Pedagogic and Psychometric Perception of Mathematics Achievement[1]

Robert E. Stake
University of Illinois at Urbana-Champaign

IN SWEDEN, IN THE U.S.A., and around the world, educational reform is complicated by contending perceptions of student achievement. In particular, teachers in the classroom and the specialists developing standardized achievement tests are not looking at achievement through the same eyes. Even when they agree on educational goals, their definitions of achievement are different. What results are incompatible notions as to how teaching and learning can be improved.

Mathematics is a subject matter spanning a dozen years of elementary and secondary schooling.[2] By professionals and lay people alike, mathematics is recognized as a subject matter taught similarly in most classrooms at a given grade level. Although there are many who advocate reform (NCTM, 1989), tasks and content identified in goal statements and syllabi appear to mean the same to curriculum coordinators and committees. But those perceptions are not necessarily the ones directing test development nor teaching in the classroom.

Instruments and procedures for assessing student achievement vary considerably from country to country, moreso than does the teaching. Primary attention is given in this paper to norm-referenced standardized testing so common as state-mandated assessment in the United States. But the differences in testing do not obscure differences in perception across teacher and testmaker communities. It is my claim that informal assessment of mathematics achievement of students is similar in many ways among teachers around the world. And psychometricians are particularly uniform in their perspective. Furthermore, formal assessment using criterion-referenced testing, performance testing, and standardized open-ended examinations is sufficiently similar across language and geographic boundaries, I believe, to make my claims here worthy of attention.

Mathematics teachers across the United States report strong and still growing emphasis on standardized achievement testing in their schools.[3] Many recognize a need to distinguish between valid and invalid uses of testing. So do qualified teachers around the world. In the eyes of experts, testing—as an activity—and individual tests are neither valid nor invalid until the results are interpreted in some way. It is the interpretations of test scores in particular situations that can be said to be valid or invalid (Cronbach, 1980; Jaeger and Tittle, 1980; Linn, 1989; Messick, 1989). The situation can be national improvement of the national education system. Especially there, the validity of using testing for improving education depends on the perceptions as to what constitutes achievement.

Standardized math tests are used in many situations to obtain valid indication of which students are better at solving the types of problems included in the test. Student motivation for scoring well (Raven, 1992), time limits of the test (Shohamy, 1984), unfamiliarity of the language and format of the test and other features of test production and administration (Traxler, 1951) can contribute to an invalidity of conclusions. But good achievement tests, properly administered, with scores cautiously interpreted, are an appropriate component of mathematics education. Test scores contribute to teacher and public awareness of how students compare, one to another and group by group.

To know the ordering of students as to proficiency is not necessarily to know what students know. Standardized math test scores are not a sound basis for indicating how well students are becoming educated in mathematics. To be educated in mathematics means

---

[1] A chapter in Donald Broady, editor, Education in the Late 20th Century, essays presented to Ulf P. Lundgren on the occasion of his fiftieth birthday. Stockholm Institute of Education Press, 1992.

acquaintance with a range of knowledge, mastery of a small sector. Education is not so much an achieving of some fixed standard. In a true sense, it requires *unique* and personal definition for each learner. Part of the established meaning among educators and others is that education is a personal process and a personally unique accomplishment. For each student, experience is different; thus the formal and informal meanings of arithmetic, algebra, geometry and all of mathematics are different from student to student.

Individualized education is one thing. The limited sampling of mathematical knowledge by tests is another. Test scores that do a nice job of indicating which students are doing best and which are doing relatively poorly on common tasks do not necessarily provide a valid indication of subject matter mastery, not on some common universe of mathematical knowledge and certainly not on personalized experiences with mathematics. One test *alone* will not provide valid measurement of the math achievement of individual students or a group as a whole. Test content almost always is too narrow. Just as a few students do not represent all the students in a school; just as a few books do not represent all the books in the library, twenty or thirty test items do not represent the broad range of math skills and knowledge that teachers are teaching. For measurement of subject matter attained, the simplicity of testing is at odds with the complexity of teaching and learning.

When any of us talk about mathematics education, we allude to long-standing and well-deliberated meanings, many of them seldom put into words, many of them greatly resisting formal statement. With good reason, we are reluctant to replace our notions of education with newer definitions offered by political policy, curriculum heuristic or testing technology. Furthermore, each person's understandings are personally constructed and somewhat tuned to cultural experience. Similarly, the ideas taught in the classroom and stimulated in the minds of students differ from teacher to teacher. Of course we speak of standard courses of study, common objectives, and shared understandings of mathematics—but the education of youth includes interpretation of mathematics against many unique experiences of past and future. Part of the invalidity of achievement testing is the constraint of standardized tests to common aims and particularly to the meanings of achievement held by test specialists. There are other constraints as well.

Much of mathematics education is beyond accurate assessment. That does not, of course, relieve teacher and society of responsibility for scratching for good evidence that education is happening. Much good evidence of math achievement comes from reflective interaction with individual students and the class as a whole, interaction during recitation, exercises, projects, and testing. Test scores alone are a flimsy indicator of the mathematics that students have learned.

In this paper, I will describe the mismatch between the highly abstract but simple constructs of mathematics held by developers of standardized tests and the situational and much more compound conceptualization of mathematics held by teachers; a difference that extends, of course, to interpretations of math achievement. The problem is common to other subject matters as well.

For any particular use of a standardized achievement test, validity of the measurement indicates the quality of information conveyed about how well the students are achieving—for a lifetime-so-far, for the year, or even for the chapter. The key distinction of this paper is between a generic, single dimension concept of achievement—a view promoted by test specialists—and a complex, experiential conceptualization of achievement detailing the many steps, the many differentiations, of content and skill—a view held by teachers. My conclusion will be that these two views are so different that the panorama of achievement which mathematics teachers regularly scan cannot be measured validly with the standardized achievement tests in use today.

Mathematics Education

Everyone knows what school mathematics is. Such a common experience seems not to need definition. What mathematics educator Tom Romberg said it is viewed as is "a vast collection of vaguely related concepts and skills to be mastered in strict order, with the sole objective of becoming competent at carrying out some algorithmic procedure in order to produce correct answers on sets of stereotyped exercises" (1990, p.2). Mathematics indeed embraces a vast array of concepts and operations but the great aspiration for the teaching of mathematics—as currently represented by the popular National Council of Teachers of Mathematics' Standards[4] —is to seek understanding more than calculation. Still, the detail and the realm of mathematics education outreach the best of definitions.

The Underperception of Mathematics Teaching and Learning. Neither good nor bad teachers stick to the point. Good teachers roam the adjacent terrain, point out and extend major connections, introduce concrete situations of relevance. Mathematics knowledge and skill are not collections of discrete elements. Each problem type and algorithm is linked into various networks of knowledge, various traits, various systems of thinking (Scheffler, 1975; Romberg and Carpenter, 1987). One digit addition and two digit addition are closely linked, one digit addition and two digit multiplication are less close, yet linked in several ways. The several ways become multitudinous when applications are acknowledged. The applications of mathematics quickly become too numerous to itemize in tables of content, lists of objectives, lesson plans—yet the practicing teacher, not only deliberately but subtly and unconsciously, reveals further dimensions of meaning for each operation and concept.

The chapter titles of a mathematics textbook seem simple enough. For example, the chapters of the book used by the upper sixth grade in the Duxbury (Massachusetts) Intermediate School in 1989 are listed in Figure 1. A quick review of Chapter I (Addition and Subtraction of Whole Numbers) finds further subdivision into the topics of: place value, reading and writing groups of 3 numerals, one and two digit addition and subtraction, properties of addition, three and four digit addition, money units, missing numbers, five digit addition, three to six digit subtraction; subtraction with zeros, comparing numbers, greatest and least numbers, rounding numbers, estimation of sums and differences, Roman numerals—with some special attention to consumer skills, career interests and problem solving. And each of these topics could be further subdivided. The inventory of topics derivable from all fifteen chapters is large.

Take the key subtopic of chapter 3, place value. A teacher expects children to be puzzled by the difference between 0.1, 0.01 and 0.001. Each is repeatedly called one-tenth. Differentiating between these three values is a specific learning target for many arithmetic teachers, not adequately subsumed or represented by a "concept of place value" in what they pay attention to. Legend has it that an item based on the magnitudes of 0.1, 0.01, and 0.001 was dropped from the PSAT because it did not correlate with total test scores—even though adequately difficult and seen by teachers as an important achievement.

Just as the textbook authors did, teachers classify mathematics into various domains, sometimes for teaching and testing (Collis and Watson, 1989). Their written-down classification headings are useful conceptual structure but draw too much attention to the well-known "main topics" of mathematics education. Educational researcher Lauren Resnick

Figure 1. Chapter Titles of a Middle School Mathematics Textbooks

1. Addition and Subtraction of Whole Numbers
2. Multiplication and Division of Whole Numbers
3. Decimals
4. Multiplication and Division of Decimals
5. Geometry
6. Factors and Multiples

7. Addition and Subtraction of Fractions
8. Multiplication and Division of Fractions
9. Probability
10. Statistics and Graphing
11. Ratio, Proportion, & Percents
12. Measurement
13. Perimeter, Area and Volume
14. Integers
15. Using Triangles

has said, "The range of mathematical concepts to be learned [is] much broader and only a few of these crafts have been intensively studied" (1989, p. 164). Furthermore, learnings within a class or even within a small subclass are related to each other in many more ways than indicated by any classification scheme. And many of the tasks and concepts within a subclass have important uniquenesses. To see this again, we can examine the seven items of Figure 2, seven problems not unlikely to appear in a forty minute activity of an algebra class.

Figure 2. A Family of Seven Mathematics Test Items

1. 20 x 1.8 + 32= ?
2. (1/5)(20)(9) + 32 = ?
3. $\frac{2 \times 20 \times 9}{10} + 32 = ?$
4. y = 1.8x + 32. Solve for y if x = 20.
5. Convert 20°C to Fahrenheit. F = (9/5)C + 32.
6. Convert 20°C to Fahrenheit. C = (5/9)(F - 32).
7. Ann wants to know today's temperature on the Fahrenheit scale. Her thermometer reads 20 degrees Celsius. What is the Fahrenheit temperature?

These seven items cut across several content domains (Hively, Maxwell, Rabehl, Sension and Lundin, 1973), yet a teacher might include all of them within a single lesson, within a single objective, or refer the solution of each of them to a single page of textbook. Each item is unique, a special variation on the others. Each will be more or less well understood by students and thus more or less difficult. Still, different in form and notwithstanding transformations, they belong to a family. The family here is not defined by mathematical operations as much as by the practical problem of dealing with two temperature scales, Celsius and Fahrenheit. To a teacher all seven need teaching and testing. To a test specialist, one probably will be sufficient.

Inventories. In the backs of our heads, we all have *epistemological inventories* of mathematics education running deep into and beyond the families of quantification. These inventories are particularly broad when we include the many applications of mathematics. The inventories often are organized around a conceptual super structure, such as that proposed by Ed Haertel and Dave Wiley (1990, see also Henderson, 1963). Such classification schemes gravitate toward a powerful simple structure. Few of them reflect the complexity of mathematics teaching to be found even at each grade level. Inventories of teaching and learning, were they actually recorded, would show that complexity. They would reveal each teacher's complex conception of the nature of mathematics education.

For understanding teacher perception of mathematics achievement, we need *inventories of content*, not just categories of topics and principal kinds of mathematical activity, but detailed inventories reflecting existing definitions of education, reflecting the complexity of teaching and learning in existing classrooms. Even people who know little mathematics can identify quite a few categories. But teachers go much beyond the categories, particularly in their choices of what and how to teach. Their conceptualization of the content of mathematics education is vast and detailed.[6] They decide, for example, whether to treat vertical addition the same as horizontal.

Most treat subtraction with borrowing different from subtraction without borrowing. Some would treat multiplication of decimal numbers with zeros immediately following the decimal point as a special learning. Many of these distinctions are not verbalized by teachers. One sees them by observing teachers in action. Mathematics educators have been diligent in classifying what teachers do but a full inventory of what they do creates a host of additional subclassifications and loads individual cells with diverse content.

Figure 3. Two Problem-Solving Exercises

1. Three children wish to divide two oranges evenly among themselves. Carefully peeled, one orange is found to have 12 sections, the other 13. What should they do?

2. Three children wish to divide two oranges evenly among themselves. Unpeeled, one orange weighs 8 1/8 ounces and has 12 sections. The other weighs 7 3/4 ounces and has 13 sections. What should they do?

Consider next the two problem-solving exercises in Figure 3. Now, what if two sections of an orange are withered? And what if juice is lost by cutting? At what point does inequality not matter? We pause reflectively before placing these two exercises in the same cell in our inventory. As teachers, we associate teaching strategy with exercise type: Should we schedule peer group dialogues (Easley and Easley, 1992) and cooperative learning (Johnson and Johnson, 1991)? Is it the right time to refer to spherical geometry? In comparing pedagogic and psychometric views in this paper, I raise questions of logic, pedagogy, learning activity, difficulty, and utility. Already it should be troubling to suppose that any one task *represents* its category nicely. A teacher has doubts that performance on any one item reveals how the same student would perform on another.

Many transformations of a mathematics problem go beyond mere restatement, examples of which are shown in Figure 2, into multidimensional extension, exemplified by five questions opening the paragraph following Figure 3. The various forms and language of teacher presentation are part of what is learned by the child. Children have a considerable capacity for recognizing item type and transformation. Capacity grows as experience grows. The teacher contributes not only by drawing attention but by engaging students in expository discourse about both large and small transformations.[7]

In the mathematics classroom, such transformations arise, over and over, minute by minute. Some are simple, some are complex (Scheffler, 1973; Romberg, 1990). Though many transformations are deliberate, mathematics teaching takes the envelope of transformation largely for granted. So do students, parents, administrators, policy makers. All find *simple* ways of representing inventories of operations and tasks. The labels that people, particularly test development people, use for identifying the domains or topics or families of mathematics items suggest a homogeneity and generality that lead us to summarize performance by a single test score. Test scores seriously understate the diversity and complexity of teachings and learnings. Mathematics education then appears to be more coherent and simple-structured than it is.

The Artificiality of Mathematics Achievement as a Construct.[8] It is not artificial for a closely observing teacher to describe how well a student has worked a mathematics exercise or project. It is not artificial to indicate how many problems or test items were answered correctly. It is not artificial to conclude that a student has achieved a level of mastery, at least for the time being, over a particular group of exercises. But, to generalize broadly about achievement is artificial. It is common for people to treat *mathematics achievement* as real[9]—but it is risky. It is artificial and risky to conclude that a student has achieved proficiency over a type of exercise such as addition of fractions or the binomial theorem. It is artificial and risky to allude to achievement of a content so vague as sixth grade mathematics.

To speak of mathematics achievement, one alludes to a selection of mathematics learned. As indicated in the previous section, I choose to call the selection an inventory. [10] There are lots of different maths. Just what math are we talking about? It is identified by the inventory. The inventory will not include "all of mathematics," much of which even the ablest mathematician does not know. We are thinking of all the math of *interest just now*. It could be that body of math taught in this school in sixth grade or that body of math needed to be in reasonably good shape when entering an engineering program at the university. It need not be a fully itemized inventory but must have some substance, some structure, some boundaries, and lots of detail.

If the concept of mathematics achievement is to be useful, the inventory of mathematics potentially achievable needs to be to some extent delimited and realized. Does the inventory include multiplication of fractions, simplest uses of a hand calculator, applications of the Pythagorean Theorem, a notion of the work of Bertrand Russell, orienteering? Specification of the contents of the inventory need not be formal; it can be experiential and intuited. If among people there is to be meaningful talk of math achievement, there should be some shared meaning of the inventory. The inventory items could be graded, e.g. essential and desirable or given a probability of being included however conceived.

Each inventory may relate to goal statements, textbook exercises, and item pools but subsidiary domains and boundaries are inevitable and inexact. Each formulation of math learning, each formalized inventory, is an umbrella for a vast array of tasks, habits, skills, and knowledge. Such formulations are great understatements of the mathematics intuitively and properly included in the practicing inventories of teachers.

Were there strong knowledge relationships among the many domains of mathematics, the need for an elaborate inventory would diminish. Were advanced skills simply comprised of "prerequisite" skills, as claimed once by Gagné (1967), the inventory could be easily specified. Were knowledge of calculus derivable from knowledge of trigonometry, we could use the latter to indicate the former. How much of a definition of education can be derived? We cannot directly measure internal diagonals of a diamond stone, yet by measures of external angles and surfaces we can calculate internal dimensions. When we have a well-developed set of relationships about the construction of an entity and if we can directly measure some of the parts, we can calculate other characteristics. The derived measures are not artificial.

But, in spite of its reputation as logically coherent aggregate of learning, in spite of the common view that advanced math skills are determined by prerequisite skills, we have no set of formal relationships binding together the many domains of mathematics and math achievement. Not only does understanding simultaneous equations remain largely independent of understanding permutations, even the knowledges of fractions and decimals remain largely independent. I presume no universal system is possible. For now, at least, the best relationships we have are few, partial, and hypothetical. It is obvious that successful long division requires some competence in subtraction but mastery of many subtraction problems cannot be assumed given mastery of the main types of long division. The field of mathematics is too complex and dissociated to permit calculation of one aspect of mathematics achievement from knowledge of another aspect.

The Dissociation of Mathematics Topics. A person has many math knowledges and skills—each simultaneously being acquired and being forgotten, forever incomplete. Many knowledges and skills are related but few are highly interdependent, the dependence complicated by the incompleteness. When mathematics is considered in the broadest sense, a surface of personal achievement stands near zero in many places, rises irregularly and not very predictably in others.[11]

As said before, we seldom are thinking of all mathematics. We usually limit our thoughts of math achievement to those things covered by certain goals or particular chapters or the teaching in this grade in this school, a terrain much more circumscribed. In spite of math's reputation as highly integrated, i.e., a succession of prerequisite learnings, the topics of mathematics are quite dissociated. The same is true of mathematics education. Generalizing from one aspect of achievement to another is problematic. For example, though many people understand both, understanding symmetry remains independent of understanding orthogonality. Often we care more about the content of mathematics and less about some generalized notion of mathematics ability. The content contained within a goal or within a lesson is too heterogeneous for a few items to represent the rest—with precision.

Often we do not require precision. There are times we want just a rough indication of how much a youngster has achieved. A teacher's recitation questions, chapter tests, and mid-term grades, for example, do serve as rough indicators of achievement. These approximates are bolstered by teacher knowledge of what has happened in the classroom. Achievement is assessed with reference to an inventory. (For outsiders, and even for the students, much of the teacher's inventory of "math covered" remains vague.) Especially among experienced teachers, the mathematics content is shared through custom and conversation. Still, precision of assessment is out of reach.[12] Each teacher's inventory is different. *It is important to recognize that as used by even the most knowledgeable teachers and testing people, the concept of mathematics achievement is artificial and imprecise, suitable at times for casual reference but a questionable basis for indicating how much mathematics a person knows.*

The concept of mathematics ability, sometimes deduced from performance on math achievement tests, suffers from the same lack of common inventory of mathematics covered. Ignoring content, the construct, math ability, is useful as indication—relative to other learners—of how much learning time or teaching effort will be required in subsequent courses. Relative standings remain quite stable for a fixed group of students, stable usually even as individuals pass from one group into other comparable groups.

Notwithstanding useful predictions, the concept of math ability has become hurtfully contrived. Many test specialists, especially those advocating item response theory (see the fine summary by Hambleton, 1989), force disparate aspects of mathematics into a single indicator, omitting from their definitions of mathematics those tasks and knowledges which do not nicely fit their scales. The hurt comes when teachers, failing to see certain mathematics topics included in the tests, drop those topics from inventories to be taught. Topical items most useful for predicting math achievement are not necessarily good for defining it. Mathematics ability and mathematics achievement are not interchangeable concepts.[13]

When it is necessary for us to estimate, to generalize over unknown terrain, to presume the nature of the whole, to work with entities whose contents have not been specified, our measures of the whole are artificial. Except in the simplest situations, the formal measurement of mathematics achievement is artificial.

Teacher Conceptualization of Mathematics Education. As every teacher knows, there are shortcomings in education—including teacher shortcomings.[14] But the fact of the matter is that, at least in Sweden and the United States, the teachers are one of the stronger assets of the system, much stronger, in my opinion, than the administrators, the textbooks, the willingness of students to be students,[15] and the tests.[16] All could be better, of course. And each of us believes, “If I were to yield to the pressures to change, conditions would become even worse." Our educational systems are more-or-less locked in. We are appalled at the prospect of further deterioration. Tests which show national achievement to be poor are usually not wrong but as part of the problem the tests push schooling toward standardized authenticated mediocrity.

Current teachers are an asset mainly because they have a long developed and far-reaching conceptualization of the connection[17] of ideas and behaviors that constitute a high school course or the year long lessons for a particular elementary school grade. An experienced mathematics teacher has a strong idea of what topics should be covered, the calendar and time allotments, the relationship and interdependence of topics, the nuances and subclassifications of topics, diverse applications of topics, the relevance of topics to standardized testing, opportunities for enrichment and cooperative learning, nurturing independent thinking and self-directed learning, ways of increasing motivation and decreasing discouragement, what will be the stumbling blocks, how socialization and conflict pre-empt academics, what experience the students will bring, the expectations of students and parents and other teachers. The work of teaching is complex.

I want to pictorialize the complexity of mathematics education. Somewhere in the mind of each math teacher is an inventory of topics to teach. Each topic alone is as intricate as a tree with bigger and smaller branches with twigs and lacy buds and leaves individually dispensable but collectively vitalizing the tree. The parts are comprehensive and capable of personal interpretation. Some writers would use trunk and main limbs to represent classifications of goals objectives chapters and types of problems extremely important as structure and discipline for the emergent learnings of mathematics. Others would use trunk and limbs to represent the relationship among ideas. And still others would emphasize the connection of mathematics to experience (playing store or calculating water need for a camp-out), to other subject matters and to preparation for livelihood too. However the representation somehow *actually* teaching the teacher represents the categories relationships and connections. More than anything else it is the teacher's comprehension of the subject matter as manifest in action that distinguishes between a teacher view and a testing view of mathematics achievement.

The inventory of mathematics to be taught is the critical epistemology of education. Comprehensiveness integrity and topical uniqueness is not to be found in the coverage of tests. Neither teacher nor testmaker is reading John Dewey. Few math teachers or test specialists know a mathematician to chat with. The authority of subject matter has been taken over by the syllabi the textbooks and the tests---each with a leaning toward simplifying increasingly myopically bent on raising those (potentially) embarrassing achievement scores. But it's give and take. Within a still vital autonomy in most classrooms the complexity of teaching continues. It draws from a stronghold of mathematics education in the minds of the teachers a cache half empty but a precious asset.

Figure 4. Goals for Education in Georgia

> The Georgia Board of Education has adopted student goal statements that identify the ideal skills and attitudes a graduate of Georgia's educational system should strive to achieve through instructional programs in the state public schools. The State board believes that the instructional program in the public schools should provide each individual with opportunities to develop abilities so that he or she
>
> communicates effectively
> uses essential mathematics skills
> recognizes the need for lifelong learning
> has the background to begin career pursuits
> participates as a citizen in our democratic system
> etc.
>
> [And for mathematics in Georgia:] The mathematics section of the Quality Core Curriculum consists of objectives relating to concepts, process skills and problem solving at each grade level, kindergarten through eighth. In grades 9-12 objectives are given for each mathematics course. ... Mathematics began and continues to be a way of organizing one's world, through the study of quantity and space, their properties and the relationship[s] within and between these concepts. Mathematics is first experienced as a language created to describe the world, accompanied by rules that govern its use. Exploring mathematics results in ...

[And for Algebra I the Topics / Concepts are identified as:]

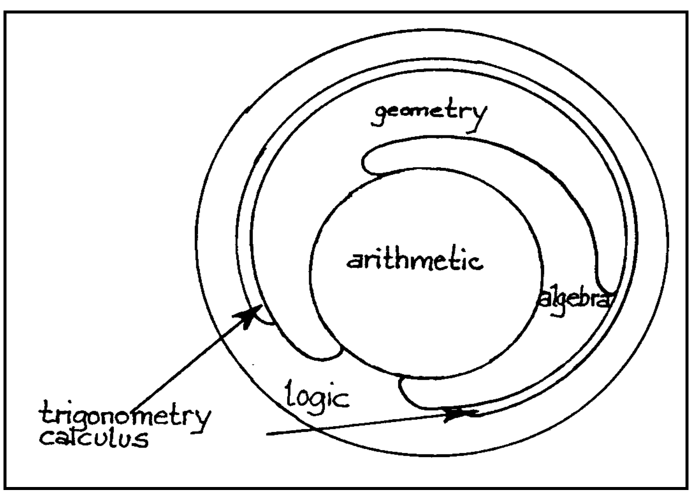
E. Polynomials
13. Identifies polynomial expressions
14. Adds and subtracts polynomials
15. Uses of laws of exponents necessary to perform polynomial operations
etc.

Representing Education. As do all people, teachers use simple representations. Their course outlines and lesson plans briefly list topics and activities. To satisfy the requirements of administrators or to talk to visiting parents, they sometimes refer to lists of objectives such as those for the state of Georgia abbreviated in Figure 4. But in thinking how and what they will be teaching, teachers work at a much higher level of complexity.

Complexity of teacher thinking was illustrated earlier in Figure 2 by seven mathematics items. To a testing person, these items are points on a single scale; they measure essentially the same thing. To a teacher, each is unique. The statistical correlation among the seven would run high but each item requires its own understanding of terms and operations. The math teacher extends instruction to the details of each item. Getting any six of the items right does not assure getting the seventh right. To a teacher, math achievement is not just getting the best score on the test, it is understanding and performing the work.

Within mathematics education, far more interweaving and interdependence of meaning occur than is apparent on a list or content behavior grid (Wilson, 1971, p. 646). To move beyond, what if we tried to represent the similarity (proximity) and sequentiality (directions) of math topics? The hypothetical map in Figure 5 might stimulate our thinking. Here, for example, the topics of *trigonometry* appear closer to *geometry* than to *arithmetic*. If we had more detail,

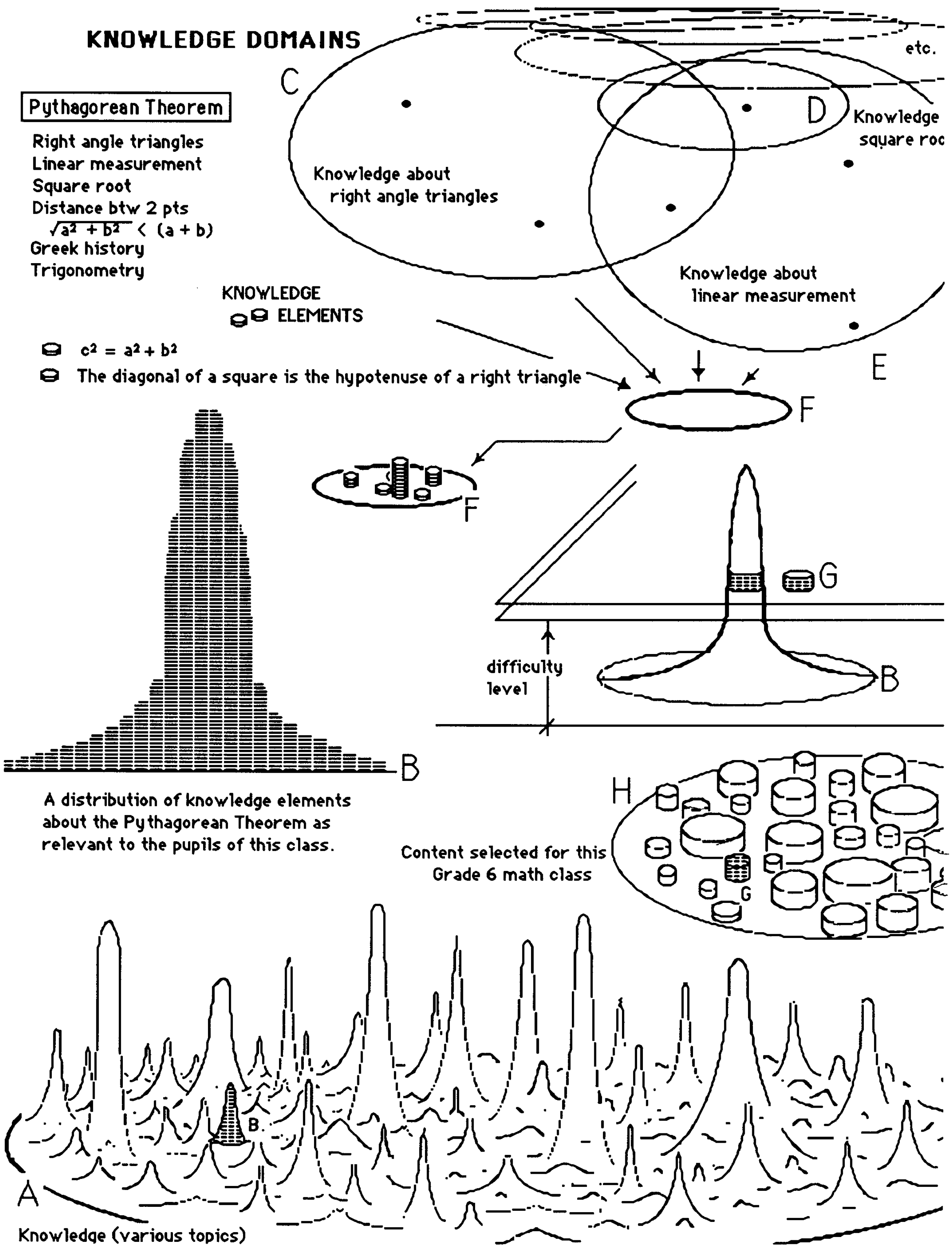
Figure 5. A Map of Mathematics Education Content.



we would expect to see *percentages* lying closer to *fractions* than to *probability*. A two-dimensional map raises many questions but turns out to be almost as unsatisfying as a list. The relationships overwhelm the mapping.

Yet when we analyze what a teacher is doing, we find topics and activities connected in logical ways as if all were mapped there in the teacher's mind. When we ask for it, the teacher cannot produce the graphics, just certain projection points as to what teaching fits where. Indirectly more than directly, the teacher has transformed complex epistemological relationships

Figure 6. An Impressionistic Representation of a Teacher's Approach to Teaching the Pythagorean Theorem



into course schedule and on-the-spot responsiveness. And when we analyze the thrust, we find teaching not aimed at developing some general mathematical ability but at developing knowledge of specific topics and skills in solving specific kinds of problems. The inventory is the tacit map by which the pursuit of knowledge is rationalized.

Mathematics teachers incorporate student performance into instruction. They allocate a great portion of time to operations and exercise work. Their conceptualization of mathematics teaching is process-oriented more than outcomes-oriented. The teachers strive for high quality experience, immersion in the topic, honing of the particular operation. Few mathematics teachers think first of making children "numerate" or (unless harassed) pressing for better scores on an achievement test. Their first aim is to help children gain command of a far-reaching, unsaid inventory of subject matter, outlined perhaps as the National Council of Teachers of Mathematics (1989) proposed but extending to a network of detail itself as salient as the major classifications.

How does a sixth grade teacher approach the lesson? Figure 6 is my impressionistic representation of choices made by a teacher as to what to teach perhaps tomorrow about the Pythagorean Theorem.[18] The topic is mentioned in the state's list of learner objectives and identified in the district curriculum guide and the textbook the teacher is using. To a degree, the textbook author defines what will be taught but, especially in recitation, the teacher modifies course content to fit the situation, noting especially the frame of mind of the present student group. Reflecting on the many pertinent topics for the class (topography A, bottom of the page), the teacher considers the facts, concepts, relationships, and applications of the Pythagorean Theorem. Distribution B represents a closer look at what is most relevant for these particular sixth graders. The teacher draws elements from several knowledge bases (circles C, D, and E) to obtain a small selection to teach (plate F), then thinks about learning difficulty (cylinder G) and inserts it as content (tray H) for this class.[19] The teacher anticipates a small presentation with graphics, reading, seatwork and homework. When it happens, ideas are modified as the conversations of instruction occur, shaped, of course, by the teacher's overall conceptualization of mathematics education.

The naiveté of Figures 5 and 6 is obvious, no less so that of the lists of Figures 1 and 4. Graphic technology to represent pedagogy and epistemology is not highly developed.[20] Classification systems and content-skill grids are common in curriculum offices but there are few devices to represent conceptual links between topics and to guide pedagogical moves from one content to another. Yet, just as ancient travelers reached destinations before there were maps, teachers teach without maps, build without blueprints. Intuitively, good teachers merge topical paths, capitalize on personal experience, draw out and preserve the youngster's line of thought.

In the pages so far, I have described teaching and learning of mathematics as enormously detailed. From their words and activities in the classroom, it is apparent that the conceptualizations of teachers as to what constitutes a course greatly influence their planning, instructional strategy, and assessments. Although the writings of mathematics educators, school district syllabi, textbooks and tests can be said to be built intellectually on a more powerful structure, these formal conceptualizations of mathematics education do not identify a great many characteristics of math achievement important to teachers. The conceptual schemes available from standardized math achievement tests will be dealt with next.

Psychometric Perception of Mathematics Achievement

The public does not understand how there could be sincere objection to using standardized achievement tests to represent what should be taught. People think the tests measure what students know and do not know. They correctly see those tests as nationally based and technically elegant. They presume that tests valid for one educational purpose would be valid for others. As students for many years themselves, they have experienced teaching and testing; the question of alignment almost never came up. Now, when a mismatch is claimed, often they presume that the teachers are wrong. Most administrators, counselors, and officials do not question testing. Many citizens interpret objection to testing as unwillingness to acknowledge shortcomings in the educational system. Dealing with these problems requires a careful look at the standardized achievement test as an information-gathering instrument.

Testing, broadly considered, is the presentation of certain challenges with responses judged as right and wrong or above and below criterion. Achievement testing is an activity which generates student performances to be interpreted as above or below pedagogical criteria. A test score indicates performance on a collection of items. Test makers conceptualize what constitutes a good performance. The score is taken as an index of achievement, a datum, a bit of information. For most people, testing is seen mainly as an information-gathering activity. After the information is put to use, we can speak of the validity of the interpretations.

Information from testing can be treated both as measurement data and as data for pondering a problem (Lorge, 1951). When we have measurement data, we think as if we have captured a dimension of *something real* something substantial, such as the measurement of age or hat size. For people having measurement appetites, and that includes most of us, standardized tests are expected to provide a trustworthy indicator of student achievement (Shavelson, McDonnell, Oakes and Carey, 1987; Burstein, Baker, Aschbacher, and Keesling, 1985). Even if the scores are not entirely accurate, there is expectation that the amount of something real is being expressed.

Especially when we don't know whether or not the measurements are precise, the information from testing can be "problem-pondering" data, provocative of thought, generative of hunches, actually helping to shape strategy because strategy is based partly on subjective judgment. These are formative data, potentially useful for redeveloping an idea or practice. It is important for a teacher to consider both possibilities when reviewing test results. Much of achievement testing will fail to provide teachers with dependable measurements, yet be useful for tactical review and reconceptualization—far more than a guess, far less than a causal link.

Testing is more than information gathering. It is a management control mechanism as well. It is used to announce purpose and priority. It is scheduled in advance by administrators so that effort is shifted toward particular goals. It is an instrument of reward and punishment. Testing and other forms of assessment are widely seen as essential to accountability and educational reform. The effectiveness of the testing process is often seen in terms of contribution to management of classroom, school, and school system.

Generalizability of Mathematics Knowledge from Standardized Tests. It is widely supposed that a good mathematics test will indicate amount of student knowledge of the mathematics broadly represented by the test items. Actually, as stated earlier, standardized tests indicate very little as to how much mathematics a student knows (McLean, I 982b; Haertel, 1985). They do not directly measure how well educated in mathematics the student is becoming. They do not identify the cognitive structures of children's thinking (Piaget, 1929; Easley, 1974).[21] Some indirect inferences can be made by teachers having a good understanding of mathematics as subject matter and how students do mathematics but the tests add little to what the teacher already knows. Unfortunately, the teacher is almost invited to make the mistake of concluding that the best learners know all that has been taught and the slowest learners little.

When used with an ordinary group of students, i.e., a heterogeneous group from the school's catchment area, many standardized achievement tests do effectively indicate which students are the best learners, the ones whose achievement is superior, the ones becoming best educated in the core curriculum of schools. Middle-range and low performing students also are consistently identified. Tests are poor indicators of future performance of poorly motivated students who subsequently become inspired or of zesty students who in midstream lose interest in academics. But there are few of either of these; approximately the same ranking of students will be found in their courses next year. *Test score interpretation often is valid when predicting standing in the same or an equivalent group at a later time.* The tests can profitably be used to indicate which students probably can handle an accelerated or advanced mathematics course.

With high correlation having come to be expected among different math tests, test specialists as well as educators and the general public have come to expect a good math test to indicate attainment of knowledge as well as student ranking. With reference to the items of Figure 2 (shown earlier), these people would expect good performance on item #1 to indicate a holding of knowledge needed to do item #2, and to a smaller extent, the knowledge needed to do the five less similar items.

Yes, the students who do best on item #1 will tend to be the ones who do best on items #2 through #7. But how well a given student will do on those six items is not indicated by performance on the first. Performance on the last item does not indicate how well a student will do on the first six. ("Doing well" here refers to level of performance on the task regardless of how other students perform.) *An item belonging to a topical family is not necessarily a good indicator of how well students will do on other items within the family and so, of course, such items are not good indicators of how well students will perform on math items generally.*

For a particular group, item #1 might be easy, let's say that 90 percent get it right. Even when the items were selected to have a high inter-item correlation, such information about item #1 gives us no idea about how difficult another item is. Even highly experienced teachers often make poor estimates of task difficulty so, even for them, a group of test items do not provide achievement information on tasks not tested.[22]

Actually, a little more can be deduced. Specialists assembling items for standardized tests want some range of difficulty among items but not too much. To maximize discrimination and to enhance the validity of predictions, they seek items that half the students will get right, half will get wrong. Still, they do not want to discourage examinees by initial items too difficult nor to embarrass educators by tests that appear too easy. They try to select a majority of items of about the same difficulty, headed by a few easy items, ending with a few difficult ones. With this knowledge, someone examining the test items, when given the difficulty of a few

Fig. 7. Skill Clusters in the SRA Standardized MathematicsAchievement Test, Level 35, Form P

Computation	Whole Numbers & Money	17 items	40 minutes for the section
	Decimals	12 items	
	Fractions, Mixed Numbers	10 items	
Concepts	Whole Numbers & Money	8 items	22 minutes for the section
	Fractions, Ratios, Proportions & Percents	7 items	
	Decimals	6 items	
	Prealgebra	4 items	
	Geometry and Measurement	5 items	
	One-Step Problems	7 items	
Problem Solving	Multiple-Step Problems	4 items	32 minutes for the section
	Rates, Proportion, Percents	4 items	
	Geom, Measmnt, Statistics	10 items	
	Problem-Solving Skills	6 items	

items, can make some guesses as to the difficulty of others. But this is only a basis for estimating achievement on other items on the test. It is not a basis for estimating whether or not the examinees would do well on problems not on the test.

As indicated above, student performance on a great range of math items is remarkably correlated. More-able students tend to find almost all items easier than the less-able do. A group of students will show achievement about as high on all math items of similar difficulty. Were there to be an inventory of math items having average difficulty equal to the difficulty of items on the test, a group of students would perform about as well en bloc as they performed on the test. It is possible to conceptualize a curricular inventory of math achievement specified only as math having a certain mean difficulty. A teacher noting a math test mean score could generalize as to achievement on this generic inventory.

But that would be a fanciful exercise.[23] If we are interested in the education of youth, we are interested in it becoming knowledgeable and skillful as to particular mathematics. There is no set value to any domain of mathematics; yet each domain is more or less valuable, not equally valuable, to a robust concept of mathematics education. The inventory—however poorly specified—is where teaching starts. Our lessons, our textbooks, our tests need to be aligned with what we want mathematics education to be. Mathematics education is largely what teachers teach, what learners learn. We should not delude ourselves into thinking that domains identified by experts (other than teachers, perhaps) capture the essence of mathematics education. Valid interpretations of achievement scores can reflect different definitions of mathematics education; teacher interpretations will reflect teacher definitions.

Neither specialists in curriculum nor technologists of testing have suitably refined and reported the inventories in *their* heads,[24] much less those in teachers' heads. The categories of items of the standardized math achievement test used in the upper sixth grade at Duxbury Intermediate are shown in Figure 7.

The authors go further to identify each item as to number of digits, operation, and units (if any) but it is safe to say that they do not know how to further map the item into the rest of mathematics. None of us do. None of us have the language or graphics to show the detail of similarities we feel between decimals and fractions, to show how we draw upon our understanding of one step problems to do two step problems, to illustrate how we use scale conversions in problem solving. We can give examples, we can demonstrate, we can show how we would teach, but we lack a language to represent those similarities, textures, and relationships. Test producers and others have little linguistic technology for sharing with test users their definitions of mathematics content. The users of math achievement tests are pretty much on their own to decide what mathematics, other than those items actually on the test, is being referred to when we conclude that an examinee is a high achiever.

For a teacher looking closely at the particular items of the test, scores provide a rough indication of how well students would perform subsequently on similar math items. Even the best teachers, however, are often wrong in presuming what content is "similar" in difficulty. *Standardized mathematics achievement tests are pretty good at indicating which students are best and poorest at learning school mathematics, pretty good at indicating (for those teachers who find it useful to believe in) a general mathematics ability, but quite poor at indicating which knowledge and skill the students have actually attained.*

Figure 8. District Outline of Content and Textbook Chapter Titles for a Sixth Grade Math Class

District Outline	Grade Six Text
1. Addition and Subtraction Whole Numbers	1. Addition and Subtraction of of Whole Numbers
2. Multiplication and Division of Whole Numbers	2. Multiplication and Division of Whole Numbers
3. Decimals	3. Introduction to Decimals

4. Multiplication and Division of Decimals	4. Multiplication and Division of Decimals
5. Geometry	5. Number Theory
6. Factors and Multiples	6. Addition and Subtraction of Fractions
7. Addition and Subtraction of Fractions	7. Multiplication and Division of Fractions
8. Multiplication and Division of Fractions	8. Geometry
9. Probability	9. Percent
10. Statistics and Graphing	10. Probability
11. Ratio, Proportion and Percents	
12. Measurement	
13. Perimeter, Area and Volume	
14. Integers	
15. Using Triangles	

Alignment of Curriculum and Testing. Various writers have noted the difference between the curriculum outlined by official goals or syllabi and the curriculum taught by teachers (Eraut, Goad and Smith, 1975, Aoki, 1983). These differences are natural and substantial. Even when coordinators and teachers share both aims and concepts of teaching method, there is no way for coordinators to state precisely what the teachers should be teaching. Even when the teaching is not very good, the profundity of teaching exhausts the describer. Stating objectives and teaching in the classroom are two very different media for defining education. In good times, the two provide a dialectic which serves to refine both aim-stating and teaching. In bad times, they fight, they embarrass, they deceive. Differences, of course, are often more than differences in media. Official goals and actual practice often are differently aimed. Values, needs, and conceptualizations can vary widely. In most efforts to reform education, there is a presumption that education would be better if the stated curriculum and the actual curriculum were more similar.

In the same vein, no textbook perfectly reflects a school's official goals. The outline of content for sixth grade mathematics in Duxbury in 1989 is shown in Figure 8 alongside the textbook chapter titles (previously shown in Figure 1). The title pairs match word for word for over half of the book. The detail that guide authors and text authors had in mind differs. For example, the Duxbury guide also lists 33 anticipated student outcomes categorized as Knowledge, Skills, and Attitudes. Almost nobody expects a closer match than shown here but it is important to recognize that it is only the headings that match, a great body of detail is free not to match. An example of a mismatch in teaching fractions would be for the district outline to call for division of sets into subsets (two out of six) as the dominant representation and the textbook to rely almost exclusively on pie charts. A mismatch might be troublesome when the district outline calls for two weeks on geometry and the textbook has but five pages. It might be a serious mismatch if the textbook only had exercises in the construction of graphs and district objectives emphasized interpretation of graphs. However, the more the curriculum is in the hands of the teachers, the less we need worry about mismatch in materials.[25] Usually the teachers take the discrepancy in stride, indifferent to, sometimes too indifferent to, the thrust and boundaries set by both textbook and guide.

And similarly, testing will not and cannot cover precisely the same ground as textbook, syllabus, and especially, teaching practice. Particularly following traditions of local control and teacher autonomy, teaching practice will have different thrusts and boundaries than standardized tests. Here again, difference in language media prevent perfect agreement but the differences are likely to be greater than that attributable to media. Test authors and teachers just differ in their definitions of achievement.

At some level of generality, the curriculum to be offered and the achievement testing should share an inventory of mathematics content. The word used by researchers in the United

States to indicate the match between the inventory of teaching and the inventory of testing is "alignment." It seems to most people that teaching and testing should be aligned (Freeman, Kuhs, Porter, Floden, Schmidt & Schwille, 1983). This is not to say students should never be tested to determine understanding of math operations and concepts beyond those taught in the classroom. There are many opportunities to learn mathematics outside the classroom and many teachers exploit them. When the purpose of the testing is primarily to increase understanding of the extent to which a youngster is becoming sophisticated, then the testing inventory should not be limited to school mathematics. *But when the purpose of testing is to increase understanding of the attainment of school math, the curriculum and the testing should be aligned.*

Alignment is difficult to measure. As I have already mentioned, it is easy to point to examples of misalignment. Some test items will not only be unlike exercises assigned but depend on learnings from domains untaught. Some goals such as "The student will apply the theory and laws governing our number system as known at this level,"[26] are not easily tested. The test, or its extension, the total item pool, is a weak representative of the voluminous inventory of intended learnings. The formal curriculum guide, though usually considerably more detailed than an item pool, also is an understatement of the inventory of desired mathematical achievement. Compared to the curriculum guide, the coverage of the test will fall short. Observation that the test items only retate directly to, say, 10 percent of the sections in the curriculum guide indicates mostly that the test is much shorter, not necessarily that there is misalignment between inventories. Matching is problematic. We do not have a good way of measuring the alignment between tests and curricula.[27]

Still, it is useful for teachers to compare the two. It will become apparent that the test is attentive to some domains and not others. It may become apparent that the curriculum guide fails to include some formats in which problems are presented on the test. Usually teachers will decide that students should be expected to solve some problems other than those they actually worked before. A careful review of achievement test items and student performances can broaden and deepen understanding of the complexities of education. Testing can contribute to a provocation of thought about what we mean by mathematics education. On the other hand, a careful review sometimes leads to mindless *teaching for the test*. Alignment usually can be improved by simplifying the curriculum. Whether or not high alignment is good needs to be decided with reference to our basic definitions of education.

My assistant, Giordana Rabitti, studied the alignment among curriculum guide, textbook and standardized test for upper sixth grade mathematics in Duxbury. All of these have been outlined in Figures 7 and 8. She concluded that comparison of category headings was insufficient and that comparison of the detail required extensive and highly subjective judgments. From the documents before her, her conclusions could be no more than impressionistic. She found these materials capable of a high degree of alignment but that careful attention by teachers to guide, text, and test would not assure alignment. The curriculum-as-taught depends on the interpretations of individual teachers who probably will continue to vary widely in their pedagogical methods and inventories of content.

Accommodating to Different Perceptions

It has not helped upgrade education to specify standards—academic skills and curricular topics—for all to master. There are other roads to reform, some with increased opportunity for children to experience intellectual problems, to voice perplexity, and to propose explanation. Many of us see it as essential that individual children are helped to relate their studies to personal (uncommon) experience. In trying to raise standards, state and national school reform efforts have overly relied on common goals and common test performance.[28] We could not do without common aspiration and expectation but great also is the need for unique teaching and arrangement for personal interpretation by each child. Many of our teachers are capable of providing it and do. Overemphasis on common goals diverts their efforts.

My field studies of American classrooms (Stake and Easley, 1978; Stake, Raths, Denny, Stenzel, and Hoke, 1986; Stake, Bresler and Mabry, 1991) tell me that the American teacher remains a major asset, not as capable as we would like, not all that children deserve, but largely pleasing to the local community and to school authorities, more the artist and even more the technician than reformist agitation suggests. Most teachers have heard the calls to reform, are sympathetic to them, have helped initiate some, and are hopeful of contributing to improved student assessment. Many are troubled by instructional time diverted to preparation for testing. Most do not see that mandated assessment already is changing the nature of education in America.

Education is being redefined away from the perceptions of teachers and toward the perception of assessors. Though difficult to measure, standardized testing, intentionally, with noticeable effect, often with harmful effect, does change education (Shepard, 1991; Smith, 1991; Wilson & Corbett, 1991; Haney & Madaus, no date). Teachers report that with increased testing and curriculum standardization, they attend more to the so-called "basics" (the most elementary knowledge and skills) and attend less to deep understanding of even a few topics. According to George Madaus (1991), the dangers in current school reform are several: overstandardization, oversimplification, overreliance on statistics, student boredom, increased dropouts, a sacrifice of personal understanding and, probably, a dimunition of the diversity of intellect among people. These dangers are smaller when the curriculum taught is the curriculum conceptualized by teachers.

Traditionally, education has been a matter of understanding based on knowledge, with each person's knowledge and understanding different because of the impossibility (and undesirability) of completely shared experience. To deal with the complexities of education for the few and the masses, not only schools but school systems were created. Just as the budget of schools is high on the list of social costs, the management of schools is one of the most comprehensive of collective endeavors. As authority and accountability of the schools are challenged, school officials look for additional ways of exerting control, not only over learners but over teachers, parents and taxpayers (Newman, 1991, Smith and O'Day, 1990). The math teachers surveyed for this monograph confirmed that testing is an instrument of management. Those who control the tests provide much of the perception that forms the definition of learning, teaching, and education.

The ostensible purpose of achievement testing is to measure the learnings of students. To an extent, this occurs. As to measuring mathematics learning, stable rankings among students are obtained. These rankings are not very different from those which—from classroom observations and assignments—teachers have generated informally and more extensively. Test scores confirm and authenticate pedagogical assessments. As to which mathematical knowledge has been attained, standardized achievement tests provide very little valid information.

Test scores provide a stable basis for comparing schools or nations; unfortunately, those comparisons are often distractive, sometimes pernicious. As illustrated in this monograph, they turn teachers and administrators to a lesser task. Comparisons make some deficiencies public but only on the rarest occasion do they point to weakness not already recognized by teachers and representatives of the public. Seldom do they provide insight or diagnostics for the remedy of deficiency. Nor are achievement test scores indicators of quality of teaching.

To a certain extent, standardized achievement tests should be aligned with the curriculum as planned and with the curriculum as taught. The rests should be in harmony with expectations of parents and the state. Obviously, there is no way to match fully these at least somewhat disparate obligations and expectations. We may someday improve the technology of representing curricular priorities, recognizing with precision what different people want teaching to be, but any reduction in disparity is likely to be more a matter of oversimplifying the wants than of drawing teachers and others into consensus.

The important reasons for lack of alignment between mathematics teaching and standardized achievement testing are the brevity of tests and the personalization of teaching and learning. Contrary to popular and technical opinion, the items of the test do not nicely represent classroom teachings, even when test items and an abundance of classroom exercises fall within the same goal statement. The curriculum is driven by two perceptions, education as experience with a vast territory of disciplinary content (held by teachers while teaching) and education as condensation of abilities (held by the psychometrician and increasingly by others distant from teaching). In Sweden, in the U.S.A., and around the world, educational reform is confounded by these contending perceptions.

Endnotes

1. Research for this paper was supported by the National Center for Research in Mathematical Sciences Education at the University of Wisconsin. At the time of this writing an earlier paper (completed in September 1991) entitled "Validity and invalidity of mathematics achievement testing" was under consideration for inclusion by the Center (under the editorship of Thomas A. Romberg, director) in a written presentation to the National Academy of Sciences.

2. For argument about perceptions of achievement in this paper mathematics will be the subject matter of reference. I believe that pedagogic and psychometric perceptions of student achievement will differ even more with most other subject matters.

3. Recent survey data from U.S. mathematics teachers are presented in the paper cited in footnote 1.

4. The current NCTM *Standards* (1989) are part of a long-running campaign by mathematics teacher educators to get teachers to conceptualize less according to topical content and more according to problem-solving and experiential learning (Carl 1991). The purpose of the present paper is not to argue for one or the other but to examine test validity in terms of teacher and test-developer conceptualizations of mathematics achievement.

5. *Mathematics¾Essentials and Applications*, by Glen Vannatta and John Stoeckinger published by the Charles E. Merrill Publishing Company, 1980.

6. In speaking of a vast and detailed content that mathematics teachers bring to the classroom I do not mean to say that as a group they put content learning higher than other learning. Clearly teachers differ. My own acquaintance with mathematics teachers is nicely reflected in the work of sociologist Robert Connell (1985) who found teachers preferring one of four emphases: intellectual growth, personal development, skill learning, and honoring custom. Those holding intellect in highest esteem take special pains in choosing content to teach but whether articulated or not and whether sophisticated or not all teachers have elaborate conceptualizations of subject matter.

7. Such discourse is at the heart of some definitions of teaching. Speaking of the teacher, Sylvia Ashton-Warner (1967) said: "From the teacher's end it boils down to whether or not she is a good conversationalist; whether or not she has the gift or the wisdom to listen to another; the ability to draw out and preserve that other's line of thought."

8. "Artificial" means that it is a construction of human interpretation and judgment not a direct inventory or objectively derived calculation from direct measurement. Math achievement is artificial because it alludes to a body of mathematics only vaguely specified and largely intuited. Artificial does not mean it is bad. Artificial constructs are central to all science. Such constructs as "energy" and "susceptibility to disease" are sometimes objectively defined but are used intuitively and practically by scientists and others. The value of artificial constructs to practitioners depends on how well rooted the concept is in action and discourse.

9. Science, particularly inductive science, is built upon constructs, aggregating through relationship into theories. Testing researchers such as Jim Popham (1978) and Ed Haertel and Dave Wiley (1990) speak of domains, traits, and achievements with confidence that these constructs interchange with the constructs educators develop from experience. The more artificial the construct, the greater the need for validation, by researcher and educator alike.

10. The emphasis here is on knowledge. Most psychologists prefer to identify the selection of mathematics learned as made up of abilities or competencies. Haertel and Wiley, for example, said ''In this paper the term 'ability' encompasses that which is commonly classified as knowledge and

skill" (1990). Such terms draw one toward thinking of education as a collection of skills and away from thinking of education as understanding of knowledge. Both are part of education but the two definitions move thinking about education in different directions.

11. To appreciate the complexity and lack of interdependence, one might think of an inventory of mathematics the same way one thinks of physical surfaces of land mass. Two dimensional space could represent knowledge and skill and the elevation could represent conceptual attainment by an individual or group. To each square kilometer, we could assign a learning task. The entire plot might cover a territory as large as a country. For one person, achievement across tasks might be as irregular as the terrain of Sweden, for another it might be as flat as Denmark. Predictions of ground elevation from one part of the country to another would be risky. One cannot indicate ground elevation of Arvika from knowledge of elevation of the railway station in Stockholm. And one does not have a very good idea of the elevation of all of Sweden by sampling elevation at thirty points. Achievement usually will be similar for nearby tasks but attainment of distant tasks is unpredictable.

12. False precision is regularly implied by teachers who grade in percents, Implying that 100 percent correct refers to a totality meaningful beyond the items actually administered.

13. Decades ago, most psychometricians abandoned testing for ''intelligence'' and reconceptualized their target as "scholastic aptitude." Still a form of intellectual power, scholastic aptitude indicates predictable relative achievement on common classroom assignments. Math ability is a specific scholastic aptitude.

14. People will disagree as to who the delinquent teachers are. Shortcomings in subject matter competence, behavior control, punctuality, dress, and test score production, any one or more can be excused by some people when other qualifications run strong. Evaluation of teacher merit is not just a measurement problem, it is confounded by ideological diversity in the school and in the community (Simons and Elliott, 1989).

15. Moment by moment, through the day, many American students persist in the view that there is little of importance in what the teachers are teaching today. "If it turns out to be important, I can get it later". The indignation of many parents about the schools feeds youngster resistance to being taught (Farrell, 1990).

16. I note also the shortcomings of educational researchers but to list them in this sentence would imply that they influence what happens in schools.

17. Ulf Lundgren (1972) conceptualized the conditions of instruction monitored by teachers. Carl Bereiter (1991) gave us opportunity to rethink the question of how teachers can be effective contributors to student achievement even when they cannot vocalize the rules by which they teach. Seymour Sarason (1971) wrote cogently on informal assessment of social conditions in the classroom. See also Lieberman, 1984; Connell, 1985; and Lampert, 1988.

18. Figure 6 is not a research finding, it is merely a euphemism, a view representing what experienced teachers have appeared to do in my observation. When asked, they seldom claim to be involved in such detailed analysis. And yet the effect of such selection can be observed. The point again is that the intuitive working of teaching is highly complex with far greater texture than the goals stated in Figure 4.

19. The learning terrain for each child, of course, is different.

20. Some of the best works to date: Rosalind Driver, 1973; Bob Gowin, 1990; D. H. Jonassen, 1982; Takahiro Sato, 1991; School Mathematics Study Group, 1961. These works analyze either instruction, epistemology or cognitive development; they do not adapt nicely to the "conversational" exchanges of American and Swedish classrooms.

21. Other than to note Easley's statement, "It seems absurd to pretend that one knows how to measure cognitive competences by administering standardized lists of questions when no validating clinical interviews—and certainly no explicit structural analyses—have been published," I will not use this chapter to examine the inability of achievement tests to represent student mathematical thinking. Here I concentrate on the disparity between the inventory of achievement conceptualized by math teachers and the collection of mathematics aptitude items pooled by test makers.

22. Interpreting performance with reference to how other examinees perform, usually using "percentile ranks," is called "norm-referencing". The sophistication of test technology for norm referencing is very high. But, as Bob Glaser (1963) Jay Millman (1974) and many educational researchers have said, instruction needs "criterion referencing" interpretation with strong reference to the content of the task. With course content much more rooted in teacher informal conceptualization

than in formal epistemological analysis, the sophistication of criterion-referenced test technology is not very high.

23. Because item difficulties for individual persons are much less stable, this reasoning makes even less sense for interpreting an individual's score.

24. More than in any other subject matter field, curriculum developers in mathematics have classified content, problems, and skills as to both what might be taught and what should be taught. In their communications district supervisors and curriculum committees have tended to use only the broad headings (roughly equivalent to textbook chapter titles), much less detailed and without the interdependencies that characterize teacher conceptualizations of what needs to be taught.

25. Lawrence Stenhouse said "Good teachers are necessarily autonomous in professional judgement. They do not need to be told what to do. They are not professionally the dependents of researchers or superintendents or innovators or supervisors. This does not mean that they do not welcome access to ideas created by other people at other places or in other times. Nor do they reject advice, consultancy or support. But they do know that ideas and people are not of much real use until they are digested to the point where they are subject to the teacher's own judgement. In short, it is the task of all educationalists outside the classroom to serve the teachers; for only the teachers are in the position to create good teaching."

26. Which was one of the “anticipated student outcomes" for sixth graders in Duxbury.

27. Ken Komoski of the EPIE Institute led the development of an alignment technology (Komoski no date).

28. The words of Andy Porter are instructive: ''Simply telling teachers what to do is not likely to have the desired results. Neither is leaving teachers alone to pursue their own predictions. But it might be possible to shift external standard setting away from reliance on rewards and sanctions (power) and toward reliance on authority ... One approach to building authoritative standards would be to involve teachers seriously in the business of setting standards ... Through the process of teacher participation the standards would take on authority" (1989, p. 354). See also Dan Koretz 1989; National Council of Teachers of Mathematics, 1989; and Harry Torrance, forthcoming.

Bibliography

Aoki, T. (1983), "Curriculum implementation as instrumental action and as situational praxis." In Aoki, T. and others, (Eds.), "Understanding situational meanings of curriculum in-service acts: Implementing, consulting, in-servicing," *Curriculum Praxis Monograph Series*, 98. Edmonton: University of Alberta.

Ashton-Warner, S. (1967), *Teacher*. New York: Bantam.

Bereiter, C. (1991), "Implications of connectionism for thinking about rules." *Educational Researcher*, April, pp. 10-16.

Carl, I. (1991), "Better Mathematics for K-8". Streamlined Seminar 9, 4. National Association of Elementary School Principals, March 1991. 1-17.

Collis, K. F., & Watson, J. M. (1989), "ASOLO mapping procedure". *Proceedings of the Thirteenth International conference for the Psychology of Mathematical Education*. Paris: University of Paris.

Connell, R. W. (1985), *Teachers' Work*. Sydney: George Allen and Unwin.

Cronbach, L. J. (1980), "Validity on parole: How can we go straight?" *New Directions for Testing and Measurement—Measuring Achievement over a Decade*. Proceedings of the 1979 ETS Invitational Conference, pp. 99-108. San Francisco: Jossey-Bass.

Driver, R. (1973), *The Representation of Conceptual Frameworks in Young Adolescent Science Students*, University of Illinois, dissertation.

Easley, J. (1974), “The structural paradigm in protocol analysis." *Journal of Research in Science Teaching*. 2, 2 281-290.

Easley, J., & Easley, E. (1992), "Changing the Teaching of Mathematics". In *Elementary Classrooms*. Rob Walker (Ed.), Series on Classroom Research. Geelong: Deakin University.

Eraut, M., Goad, L., & Smith, G. (1975), *Analysis of Curriculum Materials*. University of Sussex Education Area Occasional Paper #2.

Farrell, E. (1990*), Hanging in and Dropping Out: Voices of At-Risk High School Students*. New York: Teachers College Press.

Freeman, D. J., Kuhs, T. M., Porter, A. C., Floden, R. E., Schmidt, W. H., & Schwille, J. R. (1983), "Do textbooks and tests define a national curriculum in elementary school mathematics?" *The Elementary School Journal*, 83, 3, pp. 500-513.

Gagné, R. M. (1967), "Curriculum research and the promotion of learning." *AERA Monograph Series on Curriculum Evaluation*, 1. Chicago: Rand McNally, 19-38.

Glaser, R. (1963), instructional technology and the measurement of learning outcomes." *American Psychologist*, 18, 519-521

Gowin, B. (1990), "Epistemic elements in evaluation research." *Studies in Educational Evaluation*, 16, 319-333;

Haertel, E. H. (1985), "Construct validity and criterion-referenced testing." *Review of Educational Research*, Spring, 55, 1, pp. 23-46.

Haertel, E. H., & Wiley, D. E. (1990), *Poset and Latice Representations of Ability Structures: Implications for Test Theory*. Paper delivered at the annual meeting of the American Educational Research Association, Boston.

Hambleton, R. K. (1989), ''Principles and selected applications of item response theory." In Linn, (Ed.), *Educational Measurement*, 3, pp. 147-200.

Haney, W. & Madaus, C. (no date), *Effects of Standardized Testing and the Future of the National Assessment of Educational Progress*. Working paper. Chestnut, MA: Center for the Study of Testing, Evaluation & Educational Policy.

Hively II. W., Maxwell, G., Rabehl, G., Sension, D., & Lundin, S. (1973), "Domain-referenced curriculum evaluation: A technical handbook and a case study from the Minnemast Project." CSE Monograph Series in Evaluation, 1. Los Angeles: Center for the Study of Evaluation, UCLA.

Jaeger, R. M. & Tittle, C. K., (Eds.) (1980), *Minimum Competency Achievement Testing. Motives, Models, Measures, and Consequences*. Berkeley: McCutchan.

Johnson, D. W. & Johnson, R. T. (1991), *Learning Together and Alone*, third edition. Englewood Cliffs: Prentice Hall.

Jonassen, D. H., (Ed.) (1982), *The technology of text*.. Englewood Cliffs, Educational Technology Publications.

Komoski, P. K. (no date), *Curriculum Analysis Report. (Using the Integrated Instructional Information Resource Data Base in the Battle Creek Public Schools)*. Watermill, NY: EPIE Institute.

Lampert, M. (1988), *The Teacher's Role in Reinventing the Meaning of Mathematical Knowing in the Classroom*. (Research Series # 186). Michigan State University: Institute for Research on Teaching.

Lieberman, A. (1984), *Teachers, Their World and Their Work*. Alexandria: Association for Supervision and Curriculum Development.

Linn, R. L., (Ed.) (1989), *Educational Measurement* (3rd edition). NY: Macmillan.

Lorge, I. (1951), "The fundamental nature of measurement." In E. F. Lindquist, (Ed.), *Educational Measurement*. Washington, D.C.: American Council on Education.

Lundgren, U. P. (1972), *Frame Factors and the Teaching Process*. Stockholm: Almqvist & Wiksell.

Madaus, G. F. (1991), *The Effects of Important Tests on Students: Implications for a National Examination or System of Examinations*. AERA Conference on Accountability as a State Reform Instrument: Impact on Teaching, Learning, Minority Issues and Incentives for Improvement, Washington, D.C., June 5, 1991.

McLean, L. D. (1982), "Achievement testing—Yes! Achievement tests—No." *E+M Newsletter*, 39, 1-2.

Messick, S. A. (1989), "Validity." In Robert L. Linn, (Ed.), *Educational Measurement*, 3.

Millman, J. (1974), "Criterion referenced measurement." In W. James Popham, (Ed.), *Evaluation in Education*. Berkeley: McCutchan, pp. 309-398.

National Council of Teachers of Mathematics (1989), *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: The Council.

Newman, F. M. (1991), "Beyond common sense in educational restructuring: The issues of content and linkage." Draft. Madison: University of Wisconsin Center on Organization and Restructuring of Schools. October 29, 1991.

Piaget, J. (1929), *The Child's Conception of the World*. London: Routledge & Kegan Paul.

Raven, J. (1992), "Toward a model of competence, motivation, and behavior and a paradigm for assessment." In Berlak, Adams, Archbald, Burgess, Newmann, Raven & Romberg.

Resnick, L. (1989), Developing mathematical knowledge. *American Psychologist*, 44, 2, pp. 162-169.

Romberg, T. A. & Carpenter, T. P. (1986), "Research on teaching and learning mathematics: Two disciplines of scientific inquiry." In Merlin C. Wittrock, (Ed.), *Handbook of research on teaching*. New York: Macmillan.

Sarason, S. (1971), *The Culture of the School and the Problem of Change*. Allyn & Bacon.

Sato, T. (1991), "Development of instructional materials in NEC Technical College: Concepts and practices." Tokyo: NEC Corporation

Scheffler, I. (1973), *Reason and Teaching*. Indianapolis: Bobbs-Merrill.

Scheffler, I. (1975) "Basic mathematical skills: some philosophical and practical remarks." *The NIE Conference on Basic Mathematical Skills and Learning*. Volume 1, pp. 182-189. Los Alamitos: Southwest Regional Laboratory.

School Mathematics Study Group (1961*), First Course in Algebra: Units 9 and 10*. New Haven: Yale University Press.

Shavelson, R., McDonnell, L. M., Oakes, J., & Carey, N. B. (I987), *Indicator Systems for Monitoring Mathematics and Science Education*. Santa Monica: Rand Corporation.

Shepard, L. (1991), *Will National Tests Improve Student Learning?* AERA Conference on Accountability as a State Reform Instrument: Impact on Teaching, Learning, Minority Issues and Incentives for Improvement, Washington, D.C., June 5, 1991.

Shohamy, E. (1984), "Does the testing method make a difference? The case of reading comprehension." *Language Testing*, 1, 2, pp. 147-170.

Simons, H., & Elliot, J. (Eds.) (1989), *Rethinking Appraisal and Assessment*. Open University Press.

Smith, M. S., & O'Day, J. (1990), "Systemic school reform." *Politics of Education Association Yearbook*, pp. 233-267.

Smith, M. L. (1991), "Put to the test: The effects of external testing on teachers." *Educational Researcher*, 20, 5, pp. 8-11.

Stake, R., Bresler, L., & Mabry, L. (1991), *Custom and Cherishing: The Arts in Elementary Schools.* Council for Research in Music Education. Urbana, IL.

Stake, R., & Easley, J. (Eds.) (1978), *Case Studies in Science Education*. Urbana: Center for Instructional Research and Curriculum Evaluation.

Stake, R., Raths, J., Denny, T., Stenzel, N., & Hoke, G. (1986), "Evaluation study of the Indiana Department of Education Gifted and Talented Program." Urbana: Center for Instructional Resources and Curriculum Evaluation.

Traxler, A. E. (1951), "Administering and scoring the objective test." In Lindquist, E. F. (Ed.), *Educational Measurement*. Washington, D.C.: American Council on Education.

Wilson, B. L., & Corbett, H. D. (1991), "Two state minimum competency testing programs and their effects on curriculum and instruction." *Advances in Program Evaluation*, 1B, pp. 7-40.

Wilson J. (197l) "Evaluation of learning in secondary school mathematics." In Bloom, B., Hastings, T., & Madaus, G. (Eds.), *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw Hill, 645-696.

10