1 ❖ Issues Related to the Development of an Authentic Assessment System for School Mathematics

Thomas A. Romberg and
Linda D. Wilson

In 1990 the president of the United States and the National Governors Association announced their unprecedented agreement on national educational goals. For the nation to achieve those goals, it has become apparent that the American education system must be restructured. The strategy now being followed involves a series of steps to produce: a detailed set of content standards in English, mathematics, science, history, and geography; a set of standards describing how best to instruct students toward the attainment of each of those content standards; a set of procedures to assess student progress in meeting the content standards; and a set of standards to describe the responsibility of the professionals who will assist students in reaching those standards. The framework for restructuring schools via the specification of content standards, teaching standards, performance standards, and their interrelationships is shown in figure 1.1.

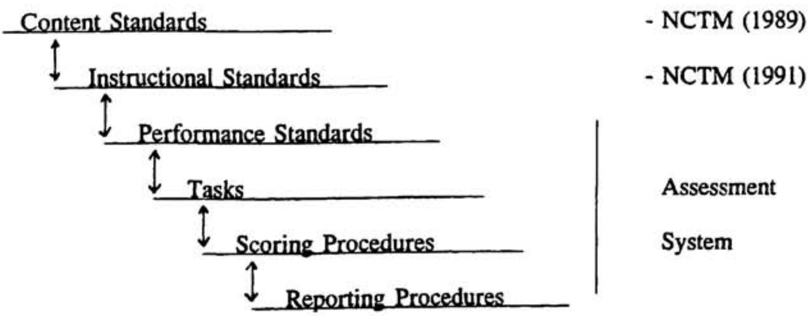

Figure 1.1. Relationships between content, instructional, and performance standards and an assessment system.

As shown in the figure, the initial stages of the framework have been addressed in mathematics. The *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics, 1989) presents a consensual vision of the mathematical content that all students should have an opportunity to learn—the *content standards* in figure 1.1. Furthermore, the *Professional Standards for Teaching Mathematics* (National Council of Teachers of Mathematics, 1991) describes the means for assisting students to learn that content—the *instructional standards* in figure 1.1. In addition, some of the needed elements for the development of an assessment system are described in the *Curriculum and Evaluation Standards.* NCTM has now taken on the task of producing "assessment standards." They are due to be published in 1995.

The progress made in the content area of mathematics is being held up as exemplary for the other four disciplines. Nevertheless, a number of critical issues need to be addressed if the assessment system is to fit with the vision of the work done so far. In addition, a great deal of work remains to be done in the development of performance standards, assessment tasks, scoring, and reporting procedures. In this chapter, we identify a number of issues that we see as especially significant. Beginning with some assumptions upon which the reform movement in mathematics is constructed, we discuss what needs to be considered at each stage of development of an assessment system for mathematics for the result to be considered "authentic."

ISSUE 1: UNDERLYING ASSUMPTIONS ABOUT THE NATURE OF MATHEMATICS

An authentic assessment system for school mathematics should begin with a vision about the nature of mathematics that is aligned with current thinking. An assumption underlying the development of any assessment system is that the responses a student makes to a set of test items or tasks will be a valid indicator of that student's understanding of some aspect of a domain of mathematics. There are three fundamental problems with this assumption. First, as Antoine Bodin (1993) has argued, one can never know what a student truly understands. One can only make inferences based on the responses a student makes to the tasks administered. This implies that the creation and selection of tasks is critical to the assessment process; in particular, they must reflect important aspects of mathematics a student has had an opportunity to learn. The second problem involves the reliability of the responses to those tasks so that a reasonable indicator of a student's understanding can be inferred. Together these lead to the final problem: What does one mean by *an understanding of mathematics?* The new and emerging answer to this question is at the heart of the calls to develop an "authentic" assessment system. To clarify this issue, we have chosen to describe the classical testing paradigm followed in the United States and point out its weaknesses with respect to current notions about the nature of mathematics.

Traditional norm-referenced standardized achievement tests for mathematics are created by following a particular measurement model. Such tests are made up of an assortment of independent, discrete questions that can be responded to quickly; all items are assumed to be equivalent; answers (usually derived by choosing among alternatives) are judged to be either correct or incorrect; and responses should be internally consistent, reflect important variations in responses between students, and be fair to all examinees. Such tests resolve the three problems mentioned previously by selecting or creating items that reflect specific concepts or procedures that appear in widely used textbooks; carefully considering a logical, hierarchical sequence of concepts and procedures; and having a group of teacher and mathematics educators judge their face validity. Reliability is established first by eliminating items that are too easy, too hard, or do not correlate with other items; and then an internal consistency coefficient is calculated. Finally, counting the number of correct responses on a test constructed in this manner is assumed to be a reasonable indicator of a student's knowledge, and differences in the number of correct responses among students is assumed to reflect differences in knowledge.

The calls to develop an "authentic" assessment system are based on the conviction that counting the number of correct answers to a series of brief questions contradicts current views of mathematics as an intellectual discipline. Ernest (1991), for example, argues that mathematics cannot be described by a single unique hierarchical structure and that mathematics cannot be represented as a set of discrete knowledge components. The mathematician William Thurston (1990) uses the metaphor of a tree to describe mathematics: "Mathematics isn't a palm tree, with a single long straight trunk covered with scratchy formulas. It's a banyan tree, with many interconnected trunks and branches—a banyan tree that has grown to the size of a forest, inviting us to climb and explore" (p. 7). A valid system for assessment in mathematics must reflect these notions—that mathematics is a set of rich, interconnected ideas. To be in line with current thinking, it must view mathematics as a dynamic, continually expanding field of human creation, a cultural product (Ernest, 1988).

In the NCTM *Standards* (1989), the development of mathematical power is presented as the central goal of school mathematics. *Mathematical power* is defined as the ability to "explore, conjecture, and reason logically, as well as the ability to use a variety of mathematical methods effectively to solve nonroutine problems" (NCTM, 1989, p. 5). The term is based on a recognition that mathematics is more than a static collection of discrete concepts and skills to be mastered. Doing mathematics includes such dynamic and integrative activities as discovering, exploring, conjecturing, sense making, and proving. Students who possess mathematical power should be able to investigate and reason, communicate ideas, and take real contexts of problems into account. The descriptive verbs used in the *Standards* evoke images of mathematics as a progressive human activity.

If one considers mathematics to be a static, linearly ordered set of discrete facts, then the logical choice for a valid assessment system is the traditional standardized achievement test. On the other hand, if one views mathematics as a dynamic set of interconnected, humanly constructed ideas, then the assessment system must allow students to engage in rich activities that include problem solving, reasoning, communications, and making connections.

ISSUE 2: UNDERLYING ASSUMPTIONS ABOUT THE LEARNING OF MATHEMATICS

It is critical that in addition to being based on certain beliefs about the nature of mathematics an assessment system be built on current views of learning mathematics. A recent study by Shepard (1991) showed that approximately half of all district testing directors in the United States hold beliefs about the alignment of tests with curriculum and teaching that are based on behaviorist learning theory, which requires sequential mastery of constituent skills and behaviorally explicit testing of each learning step. Such a learning theory was prevalent for several decades, but is now out of date with current research.

Indeed, as Romberg, Zarinnia, and Collis (1990) noted, the values and forces that dominated mathematics education for the past century (e.g., behaviorism) are embedded in the theoretical structures of prevailing methods of assessment. Tests built on behavioral objectives and a content-by-process matrix are based on behaviorist ideas about learning: that content can be broken down into small segments to be mastered by the learner in a linear, sequential fashion.

Yet a substantial body of evidence from cognitive psychology shows this hierarchical model of learning to be obsolete. The metaphor of the learner as a passive absorber of linearly ordered bits of information is contradicted by research findings from psychology. Resnick (1987) has argued that learning does not occur by passive absorption alone, but rather in many situations learners approach a new task with prior knowledge, assimilate new information, and construct their own meanings. Ernest (1991) has shown that the uniqueness of learning hierarchies in mathematics is not confirmed theoretically nor empirically. Furthermore, he argues against the notion that concepts in mathematics can be either "possessed" or "lacking" in a learner.

The shift in learning theory can best be summarized as a move from behaviorism to constructivism. Though there is not total agreement in the mathematics education community about exactly what a "constructivist" theory of learning entails, Peterson (in press) has described four basic assumptions that form the foundation for current theory, research, policy, and practice in mathematics education:

- Learners are knowledgeable "sense makers."
- Learning involves the negotiation of shared meaning.
- Knowing is contextualized or situated.
- Assumptions about knowledge influence learning.

A more appropriate metaphor for learning may be an image that is gradually brought into sharper focus as the learner makes connections, or perhaps like a mosaic, with specific bits of knowledge situated within some larger design that is continually being reorganized or redesigned in an organic manner. In either case, the emphasis is on knowing, rather than "knowing that." The *Standards* express it as a process: " 'Knowing' mathematics is 'doing' mathematics. A person gathers, discovers, or creates knowledge in the course of some activity having a purpose. This active purpose is different from mastering concepts and procedures. We do not assert that informational knowledge has no value, only that its value lies in the extent to which it is useful in the course of some purposeful activity. It is clear that the fundamental concepts and procedures from some branches of mathematics should be known by all students. Established concepts and procedures can be relied on as fixed variables in a setting in which other variables may be unknown. But instruction should persistently emphasize 'doing' rather than 'knowing that' " (NCTM, 1989, p. 7). Assessment, then, should be based on a view of the learning of mathematics as a socially constructed process, not a fixed hierarchy of skills and concepts to be mastered.

ISSUE 3: THE NEED FOR NEW PSYCHOMETRIC MODELS

"It is only a slight exaggeration to describe the test theory that dominates educational measurement today as the application of twentieth century statistics to nineteenth century psychology" (Mislevy, 1990, abstract). When Mislevy wrote those words in 1990 he was calling for the field of psychometrics to "catch up" with the advances in cognitive psychology. As noted earlier in Shepard's (1991) work, many psychometricians are still operating under theories of learning and measurement that are out of date. New knowledge of how learning takes place must be accounted for in psychometric theory. "Learners become more competent not simply by learning more facts and skills, but by reconfiguring their knowledge; by 'chunking' information to reduce memory loads; and by developing strategies and models that help them discern when and how facts and skills are important. Neither classical test theory nor item response theory (IRT) is designed to inform educational decisions conceived from this perspective" (Mislevy, Yamamoto, & Anacker, 1992). Just as an assessment system must be built on current learning theory, so must the psychometric measurement theories that support such a system be designed with cognitive psychology as its base. Fortunately, work toward new theories of testing is being accomplished, as some psychometricians are now realizing. Wilson (1992) expressed the need this way: "The consequence of this view of learning [constructivism] is that we can no longer use an atomistic model for assessment. We must assess the level of complexity of student understanding, not just the number of facts that students can pick out of a multiple-choice test" (p. 123). New models, such as those described by Mark Wilson in chapter 7, are being constructed that begin to capture more of the complexity of learning than was allowed for by standard test theory. Although no new model claims to describe all of the nuances of current learning theory, with the support of more powerful technologies progress is being made (Glaser, Lesgold, & Lajoie, 1987; Mislevy, Yamamoto, & Anacker, 1992). It is critical that an authentic assessment system take such work into account in its design.

ISSUE 4: ALIGNMENT WITH THE REFORM CURRICULUM

As described in figure 1.1, the first stages of the building of an assessment system for mathematics—that is, setting content and instructional standards—has been accomplished. Consensus has been reached in the mathematics education community about the content that all students should be given the opportunity to learn and the pertinent means of instruction. As the next four stages (setting performance standards, developing tasks, adopting scoring, and reporting procedures) are undertaken, it is critical that the outcomes be in alignment with the conceptualizations of curriculum and instruction set forth in the *Standards*.

The *Standards* are built on a set of assumptions about the nature of mathematics, about learning, and about teaching. As described earlier, mathematics is viewed as a progressive human activity. To know mathematics is to engage in the activities of doing mathematics, such as conjecturing, sense making, and communicating mathematical arguments. Another fundamental assumption is that school mathematics is not solely for the elite, but for *all* students. All students come to school with certain mathematical concepts already forming, and the role of the teacher is to build on that knowledge so that students gain increasing mathematical power. The teacher's role is no longer that of a deliverer of knowledge, but that of a guide and facilitator for student growth.

To be in accord with the work completed thus far on the mathematics curriculum and methods of instruction, the next stages in the development of an assessment system must take these fundamental assumptions into account. This implies, for example, that performance standards should be based on students solving nonroutine problems rather than performing conventional computational procedures. Assessment tasks should allow students the opportunity to demonstrate their mathematical power. They should require the active engagement of students in doing mathematics rather than making a passive response to routine questions. Also scoring and reporting methods should be designed to inform individual students about their own learning rather than to rank students in groups.

ISSUE 5: SPECIFICATION OF PERFORMANCE STANDARDS

The NCTM *Curriculum and Evaluation Standards* (1989) describe what students should have an opportunity to learn. But to establish an assessment system for school mathematics that is aligned with that vision, performance standards must be set that will describe what students are supposed to be know and be able to do in mathematics. Making the connection between curriculum standards and performance standards is a difficult task, but one that needs to be confronted.

A conventional approach to testing specifies both content and levels of performance; it then crosses them to form a "content-by-process" matrix. Although this approach affords test developers the assurance that each content area of mathematics and each level of performance (or type of process required) will be "covered" by the test, the design may also work against the assumptions about mathematics mentioned earlier. That is, separating mathematics into individual cells of "Number" as content and "Computation" as process, for example, sets up a situation for test writers to write items that fit neatly into those cells. The design does not easily allow for items that require more than one content area or more than one process in their solution. As an example, consider this item, similar to items found on recent standardized achievement tests (Romberg & Wilson, 1992):

> Which is best to use to find an estimate for 791 ÷ 19?
>
> A. 700 ÷ 10
> B. 700 ÷ 20
> C. 800 ÷ 10
> D. 800 ÷ 20

This item would fit (all too neatly) into a content area of Number and a process area of Computation/Estimation. When tests are composed primarily of items like these, the underlying assumption is that mathematics is a collection of discrete content areas and that the doing of mathematics occurs in a separate, compartmentalized, hierarchical fashion.

For a task to be considered "authentic," it should not easily fit into neat categories of single content areas and single processes. Solving nonroutine problems usually involves multiple processes and cuts across mathematical domains. Making connections necessarily involves blurring the lines between content and processes. The task (NCTM, 1989, p. 141) presented in figure 1.2 illustrates the interconnectedness of problem solving, communication, and reasoning and involves the content areas of geometry and discrete mathematics.

To build an assessment program in alignment with the NCTM *Standards* implies that all the elements of the program incorporate the four major strands of emphasis: problem solving, communication, reasoning, and connections. This does not imply that every task would necessarily include all four standards, but that the assessment program as a whole would incorporate all four of them at each level. It would also be essential that these four strands not be separated into distinct "process categories," but rather that the items reflect their necessary overlap and interrelatedness.

ISSUE 6: DEVELOPING AUTHENTIC TASKS

Increasing attention is being given to notions of "authentic assessment." Definitions or criteria for authentic assessment are being developed that are built on the framework of the reform curriculum in mathematics education. For an assessment system to be considered "authentic," it must acknowledge these criteria.

Archbald and Newmann (1988) consider three criteria to be critical to authentic assessment tasks: (1) disciplined inquiry; (2) integration of knowledge; and (3) value beyond evaluation.

Nine robots are to perform various tasks at fixed positions along an assembly line. Each must obtain parts from a single supply bin to be located at some point along the line. Investigate where the bin should be located so that the total distance traveled by all of the robots is minimal.

Figure 1.2. An example of a task for grades 9–12 (Reproduced with permission from *Curriculum and Evaluation Standards for School Mathematics*, copyright 1989, by National Council of Teachers of Mathematics, p. 141).

Disciplined inquiry refers to the production of new knowledge, such as that created by scientists or historians. It depends on prior conceptual and procedural knowledge, it develops in-depth understanding of a problem, and it "moves beyond knowledge that has been produced by others" (p. 2). *Integration of knowledge* means that authentic tasks must consider the content as a whole, rather than as a collection of knowledge fragments. Students must "be challenged to understand integrated forms of knowledge" and "be involved in the production, not simply the reproduction, of new knowledge, because this requires knowledge integration" (p. 3). The third criterion, *value beyond evaluation*, refers to the idea that authentic tasks should possess attributes that make them worthwhile activities beyond their use as evaluative tasks. An example would be a task that results in discourse, an object, or a performance. An authentic task might also have value for the collaborative opportunities it provides.

Although these criteria are more broadly based, Lajoie in chapter 2 develops a set of criteria for authentic assessment specifically in mathematics. This framework does not contradict the more general notions of Archbald and Newmann, but makes more explicit the ways in which the content of mathematics influences the design of assessment tasks. It is built on two primary foundations: the NCTM *Standards* (1989) and current learning theory. The *Standards* are predicated on two basic assumptions: the first, that knowing mathematics is doing mathematics, and the second, that there should be four goals for school mathematics content—problem solving, communication, reasoning, and connections. From current learning theory, Lajoie (1991) chooses situated cognition and social constructivism to form a foundation for the definition of authentic assessment.

Building on concepts of mathematics in the *Standards* and on learning theory, Lajoie defines seven principles for a definition of authentic assessment:

1. It must provide us with multiple indicators of the learning of the individual in the cognitive and conative dimensions that affect learning. The cognitive dimensions include content knowledge, how that knowledge is structured, and how information is processed with that knowledge. The conative dimensions should address students' interest in and persistence on tasks, as well as their beliefs about their ability to perform.

2. It must be relevant, meaningful, and realistic. It must be instructionally relevant, as indicated by its alignment with the

NCTM *Standards*. It must relate to pure and applied tasks that are meaningful to students and that provide them with opportunities to reflect, organize, model, represent, and argue within and across mathematical domains.

3. It must be accompanied by scoring and scaling procedures that are constructed in ways appropriate to the assessment tasks.

4. It must be evaluated in terms of whether it improves instruction, is aligned with the NCTM *Standards*, and provides information on what the student knows.

5. It must consider racial or ethnic and cultural biases, gender issues, and aptitude biases.

6. It must be an integral part of the classroom.

7. It must consider ways to differentiate between individual and group measures of growth and to provide for ways of assessing individual growth within a group activity (pp. 30–31).

This set of criteria, in defining authentic assessment in mathematics, could serve as a guideline for an authentic assessment system for school mathematics. It incorporates current learning theories and emphasizes the necessary alignment with the reform curriculum.

The format of authentic tasks may vary. In fact, in keeping with the need for multiple sources of information, no assessment system should be limited to a single form. In chapter 4, de Lange discusses the various formats of mathematical tasks, from multiple choice to portfolios, and sheds some light on what is meant by an "open" item.

ISSUE 7: MEASURING STATUS, GROWTH, OR A COMBINATION

It is clear that all forms of assessment, including traditional standardized tests, are designed to measure the present status of student thinking. Traditional measuring instruments were created to yield highly reliable scores on a single dimension, with the ultimate purpose of linearly ranking students on that dimension. This factorylike image of education belongs to an earlier age when behaviorist theories of learning held sway. Constructivist approaches to assessment require greater emphasis on a developing picture of individual growth. The emphasis has shifted from an industrial model of quality control to an effort to describe an individual's attainment of mathematical power. There is a need for more than status information; instead of a static score, what is needed are profiles of growth over time.

A single score, although useful for ranking students at a fixed point in time, places the emphasis of the assessment on the measure used. When assessment is seen an a means to understanding a student's growth over time, the emphasis shifts to the process of learning. Such a change vastly increases the utility of the information gained for all the audiences involved. Students, no longer limited to information focusing only on comparisons with peers on a single measure, gain an understanding of their own learning. Measures of growth over time are immensely consequential for teachers in planning instruction. As students gain more sophistication in their problem-solving strategies, assessment can best inform students and teachers by describing growth in that ability longitudinally. Parents and administrators also gain a deeper understanding of student learning when information is provided that goes beyond a single static score.

ISSUE 8: SCORING—BY WHOM AND IN WHAT FORM?

Scoring on traditional standardized tests has historically been done by machine, which works well with multiple-choice, single-answer items. But different forms of assessment with open-response items, for example, require professional judgment to score. The issue then is, Can we trust teachers to reliably mark their own students' work, and can they be trained to do so? Other countries have struggled with this issue as well and have responded with a variety of strategies and results.

In The Netherlands, an experiment was conducted to test the reliability of teachers' judgments. Fifteen teachers were asked to score the work of five students on an extended open-ended task. The teachers were given no information on the students, no information on the results of each student's previous work, and no indication on how to score the tasks, other than to use a ten-point scale. The responses yielded high interrater agreement among the fifteen teachers. In 81 percent of the cases, two scores of a student's task lay within 1.0 points of each other. When all the averages of any two scores and the average of all scores (considered the "correct grade") were calculated, roughly 90 percent of the averages of any two scores lay within a half point of the "correct" grade (de Lange, 1987, pp. 209–220). Presumably, teachers given a rubric (or entrusted to develop one) and trained in its use would yield even higher rates of agreement.

There are models in other countries of statewide mandated examinations that rely on the expertise of classroom teachers for scoring, while incorporating strategies for external verification. In Victoria, Australia, an external verification process has been used for the Certificate of Education exams, which are composed of four different kinds of tests, ranging from multiple choice to extended projects. The process, which checks the scoring of teachers and ensures reliability, has been found to have significant professional development benefits for the teachers involved as well. Before the examinations are undertaken, there is a training activity at which student work from previous years is examined in an effort to bring teachers to a common understanding of the desirable attributes of student reports and the criteria for assigning grades. After teachers have assigned grades, they submit typical examples of student work and difficult or ungraded cases to a regional review panel. Eventually, examples of work from each region are forwarded to a statewide panel for review. The review panels suggest such alterations to grades as seem appropriate, and teachers may then reassess students' work, taking the panel's advice into account.

On a final verification day, all students' work is brought to a regional meeting. Teachers are divided into verification teams under the direction of a review panel member. These teams reassess a number of reports selected by the panel member. Teachers do not reassess work from their own schools. If significant variation is found between the initial and second grades, further sampling is done and a whole class might be re-marked by at least two members of the verification team. In the trials, the grades assigned by verification teams in this way have been "remarkably consistent because sufficient professional development had taken place to ensure a common understanding of the grading process. The grades resulting from this process are as comparable as one can reasonably expect short of double marking the entire collection of reports" (Stephens & Money, 1991, p. 5).

Wilson (1992) has offered another strategy for combining the expertise that teachers have regarding individual students in their class with the more tightly controlled ratings of external examiners. He offers an analytical scheme that combines the two types of scores statistically in a manner that can give credence to both types of assessment.

The Victoria example illustrates that such processes, which rely on teachers to score the tests with sufficient checks in place, can be put into practice and that there are supplementary benefits in the professional development of teachers. Teachers who meet to discuss the rubrics and expectations for student work can develop a common language for the assessment of these tasks, but they can also take back to their classrooms a clearer vision of the kinds of mathematical activities valued. Teachers who cooperate to create and use agreed rubrics for evaluating student work gain valuable experience at using alternative forms of assessment. In a context in which they are reasonably sure of the reliability of their judgments against those of other teachers, they can in turn feel more confident in their own grading for instructional decision making. At the same time, formal rubrics and strategies for accomplishing interjudge agreement can support the development of alternative assessment tasks, prompt efforts to improve the mathematics curriculum to enable students to achieve those goals, and help instill public confidence in the use of school-based information for accountability.

ISSUE 9: MAKING REPORTS OF RESULTS UNDERSTANDABLE TO THE PUBLIC

Results of student performance on an examination need to be reported to several audiences: students, parents, teachers, administrators, and policy makers. The form and substance of these reports will necessarily vary according to the audience. Nevertheless, it is essential that they are designed to be easily understood by their constituents, at the same time that they preserve the richness of the information. It would do little good to replace traditional tests with nontraditional formats if the conceptually abundant information gathered is collapsed into a single numerical score or left in an uninterpreted form.

The first question that must be addressed in the design of reports is, What kinds of information does each audience (students, parents, teachers, administrators, or policy makers) need to make informed decisions? Because even the unit of analysis is different for each audience, the information required will likewise vary. Once that element is decided, the appropriate data sources and means of analysis can be determined.

One model for reporting is being developed by Lesh, Lamon, Gong, and Post (1992). Their "learning progress maps" are computer based, interactive, multidimensional, and decision specific, yet relatively simple in design. The maps report student progress along three dimensions. The vertical axis represents the most important conceptual models and reasoning patterns that students are encouraged to construct at a given grade level. On the horizontal axis are the basic mathematical strands (such as patterns, quantities); and the depth axis corresponds to the increasing structural complexity of the underlying conceptual systems. The result is a visual image of student learning, in the form of peaks and valleys.

Another proposed framework for reporting student progress in mathematics learning (Romberg, 1987) utilizes Vergnaud's notions about "conceptual fields." The idea is that, rather than breaking down mathematics into two dimensions of content and processes, a vast number of different forms of problem situations in mathematics can be represented by a small number of symbols and symbolic statements. For example, the related mathematical concepts of addition and subtraction of whole numbers has been defined by Vergnaud as the conceptual field "additive structure." Developing such fields can yield a map of a domain of knowledge. Such maps could free test constructors from the bind of filling in cells of a matrix.

A common challenge to current efforts at assessment reform is the development of profiles of student learning that are meaningful, concise, valid, and reliable, at the same time that they are based on a framework built on current notions of learning mathematics. One such attempt is the scheme offered by Zarinnia and Romberg (1990), shown in figure 1.3. The subcategories under *Doing Mathematics* are not the usual mathematical terms, such as space or logic. The idea is to try to capture a more realistic picture of genuine mathematical activity. The language of the chosen terms emphasizes mathematics in terms of active engagement, creative reflection, and productive effort.

Doing Mathematics	Representing & Communicating Mathematics	Mathematical Community	Mathematical Disposition
Locating Counting Measuring Designing Playing Explaining	Mental and Represented Facility in Communicating: verbally visually graphically symbolically	Individual activity Collaborative activity	Valuing math confidence Beliefs about the mathematical enterprise Willingness to engage and persist

Figure 1.3. Recommended reporting categories (Zarinnia & Romberg, 1990, p. 31).

What each of the three alternatives described briefly here has in common is an attempt to respond to current thinking about the learning of mathematics. No longer will a single numerical score suffice to describe the complex processes involved in engaging in the kinds of mathematical activity described by the *Standards*. A reporting system that seeks to support, not undermine, an authentic assessment system will have to be sophisticated enough to embrace a more complex view of the learner and an enlightened view of what it means to do mathematics at the same time that it generates information that is useful to students, teachers, parents, administrators, and the public for decision-making purposes.

The list of issues discussed in this chapter is not meant to be exhaustive, nor have we tried to resolve all of the dilemmas of each one. The other chapters of this book will elucidate some of them. Our intent is to bring to light some of the important matters that need to be addressed at each of the stages in building an assessment system for mathematics, from the initial assumptions made about mathematics and learning to the reporting schemes used. Only with careful attention to each stage of the process, and a clear vision of the overall framework, can a coherent authentic assessment system be constructed.

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.

Bodin, A. (1993). "What does to assess mean? The case of assessing mathematical knowledge." In M. Niss (ed.), *Investigations into assessment in mathematics education: An ICMI study*, pp. 113–141. Dordrecht: Kluwer Academic Publishers.

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

Ernest, P. (1988). "The impact of beliefs on the teaching of mathematics." Paper presented at the International Congress on Mathematics Education VI, Budapest.

———. (1991). *The philosophy of mathematics education.* London: Falmer Press.

Glaser, R., Lesgold, A., & Lajoie, S. (1987). "Toward a cognitive theory for the measurement of achievement." In R. Ronning, J. Glover, J.C.

Conoley, & J. Witt (eds.), *The influence of cognitive psychology on testing and measurement: The Buros-Nebraska Symposium on Measurement and Testing*, vol. 3. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Lesh, R., Lamon, S., Gong, B., & Post, T. (1992). "Using learning progress maps to improve instructional decision making." In R. Lesh & S. Lamon (eds.), *Assessments of authentic performance in school mathematics*, pp. 343–378. Albany: SUNY Press.

Mislevy, R. (1990). "Foundations of a new test theory." In N. Frederiksen, R. Mislevy, & I. Bejar (eds.), *Test theory for a new generation of tests*, pp. 19–40. Hillsdale, NJ: Lawrence Erlbaum Associates.

———, Yamamoto, K., & Anacker, S. (1992). "Toward a test theory for assessing student understanding." In R. Lesh & S. Lamon (eds.), *Assessments of authentic performance in school mathematics*, pp. 293–318. Washington, DC: American Association for the Advancement of Science.

National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

———. (1991). *Professional standards for teaching mathematics*. Reston, VA: Author.

Peterson, P. (In press). "Learning and teaching mathematical sciences: Implications for inservice programs." In *Teacher enhancement for elementary and secondary science and mathematics: Status, issues, and problems*. Arlington, VA: National Science Foundation.

Resnick, L. (1987). *Education and learning to think*. Washington, DC: National Academy Press.

Romberg, T. A. (1987, November). "The domain knowledge strategy for mathematical assessment" (Project Paper #1). Madison, WI: National Center for Research in Mathematical Sciences Education.

——— & Wilson, L. (1992). "Alignment of tests with the *Standards*." *Arithmetic Teacher* 1, no. 40, 18–22.

———, Zarinnia, A., & Collis, K. (1990). "A new world view of assessment in mathematics." In G. Kulm (ed.), *Assessing higher order thinking in mathematics*, pp. 21–38. Washington, DC: American Association for the Advancement of Science.

Shepard, L. (1991). "Psychometricians' beliefs about learning." *Educational Researcher*, 20, no. 7:2–16.

Stephens, M., & Money, R. (1991, April). "The range of performance assessed: The interaction between new developments in assessment for the senior secondary years and curriculum change in mathematics." Paper presented at the International Commission on Mathematics Instruction, Calonge, Spain.

Thurston, W. (1990). "Letters from the editors." *Quantum*, 1, no. 1:7–8.

Wilson, M. (1992, January). "Measurement models for new forms of student assessment in mathematics education." Paper presented at the meeting of the Australian Council for Educational Research, Melbourne (Models of Authentic Performance Working Group paper). Madison, WI: National Center for Research in Mathematical Sciences Education.

Zarinnia, A., & Romberg, T. A. (1990, February). *A framework for the California assessment program to report students' achievement in mathematics*. Madison, WI: National Center for Research in Mathematical Sciences Education.