Missouri Department of Elementary & Secondary Education

2016 Alignment Study Report

Prepared by
John Smithson, Ph.D., Director, Surveys of Enacted Curriculum,
Wisconsin Center for Education Research at the University of Wisconsin-Madison
January, 2016

Table of Contents

Summary of Alignment Results 2

Dimensions of Alignment 4

Topic Coverage 5

Cognitive Demand..... 5

Balance of Representation 6

Overall Alignment 8

Findings by Subject 9

Mathematics 9

Topic Coverage *9*

Cognitive Demand *11*

Overall Alignment *12*

Language Arts 13

Topic Coverage *14*

Cognitive Demand *15*

Overall Alignment *16*

Discussion 18

What is "Good" Alignment 18

Discussion of Findings 19

Concluding Summary 20

About The Author 21

References 21

Appendix A: Content Analysis Procedures (separate file)

Appendix B: Mathematics Content Viewer (separate file)

Appendix C: Language Arts Content Viewer (separate file)

Introduction

In November, 2015 the Missouri Department of Elementary and Secondary Education commissioned a study of the alignment of the Mathematics and Language Arts summative assessment program for grades 3-8 compared to the state's academic content standards for these subjects. This report focuses on the results of that study, examining the alignment relationship between each grade of the state summative assessment compared to the state's content standards for that grade and subject.

The analytical approach taken for this study utilizes the Survey of Enacted Curriculum (SEC) methodology for conducting content and alignment analyses. The SEC alignment methodology has been widely published, having been employed by numerous researchers in studies of the relationship (or alignment) between various state standards, assessments, curriculum materials and classroom instruction (Porter, et. al. 2001, 2006, 2008, 2009). The approach has been similarly used by other states to meet federal requirements for third-party review of the alignment of state assessments to state standards. The analyses presented here were conducted by Dr. John Smithson, co-developer of the SEC methodologies and Director of the Surveys of Enacted Curriculum at the Wisconsin Center for Education Research.

The SEC approach to alignment involves a two-stage process, beginning with the collection of content descriptions for the alignment target(s) and the assessments for which alignment is to be measured. This first stage of data collection involves a standardized set of procedures for conducting content analysis of both standards and assessments. The process is described in more detail in Appendix A, but in short involves a team of content experts that review standards and assessment documents to collect in a systematic and quantifiable manner, descriptions of the academic content embedded in one or another document. This process yields a set of descriptive results that provide a 'picture' of the academic content embedded in one or another analyzed document. These descriptive results can be presented graphically using the unique 'content map' data displays that SEC is known for, providing the reader a graphical depiction of academic content arrayed across three dimensions; topic coverage, cognitive demand, and relative emphasis; essentially reporting descriptions of the 'what', 'how', and 'how much' of academic content.

The second stage in the SEC alignment process involves a detailed, statistical comparison between two selected content descriptions in order to address the alignment relationships thereby revealed. While alignment results can be reduced to a single alignment measure using this approach (referred to herein as 'overall' alignment), that measure compiles information from the three underlying dimensions that the content descriptions are themselves based upon. Thus in addition to the summary 'overall' alignment index (OAI), three sub-measures of alignment are also be reported to specifically address the degree of correspondence or alignment between the two content descriptions in terms of each of the three underlying dimensions; topic coverage(TC), cognitive demand(CD), and balance of representation(BR).

It should be noted that the SEC methodology provides a systematic and quantifiable method for examining the *degree* of alignment. It is not a qualitative examination of alignment, and thus does not attempt to define *good* alignment. By convention the author sets a bar of 0.5 to indicate an acceptable level of alignment. An assessment that falls below this mark is not considered 'out of alignment' unless it falls below 0.4. Results that fall between 0.4 and 0.5 are considered aligned, but deserving further attention in order to inform decisions relative to the addition of new items and dropping of old designed to increase alignment results in future iterations of the assessment. The detailed data results provided in

the Appendices of this report offer access to the fine grain data results to support this type of diagnostic analysis.

Based upon the many hundreds of documents analyzed, an alignment of 0.5 represents an 'above average' level of alignment. Conceptually, the 0.5 benchmark also recognizes the limitations in the amount of content an assessment can reasonably cover, and the psychometric requirements a test must meet. Moreover, the SEC method yields very detailed content descriptions that inevitably capture content that goes beyond just the 'targeted' content of a given assessment item, revealing related or implicit content that a student must 'know and be able to do', to provide a correct answer or to meet the goal of one or another item or standard. This suggests that there is also an upper level of what we would consider 'good' alignment. Precisely where this 'upper limit' should be located is a matter for additional research, but an alignment measure of 1.00 is certainly considered neither desirable nor possible using the SEC approach, given the limitations of any large scale assessment program.

In addition to providing summary measures reporting the degree of alignment of one or another test to a given alignment target (standard), the descriptive and analytic results included in the Appendices provide a resource to support continued assessment development, providing the assessment developer with information that can be used to target the development of new items or modifications of old ones. Thus the results reported here provide both summative and formative data for the reader's consideration.

Summary of Alignment Results

The table below reports the summary alignment indicator results for each subject and grade of the state assessment compared to the relevant content standards (CCSS) using four summary indicator measures.

The four alignment indicators are labeled as follows: topic coverage (TC), balance of representation (BR), cognitive demand (CD), and the summary overall alignment index (OAI). Note that the criteria for alignment are most stringent in the OAI measure, requiring agreement across all three component dimensions, rather than simply one. Thus achieving the 0.5 benchmark on the OAI measure represents a notable accomplishment.

The results reported in table 1 below show some variation across the summary measures and subject areas. One pattern that persists is that the balance of representation indicator (BR) is the lowest indicator measure reported for each subject. This is largely due to the inherent limitations of comparing a relatively small number of assessment items to the full breadth of grade-specific knowledge and skills identified in the state's standards for either academic subject.

For example, the BR measure of 0.33 indicated for grade 4 language arts in Table 1 is arrived at from the following calculation: 91% of the items in the grade 4 language arts assessment target 36% of the content emphasized in the state's learning standards. Clearly a large portion of the assessment covers standards-based content (91%). However with fewer than 100 items, and the need to devote more than a single item to any given content element, the assessment addresses only 36% of the content emphasized in the standards ($.91 * .36 = .33$). While another 100 items would no doubt go a long way in substantially increasing the balance of representation measure, the cost in terms of student burden and the additional resources required for development and administration make such a strategy untenable.

Table 1: Alignment Summary
Missouri State Assessments, Grades 3-8
Aligned to
Missouri Learning Standards

Mathematics	OAI	BR	TC	CD
Grade 3	**0.53**	***0.59***	***0.83***	***0.53***
Grade 4	**0.56**	***0.68***	***0.79***	***0.63***
Grade 5	**0.53**	***0.49***	***0.71***	***0.65***
Grade 6	**0.58**	***0.44***	***0.80***	***0.59***
Grade 7	**0.50**	**0.36**	**0.60**	**0.74**
Grade 8	**0.57**	**0.45**	**0.82**	**0.62**
Lang.Arts	OAI	BR	TC	CD
Grade 3	**0.53**	**0.37**	**0.68**	**0.79**
Grade 4	**0.49**	**0.33**	**0.69**	**0.68**
Grade 5	**0.51**	**0.42**	**0.74**	**0.66**
Grade 6	**0.53**	**0.46**	**0.79**	**0.62**
Grade 7	**0.54**	**0.40**	**0.79**	**0.65**
Grade 8	***0.53***	***0.44***	***0.79***	***0.65***

Looking beyond the BR measures of Table 1, it is clear that with the exception of balance of representation, the large majority of the indicator measures exceed the 0.5 benchmark. Even the 0.49 measure for grade 4 language arts falls within the measurement error for the 0.5 threshold, and it is no coincidence that the lowest OAI measure is also where the lowest BR measure is reported. Indeed, looking at the pattern of BR and OAI measures, it is clear they are strongly correlated, indicating that a lot of the variation in alignment results is related to balance of representation. The correlation is not perfect though, and a more careful examination of the patterns in Table 1 reveals that cognitive demand and topic coverage also play a role in the overall results. Understanding each of the sub-scales underlying the overall alignment measure is important to understanding the complex nature of alignment and the impact of assessment design.

Dimensions of Alignment

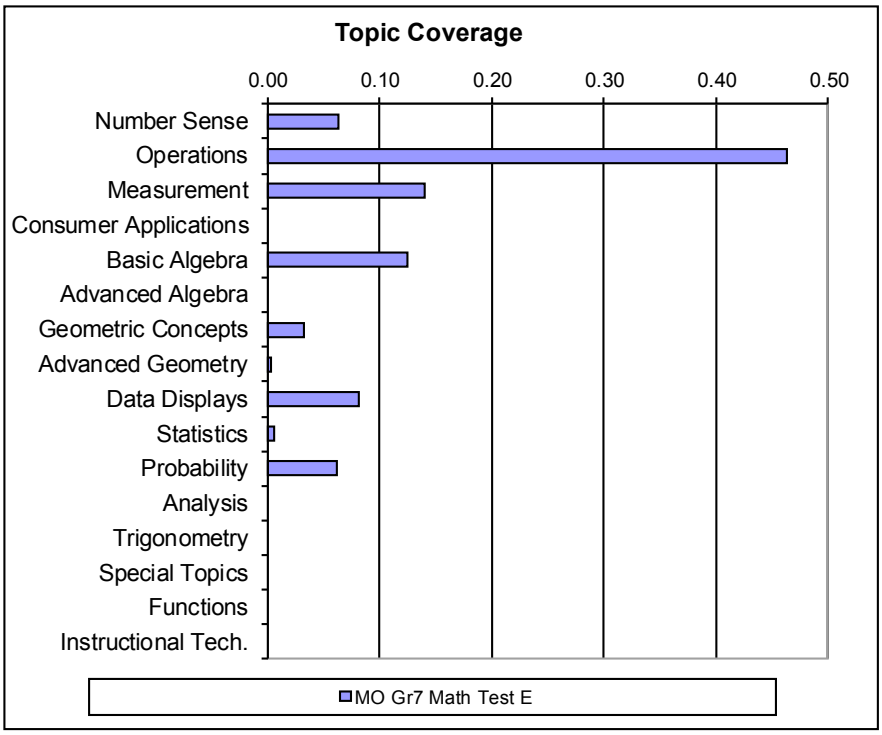
As already stated, SEC content descriptions and the alignment indicators they yield are based on three dimensions for describing academic content. The SEC content taxonomies provide a neutral framework for describing academic content that is independent of any one set of standards. This neutral framework makes it possible to describe many different types of curriculum related materials, even allowing the comparison of one set of standards to another; whether the standards of another state or another country, and much has been published along these lines (e.g. Porter, et.al. 2006, 2008).

The taxonomies used to describe mathematics and language arts content are based upon a two dimensional construct for describing content. Those two dimensions are most commonly defined by the oft-repeated "what students should know and be able to do". The 'what students should know' dimension is commonly referred to as topic coverage, while the 'be able to do' is captured by the cognitive demand dimension of the taxonomy.

Topic Coverage

Each of the SEC subject area taxonomies have a topic list ranging from roughly 200 (mathematics), to more than 350 (social studies) topics. These fine grain topics are organized into clusters of topic areas. For example the nearly 200 mathematics topics are arranged into sixteen topic areas such as Number Sense and Relationships, Operations, Measurement, Basic Algebra, Geometric Concepts, etc. Language Arts has 14 topic areas, science 27, and social studies 29. These topic areas serve as 'buckets' for organizing the more detailed 'fine grain' topics, and are used in coarse grain depictions of content descriptions, such as the one presented in Figure 1 below.

Figure 1: MO Gr. 7 Mathematics Assessment Topic Coverage



(Horizontal axis = proportion of assessment items.)

Figure 1 presents a coarse grain description of topic coverage, with each bar in the chart reporting the proportion of content associated with a given topic area; in this case depicting the grade 4 mathematics assessment. Since this is a coarse grain chart covering all mathematics content, the sum of each bar in the chart will sum to 1.00.

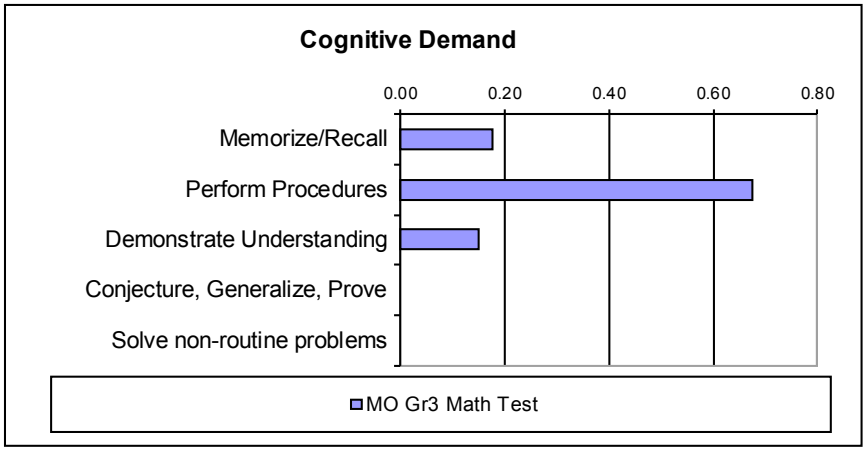
Cognitive Demand

The second dimension encoded into the SEC content taxonomies is cognitive demand. Cognitive demand (as defined in the SEC) represents a kind of middle ground between the well-known Blooms Taxonomy, and Norman Webb's (Webb, 1997) 'Depth of Knowledge' (DOK) measure used by many states and assessment developers for describing the level of cognitive challenge posed by a given test item or standards goal. In SEC this domain is referred to as cognitive demand, and consists of five categories for describing the skills that students are 'able to do'. While the specific labels for each category change from one subject to another, all four subjects draw from a common 5-category conceptual model for describing cognitive skills, i.e. what students are able to do. For example, the first category regardless of

subject references memorization, followed by a category to denote procedural skills (e.g. 'perform procedures' or 'conduct investigations'), and a third category denoting conceptual understanding (though qualified by some behavioral term such as 'demonstrate', 'communicate', or 'generate/create'. For all subjects the fourth category references analytic skills, though in mathematics this is labeled as 'conjecture/generalize'. The fifth category references the notions of integration, synthesis and novel problem solving.

Figure 2 reports the proportion of the grade 3 math test focused on one or another of the five categories of cognitive demand.

Figure 2: MO Gr. 3 Mathematics Assessment Topic Cognitive Demand



(Horizontal axis = proportion of assessment items.)

Results indicate that more than 50% of the items on the grade 3 test call for some evidence of procedural/computational skills, while the remaining items call for memorization or evidence of conceptual understanding. This emphasis on computation for a mathematics test is not surprising, though the lack of items judged to assess conjecture, generalization or mathematical proofs might be surprising in an environment emphasizing 'rigor' of mathematical content. This issue will be further addressed in the discussion section below.

Balance of Representation

Note that each of the charts above sum to 1.00 across the elements within each chart; that is, the results describe 100% of the test on one or the other dimension. The bar charts presented in Figures 1 & 2 thus provide a profile of the relative emphasis of the topics (Figure 1) and cognitive skills (Figure 2) indicated by the content analysis conducted on the document; in this case the Grade 3 mathematics assessment. This ability to report these proportional profiles or marginal measures implicitly draw upon yet another dimension of alignment underlying the SEC model, that attending to the balance of representation. Thus in some sense balance of representation is already embedded in these marginal results insofar as those descriptions incorporate the notion of relative emphasis or 'how much'.

At the coarse grain level balance or representation is defined as the product of two measures; the proportion of assessment items that target standards-based content, multiplied by the proportion of the standards that content represents. To repeat the example from the introduction: "91% of the items in the

grade 4 language arts assessment target 36% of the content emphasized in the state's learning standards" . Table 2 below provides a convenient summary of the balance of representation measures for each subject and the underlying interaction each is defined by.

As discussed above, it is difficult for a 100 item test to cover the full spectrum of academic content expected for a given grade level. This can be seen in the right-of-center column in Table 2, reporting the percent of standards represented on the test. The table indicates that each grade level assessment covers, in the case of mathematics, between about half and three-quarters of the content emphasized in the standards. For language arts, the percentage of the standards represented in the test tends to range between a third and a half of the language arts content expected for a given grade level.

The left-of-center column reports the percentage of the assessment items that target standards-based topics. Here we see that typically more than 80% of the mathematics and language arts assessments are targeting standards-based content. While one might expect that 100% of the standards should target standards-based content, it should be noted that the fine grain nature of the SEC content analysis will often identify incidental content embedded in an assessment item that, though perhaps not the primary target of an assessment item nor explicitly identified in the standards for that grade level, remains a relevant description of the content assessed by that item.

Table 2: Balance of Representation Detail

Mathematics	Pct. Test on Stands	Pct. Stnds. On Test	BR
Grade 3	**89.00**	***67.00***	***0.59***
Grade 4	**90.00**	***76.00***	***0.68***
Grade 5	**73.00**	***68.00***	***0.49***
Grade 6	**84.00**	**53.00**	**0.44**
Grade 7	**85.00**	**42.00**	**0.36**
Grade 8	**78.00**	**58.00**	**0.45**
Lang.Arts	Pct. Test on Stands	Pct. Stnds. On Test	BR
Grade 3	**0.88**	***0.42***	**0.37**
Grade 4	**0.91**	***0.36***	**0.33**
Grade 5	**0.98**	***0.43***	***0.42***
Grade 6	***0.88***	***0.42***	***0.46***
Grade 7	***0.94***	***0.43***	***0.40***
Grade 8	***0.89***	***0.49***	***0.44***

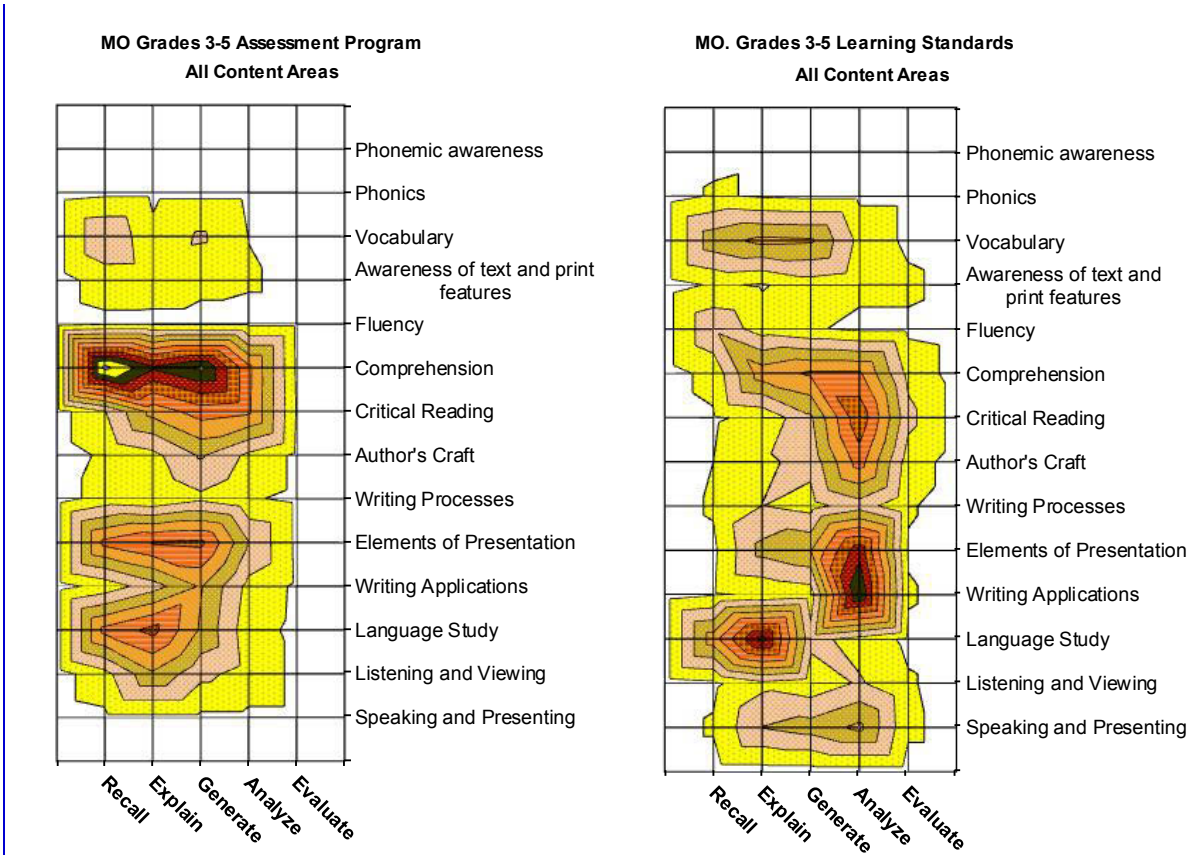
The important finding to be noted from Table 2 is that full coverage of the content emphasized for a given grade level is impractical. The question is really a matter of how much of the standards can be covered while accommodating the psychometric needs of the assessment. Note that the language arts assessment typically covers a smaller portion of the knowledge and skills in its standards than is reported for mathematics. As a subject, mathematics simply offers a more focused curriculum from which to target assessment items, while language arts by its very nature addresses a relatively larger array of knowledge and skills.

This accounts for much of the difference noted between the two subjects on balance of representation, and makes the question of balancing the needs of the assessment to provide psychometrically valid measures of student proficiency against the impetus to cover as broad a spectrum of the knowledge and skills targeted by state standards. The broader the terrain an academic subject covers the more difficult this question becomes, and the more likely that alignment for that subject area will likewise be somewhat lower than the results noted for another discipline will a more narrow focus of knowledge and skills to attend to.

Overall Alignment

When all three of these dimensions are brought together graphically, the resulting descriptions are displayed using a surface area chart format that accepts the three dimensional input of data. The result looks similar to a topographic map, and is referred to as a content map. When two such maps are displayed side by side, one can begin to discern the multiplicity of relationships existing between the two descriptions, setting the stage for the calculation of overall alignment. Figure 3 below provides an example of two such content maps, one depicting the elementary grade band results for the MO language arts testing program (left) to the grade-band learning standards.

Figure 3



Overall alignment is the most rigorous of the alignment measures as it measures alignment or agreement at the *intersection* of all three alignment dimensions. That is, to count toward overall alignment a match only counts to the extent it matches on all three dimensions; topic, cognitive demand and relative emphasis (balance of representation at the fine grain level). To achieve a measure of 0.5 is

therefore noteworthy, particularly since it indicates that the assessment is competitive in terms of its alignment characteristics relative to other states' own assessment programs, and exceeds the *average* degree of alignment we have found across the many dozens of assessments that have been analyzed. The OAI alignment result for the comparison made in Figure 3 is 0.54 (OAI).

Findings by Subject

Table 4: Mathematics Alignment Summary

Mathematics	OAI	BR	TC	CD
Grade 3	**0.53**	***0.59***	***0.83***	***0.53***
Grade 4	**0.56**	***0.68***	***0.79***	***0.63***
Grade 5	**0.53**	***0.49***	***0.71***	***0.65***
Grade 6	**0.58**	***0.44***	***0.80***	***0.59***
Grade 7	**0.50**	***0.36***	***0.60***	***0.74***
Grade 8	**0.57**	***0.45***	***0.82***	***0.62***

Table 4 below recounts the grade-specific summary results for mathematics reported in the Summary section above. In the sections to follow, more detailed results comparing the assessments to standards in terms of marginal measures for topic coverage and cognitive demand as well as overall alignment depicted by coarse grain content maps are presented for each subject. For brevity, and since the results are generally positive and similar for each grade, the results are aggregated to upper elementary (grades 3-5) and middle grades (grades 6-8) for each subject. The aggregation of results to these grade-bands provides a means for discerning the essential characteristics of the alignment issues that would be revealed by a longer grade-specific presentation of results, without the monotonous repetition of grade specific chart after grade-specific chart of fairly similar results. Indeed the aggregation to upper elementary and middle grades provides a better lens for discerning the shifts in content emphasis from upper elementary to middle grades. Nonetheless the full set of results are available in the Appendices, where the reader can easily access specific and detailed results for any grade and subject represented in this report.

Mathematics

The alignment results for the mathematics assessments reveal strong alignment results across all grades, and especially high alignment results for topic coverage, indicating a large degree of alignment across the sixteen content areas of mathematics.

Topic Coverage

Figure 5 provides a side by side comparison of aggregated topic coverage results for the grades 3-5 mathematics compared to the same grade span of learning standards. The upper bar in each pair in Figure 4 reports results for the assessment, while the bottom bar reports the comparable results for the standards. Where one bar extends beyond the other indicates topics that the assessments over- or under- emphasize (depending on which bar is longer) relative to the learning standards for mathematics.

Figure 5: Elementary (Gr.3-5) Topics Assessments to Standards

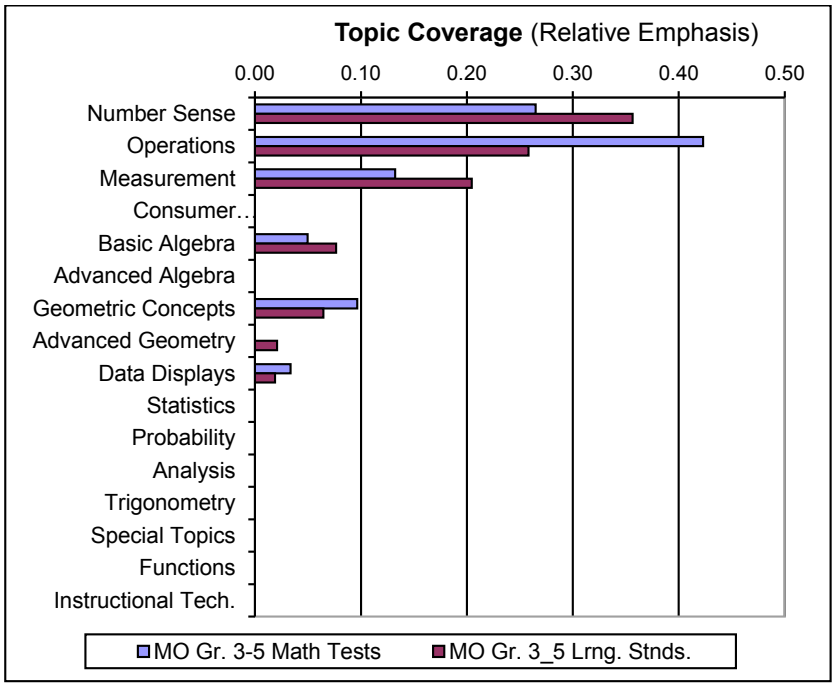
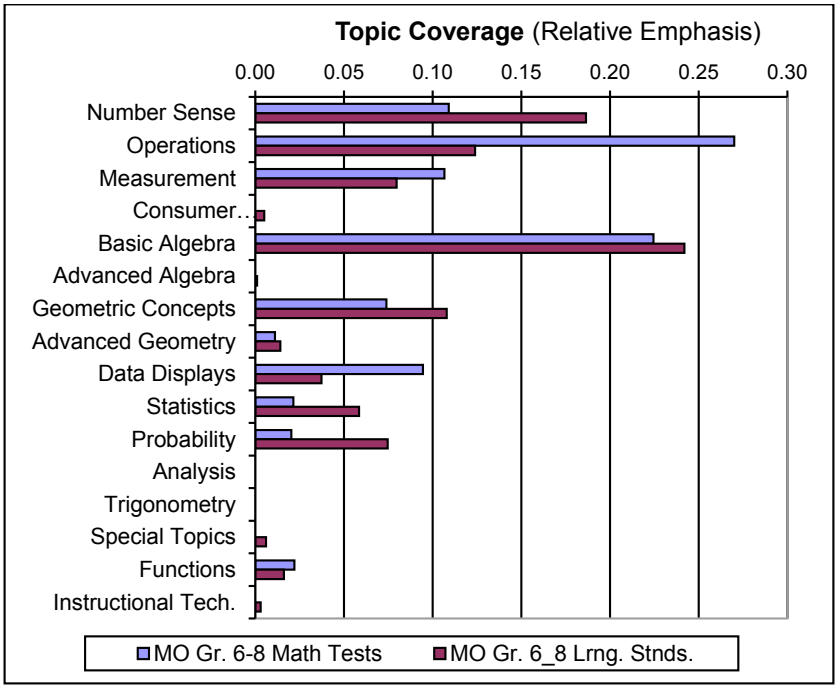


Figure 6



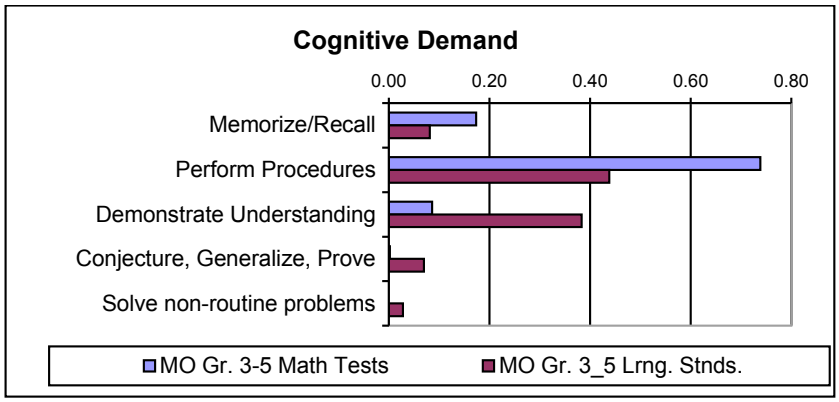
In light of the relatively high topic coverage results reported for mathematics, the differences in emphasis noted in the charts above are not especially concerning. Nonetheless, the chart does provide clues to the types of shifts in topic coverage that would lead to increased alignment. Examination of the

results above suggest some over-emphasis in the upper elementary assessments on Operations at the apparent expense of assessing topics related to Number Sense and Measurement, both of which are under-emphasized relative to the standards. For the middle grades, topics related to basic algebra represent a much larger portion of the mathematics curriculum. The assessments show a concurrent rise in the assessment of topics related basic algebra. However the emphasis on operations remains on the assessment despite a notable decline in emphasis for this topic area in the standards, leading to an apparent over-emphasis on Operations for grades 6-8. Again, keep in mind that the alignment results for these grades are quite acceptable, so the differences talked about here are nuances designed more to inform future development than to suggest any changes necessary for the current assessment instrumentation.

Cognitive Demand

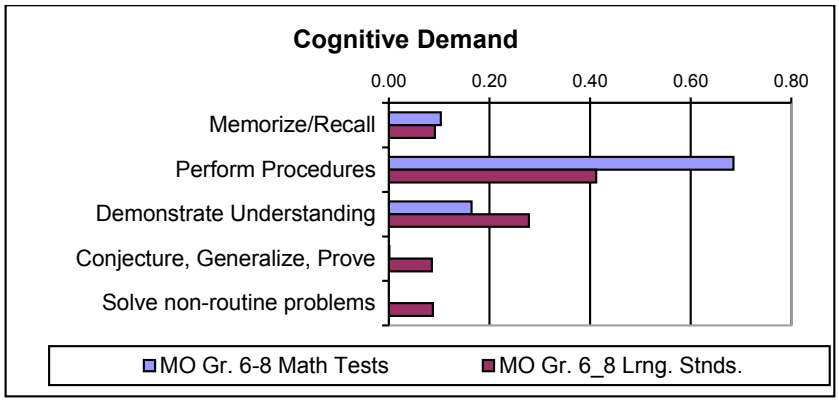
Figure 7 reports the cognitive demand results for grades 3-5 mathematics. Results indicate that there is a fair amount of over-emphasis on procedural skills, combined with an under-emphasis on conceptual understanding. With a summary TC measure of .55 there is no compelling need to alter the distribution of items across the cognitive demand categories, but if developers were interested in increasing this measure further, shifting from recall items to assessing conjecture and analysis would accomplish that goal.

Figure 7: Grades 3-5 Mathematics Cognitive Demand



It is interesting to compare the pattern of shifts in cognitive demand depicted in Figures 7 & 8. For the standards we see a shift toward more emphasis at the non-routine problem solving level in middle grades. By comparison, the shift in the assessments between upper elementary and middle level grades is characterized by greater emphasis on conceptual understanding in conjunction with a reduced emphasis on memorization and procedural skills. Across grades 3-8 it is notable that the assessments fail to assess the more complex cognitive skills related to conjecture, generalization and non-routine problems (read integrative/synthetic/creative cognitive skills). While the assessments do provide quite acceptable, if not impressive alignment results, the call for more rigor in the instruction of mathematics, in part represented by the standards reported emphases on the cognitive categories of conjecture and non-routine problems, puts the writing on the wall for assessment developers; states are going to expect to see assessment instruments shifting towards items that address these categories of cognitive demand.

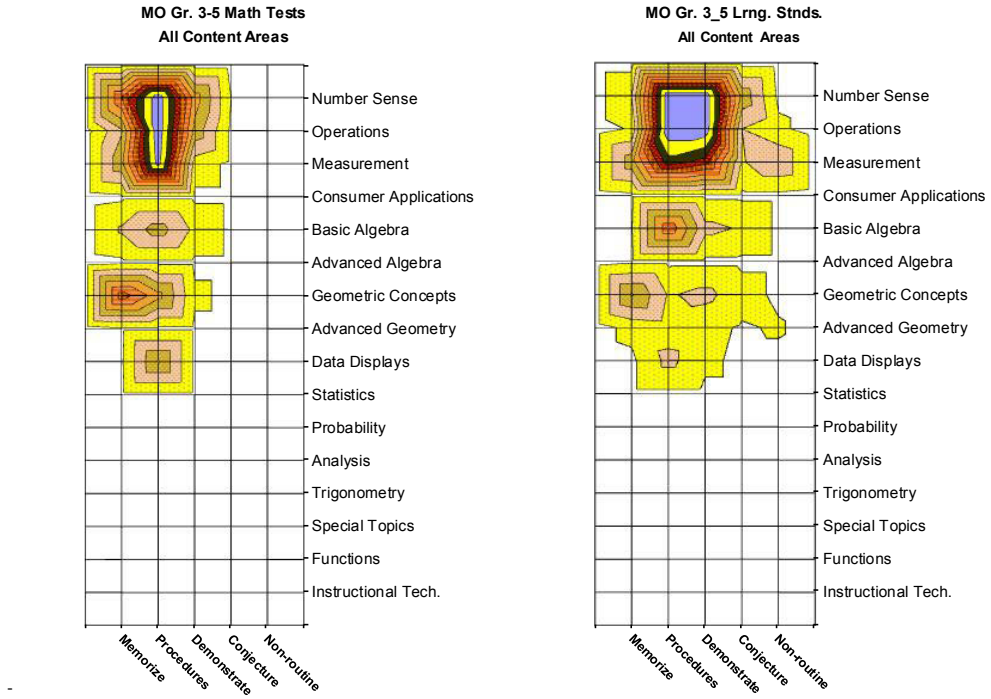
Figure 8: Grades 6-8 Mathematics Cognitive Demand



In defense of assessment developers it seems reasonable to point out that such assessment items, particularly within the context of large-scale standardized assessments that favor multiple choice response options, are challenging to create. Moreover, the assessment providers may feel that these categories of cognitive skills are being addressed with the current items. While that issue is a bit more defensible in speaking to the language arts results, where the analysts themselves did not always agree on the cognitive category being assessed (see the language arts section below), the mathematics team members were quite consistent in their judgement of the cognitive demands made upon students in the mathematics assessments.

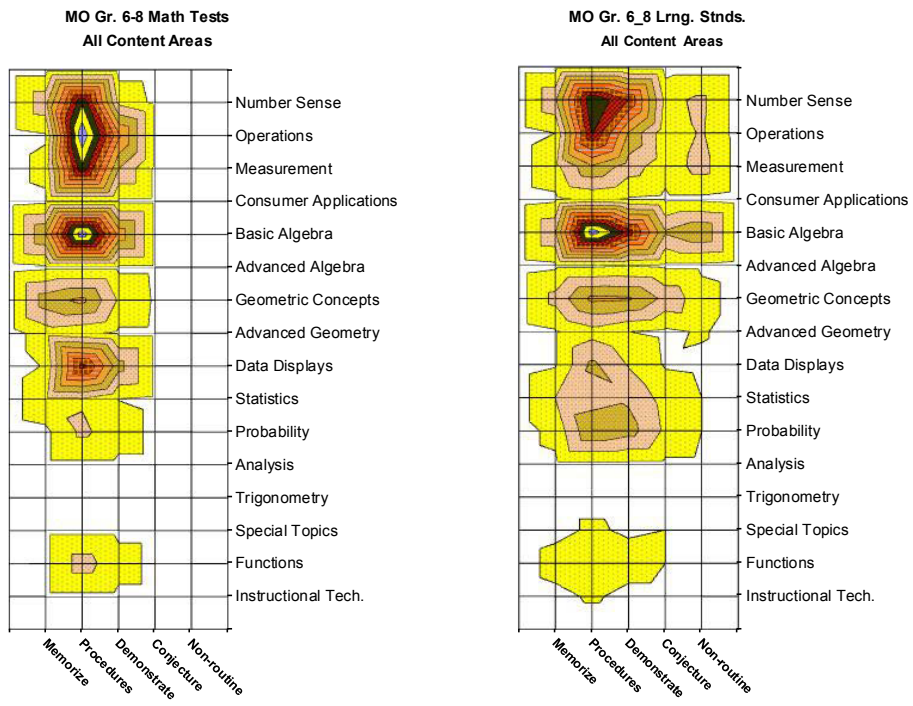
Overall Alignment

Figure 9



In Figure 9 we see the interaction of highly aligned topics with a more narrow set of cognitive skills assessed compared to the standards play out in the three-dimensional topographic depictions of the content descriptions of the assessments and standards side-by.

Figure 10



The contrast is even more apparent for the middle grades, as seen in Figure 10, where the standards show a marked movement toward the right side of the chart (the side where the more rigorous cognitive demand descriptors are found) across most topic areas. This is most apparent for the foundational topics of Number Sense, Operations, Measurement and Basic Algebra; the heart of middle grade mathematics content.

Language Arts

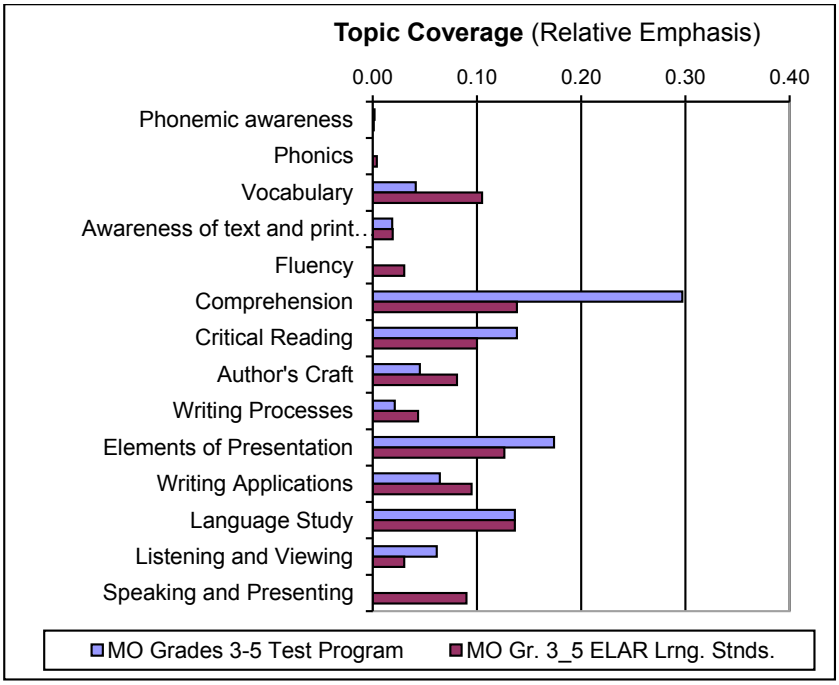
Table 5: Language Arts Alignment Summary

Language Arts	OAI	BR	TC	CD
Grade 3	***0.53***	***0.37***	***0.68***	***0.79***
Grade 4	***0.49***	***0.33***	***0.69***	***0.68***
Grade 5	***0.51***	***0.42***	***0.74***	***0.66***
Grade 6	***0.53***	***0.46***	***0.79***	***0.62***
Grade 7	***0.54***	***0.40***	***0.79***	***0.65***
Grade 8	***0.53***	***0.44***	***0.79***	***0.65***

Not unlike the pattern seen for mathematics, the results for Language Arts show relatively good alignment results for all measures and grades, excepting balance of representation. With language arts these BR measures are a noticeably lower than reported for mathematics. Results for topic coverage and cognitive demand are very similar to the results for mathematics, indicating again strong alignment across measures and grades.

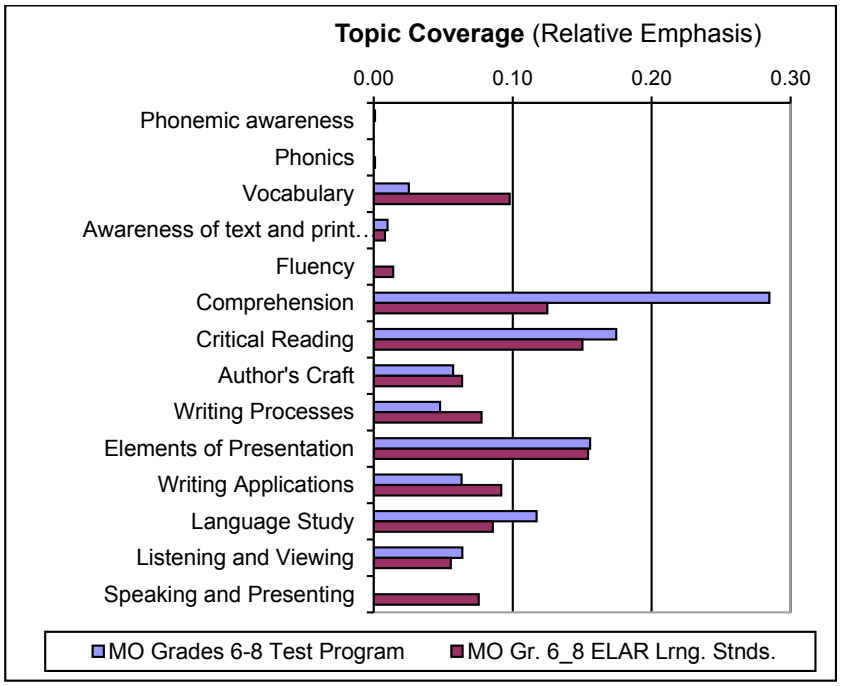
Topic Coverage

Figure 11: Grades 3-5 Language Arts Topic Coverage



The most obvious attribute noticeable in Figure 11 is the relative over-emphasis on comprehension type assessment items. This is a common characteristic of language arts assessments, with comprehension typically representing the most heavily emphasized topic area noted across language arts assessments. The other two noteworthy results reported in Figure 11 concern 'Speaking & Presenting' (not assessed) and 'Vocabulary' (under-emphasized in the assessment). There are understandable limitations to even computer-based assessments when it comes to assessing students' knowledge and skills in speaking and presenting, but the same would not be said for vocabulary content. There are other topic areas that also report a under-emphasis relative to the standards, no doubt as a consequence of the heavy emphasis on comprehension.

Figure 12: Grades 6-8 Language Arts Topic Coverage

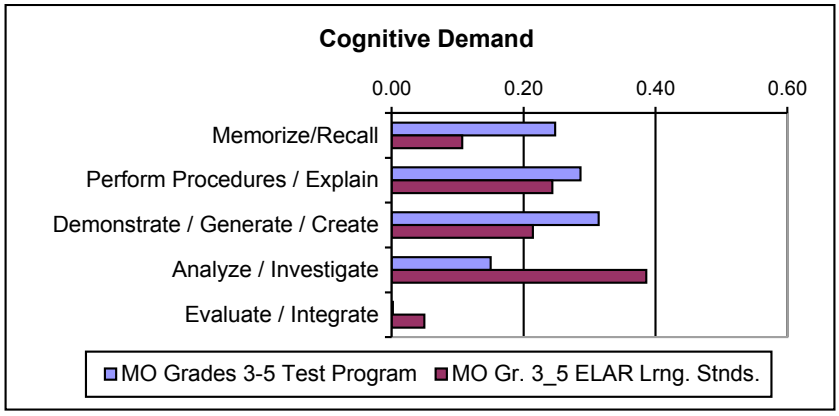


The pattern is almost identical for middle grades (see Figure 12), if anything the gaps between comprehension and vocabulary increase slightly between the assessments and standards, however the difference is a small one at best.

Cognitive Demand

The cognitive demand, or skills assessed by the upper-elementary language arts standards show some emphasis on each of the five categories of cognitive demand. The assessments address all of these except Evaluate/Integrate, which is also the least emphasized cognitive demand category for the assessment, accounting for about 5% of standards content.

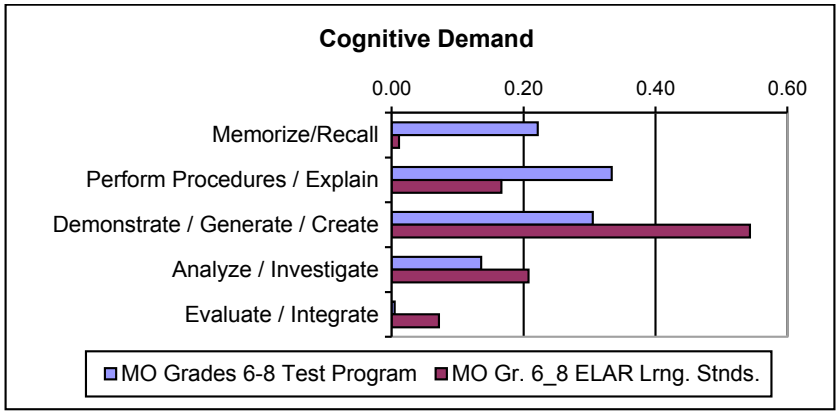
Figure 13



Among the other four other cognitive demand categories the one with the greatest gap between the assessment and standards is Analyze/Investigate, with 15% of the assessment emphasizing this skill compared to 39% for the standards. The next most noticeable difference between assessment and

standards is found with the Memorize/Recall category, which is emphasized in 11% of the standards, while representing 25% of the assessment.

Figure 14

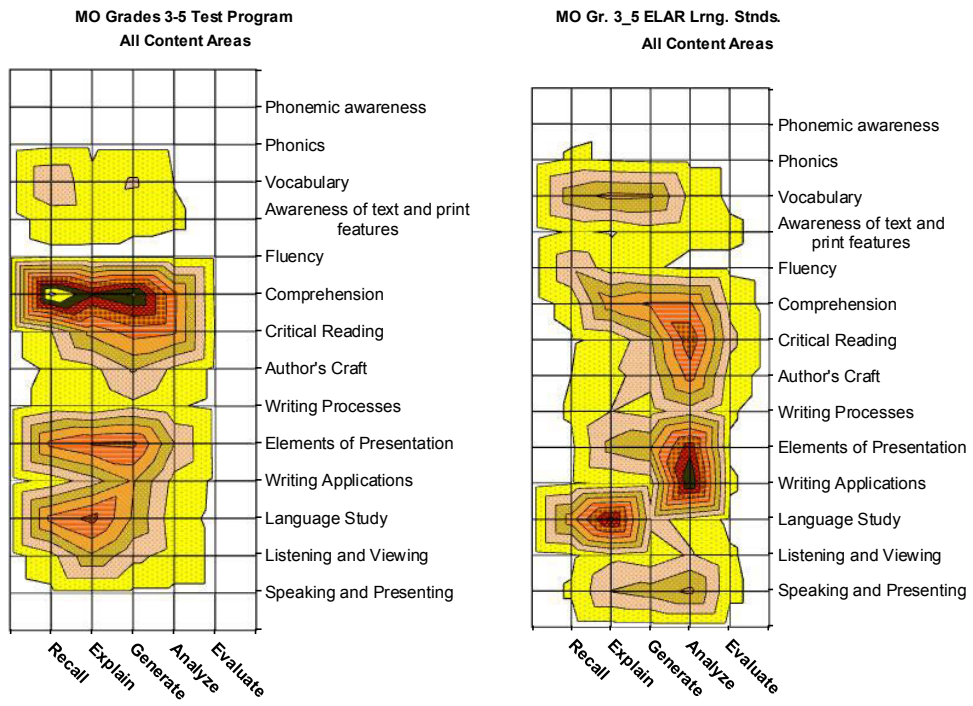


Looking at the shifts from upper elementary to middle grades, the general profile of cognitive demand emphasis stays relatively stable for the assessment, with very similar levels of emphasis for both grade spans. By contrast, the standards shift from a primary emphasis on analysis to a significant increase in the relative proportion of assessment items focused on 'generating' or 'creating' evidence to 'demonstrate' conceptual understanding. The middle grade standards also de-emphasize memorization noticeably, which in conjunction with the assessments maintaining memorization for a bit more that 20% of the test, makes memorize/recall the cognitive demand category with the greatest difference between the middle grades standards and assessments.

Overall Alignment

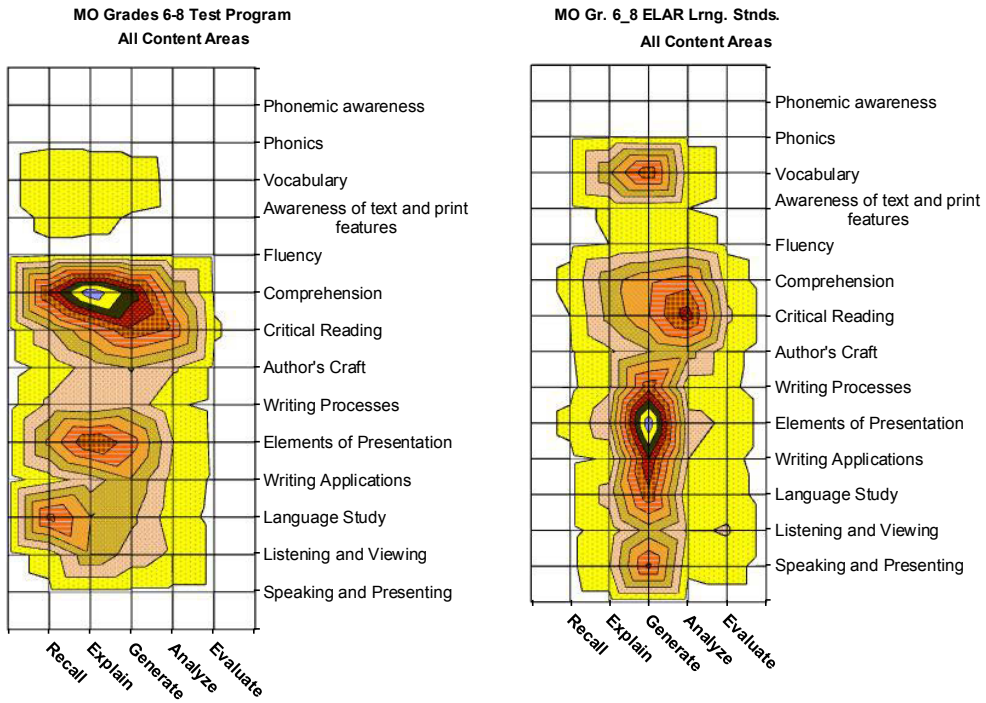
The content maps displayed in Figure 15 provide graphic examples of the complexity involved in comparing the content of the state's assessment program to its learning standards. While both the standards and assessments cover generally the same broad band of topics and cognitive demand categories, it is clear that the patterns within the two descriptions are quite different from one another. In short, alignment for language arts is a complex mix of issues related to balance of representation, topic coverage *and* cognitive demand.

Figure 15



The content descriptions for the grade 3-5 and grade 6-8 assessments look quite similar, while the shift in the standards is toward a more distinct focus on the cognitive demand category of generate, create, demonstrate, though comprehension and critical reading maintain a relatively strong emphasis analysis.

Figure 16



Discussion

What is "Good" Alignment?

Qualitative analyses of alignment set out to make explicit judgments about the quality of alignment; typically between assessments and standards though there is a growing interest in the alignment of textbooks and other curriculum materials to standards. In these analyses judgments about how 'well' assessments or textbooks are aligned to standards is the explicit goal of the analysis, based on the considered judgment of the analysts. Such studies invariably require content experts to make relatively high level inferences about the degree and quality of alignment of these documents relative to the standards they are designed to target. The process itself, by its very nature places the focus of analysis on the qualities that make up 'good' alignment, and the analysts are repeatedly making professional judgments based on criteria they chose or were given, and their expert judgment of the document's adequacy in meeting those criteria. The criteria thus provide an inherent framework for making judgments about alignment quality.

Thus in qualitative alignment studies the criteria for good alignment is explicitly stated. One may agree or disagree with those criteria, but for the purposes of analysis, they are the foundation stones for the analysis. Determining whether one or another assessment or textbook meets these criteria becomes the focus and challenge of the analysis, with consensus among analysts typical serving the role of quality assurance. In these analyses the criteria for judging alignment is explicit and defensible, while the judgments of the analysts require high level inferences and often require negotiation among analysts to reach consensus.

By contrast, determining the degree of alignment between any two SEC content descriptions is result of a relatively simple calculation. Determining the degree of alignment using this approach is straightforward and easy to justify based on the statistical procedures employed. However, justifying just what specific degree of alignment is 'acceptable' and should be set as the criteria for good alignment is much more difficult.

Since the results are not based on qualitative analyses, there are limitations to the extent to which the SEC approach can assert that one or another alignment measure constitutes 'good' alignment. Instead, the methodology offers a conceptual rationale for setting a threshold measure for indicating an acceptable level of alignment. Nonetheless results must be interpreted within the context of assessment design, and psychometric requirements in basing policy decisions about assessment development and deployment on SEC alignment results. That said, the SEC data results provide a rich set of analytic tools to support diagnostic analyses at a much greater level of detail than available through any other methodology.

While further studies incorporating achievement results may provide the necessary empirical evidence to justify determination of an optimal range for 'good' or 'effective' alignment, the current state of the research does not provide that evidence. Lacking the necessary empirical evidence, and as an interim alternative, a conceptually based rationale is used for determining the minimal degree of alignment considered acceptable for an assessment program. The conventional measure used to indicate that threshold is 0.5. The rationale behind this value is that standards offer a set of goals or objectives for student learning, but they do not address how to achieve those goals, or detail an exhaustive set of curricular content necessary to achieve those goals. Moreover, students come to classrooms with differing levels of knowledge and skills with which to assimilate the goals of content standards.

Assessments likewise operate within a context of limitations and competing goals. Add to this a language for describing content at a very detailed level of specificity, and the challenge for attaining high alignment measures increases dramatically. As a simple rule of thumb, the SEC methodology sets 0.5 (the center point of the alignment index range), as a reasonably challenging expectation for aligned content. Without evidence to support whether the contextual limitations of a sound assessment program or the need to cover the breadth of standards content should outweigh the other, the midpoint of the range seems most justifiable. The empirical fact that this measure is somewhat 'above average' relative to the range of alignment indices that have been reported for both mathematics and language arts. At the same time this leaves room for contextual constraints, student/assessment needs, and incidental content while establishing a minimal threshold (0.5) for insuring that assessment content at least equally attends to the expectations of content standards as measured along each of the SEC alignment dimensions.

Discussion of Findings

While the alignment results reported here reveal quite satisfactory alignment results, certain nuances revealed by the analyses suggest a small number of alignment issues worth comment.

The balance of representation results, particularly with respect to the proportion of the standards content represented on any given grade level assessment is notably low, particularly for language arts and the middle grades of mathematics. While there are legitimate considerations that does put a limit on the breadth of content represented in the assessment, and the overall results do exceed the 0.5 threshold, the analyses also suggest areas where a shift in content focus would improve the balance of representation results. For language arts a shift away comprehension topics and toward more items assessing vocabulary knowledge and skills would be the most obvious strategy for improving alignment results, not just on the BR measure, but as a result of the ripple effect of balance of representation (it plays a role in topic coverage and cognitive demand results as well) would almost certainly lead to increases on all of the summary measures.

Another area that shows up for language arts as being notably under-emphasized are topics related to speaking and presenting. Given the nature of the assessments, it seems a reasonable judgement to simply not require the assessment to test content related to this topic area. If the speaking and presenting content were dropped from the standards description, the results for language arts would improve noticeably, and again across the board for the alignment measures. The point however is not whether not assess content related to speaking and presenting is justifiable, but that it is a worthwhile conversation for the department to have. After all, it may well be the case that a conversation around this issue has not actually occurred, and so to say that neglecting content related to speaking and presenting is OK would presumptuous.

On the other hand, if the state does accept the argument that content related to speaking and presenting is either not necessary or not feasible, then the BR measures for language arts would increase noticeably, making the discussion around vocabulary an even more nuanced, less compelling issue to be concerned with.

With mathematics there is some over-emphasis on Operations content noted for the assessments at both the upper elementary and middle grades, with a concomitant under-emphasis for content related to Number Sense. With regard to cognitive demand, the mathematics assessments are most notable for their absence of content designed to assess cognitive skills of conjecture/generalize or non-routing problem

solving (e.g. synthetic and creative cognitive skills). While such assessment items might be challenging to devise, the call for more 'rigorous' assessments places a premium on just these types of assessment items.

Operations on the topic dimension, and performing procedures (e.g. calculation) on the cognitive demand dimension are the bread and butter of mathematics assessments, much like comprehension and conceptual understanding are the bread and butter of language arts assessments. It is not surprising to find a lot of items focused in these areas for these subjects. But the nuanced result being presented in the case of mathematics is that a shift toward more content focused on number sense and the addition of assessment of the more complex cognitive skills, suggest an area of developmental focus that would certainly pay dividends in further increasing the alignment of the mathematics assessments while attending to the call for more rigorous assessments.

Concluding Summary

The study reported above provides results of alignment analyses conducted on Missouri's assessment program for mathematics and language arts for grades 3 through 8. A full set of both coarse grain and fine grain alignment results are available in the Appendices of this report. These detailed results can be used to conduct more detailed diagnostic analyses to identify specific areas for adjustments in moving forward with further assessment design and development work.

Results indicate that the Missouri testing program for mathematics and language arts demonstrate high levels of alignment to the Missouri Standards for Learning, with all assessments meeting or exceeding the 0.5 alignment benchmark[1].

Balance of Representation results fall below the 0.5 for all of the language arts assessments and half of the mathematics assessments. This is largely due to the breadth of standards content, particularly for language arts, but also for middle grade mathematics compared to the elementary math curriculum. It is also a factor, as discussed above, of the requirements of providing a sufficient number of items to provide psychometric justification for assessing student proficiency on the particular knowledge and skills to be assessed.

The mathematics assessments report particularly strong alignment on Topic Coverage, while the cognitive demand measures indicate that the more complex cognitive demand categories of conjecture and non-routine problem solving are largely absent from the math assessments. Despite quite acceptable alignment results, even on the cognitive demand measures, this lack of targeting assessments with these types of cognitive demand is an area that would deserve more attention in future iterations of the assessments.

The language arts assessments also report relatively high levels of alignment for topic coverage and cognitive demand, with breadth of representation reporting notably lower results as already discussed. As with mathematics, all overall alignment results meet the 0.5 benchmark.

[1] The 0.49 result for grade 4 language arts is within the measurement error for these alignment indices, and is thus considered to have met the 0.5 threshold requirement.

About the Author

Dr. Smithson received his Ph.D. from the University of Wisconsin-Madison School of Education in 1998. He led the design and development of the SEC online data collection and reporting system, and has managed the operation of that web-site since it initially went online in 2000. Dr. Smithson has led content analysis work on several hundred documents over the past ten years, including analyses of the Common Core State Standards for mathematics and science, the NAEP, TIMSS, and PISA item pools, along with content standards and assessments across more than two dozen states. Dr. Smithson has participated as a co-PI in several NSF funded projects, and his content analysis work has been used by numerous researchers investigating policy and programmatic impacts of standards and K-12 efforts to implement standards.

References

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. Educational Researcher, 31(7), 3-14.

Porter, A.C., Polikoff, M.S. & Smithson, J.L. (2009). Is there a de facto national intended curriculum? Evidence from state content standards. Educational Evaluation and Policy Analysis, 31(3), 238-268.

Porter, A.C., Polikoff, M.S., Zeidner, T., Smithson, J.L. (2008). The quality of Content Analyses of State Student Achievement Tests and Content Standards. *Educational Measurement: Issues and Practice,* v27 n4 p2-14, Winter, 2008.

Porter, A.C., Smithson, J.L., Blank, R., Zeidner, T. (2006). Alignment as a Teacher Variable. *Applied Measurement in Education.*

Porter, A.C. & Smithson, J.L. (2001). Defining, developing, and using curriculum indicators. CPRE Research Report Series RR-048. Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.

Webb, N.L. (1997). Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education. Research Monograph No.6. Washington, DC: Council of Chief State School Officers.