State C

Alignment Between Standards and Assessments in Science for Grades 4 and 8 and Mathematics for Grades 4 and 8

Norman L. Webb

National Institute for Science Education
University of Wisconsin-Madison

Council of Chief State School Officers
Washington, DC

April 30, 1999

This Alignment Study was supported by a grant to the Council of Chief State School Officers provided by the National Science Foundation (Award Number REC-9803080). And by the National Institute for Science Education, in cooperative agreement between the National Science Foundation and the University of Wisconsin-Madison (Cooperative Agreement No. RED-9452971). The opinions, findings, and conclusions that are expressed in this report do not necessarily reflect those of the supporting agencies.

Alignment Institute Participants

David Bahna
Science Education Assessment
South Carolina Dept. of Education

Rolf Blank
Director of Education Indicators
Council of Chief State School Officers

Jennifer Falls
Mathematics Education
Louisiana Dept. of Education

Mary Gromko
Science Education
Colorado Department of Education

Michael Kestner
Mathematics Education
North Carolina Dept. of Public Instruction

Gerald Kulm
Mathematics Education
American Association for Advancement of Science

Michael Lower
Mathematics Education Assessment
South Carolina Dept. of Education

Megan Martin
Science Assessment Consultant
California

Curtis McKnight
Dept. of Mathematics
University of Oklahoma

Andrew C. Porter
Director
Wisconsin Center for Education Research
University of Wisconsin-Madison

Harold Pratt
Science Education
National Research Council
Center for Science, Mathematics, Engineering Education

Senta A. Raizen
Director
National Center for Improving Science Education/ West Ed.

Eleanor Sanford
Assessment
North Carolina Dept. of Public Instruction

John Smithson
Research
National Institute for Science Education
University of Wisconsin-Madison

Norman Webb
National Institute for Science Education
University of Wisconsin-Madison

Linda Wilson
School of Education, Mathematics
University of Delaware

Contributors to the Report

Norman L. Webb	*Project Director*
John Smithson	*Research & Data Analysis*
Margaret H. Powell	*Editor*
Lynn Lunde	*Secretary*

Executive Summary

State C
Alignment Between Standards and Assessments in Science for Grades 4 and 8 and in Mathematics for Grades 4 and 8

At the end of June 1998, the Council of Chief State School Officers (CCSSO) in cooperation with the National Institute for Science Education (NISE) conducted an alignment institute. Seven reviewers in mathematics and six reviewers in science analyzed the alignment between standards and assessments in mathematics and science from four states. The reviewers were content area experts and included persons not associated with the participating states, as well as curriculum and assessment staff from the participating states. Using coding forms customized for each state's standards and assessment instruments, reviewers coded the objectives for a standard each assessment item/activity was related to. Reviewers could assign an assessment item/activity to more than one objective. The results of the reviewers' codings were aggregated and summary statistics were produced using Excel spread sheet software. Alignment between the standards and tests was reported on four criteria, including categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance of representation.

Grades 4 and 8 science and mathematics standards and assessments were analyzed for State C. Only draft assessments were used. For science, 14 assessment items/activities were analyzed in relation to five standards for grade 4 and grade 8. The 14 draft items were less than one-third of the 45-item assessment that will be used in 2000 for State C's first comprehensive analysis, so a complete alignment analysis could not be performed. The analysis did indicate that the 14 grade 4 items were nearly evenly distributed among the five standards. It was judged that full alignment could be reached if items of a comparable depth-of-knowledge level and similarly distributed among the standards and each additional item corresponded to a unique objective. For grade 8, too low a percentage of items had a depth-of-knowledge level at or above the objectives for Standard I, indicating one area in need of attention. The number of stated objectives (benchmarks), 86, is large in relation to the less than 45 items to be included on the assessment. This means that the range-of-knowledge correspondence and balance-of-representation criteria will be difficult to attain unless robust assessment items are used that measure knowledge on more than one objective.

In mathematics, the grade 4 assessment of 74 items/activities and six standards were in alignment on three criteria—categorical concurrence, depth-of-knowledge consistency, and balance of representation. The range-of-knowledge correspondence criterion was met by only two of the six standards (Standards I and IV). The grade 8 assessment of 68 items/activities and standards were only partially aligned. Three of the six standards (III, IV, and VI) met three of four alignment criteria—categorical concurrence, depth-of-knowledge consistency, and balance of representation. None of the six standards were fully aligned with the assessment on the range-of-knowledge correspondence. Each of three standards failed to fully achieve alignment on one of the other three criteria. Standard II (Algebra) had too few of items to fully meet the categorical

concurrence criterion. Standard V (Data Analysis, Probability, and Discrete Math) did not have a sufficient percentage of items at a level of knowledge in relationship to the related objectives and with respect to the reviewer variation to be judged aligned in depth-of-knowledge consistency. Standard I (Number and Number Relations) failed to meet the balance-of-representation criteria.

Table of Contents

State C

Alignment Between Standards and Assessments in Science for Grades 4 and 8 and in Mathematics for Grades 4 and 8

Norman L. Webb
National Institute for Science Education

Introduction

Alignment of expectations for student learning and assessments for measuring students' attainment of these expectations is an essential attribute for an effective standards-based education system. Alignment is defined as the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide the system toward students learning what they are expected to know and do. As such, alignment is a quality of the relationship between expectations and assessments and not an attribute of any one of these two system components. Unlike validity and reliability, which are specific qualities of an assessment instrument, alignment describes the match between expectations and assessment that can be legitimately improved by changing either student expectations or assessments. As a relationship between two or more system components, alignment needs to be determined by using the multiple criteria described in detail in a National Institute for Science Education (NISE) research monograph, *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education* (Webb, 1997).

A four-day Alignment Analysis Institute was conducted June 29 through July 2, 1998. Sixteen people, including state content specialists, state assessment consultants, content experts, and researchers, attended the institute, which was coordinated by the Council of Chief State School Officers (CCSSO) with cooperation of the National Institute for Science Education (NISE). Prior to this institute, most participants attended a one-day meeting in Washington, DC, on April 29, to be introduced to the process and to the alignment criteria to be used at the institute. At the summer institute, six of the participants rated mathematics standards and assessments; seven rated science standards and assessments; and three coordinated the process. Four states volunteered to have their mathematics standards and assessments analyzed for alignment for two grade levels. Three of these states agreed to have their science standards and assessments analyzed for two or three grade levels.

A major goal of the institute was to develop a systematic process and analytic tools for judging the alignment between standards and assessments based on the criteria developed in conjunction with CCSSO and NISE (Webb, 1997) that are listed in Appendix A. Because of this, reviewers were not given lengthy training in applying the criteria, but were expected to help refine the process over the duration of the institute. One outcome of the institute is a refined process that can be used under more controlled conditions to make a judgment on the alignment

of standards and assessments. Reviewers were instructed to attend to the alignment between the state standards and assessments. There was no opportunity for reviewers to offer their opinion on either the quality of the standards or of the assessment activities/items. The results produced from the institute pertain only to how the state standards and the state assessment are in agreement and do not serve as external verification of the general quality of a state's standards or assessments. The results of the Alignment Analysis Institute do provide judgments of content area experts, independent of any of the participating states, who are very familiar with state and national standards. The means of the reviewers' coding were used to determine whether the alignment criteria were met. When reviewers did vary in their judgments, the means lessened the error that might result from any one reviewer. The standard deviations are reported, which give one indication of the variance among reviewers.

This report describes the results of an alignment study of standards and grade level tests in science and mathematics for one state, identified here as State C. The study addressed specific criteria related to consideration of the content agreement between the state standards and grade level assessments. Four criteria received major attention: categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance of representation. Other criteria such as articulation across grades and ages, equity and fairness, and pedagogical implications were given less emphasis.

Initial Methodology Developed at the Institute for the Analysis of Alignment Criteria

From three to five reviewers analyzed the alignment between the standards and the assessment. Prior to analyzing the documents, the reviewers were only given some very general instructions and broad definitions for the levels as a basis for rating the depth of knowledge required to satisfy a standard and to successfully respond to assessment activity. One purpose for conducting these alignment studies is to enable us to better specify the level of training needed by reviewers if they are to reliably code assessment activities and standards. Reviewers were given the following levels to judge depth of knowledge:

Level

1. *Recall*
   Recall of a fact, information, or procedure

2. *Skill/Concept*
   Use of information, conceptual knowledge, procedures, two or more steps, etc.

3. *Strategic Thinking*
   Requires reasoning, developing a plan or a sequence of steps; offers some complexity; more than one possible answer; generally takes less than 10 minutes to complete.

4. *Extended Thinking*
   Requires an investigation, time to think and process multiple conditions of the problem or task; and requires more than 10 minutes to complete other-than-routine manipulations.

Reviewers within a content area were encouraged to refine these levels or to add greater clarification, providing they all came to some agreement. One of the outcomes of this alignment study will be the capacity to specify the levels with greater clarity. The revised levels are provided in the summary report of the study (Webb, 1999).

Different states use different terminology to label expectations for what students are to know and do. Some states label the large categories of student expectations as "strands." Other states call these expectations "competency goals." Still others refer to state expectations as "benchmarks." To improve the interpretation of results, the same convention was used in this analysis to label the different levels of expectations. Standards refers to the most general expectations for a grade and content area. The number of standards in the four states that participated in this analysis ranged from four to ten. Goal refers to the next level of specificity of expectations. Generally, the set of goals for a standard covers the full range of knowledge specified by the standard. The number of goals for a standard in this analysis went as high as 20. Objective refers to the third level of specificity. Objectives further delineate expectations stated as a goal. The number of expectation levels can vary. In this analysis, a maximum of three levels of expectations was included. If a state only used two levels of expectations, then the most general level is called standards and the second level is called objectives.

Prior to the Alignment Analysis Institute, reviewers were sent copies of the standards and were asked to become familiar with them. At the institute, reviewers began their analysis by assigning a depth-of-knowledge level for each standard. Achieving one objective could require students to know the content at more than one depth-of-knowledge level.1. The assigned level was to represent the highest level of knowledge expected for a student to satisfactorily demonstrate attainment of the standard. All of the reviewers were to reach consensus on the assigned level for each objective (third level of expectations). This activity served two purposes. First, reviewers became more familiar with what students were expected to know and do to meet each standard. Second, the assigned levels were used as benchmarks to compare the depth-of-knowledge level of individual assessment items/activities.

Reviewers recorded the depth of knowledge for each objective on a coding matrix prepared prior to the institute. The coding matrix listed all of the objectives for student learning in the set of standards for the grade level, or range, and content area (science or mathematics). These expectations were listed in rows in the same order and using the same organization as that used in State C's standards document. For one standard, the first row specified the standard, the second row a goal, and the third row objectives, and so on. The same was done for the second

1 Objective as used in this analysis should not be confused with a behavioral objective designed to express one specific behavior and one depth-of-knowledge level.

through the final standards. (See the example below.) A numerical-alpha coding system was used to give each standard, goal, and objective a separate code. The depth-of-knowledge level for each objective was recorded in a designated column. For example, the grade 8 standard and benchmark from above were given these depth-of–knowledge codes:

		Description	Depth-of-Knowledge Level
IV.	Strand:	Earth and Space Science	
	Standard:	The students will develop an understanding of the properties of earth materials, the structure of the Earth system, the Earth's history, and the Earth's place in the universe. As students in Grades 5-8 extend their knowledge, what they know and are able to do includes:	
IV. B	Benchmarks:	B. Earth History	
IV. B.1	ESS-M-B1	Investigating how fossils show the development of life over time;	2
IV. B.2	ESS-M-B2	Devising a model that demonstrates supporting evidence that the Earth has existed for a vast period of time;	3
IV. B.3	ESS-M-B3	Understanding that earth processes such as erosion and weathering affect the Earth today and are similar to those which occurred in the past.	2

One column was included on the coding matrix for each assessment item/activity. Reviewers, after reaching consensus on the depth-of-knowledge code for each objective, individually judged and wrote in the corresponding objective-item cell the depth-of-knowledge code for the assessment item/activity. Reviewers assigned the assessment item/activity one level of depth of knowledge and noted this level in the item/activity's column for each objective, if a student's response to the item/activity provided information about what the student knew or could do with respect to the objective. Each objective coded for an item was called a hit. Multiple hits were allowed for any one assessment item/activity. Initially, reviewers were not given any specifications about limits on the number of hits for any one assessment activity/item. After discussion with other reviewers following the coding of each test, reviewers developed guidelines for multiple hits. As the reviewers gained more experience, this lowered the instances of reviewers marking multiple hits for a single item/activity. The number of multiple hits was one source of variation among reviewers. Reviewers did converge in their application of multiple hits as they became more familiar with the process and developed agreed-upon rules.

Reviewers were asked to code the assessment items/activities independently for each test, with little or no interaction. After all of the reviewers completed coding the instruments, they were asked to select a sample of items and compare their results. The primary purpose of this discussion was to improve the reliability among the reviewers in coding assessment

items/activities on the next and subsequent instruments. Reviewers could make changes as they calibrated their work with the other reviewers, if they felt it was appropriate. Reviewers discussed both what items/activities were assigned to what objectives and the depth-of-knowledge code assigned to each item.

The codings for all of the reviewers were entered onto a spreadsheet to compute summary statistics. For each assessment instrument and standards document, the codes for each review were tabulated by frequency of hits and depth-of-knowledge level for each objective. Data for all of the objectives for one standard were aggregated to provide frequencies for each standard. The results were then reported by individual standard.

Statistics for each standard were computed for four alignment criteria for content focus: categorical concurrence, depth-of-knowledge consistency, range of knowledge, and balance of representation. The mean number of hits was used to judge the categorical concurrence between the assessment instrument and the standards. The frequencies of hits aggregated across the objectives for each standard and by depth-of-knowledge levels were used to compute the percentage of hits that were below, at the same level as, or above the level for the objective. The percentage of the objectives hit within a standard was used to judge the range-of-knowledge correspondence within the standard. The distribution of hits among the objectives for a standard with at least one hit was used to compute the balance of representation for a standard. This analysis is based on the assumption that the set of objectives for a standard spans the entire domain of knowledge and skills a student should demonstrate to fully meet the standard.

Reviewers were asked for their comments on other alignment criteria, including articulation across grades, pedagogical implications, and equity. Some offered their comments, but, because of severe time pressures, systematic procedures were not used to gather information on these criteria. Reviewers reacted to the overall process and made suggestions in a debriefing session held at the end of the institute. Reviewer reactions are incorporated into the analysis of the summary report of this study, *Alignment of Science and Mathematics Standards and Assessments in Four States* (Webb, 1999).

All of the statistics were computed for each reviewer. The mean for each statistic was computed, using the results for only the reviewers who completed coding all of the items—at least two reviewers, and up to six for some tests. It is important to note that only two reviewers coded some instruments. The mean among reviewers on each statistic is a reasonable approximation for the summary information, which lessens the error of any one reviewer in coding. Of course, statistics based on coding by a greater number of reviewers will be more accurate. Standard deviations, reported along with the mean, provide one indication of the variation among reviewers. Low standard deviations indicate that there was minimal variation among reviewers on the marginal statistics for a standard. Relatively high standard deviations indicate that the reviewers disagreed more on the marginal statistics for a standard. The total number of objectives and the total number of hits for a standard also have to be considered in judging the agreement among reviewers.

Alignment Criteria Used for This Analysis

This analysis judged the alignment between the standards and the assessments using four criteria. For each criterion, an acceptable level was defined based on what would be required to assure that students have attained the standards.

Categorical Concurrence

One aspect of alignment between standards and assessments is if both address the same content categories. The categorical concurrence criterion provides a very general indication if both documents incorporate the same content. *The criterion of categorical concurrence between standards and assessment is met if the same or consistent categories of content appear in both documents.* This criterion was judged by determining whether the assessment included items measuring content from each standard. The analysis assumed that the assessment had to have at least six items measuring content from a standard in order for there to be an acceptable categorical concurrence between the standard and the assessment. The number of items, six, is based on estimating the number of items that could produce a reasonably reliable sub-scale for estimating students' mastery of content on that sub-scale. Of course, many factors have to be considered in determining what a reasonable number is, including the reliability of the sub-scale, the mean score, and cutoff score for determining mastery. Using a procedure developed by Subkoviak (1988) and assuming the cutoff score is the mean and the reliability of one item is .1, it was estimated that six items would produce an agreement coefficient of at least .63. This indicates that about 63% of the group would be consistently classified as masters or non-masters if two equivalent test administrations were employed. The agreement coefficient would increase if the cutoff score is increased to one standard deviation from the mean to .77 and, with a cutoff score of 1.5 standard deviations from the mean, to .88. None of the four states included in the analysis reported student results by standards or required students to achieve a specified cutoff score on sub-scales related to a standard. If a state did do this, then the state would want a higher agreement coefficient than .63. Six items were assumed as a minimum for an assessment measuring content knowledge related to a standard, and as a basis for making some decisions about students' knowledge of that standard. If the mean for six items is 3 and one standard deviation is one item, then a cutoff score set at 4 would produce an agreement coefficient of .77. Any fewer items with a mean of one-half of the items would require a cutoff that would only allow a student to miss one item. This would be a very stringent requirement, considering a reasonable standard error of measurement, on the sub-scale.

Depth-of-Knowledge Consistency

Standards and assessments can be aligned not only on the category of content covered by each, but also on the complexity of knowledge required by each. *Depth-of-knowledge consistency between standards and assessment indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards.* For consistency to exist between the assessment and the standard, as judged in this analysis, at least 50% of the items corresponding to an objective had to be at or above the

level of knowledge of the objective. The level of 50% is a conservative level and is based on the assumption that a minimal passing score for any one standard of 60% or higher would require the student to successfully answer at least some items at or above the depth-of-knowledge level of the corresponding objectives.

Range-of-Knowledge Correspondence

For standards and assessments to be aligned, the breadth of knowledge on both should be comparable. *The range-of-knowledge criterion is used to judge whether a comparable span of knowledge expected of students by a standard is the same as, or corresponds to, the span of knowledge that students need in order to correctly answer the assessment items/activities.* The criterion for correspondence between span of knowledge for a standard and the assessment considers the number of objectives within the standard with one related assessment item/activity. At least 50% of the objectives for a standard had to have at least one related assessment item in order for the alignment on this criterion to be judged acceptable. This level is based on the assumption that students' knowledge should be tested on content from over half of the domain of knowledge for a standard. This assumes that each objective for a standard should be given equal weight. Depending on the balance in the distribution of items and the need to have a low number of items related to any one objective, the requirement that assessment items need to be related to more than 50% of the objectives for a standard increases the likelihood that students will have to demonstrate knowledge on more than one objective per standard to achieve a minimal passing score. As with the other criteria, a state may choose to make the acceptable level on this criterion more rigorous by requiring an assessment to include items related to a greater number of the objectives. However, any restriction on the number of items included on the test will place an upper limit on the number of objectives that can be assessed. Range-of-knowledge correspondence is more difficult to attain if the content expectations are partitioned among more standards and a large number of objectives.

Balance of Representation

In addition to comparable depth and breadth of knowledge, aligned standards and assessments require the knowledge to be distributed equally in both. The range-of-knowledge criterion only considers the number of objectives within a standard hit (a standard with a corresponding item), but does not take into consideration how the hits (or assessment items/activities) were distributed among these objectives. *The balance-of-representation criterion is used to indicate the degree to which one objective is given more emphasis on the assessment than another. An index is used to judge the distribution of assessment items*. This index only considers the objectives for a standard that have at least one hit—i.e., one related assessment item/objective. The index is computed by considering the difference in the proportion of objectives and the proportion of hits assigned to the objective. An index value of 1 signifies perfect balance and is obtained if the hits (items/assessment) related to a standard are equally distributed among the objectives for the given standard. Index values that approach 0 signify that a large proportion of the hits (items/assessment) were on only one or two of all of the objectives hit. Depending on the number of objectives and the number of hits, a unimodal distribution (most

items related to one objective and only one item related to each of the remaining objectives) has an index value of less than .5. A bimodal distribution has an index value of around .55 or .6. Index values of .7 or higher indicate items/activities are distributed among all of the objectives at least to some degree (e.g., every objective has at least two items) and is used as the acceptable level on this criterion.

State C: Alignment of Standards and Assessments in Science for Grades 4 and 8

Organization of Standards for Science

State C's Board of Elementary and Secondary Education adopted its Science Framework based on the National Science Education Standards (NRC, 1995). The Framework provides a master plan for developing scientific literacy; meets the specific needs of the state; serves as a catalyst for discussion of the nature of science education; and outlines a vision of science instruction and assessment reform for local districts. Among its intended uses, the Framework is envisioned as a guide for teachers and curriculum developers to use for planning curriculum, instruction, and assessment. Another intended use is for it to serve as a framework to guide assessment specialists and test developers in creating assessment frameworks that will measure students'scientific understanding and ability more effectively. An accompanying document, Educational Assessment Program Teacher's Guide to Statewide Assessment in Science (TGSAS), introduces a plan for developing a state assessment system; outlines the structure of standards-based assessment; and provides illustrative examples of grade-level benchmarks, performance expectations, and assessment items. One purpose of the Guide is to provide testing contractors with a tool for working with local educators and Department of Education staff to develop State C's criterion-referenced science assessments. Only draft assessment items were used in this alignment study.

State C has articulated five foundation skills that apply to all students in all disciplines: Communication; problem solving; resource access and utilization; linking and generating knowledge; and citizenship. The Science Framework is divided into five strands clustered in three grade ranges—K-4, 5-8, and 9-12—that identify what students are expected to know and do. These same strands are used in the teacher's assessment guide (TGSAS). The five strands, each the same for all three grade-level clusters, are further specified by benchmarks. The general headings of the benchmarks varied by grade-range clusters and they varied in number from 2 to 12. The benchmarks represent what students should be able to do with their scientific knowledge and skills within each strand. The five strands and sample of the general benchmarks for the K-4 grade level cluster are:

Science as Inquiry

The students will do science by engaging in partial and full inquiries that are within their development capabilities.

Benchmarks:

The Abilities Necessary to do Scientific Inquiry (All three grade level clusters)

Understanding Scientific Inquiry (All three grade level clusters)

Physical Science

Students will develop an understanding of the characteristics and interrelationships of matter and energy in the physical world.

Benchmarks:

Properties of Objects and Materials

Position and Motion of Objects

Forms of Energy

Life Science

The students will become aware of the characteristics and life cycles of organisms and understand their relationships to each other and to their environment.

Benchmarks:

Characteristics of Organisms

Life Cycles of Organisms

Organisms and their Environments

Earth and Space Science

The students will develop an understanding of the properties of earth materials, the structure of the Earth system, the Earth's history, and the Earth's place in the universe.

Benchmarks:

Properties of Earth Materials

Objects in the Sky

Science and the Environment

In learning environmental science, students will develop an appreciation of the natural environment, learn the importance of environmental quality, and acquire a sense of stewardship. As consumers and citizens, they will be able to recognize how our personal, professional, and political actions affect the natural world.

Benchmarks

(All benchmarks are grouped under the standard. For this standard, no individual benchmarks were identified.)

For this alignment analysis, the five science strands and overarching standards were considered the standards. The general headings for the grouping of benchmarks were considered the goals. The specific benchmarks were considered the objectives. In State C, three levels of expectations were incorporated in the analysis:

Label for Analysis	Label by State C
Standard	Strand/Standard
Goal	Benchmarks-General
Objective	Benchmarks-Specific

For example, the following levels of expectations are stated for grades 5-8 science under Earth and Space Science Strand and for the Earth History benchmark grouping:

Strand:	Earth and Space Science
Standard:	The students will develop an understanding of the properties of earth materials, the structure of the Earth system, the Earth's history, and the Earth's place in the universe. As students in Grades 5-8 extend their knowledge, what they know and are able to do includes:
Benchmarks:	B. Earth History
ESS-M-B1	investigating how fossils show the development of life over time;
ESS-M-B2	devising a model that demonstrates supporting evidence that the Earth has existed for a vast period of time;
ESS-M-B3	understanding that earth processes such as erosion and weathering affect the Earth today and are similar to those which occurred in the past.

Assessments for Science

Analyses were performed for two grades in science, grades 4 and 8. Only draft assessment items were analyzed. A draft of the State C assessment program's Teachers' Guide for grades 4, 8, 10 was sent to school district at the beginning of the 1997-98 school year. The assessment guide uses the term "dimensions" to represent parameters within the strand and to state what students should be able to do with their scientific knowledge and skills. The dimensions, in part, describe the cognitive demands. The Science as Inquiry strand has two distinct dimensions:

Questioning, Planning, Doing, and Recording
Interpreting and Communicating.

The other four strands each have three dimensions:

Understanding Essential Content and Concepts
Explaining, Reflecting, and Connecting
Applying and Using Knowledge and Technology.

The guide references the benchmarks under each standard that are related to each dimension. In the appendices, the guide describes the blueprint for a 100-115-minute state assessment in science. The test is to be divided into one 50-minute block and one 50-65-minute block. The complete assessment is planned to consist of 30 multiple-choice questions, seven short answers questions, and one comprehensive scientific task with approximately seven short answer and one long essay embedded within the task. Each question should be focused on only one strand and dimension. The total number of 45 items and parts should be nearly equally distributed among the five strands.

Only 14 draft items were used in this analysis for both grade 4 and grade 8 science. This is less than half of the number that would be included on the actual tests, too few to reach any conclusion about alignment between the tests and the standards. In this report, the alignment

between the 14 draft items and the standards are discussed for each of four criteria in relation to the additional items and characteristics that would be needed to achieve full alignment.

Alignment of Standards and Assessment for Grade 4 Science

Categorical Concurrence

The criterion of categorical concurrence between standards and assessment is met if the same or consistent categories of content appear in both documents. Five reviewers coded the 14 items in relationship to the five State C standards for grade 4 science (Table CS4-1). The 14 items were nearly evenly distributed among the five standards. In order for a test and standard to have categorical concurrence, the test needs to include at least six items for that standard. The sample of items analyzed has the potential of achieving full categorical concurrence on all five standards if each standard had six corresponding test item or activities. This could be achieved by adding to the set of items four items related to Standard I (Science as Inquiry) and Standard II (Physical Science); three items related to Standard III (Life Science) and Standard V (Science and the Environment); and two items related to Standard IV (Earth & Space Science). Nearly all of the items were judged to relate to only one of the standards, not requiring students to draw upon ideas related to two or more standards. If this pattern continued, then the remaining items would have to be sorted precisely as these. There was strong agreement among the five raters in the number of items related to each standard. The five relaters had complete agreement on the number of items related to two of the standards (Standards III and V) and only varied by one item on the number of items related to the other four standards.

Depth-of-Knowledge Consistency

Depth-of-knowledge consistency between standards and assessment are aligned if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards. For there to be consistency between the assessment and the standard, as judged in this analysis, at least 50% of the items corresponding to an objective had to be at or above the level of knowledge of the objective.

In State C's grade 4 science assessment, 14 draft items varied in the depth-of-knowledge level in relationship to the level of the objectives (benchmarks) (Table CS4-2). Raters, on the average, indicated that on four of the five standards had more that half of the items at the level of knowledge or higher than the level of the related objective. Only Standard IV (Earth & Space Science) had, on the average, more than half of the three or four items reviewers coded as related to this standard at a level below the level of the related objective. This indicates that the set of draft items reflected adequately the depth-of-knowledge level of the corresponding objectives. It should be noted that nearly three-quarters of the 60 objectives were rated to be at Level 2 (Skill/Concept). Only 20% were rated to be at Level 3 (Strategic Thinking). For items to adequately reflect the objectives would require most of the items to measure skills and conceptual knowledge. If the remaining items were comparable to the set of existing 14 items on

the depth-of- knowledge level, then the test would be considered to have achieved depth-of-knowledge consistency with the standards.

Reviewers had high agreement in judging the relationship between the depth-of-knowledge level for the item corresponding to the objective for Standard IV. The variation among raters, represented by standard deviations of 9%, was only due to three reviewers assigning four items as related to this standard and two reviewers assigning two items. Reviewers varied the greatest on assigning the depth-of-knowledge level to items related to Standard I (Science as Inquiry). On the other three standards, reviewers were consistent in indicating the items were at or under the level of knowledge of the corresponding objectives for Standards II (Physical Science) and V (Science and the Environment) and at or above the level of knowledge of the corresponding objectives for Standard III (Life Sciences). Of course, the number of items is too small to get any stable measures of inter-reviewer agreement. The agreement among reviewers could change considerably with the rating of more items.

Range-of-Knowledge Correspondence

The range-of-knowledge criterion is used to judge whether a comparable span of knowledge expected of students for a standard is the same as, or corresponds to, the span of knowledge that students need in order to do the assessment items/activities correctly. The criterion for correspondence between span of knowledge for a standard and the assessment is the number of objectives within the standard that has at least one related assessment item/activity. Only Standard V (Science and the Environment) had a sufficient number of related items from the 14 items that corresponded to more than half of the standard's five objectives (Table CS4-3). Range-of-knowledge correspondence was met for Standard V with such a low number of items analyzed because it had a small number of objectives, five compared to 13 to 16 for the other four standards. The number of objectives hit is nearly identical to the same number of items related to each standard. This indicates that each item a reviewer felt was related to a standard corresponded to a unique objective (benchmarks). Each of the 14 items was coded as related to a different objective (benchmark) within the standards. If this pattern continues, then there is a strong likelihood that the test and standards will have range-of-knowledge correspondence. One problem that State C faces, however, in achieving range-of knowledge correspondence is distributing 30 assessment items among 60 objectives. This implies that every item would have to address a distinct objective and that the items would have to be evenly distributed among the five standards in order to have at least half of the objectives with related items. Reviewers' agreement in the number of objectives with corresponding items was high, the same as the agreement among reviewers in the number of standards with corresponding items.

Balance of Representation

The balance-of-representation criterion is used to indicate the degree to which one objective is given more emphasis on the assessment than another. An index is used to judge the distribution of assessment items. This index only considers the objectives for a standard that has at least one hit—e.g., one related assessment item/objective. Because the analysis only included

14 items, judging the balance of representation does not have additional meaning beyond the range-of-knowledge correspondence. All of the reviewers coded items related to distinct objectives, except for two instances. Two raters assigned two items related to one objective for Standard III (Life Sciences). This accounts for the mean balance index not being equal to 1.0 for Standard III (Table CS4-3). Again, because of the high number of objectives and the target of a 30-item test, the balance criterion will automatically be met if the range criterion is achieved.

Summary for Grade 4 Science Alignment

Only 14 draft assessment items, less than 50% of the total number of items on the grade 4 science test, were used in this analysis. This is too few items to permit a decision on whether the four alignment criteria were met or not met. The items analyzed were nearly evenly distributed among the five standards; in reasonable proportion to depth-of-knowledge levels related to the corresponding objectives; and each associated with a distinct objective. Full alignment would be reached by the full test of 45 items, if the remaining items continued the pattern set by the 14 items—i.e., proportionally distributed among the standards, and each only corresponding to one objective.

Alignment of Standards and Assessment for Grade 8 Science

Categorical Concurrence

Categorical concurrence between standards and assessment requires the same or consistent categories of content to appear in both documents. Five reviewers coded the 14 draft items for grade 8 science and compared these to the standards for State C. The total number of average hits by reviewers, 16.20, indicates that two or three of the 14 items were classified as related to more than one objective (Table CS8-1). One of the five reviewers coded 20 hits that were noticeably higher than the 14 to 17 hits of the other four reviewers. This reviewer appears to have double-coded five or six of the items as related to Standard I (Science as Inquiry), along with another standard. This is most likely related to viewing Standard I more as a process standard that can interact with the other more content-oriented standards. Most of the items were judged to measure knowledge related to one objective (benchmark). The 14 items were found to be nearly evenly distributed among the five standards. If the remaining items followed the same pattern then the assessment and standards would have categorical concurrence. Except for Standard I and the coding by one reviewer, reviewers had strong agreement in the number of items they coded as related to each standard. Only one reviewer differed by one number from the other four reviewers in assigning items as related to Standards III, IV, and V. The standard deviation of .45 was the lowest possible standard deviation other than zero for perfect agreement. Reviewers ranged in the number of items coded as related to Standard II (Physical Science) from three to five. For Standard I (Science as Inquiry), four of the reviewers had close agreement and ranged in the number of items from one to four. As has already been noted, one reviewer varied considerably from the others and coded seven items as related to Standard I. This is the reason for the very large standard deviation for this standard, 2.39.

Depth-of-Knowledge Consistency

Depth-of-knowledge consistency between standards and assessment indicates alignment, if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards. This analysis used as a target for an acceptable level that at least half of the items should be at or above the depth-of-knowledge level of the corresponding objective. As in the case of grade 4, a high percentage of the objectives (benchmarks), 77%, were coded as having a depth-of-knowledge Level of 2 (Skill/Concept) (Table CS8-2). Only 16% of the objectives were judged to be Level 3 (Strategic Thinking). A high percentage of the 14 draft items, 65%, were coded, on the average, to be at the same level as the related objective. Except for Standard I, 80% or more of the items were coded at a level of knowledge at or above the level of the related objective. For Standard I, the standard with the highest percentage of level 3 (Strategic Thinking) objectives, 53%, or more than half of the related items, was judged to be under the level of the corresponding objective. Projecting to a test with the 30 items, depth-of-knowledge consistency will be achieved for four of the five standards if the same pattern continues. More items related to Standard I (Science as Inquiry) need to be added that are at or above the corresponding objectives. Reviewers did vary in assigning the depth-of-knowledge levels. Reviewers had the strongest agreement comparing the knowledge levels of items and objectives for Standard V, as indicated by the lowest standard deviations. Reviewers disagreed most on comparing levels of items and objectives related to Standard I (Science as Inquiry) and Standard II (Physical Science). The variation among the reviewers in coding these variables implies they needed more training in order to reach higher agreement. Also, the number of items is too small to form any strong conclusion about the variation among reviewers.

What is considered acceptable depth-of-knowledge consistency can vary depending on the purposes for the assessments and standards and the consequences to students and schools. The depth-of-knowledge rating for each objective represented the highest level of cognitive demand for the objective. Sometimes an assessment may measure knowledge or skills related to an objective at a level lower other than the highest level. This is reasonable and to be expected. Because most objectives are broad, to demonstrate full understanding of an objective will require that students have a broad spectrum of knowledge and skills, including both lower- and higher-level skills. There also are reasons why students may be assessed at levels higher than the cognitive demand for an objective--to determine, for example, whether students can exceed expectations. However, on a high-stakes test, the proportion of items above the cognitive level of the objectives and standards tested should be limited.

Range-of-Knowledge Correspondence

The range-of-knowledge criterion is used to judge whether a comparable span of knowledge expected of students for a standard is the same as, or corresponds to, the span of knowledge students need in order to correctly do the assessment items/activities. The number of objectives for each grade 8 standard for science varied from 10 to 23 (Table CS8-3). The total number of objectives for the set of grade 8 standards is 86. The large number of total objectives

and the objectives for each standard makes it impossible for a 30-item test to measure content associated with an adequate number of objectives (50% for this analysis), unless the items are robust and measure content for more than one objective. Most of the 14 draft items reviewed need not meet this condition of measuring content for more than one objective. This implies that if the additional items are similar, the assessment and the standards will have difficulty obtaining range-of-knowledge correspondence. Most of the 14 items related to different objectives. Thus, the objectives did cover a range of objectives for the number of items. However, the total number of objectives is too large to have items measuring at least 50% of the objectives. Reviewers had strong agreement in the number of objectives that had corresponding items. Reviewers only varied by one objective on Standards III, IV, and V. They varied by two objectives on Standard II. On Standard I, as discussed under categorical concurrence above, one reviewer double-coded a number of items, whereas the other four reviewers did not. This reviewer identified five objectives for Standard I that had related assessment items whereas the other reviewers only identified one to three of the objectives that had corresponding items. The lack of agreement by reviewers in coding items related to this standard appears to be attributable to one reviewer, rather than to systematic disagreement among all of the reviewers.

Balance of Representation

The balance-of-representation criterion is used to indicate the degree to which one objective is given more emphasis on the assessment than another does. An index was used to judge the distribution of assessment items. This index only considered the objectives for a standard that had one hit—e.g., one related assessment item/objective.

The low number of items prohibits any meaningful analysis of balance between the assessment and standards for grade 8 science other than the information gained from the range criterion (Table CS8-3). Nearly all of the reviewers coded the 14 items as related to distinct objectives. Because of the high number of objectives and the low number of items in the final assessment, the remaining items also will need to continue the pattern where each item measured knowledge associated with a distinct objective not measured by any other assessment item. The mean balance index other than one for Standards I and V indicate that these are the only two standards where reviewers coded more than one item related to the same objective. This was done by two of the five reviewers on Standard I and one of the five reviewers on Standard V.

Summary for Grade 8 Science Alignment

As for grade 4 science, the small number of assessment items prevents any complete analysis of the alignment between the assessment instrument and the standards. In using only 14 assessment items, these items were reasonable in their distribution among the five standards. Too low a percentage of the items had a depth-of-knowledge level at or above the objectives for Standard I, indicating one area in need of attention. The number of objectives, 86, is so large in relation to the number of items on the assessment that to obtain range and balance will be difficult without including items that measure knowledge for more than one objective.

State C: Alignment of Standards and Assessments in Mathematics for Grades 4 and 8

Organization of Standards for Mathematics

In May 1997, State C's Board of Elementary and Secondary Education adopted the Mathematics Framework, State Standards for Curriculum Development. Four months later the Teachers' Guide to Statewide Assessment in Mathematics was released presenting the design of the state's criterion-referenced tests for grades 4, 8, and 10. The first year that the new grades 4 and 8 tests will be administered will be 1999. The Mathematics Framework specifies standards clustered into three grade ranges—K-4, 5-8, and 9-12. The same six standards are specified for each of the three clusters. Each standard is labeled by a content topic called a strand. The Teachers' Guide provides the assessment frameworks for grades 4, 8, and 10. For each grade level, under each standard, are listed the benchmarks on which students are to be assessed at that grade level. Each benchmark is further elaborated by a list of the essential skills and knowledge to be tested. These skills and knowledge statements are prefaced by "what students should know and be able to do." The six strands and standards are:

I. N: Number and Number Relations
   Standard: In problem-solving investigations, students demonstrate an understanding of the real number system and communicate the relationships within that system using a variety of techniques and tools.

II. A: Algebra
   Standard: In problem-solving investigations, students demonstrate an understanding of concepts and processes that allow them to analyze, represent, and describe relationships among variable quantities and to apply algebraic methods to real-world situations.

III. M: Measurement
   Standard: In problem-solving investigations, students demonstrate an understanding of concepts, processes, and real-life applications of measurement.

IV. G: Geometry
   Standard: In problem-solving investigations, students demonstrate an understanding of geometric concepts and applications involving one-, two-, and three- dimensional geometry, and justify their findings.

V. D: Data Analysis, Probability, and Discrete Math
   Standard: In problem-solving investigations, students discover trends, formulate conjectures regarding cause-and-effect relationships, and demonstrate critical thinking skills in order to make informed decisions.

VI. P: Patterns, Relations, and Functions
   Standard: In problem-solving investigations, students demonstrate an understanding of patterns, relations, and functions that represent and explain real-world situations.

Benchmarks and essential skills and knowledge further explicate what students are to know and be able to do. These vary are different for each grade level. For example, the Algebra Standard and its benchmarks for grade 4 are:

II. Strand A: Algebra

Standard: In problem-solving investigations, students demonstrate an understanding of concepts and processes that allow them to analyze, represent, and describe relationships among variable quantities and to apply algebraic methods to real-world situations.

	A.1	Demonstrating a conceptual understanding of variables, expressions, equations, and inequalities (e.g., use letters or boxes to represent values; understand =, ≠, <, and > symbols)
B.	A.2	Modeling and developing strategies for solving equations and inequalities
C.	A.3	Recognizing the connection of algebra to the other strands and to real-life situations (e.g., number sentences or formulas to represent real-world problems)

Grade 8 has the same algebra standard and the first two benchmarks are nearly the same as for grade 4. In addition to these two, grade 8 has three other benchmarks. All five benchmarks for the grade 8 algebra standard are:

A.	A.1	Demonstrating a conceptual understanding of variables, expressions, equations, and inequalities (e.g., symbolically represent real-world problems as linear terms, equations, or inequalities)
B.	A.2	Modeling and developing methods for solving equations and inequalities (e.g., using charts, graphs, manipulatives, and/or standard algebraic procedures)
C.	A.3	Representing situations and number patterns with tables, graphs, and verbal and written statements, while exploring the relationships among these representations (e.g., multiple representations for the same situation)
D.	A.4	Analyzing tables and graphs to identify relationships exhibited by the data and making generalizations based upon these relationships.
E.	A.5	Demonstrating the connection of algebra to the other strands and to real-life situations.

The Teachers' Guide to Statewide Assessment explicates more explicitly what students are to know and to be able to do by listing the essential skills and knowledge to be tested and other information on the assessment item content. The skills and knowledge listed are not intended to be exhaustive, but to convey the focus for the state criterion-referenced assessment.

For the purposes of this alignment analysis and in order to apply uniform terms across all four participating states and to each grade analyzed, State C's six strands/standards are designated as six standards and are labeled by Roman numerals. The benchmarks identified for each standard are designated as goals and are labeled by capital letters. The essential skills and knowledge to be tested are designated as objectives and are labeled by Arabic numerals.

Label for Analysis	Label by State A
Standard	Strands
Goal	Benchmarks
Objective	Essential skills and knowledge

For example, the following levels of expectations are stated for grade 4 mathematics for the Measurement Strand/Standard, for the nine goals, and for the assigned depth-of-knowledge level:

	Description	Depth-of-Knowledge Level
Standard III	Measurement In problem-solving investigations, students demonstrate an understanding of the concepts, processes, and real-life applications of measurement.	
Goal A	Applying (measure or solve measurement problem) the concepts of length (inches, feet, yards, miles, millimeters, centimeters, decimeters, meters, kilometers), area, volume, capacity (cups, liquid pints and quarts, gallons, milliliters, liters), weight (ounces, pounds, tons, grams, kilograms), mass, time (seconds, minutes, hours, days, weeks, months, years), money, and temperature (Celsius and Fahrenheit) to real-world experiences. What students should know and be able to do includes:	
Objective 1:	Understand the concepts of linear measure (length, perimeter), area (square units), capacity, and weight/mass.	2
Objective 2:	Recognize the names of standard units of measure in the customary (English) and metric system, and related abbreviations (e.g., “cm” or “sq.”), in the context of word problems.	1
Objective 3:	Solve mathematical problems involving measurements in the customary and metric systems, including use of the following units: Linear Capacity Weight/Mass	3

	Temperature Time	
Objective 4:	Measure length and read linear measurements accurately to the nearest centimeter or half inch, using a ruler demarcated in one or both systems.	2
Objective 5:	Understand the concept of measuring time, including 1 hour = 60 minutes, 1 minute = 60 seconds, and using calendar dates to measure time.	2
Objective 6:	Understand the Fahrenheit and Celsius scales as different approaches to measuring temperature, and recognize the practical range of each scale.	2
Objective 7:	Use addition to find perimeter of a geometric shape, given lengths of sides in a labeled illustration or word problem (customary or metric units).	2
Objective 8:	Determine area of an illustrated square or rectangle by counting square units.	2
Objective 9:	Use multiplication to find area of a rectangle, given lengths of one side (square) or two sides (other rectangles) in a labeled illustration or word problem.	2

This alignment analysis matched assessment items to the objectives. In grade 4, State C had 107 objectives. The Teachers' Guide noted that not all of the objectives (benchmarks) would be directly assessed on the state test because it would not be possible to determine which method the student used to answer the question. For example, Goal G of Standard I states students are to construct, use, and explain procedures to compute and estimate with whole numbers (e.g., mental mathematics strategies). This was the case for two goals for Standard I (G and H), one goal for Standard IV (D), and one goal for Standard V (D). These four goals, none with any further refined statement of knowledge and skills, were included in the total of the 107 objectives. In the Teachers' Guide, State C exempted five of the 105 grade 8 objectives from being directly assessed on the state test, including two goals for Standard I (F and G), one for Standard II (C), one for Standard IV (D), and one for Standard VI (B). These five goals, none with any further refined statement of knowledge and skills, were included in the total of the 105 objectives. Reviewers were able to find at least one objective related to each of the items on both the grade 4 and grade 8 assessments and did not have to code any item as related to a standard but not to a specific goal or objective for that standard.

For grade 8, an example of the standard, one benchmark, essential skills and knowledge, and the assigned depth-of-knowledge levels for the Algebra Standard and first benchmark are:

	Description	Depth-of-Knowledge Level
Standard II:	Algebra In problem-solving investigations, students demonstrate an understanding of concepts and processes that allow them to analyze, represent, and describe relationships among variable quantities and to apply algebraic methods to real-world situations.	
Goal A:	Demonstrating a conceptual understanding of variables, expressions, equations, and inequalities (e.g., use letters or boxes to represent values; understand =, ≠, <, and > symbols). What students should know and be able to do includes:	
Objective 1:	Understand the concept of the "unknown," using empty boxes or letters to represent unknown values.	1
Objective 2:	Understand and use symbols (=, ≠, <, and >) to express algebraic relationships.	2
Objective 3:	Use number sentences or formulas containing a variable (letter) to represent real-world problems.	2
Objective 4:	Given an algebraic sentence, write a related story problem (or choose the corresponding story problem), and explain their thinking.	2

Assessments for Mathematics

Analyses were performed in mathematics for grades 4 and 8. The tests analyzed were under development. For each grade, the assessment instrument consisted of two parts—Part A, multiple-choice (about one half without a calculator and about one half with a calculator) and Part B, four constructed response items. Part B for each grade consisted of four extended mathematical tasks that required students to supply a numerical answer, a short written answer, or some other type of constructed response. Reviewers analyzed 74 grade 4 items and parts of items of the 4M2 form—60 Part A and 14 Part B. Each of the four items in Part B had multiple questions. Each part of an item in Part B for this analysis was treated as a separate item. Also, reviewers coded the relationship of the entire item to the standards. This meant for Part B, a total of 14 items were coded—1a, 1b, 1c, 1, 2a, 2b, 2c, 2, 3a, 3b, 3, 4a, 4b, and 4.

Reviewers coded 68 items of Form 8M2 for grade 8. This test consisted of 57 items in Part A and four multiple part items in Part B. For grade 8, only the parts of the items in Part B were coded, not the entire items. The total number of items coded in Part C were 11—1a, 1b, 1c, 2a, 2b, 2c, 3a, 3b, 3c, 4a, and 4b. The Teachers' Guide did not indicate how much time students

would have to complete the tests. The tests do not require students to recall formulas or unit conversions from memory. The needed information to complete items of these types is provided. For both grades 4 and 8 students are to be given a paper ruler and are allowed to use calculators provided by the schools on any item where the use of calculators is not restricted. The Teachers Guide identified the following approximate distribution of items across the strands on Part A:

Strand	Grade 4	Grade 8
Number & Number Relations	40%	20%
Algebra	5%	15%
Measurement	10%	15%
Geometry & Spatial Sense	20%	20%
Data Analysis, Probability, Discrete Math	10%	20%
Patterns, Relations, & Functions	15%	10%

Alignment of Standards and Assessment for Grade 4 Mathematics

Categorical Concurrence

The criterion of categorical concurrence between standards and assessment is met if the same or consistent categories of content appear in both documents. Three reviewers coded the depth-of-knowledge of 107 objectives for the six standards and 74 assessment activities/items (Table CM4-1). The teachers' guide noted that two objectives would not be assessed on the state test, but should be assessed by teachers in their classrooms. One was the fourth geometry objective (benchmark) (Drawing and constructing models, . . .) and one was the fourth data analysis, probability, and discrete mathematics objective (benchmark) (Exploring, formulating, and solving sequence-of-patterns problems . . .). All of the reviewers were able to find at least one of the objectives related to each of the items. No additional objectives were added for any of the six standards.

All of the six standards had a sufficient number of related assessment activities/items (six or more) to make some judgment about a student's knowledge and skills related to each standard. Standard II (Algebra) had the lowest number of related items, according to the blueprint for the test. Standards IV, V, and IV were coded to have on the average about 11 or 12 items. Standard III (Measurement) and Standard I (Number and Number Relations) had the highest number of related items, 18 and 32 respectively. The proportion of the items coded as related to each standard reflected the test blueprint's distribution of items by standards, but did not match this distribution exactly. The alignment analysis indicated that the percentage of items related to Standard III was greater than projected and the percentage of items related to Standard IV was a little lower than projected:

Strand	Projected Grade 4	Alignment Analysis
Number & Number Relations	40%	34%
Algebra	5%	8%
Measurement	10%	20%
Geometry & Spatial Sense	20%	13%
Data Analysis, Probability, and Discrete Math	10%	13%
Patterns, Relations & Functions	15%	12%

Reviewers had high agreement in the number of items coded as related to five of the six standards. On these standards, the three reviewers varied by no more than four items as related to Standard V (Data Analysis). Two coded 12 items and one coded 13 items related to this standard. Reviewers varied the greatest on assigning items to Standard I (Number and Number Relations), the standard with the highest number of related items. One coded 27 items for this standard, one coded 32 items, and one coded 36 items. Numbers are used in many items, so the greater variation on this standard is not surprising. Reviewer agreement could be improved by distinguishing an item that measures knowledge related to students' understanding of number from an item that uses number mainly in measuring knowledge related to another standard.

Depth-of-Knowledge Consistency

Depth-of-knowledge consistency between standards and assessment are aligned if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards. Depth-of-knowledge consistency was attained by all six standards. For all standards, the percent of items with a depth-of-knowledge level that was the same as or above the level of the related objective was more than 50% of the items (Table CM4-2). Reviewers rated over 90% of the objectives to be at a depth-of-knowledge level of 2 (Skill/Concept) or 3 (Strategic Thinking). Reviewers rated, on the average, 72% of the items to be at or above the depth-of-knowledge level of the corresponding objective. The relatively low standard deviations indicate strong agreement among the reviewers in assigning the depth-of-knowledge level of items. On rating 11 items related to Standard VI, reviewers had perfect agreement in assigning the level of knowledge.

Range-of-Knowledge Correspondence

The range-of-knowledge criterion is used to judge whether a span of knowledge expected of students for meeting a standard is the same as, or corresponds to, the span of knowledge students need in order to correctly respond to the assessment items/activities. The correspondence between span of knowledge for a standard and the assessment relates to the number of objectives within the standard with at least one related assessment item/activity. Two of the grade 4 standards met the criterion with items that corresponded to 50% or more of the objectives, Standard I (Number and Number Relations) and Standard VI (Patterns, Relationships,

and Functions) (Table CM4-3). Two of the standards nearly met the criterion, Standards II and III, with items relating to 44% and 46% of the objectives, respectively. The items judged to be related to Standards IV (Geometry and Spatial Sense) and Standard V (Data Analysis, Probability, and Discrete Math) only corresponded to less than one-fourth of the objectives for each of these standards. Even though both of these objectives had, on the average, 12 related items, the items only matched about five of the over 20 objectives for each standard. This coverage is too narrow for an adequate range of knowledge on these standards to be assessed by the test.

Reviewers had high agreement on the number of objectives hit and the percent of objectives hit within a standard for four of the six standards. On the other two standards, two of the three reviewers agreed with each other that items were related to three of the objectives, whereas the third reviewer coded that items were related to six of the objectives. For each of the six standards, two of the three reviewers had exact agreement on the number of objectives with related items. On this criterion, reviewers were very consistent--increasing the confidence that can be given to these results.

Balance of Representation

The balance-of-representation criterion is used to indicate the degree to which one objective within a standard is given more emphasis on the assessment than another does. An index is used to judge the distribution of assessment items. This index only considered the objectives for a standard that had one hit—e.g., one related assessment item/objective.

Items were distributed evenly among objectives with hits for five of the six standards (Table CM4-3). The balance of representation criterion was not met by Standard I (Number and Number Relations). Standard I had the highest number of related items of all of the standards, but each of the reviewers coded 34% to 44% of the items related to that standard as corresponding to only one objective, F2 (Apply computational skills in the context of word problems). A large number of the objectives were hit, over 50%, but in nearly all cases, the reviewers related only one item to each objective. This resulted in an imbalance in the distribution of items among the objectives with over one-third of the items assigned to only one objective and the remaining spread over a number of other objectives. Reviewers had fairly high agreement on the balance index on five of the six standards. On these the standard deviations on the mean balance index were less than .10. On Standard II, two reviewers had the same balance index of .83, but the third reviewer had a balance index of .58. This third reviewer assigned nearly all of the items on that one standard to one objective (A3—use number sentences or formulas containing a variable (letter) to represent real-world problems), whereas the other two reviewers distributed the items more evenly among the objectives. This difference may be due to the third reviewer using more multiple codings. The analysis is not detailed enough to explain fully the reason for the difference among reviewers on this standard.

Summary for Grade 4 Mathematics Alignment

State C's grade 4 mathematics and grade 4 assessment of 74 items/activities, including both multiple-choice and constructed-response items, were found to have alignment on three of the four criteria (Table CM4-4). The standards and assessment had categorical concurrence, depth-of-knowledge consistency, and balance of representation. The range-of-knowledge correspondence criterion was only met by two of the six standards (Standards I and VI). In order to improve the alignment between the standards and the assessment on this criterion, some of the items need to be replaced by those that are related to more of the objectives for each of the standards. For two of the standards (II and III), only 2 or 3 items would need to be replaced. For Standards IV and V, about half of the items would need to be replaced in order to correspond to a greater number of the objectives. State C may have other strategies for assessing additional objectives for these standards, other than the test included in this analysis. The results reported here are only based on the information that was provided in the Teachers' Guide and the mathematics framework. Overall, the grade 4 standards and grade 4 test proved to be aligned.

Not all of the information on the assessment and standards was available for this analysis. Some of the objectives could be assessed by teachers in the classroom rather than on an on-demand assessment. Or, the state may have other plans for determining students' knowledge of the objectives not tested. The findings from this analysis provide some indication of where there may be alignment issues that can be resolved with a greater understanding of the assessment environment in State C.

Alignment of Standards and Assessment for Grade 8 Mathematics

Categorical Concurrence

The criterion of categorical concurrence between standards and assessment is met if the same or consistent categories of content appear in both documents. Four reviewers coded the depth-of-knowledge of 105 objectives for the six standards and 68 assessment activities/items (Table CM8-1). Reviewers coded nearly all of the 68 assessment items/activities as related to one objective. This is signified by the fact that the total mean of hits was 73.75, only slightly higher than the total number of assessment items. On the average among the four reviewers, at most four assessment items were coded as related to more than one objective.

Five of the six standards had a sufficient number of related assessment activities/items (six or more) to make some judgment about a student's knowledge and skills related to each standard. Standard II (Algebra) only weakly met this condition by having a mean number of hits of 5.75. One of the four reviewers coded five items as related to Standard II, while the other three indicated that six items corresponded to this standard. On all of the other standards, all four reviewers coded at least six items as related to objectives of the standard. The two standards with the greatest number of corresponding items were Standard I (Number and Number Relations) and Standard V (Data Analysis, Probability, and Discrete Mathematics). The distribution of items by

standards as coded by the reviewers varied some from the projected distribution in the Teachers' Guide. As coded, Standards I and II were over-emphasized and Standards II and IV were under-emphasized. This cannot be explained by multiple hits for one item, since this rarely occurred.

Strand	Projected Grade 8	Alignment Analysis
Number & Number Relations	20%	34%
Algebra	15%	8%
Measurement	15%	11%
Geometry & Spatial Sense	20%	13%
Data Analysis, Probability, and Discrete Mathematics	20%	25%
Patterns, Relations & Functions	10%	9%

Reviewers had high agreement on the number of items they assigned to four of the six standards. On these standards, the standard deviations ranged between .5 to .82. A standard deviation of .5 indicated the four reviewers only varied by one item. A standard deviation of .82 indicated that the four reviewers only varied by two items. Reviewers differed the most in the number of items coded as related to Standard IV (Geometry) and Standard V (Data Analysis, Probability, and Discrete Math). On these standards, the reviewers varied by three and five items, respectively. The mean among the reviewers takes into consideration the variation among reviewers and provides a more stable statistic. More training could reduce the variation among the reviewers. Even with very little training, there was strong agreement among the reviewers.

Depth-of-Knowledge Consistency

Depth-of-knowledge consistency between standards and assessment are aligned if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards. Depth-of-knowledge consistency was attained by five of the six standards and weakly by the remaining standard (Table CM8-2). For five of the standards, over 50% of the related items had a depth-of-knowledge level at or above the depth-of-knowledge level of the corresponding objective. Standard V (Data Analysis, Probability, and Discrete Math) only barely met the criterion by having 45% of the items rated at the level of knowledge and 5% above the level of knowledge of the corresponding objectives. This standard only weakly met the criterion because of the variation in coding by reviewers. Two of the reviewers rated the items as meeting the 50% criterion, whereas the other two did not. Reviewers did vary in their coding of the depth-of-knowledge level as indicated by two digit standard deviations. However, the percentage of the items for five of the standards at or above the level of knowledge of the related objectives was high enough that even with the variation among reviewers, standards still met the necessary criterion. Thus, except for Standard V, the standards and the test had depth-of-knowledge consistency. The depth-of-knowledge consistency could be improved on Standard V by replacing two or three of the items with items measuring a higher level of knowledge.

Range-of-Knowledge Correspondence

The range-of-knowledge criterion is used to judge whether the span of knowledge expected of students for meeting a standard is the same as, or corresponds to, the span of knowledge students need in order to correctly answer the assessment items/activities. The correspondence between span of knowledge for a standard and the assessment relates to the number of objectives within the standard that has at least one related assessment item or activity. None of the standards met the criterion of having items correspond to 50% or more of the objectives (Table CM8-3). For all six standards, the mean of the percent of objectives with corresponding items was below 50%. The mean percent for Standard V was 49%, which nearly met the criterion. This was judged to have weakly met the criterion because one goal was counted that was not to be assessed on the test. The other mean percentages ranged from 23% to 44%. With a 68-item test and 105 objectives, nearly every item would have measured knowledge related to a different objective unless more robust items were used that measured knowledge related to more than one objective. Students on the test analyzed could achieve satisfactory scores, but only demonstrate knowledge on less than half of the objectives for any one standard.

Reviewers had strong agreement on the mean of the objectives for each standard with related items, except for Standard VI. On the other standards, the standard deviations ranged from 2 to 5. This corresponded to a variation of four to ten percentage points for the means for each reviewer. On Standard VI, one reviewer only related six items to two objectives, whereas the other three reviewers rated seven or eight items related to this standard to four or five objectives. Still three of the four reviewers only coded items related to less than 50% of the objectives. Even with the variation among reviewers, nearly all of the reviewers on all of the standards indicated that the criterion for range-of-knowledge consistency was not met.

Balance of Representation

The balance-of-representation criterion indicates the degree to which one objective within a standard is given more emphasis on the assessment than another objective. An index is used to judge the distribution of assessment items. This index only considered the objectives for a standard that had one hit—e.g., one related assessment item/objective. One on the index indicates perfect balance. Values of .65 or lower indicate some lack of balance, with one or two objectives having a disproportionately higher percentage of relating items than the other objectives. All but one of the six standards had an acceptable level of balance of representation. The index values for all of these standards was .76 or higher. The index value for Standard I (Number and Number Relations) was .60 (Table CM8-3). The low standard deviation of .03 for this standard indicates strong agreement among the four reviewers in their distribution of items among the objectives. All of the four reviewers coded that two-thirds or more of the items related to that standard corresponded to the three objectives (I. E. 2—Solve problems involving positive and negative integers, fractions, decimals, and percents, I. E. 3—Find unit cost and percent of increase and decrease (percent of discount), and I. H.1—Demonstrate a conceptual understanding and applications of proportional reasoning). Even though reviewers found items related to 8 to 10 of the objectives for Standard I, too high a percentage of the items were related to only three of the

objectives for there to be balance among the objectives. Possibly the lack of balance with regards to this standard was intended. However, this means that students are only being assessed significantly on three of 20 objectives, or 15% of the objectives.

Reviewers had high agreement on the mean balance index for five of the six standards. The standard deviations ranged from .01 to .08, which represented variation in the index among the four reviewers from .02 to .17. For Standard IV, the index for the four reviewers ranged from .75 to 1.00, all at an acceptable level. Variation among the reviewers does not change any of the results on this criterion. Five of the six standards and the test showed balance of representation.

Summary for Grade 8 Mathematics Alignment

State C's grade 8 mathematics standards and the grade 8 assessment of 68 items/activities was partially aligned. Three of the six standards (III, IV, and VI) met three of four alignment criteria—categorical concurrence, depth-of-knowledge consistency, and balance of representation (Table CM8-4). None of the six standards fully were aligned with the assessment on the range-of-knowledge correspondence. Each three standards failed to fully achieve alignment on one of the other three criteria. Standard II (Algebra) had too few items to fully meet the categorical concurrence criterion. Standard V (Data Analysis, Probability, and Discrete Math) did not have a sufficient percentage of items at a level of knowledge for the related objectives, and with respect to the reviewer variation, to be judged aligned in depth-of-knowledge consistency. Standard I (Number and Number Relations) failed to have balance of representation. In order for the standards and test to be fully aligned, items related to each of the standards need to be replaced by those measuring different objectives for each standard. In addition, one or two items need to be added that measures objectives related to Standard II (Algebra). One or two items measuring higher level of knowledge related to objectives for Standard V (Data Analysis, Probability, and Discrete Mathematics) need to be replaced. And, items measuring one of the three objectives for Standard I (Number and Number Relations) need to be replaced by items measuring knowledge related to other objectives for that standard.

Not all information for developing the grade 8 mathematics test for State C may have been available to the reviewers who did the alignment analysis. The state may have placed greater priority on certain specific objectives over other objectives. If more information were available from the state, the range-of-knowledge consistency may not be as important an issue, or the lack of balance on objectives for Standard I may prove appropriate. Some strategy may be in place that will assure that students' knowledge is tested on the other objectives either by teachers in their classrooms or by including items on these objectives on other assessment administered at other times. It also may be appropriate to adjust the acceptable level for each alignment criterion on the basis of how students' work will be scored and how teachers will use the information from the assessment. Based on the criteria and the acceptable levels used in this analysis, the test and the standards are aligned to some degree, but this alignment could be improved.

Conclusions

A complete alignment analysis was only performed for grades 4 and 8 mathematics. Too few of the draft science assessment items were available to complete the analysis for science. The mathematics standards and assessments were found to be aligned except on the range-of-knowledge correspondence criterion. All of the grade 4 and 8 mathematics assessments and standards except one attained an acceptable level for categorical concurrence. That is, the assessments had six or more items corresponding to all but one standard. The same proportion of the assessments and standards had depth-of-knowledge consistency by having 50% or more of the assessment items related to each standard at or above the knowledge level of the corresponding objective. For all but two standards, one at grade 4 and one at grade 8, the standards and assessments had an acceptable balance of representation, or items were evenly distributed among the objectives that were hit.

Only two standards, both at grade 4, attained an adequate range-of-knowledge correspondence with the assessments. This criterion was not met by the other ten standards and assessments. For these standards, less than 50% of the objectives had items that measured related content. State C had the highest number of objectives for both grade 4 and grade 8 of any state analyzed, 107 and 105, respectively. The number of assessment items analyzed, 74 and 68, was sufficiently large for the range-of-knowledge correspondence to be met for all standards. However, the assessment items measured too narrow a range of all possible objectives for the criterion to be met. State C, by delineating its expectations in more detailed objectives, increased its challenge to measure a greater range of the objectives in order to have sufficient coverage of its expectations. Better alignment on range-of-knowledge correspondence could be obtained by replacing duplicate items measuring similar content on the same objective by an item that measures content on an objective not now measured, or by replacing items with more robust items that measure content knowledge from more than one objective.

Overall, State C assessments and standards show high alignment that could only be improved by incorporating more assessment items or replacing existing assessment items with those that measure objectives not currently being assessed.

References

National Research Council. (1995). *National science education standards.* Washington, DC: National Academy Press.

Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement, 25* (1), 47-55.

Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states.* Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 18. Madison: University of Wisconsin.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education.* Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6. Madison, WI: University of Wisconsin, Wisconsin Center for Education Research.

Tables

State C

Grade 4 Science Alignment Analysis Tables

Table CS4-1
Categorical Concurrence Between Standards and Assessment As Rated by Five Reviewers
State C--Grade 4 Science
(Number of Assessment Draft Items-14)

Standards			Level by Objective			Hits		Categorical Concurr. Acceptable
Title	Goals #	Objs #	Level	# of objs by Level	% w/in std by Level	Mean	S.D.	
I. Science as Inquiry	2	13	1 2 3	1 6 6	8 46 46	2.20	.45	Undetermined
II. Physical Science	3	16	2	16	100	2.40	.55	Undetermined
III. Life Sciences	3	13	1 2 3	2 10 1	15 77 8	3.00	0	Undetermined
IV. Earth & Space Science	2	13	1 2 3	2 6 5	15 46 39	3.60	.55	Undetermined
V. Science and the Environment	0	5	2	5	100	3.00	0	Undetermined
Total	10	60	1 2 3	5 43 12	8 72 20	14.20	.45	

Table CS4-2
Depth-of-Knowledge Consistency Between Standards and Assessment
As Rated by Five Reviewers
State C--Grade 4 Science
(Total Number of Assessment Draft Items—14)

Standards			Level by Objective			Hits		Level of Item w.r.t. Standard						Depth-of-Knowledge Consistency Acceptable
								% Under		% At		% Above		
Title	Goals #	Objs #	Level	# of objs	%/std	M	S.D.	M	S.D.	M	S.D.	M	S.D.	
I. Science as Inquiry	2	13	1 2 3	1 6 6	8 46 46	2.20	.45	40	42	53	36	7	15	Undetermined
II. Physical Science	3	16	2	16	100	2.40	.55	17	24	73	25	10	22	Undetermined
III. Life Sciences	3	13	1 2 3	2 10 1	15 77 8	3.00	0	5	11	72	18	23	22	Undetermined
IV. Earth & Space Science	2	13	1 2 3	2 6 5	15 46 39	3.60	.55	57	9	43	9	0	0	Undetermined
V. Science and the Environment	0	5	2	5	100	3.00	0	33	24	53	18	13	18	Undetermined
Total	10	60	1 2 3	5 43 12	8 72 20	14.20	.45	31	29	59	24	11	18	

Table CS4-3
Range-of-Knowledge Correspondence and Balance of Representation Between Standards and Assessment As Rated by Five Reviewers
State C--Grade 4 Science
(Total Number of Assessment Draft Items—14)

Standards			Level by Objective Level 1=Recall Level 4=Complex Reasoning			Hits		Range of Objectives				Range of Knowledge Acceptable	Balance Index (1 perfect-0 no balance)				Balance of Representation Acceptable
								# Objs Hit		% of Total			% Hits in Std/Ttl Hits		Index		
Title	Goals #	Objs #	Level	# of objs	%/std	Mean	S.D.	Mean	S.D.	Mean	S.D.		Mean	S.D.	Mean	S.D.	
I. Science as Inquiry	2	13	1 2 3	1 6 6	8 46 46	2.20	.45	2.20	.45	17	4	Undetermined	15	3	1.00	0	Undetermined
II. Physical Science	3	16	2	16	100	2.40	.55	2.40	.55	15	3	Undetermined	17	4	1.00	0	Undetermined
III. Life Sciences	3	13	1 2 3	2 10 1	15 77 8	3.00	0	2.60	.55	20	4	Undetermined	21	1	.93	.09	Undetermined
IV. Earth & Space Science	2	13	1 2 3	2 6 5	15 46 39	3.60	.55	3.60	.55	28	4	Undetermined	25	4	1.00	0	Undetermined
V. Science and the Environment	0	5	2	5	100	3.00	0	3.00	0	60	0	Undetermined	21	1	1.00	0	Undetermined
Total	10	60	1 2 3	5 43 12	8 72 20	14.20	.45	2.76	.66	28	17		20	4	.99	.05	

Table CS4-4
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria
State C—Grade 4 Science
(Total Number of Assessment Draft Items--14)

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
I. Process Skills	Undetermined	Undetermined	Undetermined	Undetermined
II. Plan and Conduct Investigations	Undetermined	Undetermined	Undetermined	Undetermined
III. Area I: Living Things	Undetermined	Undetermined	Undetermined	Undetermined
IV. Area II: Earth and Space Systems	Undetermined	Undetermined	Undetermined	Undetermined
V. Area III: Matter and Energy	Undetermined	Undetermined	Undetermined	Undetermined
VI. Area IV: Applications	Undetermined	Undetermined	Undetermined	Undetermined

State C

Grade 8 Science Alignment Analysis Tables

Table CS8-1
Categorical Concurrence Between Standards and Assessment As Rated by Five Reviewers
State C--Grade 8 Science
(Number of Assessment Draft Items – 14)

Standards			Level by Objective			Hits		Categorical Concurr. Acceptable
Title	Goals #	Objs #	Level	# of objs by Level	% w/in std by Level	Mean	S.D.	
I. Science as Inquiry	2	15	1 2 3	1 6 8	7 40 53	3.20	2.39	Undetermined
II. Physical Science	3	22	1 2 3	2 19 1	9 86 5	3.60	.89	Undetermined
III. Life Science	4	16	1 2	1 15	6 94	2.80	.45	Undetermined
IV. Earth and Space Science	3	23	1 2 3	2 17 4	9 74 17	2.80	.45	Undetermined
V. Science and the Environment	0	10	2 3	9 1	90 10	3.80	.45	Undetermined
Total	12	86	1 2 3	6 66 14	7 77 16	16.20	2.49	

Table CS8-2
Depth-of-Knowledge Consistency Between Standards and Assessment As Rated by Five Reviewers
State C--Grade 8 Science
(Total Number of Assessment Draft Items—14)

Standards			Level by Objective			Hits		Level of Item w.r.t. Standard						
								% Under		% At		% Above		
Title	Goals #	Objs #	Level	# of objs	%/std	M	S.D.	M	S.D.	M	S.D.	M	S.D.	
I. Science as Inquiry	2	15	1 2 3	1 6 8	7 40 53	3.20	2.39	54	36	43	38	3	7	Undetermined
II. Physical Science	3	22	1 2 3	2 19 1	9 86 5	3.60	.89	20	30	69	41	11	15	Undetermined
III. Life Science	4	16	1 2	1 15	6 94	2.80	.45	17	24	70	30	13	30	Undetermined
IV. Earth and Space Science	3	23	1 2 3	2 17 4	9 74 17	2.80	.45	0	0	87	30	13	30	Undetermined
V. Science and the Evironment	0	10	2 3	9 1	90 10	3.80	.45	15	22	57	20	28	18	Undetermined
Total	12	86	1 2 3	6 66 14	7 77 16	16.20	2.49	20	30	65	33	15	22	

Table CS8-3
Range-of-Knowledge Correspondence and Balance of Representation Between Standards and Assessment As Rated by Five Reviewers
State C--Grade 8 Science
(Total Number of Assessment Draft Items—14)

Standards			Level by Objective Level 1=Recall Level 4=Complex Reasoning			Hits		Range of Objectives				Range of Knowledge Acceptable	Balance Index (1 perfect-0 no balance)				Balance of Representation Acceptable
								# Objs Hit		% of Total			% Hits in Std/Ttl Hits		Index		
Title	Goals #	Objs #	Level	# of objs	%/std	Mean	S.D.	Mean	S.D.	Mean	S.D.		Mean	S.D.	Mean	S.D.	
I. Science as Inquiry	2	15	1 2 3	1 6 8	7 40 53	3.20	2.39	2.60	1.52	17	10	Undetermined	19	11	.92	.11	Undetermined
II. Physical Science	3	22	1 2 3	2 19 1	9 86 5	3.60	.89	3.60	.89	16	4	Undetermined	23	6	1.00	0	Undetermined
III. Life Science	4	16	1 2	1 15	6 94	2.80	.45	2.80	.45	18	3	Undetermined	17	3	1.00	0	Undetermined
IV. Earth and Space Science	3	23	1 2 3	2 17 4	9 74 17	2.80	.45	2.80	.45	12	2	Undetermined	18	4	1.00	0	Undetermined
V. Science and the Environment	0	10	2 3	9 1	90 10	3.80	.45	3.60	.55	36	5	Undetermined	24	3	.97	.07	Undetermined
Total	12	86	1 2 3	6 66 14	7 77 16	16.20	2.49	3.08	.91	20	10	Undetermined	20	6	.98	.06	

Table CS8-4
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria
State A Grade 3 Science
(Total Number of Assessment Items--70)

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
I. Process Skills	Undetermined	Undetermined	Undetermined	Undetermined
II. Plan and Conduct Investigations	Undetermined	Undetermined	Undetermined	Undetermined
III. Area I: Living Things	Undetermined	Undetermined	Undetermined	Undetermined
IV. Area II: Earth and Space Systems	Undetermined	Undetermined	Undetermined	Undetermined
V. Area III: Matter and Energy	Undetermined	Undetermined	Undetermined	Undetermined
VI. Area IV: Applications	Undetermined	Undetermined	Undetermined	Undetermined

State C

Grade 4 Mathematics Alignment Analysis Tables

Table CM4-1
Categorical Concurrence Between Standards and Assessment As Rated by Three Reviewers
State C--Grade 4 Mathematics
(Number of Assessment Items-60 in Part A and 4 + 10 parts in Part B for total of 74 items in Form 4M2)

Standards			Level by Objective			Hits		Categorical Concurr.
Title	Goals #	Objs #	Level	# of objs by Level	% w/in std by Level	Mean	S.D.	
I. Number and Number Relations	9	25	1 2 3	3 17 5	12 68 20	31.67	4.51	Yes
II. Algebra	3	9	2 3	7 2	78 22	7.67	1.53	Yes
III. Measurement	5	23	1 2 3	1 17 5	4 74 22	18.00	1.73	Yes
IV. Geometry	7	22	1 2 3	2 12 8	9 55 36	11.67	1.15	Yes
V. Data Analysis, Probability, and Discrete Math	6	21	2 3 4	8 10 3	39 48 14	12.33	.58	Yes
VI. Patterns, Relations, and Functions	3	7	2 3	4 3	57 43	10.67	2.31	Yes
Total	33	107	1 2 3 4	6 65 33 3	6 61 31 3	92.00	7.00	

Table CM4-2
Depth-of-Knowledge Consistency Between Standards and Assessment As Rated by Three Reviewers
State C--Grade 4 Mathematics
(Number of Assessment Items-60 in Part A and 4 + 10 parts in Part B for total of 74 items in Form 4M2)

Standards			Level by Objective			Hits		Level of Item w.r.t. Standard						Depth-of-Knowledge Consistency Acceptable
								% Under		% At		% Above		
Title	Goals #	Objs #	Level	# of objs	%/std	M	S.D.	M	S.D.	M	S.D.	M	S.D.	
I. Number and Number Relations	9	25	1 2 3	3 17 5	12 68 20	31.67	4.51	29	4	71	4	0	0	Yes
II. Algebra	3	9	2 3	7 2	78 22	7.67	1.53	8	14	92	14	0	0	Yes
III. Measurement	5	23	1 2 3	1 17 5	4 74 22	18	1.73	15	5	64	6	21	3	Yes
IV. Geometry	7	22	1 2 3	2 12 8	9 55 36	11.67	1.15	34	16	58	5	8	11	Yes
V. Data Analysis, Probability, and Discrete Math	6	21	2 3 4	8 10 3	39 48 14	12.33	.58	46	5	36	7	18	11	Yes
VI. Patterns, Relations, and Functions	3	7	2 3	4 3	57 43	10.67	2.31	33	0	67	0	0	0	Yes
Total	33	107	1 2 3 4	6 65 33 3	6 61 31 3	92.00	7.00	28	15	64	18	8	11	

Table CM4-3
Range-of-Knowledge Correspondence and Balance of Representation Between Standards and Assessment As Rated by Three Reviewers
State C--Grade 4 Mathematics
(Number of Assessment Items-60 in Part A and 4 + 10 parts in Part B for total of 74 items in Form 4M2)

Standards			Level by Objective Level 1=Recall Level 4=Complex Reasoning			Hits		Range of Objectives				Range of Knowledge Acceptable	Balance Index (1 perfect-0 no balance)				Balance of Representation Acceptable
								# Objs Hit		% of Total			% Hits in Std/Ttl Hits		Index		
Title	Goals #	Objs #	Level	# of objs	%/std	Mean	S.D.	Mean	S.D.	Mean	S.D.		Mean	S.D.	Mean	S.D.	
I. Number and Number Relations	9	25	1 2 3	3 17 5	12 68 20	31.67	4.51	12.67	.58	51	2	Yes	34	2	.60	.06	No
II. Algebra	3	9	2 3	7 2	78 22	7.67	1.53	4.00	1.73	44	19	No	8	1	.75	.14	Yes
III. Measurement	5	23	1 2 3	1 17 5	4 74 22	18	1.73	10.67	.58	46	3	No	20	1	.76	.06	Yes
IV. Geometry	7	22	1 2 3	2 12 8	9 55 36	11.67	1.15	5.33	.58	24	3	No	13	2	.71	.04	Yes
V. Data Analysis, Probability, And Discrete Math	6	21	2 3 4	8 10 3	39 48 14	12.33	.58	4.67	.58	22	3	No	13	1	.73	.08	Yes
VI. Patterns, Relations, and Functions	3	7	2 3	4 3	57 43	10.67	2.31	4.00	1.73	57	25	Yes	12	3	.74	.09	Yes
Total	33	107	1 2 3 4	6 65 33 3	6 61 31 3	92.00	7.00	6.89	3.68	41	18		17	9	.71	.09	

Table CM4-4
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria
State C--Grade 4 Mathematics
(Number of Assessment Items-60 in Part A and 4 + 10 parts in Part B for total of 74 items in Form 4M2)

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
I. Process Skills	Yes	Yes	Yes	No
II. Plan and Conduct Investigations	Yes	Yes	No	Yes
III. Area I: Living Things	Yes	Yes	No	Yes
IV. Area II: Earth and Space Systems	Yes	Yes	No	Yes
V. Area III: Matter and Energy	Yes	Yes	No	Yes
VI. Area IV: Applications	Yes	Yes	Yes	Yes

State C

Grade 8 Mathematics Alignment Analysis Tables

Table CM8-1
Categorical Concurrence Between Standards and Assessment As Rated by Three Raters
State C--Grade 8 Mathematics
(Number of Assessment Items-57 in Part A and 4 (11 parts) in Part B for a total of 68 in Form 8M2)

Standards			Level by Objective			Hits		Categorical Concurr. Acceptable
Title	Goals #	Objs #	Level	# of objs by Level	% w/in std by Level	Mean	S.D.	
I. Number and Number Relations	7	20	1 2 3 4	3 10 6 1	15 50 30 5	25.00	.82	Yes
II. Algebra	5	12	1 2 3 4	1 8 2 1	8 70 17 8	5.75	.50	Weak
III. Measurement	6	18	1 2 3 4	3 11 3 1	17 61 17 6	8.25	.50	Yes
IV. Geometry	6	22	1 2 3 4	7 4 9 2	32 18 41 9	9.50	1.29	Yes
V. Data Analysis, Probability, and Discrete Math	6	23	1 2 3 4	1 5 11 6	4 22 48 26	18.25	1.71	Yes
VI. Patterns, Relations, and Functions	4	10	2 3 4	6 3 1	60 30 10	7.00	.82	Yes
Total	34	105	1 2 3 4	15 44 34 12	14 42 32 12	73.75	2.75	

Table CM8-2
Depth-of-Knowledge Consistency Between Standards and Assessment
As Rated by Four Reviewers
State C--Grade 8 Mathematics
(Number of Assessment Items-57 in Part A and 4 (11 parts) in Part B for a total of 68 in Form 8M2)

Standards			Level by Objective			Hits		Level of Item w.r.t. Standard						Depth-of-Knowledge Consistency Acceptable
								% Under		% At		% Above		
Title	Goals #	Objs #	Level	# of objs	%/std	M	S.D.	M	S.D.	M	S.D.	M	S.D.	
I. Number and Number Relations	7	20	1 2 3 4	3 10 6 1	15 50 30 5	25.00	.82	39	19	59	20	3	6	Yes
II. Algebra	5	12	1 2 3 4	1 8 2 1	8 70 17 8	5.75	.50	8	17	78	20	14	14	Yes
III. Measurement	6	18	1 2 3 4	3 11 3 1	17 61 17 6	8.25	.50	21	9	62	22	17	18	Yes
IV. Geometry	6	22	1 2 3 4	7 4 9 2	32 18 41 9	9.50	1.29	22	9	63	9	15	2	Yes
V. Data Analysis, Probability, and Discrete Math	6	23	1 2 3 4	1 5 11 6	4 22 48 26	18.25	1.71	51	14	45	13	5	4	Weak
VI. Patterns, Relations, and Functions	4	10	2 3 4	6 3 1	60 30 10	7.00	.82	0	0	90	14	10	14	Yes
Total	34	105	1 2 3 4	15 44 34 12	14 42 32 12	73.75	2.75	31	21	61	21	8	11	

Table CM8-3
Range-of-Knowledge Correspondence and Balance of Representation Between Standards and Assessment As Rated by Three Raters
State C--Grade 8 Mathematics
(Number of Assessment Items-57 in Part A and 4 (11 parts) in Part B for a total of 68 in Form 8M2)

Standards			Level by Objective Level 1=Recall Level 4=Complex Reasoning			Hits		Range of Objectives				Range of Knowledge Acceptable	Balance Index (1 perfect-0 no balance)				Balance of Representation Acceptable
								# Objs Hit		% of Total			% Hits in Std/Ttl Hits		Index		
Title	Goals #	Objs #	Level	# of objs	%/std	Mean	S.D.	Mean	S.D.	Mean	S.D.		Mean	S.D.	Mean	S.D.	
I. Number and Number Relations	7	20	1 2 3 4	3 10 6 1	15 50 30 5	25.00	.82	8.75	.96	44	5	No	34	1	.60	.03	No
II. Algebra	5	12	1 2 3 4	1 8 2 1	8 70 17 8	5.75	.50	2.75	.50	23	4	No	8	1	.88	.08	Yes
III. Measurement	6	18	1 2 3 4	3 11 3 1	17 61 17 6	8.25	.50	7.25	.50	40	3	No	11	0	.87	.06	Yes
IV. Geometry	6	22	1 2 3 4	7 4 9 2	32 18 41 9	9.50	1.29	6.75	.96	32	5	No	13	2	.76	.04	Yes
V. Data Analysis, Probability, and Discrete Math	6	23	1 2 3 4	1 5 11 6	4 22 48 26	18.25	1.71	11.25	.50	49	2	Weak	25	2	.77	.01	Yes
VI. Patterns, Relations, and Functions	4	10	2 3 4	6 3 1	60 30 10	7.00	.82	3.75	1.26	38	13	No	9	1	.84	.11	Yes
Total	34	105	1 2 3 4	15 44 34 12	14 42 32 12	73.75	2.75	6.75	3.03	38	10		17	10	.79	.11	

Table CM8-4
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria
State C-- Grade 8 Mathematics
(Number of Assessment Items-57 in Part A and 4 (11 parts) in Part B for a total of 68 in Form 8M2)

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
I. Process Skills	Yes	Yes	No	No
II. Plan and Conduct Investigations	Weak	Yes	No	Yes
III. Area I: Living Things	Yes	Yes	No	Yes
IV. Area II: Earth and Space Systems	Yes	Yes	No	Yes
V. Area III: Matter and Energy	Yes	Weak	Weak	Yes
VI. Area IV: Applications	Yes	Yes	No	Yes