Alignment of Assessments, Standards and Instruction Using Curriculum Indicator Data

Andrew C. Porter
John L. Smithson

The approach to alignment discussed here uses a systematic and detailed language to describe curricular content in a manner that can be applied to multiple components of the standards-based, accountability-driven educational system. By employing a rich but systematic language for describing curricular content, it becomes possible to make detailed, quantitative comparisons of the instructional content emphasized in standards, assessments and instruction. The comparison process employed yields a measure that is best described as representing the "intersect" of the two descriptions being compared. Thus, an alignment measure of '1' would indicate the two descriptive 'sets' were identical to one another, while a '0' would indicate the two had no content emphasis in common at all.

In order to better understand the power of this approach for considering alignment between key elements of the standards-based system, the nature of the descriptive language is first described, and then a real-world example of an alignment analysis is considered.

Content languages have been developed for mathematics and science. Each language employs a two-dimensional grid for organizing content descriptors. A particular content unit is described in terms of 1) a topic and 2) an 'expectation for student performance'.

Examples of topics for middle school math include "place value", "fractions", "decimals", "use of measuring instruments". The number of topics employed in a given language varies by subject and grade level. In mathematics there are 57 topic descriptors at the elementary level, 90 at the middle school level (grades 6-8), and 160 at the high school level. For science, the number of topics employed is consistently higher, ranging from 80 in elementary science, to 200 in high school science. Topics are organized into content 'areas' representing logical groupings of topics. The number of broad content areas also varies by grade and subject, ranging from a minimum of six (middle school mathematics) up to twenty-five content areas (high school science).

The "expectations for student performance" dimension (sometimes called cognitive demand), consists of five categories for both mathematics and science, though the descriptors used in each subject differ. In mathematics the five categories are: 1) Memorize facts, definitions, formulas, 2) Perform procedures, 3) Demonstrate Understanding of Mathematical Concepts, 4) Conjecture, Generalize, Prove, and 5) Solve non-routine problems, make connections. In science the five categories are; 1) Memorize facts, definitions, formulas, 2) Perform procedures, conduct investigations, 3) Communicate understanding of science concepts, 4) Analyze information, and 5) Apply Concepts/Make connections. Each expectation is defined using 3 to 8 examples.

Grain Size

Content descriptions are reported at two levels, or 'grain size' for mathematics, and three levels for science (due to the more robust topic list pertinent to science instruction). Grain size determines the number of cells used to present results. Thus in the coarse grain descriptions of mathematics, six topic areas are arrayed against the 5 categories of expectations for students, yielding a matrix of 30 cells (6 rows by 5 columns). The fine grain matrices use the individual topics in the taxonomy. The fine grain matrix for middle school mathematics consists of 95 topics arrayed across 5 cognitive demand categories, yielding a matrix with 475 cells. While a fine grain map of the full taxonomy can be presented (see figure 4), it is more common to use the fine grain maps to display content descriptions for a specific content area.

Data Collection & Reporting

Descriptions of assessment or standards content are gathered through a process of content analysis. A panel of content specialists review the documents of focus using a systematic set of procedures to "code" each descriptive unit into the two-dimensional content language. For assessment instruments the descriptive unit is an assessment item. For standards and curriculum frameworks, the descriptive unit is determined by the organizational structure of the particular documents analyzed, but typically consist of the smallest unit used to organize and present the content standards.

Descriptions of instructional content are collected using a survey instrument designed to focus teacher reflection on instructional practice using a three-step process. The teacher first identifies from the topic list, those topics that are *not* covered in their instruction for a pre-selected 'target' class. For those remaining topics (i.e., topics taught to some extent), teachers are asked to estimate the number of class periods that students spent engaged with the topic during the time period reported on (e.g. instructional unit, semester, or school-year). The third step in the process asks the teacher to estimate the relative emphasis allotted to each of the five categories of cognitive demand for each taught topic.

In each case, the resulting descriptions of curricular content are arrayed into a matrix format, with topics organized into rows and expectations organized into columns. Each cell of the matrix receives a number representing the proportion of the overall content description represented by the cell of the matrix. The exact definition of this proportional data varies somewhat by type of description. For descriptions of instructional content, the proportion represents instructional time spent on content represented by a given cell of the matrix; for descriptions of assessment instruments, it is the proportion of assessment score points, and proportion of 'relative textual emphasis' in the case of standards and curriculum frameworks.

These descriptive results are visually presented in two formats. "Content maps" use surface area mapping to present the data as hills and valleys in a topographic map layout. "Content graphs" use a series of traditional bar charts, arrayed in a matrix layout, to present data for each content "cell". Content maps provide a powerful graphic image of the curricular content being portrayed (for one familiar with topographic maps).

Content graphs (in addition to being easier for some to read), are useful in comparing two or more content descriptions, since the bars for each description can be placed side-by-side in each cell of the matrix. The four figures below show examples of three content maps and a content graph. Note that the content graph reports the same data as presented by the three content maps.

Alignment

While some level of comparison is possible using either content maps or graphs, alignment analyses provide a more precise, mathematical procedure for calculating the degree of 'alignment' or similarity between any two descriptions employing the same descriptive language.

The procedure we use to calculate 'alignment' is based on cell by cell comparisons between each pair of corresponding cells from two sets of data matrices (representing two descriptions of content). The smaller of the proportional numbers in each of the two corresponding cells is aggregated across all the cells of the matrix. This sum becomes the alignment measure. This procedure yields the same mathematical result as a more complicated computation we have reported in the past. That method aggregates the absolute value of the difference between each of the two corresponding cells across all the cells of the matrix, divides that number by two and then subtract from one to calculate the alignment measure. Both methods yield the same mathematical result, which can be thought of as indicating the 'intersect' of the two data sets (see figure 3).

Advantages

There are several advantages to the approach to alignment described above. First, because the procedures are based on a common language that is neutral with respect to particular standards or reform initiatives, the approach can be used to evaluate alignment across state, national, and other standards. Further, this approach can be applied to instruction as well as policy tools such as standards and assessments, providing a quantitative measure of alignment that, measured longitudinally, can be used to examine the effects of reform policies on practice.

The process is also efficient. Raters can be trained in the content analysis process, and have time to code and discuss sample items, in about two hours. A typical state mathematics or science assessment instrument, containing something short of 100 items, takes less than two hours per test form to code. Multiple raters are used to increase the stability of the estimate of content, and inter-rater reliability is easily calculated. Teachers can report on a full school-year (or semester) of practice using a survey instrument that can be completed in about one hour. This makes economical large scale collection of teacher reports that provide useful indicator data in examining issues of alignment and reform policy efficacy.

Finally, the process provides a process for third-party, independent evaluation of state efforts to align policy tools, as well as examination of policy effects on practice. The following example draws together data from three separate studies, and provides a graphic example of the manner in which these tools and procedures can be utilized by educators and researchers.

Field Example

The following example draws together three separate data collection efforts that were recently conducted in one district in one state (referred to here as “State H”). Analyses of State H’s content standards for grade 7 mathematics were conducted by the American Institutes for Research as part of their NSF funded evaluation of GOALS 2000. Content analyses of State H’s mathematics assessments for grades 6-8 were conducted as part of the NSF funded Data on Enacted Curriculum (DEC) project. Teacher reports of mathematics instruction were also collected as part of the DEC project. It should be noted that the analyses presented here are illustrative. While the analyses of content standards and state assessments were conducted by subject area specialists, and used multiple raters (two raters for state standards and three raters for the mathematics assessments), reports of instruction were collected in only a handful of schools. Fourteen seventh grade mathematics teachers, across 10 schools participated in the survey. Thus the reports of teacher practice can not be taken as representative of the district. This caveat aside, the results of these three data collection efforts offer a practical example of the type of results and analyses available using the SEC tools and procedures.

Figure 2 reports results at the coarse grain level for instruction, assessment and standards. The maps and graphs reveal distinct features. Without doing further analysis, it is clear that the state assessment and state standards are more like one another than instruction is to either. Instruction is typified by broad content coverage, with topic emphasis (for this sample of teachers) on memorization of number sense, properties and relationships, and performing procedures with algebra topics. State H’s standards and assessments are similar in emphasizing procedural knowledge, though the state standards emphasize more non-routine problem solving, application, and synthesis than does the assessment.

While the coarse grain maps are revealing about key descriptive characteristics, fine grain maps provide a more detailed picture. Figure 4 reports the same data at the fine grain level. In the case of figure 4, the entire taxonomy is presented, though a more typical use would be to focus on a particular content area (e.g. data analysis, or algebraic concepts; see figure 5). At the fine grain level, the most distinguishing characteristic of both the standards and assessment maps is the amount of "white space" (no content emphasis) in each map. Standards are meant to focus instruction. By their very nature, assessments can only cover a sample of the curriculum, or they would quickly become too long and burdensome. Instructional practice, by contrast shows emphasis across the full range of topics in middle school mathematics. No doubt this is in part due to the aggregation of data across teachers, but also reflects the felt need of teachers to cover a broad curriculum.

These fine grain maps address some of the key concerns voiced about standards-based reform. One concern is that high stakes tests will cause teachers to 'narrow' the curriculum they present to students. Looking at the content profile of the state assessment as revealed by the fine grain content map, this could be a concern. If teachers did indeed decide to focus more narrowly on content tested, we would expect to see the content maps for teachers to become more narrowly focused on procedural knowledge. According to the data from our fourteen teachers, this has not happened yet. However, periodic sampling of teacher instructional practice over time would allow the district to track and correct any undesirable narrowing of the curriculum that might occur.

Ideally, standards not assessments, are the 'target' for instruction. But instruction remains broader than the standards as well. Whether this is good or not can be argued. The point for this discussion is that content maps and alignment analyses provide tools for examining and debating curricular decisions and describing the effects of policies.

While a visual review of the content maps provides a good deal of useful descriptive information, and some insight into alignment issues, alignment analyses offer a quantitative means of measuring agreement among policy tools (e.g. standards and assessments) and practice. Two tables in the bottom right of figure 4 report results of alignment analyses at the coarse grain and fine grain level. We recommend, and in practice, report only fine grain alignment measures, as these are the most descriptively accurate and analytically powerful. Indeed, fine grain alignment measures between instruction and assessment have been shown to have a strong correlation with student achievement, after controlling for prior achievement and socio-economic-status (see Porter and Smithson, 2001). Coarse grain alignment analyses results are reported here for comparative purposes only.

As both alignment tables indicate, standards and assessments in state H are more aligned to one another than instruction is to either. Clearly however, the results in coarse grain table report more alignment than the fine grain table. For the coarse grain analyses, results are aggregated up to the level of content area. Since there are six content *areas* in the middle school taxonomy, this means that the coarse grain alignment analyses are based on comparisons of data matrices with only 30 data points (6 content areas by 5 cognitive demand categories). This yields a much higher alignment number. By contrast, the fine grain analyses are based on 475 data points. Too often, when states claim to have assessments that are "aligned" to state standards, they are basing their notion of alignment on a coarse grain evaluation of content. While the specific procedures used to determine alignment vary by state, and are not identical to what we are calling "coarse grain" here, they tend to start with the standards, which inevitably leads to a coarser grain size for determining agreement than provided by the detailed taxonomic approach we advocate. ESEA requires all states to have assessments that are aligned to standards. The approach described here provides a useful tool for building and measuring alignment among these policy tools and instruction.

References

Porter, A.C., and Smithson, J.L. (2001). "Are content Standards Being Implemented in the Classroom?: A methodology and some tentative answers." In, *From the Capitol to the Classroom: Standards-Based Reform in the States.* S.H. Fuhrman (ed.) Chicago: National Society for the Study of Education.

Figure 1
Elementary School Mathematics
Content Matrix Layout

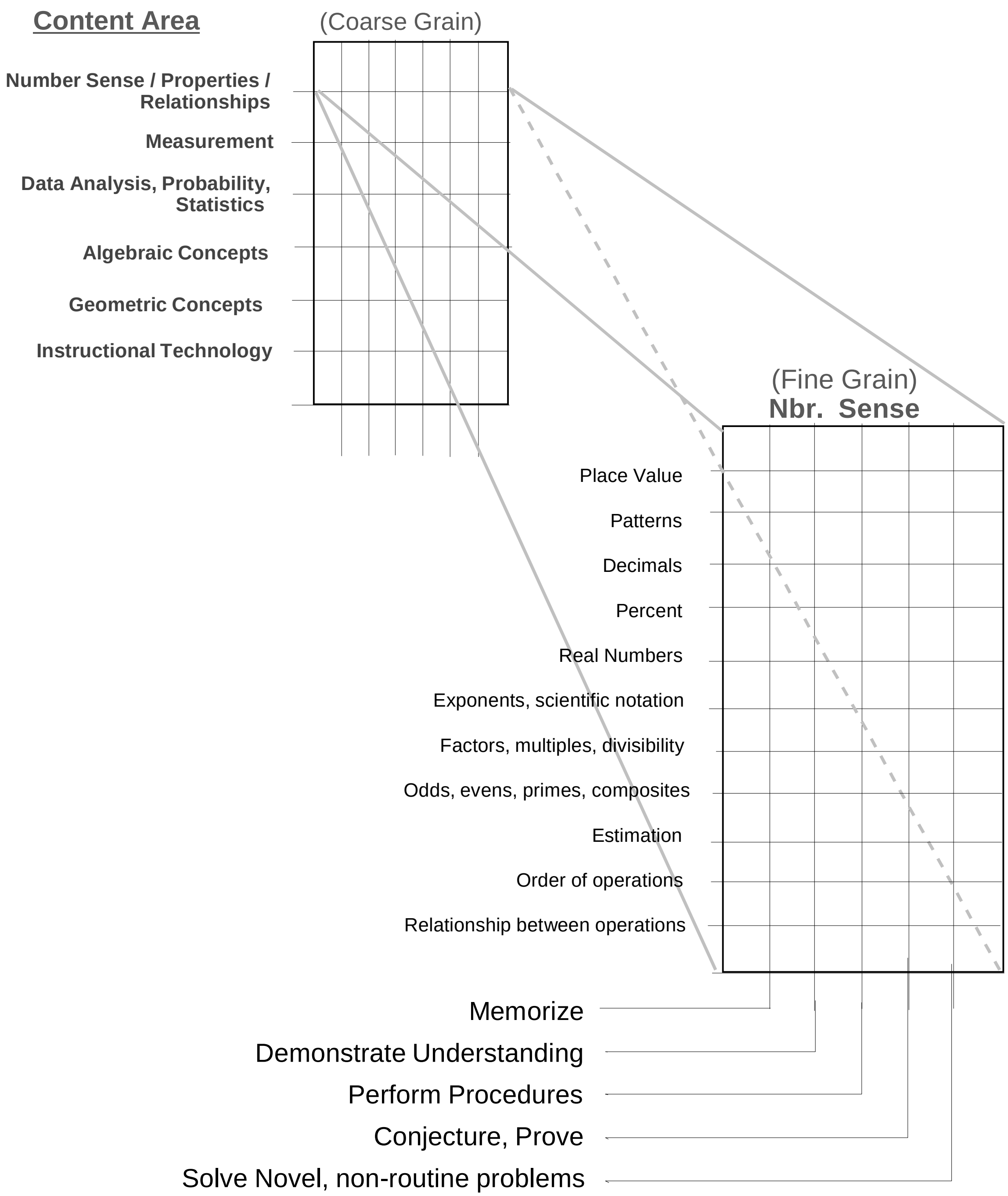


Figure 2
Content Maps & Graphs

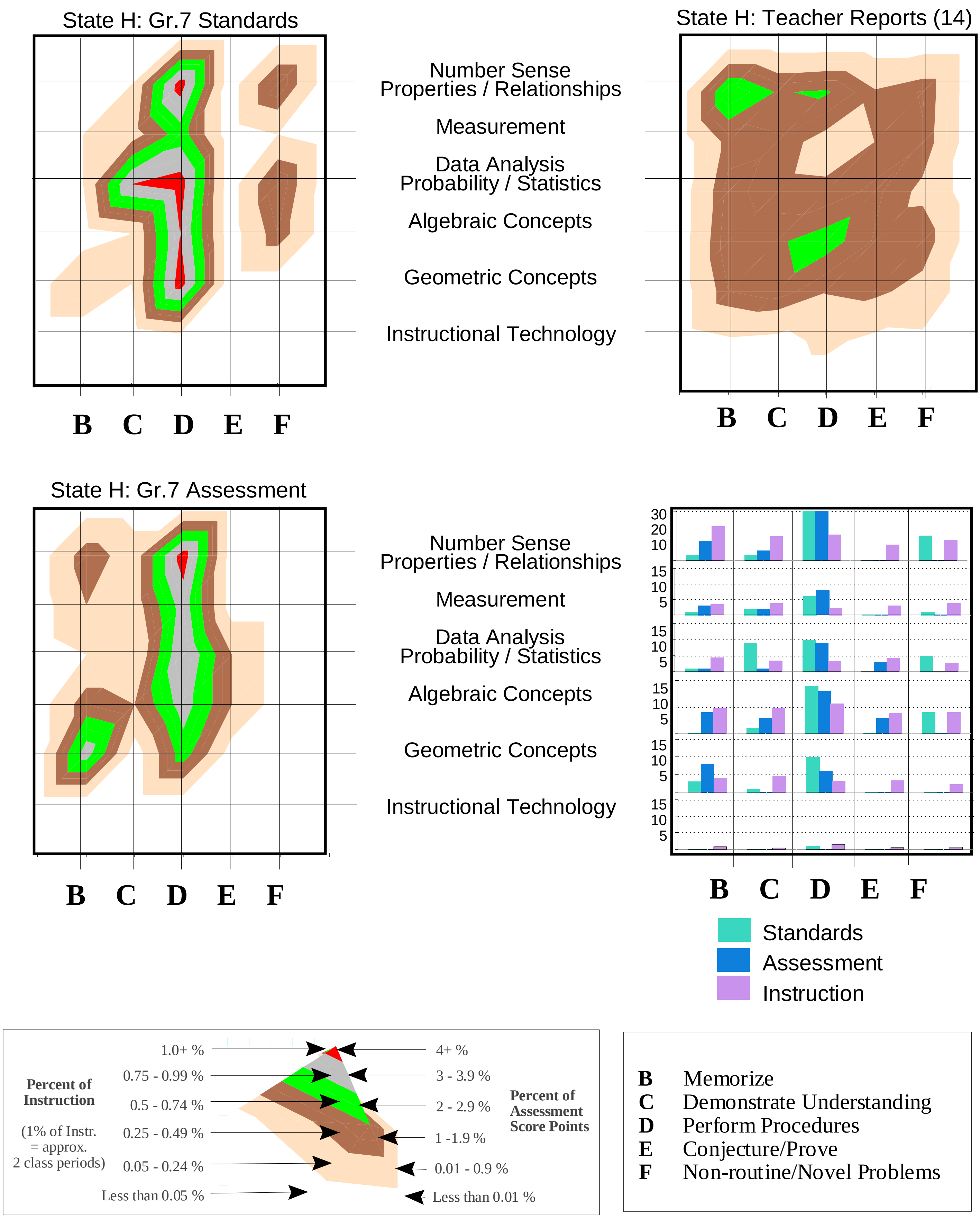


Figure 3

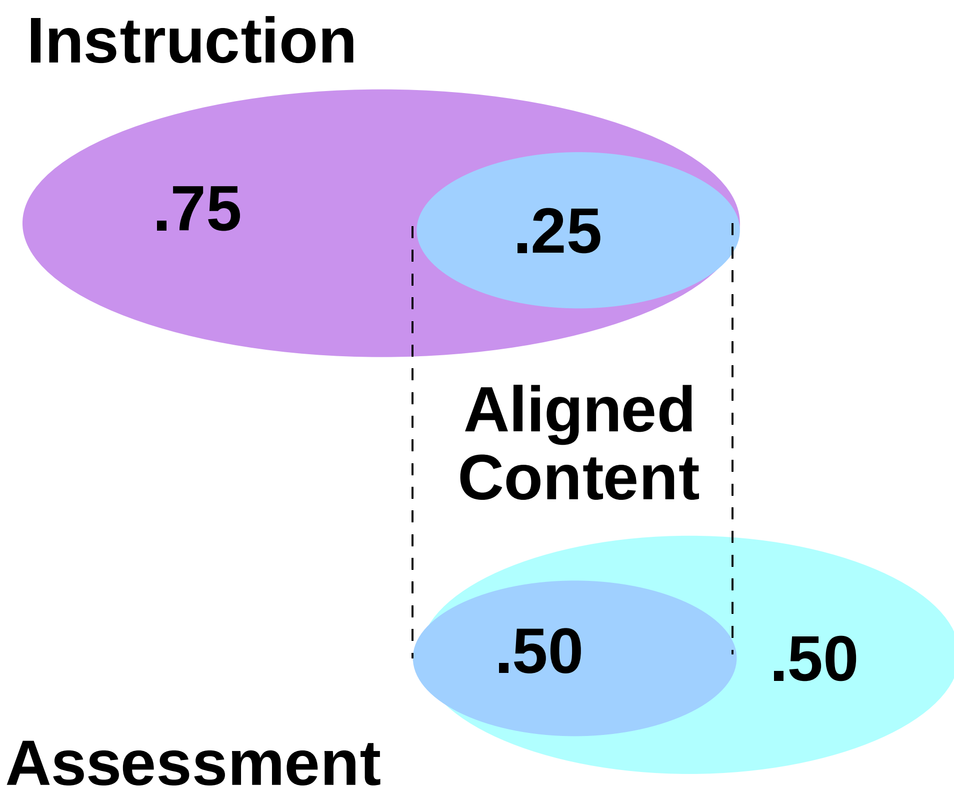
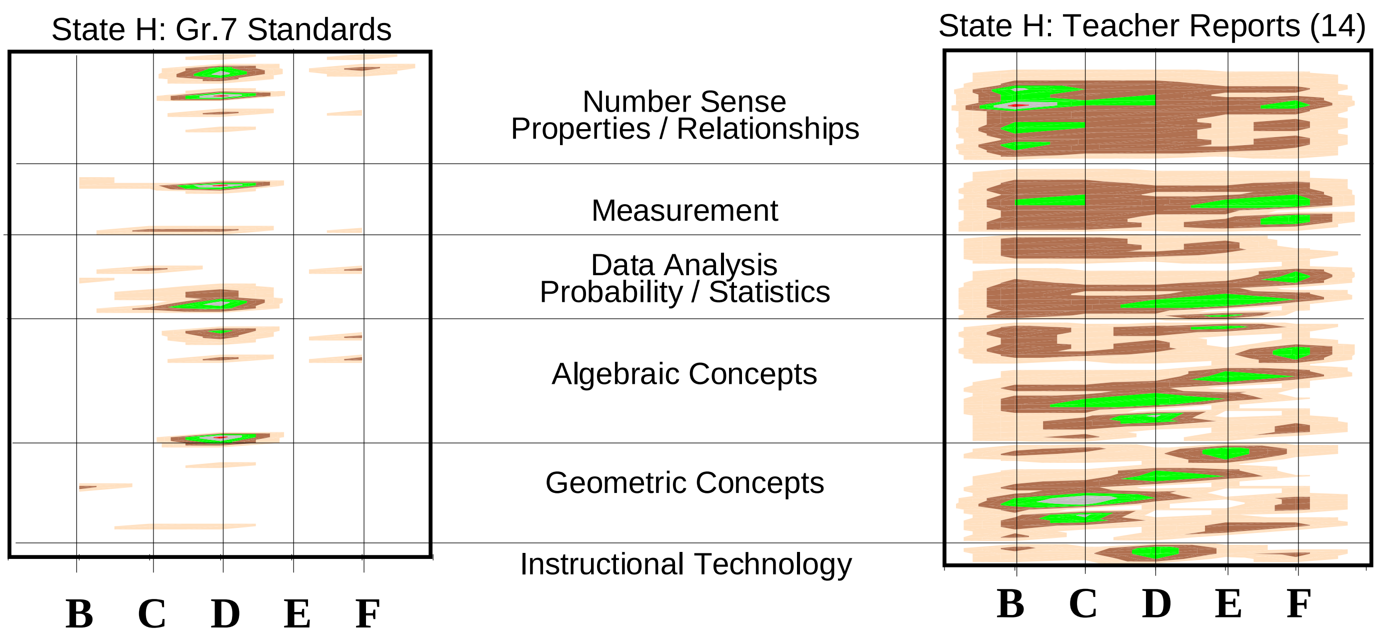
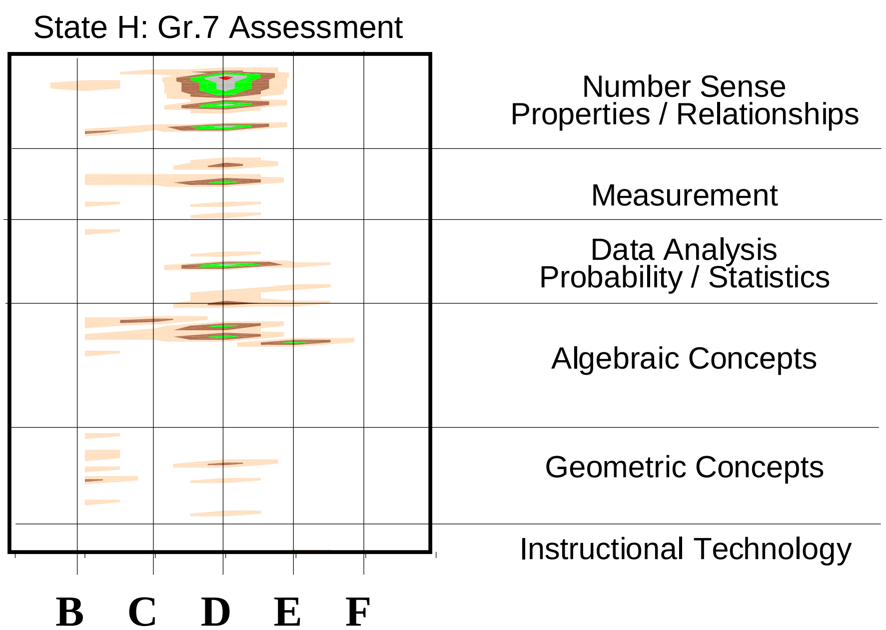


Figure 4
Fine Grain Content Maps & Alignment Analyses

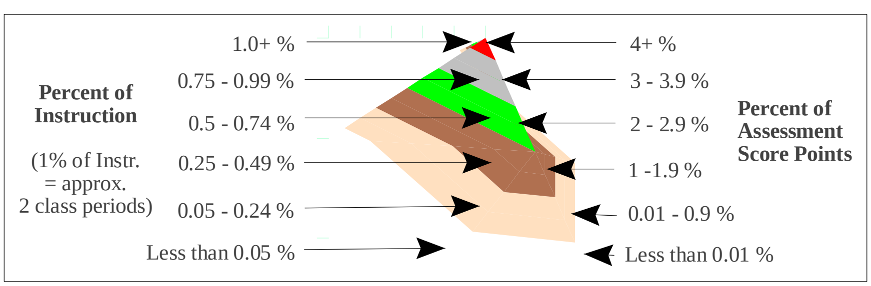




Alignment Analyses

Coarse Grain	Standards	Assessment	Instruction
Standards	**1.00**		
Assessment	**.60**	**1.00**	
Instruction	**.50**	**.49**	**1.00**

Fine Grain	Standards	Assessment	Instruction
Standards	**1.00**		
Assessment	**.29**	**1.00**	
Instruction	**.21**	**.20**	**1.00**



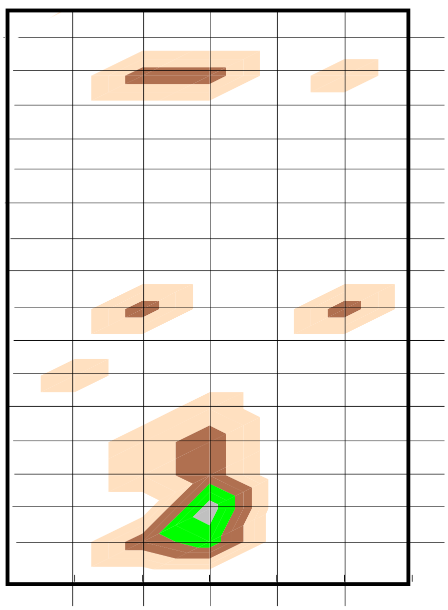
B Memorize
C Demonstrate Understanding
D Perform Procedures
E Conjecture/Prove
F Non-routine/Novel Problems

Figure 5

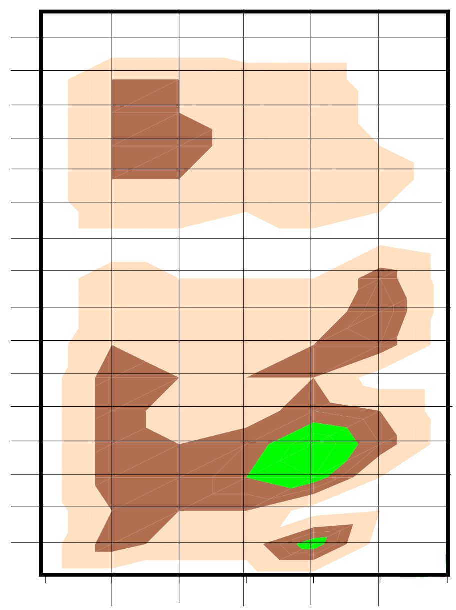
Mathematics Content Map AA3

Data Analysis, Probability, Statistics

State H: Gr.7 Standards

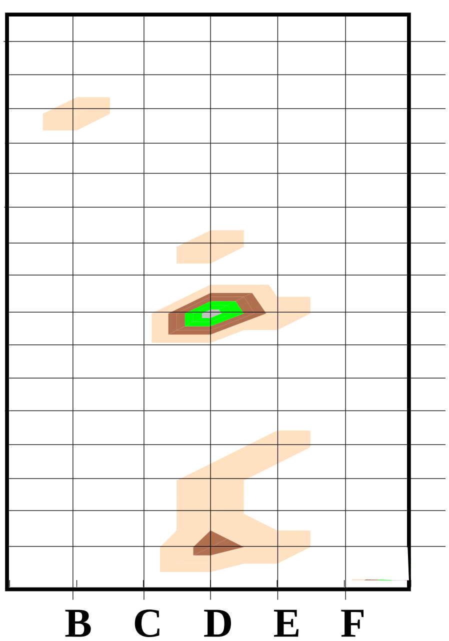


State H: Teacher Reports (14)

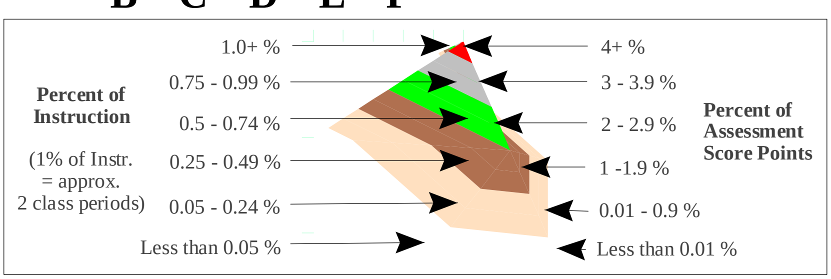


DATA ANALYSIS
Bar graph, histogram
Pie charts, circle graphs
Pictographs
Line graphs
Stem & Leaf Plots
Scatter Plots
Box Plots
Mean, median, mode
Line of best fit
Quartiles, percentiles
Sampling, sampling spaces
Simple probability
Compound probability
Combinations & permutations
Summarize data in a table or graph

State H: Gr.7 Assessment



DATA ANALYSIS
Bar graph, histogram
Pie charts, circle graphs
Pictographs
Line graphs
Stem & Leaf Plots
Scatter Plots
Box Plots
Mean, median, mode
Line of best fit
Quartiles, percentiles
Sampling, sampling spaces
Simple probability
Compound probability
Combinations & permutations
Summarize data in a table or graph



B Memorize
C Demonstrate Understanding
D Perform Procedures
E Conjecture/Prove
F Non-routine/Novel Problems