A Report to the

North Carolina Department of Public Instruction

On the Alignment Characteristics of State Assessment Instruments

Covering Grades 3-8, and High School in Mathematics, Reading and Science

September, 2015

Prepared by John L. Smithson, Ph.D., Researcher, University of Wisconsin-Madison

TABLE OF CONTENTS

North Carolina 2015 Alignment Study Report

Introduction

In September, 2014 the North Carolina Department of Public Instruction commissioned the Wisconsin Center for Education Research to conduct an in-depth study of the alignment of the state's newly developed assessments for mathematics, reading and science to new standards as part of a larger effort to make a systemic examination of the state's standards-based reform efforts. The current report focuses explicitly on the relationship between new *assessments* and their respective *content standards* or curricular goals. Phase 2 of the study will examine the relationship between *instructional practice* and relevant content standards based upon a randomly selected representative sample of teachers in the state, while Phase 3 will examine the impact of students' opportunity to learn standards-based content on student achievement. The completed study will provide the state with a unique data set for modeling the performance of the standards-based system as depicted by the various data collection and analysis strategies employed for the study.

Specifically, the current report focuses on describing the alignment characteristics of the assessment program in North Carolina based upon analyses of 42 assessment forms covering state mathematics and reading assessments for grades 3, 4, 5, 6, 7, 8, and HS, as well as state science assessment forms for grades 5, 8, and HS Biology. The science analyses are based upon content analyses of the states Essential Science Standards for grades 5 & 8, and high school Biology I end-of-course standards. Mathematics and reading alignment are based on the state's core standards for mathematics and reading.

For purposes of clarity, this report will focus upon the grade level *assessment program* for each subject analyzed as represented by multiple assessment forms. Results for individual assessments are not included in this report, but can be accessed through the data sources found in the appendices. The results presented below are based on averages of the results for each of three assessment forms analyzed at each grade analyzed.

Rationale

Standards-based educational reform has been *the* fundamental education model employed by states, and to a growing extent federal policy-makers for twenty-plus years. Emerging out of the systemic research paradigm popular in the late eighties and early nineties, the standards-based model is essentially a systemic model influencing educational change. The standards-based system is based upon three fundamental propositions: 1) standards will serve as an explicit goal or target toward which curriculum planning, design and implementation will move, 2) accountability for students, teachers and schools can be determined based upon student performance on 3) standardized tests that are aligned to the state standards. Weaving through these propositions is the notion of alignment, and the importance of it to the standards-based paradigm.

While examination of instructional alignment can help answer the first, and alignment studies of assessments can help assure the third, neither alone or in separation can address whether the assumptions of the second are justified. To do this, one must look at the role of both in explaining student achievement. Moreover, in order to address the overall effectiveness of the standards-based *system* as implemented in one or another location, one must be able to bring together compatible alignment indicators that span the domains of instruction, assessment and student performance. The Surveys of

Enacted Curriculum (SEC) is unique among alignment methodologies in that in allows one to examine the inter-relationships of instruction, assessments *and* student performance using an objective, systematic, low-inference, quantifiable approach to examining alignment issues.

The surveys of enacted curriculum (SEC), though best known for its tools for describing instructional practice, provide a methodology and set of data collection and analysis procedures that permit examination of all three propositions in order consider the relationships between each in order to look at the standards-based system as a whole to determine how well the system is functioning.

This document reports on Phase I of a three phase study commissioned by North Carolina's Department of Public Instruction to examine the effectiveness of the state's efforts to implement a newly structured standards-based system in the state. Phase I focuses on alignment of new assessments developed for mathematics and reading in grades 3-8, as well as one high school end-of-course exam administered by the state. Phase II will focus on instructional alignment, and Phase III will examine student performance in light of students' opportunities to learn standards-based content given the assessments used to generate achievement results. Once all three phases have been completed, the state will have been provided an in-depth look at the state's standards-based system as currently implemented, and a wealth of information in considering its continuing efforts at provide quality educational opportunities to the state's K-12 population.

Structure of the Report

The current report is laid out in a manner most conducive to providing the reader with the essential information necessary to convey an overall sense of the alignment results for this study without delving into the underlying structure and details of interpretation until later. By the end of the first section the reader should have a good overall picture of the alignment results and the general structure by which those results are reported.

For the reader interested in better understanding how the summary measures presented in Section I are derived, or the justification for the 0.50 threshold measure used, amidst a broader discussion of determining 'good' alignment and selecting the most appropriate alignment 'target', Section II will be of interest.

What would be Section III is instead presented as Appendix A. The reason for this is to keep the reporting of the alignment study results separate from the descriptive data upon which those results are based. It is hoped that this will make the explication of both simpler. Appendix A is intended to prepare the reader for exploring the underlying descriptive data behind the alignment results. Here the reader will learn more about content descriptions, how they are collected, processed and reported. These descriptive results further provide the reader with a detailed 'picture' of the content embedded in each of the documents analyzed for the study, and used to conduct the type of diagnostic analyses described in Appendix B. Appendix C provides a detailed description of the content alignment process and the materials used in the process. Appendices D-F provide subject-specific content viewers for generating the charts and maps introduced in Appendix A, and used to conduct fine grain analyses as described in Appendix B.

Section I: What Is Alignment Analysis?

Alignment, whether talking about the wheels on your car, or as a characteristic of assessment and instruction, is inherently a question about relationships. How does 'A' relate to 'B'? However, that also means alignment is inherently an abstraction… it's not a thing you can easily measure. Moreover, as with most relationships, the answers aren't simple 'yes' or 'no's, but rather a matter of degree. Relationships also tend to be multi-dimensional; they have more than a single aspect, dimension, or quality that is important to fully understand the nature of the alignment relationship. All of these factors make alignment analyses a challenging activity, though of critical importance to the operational efficiency of both schools and cars.

To understand how alignment is calculated using the SEC approach it is important to understand that alignment measures in SEC are *derived* from content descriptions. That is alignment analyses report on the relationship between two multi-dimensional content descriptions. Each dimension of the two descriptions can then be compared, using procedures described below, to derive a set of alignment indicator measures that summarize the quantitative relationship between any two content descriptions on any of the dimensions used for describing academic content. In addition to examination of each dimension independently, the method allows for examination of alignment characteristics at the intersection of all three dimensions employed, producing a summative 'overall' alignment indicator that has demonstrated a predictive capacity in explaining the variation of students' opportunities to learn assessed content, otherwise referred to as predictive validity.

Content descriptions are described in more detail in Section III. Keeping the focus for the moment on alignment, the reader is asked to accept for the moment that we have collected two descriptions of academic content in order to calculate and report alignment results; one description of the content covered across a series of assessment forms for a particular grade level, and the other a description of the relevant academic content standards for the assessed grade and subject.

These *content descriptions* are systematically compared to determine the alignment characteristics existing between the two descriptions, using a simple iterative algorithm that generates an alignment measure or index based on the relevant dimension(s) of the content being considered.

As mentioned, there are three dimensions to the content descriptions collected, and hence three dimensions upon which to look at the degree of alignment the analyses indicate. These indicator measures can be distilled further to a single overall alignment index (OAI) that summarizes the alignment characteristics of any two content descriptions at the intersection of the three dimensions of content embedded in the SEC approach. What these dimensions are, and the alignment indicators they yield are described next.

The Dimensions of Alignment

SEC content descriptions are collected at the intersection of three dimensions: (1) topic coverage (2) performance expectation and (3) relative emphasis. These parallel the three alignment indices that measure the relationship between the two descriptions on one or another of these three dimensions: (1) Topical Coverage (TC); (2) performance expectations (PE); and (3) balance of representation (BR). When considered in combination with one another; that is when all three dimensions are included in the alignment algorithm, a fourth, summary measure of 'overall alignment' can be calculated. The procedure for calculating alignment is discussed further on in the report, as is a discussion of what constitutes 'good'

alignment using the SEC approach. In short, each alignment indicator is expressed on a scale with a range of 0 to 1.0, with 1.0 representing identical content descriptions (perfect alignment) and 0 indicating no content in common between the two descriptions, or perfect misalignment. For reasons discussed further below, a threshold measure is set at 0.5 for each of the four summary indicator measures. Above this threshold alignment is considered to be at an acceptable level. Below that and the alignment measure is considered weak or questionable; indicating that a more detailed examination related to that indicator measure is warranted. Much like the results for medical tests, results that fall outside the range of 'normal limits' indicate that further investigation is warranted, but does not necessarily mean that the patient is in ill-health, or that a given assessment is not appropriately aligned. It means more information is needed. That additional information is available in the form of fine grain alignment analyses which serve a diagnostic, as opposed to summative purpose. An example of the fine grain analysis process is provided in Appendix B, and the necessary data and reporting tools necessary to conduct those analyses for any documents analyzed are also supplied in the Appendices.

Balance of Representation

Of the three content dimensions on which alignment measures are based, two are directly measured and one is derived. That is two of the content dimensions are based upon observer/analyst reports of the occurrence of one or another content description. The derived measure concerns 'how much' and is based on the number of reported occurrences for a specific description of content relative to the total number of reports making up the full content description. This yields a proportional measure, summing to 1.00. The SEC refers to this 'how much' dimension as 'balance of representation' (BR).

As a summary indicator, (BR) is calculated as the product of two values; the portion of the assessment that targets standards-based content, multiplied times the portion of standards-based content represented in the assessment. For example, if 90% of an assessment (i.e. 10% of the assessment covers content not *explicitly* referenced in the standards) covered 40% of the standards for a particular grade level (i.e. 60% of the content reflected in the standards was not reflected in the assessment), the BR measure would be 0.36. As with all the summary indicator measures, reported here, the 'threshold' for an acceptable degree of alignment is 0.50 or higher. Our example would thus reflect a weak measure of alignment, given this threshold measure. The rationale for this 0.5 measure is discussed in Section II.

The influence of BR runs through all of the alignment indices, since the relative emphasis of content is the value used in making comparisons between content descriptions. In a very real sense the dimensions of topic and performance expectation provide the structure for looking at alignment, while the balance of representation provides the *values* that get placed in that structure. This will become more apparent in the discussion on the calculation of alignment presented in Section II.

For assessments, relative emphasis is expressed in terms of the proportion of score points attributed to one or another topic and/or performance expectation. When talking about standards, relative emphasis refers to the number of times a particular topic and/or performance expectation is noted across all the strands of a standard presented for a given grade and subject.

Table 1

Balanceof Representation Index by Grade and Subject

(0.50 or greater = well-aligned)

Grade	3	4	5	6	7	8	HS
ELAR	0.59	0.71	0.70	0.66	0.64	0.67	0.70
MATH	0.57	0.81	0.78	0.87	0.84	0.81	0.69
SCIENCE			0.77			0.54	0.83

Without exception, all of the summary measures on BR for the assessed grades exceed the 0.5 threshold. This one measure alone however provides insufficient information for making a judgment regarding alignment. It tells only part of the alignment story. The other indicators provide other perspectives for viewing alignment that help to fill out the full picture of the alignment relationship existing between assessments and standards.

Topic Coverage

The first dimension considered in most, if not all alignment analyses, regardless of the methodology employed, concerns what Norman Webb (1997) calls *categorical concurrence*. For convenience and to better fit the SEC terminology, this indicator is simply referred to as *topic coverage* (TC) and measures a seemingly simple question; does the topic or sub-topic identified in one description match a topic or sub-topic occurring in the other description?

Actually, there are a series of questions implied here, each relevant to a comparison of the topics covered in an assessment with those indicated in the relevant target standard. 1) Which topics in the assessment *are also* in the standards? 2) Which topics in the assessment *are not* in the standards? 3) Which topics in the standards *are* in the assessments? and 4) Which topics in the standards *are not* in the assessment? Each of these represents a distinctly different question that can be asked when comparing topic coverage. The algorithm used to calculate topical concurrence is sensitive to each of these questions, with the resulting index representing, in effect, a composite response to all four questions. Table 2 provides the summary alignment results for TC for each of the assessed grades in mathematics and reading analyzed for this study.

Table 2

Topic Coverage Index by Grade and Subject

(0.50 or greater = well-aligned)

Grade	3	4	5	6	7	8	HS
ELAR	0.65	0.64	0.64	0.72	0.86	0.81	0.88
MATH	0.68	0.67	0.64	0.73	0.72	0.74	0.61
SCIENCE			0.73			0.63	0.70

Once again the summary measures for this dimension indicate above-threshold alignment results, suggesting that the assessments are well aligned to the standards with respect to topic coverage.

Performance Expectations

The SEC taxonomies enable descriptions of academic content based on two dimensions ubiquitous to the field of learning: knowledge and skills. When referencing standards this is frequently summarized with the statement "what students should know and be able to do". The '*what students should know*' part refers to topics, while the '*be able to do*' reference expectations for student performance, or performance expectations for short. The SEC taxonomies enable the collection of content descriptions on both of these dimensions, and together form the alignment 'target' for both assessments and curriculum.

Just as we can examine alignment with respect to topic coverage only, we can similarly examine the descriptions of performance expectations embedded in the content descriptions of assessments and standards. This alignment indicator is referred to as performance expectations (PE), and is based on the five categories of expectations for student performance employed by the SEC. While the labels vary slightly from subject to subject, the general pattern of expectations follows this general division: 1) Memorization/Recall, 2) Procedural Knowledge, 3) Conceptual Understanding, 4) Analysis, Conjecture and Proof, and 5) Synthesis, Integration and Novel Thinking.

Table 3 reports the summary alignment measures across assessed grade levels for mathematics and reading. Once again the summary measure for this dimension is expressed in an index with a range of 0 to 1, with 0.50 indicating acceptable alignment.

Table 3
Performance Expectations Index by Grade and Subject
(0.50 or greater = well-aligned)

Grade	3	4	5	6	7	8	HS
ELAR	0.86	0.59	0.67	0.83	0.66	0.64	0.65
MATH	0.41	0.72	0.70	0.63	0.58	0.77	0.83
SCIENCE			0.75			0.74	0.72

As can be seen from Table 3, all but one subject/grade easily surpass this threshold. The results for grade 3 mathematics indicate weak alignment, but based upon assessment design decisions. may nonetheless represent an acceptable degree of alignment (see 'what is good alignment' below). The fine grain results for grade 3 mathematics are examined more closely in Appendix B, where diagnostic analyses are presented to indicate particular areas of weak alignment that explain the relatively low alignment results.

Alignment at the Intersection

While the SEC approach to alignment allows reporting and consideration of the data results along each of these three dimensions, the most powerful alignment measure results when all three dimensions are combined into an index measure that is sensitive to the dynamic interplay of all three dimensions by comparing content descriptions at the intersection of all three dimensions.

Table 4

Overall Alignment Index by Grade and Subject

(0.50 or greater = well-aligned)

Grade	3	4	5	6	7	8	HS
ELAR	0.58	0.47	0.52	0.62	0.64	0.62	0.57
MATH	0.40	0.59	0.54	0.55	0.46	0.64	0.57
SCIENCE			0.55			0.56	0.52

The resulting alignment index, just like the summary indices for each dimension reported separately, has a range of 0.00 to 1.00 with 0.50 or higher indicating adequate overall alignment. Overall alignment results are reported in Table 4. Once again we see grade 3 mathematics indicating weak alignment, as well as slightly below-threshold results for grade 7 mathematics and grade 4 reading.

Table 5

Overall Alignment Index by Grade and Subject

(0.50 or greater = well-aligned)

Grade	OAI	BR	TC	PE
Grade 4 Reading	0.47	0.71	0.64	0.59
Grade 3 Math	0.40	0.94	0.67	0.41
Grade 7 Math	0.46	0.76	0.72	0.58

Table 5 reports all four indicators for the three assessments with below-threshold measures. Based on those results, it appears that in each case alignment issues mostly concern performance expectations. Grade 4 reading and grade 7 math both appear more borderline acceptable insofar as each of the sub-measures are above 0.5, but the PE measure for both is noticeable lower than TC and BR, again suggesting that any alignment issues related to these assessments will likely center around performance expectations.

To determine just what shifts in performance expectations these analyses would recommend requires a more fine grain analysis as described in Appendix B using the grade 3 results as an example. To summarize those results here, fine grain analyses indicate that both assessments would benefit from a shift toward more evidence of conceptual understanding of mathematical ideas and less focus on computational proficiency. However, those results must be understood and interpreted within the context of other requirements placed on assessments as addressed in the discussion of findings presented below.

Summary of Findings

To summarize this first section, assessment forms for all assessed grades in mathematics, science and reading making up the state's assessment program were examined using the SEC approach to alignment analyses. Results were presented across the three dimensions of content inherent to the SEC descriptive languages or taxonomies. By and large the assessments reveal strong alignment characteristics relative to state standards for the appropriate subject and grade level. Two grades in mathematics and one grade in Reading do show some alignment weaknesses discussed further in the discussion of findings presented below and addressed in more detail in Appendix B.

The next section examines more closely how alignment is calculated, the rationale behind the 0.5 threshold, and a discussion of circumstances under which it is appropriate to limit or narrow the scope of the alignment target in order to better reflect the intent of the state's assessment program.

Section II: Conducting & Interpreting Alignment Analysis

Using SEC, alignment is a computational algorithm. While human decisions and judgments are made on the numbers put into the algorithm, and humans again must interpret the results that emerge, alignment itself is merely a systematic comparison of two data arrays. The difference between alignment indicators is mostly a matter of the array (dimension) chosen and the number of iterative comparisons required to complete the calculation.

Note that alignment analyses can be conducted and reported at two distinct levels of granularity or detail. For our purposes we will refer to them as coarse grain and fine grain results. Coarse grain analyses are most appropriate for summarizing the alignment results for a given assessment to a given standard. They provide a concise set of measures that nonetheless address the full span of alignment dimensions and characteristics. Fine grain alignment provides a more detailed examination of the alignment characteristics of a given assessment that can be used to support diagnostic analyses intended to provide strategies for improving the alignment characteristics of one or another assessment instrument. The results reported above and discussed below are based upon coarse grain analyses. The subject-specific data to support fine grain analyses are provided in Appendices D-F. Appendix A introduces the measures and data displays used to present the descriptive results while Appendix B provides an example of a fine grain diagnostic analysis using both alignment and descriptive data to inform improvement strategies.

Regardless of grain size, the process and underlying algorithm can best be understood visually through the use of content matrices.

Content Matrices & Grain Size

To better understand the alignment algorithm and the indices it yields it will be helpful to explain the source of the numbers used to compute the results. To understand those sources is to understand content matrices and marginal measures.

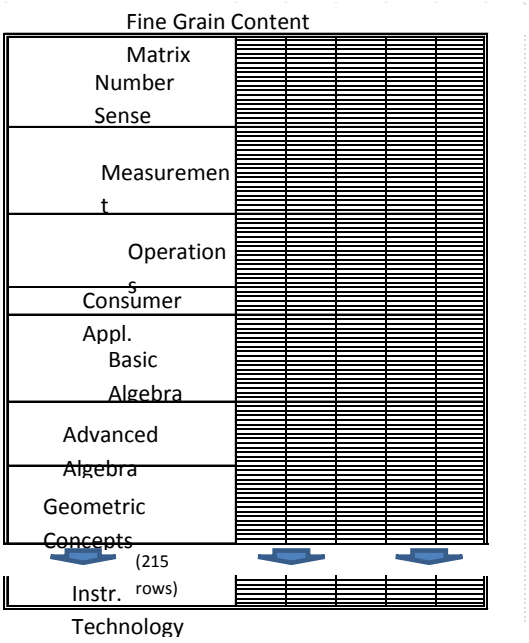
A content matrix is a two dimensional table used to display the quantitative results of a content description. This is an important construct for understanding alignment within the SEC framework, as it forms the basis for all calculations to follow. A content description is essentially a two-dimensional table of proportional values. That is, if you sum up all of the cells in the table, the result will be 1.00. Content matrices come in two sizes: coarse grain and fine grain.

Both coarse grain and fine grain content matrices have five columns of data (plus the label column), one column for each category of performance expectation. In a coarse grain matrix/table each row references a specific topic area. Mathematics has 16 topic areas, science has 28, and English, language arts and reading (ELAR) has 14, of which 6 are relevant to calculating reading alignment. Thus coarse grain content matrices have 16 (mathematics), 28 (science), 14 (ELAR) or 6 (Reading) rows by 5 columns of performance expectations. Each cell in the table references the intersection of a particular column and row to report a topic-by-performance expectation value for the specific content connoted by that cell in the table. Fine grain matrices have one row for each sub-topic listed in the content taxonomy, resulting in 215 rows/subtopics for mathematics, 328 for science and 161 ELAR, 72 of which apply to Reading.

Figure 1

Coarse Grain Content Matrix

Number Sense					
Operations					
Measurement					
Consumer Applications					
Basic Algebra					
Advanced Algegra					
Geometric Concepts					
Advanced Geometry					
Data Displays					
Statistics					
Probability					
Analysis					
Trignonmetry					
Special Topics					
Functions					
Instr. Technology					



These content matrices incorporate all three dimensions used to describe content and alignment characteristics. Again the rows represent topics (coarse grain) and sub-topics (fine grain), while the columns represent the categories of performance expectations. Balance of representation or relative emphasis is indicated by the value for a given cell in the table. Summing across the columns for any one row (whether in a fine grain or coarse grain matrix) yields the *marginal* topic/sub-topic measure, indicating the portion of the content description pertaining to that topic without respect to the categories of performance expectation. Similarly, summing the values for any column in either a coarse grain or fine grain table will yield the marginal measures for performance expectation without consideration to topic/sub-topic. These then yield two marginal measures, one for topic and one for performance expectations. (see Figure 2).

Figure 2

Topic Marginals Table (coarse grn.)

Number Sense	
Operations	
Measurement	
Consumer Applications	
Basic Algebra	
Advanced Algegra	
Geometric Concepts	
Advanced Geometry	
Data Displays	
Statistics	
Probability	
Analysis	
Trignonmetry	
Special Topics	
Functions	
Instr. Technology	

Performance Expectation Marginals

Memorize/Recall	
Perform Procedures	
Demonstrate Understanding	
Analyze, Conjecture,Prove	
Solve novel Problems	

The indicator measures of TC and PE are calculated using results from marginal tables like those depicted in Figure 2, while overall alignment indicator measures are drawn from the two dimensional coarse grain or fine grain tables, depending on the level of analysis being conducted.

Selecting the Appropriate Alignment Target

While generally speaking the appropriate target for an assessment is the set of relevant grade-level academic content standards adopted by the state, there are some instances where a more targeted selection of subject matter content is warranted.

Consider the case of reading assessments in the state. Reading encompasses only a small portion of the complete set of content standards for language arts. Yet the state has explicitly chosen to not assess writing and other language skills as part of their reading assessment program. Even at the secondary level, where the state does include open-ended response items in its end of course assessments, these items are not scored for writing standards. Thus holding a reading assessment accountable to the full scope of language arts standards would inevitably result in low alignment results relative to the larger scope of language arts content.

In order to make the alignment measures for the reading assessment more appropriate, given the intended focus of those assessments, the results reported for reading below are based on alignment across the following language arts content areas represented in the SEC taxonomy: Vocabulary, Text & Print Features, Fluency, Comprehension, Critical Reading and Author's Craft.

There are other reasons that call for a narrowing of the alignment target from the full set of standard strands. For example, assessment design requirements may create limitations on the assessment that preclude the assessment of specific topics or performance expectations. Some performance expectations, such as Analyze and Apply can be difficult or expensive to assess for some subjects or grade levels.

The mathematics results reported below *do* use the full breadth of mathematics content represented by the state's mathematics content standards, though whether this is appropriate for all assessments at all grades in mathematics would be a judgment call better left to the department to determine. For most cases in mathematics the point is largely moot, as the results show strong levels of alignment for the majority of grades and indicator measures. In those few cases of borderline or weak alignment, the fine grain analyses will reveal the specific areas of assessment that are contributing to the lower alignment measures. At that point the department will be better positioned to determine if the alignment target for those grades is appropriate, and/or whether adjustments are called for.

Calculating Alignment

The simplest alignment calculation (i.e. the one involving the fewest number of variables), pertains to the PE indicator measure. Using the grade 6 reading assessment and standards as an example, their respective topic marginal results are reported in Table 6.

The third column in table 6 reports the lower of the two values indicated for a given row. It is this algorithm that encapsulates the fundamental alignment indicator: a simple comparison of two values, each describing a common point of content across two different content descriptions. The lower of the two values is in effect put into a counter that when summed across all the relevant comparisons yields the selected alignment result. That result is an indicator number that represents the 'in common' portion of the two descriptions; or using set theory terminology, the 'intersect' of the two descriptions on the

dimension or dimensions being considered. Summing the values of this third column yields the PE measure, in this case 0.83 for the grade 6 reading assessment.

Table 6

Performance Expectation Marginals

	Test	Standards	Aligned
Memorize/Recall	0.12	0.02	0.02
Perform Procedures	0.29	0.22	0.22
Demonstrate Understandin	0.26	0.32	0.26
Analyze, Conjecture,Prove	0.33	0.36	0.33
Solve novel Problems	0.00	0.07	0.00
		PE=	**0.83**

Table 7 shows a similar table for topic coverage (TC), again for grade 6 reading. The process is identical for mathematics, except in that case the topic matrix would contain 16 rows.

Table 7

Topic Marginals Table (coarse grn.)

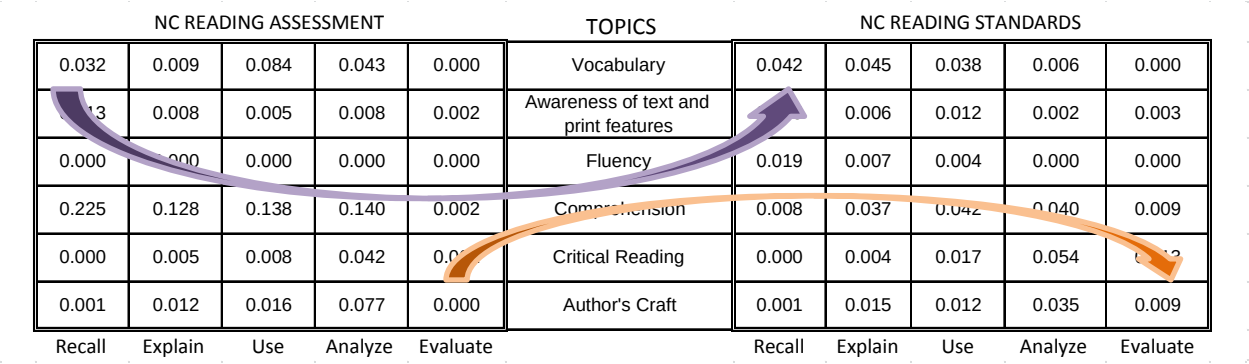
Topics	Test	Standards	Aligned
Vocabulary	0.14	0.21	0.140
Text features	0.00	0.02	0.000
Fluency	0.00	0.03	0.000
Comprehension	0.47	0.28	0.280
Critical Reading	0.16	0.33	0.160
Author's Craft	0.22	0.14	0.140
			0.72

When calculating overall alignment the calculation is complicated only by the fact that now one must compare two 2-dimensional tables or matrices. This means that the cell by cell comparisons take place across a series of rows *and* columns. Again, each cell reports the proportion of emphasis for that specific topic and performance expectation indicated for that cell, such that the sum of values across all cells of one content matrix will sum to 1.00. Now however, rather than simply comparing 5 or 6 values to complete the calculation, it is necessary to make 30 (e.g. 5 expectations by 6 topics) comparisons, to calculate an overall alignment index for reading, or 80 comparisons (5 expectations by 16 topics) to calculate OAI for mathematics.

Figure 3 displays the coarse grain content matrices for grade 6 reading standards and assessments. The two arrows overlaying the table indicate the first and last comparisons made across each cell of the tables.

In order for the alignment index for these two documents to be 1.00, both tables would need to have identical numbers in each of the corresponding cells across the two tables. That is the content matrices would have to be identical. Since the sum across cells always equals 1.00, any change of value in one cell will have a corresponding change in some other cell(s) so that the total continues to equal 1.00, and those changes will in turn lead to a less than perfect (1.00) alignment index. Using the SEC alignment algorithm to calculate the alignment of the two matrices presented in figure 3 the result comes to 0.62.

Figure 3



NC READING ASSESSMENT					TOPICS	NC READING STANDARDS				
0.032	0.009	0.084	0.043	0.000	Vocabulary	0.042	0.045	0.038	0.006	0.000
[illegible]13	0.008	0.005	0.008	0.002	Awareness of text and print features	[illegible]	0.006	0.012	0.002	0.003
0.000	[illegible]00	0.000	0.000	0.000	Fluency	0.019	0.007	0.004	0.000	0.000
0.225	0.128	0.138	0.140	0.002	Comprehension	0.008	0.037	0.042	0.040	0.009
0.000	0.005	0.008	0.042	0.0[illegible]	Critical Reading	0.000	0.004	0.017	0.054	0.[illegible]2
0.001	0.012	0.016	0.077	0.000	Author's Craft	0.001	0.015	0.012	0.035	0.009
Recall	Explain	Use	Analyze	Evaluate		Recall	Explain	Use	Analyze	Evaluate

Which of course begs the question: "does this indicate good alignment?" The section to follow discusses the challenges in determining what constitutes 'good' alignment and presents a rationale for the threshold measures employed for the purposes of this study to indicate 'sufficiently good' alignment characteristics.

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 types analyses judgments about how 'well' assessments and 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 to the relevant standards. 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 based on a relatively simple calculation. Determining the degree of alignment in this approach is straightforward and easy to justify based on the statistical methodology employed. However, justifying just what specific degree of alignment is 'acceptable' and should be set as the criteria for good alignment is 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,

psychometric requirements, and noise in the form of measurement error, 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 the summary results appropriate to this report. The data set to perform these analyses are provided in the first three subject specific appendices of this report, while Appendix B provides some example diagnostic analyses drawn from the current data set in order to model for the Department how the data-set might be used to inform decisions about maintaining or improving alignment as new items are cycled into and out of the current assessment forms. In other words, analyses designed to support further growth of the state's assessment program.

While further studies incorporating achievement results with alignment characteristics of assessment materials and instructional practice may provide the necessary empirical evidence to support a sound argument for determining the 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 0.50. The rationale behind this value follows.

Standards offer a set of goals or objectives for student learning. 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 objectives 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 using SEC increases dramatically. As a simple rule of thumb, the SEC sets 0.5, the middle of the alignment index range, as a reasonable expectation for aligned content. It leaves room for contextual constraints, student/assessment needs, and incidental (see content descriptions section) content, while establishing a minimal threshold (half) for assessments' content to hold in common with the content of the relevant standards along each of the alignment dimensions.

Discussion of Findings

As indicated by the results presented above, with few exceptions, the assessments used by the state across the grades and subjects covered by this study reveal strong levels of alignment. The results make clear that the design of the assessments attend to the content embedded in the standards, and the implementation of that design yielded assessment instruments with good alignment characteristics across the board as measured by the SEC methodology.

There are a number of mediating contextual issues that should be considered in making final determination of any alignment result. For example, the selection of an appropriate alignment target may justify a narrowing of the standards content considered for alignment purposes (discussed in more detail below). Moreover, while the threshold measure provides a convenient benchmark against which to compare results, it is at the end of the day, a measure selected by convention, and the reader would be well-advised to use these measures as *indicators* of alignment that must be considered within the real-world contexts of assessment validity and economic feasibility.

The reading assessment alignment results are very strong, with 27 of 28 indicators across all grade levels easily exceeding the 0.5 threshold. The one exception is for OAI at grade 4. Fine grain results, available through Appendix E indicate two separate alignment issues related to the grade 4 assessment. One concerning the breadth of sub-topics assessed within Vocabulary (a topic coverage issue) and the other concerning the performance expectations targeted for reading content associated with Comprehension (a performance expectation issue). Within Vocabulary results indicate that the assessment touches on only one Vocabulary topic among 13 touched on by the grade 4 standards. Within content associated with Comprehension fine grain results indicate that alignment would be improved with a shift in performance expectations from Recall and Explain to Use and Analyze. For further explanation about performance expectations, see the Content Descriptions section of the report.

In mathematics, all assessments were held to the full span of mathematics content, regardless of whether a particular content area was actually targeted as part of the assessment program for a given grade level. This sets a more challenging alignment criterion for the grade-specific mathematics assessments. Nonetheless, in only three of twenty-one instances did the indicator results dip below the 0.50 threshold. Relatively weak alignment measures are noted for the grades 3 and 7 overall alignment indices (OAI), the most sensitive and demanding of the alignment indicators, as well as the performance expectation (PE) indicator for grade 3. All other indicators for mathematics at all other grades exceeded the 0.50 measure.

Whether the weak results reported for grade 3 and 7 mathematics, or grade 4 reading assessments are significant will ultimately be a decision for the state to make. If decisions in the assessment design for any of these tests preclude one or another topic area from being assessed, then the indicator measure reported here may have under-reported the level of alignment to the more narrowly targeted subset of grade 3 or 7 standards. However in both cases it appears that it is performance expectations more than topic coverage that underlies the low OAI measure. This point to a shift in question format to assess more challenging performance expectations such as analysis, application, evaluation and integration. These can be challenging performance expectations to address in a standardized multiple choice assessment format, and while other formats are possible they are expensive and present their own challenges, including scoring reliability and validity.

To better understand the nature of these weak measures, Appendix B provides an example of a fine grain diagnostic analysis using grades 3 mathematics as an example to demonstrate what a diagnostic analysis looks like. Such analyses are intended to provide additional information to help the state in determining the adequacy of the alignment measures for grade three in particular, but potentially for any grade-level, subject or assessment form that one might want to examine more closely.

Once student performance data has been collected (Phase III of the study), additional information will be available regarding the impact of the assessments' alignment characteristics on student performance, controlling for students' opportunity to learn standards-based (and/or) assessment-based content. Such analyses may provide additional data to assist state leaders in determining the adequacy of the state's assessment program.

The results reported here mark a good beginning for the larger study of which this alignment study represents only one part. With the collection of instructional practice data to be provided in the coming months (Phase II) along with results of student performance on the assessment examined here (Phase III), the analysis team will have the necessary data to better understand and describe the impact of

instructional practice and assessment design on student achievement, thereby providing the means to determine the relative health of the state's assessment and instructional programs.

Perhaps more importantly, the results from the full study will provide both teachers and others with valuable information regarding the curriculum and assessment strategies employed in classrooms around the state and their impact on student learning.

Conclusion

This study collected and examined a comprehensive set of content descriptions covering the full span of the assessment instruments for mathematics and reading in grades 3 through 8, as well as one end of course assessment each for high school mathematics and reading. The resulting content descriptions provide a unique set of visual displays depicting assessed content that and provide the Department a rich descriptive resource for reviewing and reflecting upon the assessment program being implemented throughout the state.

Alignment analyses indicate the math and reading assessments administered by the state are for the most part very well aligned. Marginally low alignment measures were noted for grade 3 and to a lesser extent, grade 7 mathematics and grade 4 reading. Appendix B provides a detailed analysis of the grade 3 alignment results in order for the Department to better understand the alignment issues identified there while also providing an example of how diagnostic analyses can be conducted using SEC data.

The appendices attached to this report provide the complete set of detailed content descriptions of the state's assessments in mathematics and reading (soon to be followed by science). Appendices D, E, & F provide 'content viewers' of mathematics, reading, and science respectively. Each content viewer is an interactive Excel file that allows the use to select any specific assessment form, or the grade-specific aggregate description based on all three forms analyzed for a given grade. The content viewers also provide alignment results and summary marginal results for any description selected, as well as a complete set of content descriptions for the state's assessments and standards for all grade levels analyzed for the study. The content viewers require Excel version 2010 or newer, and must be running with Macros-enabled on a PC platform to function properly.

References

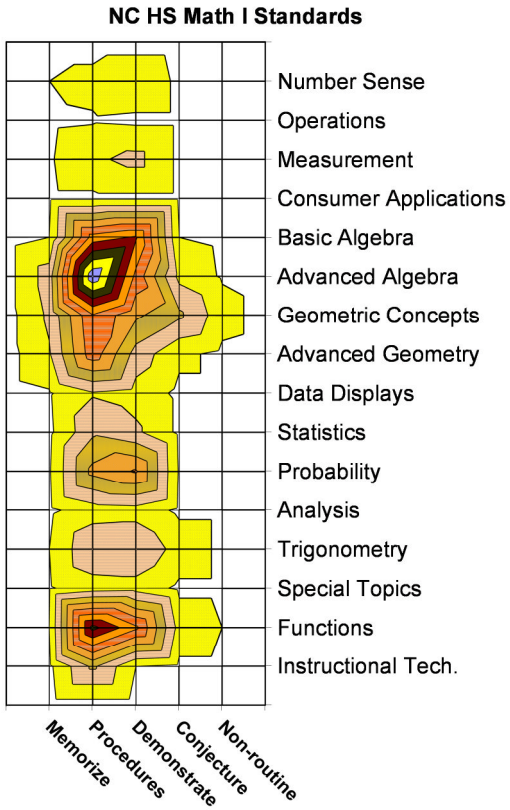
Porter, A.C. & Smithson, J.L. (2001). Defining, developing, and using curriculum indicators. CPRE Research Report Series RR-048. Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.

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.

Appendices

Appendix A: Describing Academic Content

Figure 1



Alignment analyses conducted using the SEC approach first requires the collection of content descriptions for each standard and assessment to be analyzed. At the heart of this approach to describing content is a multidimensional taxonomy that offers a systematic, quantifiable and nuanced way to describe content at a fine-grain level of detail to describe content knowledge and skills.

The process for collecting content descriptions is described below, while figure 1 at left provides an example of a content map used to display visually informative descriptions of the academic content embedded in assessment and standards documents. The map presented in figure 1 in this case describes the content standards for high school mathematics recently adopted by North Carolina.

The Three Dimensions of Content

Note that the content description provided in figure 1 is presented along three axes or dimensions; the Y-axis, represented by the list of math topics presented to the right of the image, the X-axis represented by the five categories of performance expectations running across the bottom of the image, and the Z-axis (displayed by contour lines and color bands), indicating the relative emphasis for each intersection of topic and performance expectation. These three dimensions form the foundational structure for describing and analyzing content using the SEC approach. Academic content is described in terms of the interaction of topic and performance expectations. By measuring each occurrence of some element of content (topic by performance expectation) a measure of the relative emphasis of each content topic as it appears in the content description can be obtained.

For example, figure 1 indicates that the topics with the strongest emphasis in North Carolina's high school standards are in the areas of Advanced Algebra and Functions, particularly at the performance level of procedures (equivalent to DOK level 2). figure 1 also indicates that the most challenging performance expectation in the high school mathematics standards address Geometric Concepts at the level of non-routine or novel problem-solving (equivalent to DOK level 4).

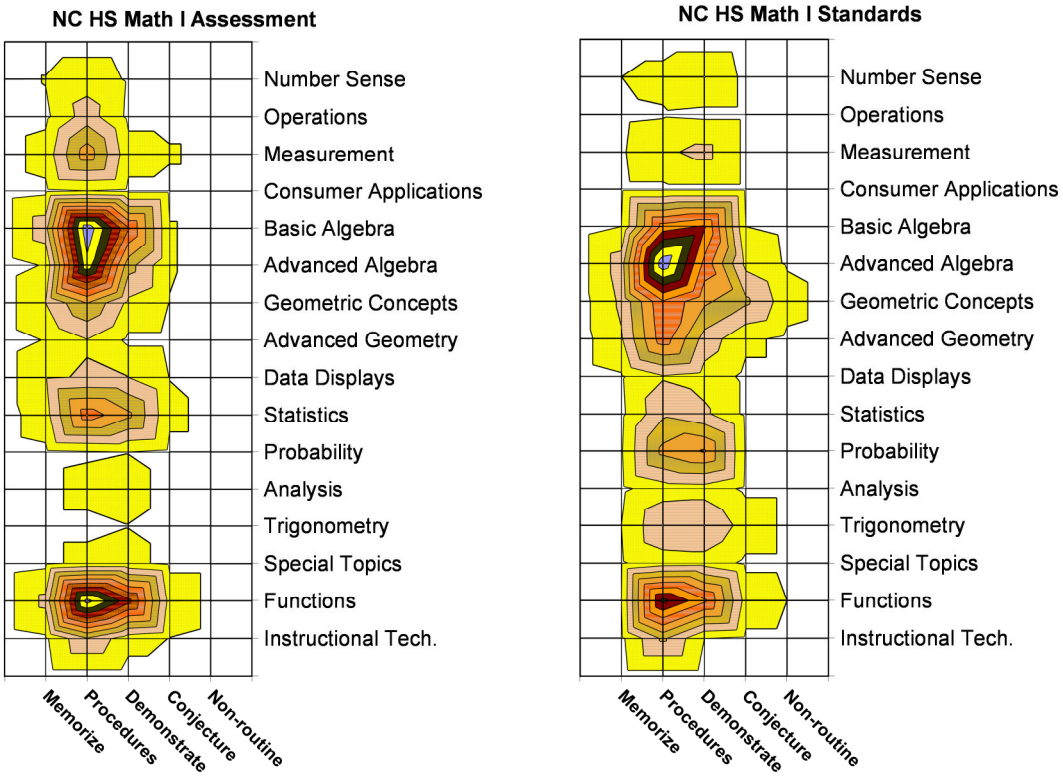
Content Analysis Workshop

Content descriptions used to generate visual displays like figure 1 are collected using a particular type of document analysis referred to as content analysis. All content analysis work is conducted using teams of content analysts (educators with K-12 content expertise) that receive a half-day of training at content analysis workshops where specific documents are then analyzed by content analysis teams over a one or two day period.

North Carolina hosted a content analysis workshop as part of the alignment study in January, 2015 at the McKimmon Conference and Training Center in Raleigh, North Carolina. There 10 subject-based teams of content analysts were formed from more than 30 teachers and other content specialists and trained to conduct independent analyses of 51 assessment forms for mathematics, reading, and science for all assessed grades. Each team was led by a veteran analyst familiar with the process and able to facilitate the conversations among team members. The process involves both independent analysis and group discussion, though group consensus is not required. The coding process is described in more detail in the next section.

The alignment analyses of any two content descriptions are based on detailed comparisons of the descriptive results collected during the content analysis process. While alignment results are based on a straightforward computational procedure and provide precise measures of the relationship between two descriptions, simple visual comparison of two content maps are often sufficient to identify the key similarities and differences between any two descriptions. For example, a simple visual comparison of the two maps presented in Figure 2 suggest that while distinctions can be identified, there is a generally similar structure to both that suggest reasonably good alignment of the two descriptions.

Figure2



A careful review of the maps presented in Figure 2 would essentially reveal the same similarities and differences that would be notable in the alignment results examining the relationships depicted above. The alignment algorithm(s) however provide a simple and systematic means for reporting detailed measures of these comparisons much more precisely than possible with a visual review. Quantitative

analyses also provide a basis for establishing guidelines for determining whether two descriptions are adequately aligned, and permit succinct reporting of alignment results in simple tables.

The detailed procedures for conducting content analyses are outlined in Appendix E, though a closer look at the actual coding process is provided through an example below. The process is systematic, objective, and quantifiable, yielding data results that support the types of three dimensional descriptions depicted in content maps like the ones presented in figures 1 and 2 above.

Anatomy of a Content Description

In order to provide the reader a better understanding of how content descriptions are collected it will be useful to provide an example, based a single assessment item. A detailed description of the content analysis process can be found in Appendix E. For the current purpose, suffice it to say that a content description consists of a topic code (a number) combined with a (letter) code indicating a specific performance expectation. These number and letter codes are retrieved from a reference document, a portion which is displayed in figure 3.

Each analyst first conducts an individual review of the document to be analyzed, using content 'codes' to describe the content item by item (for tests) or standard strand by standard strand (for standards). After the individual 'coding' analysts engage in team discussions around the rationales used to justify one or another set of content codes to describe a specific section of the relevant document. While consensus is not required, analysts have the opportunity at this point to alter their descriptions. Results of all analysts are processed and averaged to produce the final description.

Each content description is intended to describe what a student needs to know (topic) and be able to do (performance expectation) in order to correctly answer a given assessment item. Using the SEC methodology a content analyst has the option of entering from 1 to 3 content descriptions (number-letter pairs) for any one assessment item. While some assessment items can be sufficiently described to the analyst's satisfaction with a single number/letter code entry, other items require multiple entries to adequately capture what a student needs to 'know and be able to do' in providing a correct response.

Figure 3

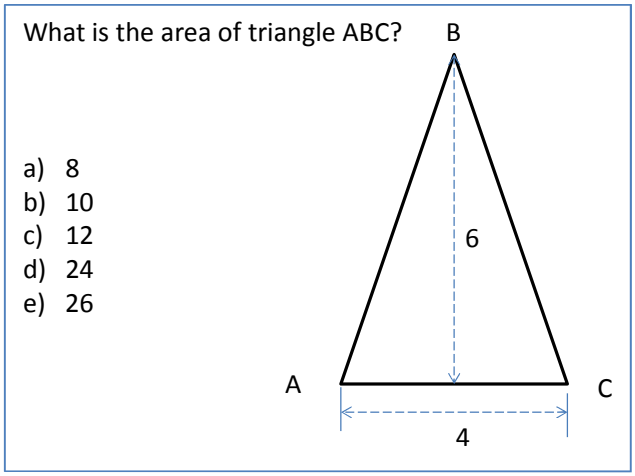
K-12 Mathematics Taxonomy

300	**Measurement**
301	Use of measuring instruments
302	Theory (arbitrary, standard units and unit size)
303	Conversions
304	Metric (SI) system
305	Length and perimeter
306	Area and volume
307	Surface Area
308	Direction, Location, Navigation
309	Angles
310	Circles (e.g., pi, radius, area)
311	Mass (weight)
312	Time and temperature
313	Money
314	Derived measures (e.g., rate and speed)
315	Calendar
316	Accuracy and Precision
390	Other
400	**Consumer Applications**
401	Simple interest
402	Compound interest
403	Rates (e.g., discount and commission)
404	Spreadsheets

900	**Data Displays**
901	Summarize data in a table or graph
902	Bar graph and histograms
903	Pie charts and circle graphs
904	Pictographs
905	Line graphs
906	Stem and Leaf plots
907	Scatter plots
908	Box plots
909	Line plots
910	Classification and Venn diagrams
911	Tree diagrams
990	Other
1000	**Statistics**
1001	Mean, median, and mode
1002	Variability, standard deviation, and range
1003	Line of best fit
1004	Quartiles and percentiles
1005	Bivariate distribution
1006	Confidence intervals
1007	Correlation
1008	Hypothesis testing
1009	Chi Square

For example, consider an assessment item that requires a student to recall which formula is needed to calculate the area of a triangle, and apply that formula to compute the alignment of a given triangle. There are at least four potential content descriptions available to an analyst considering this item, built from the possible combination of two performance expectations (recall [B] and computation [C]) with two topic codes: Area (topic[306]), and Triangles (topic[707]), though if a graphic with a picture of a triangle with numbers to indicate the size of each side and height of the triangle, as in the example to the right more content descriptions would become relevant depending on the performance level(s) an analyst attached to the student's need to interact with the graphic display (topic[901]) in answering the item.

Figure 4



Thus a description of this example item could reasonably consist of any combination of the following content 'codes'; 306B, 306C, 306D, 707B, 707C, 707D, 901B, 901C, or 901D, and possibly others depending on the grade level of the student answering the item. However the content analyst is limited to no more than 3 content code combinations to describe a single assessment item and thus would have to make choices among the various options to select those that best describe the knowledge and skills required to solve the problem.

When describing standards, analysts may use up to six content codes to describe a single standard strand. Content analysts discuss their coding decisions with their colleagues on 3-6 person teams of analysts. While analysts may change their decisions after the discussion, consensus is not required and results are averaged across analysts. This allows a surprisingly rich descriptive language for content analysts to use in describing what students need to know and be able to do to be successful on a given assessment.

Incidental Content

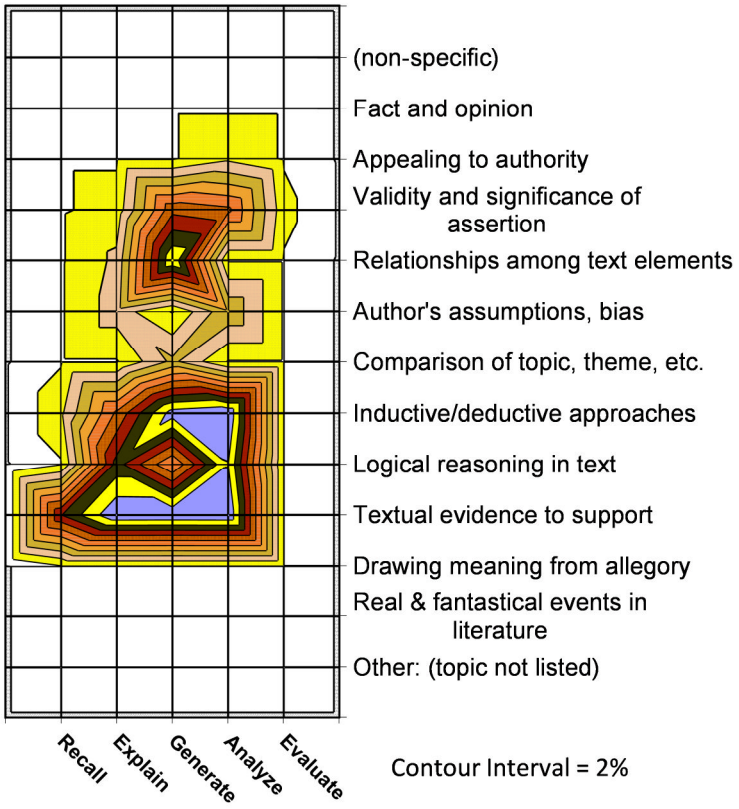
With such fine grain descriptive detail (and descriptive power) available using the SEC approach, it is clear that a good deal of 'incidental' content gets included in assessment items. 'Incidental' in the sense that they represent either prior knowledge, or involve additional performance expectations than may have been the primary assessment target of the item. As a result, the conventional criteria set for judging alignment is based on the assertion that the resulting content description should have at least half of its content aligned with the content emphasized in the relevant target (standards). In terms of the alignment indices used for this report, this is equivalent to a measure of 0.50 for any of the four alignment indices addressed in the study.

Fine Grain Content Maps

The topic codes used in the example above were at the level of sub-topic, and it is at this level that content descriptions are collected. Thus for each topic noted in a coarse grain content map like those presented in Figures _ & _ a fine grain content map can be generated to display the individual sub-topics and the distribution of relative emphasis across the topics and performance expectations making up that

topic area. Figure 5 shows a fine grain map for Critical Reading as described for grade 8 reading assessment.

Figure 5

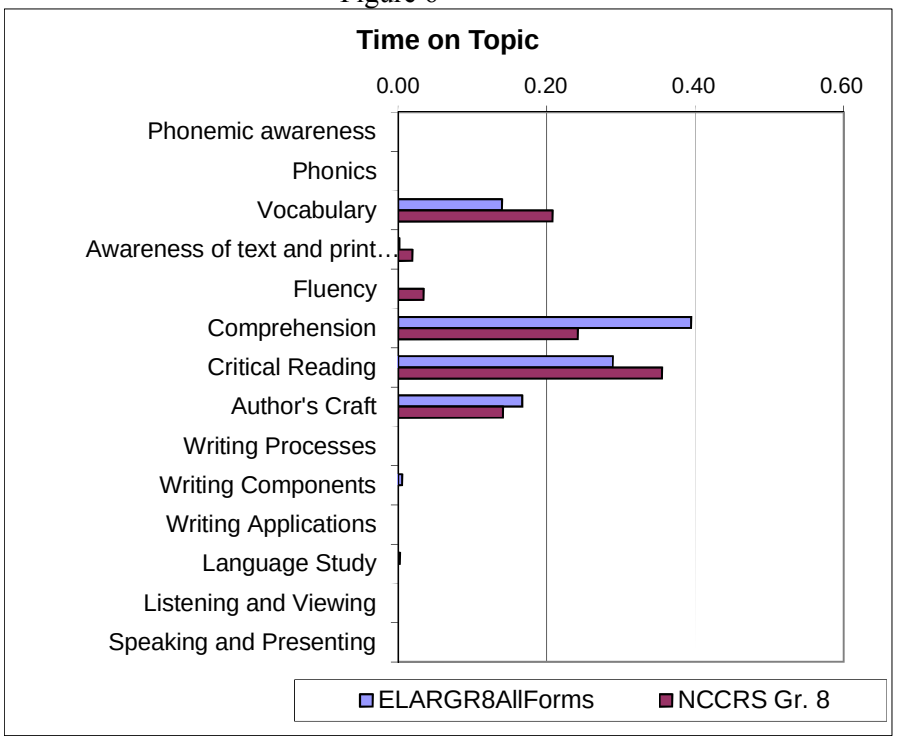


The map at right (figure 5) displays the results of a fine grain content matrix. Each intersecting line in the map corresponds to a cell in the content matrix table. The image of the map that results is generated from the data points at these intersection lines. To read the value for specific content noted at a specific interval, count the number of contour lines between the intersecting point and the nearest border and then multiply by the contour interval (in this case 0.2% (.002). For example, looking at fine grain topic 'textual evidence to support' the values explain, generate and analyze all have values of 2% or higher. On a fifty item test this would be the equivalent of one test item for each intersection in the blue portion of the map. Thus the equivalent of three items on a fifty item test would be focused on textual evidence to support an argument. As a practical matter it is more likely that textual evidence plays some role in twice as many items, shared with other content descriptions for the same items. This provides some sense of the level of detail the SEC languages permit in describing the complexities of detailed content descriptions. Such detail is generally more than is necessary for summary descriptive purposes, where coarse grain maps and alignment results are sufficient. It is very much like looking at something under magnification, for most purposes it is more detail than is useful, but for certain purposes the additional magnification is useful for more careful scrutiny, such as to adjust assessments to improve alignment (see Appendix D).

Marginal Charts

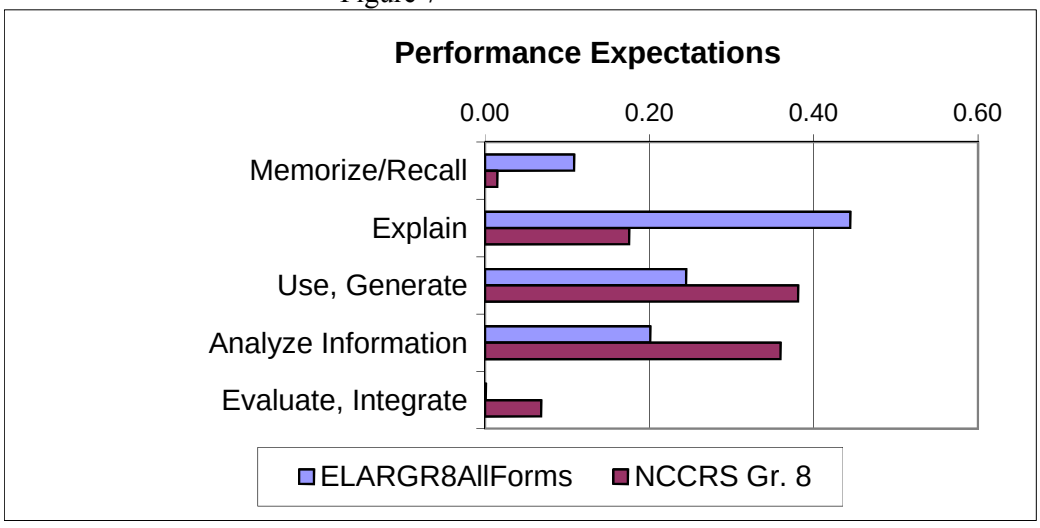
The reader was introduced to marginal measures in the alignment analysis section of the main report. These marginal results can also be displayed in bar charts to provide a visual depiction of the data results. The bar chart displays allow for easy comparison across the two descriptions being examined for alignment, providing a quick sense of the distribution of relative emphasis across topics or performance expectations. Figures 6 and 7 provide examples for grade 8 reading. The X-axis reports the proportion of score points for each topic (figure 6) or performance expectation (figure 7).

Figure 6



Topic marginal measures provide a quick and easy look at topic coverage across the two descriptions that quickly makes apparent the differences in the relative emphases of the two descriptions. This data can be particularly helpful in conducting diagnostic analyses where the goal is to identify strategies to *improve* alignment. Results in figure 6 indicate a strong emphasis on content related to comprehension of the test, compared to the standards descriptions which appear to place more emphasis on content related to critical reading. Such differences in relative emphasis are to be expected, and in some cases even necessary to assure that test specifications are adhered to or other psychometric needs are addressed. Only when the summary measures weak alignment is it recommended that such results be used to inform decisions on shifting an assessment's focus.

Figure 7



Similarly marginal measures for each of the five performance expectations provide a quick summary of the performance expectations represented in the two descriptions, providing a profile of the performance expectations of each and indicating the differences in emphasis represented by both. If the data in figure 7 were being employed for improvement purposes, figure 7 would suggest that grade 8 assessments shift the focus of performance expectations from performing procedures to analyze information.

Maps and charts like those described above can be accessed for any of the documents analyzed for this study using interactive content viewers provided in Appendices B (math), C (reading) and E (science). These content viewers are macro-enabled Excel files. Macros must be enabled in a PC-based Excel application for the interactive features to function.

APPENDIX B: Diagnostic Analyses Using SEC Data

Purpose

Just as with assessment data, SEC results can be used for either summative or formative purposes. The primary goal of the main report was to provide summative results of our alignment analyses and determine how well state assessments meet the alignment requirements of a standards-based accountability system. The purpose of this section is to illustrate how the SEC data can be used for formative purposes, i.e. to *improve* the alignment of one or another component of the state's assessment program. This diagnostic analysis utilizes fine-grain descriptive and alignment results in order to identify areas of weak alignment and suggest strategies for possible improvement..

Selection

As noted in Table 5 of the study report, two grade levels of mathematics assessments and one reading assessment reported relatively weak, overall alignment (OAI) measures. This is the most demanding of the alignment measures, as it incorporates all three dimensions together. Among the sub-measures based on individual alignment dimensions, only one assessment reported a below-threshold measure. That assessment, grade 3 mathematics will serve as the focus of this example of conducting a more fine-grain 'diagnostic' analysis designed to identify specific areas of weak alignment in order to inform strategies for improving alignment in future versions of the assessments.

Note that the necessary tools and information needed to conduct a diagnostic analysis of any grade, or grade-specific form is possible using Appendices D – F (requires Excel 2010 or newer with macros enabled running on a PC platform). The following analyses provide a step-by-step approach that could be replicated for any given selection of alignment targets and assessment descriptions (either grade-level or form-specific) using the data sets supplied in the Appendices.

It is worth noting that the procedures for conducting formative analyses of assessment alignment and instructional alignment are similar in that both use alignment indicators and content descriptions to identify strategies for improvement. Due to similarities in the diagnostic process, the formative analysis of assessment alignment presented in this report will have many parallels to the structure of a teacher personal alignment report.

Formative Analyses

Formative and summative analyses alike combine descriptive and analytic results. The primary difference between summative (evaluative) and formative (diagnostic) analyses of SEC data relate to the level of detail at which the analyses are conducted. In SEC terminology this distinction is referred to as coarse grain and fine grain. Coarse grain results provide an efficient summary of alignment characteristics that offer a reasoned, evidence-based approach to assessing the adequacy of state efforts to implement an aligned standards-based system.

Behind these summative results lies a large body of empirical, descriptive data concerning the specific assessments and forms through which the state assessment program is delivered. These descriptive data provide the basis for the detailed analyses of the alignment characteristics of the state's mathematics and reading assessments. These data can support formative analyses designed to identify strategies most likely to lead to improved alignment between any two content descriptions.

Summative alignment results generally provide the starting point for narrowing the focus in searching for areas of misalignment as is done in formative analyses. While the alignment indicators at both coarse grain and fine grain levels assist in targeting those areas where improvements in alignment are needed,

the ultimate goal is to identify or target the specific descriptive results that will provide the most information for informing alignment improvement efforts.

Thus we begin with Figure 1, which contains the summative alignment results for grade 3 mathematics from the body of this report. As can be seen in Figure 1, of the four indicators presented, two are below the 0.50 threshold and two above. The focus here will be on the two below the threshold because these weak summary measures suggest where the majority of the misalignment occurs.

Figure 1

Grade 3 Mathematics
Coarse Grain Alignment Indices Summary
(0.50 or greater = well-aligned)

	BR	TC	CC	OAI
AI	0.68	0.68	0.41	0.40

BR Balance of Representation
TC Topical Concurrence
CC Cognitive Complexity
OAI Overall Alignment Indicator

As discussed in the main report, the OAI is a composite measure, sensitive to variation on three dimensions of alignment: 1) balance of representation (BR), 2) topic coverage (TC), and 3) performance expectations (PE). Setting OAI aside for the moment, we see that only one of the three underlying alignment measures is low - the indicator for performance expectations. This strongly suggests that the fundamental alignment problem for grade 3 mathematics is going to relate to non-alignment in the area of performance expectations.

This is borne out in examining the descriptive results underlying the CC indicator . Figure 2 shows the marginal) values for the five performance expectations (i.e., the overall emphasis of the standards and assessments on cognitive demand categories without regard to topic emphasis) for the content descriptions of the state's grade 3 mathematics standards (in red) and assessments (in blue)

The chart in Figure 2 indicates that the assessment lags one performance category behind the standards; as can be seen by imagining how similar the bars would look if the blue bars were simply moved one level down in the chart. The two descriptions would then be very highly aligned with regard to performance expectations. In fact under those circumstances the coarse grain cognitive demand indicator measure would be a remarkable 0.88.

Figure 2

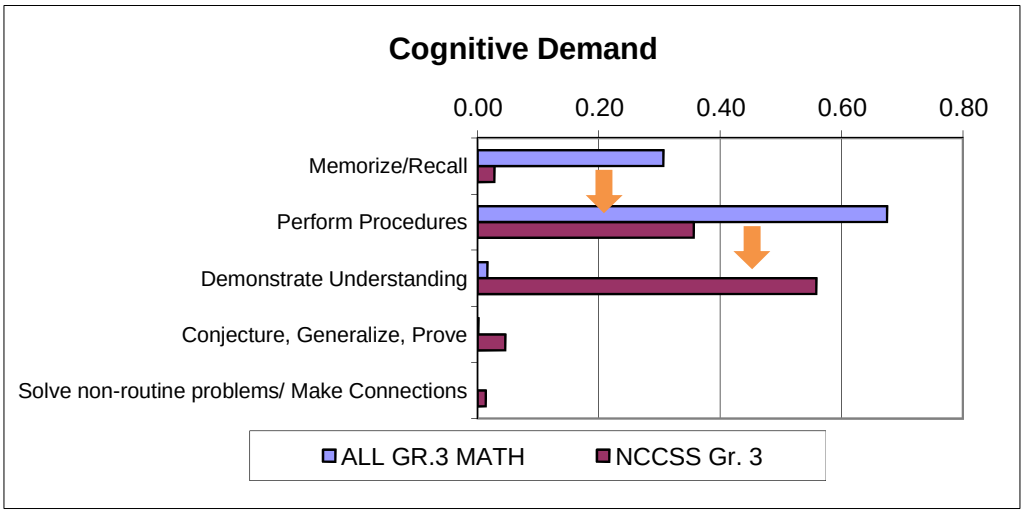
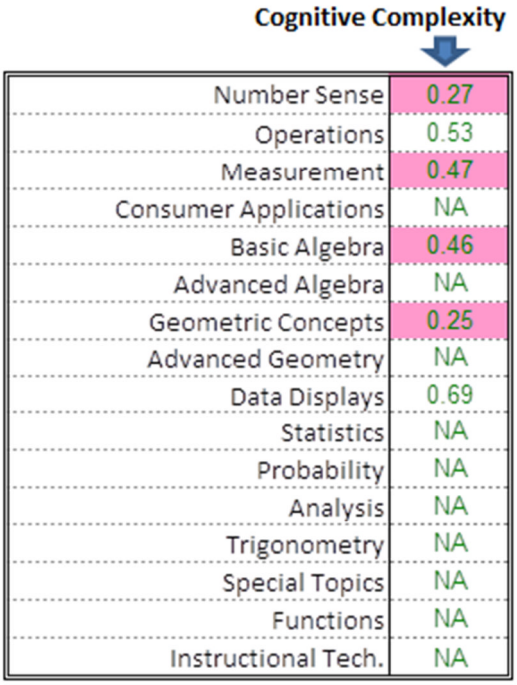


Figure 3



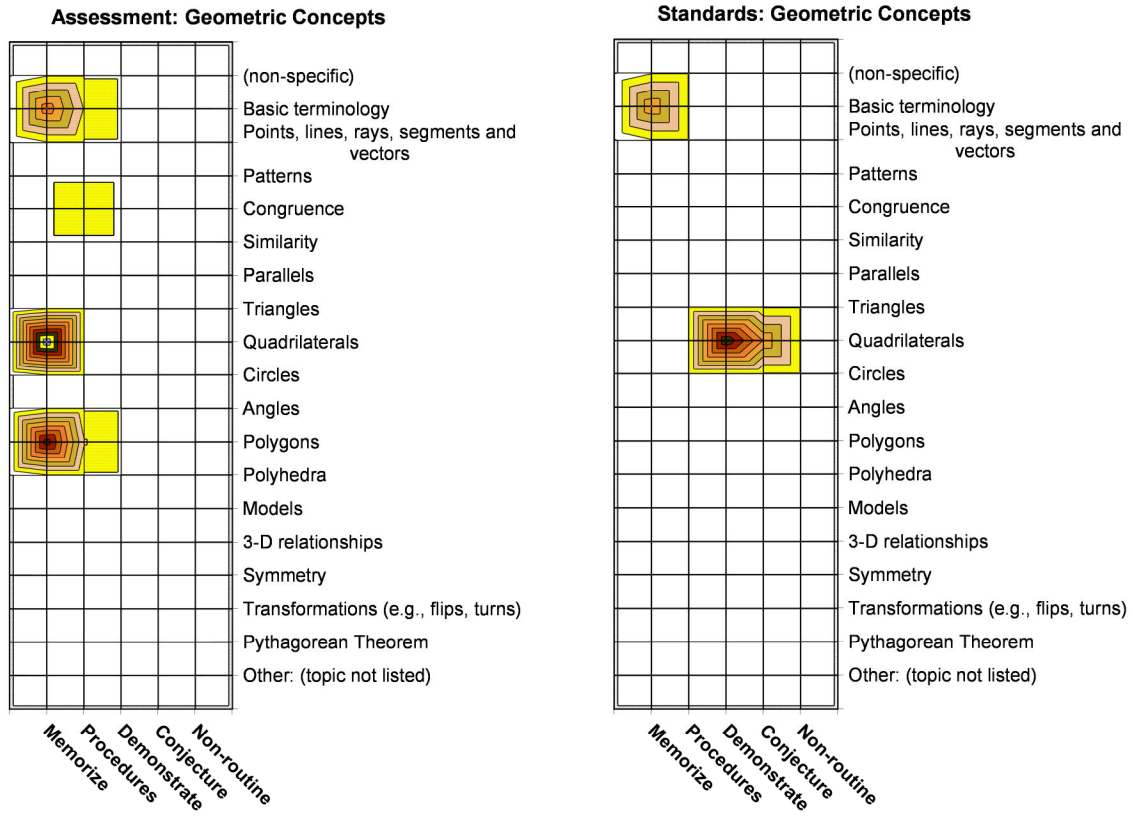
While the results reported in Figure 2 make clear that an insufficient number of items target the performance expectation 'demonstrate understanding', it does not give us any information about which topics would be best targeted for this shift in performance expectation.

It is possible to drill down on any one of the four summative indicators to get a more detailed look at the relevant alignment indicator measure for each topic area. Doing so allows one to identify those topic areas with low indicator measures to guide the next steps in the process.

Figure 3 reports the fine grain, topic area results for performance expectations.

As can be seen, many of the topic areas listed in figure 3 are not applicable (NA), indicating that neither the standards nor assessment descriptions contain content for that area. Of the six topic areas assessed, two exceed the 0.5 threshold, two additional measures are quite close, and two (Number Sense, and Geometric Concepts), reveal very low performance expectations measures. It is to these two topics that we turn next.

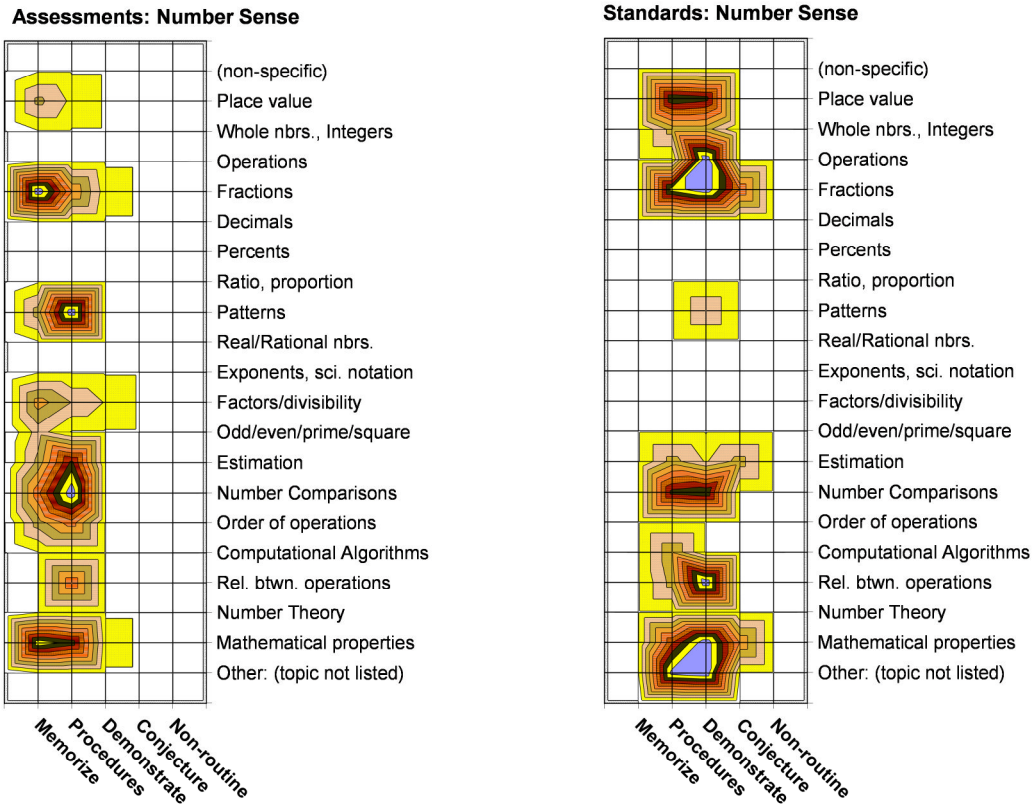
Figure 4



At this point in the analysis it is useful to return to content maps to get a visual display of the fine grain content embedded within the grade 3 standards and assessments with respect to Geometric Concepts and Number Sense. The maps displaying these results are reported in Figures 4 and 5 respectively.

As can be seen by the fine grain content maps for Geometric Concepts in Figure 4, only two sub-topics are addressed in the grade 3 standards; basic terminology and quadrilaterals. The assessment map on the left shows a good match on basic terminology at the recall level, but the assessment also touches on a couple of topics not in the standards (Congruence, and Polygons) and targets quadrilaterals at a different performance level than is addressed in the grade 3 standards (Recall instead of Demonstrate Understanding). Here is a good example of an assessment addressing the correct sub-topic at the wrong performance level. As demonstrated in this analysis, this is exactly the type of mismatch to which we were alerted by the SEC summary indicators presented in Figure 1.

Figure 5



The content descriptions for Number Sense, displayed in Figure 5, are more complex, as Number Sense is clearly an important topic area for grade 3. Once again we see a pretty good match up on sub-topics. One can identify an occasional topic that is reflected in one description but not the other, but generally topics covered in the standards are also covered in the grade 3 assessments. There are six sub-topics in the standards that show a moderate to strong degree of emphasis; place value, operations, fractions, number comparisons, relations between operations, and mathematical properties. By comparison, the assessment touches on content for five of these six sub-topics; operations being the lone sub-topic of the six not assessed to some degree.

The mismatch that occurs once again is with the performance expectations addressed for specific sub-topics. With the lone exception of Number Comparisons, the performance expectations patterns of emphases for assessments tend to be pitched one or two levels below that of the relevant standards.

Based on this analysis, if the state wanted to achieve relatively large gains in alignment while making relatively few changes to the mathematics assessments, the most efficient strategy would be to replace Recall items with items that require students to provide evidence that they understand the concepts at issue in the assessment item.

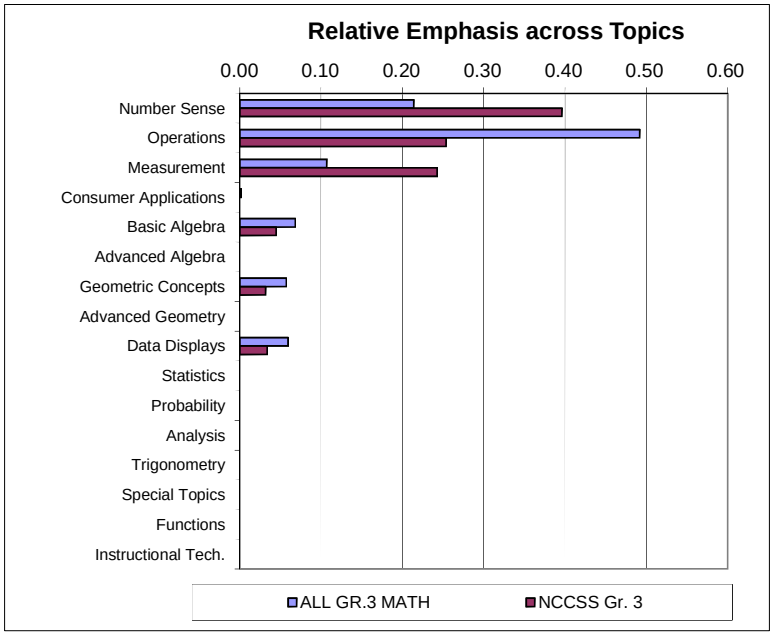
Balance of Representation

Thus far our analyses have focused on topic and cognitive demand in comparing the grade 3 assessments to the grade 3 standards. It will be recalled that BR refers to the relative emphasis placed on one or another topic. When examining BR at the fine grain level of detail it is important to understand that what constitutes relative emphasis changes, depending on whether one is talking about assessment, standards or instruction. For assessments relative emphasis is defined in terms of score points, i.e. what portion of a student's score is based on this content. For standards relative emphasis is more a function of the relative number of times a given bit of content as defined by SEC is addressed across the content strands of the standards. With instruction, relative emphasis is simply the proportion of instructional time spent on a given topic.

While the overall indicator measure for BR was good (0.68), it is worth looking at the fine grain results, topic by topic, to identify instances of especially low alignment. Figure 6 shows the relative emphasis given to each topic area for both the grade 3 assessment and standards..

As revealed in Figure 6, both the grade 3 mathematics assessments and standards show a strong emphasis on three primary topic areas: Number Sense, Operations, and Measurement. However Figure 6 also reveals that the relative emphases each description gives to these three topic areas is quite different. While the state grade 3 math standards appear to place the greatest emphasis on number sense (0.40), the assessment places the greatest emphasis on Operations (0.50). This over-emphasis on Operations (relatively speaking) comes largely at the expense of emphasis on Number Sense and Measurement.

Figure 6



While the emphasis on Operations may be based on a legitimate assessment design decision, these results indicate that practically any increase in the assessment of content related to Number Sense or Measurement would lead to improved alignment..

Discussion

The SEC data set provides a sound framework for collecting content descriptions that enable coarse- and fine-grain , quantified comparisons based upon a simple iterative algorithm that provides alignment indicators to summarize alignment characteristics and relationships. While thresholds can be set at reasonable levels based on rational argument (e.g. 0.50) or normative findings (i.e. above or below average), the implications for assessment design or classroom instructional practice remains more art than science. In other words, SEC data is designed and intended to *inform* policy decisions, not *make* them. In any analysis it is the reader, or the organization's decision-makers that must digest the results, and along with other factors, make decisions intended to improve the learning opportunities of those students under their purview. This is the case whether the data is being examined by a policy-maker, assessment director, district/school administrator, or teacher. In each case it is essentially the same analytic process.

While the analyses presented in this section make clear what strategies *could* be employed to move toward better alignment of the grade 3 assessments; nothing reported here can alone answer the policy questions of what strategies *should* be used to improve alignment of these or any other assessments; or even whether the alignment of these assessments need to be improved at all. The reader must assimilate the descriptive results and consider them within the contexts of practical feasibility.

Though not explicit, there is a logic and procedure underlying the analysis done in this section that can be generalized for use with any two content descriptions of interest. Those procedures are outlined below to enable relevant department staff to review the study results in more detail, using the interactive appendices A-C. Thus in describing the general procedures for conducting detailed alignment analyses reference will be made to particular charts, tables or content maps to be found in the relevant subject-based appendices.

Analyses begin with the summative coarse grain results for the selected target and assessment. It will be recalled that there are four summative measures; a general overall indicator (OAI) that is based upon, or sensitive to variation on each of three foundational 'dimensions' of alignment. These are Balance of Representation (BR), Topic Coverage (TC), and Performance Expectations (PE). By convention WCER researchers have set 0.50 as a threshold measure for adequate alignment; implying that measures below 0.50 deserve further attention, particularly if alignment improvement is a goal. The OAI measure serves to identify any assessments that might deserve a deeper look. For this study, alignment results indicated that the grade 3 math assessment had the weakest alignment scores among the assessments examined. For that reason, grade 3 math was selected to use as an example for this description of diagnostic analyses.

At whatever level of examination being conducted, analyses begin with the OAI measure, followed (for those cases that fall below the selected threshold) by review of the marginal measures for each alignment dimension, i.e., BR, TC, and CC. Here again the results are reviewed against a desired threshold (presumably, but not necessarily 0.50), with deeper examination guided by those indicators that fall below the threshold.

For each alignment indicator measure there is at least one, if not two descriptive displays available as an aid to understanding the data represented by that indicator. Depending on the measure, descriptive results are reported in bar charts, or content maps, or both. In the subject specific Excel-based content viewers that accompany this report, the worksheets display the results necessary for both coarse grain and fine grain analyses of alignment results. The 'Alignment Table' worksheet supplies all of the alignment indicators, both coarse grain and fine grain, organized into an alignment table that summarizes all of the alignment results for a given pair of content descriptions (consisting of an alignment target and the assessment (or instruction) being examined relative to that target).

Fine grain alignment indicators are reported by topic area in the alignment table. Both coarse and fine grain content maps content descriptions are reported in the 'ContentMap' worksheet. Finally, bar charts

detailing Balance of Representation and Performance Expectations are reported in the Marginals worksheet. The ContentMap worksheet is used to select which content descriptions are to be analyzed.

Once a selection is made from among the BR, TC, and PE indicator results, the relevant column for the selected indicator can be reviewed in the alignment table to guide selection of specific topic areas to review more closely. It is the review of the descriptive results at this fine grain level that the alignment analysis procedures are designed to facilitate, assisting one in targeting weak areas of alignment in order to consider strategies for improving alignment, or at least better understand the nature of the alignment relationship between the descriptions being reviewed. The overall process is graphically summarized in Figures 7 through 10 below. The table displayed in Figure 8 is an alignment table, from which all alignment indicators are drawn. The content viewers making up Appendices A-C all have interactive alignment tables that allow for the reporting of all the alignment indicators for all of the grade specific assessment forms analyzed for the study. Figure 8 highlights two rows in the table with low alignment indices; Number Sense and Geometric Concepts one is labeled 'A' and the other 'B', to indicate which results from Figure 8 are being detailed in Figures 9 & 10. The maps displayed in Figures 9 & 10 can similarly be accessed through the content viewers, which can be used to generate both coarse grain and fine grain content maps depicting the content of any assessment or standards target selected.

Figure 7

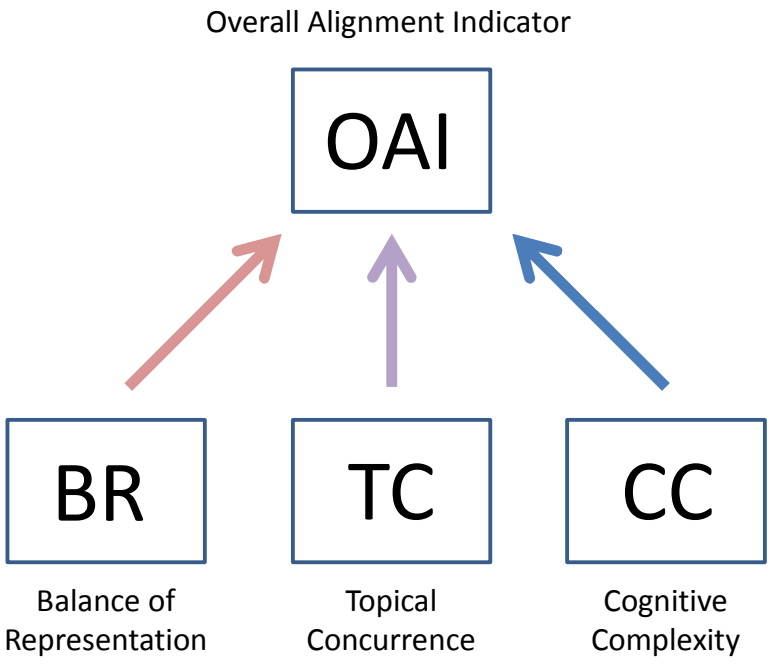


Figure 8

Alignment Analysis Summary Table

Coarse Grain	0.40	0.68	0.68	0.41
	OAI	BR	TC	CC
			(Topics)	(Cog. Dmnd.)
ALL GR.3 MATH TO: NCCSS Gr. 3	Alignment	Balance of Representation	Categorical Concurrence	Cognitive Complexity
Number Sense	0.15	-0.18	0.51	0.27
Operations	0.44	0.24	0.57	0.53
Measurement	0.32	-0.14	0.68	0.47
Consumer Applications	NA	0.00	NA	NA
Basic Algebra	0.40	0.02	0.65	0.46
Advanced Algebra	NA	0.00	NA	NA
Geometric Concepts	0.17	0.03	0.67	0.25
Advanced Geometry	NA	0.00	NA	NA
Data Displays	0.62	0.03	0.82	0.69
Statistics	NA	0.00	NA	NA
Probability	NA	0.00	NA	NA
Analysis	NA	0.00	NA	NA
Trigonometry	NA	0.00	NA	NA
Special Topics	0.00	0.00	NA	NA
Functions	NA	0.00	NA	NA
Instructional Tech.	NA	0.00	NA	NA
Fine Grain	***0.30***	***0.68***	***0.56***	***0.40***

A

B

Figure 9

Fine Grain Content Descriptions

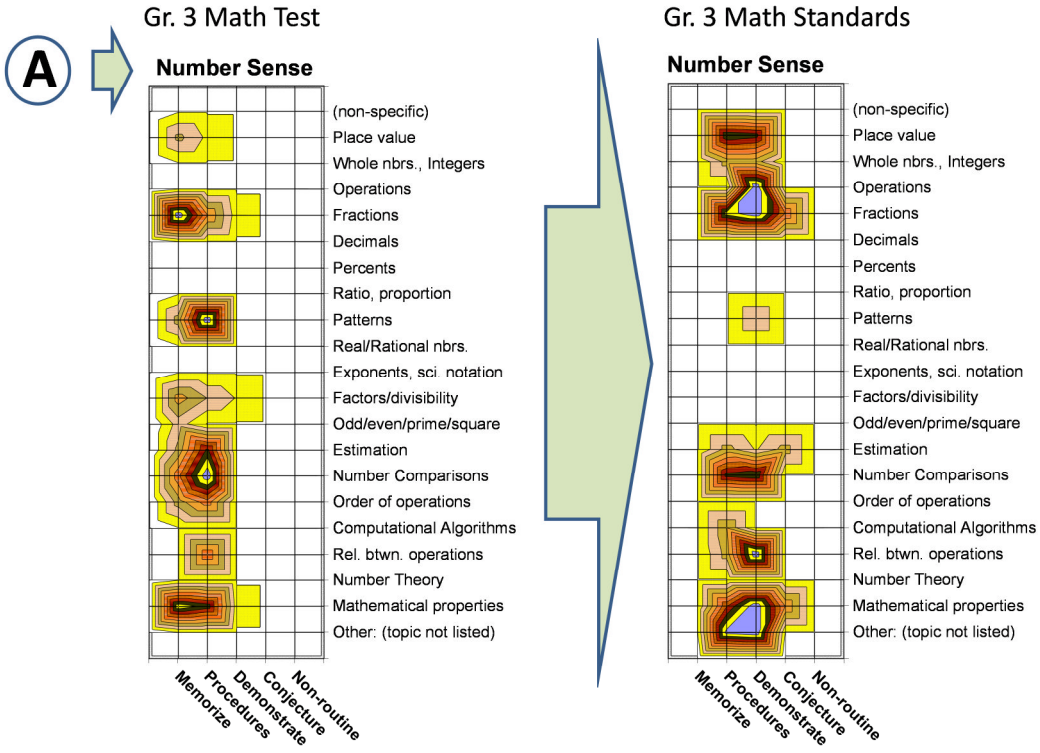
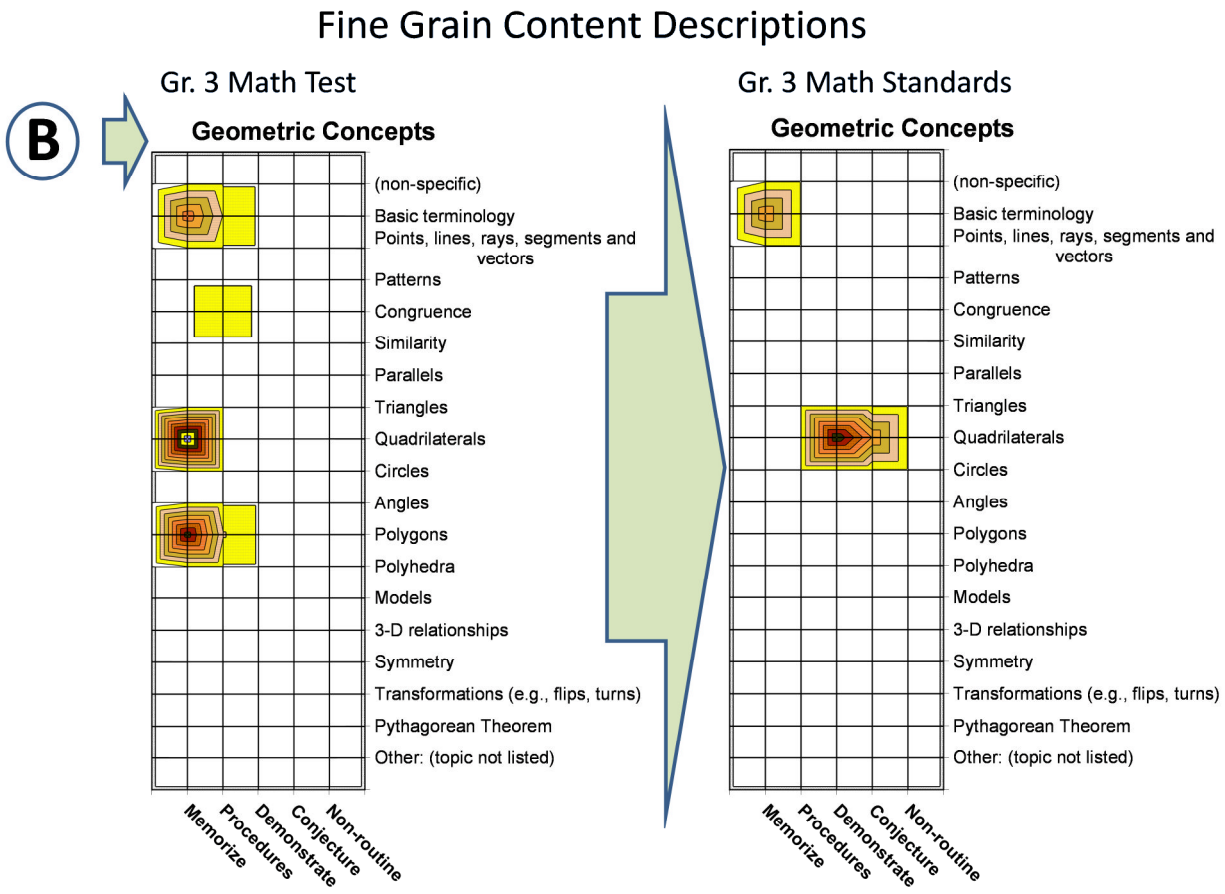


Figure 10



Using Appendices D-F

Appendices D-F house subject-specific 'content viewers'. These are macro-enable Excel-based files that provides an interactive user interface to generate alignment tables, content maps and charts with marginal measures (all the data sources needed to conduct a fine-grain analysis for any given subject and grade level, or even grade level form.

The files must be run on PC-platforms with macros-enabled in order for the interactive features to work.

Each viewer contains an introduction and set of instructions to get the user started. That with the examples of Appendix B and familiarity with the descriptive displays explained in Appendix A provides the necessary background information to examine the full data set used for the study in-depth.

Questions and technical issues may be addressed to the author.