Standards-Based Mathematics Assessment in Middle School

Rethinking Classroom Practice

Thomas A. Romberg
EDITOR

Teachers College, Columbia University
New York and London

Chapter 13

Classroom Assessment as a Basis for Teacher Change

David C. Webb
Thomas A. Romberg
University of Wisconsin–Madison

Truus Dekker
Jan de Lange
Mieke Abels
Freudenthal Institute, Utrecht University

In this chapter, we outline some of the important features leading toward a professional development program focused on reform-based assessment practices that we believe will help teachers make the necessary transitions. We then describe the Classroom Assessment as a Basis for Teacher Change (CATCH) project, where we are studying the potential of an assessment-driven professional development effort in two urban school systems.

As teachers move toward teaching for understanding, they also need to begin to assess for student understanding. To do so, teachers need to critically examine their classroom assessment practices and the familiar conventions of testing, scoring, and grading—practices that have been developed largely to monitor student mastery of skills and procedures. In this book, researchers and classroom teachers describe the teachers'

struggle in making the transition from traditional instruction toward teaching mathematics for understanding. When teachers use a reform curriculum, with its potential to promote student understanding, initially they are often unaware of the need to reconsider their current assessment practices in light of the rich evidence generated through complex, real-world mathematics problems. However, as portrayed in Chapters 2, 3, and 4, and particularly in Chapters 11 and 12 by teachers Ann Frederickson and Teresa Her, considerable instructional conflict is generated when teachers use a limited range of assessment practices to assess more substantive learning goals. Getting teachers to shift their assessment practices toward assessing student understanding has the potential of invoking real instructional change, which is key to reaching the overall reform goals for school mathematics.

Although research supports the contention that formative assessment benefits student learning and can be used to facilitate learning with understanding, many mathematics teachers (as described in several chapters here) show limited understanding of the ways in which formative assessment can be incorporated into their classroom practices. As a result, teachers often have difficulties in making didactical decisions based on their students' work and therefore defer instructional decisions to the sequence of activities in a textbook. Students in such classrooms are often left with incomplete information about their progress. They frequently find themselves at a loss to self-assess what they know or don't know, and they continue to apply and reinforce faulty mathematical conceptions.

Both the literature and our experience indicate that assessing for understanding is a critical component of teaching for understanding. Authors of the chapters in this book have suggested a number of design issues that teachers should consider when assessing for student understanding: What is the purpose of an assessment task? What type of response format should be expected from students? Does the problem context support or impede student problem solving? Does the current assessment program allow students opportunities to demonstrate understanding as well as procedural competence? The insight stories in this book have noted particular aspects of teachers' classroom practice and their role in teaching for understanding.

SUPPORTING CHANGE IN TEACHERS' CLASSROOM PRACTICES

The standards-based reform approach to instruction assumes that teachers will use evidence from several sources to inform instruction,

but in order to do this effectively, teachers require support in developing their ability to monitor student progress. In the studies reported in this volume, we found that teachers needed technical assistance with assessment design and that they sought tools and methods to further develop their capacity to assess student learning. The studies also showed that teachers could learn to use such practices in their classrooms, that they needed the support of appropriate professional development to do so, and that, as a result, their students' achievement improved (Fennema & Nelson, 1997; Webb et al., 2001).

The assessment methods used by teachers at the sites initially were grounded in practices that focused on mastery of skills and procedures. Despite the efforts of research staff to promote alternative assessment practices, we found that in the absence of additional on-site support, teachers faced difficult challenges in selecting appropriate assessment tasks and adopting the questioning techniques intended by the developers of *Mathematics in Context* (MiC; National Center for Research in Mathematical Sciences Education & Freudenthal Institute, 1997–1998). Teachers saw a need for change in their own assessment practices only after they saw the quality of students' work (including students' ability to construct reasonable justifications for their assertions) that was not being captured by conventional quizzes and tests. Over time, teachers developed a more comprehensive view of assessment as an ongoing process and regularly began to use a wider range of assessment tasks and strategies. The increased attention given to student learning via assessment motivated teachers to study further the relationship among mathematics content, instruction, and the evolution of student understanding as students progressed from informal to formal reasoning in mathematical domains.

Teachers' concerns about assessment are not strictly an issue of resolving problems related to tests and quizzes. Emerging research of teachers' instructionally embedded assessment practices has revealed that teachers are interested in exploring the use of assessment in a variety of instructional contexts (Webb, 2001). As described by teacher Teresa Her in Chapter 12, more purposeful instruction emerged after she restructured and redefined her grading system. These changes motivated her further study of the students' ways of representing and communicating their understanding of mathematics *during* instruction. It follows that when teachers explore and reflect on their own ways of formally assessing student understanding, their inquiry is likely to influence the instructional activities they choose, the questions they ask students, and the content of the classroom discussions they guide. Teachers' motivation to explore student thinking also may require

For a group of teachers, what kind and level of resources and professional development activities are needed to promote teacher interest, inquiry, and reflection toward the development of principled methods for assessing student understanding? Our effort to investigate these and other related research questions was realized as the CATCH project.

In 2001, we implemented the CATCH project in two school districts (Philadelphia, PA, and South Milwaukee, WI) to investigate further how principles of classroom assessment could be disseminated to a significantly larger group of schools and teachers. The research questions that we sought to address were largely questions of professional development and change in teacher practice, coupled with the effect of such changes on student performance. Our work included (1) the identification of materials to initiate teacher inquiry and to support the development of teachers' assessment practices, and (2) the documentation of features of school context that supported teacher change. Our documentation involved gathering information about teachers' conceptions of assessment, teachers' instructional decision making, change in teacher practice, and the relationship between teachers' assessment practices and student achievement. In particular, we were interested in answers to such questions as: How do teachers make decisions about what assessment tasks they use? What reasons motivate their choices? What kind of approach is needed to assist teachers in shifting their instructional focus (as reform curricula require) from covering topics and helping students master a collection of skills to fostering student learning with understanding in specific mathematical domains? How do teachers' assessment practices change as a result of their participation in this professional development program? How are changes in teachers' assessment practices reflected in their students' achievement?

The CATCH professional development program is designed to initiate teachers' critical examination of conventional assessment practices and support integration of teachers' assessment methods and instructional goals. Teachers are introduced to a framework for assessment design to evaluate and critique commonly used classroom assessments. An essential design principle that underpins professional development in CATCH is the pyramid model for assessing three levels of student thinking (see Chapter 1). The three levels in this model, as suggested by the three layers of the pyramid, are described as:

- Level 1: Reproduction, procedures, concepts, and definitions.
- Level 2: Connections and integration for problem solving.
- Level 3: Mathematization, mathematical reasoning, and generalization.

The relative size of the three levels in the pyramid exemplifies the number of items at each level that are required in order to assess student understanding. Assessment tasks used to describe student performance should, over time, fill the pyramid.

As teachers compare their own classroom assessments with the pyramid model, they quickly recognize that most of the assessments that they are using consist of items to assess basic skills, facts, and routine procedures (Level 1), often in multiple-choice or short-answer format. Teachers find that they rarely use problems designed to deepen student knowledge and understanding; more often, students learn how to use basic skills and routine procedures in unfamiliar contexts or choose appropriate mathematical tools to solve problems (Level 2). Nonexistent in teachers' classroom assessments is the use of questions to encourage generalization, mathematical reasoning, and argumentation (Level 3).

Through CATCH institutes and related activities, teachers concluded that they were not giving students opportunities to gain ownership of the mathematical content and were only asking students to reproduce what they had been practicing. Teachers who strive to teach for understanding find that they need to design tasks to assess goals at Levels 2 and 3 in order to assess for understanding. As teachers operationalize this assessment design model, they begin to rethink the learning objectives of their curricula and the questions that they use during instructional activities. Worth noting, these design principles have been applied successfully on large-scale assessments in the Dutch national alternative assessment used in conjunction with the Third International Mathematics and Science Study (Kuiper, Bos, & Plomp, 1997) as well as in the Programme for International Student Assessment (Organisation for Economic Cooperation and Development, 1999).

Underlying the design of resources and the organization of professional development experiences in CATCH is the emphasis on reconnecting formative assessment and teachers' instructional decisions to the development of big ideas in mathematics and assessment of student understanding over time (Shafer, 1996; Shafer & Romberg, 1999). To promote change in teachers' beliefs about assessment, teachers must experience authentic, nonroutine problem solving and need examples of tasks that promote mathematical thinking and reasoning. Achieving student learning with understanding, however, requires more than a collection of engaging classroom activities or thought-revealing assessment tasks. Rather, instructional activities need to be considered in relation to students' current conceptions of mathematics and related

activities to further develop those conceptions. The mapping of instructional activities and assessment tasks onto a learning sequence for specific mathematical domains is based on the concept of *hypothetical assessment trajectories* (de Lange, 1999), which are loosely sequenced sets of performance benchmarks for student learning in a content domain. The notion of learning lines within content domains is used as an organisational framework for teachers to select, adapt, and design assessments. There are practical issues teachers must consider, however: Tools and practices used by teachers are limited by the extent to which they can reasonably assess individual and collective learning within a classroom setting. (For additional examples of using learning trajectories as an organizational tool for instruction, see van den Heuvel-Panhuizen [2001], Fosnot and Dolk [2002], and Romberg, Carpenter, and Kwako [in press].) As demonstrated by Ann Frederickson and Teresa Her, when teachers view student learning in terms of learning and assessment trajectories, they are more apt to organize and sequence classroom activities that build from students' prior knowledge.

Attention to content goals and students' mathematical reasoning in the selection and design of classroom activities permits rich opportunities for formative assessment. As teachers broaden their conceptions of classroom assessment, use assessment trajectories to select and design assessment tasks, and make greater use of instructionally embedded assessment, they become better prepared to base their instructional decisions on the student thinking that they listen to and observe. As argued by Kilpatrick, Swafford, and Findell (2001) in *Adding It Up*, "Learning with understanding involves connecting and organizing knowledge, learning builds on what children already know, and formal school instruction should take advantage of children's informal everyday knowledge of mathematics" (p. 342). By improving the alignment among student thinking, instructional decisions, and classroom assessment, learning activities will more likely result in improved student achievement (Bransford, Brown, & Cocking, 1999). Whereas conventional classroom assessment focuses primarily on student outcomes and student recall of formal knowledge and procedures, assessing for student understanding requires that teachers attend to students' incoming knowledge and the way in which evidence for student thinking emerges through informal, preformal, and formal representations. As students learn, teachers must continue to monitor their progress, not in terms of correct or incorrect answers on some percentage scale, but in the broader and deeper sense of their conceptions of mathematical content and their growing ability to adapt what they understand to solve unfamiliar problems embedded in new contexts.

PROFESSIONAL DEVELOPMENT PROGRAM

To support teachers' efforts in implementing new assessment instruments and practices in their classrooms, the professional development activities for CATCH include opportunities for teachers to collaborate with district administrators and engage in shared planning of future professional development activities. During the introductory CATCH institute, teachers and district administrators engage in a range of assessment activities and discuss characteristics of assessment tasks. As part of the institute, teachers and administrators later identify the salient aspects of classroom assessment that require change and begin to prioritize and plan local professional development initiatives to support such changes.

Over a 2-year period, attention is given to supporting implementation of new classroom assessment tasks and practices and to developing the leadership capacity of teachers. After the initial CATCH institute, teachers implement new assessment tools and share the results with their district team. These resources and lead teachers' experiences in developing principled formative assessment methods form the practical basis that district teams use to organize and conduct local CATCH institutes for a new cohort of teachers, beginning with colleagues from the schools of lead teachers. *Framework for Classroom Assessment in Mathematics* (de Lange, 1999) serves as the theoretical basis for these institutes, and *Great Assessment Problems* (Dekker & Querelle, 2002) provides practical examples to promote teacher discussion of design principles for classroom assessment. (An outline of *Great Assessment Problems* is available online at http://www.fi.uu.nl/catch/products/GAP_book/intro.html.)

Underlying the organization of CATCH activities are four related goals for teacher engagement to support the assessment of student understanding. These categories are identified as *initiate*, *investigate*, *interpret*, and *integrate*. Figure 13.1 portrays these four interrelated categories of teachers' formative assessment practices as a *professional development trajectory*, noting for each the objective, activities provided, and intended outcomes for students and teachers. It is important to note that this trajectory does not represent a lockstep sequence of activities. The collective beliefs and interests of participating teachers in each district inform the design of professional development activities and the categories of formative assessment that need to be addressed. The intent of Figure 13.1 is simply to highlight the shift in program emphasis as a teacher cohort grows in understanding and practice.

Figure 13.1. Professional development trajectory.

Initiate	Investigate	Interpret	Integrate
Initiate teacher understanding and critique of conventional assessment practices	Engage teachers in selection and design of principled assessment instruments	Support teachers' principled interpretation of student work	Support teachers' principled instructional interventions
Tools & Activities			
• Critique of "expert" assessments (i.e., common standardized tests) and common classroom assessments • Experience with other assessment methods (i.e., informal classroom assessment, use of Level 2 and Level 3 tasks, and other example from *Framework for Classroom Assessment* and *Great Assessment Problems*)	• Assessment pyramid • Teacher use of *AssessMath!* and other assessment resources • Examples of Level 1 and Level 2 tasks • Technical support for design and adaptation of assessment instruments	• Domain-based scoring institutes • Teacher design and use of analytic and holistic scoring rubrics • Teacher collaboration on rubric design and scoring student work	• Investigation of domain-based assessment trajectories • Use of student verbal and written reasoning • Use of student misconceptions to guide instruction • Discussion of video case studies • New development cycle of balanced assessment materials • Analysis of student representations to discern student capacities and address the needs of more students
Outcomes			
• Teacher dissatisfaction with current methods • Teacher classroom experimentation • Teacher use of other formative assessment methods • Teachers developing a sense of the pros and cons of using other methods in *their* classrooms • Beginning of student learning with understanding	• Teacher categorization of tasks • Teachers designing Level 2 and Level 3 tasks • Teachers designing and using classroom instruments to assess a wider range of student thinking • Teacher development of nonconventional assessment instruments • Teachers gaining classroom experience with Level 2 and Level 3 tasks • Students learning to reason mathematically, use mathematical models, and generalize	• Teachers learning to interpret student reasoning and use different student work as starting point for instruction (i.e., teaching for student understanding) • Increased use of student argumentation • Student opportunities to "show what they can do"	• Greater instructional flexibility • Student understanding, rather than just reproduction, of mathematical content • Greater student involvement in the assessment process • Greater use of peer assessment • Opportunities for student self-assessment • Increased use of student argumentation • Increased learning opportunities for lower-achieving students

Initiate

Professional development activities in this category are oriented toward initiating teacher understanding and the critique of conventional assessment practices. Teachers critique "expert" assessments, such as commonly used standardized tests and conventional classroom assessments. Teachers also engage in other assessment methods, as students, and respond to tasks that require Level 2 and Level 3 reasoning and to other examples provided in *Framework for Classroom Assessment in Mathematics* (de Lange, 1999) and *Great Assessment Problems* (Dekker & Querelle, 2002). The expected outcomes for this category are teacher dissatisfaction with current assessment methods and reflection on the pros and cons of their own assessment methods, both of which lead to experimentation with formative assessment techniques. As teachers change their classroom assessment goals, an expected outcome is students' shift toward learning mathematics with understanding.

Investigate

Professional development activities in this category are designed to engage teachers in the investigation, selection, and design of principled assessment techniques. Using a model for categorizing tasks and examples of Level 1, 2, and 3 questions (for example, see *AssessMath!* [Cappo, de Lange, & Romberg, 1999]), teachers develop practical expertise in selecting assessment tasks and experiment with designing tasks and balanced tests to assess student understanding. The *AssessMath!* software includes an interactive collection of assessment tasks classified by content, competency levels, and grades. As teachers design and adapt assessment instruments to assess particular content, they receive technical support from colleagues and the research team. Expected teacher outcomes for this category include teacher classification of tasks, greater use of Level 2 and 3 tasks, design and use of assessments with a greater balance across reasoning levels, and use of other assessment instruments (e.g., two-stage tasks, projects, writing prompts). The expected student outcome is that students will learn to reason mathematically, use mathematical models, and generalize.

Interpret

Professional development activities in this category support teachers' principled interpretation of student work. Teachers' development

of shared knowledge is promoted through scoring institutes, design and application of holistic and analytic scoring rubrics, and discussions of student representations. To support domain-based formative assessment, activities are structured to promote teacher discussions of student work (e.g., scoring student work). Through these activities, teachers improve their ability to interpret student reasoning and begin to use student work as a starting point for instruction (i.e., teach for understanding). This leads to increased use of student argumentation and the opportunity for students to "show what they know and can do."

Integrate

Professional development activities in this category are designed to support teachers' principled instructional interventions. Teachers investigate student representations with respect to hypothetical assessment trajectories. Video selections of classroom practice are used to broaden teachers' awareness of assessment opportunities such as instructionally embedded assessment. For example, the *Modeling Middle School Mathematics* (Bolster, 2002) video series provides useful examples of teacher–student interaction in classrooms using middle grades reform curricula. Teachers investigate ways to devolve a greater share of the assessment process to students, through principled use of peer and student self-assessment. Workshops and monthly meetings offer opportunities for teachers to further develop domain-based assessment trajectories. By integrating assessment and instruction and investigating ways to use student written and verbal thinking to inform instruction, teachers develop an informed basis for making instructional decisions and eventually show greater adaptability in their lessons, transcending the textbook's lesson to address connections among mathematical concepts in students' terms. These developments lead to greater student understanding of mathematics and improved student achievement.

ENDING NOTE

Our experience shows that teachers benefit greatly from exploring such tacit features of classroom assessment as the design of assessment tasks, the interpretation of students' written and verbal responses, and strategies for eliciting or responding to student ideas during the course of instruction. Yet changing teachers' assessment practice requires more

than providing them with a new set of assessment tasks or a new scoring rubric. Teachers must be motivated to change.

- They must recognize the limited information their current assessment practices provide.
- They must realize the necessity of using tasks and practices that can reveal student understanding.
- *And* they must view teaching for student understanding as an important goal.

In the Research in Assessment Practices study, we found that teachers could learn to use formative assessment practices as a consequence of appropriate professional development and, over time, both develop a more comprehensive view of assessment as an ongoing process and use a wider range of assessment strategies. We also have found that administrators and teachers essentially agree on the value of a professional development program oriented toward improvement of classroom assessment. Those involved in our studies see the CATCH project as an opportunity to promote student-centered practice and improve student understanding of mathematics through the project focus on interrelated principles of assessment design, interpretation of student work, and instructional decision making.

In the CATCH program, professional development activities are grounded in the theory and practice of classroom assessment. Teachers developed a broader view of assessment and used guiding principles for teaching for student understanding in their selection of assessment tasks, instructional activities, and the questions they asked during instruction. Teacher inquiry of student understanding, through the exploration of classroom assessment principles and practices and supported by ongoing collaboration with colleagues, provided teachers the beginnings of a theoretical foothold to construct classroom assessment practices more conducive to student learning. CATCH teachers and administrators also participated in regular discussions of practical ways to monitor the implementation of their district curriculum standards and developed reasoned ways of judging and interpreting student performance on district assessments and state-mandated standardized tests. We note that the administrators and teachers involved in CATCH were *motivated* to make those changes.

To replicate this type of assessment-based professional development program, teachers and administrators must be receptive to a long-term, collaborative endeavor that focuses on assessment design, interpretation of student work, and formative assessment practices to support

student learning. By *long-term*, we mean both that the program should last several years—until the practices become *self-sustaining*—and that the goals for students are long term. By *collaborative*, we mean that there is a mutual relationship between the professional development staff and the teachers and that information and resources provided by the professional development staff respond to the needs and gradual development of the teachers as they change their formative assessment practices.

Participating teachers also need to commit themselves to participating in workshops and school-based professional development activities, documenting changes in their formative assessment practices, and reflecting on the impact of these changes on other instructional practices. Teachers need to be willing to share the assessment instruments they use, participate in the development of assessment materials, and reflect on and critique their own developing practices and those of their colleagues.

As Lorrie Shepard (2000) argued in her presidential address to the American Educational Research Association, "In order for assessment to play a more useful role in helping students learn, it should be moved into the middle of the teaching and learning process instead of being postponed as only the end-point of instruction" (p. 10). Assessment, particularly when situated as a necessary component of the teaching and learning process, *can* be used to bridge the often separate processes of educational practice, but it is unrealistic to expect teachers to become spontaneous assessment designers. Teachers need appropriate professional development support. A professional development program oriented toward the study and development of classroom assessment will need to become a necessary component of any district's effort in order for teachers to achieve teaching for understanding.