NCRMSE Research Review
from the National Center for Research in Mathematical Sciences Education

Volume 2, Number 3
Fall 1993

In This Issue:
- What Gets Graded Is What Gets Valued
- *Review of NCRMSE Research*

Assessment and School Mathematics

Assessment is a necessary part of school mathematics curriculum and instruction. Traditional mathematics assessment procedures test a limited set of procedural skills. Such assessment procedures do not reflect the expanded view of school mathematics that the reform movement seeks. If reform goals are to be reached, they must be changed.

Researchers at the National Center for Research on Mathematical Sciences Education (NCRMSE) began to examine issues related to mathematics assessment nearly a decade ago. They looked at the assessment procedures used by United States teachers as well as at their test-preparation practices. Using actual test items, they studied the alignment of widely used tests with the NCTM *Curriculum and Evaluation Standards for School Mathematics* (1989). Their results demonstrated that the assessment procedures commonly used in schools were not only inadequate but should be viewed as a major barrier to the reform of school mathematics. This work is reported in *Mathematics Assessment and Evaluation: Imperatives for Mathematics Educators* (Romberg, 1992) a book published last year.

In 1990, several of the same researchers formed one of the NCRMSE Working Groups, Model of Authentic Assessment. The group, led by NCRMSE Director Thomas A. Romberg, made plans to develop criteria for authentic assessment models and to identify, test, and implement alternative assessment practices for school mathematics. The Working Group designed its research agenda to include a few major research investigations and a series of smaller explorations, as well as several collaborative activities.

Major Investigations

The major investigations of the Working Group include an adaptation of a Dutch curriculum-embedded testing model for use by teachers during instruction, the use of "super-items" that are based on a developmental model of reasoning, and the development of technology-based (video and computer) assessment models.

Curriculum-Embedded Model

Jan de Lange of the Freudenthal Institute in The Netherlands and Thomas A. Romberg are preparing an assessment plan for a middle school mathematics curriculum. Their goal is to build closer connections between assessment and instructional activities. They plan to develop techniques for scoring, aggregating, and reporting information on student performances. They, with the assistance of other staff members, will design, develop, and test assessment activities in varied formats, gather data on the relation between different formats and learning goals, and develop reliable scoring rubrics.

Superitems

Kevin Collis, with other group members, is completing a set of superitems based on the Structure of Observed Learned Outcomes (SOLO) Taxonomy for Grade 8. The items are being field tested and another group member, Mark Wilson, will analyze the data obtained during the field test using his design for a psychometric model. Working Group members are beginning to develop assessment items in the areas of probability and statistics. The items are designed to obtain data that will answer research questions on the teaching and learning of statistics. The data will be used to portray student understanding as it develops over the course of K-12 schooling, using the framework of the SOLO taxonomy.

Assessment and Technology

Susanne Lajoie, one of the group's principal investigators, has used both video and computer technologies to prepare a library of exemplars of student products that would convey expectations to students and to devise ways to use the computer to record data on students' problem-solving activities. She is continuing to develop scoring rubrics for computer-derived data on student work and to collect verbal protocols that illustrate student development in the areas of problem solving, communication, reasoning, and connections.

Classroom Explorations

Two of the smaller explorations of the Working Group were carried out in classrooms. Linda Wilson studied the way in which a high school mathematics teacher, in a school that has stressed assessment reform, implemented reform ideas in her classroom. Her findings are reported in the next article in this newsletter, What Gets Graded Is What Gets Valued. M. Elizabeth Graue and Stephanie Smith conceptualized assessment from an instructional perspective. The took the position that assessment should grow from the "dimensions of learning that are the focus of instruction." They then tested the utility of their model in a middle school classroom. Their findings are reported in the last article, Review of NCRMSE Research, in this issue of the newsletter. David Clarke and Max Stephens along with Margarita Wallbridge (1993) checked to see whether changing the assessment practices at the senior secondary grades had an impact on instruction and assessment in the earlier grades. They carried out an analysis of curriculum documents, work samples from teacher and student interviews, and a questionnaire. From these data, changes in policies, curriculum and teaching, assessment, and reporting practices could be inferred. Seven members of the Working Group presented papers about their assessment research at a symposium, Moving Beyond the Rhetoric About Authentic Assessment in

Mathematics, held at the annual meeting of the American Educational Research Association in April 1993 in Atlanta, GA.

Collaborative Activities

Working Group members have developed collaborative relationships and provided information on assessment to national, state, and local groups. Richard Kitchen is collaborating with The Consortium for Mathematics and Its Applications (COMAP) to test high school units developed for the ARISE Project. He will focus on the assessment practices of several teachers who use the units in actual classrooms to extend knowledge of the ways assessment can inform instruction. Other group members have met with several state education agencies, including those in Wisconsin and Delaware, that are developing frameworks for alternative assessment programs, participated in New Standards Project meetings, and made presentations to participants attending conferences on assessment at the national level.

The assessment activities of NCRMSE have led policy makers and legislators at state and national levels to request assistance from the Working Group chair or its members. They were asked, in 1991, to prepare a document for the National Center for Education Statistics, Improving Mathematical Performance: Reflections and Suggestions Based on the Results of NAEP's Twelfth-Grade Assessment. The document is designed to provide teachers and other educators with an analysis of a report on the matematics portions of the National Assessment of Educationl Progress (NAEP). The first section of the document provides an overview of the NAEP report, including its purpose, framework used for developing test items, and a description of the 1990 level of performance for Grade 12. The remaining portions of the report take a closer look at student performance and at released test items for each of five content-reporting categories. The final section provides suggestions for improving the mathematical performance of secondary students. This report was followed by another that contains an examination of the NAEP items and their relation to the NCTM *Curriculum and Evaluation Standards, In Light of the NCTM Standards, Does the NAEP Mathematics Assessment Measure Up? A Look at the 1990 NAEP 12th-Grade Items.* The authors (Romberg, Smith, Smith, & Wilson, 1992) found that the NAEP items did not relate to the traditional high school curriculum nor did they relate to the NCTM *Standards*. Large numbers of items, they concluded, covered content typically included in a traditional Grades 5-8 curriculum.

The National Assessment Governing Board (NAGB) asked the authors of the earlier reports to reflect on setting achievement levels for the NAEP mathematics assessment and to examine the feasibility of using existing international data to validate the derived levels. On the basis of this research, using both 1990 and 1992 NAEP mathematics items, the authors concluded that efforts to set achievement levels of the NAEP assessment were flawed, primarily because of the lack of validity of the items, but also because of the level-setting procedure. After additional analyses of items from the international examination, their study concluded it was not feasible to calibrate scores from either of the international tests with the NAEP.

In 1992 NCRMSE Director Thomas A. Romberg, chair of the Models of Authentic Assessment Working Group, was appointed to the United States Department of Education's Advisory Committee on Testing in Chapter 1 by the Secretary of Education. The committee was asked to

examine the adequacy of standardized tests in measuring the academic achievement of Chapter 1 students. The committee developed a lengthy report on Chapter 1 testing and concluded that a new policy balance was necessary, that "Chapter 1 testing and evaluations should become less concerned with large-scale evaluation…and more oriented toward enhancing effective classroom instruction and elevating student achievement." After the committee completed its work, its members, including NCRMSE director Romberg, were asked to testify before the United States House Subcommittee on Elementary, Secondary, and Vocational Education on the Testing Committee's recommendations.

Some of the findings of the members of the Models of Authentic Assessment Working Group are being prepared for use by teachers and will be included in a 1994 book, *Reform in School Mathematics and Authentic Assessment*, to be published by SUNY press.

Assessment Standards

Early in 1993 NCRMSE Director Romberg and Working Group members Jeremy Kilpatrick, Susanne Lajoie, Sandra Marshall, Marvin Smith, Norman Webb, and Linda Wilson were asked to assist with the preparation of assessment standards by the National Council of Teachers of Mathematics. The new set of standards is designed to complement the *Curriculum and Evaluation Standards for School Mathematics* (1989) and *Professional Standards for Teaching School Mathematics* (1991). The first draft of Assessment Standards for School Mathematics was completed by a group of mathematics assessment specialists during the summer of 1993. It will be circulated for comment to members of the mathematics education community during the 1993-1994 academic year. The draft will be revised for publication during the summer of 1994.

For additional information about the Working Group, contact Thomas A. Romberg by telephone at (608) 263-3605. NCRMSE has prepared an annotated bibliography of its working papers. Readers can request a copy of the bibliography from Donald Chambers, Director of Dissemination, NCRMSE, 1025 W. Johnson Street, Madison, WI, 53706, phone (608) 263-0761, E-mail dlchmbrs@vms.macc.wisc.edu.

What Gets Graded Is What Gets Valued
By Linda Dager Wilson

Rhetoric About Authentic Assessment

The term authentic assessment arose from educators' need to distinguish their new ideas on assessment from more traditional ideas on testing and testing instruments. When educators describe the current status of a person's capabilities within a specific conceptual, procedural, or developmental domain, they are assessing that person.

Assessing has been confused with testing, measuring, or evaluating. A teacher's assessment practices include all of the ways she/he determines what students know or can do. Testing can be contrasted with assessment in that it involves creating a situation that will inform decisions.

Measurement involves specifying "how much" capability exists, and evaluation involves assigning a value to it (Lesh & Lamon, 1992).

Today the term authentic assessment encompasses a broad range of criteria, but performance remains an essential aspect. Proponents of authentic assessment argue that to gain a true description of what students know or can do in a discipline or domain, they need to opportunity to perform work in that discipline. A traditional standardized achievement test item would have asked, What is the most appropriate unit of measure for the length of a room: centimeters, meters, or kilometers? A more authentic task would ask the student to measure the length of a room while the teacher observes the student's performance.

Authentic assessment must build from authentic assessment tasks. To be considered authentic, assessment tasks will, according to Archbald and Newmann (1988), meet the criteria of: 1) disciplined inquiry; 2) integration of knowledge; and 3) value beyond evaluation. Disciplined inquiry depends on prior conceptual and procedural knowledge but it develops an in-depth understanding of a problem and "moves beyond knowledge that has been produced by others" (Archbald & Newmann, 1998, p. 2). Its objective is the production of new knowledge such as that created by scientists or historians. If assessment tasks are to reveal students' integration of knowledge, they must address the content as a whole rather than as a collection of fragments. Students are "challenged to understand integrated forms of knowledge…" when they are "involved in the production, not simply the reproduction of new knowledge," because production requires the integration of knowledge (p. 3). Features that characterize tasks that possess a value beyond evaluation produce either discourse, material results, or performances, and require the flexible use of time and collaboration with others.

Examples of assessment tasks that meet the three criteria specified above help to clarify the concept of authenticity. A paper-and-pencil task from the National Assessment of Educational Progress's Pilot Study of Higher-Order Thinking Skills Assessment Techniques in Science and Mathematics is cited by Archbald and Newmann as meeting their three criteria (1988). The task asks children to examine data about five children competing in three athletic events and then to decide which of the five would be the all-around winner. The task is "open-ended" in that students must devise their own solution strategies and justify their answers rather than seeking the one "right" answer. The exhibitions of mastery required by some high schools as a diploma requirement are also cited by Archbald and Newmann. Students demonstrate competence in multiple disciplines and produce projects that reveal their integration of knowledge with the exhibitions.

Grant Wiggins (1989) develops a similar set of criteria for authentic tasks: A task is deemed authentic if it requires "the performance of exemplary tasks" (p. 703) and is "responsive to individual students and to school contexts" (p. 704). Authentic tasks can, according to Wiggins, "reveal achievement on the essentials" (p. 704). In the portfolio-based assessment program in writing and mathematics that currently is being implemented in Vermont, open-ended tasks such as the one described earlier, and an oral history project are examples of tasks that fit his criteria for authenticity.

While the two sets of criteria described above deal with authentic assessment in several content areas, some educators have focused more specifically on assessment in mathematics (Webb & Romberg, 1988; de Lange, 1987, Lajoie, 1991). These educators detail criteria that are specific to mathematics but echo the broadly based criteria of Archbald and Newmann, and Wiggins. Lajoie (1991), for example, calls for assessment that will:

- Obtain multiple indications of individual's knowledge, performance, and disposition.

- Use tasks that are instructionally relevant, meaningful to students, and realistic for the discipline.

- Use scoring and scaling procedures that are appropriate to the tasks being assessed.

- Align tasks with both curriculum and instruction and design them to show what students know.

- Consider racial/ethnic, cultural, gender, and aptitude fairness.

- Be integral to the classroom environment.

- Provide for assessing individual growth when part of a group activity.

Classroom teachers are central to the implementation of these new approaches to assessment. A recent case study of a high school mathematics teacher's assessment practices was designed to examine how teachers are interpreting these approaches to assessment in their classrooms. The remaining portions of this report are developed from the case study.

Assessment reform at Valley High

Valley High is a typical small-town United States high school. Its nearly 800 students choose from a wide variety of courses from the English, mathematics, foreign language, science, art, music, and social studies departments. The mathematics department offers classes from a general mathematics track and a calculus/college preparatory track.

While, on surface, Valley High appears similar to other middle class schools in this country, it is beginning to feel the winds of change. Two years ago, the Valley School District began to initiate a series of district-wide reforms. It developed a Strategic Planning Commission of 150 members from the community and the school faculty, administration, and board and changed it with deciding what shape the reforms should take. The commission decided that its first goal would be implementing authentic assessment, based on its belief that assessment reform held promise as a catlyst for broader school reform.

A 1992 television show that focused on assessment highlighted Vally High as a "school that works." Members of the Valley School District board and the principal of Valley High present the school system as embracing authentic assessment as a central tenet. During the TV program, they claim the district is reevaluating its curriculum in light of its assessment philosophy.

Valley School District clearly has begun an ambitious assessment reform effort. At the elementary school, teachers have the option of using portfolios in place of report cards and grades. Extended parent-teacher conferences allow teachers to demonstrate the work in the portfolios to parents. Some teachers have shown parents videotapes of students at work. Valley High is more cautious about eliminating letter grades. According to its plan, by 1996 graduation will not depend on Carnegie units but on an autobiography, a thesis, a commencement project, and a portfolio.

Ms. League, the subject of this case study, has taught at Valley High for five years. She teaches a total of six classes per day with preparations in geometry, algebra II, and technical mathematics. In her overcrowded school, she has not yet been assigned a classroom of her own but carts her teaching materials down corridors to a variety of rooms. Her teaching credential and bachelor's degree in mathematics were completed after an earlier 20-year career in business.

While other members of her mathematics department were working to maintain the status quo, Ms. League, who likes to think of herself as a maverick. Fought to keep mathematics a central part of her district's new vision and to change to a less traditional mathematics textbook series. She endorsed the reforms and was the only representative of her department on the district's Strategic Planning Commission and on the Gateway Planning Committee, which was directing the Valley High graduation plan.

Assessment in the Classroom

During the three spring months of 1992, Ms. League was observed two to four days each week as she taught a class in algebra II. She was also observed a few times during the same interval as she taught geometry and technical mathematics classes. In additions to the observations, data were collected during four formal interviews with Ms. League as well as during daily informal discussions and interviews with 10 of her 22 students. Field notes and transcripts of audiotapes of the observations and interviews were analyzed to describe the ways this teacher assessed what her students knew and could do in mathematics and to examine the relationship between her beliefs and the assessment practices she used. To set up a contrast, as series of six observations of algebra I classes taught by a second Valley High teacher, two formal interviews, and several informal discussions were completed. Both teachers shared their student materials and grade books with the researcher.

A Daily Routine

On an ordinary day, Ms. League began her class by asking students about their homework assignment. A few responded with questions about particular textbook problems.

Ms. League demonstrated solutions to the linear and quadratic equations in about 20 minutes. She then assigned a problem for the student "to try" in their "teaching pairs," which are the prearranged small groups or pairs that work together on this type of assignment. Members of these groups sat near one another, pulling their desks closer to work. This problem was more

abstract and complicated than the problems the students had completed for homework. Ms. League suggested that they do it by sketching a graph and "making decisions" from the graph.

As students worked on the problem, Ms. League circulated and gave help where it was needed. At one point, she paused and interrupted their work by saying, "If you're having trouble with this, you need to write down, I need to study how to graph hyperbolas. Be sure to write that down."

Before she demonstrated the solution to the problem, she said, "When I was walking around, I saw all levels of thinking. Some of you were struggling with the graph. Others had trouble with the procedures. Remember that the test is on Friday, and while some of you did OK working together today, remember that on the test you will have to work alone. Be sure you know how much you can do alone before the evaluation on Friday."

Students continued working on problems in pairs until the end of the class, when Ms. League gave them their homework assignment for the next day (six more problems from the textbook similar to the problems they had worked on in class). Then she added, "Your other assignment is to look at what you wrote down today about what you need to do. For example, 'I need to review parabolas.' So do that tonight also."

Two aspects of this classroom are especially interesting from an assessment perspective. The first is that Ms. League routinely observed and interviewed her students as they worked in their teaching pairs, and she also emphasized ways students could self-assess. Ms. League obtained knowledge about what her students know and can do in mathematics by interviewing them and observing them as they worked in pairs. She knew her students and could discuss their strengths and weaknesses at length. During one interview, Ms. League explained the marks in her grade book for the current term. She spoke at length about one student whose name had been selected at random from those in the grade book.

This informal method of assessment is not uncommon among high school mathematics teachers, but in the case of Ms. League, it is especially well-developed. It is also central to her beliefs about teaching, that is, she believes that knowing her students is fundamental to her teaching. She views this practice of observing and interviewing her students as they work as one of the more "authentic" aspects of her assessment practice. Indeed, according to Lajoie's criteria for authentic assessment (1991), his practice could be seen as one of the "multiple indicators of knowledge."

Ms. League asked her students to remind themselves of their strengths and weaknesses as learners. She frequently asked them to write comments to themselves about the concepts on which they needed work or the procedures on which they needed practice. An increasing number of sources agree that self-evaluation is an aspect of authentic assessment and advocate that students keep journals on their progress, or invite students to be active participants in the assessment of their learning.

Ms. League would often ask students to respond to higher-order thinking questions that went beyond the procedural aspects of a lesson. In the first lesson on logarithms, for example, she

asked students to think about why there can be no log of a negative number. She made it clear that she expected students to write individual explanations. Sometimes these higher-order explanations were to be finished in class, and sometimes they were to be part of a regular homework assignment that included textbook problems. It was evident, from the frequency of these assignments, that Ms. League valued her students' conceptual knowledge, not solely their ability to carry out algebraic procedures.

An Interpretation

Ms. League used a number of authentic assessment practices, including observing, interviewing, and asking students to write about mathematical concepts and their learning of them. One would expect her students to gain the sense that both carrying out routine algorithms and talking and writing about mathematics were valued. Interviews with students indicated, however, that was not the case, as the following excerpt suggests:

> **LW:** Have you ever had to write anything in mathematics class? Write a paragraph or explain something?
> **Sara:** Yes. A report. I did a report once. That was for extra credit, though.
> **LW:** For what class was that?
> **Sara:** Algebra I. That was a while ago.
> **LW:** So you never had a situation where a teacher said, "Here's a problem. Tell me what the answer is. Now explain how you did it."
> **Sara:** Yes. I have had that one a couple of tests before, too. Like—
> **LW:** In this class?
> **Sara:** Well, not in this class.

Why is it that students in this class did not recognize that they routinely were given writing assignments? The answer seems to rest in the system Ms. League used to develop grades for her students. At the end of each quarter, students received a traditional A, B, C, or D grade based on the quizzes, tests, and examinations they took during that quarter. Homework was checked occasionally and points were given for its completion. Homework checks, however, did not include the "extra" assignments designed to foster self-evaluation or higher-order thinking. Thus the "extra" assignments did not receive points.

The tests and quizzes written by Ms. League were based on textbook problems. Items were identical to the textbook problems except that they included different numbers that were assigned for homework. The tests were used to determine whether students could carry out the procedures or algorithms they had been doing in class for the previous week of ten days. They did not include the higher-order thinking questions, nor did they include questions related to self-evaluations. Students were not allowed to collaborate on the tests and quizzes, but completed them individually.

The techniques described earlier as components of an authentic assessment system were not part of the grading system in this class. Students, for the most part, ignored any activities that did not "count" toward their grade. On any given day, roughly half of them did not complete their homework. When Ms. League asked them to do writing assignments in class or for homework,

most of them ignored her. They did not do the self-evaluations nor did they answer the higher-order thinking questions. In fact, they did not recognize that writing was a part of what they did in their algebra classroom.

Students counted as worthy of their energy in this class only those tasks or activities that were reviewed and then recorded in the grade book. They paid little attention to any assessment activities other than quizzes, tests, examinations, and an occasional homework assignment. Because they were graded, these activities were valued by students.

What counted as mathematical knowledge--for Ms. League's students--was correctly carrying out procedures for solving decontextualized problems such as a system of equations. Ms. League had a different notion of what counted as doing mathematics, illustrated by her higher-order questions, and her exhortations to students to think, write, and work collaboratively. The non-traditional activities had more to do with reasoning, reflecting, and communicating than they did with routine procedures. Since the non-traditional activities were not incorporated into the grading system, they were not valued by students. In effect, this teacher's expectations for her students were lowered by her grading system.

Ms. League was teaching in a school that publicly has embraced notions of authentic assessment and has initiated district-wide reforms to that end. The principal of her high school is a visible proponent of reform. And Ms. League was and continues to be involved extensively in those efforts. She has, unlike many of her co-workers, serve on several key reform committees and openly endorsed their cause. She did implement several authentic assessment practices, despite her challenging working conditions—e.g., six classes with a total of 150 students in three different courses, and no classroom of her own. She was given no assistance by her traditional text for implementing non-traditional teaching techniques; nor was she given any time for developing supplements.

Her initial efforts toward more authentic assessment practices were in large part failures because they were no incorporated into her grading system. This account suggests that the issue of grading systems will become more critical as the assessment reform movement makes progress. Grades, especially at the secondary level, are a powerful force that define what it means to know and do mathematics in the classroom.

(Linda Dager Wilson is an Assistant Professor of Education at the University of Delaware. Her research was supported by the NCRMSE Models of Authentic Assessment Working Group. She can be reached at the College of Education, University of Delaware, Newark, DE 19716).

References

Archbald, D., & Newmann, F. (1988). *Beyond standardized testing: Assessing authentic academic achievement in the secondary school*. Reston, VA: National Association of Secondary School Principals.

de Lange, J. (1987). *Mathematics, insight and meaning*. Utrecht, The Netherlands: OW & OC.

Lajoie, S. (1991, October). A framework for authentic assessment in mathematics. *NCRMSE Research Review*, pp. 6-11.

Lesh, R., & Lamon, S. (Eds.). (1992). *Assessments of authentic performance in school mathematics*. Waldorf, MD: AAAS Books.

Mathematical Sciences Education Board. (1991). *For good measure: Principles and goals for mathematics assessment*. Washington, DC: Author.

Webb, N., & Romberg, T.A. (1988, April). *Implications of the NCTM Standards for mathematics assessment*. Paper presented at the annual meeting of the American Educational Research Association in Atlanta, GA.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70(9), 703-713.

Review of NCRMSE Research

A Model for Classroom Assessment

Is measurement a good model for classroom assessment? In their paper, Conceptualizing Assessment From an Instructional Perspective, researchers M. Elizabeth Graue and Stephanie Z. Smith (1993) take the position that classroom assessment should use models that are more like instruction than measurement. The basic differences between classroom practice and the theory of educational measurement have been explored in recent work in the areas of cognitive sciences, curriculum, policy, and teaching. Teaching, learning, and cognition increasingly take a perspective that emphasizes local activity and the making of meaning. Educational measurement grew from psychological models with roots in behaviorist theory and the distributional characteristics of large numbers of individuals (Shepard, 1991).

Graue and Smith have proposed a dynamic and multi-faceted model of classroom assessment. Their model is based on the premise that assessment should grow from the dimensions of learning that are the focus of instruction. This model builds from three assumptions:

1) Assessment should be seen as a system of complementary parts that are considered in concert. An integrated approach that incorporates data form a variety of sources provides a more stable picture of student learning than single-administration data sources.

2) Information gathered during instruction is the closest to student learning and should be given the most weight in assessment. Contextualized evidence of learning across time is of particular interest.

3) Values for instruction should guide the development of assessment activities. In the case of mathematics, the NCTM *Standards* (NCTM, 1989, 1991) can be used as a guide.

According to the model, three dimensions of learning are of interest to instructional assessment in mathematics: content, process, and dispositions. Content includes concepts, procedures, representational strategies, and problem situations. Process includes problem solving, communication, reasoning, and connections. Dispositions includes beliefs, autonomy, and excitement. A more detailed description of the model will appear in a future issue of *Educational Assessment*.

Examination of the Model

This article summarizes an examination of the utility of the Graue/Smith model made with four mathematics teachers. The data were collected from four middle school teachers as they implemented Expressions and Formulas, a unite from the 40-unit curriculum, Mathematics in Context. The curriculum is designed to focus on mathematical content that is situated within specific contexts. The contexts provide meaning and inspire activities. The unit is designed for 6^{th} grade students and introduces the concept of a formula and the need for an agreed-upon order of operations. Students are provided experiences with notation and formulas and begin to develop an understanding of conventions.

Using data collected through classroom observations and one-on-one interviews, Graue and Smith sought to understand how four teachers learned what students do (content), what students learn (process), and what students feel (dispositions). They asked how teachers' thinking about assessment would changed if teaching and learning were their focus and they began to move away from a traditional testing perspective. Excerpts from their examination include comments from teachers that illustrate the evidence of change.

Content

Content typically receives the most attention during mathematics assessment. Students, according to the Graue/Smith model, should show growth in situational, conceptual, and procedural understanding and develop representational strategies that move from the simplistic to the more sophisticated. In general, the four teachers learned about their students' understanding of content by asking them questions—one-one-one, through instructional tasks, or in quiz situations. During the course of the unit, teachers talked about these sources of information. One of the teachers described his best source of information—talking with students in a one-on-one format (Graue & Smith 1993)—as follows:

It's easier if I can sit down or rove around the room and talk with kids individually and really ask them specific questions. It's hard for me to look at a piece of homework, because you don't really…sometimes you don't know how much help they've had at home and that sort of thing. Or they might arrive at an answer, but how did they arrive at it? Did they really understand what's going on here?

Another teacher thought his assessment task had been easier when students did more computational activities. When he described his approach, he realized that he may previously "have fallen down on the job where math was concerned." He hadn't figured out how to gather

information about students' learning with the new instructional format. He thought that portfolios of student work could be the answer (Graue & Smith).

Maybe this was an advantage, I don't know, but in doing math traditionally, what often happens is there's a homework correcting time built into the period so what you do is you take advantage of the kids. You just have them exchange papers and if there are 50 long division problems or whatever, it's pretty easy objectively to correct those things and then the kids hand them in and all you've got is a number…So you thump that down and you're done. Now, if I have a kid do anything remotely creative or anything that involves actual thought, then somewhere I have to look at is and, to be perfectly honest, I've fallen down on the job there where math is concerned. I mean, I've struggled and I've stayed somewhat abreast, but I have a number of things that I just haven't looked at and now, in this week before grades, I've got a number of things at home that, maybe I collected this a month ago, and I'm thinking, "Oh God, what was this?" I'll figure it out and I'll do the kids justice, but I would be more comfortable if I were able, directly after the kids turn something in, to sit down and think about it. But to be perfectly honest, in the day of a 6th grade teacher, that's rare. I mean, you can sometimes, but other days you just have to put it on the back burner, and then what happens, happens. (p.11)

He tied looking at student work to a common constraint of middle school teachers—finding enough time to do the kind of evaluation that new forms of curriculum require.

Process

The notion of process in the Graue and Smith examination included problem solving, communication, reasoning, and connections. The *Mathematics in Context* unit was structured to provide experiences in the four processes. Teachers were able to track students' thinking as the student's communicated their ideas about mathematical processes. Teachers were able to track students' thinking as the student's communicated their ideas about mathematical processes. When asked how a quiz went, one teacher told the researchers:

I think it went real well. Again it was the sort of thing where I could look at it and figure out essentially what kind of thinking kids were doing. They weren't just rote answers where you really had no clue as to what the thinking process was. (p. 13)

Another teacher agreed:

They had to really show proof on quite a few of them, and you could sort of see, by their work, how they were thinking—certain steps that they were taking—which was easier than some of the assignments we were trying to give. And that helped me too. (p. 13)

But another teacher had concerns:

The kids did better than I thought, because it's a hard test…There were too many words in the test. I mean, you had to read too much to get to the meat and potatoes, and some of the kids, like Jeffrey, that was too much of an effort for him. You know, given a formula and just plug in these numbers to this formula, he can knock it out cold. They've got those formulas down.

They really do…But hiding the numbers in a story problem—particularly as detailed as Tim's story problems were—that was an effort for some of them. (p. 14)

Situating mathematical content in real-life contexts has allowed students to build understanding. The stage for problems typically is set by language-rich descriptions of situations. For one teacher, the context provided a window, but in the view of another, the language used to build the context became a barrier for some students. The students who were used to substituting numbers for letters in formulas did not want to make their way through words, and some gave up. In terms of evidence of learning, traditional forms of teaching and assessing mathematics rewarded students who were skilled at completing non-contextual items. In the view of the authors, the real world does not come in simplistic problem formats—in fact, half the battle may be framing the problem rather than coming up with an answer. Adding additional language to mathematical activities is a tool for some and a hurdle for others.

Dispositions

The NCTM *Curriculum Standards* (1989) emphasize developing positive dispositions toward mathematics. The Graue/Smith model used three components to track students' learning in this area: beliefs about mathematics, autonomy, and excitement. Beliefs, in their view, were indicated by what students think mathematics is, as well as the value they place on it in various contexts. They theorized that autonomy, which develops as students become more expert and sophisticated in their mathematical thinking, can be evidenced by the students' feelings of perseverance, control, and confidence. Excitement can be seen in the reactions of students in the classroom and their willingness to try new tasks.

Teachers gained information on student's dispositions in a variety of ways: by observing their reactions to instruction, their problem-solving activities, their written work, and their interaction with others. One of the teachers described how the new approach to teaching used in the *Mathematics in Context* project provided rich information about students (Graue & Smith, 1993):

With the sort of more innovative approaches that we have with math now, linking with language and doing a lot of writing, there's more of an effort component that's possible to measure than there used to be. I can watch a kid in class here and see if he cooperates with other people, see if he's on task, whereas the traditional seat work kind of stuff, the only thing you could tell if you gave a kid 5 division problems is whether you had the answers right. Here you see the kid operating in a greater variety of ways and so it's easier to get some sort of feeling for what he's putting out.

If teachers are to monitor student dispositions, they need access to evidence of those dispositions. These teachers frequently gathered this information through observation and by looking at written work. When the issue of participation in class discussions came up, teachers found that they could not rely on calling on the students who volunteered in class because this group tended to include only a small number of the students. One teacher mused that she had more information on her "high flyers" because they participated more (Graue & Smith, 1993):

There's probably a third who always have their hands up. You really have to watch that as a teacher. You can't always be calling on those kids, and I think the more experience you have as a teacher you just don't do that. You say, I'll give you one chance to answer something kid, but not another one. It's going to have to be someone else in the room. So I try to rotate that a lot if I can. (p. 17)

These teachers realized that their access to information was related directly to student participation and they found that they could not leave that up to the students. Student participation became a different kind of activity, a source of assessment information. In addition to observing, the teachers also tracked students dispositions through self-evaluations during units.

The teachers struggled against student and parental beliefs about the nature of mathematics and 6th grade mathematics instruction. Some parents seemed to feel that the curriculum did not challenge the most capable students. Elementary schools that fed into the middle school had accelerated some groups of students by placing them into the next grade level of mathematics classes—e.g., 4th-grade students were placed in 5th-grade mathematics classes. In this study, students were not ability-grouped and almost every student participated in the *Mathematics in Context* project. Teacher discussed their concerns regarding student beliefs about mathematics that related to assessment:

The problem, of course, too, with math is that there's this notion that kids have of whether they're good in math or not. And if they are, they are supposed to get A's. And if they aren't, they're not. So, all of a sudden you take into account a kid's work habits and some of these other subjective things and he's not doing so well, and then you give this kid a B, and this kid comes and says, "Wait, I'm the smartest math student in here." Well, whatever that means, you do well on a standardized test, I suppose. But you do run into that. Especially with 6th graders who come from elementary schools that have been ability grouped.

One of the problems we've had here, really, is convincing kids that it was OK to be in a math class with everybody, and often times what they'll do, they'll claim that your curriculum is too demeaning or the 6th-grade work course is boring…But what I found is that if you take kids and pull them out of class and put them in a small group with 3 or 4 other kids whom they consider to be mathematically bright and give them the same things to do that we do in here, then it's OK. Then the stuff is great. It's only if they're in this group and they've been used to this special treatment that they fell that they're being somehow ill served. (p. 18)

Students felt that their math ability, as shown on tests given earlier and their elementary school placements in the high-ability group, was their ticket to an A. One teacher corroborated that view and added that the elitist attitude seemed to be fostered by the practices in some elementary schools:

Some of this attitude is coming from some of the elementary schools, I think, where they're sitting off in the choir; they're a little elite group and that type of thing. "Go ahead and do the 6th-grade and the 7th-grade math book."…The parents push that…Some of them don't care and

they just say, "Hey, my kid is ready for 7th-grade math and I want him in that. I want him in a special group. No matter what you say you don't get anywhere. (p. 19)

Teachers felt parents and students think that working in heterogeneous groups diminishes that quality of the classroom experience. One teacher in the study was especially troubled by mixing students of different abilities during assessment activities:

I feel real guilty if I take a brighter student who's catching on and pair him up with a kids whose head is somewhere else and then try to assess them. And they don't like that either. They sometimes never get past the fact that I'm sitting next to so and so and I don't want to be sitting next to so and so, and no matter what you say, I really don't want to be doing that. And so, I try to avoid that in the classroom. Those are problems I think I can control…you might have noticed, Doug and I do less grouping now than we did earlier in the year. (p. 19)

The teachers who explored this new approach to teaching mathematics found that their information gathering had changed and so had their evaluation of that information. Two of the teachers found they were looking at their students' learning very differently from the way they had viewed it previously, and that translated into a major change in their grading practices:

What I have had to do is sort of change my grading system. You know how all math teachers want to have numbers? I don't do that any more with those tests. I grade it like it was a history exam, so I'm more subjective, anyway, and I like that better actually. I don't have to put a number. I don't have to put 40% on the top of it. It allows a lot of freedom that way if I have some kid that I know is struggling, I know doesn't handle it conceptually, but has done something right—I can give that kid a letter grade and not have to justify it on the terms of the percentage. I might give a kid a C who got a 30. I don't have to put the 30% on there. I can just put this letter grade and I can talk to them about what they did well and what they didn't do so well. So in some ways it's kind of freed me from the shackles of number and I like that a lot. (p. 20)

Another teacher described what he considered an overly simplistic view of evaluation that lulled people into thinking that grades mean more than they actually do:

I've never relied completely on numbers. I think that's a very comfortable thing to do, especially math teachers do it, because you can reduce almost everything to a number and you can kind of kid yourself into thinking that the number means more than it actually does. Maybe it's because I've been around math enough to know that…I've seen teachers do things like give points for assignments. Ever math assignments. Points weren't necessarily even based on how many problems were right. They might have been based on whether the kid's name was on the paper and whether they were neat or not, and they give every assignment 50 points or something, and rate these assignments. And some of the assignments may have been long, some short, some easy, some hard, very diverse things. They all counted the same…And then, they'll apply those number to some sort of preconceived grading scale where they'll simply say, well, 93-100 is an A, 86-93 is a B, and all of that is, of course, completely arbitrary, but people don't know that, or they don't think about it, or they don't care. (p. 21)

Conclusion

This examination of classroom assessment from the perspective of classroom instructional practice identified several themes for further study. Strategies for gathering contextualized information such as observations, interviews, and project work were not yet fully developed or integrated into the work of the four teachers, although they were recognized as valuable sources of information about student growth. These teachers readily used observations to obtain information about student dispositions, but they were not comfortable using the more informal techniques to find out about content or process growth. In the future, informal strategies for obtaining information on content and process growth could be included in mathematics units. Rather than providing a short quiz at the end of units, teachers could be reminded of the kinds of learning and thinking they should be looking for as students progress through unit activities that relate to instructional goals.

(M. Elizabeth Graue is an Associate Professor of Curriculum and Instruction and Stephanie Z. Smith is a Project Associate at the University of Wisconsin-Madison. Their research was supported by the NCRMSE Model of Authentic Assessment Working Group).

References

Graue, M.E., & Smith, S.Z. (1993, April). *Conceptualizing assessment from an instructional perspective.* A paper presented at the annual meeting of the American Educational Research Association in Atlanta, GA.

Romberg, T.A. (in preparation). *Mathematics in context: A connected curriculum for Grades 5-8*. Chicago, IL: Encyclopedia Britannica Educational Corporation.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching school mathematics.* Reston, VA: Author.

Shepard, L.A. (1991). Psychometricians' beliefs about learning. *Educational Researcher*, 20 (7), 2-16.