State D

Alignment Between Standards and Assessments in Science for Grades 3, 7, and 10 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*

Table of Contents

Executive Summary

State D
Alignment Between Standards and Assessments in Science for Grades 7 and 10 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 the 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, categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance of representation.

Grades 3, 7, and 10 science and grades 4 and 8 mathematics standards and assessments were analyzed for State D.

State D

Alignment Between Standards and Assessments in Science for Grades 7 and 10 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 D. 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, 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

Six or seven 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 intended outcomes for this alignment study will be the capacity to specify the levels with greater clarity. The revised levels are given in the Summary Report.

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 other states call these expectations "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 in 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 objective for a 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.

---

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.

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 D's framework document. For one standard, the first row specified the standard (strand), the second row a goal (defining elements), and the third row objectives (what students are to know and to do), and so on. The same was done for the second 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 7 science standard and benchmark from above were given these depth-of–knowledge codes:

	Description	Depth-of-Knowledge Level
IV. Strand:	Universe	
IV.A. Defining Element	Characteristics of the Universe	
IV.A.1. To Know	The universe is so large that its distances are expressed in special units (i.e., light years, astronomical units).	
IV.A.1.a. To Do	Use visual and mathematical aids to determine the approximate locations of planets in the solar system. (State D Process Standards: 1.4; 2.2)	2
IV.A.1.b. To Do	Create a model in which the same scale is used to depict the distances between objects and calculate the time required to travel a direct path to them from Earth. (State D Process Standards: 1.6; 2.1)	3
IV.A.1.c. To Do	Interpret and evaluate information related to distances from our solar system to other points in our galaxy and the universe. (State D Process Standards: 1.2; 1.7; 2.7; 3.5; 4.1)	2
IV. A. 2. To Know	Celestial objects possess both similarities and differences.	
IV.A.2.a. To Do	Use a variety of resources to compare and contrast the physical properties of planets. (State D Process Standards: 1.8; 3.5; 2.3)	2

IV.A.3. To Know	Our Solar System is part of the Milky Way Galaxy, one of many galaxies in the universe.	
IV.A.3.a. To Do	Use a variety of visual aids to locate the position of the Solar System in the Milky Way Galaxy. (State D Process Standards: 1.5; 1.6; 2.2; 2.3; 4.1)	1

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 code 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 the depth-of-knowledge levels

were used to judge the consistency by considering 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, an assumption not always met.

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 in 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 review. 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 any one reviewer may have made 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 reliability among reviewers. Low standard deviations indicate that there was little 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 assessment 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 whether 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 nor 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 cut-off score set at 4 would produce an agreement coefficient of .77. Any fewer items with a mean of one-half of the items and a standard deviation of one item would require a cut-off 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. Fifty percent, a conservative cutoff point, 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. For example, assume an assessment included six items related to one standard and students were required to answer correctly four of those items to be judged proficient—i.e., 67% of the items. If three, 50%, of the six items were at or above the depth-of-knowledge level of the corresponding objectives, then for a student to

achieve a proficient score would require the student to answer correctly at least one item at or above the depth-of-knowledge level of one objective. Some leeway was used in this analysis on this criterion. If a standard had between 40% to 50% of its corresponding items at or above the depth-of-knowledge levels of the objectives, then it was reported that the criterion was "weakly" met.

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 that 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 has at least one hit—e.g., 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 D: Alignment of Standards and Assessments in Science for Grades 3, 7 and 10

Organization of Standards for Science

State D established frameworks for curriculum development to provide districts with a "frame" for building curricula based on the state's standards as a foundation. The framework in six content areas uses a three-column format—What All Students Should Know, What All students Should Be Able to Do, and Sample Learning Activities. The frameworks also provide indicators of what students should know and be able to do by the end of grades 4, 8, and 12.

The set of standards for State D consists of two parts. The first part consists of Performance (Process) Standards that apply to all content areas. The state labels each of these as a goal and then lists under each from seven to ten objectives for what students will demonstrate within, and integrate across, all content areas. The four Performance Standards are:

> Goal 1: Students in [State D] public schools will acquire the knowledge and skills to gather analyze and apply information and ideas.
> Goal 2: Students in [State D] public schools will acquire the knowledge and skills to communicate effectively within and beyond the classroom.
> Goal 3: Students in [State D] public schools will acquire the knowledge and skills to recognize and solve problems.
> Goal 4: Students in [State D] public schools will acquire the knowledge and skills to make decisions and act as responsible members of society.

Knowledge (content) standards for six content areas form the second part. Eight standards are listed for science:

> In Science, students in [State D] public schools will acquire a solid foundation that includes knowledge of

SC1 properties and principles of matter and energy
SC2 properties and principles of force and motion
SC3 characteristics and interactions of living organisms
SC4 changes in ecosystems and interactions of organisms with their environments
SC5 processes (such as plate movement, water cycle, airflow), and interactions of Earth's biosphere, atmosphere, lithosphere, and hydrosphere
SC6 composition and structure of the universe and the motions of the objects within it
SC7 processes of scientific inquiry (such as formulating and testing hypotheses)
SC8 impact of science, technology, and human activity on resources and the environment

The overview of the State D science framework states that the central purpose and primary tasks of science education are to awaken in students at all levels a sense of joy and wonder in the excitement and intellectual power of science. Four key principles are provided:

1. All students will have the opportunity to attain high levels of scientific literacy and learn the rigorous content required.
2. All students will develop rigorous knowledge and understanding of science so they can use that knowledge in relationship to scientific, social, personal, and historical perspectives.
3. For all students to understand more science, more resources (i.e., time, personnel, and materials) will be devoted to science education.
4. Science learning will have an active focus.

The science core content is organized into eight strands corresponding to the standards given above (but in a different order) and defining elements for each standard. These are:

I. Scientific Inquiry
   A. Processes
   B. Investigations
II. Scientific Relevance
   A. Nature of Technology
   B. Historical Perspective
   C. Science as a Human Endeavor
III. Matter and Energy
   A. Properties, Characteristics and Structure of Matter
   B. Characteristics, Forms and Sources of Energy
   C. Interactions of Matter and Energy
IV. Force, Motion and Mechanical Energy
   A. Relative Motion

B. Types and Properties of Forces and Motion
C. Interactions of Forces and Motion

V. The Universe
A. Characteristics of the Universe
B. Motion of the Universe
C. Tools of Space Exploration

VI. Earth Systems
A. Physical Systems
B. Processes of Systems

VII. Living Systems
A. Structure/Function/Characteristics of Living Organisms
B. Life Processes
C. Diversity/interdependence
D. Reproduction/Heredity
E. Adaptation/Evolution

VIII. Ecology
A. Interactions of Ecosystems
B. Changes in Ecosystems

The science framework also cites powerful ideas that are grouped into four categories: systems, models, constancy and patterns of change, and scale. The framework then lists, in a three-column format, for each of three grade ranges—K-4, 5-8, and 9-12—and for each defining element what all students should know, what all students should be able to do (corresponding to each knowledge expectation), and sample learning activities. What all students should be able to do are nested under what students should know and are given only for the last grade in the grade range. Thus, every expectation for what students are to do in science corresponds to an expectation for what students are to know.

For the alignment study in science for State D, three levels of expectations were incorporated in the analysis:

Label for Analysis	Label by State A
Standard	Strand/Standard
Goal	Defining element
Objective	What all students should be able to do

For example, expectations given in State D's curriculum framework for its grades 9-12 Matter and Energy strand and Interactions of Matter and Energy defining elements were labeled as the Standard III and Goal C for the grade 10 assessment analysis. Since each statement of what all students should know had one or more corresponding statements of what all students should be able to do, the latter were listed as the objectives. The

expectations for this strand and their defining elements, along with the depth-of-knowledge level, had the following labels:

	Description	Depth-of-Knowledge Level
III. Standard (Strand)	Matter and Energy	
C. Goal	Interactions of Matter and Energy	
1. Objective	Describe how energy is involved in chemical, physical, and nuclear changes.	2
2. Objective	Distinguish the direction of thermal energy in natural processes.	2
3. Objective	Investigate the relationship between heat and work.	2
4. Objective	Investigate phase changes that are induced by adding/subtracting heat energy and explain, using the particulate model, how the interaction of atoms or molecules during a change of state affects the properties of the substance.	2
5. Objective	Describe the interaction of energy waves with the materials of man-made substances.	2
6. Objective	Analyze the amount of energy contained in the mass of substances.	2
7. Objective	Explain how the transfer of energy by air and ocean currents regulate the physical environment of the Earth.	2

Assessments for Science

Even though the Science Curriculum Frameworks give the benchmark for grades 4, 8, and 12, the students are assessed in science at grades 3, 7, and 10. State D created a second document with the input of practicing classroom teachers and administrators entitled *Assessment Annotations for the Curriculum Frameworks: Science: Grades 3, 7, and 10*. The purpose of this document was to provide guidance to the assessment contractor who developed the tests and to indicate to teachers what was "fair game" for assessment. This document used a three-column format organized by content strands. The first two columns are identical to the first two columns in the curriculum framework and state What All Students Should Know and, corresponding to each of these knowledge expectations, What All Students Should Be Able To Do. In the third column, labeled "Grade [3, 7, or 10] Assessment Notes," an indication is given regarding what it is possible to test at the grade level, or is beyond the grade level, by the statement, "Grade [3, 7, or 10] state assessment" or "Beyond grade [3, 7, or 10] assessment." Teacher work

groups met in late 1996 and early 1997 and identified which of the 73 State D "standards" should be assessed. These groups indicated that all of the science knowledge standards should be assessed, eight of 33 process standards for grade 3, and 16 of 33 process standards for grades 3, 7, and 10.

The alignment between standards and the assessment was analyzed in science for grades 3, 7 and 10. All of the tests were constructed by a commercial company specifically for State D. The grade 3 science assessment is designed to be administered in three sessions. Session I consisted of 12 assessment activities. Four of these activities used a fixed-choice format that required students to match the appropriate description with a picture of a process, or select among a fixed set of alternatives. The remaining eight items used a constructed response format, including some short-answer items. Session II for grade 3 consisted of 10 activities, one fixed-choice and nine constructed-response. Session III consisted of 20 multiple-choice items with four choices each and eight constructed-response activities. Overall, the grade 3 assessment contained 50 items, including five fixed-choice, 20 multiple-choice, and 25 constructed-response.

The grade 7 science assessment is designed to be administered in three sessions. Session I consists of 14 constructed-response items that required students to produce a written answer to questions with factual or conceptual answers. In Session II, 6 constructed questions require students to supply information related to doing an experiment or reading data where more that one answer would be acceptable. Session III consists of 25 multiple-choice questions and four constructed-response items. The total number of 49 activities and items are on the grade 7 assessment.

The grade 10 science assessment also is given in three sessions. Session I consists of 11 constructed-response items that require students to produce a written answer to questions with factual or conceptual answers. In Session II, 5 constructed questions require students to supply information about an investigation or some other real-world situation. Session III consists of 25 multiple-choice questions and five constructed-response items. The total number of 46 activities and items are on the grade 10 assessment.

Alignment of Standards and Assessment for Grade 3 Science

Categorical Concurrence

The analysis of grade 3 science standards and assessment for State D was the first one conducted by the science reviewers at the institute. Reviewers were given very little prior training, with the expectation that they would help develop the process. Even though the grade 3 science assessment had 50 activities and items, the mean total hits was 42.67 (Table DS3-1). All of the reviewers coded at least some of the assessment items as not applicable to measuring the science content represented in the standards. The number of

non-applicable assessment items identified by the six reviewers was 4, 9, 16, 8, 2, and 7. This indicated that, at the beginning of the institute, reviewers differed some in their understanding of what assessment activities and science items measured the specified science content knowledge. Three of the reviewers agreed that about 8 of the items were not applicable. Reviewers also found that most of the assessment items measured knowledge related to only one objective and did not require students to demonstrate their understanding of multiple objectives. Only two of the reviewers coded up to five of the items as corresponding to more than one objective. The other four reviewers coded nearly all of the items they found as corresponding to one and only one objective.

The criterion of categorical concurrence between standards and assessment is met if the same or consistent categories of content appear in both documents. The six reviewers coded the depth-of-knowledge of 86 objectives for the eight standards and 50 assessment activities/items (Table DS3-1). The assessment instrument and three of the standards (I, III, and VII) achieved acceptable categorical concurrence. These standards had, on the average, six or more assessment items corresponding to objectives for each standard. Reviewers only found one or two assessment items corresponding to objectives for Standard II (Scientific Relevance). The other four standards had, on the average, between 3 and 5 items corresponding to objectives. Given that this was their first analysis at the institute, reviewers had fairly high agreement on the number of items they coded as corresponding to each objective. On Standard III (Matter and Energy), all of the reviewers identified either 7 or 8 items corresponding to objectives for that standard. The standard deviation of .75 is relatively low. Four of the six reviewers had high agreement on the number of hits for Standard IV (Force, Motion, and Mechanical Energy) by coding either five or six hits for objectives for this standard. One reviewer coded three items and one reviewer coded eight items corresponding to objectives for Standard IV. Reviewers varied greatly only on the number of hits on Standard I (Scientific Inquiry) and Standard VII (Living Systems).

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 only two standards, weakly attained by one standard, and not attained by five standards (Table DS3-2). Most of the objectives, 86%, were judged to have a depth-of-knowledge level of 2 (conceptual and procedural skills) or higher. Reviewers judged that all five of the objectives for Standard II (Scientific Relevance) were at a Level 4 (extended thinking). It would be very difficult on an on-demand assessment to have items that measured knowledge at this highest level. This may explain why reviewers only found one or two items corresponding to objectives for this standard, all with a depth-of-knowledge level below the level of the related objectives. Six of the standards (Standards I, II, III, IV, VI, and VII) had over 50% of the items, on the average,

at a level of knowledge below the level of knowledge of the corresponding objectives. This was judged to be too high a percentage for the assessment to provide sufficient information on a student's attainment of the standard. One of the standards (VII) was on the borderline and only missed attaining an acceptable level due to one or two items being rated below the level of the corresponding objectives. This standard was judged to have weakly met an acceptable level on this criterion.

Reviewers had relatively high agreement in the number of items judged to be below, at, or above the depth-of-knowledge levels of the corresponding objectives. On six of the standards (I, II, V, VI, VII, and VII), five of the six reviewers had varied by no more than 25 points on the percentage of items coded as below the depth-of-knowledge level of the corresponding objective. These also included the standards with the greatest number of hits.

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 have at least one related assessment item or activity. None of the eight standards had an acceptable level of 50% or more of the objectives with corresponding assessment items or activities (Table DS3-3). On two standards (IV and VIII), with the highest percentage of their objectives assessed, there was correspondence between only 40% of their objectives and assessment items. The percentages of objectives for a standard with corresponding assessment items or activities ranged from 24% (Standard III) to 40% (Standards IV and VIII). The total number of items on the assessment, assuming all of the items were applicable, would be adequate to meet this criterion provided the items were judicially selected. However, the criterion for range-of-knowledge correspondence was more difficult to achieve because of the number of items (about 16%) that were judged to be not applicable. Three or more of the reviewers had exact agreement on the percentage of objectives with corresponding items for four of the eight standards (II, III, V, and VIII). This represents fairly high agreement, given that the reviewers had no formal training in using the coding scheme. Reviewers varied the most in the percentage of objectives they found that had corresponding items for Standard I.

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. All eight standards had an acceptable level of balance of representation (Table DS3-3). Reviewers did vary in their coding in the distribution of items across the objectives. This

is particularly evident in those standards where the standard deviation is the highest for the index value, Standards V and VIII. This indicated that one or two of the reviewers had a skewed distribution of coded items on the objectives for these standards. The other reviewers distributed the items they found related to the objectives for a standard more evenly among the full range of the objectives, attaining at least one hit.

Summary for Grade 3 Science Alignment

State D's grade 3 science standards and the grade 3 assessment of 50 items and activities were only fully aligned on one criterion, balance of representation, and only partially aligned on two of the criteria, categorical concurrence and depth-of-knowledge consistency. The assessment and standards had no alignment on range-of-knowledge correspondence. These findings have to be qualified by the fact that this was the first analysis done by the reviewers at the institute. This meant that reviewers varied some simply because they lacked experience with the coding system. Reviewers also found a number of items that were non-applicable to measuring the science content knowledge described in the standards. This made it more difficult to attain range-of-knowledge correspondence. For the assessment and standards to achieve full alignment, according to the acceptable levels used in this analysis, about 10 additional assessment items would need to be added to the assessment that correspond to those standards that currently have the fewest number of related items. Greater alignment could be achieved by replacing the non-acceptable items with items measuring content related to those standards less emphasized on the current assessment instrument. One or two of the items relating to four of the standards need to be replaced by those at a higher depth-of-knowledge level. Any new or replacement items should be those that measure content related to objectives not measured by existing items.

Not all of the information was available to the reviewers. State D may make other provisions to assess the objectives that were not assessed by the assessment instrument analyzed. If this is the case, then the acceptable levels for each of the alignment criteria could be adjusted appropriately to more fully reflect the context under which students' knowledge is to be measured.

Alignment of Standards and Assessment for Grade 7 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. Six reviewers coded the depth-of-knowledge of 93 objectives for the 8 standards and 49 assessment activities/items (Table DS7-1). The assessment was evenly divided between multiple-choice items and constructed-response activities. Very few of the assessment items were coded as relating to two or more objectives. The average number of total hits (count of

objectives corresponding to an assessment item/activity) of 54.83 was only about 6 over the total number of assessment items. The range in the total number of hits among the reviewers was 51 to 61. At most, reviewers found that less than one-fourth of the items corresponded to more than one objective. Thus, most of the items mainly measured students' knowledge in one area and were not robust items that required students to integrate their knowledge among objectives. The assessment was founded to be weighted more toward items measuring Standard I (Scientific Inquiry) (22%) and Standard VII (Living Systems) (18%) and less toward items that measured Standard II (Scientific Relevance) (7%), Standard V (Universe) (6%), and Standard VI (Earth Systems) 89%.

Five of the eight standards had a sufficient number of corresponding assessment items/activities to meet the criterion of categorical concurrence, six or more items. Three of the standards (Standard II, Scientific Relevance, Standard V, The Universe, and Standard VI, Earth Systems) were coded by the reviewers as having, on the average, less than five corresponding items. The reviewers had high agreement in the number of items assigned to three of the eight standards (Standards III, IV, and V). The standard deviations for the distribution of hits among the reviewers for these were very low, indicating that most reviewers varied by only one item. But even with variation among the reviewers in the number of items judged to correspond to the other standards, the variation was not great enough to bring into question the three standards that failed to meet the criterion. At least four of the six reviewers found less than six items related to these 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. Depth-of-knowledge consistency was attained by four standards, weakly attained by two standards, and not attained by two standards (Table DS7-2). The depth-of-knowledge level for nearly two-thirds of the objectives was judged to be Level 2 (Skill/Concept). Four of the standards (Standards III, IV, V, and VIII) had over 50% of the items, on the average, at the level of knowledge or above the level of knowledge of the corresponding objectives. On the average, reviewers found a high percentage of the items corresponding to objectives for Standards I and II to be under the level of knowledge of the objectives, 70% and 61%, respectively. These two standards, along with Standard VII, had the lowest percentage of objectives judged to be at Levels 1 (Recall) or 2 (Skill/Concept), 57% and 50%. This meant it was more difficult for these standards to achieve an acceptable range on the depth-of- knowledge criterion. Reviewers varied greatly on assigning depth-of-knowledge levels as indicated by the relatively high standard deviations. Reviewers did this analysis early in the four-day institute and were still developing a common understanding of the depth-of-knowledge levels. Four of the six reviewers coded more than 70% of the corresponding items to Standard I objectives as under the depth-of-knowledge level of the

objective and three reviewers coded more than 65% of the corresponding items to Standard II objectives as under the level. Even though the reviewers needed more training, there was some agreement that too great a percentage of items corresponding to these two standards were beneath an acceptable level to meet this criterion.

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. Only two of the eight standards met the criterion of having items correspond to 50% or more of the objectives (Table DS7-3). Over 50% of the seven objectives of Standard I (Scientific Inquiry) and of the eight objectives of Standard VIII (Ecology) had corresponding assessment items. The total number of objectives was nearly twice the number of assessment items/activities. Because less than one-fourth of the items were rated to correspond to more than one objective, it would be impossible for all of the standards to meet the acceptable level for the range criterion. The mean percentage of objectives with related assessment items for six standards varied from 20% (Standards V and VI) to 38% (Standard VII). The six reviewers varied some in coding the number of objectives with hits for five of the standards, as indicated by standard deviations of 10 or more. They had relatively high agreement in the number of objectives with hits for three of the standards (III, IV, and V).

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. All of the eight standards had an acceptable level of balance of representation (Table DS7-3). All eight of the standards had index values of .72 or higher. This means that the items were evenly distributed among the objectives that were hit. The six reviewers also had high agreement on the index value as indicated by standard deviations of .13 or less.

Summary for Grade 7 Science Alignment

State D's grade 7 science standards and the grade 7 assessment of 49 items and activities were found to be only partially aligned. The assessment and only five of the eight standards met the categorical concurrence criterion; four standards had depth-of – knowledge correspondence with the assessment, two weakly; and only two standards met the range-of-knowledge correspondence. The items corresponding to a standard were distributed among the objectives with a sufficient number of hits to indicate that all eight standards met the balance-of-representation criterion. A critical factor in the lack of

alignment is the relatively small ratio between the number of items on the assessment and the number of objectives, almost 1:2. The items were not robust enough to measure knowledge related to more than one objective and there were too few items to achieve acceptable levels on three of the criteria. This analysis only used one assessment instrument. The curriculum framework indicated that some objectives should be assessed by the teacher. State D may have plans for assessing students' knowledge in other ways, in which case the misalignment is acceptable given the context of additional assessment. To achieve categorical concurrence, about six items measuring Scientific Inquiry either need to be replaced with items related to other standards, or additional items need to be added. To achieve depth-of-knowledge consistency, only one or two assessment items need to be replaced by items at a higher level of depth-of-knowledge for three of the standards. Finally, any additional or replaced items need to correspond to a greater number of the objectives.

Alignment of Standards and Assessment for Grade 10 Science

Categorical Concurrence

Six reviewers coded the depth-of-knowledge level of 72 objectives for the eight grade 10 standards and 46 assessment activities/items (Table DS10-1). The assessment was divided between 25 multiple-choice items and 21 constructed-response activities. Very few of the assessment items were coded as relating to two or more objectives. The six reviewers averaged a total of 50 hits, only four more than the actual number of assessment items. The reviewer with the highest number of hits, 58, coded at most 12 items as corresponding to more than one objective. This indicates that the reviewers found that most of the 46 assessment items measured content related to only one objective.

Five of the standards had an acceptable number of items that were judged to have categorical concurrence with the assessment (Table DS10-1). Each of these standards (I, II, III, VI, and VII) was judged by reviewers, on the average, to have six or more assessment items corresponding to one of its objectives. Three of the standards (IV, V, and VIII) failed to meet an acceptable level of six or more items. The average number of hits by reviewers on these three standards was less than four. Agreement among reviewers in assigning the number of items that corresponded to objectives of a standard varied. Either five or six of the reviewers had high agreement on the number of hits related to Standards I, III, V, and VI. Four reviewers had high agreement on the number of hits related to Standards II and VII. Only three reviewers had high agreement on the number of hits related to Standards IV and VIII. On the three standards that failed to achieve an acceptable number of related assessment items, reviewers were consistent in coding fewer than six items as corresponding to the objectives of these standards. On Standard V, all six reviewers coded four or fewer items; on Standard VIII, five of the reviewers did; and on Standard IV, four of the reviewers did.

Depth-of-Knowledge Consistency

Depth-of-knowledge consistency between standards and assessment is indicated 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 only one standard, weakly attained by one standard, and not attained by six standards (Table DS10-2). On six standards, the percentage of items with a depth-of-knowledge level below the corresponding objective was, on the average, more than 50%. These findings have to be cautiously interpreted because of the great range in agreement among the reviewers in assigning the depth-of-knowledge level of the assessment items. Reviewers only had strong agreement on Standard I and moderate agreement on Standard VIII. The range in assigning the depth-of-knowledge level to assessment items in relation to the corresponding objectives for the other six standards was high. This indicates the need for more significant training of the reviewers and discussion on how to interpret the four depth-of-knowledge levels used in this analysis.

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. Only one of the eight standards met the criterion that items correspond to 50% or more of the objectives (Table DS10-3). Achieving an acceptable level of range-of-knowledge correspondence is more difficult the greater the number of objectives in relation to the total number of assessment items. For each standard to have an acceptable range-of-knowledge correspondence, it is possible for 46 assessment items to be distributed among the 72 objectives and eight standards. However, reviewers judged that on average assessment items only corresponded to from two to four objectives on any one standard. Since most standards had seven or more objectives, the items were too concentrated on only a small range of objectives to indicate adequate coverage of the span of content within a standard. On seven of the standards, three or more of the reviewers coded an assessment item or activity related to the standard, but not to any of the specific objectives listed under that standard. This implies that the set of objectives for these standards did not cover the full span of content specified in the standard. This could be a problem in the wording of the standard as much as an issue in the selection of items on the assessment. Reviewers' agreement was moderate, with only high agreement on three standards (II, III, and V).

Balance of Representation

All eight of the standards had an acceptable balance of representation (Table DS10-3). The items were evenly distributed over the objectives within a standard that had one or more related assessment items/activities. All of the standards had a balance index of .75 or higher, well above the minimum desired value of .70. Raters had high agreement on the balance index on four of the standards (II, III, IV, and VIII). They had moderate agreement on the remaining four standards.

Summary for Grade 10 Science Alignment

State D's grade 10 science standards and the grade 10 assessment of 46 items and activities were fully aligned on one criterion—balance of representation—and partially aligned on a second criterion—categorical concurrence. On the latter, five of the eight standards met an acceptable level for the criterion. The standards and assessment were found not to be aligned on depth-of-knowledge consistency and range-of-knowledge correspondence. For there to be greater alignment, two or three items should be added to the assessment related to each of three standards. These items should be at the same depth-of-knowledge level as the corresponding objectives. Some of the items related to the standards with the highest number of corresponding items (Standards I and II) could either be replaced or modified to measure knowledge related to that knowledge required by other standards. Finally, one or two new items should be added or existing items replaced to expand the number of objectives with corresponding items. Of course, not all information was available for this analysis. If State D assesses content knowledge in other ways in addition to the test analyzed, then the acceptable levels applied here are too rigorous. This should be taken into consideration in interpreting these findings.

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

Organization of Standards for Mathematics

State D established frameworks for curriculum development to provide districts with a "frame" for building curricula using the state's standards as a foundation. The frameworks in six content areas use a three-column format—What All Students Should Know, What All students Should Be Able to Do, and Sample Learning Activities. The frameworks also provide indicators of what students should know and be able to do by the end of grades 4, 8, and 12.

The standards for State D consist of two parts. The first part, Performance (Process) Standards, apply to all content areas. The state labels each of these as a goal and then lists under each from seven to ten objectives for what students are to demonstrate within, and integrate across, all content areas. The four Performance Standards are:

Goal 1: Students in [State D] public schools will acquire the knowledge and skills to gather analyze and apply information and ideas.
Goal 2: Students in [State D] public schools will acquire the knowledge and skills to communicate effectively within and beyond the classroom.
Goal 3: Students in [State D] public schools will acquire the knowledge and skills to recognize and solve problems.
Goal 4: Students in [State D] public schools will acquire the knowledge and skills to make decisions and act as responsible members of society.

Knowledge (content) standards for six content areas form the second part. Six standards are listed for mathematics:

In mathematics, students in [State D] public schools will acquire a solid foundation that includes knowledge of

MA1 addition, subtraction, multiplication and division; other number sense, including numeration and estimation; and the application of these operations and concepts in the workplace and other situations
MA2 geometric and spatial sense involving measurement (including length, area, volume), trigonometry, and similarity and transformation of shapes
MA3 data analysis, probability and statistics
MA4 patterns and relationships within and among functions and algebraic, geometric and trigonometric concepts
MA5 mathematical systems (including real numbers, whole numbers, integers, fractions), geometry, and number theory (including primes, factors, multiples)
MA6 discrete mathematics (such as graph theory, counting techniques, matrices)

The mathematics framework is prefaced by a statement indicating that it is intended to provide policymakers, instructional leaders, teachers, and community members with the support, definition, and direction necessary to reconstruct mathematics education in the school setting. A goal of the framework is to prepare students to become mathematically literate citizens. The organization of the mathematics standards is nearly identical to that of the NCTM *Standards* (1989). Ten strands are identified. The first four are performance strands—problem solving, communication, reasoning, and connections. They are described as the major areas of emphasis for the teaching of mathematics and the ultimate goals for all students. The other six strands are the knowledge (content) strands enumerated above and are labeled:

Number Sense
Geometric and Spatial Sense
Data Analysis, Probability and Statistics
Patterns and Relationships

Mathematical Systems and Number Theory
Discrete Mathematics

The mathematics framework continues with K-12 content overviews for each of the ten strands. This is followed by a three-column format for each strand, listing what all students should know, what all students should be able to do, and benchmark assessment notes for three grade ranges—K-4, 5-8, and 9-12. The assessment notes pertain to grade 4 for the K-4 range, grade 8 for the 5-8 range, and grade 10 for the 9-12 range. For example, for the K-4 grade range and by the end of grade 4, under Patterns and Relationships, all students should know that

1. Mathematical ideas may be represented with visual models.
2. Mathematical symbols can be used to represent real-world situations.
3. Patterns and relationships can be represented in a variety of ways.
4. Information can be organized to look for a pattern or relationship.
5. Patterns can be geometric and/or numeric.

What all students should be able to do (which is not nested under what they should know) is listed in the middle column:

a. create, recognize, describe, and extend a wide variety of patterns (NCTM Standard 13; State D Process Standards 1.6, 1.8, 2.1, 3.3);
b. represent and describe mathematical relationships (NCTM Standard 13; State D Process Standards 1.6, 1.8, 3.3);
c. investigate the use of variables and open sentences in expressing relationships (NCTM Standards 13; State D Process Standards 1.6, 1.8, 3.3).

The third column, labeled Eighth Grade Assessment Notes, indicates for each expectation of what students should be able to do whether students may be tested on it by the grade 8 state assessment. For some of the standards and what students are to be able to do, the framework indicates that students are to do a local assessment for the entire expectation, or for part of the expectation. For example, under the 5-8 communication strand, all students should be able to model situations using oral, written, concrete, pictorial, graphical, technological, and algebraic methods. According to the assessment notes, oral and technological skills are to be assessed locally; otherwise, they are to be measured on the grade 8 state assessment.

For the purposes of this alignment analysis, in order to apply uniform terms across all four participating states and to each grade analyzed, State D's four performance (process) strands and six knowledge (content) strands are designated as ten standards and are labeled by Roman numerals. What is listed as benchmarks under the column labeled "What students should be able to do" are designed as the objectives and are labeled by

Arabic numerals. What all students should know was not included in this analysis, since the assessment notes indicated that the assessments for grades 4 and 8 would test what students should be able to do.

Label for Analysis	Label by State D
Standard	Strands
Goal	(not used)
Objective	Benchmarks

For example, the following objectives stated for grade 4 under Standard VIII, Patterns and Relationships, along with the assigned depth-of-knowledge level, are:

Expectation	Description	Depth-of-Knowledge Level
VIII: Standard	Patterns and Relationships	
1. Objective	Create, recognize, describe, and extend a wide variety of patterns (NCTM Standard 13; State D Process Standards 1.6, 1.8, 2.1, 3.3)	4
2. Objective	Represent and describe mathematical relationships (NCTM Standard 13; State D Process Standards 1.6, 1.8, 2.2, 3.3)	3
3. Objective	Investigate the use of variables and open sentences in expressing relationships (NCTM Standard 13; State D Process Standards 1.6, 1.8, 3.3)	4

An example of a standard and objectives, along with the assigned depth-of-knowledge levels, for grade 8 under Standard IX, Mathematical Systems and Number Theory, is:

Expectation	Description	Depth-of-Knowledge Level
IX. Standard	Mathematical Systems and Number Theory	
1. Objective	Evaluate the need and applications of numbers not contained in the set of whole numbers.	3
2. Objective	Develop an understanding of and explain order and relationship among integers, fractions, and decimals.	3
3. Objective	Use real-world and mathematical problem situations to develop and apply number theory concepts (such as primes, factors, and multiples).	4
4. Objective	Realize the dynamic nature of mathematics and how different mathematical systems apply to current and developing real-world situations.	4

5. Objective	Apply commutative, associative, and distributive relationships in computation and estimation situations.	2
6. Objective	Recognize the connection of irrational numbers and the real world.	2

Assessments for Mathematics

State D's *Assessment Annotations for the Curriculum Frameworks: Mathematics: Grades 4, 8, and 10* indicates what the state considers as "fair game" on its grades 4, 8, and 10 assessments in mathematics. This document uses a three-column format organized by content strands. The first two columns are identical to the first two columns in the mathematics curriculum framework and state what all students should know and be able to do. In the third column, labeled "Grade [4, 8, or 10] Assessment Notes," an indication is given for each grade level of what may be tested at the grade level by the state assessment, or what should be tested by a local assessment. For example, at grade 4 under the Number Sense Standard, the objective that all students should be able to model, explore, develop, and explain number operations for whole numbers may be tested on the grade 4 state assessment, whereas the objective to use technology to explore numbers is designated for testing by a local assessment.

The alignment between standards and assessment in mathematics was analyzed in this study only for grades 4 and 8. All of the mathematics tests were constructed by a commercial company specifically for State D. The grade 4 mathematics assessment was given in three sessions. Icons were used to indicate those activities on which students could use a ruler, coins, or pattern blocks. Session I consists of 13 constructed-response items that require students to produce a written answer or draw a picture in response to a variety of questions, some with multiple parts. Two of these activities are worth one point, eight worth up to two points, one up to three points, and one up to four points. In Session II, nine constructed-response questions require students to produce an answer to a problem or a situation. Students are required to show their work and on some of the activities explain their answers in words. Three of these activities are worth one point, four up to two points, one up to three points, and one up to four points. Session III consists of 32 multiple-choice questions. All of the multiple-choice items are worth one point. The total number of 54 activities and items were assessed in grade 4.

The grade 8 mathematics assessment is also given in three sessions. Session I consists of 10 constructed-response items that require students to produce a written answer to problems and real-world situations. One of these activities is worth one point, eight of these activities are worth two points, and one is worth up to four points. In Session II, there are 10 constructed-response questions similar in format to those in the first session. One of these activities is worth one point, six are worth two points, two of these up to three points, and one up to four points. Session III consists of 31 multiple-

choice questions. All of the multiple-choice items are worth one point. A total of 51 activities and items are assessed in grade 8.

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. Six reviewers coded the depth-of-knowledge of 56 objectives for the 10 standards and 54 assessment activities/items (Table DM4-1). The assessment activities are very robust and most were coded as corresponding to more than one objective. All but one of the 10 standards had a sufficient number of related assessment activities/items (six or more) to make some judgment about a student's knowledge and skills as related to each standard. Standard 10 (Discrete Mathematics) only had an average of less than three hits, far too few to be considered a valid test on the assessment instrument. Standard 5 (Number Sense) had the highest number of hits, as could be expected for grade 4 mathematics. Standard 9 (Mathematical Systems and Number Theory) barely met the criterion of six items.

The average number of total hits per reviewer was 136, nearly two and a half times greater than the 54 assessment items. On the average, reviewers coded each assessment item as corresponding to two to three objectives. This is explained by the fact that the standards included both process standards and knowledge standards and because the assessment included some open-ended activities. The six reviewers varied in the total number of hits coded, ranging from 82 to 194. This variation is due in part to the different interpretations reviewers applied in the multiple codings. The three reviewers with the highest number of hits, all over 175, coded a large number of hits on the four process standards. One of these reviewers coded 34 hits on Standard III (Connections), whereas another reviewer coded only six hits for this standard and four coded no hits. The four reviewers with the lowest number of hits all had less than 115 hits. This analysis was one of the first analyses done by the reviewers with no training. The more coding the reviewers did, the more consistent they became. The reviewers were very consistent on the number of items assigned to each of the content standards. Reviewers varied in the number of hits for each standard as indicated by the relatively large standard deviations--more than half of the mean number of hits. Reviewers disagreed more on the number of hits for the four process standards than they did on the number of hits for the six knowledge standards. Reviewers could agree on the knowledge measured by an assessment item, but had more difficulty in agreeing if an assessment activity measured problem solving, communication, reasoning, or connections. Reviewers had the highest disagreement on assigning items to the connections standards. Reviewers were not given any guidelines or training on what would constitute a hit for any of the standards. Clearly, some reviewers took more liberty in assigning multiple hits than did others. This issue

most likely could be resolved by more specific training for the reviewers before they begin doing the analysis.

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 four standards, weakly attained by one standard, and not attained by five standards (Table DM4-2). The depth-of-knowledge levels for the objectives were high. None of the objectives were rated with a depth-of-knowledge level of 1 (Recall) and only 21% of the objectives were rated at a knowledge level of 2 (Skill/Concept). Four out of five of the objectives were rated as a Level 3 (Strategic Thinking), or a Level 4 (Extending Thinking). Only four of the standards, including three process standards and Standard X that had very few hits, had over half of the corresponding hits with a depth-of-knowledge level at or above the knowledge level of the standards. Three of the content standards (VII, VIII, and IX) had over 80% of the corresponding items with a depth-of-knowledge level below the knowledge level for the standards. Most of the assessment items had a depth-of-knowledge level of two. Only those standards with some objectives at this level met the criterion.

Reviewers had high agreement in assigning items to standards with depth-of-knowledge levels below the objectives. This was particularly true for the content standards. Reviewers disagreed more in rating the depth-of-knowledge level of items that corresponded to the process standards. As in the case of Categorical Concurrence, training before coding would most likely improve reviewer agreement.

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 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 that has at least one related assessment item or activity. Eight of the ten standards met the criterion of having items correspond to 50% or more of the objectives (Table DM4-3). The number of multiple hits by any one assessment activity increased the likelihood that the range criterion would be met by a large number of standards. For Standard IV (Connections) only one of its four objectives, on the average, had a corresponding item. This means that the reviewers assigned items related to this standard to only one of the possible four items. There were too few items related to Standard X (Discrete Mathematics) for them to cover a sufficient number of the objectives.

Reviewers had high agreement on the percentage of objectives with corresponding items for four of the ten standards, Standards V, VI, VII, and X. They had reasonable agreement on five of the other standards. They varied greatly on the percentage of objectives with corresponding items on Standard IV (Connections).

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. 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 by one or two objectives with a disproportionately higher percentage of relating items than the other objectives. All of the 10 standards had an acceptable level of balance of representation (Table DM4-3). Nine of the standards had index values of .7 or higher. Standard II had an index value of .67, which indicates some stacking of items on two objectives. Standard II was judged to only weakly meet the acceptable level. The raters agreed with each other in the distribution of items among the objectives for a standard as indicated by standard deviations of less than .20 for all of the ten standards.

Summary for Grade 4 Mathematics Alignment

State D's grade 4 mathematics standards and the grade 4 assessment of 54 items and activities were aligned on three of the four criteria—categorical concurrence, range-of-knowledge, and balance of representation. The assessment and standards were only partially aligned on depth-of-knowledge consistency. Too many of the assessment items were at a depth-of-knowledge level below that of the corresponding standards. A high percentage of the objectives for the standards required extended thinking such as applying knowledge to a real-world problem. It could be that the intent by State D is that the students' ability to meet these objectives would be measured by local assessments rather than by the state assessment. If this is the case, then the misalignment here can be explained and is presumably handled elsewhere in the assessment system. This information was not available for this analysis. If the intent that the on-demand state assessment should measure students' ability to do the mathematics specified in the objectives, then a few items—3 or 4 of the items corresponding to most standards—need to be replaced by items at a higher depth-of-knowledge level. Most likely this would require items at level 2 being replaced by level-3 items. Reviewers did vary some in assigning items to specific objectives. Disagreement among reviewers was exacerbated as a result of the fact that process standards were included along with knowledge standards. It is clear that the items are robust and measured knowledge and processes associated with more than one standard.

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. Seven reviewers coded the depth of knowledge of 63 objectives for State D's 10 standards and 51 assessment activities/items (Table DM8-1). Unlike the grade 4 mathematics alignment analysis, the average number of multiple hits was 18, or for at most one third of the items, and about the same number as the constructed-response activities. The grade 8 analysis was performed after the grade 4 analysis and after the reviewers had experience coding one assessment and set of standards. The reviewers refined their coding process by clarifying in more detail what was considered a match between an assessment item and an objective. An assessment item was related to an objective only if the item measured primarily the content knowledge as specified in the objective. This operational definition helped reviewers to decide between knowledge that *had to be used* to answer an item correctly from knowledge that *could be used*. Under this rule, reviewers coded hits corresponding to the four process standards (Standards I, II, III, and IV) less frequently and not unless the process was central to the assessment item.

The grade 8 standards and assessment attained categorical concurrence for five of the ten standards. For Standards I, V, VI, VII, and IX, the average number of hits was six or more. Standard V (Number Sense) had the highest average number of hits, 21.00, 28 % of the hits. All of the other five standards had average hits of less than five. Standard X (Discrete Mathematics) had an average hit of .57, less than one corresponding assessment item. Three of the five standards that did not meet the criterion of categorical concurrence were process standards. Whereas reviewers coded on the average nine items corresponding to Standard I (Problem Solving), they coded on the average less than four items related to the process standards, Communication, Reasoning, and Connections.

Reviewers varied in the number of hits they assigned to each standard. On two of the process standards (I and III), the standard deviation was more than half of the mean number of hits. For Standard III, one reviewer coded a large number of hits compared to the other six. On Standard III (Reasoning), one reviewer coded 12 hits compared to between 0 and 4 by the other reviewers. This was a systematic variation by one reviewer that could be related to the need for more training. Even with variation among the reviewers, still no more than one of the seven coded six or more hits for any of the five standards failing to meet the criterion. The variation among reviewers did not affect the final results that too few assessment items were found that corresponded to five of the 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. Depth-of-knowledge consistency was attained by four standards, weakly attained by two standards, and not attained by three standards (Table DM8-2). No reviewer coded any assessment items corresponding to Process Standard IV (Connections).

As was true for grade 4 mathematics, the depth-of-knowledge level for the objectives was high in grade 8 mathematics. None of the objectives were rated with a depth-of-knowledge level of 1 (Recall) and only 17% of the objectives were rated at a knowledge level of 2 (Skill/Concept). More than four out of five of the objectives were rated as a Level 3 (Strategic Thinking) (38%), or a Level 4 (Extended Thinking) (44%). The standards are very demanding in the level of knowledge they require. Two of the process standards, Standard II (Communication) and Standard III (Reasoning), and two of the content standards, Standard V (Number Sense) and Standard X (Discrete Mathematics), had over half of the corresponding hits at a depth-of-knowledge level at or above the knowledge level of the standards. Standard X had only four hits in total by all of the seven reviewers, of which one hit was rated under the level of knowledge of the objective. Three standards (I, VII, and VIII) had over 65% of the corresponding items with a depth-of-knowledge level below the knowledge level for the standards.

Reviewers varied in their agreement in assigning items to standards with depth-of-knowledge levels below the objective for most standards. Only on Standards VI and IX was there strong agreement among the reviewers, as indicated by a standard deviation for the percent under category of at most 15, or less than one third of the mean. The relative high standard deviations indicate that more training was needed to gain higher agreement among reviewers.

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 required for students 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. Only three of the ten standards met the criterion of having items that corresponded to 50% or more of the objectives (Table DM8-3). Standard VII (Data Analysis) and Standard VIII (Patterns and Relationships) weakly met this criterion. Because the total number of objectives, 63, exceeded the total number of assessment items and is only slightly less than the average number of hits, attaining a correspondence in range-of-knowledge would be difficult unless the items were evenly distributed among

the standards. This was not the case. Nearly three-quarters of the hits were assigned to four standards, Standards I, V, VI, and VII.

Reviewers had relatively high agreement on the average number of objectives with at least one corresponding hit. The standard deviations were 20 or less. The reviewers only varied greatly on the standards where there were very few corresponding items.

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. 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 by one or two objectives having a disproportionately higher percentage of relating items than the other objectives. Eight of the ten standards had an acceptable level of balance of representation (Table DM8-3). Two of the standards (I and V) had index values of .65, or slightly lower, which indicates some overweighing of one objective compared to the others, and thus can only be considered to have weakly met the criterion. The reviewers had high agreement on the distribution of items among the objectives with at least one corresponding assessment item or activity. Most of the standard deviations were below .10.

Summary for Grade 8 Mathematics

State D's grade 8 mathematics standards and the grade 8 assessment of 51 items and activities were aligned on one of the four criteria—balance of representation. On this criterion, there was acceptable agreement between the assessment items and objectives on eight of the ten standards. The assessment and standards partially met the categorical concurrence criterion. There were an acceptable number of items for half of the standards. Too high a frequency of assessment items had a depth-of-knowledge level below that of the corresponding objective for four of the standards. Only for three of the standards did a sufficient number of objectives have corresponding items to represent an adequate range of coverage of the standard on the assessment. Six standards did not. Greater alignment between the assessment and standards could be achieved by replacing more assessment items with those that required students to employ basic mathematical processes of communication, reasoning, and connections. Also, one or two assessment items corresponding to half of the standards should be replaced by assessment items requiring a higher level of knowledge, at least at the level of the corresponding objective. Finally, some items need either to be added or are needed to replace existing items that measure knowledge on a greater range of the objectives for the standards. This was particularly true for the four process standards.

It could be the intent of State D that the students' ability to meet many of the objectives of the standards would be measured by local assessments rather than by the state assessment. If this is the case, then the misalignment here can be explained and presumably addressed elsewhere in the assessment system. This information was not available at the time of this analysis. There was disagreement among reviewers, however not as great as for the grade 4 assessment. However, there was adequate agreement among the seven reviewers to have stability in judging an acceptable level on each of the criteria.

Conclusions

Alignment of State D's standards and assessments was analyzed for three grades in science and two grades in mathematics for a total of five analyses. Using four criteria to judge the alignment, the standards and assessments were found to be only partially aligned. Only the standards and assessments for grade 4 mathematics were found to have adequate alignment on three of the four criteria—categorical concurrence, range-of-knowledge correspondence, and balance of representation. For the other four analyses conducted, nearly 40% or more of the standards and assessments failed to meet the set acceptable level on three of the criteria. In all of the five analyses, at least 70% of the standards and assessments met the balance-of-representation criterion.

No set patterns were found among the grade levels and content areas as to which criteria were met and which were not. The standards and assessments for all five analyses were at least moderately low on depth-of-knowledge consistency. This was attributed in part to State D specifying that a high percentage of its standards meet a depth-of-knowledge level of 3 (Strategic Thinking) or 4 (Extended Thinking). Four of the five analyses found the standards and assessments failed to have categorical concurrence on one-third or more of the items. Too few items were included on the assessment for these standards. One contributing factor to this is that State D had two types of standards—process and content. On only one analysis, 80% of the standards and assessments attained range-of-knowledge correspondence (grade 4 mathematics). On the other four analyses, less than one-third of the standards and assessments attained this criterion. The standards and assessments were less aligned on this criterion than on any of the other three criteria.

The analysis of State D was the first one done at the institute, during the time when the reviewers were still learning the process and developing decision rules. This could have affected the results of the analysis, but it is difficult to determine exactly how. On some of the criteria, the reviewers had fairly high agreement. State D had the largest number of standards for each grade level and content area, but also had one of the lowest numbers of assessment items and activities. It also included a higher proportion of constructed-response items. Finally, no one from the state was present for the analyses to answer questions that arose. State D may have other strategies for assessing students' attainment of some of the expectations such as by teachers in their classrooms. Thus, the

results from this analysis should only be regarded as advisory and considered in the context of the factors just indicated.

References

Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. In *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 D

Grade 3 Science Alignment Analysis Tables

Table DS3-1
Categorical Concurrence Between Standards and Assessment As Rated by Six Reviewers
State D--Grade 3 Science
(Total Number of Assessment Items—20 Multiple Choice. 5 Fixed Choice, and 25 Constructed Response Items—Total 50)

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. Scientific Inquiry	2	7	2 3 4	3 2 2	43 28 28	6.33	3.08	Yes
II. Scientific Relevance	3	5	4	5	100	1.50	.84	No
III. Matter and Energy	3	19	1 2 3	3 13 3	16 68 16	7.17	.75	Yes
IV. Force, Motion, and Mechanical Energy	3	11	1 2 3	3 5 3	27 45 27	5.50	1.64	No
V. Universe	3	12	1 2 3	3 7 2	25 58 17	5.33	1.97	No
VI. Earth Systems	2	11	1 2 3	1 5 5	9 45 45	3.50	1.76	No
VII. Living Systems	5	16	2 3 4	13 2 1	81 13 6	9.17	3.37	Yes
VIII. Ecology	2	5	1 2	2 3	40 60	4.17	1.47	No
Total	23	86	1 2 3 4	12 49 17 8	14 57 20 9	42.67	4.93	

Table DS3-2
Depth-of-Knowledge Consistency Between Standards and Assessment As Rated by Six Reviewers
State D--Grade 3 Science
(Total Number of Assessment Items—20 Multiple Choice, 5 Fixed Choice, and 25 Constructed Response Items—Total 50)

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. Scientific Inquiry	2	7	2 3 4	3 2 2	43 28 28	6.33	3.08	90	15	10	15	0	0	No
II. Scientific Relevance	3	5	4	5	100	1.50	.84	83	41	0	0	0	0	No
III. Matter and Energy	3	19	1 2 3	3 13 3	16 68 16	7.17	.75	61	31	39	31	0	0	No
IV. Force, Motion, and Mechanical Energy	3	11	1 2 3	3 5 3	27 45 27	5.50	1.64	73	17	20	16	7	10	No
V. Universe	3	12	1 2 3	3 7 2	25 58 17	5.33	1.97	31	17	67	19	3	7	Yes
VI. Earth Systems	2	11	1 2 3	1 5 5	9 45 45	3.50	1.76	67	19	21	23	13	21	No
VII. Living Systems	5	16	2 3 4	13 2 1	81 13 6	9.17	3.37	57	13	41	12	2	3	Weak
VIII. Ecology	2	5	1 2	2 3	40 60	4.17	1.47	15	13	41	31	44	37	Yes
Total	23	86	1 2 3 4	12 49 17 8	14 57 20 9	42.67	4.93	60	32	33	28	7	20	

Table DS3-3
Range-of-Knowledge Correspondence and Balance of Representation Between Standards and Assessment As Rated by Six Reviewers
State D--Grade 3 Science
(Total Number of Assessment Items—20 Multiple Choice, 5 Fixed Choice, and 25 Constructed Response Items—Total 50)

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. Scientific Inquiry	2	7	2 3 4	3 2 2	43 28 28	6.33	3.08	2.50	1.38	36	20	No	15	6	.93	.09	Yes
II. Scientific Relevance	3	5	4	5	100	1.50	.84	1.33	.82	27	16	No	4	2	1.00	.00	Yes
III. Matter and Energy	3	19	1 2 3	3 13 3	16 68 16	7.17	.75	4.50	1.76	24	9	No	17	2	.80	.10	Yes
IV. Force, Motion, and Mechanical Energy	3	11	1 2 3	3 5 3	27 45 27	5.50	1.64	4.50	1.05	41	10	No	13	5	.88	.06	Yes
V. Universe	3	12	1 2 3	3 7 2	25 58 17	5.33	1.97	4.17	1.33	35	11	No	12	4	.82	.22	Yes
VI. Earth Systems	2	11	1 2 3	1 5 5	9 45 45	3.50	1.76	2.83	1.47	26	13	No	8	4	.93	.08	Yes
VII. Living Systems	5	16	2 3 4	13 2 1	81 13 6	9.17	3.37	5.33	1.97	33	12	No	21	6	.78	.14	Yes
VIII. Ecology	2	5	1 2	2 3	40 60	4.17	1.47	2.00	.63	40	13	No	10	5	.76	.26	Yes
Total	23	86	1 2 3 4	12 49 17 8	14 57 20 9	42.67	4.93	3.40	1.84	33	14		13	7	.86	.15	

Table DS3-4
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria
State D Grade 3 Science
(Total Number of Assessment Items--50)

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
I. Scientific Inquiry	Yes	No	No	Yes
II. Scientific Relevance	No	No	No	Yes
III. Matter and Energy	Yes	No	No	Yes
IV. Force, Motion, and Mechanical Energy	No	No	No	Yes
V. Universe	No	Yes	No	Yes
VI. Earth Systems	No	No	No	Yes
VII. Living Systems	Yes	Weak	No	Yes
VIII. Ecology	No	Yes	No	Yes

State D

Grade 7 Science Alignment Analysis Tables

Table DS7-1
Categorical Concurrence Between Standards and Assessment As Rated by Six Reviewers
State D--Grade 7 Science
(Total Number of Assessment Items—25 Multiple Choice and 24 Constructed Response Items—Total 49)

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. Scientific Inquiry	2	7	2 3 4	4 2 1	57 28 4	12.17	2.86	Yes
II. Scientific Relevance	3	6	1 2 3	2 1 3	33 17 50	3.83	2.04	No
III. Matter and Energy	3	17	1 2 3	5 11 1	29 65 6	7.83	1.17	Yes
IV. Force, Motion, and Mechanical Energy	3	10	2	10	100	5.67	.82	Yes
V. Universe	3	13	1 2 3 4	3 8 1 1	23 61 18 18	3.50	.84	No
VI. Earth Systems	2	15	2 3	10 5	67 33	4.17	1.94	No
VII. Living Systems	5	17	2 3 4	10 6 1	59 35 6	10.17	3.06	Yes
VIII. Ecology	2	8	2 3	6 2	75 25	7.50	1.64	Yes
Total	23	93	1 2 3 4	10 60 20 3	11 64 22 3	54.83	4.17	

Table DS7-2
Depth-of-Knowledge Consistency Between Standards and Assessment As Rated by Six Reviewers
State D--Grade 7 Science
(Total Number of Assessment Items—25 Multiple Choice and 24 Constructed Response Items—Total 49)

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. Scientific Inquiry	2	7	2 3 4	4 2 1	57 28 4	12.17	2.86	70	21	29	23	2	5	No
II. Scientific Relevance	3	6	1 2 3	2 1 3	33 17 50	3.83	2.04	61	24	31	17	8	14	No
III. Matter and Energy	3	17	1 2 3	5 11 1	29 65 6	7.83	1.17	30	25	53	26	17	15	Yes
IV. Force, Motion, and Mechanical Energy	3	10	2	10	100	5.67	.82	2	5	98	5	0	0	Yes
V. Universe	3	13	1 2 3 4	3 8 1 1	23 61 18 18	3.50	.84	23	24	77	24	0	0	Yes
VI. Earth Systems	2	15	2 3	10 5	67 33	4.17	1.94	55	18	45	18	0	0	Weak
VII. Living Systems	5	17	2 3 4	10 6 1	59 35 6	10.17	3.06	56	22	37	15	7	12	Weak
VIII. Ecology	2	8	2 3	6 2	75 25	7.50	1.64	50	23	45	25	6	14	Yes
Total	23	93	1 2 3 4	10 60 20 3	11 64 22 3	54.83	3.84	47	29	47	29	5	11	

Table DS7-3
Range-of-Knowledge Correspondence and Balance of Representation Between Standards and Assessment As Rated by Six Reviewers
State D--Grade 7
(Total Number of Assessment Items—25 Multiple Choice and 24 Constructed Response Items—Total 49)

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. Scientific Inquiry	2	7	2 3 4	4 2 1	57 28 4	12.17	2.86	4.17	.98	54	11	Yes	22	5	.74	.09	Yes
II. Scientific Relevance	3	6	1 2 3	2 1 3	33 17 50	3.83	2.04	2.33	1.21	33	17	No	7	3	.95	.08	Yes
III. Matter and Energy	3	17	1 2 3	5 11 1	29 65 6	7.83	1.17	4.50	1.05	25	6	No	14	2	.75	.08	Yes
IV. Force, Motion, and Mechanical Energy	3	10	2	10	100	5.67	.82	3.50	.55	35	5	No	10	2	.82	.04	Yes
V. Universe	3	13	1 2 3 4	3 8 1 1	23 61 18 18	3.50	.84	2.67	.82	20	5	No	6	2	.88	.06	Yes
VI. Earth Systems	2	15	2 3	10 5	67 33	4.17	1.94	3.17	.98	20	6	No	8	3	.89	.13	Yes
VII. Living Systems	5	17	2 3 4	10 6 1	59 35 6	10.17	3.06	6.67	2.34	38	14	No	19	6	.77	.10	Yes
VIII. Ecology	2	8	2 3	6 2	75 25	7.50	1.64	5.17	1.47	59	16	Yes	14	3	.85	.10	Yes
Total	23	93	1 2 3 4	10 60 20 3	11 64 22 3	54.83	3.84	4.02	1.79	36	17		13	6	.83	.11	

Table DS7-4
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria
State D Grade 7 Science
(Total Number of Assessment Items--49)

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
I. Scientific Inquiry	Yes	No	Yes	Yes
II. Scientific Relevance	No	No	No	Yes
III. Matter and Energy	Yes	Yes	No	Yes
IV. Force, Motion, and Mechanical Energy	Yes	Yes	No	Yes
V. Universe	No	Yes	No	Yes
VI. Earth Systems	No	Weak	No	Yes
VII. Living Systems	Yes	Weak	No	Yes
VIII. Ecology	Yes	Yes	Yes	Yes

State D

Grade 10 Science Alignment Analysis Tables

Table DS10-1
Categorical Concurrence Between Standards and Assessment As Rated by Six Reviewers
State D--Grade 10 Science
(Total Number of Assessment Items—25 Multiple Choice and 21 Constructed Response Items—Total 46)

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. Scientific Inquiry	2	7	3 4	4 3	57 43	9.00	1.79	Yes
II. Scientific Relevance	3	5	3 4	3 2	60 40	8.00	2.97	Yes
III. Matter and Energy	3	15	1 2 3	1 13 1	7 86 7	6.50	1.76	Yes
IV. Force, Motion, and Mechanical Energy	3	7	2 3	4 3	57 43	3.83	2.23	No
V. Universe	3	6	2 3	3 3	50 50	3.17	.98	No
VI. Earth Systems	2	10	2 3 4	4 4 2	40 40 20	7.50	2.26	Yes
VII. Living Systems	5	15	2 3	10 5	67 33	8.50	1.64	Yes
VIII. Ecology	2	7	2 3	6 1	86 14	3.83	2.32	No
Total	23	72	1 2 3 4	1 40 24 7	1 56 33 10	50.33	4.97	

Table DS10-2
Depth-of-Knowledge Consistency Between Standards and Assessment As Rated by Six Reviewers
State D--Grade 10 Science
(Total Number of Assessment Items—25 Multiple Choice and 21 Constructed Response Items—Total 46)

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. Scientific Inquiry	2	7	3 4	4 3	57 43	9.00	1.79	92	6	6	6	2	4	No
II. Scientific Relevance	3	5	3 4	3 2	60 40	8.00	2.97	53	27	40	27	8	19	Weak
III. Matter and Energy	3	15	1 2 3	1 13 1	7 86 7	6.50	1.76	59	32	41	32	0	0	No
IV. Force, Motion, and Mechanical Energy	3	7	2 3	4 3	57 43	3.83	2.23	66	26	33	25	1	3	No
V. Universe	3	6	2 3	3 3	50 50	3.17	.98	71	30	29	30	0	0	No
VI. Earth Systems	2	10	2 3 4	4 4 2	40 40 20	7.50	2.26	79	18	18	15	3	8	No
VII. Living Systems	5	15	2 3	10 5	67 33	8.50	1.64	68	24	27	24	5	7	No
VIII. Ecology	2	7	2 3	6 1	86 14	3.83	2.32	4	10	74	39	22	40	Yes
Total	23	72	1 2 3 4	1 40 24 7	1 56 33 10	50.33	4.58	65	33	31	31	4	17	

Table DS10-3
Range-of-Knowledge Correspondence and Balance of Representation Between Standards and Assessment As Rated by Six Reviewers
State D--Grade 10 Science
(Total Number of Assessment Items—25 Multiple Choice and 21 Constructed Response Items—Total 46)

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. Scientific Inquiry	2	7	3 4	4 3	57 43	9.00	1.79	3.33	1.75	42*	22	No	18	3	.75	.15	Yes
II. Scientific Relevance	3	5	3 4	3 2	60 40	8.00	2.97	3.67	.52	65*	10	Yes	16	5	.85	.06	Yes
III. Matter and Energy	3	15	1 2 3	1 13 1	7 86 7	6.50	1.76	3.83	1.17	24*	7	No	13	3	.77	.07	Yes
IV. Force, Motion, and Mechanical Energy	3	7	2 3	4 3	57 43	3.83	2.23	2.67	1.51	35	18	No	8	5	.92	.10	Yes
V. Universe	3	6	2 3	3 3	50 50	3.17	.98	2.83	.75	43*	12	No	6	2	.90	.13	Yes
VI. Earth Systems	2	10	2 3 4	4 4 2	40 40 20	7.50	2.26	3.67	1.51	34*	15	No	15	3	.80	.16	Yes
VII. Living Systems	5	15	2 3	10 5	67 33	8.50	1.64	4.67	2.16	29*	14	No	17	3	.79	.14	Yes
VIII. Ecology	2	7	2 3	6 1	86 14	3.83	2.32	2.33	1.21	31*	15	No	8	5	.93	.08	Yes
Total	23	72	1 2 3 4	1 40 24 7	1 56 33 10	50.33	4.58	3.38	1.48	38	18		13	6	.84	.13	

*One half or more of the reviewers assigned one or more items to the standard but not to a specific objective.
The mean was computed using the total objectives plus one.

Table DS10-4
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria
State D Grade 10 Science
(Total Number of Assessment Items--46)

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
I. Scientific Inquiry	Yes	No	No	Yes
II. Scientific Relevance	Yes	Weak	Yes	Yes
III. Matter and Energy	Yes	No	No	Yes
IV. Force, Motion, and Mechanical Energy	No	No	No	Yes
V. Universe	No	No	No	Yes
VI. Earth Systems	Yes	No	No	Yes
VII. Living Systems	Yes	No	No	Yes
VIII. Ecology	No	Yes	No	Yes

Table DM4-1
Categorical Concurrence Between Standards and Assessment As Rated by Seven Reviewers
State D--Grade 4 Mathematics
(Number of Assessment Items–32 Multiple Choice and 22 Constructed Response—Total 54)

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. Problem Solving	1	6	3 4	4 2	67 33	18.71	11.34	Yes
II. Communication	1	5	2 3	1 4	20 80	21.57	16.15	Yes
III. Reasoning	1	4	3 4	2 2	50 50	9.00	6.68	Yes
IV. Connections	1	4	2 4	1 3	25 75	8.14	13.02	Yes
V. Number Sense	1	9	2 3 4	3 3 3	33 33 33	31.57	8.08	Yes
VI. Geometric and Spatial Sense	1	9	2 3 4	3 2 4	33 22 44	16.43	5.65	Yes
VII. Data Analysis, Probability and Statistics	1	4	3 4	1 3	25 75	11.57	2.37	Yes
VIII. Patterns and Relationships	1	3	3 4	1 2	33 67	9.14	3.29	Yes
IX. Mathematical Systems and Number Theory	1	5	3 4	4 1	80 20	6.71	5.31	Yes
X. Discrete Mathematics	1	7	2 3 4	4 2 1	57 29 14	2.71	2.14	No
Total	10	56	2 3 4	12 23 21	21 41 38	135.98	45.21	

Table DM4-2
Depth-of-Knowledge Consistency Between Standards and Assessment As Rated by Seven Reviewers
State D--Grade 4 Mathematics
(Number of Assessment Items–32 Multiple Choice and 22 Constructed Response—Total 54)

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. Problem Solving	1	6	3 4	4 2	67 33	18.71	11.34	60	24	33	13	7	13	No
II. Communication	1	5	2 3	1 4	20 80	21.57	16.15	23	14	47	18	29	14	Yes
III. Reasoning	1	4	3 4	2 2	50 50	9.00	6.68	34	27	40	30	12	16	Yes
IV. Connections	1	4	2 4	1 3	25 75	8.14	13.02	37	47	5	10	0	0	Yes
V. Number Sense	1	9	2 3 4	3 3 3	33 33 33	31.57	8.08	58	10	33	9	9	6	No
VI. Geometric and Spatial Sense	1	9	2 3 4	3 2 4	33 22 44	16.43	5.65	53	20	32	16	15	14	Weak
VII. Data Analysis, Probability and Statistics	1	4	3 4	1 3	25 75	11.57	2.37	86	17	11	15	3	3	No
VIII. Patterns and Relationships	1	3	3 4	1 2	33 67	9.14	3.29	86	18	12	16	2	5	No
IX. Mathematical Systems and Number Theory	1	5	3 4	4 1	80 20	6.71	5.31	83	0	17	24	0	0	No
X. Discrete Mathematics	1	7	2 3 4	4 2 1	57 29 14	2.71	2.14	0	0	55	52	45	52	Yes
Total	10	56	2 3 4	12 23 21	21 41 38	135.98	45.21	54	35	33	27	12	22	

Table DM4-3
Range-of-Knowledge Correspondence and Balance of Representation Between Standards and Assessment As Rated by Seven Reviewers
State D--Grade 4 Mathematics
(Number of Assessment Items–32 Multiple Choice and 22 Constructed Response—Total 54)

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. Problem Solving	1	6	3 4	4 2	67 33	18.71	11.34	4.14	1.35	71	25	Yes	14	7	.71	.09	Yes
II. Communication	1	5	2 3	1 4	20 80	21.57	16.15	3.71	1.11	74	22	Yes	14	7	.67	.16	Weak
III. Reasoning	1	4	3 4	2 2	50 50	9.00	6.68	2.86	1.46	71	37	Yes	6	3	.81	.17	Yes
IV. Connections	1	4	2 4	1 3	25 75	8.14	13.02	.71	.95	18	24	No	5	7	.89	.19	Yes
V. Number Sense	1	9	2 3 4	3 3 3	33 33 33	31.57	8.08	7.29	1.25	81	14	Yes	25	9	.73	.04	Yes
VI. Geometric and Spatial Sense	1	9	2 3 4	3 2 4	33 22 44	16.43	5.65	7.29	.95	81	11	Yes	13	4	.79	.05	Yes
VII. Data Analysis, Probability and Statistics	1	4	3 4	1 3	25 75	11.57	2.37	2.86	.69	71	17	Yes	9	4	.71	.11	Yes
VIII. Patterns and Relationships	1	3	3 4	1 2	33 67	9.14	3.29	2.14	.90	71	30	Yes	7	3	.78	.16	Yes
IX. Mathematical Systems and Number Theory	1	5	3 4	4 1	80 20	6.71	5.31	3.14	1.21	63	24	Yes	4	3	.76	.11	Yes
X. Discrete Mathematics	1	7	2 3 4	4 2 1	57 29 14	2.71	2.14	2.00	1.00	29	14	No	2	2	.92	.15	Yes
Total	10	56	2 3 4	12 23 21	21 41 38	135.98	45.21	4.16	2.30	66	30		10	8	.78	.15	

Table DM4-4
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria
State D Grade 4 Mathematics
(Total Number of Assessment Items--54)

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
I. Problem Solving	Yes	No	Yes	Yes
II. Communication	Yes	Yes	Yes	Weak
III. Reasoning	Yes	Yes	Yes	Yes
IV. Connections	Yes	Yes	No	Yes
V. Number Sense	Yes	No	Yes	Yes
VI. Geometric and Spatial Sense	Yes	Weak	Yes	Yes
VII. Data Analysis, Probability and Statistics	Yes	No	Yes	Yes
VIII. Patterns and Relationships	Yes	No	Yes	Yes
IX. Mathematical Systems and Number Theory	Yes	No	Yes	Yes
X. Discrete Mathematics	No	Yes	No	No

Table DM8-1
Categorical Concurrence Between Standards and Assessment As Rated by Seven Reviewers
State D--Grade 8 Mathematics
(Number of Assessment Items–31 Multiple Choice and 20 Constructed Responses—Total 51)

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. Problem Solving	1	5	3 4	2 3	40 60	9.43	4.79	Yes
II. Communication	1	7	3 4	3 4	43 57	2.29	1.11	No
III. Reasoning	1	5	3 4	4 1	80 20	3.71	3.90	No
IV. Connections	1	4	4	4	100	0	0	No
V. Number Sense	1	9	2 3 4	4 3 2	44 33 22	21.00	4.55	Yes
VI. Geometric and Spatial Sense	1	9	2 3 4	3 1 5	33 11 56	10.71	2.14	Yes
VII. Data Analysis, Probability and Statistics	1	9	3 4	4 5	44 56	10.00	2.08	Yes
VIII. Patterns and Relationships	1	5	2 3 4	1 3 1	20 60 20	4.14	1.86	No
IX. Mathematical Systems and Number Theory	1	6	2 3 4	2 2 2	33 33 33	7.14	1.95	Yes
X. Discrete Mathematics	1	4	2 3 4	1 2 1	25 50 25	.57	.98	No
Total	10	63	2 3 4	11 24 28	17 38 44	69.00	9.97	

Table DM8-2
Depth-of-Knowledge Consistency Between Standards and Assessment As Rated by Seven Reviewers
State D--Grade 8 Mathematics
(Number of Assessment Items–31 Multiple Choice and 20 Constructed Responses—Total 51)

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. Problem Solving	1	5	3 4	2 3	40 60	9.43	4.79	74	35	19	19	7	19	No
II. Communication	1	7	3 4	3 4	43 57	2.29	1.11	45	30	43	36	12	21	Yes
III. Reasoning	1	5	3 4	4 1	80 20	3.71	3.90	31	37	26	24	29	37	Yes
IV. Connections	1	4	4	4	100	0	0	0	0	0	0	0	0	No
V. Number Sense	1	9	2 3 4	4 3 2	44 33 22	21.00	4.55	48	28	40	21	13	10	Yes
VI. Geometric and Spatial Sense	1	9	2 3 4	3 1 5	33 11 56	10.71	2.14	55	13	26	15	19	12	Weak
VII. Data Analysis, Probability and Statistics	1	9	3 4	4 5	44 56	10.00	2.08	67	28	33	28	0	0	No
VIII. Patterns and Relationships	1	5	2 3 4	1 3 1	20 60 20	4.14	1.86	71	29	19	30	10	16	No
IX. Mathematical Systems and Number Theory	1	6	2 3 4	2 2 2	33 33 33	7.14	1.95	58	14	16	10	25	9	Weak
X. Discrete Mathematics	1	4	2 3 4	1 2 1	25 50 25	.57	.98	7	19	14	38	7	19	Yes
Total	10	63	2 3 4	11 24 28	17 38 44	69.00	9.97	57	34	31	26	12	19	

Table DM8-3
Range-of-Knowledge Correspondence and Balance of Representation Between Standards and Assessment As Rated by Seven Reviewers
State D--Grade 8 Mathematics
(Number of Assessment Items–31 Multiple Choice and 20 Constructed Responses—Total 51)

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. Problem Solving	1	5	3 4	2 3	40 60	9.43	4.79	2.14	.38	43	8	No	14	7	.65	.16	Weak
II. Communication	1	7	3 4	3 4	43 57	2.29	1.11	1.71	.76	24	11	No	3	2	.96	.09	Yes
III. Reasoning	1	5	3 4	4 1	80 20	3.71	3.90	1.71	.95	33	19	No	5	5	.88	.08	Yes
IV. Connections	1	4	4	4	100	0	0	.0	0	0	0	No	0	0	0	0	Yes
V. Number Sense	1	9	2 3 4	4 3 2	44 33 22	21.00	4.55	5.43	2.23	57	21	Yes	30	6	.62	.08	Weak
VI. Geometric and Spatial Sense	1	9	2 3 4	3 1 5	33 11 56	10.17	2.14	5.14	.69	55	7	Yes	16	3	.75	.04	Yes
VII. Data Analysis, Probability and Statistics	1	9	3 4	4 5	44 56	10.00	2.08	4.43	1.40	49	16	Weak	15	2	.74	.09	Yes
VIII. Patterns and Relationships	1	5	2 3 4	1 3 1	20 60 20	4.14	1.86	2.43	.53	47	10	Weak	6	3	.83	.09	Yes
IX. Mathematical Systems and Number Theory	1	6	2 3 4	2 2 2	33 33 33	7.14	1.95	3.57	.53	54	8	Yes	10	2	.77	.05	Yes
X. Discrete Mathematics	1	4	2 3 4	1 2 1	25 50 25	.57	.98	.43	.79	11	20	No	1	1	1.00	0	Yes
Total	10	63	2 3 4	11 24 28	17 38 44	69.00	9.97	2.70	2.03	37	23		10	9	.82	.15	

Table DM8-4
Summary of Attainment of Acceptable Alignment Level on Four Content Focus Criteria
State D Grade 8 Mathematics
(Total Number of Assessment Items--51)

Standards	Alignment Criteria			
	Categorical Concurrence	Depth-of-Knowledge Consistency	Range of Knowledge	Balance of Representation
I. Problem Solving	Yes	No	No	Weak
II. Communication	No	Yes	No	Yes
III. Reasoning	No	Yes	No	Yes
IV. Connections	No	No	No	Yes
V. Number Sense	Yes	Yes	Yes	Weak
VI. Geometric and Spatial Sense	Yes	Weak	Yes	Yes
VII. Data Analysis, Probability and Statistics	Yes	No	Weak	Yes
VIII. Patterns and Relationships	No	No	Weak	Yes
IX. Mathematical Systems and Number Theory	Yes	Weak	Yes	Yes
X. Discrete Mathematics	No	Yes	No	Yes