Standards-Based Mathematics Assessment in Middle School

Rethinking Classroom Practice

Thomas A. Romberg
EDITOR

Teachers College, Columbia University
New York and London

Chapter 12

Retracing a Path to Assessing for Understanding

Teresa Her
Savannah Oaks Middle School, Verona, WI

David C. Webb
University of Wisconsin–Madison

To what can we attribute primary credit for the instigation of a major change in educational perspective or practice? What patterns or paths exist to facilitate successful learning and growth in practicing educators? What catalysts can we identify to motivate and support pedagogical reform?[1] These are critical questions for any who are invested in the classroom experiences that shape student learning, and the simplicity of an anecdote can provide powerful insight toward potential solutions. This chapter retraces my experiences and experiments with developing assessment tools and practices consistent with my goal to teach for student understanding. I hope that these reflections on the process of change in practice can reveal potential models for practical pedagogical change.

A DEVELOPING PERSPECTIVE ON THE ROLE OF ASSESSMENT

As a preservice teacher, I was inundated with books and articles about "new assessment" techniques (which primarily emphasized the use of portfolios) but was still unprepared for actually assessing and assigning grades to my first class of students. Student teaching experiences

Standards-Based Mathematics Assessment in Middle School. Copyright © 2004 Teachers College, Columbia University. All rights reserved. ISBN 0-8077-4481-6 (paper), ISBN 0-8077-4482-4 (cloth). Prior to photocopying items for classroom use, please contact the Copyright Clearance Center, Customer Service, 222 Rosewood Drive, Danvers, MA, 01923, USA, telephone (978) 750-8400.

in elementary school exposed me to a variety of report forms that were multicategorical and descriptive, unlike the single letter grade and prewritten comment I was expected to select as a middle school teacher. While I disregarded the elementary format as time-consuming and subjective, I had no solid framework of my own from which to build an assessment system. I did not have the time, energy, or confidence as a first-year teacher to deviate from the norm. As a result, initially I relied on the system used by my colleagues that converted total points in each of five categories (tests and quizzes, homework, classwork, participation, and behavior) into percentages that were weighted and averaged for a final percent-based grade. The district grading scale completed the process, translating the percentages to letter grades. Assured that I was in good company with this grading procedure, I tried to ignore the uneasiness I felt while completing each grading sheet with a single letter to summarize a student's entire quarter of work and learning.

During my second year of teaching, I transferred to a newly created position in a seventh- and eighth-grade class of 25 at-risk students. The feelings of resentment and failure that students brought to my class, coupled with my uneasiness with my grading practices, motivated me to find ways to rekindle student interest in learning through innovative changes in assessment. After much debate and research into successful "at-risk" teaching practices, my teammate and I sought to engage our students through an assessment model that used student conferences to negotiate individual goals, and portfolios to demonstrate progress toward those goals for a pass-or-fail grade. This attempt had mixed results: Some students thrived in the freedom of a gradeless environment and made honest attempts to attain their goals; others manipulated the system by resisting goal setting from the start or by producing empty tokens of achievement to meet minimum portfolio requirements. In addition, setting individual student goals (and the accompanying individualized instruction) presented a significant challenge. The class was disbanded after a year, and I spent the summer preparing to teach 80 sixth graders again.

Unwilling to return to sixth grade with the same assessment practices, I sought the help of our school's learning resource coordinator, who enrolled me in conferences and showered me with books and articles. She also served as a sounding board and critical evaluator of ideas and solutions, while at the same time encouraging me to learn from experience. For example, one such attempt at changing practice found my teammate and me experimenting with portfolios and goal setting as a means to determine grades. Rather than averaging a collection of scores,

we required students to document attainment of mutually agreed-upon goals in a portfolio format for a pass-or-fail grade. We found this approach had similar mixed results.

More than anything, I wanted to be fair and honest with the students and respectful of the work that they were doing, honoring progress as well as finished products. During the school year, I had open discussions with students to ask what they thought about school, grades, and report cards, and what they would like to see changed. I was eager to use their ideas to increase their sense of ownership. I wanted to create a unique assessment system that combined aspects of goal setting and self-assessment with more uniform standards of achievement. Inherent in this system was the recognition of learning as a process and greater attention to student improvement. Improvement and growth were measured by comparing scores on a single assessment piece for each unit, administered first as a pretest and again as a posttest.

Despite noticeable progress in garnering student interest and learning, the progress-based system began to lose its appeal after the first year of implementation. For one, I recognized that I still depended primarily on percentages. The goals that students set were arbitrary in that they were number amounts rather than actual items or skills they wished to learn or improve. Furthermore, this approach did not detail where the students had improved, nor did it provide adequate feedback to parents, students, or myself.

I remained dissatisfied.

AN OPPORTUNITY TO RETHINK CLASSROOM PRACTICE

Some of my colleagues, who both followed and challenged me in my ongoing search for alternative grading and instructional practices during the previous year, were involved with a university design collaborative whose aim was to improve student understanding of and performance in mathematics at the middle level (Romberg, Webb, Burrill, & Ford, 2001). After piloting new instructional materials and meeting with project staff, colleagues shared their successes and encouraged me to join their discussions. As a result, I attended several meetings and was introduced to *Mathematics in Context* (MiC; National Center for Research in Mathematical Sciences Education & Freudenthal Institute, 1997–1998). I spent the summer evaluating the materials and comparing them with the district scope and sequence for my grade level to identify overlaps or gaps in content. I found that MiC adequately

covered the required content and often went well beyond grade-level expectations. I was intrigued with the unique problems that were designed to emphasize questioning, reasoning, and communication. Inherent across MiC units were opportunities to use students' informal knowledge as scaffolding to develop well-connected understanding of formal knowledge and algorithms. I decided to use MiC the following year.

During my fifth year of teaching, I grappled with the changes in pedagogy necessitated by a curriculum that emphasizes teaching for understanding. Through frequent observations, one-on-one discussions, and small-group meetings, colleagues and university researchers constantly posed questions that compelled me to defend, rethink, explain, or define what I was doing and why I was doing it: What evidence of student learning is available during the lesson? What questions could be asked to gather evidence of student understanding? What is the purpose of asking students to explain their thinking? What is a good explanation? And so forth.

During monthly meetings, we struggled through challenging mathematical problems and were encouraged to develop an appreciation for student representations and mathematical communication. Teamwork, communication, and rich conversations moved from being simply "warm and fuzzy" teaching methods to being valued as necessary components in the construction of students' understanding of mathematics. We also came to view classroom discussions as a legitimate opportunity for classroom assessment.

Several members of the research team also challenged the tacit limitations I had established for student achievement. The questions and problems I routinely used, assessed students' recall knowledge but were insufficient for assessing students' conceptual understanding. The researchers encouraged me to increase the quality and complexity of the content that I introduced, even as they taught me the mathematics necessary to do so. I was equally affected by the uniqueness of the assessments in MiC in terms of the ways that they documented student achievement and assessed student progress in understanding algebraic concepts. Getting together with colleagues and researchers to discuss our interpretation of student responses was a learning experience that deeply affected my concept of assessment. Rather than being marked right or wrong, individual problems or related problem sets were evaluated for overall understanding, and results were assigned a value according to problem-specific rubrics. It was the first time I had been encouraged to give points for something beyond just a correct or incorrect answer. The scoring rubrics that we used defined point values

for student communication and progress toward algebraic understanding, sometimes in spite of the minor computation errors that students demonstrated. We also used strategy code rubrics to categorize the various strategies that students used for particular problems.

FROM GRADES TO ASSESSMENT

These collective experiences and experiments in practice eventually led me to seriously contemplate the role that informal assessment techniques should play in instructional planning and classroom assessment. I searched for alternative ways to collect information from students and tried to become more efficient and selective with the information I recorded and my methods for recording it. In part, this shift was motivated by a desire to ease the burden of evaluating and recording every task I required of students. Research staff as well as veteran colleagues made convincing arguments that a greater quantity of scores does not guarantee a greater understanding of students' knowledge (although it can guarantee greater headaches). *I came to realize that I was not morally obligated to assess every task.* The concept that some tasks could be simply "learning tasks" (i.e., tasks used to gather information about students without scoring or grading) was a difficult concept for me to grasp. I needed to let go of the fear that students would feel cheated or let down if the work they had done was not evaluated. If I truly wanted students to own and value their *own* learning, I needed to start practicing the same thing by devoting more time to learning from students instead of giving so much attention to evaluating them.

I also realized that up to this point my questions about assessment had focused not on collecting information but on reporting grades. Until I began caring more about what students actually could do and what they actually knew, I never questioned whether scores and points could tell the story I wanted to tell about each student. My interest in teaching for understanding necessitated a change in what I collected because it reoriented what I valued, what I expected students to show, and what I wanted parents to know. Yet I still held on to prior methods for assigning student grades and, for a time, maintained dual records: one for the abilities that students were able to demonstrate and one for points so that I could compute a grade. If, however, I wanted to find out what students understood and where they were struggling in order to inform my instruction, then why did I need to collect another data set in the form of points? Looking back,

I realized the redundancy of assessing students twice was a manifestation of my unresolved conflict between assessment and grading.

The final catalyst to yet another change in assessment practice was my frustration with a computerized grading program that I had been using to record and calculate student grades. Although it saved time every grading period, it was more difficult to modify for situations in which points did not "add up" to tell the whole story. For example, I refused to give an F to a student who was working hard, making progress, and learning to the best of his or her ability, regardless of what the points added up to. To me, an F represented failure to learn or make progress. In the grading program, though, an F was assigned to any student who attained at or below 69%. What did the 69% mean? I had no recollection of what the points stood for. If I had added a few more easy items to a particularly challenging test, a larger base of possible points could have boosted the overall percentage. It all seemed so arbitrary! I was growing frustrated with adjusting points to tell a more honest and complete story. When I began changing my instruction to teach for understanding—and, by necessity, began wanting to capture students' understanding—I recognized that the percent-based scoring method was the source of my problem. By the end of each grading period, the total points that students had earned told me little about their understanding. Theodore Sizer, in the foreword to *Authentic Assessment in Action* (Darling-Hammond, Ancess, & Falk, 1995), put my struggle into words when he wrote:

> The quality of that feedback must be incisive and apt. Telling the kid, "You got a 57 on Friday's test . . . you gotta do better. . . ." is not much help. Indeed, that 57 may tell us more about the test than the test taker. Understanding that one has not done well on a test is the barest beginning of why one did not do well. The learning is in the substance and barely in the score. (p. vii)

Points and percentages could be assigned and earned so arbitrarily (e.g., correct answers on work that had nothing to do with unit objectives, extra credit, completion of a worksheet, good behavior with a sub, a reward for classroom jobs). Points also could be lost just as arbitrarily (e.g., too much time spent in the bathroom, a simple mistake on a test, a penalty for late work). I was aware of several teachers who routinely subtracted anywhere from 10 to 50% of the total points for assignments that were a day late, regardless of student performance. The final percentages rarely told an accurate story about the mathematics concepts that students actually knew. A student who had little experience with

a concept and struggled on quizzes, but through test retakes, after-school tutoring, and hard work managed to learn the material and demonstrate it on the final test, might still receive a low grade because the one high score was not enough to raise the collection of low scores.

Furthermore, the computerized printouts that listed possible and earned scores for every task left me constantly haggling with parents and students over scores on assignments and tests that were weeks or even months old. The discussions all seemed to revolve around getting enough points for the desired grade and rarely centered on what actually had been learned. When I realized this, a second and more powerful realization was not far behind: If a parent *had* asked what his or her student actually understood or needed help with, I wouldn't have been able to answer. At that point, my growing need and desire to assess what students understood, replaced my need to simply assign a grade for what they had done.

ASSESSING FOR UNDERSTANDING

When I stopped wanting points and started wanting information, I hit a wall. I was learning and applying new ways to gather such information but how could it possibly be recorded? I wanted something simple and efficient that I could use in a typical teaching day. It also had to be accessible to students and to parents because I felt that their involvement was critical.

After a long day of teaching, I sat in a student chair and tried to harness all the assessment-related ideas and frustrations that were swirling in my head. As I corrected a set of tests, I found I could see what the students understood by studying their answers, work, and patterns of correct and incorrect responses. When I attempted to put all of those insights into a composite numerical score, however, the insights and meaning were lost. So I began to wonder: Why combine? Why not eliminate the combining step and retain the distinct insights? If the process of aggregating scores on all the responses together causes the reporting to be meaningless, then why aggregate? I recalled the assessments designed by the university design collaborative that used points, not to total a cumulative score, but to indicate performance of separate skills. I realized I wanted a similar way to use scores to note the skills and concepts that students understood on any given test. To do that, I needed to identify the particular skills, concepts, and "understandings" I was looking for. That thought begged the question, "What *should* I be looking for?"

I turned to my experience on the district Standards and Assessment Steering Committee (SASC). One very useful result of this committee was that each staff member received a list of specific academic standards for his or her subject areas and grade level. While we shared rubrics, watched videos, and studied various epistemologies and methods, we discussed effective and desirable ways to communicate student progress.

As I struggled to pull these pieces into a unified program, I was aided by the listening ear and creative thinking of the co-author of this chapter—a university collaborative project assistant with a passion for assessment who allowed me to think through my ideas verbally and provided additional materials, ideas, and suggestions. We identified other methods that seemed practical and useful: a colleague who marked student work according to a three-point scale of 0 for "clueless," 1 for "getting it," or 2 for "nailed it"; a model for unit planning that had been shared in a recent SASC meeting and emphasized beginning with standards and designing assessments to guide task selection and instructional decisions; the RME trilevel assessment pyramid that served as a model for the collection of tasks that should be used for assessing students (see Figure 1.3 in Chapter 1); and the three-tiered list of unit objectives at the beginning of each MiC teacher guide, which outlined the tasks that could be used to assess each level of the model assessment pyramid (Figure 12.1).

I combined these components to produce a system that met the following goals for my latest method of assessment:

- Identify objectives for three reasoning levels, drawn from standards and unit content, to provide appropriate challenge for a wide range of student abilities.
- Allow for both informal and formal assessment results to be recorded separately for each objective, using marks for "not yet," "in progress," or "demonstrated."
- Permit students to make multiple attempts at improving and demonstrating understanding; do not give definitive marks until several opportunities have been offered.
- Base letter grades on a performance rubric, not percentages.
- Empower students and parents by making the content objectives as well as the marking and grading process explicit.

I spent a good deal of time fine-tuning a grading system that would transcend percentages by defining the meaning of each letter grade, so that an A truly represented high-level achievement, a C matched

Figure 12.1. Objectives for the More or Less unit. From "More or Less" (pp. xvi–xvii), by R. Keijzer, M. van den Heuvel-Panhuizen, M. Wijers, J. Shew, L. Brinker, M. A. Pligge, M. Shafer, & J. Brendefur, 1998, in National Center for Research in Mathematical Sciences Education & Freudenthal Institute (Eds.), (1997–1998), *Mathematics in Context*, Chicago: Encyclopædia Britannica. Copyright © 1998 by Encyclopædia Britannica. Reprinted with permission.

In the *Mathematics in Context* curriculum, unit goals, categorized according to cognitive procedures, relate to the strand goals and to the National Council of Teachers of Mathematics Curriculum and Evaluation Standards. Additional information about these goals is found in the *Teacher Resource and Implementation Guide*. The *Mathematics in Context* curriculum is designed to help students develop their abilities so that they can perform with understanding in each of the categories listed below. It is important to note that the attainment of goals in one category is not a prerequisite to attaining those in another category. In fact, students should progress simultaneously toward several goals in different categories.

Conceptual and procedural knowledge

1. Use estimation strategies to multiply fractions and decimals.
2. Multiply fractions and decimals.
3. Use number sense to multiply two decimal numbers.
4. Find a percent of a number.
5. Calculate discount and sale price.
6. Compute the total cost with tax.

Reasoning, communicating, thinking, and making connections

7. Relate percents to fractions and decimals.
8. Find an original price using the sale price and the percent discount.
9. Develop number sense.

Modeling, nonroutine problem solving, critically analyzing, and generalizing

10. Solve percent increase and decrease problems.
11. Explore percents as operators.

average performance on grade-level work, and an F was reserved for cases in which the student failed to demonstrate any learning or progress. This redefined grading system combined elements of de Lange's trilevel assessment pyramid and reasonable expectations for student work habits.

I tested the new system with a representative sample of 15 students. I piloted the "rubric system" concurrently with the "total points system" to compare the final grades from each. This way, I could decide which system yielded a more accurate portrait of student achievement. I also hoped to work out any unforeseen kinks before introducing the new system to all of my students.

I was pleased with the results. Students who received As in the new system because they understood all the material and had pushed themselves "above and beyond" in their learning also received As in the total points system. Students who received As because they had accumulated enough points or were well behaved but did not demonstrate a high degree of understanding, however, received a slightly lower grade in the new system. Conversely, very few of the students who previously had received Fs based on total points received Fs in the new system because so few fit the new definition. Moreover, when student performance was inconsistent and did not fit neatly within the confines of one grade category, I circled the components of each definition demonstrated by that student and assigned a grade that fell in the middle of the range. Even though I would much rather have let the circled definitions tell the story themselves and not have been bothered with trying to pick a letter grade in the middle, I was still required by the district to give a final grade to summarize student achievement. At least, I reasoned, with the rubric attached to the report card, parents would have a clearer picture of what contributed to the grade received. By having a list of objectives, parents and students could pinpoint the mathematics skills and concepts that required additional attention.

Satisfied with the results of the pilot, I began to create a list of objectives for the next MiC unit. I notified parents of the new grading system and explained to students the rationale behind my decision. Most students were interested in giving the new system a try. They enjoyed seeing up front the concepts that they would be investigating, and the process for determining their grades was now demystified. Some students were motivated to set personal goals and familiarize themselves with the requirements for earning the grade that they desired. Each student was given a copy of the unit objectives and the

grading rubric, so students could keep track of their progress throughout the unit (goals they had demonstrated, what remained yet to be demonstrated, and their current grade). An example of the student record sheet is shown in Figure 12.2.

I organized my grade book by main content objectives, with space to record student performance as 0 (not yet), 1 (in progress), or 2 (demonstrated). I heeded caution to keep the assessment load manageable by limiting the number of times I formally assessed each objective (one or two tasks or opportunities per objective), especially as I had made a commitment to trust in-class observation and other informal assessment strategies as valid information. Also built into the system was the opportunity for retakes or "prove-its," which were opportunities for students to conference with me about what was incorrect and to redo a similar task to prove that the concept was understood. This approach placed more of the burden of proof on the student to change unsatisfactory performance and less on me to test each item repeatedly for progress.

IMPACT AND CHALLENGES OF IMPLEMENTATION

Once the system had been implemented, the students and I continued to dialogue about its impact on their learning so that minor adjustments could be made if or when they were needed. David Webb sought additional student responses by conducting a series of confidential student interviews. David and I also continued to share observation notes and discuss the progress and changes that I was experiencing both in the classroom and in time spent planning, preparing, and grading. Based on these discussions, interviews, and observations, I identified six areas of major impact related to the implementation of this new system.

1. Increased student involvement, accountability, and self-advocacy.
2. Change in the content and quantity of teacher–student communication.
3. Raising of the academic "bar."
4. Increased attention to planning instructional units.
5. Major revisions in the design and content of formal assessments.
6. Ongoing adjustments to the goals and methods of in-class instruction.

I describe these changes in more detail in the sections that follow.

Figure 12.2. Student record sheet for the More or Less unit.

More or Less Student Record Sheet

0	1	2
Not yet demonstrated	In progress	Demonstrated

Basic Skills Level (Recognize, recall, identify, accurate use of basic rules)			
1. Reasonable estimations (multiply fractions and decimals)	0	1	2
2. Accurately multiply fractions and decimals	0	1	2
3. Find the percent of a number	0	1	2
4. Calculate discount and sale prices	0	1	2
5. Compute total cost with tax	0	1	2

Application Level (Reason, communicate, connect, apply, solve)			
6. Articulation of estimation and calculation strategies	0	1	2
7. Relate percents to fractions and decimals	0	1	2
8. Work backward to find starting price or discount	0	1	2

Analysis and Extension Level (Interpret, analyze, draw, and justify conclusions, construct informed opinions, extend and generalize)			
9. Compound interest (rule of 72)	0	1	2
10. Refuting common misperceptions with percent	0	1	2
11. Finding a percent of a percent (forward and backward)	0	1	2
12. Extended thinking	0	1	2

Work Habits			
Timeliness	0	1	2
Quality	0	1	2

Participation			
Math chats; class discussion	0	1	2
Readiness	0	1	2

Grading Rubric

A	B	C	D	F
• All basic skills demonstrated; most applications demonstrated	• All basic skills demonstrated; most applications at least in progress	• Most basic skills demonstrated	• Most basic skills at least in progress	• No evidence of any progress or understanding at any level
• Some extensions at least in progress; work habits and participation are exceptional	• Work habits and participation are consistent	• Work habits and participation are somewhat consistent	• Work habits and participation are inconsistent	• Work habits and participation are nonexistent

Increased Student Involvement, Accountability, and Self-Advocacy

One immediate benefit of shifting from an emphasis on grades to an emphasis on assessing understanding was the impact it had on student communication and involvement. The student record forms, progress reports, and class conversations about assessment played a major role in increasing student fluency in the language of mathematics so that students were able to communicate the specific skills that they were learning and the concepts that they did not understand. They also were able to communicate what they were learning to parents and others. Even without the benefit of the forms in front of them, many students were able to articulate a clear understanding of and appreciation for the assessment process. The student responses in the following interview excerpts illustrate some student perceptions of this approach:

Interviewer: Describe how things have changed between the old system and the new system.
Student: In the old system, she would like give you sheets and stuff to work on, and then she would score it. So if it was late, you would still get a point off whether or not everything was right on it. And now she is looking for how well you have demonstrated your skills. . . .

Student: Sometimes before, in math, I would—like—not understand things . . . I would understand them but I just didn't understand the problem or something. But now, if I mess up on something, and she already knows that I have proved it, and she knows that I can do it, it is not going to reflect badly on the rest of my stuff.
Interviewer: It is just sort of one mistake.
Student: One mistake. Our old grading system was like—the whole thing was like earning certain points, and then you see how much you can get, and then it ruins your score if you mess up on something.

Student questions like, "What is my percentage today?" or "How many extra credit points do I need to get an A?" were replaced with, "When can I demonstrate that I now can subtract fractions with unlike denominators?" or "How can I extend my thinking to go beyond the basic skills?" Parents stopped calling to haggle over points for lateness or lower-than-expected test scores. Students started advocating for themselves

more, and they wanted to learn and demonstrate specific skills. As one student remarked, "We are more concentrated on learning and not on getting a good grade."

Change in the Content and Quantity of Teacher–Student Communication

The student self-advocacy mentioned above was one unexpected but much welcomed change in teacher–student communication. Another was the increase in communication necessitated by students' desire to receive help and to demonstrate the objectives they missed on formal assessments. Previous quests for points or percentages never elicited the requests that I received from students to demonstrate learning goals. As a result, I decided to institute "math chats."

Similar to the fun book chats my language arts teammate had developed as a more interactive alternative to regular book reports, math chats were organized during the mandatory tutorial period or after school as informal opportunities for students to demonstrate achievement of learning goals and verbalize higher-level thinking. During a session, I would sit with a small group of students to discuss either an area of mathematics they were struggling with or an enrichment activity they wanted to explore further. Students established the goal of the session. Sometimes students selected the math problems they wanted to discuss, and other times I would select "brain stretchers" to push students to extend and apply their mathematical knowledge in new ways. I quickly found that math chats gave me additional insight into students' math thinking, abilities, struggles, and comprehension in lieu of constant test writing and grading. Math chats also were an opportunity to try out questioning techniques and instructional strategies to help students overcome challenging content, techniques that I would use in later lessons. For example, for a group of students who were struggling with addition of fractions, with only a few questions I found that their struggle was due to a limited understanding of the meaning of numerator and denominator. After briefly discussing a fair-share activity with sub sandwiches, students had a better sense of why denominators could not be added. This principle of briefly reviewing fundamental concepts was used in later lessons and other classes to the benefit of other students. Surprisingly, math chats became a popular "hangout" for many students, and most students attended them at least once a week.

Although the move to incorporate math chats into my day amounted to an additional investment of time, often cutting into my prep and lunch periods, it was an instructional opportunity that held great benefits not just for the students but also for me as their teacher.

The small size of the group led to a style of communicating that was much more personal and revealing than in whole-class discussions. The additional time spent with students formed bonds that enhanced instructional time with the whole class. Not only did I get to know the students' math needs better, informing my instructional decisions, questioning, and pacing, but I got to know students better as people. I found that students were more connected with me and with the material I was teaching, and the improved behavior in the classroom was a testament to the increased respect that had grown out of our time together. Likewise, I felt a greater sense of connection with and understanding of their concerns, struggles, strategies, and successes, and I was able to incorporate them into class discussion for a richer dialogue and more personalized instruction.

Raising the Academic Bar

Although increased student involvement and a positive rapport with students contribute to a productive learning environment, students' interest in learning challenging material gave me further motivation to raise the academic bar for all students. Once I determined the three levels of reasoning I wanted to address, I was obligated to plan tasks and activities that could elicit such thinking. I was required, as a result, to raise my own awareness and knowledge of math content so that I could offer challenging analysis-level extension tasks to students who wanted to press beyond basic grade-level requirements. Students continuously rose to meet these challenges and attempted and achieved more than ever before.

Interviewer: What is the difference between a basic skills question and an extension question?

Student: Basic skills is where you have to know it and it is something that the unit is spending a lot of time on. And an extension question is [where] Miss Her has just briefly touched it or there has been a question on the test to see if you did it. An extension is beyond what she expects but for people who want something to challenge them. I think everybody should at least try the extension question because you might surprise yourself, like I did.

Interviewer: And so you tried extension problems?

Student: I always tried them. I didn't get compound interest but then I wanted to show her that I understand it now. . . . With extensions, I think everybody should try them because the ones I didn't think I could get, it turns out that I did understand it.

Significant Change in Planning Instructional Units

Beginning unit planning by listing the desired objectives and outcomes may sound like a reasonable and organized approach, but, to be honest, I was not in the habit of doing it. As a preservice teacher in college, I remember being required to list objectives for the lessons I designed, but more often than not those objectives were filled in after the bulk of the lesson had been designed, so that they would define and support the tasks that had already been chosen. Not long into my first year of teaching—having spent long nights writing out every single lesson with separate purposes, methods, and activities—I chose to forego the tedious lesson-planning process and rest in the assurance that I knew generally what I was teaching, what the district expected, and the outcomes I wanted. I also felt that I could produce this information on demand. The shift toward naming the skills and objectives of a unit before choosing the activities—with the intent of directing all efforts and activities to the end goal of developing or mastering these objectives—marked a significant change in my practice. The front-end work (i.e., research and preparation prior to the teaching of any lesson) was time-consuming and daunting. The first unit I tried with this new assessment-driven planning method was the More or Less unit (Keijzer, van den Heuvel-Panhuizen, et al., 1998), a number strand unit that emphasizes the application of percentages, fractions, and decimals. In this first run, the knowledge and support of the research team were instrumental in my selection of assessment items to address the three levels of objectives reflecting district and state standards. To my great surprise and pleasure, however, once under way, the days of instruction and assessment flew by smoothly with very little maintenance, and evenings were almost completely devoid of concerns over planning the next day's lesson.

Planning additional units by articulating the learning goals up front was easier the following year. I was more familiar with the curriculum and standards, had a better idea of how students might respond to problems, and had completed the process for several MiC units. I also spent time during the summer making minor adjustments to the previous year's unit goals and collaborated with another sixth-grade teacher to plan other units using this system. In comparison, I noticed that many of the supplementary lessons I taught under the "old system" of planning and assessment no longer met the recently identified goals and objectives. Some key concepts were not addressed, so I included activities to ensure that students had the opportunity to learn important unit goals. In most cases, the lessons I eliminated involved busywork projects,

repetitive skill-building tasks, and nonessential problems from the unit. Lessons or tasks that needed to be added to what I previously had taught were primarily extension activities and challenges that students needed to demonstrate in order to earn higher grades.

Major Revisions in the Design and Content of Formal Assessments

Formal assessments took on a new look. At the top of each quiz or test, I indicated the objectives that were being tested. Because single questions often offered the opportunity to demonstrate more than one objective, distinguishing the objectives from the questions was an important organizational step and made recording much easier. It also helped to have the unit objectives identified up front so students could connect what they had been learning to what the assessments were asking for. As a result, the purpose of the assessments became clearer than ever—both for the students and for me. I found myself re-evaluating and rewriting assessments to focus on key concepts, to offer legitimate extension problems, and to eliminate any tasks that did not fit the desired objectives. This process made most of my "old system" tests much shorter, more focused, and more concise. It became possible to assess and record precisely what I valued in terms of mathematical content and process, and because the assessment criteria were available to all, students knew ahead of time exactly what was expected (see Figure 12.3).

Ongoing Adjustments to the Goals and Methods of In-Class Instruction

Knowing my objectives clearly from the start focused my instructional decision making so that it was more certain and purposeful. I was able to more quickly assess and choose the most productive paths of discussion, debate, and exploration because I had a better understanding of the core skills and concepts for the unit. This knowledge served as an interpretive background for making instructional decisions. I found myself asking and able to answer such questions as these: Will this sidebar or line of questioning help further our goals? Will this lead to or lay groundwork for future skills? Will further exploration of this question distract us from our primary goals? How can I redirect or use this line of thinking to connect it with concepts emphasized in this unit or with bigger mathematical ideas?

My increased desire for informal indicators of understanding also affected the instructional period. I sought less to give answers and more

Figure 12.3. First skills list for the More or Less unit.

Primary Strand(s): Number operations and relationships (Strand B)

Standards:

Grade 6—Mastery

- Add, subtract, multiply, and divide whole numbers and decimals.
- Use a variety of estimation strategies to solve and check reasonableness of results of computation problems with whole numbers, fractions, and decimals.

Grade 7—Introduction

- Multiply and divide decimals, fractions, and mixed numbers.
- Use proportional reasoning to solve mathematical and real-world problems (e.g., unit rates, equivalent fractions, equal ratios, constant rate of change, proportions, percentages).

Grade 8—Introduction

- Understand how different algorithms work for arithmetic computations and operations. Use appropriate computational methods (e.g., mental, paper and pencil, calculator, computer, spreadsheet) for situations with rational numbers.
- Perform operations on rational numbers (add, subtract, multiply, divide . . .). Understand the concept of proportion and the applications of proportional reasoning (e.g., scale, similarity, percentage, rate).
- Apply proportional thinking in a variety of problem situations that include
  - ➢Ratios and proportions (e.g., rates, scale drawings, similarity).
  - ➢Percents, including greater than 100% and less than 1% (discounts, rates of increase and decrease, sales tax).

Curricular Goals and Objectives:

Basic Skills Level

1. Reasonable estimations (multiplying decimals and fractions).
2. Accurately multiply fractions and decimals.
3. Find the percent of a number.
4. Calculate discount and sale prices.
5. Compute total cost with tax.

Application Level

6. Relate percents to fractions and decimals.
7. Work backward to find starting price or discount.
8. Articulation of various estimation and calculation strategies.

Analysis and Extension Level

9. Compound interest (Rule of 72).
10. Refuting common misperceptions with percentages.
11. Finding a percent of a percent (forward *and* backward).
12. Extended thinking.

to hear strategies and approaches in the hope of identifying students' discoveries and struggles. Conversations and student-led justifications and explanations became more than occasional practices employed for the sake of variety. Instead, they became my primary methods for gauging and documenting student understanding.

REFLECTIONS AND FUTURE IMPLICATIONS

Let me state, first and foremost, that I do not in any way wish to promote the product described herein as the definitive assessment system. Rather, this chapter recounts a journey of experimentation with assessment models and practices that is by no means over. Assessing for understanding and teaching for understanding both rely on dynamic interaction and construction of meaning rather than static procedures and formulas. No system can be canned and reproduced without inviting the resentment from teachers that might result from top-down directives. The process described herein, which involved years of questioning, trial and error, experimentation, conversation, readings, failures, successes, and communication with parents, students, administrators, and colleagues, was driven by a personal quest to assess student understanding.

If it is not desirable to replicate a product, then how might one replicate the process? Although it is difficult to isolate one catalyst as the primary factor for the change outlined in the path I have retraced, I believe I can identify at least three critical contributing components.

Dissatisfaction with the Status Quo

A recurring theme throughout my years of teaching has been continuous experimentation with change motivated by regular re-evaluation of and reflection on what led to a sense of frustration or dissatisfaction with the methods I was using. Prawat (1989), in his writing on teaching for understanding, observes that "[students] must first be dissatisfied with their preconceptions before being receptive to alternative explanations" (p. 321). I would take this observation a step further to assert that for any of us to be receptive to alternatives to the status quo, we first must be dissatisfied with it. Encouraging other teachers to examine their practices critically and engage in active reflection, so as to desire change, is by far preferable and more likely to be maintained than for someone else to be dissatisfied and require teachers to change something that they may be comfortable with. Inherent in this, is the issue of ownership; the intrinsic desire for change that results from personal discovery will always outperform the extrinsic motivation of a top-down mandate.

A Philosophy of Respect for Students

Because true educational reform above all should benefit students, a desire to improve the learning opportunities for students will not be made willingly unless one is motivated by a concern for their welfare. It takes little effort as a teacher to maintain teacher-centered practices. Student-centered teaching and assessing, however, require great effort and perseverance. They also require the willingness to take risks. Unless teachers see improved student performance and engagement as valid rewards, they will have little reason to take on the extra work concomitant with changing familiar practices.

Presence of Support Systems

Even the most motivated teacher can become quickly and easily burned out when attempting to innovate practice in isolation. In my own story, I acknowledge four key aspects of outside support without which I would still be dissatisfied with the status quo and desiring to do better by my students, but trapped in old and unproductive methods of teaching and assessing.

1. *Administrative support.* The administration provided valuable support, granting me permission to experiment with assessment systems, albeit in the confines of my own classroom. Such flexibility was necessary fertilizer for innovative growth. The money and time to attend various conferences and workshops inspired me and fed my ideas for reform.
2. *Collegial support.* Many of the ideas that collided to create my most recent system of assessment were borrowed or built using the advice and work of admired colleagues. Fellow teachers with a love for students shared rubrics they had created or spent time listening to my ideas and sharing their experiences and wisdom. Others who took similar risks and advocated for departmental reforms in tracking, grouping, and instruction inspired my efforts further. Darling-Hammond and colleagues (1995) acknowledged the power of peers when they witnessed that resistant or skeptical faculty were "won over to innovative practices by three factors: the improved quality of student learning and performance they have witnessed; the persuasive arguments of a critical core of peers; and opportunities to collaborate with colleagues" (p. 266).
3. *A "big brother" or, in this case, university support.* The role of the university collaborative in my later efforts at reforming practice was invaluable. Videotaping and personal observations were instrumental

to my growth, making it possible to view my practices with a critical and curious eye. The university research assistants who worked closely with the teachers also provided critical support. Through the mathematics problems they introduced and encouraged us to solve, through their availability to answer questions over the phone and in person regarding the content and skills promoted in MiC, we grew in our knowledge of content and knowledge of student learning—two critical perspectives needed when teaching and assessing for understanding (Ball, 1993). A final benefit of the university collaborative's involvement was the ability of its staff to show us the "big picture" by connecting us with articles, research, and other examples of reform in the mathematics community. As outsiders, they also helped set limits and reasonable goals in order to avoid the dangers of burnout.

4. *Curricular support.* Although quality materials are not a panacea for improvements in math education, they definitely support change in classroom practice. The open-ended and well-designed tasks in MiC lent themselves to richer discussions and explorations than might have been possible with more traditional materials. By modeling and requiring students to make use of alternative strategies and nonroutine problem solving, the materials helped to create an environment that encouraged teachers to do the same.

For me, the quest to change classroom assessment was a complex journey requiring a great deal of reflection, planning, and experimentation. There were potholes and dead ends along the way. The journey was not overnight, but one that spanned many years and continues to unfold. And while others might be willing to brave such an adventure solo, in my experience it would be wise to consider sharing the journey with interested colleagues, who can provide alternative perspectives for proposed ideas and offer insight on practical ways to put those ideas into practice.

NOTE

1. Although co-authored, this chapter is written in the voice of teacher Teresa Her.