

## **Missouri Assessment Program-Alternate (MAP-A) Alignment Review Communication Arts: Technical Report**

**Leslie R. Taylor  
Hilary L. Campbell  
Rebecca L. N. Dvorak  
Richard C. Deatz  
Lisa E. Koger  
Milton E. Koger  
D. E. (Sunny) Becker  
Arthur A. Thacker**

*Prepared for:* Missouri Department of Elementary and Secondary Education  
205 Jefferson Street  
P. O. Box 480  
Jefferson City, Missouri 65102

*Prepared under:* Contract No: C308004001-002

December 14, 2009



## **Missouri Assessment Program-Alternate (MAP-A) Alignment Review Communication Arts: Technical Report**

**Leslie R. Taylor  
Hilary L. Campbell  
Rebecca L. N. Dvorak  
Richard C. Deatz  
Lisa E. Koger  
Milton E. Koger  
D. E. (Sunny) Becker  
Arthur A. Thacker**

*Prepared for:* Missouri Department of Elementary and Secondary Education  
205 Jefferson Street  
P. O. Box 480  
Jefferson City, Missouri 65102

*Prepared under:* Contract No: C308004001-002

December 14, 2009

## EXECUTIVE SUMMARY

### *Purpose and Scope of Work*

The Missouri Department of Elementary and Secondary Education (DESE) requested an external independent alignment study of the Missouri Academic Program-Alternate (MAP-A) in Communication Arts for students with significant cognitive disabilities. Specifically, DESE wanted an evaluation of the alignment between the MAP-A portfolio assessment, the extended content standards (or Alternate Grade-Level Expectations<sup>1</sup>), and the Missouri Show-Me Standards<sup>2</sup>. Missouri uses the MAP-A portfolio assessment in the federal and state accountability programs. DESE awarded Human Resources Research Organization (HumRRO) the contract to conduct this alignment study, and work began on February 1, 2009.

DESE requested the alignment study to meet both state and federal requirements. The federal requirement of the U.S. Department of Education (USDE) stems from the *No Child Left Behind Act* (NCLB) 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 its 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 a state's assessment system. Alignment results should demonstrate that the assessments represent the full range of content standards, and that these assessments measure student knowledge in the same manner and at the same level of complexity as expected 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.

Alternate assessments are included in the federal requirements. The federal government has established regulations for students with significant cognitive disabilities in the calculation of school and district AYP determinations, often referred to as the "1% rule" (U.S. Department of Education, 2005). This rule allows the state to accommodate students with significant cognitive disabilities in its AYP calculations by setting different performance expectations for up to 1% of the student population. As a result, states can develop alternate content

---

<sup>1</sup> Missouri Alternate Grade-Level Expectations can be found at:  
[http://www.dese.mo.gov/divimprove/assess/mapa\\_resources.html](http://www.dese.mo.gov/divimprove/assess/mapa_resources.html)

<sup>2</sup> Missouri Show-Me Standards can be found at: <http://dese.mo.gov/standards/>

standards (often referred to as extended standards), achievement standards, and assessments designed to more fairly demonstrate the knowledge of these students. However, the content on which these students are assessed must be academic (rather than reflect life function skills), and the achievement of these students must continue to reflect challenging academic goals. As such, states must show that the extended standards and alternate achievement standards for these students link to the grade-level expectations, although the breadth and depth of these expectations can be reduced (USDE, 2005).

### ***Methodology***

HumRRO convened panels of Missouri educators and national content experts to review the MAP-A portfolio assessments. These panelists included current and former teachers, administrators, and curriculum specialists/district coordinators.

Six panelists met to review the Communication Arts portfolios relative to the Missouri Alternate Grade-Level Expectations (AGLEs). The panel consisted of five in-state Missouri panelists and one out-of-state panelist. Each panelist evaluated portfolios for all levels under review per content area.

HumRRO used the Links for Academic Learning alignment method (referred to as the LAL method in this report) developed by the National Alternate Assessment Center (NAAC) to conduct the reviews and analyze the results (Flowers, Wakeman, Browder, & Karvonen, 2007). This method requires panelists to rate the content standards and assessments on multiple dimensions. Ratings are then analyzed and interpreted based on seven criteria. These criteria are listed below (adapted from Flowers et al., 2007):

**LAL Criterion 1: Academic** - The content is academic and includes the major domains/strands of the content area as reflected in state and national standards (e.g., Mathematics, Reading, Communication Arts).

**LAL Criterion 2: Age Appropriate** - The content is referenced to the student's assigned grade level (based on chronological age).

**LAL Criterion 3: Standards Fidelity**

- a. **Content Centrality** - The target content maintains fidelity with the content of the original grade-level standards.
- b. **Performance Centrality** - The focus of achievement maintains fidelity with the specified performance in the grade-level standards.

**LAL Criterion 4: Content Coverage** (Webb alignment indicators) - The content differs from grade level in range, balance, and depth-of-knowledge (DOK), but matches high expectations set for students with significant cognitive disabilities.

**LAL Criterion 5: Content Differentiation** - There is some differentiation in content across grade levels or grade bands.

**LAL Criterion 6: Achievement** - The expected achievement for students is for the students to show learning of grade referenced academic content.

**LAL Criterion 7: Performance Accuracy** - The potential barriers to demonstrating what students know and can do are minimized in the assessment to increase measurement accuracy of student performance.

Under LAL Criterion 4 above, we refer to the “Webb alignment indicators.” Dr. Norman Webb (2005) developed an alignment procedure involving an evaluation 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. Webb’s method generally has been applied to regular general education assessments, and some special education researchers (i.e., Flowers et al., 2007) consider this approach to be limited as a primary alignment method for alternate assessments. However, the Webb alignment indicators provide important information regarding content coverage. Thus, the LAL method includes the following Webb alignment indicators:

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

The outcomes of the analyses on the LAL criteria and Webb alignment indicators are evaluated against decision rules to judge their acceptability. However, because the MAP-A is a portfolio assessment, with only four entries per student to assess a large number of Alternate Grade-Level Expectations (AGLEs), we can expect Webb’s Indicators 1–3 referenced above to not be met. The criterion for meeting the categorical concurrence requirement would be that the assessment include six items per content strand (e.g., Reading Processes), so even if all four entries assessed a single strand (which is highly unlikely), the criterion still could not be met. Similarly, the requirement for acceptable range-of-knowledge representation is that at least one item on the assessment relate to 50% of the indicated standards. Because there are so many Communication Arts AGLEs per strand (more than 20), this criterion also could not be achieved. Finally, the requirement for balance-of-knowledge correspondence is

inappropriate. Each Communication Arts portfolio includes four performance tasks (or portfolio entries). Each grade span requires that the portfolio be used to assess three strands (two reading strands and one writing strand). The typical portfolio will contain entries for each assessed strand, with one strand (usually the writing strand) assessed using two entries. The design of the system typically results in a portfolio that has only one entry for a given AGLE and only two entries within any strand. The balance index is inappropriate because it would be misleading. Since each portfolio is likely to have only 1–2 AGLEs per strand, the index might be high, but still not be informative regarding the distribution of AGLEs within strands on the assessment.

Webb’s DOK consistency requirement is appropriate for this study. The match between the DOK of the AGLEs and the DOK indicated by the portfolio entries can be ascertained and reported. For the other indicators, HumRRO chose to describe the distribution of the AGLEs across portfolios to provide DESE information about which AGLEs were assessed within strands and whether certain AGLEs were favored while others were avoided. We determined that if there was a reasonable distribution of AGLEs across portfolios within a grade span (e.g., Grades 3, 4 and 5), a student might have the opportunity to receive instruction across several AGLEs within each strand over three years. It is still possible that a single student might be instructed on the same four AGLEs for all three years, but if there is a wide distribution of AGLEs assessed among the sample of portfolios and if no AGLE dominates the assessments, this seems unlikely. The ratings included in this report for Webb’s Criteria 1–3 (noted above) do not represent acceptability using Webb’s interpretations.

### ***Summary Alignment Results***

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

The results of the alignment reviews provide positive support for the content validity of the MAP-A assessment based on several outcomes. First, the panelists found nearly all of the grade-level AGLEs for Communication Arts to be academic rather than functional. Second, nearly all MAP-A portfolio entries across grades were rated as matched to AGLEs. Third, panelists determined that the AGLEs and assessments are accessible to a wide range of students with various physical and cognitive disabilities. Finally, the alignment review of the achievement standards to portfolio entries suggests that the Communication Arts MAP-A assessments are designed to reflect DOK levels similar to the DOK levels indicated by the AGLEs.

As with most alignment reviews, the findings also point to some areas where content and performance alignment could be strengthened over time.

## AGLEs to Missouri Show-Me Standards

Table 1 displays the summary conclusions regarding content alignment between the AGLEs and Missouri Show-Me Standards for Communication Arts. These judgments are based on whether the AGLEs achieved acceptable levels of linkage with the full content standards for each grade level. The minimum level for each of the criteria in Table 1 is 90%.

- High linkage - most of standards are acceptable (at least 90%)
- Partial linkage - some standards are acceptable (50-89%)
- Weak linkage - few to no standards are acceptable (less than 50%)

**Table 1. Summary Conclusions on Alignment of AGLEs to Missouri Show-Me Standards for Communication Arts on LAL Criteria 2, 3, and 5**

Grade Span	LAL Criterion 2	LAL Criterion 3		LAL Criterion 5
	Age Appropriate	Content Centrality	Performance Centrality	Content Differentiation
	Is content referenced to student's assigned grade level?	Do the extended standards link to the target content in the grade-level standards?	Does the performance of the extended standards link to expectations of the grade level standards?	Do the extended standards show appropriate increases between grade levels?
3–5	High	Weak	High	Weak
6–8	Partial	Weak	High	Weak
HS	Partial	Weak	High	Weak

The content alignment conclusions in Table 1 indicate that the grade-level AGLEs link well to the Missouri Show-Me Standards in terms of performance, but only weakly in terms of content. Panelists judged that while alternate assessment students might perform similar tasks to the regular education students, the content of those tasks was only weakly related. Panelists also found little evidence of differentiation across grades. In addition, this content is well differentiated, or vertically aligned, between grades overall.

Table 2 displays the overall conclusions pertaining to LAL Criterion 7 - Performance Accuracy (content accessibility) for the AGLEs. For this criterion, conclusions reflect overall judgments of acceptability based on the following categories<sup>3</sup>:

<sup>3</sup> Adapted from universal design ratings used by the National Center on Educational Outcomes (NCEO). See Thompson et al. (2005).

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

**Table 2. Summary Conclusions on Performance Accuracy (LAL Criterion 7) of Communication Arts AGLEs**

Criterion 7		
Grade Span	Performance Accuracy (Potential Barriers to Accessibility)	
	Is the content appropriate for students at different levels of communication?	Is the content accessible to different disability groups?
3–5	Questionable	Good
6–8	Questionable	Good
HS	Questionable	Good

The conclusions on Performance Accuracy clearly are disparate. Although seemingly in conflict, the two ratings on access address quite different aspects. The conclusions on communication reflect panelists' judgments that students with limited symbolic ability may have difficulty comprehending or demonstrating the content in some AGLEs. In contrast, panelists felt that students with varying physical or cognitive impairments (i.e., difficulty with instructions, attention, and sensory integration) can typically access and demonstrate this knowledge.

### MAP-A Tasks to AGLEs

Table 3 provides summary conclusions on the alignment of the MAP-A assessments to AGLEs. The conclusions are based on the following criteria:

- High linkage - most of tasks are acceptable (at least 90%)
- Partial linkage - some tasks are acceptable (50%-89%)
- Weak linkage - few to no tasks are acceptable (less than 50%)

**Table 3. Summary Conclusions on Alignment of Communication Arts MAP-A Portfolios to AGLEs for LAL Criteria 1, 2, 3, 4, and 5**

	LAL Criterion 1	LAL Criterion2	LAL Criterion 3		LAL Criterion 4		LAL Criterion 5
Grade Span	Academic Content	Age Appropriate	Content Centrality	Performance Centrality	Content Coverage		Content Differentiation
	Are students assessed on academic content?	Is task content referenced to student’s assigned grade level?	Do tasks link to the target content in the AGLEs?	Does the performance of task link to expectations of the AGLEs?	Do the tasks assess students at the appropriate breadth of knowledge? <sup>a</sup>	Do the tasks assess students at the appropriate depth-of-knowledge? <sup>b</sup>	Do the assessments show appropriate increases between grade levels?
<b>3–5</b>	<b>High</b>	<b>High</b>	<b>Partial</b>	<b>High</b>	<b>Weak</b>	<b>High</b>	<b>Weak</b>
<b>6–8</b>	<b>Partial</b>	<b>Partial</b>	<b>Partial</b>	<b>High</b>	<b>Weak</b>	<b>Partial</b>	<b>Weak</b>
<b>HS</b>	<b>Partial</b>	<b>Partial</b>	<b>Partial</b>	<b>High</b>	<b>Weak</b>	<b>Partial</b>	<b>Weak</b>

<sup>a</sup> These conclusions are based on a summary judgment across the Webb statistics of Categorical Concurrence, Range of Knowledge, and Balance of Knowledge, which may be inappropriate for this assessment.

<sup>b</sup> These conclusions are based on the results from the DOK consistency analyses.

As Table 3 illustrates, the 2009 MAP-A assessments linked well to the AGLEs on the majority of dimensions in the lower grades. Evidence for middle and high school grades was weaker, but typically at least partially linked. The exceptions to this rule were for content coverage and content differentiation across grades. Under LAL Criterion 4, indications of alignment were restricted by the numbers of portfolio entries compared to AGLEs as described above. These criteria may not be appropriate for this assessment, but it does seem clear that there are more AGLEs than can be adequately assessed from a single portfolio. On depth-of-knowledge assessed, panelists determined that at least half of all tasks assessed within Grades 6–8 and high school under Writing Processes assessed students at a lower level of cognitive complexity than expected in the AGLEs. All other strands for all grades were represented by at least 50% of portfolio entries at or above the DOK level indicated by the AGLEs. Regarding LAL Criterion 5, Content Differentiation, panelists did not find the content between each grade-level assessment to demonstrate very much content differentiation in terms of breadth or depth. Panelists found that much of the content was identical across grade levels.

Table 4 includes results related to Criteria 6 and 7 of the LAL method. These rating questions asked panelists to determine whether the assessment tasks are designed in such a way that students can demonstrate knowledge at various levels of functioning and ability. Ratings in this case are based on evaluations of accessibility, rather than on content alignment<sup>4</sup>:

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

**Table 4. Summary Conclusions on Accessibility (LAL Criteria 6 and 7) of MAP-A Communication Arts Portfolios**

	Criterion 6	Criterion 7		
Grade Span	Achievement	Performance Accuracy (Potential Barriers)		
	Does the assessment allow for accurate inference about student learning?	What level of symbolic communication does task require?	Is task accessible to different disability groups?	Can task be modified/supports provided without changing meaning or difficulty?
3–5	Questionable	Questionable	Excellent	Excellent
6–8	Questionable	Questionable	Good	Excellent
HS	Questionable	Questionable	Excellent	Excellent

<sup>4</sup> Alignment refers to overlap in content expectations. In this case, the goal is not to measure the test against the content expectations but to evaluate the level of accessibility.

The most noticeable issue regarding accessibility for the MAP-A Communication Arts portfolio assessments concerns the measurement of achievement (LAL Criterion 6). Panelists indicated that the assessment accuracy could be high, but that making inferences about student achievement regarding degree of new learning, generalizability, and program indicators was problematic. Panelists were concerned that new learning might be difficult to measure since the content of the portfolio could change radically from year to year and because assessed strands changed based on grade spans. Panelists were also concerned that the portfolio entries might not generalize across multiple AGLEs within a strand. Finally, the panelists were concerned that the portfolios did not provide a strong indication of program quality.

Results on LAL Criterion 7 were mostly positive, with the exception of level of symbolic communication. It should be noted that the panelists were rating portfolios designed for particular students rather than the system for creating portfolios itself. It might not be surprising that most panelists judged that creating these particular portfolio entries would require symbolic communication beyond pre-symbolic. All grades showed otherwise acceptable results for this criterion.

## Recommendations

HumRRO makes the following recommendations to strengthen the linkage between the components of the Missouri alternate assessment system.

### AGLEs for Communication Arts

- (1) **Review the access points for the AGLEs at each grade level.** For each grade level, panelists identified some AGLEs that may limit access only to those students with higher symbolic abilities, thus excluding a portion of students from the assessment system. Reviewing the AGLEs may involve additional bias reviews to modify the current expectations or additional explanation (e.g., content limitations, examples) within the MAP-A Test Specifications document may be sufficient to better illustrate how teachers might make these content expectations more appropriate for students with lower symbolic abilities.
- (2) **Review AGLEs for links to academic content contained in the Missouri Show-Me Standards.** The link between the regular and the alternate standards was not as strong as expected by panelists in terms of content. There may be qualitative differences in the expectations for alternate assessment students despite clear indications that similar performance is expected.
- (3) **Review AGLEs for differentiation from grade to grade.** Panelists found little new, broader, or deeper knowledge requirements associated with moving from grade to grade. If there are differences

that are not specifically described in the AGLEs, those differences should be made explicit.

### **MAP-A Portfolio Performance Tasks**

- (1) ***Review performance tasks for age appropriateness for Grades 6–8 and high school.*** A portion of the portfolio tasks (14-19%) for these grades was judged by panelists to be age inappropriate. While not a large percentage, this may indicate the need for additional training of teachers or that clearer instructions be provided for preparing the portfolio performance task.
- (2) ***Review performance tasks for content centrality for all grades.*** A portion of the portfolio tasks (20%-34%) was judged by panelists to be weakly linked or not linked to the AGLEs. This may indicate the need for additional training of teachers or that clearer instructions be provided for preparing the portfolio performance tasks to target specific AGLEs.
- (3) ***Consider the implications of using a portfolio system to assess standards with so many AGLEs per strand.*** Webb's alignment criteria related to breadth of content knowledge were inappropriate for this assessment because of the small numbers of portfolio entries compared to the relatively large numbers of AGLEs. There are simply too many AGLEs to adequately measure with the limited number of portfolio entries.
- (4) ***Consider the inferences expected to be made from Communication Arts MAP-A scores by schools and teachers in relation to the test design.*** Panelists indicated concerns that the Communication Arts MAP-A scores might not provide adequate information about students' new learning, that they might not generalize beyond the specific content of the portfolio performance task, and that the scores might not reflect program quality.

**MISSOURI ASSESSMENT PROGRAM-ALTERNATE (MAP-A) ALIGNMENT  
REVIEW: TECHNICAL REPORT**

**TABLE OF CONTENTS**

**List of Tables.....xii**

**Chapter 1 Introduction ..... 1**

    Alignment Requirements for State Assessment Systems..... 1

    Structure of Missouri’s Alternate Assessment System .....2

**Chapter 2 Alignment Study Design and Methodology .....5**

    Alignment of Assessments and Standards on Content and Accessibility .....5

*Links for Academic Learning (LAL) Alignment Method* .....6

*Panelists*.....8

*Materials*..... 10

*Procedures*..... 10

**Chapter 3 Results for MAP-A Communication Arts ..... 13**

    Alignment of Communication Arts AGLEs to Missouri Show-Me Standards . 13

    Alignment of Communication Arts Portfolio Assessments to AGLEs.....22

    Inter-Rater Agreement Results .....39

**References .....42**

**Appendix A Alignment Results per Grade Span Assessment ..... 1**

**Appendix B Workshop Instructions ..... 1**

**List of Tables**

Table 1. Summary Conclusions on Alignment of AGLEs to Missouri Show-Me Standards for Communication Arts on LAL Criteria 2, 3, and 5 ..... vi

Table 2. Summary Conclusions on Performance Accuracy (LAL Criterion 7) of Communication Arts AGLEs ..... vii

Table 3. Summary Conclusions on Alignment of Communication Arts MAP-A Portfolios to AGLEs for LAL Criteria 1, 2, 3, 4, and 5.....viii

Table 4. Summary Conclusions on Accessibility (LAL Criteria 6 and 7) of MAP-A Communication Arts Portfolios..... ix

Table 1.1. Design of the Communication Arts MAP-A Assessment.....2

Table 2.1 Professional and Demographic Characteristics of Communication Arts MAP-A Alignment Panelists .....9

Table 3.1. Mean Number of Communication Arts AGLEs Rated as Academic by Panelists ..... 13

Table 3.2. Communication Arts AGLEs at Various Levels of Age Appropriateness .....	14
Table 3.3. Mean Percent of AGLEs Linked to Show-Me Standards .....	15
Table 3.4. Mean Number of AGLEs at Various Levels of Content Centrality .....	16
Table 3.5. Mean Number of AGLEs at Various Levels of Performance Centrality .....	17
Table 3.6. Consensus Ratings on Content Differentiation between Grade Span MAP-A AGLEs for Communication Arts.....	18
Table 3.7. Mean Number of AGLEs Rated at Each Level of Symbolic Communication.....	20
Table 3.8. Communication Arts AGLEs Rated as Accessible to All Students .....	21
Table 3.9. Communication Arts AGLEs Rated as Amenable to Accommodations for All Students .....	21
Table 3.10. Mean Number of Communication Arts MAP-A Portfolio Tasks Rated as Academic by Panelists .....	23
Table 3.11. Communication Arts MAP-A Performance Tasks at Various Levels of Age Appropriateness .....	24
Table 3.12. Mean Percent of Tasks Linked to AGLEs .....	25
Table 3.13. Mean Percent of Performance Tasks at Various Levels of Content Centrality.....	25
Table 3.14. Mean Percent of Performance Tasks at Various Levels of Performance Centrality .....	26
Table 3.15. Summary of Breadth Across Content Categories Results for MAP-A Communication Arts Portfolios by Grade Level .....	30
Table 3.16. DOK Consistency for Communication Arts MAP-A, Grades 3–5: Mean Percent of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives.....	31
Table 3.17. DOK Consistency for Communication Arts MAP-A, Grades 6–8: Mean Percent of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives.....	32
Table 3.18. DOK Consistency for Communication Arts MAP-A, High School: Mean Percent of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives.....	32
Table 3.19. Mean Percentage of Performance Tasks at Each DOK Level .....	33
Table 3.20. Consensus Ratings on Content Differentiation between Grade Span MAP-A Assessments for Communication Arts.....	35
Table 3.21. Degree of Inference Evident on Student Learning in Communication Arts MAP-A Assessments.....	37
Table 3.22. Mean Number of Tasks Rated at Each Level of Symbolic Communication.....	38
Table 3.23. MAP-A Communication Arts Portfolio Entries Rated as Amenable to Accommodations for All Students .....	39
Table 3.24. Pair-wise Comparisons on AGLE Ratings per Grade Level .....	40
Table 3.25. Pair-wise Comparisons on Portfolio Ratings per Grade Level .....	41
Table A- 1. Content Representation for Communication Arts MAP-A: Number of Portfolio Entries per Strand.....	A-3

Table A- 2. Depth-of-Knowledge Consistency for Communication Arts MAP-A, Grades 3-5: Mean Percent of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives..... A-4

Table A- 3. Depth-of-Knowledge Consistency for Communication Arts MAP-A, Grades 6-8: Mean Percent of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives..... A-4

Table A- 4. Depth-of-Knowledge Consistency for Communication Arts MAP-A, High School: Mean Percent of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives ..... A-5

## MISSOURI ASSESSMENT PROGRAM-ALTERNATE (MAP-A) ALIGNMENT REVIEW: TECHNICAL REPORT

### Chapter 1 Introduction

The Missouri Department of Elementary and Secondary Education (DESE) requested an external independent alignment study of the Missouri Academic Program-Alternate (MAP-A) in Communication Arts for students with significant cognitive disabilities. Specifically, DESE wanted an evaluation of the alignment between the MAP-A portfolio assessment, the extended content standards (or Alternate Grade-Level Expectations<sup>5</sup>), and the Missouri Show-Me Standards<sup>6</sup>. Missouri uses the MAP-A portfolio assessment in the federal and state accountability programs. DESE awarded Human Resources Research Organization (HumRRO) the contract to conduct this alignment study, and work began on February 1, 2009.

#### *Alignment Requirements for State Assessment Systems*

DESE requested the alignment study to meet both state and federal requirements. The federal requirement of the U.S. Department of Education (USDE) stems from the *No Child Left Behind Act* (NCLB) 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 its 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 a state's assessment system. Alignment results should demonstrate that the assessments represent the full range of content standards, and that these assessments measure student knowledge in the same manner and at the same level of complexity as expected 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.

Alternate assessments are included in the federal requirements. The federal government has established regulations for students with significant cognitive disabilities in the calculation of school and district AYP determinations, often referred to as the "1% rule" (U.S. Department of Education, 2005). This rule

---

<sup>5</sup> Missouri Alternate Grade Level Expectations can be found at:  
[http://www.dese.mo.gov/divimprove/assess/mapa\\_resources.html](http://www.dese.mo.gov/divimprove/assess/mapa_resources.html)

<sup>6</sup> Missouri Show-Me Standards can be found at: <http://dese.mo.gov/standards/>

allows the state to accommodate students with significant cognitive disabilities in its AYP calculations by setting different performance expectations for up to 1% of the student population. As a result, states can develop alternate content standards (often referred to as extended standards), achievement standards, and assessments designed to more fairly demonstrate the knowledge of these students. However, the content on which these students are assessed must be academic, and the achievement of these students must continue to reflect challenging academic goals. As such, states must show that the extended standards and alternate achievement standards for these students link to the state standards, although the breadth and depth of these expectations can be reduced (USDE, 2005).

### ***Structure of Missouri's Alternate Assessment System***

Missouri chose to develop a portfolio system for those students with the most significant cognitive disabilities. A portfolio assessment is unique to each student and based on the student's Individualized Education Plan (IEPs), as opposed to a fully standardized assessment with common items or tasks across students. As part of this alternate assessment system, Missouri also constructed Alternate Grade-Level Expectations (AGLEs) on which the portfolio assessments must be based. Per the federal requirement, these content expectations correspond with the Missouri Show-Me Standards, although reduced in breadth and depth.

For Communication Arts, portfolios are assessed at Grades 3–8 and 11, but each might include content learned in prior grades. The Communication Arts MAP-A assesses six content strands and two process strands. The design of the assessment is summarized in Table 1.1 below. For each student, teachers submit four entries with each grade-level assessment for both Communication Arts and Mathematics. For Communication Arts each entry assesses one content strand. Writing conventions are assessed at Grades 3-5. Writing process strands are assessed at Grades 6–8 and 11. Reading strands are assessed at all tested grades.

***Table 1.1. Design of the Communication Arts MAP-A Assessment<sup>7</sup>***

---

Grade Focus	Title of Strand
3-8 and 11	Develop and apply skills and strategies to the reading process (RD and/or RP)
3-5	Compose well-developed text using standard English conventions (WC)
6-8 and 11	Apply a writing process in composing text or write effectively in various forms and types of writing (WP)

---

<sup>7</sup> Information taken from DESE's MAP-A informational PowerPoint presentation, prepared in August 2009.

### ***Organization and Contents of the Report***

This report contains three chapters. Chapter 2 describes the alignment method and test review details, including panelist characteristics, materials, and procedures. Chapter 3 provides alignment results for Communication Arts on the alignment of the AGLEs to the full Missouri Show-Me Standards, alignment of the MAP-A portfolio performance tasks to the AGLEs, and accessibility of AGLEs and MAP-A portfolios to those students who take this assessment.

Additional information is provided in the appendices of this report. Appendix A contains tables providing more detail on the content alignment results per content area and grade-level test form. Appendix B includes task descriptions and definitions of terms used by panelists who provided the ratings on which this report is based.



## 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 Missouri study.

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

The term *alignment* in this context refers to the degree of consistency evident in instruction and measurement of the state's academic content standards. School curricula should 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. In essence, all alignment evaluations link to the state 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?

Several alignment methods are currently in use for general education and alternate assessments. Most of these methods involve rating various aspects of test items or performance tasks relative to the content standards. Generally, education experts serve as panelists who review and rate the assessments on several measures of content breadth and depth to determine the extent of alignment.

Alignment studies of alternate assessments often require review of additional aspects of alignment unique to those assessments. These dimensions include: (a) accessibility of the assessment system to students with a variety of disabilities, (b) the extent to which test content is academic, and (c) the extent to which alternate content standards are linked with the state's general academic standards. Alternate assessments differ from general state assessments in form

and structure; thus, an alignment methodology must be responsive to these differences.

### **Links for Academic Learning (LAL) Alignment Method**

For the current alignment study, HumRRO applied the Links for Academic Learning alignment method (LAL) developed by the National Alternate Assessment Center to conduct the content alignment reviews and analyze the results (Flowers, Wakeman, Browder, & Karvonen, 2007). This method requires panelists to rate the content standards and assessments on multiple dimensions. Ratings are then analyzed and interpreted based on the following seven criteria (adapted from Flowers et al, 2007):

**LAL Criterion 1: Academic** - The content is academic and includes the major domains/strands of the content area as reflected in state and national standards (e.g., Mathematics, Reading, Communication Arts).

**LAL Criterion 2: Age Appropriate** - The content is referenced to the student's assigned grade level (based on chronological age).

#### **LAL Criterion 3: Standards Fidelity**

**a. Content Centrality** - The target content maintains fidelity with the content of the original grade-level standards.

**b. Performance Centrality** - The focus of achievement maintains fidelity with the specified performance in the grade-level standards.

**LAL Criterion 4: Content Coverage** (Webb alignment indicators) - The content differs from grade level in range, balance, and depth of knowledge (DOK), but matches high expectations set for students with significant cognitive disabilities.

**LAL Criterion 5: Content Differentiation** - There is some differentiation in content across grade levels or grade bands.

**LAL Criterion 6: Achievement** - The expected achievement for students is for the students to show learning of grade referenced academic content.

**LAL Criterion 7: Performance Accuracy** - The potential barriers to demonstrating what students know and can do are minimized in the assessment to increase measurement accuracy of student performance.

The LAL method is appropriate for alignment of multiple assessment formats (i.e., test forms, performance tasks, performance portfolios, and checklists). The method allows for comparison of the assessment to extended standards, as well as alignment of extended standards to full state content standards. In addition, the LAL method includes steps to evaluate the

accessibility of the extended standards and the assessment to students. The review of assessments to standards, such as the MAP-A portfolio assessment to the AGLEs, includes all of the LAL Criteria 1 through 7. The LAL Criteria 1, 2, 3, 5, and 7 apply to a review of extended standards, while all seven criteria apply to a review of the assessment.

Under LAL Criterion 4 above, we refer to the “Webb alignment indicators.” Dr. Norman Webb (2005) developed an alignment procedure involving an evaluation 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. Webb’s method generally has been applied to regular general education assessments, and some special education researchers (i.e., Flowers et al., 2007) consider this approach to be limited as a primary alignment method for alternate assessments. However, the Webb alignment indicators are still informative regarding content coverage. Thus, the LAL method includes the Webb alignment indicators. These alignment indicators include:

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

The outcomes of the analyses on the LAL criteria and Webb alignment indicators are evaluated against decision rules to judge their acceptability. However, because the MAP-A is a portfolio assessment, with only four entries per student (each designed to assess one AGLE for Communication Arts), we can expect that Webb’s Indicators 1–3 referenced above will not be met. The criterion for meeting the categorical concurrence requirement would be that the assessment includes six items per content strand (e.g., Reading Comprehension), so even if all four entries assessed a single strand, the criterion still could not be met. Similarly, the requirement for acceptable range-of-knowledge representation is that at least one item on the assessment relate to 50% of the indicated standards. Because there are so many Communication Arts AGLEs per strand, this criterion could also not be achieved. Finally, the requirement for balance-of-knowledge correspondence is also inappropriate. Each Communication Arts portfolio is expected to assess four strands (there are two Reading strands assessed at every tested grade and two writing strands, one of which is assessed in elementary grades, the other in middle and high school). The balance index is inappropriate because it could be misleading.

Since each portfolio is likely to have only 1–2 AGLEs per strand, the index could be high (about the same number of AGLEs per strand), but this would not provide useful information about the distribution of the AGLEs by strand.

Webb’s DOK consistency requirement is appropriate for this study. The match between the DOK of the AGLEs and the DOK indicated by the portfolio entries can be ascertained and reported. For the other indicators, HumRRO chose to describe the distribution of the AGLEs across portfolios to give DESE information about which AGLEs were assessed within strands and whether certain AGLEs were favored while others were avoided. We determined that if there was a reasonable distribution of AGLEs across portfolios within a grade span (e.g. Grades 3, 4, and 5) that a student might have the opportunity to receive instruction across several AGLEs within each strand over three years. It is still possible that a single student might be instructed on the same four AGLEs for all three years, but if there is a wide distribution of AGLEs assessed among the sample of portfolios and if no AGLE dominates the assessments, this seems unlikely. The ratings included in this report for Webb’s Criteria 1–3 above do not represent acceptability using Webb’s interpretations.

For the MAP-A alignment review using the LAL method, Missouri and out-of-state educators performed multiple ratings to carry out the two primary alignment tasks: (a) comparison of the grade-span AGLEs to the Communication Arts Missouri Show-Me Standards, and (b) comparison of the Communication Arts MAP-A performance tasks (per grade test) to the AGLEs. These tasks served as the basis for the content alignment evaluation of the AGLEs and assessments relative to the full Missouri Show-Me Standards, as well as the content accessibility evaluation of the AGLEs and assessments relative to the population of students for whom the alternate assessment was designed.

## **Panelists**

HumRRO convened a panel of Missouri educators and national content experts to review the Communication Arts MAP-A portfolio assessments. These panelists included current and former teachers, administrators, and curriculum specialists or district coordinators. The panel consisted of six members; five in-state Missouri panelists and one out-of-state panelist. Each expert evaluated portfolios for all grade levels under review. Table 2.1 presents the characteristics of the panelists.

**Table 2.1 Professional and Demographic Characteristics of Communication Arts MAP-A Alignment Panelists**

Professional Position	Number of Panelists		Gender		Region of Origin in Missouri (based on Missouri RPDCs)								
	Missouri	Out of State	M	F	1- SE	2- Heart	3- KC	4- NE	5- NW	6- SC	7- SW	8- STL	9- Central
Comm Arts: Grades 3-5, 6-8, HS	5	1	0	5									
Teacher	3						1	1			1		
Administrator	1									1			
Curriculum Specialist	1											1	

## Materials

Panelists evaluated the alignment of the MAP-A performance tasks with the Missouri Show-Me Standards and AGLEs using forms for both the Webb and LAL alignment methods. A description of the forms, rating scales, and operational definitions is provided in Appendix B. All ratings were then coded into a Microsoft Excel spreadsheet to facilitate analyses.

**Test Forms.** Reviewers evaluated a sample of 60 2008 Communication Arts MAP-A portfolios per grade level. Table 1.1 from the previous chapter describes the content and structure of the portfolios. Briefly, each portfolio included four entries (performance tasks). Each entry was designed to measure one of four content stands per grade level.

**Rating Forms and Instructions.** Panelists completed three rating forms individually and an additional three rating forms via group consensus (see Appendix B for descriptions of each form). Panelists received instruction sheets enumerating the alignment tasks that they needed to complete as well as code sheets listing the depth-of-knowledge ratings and other possible ratings for each task (see Appendix B).

## Procedures

HumRRO conducted this alignment review at the Assessment Resource Center (ARC) at the University of Missouri, Columbia, in Columbia, Missouri, in fall 2009. The workshops began with introductions of staff and observers. Next, panelists read and signed affidavits of nondisclosure for the secure materials they would review during the workshop. HumRRO staff then gave a brief presentation to describe alignment studies and to introduce tasks the reviewers would complete.

Following the general introduction, panelists began working within their content groups. A single group of six panelists reviewed all Communication Arts standards documents and assessment materials. Other groups at the workshop reviewed Mathematics and Science materials.

Within the small group, a HumRRO staff member further trained reviewers using sample standards and assessment tasks. Regarding instructions on how to rate standards and items, the HumRRO staff member provided general suggestions and comments when appropriate; however, they emphasized to reviewers that staff would not provide explicit direction on how to rate standards or items because reviewers were valued as content experts. The HumRRO staff member provided brief instructions about how to use each rating form.

After reviewing sample DOK evaluations as a group, reviewers rated the Standards from the Missouri Show-Me Standards relevant to each grade-level

test. Panelists first made independent evaluations without discussion. Once all reviewers had completed their ratings, groups discussed their ratings to achieve consensus DOK ratings for each Standard; a voluntary scribe within each group recorded these consensus ratings. Next, reviewers followed the same process to rate the DOK of the AGLEs, first individually and then to reach consensus.

Next, reviewers rated the AGLEs on a variety of factors, including (a) whether the Standard listed is the best match, (b) how well the AGLE links to the Standard, (c) whether the AGLE measures student performance of the Standard, (d) whether the AGLE is appropriate for the chronological age at which it is measured, (e) the level of symbolic communication required of students to demonstrate its content, and (f) whether the content expectation of the AGLE is accessible to various disability groups. These ratings were made individually; no consensus ratings were obtained.

Reviewers then received more specific instructions for rating portfolio performance tasks. For training, HumRRO staff facilitated reviewers in evaluating and discussing sample items as a group. After completing sample items, reviewers individually rated performance tasks on rating forms. The panelists rated the items on the same dimensions that they rated each AGLE (as described above). Reviewers in LAL alignment studies are typically instructed to assign a *primary Standard* to an item based on a judgment that an item clearly measured this Standard. Furthermore, reviewers could assign an *additional Standard* only in cases when the item seemed to assess another Standard as clearly as the primary Standard. For the Communication Arts MAP-A, the standard (AGLE) was clearly indicated by the teacher on each performance task. Teachers justified the event as a measure of a particular AGLE as part of the portfolio documentation. Therefore, reviewers verified that the teachers' indications of the standards the entries assessed were accurate and appropriate rather than matched the entry to standards themselves. Because this approach was confirmatory, results among reviewers indicated near exact agreement of which standard each portfolio entry assessed. Reviewers also indicated whether the content of the performance task was academic and whether it could be modified or supports be provided without changing its meaning.

Finally, panelists worked in their small groups to develop consensus ratings for three additional aspects of the MAP-A Communication Arts assessment. HumRRO staff trained panelists on each task, and then the voluntary scribe from within the small group recorded the group's consensus ratings on rating sheets. The first consensus task required panelists to rate whole test barriers, or aspects of the Communication Arts MAP-A as a whole that might prevent students with various disabilities from fully participating (with or without supports or accommodations). The second consensus task asked panelists to rate the extent to which the scoring rubric and achievement standards allow for the demonstration of student learning. Lastly, reviewers developed consensus ratings of the extent to which content differs across grades.



## Chapter 3 Results for MAP-A Communication Arts

### *Alignment of Communication Arts AGLEs to Missouri Show-Me Standards*

**LAL Criterion 1: Academic** - *The content is academic and includes the major domains/strands of the content area as reflected in state and national standards (e.g., Reading, Mathematics, Communication Arts).*

Per the USDE (2005), alternate assessments counting towards Title I must assess students only on academic content, as opposed to functional life skills. Panelists judged the grade-level Communication Arts assessments as to whether each AGLE focuses primarily on academics (similar to the Missouri Show-Me Standards). Results of this analysis are presented in Table 3.1. In terms of acceptability, at least 90% of tasks should be rated as academic.

**Table 3.1. Mean Number of Communication Arts AGLEs Rated as Academic by Panelists**

Grade Spans	Number of AGLEs at Grade Span	Mean Number of AGLEs Academic		Mean Number of AGLEs Functional		Mean Percentage of AGLEs Rated Academic	Number of Panelists Rating More than 90% of Rated AGLEs Academic <sup>a</sup>
		M	SD	M	SD		
3–5	150	136.7	4.6	13.3	4.6	91.1%	4 of 6
6–8	179	165.0	4.9	14.0	4.9	92.2%	4 of 6
HS	193	179.8	4.8	13.2	4.8	93.2%	5 of 6

<sup>a</sup> Percentages are based on actual ratings. Missing data were excluded from the numerator and denominator.

As Table 3.1 demonstrates, all three grade spans met the criterion for Academic; on average, panelists rated more than 90% of Communication Arts AGLEs as academic in nature at each grade span. Although some panelists rated slightly more than 10% of MAP-A Communication Arts AGLEs as measuring functional skills, the general consensus across panelists was that the AGLEs were primarily focused on academic skills.

**LAL Criterion 2: Age Appropriate** - The content is referenced to the student's assigned grade level (based on chronological age).

This criterion pertains to the developmental level of the content included in the AGLEs. For this evaluation, panelists were asked whether the content of the Communication Arts AGLEs is appropriate for the age and grade level indicated. Response options for this scale included:

- Adapted - Linked to grade level content.
- Inappropriate - Content is off-grade level.
- Neutral - Content is not age-bound and is appropriate at any age.

Table 3.2 includes the results of panelists' ratings. Column 3 lists the rating categories, while the 'Mean' in Column 4 refers to the mean number of AGLEs receiving that rating across panelists. Column 6 represents this same mean as a percentage of the total number of AGLEs per grade. Acceptability for this criterion is that at least 90% of AGLEs are rated as 'adapted' or 'neutral'<sup>8</sup>.

**Table 3.2. Communication Arts AGLEs at Various Levels of Age Appropriateness**

Grade Span	Number of AGLEs at Grade Span	Age Appropriateness Rating	Mean	SD	Mean Percentage of AGLEs per Rating	Number of Panelists Rating at Least 90% of Rated AGLEs Adapted or Neutral <sup>a</sup>
3–5	150	Adapted	102.3	26.7	68.2%	4 of 6
		Neutral	36.3	21.3	24.2%	
		Inappropriate	11.3	13.5	7.6%	
6–8	179	Adapted	117.2	49.6	65.5%	4 of 6
		Neutral	32.5	18.2	18.2%	
		Inappropriate	29.3	41.6	16.4%	
HS	193	Adapted	125.7	61.0	65.1%	4 of 6
		Neutral	31.7	21.4	16.4%	
		Inappropriate	35.7	49.5	18.5%	

<sup>a</sup> Percentages are based on actual ratings. Missing data were excluded from the numerator and denominator.

As Table 3.2 displays, panelists varied considerably in their ratings of the age appropriateness of the MAP-A Communication Arts AGLEs. On average, the highest percentage of adapted or neutral content occurred for the 3–5 grade span, which met the acceptability criterion of at least 90%. Average ratings for Grades 6–8 and high school were slightly below the 90% threshold for adapted or neutral content, with both falling between 80% and 85%. A closer look at the

<sup>8</sup> The LAL method does not specify a minimum for Criterion 2. This minimum level was established by HumRRO.

panelists' ratings reveals that two panelists (compared to the other four panelists) consistently rated a larger proportion of the items as inappropriate. For instance, for the high school grade span, these raters indicated approximately 100 AGLEs were inappropriate; the highest number rated as inappropriate by another rater was 9. Clearly, raters were divided on the proportion of AGLEs that was age appropriate.

**LAL Criterion 3: Standards Fidelity**

- a. Content Centrality** - *The focus of achievement maintains fidelity with the content of the original grade level standards.*

To evaluate this criterion, panelists provided ratings indicating their judgments on the degree of content match between the AGLEs and Missouri Show-Me Standards for Communication Arts. First, we asked panelists to provide a simple evaluation (yes or no) of whether the Show-Me Standard listed as linked with the AGLEs did, in fact, match. For those AGLEs judged as matched to the designated standard, we then asked panelists to provide a second rating to indicate *how well* the AGLE linked to the standard.

Concerning overall content match, Table 3.3 shows that almost all the Communication Arts AGLEs were matched to the Show-Me Standards.

**Table 3.3. Mean Percent of AGLEs Linked to Show-Me Standards**

Grade Span	Percentage of AGLEs Linked to AGLEs
3–5	98.4%
6–8	98.7%
HS	99.0%

For the second evaluation, panelists reviewed each grade-level AGLE for the degree of link to the central content targeted by the standards. In this case, panelists used the following 4-point scale to determine how well the AGLE reflects the standard content:

1	2	3	4
No Link	Weak Link	Moderate Link	Close Link

In terms of an acceptable level for this criterion, at least 90% of extended standards should be rated as 'moderate' or 'close' to the full standards. Table 3.4 shows that each set of grade-level AGLEs failed to meet this minimum. Panelists found over half the AGLEs to link weakly with the standards; at each grade level, the large majority of panelists rated AGLEs as weakly or moderately linked with the standards. All panelists rated fewer than 90% of AGLEs as moderately or closely linked.

**Table 3.4. Mean Number of AGLEs at Various Levels of Content Centrality**

Grade Span	Number of AGLEs at Grade Span	Content Centrality Rating	Mean	SD	Percentage of AGLEs per Rating	Number of Panelists Rating at Least 90% of AGLEs Moderate or Close <sup>a</sup>
3–5	150	No link	1.8	2.8	1.2%	0 of 6
		Weak link	83.7	16.6	55.8%	
		Moderate link	46.2	11.2	30.8%	
		Close link	18.3	20.9	12.2%	
6–8	179	No link	2.5	2.6	1.4%	0 of 6
		Weak link	95.3	19.9	53.3%	
		Moderate link	59.2	15.6	33.1%	
		Close link	22.0	22.9	12.3%	
HS	193	No link	1.8	2.4	0.9%	0 of 6
		Weak link	100.8	25.5	52.2%	
		Moderate link	64.3	13.8	33.3%	
		Close link	26.0	27.1	13.5%	

<sup>a</sup> Percentages are based on actual ratings. Missing data were excluded from the numerator and denominator.

***b. Performance Centrality - The focus of achievement maintains fidelity with the specified performance.***

We asked panelists to directly compare the performance expectations in the AGLEs with the full content standards. Panelists evaluated the language of each AGLE to terms of whether the expectations are the same, partly similar, or differ entirely from what is expected in the corresponding standards. For example, if the standard requires students to ‘compare and contrast’ traits, and the AGLE asks students to ‘group’ or ‘categorize’ based on traits, these expectations are parallel. If a standard expects students to ‘identify and explain’ while the AGLE asks students to ‘identify’ only, these expectations are partly similar. When students are asked to ‘distinguish between’ in the standard but the AGLE requires students to ‘recognize’, then the expectation for demonstrating knowledge is different. Table 3.5 shows the results of this comparison. For acceptability, at least 90% of the AGLEs should be rated as ‘partly similar’ or the ‘same’ when compared with the full content standards.

**Table 3.5. Mean Number of AGLEs at Various Levels of Performance Centrality**

Grade Span	Number of AGLEs at Grade Span	Content Centrality Rating	Mean	SD	Percentage of AGLEs per Rating	Number of Panelists Rating at Least 90% rated Similar or Same <sup>a</sup>
3–5	150	Differ Entirely	2.7	2.6	1.8%	6 of 6
		Partly Similar	15.3	15.5	10.2%	
		Same	132.0	14.7	88.0%	
6–8	179	Differ Entirely	2.3	2.0	1.3%	6 of 6
		Partly Similar	15.0	15.5	8.4%	
		Same	161.7	14.3	90.3%	
HS	193	Differ Entirely	2.5	2.3	1.3%	6 of 6
		Partly Similar	16.2	16.9	8.4%	
		Same	174.2	16.7	90.2%	

<sup>a</sup> Percentages are based on actual ratings. Missing data were excluded from the numerator and denominator.

The AGLEs for each grade surpassed the minimum level of acceptability for all raters across grade spans, with the majority of the content expectations rated as requiring the same or a similar type of performance as the standards. All panelists independently arrived at this acceptable level.

***LAL Criterion 5: Content Differentiation - There is some differentiation in content across grade levels or grade bands.***

This criterion focuses on whether the content expectations change appropriately between grade spans (e.g., whether the AGLEs for Grades 6–8 are sufficiently differentiated from the AGLEs for Grades 3–5 and the AGLEs for high school). For this reason, the evaluation of content differentiation involves a comparison *between* grade span-level content expectations. Panelists rated the AGLEs between grade spans as to whether they evidenced broader, deeper, and newer knowledge, as well as if certain expectations represented prerequisite skills (see Appendix B for a more detailed explanation of the categories). Across these categories, panelists indicated whether the content differentiation of AGLEs between grade spans was clear (C), partial (P), limited (L), or None (N). These ratings were reached collaboratively among panelists to achieve consensus evaluations. According to the LAL method, content expectations should show evidence of at least partial differences in content between grades on the dimensions of Broader, Deeper, Prerequisite, and New. Table 3.6 presents these results.

**Table 3.6. Consensus Ratings on Content Differentiation between Grade Span MAP-A AGLEs for Communication Arts**

Criterion	3–5	6–8	HS	Selected Notes from Panelists
Broader	L	L	L	AGLEs are repeated at each grade level. Some variability exists. For instance, RD2 does not show breadth or depth across grade spans, but WP3 shows increasing breadth.
Deeper	L	L	L	Some strands and sub-strands (e.g., RD5, RC4) show good progression, but some AGLEs do not increase in depth across grade spans (e.g., RC3.2)
Prerequisite	N	N	N	Not a spiral curriculum. Prerequisites are repeated through grade span.
New	L	L	L	Some sub-strands show new content, but less than 25% progression if all the repetition was taken out.
Identical <sup>a</sup>	L	L	L	Clearly identical, limited differentiation. <sup>b</sup>

<sup>a</sup> None (N) is an appropriate rating for this dimension because it indicates that no identical content is evident between grades.

<sup>b</sup> Given raters' comment, we believe they misunderstood the reverse-coding for this dimension and meant to indicate limited differentiation, not limited existence of identical content expectations.

As indicated in Table 3.6, panelists observed a considerable amount of repetition in the Communication Arts AGLEs across grade spans. Overall, their ratings indicate a consistent lack of differentiation from earlier grade spans to later grade spans. Panelists indicated that many of the AGLEs were clearly identical. Although they were able to indicate some specific examples of strands and sub-strands that demonstrated differentiation across grade spans, generally panelists considered differentiation across the Communication Arts AGLEs to be limited.

***LAL Criterion 7: Performance Accuracy - The potential barriers to demonstrating what students know and can do are minimized in the assessment to increase measurement accuracy of student performance.***

Panelists evaluated whether students could reasonably demonstrate the content and performance expected in the AGLEs by providing several different ratings. First, we asked panelists to determine the level of communication required by each AGLE in order for students to demonstrate knowledge. The common categories applied, according to the LAL method, include the following three ability levels for students with significant disabilities<sup>9</sup>:

<sup>9</sup> In addition to rating descriptions in the LAL manual, these definitions for communication levels have been expanded for clarity based on descriptions in a document published by the North Carolina Department of Public Instruction, Exceptional Children Division:  
<http://www.ncpublicschools.org/docs/ec/instructional/extended/extendedcontentstandards.ppt>

- Pre-symbolic - student may demonstrate intentionality by showing interest, focus, or desire for a result through behavior; can use idiosyncratic gestures, sounds, or purposeful movements but no discrimination between pictures or other symbols.
- Early symbolic - student demonstrates emerging knowledge of symbols with some recognition of symbol-object relationships.
- Symbolic - student has broad knowledge of and can communicate consistently with symbols (e.g., pictures) or words (e.g., speech, assistive technology, signs).

In general for extended standards and alternate assessments, it is expected that teachers and test administrators modify the content to instruct and assess students at the appropriate level based on their IEPs. However, if the level of communication required in the extended standards document is always at the highest level (symbolic), it becomes more difficult for accommodations and supports to be provided to students at the more basic levels of communication and still retain comparability in content and performance. Teachers and assessment administrators find it much less problematic to increase the scope of content and performance expected for individual students than attempt to pare down. For these reasons, HumRRO's position on this issue is that it is preferable that the access point of most extended standards (and assessment tasks) be pre-symbolic<sup>10</sup>. Thus, the minimum level of acceptability is that the access point for at least 90% of the AGLEs should be pre-symbolic.

Table 3.7 presents panelists' mean ratings on the communication levels needed to demonstrate content knowledge for each set of grade span AGLEs.

---

<sup>10</sup> The authors of the LAL method suggest a different perspective that focuses more on symbolic communication. For more information, please refer to the *Links for Academic Learning: An Alignment Protocol for Alternate Assessments Based on Alternate Achievement Standards*.

**Table 3.7. Mean Number of AGLEs Rated at Each Level of Symbolic Communication**

Grade Span	Number of AGLEs at Grade Span	Level of Symbolic Communication Required	Mean	SD	Mean Percentage of AGLEs per Rating	Number of Panelists Rating at Least 90% of Rated AGLEs at Pre-symbolic <sup>a</sup>
3–5	150	Pre-symbolic	10.7	2.9	7.1%	0 of 6
		Early symbolic	24.8	19.9	16.6%	
		Full symbolic	114.5	19.2	76.3%	
6–8	179	Pre-symbolic	11.5	2.9	6.4%	0 of 6
		Early symbolic	24.0	17.4	13.4%	
		Full symbolic	143.3	18.0	80.1%	
HS	193	Pre-symbolic	11.0	2.8	5.7%	0 of 6
		Early symbolic	25.3	19.6	13.1%	
		Full symbolic	156.5	19.8	81.2%	

<sup>a</sup> Percentages are based on actual ratings. Missing data were excluded from the numerator and denominator.

<sup>b</sup> Data from only 6 raters are used in this analysis; one rater asked to have her data removed for this criterion because she realized she misunderstood the directions and consistently misrated her AGLEs.

Based on these panelists' ratings, none of the grade spans met the minimum acceptability requirement of 90%. Panelists rated the majority of Communication Arts AGLEs at all three grade spans to be accessible primarily at the fully symbolic level. These ratings were consistent across panelists; although there was some variability in specific ratings, all of the raters assigned considerably fewer than 90% of the Communication Arts AGLEs to the pre-symbolic level. In fact, none of the raters assigned more than 10% of the AGLEs to the pre-symbolic level at any grade span. The high percentage of AGLEs requiring full symbolic communication for access represents a threat to performance accuracy for the Communication Arts MAP-A.

The second type of rating performed by panelists focused on general accessibility to students based on various types of disabilities (beyond communication abilities). For example, can students with visual impairments, an inability to follow instructions, or a need for assistive technology demonstrate the knowledge expected by these AGLEs? Panelists provided simple 'yes' (accessible to all) or 'no' (not accessible to some groups) responses to indicate their judgments. If they gave a 'no' rating, we asked panelists to provide some explanation of which groups would be disadvantaged and why in a Comments section. Table 3.8 includes the percentage of AGLEs that were judged as accessible to all groups.

**Table 3.8. Communication Arts AGLEs Rated as Accessible to All Students**

Grade Span	Number of AGLEs at Grade Span	Mean	SD	Mean Percentage of AGLEs Rated Accessible	Number of Panelists Rating at Least 90% of AGLEs Accessible <sup>a</sup>
3–5	150	141.5	11.1	94.3%	4 of 6
6–8	179	170.0	11.8	95.0%	5 of 6
HS	193	185.0	10.2	95.9%	5 of 6

<sup>a</sup> Percentages are based on actual ratings. Missing data were excluded from the numerator and denominator.

Panelists generally indicated that a high percentage of Communication Arts AGLEs were accessible to all students. For all three grade spans, panelists on average rated more than 90% of the AGLEs as accessible to all students. In the instances where individual raters did not judge at least 90% of the AGLEs to be accessible to all students, they never judged less than 80% to be accessible. Overall, these results indicate good support for the accessibility of the Communication Arts AGLEs.

The third type of rating performed by panelists focused on the general extent to which accommodations or supports could be provided to enable students with various types of disabilities to access the content. For example, can students with hearing impairments or a need for assistive technology demonstrate the knowledge expected by these AGLEs if appropriate supports were provided? Panelists provided simple ‘yes’ (accessible with accommodations to all) or ‘no’ (not amenable to accommodations or supports for some groups) responses to indicate their judgments. If they gave a ‘no’ rating, we asked panelists to provide some explanation of which groups would be disadvantaged and why in a Comments section. Table 3.9 includes the percentage of AGLEs that were judged as accessible to all groups.

**Table 3.9. Communication Arts AGLEs Rated as Amenable to Accommodations for All Students**

Grade Span	Number of AGLEs at Grade Span	Mean	SD	Mean Percentage of AGLEs Rated Amenable	Number of Panelists Rating at Least 90% of AGLEs Amenable <sup>a</sup>
3–5	150	145.3	9.5	96.9%	5 of 6
6–8	179	172.2	10.8	96.2%	5 of 6
HS	193	188.8	8.8	97.8%	5 of 6

<sup>a</sup> Percentages are based on actual ratings. Missing data were excluded from the numerator and denominator.

As Table 3.9 demonstrates, panelists also rated the Communication Arts AGLEs to be highly amenable to accommodations or supports for all three grade spans. On average, panelists indicated more than 90% of the AGLEs at all three grade spans were amenable to accommodations and/or supports for a wide range of student disabilities. Because Missouri teachers select AGLEs to build

tasks for their own students, they might be particularly aware of ways to accommodate the AGLEs to enable access for a wide variety of students.

Following their ratings of individual AGLEs, panelists were asked to reflect holistically on the entire set of AGLEs and reach consensus on the extent to which they include potential barriers that limit student learning. Panelists indicated some of the Communication Arts AGLEs would be problematic for students with visual impairments or students nonverbal students who communicate through pictures. Consistent with their individual AGLE communication ratings, they also indicated it would be challenging for students without clear, intentional communication to demonstrate their knowledge and abilities for AGLEs other than those dealing with attention.

Finally, we asked panelists to evaluate whether accommodations, modifications, and supports were sufficiently defined to enable standardized administration. They noted that accommodations, modifications, and supports are based on individual student needs, so they are not sufficiently defined to maintain standardization. Given Missouri's MAP-A system, in which teachers select AGLEs and design tasks accordingly, it is unsurprising that panelists would provide high ratings for accessibility and being amenable to accommodations or supports but suggest that there is limited standardization across administrations.

### ***Alignment of Communication Arts Portfolio Assessments to AGLEs***

***LAL Criterion 1: Academic*** - *The content is academic and includes the major domains/strands of the content area as reflected in state and national standards (e.g., reading, mathematics, Communication Arts).*

Per the USDE (2005), alternate assessments counting towards Title I must assess students only on academic content, as opposed to functional life skills. Panelists were asked to judge the grade-level Communication Arts assessments as to whether each task focuses primarily on academics. Results of this analysis are presented in Table 3.10. To be considered acceptable, at least 90% of tasks should be rated as academic.

**Table 3.10. Mean Number of Communication Arts MAP-A Portfolio Tasks Rated as Academic by Panelists**

Grade Spans	Number of Tasks Rated at Grade Span	Mean Number of Tasks Academic		Mean Number of Tasks Functional		Mean Percentage of Tasks Rated Academic	Number of Panelists Rating More than 90% of Rated AGLEs Academic <sup>a</sup>
		M	SD	M	SD		
3–5	60	54.3	3.3	5.3	3.2	91.1%	4 of 6
6–8	60	45.2	5.8	14.8	5.8	75.3%	1 of 6
HS	60	46.3	6.8	13.2	6.8	77.9%	1 of 6

<sup>a</sup> Percentages are based on actual ratings. Missing data were excluded from the numerator and denominator.

As Table 3.10 shows, grade span 3-5 met the criterion of at least 90% of portfolio entries being rated as academic, but ratings for the other grade spans fell below this level. For Grades 6–8 and high school, panelists indicated that approximately one-fourth of the portfolio entries measured functional or foundational skills rather than academic content. Because only one of the six raters at these grade spans indicated at least 90% of the entries were academic in nature, it is clear these ratings are not the result of one or two panelists with aberrant ratings but rather result from a more consistent pattern across panelists. Missouri may wish to review the portfolio entries for the middle school and high school grade spans to ensure they are assessing academic content.

***LAL Criterion 2: Age Appropriate*** - The content is referenced to the student's assigned grade level (based on chronological age).

Panelists evaluated the performance tasks on whether the content and performance assessed students at an appropriate level linked to their assigned grade. Table 3.11 shows the mean number and percentage of tasks judged as adapted (linked) to grade level, inappropriate (off-grade), and neutral (not age-bound). For acceptable linkage, at least 90% of tasks must be judged adapted or neutral. The Grade 3–5 Communication Arts MAP-A portfolio entries surpassed the minimum requirement of 90% rated as “adapted” or “neutral”. At Grades 6–8 and high school, the 90% was not quite achieved; 86% of tasks in Grades 6–8 and 87% of tasks for high school were rated as adapted or neutral. However, it is worth noting that, for both of these grade spans, one single rater rated a disproportionate number of the tasks as inappropriate, pulling the overall group average down. For Grades 6–8, only half the raters indicated at least 90% of the entries were adapted or neutral, but for high school this single rater had a more substantial impact on scores. Ultimately, most panelists rated most portfolio entries as age appropriate, particularly for Grades 3–5 and high school.

**Table 3.11. Communication Arts MAP-A Performance Tasks at Various Levels of Age Appropriateness**

Grade Span	Number of Tasks Rated at Grade Span	Age Appropriateness Rating	Mean	SD	Mean Percentage of Tasks per Rating	Number of Panelists Rating at Least 90% of Rated Tasks Adapted or Neutral <sup>a</sup>
3–5	60	Adapted	49.5	10.3	82.5%	5 of 6
		Neutral	4.7	5.7	7.8%	
		Inappropriate	5.8	8.8	9.7%	
6–8	60	Adapted	48.2	8.6	80.5%	3 of 6
		Neutral	3.5	2.7	5.8%	
		Inappropriate	8.2	8.4	13.6%	
HS	60	Adapted	51.7	12.0	86.1%	5 of 6
		Neutral	0.7	0.8	1.1%	
		Inappropriate	7.7	11.7	12.8%	

<sup>a</sup> Percentages are based on actual ratings. Missing data were excluded from the numerator and denominator.

Ratings involved evaluation of the portfolio assessment relative to the AGLEs on all seven LAL criteria. As with comparisons of the AGLEs to the Missouri Show-Me Standards, most results for Communication Arts reflect mean ratings on a series of scales. Mean ratings were derived from frequency counts (per panelist) of how many performance tasks were rated at each response option. From these counts, we then calculated the mean number of performance tasks per response option (across panelists) for each rating scale. At least 90% of performance tasks must achieve acceptable ratings to demonstrate linkage to grade-level content for each LAL criterion.

### **LAL Criterion 3: Standards Fidelity**

#### **a. Content Centrality** - *The focus of achievement maintains fidelity with the content of the original grade level standards.*

Panelists rated tasks for content match to the AGLEs to determine the extent to which the tasks assess grade-level content. Several analyses were performed on these ratings. First, panelists reviewed the number of tasks that were linked to at least one AGLE. The panelists for each grade span felt that most of the tasks were aligned to the Communication Arts AGLEs as shown in Table 3.12. However, at each grade span some tasks were rated to have no alignment. On average, nearly 10% of the tasks were rated as not linked to the AGLEs for Grades 6–8 and more than 10% of the tasks were rated as not linked to the AGLEs for high school. It is advisable to examine the tasks at these grade spans for relevance to the AGLEs.

**Table 3.12. Mean Percent of Tasks Linked to AGLEs**

Grade Span	Percentage of Tasks Linked to AGLEs
3–5	96.9%
6–8	90.6%
HS	87.8%

We also asked panelists to evaluate *how well* the tasks targeted the AGLEs. For acceptability, at least 90% of tasks should be judged as moderately to closely linked with the AGLEs. Table 3.13 presents the mean number and percentage of tasks that fell into each category based on panelists' ratings.

**Table 3.13. Mean Percent of Performance Tasks at Various Levels of Content Centrality**

Grade Span	Number of Performance Tasks at Grade Span	Content Centrality Rating	Mean	SD	Percentage of Performance Tasks per Rating	Number of Panelists Rating at Least 90% of Performance Tasks Moderate or Close <sup>a</sup>
3–5	60	No link	0.7	1.2	1.1%	2 of 6
		Weak link	11.8	10.8	19.7%	
		Moderate link	18.8	15.5	31.4%	
		Close link	28.7	24.9	47.8%	
6–8	60	No link	2.2	2.0	3.6%	1 of 6
		Weak link	18.3	15.4	30.6%	
		Moderate link	14.0	10.3	23.3%	
		Close link	25.5	21.9	42.5%	
HS	60	No link	2.5	2.9	4.2%	0 of 6
		Weak link	18.3	11.2	30.6%	
		Moderate link	18.0	10.9	30.0%	
		Close link	21.2	18.5	35.3%	

<sup>a</sup> Percentages are based on actual ratings. Missing data were excluded from the numerator and denominator.

Panelists rated the majority of tasks across grades as linked sufficiently to the target content of the AGLEs. However, most panelists rated fewer than 90% as moderately or closely linked. These tasks may require review to determine if the content link could be improved with edits to the tasks. It is possible that improved teacher training could result in more closely linked tasks. This is most problematic for Grades 6–8 and high school, where, on average, approximately 20 of 60 tasks were rated as having 'no link' or a 'weak link.' At the high school

level, all panelists rated fewer than 90% of tasks as moderately or closely linked; at grade span 6–8, only one panelist rated 90% of tasks as moderately or closely linked.

***b. Performance Centrality*** - *The focus of achievement maintains fidelity with the specified performance.*

In addition to the targeted content, the alternate assessment tasks should retain the performance intended by the full content standards to some extent. For example, if the full content standards require students to ‘compare and contrast’ content, the AGLEs should require students to make some type of distinction. Table 3.14 shows the mean number of tasks rated as the same, partly similar, or differing entirely to the performance expectations of the corresponding standards. For acceptability, at least 90% of tasks should receive ratings of ‘Partly Similar’ or ‘Same.’

**Table 3.14. Mean Percent of Performance Tasks at Various Levels of Performance Centrality**

Grade Span	Number of Performance Tasks at Grade Span	Content Centrality Rating	Mean	SD	Percentage of Performance Tasks per Rating	Number of Panelists Rating at Least 90% of Performance Tasks Similar or Same <sup>a</sup>
3–5	60	Differ Entirely	1.3	1.8	2.2%	6 of 6
		Partly Similar	25.7	27.2	42.8%	
		Same	33.0	28.2	55.0%	
6–8	60	Differ Entirely	2.5	2.4	4.2%	6 of 6
		Partly Similar	23.0	25.6	38.5%	
		Same	34.2	26.7	57.3%	
HS	60	Differ Entirely	2.7	3.1	4.5%	5 of 6
		Partly Similar	19.7	21.7	33.1%	
		Same	37.0	25.3	62.4%	

<sup>a</sup> Percentages are based on actual ratings. Missing data were excluded from the numerator and denominator.

At grade spans 3–5 and 6–8, all panelists concluded that the minimum level of acceptability (90%) of tasks assessing students on at least some of the same performance expectations was met. Overall, ratings at the high school grade span also met this criterion.

**LAL Criterion 4: Content Coverage** (*Webb dimensions*) - *The content differs from grade level in range, balance, and DOK, but matches high expectations set for students with significant cognitive disabilities.*

For most alternate assessments, LAL Criterion 4 incorporates the Webb alignment statistics. These measures reveal the degree of content coverage, along with the extent of cognitive processing expected by the assessment. For example, one measure determines *which* content strands are covered by the assessment (referred to as Categorical Concurrence) based on panelists' judgments about the content targeted per test item or task. Results should correspond well with the state content standards and test blueprint documents. Panelists do not receive the test vendor's information about intended content target while giving their own ratings to retain an independent process.

The structure of the MAP-A as a portfolio-based assessment made the use of the Webb alignment statistics inappropriate. Many portfolio-based alternate assessments allow teachers to select the standard and corresponding entry (task) for their students. This information is included in the full portfolio as part of the assessment, and frequently the scoring process incorporates the correspondence between the standard and task selected by teachers as part of the score. For the MAP-A alignment review, panelists received intact portfolios, which included identification of the targeted AGLEs along with the performance task description. Thus, no differences exist between reviewers regarding the AGLE matched to each task. Furthermore, it would have been difficult to impossible for reviewers to determine independently the AGLEs intended for assessment by teachers without some context for three reasons: (a) teachers provide explanation or rationale within the portfolio for how the task should relate to the AGLE; (b) the Missouri AGLE document includes a voluminous number of content standards, often overlapping, per strand at each grade; and (c) some performance tasks, unfortunately, were poorly written by teachers or were vague.

Despite differences in assessment structure, the goals of the Webb alignment indicators remain relevant; it is still important to determine whether teachers adequately cover the breadth and depth of the AGLEs. Although panelists had knowledge of the target content, they still had the ability to confirm or reject the appropriateness of teachers' selections. As a result, HumRRO addressed the following core alignment issues by examining the content and distribution of performance tasks relative to the AGLEs using simple frequency counts based on teachers' selections, as opposed to reviewers' ratings.

- How many content categories were covered by performance tasks (comparable to goal of *categorical concurrence* measure)?
- How many standards (AGLEs) within content categories were targeted (comparable to goal of *range-of-knowledge correspondence* measure)?
- Which standards per content categories were targeted most often by teachers? In other words, do teachers tend to distribute

performance tasks across AGLEs, or do they write them for some AGLEs much more than others (comparable to goal of *balance-of-knowledge representation* measure)?

- Does the DOK of the entry match the DOK of the standard (true Webb measure of *depth-of-knowledge consistency*)?

Results for the Webb method are reported at the content strand level. The frequencies reported in tables below indicate the number of performance tasks that target each Communication Arts Strand based on teachers' selections identified in the portfolios. *If reviewers disagreed with teachers on the content targeted by performance tasks, we point to these discrepancies in the discussion below the reported results.*

***Breadth Across Content Categories (similar to Categorical Concurrence).*** In the previous section on Content Centrality under LAL Criterion 3, we presented results on whether, and how well, each task matched to content expectations. For this analysis, we report on *which* AGLEs were assessed by teachers. As a minimum criterion to reflect adequate coverage per strand, we considered the assessment to reflect adequate breadth if at least half (50%) of the total AGLEs within a strand were represented at least once in the sample of portfolios. This minimum criterion is less stringent than that established by Webb as well as the typical LAL alignment evaluation. We established a new criterion as Missouri's system would not meet the requirements of either because of the large numbers of AGLEs per strand and the small number of portfolio entries per subject area.

Table 3.15 summarizes the MAP-A alignment results for breadth across content categories. As Table 3.15 indicates, certain strands are assessed at each grade span. Shaded areas represent content not assessed at that grade level. Reading strands are assessed at all grade levels. Each portfolio entry is expected to address one of the content strands. The '% accurate' column represents the panelists' agreement that the content indicated by teachers assessed the content of the standards. The final column indicates the number of standards represented within each strand in the sample of portfolios compared to the total number of standards included in that strand. For example, among the 25 portfolio entries coded for Reading Processes, 5 of the 21 standards within that strand were represented. Only the strand for Writing Conventions, assessed only at Grades 3–5, met our minimum criterion for breadth across content categories. The middle and high school grades often had more AGLEs represented than were in a strand. Teachers routinely listed AGLEs indicated at earlier grades as the assessed content. For this reason, the standards were considered cumulative (or partly so) for the upper grade levels. No acceptability indication is given because of this anomaly in the data. No strand would have been determined to be acceptable had ratings been given.

It should be noted that this minimum criterion was established under the assumption that students would be instructed across multiple years and that the portfolio entries would be distributed across AGLEs representing various students' instructional programs during the grade span. The criterion we established does not refer to individual student portfolios for a given year. If it did, the minimum criterion for breadth across content categories would not have been met for any strand at any grade. This new criterion was developed to provide Missouri with useful information about which strands are being assessed more completely than others. The criterion may be more similar to Webb's range-of-knowledge correspondence indicator than to categorical concurrence. Evidence from this table also is presented in the range-of-knowledge section.

**Table 3.15. Summary of Breadth Across Content Categories Results for MAP-A Communication Arts Portfolios by Grade Level**

Title of Strand	Number of Tasks per Strand						Summary		
	Grades 3–5		Grades 6–8		High School		Standards with at Least One Task/Total Standards per Strand (Yes or No to Indicate if Minimum Criterion is Met)		
	Tasks Matched	% Accurate	Tasks Matched	% Accurate	Tasks Matched	% Accurate	Grades 3–5	Grades 6–8	High School
Reading Processes (RP)	25	97	19	93	21	94	5/21 (No)	8/29*	7/33*
Reading Development (RD)	5	100	11	94	9	73	16/44 (No)	11/54*	11/58*
Writing Conventions (WC)	30	97					17/22 (Yes)		
Writing Processes (WP)			30	88	30	88			13/35* 14/37*

\* AGLEs from prior grades were repeated at higher grades, indicating that the AGLEs for Communication Arts were considered cumulative. No acceptability ratings are provided for these grade spans.

**Depth-of-Knowledge Consistency.** Depth-of-knowledge (DOK) consistency measures the type of cognitive processing required by each performance task compared to the requirements implied by the content objectives. In this case, ratings on the MAP-A portfolios can be analyzed using the common DOK consistency measure established by Webb because reviewers made these judgments independently for the performance tasks and the AGLEs. Teachers do not identify DOK levels per performance task.

As part of the rating process, reviewers first determined the DOK level for each AGLE using a rating scale (see Appendix B for the LAL DOK level descriptions). Next, as they reviewed performance tasks, panelists rated the level of processing needed to perform the task using the same DOK rating scales. We compared these separate ratings on cognitive complexity to determine the extent to which the assessed performance matched the content expectations specified in the AGLEs. Tables 3.16–3.18 summarize the DOK consistency results for each grade level of the Communication Arts MAP-A assessment. Since reviewers evaluated DOK at the most specific level of the standards document (AGLEs), the table refers to consistency between the performance tasks and the AGLEs to which they were matched. Results are summarized in terms of the percentage of AGLEs assessed by performance tasks at or above the same cognitive level. Webb’s suggested criterion for this alignment indicator is the same as for a regular assessment—that is, at least 50% of the tasks should have complexity ratings at or above the level of the corresponding AGLE per content strand. The minimum criterion for DOK consistency was met for all strands except Writing Processes for Grades 3–5 and high school.

**Table 3.16. DOK Consistency for Communication Arts MAP-A, Grades 3–5: Mean Percent of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives**

Title of Strand	Mean Tasks per Strand	DOK Consistency						DOK Consistency Target Met
		% Tasks Below		% Tasks Same Level		% Tasks Above		
		M	S.D.	M	S.D.	M	S.D.	
Reading Processes (RP)	25	25.3	0.80	42.0	0.84	32.7	1.26	Yes
Reading Development (RD)	5	30.0	1.64	30.0	2.13	40.0	1.12	Yes
Reading Comprehension (RC)	0							
Writing Conventions (WC)	30	43.9	0.91	36.1	0.69	20.0	0.93	Yes
Writing Processes (WP)	NA							
Percent of strands with 50% of item DOK at or above objective DOK:								100%

**Table 3.17. DOK Consistency for Communication Arts MAP-A, Grades 6–8: Mean Percent of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives**

Title of Strand	Mean Tasks per Strand	DOK Consistency						DOK Consistency Target Met
		% Tasks Below		% Tasks Same Level		% Tasks Above		
		M	S.D.	M	S.D.	M	S.D.	
Reading Processes (RP)	19	10.5	.079	51.8	0.78	37.7	1.43	Yes
Reading Development (RD)	11	42.4	1.04	47.0	2.05	10.6	0.79	Yes
Reading Comprehension (RC)	0							
Writing Conventions (WC)	0							
Writing Processes (WP)	30	56.1	0.99	22.8	1.58	21.1	0.96	No
Percent of strands with 50% of item DOK at or above objective DOK:								66.7%

**Table 3.18. DOK Consistency for Communication Arts MAP-A, High School: Mean Percent of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives**

Title of Strand	Mean Tasks per Strand	DOK Consistency						DOK Consistency Target Met
		% Tasks Below		% Tasks Same Level		% Tasks Above		
		M	S.D.	M	S.D.	M	S.D.	
Reading Processes (RP)	21	11.1	0.80	57.9	0.83	31.0	1.08	Yes
Reading Development (RD)	9	38.9	1.38	25.9	1.79	35.2	1.31	Yes
Reading Comprehension (RC)	0							
Writing Conventions (WC)	0							
Writing Processes (WP)	30	62.2	1.09	21.7	1.48	16.1	0.75	No
Percent of strands with 50% of item DOK at or above objective DOK:								66.7%

Table 3.19 shows how DOK ratings were distributed for Communication Arts. Ratings were from 1–6 indicating attention, memorize/recall, performance, comprehension, application, and analysis/synthesis/evaluation, respectively. The mean, standard deviation, and range are presented to provide an indication that

typical DOK level portfolio entries are designed to assess. As can be seen in the table, portfolio entries for Communication Arts represented all potential DOK levels and were centered around the performance level. The standard deviations were fairly high, indicating considerable variance of DOK among the entries at each grade span.

**Table 3.19. Mean Percentage of Performance Tasks at Each DOK Level**

Grade Span	Mean	SD	Range
3-5	3.06	1.34	1-6
6-8	3.18	1.72	1-6
HS	3.08	1.41	1-6

**Breadth within Content Categories (similar to Range-of-Knowledge Correspondence).** Webb’s range-of-knowledge measure indicates how fully the performance tasks cover each of the AGLEs within each major content category. The assessed AGLEs within a strand should be linked with at least one performance task. Webb’s minimum level of acceptability for range-of-knowledge correspondence is that a mean of 50% of standards per content category should be matched to at least one assessment task. For the calculation of range for the MAP-A, we determined the frequency of performance tasks matched to each content strand by teachers. The minimum level of acceptability in this case was set at 50% of AGLEs per content strand matched to at least one performance task per grade span across all portfolios in the sample. We used the same criterion to indicate breadth across content categories under the assumption that students would have multiple learning opportunities across a grade span and that the portfolios represent a sampling of them. This assumption allows us to aggregate across portfolios and grade levels. A conventional interpretation of Webb’s breadth criterion is inappropriate because no single student’s portfolio (coded for four AGLEs total) could meet the minimum requirement. The final columns in Table 3.15 above contains this information as well as a summary indication of whether this minimum criterion was met. Only the strand for Writing Conventions (assessed only in Grades 3–5) met our minimum criterion for range (the same criterion used to indicate breadth across content categories above).

**Distribution of Assessment Tasks Among AGLEs (similar to Balance-of-Knowledge Representation).** The fourth measure of alignment included in the Webb method is *balance-of-knowledge representation*. This measure indicates the number of tasks linked to each standard per strand. The number of tasks should be distributed rather evenly among the standards for each strand to achieve a balanced assessment. The content balance is determined by calculating an index, or score, for each strand<sup>11</sup>. The large number of AGLEs per strand and the relatively small number of portfolio entries make this indicator

<sup>11</sup> The exact formula for calculating the balance index is explained in detail in Norman Webb’s (2005) alignment training manual: <http://www.wcer.wisc.edu/WAT/index.aspx>.

inappropriate for the MAP-A. Each portfolio typically assesses four different strands, and consequently four different AGLEs. The balance indicator would indicate that the content is evenly represented across strands, but this indication would not have meaning in this context. Table 3.15 shows that among the 60 portfolios sampled per grade level, about half were attributed to Writing strands and half to Reading strands. No single standard (AGLE) was ever represented by more than 5 portfolios among the 60. Most of those represented (and many were not represented at all) were assessed only on a single portfolio.

***Summary and Recommendations for Content Coverage.*** The MAP-A Communication Arts assessment is a portfolio system with only four entries per student. Each entry is designed to measure only one AGLE. There are as many as 58 AGLEs within a single strand and at least 21 AGLEs within each content strand. As such, typical measures of content coverage, such as Webb's alignment criteria, are not appropriate; there is simply no way to represent the breadth of the content indicated by so many AGLEs in a single portfolio. The MAP-A Communication Arts assessment does not meet traditional minimum alignment criteria for breadth of content coverage, breadth within content categories, or distribution of assessment tasks among AGLEs.

In order to provide Missouri with information regarding the relative strengths and weaknesses of the Communication Arts MAP-A, we aggregated portfolios across grade spans under the assumption that students could receive instruction and be assessed on multiple AGLEs within strands as they progressed from grade to grade. We also established new minimum criteria as a means of pointing out where the largest gaps in content coverage across portfolios occurred. These indicators are informative only, and should not be used as substitutions for more stringent alignment indicators.

The one exception to the rule for these analyses was DOK consistency. It was possible to directly compare the DOK of the portfolio entries with that indicated by the AGLEs. With the exception of Writing Processes, which is only assessed at Grades 6–8 and high school, the portfolios met Webb's criteria of 50% or more entries at or above the DOK of the AGLE. Teachers seem to be preparing portfolio tasks for students who have DOK requirements that are as high as or higher than indicated by the standards.

There are currently too many AGLEs to be represented by any student's portfolio. This would be true even if the portfolio was tripled in size (e.g., aggregated across 3 grade levels). Missouri may wish to consider either reducing the number of AGLEs to be assessed or altering the manner of assessment to include additional assessment items. There is currently no way to ensure that the assessment scores represent the scope of the Communication Arts content indicated by the standards.

**LAL Criterion 5: Content Differentiation** - *There is some differentiation in content across grade levels or grade bands.*

As with the evaluation of the AGLEs, LAL Criterion 5 focuses on whether the content increases in depth, breadth, and complexity at higher grade levels. Panelists achieved consensus ratings on the amount of content differentiation of the Communication Arts MAP-A performance tasks between grade span tests (higher and lower). Table 3.20 shows panelists' consensus ratings across the various dimensions using the rating scheme of clear (C), partial (P), limited (L), or none (N). For acceptability, each test should demonstrate evidence that it possesses at least partially different content per dimension, relative to higher or lower grade tests.

**Table 3.20. Consensus Ratings on Content Differentiation between Grade Span MAP-A Assessments for Communication Arts**

Criterion	3–5	6–8	HS	Selected Notes from Panelists
Broader	L	L	L	Same objectives were seen over and over. AGLEs are repeated at each grade level. A lot of (.1s), attending, matching, spacing, capitalization and punctuation.
Deeper	L	L	L	Extremely limited using lower level skills. Foundation skills were more widespread than academics; DOKs were lower level.
Prerequisite	C	C	N	Lower level prerequisites were being used across the grade levels.
New	N	N	L	New skills were not being chosen. Tasks were at lower grade level. All tasks could be accomplished at the lower grade levels (e.g., student read text, looked up word).
Identical <sup>a</sup>	C	C	C	Clearly identical. Limited differentiation.

<sup>a</sup> None (N) is an appropriate rating for this dimension because it indicates that no identical content is evident between grades.

As captured in Table 3.20, panelists saw only limited evidence of content differentiation across most dimensions of the Communication Arts MAP-A performance tasks. Panelists reported considerable repetition in tasks across grade spans and noted that most tasks were at a lower DOK level, appropriate for a lower grade span, or were even more foundational than academic in nature. Although panelists indicated there was limited evidence of introducing new tasks in high school, those tasks were considered to be low-level and to not build on prerequisite skills. Overall, the lack of content differentiation in Communication Arts performance tasks across grade spans seems to present an area for improvement.

**LAL Criterion 6: Achievement** - *The expected achievement for students is for the students to show learning of grade-referenced academic content.*

The sixth LAL criterion pertains to demonstration of student learning. The focus in this case is whether students are given the opportunity to demonstrate

academic skills or knowledge acquired from their coursework on the assessment. To determine the extent to which the MAP-A *enables* students to demonstrate this learning, panelists evaluated the scoring rubric and achievement level descriptors relative to the assessment. Panelists worked together to determine whether the assessment allowed for demonstration of high, low, or no evidence of student learning. These consensus ratings were made across several dimensions of learning, which are described below (adapted from Flowers et al, 2007):

- Level of accuracy - extent to which scoring makes clear distinctions in student responses.
- Level of independence - extent to which student performance is based on independent response without teacher supports.
- New learning - extent to which evidence of new learning is demonstrable based on use of baseline or pretest OR clear content differentiation between grade tests.
- Generalization across people and settings - extent to which students must demonstrate knowledge across people or settings to receive credit.
- Generalization across materials and activities - extent to which students must demonstrate knowledge across different types of materials (i.e., objects) or activities.
- Standard setting - extent to which achievement standards are distinct and based on demonstration of independent student performance.
- Program quality indicators - extent to which the inclusion of program characteristics (i.e., opportunities for instruction; access to materials; teacher qualities) is limited as part of student score.

For accurate assessment of achievement, most dimensions should receive ratings of 'high inference' regarding the ability to evaluate student learning.

Table 3.21 includes the group consensus ratings on the degree of student inference evident in the Communication Arts MAP-A assessment per grade level.

**Table 3.21. Degree of Inference Evident on Student Learning in Communication Arts MAP-A Assessments**

Dimensions	Grades 3–5	Grades 6–8	HS	Selected Notes from Panelists
Level of Accuracy	H <sup>a</sup>	H	H	Student independence across sampling fell within the mastery level ranging within 85 <sup>th</sup> to 100 <sup>th</sup> percentile on accuracy levels.
Level of Independence	L <sup>b</sup>	L	L	At least 50% show independence, however, teacher interaction and prompting unclear.
New Learning	N <sup>c</sup>	N	N	No baseline present. Weak differentiation, mastery could be made yearly.
Generalization Across People and Settings	L	L	L	Approximately 30% of portfolios were set up by teacher that student used, generalizations among various settings and teachers.
Generalization Across Materials and Activities	L	L	L	Approximately 30% of portfolios were set up by teacher that materials and activities allowed mastery.
Standard Setting	N	N	N	No standard setting was given. Standard is not listed on portfolios reviewed. MO does not set a standard for portfolios reviewed.
Program Quality Indicators	L	L	L	Portfolio reviewed APIs <sup>d</sup> were lower level. There is evidence of a child reaching independence.

<sup>a</sup> H = high student inference

<sup>b</sup> L = low student inference

<sup>c</sup> N = no student inference

<sup>d</sup> API = Alternate Performance Indicators, a subset of AGLEs

Panelists agreed that ratings for these student inference dimensions were consistent across grade spans. As Table 3.21 illustrates, panelists' ratings of the Achievement criterion for the Communication Arts MAP-A were mixed. For the Level of Accuracy dimension, panelists reported the MAP-A enabled a high degree of inference, suggesting the MAP-A requires students to demonstrate a high level of accuracy to obtain credit. For the dimensions of Level of Independence, Generalization Across People and Settings, Generalization Across Materials and Activities, and Program Quality indicators, panelists indicated the MAP-A enabled a low degree of inference. For the New Learning and Standard Setting dimensions, panelists indicated the MAP-A did not enable any inference about student learning. Among panelists' noted concerns were the lack of a baseline, the fact that similar performance could allow mastery in multiple grades, and the lack of standards for portfolios. They also indicated the level of prompting or assistance teachers provided to students for portfolio tasks was unclear.

**LAL Criterion 7: Performance Accuracy** - *The potential barriers to demonstrating what students know and can do are minimized in the assessment to increase measurement accuracy of student performance.*

This criterion is intended to evaluate the degree of accessibility of the assessment for all student groups who take it. Reduced access to the assessment tasks would decrease accurate measurement of these students' skills. As with the AGLEs, panelists rated tasks on the levels of communication required to respond and the access available to each type of student who takes the assessment. In addition, panelists evaluated each task on whether accommodations or supports can be provided for different types of students without substantially altering the target content.

Table 3.22 gives mean ratings by reviewers on the communication levels required of students in order to respond to the Communication Arts tasks. For acceptability, at least 90% of tasks should be rated as pre-symbolic for reasonable access by all students.

**Table 3.22. Mean Number of Tasks Rated at Each Level of Symbolic Communication**

Grade Span	Number of Tasks Rated Grade Span	Level of Symbolic Communication Required	Mean	SD	Mean Percentage of Tasks per Rating	Number of Panelists Rating at Least 90% of Rated Tasks at Pre-symbolic <sup>a</sup>
3–5	60	Pre-symbolic	2.2	2.0	3.6%	0 of 6
		Early Symbolic	11.3	6.7	18.9%	
		Full Symbolic	46.5	8.1	77.5%	
6–8	60	Pre-symbolic	6.0	5.0	10.0%	0 of 6
		Early Symbolic	14.7	6.2	24.4%	
		Full Symbolic	39.3	10.4	65.6%	
HS	60	Pre-symbolic	3.7	2.0	6.1%	0 of 6
		Early Symbolic	11.8	6.8	19.7%	
		Full Symbolic	44.5	8.0	74.2%	

<sup>a</sup> Percentages are based on actual ratings. Missing data were excluded from the numerator and denominator.

Based on these panelists' ratings, it is evident they believed only a small portion of the portfolio entries were accessible to students communicating at a pre-symbolic level. None of the raters at any grade span suggested the Communication Arts portfolio entries met the criterion of 90%, and the largest percentage of entries rated as pre-symbolic was 10% for Grades 6–8. In keeping with their AGLE ratings, panelists indicated that most portfolio entries were also accessible primarily at the full symbolic level.

Because portfolio tasks are designed for individual students, panelists were not asked to provide ratings of accessibility for the Communication Arts performance tasks as they did for the AGLEs. However, panelists did rate the general extent to which tasks were amenable to appropriate supports or accommodations. Panelists provided simple ‘yes’ (amenable to accommodations or supports) or ‘no’ (not amenable to accommodations or supports) responses to indicate their judgments. If they gave a ‘no’ rating, we asked panelists to explain their rationale in a Comments section. Table 3.23 includes the percentage of AGLEs that were judged as amenable to accommodations.

**Table 3.23. MAP-A Communication Arts Portfolio Entries Rated as Amenable to Accommodations for All Students**

Grade Span	Number of Tasks Rated at Grade Span			Mean Percentage of Tasks Rated Amenable <sup>a</sup>	Number of Panelists Rating at Least 90% of AGLEs Amenable <sup>a</sup>
		Mean	SD		
3–5	60	60.0	0.0	100%	6 of 6
6–8	60	59.0	2.4	100%	6 of 6
HS	60	60.0	0.0	100%	6 of 6

<sup>a</sup> Percentages are based on actual ratings. Missing data were excluded from the numerator and denominator.

Panelists were quite consistent in their judgments that Communication Arts portfolio entries were amenable to accommodations for all students. Missing ratings account for the variability for Grades 6–8; all ratings provided indicated all entries were amenable. As with the AGLEs, Missouri teachers might have been uniquely able to see the portfolio entries as amenable to accommodations because they were designed by considering particular student needs.

Following their individual ratings of the Communication Arts MAP-A performance tasks, panelists were asked to reflect holistically on the entire set of tasks and to reach consensus on the extent to which they include potential barriers that limit student learning. Consistent with their previous individual ratings, panelists did not think accessibility was a concern for assessment other than the level of communication required to respond to the performance entries.

### ***Inter-Rater Agreement Results***

We evaluated the extent to which panelists provided exactly the same ratings on items, which qualifies as a measure of absolute agreement (Shavelson & Webb, N. M., 2005; Tinsley & Weiss, 1975). Most of the LAL criteria require categorical ratings (e.g., Which AGLE is appropriate? Is AGLE accessible or not?) on the AGLEs and portfolio tasks. Several agreement measures exist to analyze categorical ratings (see Gwet, 2001; Webb, N. L., 2005). For these data, we applied a measure based on one developed by Norman Webb, which

provides a basic estimate of percent agreement between reviewers<sup>12</sup>. This analysis involves a pair-wise comparison (one-to-one) of each reviewer’s ratings with all other reviewers per item or task. Results then are averaged across reviewers per test form and evaluated as follows:

- Exact agreement                    1.00
- Good agreement                    0.70 to 0.99
- Adequate agreement                0.60 to 0.69
- Weak agreement                    0.59 or lower

As can be seen in Table 3.24, panelists agreed substantially in their ratings with good to exact agreement on six of the eight dimensions. The additional dimensions (Content Centrality and Age Appropriate Levels) had adequate agreement.

**Table 3.24. Pair-wise Comparisons on AGLE Ratings per Grade Level**

LAL Criteria	Min Agree	Max Agree	Mean across Strands	SD
Academic	50%	100%	95%	10%
API Match	67%	100%	99%	5%
Content Centrality	50%	100%	65%	17%
Performance Centrality	50%	100%	89%	13%
Age Appropriate	50%	100%	66%	17%
Communication Levels	50%	100%	82%	17%
Accessibility	50%	100%	95%	11%
Accoms_Supps	50%	100%	97%	7%

Table 3.25 presents these same types of agreement analyses on panelists’ ratings of the portfolio tasks. Again, an exact match indicates that all reviewers agreed in their ratings across the Strand level. The panelists had good to excellent average levels of agreement on five of the portfolio dimensions, and weak agreement on the remaining two dimensions.

---

<sup>12</sup> Refer to *Webb, N. L. (2005). Webb Alignment Tool (WAT): Training Manual* for a detailed discussion of the agreement analysis based on pair-wise comparisons.

**Table 3.25. Pair-wise Comparisons on Portfolio Ratings per Grade Level**

LAL Criteria	Min Agree	Max Agree	Mean across Strands	SD
Academic	50%	100%	88%	17%
API Match	50%	100%	92%	14%
Content Centrality	33%	83%	55%	10%
Performance Centrality	50%	100%	59%	12%
Age Appropriate	33%	100%	84%	17%
Communication Levels	33%	100%	79%	20%
Accoms_Supps	83%	100%	99%	3%

## References

- Flowers, C., Wakeman, S., Browder, D., & Karvonen, M. (2007). *Links for academic learning: An alignment protocol for alternate assessments based on alternate achievement standards*. Charlotte, NC: University of North Carolina at Charlotte. Retrieved from: [http://www.naacpartners.org/LAL/documents/NAAC\\_AlignmentManualVer8\\_3.pdf](http://www.naacpartners.org/LAL/documents/NAAC_AlignmentManualVer8_3.pdf)
- Missouri Department of Elementary and Secondary Education (January, 2008). *Missouri Assessment Program-Alternate (MAP-A): Instructor's guide and implementation manual 2009-2010*. ~~Roseville, MN~~ Jefferson City, MO: Missouri Department of Elementary and Secondary Education. Retrieved from: <http://dese.mo.gov/divimprove/assess/documents/2009-2010-MAP-A-Web-Instructors-Guide.pdf>
- No Child Left Behind Act of 2001. Public Law 107-110.
- North Carolina Department of Public Instruction. (unknown). *Extended content standards: Three levels of access so that all children can participate in the general education curriculum*. Charlotte, NC: North Carolina Department of Public Instruction. Retrieved from: <http://www.ncpublicschools.org/docs/ec/instructional/extended/extendedcontentstandards.ppt>
- 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 Missouri, National Center on Educational Outcomes.
- Thompson, S. J., Morse, A. B., Sharpe, M., & Hall, S. (2005). *Accommodations manual: How to select, administer, and evaluate use of accommodations for instruction and assessment of students with disabilities*. CCSSO State Collaborative on Assessment and Student Standards Assessing Special Education Students. Washington, DC. Retrieved from [http://www.osepideasthatwork.org/toolkit/accommodations\\_manual.asp](http://www.osepideasthatwork.org/toolkit/accommodations_manual.asp)
- U.S. Department of Education. (August, 2005). *Alternate achievement standards for students with the most significant cognitive disabilities*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education. Retrieved from <http://www.ed.gov/admins/lead/account/saa.html#guidance>.
- 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 mathematics and Math education*. Washington, D.C.: Council of Chief State Schools Officers.
- Webb, N. L. (1999). *Alignment of Math and mathematics standards and assessments in four states. (Research Monograph 18)*. Madison, WI: National Institute for Math Education and Council of Chief State School Officers. (ERIC Document Reproduction Service No. ED440852)



**Appendix A**  
**Alignment Results per Grade Span Assessment**

## ***Alignment Results***

The following tables include complete statistical results similar to the Webb alignment indicators (LAL Criterion 4: Content Coverage). There are several deviations from the typical Webb alignment method that were required due to the nature of the MAP-A portfolio entries. Deviations from the Webb method are explained by alignment indicator below.

### **Breadth Across Content Categories**

The breadth across content categories (similar to Webb's Categorical Concurrence) results for Grades 3-5, 6-8, and high school for the Communication Arts MAP-A are presented in Tables A- 1-3 below. Each table includes: the title of the strand (broadest analyzed category); the number of portfolio entries matched to that strand; and the percentage of portfolio entries determined by panelists to accurately represent the targeted content. This table deviates from typical categorical concurrence reports for several reasons. First, the data included in the table does not represent the panelists matching content descriptions from the standards to performance tasks. Each portfolio entry was marked with the standard it was designed to measure by the students' teachers. Panelists were asked to verify that the standards indicated by the teachers were accurate. There was therefore no variance among the panelists regarding the match to standard. Second, the number of portfolio entries per student was only 4. It would not make sense to expect each portfolio to represent all of the standards under each strand, so Webb's typical criterion of determining whether each strand is represented by at least one item is never achieved, nor reported. Similarly, because only 15 portfolios were analyzed per grade span, it would not be expected that every standard be represented across all portfolios. Also, analyses across portfolios would not be indicative of the concurrence between the standards and the assessment for any student. For this table, we chose not to include an indication of whether Webb's categorical concurrence criterion was met.

**Table A- 1. Content Representation for Communication Arts MAP-A:  
Number of Portfolio Entries per Strand**

Title of Strand	Number of Tasks per Strand					
	Grades 3-5		Grades 6-8		High School	
	Tasks Matched	% Accurate	Tasks Matched	% Accurate	Tasks Matched	% Accurate
Reading Processes (RP)	25	97	19	93	21	94
Reading Development (RD)	5	100	11	94	9	73
Reading Comprehension (RC)	0	NA	0	NA	0	NA
Writing Conventions (WC)	30	97	0	NA	0	NA
Writing Processes (WP)	NA	NA	30	88	30	88
Total	60		60		60	

### Depth-of-Knowledge Consistency

The Depth-of-Knowledge (DOK) consistency results for grade spans 3-5, 6-8, and high school of the Communication Arts MAP-A assessment are presented below. The tables present the results from the comparisons between the depth of knowledge expected in the standards 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 content standards along with the corresponding standard deviations. Results are separated by grade span. Standards with at least 50% of items at the same (or above) DOK level met the minimum criterion. The percent of strands summary percentage in the last row of the tables refers only to strands assessed by the portfolios in the sample (does not include non-assessed strands).

**Table A- 2. Depth-of-Knowledge Consistency for Communication Arts MAP-A, Grades 3-5: Mean Percent of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives**

Title of Strand	Mean Tasks per Strand	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Tasks Below		% Tasks Same Level		% Tasks Above		
		M	S.D.	M	S.D.	M	S.D.	
Reading Processes (RP)	25	25.3	0.80	42.0	0.84	32.7	1.26	Yes
Reading Development (RD)	5	30.0	1.64	30.0	2.13	40.0	1.12	Yes
Reading Comprehension (RC)	0							
Writing Conventions (WC)	30	43.9	0.91	36.1	0.69	20.0	0.93	Yes
Writing Processes (WP)	NA							
Percent of strands with 50% of item DOK at or above objective DOK:								100%

**Table A- 3. Depth-of-Knowledge Consistency for Communication Arts MAP-A, Grades 6-8: Mean Percent of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives**

Title of Strand	Mean Tasks per Strand	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Tasks Below		% Tasks Same Level		% Tasks Above		
		M	S.D.	M	S.D.	M	S.D.	
Reading Processes (RP)	19	10.5	.079	51.8	0.78	37.7	1.43	Yes
Reading Development (RD)	11	42.4	1.04	47.0	2.05	10.6	0.79	Yes
Reading Comprehension (RC)	0							
Writing Conventions (WC)	0							
Writing Processes (WP)	30	56.1	0.99	22.8	1.58	21.1	0.96	No
Percent of strands with 50% of item DOK at or above objective DOK:								66.7%

**Table A- 4. Depth-of-Knowledge Consistency for Communication Arts MAP-A, High School: Mean Percent of Performance Tasks with DOK Below, At, and Above DOK Level of Objectives**

Title of Strand	Mean Tasks per Strand	Depth-of-Knowledge Consistency						DOK Consistency Target Met
		% Tasks Below		% Tasks Same Level		% Tasks Above		
		M	S.D.	M	S.D.	M	S.D.	
Reading Processes (RP)	21	11.1	0.80	57.9	0.83	31.0	1.08	Yes
Reading Development (RD)	9	38.9	1.38	25.9	1.79	35.2	1.31	Yes
Reading Comprehension (RC)	0							
Writing Conventions (WC)	0							
Writing Processes (WP)	30	62.2	1.09	21.7	1.48	16.1	0.75	No
Percent of strands with 50% of item DOK at or above objective DOK:								66.7%



**Appendix B**  
**Workshop Instructions**

## MAP-A Panelist Task Instructions

	Rating Task	Documents Needed
1 a	DOK of Missouri Show-Me Standards (individual and consensus)	(1) Select Missouri Show-Me Standards (2) Rating Scale Code Descriptions – DOK scale
1 b	DOK of A-GLEs (individual and consensus)	(1) Alternate Grade-Level Expectations (A-GLEs) for your content area (2) Rating Scale Code Descriptions – DOK scale (3) A-GLE_DOK Rating Forms
2	Alignment Dimensions of A-GLEs (individual)	(1) Alternate Grade-Level Expectations (A-GLEs) for your content area (2) Rating Scale Code Descriptions (3) A-GLE_Alignment Dimensions Rating Forms (4) MAP-A Test Documents a. Test Administration Manual (includes Tasks and Scoring Rubric) b. Response Cards c. Presentation Pages (online version available for review)
3	Alignment Dimensions of Portfolio Entries (individual)	(1) Alternate Grade-Level Expectations (A-GLEs) for your content area (2) Rating Scale Code Descriptions (3) Portfolio_Alignment Dimensions Rating Forms (4) MAP-A Test Documents a. Instructor Administration Manual b. NCEO Considerations Guide c. Portfolios (accessible online)
4	Content differentiation across grades (consensus)	(1a) Alternate Grade-Level Expectations (A-GLEs) for your content area (1b) Rating Scale Code Descriptions – Content Differentiation (1c) Content Differentiation Rating Form_Standards (2a) Portfolios (accessible online) (2b) Content Differentiation Rating Form_Portfolios
5	Scoring criteria (consensus)	(1) MAP-A Test Documents a. Instructor Administration Manual b. Scorer Training Manual (pages 24-34)

		<ul style="list-style-type: none"> <li>c. Portfolios (accessible online)</li> <li>(2) Alternate Achievement Standards</li> <li>(3) Rating Scale Code Descriptions – Scoring Inferences</li> </ul>
6	Whole Test Considerations (consensus)	<ul style="list-style-type: none"> <li>(1) Whole Test_Rating Forms (1 per grade span)</li> <li>(2) Rating Scale Code Descriptions – Accessibility Dimensions (as reference)</li> <li>(3) NCEO Considerations Guide (as reference)</li> </ul>

## 1 Rate DOK of Standards

- A. Using the 'Select Show-Me Standards' printouts, assign a depth-of-knowledge rating to the Show-Me Standards most relevant to the MAP-A. You will provide a single rating per row on the combination of knowledge standards and goals listed. First, you will rate the standards independently. Then, we will come to consensus on the ratings (3/4 majority). Use the DOK rating scale as a guide to choosing ratings.
- B. Using the A-GLEs for your content area, assign a DOK rating to each blank (non-shaded) cell on the A-AGLE DOK Rating Form. Again, we will do individual ratings followed by consensus analysis.

Decision Rule: When rating DOK of content standards, go with the **higher** level if wavering between levels.

## 2 Rate the Alignment Dimensions of the A-GLEs (Rate ONLY Grades 3 through 12)

Using the rating scale codes on page 2 of the Rating Scale Code Descriptions, evaluate each individual API on all of the dimensions (columns) in the form. Repeat these same tasks for each grade span (3-5, 6-8, high school). All ratings will occur independently from other reviewers (if you have a question, ask a facilitator).

- A. Academic: Determine whether the content listed in the API is academic in nature (as opposed to functional or foundational skills).
- B. Standard Match: Referring to the standards listed at the bottom of the A-GLE handout, determine whether each API matches the standards listed by indicating 'Y' (yes) or 'N' (no). If the API does not match, please enter 'N' in this column. Then, refer to the full Show-Me Standards (copies available) to determine if another standard is appropriate and enter the standard number in Notes/Comments.
- C. Content Centrality: Indicate *how well* you think that the API actually links to listed standard on content. Please use a code of '1' (No Link) only when the API does not link to any standard at all.
- D. Performance Centrality: Determine the extent to which the API measures student performance expected in the standard. NOTE: If you chose a different standard, evaluate the API against that standard instead of the one listed.
- E. Age Appropriate: Evaluate whether the API is appropriate for the age (grade span) at which the content is measured.
- F. Level of Communication: Evaluate the level of communication required to demonstrate content knowledge. NOTE: Please consider the lowest functioning student who could access this API.
- G. Accessibility: Evaluate the degree of accessibility of this API for various disability groups. If the statement is accessible to all groups, enter a 'Y' (yes). If you think that the content is NOT accessible to some groups, enter 'N' (no) and provide an annotation in the Notes/Comments column to indicate those groups negatively affected. Use the Instructor's Administration Manual as references for your ratings.

- H. Accommodations and Supports: Determine whether the API is amenable to various forms of accommodations for students with various disabilities as well as supports needed by these students. Use the Instructor's Administration Manual and the NCEO Considerations Guide as references for your ratings.

### 3 Rate the Alignment Dimensions of the portfolio entries

Access the portfolios online following instructions given by ARC. Your facilitator will instruct you on which sets of portfolios to begin rating. You will be rating up to 15 portfolios per grade span (total of 45 portfolios across grade spans).

For each portfolio, examine the Task/Activity developed by the teacher relative to the API listed on the Entry - Data Form. Each portfolio contains multiple entries (4 each for Math and Comm Arts; 2 entries for Science). Rate each entry using the same rating scales as for the A-GLEs. Instructions for several scales differ; these different instructions are listed below:

- A. Academic: Determine whether the Task/Activity is primarily academic in nature.
- B. API Match: Determine whether the content/performance of the Task/Activity matches the API listed ( 'Y' or 'N'). If the content does not match, enter 'N' in this column, and review the A-GLEs to determine if another API is appropriate; enter that API in Notes/Comments.
- C. Content Centrality: Indicate *how well* you think the Task/Activity links to API on content. Use a code of '1' (No Link) only when the Task/Activity does not link to any API at all.
- D. Performance Centrality: Determine extent that Task measures student performance expected in the API. NOTE: If you chose a different API, evaluate the Task against that choice.
- E. Age Appropriate: Evaluate whether the Task/Activity is appropriate for the age (grade span) at which the content is measured.
- F. Level of Communication: Evaluate the level of communication required *for this student* to demonstrate content knowledge required by the task.
- ~~G. Accessibility: Not utilized for the portfolios.~~
- H. Accommodations and Supports: Determine whether the API is amenable to various forms of accommodations for students with various disabilities as well as supports needed by these students. Use the Instructor's Administration Manual and the NCEO Considerations Guide as references for your ratings.

### 4 Rate content differentiation across grades (A-GLEs and portfolios)

This task involves two separate global evaluations per grade span: (1) A-GLEs, (2) portfolios. Using the Content Differentiation Rating Forms, compare the content expectations of the A-GLEs between grade levels/spans (should see increases in breadth and depth). Provide a holistic judgment about the differences found between grade spans using the Rating Scale Code Descriptions. Perform the same ratings on the portfolios. You should have 6 completed rating forms (3 for A-GLEs, 3 for portfolios) when you are finished. Although you will be providing global ratings, please cite evidence from specific portfolios as often as possible.

### 5 Rate Whole Test Considerations based on the portfolios you reviewed within a given grade span. to demonstrating student knowledge

This is a global evaluation of the MAP-A. Instead of providing ratings for individual portfolios, consider your impressions of the set of portfolios you reviewed in each grade span. Use a separate Whole Test Rating form for each grade span. 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. You may perform this task as a group (if everyone else has completed their individual ratings). You should have 3 completed ratings forms (1 per grade span). Although you will be providing global ratings, please cite evidence from specific portfolios as often as possible.

**6 Rate scoring criteria (evaluate scoring rubric, rules, achievement descriptors, and instructions for administration)**

This is a global evaluation of the MAP-A. In addition to the portfolios, review the test documentation to get a sense of the extent to which students are capable of demonstrating independent learning AND whether it is possible to determine whether student work is independent or assisted. These documents *should* provide information about student performance, rather than assessment system or teacher performance. Refer to the Rating Scale Code Descriptions for explanation of codes. You should have 3 completed ratings forms (1 per grade span). Although you will be providing global ratings, please cite evidence from specific portfolios as often as possible.

<b>Rating Scale Code Descriptions</b>
---------------------------------------

**Depth of Knowledge (DOK) Scale**  
**(for use with the Full Content Standards, Alternate Standards, and Assessment)**

Level	DOK Description
0	None (no content clearly measured; too vague)
1	Attention (touch, look, vocalize, respond, attend).
2	Memorize/recall (list, describe (facts), identify, state, define, label, recognize, record, match, recall, relate).
3	Performance (perform, demonstrate, follow, count, locate, read).
4	Comprehension (explain, conclude, group/categorize, restate, review, translate, describe (concepts), paraphrase, infer, summarize, illustrate).
5	Application (compute, organize, collect, apply, classify, construct, solve, use, order, develop, generate, interact with text, implement).
6	Analysis, Synthesis, Evaluation (pattern, analyze, compare, contrast, compose, predict, extend, plan, judge, evaluate, interpret, cause/effect, investigate, examine, distinguish, differentiate, generate).

### Content and Performance Dimensions (for use with Alternate Content Standards and Assessment)

Category	Code	Description
Academic	A	Academic – content includes major domains/strands in State standards.
	F	Functional (or Foundational) – primary content focuses on practical skills, such as daily living skills (e.g., hand washing) or pre-academic (e.g., orienting a book, lines/marks on a page with pencil).
	N	Neither functional nor academic.
Standard Match		See full content standards and alternate content standards.
Content Centrality	1	No link
	2	Weak link
	3	Moderate link
	4	Close link
Performance Centrality	A	All - performance expectation is identical to content standard.
	S	Some - performance expectation partially matches content standard (content standard may include two different performance expectations, such as <i>Identify and explain</i> ).
	N	None - performance expectation is different from content standard
Age Appropriate	A	Adapted from grade-level content
	I	Inappropriate; off-grade content (should be taught/assessed at a higher or lower grade level).
	N	Neutral; content is not age-bound, but could be taught at any grade (NOT COMMON)

### Accessibility Dimensions (for use with Alternate Content Standards and Assessment)

Category	Code	Description
Levels of Communication	P	Pre-symbolic - student may demonstrate intentionality by showing interest, focus, or desire for a result through behavior; can use idiosyncratic gestures, sounds, or purposeful movements but no discrimination between pictures or other symbols.
	E	Early Symbolic - student demonstrates emerging knowledge of symbols with some recognition of symbol-object relationships.
	S	Symbolic - student has broad knowledge of and can communicate consistently with symbols (e.g., pictures) or words (e.g., speech, assistive technology, signs).
Access	Y	Yes, the standard is accessible to all students.
	N	No, some students cannot access the content of this standard or item (PLEASE provide annotation in Notes to explain).

<b>Accommodation/ Supports</b>	Y	Yes, students can access content with appropriate accommodations (e.g., audio) or supports (e.g., objects; assistive technology).
* Portfolio includes additional code	N	No, the content is not amenable to accommodations or supports.

**Content Differentiation Across Grades  
(for use to compare between grades for Alternate Standards and for  
Assessment)**

---

- (a) **broader**—higher-grade standards or items reflect broader application of target skill/knowledge;
- (b) **deeper**—higher-grade standards or items reflect deeper mastery of the target skill/knowledge
- (c) **prerequisite**—lower-grade standards or items reflects a different by prerequisite skill for mastery of the higher grade standard;
- (d) **new**—the higher-grade has a new skill or knowledge unrelated to skill/knowledge covered at prior grades; and
- (e) **identical**—higher-grade standards or items appear identical to one of the lower-grade standards.

## Scoring Inferences

(Use to evaluate level of independence evident in student work provided in portfolios)

Degree of Inference about Student Learning (based on scoring for each AA item or found in the standards setting information)

Criterion	High Student Inference Can clearly infer student showed learning	Low Student Inference Student performance mixed with educator performance	No Student Inference Can clearly infer student did not have to show any learning/ Teacher or program performance rated ("Raggedy Andy" would pass)	Rationale for Rating (provide where evidence found)
Level of accuracy	High level of accuracy (If one response; response is correct. If multiple responses, above 90% correct)	Lower level of accuracy or accuracy intermixed with teacher assistance to extent difficult to determine what student did.	Does not have to get items correct to receive credit.	
Level of independence	Only independent response receives credit (Students may receive a verbal question/ direction to respond but not told what response to make)	Credit given for responses in which student performs either without guidance after told or shown the exact response to make (verbal, model prompts, scaffolding) or are done after shown/ told exact response to make and also given some guidance to make the response (partial physical)	Credit given for responses made with hand over hand assistance	
New learning (important to AA because alternate achievement is not as clear as grade level)	Baseline or pretest provides support that this is new learning OR One time performance but clear differentiation of AA items by grade level (criteria 5)	One time performance AND grade level differentiation of AA items was not clear (criteria 5)	No baseline, pretest, and weak differentiation across grade level AA items suggest student could achieve proficiency by making same response year after year (criteria 5).	

Criterion	High Student Inference Can clearly infer student showed learning	Low Student Inference Student performance mixed with educator performance	No Student Inference Can clearly infer student did not have to show any learning/ Teacher or program performance rated ("Raggedy Andy" would pass)	Rationale for Rating (provide where evidence found)
<b>Generalization across people and settings</b> (Note: this is less important than conceptual generalization)	Tasks are demonstrated across people or settings for full credit	At least some tasks are demonstrated across more than one person or setting	Task is only demonstrated with one person in one setting	
<b>Generalization across materials and activities (conceptual generalization)</b>	Tasks are demonstrated across materials and activities or all standards have more than one task	At least some tasks are demonstrated across materials or activities; or there is more than one task for some standards	Task is only demonstrated with one specific material and activity; there is only one task per standard	
<b>Standard Setting</b>	Standard set for proficiency is based on independent student performance and high level of accuracy	Standard set for proficiency will require student show some independent responding and respond correctly above chance level	Standard set for proficiency is so low students could meet it with either chance responding or prompting that gives student the answer	
<b>Program Quality Indicators</b>	If program quality indicators are used, they are not factored into student score	If program quality indicators are used, they have minimal impact on student score (e.g., small portion of rubric)	Student score is heavily influenced by program quality indicators in rubric	