REVIEWER BACKGROUND INFORMATION AND INSTRUCTIONS
MATHEMATICS STANDARDS AND ASSESSMENT ALIGNMENT ANALYSIS
CCSSO TILSA ALIGNMENT STUDY
May 21-24, 2001
version 2.0

Overview of Alignment Study

This alignment study is designed to produce measures on five criteria. The underlying assumption of this approach is to compare the relationship between assessment instruments and standards by analyzing how these documents compare using the same criteria. The five criteria are:

1. **Categorical Concurrence**
   This criterion between standards and assessment is met if the same or consistent categories of content appear in both documents.
2. **Depth-of-Knowledge Consistency**
   This criterion between standards and assessment is met 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.
3. **Range-of-Knowledge Correspondence**
   This criterion is met if 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.
4. **Balance of Representation**
   This criterion is met if the degree to which one objective is given emphasis on the assessment is comparable to the emphasis given to the other objectives within a standard.
5. **Source of Challenge**
   This criterion is met if the primary difficulty of the assessment items is highly related to students' knowledge and skill with the content area as represented in the standards.

The primary role of each reviewer in this analysis is to do four tasks:

1. judge the depth-of-knowledge level of each objective,
2. judge the depth-of-knowledge level of each assessment item,
3. identify the one or two objectives to which each item corresponds, and
4. judge if there is a source of challenge issue with an item.

All of the information from these four tasks is to be recorded on a coding sheet, a matrix with standards as rows and item numbers as columns. The item-by-objective codings by reviewers will then be aggregated and analyzed. The analysis will produce statistics on each of the five criteria. To standardize terminology for this analysis, levels of standards will be referred to as standard (highest level), goal (next level of specificity within each standard), and objective (third or final level of specificity within each standard).

Rating Depth-of-Knowledge Levels for Each Objective

Reviewers' first task is to become familiar with the state standards to be included in the alignment analysis. This will require some study to understand the structure of the set of standards, what topics and expectations are included under each standard, how the goals and objective span a standard, and what terminology is used by the state to express what students are expected to know and to do.

Reviewers' second task is to assign a depth-of-knowledge level (as defined below) to each objective. The depth-of-knowledge level of an objective should be the level of the items most commonly used to measure if a student in that grade knows and can do what is expected by the objective. In assigning a depth-of-knowledge level to an objective, think about the domain of items that would be appropriate to measure the objective. Then identify the depth-of-knowledge level of the most common of these items. If there is a question between two levels for an objective, such as a 1 or a 2, or 2 or 3, select the higher of the two levels. The depth-of-knowledge level should be entered on the coding sheet in the column to the left of the objective statement with the title DOK. At the institute, after receiving training the team of reviewers for each state will reach consensus on the depth-of-knowledge level for each objective before coding items.

Depth of Knowledge Consistency Criterion

Standards and assessments can be aligned on the basis of 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.

Standards vary on the complexity of what students are expected to know and do. Some standards simply expect students to reproduce a fact or complete a sequence of steps while others expect students to reason, extend their thinking, synthesize information from a multiple of sources, and produce significant work over time. Alignment on depth-of-knowledge is achieved when the assessment and standards agree on the cognitive level students are expected to demonstrate and are asked to perform. It is unreasonable, and even undesirable, for every assessment item to have precisely the same depth-of-knowledge level as the corresponding standard or objective. The domain of items corresponding to most statements of outcomes would include a range of depth-of-knowledge levels. However, it is reasonable to expect that a majority of the items on an assessment correspond to the most common depth of knowledge level within the domain of items.

Four levels of depth-of-knowledge are used in this analysis. The levels represent a hierarchy based on two main factors. One factor is the mathematical sophistication and complexity. The mathematical sophistication will depend on the abstractness of the activity, the amount of mathematics that has to be recalled or drawn upon, the number of mathematical principles needed to be employed, the lack of routine, and the need to

extend or produce novel findings. The other factor is the strong likelihood that students at the grade level tested would have received prior instruction or would have had an opportunity to learn the content. Some assessment items have a low depth-of-knowledge level because the knowledge required is commonly known and students with normal instruction at a grade level should have had the opportunity to learn how to routinely (habitually) perform what is being asked.

In the alignment analysis you will be asked to specify a depth-of-knowledge level for each objective under a standard (the lowest level to be analyzed). Then when coding individual items, you will be asked to identify the depth-of-knowledge level of each assessment item and record that level in the cell in the column for the item and row of the corresponding objective.

Levels of Depth of Knowledge for Mathematics

Interpreting and assigning depth-of-knowledge levels to both objectives within standards and assessment items is an essential requirement of alignment analysis. Four levels of depth of knowledge are used for this analysis.

Level 1 Recall and Reproduction

Level 1 is the recall of information such as a fact, definition, term, or a simple procedure, as well as performing a simple algorithm or applying a formula. That is, in mathematics a one-step, well-defined, or straight algorithmic procedure should be included at this lowest level. Verbs such as “identify,” “recall,” “recognize,” “use,” “compute,” and “measure” generally represent cognitive work at the recall and reproduction level. Assessment items and expectations that require students to compute a sum, difference, product, or quotient are considered a Level 1. Simple word problems that can be directly translated into a number sentence and solved by computation are considered a Level 1. Verbs such as “describe” and “explain” could be classified at different levels depending on what is to be described and explained. Some examples that represent, but do not constitute all of Level 1 performance, are:

- Recall or recognize a fact, term, or property
- Compute a sum, difference, product, or quotient
- Represent in words, pictures, or symbols a mathematical object or relation
- Provide or recognize a standard mathematical representation for a situation
- Provide or recognize equivalent representations
- Perform a routine procedure such as measuring a length
- Evaluate an equation or formula for one of its items

Level 2 Skills and Concepts

Level 2 includes the engagement of some mental processing beyond recalling or reproducing a response. A Level 2 assessment item requires students to make some decisions as to how to approach the problem or activity, whereas Level 1 requires

students to demonstrate a rote response, perform a well-known algorithm, follow a set procedure (like a recipe), or perform a clearly defined series of steps. Keywords that generally distinguish a Level 2 item include "classify," "organize," "estimate," "make observations," "collect and display data," and "compare data." These actions imply more than one step. For example, to compare data requires first identifying characteristics of the objects or phenomenon and then grouping or ordering the objects. Some action verbs, such as "explain," "describe," or "interpret" could be classified at different levels depending on the object of the action. For example, interpreting information from a simple graph, requiring reading information from the graph, is a Level 2. Interpreting information from a complex graph that requires some decisions on what features of the graph need to be considered and how information from the graph can be aggregated is a Level 3. Caution is warranted in interpreting Level 2 as only skills because some reviewers will interpret skills very narrowly, as primarily numerical skills, and such interpretation excludes from this level other skills such as visualization skills and probability skills, which may be more complex simply because they are less common. Other Level 2 activities include making observations and collecting data; classifying, organizing, and comparing data; and organizing and displaying data in tables, graphs, and charts. Some examples that represent, but do not constitute all of Level 2 performance, are:

- Specific and explain the relationship between facts, terms, properties, or operations
- Describe and explain examples and non-examples of mathematical concepts
- Describe how different representations can be used for different purposes
- Represent a situation mathematically in more than one way
- Coordinate different representations depending on situation and purpose
- Select a procedure according to specified criteria and perform it
- Formulate a routine problem given data and conditions
- Compare statements such as definitions, examples, or arguments
- Compare given strategies or procedures
- Solve a routine problem that requires some interpretation with multiple steps
- Provide an *informal* justification of one or more steps in a routine procedure

Level 3 Strategic Thinking

Level 3 requires reasoning, planning, using evidence, and a higher level of thinking than the previous two levels. In most instances, requiring students to explain their thinking is a Level 3. Requiring a very simple explanation should be a Level 2. Activities that require students to make conjectures are also at this level. The cognitive demands at Level 3 are complex and abstract. The complexity does not result only from the fact that there are multiple answers, a possibility for both Levels 1 and 2, but because the task requires more demanding reasoning. An activity, however, that has more than one possible answer and requires students to justify the response they give would most likely be a Level 3. Other Level 3 activities include drawing conclusions from observations; citing evidence and developing a logical argument for concepts; explaining phenomena in terms of concepts; and using concepts to solve non-routine problems. Some examples that represent, but do not constitute all of Level 3 performance, are:

- Analyze similarities and differences between procedures
- Analyze similarities and differences between problem-solving strategies
- Formulate an original problem given a situation
- Provide *formal* justification for the steps in a solution process
- Solve non-routine problems
- Formulate a mathematical model for a complex situation
- Analyze the assumptions made in a mathematical model
- Analyze a deductive argument, including proofs of various types

Level 4 Extended Thinking

Level 4 requires complex reasoning, planning, developing, and thinking most likely requiring an extended period of time. The extended time period is not a distinguishing factor if the required work is only repetitive and does not require applying significant conceptual understanding and higher-order thinking. For example, if a student has to take the water temperature from a river each day for a month and then construct a graph, this would be classified as a Level 2. However, if the student conducts a river study that requires taking into consideration a number of variables, this would be a Level 4.

At Level 4, the cognitive demands of the task should be high and the work should be very complex. Students should be required to make several connections—relate ideas *within* the content area or *among* content areas—and have to select one approach among many alternatives on how the situation should be solved, in order to be at this highest level. Many on-demand assessment instruments will not include any assessment activities that could be classified as Level 4. However, standards, goals, and objectives can be stated so as to expect students to perform extended thinking. “Develop generalizations of the results obtained and the strategies used and apply them to new problem situations,” is an example of a Grade 8 objective that is a Level 4. Many, but not all, performance assessment and open-ended assessment activities requiring significant thought will be Level 4. Some examples that represent but do not constitute all of Level 4 performance are:

- Develop a generalization from a mathematical situation
- Apply mathematics to model and illuminate a practical problem or situation
- Conduct a project requiring specifying a problem, identifying a number of solution paths, selecting the most effective solution path, solving the problem, and reporting the results
- Prove an original theorem
- Design a mathematical model to inform and solve a practical or abstract situation

Source of Challenge Criterion

The Source of Challenge criterion is only used to identify items where the major cognitive demand is inadvertently placed and is other than the targeted mathematical idea. Excessive reading demands, culturally bias, or specialized knowledge could be reasons for an item to have a source of challenge problem. Such items characteristics may cause some students to not answer an assessment item or answer an assessment item

incorrectly even though they have the mathematical understanding and skills being assessed. Items with an appropriate source of challenge level will discriminate between those students who have the mathematical knowledge the assessment item intends to measure from those students who do not have this knowledge.