

# **Mathematics and Science Teachers' Experiences with Assessment as an Instructional Design Process**

**by**

**Vicky Kouba, University at Albany, SUNY; Scott Stowell,  
Spokane Public Schools; David W. Champagne, University of  
Pittsburgh; Audrey B. Champagne, Susan Sherwood, Özlem  
Çeziktürk, University of Albany, SUNY**

*This paper is based on research supported in part under the Research and Development centers Program (award number R305A60005) as administered by OERI, as part of the Center on English Learning & Achievement (CELA). However, the contents do not necessarily represent the positions or policies of the Department of Education, OERI, or CELA.*

*The work reported here derives from Project 2.23 (<http://cela.albany.edu/research/project2.23.htm>), part of CELA's research into "The Role of the Academic Disciplines in Fostering Literacy in Middle and High School." This paper was presented in April 2001 at the AERA Annual Meeting."*

# **Mathematics and Science Teachers' Experiences with Assessment as an Instructional Design Process**

*"I feel more confident when interpreting the ... standards. My ability to set criteria, design appropriate tasks and develop scoring based on these standards was non-existent two weeks ago. Today, I have set criteria, modified tasks and developed scoring guides all based on two related ... standards." (East Coast teacher remarks)*

## **Introduction**

Changing the usual way of doing anything is a difficult process especially if one has years of experience with the "usual" rather than the "new" way. In our work on altering perspectives on assessment, teachers on the east and west coasts of the United States reported that using assessment as a tool for instructional design is always challenging, initially difficult, and often frustrating. However, as indicated in the above remark, many teachers who work to change their use of assessment from a narrow, end-of-unit and end-of-year summation to assessment in the service of improving the design and on-going refinement of instruction find the process worth the time and effort.

In the summer of 1998 we embarked on a program to help teachers use assessment as a means to improve the design of standards-based instruction in science and mathematics. What we report here is exploratory in nature and addresses the following goals:

To--

- ?? Describe the general nature of mathematics and science teachers' knowledge of assessment terms;
- ?? Document science and mathematics teachers' experiences in professional development courses or programs that focused on the use of assessment in the design of instruction; and
- ?? Identify science and mathematics teachers' stated concerns and attitudes regarding the use of assessment in the design of instruction.

## **Background and Perspective**

The use of assessment to ensure a better instructional design is not a new idea. Gagne, Briggs and Wager (1992) stated that one of the main purposes of assessment is to evaluate and improve the instruction itself. What is relatively new is the movement to engage teachers in using assessment as a means for interpreting standards, which includes:

- ?? aligning standards and curriculum;
- ?? identifying evidence that standards have been met; and
- ?? designing instruction to insure that students have an opportunity to learn what is needed to produce convincing evidence that they have achieved the standards.

Wiggins (1998) documents historical and current thoughts on the use of assessment in the design of both instruction and curriculum, and provides an excellent bibliography for further reading in the area of what he calls "educative assessment." Our efforts with the east and west coast teachers were carried out in a similar spirit and with the same beliefs that Wiggins espouses

-- that an "educative assessment system is designed to teach" and "to provide useful feedback..." (p. 12)

Our philosophy may best be summarized by the phrase, "assessment in the service of instruction." Wiggins and McTighe (1998) call the approach of using assessment as a tool for design, "backward design," and recommend it as a means for avoiding "common inadequacies in curriculum and assessment planning." (p. 3). The process of using assessment as the guide for the design of instruction is "backward" because it begins with identifying the end goals to be achieved, which are based on national or state standards and can be perceived as an interpretation of the standards. Then the acceptable evidence or criteria for judging the extent to which a standard or goal is achieved is determined. The final stage is planning learning experiences and instruction that provide all students the opportunity to learn what will enable them to generate the acceptable or convincing evidence. In summary, professional development from the perspective of making the best use of both formative and summative assessment is a multi-step process consisting of

- 1) consideration of the standards or goals students are to achieve;
- 2) identification of what teachers will accept as evidence that the students have achieved the standards or goals;
- 3) identification of the criteria for interpreting and valuing the expected evidence; and
- 4) design of the instruction that will provide students with both the knowledge base and opportunity to provide the evidence.

Traditional education has long treated assessment as the last step in the design of curriculum and instruction, rather than the first and guiding step. Also, in our experience with teachers, we found that most equated "assessment" primarily with "testing." Both of these aspects suggest that to ask teachers to think differently about the role of assessment in the design of instruction is to ask them to make a conceptual change. Research indicates that conceptual change is difficult for adults, and requires careful guidance in professional development experiences (e.g., Borko, 1997; Loucks-Horsley, Hewson, Love, & Stiles, 1998; Birman, Desimone, Porter, & Garet, 2000). However, we had no hard data on teachers' experiences in a professional development venue that asked them to make deliberate use of assessment as a more formative tool in the design of instruction. We felt that a descriptive exploration of teachers' experience in such a venue was a necessary step in our understanding of how to help teachers become more effective when teaching science and mathematics.

### **Mode of Inquiry and Evidence**

Our investigation consisted of two phases. Phase 1 involved 11 science and mathematics teachers from 11 different school districts on the east coast of the United States. The teachers were students in an assessment and instructional design course that met for 45 hours across 8 days over a two-week period in the summer. Phase 2 involved 19 science and mathematics teachers from a single school district on the west coast of the United States. These teachers were involved in a professional development program that met for one to two weeks each summer over five years, with follow-up support and meetings throughout each school year.

## Phase 1

The instruction for the East Coast group of teachers centered on a "backward design" of instructional design in which teachers began with state and national standards and planned a curriculum and instruction unit that would provide students the opportunity to generate (using written extended response types of tasks) evidence of having attained state or national standards.

- ?? Days 1 and 2 of the course had the following focus: Overview of the course; background and context of assessment and of standards as targets of assessment; identification of criteria for forms of scientific and mathematical explanations and relationship to content expectations.
- ?? Days 3 and 4 had the following focus: Process of moving from standards to extended written response items; examples and plans for a unit; analysis of criteria for assessing; analysis of student responses and refinement of criteria; reflection on how extended response assessment fits into larger assessment plans.
- ?? Days 5 and 6 had the focus: Reflection on district/school assessment plans and relationship of extended response assessment to those; continued use of examples for analysis of criteria for assessment; analysis of student responses and refinement of criteria and plan for achieving criteria.
- ?? Days 7 and 8 had the focus: Implications of use of extended response assessment for classroom practice; final development and refinement of unit.

The class materials from the multi-district East Coast Group used in the Phase 1 analysis included:

- ~~///~~ a questionnaire on knowledge of assessment terms,
- ~~///~~ daily response/feedback forms,
- ~~///~~ weekly reflective journals,
- ~~///~~ a report on home district assessment plan, and
- ~~///~~ a culminating assessment planning project.

All of the materials were read and analyzed by two investigators who identified 10 recurring themes across the materials and across the individual teachers.

The materials from the course were re-read and coded according to ten themes (which are not mutually exclusive):

- ?? Language of Assessment,
- ?? Understanding of National and State Standards;
- ?? Teacher Independence;
- ?? Working with Standards;
- ?? Importance of State Exams;
- ?? Assessment Planning Process;
- ?? District Assessment Plan; and
- ?? Changes in Attitude, Main Concerns, and Areas of Difficulty.

## Phase 2

The analysis of Phase 1 materials was used to redesign the questionnaire on knowledge of assessment terms and attitudes or beliefs on different aspects of assessment and the use of assessment in an instructional design process. The questionnaire was given to the West Coast teachers in the second year of a five-year extensive professional development experience partially focused on the review of instructional materials and the development of assessment items. The questionnaire was ancillary to the goals and materials of the West Coast teacher enhancement/professional development project. The West Coast group was chosen as the follow-up or “check-point” group because a portion of the groups’ professional development experience involved the use of assessment to design and improve instruction and to design appropriate open-response or extended response items. Other portions of the West Coast project focused on the alignment of standards, goals, curricula and assessment; and the development of a district science learning framework; the improvement of classroom-based assessment and teachers content knowledge. The results of that questionnaire were analyzed and used to elaborate on aspects of the results from the East Coast teachers.

### **Results and Discussion**

#### Language of Assessment

As part of the first day activities, the East Coast teachers were given a questionnaire designed to reveal their knowledge of particular terms used in assessment and in the design of instruction. They were asked to define the terms:

- ~~✍~~ alternative assessment,
- ~~✍~~ authentic assessment,
- ~~✍~~ performance assessment.

They also were asked to explain how the “testing” and “assessment” were related. After they had written their responses to these and other items, the responses were used as the basis for a discussion of terms and the role of assessment in the design of instruction. At the end of class, the teachers were asked to write in their response journals their reactions to what had occurred.

Our analysis of the East Coast questionnaire responses revealed that the teachers did not have common definitions of assessment terms. The responses given for the terms “alternative assessment,” “authentic assessment,” and “performance assessment,” reveal that some of the variation in definitions is rooted in a genuine lack of experience with or knowledge of assessment terms. As shown in Table 1, although the teachers agreed that “alternative assessment” is some “other” kind of assessment, they differed on whether it was an alternative to traditional multiple-choice, true/false or short answer tests; to a standardized test; or to testing in general. The teachers also did not demonstrate uniform knowledge of what the alternative might be. Definitions for “authentic assessment” and “performance assessment” showed a dearth of experience with assessment alternatives and with terms used in assessment.

Table 1 East Coast Teachers' Definition of Assessment Terms

<b>Teacher</b>	<b>Alternative Assessment</b>	<b>Authentic Assessment</b>	<b>Performance Assessment</b>
TE1	"Some type of assessment other than your standard multiple choice, true/false, etc. Something different."	"A type of alternative assessment which involves "hands-on" and/or real life application of knowledge."	"A type of authentic assessment which involves mainly "hands-on."
TE2	"Other than traditional tests for purpose of getting richer more accurate testing."	"Real life."	"A real life activity."
TE3	"Range of assessments other than traditional written exams to assess multiple ways in which learning takes place."	"Assessment aimed at tapping various models or ways in which we learn, i.e., running a business."	"Doing tasks, evaluating problems, designing tasks as assessment rather than spitting back information."
TE4	"Using method other than testing to discover what a student has learned."	"Similar to alternative assessment, uses real life application."	"A method in which students show or demonstrate the skills they have learned."
TE5	"A way in which to evaluate student understanding different from norm (i.e. tests)"	"A way in which to accurately assess student understanding."	"A way in which to evaluate students' performance of tasks."
TE6	"Using different means to gather information about students' progress and learning – instead of using multiple-choice, standardized test – grasp information about students using portfolios, writing samples, running records, etc."	"Using samples of students' work throughout the year – portfolios, writing samples, tests that show students progress and learning – more holistically centered not focusing on right vs. wrong but process and skills."	"Using written and performance based tasks where students show their learning and skills through the different tasks, i.e., manipulation portion of state test."
TE7	"Assessment of students' experience in non-traditional methods."	--blank	"Hands-on."
TE8	"An assessment other than traditional (matching, true/false, short answer, multiple choice. Alternative assessment adds to the traditional."	"Assessing a real life or some aspect of real situation set up for assessment purposes."	"An on-going assessment of development"
TE9	"A different way of evaluating materials or course."	"Using data/materials of testing to perform an assessment."	"Using student materials like a portfolio to use in assessment."
TE10	"Evaluation of acquired knowledge that is not based on multi-choice, short answer response for more accurate, richer, more useful assessment."	"Evaluation based on true understanding of subject not rote response."	"Evaluation based on actual activity such as a lab experiment."
TE11	"Other than standardized testing – portfolios, etc."	"Assessment in relevant situation; true; unbiased."	"Performance of tasks or activities students are engaged in, and extended task including psychomotor skill being evaluated."

Differences existed even on items where we expected agreement. For example, on the comparison of testing and assessment, we expected teachers to agree among themselves that testing is a form of assessment often used to assign grades. Some teachers said just that:

*“Testing is one type of assessment.”*

*“Testing is one form of assessment.”*

*“Test is a method of assessment.”*

*“Assessment is determining what and how much a student has learned.*

*Testing is one form of assessment.”*

*“Testing is one form of assessment used for grading. Assessment is more general and used for diagnostic purposes.”*

Others gave responses that, although vague, seemed to imply a meaning similar to what we expected:

*“Testing – scoring purposes”*

*“Testing is used in assessment.”*

However, some teachers seemed to view tests as the means by which to do assessment:

*“Tests are used to assess student performance.”*

*“Assessment is the information gained from using different tests and tasks”*

And two teachers gave responses that seemed to view assessment and testing as related but separate:

*“Both have similar meanings. Ways to determine the level of understanding in the students.”*

*“Assessment – what you do know; Testing—what you don’t know.”*

The teachers’ responses also differed in that some attempted a definition of assessment, whereas others used the term but did not define it. To determine what teachers would say if directly asked to define assessment, on the Phase 2 West Coast questionnaire, we asked teachers to define assessment. Because they had already been working together on assessment, we expected some uniformity of definition. Although the responses had a similar focus – that of looking at student understanding-- teachers varied on their definition of the nature of assessment:

~~✍~~ Some teachers viewed assessment as a **process**:

- *“The method by which we determine students’ understanding and knowledge;”*
- *“A procedure to gather achievement and performance data regarding student attainment of intended learning targets;”*

whereas others viewed as **the information itself**:

- *“Lots of information about student learning: can be verbal, written, diagrams, etc;”*
- *“Data collection about student learning).*

~~✍~~ Some teachers viewed assessment as a **measurement**:

- *“A way of measuring what has been learned;”*

whereas others viewed it as **an estimation**:

- *“An estimation of how well students understood material.”*

~~✍~~ Some viewed assessment as an end **evaluative yes/no means of determining whether a goal had been reached**:

- *“An educational tool used to collect data on students to determine whether or not they have achieved a certain learning goal;”*
- *“A way to examine the results of individual knowledge and concepts on learned information on relevant or any topic;”*
- *“An evaluation of the ability of the students to understand and apply the principle concepts of the subject taught;”*
- *“A method of evaluating student attainment of learning targets;”*

whereas others viewed it as a means to evaluate the **level, extent, or degree of understanding**:

- *“A method used to discover the level of understanding the students have attained;”*
- *“A means of determining the degree to which a student understands a particular concept;”*
- *“To determine the level of understanding a person has about certain knowledge;”*
- *“A means of discovering how a student has grasped pertinent information and the degree of integration and understanding attained.”*

~~✍~~ Some teachers gave **short or unclear** definitions:

- *“Student feedback; a check on understanding;”*
- *“Collecting data that will validate student learning/understanding;”*

whereas others gave **elaborate** definitions:

- *“A tool used to check a student’s understanding of some particular material. May be used to assign grades but may not be graded at all;”*
- *“A mechanism by which student understanding can be evaluated—may include a range of means such as teacher observation, examination, daily work, etc;”*
- *“Assessment is being able to check kids’ abilities to know, understand, and be able to do tasks in compliance with standards according to individual curricular areas. May be formative, summative, or diagnostic in nature. Also may be performance based.”*

The East Coast teachers’ comments in their response journals raised some important points regarding the language of assessment and teachers’ experience with assessment. First, many teachers may not be aware of the extent to which their definitions and those of others differ. Although the East Coast teachers expressed at the beginning of the discussion of their definitions of terms that they expected some differences, the differences were more profound and more varied than they expected:

- *“I am amazed at the different definitions that are floating around for educational vocabulary;”*
- *“The discussion on comparing words was very interesting – just when I thought it was cut and dry – it was not.”*

To check our belief that teachers expect some disagreement over terms, we asked the West coast teachers to respond on a Likert scale (strongly agree to strongly disagree) to the statement, “Educators do not have a common understanding of assessment terminology.” Eleven of the

nineteen teachers agreed with the statement, and the eight others strongly agreed. The same distribution of agree and strongly agree ratings occurred with the statement “It is difficult to collaborate when there are different interpretations of terms.” Three of the eleven East Coast teachers also remarked in their journals that they found it hard to collaborate when the teachers involved in a cooperative effort have differing interpretations of terms.

Part of the difficulty in collaboration when there are fundamental differences in the understanding of terms may lie in the fact that there are at least two types of teachers among those trying to collaborate. There are those who are fearful of revealing that they hold a different understanding; and there are those who are confident that although there are differences in definitions or understandings, it is their particular perspective that is the “true” understanding. The East Coast journal comments revealed the existence of those teachers who may be uneasy at doing something that might reveal they lack understanding. They think they are alone in their confusion (e.g., “*Many of the ‘key words’ in education are not broadly understood. It thought I was the only one who was unclear on these ‘easy’ terms. It was great to discuss the discrepancies and gain a new perspective;*” “*I would like to review the questionnaire we did at the beginning of class. I was not familiar with many terms.*” [This entry was made prior to the discussion of the terms.]) The West Coast questionnaire responses revealed the existence of the confident group. We asked the West Coast teachers to rate (on a five-point scale) their confidence in their understanding of the meanings of certain assessment terms, including “alternative assessment,” “authentic assessment,” and “performance assessment.” Seventy-two percent of the possible 57 responses were in the **confident** or **reasonably confident** rankings, twenty-one percent were **unsure** level, and only seven percent were in the **doubtful** or **no confidence at all** levels.

It was clear from the East Coast teachers’ response journals that they found value in discussing the definitions and meaning of terms. Some teachers expressed a belief that such discussion facilitated group work on the standards-based, assessment-based instructional design they were doing in class. Others saw the discussions as helpful in more individual ways:

- “*The comparison and definitions of terms was somewhat eye-opening for me. I had very little background on all those buzzwords!*”
- “*Once again, the variety of wording and meaning is extreme. It seems that there is a necessity for unified terminology for communication between groups to be most effective.*”
- “*The convergence/divergence on terms in groups was extremely useful – important to have clarity on terms.*”
- “*I’m glad that we decided to throw out some terminology. All too often we let terminology hinder our efforts.*”
- “*The first week of this course has exposed me to some terms that I felt were familiar to me such as assessment, standards, and extended responses. I have found that there is much room for interpretation about how we define these terms, and how I will need to better understand them as the new state changes come about.*”

A second important issue that emerged from our analysis of the East Coast teachers' response journals was that discussion of terms, or writing about the definition of terms, reveals important gaps in some teachers' understanding of assessment. One of the steps in working backwards through assessment to design instruction is to identify the criteria upon which to judge whether students have met an objective or an outcome. When the East Coast teachers were asked to identify criteria or indicators that they could use to judge the level to which students were meeting both the objectives of a unit and the state standards related to the unit, the teachers struggled. The level of frustration in the class that day was higher than the either of the instructors was used to experiencing. For the most part, the teachers seemed to refuse to write criteria. We did not understand what had really occurred until we read the journal responses to a prompt asking the teachers to describe the relationships among objectives, criteria, and standards. We found that many of the East Coast teachers did not distinguish criteria from objectives or did not seem to understand that criteria or benchmarks are a means by which we can systematically judge to what level students are meeting standards:

- *"To me, I think objectives and criteria are the same."*
- *"Not sure about this one [criteria] but I think I use it to mean a definite set of expectation. (I would really like to try to concept map all these terms but I'm not sure I could, at this point.)"*
- *"I really don't see the difference between outcomes and criteria."*
- *"Criteria to me in the whole sense of the word means guidelines or bits of information to be embedded into a project."*
- *"Today's task is very difficult. I don't feel prepared. It may be because I don't have a clear understanding of the relationships of the terms [criteria, standards, assessment]. I really don't see the difference between objectives and criteria. I feel like I am hunting for criteria by trial and error. Also, I suspect there may be more than one valid interpretation of criteria."*

Just as the East Coast teachers struggled with setting criteria, so, too, did the West Coast teachers struggle with setting benchmarks and indicators of essential academic learning requirements (EALR). As shown in Table 2, the West Coast teachers had varied, often murky, distinctions among the three. With both groups of teachers, the professional development facilitators had to revisit and work on building the teachers' conceptual understanding of criteria and benchmarks. Unfortunately, we did not collect data on the teacher's definitions after further work on the concepts underlying the terms. We did note that both the identification of criteria and benchmarks remained a challenge for the teachers.

The East Coast teachers engaged in a task designed to work toward a community consensus on the definition of terms. Teachers were placed in groups of three and asked to look at their descriptions of the relationship between paired words: assessment and testing; item and task; scoring guide and rubric; curriculum and instruction; teaching and learning; and standards and objectives. They were asked to record where their descriptions converged and diverged. The three instructors also worked as a group to do the activity. A whole-class discussion followed the group work. The goal for the whole class discussion was to come to a consensus of some kind about the meanings of each pair of terms. The results of the group work and the whole class discussions are summarized in Table 3. Although a consensus was not always reached and the

**Table 2** West Coast Teachers' Definitions of Terms Used in Their Extended Professional Development Experience: Essential Academic Learning Requirement (EALR), Benchmark, and Indicator

<b>Teacher</b>	<b>EALR</b>	<b>Benchmark</b>	<b>Indicator</b>
TW1	"State committee generated description of what kids need to know, understand and be able to do in any given curricular area."	"Level (1, 2, 5) of EALR's that correlate to academic progression. For science, benchmark 1 = grade 5, 2 = grade 8, 3 = grade 10"	"Elaborations inside EALR's that unpack the general EALR statement."
TW2	"Definition of curriculum"	"expectations of understanding at specific grade level (i.e., 4 <sup>th</sup> , 8 <sup>th</sup> , 10 <sup>th</sup> )"	"Some measure which can yield evidence that a specific topic has been understood."
TW3	"Statement which defines a fundamental idea or concept. This concept will serve as a base line for understanding within a certain discipline."	"Part of the EALR that is more specific and can be specialized to a particular discipline."	"What behaviors the student will demonstrate to show their understanding of a benchmark/EALR."
TW4	"The State set of objectives that all students receiving a Certificate of Mastery shall have learned. (yeah, right!!)"	"The specific objectives outlined for each level (approximately 5 <sup>th</sup> , 8 <sup>th</sup> and 10 <sup>th</sup> grade) under each EALR"	"Elaboration?"
TW5	"The learning target (goal) that all students should understand – Large concepts such as 'How the natural world works.'"	"Breaks the EALR into different levels of understanding"	"Elaborations – details of what students should understand (demonstrate) if they have met the benchmark."
TW6	"What the State has decided all students need to be able to do."	"At grade levels 4, 8 and 10, what depth of understanding students need to have attained."	"The key, focal points which will tell me whether the student is on track."
TW7	"Level of knowledge and understanding expected at certain levels of education."	"Minimum levels of understanding of certain knowledge."	"?"
TW8	"A summary statement about intended learning targets in given area of the domain."	"A point in time (5 <sup>th</sup> , 8 <sup>th</sup> , 10 <sup>th</sup> ) at which the accumulated science learning should enable a student to attain the intended targets at the benchmark level."	"Indicator/learning target level – what students are intended to learn in a given domain. Concepts & principles, reasoning/ problem solving abilities, constructs and connections in the domain of science."
TW9	"A specific learning goal set forth by the state. An EALR addresses a general piece of subject information. The body of EALR's should, hypothetically, cover all of the learning goals for students in a particular subject area."	"This is still a little hazy in my mind. A subcategory of an EALR: the benchmark is 'evidence' that a student has learned a particular concept"	"This is still a little hazy in my mind. A subcategory of a Benchmark: the indicator is a more specific piece of evidence that a student has learned a clearly-defined portion of a Benchmark."
TW10	"Big ideas --- the level of understanding that students should attain. In our case determined by the State."	The EALR's broken down to grade level. More specific..."	"More specific --- learning objectives that address the benchmarks --- these are the teaching objectives that teachers use to focus learning."

Table 2 continued

TW11	"Requirements of concepts that a student must understand to be competent in the world today as seen by 'the group.'"	"Goals students should reach by a certain time."	"Ways to determine if student knows Benchmark by understanding knowledge and concepts of certain tasks."
TW12	"A set of specific learning goals that all students are expected to master. In our state, these are written by a commission at the state level."	"An assessment given at specific times in a student's academic career to collect data on whether the student is making academic progress in relation to essential learnings. Teachers can use data to determine if, what, and how they are teaching is allowing students to progress satisfactorily."	"A specific state of what a students knows or is able to do that demonstrates satisfactory mastery of an EALR."
TW13	"A state (or district) state of 'their' selection of small parts of subject area that will be evaluated in some way for the purpose of state level competency tracking."	"A statement of key ideas or portions of a subject unit that should be addressed in teaching each subject area."	"A subject oriented short summary of the key components of a topic that should be included and covered in a unit."
TW14	"State mandated learning all students are to be given the opportunity to attain."	"Divides the EALRs into grade-level appropriate learnings."	"Elaboration of the EALR's and benchmarks, more detailed and specific."
TW15	"What a student must know about a subject in order to be recognized as competent."	"A single knowledge item or concept that schools need to teach towards."	"A statement that outlines what successful completion of EALR's by students should look like."
TW16	"A standard set by a legislative/governing body indicating what all students should know and be able to do after their academic experiences are completed."	"A specific level of understanding or attainment of some level of proficiency."	"Some specific evidence that a student has attained a specific benchmark."
TW17	"A specific statement that discusses what the student should know or be able to do in reference to a specific subject."	"A more general statement than an EALR, a benchmark is a statement that describes the level of understanding a student should have after completing the EALRs."	"A task used to show where the students are in their level of understanding for a particular subject."
TW18	"Overall statements of what students should be able to understand and do."	"Level of understanding required in a general sense at certain grade levels."	"Particular skill a student will demonstrate to satisfy benchmark and EALR."
TW19	"A short list of very general statements created through the Commission on Student Learning that are the basis for the state exams and will drive instruction."	"A set of statements at three grade levels that specify what students should know, understand, and be able to do to meet the EALR's"	"An elaboration of a benchmark that gives more detail."

Table 3  
Convergence, Divergence, and Consensus on Definitions of Paired Terms: East Coast Group

Term and Group	Summary of Results
Assessment and Testing	<p><u>Group 1</u> -- Agreed that both were evaluation tools used to gain information. They thought their definitions differed in that some thought assessment shows what you know and testing shows what you don't know, whereas others thought that testing is just one piece of an assessment.</p> <p><u>Group 2</u> -- Agreed that assessment was most inclusive, with testing as a type of assessment. They agreed that assessment is a more global term that covers a larger range of ways students learn than does the term testing. They identified no areas of divergence.</p> <p><u>Group 3</u> -- Agreed that testing is one form of assessment, which is the determination of what and how much students have learned. They had no area of divergence.</p> <p><u>Group 4</u> -- Agreed that testing is a form of assessment and had no area of divergence.</p> <p><u>Instructors</u> -- Agreed that assessment was the more general term and that testing was a form of assessment. They disagreed on whether testing was diagnostic or was used only for summative purposes.</p> <p><u>Whole Class Consensus</u> -- Agreed that assessment was the more global term and that testing was one form of assessment. Could not come to consensus on assessment beyond that it was "getting information" to be used in some way.</p>
Item and Task	<p><u>Group 1</u> -- Agreed that both were parts of an assessment. Disagreed as to whether an item differs from a task in that an item has one answer, whereas a task may have more than one possible answer. Also some group members wanted to restrict the definition of a task as having a solution using a process.</p> <p><u>Group 2</u> -- Agreed that a task can be a series of items. Members were split on whether a task is a type of item, meaning not all items are tasks.</p> <p><u>Group 3</u> -- Agreed that they had no idea of any difference between the two, but after some discussion agree that an item is a specific question and a task ia an activity that finds the answer. They identified no area of divergence.</p> <p><u>Group 4</u> -- Agreed that a task was a job or skill to be performed. They could not agree on a definition for an item. Some thought an item was an idea to gain or teach, whereas others thought an item was an assessment question.</p> <p><u>Instructors</u> -- Agreed that items and tasks were different, but could not come to agreement on the nature of that difference. Some thought an item must have only a short response, whereas others thought an item could also have an extended response. One thought that a task must involve something physically carried out, one thought it had to have an extended response, and the thought that item was a global term for a standard performance.</p> <p><u>Whole Class Consensus</u> -- After several attempts at consensus, the class agreed to disagree, with some maintaining that task and item were the same whereas others thought they were different, but could not agree on characteristics of those differences.</p>

Table 3 continued

<p><i>Scoring Guide and Rubric</i></p>	<p><u>Group 1</u> -- Agreed that they were tools used for evaluation and both provided specific expectations for students and served as outlines for students. They also agreed that the rubric gives more feedback to a student than does a scoring guide. They had not identified area of disagreement.</p> <p><u>Group 2</u> -- Agreed that a rubric is a specifically structured array of criteria with specific grading, whereas a scoring guide is a general aide to looking at responses, which does not provide a model of acceptable answers. They had no divergence.</p> <p><u>Group 3</u> -- Could not come to any agreement. Some members believed that rubrics are for subjective material, whereas other members thought rubrics could be used for any items or responses. Definitions of scoring guides ranged from assessment with point values, to no different from rubrics to any answer key.</p> <p><u>Group 4</u> -- Agreed that scoring guides and rubrics had similar definitions and similar uses. They saw both as tools of assessment that are developed prior to student activity.</p> <p><u>Instructors</u> -- Agreed that the terms were similar, but that rubrics had exemplars of expected responses and scoring guides did not.</p> <p><u>Whole Class Consensus</u> -- Agreed that a scoring guide and a rubric were means of evaluating responses, but could not come to consensus on other aspects.</p>
<p><i>Curriculum and Instruction</i></p>	<p><u>Group 1</u> -- Agreed that curriculum is what instruction is based on and curriculum is an outline for instruction. Had no areas of divergence</p> <p><u>Group 2</u> -- Agreed that curriculum was the overall design and content of an area to be taught, whereas instruction was the mode or delivery style. Had no areas of divergence.</p> <p><u>Group 3</u> -- Agreed that curriculum is the material and instruction is the process of teaching. Instruction is based on curriculum. Had no areas of divergence.</p> <p><u>Group 4</u> -- Agreed that instruction is used to get curriculum across and that curriculum influences instruction. Had no areas of divergence.</p> <p><u>Instructors</u> -- Agreed that curriculum is the whole package in a content area and instruction is the way to use or actualize the curriculum. Diverged on the role of students in instruction, with one viewing students as instruments carrying out instruction and others viewing instruction as not something students engaged in, but that teachers engaged in.</p> <p><u>Whole Class Consensus</u> -- Agreed, in a broad sense, curriculum is, at least, <u>what</u> is taught and instruction is <u>how</u> it is planned and taught.</p>

Table 3 continued

<p><i>Teaching and Learning</i></p>	<p><u>Group 1</u> -- Agreed that teaching and learning go hand in hand. We teach so that others will learn, and in the process we learn as we teach. No areas of disagreement.</p> <p><u>Group 2</u> -- Agreed that learning is the internal structuring of knowledge, whereas teaching is one mode of facilitating learning. No areas of disagreement.</p> <p><u>Group 3</u> -- Agreed that teaching is the presentation and explanation of skills, curriculum, etc. where as learning is the understanding, acceptance, and usage of skills and curriculum. Teaching facilitates learning and there is a fine line between the two. No areas of disagreement.</p> <p><u>Group 4</u> -- Agreed that learning is the product of effective teaching. No areas of disagreement.</p> <p><u>Instructors</u> -- Agreed that teaching is the organizing and guiding of learning; and that learning takes place internally in students. No areas of disagreement.</p> <p><u>Whole Class Consensus</u> -- After long discussion, agreed that teaching and learning are related and that teaching, by whomever and whatever, results in learning, which is internal to students.</p>
<p><i>Standards and Objectives</i></p>	<p><u>Group 1</u> -- Agreed that standards are the overall goals to be met and objectives are specific indicators to grade levels and ages. No areas of disagreement.</p> <p><u>Group 2</u> -- Agreed that standards are expectations that are broad whereas objectives are specific goals or means of meeting expectations. No areas of disagreement.</p> <p><u>Group 3</u> -- Agreed that standards are basic skills and knowledge needed for success. Objectives are plans to meet those standards. No areas of disagreement.</p> <p><u>Group 4</u> -- Agreed that standards are broader than objectives. Disagreed on the process of arriving at standards and objectives. Some thought that one begins with objectives and puts them together to form standards, whereas others thought that one begins with standards and interprets them by writing objectives.</p> <p><u>Instructors</u> -- Agreed that standards are global and broad, whereas objectives are more specific. Disagreed on the use of objectives. One viewed objectives as just for classroom use by teachers and others viewed objectives as useful in situations beyond the classroom.</p> <p><u>Whole Class Consensus</u> -- Agreed that standards are more general than objectives.</p>

consensus statements that were reached were quite general in nature, the teachers found the discussion itself the important part of the process. The teachers who were not willing to give up their personal definitions were willing to listen to the viewpoints of others. The teachers strongly encouraged each other to state definitions and perspectives as part of any further discussion, and also read subsequent course materials with a more critical eye, seeking to identify the definitions and perspectives of the authors.

Knowledge about the concepts behind assessment terms and practices is often essential to teachers being able to design instruction based on assessment, and to teachers being able to work together effectively even when they do not share common definitions or perspectives. Our exploration of the teachers' definitions of terms used in assessment supports the value of giving time and attention to the language of assessment as part of professional development, both as a means of identifying gaps in knowledge and as a means for filling in those gaps. Some of the East Coast teachers' comments in their journals near the end of the course indicate what they gained from a discussion of terms:

- *"I found that the problem solving activities I do in my class really do not characterize themselves as extended task responses."*
- *"My short comings came from poor assessment goals and techniques. I taught in response to ONE understanding level and ONE assessment technique."*
- *"You need to have criteria set up to make decisions. If you don't have criteria, it makes it hard a) to define a task (projects and content) and b) to evaluate the project."*
- *"Difficult and beneficial, could become frustrating. We encountered different interpretations and very different views of how to meet the standards. Talking about these was difficult because we were different, however, I think these different interpretations are necessary to really understand and use the standards."*
- *"The discussions have created community among us and a support system (albeit a temporary one) in which we can hear other options and approaches to problems and situations."*
- *"It's starting to come together. I understand the process we need to follow to ensure uniformity. (standard > evidence (criteria)>tasks). However, I still have trouble writing criteria."*
- *"It was interesting to hear different interpretation of what different terms mean. Our work and discussion have clarified that there are great inconsistencies in our system that greatly affect the efficiency."*

### Experiences, Concerns and Attitudes

After reading all the data sources from both The East Coast and West Coast groups, two investigators identified themes that seemed to emerge across all the sources. An analysis of teacher comments and responses in light of those themes gave us some insight into the experience, concerns and attitudes the teachers had as they engaged in professional development courses or programs that focused on the use of assessment in the design of instruction.

Understanding Standards. Because both the East Coast and the West Coast professional development experiences focused on using assessment to move from standards to the design of instruction, understanding standards and working with standards emerged as themes in the teachers' responses.

More than three-fourths of the East Coast teachers had read state mathematics and science standards and had studied them in some other workshop or course. However, only thirty-six percent had used the standards in planning one of their own courses and only eighteen percent had considered the standards in the planning of all of their instruction. The majority of both groups (East Coast 73%, West coast 89%) thought that aligning assessment with standards is a difficult process when done at a level of detail greater than just superficially matching a standard (via a checklist) to an activity. Eight-two percent of the East Coast teachers thought the state standards were difficult to understand. The West Coast teachers, for the most part thought they had a good understanding of their state standards, with thirty-two percent indicating that they were undecided.

One of the greatest difficulties many of the East Coast teachers had was that of starting from standards rather than starting from activities or tasks. The generality of standards and the lack of an overview in terms of what could be done specifically at each grade level to work towards achieving the standards seemed quite difficult. For example, teachers made the following comments.

- *“In working with the standards, I felt somewhat overwhelmed at first, trying to take a standard and putting [it] into objective terms.”*
- *“It was VERY difficult for me to take the standards and begin to decipher them, come up with and communicate its intent in terms of criteria for assessment.”*
- *“I find it difficult to break down the standards into tasks and then to develop criteria for a response. It is hard to know how much detail is needed.”*
- *“I’m finding the task of changing the way I think and process is very frustrating and challenging. Just when I think I’ve gained an understanding with this assessment project, I hit an obstacle. I’m having a difficult time relating or finding a task to include my criteria more specifically.”*

One teacher admitted giving up and starting with tasks, and another indicated that he or she would still work from the tasks, but with a consideration of standards:

- *“This task (process) is very difficult. I found myself trying to work backwards from the task to the standards, in hopes that some of my tasks (used currently) can be used to meet these standards.”*
- *“I can’t see myself making new tasks on a daily basis, but rather adjusting tasks to fit the standards, and my students’ needs.”*

Another teacher clearly expressed what many of the teachers were struggling with, trying to change from just doing a series of interesting activities that might map to standards and objective to actively using standards and objectives with stated criteria to design activities:

- *“I think most teachers, including myself at times, design lessons and activities which produce student work. We then collect this work and sit at our desks with this stack of papers and say, ‘OK, now I have to come up with a way to grade these.’ Clearly this is a backwards approach although it doesn’t appear backwards to many teachers. In fact, it even seems to make sense.”*

By the end of the course, this teacher felt progress had been made:

- *“I’m becoming more comfortable with outlining criteria. I’m beginning to see how we can establish some middle ground between general standards and specific tasks. This was an area of stress for me this past weekend. I’m still a bit confused but I feel as though I am inching along, fighting my way through thick brush. It’s a real battle and requires a lot of effort and energy but I do feel that I/we are making progress.”*

Many of the teachers, even those who had already had professional development or workshops with the standards, felt the need for continued work on interpreting standards and expressed concerns over the ability of teachers to help students meet standards that the teachers themselves did not understand in depth:

- *“We have a long way to go for understanding the standards.”*
- *“I realized the depth of interpretation was far greater than I had originally thought.”*
- *“The [state] standards are a step in the right direction, but teachers, for the most part, have trouble in interpreting them and will therefore have trouble meeting them.”*
- *“What this meant for teachers, was that we had to become familiar with what the standards are and then adjust our curricula accordingly. I do not believe this actually occurred in every department or at every grade level.”*
- *“In my district I really feel that there need to be some workshops and inservices devoted to the standards.”*
- *“The main request I would have for my district is to provide teachers with more time to look at the standards.”*
- *“If it is so difficult to interpret the standards, how are the majority of teachers going to meet them?”*

There was a dissenting group of teachers who felt that they could do an effective job regardless of the depth of their understanding of the standards or who indicated that the curricula they were working with had been defined and written [by a committee that looked at the standards] and that that was sufficient to change the instructional approaches:

- *“I don’t feel I need a total understanding of the standards to work with them.”*
- *“In the middle school math department, the curricula has been defined and written at the elementary and middle level according to the standards and the shift in teaching is starting to occur.”*

One of the reoccurring comments was that the process of interpreting and achieving standards is not something that teachers can do alone or in isolation. As one teacher said, *“The problem of*

*bringing about a change in the classroom can not be isolated from the school district's incentives to change."* In fact, the teachers had a great deal to say about the role of districts, as well as the role of state exams and how they felt about challenges to their independence as teachers.

District Assessment Plans, State Exams, and Teacher Independence. As part of working with assessment to design standard-based instruction or in working to develop reasonable standards-based tasks, many of the East Coast teachers expressed dismay over district assessment plans and alignment between state assessment, classroom activities, and standards. Although eighty-four percent of the West coast teachers thought that the district was working to implement standards, forty-two percent thought that there was a lack of communication within the same grade level or department. Fifty-six percent of the East Coast teachers observed that there was a lack of communication and sixty-four percent thought there was a lack of unity in the district regarding standards. The teachers who reported a lack of unity or communication made the following comments:

- *"There is little or no contact between member schools in the district... There is no formal scheme for regular contact nor does there seem to be any desire for it."*
- *"There is no continuity or consistency in curriculum between Houses. My ninth graders come into high school science with tremendous disparity in preparation because of this lack of agreed upon standards within the middle school."*
- *"...among the three of us who teach Earth Science there is no formal sharing or even discussion about the curriculum, methods, or assessments."*
- *"As a department we will mobilize quickly and effectively, but we still do it in isolation."*
- *"Teacher preparation is done on an individual level; teachers are just expected to know what to do."*
- *"I feel there is a lack of communication within my department about assessment. I teach ninth grade students and I have not seen anything to inform me of what my students have achieved in the middle school."*
- *"I feel sometimes that we are isolated to our own level therefore making it difficult to see where we need to be headed."*
- *"The culture of the separate dominates and not the collaborative. There is little collaboration between schools, between parents and district, between teachers and parents."*

One teacher's comment seemed to speak to the sense that to collaborate to implement standards and address needed changes was overwhelming. The teacher wrote, *"In a school setting it is much easier to just throw in the towel, close my door, and say, 'you do it.'"*

Teachers who saw evidence of collaboration, often indicated that individual efforts (rather than a district-level, administrative plan) sustained the collaboration. They wrote:

- *"Our middle school math has an excellent department and means of communication. Teachers at the same grade level meet every other day to*

*discuss grade level concerns and student achievement. Teachers can not teach effectively on a island with no bridges...I will continue to share ideas with my colleagues, and pick up ideas from them.”*

- *“I meet with other eighth grade math teachers at least once a week to discuss new strategies, get ideas, share my own ideas... my colleagues and I make the time to meet, because meeting the standards and developing new assessment techniques is difficult and sometime we can not generate effective strategies all on our own.”*

The East Coast teachers interviewed district administrators and other personnel in their home districts to determine what kind of district assessment plan was in place. Seventeen percent of the teachers reported no district assessment plan; twenty-five percent reported that there was an unwritten plan; another thirty-three percent reported that there was a written assessment plan, but no clear tie to standards; and twenty-five percent reported that a district team was working on standards.

The teachers made the following comments regarding unsettling matters or attitudes they discovered in their districts:

- *“It seems to be their [administrators] understanding that, by hiring only teachers with a masters that the ‘new’ Master Teacher knows all of this and will implement standards on his/her own, for the good of the class.”*
- *“Administrators appear to be stuck in their ivory towers, mired down with paperwork, and thinking the system will fix itself with the right teachers and new books.”*
- *“Money seems to be one of the major stumbling blocks.”*
- *“There are no consistent assessment measures across the grade levels or among teachers at the same grade level, meaning no consistent standards exist.”*
- *“It is important to recognize that the majority of these assessments (state tests and standardized tests) in this plan have very little to do with the new standards, yet they make up the majority of the assessment plan.”*

Although the majority of East Coast teachers found unclear, un-unified district or school assessment plans, they did remark on a number of positive aspects:

- *“An assistant principal has recently been hired in order to provide more time for the principal to focus on curriculum development and standards work.”*
- *“Within the last couple of years, the district has implemented through surveys and curriculum work, an assessment of how our current programs are meeting the state standards.”*
- *“Our district had all staff members complete a needs assessment survey ...to determine our direction for instruction for the standards...Based on these results, our district has funded summer curriculum work for teachers, levels, and departments to work on taking the standards and correlating curriculum and instruction to these standards.”*

In some districts, the stance regarding standards seemed to be to wait to see how the standards are interpreted in terms of the state exams that will be given. Teachers reported that the state exams were viewed as the main evaluation tool for assessing standards, and that assessment should be left to the state. In the discussions with administrators, the issue of using assessment as a means for interpreting standards and designing instruction was overshadowed by the view of assessment as only summative and only high stakes. For example, teachers reported district administrators as saying:

- *“It appears that there is no unity or district goals except to meet the standards ... until after we have a sampling of the test.”*
- *“Let the state provide the assessment plan.”*
- *“Assessment is a science, leave it up to the experts.”*

Some of the shift from thinking of assessment in more formative terms to thinking in summative terms may have been driven by a concern over legal issues, as for the teacher who wrote:

- *“There may be some liability issues when a student does not graduate because he didn’t pass an exam. If the question of who made the exam came up and it was a teacher-made test, then the reliability of the exam is called into question and the district may be in a tough position with the student’s family and attorney.”*

A number of teachers did see the state assessment as a strong influence on instruction and curriculum:

- *“Teachers’ effectiveness is also evaluated by the [state exams]. Teachers whose percentage passing has dropped, are required to do item analysis, and are expected to make the necessary changes in their teaching and/or curriculum.”*
- *“In one situation, Physics exam scores reflected a need for more lab time. Revisions of the next year’s schedule immediately took place including extended lab time.”*
- *“The upcoming assessments in the state are functioning as motivators for teachers to develop new techniques of teaching and assessing throughout the year.”*
- *“Since the state is assessing in a certain way, then we need to design and implement our program in a way that is consistent with the assessment methods.”*
- *“He [a district administrator] was clear that [if the state exams change to meet the standards] the program will have to be modified, but the modifications will aim to achieving success on the state exams.”*

With the East Coast group, we constantly had to bring discussion from summative state exams back to what the teachers could do with formative assessment. However, even when we were successful in making that shift, the teachers still questioned the worth of doing anything that was not reflected in the state exams.

Many teachers valued their individual academic freedom and expressed concern that the standards-based reform and the state exam might impinge on that freedom:

- *“Are we in danger of taking away teachers’ freedom to design and orchestrate their classroom? I enjoy the freedom of designing what I do with my day/job.”*

- *“The teachers do have a tremendous amount of power, however, just from casual discussion, it’s my understanding that the teachers are violently opposed to any district mandates or imposition of standards. It is perceived as whittling away at their freedom and a step toward micromanaging. The strength of the system is based on very strong, independent teachers who are given tremendous freedom as well as responsibility.”*
- *“I am somewhat put off by being told that this is the task that my students will perform. I want my freedom.”*
- *“I am uncomfortable with how much structured guidance must we provide with standards, objectives, goals, outcomes, etc. before we all feel smothered and straight-jacketed.”*
- *“Now I ask, where does teacher empowerment fit in? I realize that I am far from helpless when I decide how to teach, but how empowered am I when dealing with the standards?”*

The East and West Coasts differed as to who should have the final say in providing assessment. Nine percent of the East Coast teachers versus fifty-eight percent of the West Coast teachers indicated that the state should have the final say. Two important issues were raised regarding this theme. The first was that as the role of high-stakes tests in summative assessment increases, so too does the concern over fairness and legality. The second issue was that of the mismatch between what is called for in the standards and what is tested on non-teacher generated exams (a mismatch between what is assessed in a formative assessment by the teacher and what is assessed in a summative assessment by the state). One teacher wrote, *“It is important to recognize that the majority of these assessments (state and standardized tests) in this plan [that teacher's district plan] have very little to do with the new standards, yet they make up the majority of the assessment plan.”*

Assessment Planning, Attitudes and Difficulties. At the end of the East Coast course of professional development, the teachers created an instructional unit using the process of moving from standards to instruction, using assessment goals and criteria as the major design focus, which they called the “assessment planning process.” To help teachers focus on content and assessment, we asked them to include in their unit two extended-response tasks that they could use at any point during the unit to determine what students had learned or to make adjustments to instruction. An extended response task is one that is open to interpretation, has either multiple answers and/or multiple ways to determine an answer, and requires an explanation or justification of an answer.

As we worked with both the East Coast and the West Coast teachers on the design of extended-response tasks, it became clear that depth in understanding content knowledge aided teachers in the design of such tasks and in the generation of expected responses to the tasks. Lack of depth of content knowledge opened possibilities for students to form incomplete conceptions of relationships. For example, a teacher working on an integrated mathematics and science unit involving the planets wanted the students to use what they knew about the relative diameters of the Earth and moon (the moon’s diameter is about ¼ of the Earth’s) to pick from a container of different sized spheres one that would accurately represent the moon if the earth was

the classroom globe, and then explain the relative differences between the globe and the sphere representing the moon. The teacher had used a similar task in the past and had trouble with the students picking moons that were too large relative to the size of the model for the earth. The teacher's directions to the students were to choose the sphere that showed the "1:4 relationship in size between the moon and the Earth." We talked with the teacher about the inadvisability of using the word "size" rather than the more specific term "diameter," because "size" could also mean surface area or volume, which do not maintain the 1 to 4 relationship. The teacher did shift the language from a general use of the term "size" to "diameter," but did not address the confusion students have because they expect the volume of the moon to also be  $\frac{1}{4}$  that of the Earth's, when it is actually  $\frac{1}{64}$  (that is,  $\frac{1}{4}$  cubed). After more discussion and one-on-one work with an instructor, the teacher decided that the students would need some preliminary work with stacking cubes (how many cubes 1-in on a side are need to make a cube that is 2 inches on a side? Or three inches on a side?). Many of the teachers, began to see the need for depth of understanding on the students' part when faced with writing explanations, but seldom directly talked about the need for depth in their own understanding:

- *"I have learned from doing this project that it is not an easy task to write these items, and it takes a lot of thought. What I found most difficult was preparing a scoring guide that adequately and fairly grades students' comprehension of content."*
- *"The second and equally important outcome of this project is the new understanding I have of my own teaching responsibilities. If I am going to use extended response as an assessment technique then I have to teach the necessary skills used by students to answer in this manner."*
- *"I have come to realize that the current [state] curriculum is not conducive to raising students to standards."*
- *"The learning goal for mathematics could be represented by my new formula Competency = Comprehension + Computation. In other words, instruction must change to ensure that students know not only that the computational answer is important, but more of the reason why this is the answer, and can communicate this."*
- *"I found that the problem solving activities I do in my class really do not characterize themselves as extended task responses."*
- *"I have also learned that knowledge is not as cut and dried as it seems. I realize now that knowledge is tailored to the given situation."*
- *"My short comings came from poor assessment goals and techniques. I taught in response to ONE understanding level and ONE assessment technique."*
- *"With the extended task responses we are finding that the deeper we evaluate a student's understanding, the more directions and questions about what we expect and don't expect in an answer."*

The East Coast teachers also saw the conflict between depth and breadth in the current science and mathematics curriculum and made statements such as the following:

- *"The [national science] standards seem to combine depth and breadth and I don't think this is realistic. If we as teachers are going to focus on understanding, how can we teach so much material?"*

- *"I will definitely advocate for a shorter curriculum."*

Lack of depth of content knowledge appeared to be one of the major stumbling blocks in identifying criteria for evaluating evidence that standards had been met. Most of the East Coast teachers resisted providing ideal sample responses for the tasks that they created and would write a somewhat vague scoring guide for evaluating an explanation, such as:

Was explanation clear and supportive of work shown? (20 pts)

Was explanation attempted using some evidence from work shown, but unclear? (10 pts)

Was explanation attempted, using no evidence from work shown, but unclear? (5 pts)

With guidance, the teachers were able to write detailed criteria for different levels of responses when working together in a group (6<sup>th</sup> grade students should be able to recognize whether a pattern exists in data and describe it if it does exist; 8<sup>th</sup> grade students should be able to recognize whether a pattern exists in data, describe it if it does exist, and identify whether it increases by addition—forms a line, or by multiplication—forms a curve). However, when working alone on their units they did not make distinctions in criteria for different grade levels, and their criteria were often general or vague (e.g., *"recognition of patterns, or lack of, in the experimental results"*).

Although each teacher had difficulty delineating levels of criteria and had concerns related to issues of freedom and undue pressure to adhere to standards, they all produced units that translated standards into instruction. The degree to which they used assessment to do so varied. Comments from the East Coast teachers' journals indicated that from the first two days to the last two days of the course they experienced the following feelings regarding using an assessment planning process:

- *"somewhat optimistic" to "frustrating";*
- *"very difficult" to "still very difficult";*
- *"extremely frustrating" or "exciting but difficult" to "getting easier";*
- *"frustrating" to "not easy"; and*
- *"overwhelming" to "lovely challenge."*

One teacher's final comment reflects the process of assessment in the service of standards-based instructional design as many of us see it:

*"I started this class with some ideas regarding the standards. These ideas were vague and so were the standards to me at that time. I now feel that I am developing a systematic approach that transforms the standards into components that I can use in my classroom. Practicing writing acceptable criteria for individual standards is very difficult, but I feel that I am understanding more and getting better at writing them. So far, it seems that writing the criteria is the key to tackling the standards."*

## Conclusion and Implications

Our work with both the East Coast and West Coast groups indicates that the process of translating standards into performance expectations appears to be a difficult and challenging process, especially when teachers feel that communication among the various components and units of a district is not functioning well. The differences in individual definitions of assessment and related terms are much greater than teachers expected. For any assessment-focused professional development experience, there may be two groups of teachers who will benefit from a discussion of individual definitions of terms: those who are afraid to reveal a lack of knowledge about terms because they fear they are alone in their confusion or lack of knowledge; and those, who know that differences in understandings of terms exist, but who are confident in their particular definitions. A common, non-complex description of "assessment" seems a necessary part of working with teachers on assessment in the service of instruction. Based on the teachers' varied responses and confusion over complex or rigorous definitions of assessment, we have begun to define assessment simply as "data collection with a purpose."

Teachers also feel that freedom and time concerns are major blocks to implementing the process of using assessment to implement standards. Teachers identified a need for both the restructuring of how they work together in districts, and up-to-date professional development programs to enhance teachers' knowledge about assessment techniques, standards, new ideas from research (e.g., "backward design"), and the interplay among these.

Initially the East Coast teachers did not see the importance of specifying criteria that help to establish a clear and coherent vertical articulation of the science and mathematics curriculum from one grade to the next. When they did see the need, many felt overwhelmed by the challenge. A major difficulty that the instructors saw was that the teachers' knowledge of science or mathematics and of assessment was not adequate enough for the teachers to make distinctions both in the nature and quality of assessment materials and in the science content from one grade to the next. The West Coast facilitators reported that depth of science content knowledge appeared to have a direct effect on both the quality of the open-ended assessment items the teachers devised, and the degree to which teachers could clearly and consistently align standards with curriculum, and develop benchmarks for standards. Thus, as Loucks-Horsley, Hewson, Love & Stiles (2000) and others have indicated, professional development must also enhance teachers' content knowledge.

We also see implications for initial teacher preparation programs. Depth of science or mathematics content understanding must be part of a teacher's preparation. Teachers need a sound understanding of the science or mathematics they will be teaching – a kind of understanding that helps teachers with subtle differences in content. Teachers also need a solid enough understanding of concepts and principles to be able to help students apply the concepts and principles in contexts and to develop benchmarks or criteria for assessing the extent to which students are meeting learning standards at all grade levels.

---

## References

- Birman, B. F., Desimone, L., Porter, A. C., & Garet, M. S. (2000). Designing professional development that works. Educational Leadership, 57 (8), 28-33.
- Borko, H. (1997). New forms of classroom assessment: Implications for staff development. Theory into Practice, 36 (4), 231-238.
- Gagne, R. M., Briggs, L. J., & Wager, W. W. (1992). Principles of Instructional Design. ForthWorth, TX: Harcourt Brace Jovanovich College Publishers.
- Loucks-Horsley, S., Hewson, P.W., Love, N., & Stiles, K. E. (1998). Designing professional development for teachers of science and mathematics. Thousand Oaks, CA: Corwin Press.
- Wiggins, G. (1998). Educative assessment. San Francisco, CA: Jossey-Bass.
- Wiggins, G., & McTighe, J. (1998). Understanding by design. Alexandria, VA: ASCD.