7: Mean Balance Index per Sub-Strand.....                        | A-9  |
| Table A-19. Balance-of-Knowledge Representation for Reading MCA-MOD, Grade 8: Mean Balance Index per Sub-Strand.....                        | A-9  |
| Table A-20. Balance-of-Knowledge Representation for Reading MCA-MOD, Grade 10: Mean Balance Index per Sub-Strand.....                       | A-10 |
| Table A-21. Grade 5 MCA-MOD: Grade Span Benchmarks Matched to Items by Panelists..  | A-10 |
| Table A-22. Grade 6 MCA-MOD: Grade Span Benchmarks Matched to Items by Panelists..  | A-11 |
| Table A-23. Grade 7 MCA-MOD: Grade Span Benchmarks Matched to Items by Panelists..  | A-11 |
| Table A-24. Grade 8 MCA-MOD: Grade Span Benchmarks Matched to Items by Panelists..  | A-12 |
| Table A-25. Grade 10 MCA-MOD: Grade Span Benchmarks Matched to Items by Panelists..   | A-12 |

***List of Figures***

|   |    |
|---|----|
| Figure 3.1. Grade 5: Distribution of items per cognitive level based on Pearson Item Assignment compared to panelists’ mean item ratings. ....  | 13 |
| Figure 3.2. Grade 6: Distribution of items per cognitive level based on Pearson Item Assignment compared to panelists’ mean item ratings. ....  | 14 |
| Figure 3.3. Grade 7: Distribution of items per cognitive level based on Pearson Item Assignment compared to panelists’ mean item ratings. ....  | 15 |
| Figure 3.4. Grade 8: Distribution of items per cognitive level based on Pearson Item Assignment compared to panelists’ mean item ratings. ....  | 16 |
| Figure 3.5. Grade 10: Distribution of items per cognitive Level based on Pearson Item Assignment compared to panelists’ mean item ratings. .... | 17 |

# INDEPENDENT ALIGNMENT REVIEW OF THE READING MINNESOTA COMPREHENSIVE ASSESSMENT - SERIES III (MCA-MOD)

## Chapter 1. Introduction

The Minnesota Department of Education (MDE) requested an external independent alignment study of the Reading Minnesota Comprehensive Assessment - Modified (MCA-MOD). Specifically, MDE wanted an evaluation of the alignment of the Reading MCA-MOD for grades 5 – 8 and 10 to the Minnesota Academic Standards<sup>2</sup>. Minnesota uses the Reading MCA-MOD test in the federal and state accountability programs. The Human Resources Research Organization (HumRRO) was contracted to conduct this alignment study, which occurred Oct. 3 – 5, 2011.

MDE requested the alignment study in order to meet both state and federal requirements. The federal requirements of the U.S. Department of Education (USDE) stem from the No Child Left Behind (NCLB) Act of 2001. NCLB challenges each state to establish a coherent assessment system based on solid academic standards. This law calls for states to provide independent evidence of the validity of their assessments used to calculate Adequate Yearly Progress (AYP). All states receiving Title I funds must present evidence of establishing a fair and consistent assessment system that is based on rigorous standards, sufficient alignment between standards and assessments, and high-quality educational results.

An alignment review can provide one form of evidence supporting the validity of the state assessment system. Alignment results should demonstrate that the assessments represent the full range of the content standards and that the assessments measure student knowledge in the same manner and at the same level of complexity as specified in the content standards. All aspects of the state assessment system must coincide, including the academic content standards, achievement standards (linked to cut scores), performance level descriptors and each assessment.

### *Organization and Contents of the Report*

This report contains five chapters. Chapter 2 explains alignment methodologies. Subsequent chapters provide alignment results for comparisons between the components of the assessment system: (a) Chapter 3 presents results of the alignment comparison between the Reading assessments and the Minnesota Academic Standards; (b) Chapter 4 presents results on the accessibility of the assessments to all students; and (c) Chapter 5 provides recommendations for MDE to strengthen the alignment of the Reading MCA-MOD over time.

Additional information is provided in the appendices of this report. Appendix A contains tables providing more detail on the content alignment results for the grade-level test forms. Appendix B provides examples of rating forms and training materials used in the alignment workshops.

---

<sup>2</sup> Minnesota Academic Standards can be found at <http://education.state.mn.us/MDE/EdExc/StanCurri/K-12AcademicStandards/>

## **Chapter 2. Alignment Study Design and Methodology**

In this section, we discuss key concepts related to alignment research, followed by a description of the alignment evaluations and methods used as part of the Minnesota study.

### ***Alignment of Assessments and Standards on Content***

The term *alignment* in this context refers to the degree of accuracy evident in instruction and measurement of the state’s academic content standards. School curriculum must include appropriate content laid out by the state. Any documents developed to accompany the content standards (e.g., performance descriptors, test specifications, teaching guides) must accurately represent the expectations. Assessments must measure only the content specified in the standards, and student scores generated from these assessments should adequately reflect student knowledge of the content standards. An alignment study evaluates the strength of any or all of these relationships.

In general, alignment evaluations for any assessment reveal the breadth, or scope, of knowledge as well as the depth of knowledge, or cognitive processing, expected of students by the state’s content standards. Alignment analyses help to answer questions such as the following:

- How much and what type of content is covered by the assessment?
- Is the content in the assessment, or other standards, sufficiently similar to the expectations of the full content standards?
- Are students asked to demonstrate this knowledge at the same level of rigor as expected in the full content standards?
- Does the assessment accurately measure student knowledge of content standards?

These questions essentially can be grouped into two categories—content alignment and performance alignment. However, all alignment evaluations tie back to the state content standards.

### ***Content Alignment and Accessibility***

Several methods of alignment are in current use. Most methods involve ratings of several aspects of the assessment items relative to the content standards. The ratings are analyzed statistically to determine the extent of alignment. HumRRO used the alignment method developed by Norman Webb (1997; 1999; 2005) to evaluate the Reading MCA-MOD.

### ***Webb Alignment Method***

The Webb alignment method was originally designed for use with standard large-scale assessments. Dr. Webb has researched and refined this method over time (e.g., Webb, 1997; 1999; 2005), and his approach is supported by the Council of Chief State School Officers (CCSSO).

The Webb method includes four major criteria to evaluate alignment. These criteria link with statistical procedures used to assess how well individual portions of the assessments and

standards documents actually match. The four alignment criteria are: categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance-of-knowledge representation.

***Categorical concurrence*** is a basic measure of alignment between content standards and test items. This term refers to the proportion of overlap between the content stated in the standards document and that assessed by items on the test.

***Depth of Knowledge*** (DOK) measures the type of cognitive processing required by items and content standards. For example, is a student expected to simply identify or recall basic facts, or is the student expected to use reasoning in manipulating information or strategizing? Using Reading as an example, a student may be asked to identify a simile among several answer choices. This task should be less complex than trying to explain the concept of a simile and writing one.

The purpose of using DOK as a measure of alignment is to determine whether a test item and its corresponding standard are written at the same level of cognitive complexity. Reviewers make two separate judgments about cognitive complexity, one for the standard and one for the item. These two judgments are compared to determine whether the item is written at the same level as the standard to which it is linked. Webb refers to this comparison as *Depth-of-Knowledge consistency*.

Another measure examines the ***range-of-knowledge correspondence*** between the assessment and content standards. The range-of-knowledge measure looks in greater detail at the breadth of knowledge represented by test items. Categorical concurrence simply notes whether a sufficient number of items on the test covers each general content topic (individual strands). However, states usually lay out more specific *content objectives*, or standards, under each strand. The range indicates the number of content objectives assessed by items.

Finally, the ***balance-of-knowledge representation*** criterion focuses on content coverage in yet more detail. In this case, the number of items matched to the content objective does matter. The balance of representation determines whether the assessment measures the content objectives equitably within each standard. Based on Webb's method, items should be distributed evenly across the objectives per standard for good balance. The balance-of-knowledge representation is determined by calculating an index, or score, for each standard. Each standard should meet or surpass a minimum index level to demonstrate adequate balance.

### ***Scope of Alignment Evaluations for Reading MCA-MOD***

The alignment evaluation performed for this Minnesota study involved a comparison of the Reading MCA-MOD to the Minnesota Academic Standards. The content alignment evaluation involved a review by a panel of current and recently retired Minnesota educators highly familiar with the content standards and the assessment.

### ***Review of Content Alignment and Accessibility***

For the content alignment review, HumRRO convened panels of Minnesota educators to review the grades 5 – 8 and 10 Reading MCA-MOD. The review involved two major tasks: (a) matching the Reading items to grade level Minnesota Academic Standards for Reading, and (b) evaluating test quality with respect to students who take the test.

#### ***Panelists***

HumRRO developed two review panels with the assistance of MDE and Pearson. Panelists were recruited by Pearson from their database of Minnesota educators. Every effort was made to produce panels consisting of teachers reflecting the population of students who take the assessments. Panels were convened in facilities procured by Pearson. HumRRO directed the actual reviews independently of MDE and Pearson. Table 2.1 presents the characteristics of the panelists per grade-level of the Reading MCA-MOD.

**Table 2.1. Professional and Demographic Characteristics of Reading MCA-MOD Panelists**

| Professional Position   | Number of Panelists | Average Years of Experience | Special Certifications | Region of Origin in Minnesota |                   |              | Gender |   | Ethnicity           |                     |                        |                                |   |
|-------------------------|---------------------|-----------------------------|------------------------|-------------------------------|-------------------|--------------|--------|---|---------------------|---------------------|------------------------|--------------------------------|---|
|                         |                     |                             |                        | 7-County Metro                | Greater Minnesota | MPLS/St Paul | M      | F | White, Non-Hispanic | Black, Non-Hispanic | Asian/Pacific Islander | American Indian/Alaskan Native |   |
| <b>Grade 5 – 7</b>      |                     |                             |                        |                               |                   |              |        |   |                     |                     |                        |                                |   |
| Teacher                 | 3                   | 24 (n = 2)                  | 3                      | 0                             | 1                 | 2            | 0      | 3 | 3                   | 0                   | 0                      | 0                              | 0 |
| Administrator           | 1                   | 22 (n = 1)                  | 0                      | 1                             | 0                 | 0            | 0      | 1 | 1                   | 0                   | 0                      | 0                              | 0 |
| College Educator        | 1                   |                             | 1                      | 0                             | 1                 | 0            | 0      | 1 | 1                   | 0                   | 0                      | 0                              | 0 |
| <b>Grade 8 &amp; 10</b> |                     |                             |                        |                               |                   |              |        |   |                     |                     |                        |                                |   |
| Teacher                 | 3                   | 23.33 (n = 3)               | 2                      | 2                             | 1                 | 0            | 1      | 2 | 3                   | 0                   | 0                      | 0                              | 0 |
| Administrator           | 2                   | 10 (n = 1)                  | 1                      | 1                             | 0                 | 1            | 1      | 1 | 2                   | 0                   | 0                      | 0                              | 0 |
| College Educator        | 0                   |                             | 0                      | 0                             | 0                 | 0            | 0      | 0 | 0                   | 0                   | 0                      | 0                              | 0 |

## **Materials**

Panelists evaluated the alignment of the MCA-MOD items with the Minnesota Academic Standards using rating forms adapted from Webb (2005). All rating forms were completed electronically in Microsoft Excel. Examples of rating forms and instructions are presented in Appendix B.

**Test Forms.** Panelists evaluated the 2011 Reading MCA-MOD test form per grade. The MCA-MOD assessment was administered as a paper and pencil assessment. Table 2.2 lists the characteristics of the form for the 2011 administration for each grade-level test. Because the test form is a secure document, this report does not include any examples of items or references to specific item content.

**Table 2.2. Characteristics of 2011 Reading MCA-MOD Test Forms Reviewed**

| Grade Level | Total Items per Form | Number of Operational Items | Number of Field Test Items |
|-------------|----------------------|-----------------------------|----------------------------|
| 5           | 55                   | 35                          | 20                         |
| 6           | 55                   | 35                          | 20                         |
| 7           | 55                   | 35                          | 20                         |
| 8           | 55                   | 35                          | 20                         |
| 10          | 57                   | 35                          | 22                         |

Panelists made their content alignment ratings on a print version of screen shots of items from the online test form.

**Rating Forms and Instructions.** Panelists were given instruction sheets listing the rating tasks and forms, as well as code sheets identifying the range of acceptable codes per task (see Appendix B). Panelists completed two rating forms individually: (a) Item Rating Form, and (b) “whole test” rating form. In addition, each grade span group completed DOK Ratings of Minnesota Academic Standards through consensus (see Appendix B for samples of each).

## **Procedures**

HumRRO conducted this alignment review at meeting rooms in Minneapolis, MN on Oct. 3–5, 2011. The workshops began with introductions of staff and observers. HumRRO staff gave a presentation describing the purpose of the reviews and alignment research in general. This presentation briefly introduced the alignment tasks the panelists would be performing. Reviewers had the opportunity to practice making ratings during the large group session.

Following the general introduction, panelists began working within their content groups. The Reading MCA-MOD reviewers were split into two groups, one at each grade span (grade span 5–7, and 8 & 10). Each grade span contained five reviewers. One panelist in the 5 – 7 grade span was not able to return on the second day. HumRRO staff supervised the grade span groups.

Within their small groups, panelists first read and signed affidavits of nondisclosure for the secure materials they would be reviewing during the workshop. Then, HumRRO staff further trained reviewers with sample assessment items and by answering questions on rating criteria. Regarding instructions on how to rate standards and items, HumRRO staff provided general suggestions and comments when appropriate; however, they emphasized to reviewers that staff would not give explicit direction on how to rate standards or items because reviewers were valued as content experts. Each panelist received a laptop with rating forms already uploaded and formatted. HumRRO staff provided brief instructions about how to work with the electronic rating forms.

After reviewing sample DOK evaluations as a group, panelists proceeded to rate the benchmarks from the Minnesota Academic Standards relevant to each grade span test. For example, panelists reviewing the grade 5 test rated the benchmarks for grades 5 and 6. Panelists first made independent evaluations without discussion. Once all reviewers had completed their DOK ratings, the groups discussed their ratings to achieve consensus for each benchmark; a voluntary scribe within each group recorded these consensus ratings.

Reviewers then received more specific instructions for rating the items. For training, HumRRO staff facilitated the reviewers in evaluating and discussing sample items as a group. After completing the sample items, reviewers rated the items individually on electronic rating forms on their laptops. Panelists rated the individual items on the 2011 test forms for their group on several dimensions, including: (a) content match to the benchmarks in the Minnesota Academic Standards, (b) depth of knowledge required by the item, (c) degree of alignment (i.e., how well the item links to the benchmark), (d) content clarity (i.e., readability), and (e) quality of accompanying graphics (if applicable). Panelists assigned a *primary benchmark* to an item based on a judgment that an item clearly measured this content; however, reviewers could assign an *additional standard* if the item seemed to assess another standard equally to the primary standard. These ratings were conducted individually without consensus.

Finally, panelists worked in their small groups to rate three additional aspects of the MCA-MOD tests. HumRRO staff trained the panelists on each task, and had them record their ratings in pre-formatted Excel spreadsheets. The first consensus task required panelists to rate potential barriers for students in being able to demonstrate knowledge (aspects of the MCA-MOD as a whole that might prevent students with various disabilities or English learners from fully participating). For the second task, panelists rated the extent to which content differs appropriately across the grade level assessments.

All panelists finished tasks in approximately 2 days, although they completed their ratings at different times. Once panelists finished the review, their session ended.

### Chapter 3. Results: Content Alignment

In this chapter, we report the results of the content alignment evaluation. These analyses are based on panelists’ ratings of the Reading MCA-MOD items.

#### *Reliability Results*

In this section, we report on the comparison of panelists’ ratings of content match to the test contractor’s intended content match. The agreement levels across grades were sufficient to provide further evidence supporting the validity of the alignment process and outcomes reported here.

#### *Panelist-Test Developer Analyses*

Table 3.1 presents the agreement outcomes between panelists and Pearson on the content assessed by items per grade level. Agreement was analyzed at several levels of specificity as shown under the table heading ‘Percent Agreement with Pearson Codes’. All of the items were analyzed first for ‘Exact Match’, which indicates that panelists chose the same strand, sub-strand, and benchmark for the item as the test developer. If panelists did not show an exact match with Pearson, we determined the percent agreement at the *sub-strand* level (panelists selected the same strand and sub-strand as Pearson). Finally, for remaining items, we determined whether panelists at least chose the same *strand* as the test developer. The last column in Table 3.1 shows the percentage of ratings by panelists that did not match the Pearson coding at all on items. Because panelists could assign two content codes to a single item, we counted either code if at least one matched with Pearson. The agreement levels reported in Table 3.1 represent separate analyses; thus, percent agreement in each row adds to greater than 100%.

***Table 3.1. Percent Agreement between Panelists and Pearson on Target Content for Operational Items***

| Grade Level | Number of Operational Items per Form | Total Number of Panelist Ratings across Items | Percent Agreement with Pearson Codes |                  |              |          |
|-------------|--------------------------------------|---|--------------------------------------|------------------|--------------|----------|
|             |                                      |   | Exact Match                          | Sub-Strand Match | Strand Match | No Match |
| 5           | 35                                   | 193   | 2                                    | 67               | 100          | 0        |
| 6           | 35                                   | 165   | 0                                    | 65               | 93           | 7        |
| 7           | 35                                   | 155   | 2                                    | 63               | 95           | 5        |
| 8           | 35                                   | 187   | 6                                    | 68               | 96           | 4        |
| 10          | 35                                   | 180   | <1                                   | 74               | 100          | 0        |

As Table 3.1 indicates, panelists were not able to match Pearson in identifying the assessment target of items at the most specific (benchmark) content level. However, panelists were able to match Pearson more consistently at the sub-strand content level. Panelists were more consistent with Pearson at the highest (strand) content level. Furthermore, panelists differed completely from Pearson on content match for a small percentage of items per grade level. These

findings suggest that the operational Reading items do, in fact, measure the intended content particularly at the strand level.

### ***Webb Alignment Results***

In this section, we review the general outcomes of item analyses on the four Webb alignment indicators. These analyses only include operational items. More detailed numeric results can be found in Appendix A.

All of Webb’s measures begin with calculations for each panelist and build up to a summary of results across raters per content strand. First, we calculated the mean ratings across items for each panelist, and then we determined the mean rating across panelists per strand. Results are presented at the strand level.

### ***Categorical Concurrence***

Categorical concurrence describes the extent to which the MCA-MOD items cover the content strands in the Minnesota Academic Standards for Reading. Webb recommends a minimum of six test questions to adequately assess each content strand. This criterion serves as a guideline for reasonable content coverage. For the Reading MCA-MOD, there is only one strand that was assessed on the test. Therefore, this analysis is conducted at the sub-strand level where there are different content areas. Table 3.2 summarizes the MCA-MOD alignment results for categorical concurrence.

***Table 3.2. Summary of Categorical Concurrence Results for Reading MCA-MOD***

| Grade Level | Mean Number of Items per Sub-Strand per Form |               |             | Sub-Strands with at Least Six Items |
|-------------|--|---------------|-------------|-------------------------------------|
|             | Vocabulary Expansion                         | Comprehension | Literature  |                                     |
| 5           | 7.20   | 28.00         | <b>3.40</b> | 2 of 3                              |
| 6           | 9.50   | 25.75         | 8.00        | 3 of 3                              |
| 7           | 6.75   | 20.00         | 11.00       | 3 of 3                              |
| 8           | 6.80   | 26.80         | <b>3.40</b> | 2 of 3                              |
| 10          | 8.00   | 24.80         | <b>3.20</b> | 2 of 3                              |

As Table 3.2 indicates, the grade 5 – 8 and 10 assessments include a sufficient number of items to meet the minimum requirements for categorical concurrence on nearly all Reading content sub-strands assessed. The grade 5, 8, and 10 items cover only two of the three strands, missing the minimum requirements on the Literature content strand. In general, these results indicate that the Reading MCA-MOD adequately cover the Reading content students are expected to know across these grade levels.

In addition to identifying the benchmark assessed by each item, we asked panelists to indicate *how well* the item assessed the benchmarks. Panelists rated the extent of item alignment to the benchmarks on a 4-point scale ranging from ‘Not aligned to any benchmark’ to ‘Fully aligned to a benchmark – exemplary item’. Table 3.3 presents the mean number of items (across

panelists) at each level of alignment. For each grade assessment, panelists rated items as aligned well to the benchmarks matched to that item.

**Table 3.3. Panelists Ratings on Overall Item Alignment**

| Grade | Degree of Alignment | Mean Number of Items per Level | SD    | Percentage of Items per Level |
|-------|---------------------|--------------------------------|-------|-------------------------------|
| 5     | Not at all aligned  | 0.00                           | NA    | 0                             |
|       | Weakly aligned      | 5.00                           | 2.55  | 13                            |
|       | Highly aligned      | 25.40                          | 7.83  | 66                            |
|       | Fully aligned       | 10.25                          | 5.19  | 21                            |
| 6     | Not at all aligned  | 0.00                           | NA    | 0                             |
|       | Weakly aligned      | 1.00                           | NA    | 1                             |
|       | Highly aligned      | 32.25                          | 8.73  | 78                            |
|       | Fully aligned       | 17.50                          | 14.85 | 21                            |
| 7     | Not at all aligned  | 0.00                           | NA    | 0                             |
|       | Weakly aligned      | 1.50                           | 0.71  | 2                             |
|       | Highly aligned      | 31.75                          | 6.13  | 82                            |
|       | Fully aligned       | 12.50                          | 12.02 | 16                            |
| 8     | Not at all aligned  | 1.00                           | NA    | 1                             |
|       | Weakly aligned      | 2.80                           | 1.92  | 7                             |
|       | Highly aligned      | 23.80                          | 10.08 | 64                            |
|       | Fully aligned       | 10.60                          | 10.50 | 28                            |
| 10    | Not at all aligned  | 1.00                           | NA    | 1                             |
|       | Weakly aligned      | 1.50                           | 0.58  | 3                             |
|       | Highly aligned      | 25.60                          | 11.24 | 71                            |
|       | Fully aligned       | 11.25                          | 11.44 | 25                            |

In general, panelists across all grades and forms rated items as being ‘Highly aligned’. The grade 5 assessment is the only exception to this general conclusion. For this assessment, panelists rated 13% of the items as being ‘Weakly aligned’.

### ***Depth-of-Knowledge Consistency***

Analyses of depth of knowledge (DOK) measure the type of cognitive processing required of students by content standards. The DOK requirements implied by the benchmarks should be matched by assessment items. To confirm this match, panelists were asked to rate the benchmarks and the Reading items separately. Webb includes an alignment indicator that directly compares panelists’ DOK ratings of content standards and test items, which he refers to as *depth-of-knowledge consistency*.

To make their ratings, panelists used a rating scale (adapted from Webb, 2005) with four levels of cognitive complexity. Further information and examples of the DOK levels are found in Appendix C.

- Level 1 Recognition - simple recall of information (i.e., facts, terms); sequencing; more automatic.
- Level 2 Skills/Concepts - beyond habitual response; applying concepts; problem-solving.
- Level 3 Strategic Thinking - requires basic reasoning, planning, or use of evidence; generating hypotheses.
- Level 4 Extended Thinking - complex reasoning; evaluation of multiple sources or independent pieces of evidence; often over an extended period of time.

Table 3.4 summarizes the depth-of-knowledge consistency results for each grade level of the Reading MCA-MOD. Because reviewers evaluated depth of knowledge at the most specific level of the standards document (benchmarks), the table refers to consistency between the items and the benchmarks to which they were matched. Results are summarized in terms of the percentage of items with cognitive complexity ratings at or above (more complex than) the rating for the corresponding benchmark.

Webb’s suggested criterion for this alignment indicator is that at least 50% of the items should have complexity ratings at or above the level of the corresponding benchmark.

**Table 3.4. Summary of Depth-of-Knowledge Results for Reading MCA-MOD Operational Items**

| Grade | Percentage of Items with DOK At or Above the Level of the Benchmarks per Sub-Strand |               |            | Number of Sub-Strands Assessed Adequately | Specific Sub-Strands Assessed Inadequately |
|-------|---|---------------|------------|---|--|
|       | Vocabulary Expansion  | Comprehension | Literature |   |  |
| 5     | 81  | <b>41</b>     | 73         | 2 of 3                                    | Comprehension                              |
| 6     | 80  | 56            | 87         | 3 of 3                                    |  |
| 7     | 76  | <b>42</b>     | 78         | 2 of 3                                    | Comprehension                              |
| 8     | 83  | 57            | <b>32</b>  | 2 of 3                                    | Literature                                 |
| 10    | 86  | <b>44</b>     | 83         | 2 of 3                                    | Comprehension                              |

For grades 5, 7, and 10, panelists’ ratings using Webb DOK levels imply that for less than half of the items targeting the Comprehension content sub-strand assessed students at the appropriate cognitive complexity. For grade 8, panelists’ ratings show that a third of the items targeting Literature assessed students at the level of cognitive complexity intended in the item specifications.

Across all grades, panelists’ ratings on depth-of-knowledge consistency suggest that some of the Reading MCA-MOD items may not assess students at the level expected in the

Minnesota Academic Standards. The tables above indicate that only the grade 6 assessment met the minimum criterion of the Webb method for all sub-strands.

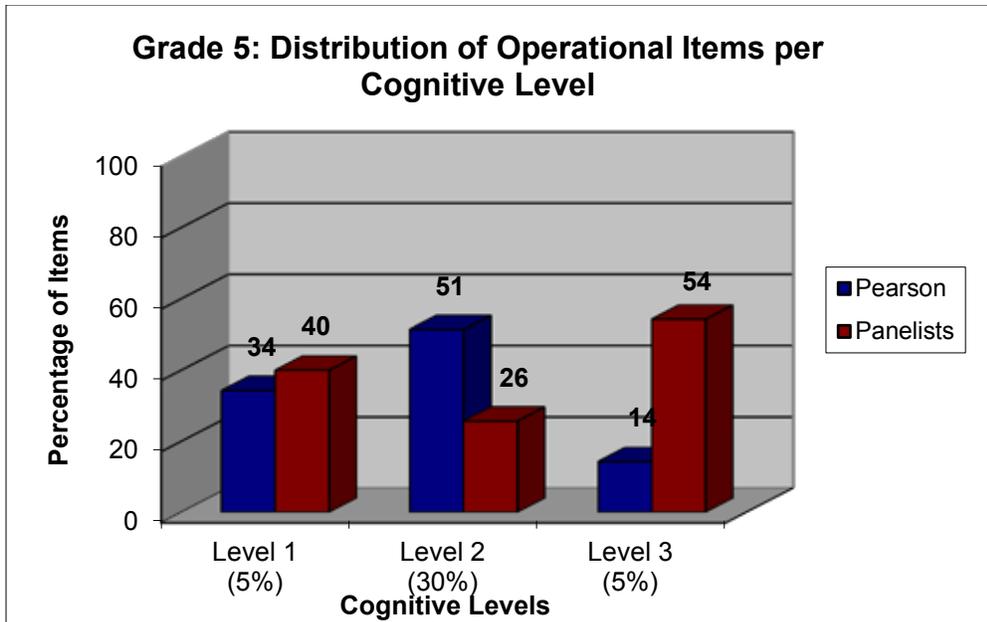
As a result of these outcomes based on the Webb method, we conducted a more in-depth review of panelists’ evaluations compared to the assessment targets intended by the test contractor. This analysis required us to map the Webb cognitive levels to the Minnesota cognitive levels. The processing distinctions made by Webb and Minnesota are comparable, and they stem from the same research on Bloom’s Taxonomy (Bloom, Englehart, Furst, Hill, & Krathwohl, 1956). However, Minnesota chose to adopt three cognitive levels, whereas Webb makes four distinctions. A comparison of these frameworks suggests that Webb’s Level 3 (strategic thinking) and Level 4 (extended thinking) can be collapsed into Cognitive Level C (MCA-MOD Test Specifications for Reading, 2010, p. 5).

The MCA-MOD Test Specifications for Reading includes the following table specifying item distributions across cognitive levels per grade test. This table indicates the minimum proportion of items per cognitive level that should be included in each administration (2010, p.9).

***Table 3.5. Cognitive Level Minimum Distribution of Items in Reading from 2010 MCA-MOD Test Specifications***

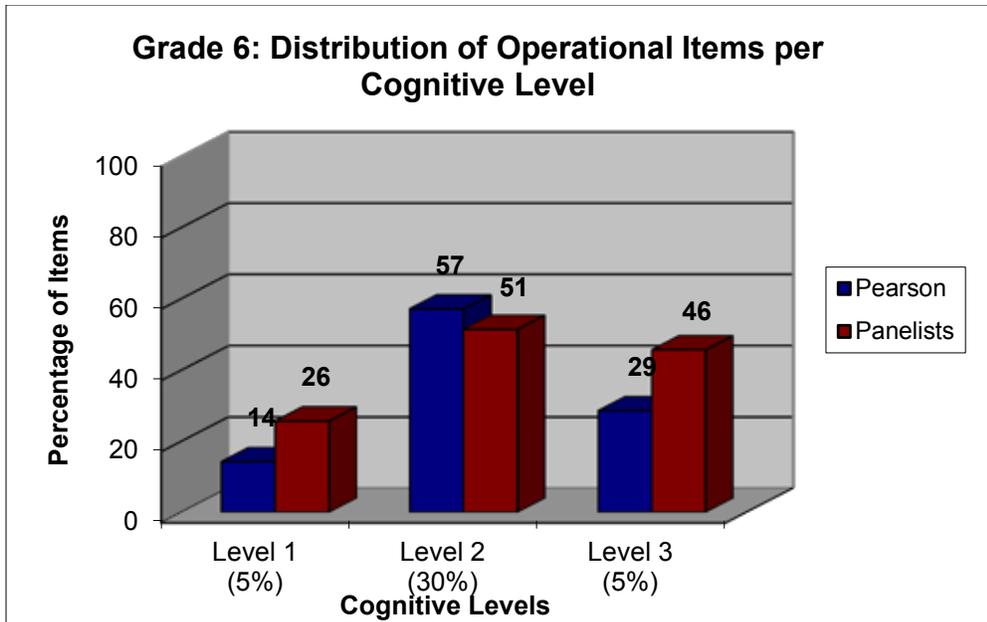
| Grades     | Distribution of Items by Cognitive Level |         |         |
|------------|--|---------|---------|
|            | Level 1                                  | Level 2 | Level 3 |
| 5 – 8 & 10 | 5%                                       | 30%     | 5%      |

Relative to these proportions, we compared the actual number of items that Pearson assigned to each cognitive level on the 2011 assessments with the mean number of items per cognitive level based on panelists’ ratings. Figures 3.1 through 3.5 display the percentage of items by Pearson and the percentage of mean items based on panelists’ ratings for cognitive level per form and grade. The x-axis includes the Webb levels (Level 1, Level 2, and Level 3). The y-axis indicates the percentage of items per level. However, it is important to note that the percentage of items for the panelists is based on the *mean* number of items in the panelists’ distribution, while the scale reflects the percentage of items assigned to each cognitive level in the Pearson item distribution.



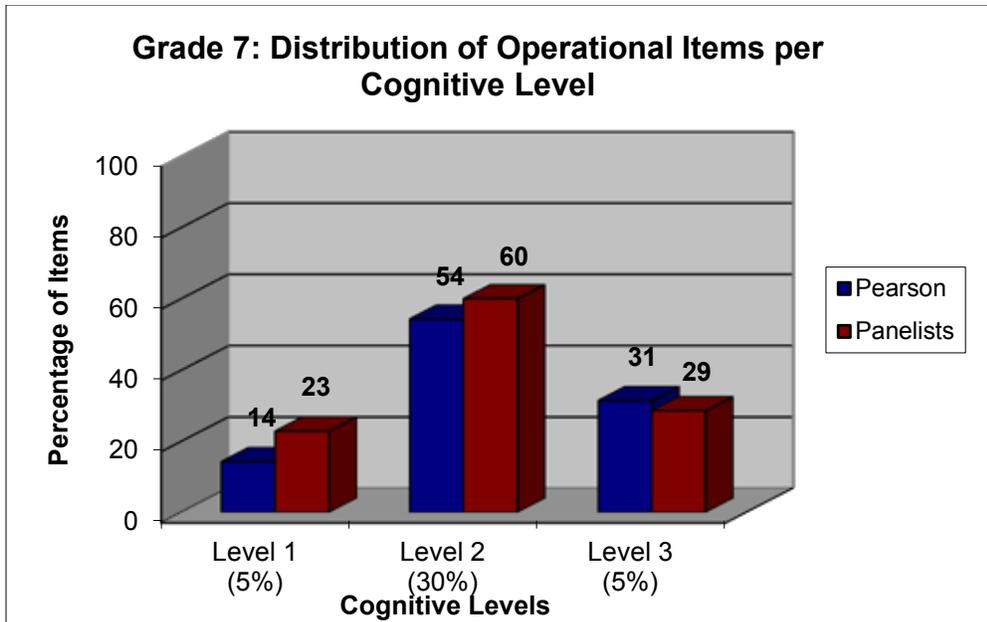
**Figure 3.1. Grade 5: Distribution of items per cognitive level based on Pearson Item Assignment compared to panelists' mean item ratings.**

For the grade 5 assessment, the panelists' cognitive ratings missed the minimum distribution outlined in the test specifications. Based on panelists' cognitive ratings, there appears to be an adequate percentage of items at Levels 1 and 3 but a slightly lower percentage of items at Level 2. Of greater importance is that panelists' cognitive ratings resulted in more than half of the items being assigned a DOK of Level 3 compared to only 14% of the items being assigned a Level 3 by Pearson.



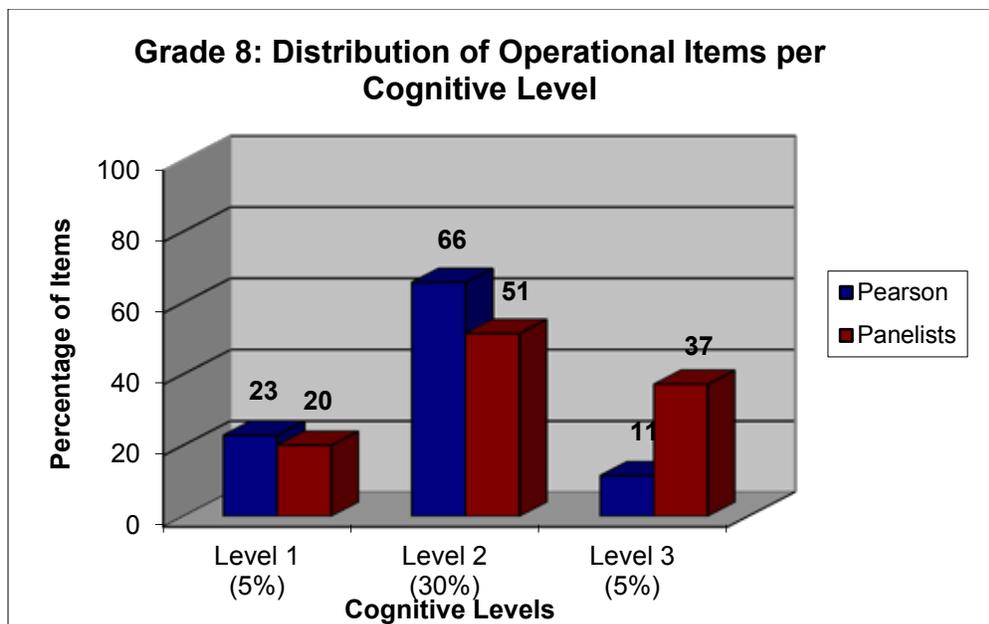
**Figure 3.2. Grade 6: Distribution of items per cognitive level based on Pearson Item Assignment compared to panelists’ mean item ratings.**

For the grade 6 assessment, the panelists’ cognitive ratings and Pearson’s cognitive ratings resulted in the minimum distribution outlined in the test specifications. There is adequate representation of items at all levels. Of greater importance is that panelists’ cognitive ratings resulted in a little less than half of the items being assigned a DOK of Level 3 compared to only 29% of the items being assigned a Level 3 by Pearson.



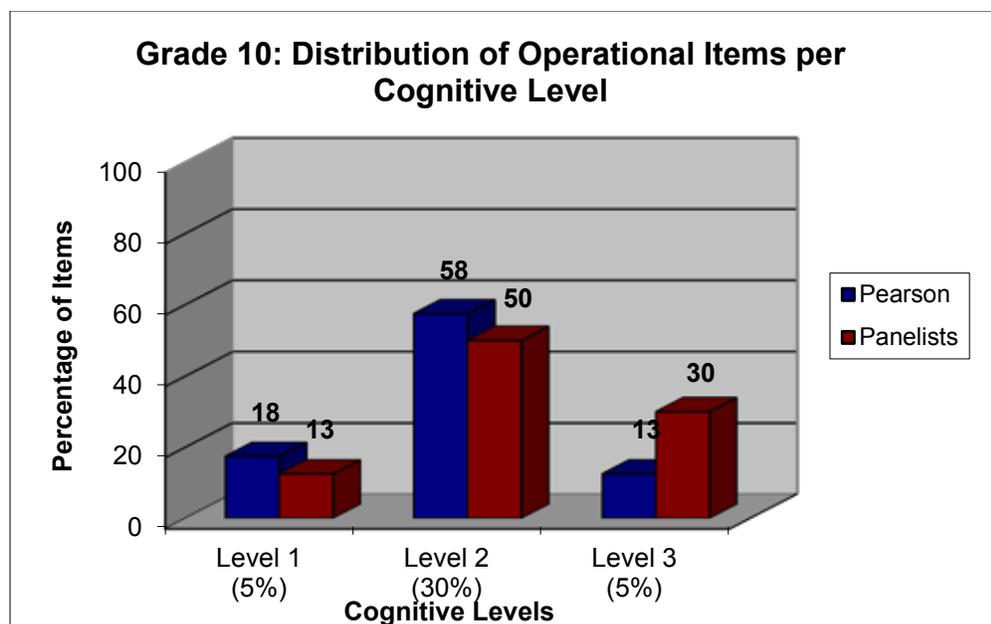
**Figure 3.3. Grade 7: Distribution of items per cognitive level based on Pearson Item Assignment compared to panelists' mean item ratings.**

For the grade 7 assessment, the panelists' cognitive ratings and Pearson's cognitive ratings resulted in the minimum distribution outlined in the test specifications. There is adequate representation of items at all levels.



**Figure 3.4. Grade 8: Distribution of items per cognitive level based on Pearson Item Assignment compared to panelists’ mean item ratings.**

For the grade 8 assessment, the panelists’ cognitive ratings and Pearson’s cognitive ratings resulted in the minimum distribution outlined in the test specifications. There is adequate representation of items at all levels. Of greater importance is that panelists’ cognitive ratings resulted in 37% of the items being assigned a DOK of Level 3, compared to only 11% of the items being assigned a Level 3 by Pearson.



**Figure 3.5. Grade 10: Distribution of items per cognitive Level based on Pearson Item Assignment compared to panelists’ mean item ratings.**

For the grade 11 assessment, the panelists’ cognitive ratings and Pearson’s cognitive ratings resulted in the minimum distribution outlined in the test specifications. There is adequate representation of items at all levels. Of greater importance is that panelists’ cognitive ratings resulted in 30% of the items being assigned a DOK of Level 3, compared to only 13% of the items being assigned a Level 3 by Pearson.

### **Range of Knowledge**

The range-of-knowledge measure examines in greater detail the breadth of knowledge covered by the assessment. In addition to evaluating which content strands are assessed, we must look at how many of the benchmarks within a strand are represented by items. The benchmarks should be linked with at least one item. Webb’s minimum level of acceptability for range-of-knowledge correspondence is that at least 50% of benchmarks per strand link with items to ensure adequate breadth of content coverage.

Table 3.6 lists the number of strands and benchmarks found in the Minnesota Academic Standards. Column 4 indicates the number of benchmarks that may be represented on the assessment. The last column also indicates the number of items available to assess these strands and benchmarks.

**Table 3.6. Number of Content Sub-Strands and Benchmarks per Grade Level Reading MCA-MOD**

| Grade Level Test | Number of Content Sub-Strands | Number of Benchmarks | Number of Benchmarks Available for Assessment | Total Items per Form |
|------------------|-------------------------------|----------------------|---|----------------------|
| 5                | 3                             | 26                   | 15  | 35                   |
| 6                | 3                             | 26                   | 13  | 35                   |
| 7                | 3                             | 27                   | 11  | 35                   |
| 8                | 3                             | 33                   | 11  | 35                   |
| 10               | 3                             | 29                   | 13  | 35                   |

To determine how many of these benchmarks were matched to items, we first computed the frequency of benchmarks covered (per strand) separately for each panelist. Next, we calculated the mean number of benchmarks linked with items across panelists. Table 3.7 summarizes the range-of-knowledge results for each grade level of the Reading MCA-MOD per content strand.

**Table 3.7. Summary of Range-of-Knowledge Results for Reading MCA-MOD**

| Grade | Percentage of Benchmarks per Sub-Strand Matched to at Least One Item |               |            | Number of Sub-Strands Assessed Adequately | Specific Sub-Strands Assessed Inadequately |
|-------|--|---------------|------------|---|--|
|       | Vocabulary Expansion   | Comprehension | Literature |   |  |
| 5     | 80   | 69            | <b>45</b>  | 2 of 3                                    | Literature                                 |
| 6     | 88   | 59            | 100        | 3 of 3                                    |  |
| 7     | 100  | 100           | 70         | 3 of 3                                    |  |
| 8     | 80   | 72            | <b>40</b>  | 2 of 3                                    | Literature                                 |
| 10    | 100  | 50            | <b>43</b>  | 2 of 3                                    | Literature                                 |

For each grade assessment, nearly all benchmarks were represented by items for each of the content strands. In grade 5, 8, and 10 the minimum range-of-knowledge criterion of 50% was not met on the Literature strand.

***Balance-of-Knowledge Representation***

The fourth measure of alignment included in the Webb method is *balance-of-knowledge representation*. This measure describes the distribution of items linked to each benchmark within

each strand. The number of items should be distributed rather evenly between the benchmarks to achieve good balance.

The content balance is determined by calculating an index, or score, for each strand<sup>3</sup>. According to Webb, the minimum acceptable index for a single strand is 70 (on a scale of 0 to 100 with 100 representing perfect balance). An index of 70 or higher suggests that items broadly assess the benchmarks for a strand instead of clustering around one or two benchmarks.

One caution should be noted regarding the balance index when interpreting the results. Only those benchmarks actually matched to items by the panelists are included in calculations of the balance index. A given strand may include more benchmarks than are actually linked to items by panelists. For example, if a particular strand includes eight benchmarks in the state content standards document but panelists found items matching to just three benchmarks, only these three benchmarks are evaluated for item distribution. Recognizing this feature of the balance index is important in cases when the range measure and balance measure produce seemingly contrasting results.

Table 3.8 summarizes the results on balance-of-content representation per grade for the Reading MCA-MOD. All of the grades assessed surpassed the minimum level of acceptability (index of 70) for demonstrating good content balance among those benchmarks matched to items for two of the three content sub-strands. Grade 7 was the only grade that surpassed the minimum level of acceptability for all content sub-strands.

***Table 3.8. Summary of Balance-of-Knowledge Representation Results for Reading MCA-MOD***

| Grade | Balance Index per Sub-Strand |               |            | Sub-Strands with Adequate Balance | Sub-Strands with Limited Balance |
|-------|------------------------------|---------------|------------|-----------------------------------|----------------------------------|
|       | Vocabulary Expansion         | Comprehension | Literature |                                   |                                  |
| 5     | 89                           | <b>56</b>     | 100        | 2 of 3                            | Comprehension                    |
| 6     | 82                           | <b>59</b>     | 74         | 2 of 3                            | Comprehension                    |
| 7     | 74                           | 70            | 79         | 3 of 3                            | None                             |
| 8     | 89                           | <b>61</b>     | 97         | 2 of 3                            | Comprehension                    |
| 10    | 100                          | <b>62</b>     | 90         | 2 of 3                            | Comprehension                    |

### ***Summary and Discussion on Webb Alignment Indicators***

The content alignment review of the Reading MCA-MOD evaluated the operational items compared to the Minnesota Academic Standards for grades 5 – 8 and 10. A test form for a

<sup>3</sup> The exact formula for calculating the balance index is explained in detail in Webb’s (2005) alignment training manual: <http://www.wcer.wisc.edu/WAT/index.aspx>.

given yearly administration should be representative of the full set of items in the pool, and, thus, should align appropriately to the content expectations. Alignment of large-scale assessments to state content standards is a requirement of the No Child Left Behind Act of 2001.

HumRRO applied the Webb alignment method to conduct the review. The overall alignment results for the Reading MCA-MOD were mixed. At grade 6, the assessment met, to the full extent, the minimum requirements for all of the Webb indicators while three of the four Webb indicators were met in grade 7. All other grades (5, 8, and 10) did not meet the minimum requirements on any of the Webb indicators. Results on alignment indicators, such as depth-of-knowledge consistency and categorical concurrence, suggest that some items represent the benchmarks in a more limited way than expected. We present summary alignment judgments for the Reading MCA-MOD in this section based on the statistical outcomes.

Summary alignment judgments are based on Webb (2005). These summary judgments focus on the percentage of content strands represented well by the assessment. Webb outlined a scale with a range of potential alignment outcomes applied to each of the four indicators:

- Fully aligned – assessments align to all content strands (91%–100%);
- Highly aligned – assessments align to the majority of strands (70%–90%)
- Partially aligned – assessments align well to some strands (50%–69%)
- Weakly aligned – assessments align to less than half the strands (below 50%).

Webb’s alignment method does not allow for a *single* judgment of overall alignment across the four alignment indicators. However, one can get a sense of overall alignment between the assessments and standards by looking at all of the alignment indicators together.

Table 3.9 presents the summary alignment outcomes for the Reading MCA-MOD based on the above scale. The table includes a summary judgment for each Webb alignment indicator per grade assessment based on the percentage of strands that met the minimum alignment criteria. This summary table is linked to the bottom row of each of Tables A-1 through A-12 in Appendix A. Thus, these summary judgments reflect a final evaluation of each grade assessment per Webb criteria *across* the strands.

As shown in Table 3.9 with green highlighting, several outcomes point to strong content alignment of the Reading MCA-MOD to the Minnesota Academic Standards. For grades 6 and 7, range-of-knowledge and balance-of-knowledge representation results suggest that items assess all available benchmarks and items seem to be distributed reasonably, at least across benchmarks matched by panelists. The results also suggest that a sufficient number of operational items are used to cover the sub-content areas. Finally, panelists’ indicated that for grade 6 the items corresponded with the DOK levels of the benchmarks.

The grade 5, 8, and 10 assessments demonstrated lower levels of alignment to the content standards on all of the Webb criteria. Table 3.9 highlights those results showing partial alignment to the content standards, indicated by the yellow highlighting. For each grade, these results are indicative of the minimum criteria being missed on one of the three content sub-strands.

**Table 3.9. Summary Alignment Outcomes on Each Webb Criterion per Grade Level for Reading MCA-MOD**

| Grade Assessment | Percentage of Strands that Met Webb Criteria |                                |                                   |                                     |
|------------------|--|--------------------------------|-----------------------------------|-------------------------------------|
|                  | Categorical Concurrence                      | Depth-of-Knowledge Consistency | Range-of-Knowledge Correspondence | Balance-of-Knowledge Representation |
| 5                | Partially aligned (67%)                      | Partially aligned (67%)        | Partially aligned (67%)           | Partially aligned (67%)             |
| 6                | Fully aligned (100%)                         | Fully aligned (100%)           | Fully aligned (100%)              | Fully aligned (100%)                |
| 7                | Fully aligned (100%)                         | Partially aligned (67%)        | Fully aligned (100%)              | Fully aligned (100%)                |
| 8                | Partially aligned (67%)                      | Partially aligned (67%)        | Partially aligned (67%)           | Partially aligned (67%)             |
| 10               | Partially aligned (67%)                      | Partially aligned (67%)        | Partially aligned (67%)           | Partially aligned (67%)             |

Panelists reviewing the grade 5 assessment found that the Literature sub-strand did not contain an adequate number of items to assess the content nor were there enough benchmarks assessed from the sub-strand. For the Comprehension sub-strand in grade 5, panelists found that a good distribution of items across the benchmarks was missing and that less than half of the items assessed students at an appropriate cognitive complexity level. This same pattern of results was seen by panelists evaluating the grade 10 assessment.

Additionally, panelists reviewing the grade 8 assessment found that for the Comprehension sub-strand a good distribution of items across the benchmarks was missing. For the Literature sub-strand at grade 8, panelists found that the assessment did not contain an adequate number of items to assess the content nor were there enough benchmarks assessed from the sub-strand. Finally, roughly one-third of the items did not assess students at the appropriate cognitive complexity level.

Suggestions for improving the alignment between the Reading assessments and Minnesota Academic Standards are discussed in Chapter 5 Summary and Recommendations.

## Chapter 4. Results: Test Quality of Reading MCA-MOD

In this chapter, we report the results of panelists’ evaluations of test quality. Alignment of assessments to the state content standards serves as one form of test validity evidence. Other areas of validity are critical as well, such as whether the assessment enables students to demonstrate what they know. For example, are test items free of biases, clear in language, and appropriate for the grade level?

All assessments should “be designed from the beginning to be accessible and valid with respect to the widest possible range of students, including students with disabilities and students with limited English proficiency” (NCLB, 2001, Section 200.2(b)(2)). The Reading MCA-MOD underwent bias reviews as part of the item development process; however, review of quality and accessibility by an independent evaluator provides further evidence of a fair process and assessment. This evaluation of test quality for the Reading MCA-MOD represented a broad review of student access to test content.

Panelists evaluated the Reading MCA-MOD on several dimensions at the item level and across each grade test as a whole. Item ratings included review of written content and figures or graphics, and were based on simple yes-no evaluations of item quality. Panelists also made “overall item quality” ratings with annotations to report the rationale for their ratings. Finally, panelists in each grade span group made ratings on specific aspects of the test as a whole. Results reported in this section include those for operational items from the 2011 MCA-MOD.

Panelists made their content alignment ratings based on a print version of screen shots for the two online 2011 test forms and a print version of the intact paper and pencil form.

### *Written Content*

Panelists rated the language used in the items for the extent to which students of various backgrounds and ability levels could access the Reading content. Ratings consisted of ‘yes’ or ‘no’ responses. Table 4.1 below indicates the mean number of items per grade test rated as accessible or not. As the table demonstrates, the majority of items were rated favorably on accessibility.

**Table 4.1. Mean Number of Items Rated as Accessible in Content to Range of Students per Grade Assessment**

| Is item content accessible to the range of students who take the assessment? |                      |      |                      |      |
|--|----------------------|------|----------------------|------|
| Grade  | Yes                  |      | No                   |      |
|  | Mean number of items | SD   | Mean number of items | SD   |
| 5  | 33.00                | 1.41 | 0.00                 | NA   |
| 6  | 33.00                | 1.41 | 0.00                 | NA   |
| 7  | 29.00                | NA   | 0.00                 | NA   |
| 8  | 24.00                | 3.08 | 2.67                 | 1.53 |
| 10   | 28.20                | 6.22 | 4.00                 | 1.41 |

If panelists responded ‘no’, we asked them to provide an explanation of their responses. Most comments pertained to confusing language in items such as unclear cue reference, use of idiomatic language (i.e., ill afford), and outdated references.

### ***Overall Item Quality***

In addition to rating items on accessibility, panelists had the opportunity to give items a general rating reflecting their judgments of quality. This rating encompassed aspects such as clarity (e.g., wording or item scenario, prompt, or response options) and appropriateness (e.g., off-grade, exceeds benchmark).

- Poor quality - item exhibits serious flaw; recommend replacement.
- Fair quality - item exhibits minor but repairable flaw.
- Good quality - item exhibits no real flaws and is typical for this type of assessment.
- Exceptional quality - item is exemplary for this type of assessment.

Table 4.3 displays the mean ratings on overall item quality per grade assessment. As the table illustrates, panelists considered the vast majority of items to be ‘good’ to ‘exceptional’ in quality.

**Table 4.3. Panelist Ratings on Overall Item Quality**

| Grade | Item Quality | Mean Number of Items per Level | SD <sup>a</sup> | Percentage of Items per Level |
|-------|--------------|--------------------------------|-----------------|-------------------------------|
| 5     | Poor         | 5.00                           | NA              | 3                             |
|       | Fair         | 6.00                           | 2.65            | 11                            |
|       | Good         | 15.80                          | 7.92            | 49                            |
|       | Exceptional  | 14.50                          | 6.61            | 35                            |
| 6     | Poor         | 0.00                           | NA              | 0                             |
|       | Fair         | 2.00                           | 1.41            | 3                             |
|       | Good         | 23.50                          | 10.47           | 71                            |
|       | Exceptional  | 17.00                          | 9.90            | 26                            |
| 7     | Poor         | 0.00                           | NA              | 0                             |
|       | Fair         | 1.00                           | 0.00            | 2                             |
|       | Good         | 21.75                          | 9.18            | 78                            |
|       | Exceptional  | 11.00                          | 12.73           | 20                            |
| 8     | Poor         | 1.00                           | NA              | 1                             |
|       | Fair         | 3.25                           | 3.86            | 10                            |
|       | Good         | 16.40                          | 7.96            | 64                            |
|       | Exceptional  | 10.67                          | 7.02            | 25                            |
| 10    | Poor         | 1.00                           | 0.00            | 1                             |
|       | Fair         | 6.00                           | 5.66            | 8                             |
|       | Good         | 21.00                          | 8.86            | 70                            |
|       | Exceptional  | 7.50                           | 4.80            | 20                            |

<sup>a</sup>Not all panelists rated an item at each item quality level.

For those items rated as ‘fair’ or ‘poor’ in quality, we asked panelists to provide comments to identify the issue and suggest improvements. Many items falling into these categories received comments regarding clarity or complexity. For example, the word ‘horses’ is used in the item stem yet the word ‘ponies’ is used in the distractors. Several comments stated that students may be able to answer the question without needing to read the passage. Still other comments noted that prior knowledge is needed to answer the question correctly. There were a couple of items that panelists noted as having more than one correct answer. Notations for other items suggested that, while the item aligned to the benchmark overall, the expectations for students to respond to the item exceeded the content expectations of the benchmarks (i.e., benchmark asks students to ‘generate a question’ while the item did not). Overall, the majority of the panelists’ ratings on overall item quality was at the ‘Good’ or ‘Exceptional’ level.

### ***Whole Test Evaluation***

At the end of the review session, after panelists had completed all other independent ratings, panelists reviewed the test as a whole to provide more global perspectives on the ability of students to demonstrate their knowledge on the assessment.

The whole test review included five questions to guide panelists’ evaluations. The panelists were expected to generate written conclusions for each question at a global level based

on the independent item ratings just completed. Table 4.4 and 4.5 present these questions, along with the responses given by each grade-span group. Even though these whole test evaluations were independent, panelists were in agreement that the Reading MCA-MOD are accessible and appropriate for Minnesota students. Some comments point to particular features of the assessments that they considered to be particularly positive or negative.

***Table 4.4. Grade 5 – 7 Reading MCA-MOD: Consensus Ratings on Whole Test Evaluation***

| Guiding Questions   | Overall Evaluation<br><br>(Yes = mostly to all,<br>No = somewhat to<br>none) | Comments Supporting Ratings   |
|---|--|---|
| Is the computer administered assessment format effective for this population of students? | No   | Make sure that questions are located as closely as possible to the selection to reduce unnecessary eye movements for these students. Poetry and other items are difficult to go back and forth for the ADHD student. Poetry sections may cause a problem with scrolling from one screen to another to answer questions. |
| Is language clear and appropriate for a Reading test?                                     | Yes  |   |
| Are graphics used clear and appropriate?  | NA   |   |
| Is the level of language proficiency expected by test items appropriate?                  | Yes  |   |
| Is this assessment accessible to all students who will take it?                           | Yes  |   |

**Table 4.5. Grade 8 and 10 Reading MCA-MOD: Consensus Ratings on Whole Test Evaluation**

| Guiding Questions   | Overall Evaluation<br>(Yes = mostly to all,<br>No = somewhat to<br>none) | Comments Supporting Ratings  |
|---|--|--|
| Is the computer administered assessment format effective for this population of students? | Yes  | The computer format has been helpful for most students. As long as students are aware ahead of time and practice using an online test.   |
| Is language clear and appropriate for a Reading test?                                     | No   | Many of the passages used obscure or unconventional language. The language and topics are presented as somewhat neutral in topic although those topics may provide an unfair advantage to rural students. The language seems to be at the correct level.   |
| Are graphics used clear and appropriate?  | NA   |  |
| Is the level of language proficiency expected by test items appropriate?                  | Yes  | There were a lot of parts of the test not only in the passages but also the items and answer choices that were at a level that was unnecessarily difficult. True only for students from language rich environments. Some of the language is archaic and may not be understandable to all students.                                       |
| Is this assessment accessible to all students who will take it?                           | No   | Because the items had many references to rural settings and cultures that are all white, if culturally biases those test takers. The material will be very difficult for ELL students, especially considering the content areas are largely unfamiliar for many students. Students familiar with rural settings would have an advantage. |

### ***Summary and Discussion of Test Quality Results***

The results of the test quality review by panelists suggest that the Reading MCA-MOD allow a wide range of students the opportunity to demonstrate their knowledge of Reading. The majority of items received positive ratings by panelists, and global judgments about test quality also emphasized this point.

Table 4.6 presents the summary outcomes on the item quality ratings. The table includes conclusions regarding the quality of the items on each assessment, along with the percentage of

items that received favorable ratings. These conclusions are based on the following decision criteria (adapted from Thompson, Johnstone, Anderson, & Miller, 2005).

- Excellent – all items are acceptable;
- Good – most items are acceptable (at least 90%);
- Acceptable -- many items are acceptable (70%-89%);
- Questionable – few items are acceptable (less than 70%).

**Table 4.6. Item Quality Ratings for Reading MCA-MOD per Grade Assessment**

| Grade | Percentage of Items with Acceptable Ratings |                      |
|-------|---|----------------------|
|       | Written Content                             | Overall Item Quality |
| 5     | Excellent<br>(100%)                         | Acceptable<br>(84%)  |
| 6     | Excellent<br>(100%)                         | Good<br>(97%)        |
| 7     | Excellent<br>(100%)                         | Good<br>(98%)        |
| 8     | Acceptable<br>(89%)                         | Acceptable<br>(89%)  |
| 10    | Acceptable<br>(88%)                         | Good<br>(90%)        |

Table 4.6 shows that none of the grade assessments included enough items with low ratings on any dimension to warrant a conclusion of questionable quality. However, each assessment included *some* items with low ratings (and corresponding annotations highlighting possible issues), as demonstrated by findings of ‘acceptable’ quality (70%-90% of items). Panelists for the grades 8 and 10 assessment in particular commented on a number of items that used idiomatic language as well as confusing and complex language. For this reason, at least those items with low ratings could be reviewed for improvement. For the grades 5 and 8 assessment, panelists commented on the overall item quality by noting that the majority of the items were ‘good’, but there was a percentage of items rated as ‘fair’, showing minor but repairable flaw.

As a whole, the independent item ratings, along with whole test reviews, suggest that the MCA-MOD function well for the majority of students who take these assessments. A small number of items on each grade assessment may require review to enhance clarity in wording or in accompanying graphics and reduce potential bias against particular student groups.

## **Chapter 5. Summary and Recommendations**

HumRRO conducted a review of the Reading MCA-MOD to examine: (a) content alignment to the Minnesota Academic Standards for Reading and (b) accessibility for all students who take these assessments. Alignment of assessments and achievement standards to the state academic content standards is a requirement of the No Child Left Behind Act of 2001.

The cumulative results provide reasonable evidence for content validity of the Reading MCA-MOD. Concerning content alignment, each assessment clearly covers the content categories specified in the Minnesota Academic Standards for Reading. Concerning accessibility, panelists determined that the majority of items are appropriate for a wide range of students.

As with most reviews of state assessment systems, these findings point to areas in which Minnesota could strengthen the alignment between the assessments and the content standards. For this reason, HumRRO makes the following recommendations to Minnesota on ways in which alignment might be improved. These recommendations focus on the more critical findings:

- ***Review the cognitive complexity (depth of knowledge) for items on the Grade 5, 7, 8, and 10 assessments.*** The panelists reviewing these assessments rated a number of items as less demanding cognitively than the Minnesota Academic Standards. Thus, the assessments may not adequately reflect the rigor of the content expectations. Finding a disproportionate number of items assessing more basic cognitive skills is not uncommon among large-scale assessments. However, such a circumstance also is not an inevitable consequence of standardized testing, particularly for Reading assessments. Given the outcomes on depth-of-knowledge ratings along with the accessibility outcomes, it is likely that increasing the complexity of the assessment would involve modifications to current operational items, rather than item replacement, because no items were rated as seriously flawed.
- ***Review the categorical concurrence (strand coverage) for items on the Grade 5, 8, and 10 assessments.*** Webb’s criterion of a minimum of six items per content strand was missed on three of the grade assessments based on panelists’ reviews. The assessment at grades 5, 8 and 10 needs to be reviewed to ensure that an adequate number of items are being assessed at each sub-strand level.
- ***Review the range of knowledge and balance of knowledge on the Grade 5, 8, and 10 assessments.*** These two Webb criteria examine the breadth of knowledge covered by the assessment and the distribution of items at the benchmark level within a content area. Missing these two criteria indicate that not all of the sub-strand content areas are being adequately assessed by the items and the items are not distributed evenly among the benchmarks being assessed within a content area. Increasing the number of items chosen to assess a content area and ensuring that items are adequately assessing all of the benchmarks instead of some and not others should provide a better range and balance of assessed content.
- ***Review those items that received the lowest ratings on test quality for possible revision.*** As noted in Recommendation 1, no items were rated as seriously flawed or requiring replacement. However, panelists found a small number of items for

each grade test that could benefit from review to increase clarity in language. There were a couple of items in grade 10 that reviewers noted as having, potentially, more than one correct answer.

## References

- Bloom, B., Englehart, M. Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, Toronto: Longmans, Green.
- No Child Left Behind Act of 2001. Public Law 107-110.
- Thompson, S.J., Johnstone, C.J., Anderson, M. E., & Miller, N. A. (2005). *Considerations for the development and review of universally designed assessments* (Technical Report 42). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- U.S. Department of Education. (April, 2004). *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education. Retrieved from <http://www.ed.gov/policy/elsec/guid/saaprguidance.doc>.
- Webb, N. L. (2005). *Webb alignment tool: Training manual*. Madison, WI: Wisconsin Center for Education Research. Available: <http://www.wcer.wisc.edu/WAT/index.aspx>.
- Webb, N. L. (1997). *Research Monograph No. 6: Criteria for alignment of expectations and assessments in Reading and Reading education*. Washington, D.C.: Council of Chief State Schools Officers.
- Webb, N. L. (1999). *Alignment of reading and reading standards and assessments in four states. (Research Monograph 18)*. Madison, WI: National Institute for Reading Education and Council of Chief State School Officers. (ERIC Document Reproduction Service No. ED440852)

## Appendix A. Content Alignment Results per Grade Level Assessment

The following tables include complete statistical results on the Webb alignment indicators, including means and standard deviations per strand for each grade Reading MCA-MOD test.

### *Categorical Concurrence*

The categorical concurrence results for grades 5 – 8 and 10 of the Reading MCA-MOD are presented below. Each table includes: the target number of items from the test blueprint; the mean number of items matched by panelists; the standard deviation among panelists’ ratings; and, the final alignment conclusion (Yes or No). The bottom row indicates the percentage of strands that met the minimum alignment criterion. Note that the total mean items matched may exceed the number of items on the assessment, as raters were able to match items to more than one strand.

***Table A-1. Categorical Concurrence for Reading MCA-MOD, Grade 5: Mean Number of Items per Sub-Strand***

|   |                               | Number of Items per Sub-Strand |                    |                                   |
|---|-------------------------------|--------------------------------|--------------------|-----------------------------------|
| Title of Sub-Strand                               | Target # Items from Blueprint | Mean Items Matched             | Standard Deviation | At Least Six Items per Sub-Strand |
| Vocabulary Expansion                              | 6-8                           | 7.20                           | 1.64               | Y                                 |
| Comprehension                                     | 14-17                         | 28.00                          | 4.64               | Y                                 |
| Literature  | 12-15                         | 3.40                           | 2.41               | N                                 |
| Total   | 35                            |                                |                    |                                   |
| Percentage of sub-strands with at least six items |                               |                                |                    | 67%                               |

***Table A-2. Categorical Concurrence for Reading MCA-MOD, Grade 6: Mean Number of Items per Sub-Strand***

|   |                               | Number of Items per Sub-Strand |                    |                                   |
|---|-------------------------------|--------------------------------|--------------------|-----------------------------------|
| Title of Sub-Strand                               | Target # Items from Blueprint | Mean Items Matched             | Standard Deviation | At Least Six Items per Sub-Strand |
| Vocabulary Expansion                              | 6-8                           | 9.50                           | 1.73               | Y                                 |
| Comprehension                                     | 14-17                         | 25.75                          | 6.99               | Y                                 |
| Literature  | 12-15                         | 8.00                           | 1.73               | Y                                 |
| Total   | 35                            |                                |                    |                                   |
| Percentage of sub-strands with at least six items |                               |                                |                    | 100%                              |

**Table A-3. Categorical Concurrence for Reading MCA-MOD, Grade 7: Mean Number of Items per Sub-Strand**

|   |                               | Number of Items per Sub-Strand |                    |                                   |
|---|-------------------------------|--------------------------------|--------------------|-----------------------------------|
| Title of Sub-Strand                               | Target # Items from Blueprint | Mean Items Matched             | Standard Deviation | At Least Six Items per Sub-Strand |
| Vocabulary Expansion                              | 6-8                           | 6.75                           | 1.89               | Y                                 |
| Comprehension                                     | 16-19                         | 20.00                          | 5.29               | Y                                 |
| Literature  | 11-14                         | 11.00                          | 6.38               | Y                                 |
| Total   | 35                            |                                |                    |                                   |
| Percentage of sub-strands with at least six items |                               |                                |                    | 100%                              |

**Table A-4. Categorical Concurrence for Reading MCA-MOD, Grade 8: Mean Number of Items per Sub-Strand**

|   |                               | Number of Items per Sub-Strand |                    |                                   |
|---|-------------------------------|--------------------------------|--------------------|-----------------------------------|
| Title of Sub-Strand                               | Target # Items from Blueprint | Mean Items Matched             | Standard Deviation | At Least Six Items per Sub-Strand |
| Vocabulary Expansion                              | 6-8                           | 6.80                           | 1.64               | Y                                 |
| Comprehension                                     | 18-21                         | 26.80                          | 2.59               | Y                                 |
| Literature  | 9-12                          | 3.40                           | 1.52               | N                                 |
| Total   | 35                            |                                |                    |                                   |
| Percentage of sub-strands with at least six items |                               |                                |                    | 67%                               |

**Table A-5. Categorical Concurrence for Reading MCA-MOD, Grade 10: Mean Number of Items per Sub-Strand**

|   |                               | Number of Items per Sub-Strand |                    |                                   |
|---|-------------------------------|--------------------------------|--------------------|-----------------------------------|
| Title of Sub-Strand                               | Target # Items from Blueprint | Mean Items Matched             | Standard Deviation | At Least Six Items per Sub-Strand |
| Vocabulary Expansion                              | 6-8                           | 8.00                           | 3.16               | Y                                 |
| Comprehension                                     | 18-21                         | 24.80                          | 3.27               | Y                                 |
| Literature  | 9-12                          | 3.20                           | 1.30               | N                                 |
| Total   | 35                            |                                |                    |                                   |
| Percentage of sub-strands with at least six items |                               |                                |                    | 67%                               |

### ***Depth-of-Knowledge Consistency***

The Depth-of-Knowledge (DOK) consistency results for grades 5 – 8 and 10 of the Reading MCA-MOD are presented below. The tables present the results from the comparison between the depth-of-knowledge expected in the content benchmarks and the depth-of-knowledge assessed by items. The tables include the mean percentage of items rated as below, at the same level, or above the DOK level of the benchmarks along with the corresponding standard deviations. Benchmarks with at least 50% of items at the same (or above) DOK level met the minimum criterion.

***Table A-6. DOK Consistency for Reading MCA-MOD, Grade 5: Mean Percent of Items with DOK Below, At, and Above DOK Level of Benchmarks***

| Title of Sub-Strand  | Mean Items per Sub-Strand | Depth-of-Knowledge Consistency |      |                    |      |               |      | DOK Consistency Target Met |
|--|---------------------------|--------------------------------|------|--------------------|------|---------------|------|----------------------------|
|  |                           | % Items Below                  |      | % Items Same Level |      | % Items Above |      |                            |
|  |                           | M                              | S.D. | M                  | S.D. | M             | S.D. |                            |
| Vocabulary Expansion   | 6.0                       | 18.8                           | 23.4 | 44.3               | 11.7 | 37.0          | 29.1 | Y                          |
| Comprehension  | 27.8                      | 58.6                           | 15.3 | 34.7               | 19.9 | 6.7           | 8.0  | N                          |
| Literature   | 1.8                       | 26.7                           | 43.5 | 50.0               | 50.0 | 23.3          | 32.5 | Y                          |
| Percentage of strands with 50% of item DOK at or above objective DOK |                           |                                |      |                    |      |               |      | 67%                        |

***Table A-7. DOK Consistency for Reading MCA-MOD, Grade 6: Mean Percent of Items with DOK Below, At, and Above DOK Level of Benchmarks***

| Title of Sub-Strand  | Mean Items per Sub-Strand | Depth-of-Knowledge Consistency |      |                    |      |               |      | DOK Consistency Target Met |
|--|---------------------------|--------------------------------|------|--------------------|------|---------------|------|----------------------------|
|  |                           | % Items Below                  |      | % Items Same Level |      | % Items Above |      |                            |
|  |                           | M                              | S.D. | M                  | S.D. | M             | S.D. |                            |
| Vocabulary Expansion   | 8.3                       | 20.1                           | 30.1 | 69.5               | 26.2 | 10.4          | 12.5 | Y                          |
| Comprehension  | 25.5                      | 43.8                           | 18.3 | 45.6               | 15.8 | 10.6          | 10.8 | Y                          |
| Literature   | 6.7                       | 13.3                           | 23.1 | 44.8               | 5.0  | 41.9          | 19.1 | Y                          |
| Percentage of strands with 50% of item DOK at or above objective DOK |                           |                                |      |                    |      |               |      | 100%                       |

**Table A-8. DOK Consistency for Reading MCA-MOD, Grade 7: Mean Percent of Items with DOK Below, At, and Above DOK Level of Benchmarks**

| Title of Sub-Strand  | Mean Items per Sub-Strand | Depth-of-Knowledge Consistency |      |                    |      |               |      | DOK Consistency Target Met |
|--|---------------------------|--------------------------------|------|--------------------|------|---------------|------|----------------------------|
|  |                           | % Items Below                  |      | % Items Same Level |      | % Items Above |      |                            |
|  |                           | M                              | S.D. | M                  | S.D. | M             | S.D. |                            |
| Vocabulary Expansion   | 4.8                       | 24.4                           | 39.9 | 58.9               | 40.2 | 16.7          | 33.3 | Y                          |
| Comprehension  | 16.3                      | 57.9                           | 21.1 | 32.7               | 25.9 | 9.4           | 7.2  | N                          |
| Literature   | 10.0                      | 22.0                           | 15.7 | 37.9               | 25.9 | 40.0          | 41.5 | Y                          |
| Percentage of strands with 50% of item DOK at or above objective DOK |                           |                                |      |                    |      |               |      | 67%                        |

**Table A-9. DOK Consistency for Reading MCA-MOD, Grade 8: Mean Percent of Items with DOK Below, At, and Above DOK Level of Benchmarks**

| Title of Sub-Strand  | Mean Items per Sub-Strand | Depth-of-Knowledge Consistency |      |                    |      |               |      | DOK Consistency Target Met |
|--|---------------------------|--------------------------------|------|--------------------|------|---------------|------|----------------------------|
|  |                           | % Items Below                  |      | % Items Same Level |      | % Items Above |      |                            |
|  |                           | M                              | S.D. | M                  | S.D. | M             | S.D. |                            |
| Vocabulary Expansion   | 6.0                       | 17.3                           | 28.9 | 78.7               | 27.2 | 4.0           | 8.9  | Y                          |
| Comprehension  | 18.2                      | 43.3                           | 8.7  | 52.4               | 10.1 | 4.3           | 4.3  | Y                          |
| Literature   | 2.8                       | 68.0                           | 29.5 | 32.0               | 29.5 | 0.0           | 0.0  | N                          |
| Percentage of strands with 50% of item DOK at or above objective DOK |                           |                                |      |                    |      |               |      | 67%                        |

**Table A-10. DOK Consistency for Reading MCA-MOD, Grade 10: Mean Percent of Items with DOK Below, At, and Above DOK Level of Benchmarks**

| Title of Sub-Strand  | Mean Items per Sub-Strand | Depth-of-Knowledge Consistency |      |                    |      |               |      | DOK Consistency Target Met |
|--|---------------------------|--------------------------------|------|--------------------|------|---------------|------|----------------------------|
|  |                           | % Items Below                  |      | % Items Same Level |      | % Items Above |      |                            |
|  |                           | M                              | S.D. | M                  | S.D. | M             | S.D. |                            |
| Vocabulary Expansion   | 7.4                       | 14.0                           | 31.3 | 79.6               | 34.7 | 6.4           | 9.8  | Y                          |
| Comprehension  | 20.0                      | 55.7                           | 20.2 | 34.0               | 20.7 | 10.3          | 7.8  | N                          |
| Literature   | 3.2                       | 16.7                           | 15.6 | 83.3               | 15.6 | 0.0           | 0.0  | Y                          |
| Percentage of strands with 50% of item DOK at or above objective DOK |                           |                                |      |                    |      |               |      | 67%                        |

***Range-of-Knowledge Correspondence***

The results for Range-of-Knowledge correspondence for grades 5 – 8 and 10 for the Reading MCA-MOD are presented below. The tables include the mean number, standard deviation, and percentage of benchmarks by content strand. For acceptable range-of-knowledge correspondence, a minimum of 50% of content benchmarks within each strand should be matched to at least one item.

**Table A-11. Range-of-Knowledge for Reading MCA-MOD, Grade 5: Mean Percent of Benchmarks per Sub-Strand Linked with Items**

| Title of Sub-Strand  | Number of Benchmarks | Mean Items per Sub-Strand | Range of Benchmarks               |      |                                      | Range-of-Knowledge Target Met |
|--|----------------------|---------------------------|-----------------------------------|------|--------------------------------------|-------------------------------|
|  |                      |                           | Benchmarks with At Least One Item |      | % of Total Benchmarks per Sub-Strand |                               |
|  |                      |                           | M                                 | S.D. |                                      |                               |
| Vocabulary Expansion   | 2                    | 6.0                       | 1.6                               | 0.5  | 80                                   | Y                             |
| Comprehension  | 9                    | 27.8                      | 6.2                               | 0.8  | 69                                   | Y                             |
| Literature   | 4                    | 1.8                       | 1.8                               | 0.8  | 45                                   | N                             |
| Total  | 15                   |                           |                                   |      |                                      |                               |
| Percentage of strands with 50% of Benchmarks linked to at least one item |                      |                           |                                   |      |                                      | 67%                           |

**Table A-12. Range-of-Knowledge for Reading MCA-MOD, Grade 6: Mean Percent of Benchmarks per Sub-Strand Linked with Items**

|  |                      |                           | Range of Benchmarks               |      |                                      | Range-of-Knowledge Target Met |
|--|----------------------|---------------------------|-----------------------------------|------|--------------------------------------|-------------------------------|
| Title of Sub-Strand  | Number of Benchmarks | Mean Items per Sub-Strand | Benchmarks with At Least One Item |      | % of Total Benchmarks per Sub-Strand |                               |
|  |                      |                           | M                                 | S.D. |                                      |                               |
| Vocabulary Expansion   | 2                    | 8.3                       | 1.8                               | 0.5  | 88                                   | Y                             |
| Comprehension  | 8                    | 25.5                      | 4.8                               | 1.0  | 59                                   | Y                             |
| Literature   | 3                    | 6.7                       | 3.0                               | 0.0  | 100                                  | Y                             |
| Total  | 13                   |                           |                                   |      |                                      |                               |
| Percentage of strands with 50% of Benchmarks linked to at least one item |                      |                           |                                   |      |                                      | 100%                          |

**Table A-13. Range-of-Knowledge for Reading MCA-MOD, Grade 7: Mean Percent of Benchmarks per Sub-Strand Linked with Items**

|  |                      |                           | Range of Benchmarks               |      |                                      | Range-of-Knowledge Target Met |
|--|----------------------|---------------------------|-----------------------------------|------|--------------------------------------|-------------------------------|
| Title of Sub-Strand  | Number of Benchmarks | Mean Items per Sub-Strand | Benchmarks with At Least One Item |      | % of Total Benchmarks per Sub-Strand |                               |
|  |                      |                           | M                                 | S.D. |                                      |                               |
| Vocabulary Expansion   | 2                    | 4.8                       | 2.0                               | 0.0  | 100                                  | Y                             |
| Comprehension  | 4                    | 16.3                      | 4.0                               | 0.0  | 100                                  | Y                             |
| Literature   | 5                    | 10.0                      | 3.5                               | 1.0  | 70                                   | Y                             |
| Total  | 11                   |                           |                                   |      |                                      |                               |
| Percentage of strands with 50% of Benchmarks linked to at least one item |                      |                           |                                   |      |                                      | 100%                          |

**Table A-14. Range-of-Knowledge for Reading MCA-MOD, Grade 8: Mean Percent of Benchmarks per Sub-Strand Linked with Items**

|  |                      |                           | Range of Benchmarks               |      |                                      | Range-of-Knowledge Target Met |
|--|----------------------|---------------------------|-----------------------------------|------|--------------------------------------|-------------------------------|
| Title of Sub-Strand  | Number of Benchmarks | Mean Items per Sub-Strand | Benchmarks with At Least One Item |      | % of Total Benchmarks per Sub-Strand |                               |
|  |                      |                           | M                                 | S.D. |                                      |                               |
| Vocabulary Expansion   | 2                    | 6.0                       | 1.6                               | 0.5  | 80                                   | Y                             |
| Comprehension  | 5                    | 18.2                      | 3.6                               | 0.5  | 72                                   | Y                             |
| Literature   | 4                    | 2.8                       | 1.6                               | 0.9  | 40                                   | N                             |
| Total  | 11                   |                           |                                   |      |                                      |                               |
| Percentage of strands with 50% of Benchmarks linked to at least one item |                      |                           |                                   |      |                                      | 67%                           |

**Table A-15. Range-of-Knowledge for Reading MCA-MOD, Grade 10: Mean Percent of Benchmarks per Sub-Strand Linked with Items**

|  |                      |                           | Range of Benchmarks               |      |                                      | Range-of-Knowledge Target Met |
|--|----------------------|---------------------------|-----------------------------------|------|--------------------------------------|-------------------------------|
| Title of Sub-Strand  | Number of Benchmarks | Mean Items per Sub-Strand | Benchmarks with At Least One Item |      | % of Total Benchmarks per Sub-Strand |                               |
|  |                      |                           | M                                 | S.D. |                                      |                               |
| Vocabulary Expansion   | 1                    | 7.4                       | 1.0                               | 0.0  | 100                                  | Y                             |
| Comprehension  | 6                    | 20.0                      | 3.0                               | 1.0  | 50                                   | Y                             |
| Literature   | 6                    | 3.2                       | 2.6                               | 0.9  | 43                                   | N                             |
| Total  | 13                   |                           |                                   |      |                                      |                               |
| Percentage of strands with 50% of Benchmarks linked to at least one item |                      |                           |                                   |      |                                      | 67%                           |

### ***Balance-of-Knowledge Representation***

The results for Balance-of-Knowledge representation for grades 5 – 8 and 10 of the Reading MCA-MOD are presented below. The tables also include the percentage of items linked to each strand. The minimum acceptable balance index is 70 out of 100.

**Table A-16. Balance-of-Knowledge Representation for Reading MCA-MOD, Grade 5: Mean Balance Index per Sub-Strand**

| Title of Sub-Strand   | Balance-of-Knowledge Representation |                                   |                           |   |                    |      |                          |
|---|-------------------------------------|-----------------------------------|---------------------------|---|--------------------|------|--------------------------|
|   | Benchmarks per Sub-Strand           | Mean Benchmarks Linked with Items | Mean Items per Sub-Strand | Mean % of Items (of total) Linked to Sub-Strand | Mean Balance Index |      | Balance Index Target Met |
|   |                                     |                                   |                           |   | M                  | S.D. |                          |
| Vocabulary Expansion  | 2                                   | 1.6                               | 6.0                       | 17  | 89                 | 14.7 | Y                        |
| Comprehension   | 9                                   | 6.2                               | 27.8                      | 78  | 56                 | 5.9  | N                        |
| Literature  | 4                                   | 1.8                               | 1.8                       | 5   | 100                | 0.0  | Y                        |
| Total   | 15                                  |                                   |                           |   |                    |      |                          |
| Percentage of standards with a balance of representation index of 70 or greater |                                     |                                   |                           |   |                    |      | 67%                      |

**Table A-17. Balance-of-Knowledge Representation for Reading MCA-MOD, Grade 6: Mean Balance Index per Sub-Strand**

| Title of Sub-Strand   | Balance-of-Knowledge Representation |                                   |                           |   |                    |      |                          |
|---|-------------------------------------|-----------------------------------|---------------------------|---|--------------------|------|--------------------------|
|   | Benchmarks per Sub-Strand           | Mean Benchmarks Linked with Items | Mean Items per Sub-Strand | Mean % of Items (of total) Linked to Sub-Strand | Mean Balance Index |      | Balance Index Target Met |
|   |                                     |                                   |                           |   | M                  | S.D. |                          |
| Vocabulary Expansion  | 2                                   | 1.8                               | 8.3                       | 22  | 82                 | 15.9 | Y                        |
| Comprehension   | 8                                   | 4.8                               | 25.5                      | 66  | 59                 | 3.3  | N                        |
| Literature  | 3                                   | 3.0                               | 6.7                       | 17  | 74                 | 8.4  | Y                        |
| Total   | 13                                  |                                   |                           |   |                    |      |                          |
| Percentage of standards with a balance of representation index of 70 or greater |                                     |                                   |                           |   |                    |      | 67%                      |

**Table A-18. Balance-of-Knowledge Representation for Reading MCA-MOD, Grade 7: Mean Balance Index per Sub-Strand**

| Title of Sub-Strand   | Balance-of-Knowledge Representation |                                   |                           |   |                    |      |                          |
|---|-------------------------------------|-----------------------------------|---------------------------|---|--------------------|------|--------------------------|
|   | Benchmarks per Sub-Strand           | Mean Benchmarks Linked with Items | Mean Items per Sub-Strand | Mean % of Items (of total) Linked to Sub-Strand | Mean Balance Index |      | Balance Index Target Met |
|   |                                     |                                   |                           |   | M                  | S.D. |                          |
| Vocabulary Expansion  | 2                                   | 2.0                               | 4.8                       | 15  | 74                 | 10.4 | Y                        |
| Comprehension   | 4                                   | 4.0                               | 16.3                      | 53  | 70                 | 8.9  | Y                        |
| Literature  | 5                                   | 3.5                               | 10.0                      | 32  | 79                 | 14.8 | Y                        |
| Total   | 11                                  |                                   |                           |   |                    |      |                          |
| Percentage of standards with a balance of representation index of 70 or greater |                                     |                                   |                           |   |                    |      | 100%                     |

**Table A-19. Balance-of-Knowledge Representation for Reading MCA-MOD, Grade 8: Mean Balance Index per Sub-Strand**

| Title of Sub-Strand   | Balance-of-Knowledge Representation |                                   |                           |   |                    |      |                          |
|---|-------------------------------------|-----------------------------------|---------------------------|---|--------------------|------|--------------------------|
|   | Benchmarks per Sub-Strand           | Mean Benchmarks Linked with Items | Mean Items per Sub-Strand | Mean % of Items (of total) Linked to Sub-Strand | Mean Balance Index |      | Balance Index Target Met |
|   |                                     |                                   |                           |   | M                  | S.D. |                          |
| Vocabulary Expansion  | 2                                   | 1.6                               | 6.0                       | 22  | 89                 | 14.7 | Y                        |
| Comprehension   | 5                                   | 3.6                               | 18.2                      | 67  | 61                 | 8.4  | N                        |
| Literature  | 4                                   | 1.6                               | 2.8                       | 10  | 97                 | 7.5  | Y                        |
| Total   | 11                                  |                                   |                           |   |                    |      |                          |
| Percentage of standards with a balance of representation index of 70 or greater |                                     |                                   |                           |   |                    |      | 67%                      |

**Table A-20. Balance-of-Knowledge Representation for Reading MCA-MOD, Grade 10: Mean Balance Index per Sub-Strand**

| Title of Sub-Strand   | Balance-of-Knowledge Representation |                                   |                           |   |                    |      |                          |
|---|-------------------------------------|-----------------------------------|---------------------------|---|--------------------|------|--------------------------|
|   | Benchmarks per Sub-Strand           | Mean Benchmarks Linked with Items | Mean Items per Sub-Strand | Mean % of Items (of total) Linked to Sub-Strand | Mean Balance Index |      | Balance Index Target Met |
|   |                                     |                                   |                           |   | M                  | S.D. |                          |
| Vocabulary Expansion  | 1                                   | 1.0                               | 7.4                       | 23  | 100                | 0.0  | Y                        |
| Comprehension   | 6                                   | 3.0                               | 20.0                      | 66  | 62                 | 3.4  | N                        |
| Literature  | 6                                   | 2.6                               | 3.2                       | 11  | 90                 | 9.1  | Y                        |
| Total   | 13                                  |                                   |                           |   |                    |      |                          |
| Percentage of standards with a balance of representation index of 70 or greater |                                     |                                   |                           |   |                    |      | 67%                      |

***Benchmarks Matched to Items by Panelists***

Tables A-21 through A-25 present the benchmarks, along with mean number of items, matched by panelists. Column 1 includes the Item Codes corresponding to the benchmarks from the MCA-MOD Test Specifications for Reading.

**Table A-21. Grade 5 MCA-MOD: Grade Span Benchmarks Matched to Items by Panelists**

| Benchmark Item Codes | Mean Number of Items per Benchmark | SD        |
|----------------------|------------------------------------|-----------|
| I.B.2                | 2.33                               | 1.15      |
| I.B.4                | 4.60                               | 1.67      |
| I.C.3                | 3.80                               | 1.92      |
| I.C.4                | 3.80                               | 4.66      |
| I.C.5                | 1.50                               | 1.00      |
| I.C.7                | 14.80                              | 2.17      |
| I.C.8                | 1.00                               | 0.00      |
| I.C.9                | 3.00                               | 0.00      |
| <b>I.C.10</b>        | <b>0.00</b>                        | <b>NA</b> |
| I.C.11               | 2.00                               | 1.00      |
| <b>I.C.13</b>        | <b>0.00</b>                        | <b>NA</b> |
| <b>I.D.2</b>         | <b>0.00</b>                        | <b>NA</b> |
| I.D.6                | 1.00                               | 0.00      |
| I.D.7                | 1.00                               | 0.00      |
| I.D.8                | 1.00                               | NA        |

**Table A-22. Grade 6 MCA-MOD: Grade Span Benchmarks Matched to Items by Panelists**

| Benchmark Item Codes | Mean Number of Items per Benchmark | SD        |
|----------------------|------------------------------------|-----------|
| I.B.2                | 6.50                               | 1.29      |
| I.B.4                | 2.33                               | 1.15      |
| I.C.1                | 2.00                               | NA        |
| I.C.3                | 14.50                              | 1.29      |
| I.C.5                | 6.00                               | 2.58      |
| I.C.6                | 2.00                               | NA        |
| I.C.7                | 1.25                               | 0.50      |
| I.C.8                | 2.33                               | 1.53      |
| <b>I.C.12</b>        | <b>0.00</b>                        | <b>NA</b> |
| I.C.13               | 2.00                               | 1.41      |
| I.D.3                | 3.33                               | 2.52      |
| I.D.5                | 1.67                               | 0.58      |
| I.D.8                | 1.67                               | 1.15      |

**Table A-23. Grade 7 MCA-MOD: Grade Span Benchmarks Matched to Items by Panelists**

| Benchmark Item Codes | Mean Number of Items per Benchmark | SD        |
|----------------------|------------------------------------|-----------|
| I.B.2                | 3.25                               | 2.63      |
| I.B.3                | 1.50                               | 0.58      |
| I.C.1                | 6.75                               | 4.50      |
| I.C.4                | 5.50                               | 4.43      |
| I.C.7                | 1.75                               | 0.50      |
| I.C.8                | 2.25                               | 0.96      |
| I.D.3                | 5.25                               | 3.10      |
| I.D.4                | 2.33                               | 1.15      |
| <b>I.D.5</b>         | <b>0.00</b>                        | <b>NA</b> |
| I.D.6                | 1.25                               | 0.50      |
| I.D.10               | 2.33                               | 0.58      |

**Table A-24. Grade 8 MCA-MOD: Grade Span Benchmarks Matched to Items by Panelists**

| Benchmark Item Codes | Mean Number of Items per Benchmark | SD        |
|----------------------|------------------------------------|-----------|
| I.B.2                | 4.80                               | 1.30      |
| I.B.3                | 2.00                               | 1.00      |
| I.C.1                | 2.40                               | 1.67      |
| I.C.4                | 12.00                              | 1.41      |
| <b>I.C.6</b>         | <b>0.00</b>                        | <b>NA</b> |
| I.C.11               | 1.60                               | 0.55      |
| I.C.14               | 3.67                               | 2.52      |
| I.D.3                | 2.00                               | 1.73      |
| I.D.4                | 1.50                               | 0.71      |
| <b>I.D.7</b>         | <b>0.00</b>                        | <b>NA</b> |
| I.D.12               | 1.00                               | NA        |

**Table A-25. Grade 10 MCA-MOD: Grade Span Benchmarks Matched to Items by Panelists**

| Benchmark Item Codes | Mean Number of Items per Benchmark | SD        |
|----------------------|------------------------------------|-----------|
| I.B.2                | 7.40                               | 3.78      |
| I.C.3                | 1.50                               | 0.71      |
| I.C.5                | 3.40                               | 0.89      |
| I.C.6                | 1.33                               | 0.58      |
| I.C.7                | 15.20                              | 5.89      |
| <b>I.C.8</b>         | <b>0.00</b>                        | <b>NA</b> |
| <b>I.C.9</b>         | <b>0.00</b>                        | <b>NA</b> |
| I.D.4                | 1.50                               | 0.71      |
| I.D.5                | 1.25                               | 0.50      |
| I.D.6                | 1.50                               | 0.71      |
| I.D.7                | 1.00                               | 0.00      |
| I.D.10               | 1.00                               | 0.00      |
| I.D.14               | 1.00                               | NA        |

**Appendix B.  
Sample Alignment Review Materials**

*Panelists received the following instruction sheet as a reference guide corresponding with verbal instructions from HumRRO facilitators.*

**MCA-MOD  
Panelist Instructions**

| <b>Rating Task</b>           | <b>Documents Needed</b>   | <b>File Format</b> |
|------------------------------|---|--------------------|
| DOK of MN Academic Standards | (1) Minnesota Academic Standards for Reading (HumRRO Coded)               | Print Copy         |
|                              | (2) DOK Codes for Reading   | Print copy         |
| MCA-MOD Reading Items        | (1) Minnesota Academic Standards for Reading (HumRRO Coded) – grade spans | Print copy         |
|                              | (2) DOK Codes for Reading   | Print copy         |
|                              | (3) MCA-MOD items (printed)   | Print copy         |
|                              | (4) MCA-Reading_ItemRating_Grade X_Oct2011                                | Excel spreadsheet  |
| Whole Test                   | (1) MCA-MOD items (printed)   | Print copy         |
|                              | (2) MCA-Reading_WholeTestRatings_Grades x-x_Oct2011                       | Excel spreadsheet  |

**1 Rate DOK of Minnesota Academic Standards at Benchmark Level**

Using the ‘Minnesota Academic Standards for Reading’ printouts, assign a depth-of-knowledge rating to each benchmark of the Minnesota Academic Standards. You may simply write down your DOK ratings next to each benchmark and HumRRO Code. First, you will rate the benchmarks independently. Then, we will come to consensus on the ratings (3/4 majority). The consensus ratings will be retained for analysis. We will repeat this process to evaluate each relevant grade-span of the standards.

**2 Rate MCA-MOD Reading items on multiple dimensions**

Open the file ‘MCA-Reading\_ItemRating\_Grade X\_Oct2011’. Rename the file using the naming conventions (dok,test,contentarea,grade#,firstinitial,lastname --- ex. dokmcam3ktaylor). Click on the worksheet with the appropriate grade level.

- A Item DOK.** Assign a depth-of-knowledge rating to each item using the same DOK codes. Rate each item on the degree of cognitive processing required of students to answer the item adequately. Enter the DOK level (number) in the spreadsheet under the column labeled Item DOK Rating next to each item number.
- B Standards Match.** Use the ‘[MN Academic Standards with HumRRO Codes](#)’ to identify the benchmark that the item targets using the numeric code found in the right-hand column.

- C Degree of Alignment.** Rate the overall match level of the item to the benchmark to indicate *how well* you think that the item actually links to the listed benchmark. Using the rating scale below, enter the appropriate rating number from the scale into your spreadsheet under the column 'Overall Alignment'.
- 1 Not aligned to any benchmark (Use ONLY if you did not assign a benchmark to the item).
  - 2 Weakly aligned to this benchmark – does not assess the content of the academic standards well.
  - 3 Highly aligned to this benchmark - targets core content reasonably well.
  - 4 Fully aligned to the benchmark - Exemplary item, clear example of standard to which it is matched.
- D Item Quality.** Rate the overall quality of the item. Is the item clear and precise? Could you understand what the item is asking students to do (NOT whether you are capable of answering the item correctly)? Use the scale below to make your judgments.
- Overall Item Quality
- 1 Item is of poor overall quality (Rating requires annotation).
  - 2 Item is of good quality, but has some easily repairable flaw (Rating requires annotation).
  - 3 Item is of good quality, typical of what you would expect on this and similar tests.
  - 4 Item is of exceptional quality (annotations encouraged).
- E Notes/Comments.** Provide annotations for any item that you give a low rating on degree of alignment (rating of 1 or 2) or on item quality (rating of 1).

*This rating task will occur at the end of Day 3. Only a few panelists may have time to complete these ratings, depending on time.*

### **3 Rate 'Whole Test' barriers to demonstrating student knowledge**

Open the Excel 'MCA-Reading\_WholeTestRatings\_Grades x-x\_Oct2011' file. Click on the appropriate grade worksheet.

Make an evaluation of the test as a whole on the dimensions listed. Consider each student group who may be taking the assessment. These evaluations only require a Y (yes) or N (no) response in each of the blank cells.

*Panelists received the following coding sheet as a reference guide for the DOK rating scale.*

## **Depth-of-Knowledge (DOK) Levels for Reading**

(adapted from *Web Alignment Tool (WAT) Training Manual*)

- **Level 1 (recall/reproduction)** item requires recall of information such as fact, definition, term or simple procedure as well as performance of a simple Reading process or procedure.

Keywords: Identify, define, determine, perform (simple procedure), list.

- **Level 2 (skill/concept)** Item calls for engagement of some mental processing beyond a habitual response. Students required to make some decisions as to how to approach a problem or activity, such as selecting procedures, describing or giving examples of Reading concepts, deciding how to display or interpret data.

Keywords: Describe, observe, classify, confirm, organize, distinguish

- **Level 3 (strategic thinking)** Items require students to use reasoning and evidence, plan, and make conjectures. Students should be able to explain phenomena in terms of scientific concepts, explain simple relationships, explain their own thinking and conclusions, solve non-routine problems, and develop research questions.

Keywords: Connect, explain, analyze, outline procedures, make conclusions, interpret.

- **Level 4 (extended thinking)** Items require student to use complex and abstract reasoning and thinking, often over an extended period of time. Students must design and plan experimental studies, select and appropriate method among alternatives, or deduct the relationship among several variables.

NOTE: Many on-demand assessment instruments will not include assessment activities that could be classified as Level 4. However, standards, goals, and objectives can be stated so as to expect students to perform thinking at this level. On-demand assessments that do include tasks, products, or extended responses would be classified as Level 4 when the task or response requires evidence that the cognitive requirements have been met.

Keywords: Design, plan, and develop experiments; make inferences from results; critique; predict; explain (complex) relationships or differences among variables.

*Panelists received the Minnesota Academic Standards for Reading coded for data entry into rating forms. The content of the standards was extracted exactly from the full Minnesota Academic Standards document. Only a portion of the coded standards is replicated below for grade 5 as an example.*

| Grade   | Strand                          | Sub-Strand  | Standard   | Benchmarks   | HumRRO ID |
|---------|---------------------------------|---|--|--|-----------|
| GRADE 5 | I. READING<br>AND<br>LITERATURE | A. Word<br>Recognition,<br>Analysis, and<br>Fluency | The student will decode unfamiliar words using phonetic and structural analysis and will read with fluency and expression. | 1. Read unfamiliar, complex and multi-syllabic words using advanced phonetic and structural analysis.  | 02511101  |
|         |                                 |   |  | 2. Read aloud narrative and expository text with fluency, accuracy and appropriate pacing, intonation and expression.  | 02511102  |
| GRADE 5 |                                 | B. Vocabulary<br>Expansion                          | The student will use a variety of strategies to expand reading, listening and speaking vocabularies.                       | 1. Acquire, understand and use new vocabulary through explicit instruction as well as independent reading.   | 02512101  |
|         |                                 |   |  | 2. Use knowledge of root words, derivations, antonyms, synonyms, idioms, homonyms and multiple-meaning words to determine word meanings and to understand texts. | 02512102  |
|         |                                 |   |  | 3. Use word reference materials, such as dictionaries, thesauruses, to understand and express word meaning.  | 02512103  |
|         |                                 |   |  | 4. Analyze word structure and use context clues in order to understand new words.  | 02512104  |

Panelists received the Minnesota Academic Standards for Reading in a rating form in which to make DOK ratings for each benchmark. Panelists entered DOK ratings (1, 2, 3, or 4) in the last column of the table next to each benchmark. The content of the standards was extracted exactly from the full Minnesota Academic Standards document. Only a portion of the standards is replicated for grade 5 as an example.

| Grade   | Strand                          | Sub-Strand  | Standard   | Benchmarks   | HumRRO ID |
|---------|---------------------------------|---|--|--|-----------|
| GRADE 5 | I. READING<br>AND<br>LITERATURE | A. Word<br>Recognition,<br>Analysis, and<br>Fluency | The student will decode unfamiliar words using phonetic and structural analysis and will read with fluency and expression. | 1. Read unfamiliar, complex and multi-syllabic words using advanced phonetic and structural analysis.  | 02511101  |
|         |                                 |   |  | 2. Read aloud narrative and expository text with fluency, accuracy and appropriate pacing, intonation and expression.  | 02511102  |
| GRADE 5 |                                 | B. Vocabulary<br>Expansion                          | The student will use a variety of strategies to expand reading, listening and speaking vocabularies.                       | 1. Acquire, understand and use new vocabulary through explicit instruction as well as independent reading.   | 02512101  |
|         |                                 |   |  | 2. Use knowledge of root words, derivations, antonyms, synonyms, idioms, homonyms and multiple-meaning words to determine word meanings and to understand texts. | 02512102  |
|         |                                 |   |  | 3. Use word reference materials, such as dictionaries, thesauruses, to understand and express word meaning.  | 02512103  |
|         |                                 |   |  | 4. Analyze word structure and use context clues in order to understand new words.  | 02512104  |

*Panelists reviewed the individual Reading MCA-MOD items using the following rating form in electronic format. The format of the rating form was identical for each grade span. The number of items listed per rating form did differ for each grade test.*

| Item Number                  | Depth Of Knowledge                                | Benchmark 1              | Benchmark 2              | Written Content                 | Figures/Graphics                | Overall Alignment       | Overall Item Quality    | Explanation  |
|------------------------------|---|--------------------------|--------------------------|---------------------------------|---------------------------------|-------------------------|-------------------------|--|
| (Number Listed in Test Form) | 1-Recall<br>2-Skill<br>3-Reasoning<br>4-Inference | (Enter Standard ID Code) | (Enter Standard ID Code) | Y=universal<br>N=needs revision | Y=universal<br>N=needs revision | (Enter Scale of 1 to 4) | (Enter Scale of 1 to 4) | Please provide if you entered an Overall Alignment rating of '1' or '2' and/or an Overall Item Quality rating of '1' |
| 1                            |   |                          |                          |                                 |                                 |                         |                         |  |
| 2                            |   |                          |                          |                                 |                                 |                         |                         |  |
| 3                            |   |                          |                          |                                 |                                 |                         |                         |  |
| 4                            |   |                          |                          |                                 |                                 |                         |                         |  |
| 5                            |   |                          |                          |                                 |                                 |                         |                         |  |
| 6                            |   |                          |                          |                                 |                                 |                         |                         |  |
| 7                            |   |                          |                          |                                 |                                 |                         |                         |  |
| 8                            |   |                          |                          |                                 |                                 |                         |                         |  |
| 9                            |   |                          |                          |                                 |                                 |                         |                         |  |
| 10                           |   |                          |                          |                                 |                                 |                         |                         |  |
| 11                           |   |                          |                          |                                 |                                 |                         |                         |  |
| 12                           |   |                          |                          |                                 |                                 |                         |                         |  |
| 13                           |   |                          |                          |                                 |                                 |                         |                         |  |
| 14                           |   |                          |                          |                                 |                                 |                         |                         |  |
| 15                           |   |                          |                          |                                 |                                 |                         |                         |  |