DOCUMENT RESUME

ED 289 709 SE 048 794

AUTHOR Romberg, Thomas A., Ed.; Stewart, Deborah M., Ed.
TITLE The Monitoring of School Mathematics: Background Papers. Volume 2: Implications From Psychology; Outcomes of Instruction. Program Report 87-2.
INSTITUTION Wisconsin Center for Education Research, Madison.
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE Mar 87
GRANT SPA-8550193
NOTE 295p.; For related documents, see SE 048 793-795.
PUB TYPE Reports - Descriptive (141) -- Collected Works - General (020)

EDRS PRICE MF01/PC12 Plus Postage.
DESCRIPTORS Abstract Reasoning; Attitude Measures; Curriculum Development; Educational Assessment; Educational Change; *Educational Psychology; *Elementary School Mathematics; Elementary Secondary Education; Learning Processes; *Learning Theories; Logical Thinking; *Mathematics Curriculum; Mathematics Education; Mathematics Instruction; Mathematics Tests; *Secondary School Mathematics

ABSTRACT

This document is the second of three related volumes. They present the rationale, background, and framework for a comprehensive monitoring system being developed for the National Science Foundation. The system is being designed to gather information about the effects of national, state, and local policy actions designed to change the teaching and learning of mathematics in the schools of America. The papers included were produced by project staff, commissioned, or reprinted from previous works. Expert reviews and critiques of sets of papers are included. In this volume the implications of psychology to the learning of mathematics is addressed, and the problems of assessing learning based on both the new mathematical fundamentals and knowledge of learning are examined. Part 1, related to implications from psychology, summarizes advances in cognitive psychology, research on intrinsic motivation, the role of intuition, as well as a synthesis of psychological research in relation to curriculum engineering. Part 2 begins to address the issue of determining a reasonable approach to assessing the outcomes of instruction in mathematics due to shifts in emphasis related to recent reforms. (PK)

Program Report 87-2

# THE MONITORING OF SCHOOL MATHEMATICS: BACKGROUND PAPERS

## Volume 2: Implications from Psychology; Outcomes of Instruction

Edited by Thomas A. Romberg and Deborah M. Stewart

Report from the School Mathematics Monitoring Center

Thomas A. Romberg, Director

Wisconsin Center for Education Research
School of Education
University of Wisconsin-Madison

March 1987

Preparation of this document was supported by a grant to the Wisconsin Center for Education Research from the National Science Foundation for the establishment of a school mathematics monitoring center (NSF grant number SPA-8550193).

# Wisconsin Center for Education Research

## MISSION STATEMENT

The mission of the Wisconsin Center for Education Research is to improve the quality of American education for all students. Our goal is that future generations achieve the knowledge, tolerance, and complex thinking skills necessary to ensure a productive and enlightened democratic society. We are willing to explore solutions to major educational problems, recognizing that radical change may be necessary to solve these problems.

Our approach is interdisciplinary because the problems of education go far beyond pedagogy. We therefore draw on the knowledge of scholars in psychology, sociology, history, economics, philosophy, and law as well as experts in teacher education, curriculum, and administration to arrive at a deeper understanding of schooling.

Work of the Center clusters in four broad areas:

- Learning and Development focuses on individuals, in particular on their variability in basic learning and development processes.
- Classroom Processes seeks to adapt psychological constructs to the improvement of classroom learning and instruction.
- School Processes focuses on schoolwide issues and variables, seeking to identify administrative and organizational practices that are particularly effective.
- Social Policy is directed toward delineating the conditions affecting the success of social policy, the ends it can most readily achieve, and the constraints it faces.

The Wisconsin Center for Education Research is a noninstructional unit of the University of Wisconsin-Madison School of Education. The Center is supported primarily with funds from the Office of Educational Research and Improvement/Department of Education, the National Science Foundation, and other governmental and nongovernmental sources in the U.S.

CONTENTS

## CONTRIBUTORS TO THIS VOLUME

Kevin T. Collis, University of Tasmania, Australia

Brian Donovan, Philip Institute of Technology, Australia

Efriam Fischbein, Tel-Aviv University, Israel

James G. Greeno, Stanford University

Giyoo Hatano, Dokoyo University, Japan

Kayoko Inagaki, Dokoyo University, Japan

Gilah Leder, Monash University, Australia

Douglas McLeod, Washington State University

Gary G. Price, Wisconsin Center for Education Research
University of Wisconsin-Madison

Thomas A. Romberg, Wisconsin Center for Education Research
University of Wisconsin-Madison

Fredric W. Tufte, University of Wisconsin-Platteville

Norman L. Webb, Wisconsin Center for Education Research

E. Anne Zarinnia, Wisconsin Center for Education Research

PREFACE

This set of papers, published in three volumes as a monograph of the School Mathematics Monitoring Center, presents the rationale, background, and framework for a comprehensive monitoring system being developed for the National Science Foundation. The system is being designed to gather information about the effects of national, state, and local policy actions designed to change the teaching and learning of mathematics in the schools of America.

To build the monitoring system three assumptions were made. First, as a society we are involved in a major economic revolution. This revolution, addressed in Chapter 2, directly affects mathematics, its use, and what is deemed fundamental. As a consequence we believe "that most students need to learn more, and often different, mathematics" (Romberg, 1984, p. xi). Second, in spite of the changes in school mathematics inherent in the first assumption, we believe that there is general concensus about the goals for school mathematics and about the kinds of changes needed to achieve those goals. Thus, to develop the framework for the system one must begin with an understanding of those goals and the ideas on which they are based. Only then can indicators be developed to see whether the goals are being reached. Third, the policy actions with respect to the specific goals set for school mathematics must be consistent with the more general educational goals for a free and democratic society.

The need to monitor changes in school mathematics was proposed at two conferences. The first was organized by the Conference Board of the Mathematical Sciences (the New Goals Conference, CBMS, 1984), and the second by the National Council of Teachers of Mathematics, the U.S. Department of Education, and the Wisconsin Center for Education Research (School Mathematics: Options for the 1990s, Romberg, 1984). One conclusion from both conferences was that information about the nature of proposed changes and their effects on schooling practices was needed. During the past 25 years the federal government has invested considerable funds to change the teaching and learning of mathematics in America's schools, and today it is in the process of funding several new projects. Unfortunately, evidence of the impact of past dollars on classroom instruction is lacking. The special evidence that exists was unsystematically gathered and is incomplete. As new monies are spent and programs developed, it is crucial that a systematic plan be adopted to gather information about the effects of these planned changes.

During the past year the staff of the Monitoring Center prepared a series of papers, commissioned additional papers, convinced some authors to allow us to reprint a paper they had recently prepared, and asked a few nationally recognized experts to review and critique sets of papers. In all we have collected some 30 papers that address the issues of a new world view, what is fundamental in mathematics, what implications recent research in psychology or sociology has for school mathematics, etc. The intent of gathering these papers was to assist the staff of the project in the design of a monitoring system for school mathematics. However, since they comprise a review of the current thinking about schooling by a number of noted educators, we have chosen to publish them in this three-volume monograph so that others may have access to this information.

The first volume addresses the need for a monitoring center, the new world view, and what is now considered a fundamental for students to know about mathematics. In the second volume the implications of psychology to the learning of mathematics is addressed, and the problems of assessing learning based on both the new mathematical fundamentals and our knowledge of learning is examined. The final volume is comprised of papers that are based on current sociological notions about schools and how that knowledge affects the role of teachers and instruction in classrooms.

## IMPLICATIONS FROM PSYCHOLOGY

One of the primary sources of research findings that support the need for reform in the teaching and learning of mathematics is psychology. During the past quarter of a century there has been a major revolution in that field. Learners are no longer considered passive recipients of information that is fixed via reinforcement. Today learners are seen as active processors of information and constructors of knowledge. To portray the importance of this research for the reform movement in school mathematics, we have solicited five chapters.

In the initial chapter in this volume, chapter 12, Jim Greeno summarizes the recent advances in cognitive psychology. Giyoo Hatano and his colleague Kayoko Inagake summarize research on intrinsic motivation in chapter 13. In chapter 14, Efriam Fischbein covers the role of intuition in mathematical reasoning. Each of these chapters, written by internationally known psychologists who have worked in the learning of mathematics, portrays important aspects of recent work that has implications for the reform movement in mathematics. In chapter 15, Tom Romberg and Fredric Tufte provide a review and synthesis of some of the recent psychological research in relationship to curriculum engineering. Chapter 16, the final chapter in this section, contains a critical review of the previous chapters that was prepared by Gary Price.

Chapter 12

# MATHEMATICAL COGNITION: ACCOMPLISHMENTS AND CHALLENGES IN RESEARCH[1]

James G. Greeno

This paper presents an overview of research about knowledge and cognitive processes in mathematical problem solving and reasoning. I discuss broad trends that I illustrate with examples; this is not thorough review of research findings.

The paper has three main sections. First I discuss research accomplishments in the decade from the mid-1970s to the present. In this period we have been successful in establishing what can be called the Knowledge Structure Program for research in mathematics education. The dominant goal of this research has been to understand knowledge that is required for successful performance of school tasks. Considerable progress has been made in the form of cognitive models that simulate cognitive structures and processes that students acquire when they are successful in the tasks that are used in instruction. Results of this research are applicable in the design of new tasks and representations that address instructional problems, and some promising preliminary projects are under way.

Next I discuss an alternative that many consider preferable to the idea of knowledge structures as goals of mathematics education. Rather than focusing on the content of mathematics, instruction could attempt to provide abilities to think mathematically and cognitive resources for reasoning in situations other than classrooms. I discuss recent research findings in cognitive anthropology and developmental psychology that support the feasibility of these deeper goals of mathematics education and suggest some features of instruction that could be effective.

Then I discuss two general theoretical concepts about knowledge that seem particularly germane to the goals of mathematics education: the situated and generative character of knowledge. I describe some research related to these concepts, including some recent and current projects in instructional

---

[1] An earlier version of this paper was presented at the annual meeting of the American Educational Reseach Association in April 1986. I am grateful for discussions with my colleagues Andrea A. diSessa, Peter Pirolli, Frederick Reif, and Alan H. Schoenfeld about these matters, including reactions to a draft of this paper.

research and development, as illustrations of research directions that could inform educational development in the service of deeper instructional objectives. Finally, I offer a few conclusions.

## The Knowledge Structure Program

### Cognitive Models as Instructional Objectives

A program of research that became feasible in the mid-1970s has turned out to be remarkably productive and successful. An idea about formulating objectives of instruction in the form of cognitive models that simulate performance in school tasks was discussed programmatically at a conference held in 1974, sponsored by the Office of Naval Research. Hayes (1976), commenting on Greeno's (1976) discussion, put the idea as follows:

> Cognitive objectives in education [are] intended to replace the more traditional behavioral objectives. To specify a behavior objective for instruction, we state a particular set of behaviors we want the students to be able to perform after instruction, e.g., to solve a specified class of arithmetic problems or to answer questions about a chapter in a history text. To specify a cognitive objective, we state a set of changes we want the instruction to bring about in the students' cognitive processes, e.g., acquisition of a particular algorithm for division or the assimilation of a body of historical fact to information already in long-term memory. (pp. 235-236)

### Relevant Advances in Cognitive Psychology

The goal of formulating instructional objectives as cognitive models seemed a feasible program at the time because of two important advances in cognitive psychology that had just emerged: a model of problem solving and a model of language understanding.

A psychological model of knowledge used in solving novel problems was published by Newell and Simon in 1972. This work established both the feasibility of using ideas developed in artificial intelligence as a basis for developing hypotheses about human cognition and the methodology of testing those hypotheses using thinking-aloud protocols obtained while individuals work on solving problems. Newell and Simon characterized general strategies of problem solving, including means-ends analysis, that are effective when an individual without special instruction in a domain is given instructions about the states and operators that can be used to solve a puzzle. An important formal notion is the use of production rules to represent knowledge for cognitive activity. In a system of production rules, each rule specifies a pattern of information and an action, which may be a physical action or a cognitive action such as a decision or an inference, and the action is performed whenever the condition is true in the situation. A later development that was important for modelling knowledge for school tasks was a model of knowledge for planning, published by Sacerdoti in 1977. Sacerdoti characterized knowledge about actions in a domain with their consequences and prerequisites so that a planner can construct sequences of actions to achieve goals.

At about the same time there were significant advances in artificial intelligence and cognitive psychology regarding knowledge and cognitive processes involved in understanding language. Winograd (1972) developed a system that takes English sentences as input and constructs programs for examining conditions in an environment and moving objects about in the environment. Schank (1972) developed a system that converts English sentences to structures of information about the actions and situations that the sentences describe. Anderson and Bower (1973), Kintsch (1974), Norman, Rumelhart, and the LNR Research Group (1975), and others developed psychological models that simulate understanding of language based on use of schematic knowledge and propositional structures to form representations of meanings of sentences and paragraphs of text. Meanings are represented as semantic networks in which concepts correspond to nodes and relations among the concepts correspond to links. Knowledge in the form of schemata provides general structures that the understander uses to construct semantic networks for the meanings of specific sentences and situations. The outcome of understanding is a knowledge base that can be used to answer questions, either by retrieving information that was included directly in information that was understood, or by retrieving information that was inferred as part of the process of understanding, or by making inferences based on information that was understood.

## Progress in a Decade of Research

The idea that several investigators began to work on in the mid-1970s is that the concepts and methods of cognitive psychology, including the concepts of production systems, schemata, and semantic networks and the methods of protocol analysis and simulation modelling, could be used in understanding what students need to learn to succeed in school instruction. Students' learning is tested by questions they are asked and by problems they are required to solve. The research effort that I call the Knowledge Structure Program takes tasks that are used in instruction and constructs models of the knowledge required to perform the tasks successfully. Data used to guide construction of the models may include detailed analyses of successful student performance, often including thinking-aloud protocols. Data also may include characteristic errors of performance or reasoning, with explicit features of the models that overcome those difficulties. Some important analyses have been based mainly on considerations of the structure of subject-matter concepts and the experience of teachers regarding student difficulties. The strongest work has combined deep insights into the structure of subject-matter concepts with empirical and theoretical analyses of students' successful performance and their difficulties of understanding and learning.

Significant progress has been made in domains of school mathematics. First, cognitive procedures for solving routine problems of calculation have been simulated for elementary arithmetic (Brown & Burton, 1980) and algebra (Sleeman, 1984). These analyses include detailed hypotheses about the incorrect cognitive procedures of students who make systematic errors as well as the structure of procedures acquired by students who succeed. Simulations of problem-solving procedures of successful students in high school geometry have also been developed (Greeno, 1978). This analysis included hypotheses about schematic knowledge of general patterns that enables flexible planning and solution of problems requiring constructions.

Ideas about language understanding have been combined with problem-solving hypotheses in models of schematic and procedural knowledge for solving word problems in elementary arithmetic (Briars & Larkin, 1984; Kintsch & Greeno, 1985; Riley, Greeno, & Heller, 1983). Students' understanding based on schemata of general quantitative relations has been simulated in domains of computational procedures (Resnick, 1983; VanLehn & Brown, 1978) and proof exercises (Greeno, 1983).

Although most of the analyses of school mathematics tasks have been simulations of performance, a promising simulation of learning has been provided in the domain of high school geometry (Anderson, 1983), and a tutoring system based in part on that model has been developed and is being tested (Anderson, Boyle, Farrell, & Reiser, 1984).

Models of knowledge required for successful performance are potentially quite important for instruction, especially when they reveal aspects of knowledge that are implicit in performance. Implicit knowledge includes the patterns of information that students need to recognize in understanding word problems or the constraints of a computational procedure and the search strategies that are used to organize problem-solving activity when working on proof exercises. An important possibility for instruction is that by making some of these usually tacit components of knowledge explicit, students who would otherwise fail to acquire the knowledge that is needed for success might be able to succeed.

The analyses that have been developed do not "cover" the school mathematics curriculum, by any means. However, the feasibility of projects that would develop models of knowledge in the remaining topics seems well established. Many important aspects of problem solving, reasoning, and understanding remain to be analyzed in topics such as rational number computation, ratios, percentages, symbolic algebra, and graphing. Cognitive analysis in these and other domains undoubtedly will require significant effort and nontrivial insight. E en so, with a reasonable investment of scientific resources, it would not be surprising if a quite complete set of analyses of the standard precollege mathematics curriculum could be assembled within five to ten years, at the theoretical level that has been achieved in the analyses that I have mentioned.

## More Ambitious Goals for Mathematics Education

### Alternative Goals and Assumptions

One consequence of having models of knowledge for tasks that describe knowledge structures specifically is a possibility of reflecting on whether that knowledge is what we want students to learn. The models that simulate students' performance in routine mathematical tasks emphasize limitations that have been noticed many times. Students can learn to solve the problems that are used in standard instruction without acquiring very deep understanding of the mathematical concepts and principles that the problems are meant to convey, and learning to solve problems in the context of instruction often fails to transfer significantly to other contexts.

Many individuals have wished for a deeper orientation in the teaching of mathematics. Davis (1984) put the point as follows:

> Mathematics is presented from a wrong point of view: it is presented as a matter of learning dead "facts" and "techniques," and not in terms of its true nature, which involves processes that demand thought and creativity: confronting vague situations and refining them to a sharper conceptualization; building complex knowledge representation structures in your own mind; criticizing these structures, revising them and extending them; analysing problems, employing heuristics, setting sub-goals and conducting searches in unlikely (but shrewdly chosen) corners of your memory. (p. 347; emphasis in the original)

On this view, the goals of instruction in mathematics should be to strengthen students' abilities to understand and reason productively about the concepts and techniques of mathematics, rather than only knowing the content of the concepts and how to perform the techniques correctly.

This is a lofty goal--in effect, it proposes that students should learn to understand and reason in mathematics as mathematicians understand and reason. Opinions differ about whether such a goal is feasible. For example, a pessimistic view was laid out by Poincaré (1956).

> We know that this feeling, this intuition of mathematical order, that makes us divine hidden harmonies and relations, can not be possessed by every one. Some will not have either this delicate feeling so difficult to define, or a strength of memory and attention beyond the ordinary, and then they will be absolutely incapable of understanding higher mathematics. Such are the majority.

> Others will have this feeling only in a slight degree, but they will be gifted with an uncommon memory and a great power of attention. They will learn by heart the details one after another; they can understand mathematics and sometimes make applications, but they cannot create. Others, finally, will possess in a less or greater degree the special intuition referred to, and then not only can they understand mathematics even if their memory is nothing extraordinary, but they may become creators and try to invent with more or less success according as this intuition is more or less developed in them. (p. 2043)

Others are more optimistic. Davis (1984) asserted that

> the trials of the 1950s and 1960s demonstrated that students are well able, cognitively or intellectually, to move ahead far faster in mathematics and to deal with a "problem-analysis" and a "heuristic" approach to mathematics. (p. 348)

And in a delightful book titled Thinking mathematically, Mason, Burton, and Stacey (1982) presented the following optimistic message for their student readers:

> ASSUMPTION 1 You can think mathematically
>
> ASSUMPTION 2 Mathematical thinking can be improved by practice with reflection
>
> ASSUMPTION 3 Mathematical thinking is provoked by contradiction, tension and surprise
>
> ASSUMPTION 4 Mathematical thinking is supported by an atmosphere of questioning, challenging, and reflecting
>
> ASSUMPTION 5 Mathematical thinking helps in understanding yourself and the world (p. v, emphasis in original)

Historically, emphasis on rote training of calculation in the curriculum has been justified by a belief that most students could not achieve understanding of mathematical concepts and principles (Cohen, 1982). On the classical associationistic conception of learning, it is assumed that basic learning is the formation of bonds between ideas or between stimuli and responses, and that simple procedures such as arithmetic calculation are relatively easy to acquire (e.g., Thorndike, 1922). In that theory, conceptual understanding is harder to account for (e.g., Greeno, James, DaPolito, & Polson, 1978), and perhaps because of that theoretical difficulty it is expected that conceptual understanding requires exceptional ability by the learner.

Evidence in Developmental Psychology

Recent findings in developmental psychology support a very different picture of cognitive capabilities of young children than that of the classical association theory. In the domain of mathematics, Gelman and Gallistel (1978) found considerable evidence that preschool children implicitly understand principles of order, one-to-one correspondence, and cardinality, rather than having only a mechanical knowledge of counting rules and procedures. A telling piece of evidence is that children can modify their counting procedure correctly when an unusual constraint is imposed. After the child counted a set of objects the experimenter selected one of them and said, "Now count them again, but make this the 'one'". On different trials different objects were selected and different numerals were associated with the selected objects. Most five-year-olds produced counting performance that complied with the novel constraints as well as the principles of counting. Because these counting procedures could not have been learned, the children's generative knowledge must have included implicit understanding of the principles.

In a related domain, Bullock, Gelman, and Baillargeon (1982) showed that preschool children make judgments about causality that reflect significant implicit understanding of principles such as temporal order (causes precede their effects), local action, and mechanism. Children also probably have implicit understanding of causal relations among quantities--for example, throwing something harder makes it travel farther. diSessa (1983) has begun to formulate a theory of implicit structures of reasoning about quantitative causality that he calls phenomenological primitives.

Carey (1985) and Keil (in press) have studied children's knowledge about living things and have shown that their understanding grows in ways that reflect a structure of concepts and principles, rather than haphazard accretion of facts and experiences. Carey (1985) argued that, between the ages of about six and ten years, children move from an understanding of activity, body parts, and functions such as eating based on psychological concepts such as intention (e.g., people eat because they get hungry) to an understanding in terms of biological principles and concepts (e.g., people eat because food is needed to stay alive and grow). Keil (in press) provided particularly compelling evidence that children acquire principles with inferential force that goes beyond simple classification by features. He showed that principles of biological origin replace features of appearance in determining children's judgments of the category that animals belong to. Children were shown pictures of two animals, a raccoon and a skunk, and were told that an animal that used to look like the raccoon had been changed by some scientists to look like the other by changing its color, the shape of its tail, and its body size. Older children, though not younger ones, said that the animal was still a raccoon, because a change in appearance does not change what an animal is. On the other hand, changes in the functional properties of artificats related to their use lead children to change their judgement of what the object is--for example, when a coffee pot's features are changed to those of a bird feeder.

These studies and others strongly suggest that children's learning should be considered as an active process in which general principles and concepts play a significant role in organizing information and procedures that the child acquires. The fact that most children acquire the procedures of arithmetic more or less correctly but without significant understanding may be the result of a perverse method of instruction, rather than of any significant limitations of the children's ability to grasp the mathematical concepts and principles that make the procedures meaningful.

Evidence in Reasoning About Quantities Outside of School Settings

Further evidence of children's ability to reason intelligently with mathematical ideas, rather than merelv learning rote procedures, has been obtained in studies of performance of young salesmen and saleswomen in street markets in Recife, Brazil. Children who sell produce or lottery tickets compute complex quantities involving novel combinations virtually without errors. As an example, in a study of produce sellers (Carraher, Carraher, & Schliemann, 1985) a customer asked a 12-year-old saleswoman the price of ten coconuts that she had said cost 35 cents each. The reply was "Three will be 105; with three more, that will be 210. I need four more. That is--315--I think it is 350". Children whose computations in the market had been observed were later given a paper-and-pencil test of problems identical to problems they had solved correctly in the market; their average score was only 74%. Performance of children who sell lottery tickets is even more impressive, because their calculations depend on the number of combinations of numbers that can win, based on numbers chosen by the bettor (Acioly & Schliemann, 1986).

The important characteristic of quantitative reasoning by the street marketeers in Brazil is its situatedness--it is richly connected to the setting in which it occurs. This also characterizes performance of adults who have been observed in tasks that involve reasoning about quantities in practical settings. Scribner (1984) studied performance in the task of preloading orders in a dairy, a poorly paid job that is done in a cold-storage room and presumably does not attract workers who have achieved high levels of academic success. The preloaders are given orders to assemble in an unusual notation: a number of cases, a + or a - sign, and a number of units. "a + b" means "a" full cases and "b" additional units, and "a - b" means "a" full cases less "b" units. Actions of the preloaders in assembling the orders were observed, and in most cases they chose an action that required minimal effort. This frequently involved use of a partially filled case and a conversion of the problem; for example, to assemble a "1 - 6" order of a product that has 16 units per case, a literal solution would be to remove six units from a full case, but if there was a half-filled case available, preloaders typically used that and added two units to it.

Lave, Murtaugh, and de la Rocha (1984) have studied quantitative reasoning of shoppers and individuals learning to control their diets. They found that calculation was involved in a significant number of decisions made by individuals shopping for groceries--about one in every six items purchased involved explicit consideration of alternatives. And virtually all of the calculations--98%--were correct. But many of the calculations also were nonstandard. In one example, the price marked on a package of cheese seemed too high, but rather than multiplying its weight by the unit price, the shopper searched for another similar package to confirm that the marked price was in error. In comparing a 32-ounce package of noodles priced at $1.12 and a 64-ounce package priced at $1.79, a shopper said, "That's two dollars for four pounds. This is a dollar. That's 50 cents a pound, and I just bought two pounds for a dollar twelve, which is 60. So there is a difference." Arithmetic is apparently used to explain or justify quantitative judgments that are made informally; in the last case, the initial approximation did not agree with a judgment that the shopper already had made (and announced), but an adjusted approximation was more satisfactory. In contrast with their accuracy in judging best buys, the shoppers in Lave et al.'s study only scored 59% on a paper-and-pencil test of arithmetic operations involving integer, decimal, and fractional numbers.

An especially clear example of generative quantitative reasoning situated in a task setting was observed in de la Rocha's study of reasoning in the kitchen. A new member of Weight Watchers was asked to work out an allotment of cottage cheese that is three-quarters of the two-thirds cup the program allows. The person filled a measuring cup two-thirds full, dumped the cottage cheese onto a cutting board, spread it into a circle, marked the circle into four quadrants, removed one of the quadrants, and served the rest. de la Rocha also found many examples in which individuals created alternatives to standard measuring procedures, honoring equivalences of units--for example, using the decoration on a drinking glass to measure an amount of milk, or a number of serving spoons of rice that is equal to a prescribed fraction of a cup.

## Directions for Research and Development

The Knowledge Structure Program discussed earlier has provided analyses of performance in standard instructional tasks. That is now an established research effort and can be continued productively with strong potential benefits for cognitive theory and educational practice.

At the same time, there are opportunities to develop new directions for research and instructional development related to deeper goals than those that currently dominate mathematics education. I now discuss two general issues for which research findings and methods are in a less developed state than the Knowledge Structure Program, but there have been some beginnings. The two issues are understanding knowledge as a resource for reasoning and instructional settings that promote conceptual growth. These issues arise from the two main features of productive knowledge seen in the research discussed above: it is situated, and it is generative.

## Understanding and Fostering Knowledge Resources for Situated Reasoning

I discussed research in the previous section that indicates that individuals, including unschooled children, reason in flexible and strong ways about quantities in practical situations. The relation of school mathematics to this situated reasoning is tenuous, at most; indeed, in the cases that have been studied it can be argued that mathematics learned in school plays no helpful role in the individuals' reasoning and problem solving. At the same time, the reasoning that has been demonstrated occurs at a low level of mathematics. As Resnick (in press) has noted, nearly all of the examples that have been observed are limited to additive compositions of quantities.

A major educational advance would be achieved if we could find ways to teach mathematics beyond the level of addition and subtraction so that it would become part of individuals' reasoning in everyday situations.[2] This goal is not an easy one to achieve, and recent theoretical analyses have begun to clarify reasons for the difficulty.

Recent analyses have focused on a crucial distinction between symbolic knowledge and knowledge for activity in physical and social situations. School instruction in mathematics and other subjects is primarily in symbolic domains. If symbolic knowledge transferred easily into physical and social situations, school-based knowledge would be applied naturally and broadly.

Two important recent discussions have emphasized the distinction between symbolic and situated knowledge in the context

---

[2] I am assuming that little everyday reasoning by most persons even includes multiplicative and proportional relations, although the evidence that I have for that involves extrapolations from laboratory studies where proportional reasoning is often problematic. If some level of arithmetic above addition and subtraction is commonly used in everyday reasoning, then my remarks would apply to a somewhat higher level of mathematics instruction.

of computer programs. Dreyfus and Dreyfus (1986) and Winograd and Flores (1986) have developed arguments that use an idea that Heidegger developed. Heidegger argued that most of the interactions we have with objects in the world are direct, rather than involving intermediate representations such as images or descriptions. Symbolic representations play a significant role in cognition when something in the world departs from what an individual expects. As an example, the action of opening a door, including reaching to the doorknob, grasping and turning it, and pushing or pulling, is ordinarily done without any significant processing of symbols. However, if the knob doesn't turn or the door is stuck, the individual may well engage in some propositional reasoning ("Is it locked? Do I have the key?") or create a mental model to help in inferring where to push or kick the door to get it to open.

Dreyfus and Dreyfus (1986) used Heidegger's idea in analyzing the acquisition of cognitive skills. They argued that rules, descriptions, and explanations play a significant role only in the early stages of acquiring a skill, and that expertise in a domain depends crucially on acquisition of knowledge for responding directly to a very large variety of patterns in complex and flexible ways, most of which is not articulable in verbal or other symbols. While this general idea has been expressed before, notably in Fitts' (1962) theory of skill acquisition, Dreyfus and Dreyfus' emphasis on the limits of symbolic representations to the early stages of skill acquisition sheds new light on the significance of the analysis.

Another recent analysis by Smith (1983) provides a framework for clarifying the problem further. Smith's analysis is also in the domain of computer programming, but like the analyses discussed previously, it apples as well to procedures that are learned by students in mathematics instruction. Smith was concerned with the semantics of programming languages and provided an integration of two previously separate ideas of meaning.

The left panel of Figure 1 shows some of the components of Smith's analysis. He considered a field of symbolic expressions and a domain of objects that the expressions can refer to. In a programming language, the symbolic field is the set of data structures that can be expressed. In mathematics, the symbolic field is the set of expressions that can be written with numerals, operators, variables, and so on. There is a mapping from the symbolic field to the denotational domain, $\phi$, in the manner of standard model-theoretic semantics. There also is a mapping within the symbolic domain, $\Psi$, which refers to the rules for transforming expressions into other expressions. In a programming language, $\Psi$ is the set of transformations that can be made using statements in the language. In mathematics, $\Psi$ is the set of transformations that can be performed with the rules that are available. An important result in Smith's analysis is a set of conditions on $\Psi$ and $\phi$ that make them coherent. It is important that the transformations on symbols do not change the denotations and truth-values of expressions, and Smith showed how an appropriate set of coherence conditions can be satisfied. (In effect, this generalizes the metamathematical concept of soundness.)

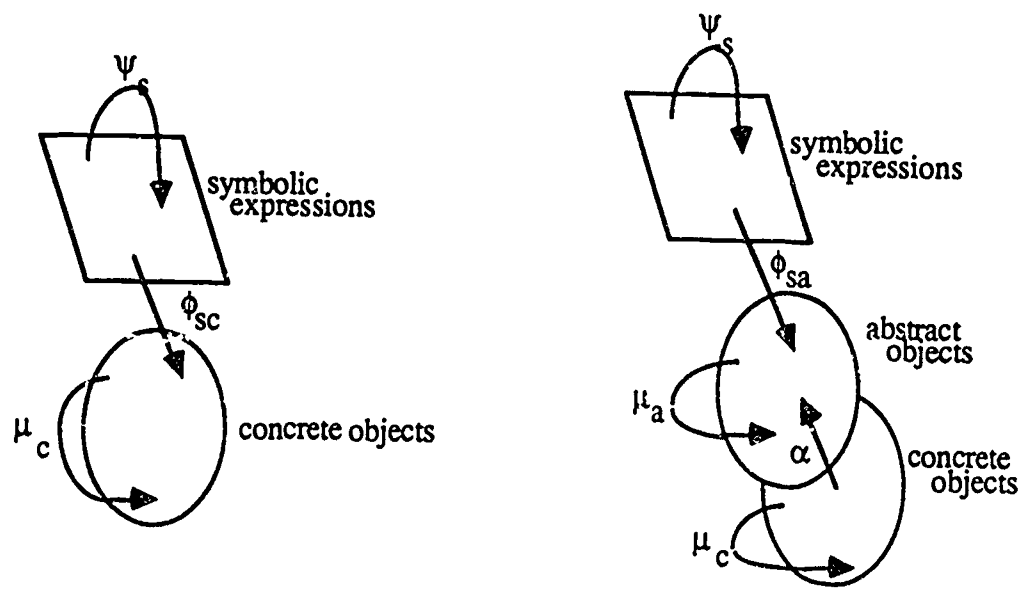



Figure 1. Components of an analysis of symbols and meanings.

I include the right side of Figure 1 to emphasize that in mathematics the denotations of expressions are primarily abstract entities--numbers, operations, functions, and so on--that can be understood as abstract structures in physical and social situations.

The mapping $\mu$ in Figure 1 refers to transformations that can be performed on the objects in a domain. $\mu_c$ refers to transformations on ordinary objects--moving them about, for example. $\mu_a$ refers to transformations of abstract entities, such as adding two numbers.

I now can state a conjecture about the reason that school mathematics learning transfers so poorly to reasoning in physical and social situations. School mathematics instruction focuses on symbolic operations, $\psi_s$. Students may even believe that the symbolic operations are a self-contained system that is unconnected with any referents in the world. (Children interviewed by Ginsburg, 1977, for example, seem to take that view.) Expert mathematicians understand that the symbols refer to the abstract entities of mathematics; that is, they have a conceptual domain containing those entities, they know the mapping $\phi_{sa}$, and they know what transformations in that domain, $\mu_a$, correspond to transformations of symbols, $\psi_s$, because of the denotational mapping $\phi_{sa}$. In contrast, children may well learn the manipulations of symbols, $\psi_s$, without connecting them to their denotations in the domain of either abstract entities or concrete objects. The quantitative reasoning of unschooled domain experts involves a manipulation of quantities, $\mu_a$, in contexts of specific domains of objects, and their lack of success in paper-and-pencil tests indicates that these operations are not connected well with symbolic expressions of arithmetic. It is reasonable to conjecture that the abstract structures that these individuals have are not as general as those that are known by experts in mathematics. Indeed, there is evidence (L. Resnick, personal communication) that the reasoning of unschooled experts is limited to a subset of numbers that occur frequently in the domain.

The question of teaching so that operations on symbols are meaningful has been a concern of many educators and cognitive psychologists. Wertheimer (1945/1959) and Dienes (1967) provided examples that became classical, involving spatial representations coordinated with formulas and proofs in geometry and algebra. The use of manipulative materials in the teaching of arithmetic has been advocated and studied at least since Brownell's (1935) well-known work, and Bransford (1986) is developing methods of providing concrete contexts for using arithmetic to solve problems that use the technology of video disks.

The mere use of concrete materials and contexts does not guarantee that children will understand the meanings of symbolic expressions and operations, of course. The framework provided by recent discussions of symbolic representations and cognitive skill may enable a clearer theoretical characterization of the conditions for such instruction to be effective. In particular, the idea would be that to understand the meanings of mathematical symbols it is important for students to acquire the appropriate mathematical concepts that the symbols denote. These are abstract structures, and they probably are not acquired automatically by experiencing connections between the symbols and specific concrete embodiments. Dienes' (1967) idea of multiple embodiments of concepts and Skemp's (1979) discussions of abstraction are clearly relevant to this task, but the various illustrations of concepts need to be carefully focused on specific conceptual targets and related systematically to symbolic expressions and operations.

The relations between alternative representations of abstract concepts is not a simple matter; some of their complexities have been discussed recently by Schoenfeld (in press, a). Recent findings by Resnick and Omanson (in press) illustrate the complexity of these matters. They conducted a systematic study of the effects of an instructional procedure developed and discussed previously by Resnick (1983) for multidigit subtraction. Resnick was concerned with students who made systematic errors in their test performance of a kind studied in detail by Brown and Burton (1980) and called "bugs," by analogy to flawed computer programs. Resnick had preliminary success with a procedure called mapping instruction, in which a procedure of subtraction with place-value blocks is taught and related in a detailed, step-by-step fashion with the paper-and-pencil procedure of subtraction with numerals. Resnick and Omanson's study applied the mapping instruction systematically to a number of children with bugs. Although a few of the children learned to subtract correctly, several did not. There was an intriguing trend in the data for those children who were remediated to talk about the quantities represented in the problems more than the children whose performance remained buggy. The trend in the data should be examined in a systematic way, but it is consistent with the conjecture that understanding involves linking symbolic expressions with abstract conceptual structures, rather than only with concrete objects.

Some recent results by Brown and Kane (1986) are suggestive about the process of acquiring general concepts involving relations between domains. Brown and Kane addressed the issue of transfer and showed that children can learn in ways that transfer to new problems when (a) they have a positive set to learn generalizations rather than solutions of specific problems, (b) they perceive the solution tool of a problem as one of many uses of the tool, and (c) the structure of analogous problems is made salient to the children. These conclusions, coupled with the suggestive trend in Resnick and Omanson's data, suggest that instruction that includes discussion as well as presentation of the general properties of quantities and their representations both in written symbols and concrete materials might be especially effective. Exploration of this possibility seems a useful target for research.

## Instruction for the Growth of Conceptual Systems

In this final section I discuss some ideas and frameworks for developing educational systems that could support the kind of deep conceptual growth that is needed for students both to understand the concepts and principles of mathematics and to use those concepts as resources for reasoning in the situations of their nonacademic lives.

An idea that may be very useful has been developed by Kitcher (1984) in an analysis of mathematical knowledge. Kitcher developed the idea of a mathematical practice, which he used to analyze significant historical changes in mathematics.[3] Components of a mathematical practice include (1) the questions that are understood as meaningful and legitimate, (2) the methods of reasoning that are accepted as supporting conclusions, and (3) a set of

[3]The idea is meant to capture valid aspects of Kuhn's (1970) concept of a paradigm while avoiding the excesses of Kuhn's concept. For example, an important part of Kitcher's accomplishment is to show in considerable detail how changes in practice occur naturally as progress within a field, not as a revolution that restructures the entire framework of inquiry, and how meaningful communication occurs between adherents of different practices as part of the process of modifying and extending knowledge.

metamathematical views that characterize goals and structures of mathematical knowledge, as well as (4) the mathematical language and (5) the statements of findings and conclusions that are accepted as established.

The educational idea that Kitcher's discussion suggests is that we could try to communicate significant components of mathematical practice to school children, rather than only communicating mathematical concepts and techniques. This idea is consistent with a view that students should learn processes of mathematical thinking, rather than only the content of mathematics. However, Kitcher's formulation of the components of mathematical practice could be a beginning of a more explicit formulation of the goal of teaching students to think mathematically.

Current instruction focuses on the fourth and fifth components of Kitcher's list, the language of mathematics and the accepted findings and conclusions. The further goals of educating students in mathematical practice would include questions, methods of reasoning, and metamathematical views. That is, we would attempt, in mathematics instruction, to educate students so they would be able to ask meaningful mathematical questions, construct and evaluate arguments, and understand the goals and structures of mathematical knowledge. All of these goals are attractive, and they have been proposed before (for example, see Brown & Walter, 1983; Kilpatrick, in press; and Schoenfeld, 1985, especially chapter 5). The question is what we can do now to make these goals more feasible and effective as guides for educational practice.

Each of these goals of education--asking questions, formulating and evaluating sequences of reasoning, and understanding metamathematical views--involves cognitive capabilities that are poorly understood. We now know how to analyze cognitive capabilities for solving problems and answering questions, and these scientific advances have potential value for developing improved instruction for problem solving and question answering. To move from this successful program of research to the deeper issues of questioning, reasoning, and metamathematics (in Kitcher's sense) would take cognitive research into territory that is almost entirely uncharted, but it would provide important opportunities to extend cognitive theory as well as potentially significant resources for changing mathematics education.

In fact, progress on achieving educational goals will be needed if we are to make progress on the theoretical questions of questioning, reasoning, and metamathematical beliefs. These deeper educational objectives are not achieved frequently in current educational practice, and therefore there are few opportunities to study the phenomena that we want to understand. To study these phenomena from a cognitive standpoint, as well as to provide examples for educational practice, we need to create environments in which students learn to ask meaningful questions, compose arguments, and come to understand metamathematical considerations.

It will require modifications of the environments in which we conduct education to achieve the deeper intellectual goals of communicating mathematical practice. Some interesting innovations have been and are being explored, and I close this essay with a brief characterization of some of their features.

The main feature of learning that is emphasized in recent research and the idea of acquiring a practice is a more active role played by learners. We are coming to understand several ways in which learning involves construction of knowledge, rather than its passive acquisition. Environments that encourage the construction of knowledge include (1) collaborative settings in which teachers and students work together to construct meanings and ideas; (2) settings in which teachers or tutors function as coaches and models of the activities the students are learning to engage in; and (3) settings in which students engage in exploration of ideas and environments.

A classic case of collaborative learning was described by Fawcett (1938), who developed a course in deductive reasoning that included geometry as well as material from everyday life such as newspaper articles and advertisements. Fawcett and his students discussed definitions of concepts, assumptions that were required for conclusions to follow, the relative advantages of different ways of proving conclusions, and other aspects of reasoning that are ordinarily not explicitly discussed in geometry courses. Lampert (in press) is providing a current example of collaborative instruction in her teaching of mathematics in the fifth grade. Lampert and her students engage in conversations about the meanings of mathematical concepts, operations, and notation, and the students play an active role in the process of making sense of mathematics. Activities of collaborative mathematical work probably offer the best chance of educating students for activities of the practice of mathematics. As Schoenfeld (in press, b) put it:

> A significant part of what I attempt to do (in my problem solving courses in particular, but increasingly in all of my mathematics instruction) is to create a microcosm of mathematical culture--an environment in which my students create and discuss mathematics in much the same way that mathematicians do. Having experienced mathematics in this way, students are more likely to develop a more accurate view of what mathematics is and how it is done. (p. 23)

A second way of organizing an instructional environment emphasizes modelling by an instructor of the kind of activity that students are attempting to acquire and then coaching the students as they carry out the activity. This is the standard method of instruction in domains that are understood primarily as domains of skill, such as athletics or musical performance. It has been less standard in school subjects, perhaps because we have understood these as consisting of knowledge, rather than skill. But if we shift our goals toward having students learn the practice of mathematics, modelling and coaching will become more appropriate as teaching methods. Modelling and coaching have been discussed especially in the context of increasing students' metacognitive skills, for example by Palincsar and Brown (1984) in reading comprehension, by Bereiter and Scardamalia (1982) in written composition, by Schoenfeld (in press, b) in mathematical problem solving, by Brown, Burton, and deKleer (1982) in electronic troubleshooting, and by Burton and Brown (1982) in strategies of an arithmetic game.

Flexible learning activities can also be encouraged in environments in which students can explore the structure of an environment, generate and test their own hypotheses, and discuss the phenomena that they experience. Exploratory environments for learning can be quite open (e.g., Papert, 1980), or they can have relatively definite structure designed to communicate quite specific ideas. Relatively structured microworlds and systems for representing problems have been developed and discussed by many individuals, for example, by Bork (1981), diSessa (1982), Greeno (in press), Schwartz (1985), and Schwartz, Yerushalmy, and Gordon (1985).

Cole and his group (Laboratory of Comparative Human Cognition, 1982) have created and are studying an environment that combines aspects of exploration, coaching, and collaboration. Their experiment is in many ways the most adventurous of the various attempts to construct new environments for learning.

## Conclusions

I have been discussing recent advances in theory and research that are relevant to some problems of long standing. The problems of teaching mathematics so that its concepts and principles are understood and so that it can be used by students in their everyday activities have been recognized for decades. These are not the kinds of problems for which we are likely to find "solutions" in the usual sense. I am impressed with another idea about problems, however, that was spelled out in a book about metaphor by Lakoff and Johnson (1980).

> An Iranian student, shortly after his arrival in Berkeley, took a seminar on metaphor from one of us. Among the wondrous things that he found in Berkeley was an expression that he heard over and over and understood as a beautifully sane metaphor. The expression was "the solution of my problems"--which he took to be a large volume of liquid, bubbling and smoking, containing all of your problems, either dissolved or in the form of precipitates, with catalysts constantly dissolving some problems (for the time being) and precipitating out others. He was terribly disillusioned to find that the residents of Berkeley had no such chemical metaphor in mind. And well he might be, for the chemical metaphor is

> both beautiful and insightful. It gives us a view of problems as things that never disappear utterly and that cannot be solved once and for all. All of your problems are always present, only they may be dissolved and in solution, or they may be in solid form. The best you can hope for is to find a catalyst that will make one problem dissolve without making another one precipitate out. And since you do not have complete control over what goes into the solution, you are constantly finding old and new problems precipitating out and present problems dissolving, partly because of your efforts and partly despite anything you do.
>
> The CHEMICAL metaphor gives us a new view of human problems. It is appropriate to the experience of finding that problems which we once thought were "solved" turn up again and again. The CHEMICAL metaphor says that problems are not the kind of things that can be made to disappear forever. To treat them as things that can be "solved" once and for all is pointless. To live by the CHEMICAL metaphor would be to accept it as a fact that no problem ever disappears forever. Rather than direct your energies toward solving your problems once and for all, you would direct your energies toward finding out what catalysts will dissolve your most pressing problems for the longest time without precipitating out worse ones. The reappearance of a problem is viewed as a natural occurrence rather than a failure on your part to find "the right way to solve it." (pp. 143-144)

The problems of understanding and reasoning in and with mathematics surely are the kind to which the chemical metaphor applies; they will not be solved in a simple way, and they probably will not go away completely. It is still reasonable, of course, to work toward improving on the solution that has already been achieved. Perhaps some reagents can be found that can cause more of these problems to go into solution without causing other problems to reappear more stubbornly.

The Knowledge Structure program of research has clarified the solution that we currently have. The models of knowledge structures that have been developed show the essential characteristics of knowledge that many students acquire in order to be successful in tasks that are used in instruction. It is likely that instruction in performing those tasks can be improved, partly because of the clearer definitions of the needed structures that cognitive models are providing.

Those models also provide a clearer view of important limitations of instruction that uses those tasks. The models reinforce our realization that students can learn to solve the problems that are used in instruction without achieving significant understanding of mathematical principles and concepts and without realizing that mathematical knowledge is a significant resource for reasoning in a broad range of nonacademic settings.

The task of moving toward a better understanding of how to teach mathematics more meaningfully is one that has attracted much research attention in the past and will continue to be an important and productive topic. Recent developments in several fields provide resources that can play a role in the next phase of this effort. These include important recent work in the study of cognitive development of children, studies of reasoning processes by children and adults in practical settings, studies of expert reasoning, progress toward a theory of meaning of symbolic representations, and significant development of new instructional settings. The detailed implications of these ideas for mathematics education are not completely clear yet, but they give considerable promise to the prospects for significant progress during the next period of research and educational development.

References

Acioly, N. M., & Schl'_mann, A. D. (1986). Intuitive mathematics and schooling in understanding a lottery game. Recife, Brazil: Universidade Federal de Pernambuco.

Anderson, J. R. (1983). Acquisition of cognitive skill. Psychological Review, 89, 396-406.

Anderson, J. R., & Bower, G. H. (1973). Human associative memory. Washington, DC: V. H. Winston & Sons.

Anderson, J. R., Boyle, C. F., Farrell, R., & Reiser, B. (1984). Cognitive principles in the design of computer tutors. Proceedings of the Cognitive Science Society, 2-11.

Bereiter, C., & Scardamalia, M. (1982). From conversation to composition: The role of instruction in a developmental process. In R. Glaser (Ed.), Advances in instructional psychology (vol. 2, pp. 1-64). Hillsdale, NJ: Erlbaum.

Bork, A. (1981). Learning with computers. Bedford, MA: Digital Press.

Bransford, J. (1986, April). The development of conceptual tools. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Briars, D. J., & Larkin, J. H. (1984). An integrated model of skill in solving elementary word problems. Cognition and Instruction, 1, 245-296.

Brown, A. L., & Kane, M. J. (1986). Analogical transfer in young children: A theoretical conundrum. Unpublished manuscript, University of Illinois at Urbana-Champaign.

Brown, J. S., & Burton, R. R. (1980). Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 4, 379-426.

Brown, J. S., Burton, R. R., & deKleer, J. (1982). Pedagogical, natural language and knowledge engineering techniques in SOPHIE I, II, and III. In D. Sleeman & J. S. Brown (Eds.), Intelligent tutoring systems (pp. 227-282). New York: Academic Press.

Brown, S. I., & Walter, M. I. (1983). The art of problem posing. Philadelphia: Franklin Institute.

Brownell, W. A. (1935). Psychological considerations in the learning and teaching of arithmetic. In The teaching of arithmetic: Tenth yearbook of the National Council of Teachers of Mathematics (pp. 1-31). New York: Columbia University Press.

Bullock, M., Gelman, R., & Baillargeon, R. (1982). The development of causal reasoning. In W. J. Friedman (Ed.), The developmental psychology of time. New York: Academic Press.

Burton, R. R., & Brown, J. S. (1982). An investigation of computer coaching for informal learning activities. In D. Sleeman & J. S. Brown (Eds.), Intelligent tutoring systems (pp. 79-98). New York: Academic Press.

Carey, S. (1985). Conceptual change in childhood. Cambridge, MA: Bradford/MIT Press.

Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1985). Mathematics in the streets and in schools. British Journal of Developmental Psychology, 3, 21-29.

Cohen, P. C. (1982). A calculating people: The spread of numeracy in early America. Chicago: University of Chicago Press.

Davis, R. B. (1984). Learning mathematics: The cognitive science approach to mathematics education. Norwood, NJ: Ablex.

Dienes, Z. P. (1967). Building up mathematics (rev. ed.). London: Hutchinson Educational Press.

diSessa, A. A. (1982). Unlearning Aristotelian physics: A study of knowledge-based learning. Cognitive Science, 6, 37-75.

diSessa, A. A. (1983). Phenomenology and the evolution of intuition. In D. Gentner & A. Stevens (Eds.), Mental models (pp. 15-33). Hillsdale, NJ: Erlbaum.

Dreyfus, H. J., & Dreyfus, S. E. (1986). Mind over machine: The power of human intuition and expertise in the era of the computer. New York: Free Press.

Fawcett, H. P. (1938). The nature of proof. The thirteenth yearbook of the National Council of Teachers of Mathematics. New York: Teachers College, Columbia University.

Fitts, P. M. (1962). Factors in complex skill training. In R. Glaser (Ed.), Training research and education (pp. 177-197). Pittsburgh: University of Pittsburgh Press.

Gelman, R., & Gallistel, C. R. (1978). The child's understanding of number. Cambridge, MA: Harvard University Press.

Ginsburg, H. P. (1977). Children's arithmetic: The learning process. New York: Van Nostrand.

Greeno, J. G. (1976). Cognitive objectives of instruction: Theory of knowledge for solving problems and answering questions. In D. Klahr (Ed.), Cognition and instruction (pp. 123-160). Hillsdale, NJ: Erlbaum.

Greeno, J. G. (1978). A study of problem solving. In R. Glaser (Ed.), Advances in instructional psychology (vol. 1, pp. 13-75). Hillsdale, NJ: Erlbaum.

Greeno, J. G. (1983). Forms of understanding in mathematical problem solving. In S. G. Paris, G. M. Olson, & H. W. Stevenson (Eds.), Learning and motivation in the classroom. Hillsdale, NJ: Erlbaum.

Greeno, J. G. (in press). Instructional representations based on research about understanding. In A. H. Schoenfeld (Ed.), Cognitive science and mathematics education. Hillsdale, NJ: Erlbaum.

Greeno, J. G., James, C. T., DaPolito, F., & Polson, P. G. (1978). Associative learning: A cognitive analysis. Englewood Cliffs, NJ: Prentice-Hall.

Hayes, J. R. (1976). It's the thought that counts: New approaches to educational theory. In D. Klahr (Ed.), Cognition and instruction (pp. 235-242). Hillsdale, NJ: Erlbaum.

Keil, F. (in press). On the acquisition of natural kind and artifact terms. In W. Demopolous (Ed.), Conceptual change. Norwood, NJ: Ablex.

Kilpatrick, J. (in press). Problem formulating: Where do good problems come from? In A. H. Schoenfeld (Ed.), Cognitive science and mathematics education. Hillsdale, NJ: Erlbaum.

Kintsch, W. (1974). The representation of meaning in memory. Hillsdale, NJ: Erlbaum.

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. Psychological Review, 92, 109-129.

Kitcher, P. (1984). The nature of mathematical knowledge. New York: Oxford University Press.

Kuhn, T. S. (1970). The structure of scientific revolutions. Chicago: University of Chicago Press.

Laboratory of Comparative Human Cognition. (1982). A model system for the study of learning difficulties. San Diego: University of California, Center for Human Information Processing.

Lakoff, G., & Johnson, M. (1980). Metaphors we live by. Chicago: University of Chicago Press.

Lampert, M. (in press). On learning, doing, and teaching multiplication. Cognition and instruction.

Lave, J., Murtaugh, M., & de la Rocha, O. (1984). The dialectics of arithmetic in grocery shopping. In B. Rogoff & J. Lave (Eds.), Everyday cognition: Its development in social context (pp. 67-94). Cambridge, MA: Harvard University Press.

Mason, J., Burton, L., & Stacey, K. (1982). Thinking mathematically. London: Addison-Wesley.

Newell, A., & Simon, H. A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Norman, D. A., Rumelhart, D. E., & the LNR Research Group. (1975). Explorations in cognition. San Francisco: Freeman.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction, 1, 117-175.

Papert, S. (1980). Mindstorms: Children, computers and powerful ideas. New York: Basic Books.

Poincaré, H. (1956). Mathematical creation. In J. R. Newman (Ed.), The world of mathematics (vol. 4, pp. 2041-2052). New York: Simon and Schuster.

Resnick, L. B. (1983). A developmental theory of number understanding. In H. P. Ginsburg (Ed.), The development of mathematical thinking. New York: Academic Press.

Resnick, L. B. (in press). The development of mathematical intuition. In M. Perlmutter (Ed.), Minnesota symposium on child psychology (vol. 4). Boston: Allyn and Bacon.

Resnick, L. B., & Omanson, S. F. (in press). Learning to understand arithmetic. In R. Glaser (Ed.), Advances in instructional psychology (vol. 3). Hillsdale, NJ: Erlbaum.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), The development of mathematical thinking. New York: Academic Press.

Sacerdoti, E. D. (1977). A structure for plans and behavior. New York: Elsevier-North Holland.

Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. Cognitive Psychology, 3, 552-631.

Schoenfeld, A. H. (1985). Mathematical problem solving. Orlando, FL: Academic Press.

Schoenfeld, A. H. (in press, a). On having and using geometric knowledge. In J. Hiebert (Ed.), Conceptual and procedural knowledge: The case of mathematics. Hillsdale, NJ: Erlbaum.

Schoenfeld, A. H. (in press, b). What's all the fuss about metacognition? In A. H. Schoenfeld (Ed.), Cognitive science and mathematics education. Hillsdale, NJ: Erlbaum.

Schwartz, J. L. (1985, April). The power and the peril of the particular: Thoughts on a role for microcomputers in science and mathematics education. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Schwartz, J. L. Yerushalmy, M., & Gordon, M. (1985). The geometric supposer: Triangles. Pleasantville, NJ: Sunburst Communications.

Scribner, S. (1984). Studying working intelligence. In B. Rogoff & J. Lave (Eds.), Everyday cognition: Its development in social context (pp. 9-40). Cambridge, MA: Harvard University Press.

Skemp, R. R. (1979). Intelligence, learning, and action. Chichester, England: Wiley.

Sleeman, D. (1984). An attempt to understand students' understanding of algebra. Cognitive Science, 8, 387-412.

Smith, B. C. (1983). Reflection and semantics in a procedural language (MIT/LCS/TR-22). Cambridge, MA: Massachusetts Institute of Technology, Laboratory for Computer Science.

Thorndike, E. L. (1922). The psychology of arithmetic. New York: Macmillan.

VanLehn, K., & Brown, J. S. (1978). Planning nets: A representation for formalizing analogies and semantic models of procedural skills. In R. E. Snow, P. A. Federico, & W. Montague (Eds.), Aptitude, learning, and instruction. Volume 2: Cognitive process analyses of learning and problem solving (pp. 95-138). Hillsdale, NJ: Erlbaum.

Wertheimer, M. (1959). Productive thinking (enlarged ed.). New York: Harper. (Original work published 1945)

Winograd, T. (1972). Understanding natural language. Cognitive Psychology, 3, 1-191.

Winograd, T., & Flores, F. (1986). Understanding computers and cognition: A new foundation for design. Norwood, NJ: Ablex.

Chapter 13

# A THEORY OF MOTIVATION FOR COMPREHENSION AND ITS APPLICATION TO MATHEMATICS INSTRUCTION

Giyoo Hatano and Kayoko Inagaki

## 1. Why Do We Need Instructional Strategies for Enhancing Motivation for Comprehension

One of the major goals of education is the acquisition of a well-organized body of knowledge through comprehension. For this reason, it is essential for educational researchers to give close attention to students' motivation for comprehension and to teachers' strategies for enhancing it. Although motivation and comprehension have been studied extensively as discrete topics, motivation for comprehension and how to enhance it have been neglected in educational research, and no well-articulated theory of instructional strategy has yet been offered. In this paper, we will argue that the study of workable instructional strategies for enhancing motivation for comprehension should be given high priority. Then, we will present an outline of our theory of motivation for comprehension. Finally, we suggest some instructional strategies, derived from the theory, which may be used in mathematics classes.

Before turning to the main issues, we will define comprehension (or understanding, a term to be used interchangeably) as it is used in this paper. Since we are concerned with comprehension in relation to mathematical and scientific problem solving, the term comprehension might be defined as apprehending "the 'how' and 'why' of the connections observed and applied in action" (Piaget, 1978). In other words, to comprehend means to achieve insight or to find satisfactory explanations for the validity of a given rule or the success of a procedure. Whether a given set of explanations is satisfactory or not may vary from individual to individual. It depends not only on logico-mathematical validity of the explanations but also on how plausible and illuminating they are in an individual's phenomenological world. For example, knowing the mathematical derivation of a theorem does not necessarily guarantee insight.

When we refer to the process of achieving insight we will use another term, comprehension activity. Comprehension activity includes generating inferences, checking their plausibility, and coordinating pieces of old and new information to build an enriched and coherent representation, which will serve as the basis for insight. Motivation for comprehension is equivalent to motivation for directed, persistent comprehension activity.

To illustrate and clarify comprehension activity, we present a hypothetical example directed to a well-defined procedure for preparing fish, making sashimi of a bonito. We have intentionally chosen this non-mathematical/scientific example, because we want to stress that comprehension activity for insight may occur in our everyday life as well as in instruction. The recipe (from Fish and vegetable cooking, by NHK publishers, 1984), starting with a big cut of a bonito, requires us:

1. to roast its skin-covered surface quickly with strong heat;
2. to put the side into ice water, and cool it for five minutes;
3. to take it out of the water, and wipe it off;
4. and finally, to cut it into slices 1 cm thick. These slices are ready to eat with soy-sauce and seasonings.

People can follow the recipe without truly comprehending what they are doing and get delicious bonito sashimi. But why does this procedure (the recipe) work? Why are these steps necessary? Suppose that you are interested in questions like these, and that you are engaged in comprehension activity. If you can generate some inferences relying on your prior knowledge, you might test them. If you cannot, you have to proceed in a trial-and-error fashion, i.e., run the procedure with one or more critical steps removed or modified. For example, you might examine how the sashimi tastes without roasting, or when roasted with mild heat. You will soon find that, without roasting, the skin of the sashimi is too tough to swallow, even after chewing it for a couple of minutes. You will also learn that "quickly with strong heat" is critical, because otherwise you have well-done bonito steak, instead of sashimi. From this experience, you can make an inference as to the next step: the ice water is needed to cool the roasted fish very quickly. You can confirm this inference by putting the bonito in water without ice or by putting it in a refrigerator.

You may be tempted to go on. You may run more experiments with varying parameters, consult cookbooks or books on ichthyology, question your family or friends, relate the set of observed facts to similar experiences, e.g., making sashimi of other fish. If you comprehend the recipe, you can modify it flexibly when you have to meet a different set of constraints, e.g., when you have no ice or no strong heat. To achieve the comprehension, you have to engage in prolonged comprehension activity, spending much time, effort, and cost. We do not claim that every comprehension activity is like this. However, it is almost always true that comprehension takes time.

Now let us return to the main issue. Why do we need instructional strategies for enhancing motivation for comprehension? Our answer is divided into two parts. First, we claim that these strategies cannot be derived from more general theories of motivation. Second, we claim that, without such strategies, it is highly unlikely that a majority of students will engage in persistent comprehension activity directed to a target rule or procedure. We will elaborate these assertions in turn.

Studies on "motivation in education," despite the progress of the past 15 years, have either ignored or paid little attention to motivation for comprehension. Many of them, having historical roots in the theory of achievement motivation, have developed a cognitive-attributional approach to motivation (Ames & Ames, 1984; Levine & Wang, 1983; Paris, Olson, & Stevenson, 1983). The studies revealed that causal attribution for success or failure influences students' motivation for achievement and thus their performances (e.g., Weiner, 1980). Students' attribution of success/failure--and conception of ability/effort underlying the attribution--constitutes a significant part of metacognition which, as we shall see later, plays an important mediating role in determining whether they engage in comprehension activity.

However, these studies have been concerned primarily not with comprehension but with achievement or problem solving competence. Dependent measures were usually based on the number of correctly solved problems. Although correct solutions may reflect comprehension, the distinction between competently solving problems using a certain procedure and understanding that procedure is critical, especially in mathematics and science instruction. To be a competent problem-solver on standard achievement tests, one needs to know how and when to apply a given procedure, but it is not necessary to demonstrate comprehension.

We can solve a great number of mathematical problems using a target procedure at the right time, without achieving insight, without enjoying "the pleasure of understanding" (Piaget, in Tanner & Inhelder, 1960). Very few of us can explain why a given mathematical procedure works, though we believe it valid and can apply it efficiently. Consider, for example, the Euclidean algorithm. It is often presented in high school algebra but its proof, which involves mathematical induction, is omitted. Students believe it valid because with little thought or effort they produce the greatest common divisor of two integers. Later, their lack of insight becomes a serious handicap when novel problems are posed for solution. For example, if asked to find integral solutions for $y$ and $z$ given that $21z - 15y = 9$, they would fail to recognize the relevance of the algorithm. This is similar to the fact that we can make delicious bonito sashimi promptly by just following the recipe, without understanding why the steps in the recipe are needed. Lack of insight becomes a serious deficit only when unusual, novel problems are posed. After having applied a procedure many times successfully, we tend to lose interest in knowing why it works. Therefore, strong motivation for achievement does not guarantee strong motivation for comprehension. Those procedures enhancing the former may not be effective for the latter.

The problem is compounded by the fact that, using Nicholls' (1983) distinction, most attributional studies have dealt with the extrinsic and ego-involvement aspects of motivation for achievement. In other words, they have dealt with attaining high achievement as a means to an end, that of external rewards, or of looking smart, or avoiding looking stupid. Although some comprehension is necessary for high achievement, the subjects' activity in these studies was not directed to knowing or understanding. Only a few studies have pursued task-involvement or intrinsic aspects of motivation, which are most important for comprehension.

A group of studies on the relationships between extrinsic and intrinsic motivations, another major stream of recent research on motivation, have revealed that extrinsic rewards tend to undermine intrinsic motivation (Deci, 1975; Lepper, 1983; Lepper & Greene, 1978; Maehr, 1976). This finding is also relevant to motivation for comprehension, since, as we shall see, it is possible that external rewards may also undermine intrinsic motivation for comprehension. However, these studies have not paid due attention to the intrinsic pleasure of understanding, nor suggested strategies for enhancing intrinsic interest in comprehension.

Now we will move to the second part of the argument supporting the need for instructional strategies to enhance motivation for comprehension. Cognitively oriented instructional psychology has been interested in the process of comprehension and strategies for presenting stimuli to enhance it, but it has neglected motivation for comprehension. In Resnick's (1981) review of instructional psychology, for example, there is no reference to motivation for knowing or understanding. Cognitive researchers use four major reasons to justify their indifference to and neglect of motivational issues related to comprehension. First, it is claimed that comprehension is performed automatically; thus, no motivation is involved. However, this is not convincing if we reflect on the example of bonito sashimi. Unlike the perceptual recognition of an object or the processing of a sentence, which is also sometimes called comprehension, comprehension as insight, or finding satisfactory explanations, is far from automatic. It requires much time and a considerable measure of conscious effort.

Second, cognitive researchers sometimes claim that active human beings are always motivated to engage in comprehension, though the process is not automatic. Therefore, no instructional strategies are needed to enhance the motivation. We believe that this claim is based on a misunderstanding of the "zeitgeist" of contemporary cognitive psychology, namely the assertion that human beings are active agents of information processing. We would agree that "active human beings" almost always try to comprehend, and comprehension gives them intrinsic satisfaction, irrespective of any accompanying external rewards. However, this does not mean that they always engage in persistent comprehension activity directed to a target rule or procedure. In fact, while many Japanese do know how to make bonito sashimi, very few comprehend how or why the recipe works. Comprehension activity may cease without producing any satisfactory explanations. Since there are so many targets to which one's comprehension activity can be directed, it has to be selective. In other words, although the zeitgeist may enable us to ignore the initiation and reinforcement questions, we are compelled to attend to the issues of persistence and choice.

There may be a practical basis for this misunderstanding: Subjects in the laboratory experiments, often college students, try hard to comprehend as soon as they are instructed to do so. However, as you will notice, students in the usual classroom are not always motivated to comprehend the target.

Third, some cognitive researchers believe that motivation is beyond the teacher's control. Therefore, although they are willing to accept the fact that students' motivation for comprehension makes a difference, it is impractical to consider it. We believe, however, that it is possible to formulate instructional strategies that are likely to enhance students' intrinsic motivation to comprehend the particular target, assuming that the students have acquired a specified set of prior knowledge. We will describe some of those strategies in more detail in the last section of this paper.

Finally, other cognitive researchers assert that efforts to enhance students' motivation are not very rewarding because it is doubtful whether enhanced motivation leads to "correct" comprehension. We believe, on the contrary, that through increasing motivation a teacher can indirectly enhance the likelihood of students' correct comprehension. Unlike students' acquisition of procedures and memory of rules, their comprehension is not amenable to a teacher's direct control, since comprehension means to find "satisfactory" explanations, which may differ from individual to individual. However, we can assume that strong motivation for comprehension usually leads to deeper comprehension; many significant inferences are generated and relevant pieces of information interrelated. Strong motivation is also likely to lead students to "correct" solutions and explanations because it makes them engage in more persistent and meticulous comprehension activity. They will check carefully whether generated inferences are harmonious with the given set of information, thus eliminating erroneous explanations. If their comprehension activity is still not sufficient for excluding all of the incorrect explanations, the teacher may intervene by giving additional information or drawing their attention to relevant information that refutes their conclusions.

In summary, instructional procedures for intrinsically motivating students to comprehend cannot be derived from any available achievement-oriented theories, and none of the arguments by cognitive researchers can justify the neglect of motivational issues related to knowing and understanding. We need instructional strategies specifically for enhancing motivation for comprehension, and educational researchers must seriously pursue this task.

## 2. Outline of Cognitive Berlynean Theory

In this section, we summarize our theory of motivation for comprehension (Inagaki & Hatano, 1986). Within the framework of recent cognitive instructional psychology, it elaborates and extends Berlyne's theory of epistemic behavior and may be called a cognitive Berlynean theory.

When seeking a groundwork on which to construct a tenable theory of motivation for comprehension, it was necessary to return to Berlyne's work of the early 1960s. Berlyne (1960, 1963, 1965a, 1965b) conceptualized the motivation inherent in epistemic behavior and suggested a number of possible instructional strategies to motivate students to acquire knowledge. Though his theory does not deal with motivation for comprehension itself, it has at least three properties indispensable to any theory of motivation for comprehension: (1) it focused on intrinsic motivation for knowing; (2) it systematically described when (or by what stimuli) such motivation is aroused, and what kind of behaviors the motivation induces; and (3) it had a prescriptive component, suggesting how we can motivate students. Recently, Malone (1981), who was interested in taking advantage of the attractiveness of computer games for educational settings, tried to conceptualize intrinsically motivating instruction relying in part on Berlyne's theory. However, he seemed to concern himself much more with the characteristics that make instruction enjoyable than with characteristics that would motivate students to deeply comprehend the target.

### Summarizing and Restating Berlyne's Theory

We will first demonstrate that Berlyne's "motivation of epistemic behavior" implicitly included "motivation for comprehension." We will then incorporate the results of recent research in order to update the theory.

Berlyne (1963) stated,

> the epistemic behavior refers to behavior whose function is to equip the organism with knowledge. . . . Epistemic behavior can be divided into three categories, namely, epistemic observation, which includes the experimental and other observational techniques of science, consultation, which includes asking other people questions or consulting reference books, and directed thinking. (p. 322)

Directed thinking is "thinking whose function is to convey us to solutions of problems" (1965a, p. 19). It should be noted that Berlyne defined critical terms like epistemic behavior and directed thinking in terms not of processes but of functions. Thus, if comprehension is regarded as achieving satisfactory explanations to the "how" and "why," then the corresponding comprehension activity is a case of epistemic behavior, more specifically, of directed thinking. Berlyne's notion of knowledge acquisition by directed thinking is very similar to what we now call "acquisition of an organized body of knowledge through comprehension activity."

According to Berlyne (1963), epistemic behavior is initiated by a specific dissatisfaction called epistemic curiosity, which is produced by conceptual conflict, and the behavior is reinforced by the reduction of epistemic curiosity, that is by relief of that conceptual conflict. Since comprehension activity is a form of epistemic behavior, we assume that it is initiated and maintained toward a specific object by strong epistemic curiosity. We add the qualifier strong because comprehension requires much time and effort. However, we do not agree with Berlyne that epistemic curiosity is a kind of discomfort drive state.

By conceptual conflict Berlyne means "conflict between incompatible symbolic response patterns, that is, beliefs, attitudes, thoughts, ideas" (1965a, p. 255). He distinguished several types of conceptual conflict--doubt, perplexity, contradiction, conceptual incongruity, confusion, and irrelevance (1965a)--and added surprise to the list when he discussed the use of conceptual conflict in educational settings (1965b).

Let us restate those constructs. First, in cognitive terms, conceptual conflict inducing strong epistemic curiosity is a state in which a person is aware that his/her comprehension is inadequate, but is within his/her reach. To avoid a behaviorist flavor, we call this state cognitive incongruity. This state motivates a person to pursue insight, to find satisfactory explanations to the target rule or procedure, by:

1. seeking further information from outside;
2. retrieving another piece of prior knowledge;
3. generating new inferences;
4. examining the compatibility of inferences more closely.

In other words, cognitive incongruity motivates a person to pursue insight through comprehension activity. Success in achieving adequate comprehension or insight would bring a stop to all this comprehension activity, and the comprehended rule or procedure is recalled and used subsequently on similar occasions more promptly and properly.

Second, Berlyne identified several types of conceptual conflict (our cognitive incongruity) that we group into two: One is the surprise type, which is induced when a person encounters an event or information that disconfirms his/her prediction based on a prior knowledge. He/she will be motivated to understand why and to seek new information by which the prior knowledge can be repaired. The other is the perplexity type, which is induced when a person is aware of equally plausible but competing ideas (predictions, assertions, explanations) related to the target object or procedure. In this case he/she seeks further information to choose one of the alternatives.

## Reformulating Berlyne's Theory

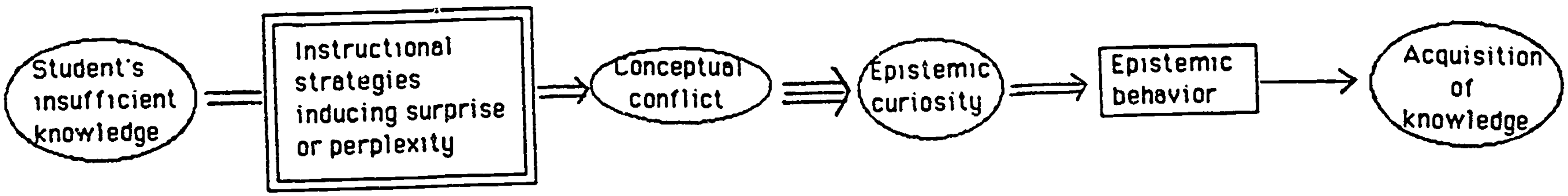
Now we propose some reformulations of Berlyne's theory. His theory about epistemic behavior was constructed in the early 1960s. Since then, as cognitive psychology has developed, a number of important ideas related to the issue of motivation for knowing and understanding have been proposed, and data have been collected based on them. To bring Berlyne's theory closer to an "ideal" theory, we incorporate four constructs. First, we append a third type of cognitive incongruity, discoordination, to the list producing strong epistemic curiosity. Second, we propose that, for cognitive incongruity to occur, students must recognize the inadequacy of their comprehension; in other words, they must be able to monitor their comprehension. Third, we believe that cognitive incongruity induces comprehension activity only when students realize the importance and possibility of comprehension about the target rule or procedure. Fourth, we argue that one is unlikely to engage in prolonged comprehension activity unless one is free from any urgent need, such as the need often produced by expecting material or other rewards. With these reformulations, the resultant theory, the cognitive Berlynean theory, can better describe stimulus conditions under which students possessing specified prior knowledge are always (or nearly always) motivated to engage themselves in comprehension activity, and without which, they are never (or almost never) motivated to do so. Figure 1 shows these reformulations schematically.

## Discoordination Induces Comprehension Activity

Since Berlyne's death, psychologists' views of human beings have changed. As Hunt (1963, 1965) aptly put it, human beings had been considered as idle under behaviorists' drive-reduction theory. Berlyne, in his attempt to "liberate" this drive-reduction theory, was not free from such a passive view of human beings.

Current cognitive psychology views human beings as active agents; it assumes that human beings actively seek pieces of information and try to organize them. A good example of this active information seeking occurs after a person has chosen a target as the object of his/her comprehension activity (e.g., Clement, in press; Collins, Brown, & Larkin, 1980; Hatano & Inagaki, 1983). We do not think the subjects of these experiments were suffering from prolonged (aversive) curiosity, or from potential danger to their survival. They certainly felt satisfaction and tension reduction when they had understood the target, but, we believe, they had enjoyed the process of performing the comprehension activity as well.

Berlyne's Theory




Cognitive Berlynean Theory

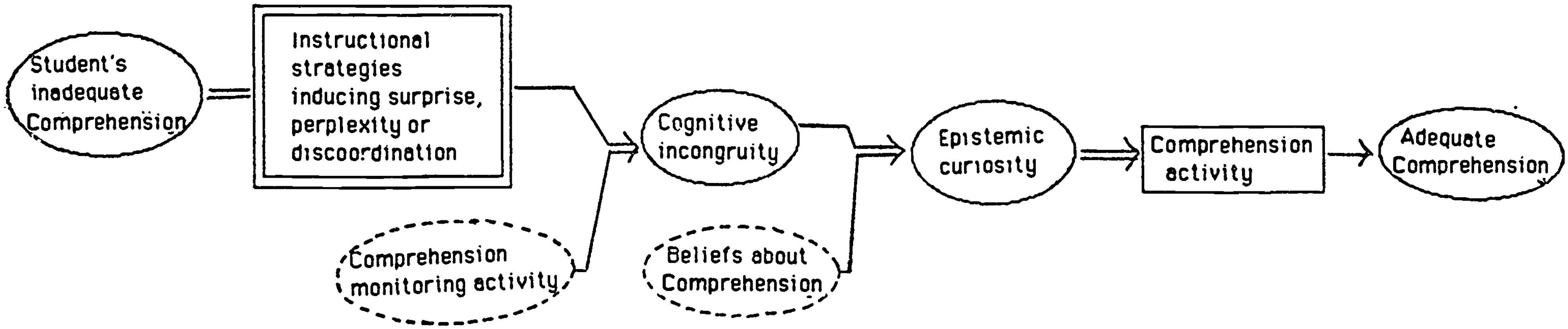





Figure 1. A Schematic Comparison between Berlyne's Theory and Cognitive Berlynean Theory

This change in perspective prompts our first reformulation of Berlyne's theory. People may try hard to comprehend without the incentive of inconsistency or incompatibility. Thus we propose that there is a third type of cognitive incongruity in addition to surprise and perplexity, namely, discoordination. This last type of cognitive incongruity is the awareness of a lack of coordination among the pieces of knowledge involved. In other words, it is the recognition that, although pieces of knowledge about the target are available, they are not well connected, or that other pieces of related information cannot be generated by transforming the existing ones. More specifically, people may be aware of the inadequacy of their comprehension in four conditions:

1. they are not yet certain whether two pieces of information they know about the target are identical or not, contradictory or not;
2. they cannot apply a known principle to concrete situations;
3. they cannot justify each step of the procedure;
4. they have rich examples but cannot abstract a rule.

## The Role of Comprehension Monitoring

Berlyne (1965a, 1965b) described some tactics to arouse conceptual conflict. However, it has been found that these operations do not always work well. According to Berlyne, for example, presenting material containing information that contradicts prior knowledge should arouse surprise, but in practice this operation induces no conceptual conflict in some students. Using our terminology, when presented with information that purports to reveal inadequacy in their comprehension, some students may fail to recognize the inadequacy and thus feel no cognitive incongruity.

Recent research on comprehension monitoring, following the pioneering work by Markman (1977, 1979), has shown that younger children fail to perceive the insufficiency or inconsistency of a given message more often than do older children or adults, but another line of research on metacomprehension has revealed that even college students tend to have this "illusion of comprehension" (Glenberg & Epstein, 1985; Glenberg, Wilkinson, & Epstein, 1982; Maki & Berry, 1984). College students often believe that they have understood a given text, though in fact they have not, at least as assessed by a multiple-choice test. This suggests a more or less general tendency among human beings to fail to recognize the inadequacy of their own comprehension.

As indicated earlier, we believe that people must be selective in directing prolonged comprehension activity, not through idleness, but because the activity requires much time and effort. This need for selection may be operant in recognition of inadequacy of comprehension as well as in the decision to pursue more adequate comprehension. In one sense, the illusion of comprehension guards people from engaging in prolonged comprehension activity too often, or in too diverse domains.

A few implications for effective strategies of motivating for comprehension may be derived from the studies in comprehension monitoring. First, students can promptly recognize inadequacy of comprehension only in domains where they have acquired rich and well-structured knowledge, in their domains of expertise. Second, to induce cognitive incongruity in less well-structured domains, it is necessary to make the inadequacy of comprehension abundantly clear, for example by ensuring that students' predictions are specific and explicit before disconfirming information is given. Any concurrent cognitive activity, which may tax the resources of less experienced people, must be removed. Third, it is desirable to provide the opportunity for students to check their comprehension in the context of another activity. Requiring children to translate what they understand into action, for example, may induce cognitive incongruity that would otherwise be not induced. Dialogical interactions, such as discussion, controversy, and reciprocal teaching, in which knowledge or comprehension is to be shared, often provide appropriate contexts for children to perceive cognitive incongruity.

## The Role of Metacognitive Beliefs About Comprehension

Will people engage in prolonged comprehension activity whenever they experience cognitive incongruity? Certainly not. Selectivity in seeking adequate comprehension operates also after cognitive incongruity is induced. Berlyne (1965a) indicated the possibility that aroused conceptual conflict neither induces epistemic behavior nor thus leads one to acquire knowledge. He proposed that "suppression" relieves conceptual conflict and thereby also precludes epistemic behaviour. We offer a more target-specific explanation: comprehension activity is induced or inhibited depending on metacognitive beliefs about comprehension of the target.

Two aspects of metacognition play an important part here. One is the belief about one's own capability of comprehending a specific target or of comprehending in general. If students have confidence in their ability to understand, they are likely to pursue comprehension. They will not be inhibited, even by an apparent deadlock. If they are not confident, however, they may suppress the motivation to comprehend, even when they feel incongruity. Studies on learned helplessness and causal attribution of success-failure (e.g., Diener & Dweck, 1978) give indirect support for the importance of students' beliefs.

The second aspect of metacognition is belief about the importance of comprehension in general or the significance of comprehending a specific target. In other words, whether or not cognitive incongruity leads to comprehension activity depends, in part, on whether or not students believe that the target is worth comprehending. When subjects experience cognitive incongruity about a target which they value (because it is relevant to their lives), they are likely to engage in comprehension activity. On the other hand, when they feel cognitive incongruity about a target of little interest or value to them, they will be reluctant to exert the mental effort required for comprehension activity.

In summary, we assume that each individual has personal "domains of interest," in which they believe comprehension to be both valuable and attainable. When individuals experience cognitive incongruity, they are willing to engage in prolonged comprehension activity within, but not outside of, those domains.

This creates a serious problem for the teacher who is trying to motivate students to comprehend a target rule or procedure outside their domains of expertise/interest. In these circumstances, students are unlikely to recognize the inadequacy of their comprehension, unlikely to engage in comprehension activity even when incongruity is aroused and, as a consequence, unlikely to acquire knowledge through comprehension. This vicious cognitive cycle can be broken only by introducing other activities, social-interactional ones in most cases. Miyake (1986), for instance, effectively demonstrated that dialogical interaction motivates people to engage in prolonged comprehension activity.

## Extrinsic Reward Reduces Motivation for Comprehension

Teachers' conventional methods of motivating students, such as grades or rewards, are based on extrinsic motivation. What effects do such extrinsic motivational methods have on epistemic behavior or comprehension activity? Berlyne (1965a) pointed out the differences between learning based on conceptual conflict and learning relying on external reinforcement but did not clarify further the relationship between extrinsic motivation and intrinsic motivation. This relationship has been conceptualized much more satisfactorily since Berlyne's death.

Studies on the so-called undermining effects of extrinsic rewards have shown that promised and/or given rewards deteriorate both the quality of performance in the task and intrinsic interest (Lepper, 1983; Lepper & Green, 1978). This suggests, indirectly, the possibility that extrinsic rewards inhibit motivation for comprehension.

In her review of the literature, Inagaki (1980) maintained that the expectation of rewards changes the goal of ongoing cognitive activity from comprehension to obtaining the reward and thus prevents learners from achieving deep understanding. Inagaki also hinted that the expectation of external evaluation--a grade based on a test score or of the right answer to be provided immediately--may have similar effects of changing the goal. Activities pursuing external rewards will not enhance motivation for comprehension.

## 3. Instructional Strategies for Enhancing Motivation for Comprehension

In this final section, relying on our cognitive Berlynean theory, we specify instructional strategies for inducing cognitive incongruity. To heighten motivation for comprehension, urgent extrinsic needs--external rewards, favorable evaluations, authorized right answers--should be removed from classroom learning. It is also necessary to help those students who are not confident in their ability to comprehend, or who do not value comprehension, to change their metacognitive beliefs about comprehension. However, we will proceed without further discussion of these issues because they have been in part pursued in general studies of motivation in education.

### Strategies for Inducing Cognitive Incongruity

Strategies for inducing cognitive incongruity may be grouped according to the types of incongruity that they are to induce. When pupils have acquired fairly rich and well-structured knowledge, which includes "erroneous" rules or procedures--called misconceptions, false mental models, bugs--we can arouse surprise by asking the pupils to make a prediction and then providing an event or information that clearly disconfirms it. For example, junior high school students usually believe that the quotient a/b must be a specific quantity. Therefore, when they are taught that 12/0 is undefined, they are surprised (Tokuda, 1975). This surprise may be strengthened by having had the pupils express a clear and specific prediction beforehand. Before the experiment is run or disconfirming information is given, students may also feel surprise by finding out in the course of peer interaction that there exists a whole range of plausible options differing from theirs.

We can induce perplexity easily by taking advantage of the fact that there are usually many different ideas generated among students in a classroom. A teacher need only tally pupils' responses to induce perplexity. For the quotient of 12/0, students' modal answers are 0 or $\infty$, but other answers are usually offered. Peer interaction, the presence of others expressing different ideas, is especially advantageous for amplifying perplexity, because the students have a chance for argumentation; it is hard to recognize as plausible those ideas that are merely read or encountered passively.

When students do not have rich and well-structured knowledge regarding the target, it is sometimes necessary to first teach a specific rule and to encourage them to apply this rule to a number of confirming cases. This approach will make students commit themselves to the given rule, because they are likely to appreciate its effectiveness. Subsequently, the students experience surprise when shown information that is dissonant with this newly acquired rule. When the teacher asks whether this rule works for another example that seems radically different from the confirming cases, or whether the rule always holds true, the students will have difficulty in deciding whether it applies; that is, they will feel doubt, a subtype of perplexity. Berlyne (1965a) reported that this type of procedure was successfully used by David L. Page to teach third-graders that the difference between the squares of two adjacent integers, $(n + 1)^2 - n^2$, is always an odd number.

Discoordination may be experienced by a student in the process of explaining why his/her views are reasonable when asked for clarification or when the views are directly challenged or disputed. Why is discoordination induced in these situations? First, in the process of trying to convince or teach other students, one has to verbalize, or make explicit that which is known only implicitly. One must examine one's own comprehension in detail and thus become aware of any inadequacies, thus far unnoticed, in the coordination among those pieces of knowledge. Second, since persuasion or teaching requires the orderly presentation of ideas, one has to better organize intra-individually what one knows. Third, effective argumentation or teaching must incorporate opposing ideas, in other words, coordinate different points of view inter-individually between proponents and opponents or between tutors and learners. Of course, it is practically impossible to coordinate all the pieces of information available at any given moment. Thus, in one sense, an "illusion of comprehension" is adaptive because it frees one from endless comprehension activity. One feels strong discoordination only when one struggles to coordinate.

## Peer Interaction Enhances Motivation for Comprehension

The above discussion suggests that peer interaction, or dialogical interaction in general, such as discussion, controversy, and reciprocal teaching, tends to induce persistent comprehension activity directed to the target. It creates and amplifies surprise and perplexity, produces discoordination, and relates the target to one's domains of expertise and interest. It also invites students to "commit" themselves to some ideas, by asking them to state their ideas to others, thereby placing the issue in question in their domains of interest. In addition, the social setting makes the enterprise of comprehension more meaningful. Unless extrinsic motivation is so strong that it supersedes motivation for comprehension, this social aspect will make comprehension activity more enduring.

Is it possible for teacher-pupil interaction to produce the same effect as peer interaction? If so, it will be more desirable, because the teacher wishes to maintain control. In principle, a teacher who has richer and better-organized knowledge about the target than any of the students can help them recognize the inadequacy of their comprehension by giving counterexamples, proposing plausible alternatives that students have not offered, or by asking questions to clarify the students' ideas. The Socratic method of teaching is a good example of such instructional strategies. Collins (1977), in his attempts to describe the Socratic method, listed 24 specific strategies teachers could use, which included a number probably effective for enhancing motivation for comprehension.

However, practically, teacher-pupil interaction as a means for enhancing motivation for comprehension has serious limitations. First, since students know that their teacher is more knowledgeable than they are, if the teacher is actively intervening, they will depend on the authorized "right" answer. This anticipation of the right answer must weaken the motivation, as mentioned in the preceding section. Second, even when the teacher tries to behave as one of the less knowledgeable students by asking questions rather than giving answers, it is almost impossible to completely eliminate artificiality. This inevitably reduces the value the students assign to the comprehension they ultimately achieve. Being a good Socratic teacher is at least as hard as functioning as a good organizer of peer interactions.

## A Concrete Example

How shall we organize peer interaction to enhance students' motivation for comprehension? Though teacher-pupil interaction has limited effectiveness in inducing cognitive incongruity, the teacher's role in enhancing motivation for comprehension by organizing peer interaction is critically important.

Deriving theory-based instructional strategies, in other words translating a theory into practice, is often not an easy task. Fortunately, in this case, we have model system of instruction that has developed independently but is harmonious with our theory. This is a Japanese science-education method called Hypothesis-Experiment-Instruction" (Itakura, 1962), originally devised by Itakura, used in science classes from elementary to high school. A few have applied the same instructional procedures to mathematics and to limited areas of social studies. From our perspective, Hypothesis-Experiment-Instruction is effective in enhancing motivation for comprehension because it maximally utilizes classroom discussion and arranges a series of problems to induce all three types of incongruity.

The procedure is as follows: (1) Pupils are presented with a question with three or four alternative answers. (2) They are asked to choose one by themselves. (3) Pupils' responses, counted by a show of hands, are tabulated on the blackboard. (4) They are encouraged to explain and discuss their choices with one another. (5) They are asked to choose an alternative once again. They may change their original choice. (6) Pupils test their predictions by observing an experiment or reading a given passage.

The response alternatives should represent a plausible idea embodying a common bug or misconception held by pupils as well as the correct response. For example, the first lesson on "buoyancy" begins with the following question, alternative answers to which are all plausible and are usually chosen by at least several students (Shoji, 1975). "Suppose that you have a clay ball on the end of a spring. You hold the other end of the spring and put half of the clay ball into water. Will the spring (a) become shorter, (b) become longer, or (c) retain its length?" Thus the right answer, e.g., (a) in the above example, often contradicts predictions of a majority of pupils at the beginning part of a topic. It is also emphasized that pupils can clearly confirm or disconfirm their predictions by observing an experiment or consulting a reference book.

If you visit a classroom in which Hypothesis-Experiment-Instruction is implemented successfully, you will be impressed by lively discussions in a large group of 40-45 students. You will recognize that the teacher, after presenting a problem, is a chairperson, who tries to stay as neutral as possible during students' discussion. Several students may express their opinions often, but a majority of them are vicariously participating in the discussion, nodding or shaking their heads, or making just brief remarks. When asked, most of them reply that they enjoy discussion and feel the method exciting.

We have done a number of studies examining the effectiveness of this method, paying special attention to its effect on motivation for comprehension (Inagaki, 1986; Inagaki & Hatano, 1968, 1977). Materials of instruction were taken from mathematics as well as from science. Each class was randomly divided into experimental and control groups. In the former, the above 6 steps were followed, while in the latter, steps 3, 4 and 5 were omitted. All the pupils were required individually during the instruction to answer a short test consisting of a few multiple-choice items and also a questionnaire about their interest. They were also given a test involving a number of comprehension or transfer items and asked about their reactions to the opinions expressed by other pupils after the instruction. In addition, the process of discussion in the experimental condition was audio-taped, and behaviors of some selected pupils were observed.

General findings were as follows: (a) Experimental subjects showed higher interest than the control subjects in testing their predictions or knowing explanations; that is they showed higher epistemic curiosity before step 6. (b) The experimental subjects offered adequate explanations of the observed fact or stated rule; that is they showed explicit understanding more often than the control group. (c) They could apply the rule or procedure more promptly and more properly to a variety of situations, in other words, showed better implicit understanding. (For the above distinction between explicit and implicit understanding, see Greeno, 1980.) (d) Epistemic curiosity and understanding were correlated even within the experimental or control group. (e) Cognitive changes among the experimental subjects occurred primarily after they tested their predictions. In other words, group discussion produced few conversions by itself but made the students more sensitive to the feedback in step 6.

While our theory is more or less universal, its application must be culture-bound. The instructional strategies described are based on several assumptions. Enhancing motivation for comprehension through peer interaction presupposes that each student is attentive to remarks made by others and tries to incorporate them into his or her cognitive structure; that is, he or she listens well to peers. Also, discussion in a large group of 40-45 pupils, with the teacher as chair, is possible only when most students behave well. Therefore, we do not suggest the application of ready-made instructional strategies to other social-cultural settings. Further studies will enable us to specify effective strategies that motivate students to engage in comprehension activity in a variety of mathematics classes.

## References

Ames, R. E., & Ames C. (Eds.) (1984). Research on motivation in education (Vol. 1). Academic Press.

Berlyne, D. E. (1960). Conflict, arousal, and curiosity. New York: McGraw-Hill.

Berlyne, D. E. (1963). Motivational problems raised by exploratory and epistemic behavior. In S. Koch (Ed.), Psychology: A study of a science (Vol. 5, pp. 284-364). New York: McGraw-Hill.

Berlyne, D. E. (1965a). Structure and direction in thinking. New York: John Wiley.

Berlyne, D. E. (1965b). Curiosity and education. In J. D. Krumboltz (Ed.), Learning and the educational process (pp. 67-89). Chicago: Rand McNally.

Clement, J. (in press). Observed methods for generating analogies in scientific problem solving. Cognitive Science.

Collins, A. (1977). Processes in acquiring knowledge. In R. C. Anderson, R. J. Spiro, & W. E. Montagne (Eds.), Schooling and the acquisition of knowledge (pp. 339-363). Hillsdale, NJ: Erlbaum.

Collins, A., Brown, J. S., & Larkin, K. M. (1980). Inference in text understanding. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension (pp. 383-407). Hillsdale, NJ: Erlbaum.

Deci, E. L. (1975). Intrinsic motivation. New York: Plenum.

Diener, C. I., & Dweck, C. S. (1978). An analysis of learned helplessness: Continuous changes in performance, strategy and achievement cognitions following failure. Journal of Personality and Social Psychology, 36, 451-462.

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 702-718.

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. Memory and Cognition, 10, 597-602.

Greeno, J. G. (1980). Forms of understanding in mathematical problem solving. Paper presented at the 22nd International Congress of Psychology, Leipzig.

Hatano, G., & Inagaki, K. (1983, September). Rikaieno naihatutekidoukizuke [Intrinsic motivation for understanding]. Paper presented at the meeting for Japanese Educational Psychological Association, Kumamoto.

Hunt, J. McV. (1963). Motivation inherent in information processing and action. In O. J. Harvey (Ed.), Motivation and social interaction: Cognitive determinants (pp. 35-94). New York: Ronald Press.

Hunt, J. McV. (1965). Intrinsic motivation and its role in psychological development. In D. Levine (Ed.), Nebraska symposium on motivation (pp. 189-282). Lincoln: University of Nebraska Press.

Inagaki, K. (1980). Effects of external reinforcement on intrinsic motivation. Japanese Psychological Review, 23, 121-132. (In Japanese with English summary).

Inagaki, K. (1986). Peer interaction and motivation for understanding. Unpublished manuscript available from Kayoko Inagaki, School of Education, Chiba University, Chiba, Japan.

Inagaki, K., & Hatano, G. (1968). Motivational influences on epistemic observation. Japanese Journal of Educational Psychology, 16, 191-202. (In Japanese with English summary).

Inagaki, K., & Hatano, G. (1977). Amplification of cognitive motivation and its effects on epistemic observation. American Educational Research Journal, 14, 485-491.

Inagaki, K., & Hatano, G. (1986). Motivation for understanding in the classroom: A cognitive Berlynean theory. Unpublished manuscript available from Kayoko Inagaki, School of Education, Chiba University, Chiba, Japan.

Itakura, K. (1962). Instruction and learning of concept force in static based on Kasetsu-Jikken-Jugyo (Hypothesis-Experiment-Instruction): A new method of science teaching. Bulletin of National Institute for Educational Research, 52. (In Japanese).

Lepper, M. R. (1983). Extrinsic reward and intrinsic motivation: Implications for the classroom. In J. M. Levine & M. C. Wang (Eds.), Teacher and student perception: Implications for learning (pp. 281-317). Hillsdale, NJ: Erlbaum.

Lepper, M. R., & Greene, D. (Eds.). (1978). The hidden cost of reward. Hillsdale, NJ: Erlbaum.

Levine, J. M., & Wang, M. C. (Eds.). (1983). Teacher and student perception: Implications for learning. Hillsdale, NJ: Erlbaum.

Maehr, M. L. (1976). Continuing motivation: An analysis of a seldom considered educational outcome. Review of Educational Research, 46, 443-462.

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. Journal of Experimental Psychology: Learning, Memory and Cognition, 10, 663-679.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. Cognitive Science, 4, 333-369.

Markman, E. M. (1977). Realizing that you don't understand: A preliminary investigation. Child Development, 48, 986-992.

Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. Child Development, 50, 643-655.

Miyake, N. (1986). Constructive interaction and the iterative process of understanding. Cognitive Science, 10, 151-177.

Nicholls, J. G. (1983). Conception of ability and achievement motivation: A theory and its implications for education. In S. G. Paris, G. M. Olson, & H. W. Stevenson (Eds.), Learning and motivation in the classroom (pp. 211-237). Hillsdale, NJ: Erlbaum.

Paris, C. G., Olson, G. M., & Stevenson, H. W. (Eds.). (1983). Learning and motivation in the classroom. Hillsdale, NJ: Erbaum.

Piaget, J. (1978). Success and understanding. London: Routledge & Kegan Paul.

Resnick, L. B. (1981). Instructional psychology. Annual Review of Psychology, 32, 659-704.

Shoji, K. (1975). Kasetsu-Jikken-Jugyo no kyoikukiroku [The summary protocol of Hypotehsis-Experiment-Instruction]. Tokyo: Seijo Elementary School.

Tanner, J. M., & Inhelder, B. (Eds.). (1960). Discussions on child development, 4. London: Tavistock.

Tokuda, S. (1975). Division by 0: [12÷0] is equal to 0, isn't it? Mathematics Classroom, 273, 77-84. [In Japanese]

Weiner, B. (1980). Human motivation. New York: Holt.

## Chapter 14

## THE INTUITIVE DIMENSION OF MATHEMATICAL REASONING

Efraim Fischbein

In any mathematical activity one may identify three basic components:

1. The formal aspect is expressed in the strictly deductive, logical structure of mathematics: axioms, definitions, theorems, proofs.

2. The algorithmic aspect, which includes standardized mathematical operations, formulae, and solving strategies, is the instrumental component of any mathematical activity.

3. The intuitive dimension refers chiefly to the dynamics of the subjective acceptance of a mathematical idea.

Let us consider, for example, elementary arithmetical operations. One must define what one means by addition, subtraction, multiplication, and division and the relations among them. One must define the laws of associativity, distributivity, and commutativity and how they apply to elementary arithmetical operations; one identifies the group properties of various sets of numbers under these operations. Such components comprise the formal aspect of mathematical activity.

At the algorithmic level, we are interested in the techniques of mathematical operations as applied to various classes of mathematical entities. Students also learn standard strategies for solving standard problems with the help of these operations (such as the famous "rule of three").

A third aspect of mathematical activity which is very often overlooked in the instructional process is the intuitive dimension. In learning mathematics, one does not deal exclusively with the logical structure of mathematical truths. One must also assimilate and integrate such truths into the fundamental schemas of mental behavior in order to apply them in problem solving. As a matter of fact, one tends to confer automatically on the various types of mathematical ideas a certain subjective interpretation--which makes these ideas directly accessible and acceptable to the individual. In other words, one confers on the respective concept or statement an intuitive meaning. Even after an individual has acquired sufficient training to consider a certain topic in a general and abstract rigorous manner, he remains dependent on primary intuitive interpretations. For example, although one knows that a point, a line, and a surface are "pure" concepts, i.e. abstract, ideal mental entities, one tends to attach to them figural, intuitive representations. This tendency may influence reasoning even when the individual is aware of the purely abstract nature of the respective entities.

The formal, the algorithmic, and the intuitive aspects of mathematical reasoning describe neither developmental levels nor learning stages, though their description may be helpful in explaining some developmental phenomena or in devising teaching programs. In our opinion, every genuine mathematical activity--no matter the age of the individual or the complexity of the mathematical concepts involved--includes all three aspects. Any attempt to reduce a child's mathematical activity to mere intuitive processes or a university student's reasoning to pure formal inferences will have a negative result.

This paper focuses on the intuitive dimension of mathematical activity.

## The Concept of Intuition

The concept of intuition has a long history. Philosophers, mathematicians, other scientists, and pedagogical specialists have all used it, and a variety of meanings, some contradictory, have been attached to the term. According to Descartes (1967) and Spinoza (1967), intuition is the initial source and the ultimate reliable guarantee of certitude. In Bergson's view (1954), intuition is the key to understanding the essence of life phenomena, of duration, of motion. Modern science philosophers, like Hahn (1956) and Bunge (1962), consider intuition a primitive, unreliable form of knowledge.

Although various definitions have been proposed, some features are commonly accepted. Intuition is always described as <u>immediate knowledge</u>, as a cognition which is accepted directly as self-evident, with a feeling of intrinsic certitude, and without any need for verification or proof.

Mathematicians and other scientists use the term intuition in two different but related ways: (a) as similar to the moment of "illumination" in a problem-solving process (the initial, global grasp of a possible solution to a problem); or (b) when referring to a statement which may be accepted as self-evident (e.g., the whole is bigger than each of its parts). Both meanings are fundamentally important for mathematics education.

The "illumination" meaning refers to the student's approach to problem solving. Shall we teach students algorithmic techniques exclusively, to enable them to identify classes of problems and to solve them? Or shall we encourage students to guess a solution before having firm grounds for accepting it? Bruner raises the question: "Should students be encouraged to guess, in the interest of learning eventually how to make intelligent conjectures?

Possibly there are kinds of situations where guessing is desirable and where it may facilitate the development of intuitive thinking to some reasonable degree. There may, indeed, be a kind of guessing that requires careful cultivation" (p. 64).

The second meaning refers to the way in which the student represents and accepts a certain concept or statement. The learning of a formal definition or a formal proof does not determine absolutely the manner in which a student understands and uses it. Obstacles to understanding, misconceptions, and inadequate solving strategies are very often the effect of intuitive influences.

Let us consider in more detail these two categories of intuition.

## Anticipatory Intuitions

Describing the problem-solving process, Hadamard (1949), following the autobiographical accounts of Poincaré (1914)--describes four stages: preparation, incubation, illumination, and verification. The moment of illumination corresponds to what we have called anticipatory intuition.

Much problem-solving solution activity is unconscious, but the unconscious segment is preceded by a preparatory stage which is conscious and purposeful. The preparatory stage refers to the activity of learning the problem, to analyzing the concepts and relationships involved. During the preparatory stage, we try to become aware of the implications and consequences of available information. We try to organize this data and to grasp a new structure to lead us to the solution. Hadamard (1949) observed that very often the path to the correct solution is blocked by choosing and following rigidly a too-narrow path: ". . . in both domains the mathematical and the experimental, the fact of not sufficiently 'thinking aside' is a most ordinary cause of failure. . . ." (p. 49).

To succeed, one must maintain a strict balance between following a chosen investigative line and keeping the mind open to all available options. The delicate equilibrium between openness and flexibility on the one hand, and stability and consistency on the other, represents what may be the most essential ability of a good problem solver. Excessive rigidity or excessive divergency during problem solving are insurmountable obstacles.

The incubation stage is largely an unconscious segment of the problem-solving endeavor. The individual, tired from his effort, changes his line of thought or rests. Between this moment and the moment of illumination-the initial grasp of the solution--something must occur because there is often a fundamental difference between the representation of the problem before and after interruption of the conscious activity; the solution seems to appear suddenly, as if the mind has continued to work in the respective interval. What kind of work is this? Is this a blind, automatic work which produces many combinations (via associations)? According to Poincaré, (1914) this combinational production does not represent a characteristic aspect of the creative process; everybody, says Poincaré, may associate blindly everything with everything, and this would not lead to any solution. The unconscious mind's essential task is to select and retain those combinations which would be plausibly useful in attaining an acceptable solution. "To invent means to discern, to choose" (Poincaré, 1914, p. 48). But good choices follow certain criteria, and Poincaré (1914) mentions several:

> . . . the mathematical facts worthy of being studied are those which, by analogy with other facts, are able to lead us to the knowledge of a mathematical law in the same manner in which experimental facts lead us to the discovery of a physical law. They are those aspects which reveal surprising affinities between different facts known for very long but which have been considered unjustly alien one to the other. (p. 49)

According to Poincaré, the most fertile combinations are those which consist of elements borrowed from very distant, very different domains. But this alone is insufficient: the number of possible combinations may be so great that a lifetime would not be enough to examine them.

Poincaré offers a second criterion for successful selection of combinations useful to mathematical invention. This he calls the feeling of mathematical beauty, an awareness of the harmony of numbers and forms of mathematical elegance. "This is a genuine aesthetic feeling known to every true mathematician" (Poincaré, 1914, p. 57). Certainly, we may disagree. Mathematicians may occasionally enjoy the harmony and elegance of a solution or a proof, but one may assume that these qualities are not always apparent. What seems to be a fundamental component of mathematical invention, however, is what Poincaré (1914) has called the intuition of mathematical order, which helps us to guess the existence of harmonies and hidden relationships (p. 7).

The third stage in the problem-solving process is illumination, or what we have called anticipatory intuition. It is characterized by suddenness and by a feeling of certainty.

Let me recall a well-known autobiographical note of Poincaré (1913) which refers to the invention in mathematics:

> Just at this time I left Caen where I was then living, to go on a geologic excursion under the auspices of the school of mines. The changes of travel made me forget my mathematical work. Having reached Coutances we entered an omnibus to go some place or other. At the moment when I put my foot on the step the idea came to me, without anything in my former thoughts seeming to have paved the way for it, that the

transformations I had used to define the Fuchsian functions were identical with those of non-Euclidean geometry. I did not verify the idea; I should not have had time, as, upon taking my seat in the omnibus, I went on with a conversation already commenced, but I felt a perfect certainty. On my return to Caen, for conscience' sake I verified the result at my leisure.

Then I turned my attention to the study of some arithmetical questions apparently without much success and without a suspicion of any connection with my preceding researches. Disgusted with my failure, I went to spend a few days at the seaside, and thought of something else. One morning, walking on the bluff, the idea came to me, with just the same characteristics of brevity, suddenness and immediate certainty, that the arithmetic transformations of indeterminate ternary quadratic forms were identical with those of non-Euclidean geometry. (p. 388)

David Tall (1980) describes his complicated efforts to solve a problem related to infinitesimal quantities.

Reconsidering the theory as a whole it now all seems so inevitable. The ideas were not invented. They were discovered. Reading about the process of discovery written in these pages it is amazing to see the number of errors made and the false intuitions which had the ring of truth. Yet such was the intensity of excitement at the time that these temporary setbacks were insufficient to cause permanent blockages. . . .

. . . A classic description of "problem solving" involves conjectures which are then checked out. Here the researcher never felt that he made "conjectures." What he saw were "truths" evinced by strong resonances in his mind. Even though they often later proved to be false, at the time he felt much emotion vested in their truth. These were no cold, considered possibilities, they were intense, intuitive certainties. Yet at the same time its contact with them often seemed tenuous and transient; initially he had to write them down even though they might seem imperfect, before they vanished like ghosts in the night.

When such "truth" later proved false, it was rarely because of a coolly considered counter-example. That usually came later still after a period of mental unease already mentioned. In fact, the researcher, when in a state of mental excitement did not wish to check the detail at all, lest he lose the thread of the overall idea. It is remarkable the number of times that there were small errors which went unnoticed at the time but later produced unease then correction. (pp. 33-34)

<u>The Nature of Anticipatory Intuition</u>

Several aspects of problem solving may be deduced from these descriptions of its stages. Anticipatory intuition is generally preceded by conscious preparatory work and by a tacit period of incubation. One assumes that many combinations are tested but that these are not produced by mere fortuitous associations. Inductive attempts may often play an important role (Polya, 1954, pp. 3-11). Before a specific general statement is identified, one checks several cases, as in the empirical sciences. But sometimes a general statement first comes to mind, and one then checks several instances before a formal proof is found.

Analogy also plays a fundamental role in mathematical invention; through analogy one guesses the common mathematical structure of different classes of entities.

It appears that after the period of conscious preparatory work, the same research process continues in the "underground," at the tacit level. The difference is that, at the unconscious level, the production of associations, the identification of analogies, and the inductive-deductive reciprocal controls are activated at a much greater speed through automatic means.

The suddenness of the illumination moment becomes apparent. In fact, it represents, the final moment of a complex process, which starts with a feeling of satisfaction, of liberation, and of tension reduction. Suddenly, one has a global picture of the solution, a picture in which formerly disparate or even contradictory elements fit together in a new, unitary, coherent, self-consistent conception. Sometimes, these solution flashes have the appearance of a positive breakthrough "accompanied by pleasurable feelings" (Tall, 1980, p. 33).

A fundamental characteristic of anticipatory intuitions is that they appear to be absolutely certain. Although they represent no more than conjectures--before a complete verification is achieved--this is masked by the appearance of definitive truth (Poincaré, 1913 and Tall, 1980). The impression, according to Tall (1980), is that these ideas were not invented, but discovered.

Eugen Rusu (1962) a Rumanian mathematician and psychologist, also emphasized these aspects:

> . . . in the unstable and undecided atmosphere of the clouds before the storm, suddenly appears a lightning. In its brief light one grasps a convergent line of facts, a structure. The proof did not yet appear in all its details. What appeared is its guiding idea and the conviction that it indicates the right direction. (p. 22)

Polya (1954) also speaks about beliefs when referring to mathematical discovery.

> A scientist deserving the name endeavors to extract the most correct belief from a given experience and to gather the most appropriate experience in order to establish the correct belief regarding a given question. (p. 3)

A belief is different from a formal conviction based on a complete proof. A belief implies incompleteness in the arguments on which the conclusion is based. We need to believe when we cannot display a complete set of arguments; we must hide the gap with our conviction that the missing elements are there but not yet identified. Referring to beliefs in mathematical reasoning suggests that mathematical reasoning, like empirical investigation, uses heuristic means, (e.g., induction, analogies, the preliminary solution of a simpler problem) by which one jumps from a limited amount of empirically gathered arguments to a general idea. This jump is the moment of illumination, the moment of anticipatory intuition.

Emergence of a solution usually cannot be the result of gradual elaboration. The process' inductive, constructive nature implies a jump from finite (a limited number of examined facts) to infinite (the universal statement); one obtains a sudden belief that one is on the right path.

A belief implies intrinsic consistency, coherence, resistance to change, imperativeness. Certainly, the first global representation of a solution must be followed by analysis and verification for it is only then that empirical belief becomes a conviction based on formal, complete justification. But even, after the formal, analytical proof has been found, a global representation remains necessary.

> . . . any mathematical argument, however, complicated must appear to me as a unique thing. I do not feel that I have understood it as long as I do not succeed in grasping it in one global idea and unhappily . . . this often requires a more or less painful exertion of thought. (Hadamard, 1949, p. 65-66).

This is no longer anticipatory intuition (more syncretic than synthetic). The final, global representation, the conclusive intuition, provides the problem solver with a concentrated summary through which, on the basis of a subtle hierarchical organization, the main line of thought becomes salient and directly convincing.

As a matter of fact, the unconscious and the conscious components of the mental work are less distinct than might be deduced from Hadamard's (1949) description. Certainly there are periods of apparent relaxation during which tacit elaboration seems to continue (as evidenced by the apparently sudden discovery of a new idea that follows) but the stages of preparation, incubation, illumination, and verification do not occur in succession, one following another like acts in a play. The search activity is a mixture of associations, analogies, inductive attempts, guesses, hopes, beliefs, and efforts of verification, in which unconscious efforts occur simultaneously with conscious or semiconscious endeavors. There is, of course, a specific direction in the solution process, but this direction is far from consistent.

Tall's (1980) autobiographical note accentuates this observation:

> I recall that my mind was buzzing with ideas--I still wasn't clear about the archimedean bit, nor completeness. . . . However I spent an hour photocopying music, including "Virginia don't go too far" (a Gershwin song). I thought about the hyperreals of Robinson 'going too far' extending to many functions. (p. 29)

> Reconsidering the theory as a whole, it now all seems so inevitable. These ideas were not <u>invented</u>, they were <u>discovered</u>. Reading about the process of discovery written in these pages it is amazing to see the number of errors made and the false intuitions which had the ring of truth. Yet such was the intensity of excitement at the time that these temporary setbacks were insufficient to cause permanent blockages. . . . Before a major "illumination" takes place there are various moments of intuitive leaps characterized by the same feeling of belief that something essentially new has been grasped, that an important break-through has occurred. These are positive, apparently successful, breaks-through. But there are also negative break-through with a vague feeling of unease, with the conscious rationalization of the error sometimes taking days or even months to register. (p. 33)

These micro-intuitions, usually based on tacit elaborations expressed at the conscious level facilitate the relatively sudden formation of apparently coherent structures in which various elements seem to fit together in a unique, meaningful picture. Their essential role is to organize ideas, and to include in the constructive search activity moments of apparent success, of apparent clarity and certitude from which the endeavor may continue with confidence. These intuitive leaps have a double function: they synthesize in new, apparently coherent and intrinsically believable representations the progress already achieved and they increase the perspective of further efforts in terms of analytical control and new avenues of exploration.

## Affirmative Intuitions

A second category of intuition, inextricably related to the first, we have termed <u>affirmative intuition</u>. Affirmative intuitions are cognitions (representation, interpretations) which are directly acceptable to the individual as certain and self-evident. Such cognitions also are associated with a feeling of <u>belief</u> which generally exceeds data at hand. Some of these beliefs are considered correct by the scientific community, while others are viewed as false and must be rejected or corrected via instruction.

Intuitive affirmatory cognitions may refer to concepts, to relations, to inferences or to operations. In all these circumstances, we deal with meanings expressed in representations or interpretations directly acceptable to the individual as clear and self-consistent.

## Intuitive Meanings of Mathematical Concepts

A person's knowledge of a formal definition or description of a mathematical object does not generally eliminate the intuitive meaning attached to that concept, and it is this intuitive meaning that makes the respective cognition directly acceptable to the individual. Such acceptance is achieved by conferring upon the respective cognitions some globally representative, behaviorally meaningful interpretation.

Let us consider several examples. In formal mathematics, the concepts of point, straight line, surface--in fact, every geometrical concept--are abstractions. They are defined by axioms or by formally established definitions, and they do not exist as objective, material realities. But one tends automatically to confer upon them intuitive meanings. It is psychologically impossible to think of a point other than as a small spot, or of a line as anything but a fine ink stripe or a well stretched string.

David Hilbert (in Reid, 1970) observed:

> Who does not always use, along with the double inequality $a > b > c$, the picture of three points following one another on a straight line as the geometrical picture of the idea "between"? Who does not make use of drawings of segments and rectangles enclosed in one another when it is required to prove with perfect rigor a difficult theorem on the continuity of functions or the existence of points of condensation? Who could dispense with the figure of the triangle, the circle with its center or with the cross of the three perpendicular axes? Or would give up the representation of the vector field or the picture of a family of curves or surfaces with its envelope which plays so important a part in differential geometry, in the theory of differential equations, in the foundations of the calculus of variation and in other purely mathematical sciences? (p. 79)

These are not mere pictorial representations with no influence on the course of mathematical reasoning. In fact, these representations wield active influence, often beyond conscious control, on reasoning strategies and solution choice.

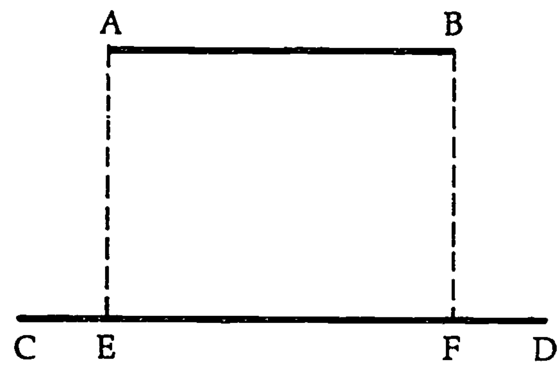



Figure 1

Comparing two sets of points--line segment AB and line segment CD--one intuitively arrives at a contradiction (Figure 1): if one agrees with Cantor that the two sets are equivalent, the intuitive reaction is that segment CD is longer.

If one draws perpendiculars AE and BF, it becomes intuitively obvious that one may establish a one-by-one correspondence between the sets of points of AB and EF. What about CE and FD? Such reasoning is correct when one considers pictorial representations rather than mathematical points. But let us attempt to eliminate the pictorial representation and to consider only the abstract mathematical notion of a point. It is very difficult to do so. How is it possible to compare quantitatively sets of 0-dimensional entities? There is a well-known proof (see Figure 2) that shows how a one-to-one correspondence may be established between two sets of points. Nevertheless, a feeling of uneasiness persists. The intuitive impression is that CD is somehow a stretched version of AB (a compromise between the original intuitive representation and the formal meaning attached to the respective concepts).

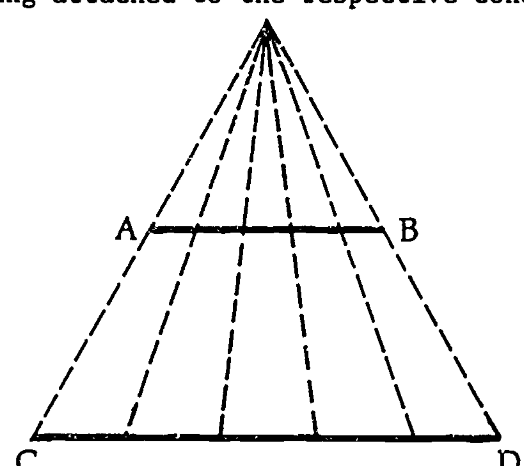



Figure 2

A child trying to overcome the contradiction affirmed: "Both segments contain the same number of points. In both there is an infinity of points. But the points in CD are bigger." The theory of infinity, as established by Cantor in the 19th century, has faced enormous difficulties because of intuitive obstacles.

A similar situation occurs with the number concept. It took hundreds of years for mathematicians to confer on the concept of negative number a formal mathematical status; negative number is intuitively a contradictory notion. The intuitive roots of the notion of number are to be found in the representation of equivalent sets. A number refers intuitively to the act of evaluating what has been called the cardinal of the set. This is an abstract notion--all equivalent sets have the same cardinal. This may be established behaviorally by establishing the bijections. Briefly speaking, the idea of number is intuitively meaningful, as long as it is related to sets of objects (or, at a higher level, to the notion of measure). But a negative number has no such practical interpretation. It is true that one may consider the absence of something, a certain deficit. One may claim, for example, that one has $5 less than is needed to buy a specific object. But to affirm that a number may absolutely represent a quantity less than nothing is something totally different. An existing quantity or a ratio between quantities (representable by numbers) which is less than nothing is intuitive nonsense--and so are operations with such numbers. What is the intuitive meaning of multiplying $(-2)\times(-5)$? For this reason, mathematicians, after discovering that one may obtain negative numbers when solving certain equations, have claimed that such curiosities are mere artifacts and must be eliminated.

The Scottish mathematician McLaurin (1698-1746) clearly understood the formal nature of mathematical entities: "It is not necessary to really describe the objects of our theories or that they should really exist. But it is essential that their relationships should be conceived clearly and deduced obviously" (in Glaeser, 1981, p. 318).

In spite of this, MacLaurin's Treatise of Algebra observed that an isolated quantity cannot be negative; that it may be so only by comparison. Rigorously speaking, a negative quantity is not less than nothing; it is not less real than a positive quantity when considered in an opposite sense (in Glaeser, 1981, p. 317). Such great mathematicians, as Descartes, Euler, Laplace, and Cauchy have struggled with these contradictions, and it was not until 1867 that German mathematician Hankel definitely solved the problem. He affirmed that negative numbers are not symbols of given realities but formal constructs, and that operations with them are governed only by formal considerations of consistency and not by practical meanings.

Today, students experience with less acuity the inner, intuitive contradictions inherent in the notion of negative numbers; they became accustomed to the concept during childhood. But the psychological difficulties reappear when dealing with the operations with negative numbers.

In order to understand the child's difficulties and successes when operating with fractions, one must know the underlying intuitive models the child has in mind. Behr and Wachsmuth (1982) describe such models. Some children use unit-fraction iteration: three-fifths is established by finding one-fifth and then performing an iterative behavior. While this procedure is sufficient for understanding the meaning of a fractional number, it does not independently support the more abstract idea of the equivalence of fractions such as the equivalence between 3/5 and 6/10 (Hunting, 1986).

Relational intuitions are expressed in self-evident, self-consistent statements: "The whole is bigger than each of its parts"; "Every number has a successor"; "Through a point outside a line, one may draw one parallel and only one to that line." Intuitively acceptable, they may become obstacles to theoretical developments that would contradict them. Indeed, the first statement above prevented mathematicians for many centuries from accepting the concept of actual infinity. If one accepts the concept of actual infinity, one must accept that a set may be equivalent to some of its proper subsets (e.g., the equivalence between the set of natural numbers and that of even numbers). In admitting the fifth postulate of Euclides as absolute and self-evident, the path to non-Euclidean geometries is closed. The development of mathematical ideas has been hindered for many centuries by such intuitively accepted statements.

Let us present another example. Carolyn Kieran, quoting various sources, has shown that for elementary and junior high school pupils the equality symbol represents an operator rather than a symbol of equivalence. Intuitively, the equality symbol represents for these subjects "a do-something signal." The sentence 3+5=8, for example, is interpreted as "3 and 5 make 8." Children rejected a sentence such as 4+5=3+6 because they expected an answer and not another problem to follow the equality symbol (Kieran, 1981, p. 319). The underlying intuitive model is that of an input-output operator, which prevents the child from interpreting the equality sign as a relation symbol, or as the symbol of equivalence with properties of symmetry, transitivity and reflexivity. When children were asked about the meaning of "3=3," a typical response was: "This could mean 6-3=3 or 7-4=3."

The problem of the intuitiveness of mathematical statements also raises important didactical problems:

1. If a statement is intuitively evident, students are reluctant to accept the necessity of a proof. The proof appears to be an unnecessary requirement which may cast doubt on the seriousness of mathematics itself. (We refer to such theorems as: "Two crossing lines determine pairs of equal opposite angles" or "If two sides of an isosceles triangle are equal, the opposite angles are also equal."

2. Self-evident statements are not absolute truths, and they may be replaced formally by statements which are counter-intuitive. For example, one may consider axiomatic systems in which Euclides' postulate is replaced by counter-intuitive axioms (e.g., through a point outside a line, one may draw an infinity of parallels to that line). Students will certainly be shocked by such statements, but the acceptance of counter-intuitive statements freely chosen or deductively proven (not leading to contradictions) is a sine-quo-non part of mathematics education.

3. Certain mathematical statements may not have a direct, intuitive meaning, but such a meaning may be created by using adequate intuitive models. The statement "if $A > B$, then $-A < -B$" has no intuitive meaning, but an intuitive model may easily be associated and understood using the number line.

4. There are many situations in which a statement has no intuitive meaning and in which such a meaning cannot be produced. The definition $a^0 = 1$, or the relation $a^{\frac{m}{n}} = \sqrt[m]{a^n}$, has no intuitive meaning, and no corresponding behavioral representation is possible. We do not recommend that effort be exerted to create artificial models for justifying such relations. The student must learn that mathematics is a formal, deductive body of knowledge in which statements are formally justified. Adequate, intuitive models may help in grasping the meaning of a concept or statement, but such intuitive means cannot always be provided.

## The Intuitive Meaning of Operations

Arithmetical operations are formally defined by axioms. Nevertheless, one tends to attach to these operations intuitive meanings which are commonly based on a corresponding practical operation. The sentence "5+3=8" intuitively means putting together two sets of elements But it may also be interpreted as counting from five on three additional elements (for instance, by using fingers). The sentence "7-3=4" may direct students to eliminate from a set of seven a set of three elements, or to build up from three to seven. If the text of the problem suggests intuitively a different operation than that which must actually be performed, the child encounters difficulties: "John has \$5. He needs \$8 to buy a pocket calculator. How much does he need?" The child must actually add, but the formal operation to be performed is subtraction.

The typical intuitive interpretation of multiplication is repeated addition, but this imposes several constraints. In formal mathematics multiplication is commutative. But if multiplication must solve a practical problem, the situation may be different. One must consider both the operator and the operand, "3 x 5" means "3 + 3 + 3 + 3 + 3" or "5 + 5 + 5." In the first interpretation, 5 is the operator and 3 is the operand. In the second interpretation, 3 is the operator and 5 is the operand. Intuitively this makes a great difference: one cannot intuitively conceive of taking a quantity 0.63 times, or 3/7 times, whereas one can easily conceive of $3 \times 0.63 = 0.63 + 0.63 + 0.63$, even if one is unable to perform the operation.

It has been shown that adults as well as children encounter difficulties when asked to solve a multiplication problem in which the operator is a decimal. A problem in which the same numbers intervene, but in which their role is changed, is solved more easily. Let us consider the following questions:

1. From 1 quintal of wheat, you get 0.75 quintal of flour. How much do you get from 15 quintals of wheat?

2. The volume of 1 quintal of gypsum is 15 $cm^3$. What is the volume of 0.75 quintal?

These are examples taken from research in Pisa, Italy, and all subjects were familiar with the term "quintal." In both problems, the solution is derived from multiplying 15 by 0.75. Grades five, seven and nine were investigated: grade five scored 79% and 57% correct on questions 1 and 2, respectively; grade seven scored 74% and 57% correct, respectively; grade nine scored 76% and 46% correct, respectively. When 0.75 was used as an operator, a dramatic deterioration of scores was observed.

A second constraint of the repeated addition model is that the product of multiplication must be larger than each of the factors. A difficulty appears if the operator is smaller than 1, since in this case the multiplication "makes smaller" (see Fischbein et al., 1985).

It also has been assumed that division is associated intuitively with two models: partitive division (sharing division) and quotative division (measurement division). The structure of the problem determines the model which is activated. In the first case, division is seen as an operation through which an object or a collection of objects is divided into equal fragments. In this interpretation, the dividend must be larger than the divisor, the divisor (the operator) must be a whole number, and the quotient must be smaller than the dividend (operand). Quotative division refers to a situation in which one seeks to determine how many times a given quantity is contained in a larger quantity. The only restriction is that the dividend must be larger than the divisor. As with multiplication, problems that violate these constraints create difficulties at various age levels (Fischbein, et al., 1985).

Thus, the intuitive meanings of mathematical operations play an important role in solution choice. Schools should develop in children an awareness of intuitive interpretations and an ability to understand and to control them.

Let us consider an additional example: A bottle of 0.75 litre of juice costs $2. What would be the price of 1 litre of juice?" The intuitive tendency is to choose multiplication as the solution operation; the idea that the correct solution is 2 ÷ 0.75 is not suggested intuitively. It is not the structure of the problem itself which creates the difficulty, but the relationship between the numerical data: the divisor is a decimal.

Let us consider the same problem with different data: "For $10, one can buy 5 litres of juice. What is the price of 1 litre?" It is intuitively clear that one must divide 10 by 5. It is not the presence of the decimal which is the main source of difficulty, but its function. In the research mentioned above, one finds the following problem: "Five friends bought together 0.75 kg. of chocolate. How much does each one get?" Even fifth graders solved the problem easily (85% correct answers).

These examples show that conflicts may arise between the formally correct solution and the tendencies supported by intuitive primitive models. We assume that in multiplication and division problems, the didactical solution is to develop proportional reasoning in pupils. According to Inhelder and Piaget (1958) proportion is one of the main operational schemas. As a matter of fact, each schema is only a potentiality. The elementary intuitive forms of proportional reasoning are present even in concrete-operational children. The challenge is to improve that intuitive background and to develop corresponding quantitative strategies. The famous "rule of three" may play an essential role in overcoming these intuitive difficulties through the use of formal strategies.

Let us return to the problem of the $2 0.75 litre of juice and the price of 1 litre. The proportionality is not intuitively evident; schema may help. One begins with a simpler problem in which the proportion _is_ evident:

6 litre --- $10

3 litre --- x dollars

The ratio between the quantities is equal to the ratio between their prices. If the quantity of juice is higher, the price is also proportionally higher. The problem becomes $6/3 = 10/x$. If the quantity of juice is one half, the price is also one half, and $x = 5$. On the other hand, the student must learn the transformations which would enable him to generalize the solution procedures.

I would like to emphasize that developing intuitive, active attitudes and teaching adequate algorithms are not opposite, didactical strategies. On the contrary, students must learn to merge the two approaches in a unitary, complex information-processing strategy on a strong, formal basis. As Vergnaud (1983) has shown, arithmetical operations must be assimilated not as isolated procedures but in the realm of complex conceptual systems.

Intuition and inferences. Some inferences seem to express intuitions, while others do not. From A = B, B = C, one concludes as a direct intuitive consequence, that A = C. Similarly, from A > B and B > C, one concludes that, evidently A > C. Such logical intuitions develop during the concrete-operational stage.

Conditional reasoning becomes more complicated. According to Inhelder and Piaget (1958), the formal-operational period is characterized by the emergence of hypothetical and combinatorial, propositional reasoning. This means that the logical structures of implication, conjunction, and disjunction should work to guarantee the adolescent's capacity to perform the logical operations requested by mathematical reasoning. In fact, things are very often not so. Even if one knows the truth table of the basic logical operations, one is not necessarily able to use these operations correctly in concrete problem-solving situations.

Knifong (1974), referring specifically to conditional reasoning, claims that children answer correctly only if the correct solution may be found by transduction, and this may occur with forms of reasonings called modus ponens and modus tollens. For example: "If this object is sugar, then it is sweet."

Modus ponens: "This object is sugar--then it is sweet.

Modus tollens: "This object is not sweet--then it is not sugar.

According to Knifong, children do not conclude correctly when denying the antecedent (the object is not sugar) or when affirming the consequent (this object is sweet). In the first case, the tendency is to deny the consequent; in the second, to affirm the antecedent. Knifong calls this relation non-directional justaposition.

In research by Galbraith (1981), pupils were asked about numbers for which the sum of the digits can be divided by 7. (Examples include 34 [3+4=7]; 185 [1+8+5=14].)

The question continues:

> If we make a list L of all such numbers which are less than 70, the start of it looks like this: 67, 16, 25, 34. Write down the next largest number on the list. Gary says: If you start with 7 and keep adding you always get a number on the list L.
> (1) Is Gary right?
> Brenda says: "Every number in the list can be found by adding 9 to the previous number. You start with 7.
> (2) Is Brenda right?" (Galbraith, 1981, pp. 9-10)

"If a rule goes for one, it will go for another"; "If it works for three, it should work." These pupils do not use implication as a logical, formal tool; their approach is an empirical one. Even after locating numbers in the list L (59 and 68) whose sum of the digits is divisible by 7, but which could not be obtained by adding successively 9, many subjects did not accept that Brenda's statement is thereby refuted (that is, if $p \rightarrow q$, then $\overline{q} \rightarrow \overline{p}$).

O'Brien et al. (1971) found that only 20% of grade 10 students were able to answer implication tests correctly. The authors concluded that this inability may explain student's failure in constructing a mathematical proof or checking its validity.

Logical schemas do not necessarily develop as actual capabilities in children and adolescents, and systematic training is requested. This training must be considered at all three levels of mathematical reasoning:

1. The formal level implies knowledge of truth tables of the most commonly used logical operations (implication, disjunction, conjunction).

2. The algorithmic level involves drill-and-practice activities referring to transformations of logical relations. Computer programs may be helpful at this level.

3. Intuitive understanding and use of logical operations may be developed by asking students to solve problems through global, direct evaluations before any systematic explicit control is performed. For example:

   If figure A is a square, its diagonals are equal. Let us suppose that one has proven that the diagonals of A are equal. Is figure A a square?

   An irrational number has an infinity of decimals. Number A has an infinity of decimals. Is it an irrational number?

In order to answer intuitively--and, not by resorting to the truth table of implication--one must imagine the situation and try to produce concrete instances which may confirm or deny the inverse implication ($q \rightarrow p$). It is essential to compare the solution deduced from the truth table with that which is produced by analyzing concrete examples. For example, the truth table indicates that the truth of q does not imply the truth of p: a number may have an infinity of decimals and, still be a rational number.

Intuition and proof. Are students aware of the profound distinction between an empirical proof and a formal (logical, mathematical) proof? Fischbein and Kedem (1982) have reported that for many high school students such a distinction is not clear cut. About 400 students in grades 10, 11, and 12 were presented with the following sentence: "Dan claims that the expression $n^3 - n$ is divisible by 6 for every n." The sentence was followed by a complete proof ($n^3 - n = (n - 1)\ n\ (n + 1)$. This expression is divisible by 2 and by 3, etc. About 81% of the subjects claimed that the proof is fully correct. The question was then asked: "Moshe claims that he has checked the number n = 2357 and has found that $2357^3 - 2357$ is not divisible by 6. What is your opinion on that matter?" Only 32% of the students claimed that it must be a mistake, or that it is impossible. Many did not explain the apparent contradiction. A portion of the subjects claimed that the theorem is true only for some classes of numbers, or that Moshe's result refutes the statement of Dan. There were subjects who claimed that one must check the theorem for various numbers or that "an exception is always possible." Most of the same students have affirmed previously that they accept the proof as fully correct.

In reality, their basic, intuitive attitude towards a general, mathematical statement was identical to that in empirical situations in which there are no universally valid proofs. Exceptions are possible and additional controls are therefore welcomed (Fischbein & Kedem, 1982). The act of learning the theory and the meaning of mathematical proofs does not necessarily change the intuitive, the deep-structure attitude of the individual. Our opinion is that special training is required which would create in the student an intuitive understanding of the meaning of a formal proof (with its absolute, universal validity).

## Summary and Didactical Suggestions

For a long time, reasoning has been analyzed largely in terms of propositional networks governed by logical rules. The modern information-processing approach--inspired by computer programming--has continued along the same line and emphasized the conceptual algorithmic structure of thinking. But since 1960, researchers have become aware of the decisive role played by cognitive components deeply rooted in our adaptive behavior, such as images, models, and beliefs. Kelly (1963) emphasized the role of beliefs and expectations; Norman (1979, 1982) analyzed the structure of models with their limitations. Paivio (1971), and more recently Shepard (1978), were concerned with the impact of images on reasoning. (This is only to recall a few from the hundreds of contributions.)

The term intuition accounts for constructs that synthesize these various aspects of problem solving in unitary cognitive structure. An intuition is a nodal moment in the flow of cognition, expressed with a stabilized, confident expectation which exceeds the data at hand. Intuitions--both anticipatory and affirmatory--represent in the stream of thoughts the apparently firm, reliable grounds that allow an individual to progress in problem solving.

But the crystallization of intuitions implies additional, often extraconceptual, elements. Pictorial and behavioral interpretations, analogies, and paradigms contribute to the imbuing of ideas with an appearance of familiarity, practicality, and direct accessibility. An anticipatory intuition may inspire a new direction for solution attempts, and affirmative intuitions may enable the student to achieve a deeper, more personal, and more productive understanding of a concept or statement.

On the other hand, the intuitive loading of a concept may omit or distort its genuine meaning. Conflicts between intuitive meaning and formal constraints may arise without either the student or the teacher becoming aware of them.

Mathematical entities do not have an external, independent existence as do the objects of empirical sciences. Mathematics involves entities whose properties are fixed by axioms and definitions; dealing with such entities requires a mental attitude that is fundamentally different from that required by empirical, materially existing realities. When one defines a category of concrete objects, one knows that the definition only approximates the knowledge of the respective category. New properties, not deducible from the definition, may be discovered. Mathematical entities owe their very existence and all of their properties to that which has been imposed by definition. This creates a new didactical situation: the student must learn to understand and to use mathematical concepts in absolute conformity with the corresponding axioms and definitions, no less and no more. This is an important and very difficult task.

Consequently, special exercises should be devised to train students to analyze concepts and definitions in order to distinguish clearly between the properties imposed by definitions and those suggested by intuitive components. Is a square a parallelogram? Certainly it is, because it corresponds to the definition of the parallelogram. May a tangent have more than one point of contact with the curve? Why not? The unicity of the point of contact is not included in the definition of the tangent (expressing the slope of the curve in a given point). Are the set of points of a line segment and the set of points of a square equivalent? If a point is identified as a small spot, the two sets are certainly not equivalent. If the point is considered zero-dimensional, there is no intuitive answer to this question; the answer is purely abstract, based on a formal proof.

One cannot eliminate the usual intuitive representations associated with mathematical concepts. We cannot eliminate these analogies, behavioral meanings, images, and paradigms because <u>this is the way we think</u>. Our thinking activity remains profoundly rooted in our adaptive, practical behavior, which implies spatiality, structurality, and fluent continuity. The main problem is to learn to live with the intuitive loading of concepts--necessary to the dynamics of reasoning--and, simultaneously, to control conceptually the impact of these intuitive influences.

Wittmann put it clearly:

> The students should gradually learn to analyze concepts, constructions, theorems and proofs. Such analyses are based on a written piece of mathematics, e.g., a proof, a small context of concepts and theorems. They aim at deeper understanding of the assumptions of a proof, of the form of inferences, of logical relationships and at the formulation of more systematic versions of the text at hand. (1981, p. 395)

What we would like to emphasize is that such analyses should habituate the student to become aware of the exact formal meaning and implications of mathematical concepts, as distinct from the implications of the underlying intuitions. Without its engine and wheels, a car could not move--but the steering wheel controls its direction.

Secondly, students should also learn to analyze and formalize their primary intuitive acquisitions. The student must learn to abstract formal structures from practical realities, to define them, to render explicit the properties of a class of entities, to produce proofs after anticipatory intuition has suggested a certain statement.

A third aspect refers to the role of heuristic attitudes in mathematical reasoning. A creative mathematical activity is a constructive process not reducible to mere deduction. In a constructive process, one must anticipate, and this implies a certain amount of guessing. Guessing in a problem-solving endeavor is not a blind trial-and-error process. Some general heuristics have been described, including the means-end strategy, intuitions based on analogy or induction, and reference to a known or more simple problem.

When one guesses, one usually does so in accordance with the lines of force determined by intuitive tendencies and not necessarily in conformity with formal constraints. The first basic recommendation for developing anticipatory intuitions is to improve the capacity to discern the formal mathematical properties beyond the intuitive representations.

Analogies seem to play a fundamental role in generating new ideas as Poincaré (1913) and Polya (1954) have emphasized. Much greater attention should be given, in our opinion, to instilling in students a sensibility for similarities, an ability to identify isomorphisms and to describe common structures. Our assumption is that if the student is consciously accustomed to proceeding this way he will develop similar capacities at a subconscious level. During his problem-solving efforts, apparently spontaneous, productive analogies will emerge automatically and will become a source of anticipatory intuitions.

We propose that the capacity to evaluate preliminary solutions and the plausibility of intuitive leaps can also be trained. This probably does not involve teaching formal problem-solving strategies. It is rather a problem of practical training in which systematic classroom discussions and evaluation of competing hypothesis may play an important role.

## References

Behr, J. M., & Wachsmuth, J. (1982). Turning the plates--size perception of rational numbers among 9 and 10 year-old children, in A. Vermandel (Ed.). PME Proceedings of the Sixth International Conference, 181-186. Universitaire Instelling, Antwerpen.

Bergson, H. (1954). Creative evolution. Translated by A. Mitchell. London: McMillan, et Co. Ltd.

Bruner, J. ( 1965). The process of education. Cambridge, MA: Harvard University Press.

Bunge, M. (1962). Intuition and science. Engelwood Cliffs, Prentice Hall, Inc.

Descartes, R. (1967). The philosophical works. Translated by E. S. Haldane and G. R. T. Ross. Cambridge, MA: The University Press.

Fischbein, E., Deri, M. M., Nello, M. S., & Marino, M. S. (1985). The role of implicit models in solving verbal problems in multiplication and division. Journal of Research in Mathematics Education, 16(1), 3-17.

Fischbein, E., & Kedem, J. (1982). Proof and certitude in the development of mathematical thinking. In A. Vermandel (Ed.), PME Proceedings of the Sixth International Conference. Antwerpen: Universitaire Insetlling.

Galbraith, P. D. (1981). Aspects of proving. A clinical investigation of process. Educational Studies in Mathematics, 12, 1-28.

Glaeser, G. (1981). Epistémology des nombres relatifs. Recherches en Didactique des Mathématiques, 2, 3, 303-346.

Hadamard, J. (1949). An essay on the psychology of invention in the mathematical field. Princeton: Princeton University Press.

Hahn, H. (1956). The crisis in intuition. In J. R. Newman (Ed.). The world of mathematics (pp. 1957-1976). New York: Simon and Schuster.

Hunting, P. R. (1986). Rachel's schemes for constructing fraction knowledge. Educational Studies in Mathematics, 17(1), 49-66.

Inhelder, B., & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. London: Routledge & Kegan Paul Ltd.

Kelly, G. A. (1963). A theory of personality: The psychology of personal constructs. New York: The Norton Library.

Kieran, C. (1981). Concepts associated with the equality symbol. Educational Studies in Mathematics, 12, 317-326.

Knifong, J. (1974). Logical abilities in young children--two styles of approach. Child Development, 45, 78-83.

Norman, D. A. (1979). What goes on in the mind of the learner? In W. J. McKeachie (Ed.), New directions in learning and teachings. San Francisco: Jossey-Bass.

Norman, D. A. (1982). Some observations on mental models. In D. Gentner and A. Stevens (Eds.), Mental models. Hillsdale, NY: Erlbaum.

O'Brien, T. C., Shapiro, B. Y., & Reali, N. C. (1971). Logical thinking--language and context. Educational Studies in Mathematics, 4, 201-219.

Paivio, A. (1971). Imagery and verbal processes. New York: Holt.

Poincaré, H. (1913). The foundations of science. (Translated: Holsted). New York: Science Press.

Poincaré, H. (1914). Science et methode. Paris: Flamarion.

Polya, G. (1954). Induction and analogy in mathematics. Princeton, NJ: Princeton University Press.

Reid, C. (1970). Hilbert. Heidelberg: Springer.

Rusu, E. (1962). Despre invatarea matematicii. Aspecte psihologice. Bucuresti, E.S.D.P.

Shepard, R. N. (1978). Externalization of mental images and the act of creation. In B. S. Randhawa & W. E. Coffman (Eds.), Visual learning, thinking and communication, 133-189. New York: Academic Press.

Spinoza, B. (1967). Ethics (A. Bogle, Trans.). London: Everyman's Library, J. M. Dent.

Tall, D. (1980). The anatomy of a discovery in mathematics research. For the Learning of Mathematics, 1, 25-34.

Vergnaud, G. (1983). Multiplicative structures. In R. Lesh and M. Landau (Eds.). Acquisition of mathematics concepts and processes, 127-174. London: Academic Press.

Wittman, E. (1981). The complementary roles of intuitive and reflective thinking in mathematics teaching. Educational Studies in Mathematics, 3, 389-397.

Chapter 15

# MATHEMATICS CURRICULUM ENGINEERING: SOME SUGGESTIONS FROM COGNITIVE SCIENCE

Thomas A. Romberg and Fredric W. Tufte

The purpose of this paper is to present some of the implications that recent research in cognitive science has for the engineering of mathematics curricula. To build a curriculum one must make several decisions about the content that is to be included, how that content is to be segmented, how the segments are to be sequenced, approximately how much time is to be spent on each segment, and what is to be considered acceptable work. These are all curriculum engineering decisions. In this paper we propose a set of principles on which such decisions should be made. The principles have been derived from recent psychological research. Since this research is not about curriculum engineering but about how people process and retain information, the principles must be considered as suggestions based on this research rather than as findings. To build a list of principles, we first describe the curriculum engineering problem being addressed; second, we briefly outline what it means to draw inference from research; third, we give a summary of cognitive science research related to how information is stored in long-term memory; and finally, from this research we draw curriculum engineering principles.

The rationale for preparing this paper is that, if significant gains are to be made in the mathematical accomplishments of school children, then as Romberg and Carpenter (1985) have argued, "researchers and curriculum developers must be attuned to a changed perception of what it means to know mathematics and to what the rapidly expanding literature from cognitive science has to say about how children, adolescents and young adults store and process information" (p. 852). In this chapter findings from cognitive psychology that appear to have application in an educational setting are presented.

Furthermore, it is a premise of this paper that, as expressed by Romberg (1983), to know mathematics is to do mathematics, and that among the essential activities involved in doing mathematics are abstracting, inventing, proving, and applying. Mathematics is not, as it is often taught, a static collection of bits and pieces, leading nowhere except to achievement on a test measuring knowledge of terminology and algorithmic procedures. The fragmentary nature of many existing mathematics programs leaves the student with an almost total inability to apply mathematics in any but routine situations and, in fact, with very little experience with mathematical thought itself. The future emphases of instruction must be on the powerful ideas of mathematics, their interrelatedness, and the development of quantitative reasoning (Romberg, 1984). To accomplish mathematics programs with these emphases new curricula will have to be developed. This paper was prepared to give direction to that work.

## CURRICULUM ENGINEERING

A curriculum is an operational plan detailing what content is to be taught to students, how students are to acquire and use that content, and what teachers are to do in carrying out that curriculum (Romberg, 1970). The key to this definition is the notion of planning and that human beings are involved in the planning effort. Romberg and Price (1983) have pointed out that such a plan is viewed differently at different levels. There will be general specifications and needs at a "board of directors" level, a package of materials at a publishers level, guidelines to teachers at a local level, and daily lesson plans at the teacher level.

Curricula also can be viewed from different content conceptualizations: an ideal curriculum as envisioned by curriculum theorists; an available curriculum as reflected in the current textbooks, curriculum frameworks, etc.; the actual curriculum that is implemented in a particular classroom; and the learned curriculum (Romberg, 1985). Because of these differing perspectives it should be clear that building an operational plan for a curriculum is a complex task. Curriculum engineering is the iterative process by which parts for the operational plan are invented and then put together into the final plan to be implemented. The process is iterative in that no product is ever viewed as a final "best" plan. Rather, changes are always anticipated, and each new model is to be an improvement over the old. In this section the traditional concerns in building a curriculum are first described, then a rationale for challenging that tradition is presented.

### Traditional Curriculum Engineering

The steps of traditional curriculum engineering have been common practice for decades. They were formalized in the 1930s by Ralph Tyler (1931). The process begins with an epistemological assumption that knowledge is external to the knower. For example, mathematics is viewed as a body of knowledge (concepts, skills, procedures) that is well defined and agreed on in the society. The goal for schools is to expose students to this body of extant knowledge. The engineering task then involves four steps:

1) The content of mathematics is organized into several agreed on categories. Typical content categories for school mathematics include arithmetic, algebra, geometry, statistics, measurement, trigonometry. These are sometimes referred to as strands (California

Statewide Mathematics Advisory Committee, 1972) or learning hierarchies (Harvey, McLeod, & Romberg, 1970).
2) The content categories are then segmented (organized into topics or chapters). Each topic is to take from two to four weeks to teach.
3) The topics or chapters are then sequenced for instruction.
4) Finally, specific activities or lessons are developed within each topic.

This approach to curriculum development puts its emphasis on the content to be covered. Only in the last step when activities are developed is any consideration given to either what learners know and are capable of doing or the work teachers are to do.

## A Challenge to Tradition

We believe there are four problems with the traditional approach to curriculum development. Based on these problems, a new approach seems warranted.

1) Student's conception of mathematics. To most students mathematics is a static collection of concepts and skills to be mastered one by one. Furthermore, each student's task is to get correct answers to well-defined problems or exercises. Most recent curricula in mathematics has been over fragmented. The use of behavioral objectives and learning hierarchies, such as advocated by Gagné (1965) and operationalized in many individualized programs, such as IPI (Lindvall & Bolven, 1967), has separated mathematics into literally thousands of pieces, each taught independent of the others. The difficulty with this approach is that, while an individual objective might be reasonable, it is only part of a larger network. It is the network (the connections between objectives) that is important. Students get as a view of mathematics isolated pieces rather than relationships.

The fragmentation and resulting emphasis on low level objectives is reinforced by the testing procedures often associated with such curricula. Multiple-choice questions on concepts and skills emphasize the independence rather than the interdependence of ideas and reward correct answers rather than reasonable procedures.

Students' conceptions of mathematics are greatly influenced by their teachers, and in the United States most teachers do not have a broad view of mathematics. Few of our teachers are familiar with the history or philosophy of mathematics or have ever worked as mathematicians. The large majority of teachers' knowledge of mathematics is what is done in schools. Therefore, it is not surprising that they see little reason either to view or to teacher mathematics in a different way. They have little sense of mathematics as a craft, as a language, and as a set of procedures to solve problems. Mathematics does not simply deal with procedures to get answers. It involves such activities as assigning numbers (measurement), building mathematical models to represent situations, and examining patterns (Romberg, 1983).

The segmenting and sequencing of mathematics has led to an assumption that there is a strict, partial ordering to mathematics. In American schools, this assumption translates into guidelines such as "you can't study geometry unless you can do arithmetic; you can't study algebra unless you can do decimals; you can't study calculus unless you have had trigonometry." A student who is having difficulty adding fractions with unlike denominators should not be denied the opportunity to study geometric relationships.

In summary, the most serious problem faced by curriculum developers is to realize that, while daily lessons (pieces of mathematics) must be taught, the interconnectedness of ideas must somehow become the focus of instruction.

2) Learning as absorption. Traditional mathematics programs have conceived of the learner as being a passive absorber of information, storing it in memory in little pieces that are easily retrievable. Note that this view of learning is consistent with the fragmentation of mathematical content.

Probably the most dramatic research findings of the past quarter century show that learning is not like that at all. Instead, individuals approach each new task with prior knowledge. They assimilate new information and construct their own meanings. These research findings are the basis of the recommendations made in this paper.

3) Deskilling of teachers. Because of concerns about trying to get teachers to adopt and use new programs, there has been a tendency to overspecify instructions for teachers. A detailed syllabus takes important teaching skills away from the teacher. Often there are no decisions left to make about what activities to use or how much time to spend. Taken to an extreme, the teacher becomes only a conduit in a system, covering the pages of a book without thinking or consideration; the emphasis in teaching is shifted from curricular content and individual learning to management; the teacher becomes a manager of resources and personnel (Berliner, 1982). As one teacher put it, "I am teaching your mathematics to my students" (Stephens, 1982/1983).

Teachers are not encouraged to adapt or change to meet local needs or conditions. They are not encouraged to relate ideas of one lesson to another. For students, mathematics becomes completing pages or doing sets of exercises with little relationship between ideas, and teachers reinforce this perspective.

Stephens (1984) has discussed this problem in more detail. He pointed out the distinctions between teacher work associated with a centrally developed curriculum, curriculum guidelines, and locally developed curricula. Unfortunately, the assumption made by many developers is that teachers are not competent to develop their own curricula; therefore, development decisions are made for them. Teachers are then unaware of the reasons for such decisions, of the values associated with various activities, and of the importance of various actions. As a result they are likely to become more technical adopters of the curriculum. This was certainly the fate of most of the modern mathematics programs in the 1960s.

4) Text as technology. Most curriculum development work has emphasized the development of textbooks. The result has been that the curriculum has been defined by the textbook. The curriculum package includes the text, which is a repository of problem lists, a set of paper-and-pencil worksheets, and a chalkboard. Children are to work independently with little opportunity to discuss, argue, build models, or try out ideas collaboratively. However, mathematics is not simply working paper-and-pencil exercises. Although many of the new books include things to read, there is very little that is interesting to read. Thus, textbook mathematics gives students little reason to connect ideas in "today's" lesson with those of past lessons.

These four difficulties, we believe, stem from a narrow mechanical concept of education. This is true of all education, but it is especially true for mathematics. Too often the acquisition of a prescribed amount of knowledge under competitive conditions and time pressures constitutes mathematics instruction. If we are going to do anything different, now is the time to consider a new approach.

We believe that information about how individuals personally construct knowledge and store it in memory should be the basis of curriculum engineering. This is a different epistemological basis for knowledge than traditional engineering. In this view all knowing is personal and idiosyncratic. Nevertheless, consensual meanings can be arrived at via negotiation. It is this perspective and the research on which it is based that have led us to the set of recommendations in this chapter.

## Research and Implications for Practice

This brief section has been included in this paper for two reasons. First, we believe that all educational decisions (including those about curricula) should be based on valid, reliable information. A primary source for such information is research. Second, because most of the research referenced in this paper was not carried out to inform the topic of this paper, curriculum engineering, the inferential procedures must be justified.

The primary purpose of any research program is to try to make sense out of a complex phenomenon. The first step in such a program is to develop some model (framework, metaphor, etc.) designed to capture what are believed to be important features of the phenomenon. All such models are of necessity incomplete. Nevertheless, they are fundamental to the investigations that follow, for it is from the model that conjectures are derived. Second, a research program is established to systematically gather and report evidence to substantiate or refute those conjectures. In this sense all research results are descriptive since the findings are about the model. Finally, it is hoped that such research eventually can provide us with an understanding of the phenomenon.

Alan Bishop (1982) has argued that there are two things one can learn from research: the researcher's view of the phenomenon (the model) and the way evidence is collected about conjectures. It is this view of research we want to stress in this paper. In particular, because the research that is summarized in this chapter clearly refutes the simplistic "learning as absorption" notions of traditional curriculum engineering, new principles based on this research seem warranted.

## COGNITIVE SCIENCE

As anyone familiar with human psychology can attest, there has been a major revolution in the field during the past decade. The variety of current models of human processing of information and learning have been labeled "cognitive science." Although there are many variants, they all are based on the metaphor of the computer, in that information is assumed to be received, stored, and processed by humans in ways that are analogous to how a computer performs those same actions. This is not the place for a review of those models. For the reader unfamiliar with this work the brief book by Phillips and Soltis (1985) is a good introduction. Howard Gardner's book _The Mind's New Science_ (1985) is a thorough discussion of the history of this revolution. Richard Anderson's treatise _The Architecture of Cognition_ (1983) is an excellent example of current theorizing in the field.

What is important for this paper is the research from cognitive science which suggests that learning occurs when information entering the senses is actively processed and related to previously learned information stored in a permanent semantic and factual knowledge base. New information is fitted or assimilated into existing cognitive structures in such a way as to provide a meaning, an explanation, an order, or a logic for the experiences being witnessed or reflected on by the learner. For example, the typical American seeing the word _pectopah_ for the first time when in Moscow is unable to suggest any meaning for the term. However, if this word in cyrillic were transliterated to the Roman alphabet as _restoran_, one would probably guess that it was the Russian word for restaurant. All kinds of images would then be available to give it meaning.

A consequence of this assimilation process is that each individual's knowledge is uniquely personal. Different individuals process and link new information in unique ways and, hence, develop cognitive structures that reflect different perspectives of the same reality. Hewson (1982, 1984) hypothesized three conditions necessary for the assimilation of new information. First, the learner must understand the new information; second, the new information must be reconcilable with existing conceptions; and third, the resulting accommodated structure must be useful. If these conditions are satisfied, the potential for learning exists.

Greeno (1980) and Greeno and Bjork (1973) have presented an information processing model of memory that is representative of those used by cognitive psychologists. In the Greeno and Bjork model, sensory information enters short-term sensory storage (STSS) where it is held momentarily. Information selected from this system is held by working memory (WM). Working memory has a small capacity to hold information; it is generally assumed to hold from five to nine "chunks" of knowledge (Miller, 1956). The information in WM is hypothesized to have a short life span, on the order of a few seconds. If the information selected for WM can be organized in some way, it is stored in short-term memory (STM) for minutes or hours. From STM the information may become integrated with the individual's existing knowledge to become a part of the semantic and factual knowledge base which is stored in long-term memory (LTM). Also, there is posited an executive control mechanism, consisting of a set of metacognitive processes, whose purpose is to enhance the exchange, storage, rehearsal, and retrieval of information, between and within memory systems. When an individual processes information, it is common to view WM not so much as a holder of information itself, but as a system of pointers which are associated with or point toward chunks of information from short-term memory and the semantic and factual knowledge system in long-term memory.

Semantic nd factual knowledge are stored in LTM in procedural and declarative schemata. It is these schemata that are often referred to as knowledge structures. A schema or knowledge structure can be envisioned as a hierarchical network consisting of nodes connected by lines representing some type of relationship. The relationship might be superset, subset, attribute, similarity, proximity, operation, antecedent, consequent, etc. For example, seeing the word restaurant triggers a "restaurant schema" from LTM which is based on an individual's past experiences of dining at restaurants. Figure 1 represents a possible schema for "quadratic equation." The concepts are nodes in the semantic network; each node may be a supernode which is itself a network of nodes.

It is the relationships that carry inference that form the basis for organizing semantic information, and it is these relationships that make it possible for people to know more than they learn (Shavelson, 1974). For example, if entity A is similar to entity B and B is similar to entity C, then it may be possible to infer that A is similar to C, at least if similarity is

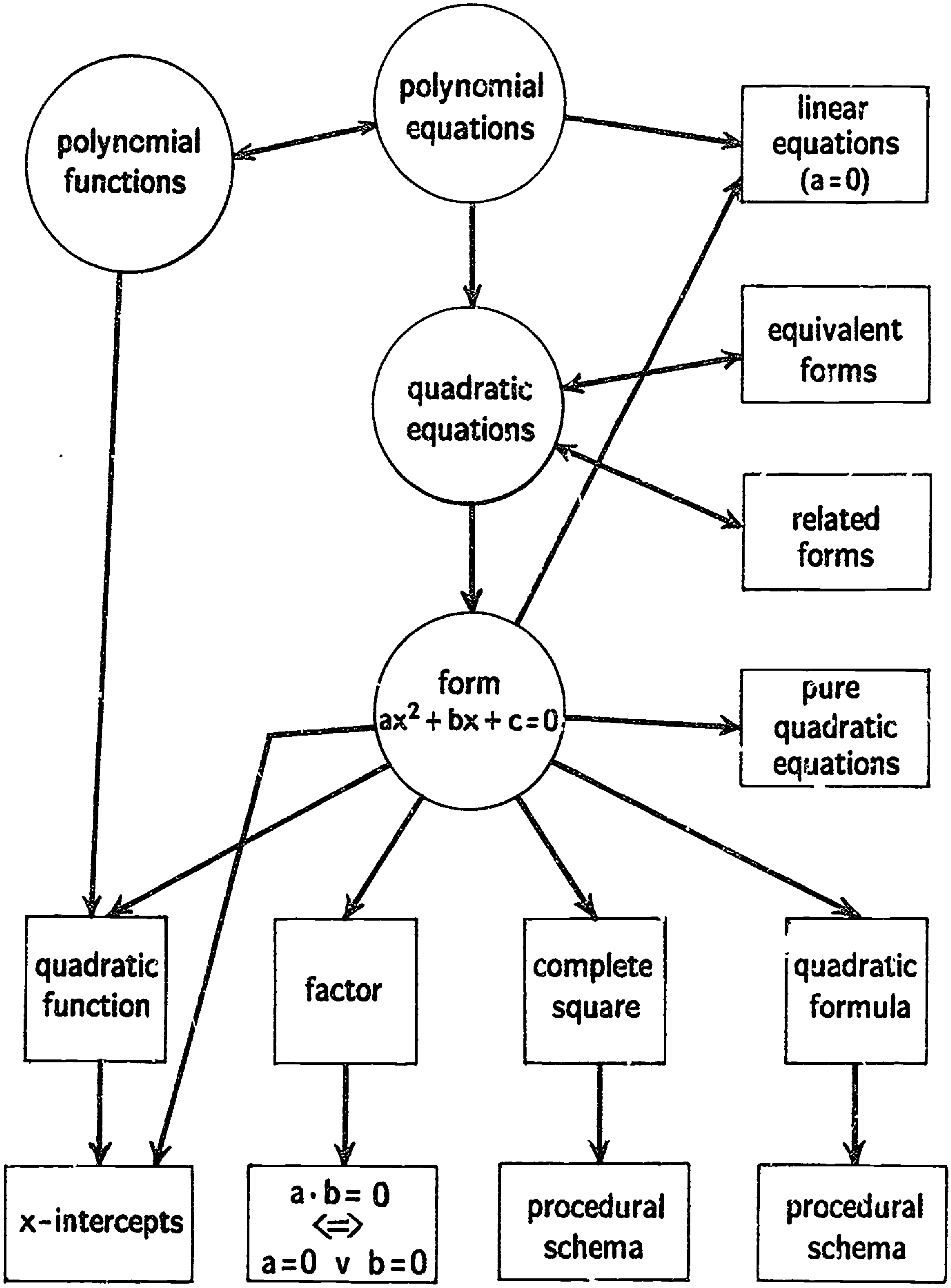



Figure 1. Schema for "quadratic equation."

transitive. Checking the transitivity schema may be "controlled" by the executive control mechanism.

Many researchers view schemata as a psychologically rational organization of information and procedures that are used to understand the world. The components or entities which comprise the schema are similar in nature to variables, in that similar situations or experiences can be interpreted through the use of the same schema. The particulars of the situation become instances of the variables. Usually there are default values which the variables may assume if particular values are not explicit to the situation or experience. To a mathematician, for example, the mental construct of a quadratic function is quite similar to what a football play construct would be to a coach. Very possibly, the mathematician views a quadratic function as a type of polynomial function. When reference is made to a quadratic, a specialized form comes to mind, associated with more specific values of variables comprising the polynomial schema. And, if no value is specifically mentioned, the mathematician might initially assume the coefficient of the linear term to be nonzero, just as a coach might assume a specific alignment or placement of players for execution of a particular football play. These default values, built up by exposure to numerous similar experiences, help to provide a coherent view of the situation or experience.

Studies conducted using nonspecific general knowledge provide conclusive evidence that appropriate well-developed schemata facilitate learning and recall. For example, Anderson, Spiro, and Anderson (1978) showed that subjects were able to recall a list of 18 food items better when it was embedded in a story about dining at a restaurant than when given alone. The greater structure provided opportunities to associate food items with various experiences in the dining episode.

In another experiment (Spilich, Vosonder, Chiesi, & Voss, 1979), subjects were given a description of a half-inning in a fictitious baseball game. Knowledgeable baseball fans were able to recall more information about the game than were low-knowledge subjects. Generic features of the game, known by knowledgeable fans, were useful in recalling information. These features provided a structure into which the specific details of the fictitious game could be incorporated, and recall of a few key details or events could then trigger a natural or logical progression of the story.

Both of these studies make use of schemata that were developed over long periods of time, by frequent exposure during everyday experiences. Such schemata can be well developed and probably explain the superior performance obtained on the many tasks employing them. More specific schemata, developed in school for the purpose of performing school tasks, may be a different matter. Such schemata must often be developed over much shorter time periods and with a more limited set of experiences, which are often contrived, and with features considered uninteresting or of dubious value to the student.

The major premise of this paper is that the mathematics curriculum should reflect the way knowledge is optimally organized in the semantic and factual knowledge base. Elaboration of this premise necessitates an understanding of the way knowledge is organized in memory and, more specifically, the type of organization that promotes both the encoding and retrieval of information.

To this end we now examine three areas of investigation that have provided knowledge about effective cognitive functioning and the organization of knowledge in the permanent memory base. First, we consider formal modeling of problem-solving protocols; second, we look at qualitative differences between the problem representations of novice and expert problem solvers in content-rich domains; and last, we discuss the results of research on the recall of lists, stories, and prose. We then relate several recent curricular innovations that have attempted to incorporate these results from cognitive science.

## Formal Models

The analysis of formal models of problem solving is fueled by the hope that these models and corresponding computer simulations of problem-solving behavior will shed light on cognitive functioning. Design of the models, reflecting problem-solving capabilities of human problem solvers, may in itself provide insight into the nature of thinking and effective and efficient problem-solving activity.

Production Systems. Growing out of the problem-solving protocols of puzzle, chess, and scientific problems are descriptive models employing the use of productions. A production is a process containing two components, a condition component and an action component (Simon, 1978). The condition component of a production is a set of tests to determine whether elements satisfy certain conditions. The action component specifies the action or actions to be performed on the elements if they meet the tests prescribed by the condition component of the production. A set of productions is called a production system.

By examining think-aloud problem-solving protocols of subjects solving simple kinematics problems, for example, it is possible to design a production system reflecting their problem-solving behavior. Both strategy and sequencing considerations can be built into the system. Suppose a production system is to model the performance of a novice solving a simple kinematics problem using means-end analysis. The condition parts of each production would test to see whether the independent variables of each equation are known and whether the dependent variable is wanted. If both of these conditions are met, the action part--solving the equation for the dependent variable--will be executed. If the dependent variable is wanted but not all independent variables are known, the action part of the production would not be executed, but the first independent variable that was not known would be put on a list of wanted variables. This latter action would be accomplished by a separate production.

Testing of production systems, to substantiate the degree to which the system reflects performance, can be done by comparing the protocols of the subjects with those of the production system on a wide variety of problems within the capabilities of the system. Researchers have been able to obtain remarkable similarities between the protocols of individuals and their corresponding production systems (Simon & Simon, 1978; Anderson, Greeno, Kline, & Neves, 1981). These good matches suggest the adequacy of the production system to model at least some of the cognitive behaviors of the subject. Furthermore, by making slight modifications or additions to the production system of a novice, it is often possible to model the problem-solving behavior of experts. Differences in the cognitive structures of experts and novices can then be studied by examining the changes that were made in the production system.

Computer Models. Probably the easiest and most reliable way of testing a production system is to use a computer program incorporating the productions of the system. A program becomes the model of cognition involved in problem solving. The computer is able to keep track of all its executions (procedural knowledge), components (factual knowledge), and the interactions between them. It provides a powerful tool for testing, developing, and refining models of problem solving. Its untiring ability at repeated execution enables the researcher to test the completeness, reliability, and consistency of proposed theories. The use of a few parameters enables a single program to exhibit problem-solving behaviors of both experts and novices as well as all stages of development in between.

For example, Anzai and Simon (1979) studied the think-aloud problem-solving protocols of a student solving and refining a solution to the Tower of Hanoi puzzle. In each of four trials, the subject used a strategy which was both a transformation of and more efficient than the strategy used on the preceding trial. Each of these strategies was first programmed as a production system. Analysis of the differences among these systems led to an understanding of the transformations made by the subject from one trial to another. This information was then used in the design of an adaptive production system with the ability to create each new strategy from the preceding one. That is, the system acquired "knowledge" about the effectiveness of its moves, depending on whether the move had favorable or unfavorable consequences, and used this knowledge to modify itself. The corresponding computer program served as a model for the "learning by doing" that was exhibited by the subject on the Tower of Hanoi puzzle.

Successes documented by researchers in artificial intelligence lend support to the notion that condition-action mechanisms play an important role in human thinking. The mathematical and logical structures normally built into computer hardware and software do not, however, reflect the informal and primitive structures used by human problem solvers.

From both a practical and theoretical perspective, it is impossible to clearly separate conceptual and procedural knowledge. We may operate procedurally and conceptually simultaneously, and in fact individuals may operate differently when confronted with the same mathematical problem. For example, when novices solve the equation $x(x - 3) = 6$ by writing $x = 6$ or $x - 3 = 6$, they are operating at a procedural level, albeit an incorrect one. The action component of a production is being triggered without a check on the conditions that make the action appropriate. Many buggy algorithms appear to be of this type. Incorrect application of the distributive property often results in student errors of the form $f(a + b) = f(a) + f(b)$. It is surprising, after exercising considerable care in introducing both the trigonometric functions and the addition formulas, that many students insist on writing $\sin(a + b) = \sin(a) + \sin(b)$. It seems that the distributive property and other mathematical formulations such as the definition of function addition set students up for committing these errors. Too many students make generalizations at a symbolic-manipulative level rather than attend to a conceptual understanding of the principles involved. As much care must be exercised in teaching the conditions under which mathematical properties and theorems can be applied as in the actual applications of these properties and theorems. And although automaticity may eventually be desired, the use of conceptual knowledge should initially guide students' activities.

Just as formal systems of logic provide powerful methods of manipulating and processing data, the idiosyncratic informal knowledge structures developed by individuals, built up as they are by exposure to numerous similar and related experiences, have dominating effects on reasoning and problem-solving abilities. Whereas formal systems are complete and consistent, the function of education is, in large measure, the process of molding or changing the informal structures of novices to more closely resemble those of the more formal systems of expert problem solvers.

## Qualitative Differences

Several lines of investigation have shed light on the informal modes of reasoning employed by human subjects. One particularly fruitful area of research has dealt with the qualitative differences between novice and expert problem solvers in context-rich domains. A primary goal of this line of research has been to make more explicit the relationship between conceptual knowledge and problem-solving strategies.

One basic difference found in novice-expert information processing of elementary physics problems is the type of strategy employed (Chi, Glaser, & Rees, 1982; Larkin, 1979; Larkin, McDermott, Simon, & Simon, 1980; Simon & Simon, 1978). Novices tend to use a means-end or "working backwards" strategy while experts use a "forward looking" strategy. For example, if the problem requires $\underline{z}$, given $\underline{u}$, $\underline{v}$ and $\underline{x}$, the novice first searches memory for an equation containing $\underline{z}$ as the dependent variable. Suppose $\underline{z} = \underline{f}(\underline{x},\underline{y})$. Because $\underline{x}$ is given, the novice must next find an equation containing $\underline{y}$, preferably containing $\underline{u}$ and $\underline{v}$ as independent variables. Suppose $\underline{y} = \underline{g}(\underline{u},\underline{v})$. Now the novice can calculate $\underline{y}$ and use the function $\underline{f}$ to calculate the original unknown $\underline{z}$. So the novice works backward from the quantity that must be found. The expert, on the other hand, concentrates on the given quantities $\underline{x}$, $\underline{v}$, and $\underline{u}$. Search is made for an equation containing the given quantities along with one unknown. Hence, the expert might first pick $\underline{y} = \underline{g}(\underline{u},\underline{v})$, solve for $\underline{y}$, and then use $\underline{z} = \underline{f}(\underline{x},\underline{y})$ to solve the unknown $\underline{z}$. Initial attention of the novice is directed at the unknown whereas the expert concentrates on the variables whose values are given. To the novice, the goal is the overriding feature of the problem, and attention is directed at that factor. As a result, it may be that insufficient attention is directed at other essential features of the problem structure. Larkin (cited in Woods & Crowe, 1984) has found that persons successful in completing a problem spend considerably more time reading the problem statement before beginning to write equations.

The two modes of processing that have been characterized by working forward and working backward were observed by Kantowski (1975) in an early study involving the use of heuristic strategies in geometry problems. Kantowski noted that students who were unsuccessful in proving geometry theorems often attempted to work backward from the conclusions. Successful subjects worked forward from the hypotheses. In fact, when the conclusions of the theorem were withheld and students were asked to obtain as many results as possible from the given hypotheses, previously unsuccessful students were often able to generate the correct conclusions.

In a related experiment, Sweller (in press) has confirmed that reducing the goal specificity in trigonometry problems enhances problem-solving skills. When Sweller's students were requested to find all unknown parts of a triangle, they were more successful than when asked for a particular part of the triangle. Solving for a particular part of a triangle is characterized by a greater processing load. The student must sort through the trigonometric functions, selecting those involving the correct variables. With objectives of less specificity, selection of an appropriate function is less critical, and the subject is free to devote processing capacity to other concerns. For example, given the right triangle ABC, it is easier to select a trigonometric function involving the pair (a,c) than it is to find one involving the triple (a,c,B). The triple involves a greater processing capacity.

It is interesting that the working forward strategy of the expert is not invarient across problem types. When problems become more difficult, experts usually revert to the same means-end analysis employed by novices. Only when experience indicates that the problem space falls within richly developed cognitive structures does the expert concentrate attention on the independent variables. As experience and confidence diminish, it appears goals and subgoals are required to direct the searches of the knowledge structures (Larkin et al., 1980). As noted previously, this attention to goal-states may place an additional burden on short-term memory. For instructional purposes, it may be that the forward working strategy is a problem-solving heuristic that should be continually emphasized.

Although difficult to define operationally, another difference between expert and novice problem solvers is the qualitative analysis applied to the problem prior to the actual retrieval of equations (Larkin, 1979; Larkin & Rainard, 1984; Simon, 1978). The greater degree of qualitative analysis used by experts appears to restructure the problem in terms of the physical principles involved. Restructuring may occur in an effort to obtain a fit between the problem and a particular knowledge structure. This suggests that knowledge structures of experts, or those components of knowledge structures which form the problem space, may in fact be structured or organized different from those of novices. Paige and Simon (1966) noted that good problem solvers were able to identify the inconsistencies in algebra word problems that contained no solutions. Less skilled problem solvers could not conceptualize the discrepancies until well into the problem solution. This suggests an additional qualitative analysis of the problem situation on the part of more skilled individuals.

Clement (1979), using clinical techniques, has documented differences in the way students interpret and view physics equations. More advanced students have much broader conceptions of the equations and corresponding variables. Richer meanings in terms of real world representations and in terms of the interactions and relationships among variables are exhibited by more advanced students.

Very possibly, formation of isomorphic real-world referents of the problem statement on the part of expert problem solvers is an attempt to obtain a match between the real-world and psychological interpretations of the variables involved. That is, when the problem representation becomes meaningful in terms of the external world, the physical forces and corresponding variables involved in the problem become more apparent. These forces and variables are thus matched with the associated representations of variables and equations from the individual's knowledge structures. Novices, on the other hand, lack these richer interpretations of the equations and variables and are more dependent on formal propositional calculus. This more deductive mechanistic solution does not require qualitative analysis of the problem.

Chi, Glaser, and Rees (1982) asked several novices and experts to classify physics problems. Results indicate that experts classify on the basis of physical principles while novices employ concepts and structural features. Words like rotation, velocity, spring, and inclined plane tend to influence categorization by novices, while conservation principles and Newton's laws determine the classifications used by experts. Hence, the knowledge structures of experts appear to be different from those of novices. Supporting this contention is the fact that experts tend to start their problem-solving protocols with such statements as "all forces sum to zero" or "$\underline{F} = \underline{ma}$," while novices initiate protocols with equations of more limited applicability (Chi, Glaser, & Rees, 1982). In fact, experts are often unable to recall many of the formulas used by novices (Simon & Simon, 1978). It appears experts' knowledge structures, or at least the problem representation which they develop, are organized around fundamental generic principles of wide applicability. These structures are also semantically rich in comparison to those of novices. For example, Simon and Simon (1978) presented the following to a novice and to an expert: A bullet leaves the muzzle of a gun at a speed of 400 meters per second. The length of the gun barrel is half a meter. Assuming that the bullet is uniformly accelerated, how long was the bullet in the gun after it was fired? The novice evoked the formula $\underline{s} = \underline{vt}$, solved for $\underline{t}$, $\underline{t} = \underline{s}/\underline{v} = (1/2)/200 - 1/400$ (although not nearly this efficiently). The expert, on the other hand, used a more general proportionality schema arguing something like the following: 200 feet in one second so 1/2 foot in how many seconds? Clearly 1/400 second. The expert used a much more general principle, applicable to the situation (using, as did both expert and novice, the average velocity). Notice also the computational efficiency of the expert protocol.

The novice-expert investigations seem to indicate that problem representations developed by novices are organized around the specific objects given in the problem, whereas the representations formulated by experts are organized around general principles and abstractions of which the objects of the problem are mere instances or exemplars. One feature that clearly differentiates novice from expert is the experience each has had within the content domain. In fact, Glaser (1984) believes that the problem-solving difficulties of novices are due primarily to inadequacies in their knowledge base as opposed to any inherent limitation in processing capacity, reasoning ability, or use of general problem-solving heuristics.

## Lists and Stories

Investigations concerning how people encode, understand, and recall lists, simple stories, and expository text provide clues about how knowledge structures might be organized to promote comprehension and retrieval of information.

Mandler (1984) reported an experiment in which subjects at three different age levels were presented a list of pictures of common objects. The list consisted of five categories of six items each. Subjects were told to either memorize the list or not, and within each of these conditions they were informed or not informed about the categorical structure of the list.

For both seven- and ten-year-olds, categorical information aided recall of the list, and performance was significantly better when the children were given the categorical information and told to memorize the list. Adults, however, performed equally well when told to memorize, whether or not they were provided the categorical information. Only when both conditions were lacking did recall significantly suffer. Adults were evidently able to invoke an organizational schema to aid recall, something that the younger children were unable to do. Categorical information aided the recall among the seven- and ten-year-olds more than did instructions to memorize. It is apparent that structure aids recall and that the ability to impose structure, even in common experiences, is age related.

Earlier in this paper we cited examples supporting the contention that incorporation of lists and stories into familiar schemata aided recall. Because both taxonomic and schematic organization improve retrieval of information, which structure is to be preferred? In an attempt to answer this question, Rabinowitz and Mandler (1983) presented college students a set of 25 phrases, each consisting of a noun and a verb. The phrases were organized into five taxonomic categories and also into five schema-related organizations. For example, one of the taxonomic categories consisted of "going places," and the places were "mountains," "Hawaii," the "theater," a "party," and the "stadium." In contrast, one of the schematic organizations involved an episode of "going skiing" and consisted of phrases dealing with experiences common to this situation. The authors found markedly superior recall for students presented the schematic organization. In addition, when other students were asked to sort the phrases into categorical or schematic groups, most students chose a schematic organization. Believing that the preferred schematic organization may have been responsible for greater recall, Rabinowitz and Mandler reconstructed their list of phrases and their taxonomic and schematic organizations. This time, even though students preferred a taxonomic classification as much as a schematic one in their own groupings, recall was enhanced when students were presented the phrases in schematically blocked groups.

An interesting aspect of the Rabinowitz and Mandler experiment was that, when students were asked to construct their own event schemata from the list of phrases, they differed in almost every case from the ones presented by the experimenters. Because no single schema was used to aid encoding and recall of particular phrases, the results cannot be attributed to a more obvious relationship of the phrases to any particular schematic or taxonomic organization. The advantage exhibited by schematic organization in the recall of phrases may be due to the larger number of, or more easily created, relational links among the objects of the schematic organization (Mandler, 1984).

Another line of research that substantiates the powerful effect that schemata have on encoding, retention, and retrieval of information concerns investigations about how people understand and recall simple stories and text. One particular effort has dealt with the structural features of stories, the associated cognitive structures, and the relationships between them.

Two central notions appear to guide much of the research on stories and narratives. First, stories consist of episodes that are themselves sequences of events or states that are causally related. That is, significant events or states in a story are causally related to subsequent events or states. Furthermore, the episodes are hierarchically arranged. For example, the protagonist of the story may break a main goal into a series of subgoals, each of which must be attained for successful completion of the main goal. Also, two or more subgoals may be linked conjunctively rather than causally. In addition to the episodes, there is usually a setting that serves to introduce the protagonist and convey information about the social, physical, or temporal context in which the episodes take place. Some research suggests that settings are among those story components that are most frequently and most accurately remembered (Stein & Trabasso, 1982). The episodes include (a) an initiating event which introduces the story line and seeks a response or formulation of a goal by the protagonist, (b) an action or series of actions by the protagonist in an attempt to attain the goal, (c) a consequence marking the concluding action of the protagonist relative to the goal, and (d) reflections or reactions by the protagonist about those actions resulting in the attainment or nonattainment of the goal.

Rumelhart (1977) has argued that readers of complex stories construct different levels of organization for story episodes. The superordinate goal, and the protagonist's attempt at attaining that goal, is at the highest level of the hierarchy. At lower levels of the hierarchy are episodes consisting of subgoals and associated attempts at their attainment, with the relationship to the superordinate goal the primary factor regulating the level of the episode within the hierarchy. Rumelhart proposed a close relationship between the level of an episode and the subsequent ability of a reader to summarize and recall the story.

Black and Bower (1980) have proposed a story memory theory that takes a problem-solving approach to story recall. Black and Bower suggest that the cognitive representation of a story is similar to the representation that is formed when an individual solves a problem. That is, the reader views the protagonist as faced with a problem, and the story changes from one state to another as the protagonist executes a series of subgoals in an effort to achieve a solution to the problem. The reader forms a cognitive structure of the story that employs causality, linking the critical events that lead from the initial state to the desired or outcome state. For a complex story, each of the story episodes forms a "path," and these are hierarchically arranged corresponding to Rumelhart's story structure. Black and Bower view the problem-solving process as traversing a series of states, with each of the states identified by subgoals. The states that must be traversed in moving from the initial state or problem state to the goal state or problem solution is called the critical path. The actions or series of actions that must be accomplished to attain each of the subgoals can be described at several levels of detail. Black and Bower provided evidence that the best remembered part of a story is the critical path; detailed events within an episode were recalled better when the episode caused a major state change, and more general, less detailed event statements were best recalled.

Meyer, Brandt, and Bluth (1980) have studied the effect of text structure on the recall of expository text. Using a structural analysis system that identifies logical connections among ideas and also hierarchical arrangements of the ideas, the authors compared the top-level structures of text with that of ninth-grade students' written recalls of the text. It was found that students who used the text's top-level structure recalled significantly more information than students who did not; however, only 22 percent of the students consistently utilized the top-level structure. Meyer, Brandt, and Bluth also found strong correlations between comprehension skills and the use of the top-level structure in text. Evidently those students using the top-level structural and organizational features of the text were able to develop a rich retrieval network which facilitated the recall of details that could be linked to the organizational structure. It was also found that the use of top-level structure was directly related to recall of the main points of the text after one week.

Because many students are unable to use the structure of text to guide encoding and retrieval of expository text, Barnett (1984) and Slater, Graves, and Piche (1985) have studied the effects of structural organizers on the recall of expository text. These studies have verified beneficial results when the organization and structure of text is pointed out to students prior to the reading of text. Structural organizers appear to aid both comprehension and recall of expository text.

## Examples

Each of the following examples illustrates the use of one or more of the results from cognitive science that have been discussed above. The first examples are related to the work of Papert (1980) and Tall (1985) and portray their attempt to develop schemata at the highest level of abstraction--schemas that could be considered generic within mathematics. Finally, a line of research by Carpenter and Moser (1982) attempted to delve into the cognitive structure of young children and suggests curricular revisions that build on existing cognitive structures.

Schemas are neither specific nor universal. Each of us interprets the world through our own highly personalized and idiosyncratic mental structures. What we learn, and in fact what we are capable of learning, depends on the mental models each of us has developed. Many of these models are built up over long periods of time, as was illustrated by our previous examples of the restaurant schema and baseball game schema. Papert (1980) provided a classic personal example involving his fascination with gears and how he would mentally "rotate circular objects against one another in gear-like motions," and how this resulted in "chains of cause and effect" (p. vi). The experiences with gears described by Papert produced a collection of models (cognitive schemata) that he could use to give meaning and interpretation to many of the mathematical problems he encountered later in life.

Papert also attributed a personal satisfaction to his thinking about gears and their motions and effects under varying conditions and configurations. His experience indicates that certain affective variables may play an important role in the development of useful cognitive models.

Although gears represented the physical manifestation of the mental structures developed by Papert, the computer with appropriate computer software is now a vehicle that he advocates for use in creating powerful mental schema. Using the LOGO language, children can actively interact with a turtle graphic and can experience those same sequences of cause and effect that Papert experienced by mentally manipulating the gears. Because the LOGO language makes use of such psychologically appealing entities as motion, user control, and immediate feedback, its use is intrinsically interesting for most students. Extensive interaction with the LOGO "microworlds" may result in the development of powerful mental schema that children can use to interpret and understand their developing world.

The Graphic Calculus used by Tall (1985) attempts to provide microworlds for the interpretation of concepts in calculus. These microworlds, which Tall calls generic organizers, are computer programs developed for the purpose of teaching concepts through the use of a wealth of specific examples. They rely on mental involvement on the part of the student and therefore differ from Papert's microworlds, which require both a physical and mental activity.

In one of Tall's programs, students enter a function and watch as the computer displays a tangent line moving along the curve and simultaneously generates the associated derived function. By analyzing the behavior of several functions the students develop a concept image that can aid in the further development and use of the derivative concept. Preliminary evidence suggests that, given the graph of some function, students who have used this program are better able to identify and construct the graph of the derived function. As a consequence, these students have developed a superior qualitative conception of the derivative concept.

Both Papert and Tall envision the development of schemata that will be particularly useful in interpreting and understanding knowledge and concepts which students will confront as they progress through life. In a sense, these generic mental structures are akin to the setting of a story; they attempt to provide the background through which the specific learning or story episodes can be interpreted.

To illustrate this point in a mathematical setting, consider for example the area schema which we wish to develop in all school children. Although the area schema overlaps other schemata such as the measurement schema, it is in itself an extremely important concept and used throughout mathematics in varying degrees of abstraction. The area concept subsumes in whole or in part many more particular schemata such as triangulation, decomposition, transformation, and limiting procedures. For example, a common technique in determining surface areas is to partition the object into constituent parts. A right circular cylindrical can, for instance, might be viewed as two circles and a rectangle. These area and partitioning schemata form a hierarchy, at the bottom of which may exist particular formulas which the student might apply. The general area schema at the top of this hierarchy is representative of the type of schema Papert and Tall are trying to develop.

Carpenter and Moser (1982) have described the rich collection of problem-solving and counting strategies used by primary school children to solve addition and subtraction word problems prior to receiving any formal instruction. At this state in their development, children analyze word problems in terms of the semantic structure of the problem and tend to model the action suggested by that structure. It is disturbing that evidence from the national assessments indicates that many children lose these natural tendencies (Carpenter, Corbitt, Kepner, Linquist & Reys, 1981; National Assessment of Education Progress, 1983). Carpenter and Moser (1982) suggest that this regression in problem-solving skill may somehow be embedded in the transition from the informal modeling and counting strategies that children initially use to the more formal use of number facts and algorithms taught in school. The mathematics curriculum is not making use of the cognitive structures or schemata possessed by children when they enter school. The mathematics they are taught is divorced from the meanings children have already developed about arithmetic operations. Methods must be found to build on the schemata students already possess.

Young children often have difficulty writing number sentences to represent word problems because the formal strategies taught in school do not reflect the informal methods of solution used by the children. For example, consider the following "join" problem.

> Jane had 3 candies. Her grandmother bought her some more. Now she has 8 candies. How many candies did Jane's grandmother buy her?

Young children normally solve this problem with a "counting on" strategy, modeling the action in the problem rather than with a "removing" strategy. Hence the solution strategy of the children more closely reflects the noncanonical number sentence $3 + \underline{x} = 8$ than the canonical form $8 - 3 = \underline{x}$ that is usually taught in the early grades. Students are required to make mental transformations in an effort to represent story problems that are more appropriately modeled using noncanonical forms. Carpenter, Bebout, and Moser (1985) have demonstrated that first-grade children can use noncanonical forms to represent and solve associated story problems, and they suggest that early instruction in writing noncanonical number sentences may be a viable approach for building on the problem-solving schemata children have previously developed. Studies investigating the early use of noncanonical forms and their effects on subsequent learning may have curricular implications if such instruction provides for the development of more powerful cognitive structures.

Recurrent forces often discourage curricular revisions. In the United States at least, the view persists at many levels that arithmetic, geometry, and algebra are separate and distinct subjects, and the content of these subjects must be taught in certain self-contained units to students of certain ages. Another viewpoint mitigating against curricular change in mathematics concerns the insistence on certain formal algorithmic procedures and the use of such procedures in problem-solving situations. For example, solution of the noncanonical number sentence $\underline{a} + \underline{x} = \underline{b}$ may require both a rethinking of the mathematics taught in the early grades and a willingness to accept solutions obtained by informal counting strategies. That is, teachers must be willing to accept a solution based on a "counting on" strategy in lieu of a transformation to a corresponding canonical form.

The fact that first-grade students can be successfully taught the use of noncanonical number sentences to represent addition and subtraction word problems suggests the extension of their action oriented problem-solving schemata to include some quite abstract mathematical concepts. It is building on existing schemata in this manner that may result in the development of rich and powerful mental structures.

## Implications to Curriculum Engineering

During the past decade several authors have written about developing curriculum units based on notions from cognitive science (Carpenter, Fennema & Peterson, 1985; Case, 1978; Wittmann, 1984). Two proposals which have or are being tried out are story-shell curriculum units (Romberg, 1983) and constructivist curriculum units (Driver & Oldham, 1985).

Story-shell Curriculum Units. Romberg (1983) suggested that the mathematics curriculum be redesigned around a sequence of curriculum units with the activities of each unit related to a "story shell." The story shell is analogous to the critical path that Black and Bower (1980) have found useful in describing those aspects of stories which are best remembered. The structure of a unit should be similar to the chapters in a Dickens' novel. His novels were written serially with a new chapter being published periodically. Thus, in each chapter characters had to be reintroduced and yet each chapter had to be complete in that a problem was introduced, and a crisis developed and was later resolved. In a similar manner story-shell curriculum units should reintroduce ideas (bring to mind current conceptions), create a crisis or conceptual conflict, and then resolve it. Thus, the unit should tell a story. It should have a beginning and an end and culminate in some knowledge deemed beneficial to the student. The important ideas, key concepts, and procedures within the unit correspond to the states along the critical path of a story. The story shell might be introduced to the students at the beginning of the unit and serve as one level of abstraction and as a structural organizer, with the students working through more detailed levels of the critical path as they proceed through the unit. The students should play a role similar to the protagonist in a story, with the student's investigations proceding from one subgoal to another. The rationale for story-shell curriculum units is as much mathematical as it is psychological (Romberg, 1983). If to know mathematics is to do mathematics, then the essential activities involved in doing mathematics consist of abstracting, inventing (discovering), proving, and applying. Mathematics is not a "static collection of concepts and skills to be mastered one by one" (Romberg, 1984, p. 7).

> In the final analysis, the importance of mathematics arises from the fact that its abstractions and theorems, for all their abstractness, originate in the actual world and find widely varied applications in the other sciences, in engineering, and in all the practical affairs of daily life; to realize this is a most important prerequisite for understanding mathematics. (Romberg, 1983, p. 127)

The story shell is intended to provide students a purpose for studying the curriculum unit and to contribute unifying constructs around which they can organize their knowledge. Traditional mathematics instruction, fragmented into a topic by topic sequence, makes it difficult for students to organize their knowledge. The curriculum often lacks unifying notions to give purpose to the topics being studied.

> The student cannot possibly appreciate the role of unification if he has no comprehension of what is being unified. More than that, because of the fact that the student does not yet know the need for, or importance of, unification, he is in effect being asked to accept the teacher's word for the fact that this is an important idea to study, one that will most

> assuredly be needed later. Thus the teaching of mathematics is carried out with the need for learning clear in the mind of the teacher, but a mystery to the student. (Fremont, 1967, p. 716)

The story shell imparts a meaning to the material being developed within the curriculum unit that the student might otherwise not have. Students often put meanings and interpretations on experiences which are not intended by the teacher (Phillips & Soltis, 1985). As a result, the cognitive structures that students develop may not be able to deal effectively with later experiences. The disaster studies are a case in point (Clement, 1979). Students often devise shortcuts, alternative methods, and ways of dealing with problems that may provide acceptable results for the problems at hand; however, these structures can be conceptually erroneous and lead to difficulty as the curriculum requires a more generalized interpretation of previously learned knowledge. The story shell can provide a framework in which students can meaningfully interpret material presented within the curriculum unit.

To conclude this section we summarize two recent curricular projects by Hewson and Posner (1984) and Curts (1985) which make very direct use of story shells of the type envisioned by Romberg (1983).

Hewson and Posner (1984) used schema theory to design instructional materials for an introductory noncalculus college physics course. The major concepts of the course were identified and found to involve the notion of "change." That is, most of the physical phenomena studied in the course involved a change from one state to another; examples included a change in position, a change in temperature, a change in magnitude, and a change in direction. The entities undergoing change, called objects of change, normally have one or more causal factors, often involving forces of one type or another. In addition, objects of change are generally functions or correlates of other associated changes. In kinematics, for instance, a change in position is associated with a change in time.

Hewson and Posner constructed change networks that involved diagrams consisting of the objects of change, the initial and final stages, the causes of change, and linkages to correlated objects of change. This network was initially presented to students using objects from the student's real-life experiences, like the color of blue jeans, the weight of a person on a diet, and the amount of money in a bank account. Those examples were intended to serve as a link, bridging the gap between the student's existing knowledge and the subsequent instruction in physics. Students were told that many of the major ideas of physics could best be understood in terms of this basic "change" model and that they should try to relate the components of the model to each of the objects of change they would study in their physics course.

It is apparent that the change networks developed by Hewson and Posner formed a framework around which students could organize their knowledge of physics. The basic network emphasized qualitative entities and relationships and was general enough to incorporate many of the major ideas of physics. It was hoped that as the basic model was assimilated, it could provide the format into which subsequent information could be organized and also could aid problem solving by serving as a set of expectations that could guide and direct the students in their investigation and selection of appropriate data. Hewson and Posner indicated that some students benefitted from their instructional materials in the intended manner, while others found it difficult to relate the change networks to the course content. The authors speculated that some students may require a model that provides more explicit associations between physics concepts and their use in problem solving.

Curts (1985) developed an introductory statistics course for beginning college level biology students that used concepts from exploratory data analysis (Tukey, 1977) to examine and model biological data sets. Curriculum units were introduced by considering data sets appearing in biological and medical journals. Students were required to read the journal articles and then were assigned several questions pertaining to both the article and the included data set. In a typical article, students were asked to (a) clearly identify the problem the researchers were investigating or attempting to answer, (b) describe the independent and dependent variables, (c) discuss the manner used to obtain the data, and (d) point out the author's conclusions.

After students had developed a qualitative understanding of the problem being addressed in the article, they were presented with data analysis techniques appropriate for the examination of the corresponding data set. Students then used those techniques to examine the relationships between variables and to construct mathematical models to summarize the behavior of the data. Students were asked to support, reject, or qualify conclusions reached by the authors of the articles and to defend their own conclusion. Particular attention was paid to the effects of outliers and to possible empirical explanations for the outliers. In short, a collection of activities was organized around the problem situation presented in the journal article.

Particular care was exercised in the selection of journal articles that were used to introduce each curriculum unit. In addition to having appeared in biological or medical journals, articles were selected because of (a) simplicity and brevity, (b) inclusion of the data base, (c) ease at which the data could be manipulated or organized in accordance with the exploratory data techniques to be taught, (d) the use of concepts mastered in prior units, and (e) ability of the articles to prepare students for future units.

The journal articles and their use by Curts clearly incorporated features of a story shell. The article provided a realistic problem situation to be investigated by the student, closely approximating a real professional situation. Students were to attack the analysis of data like a detective, searching for patterns that might provide meaningful relationships between the variables. The problem-solving atmosphere and detective work led to discussion and arguments among the students about possible solutions and their validity. The emphasis was on model building and interpretations and not on the application of standard formulas.

In addition to providing a realistic empirical situation for investigation, the journal article and associated data set made the introduction of data analysis techniques psychologically meaningful. The problem provided a microworld or schema to which new information could be adjoined, integrated, and interpreted.

The instructor of a course has a wealth of experiences with which to interpret and provide meaning to new concepts. Students are just beginning to accumulate these schema-building experiences. The story shell as used by Curts provided a common experience or medium to which both student and teacher could relate. This undoubtedly enhanced meaningful communication about the concepts and ideas that were being developed.

## Constructive Teaching Units

Rosalind Driver and her associates at the Centre for Studies in Science and Mathematics Education at the University of Leeds in England have embarked on an ambitious curriculum development project in science based on constructivist notions of learning. Their view of curriculum "is not a body of knowledge or skills but a programme of activities from which knowledge or skills can possibly be acquired or constructed" (Driver & Oldham, 1985, p. 10). The units now being developed and tested are probably similar to story-shell units; however, their emphasis is more on the role of the teacher with respect both to the development of units and to subsequent instruction. Implicit in their position is the view that, if teachers have a coherent grasp of the subject, content will be transmitted in an effective way to students. Also implicit is that all curriculum units are problematic.

When we accept the notion that the curriculum defines the program of activities from which knowledge or skills can possibly be constructed and acknowledge that what is constructed by any individual depends to some extent on what is brought to the situation, we make the suitability and effectiveness of selected learning activities an empirical problem. Teachers must determine whether students are effectively assimilating the experiences they are given. For this reason curriculum development has to be an empirical reflexive approach.

The general model being used by Driver and Oldham (1985) for the development of new curriculum materials is given in Figure 2. This figure indicates four components that influence the development and design of curriculum. The first and most conventional one is the decision on content. Here, experts specify experiences to which students should be exposed and suggest what ideas students may construct from those experiences.

Second, curriculum design is influenced by ideas that students bring to the learning situation. Driver and Oldham identified students' prior knowledge by analyzing data from a national sample of students' responses to open-ended written questions in the topic area.

Third, Driver and Oldham argued that knowing where students are starting from is not, by itself, enough to plan curricular activities. Curriculum development must make use of constructivist notions of learning. Conceptual change occurs as the result of active processing of information and attempts by the learner to impart a meaning to this information.

The last information comes from practical knowledge of students in school and classroom settings: how to organize a group of approximately 30 people to do something in about one hour; how to present a problem in an interesting way to a group of 14-year-olds; how to deal with the usual constraints of time, resources, furniture, and space. If these types of issues are not addressed in the curriculum design it is believed that long term implementation is not probable.

In addition, Driver and Oldham have found it both necessary and useful to have a model for a constructivist teaching sequence. This model is illustrated in Figure 3.

The sequence comprises five phases: orientation, elicitation, restructuring, application, and review.

The orientation phase is designed to give students the opportunity to develop a sense of purpose and motivation for learning the topic. Then instruction moves to the elicitation phase in which pupils make their ideas explicit.

This is followed by a restructuring phase which includes a number of different aspects. Once the students' ideas are "out in the open," clarification and exchange occurs through discussion (Gall and Gall, 1976; Hornsey and Horsfield, 1982). In this way, the meanings students construct and the language they use may be "sharpened up" in comparison with different, and possibly conflicting, views of others (Nussbaum & Novick, 1982; Rowell & Dawson, 1983; Stavy & Berkovitz, 1980), and inadequacies may be pointed out (Strike & Posner, 1982). The exchange of views and perspectives may lead to spontaneous challenge and disagreement among students. Alternatively, both subtle and explicit attempts may be made by the teacher to promote conceptual conflict through

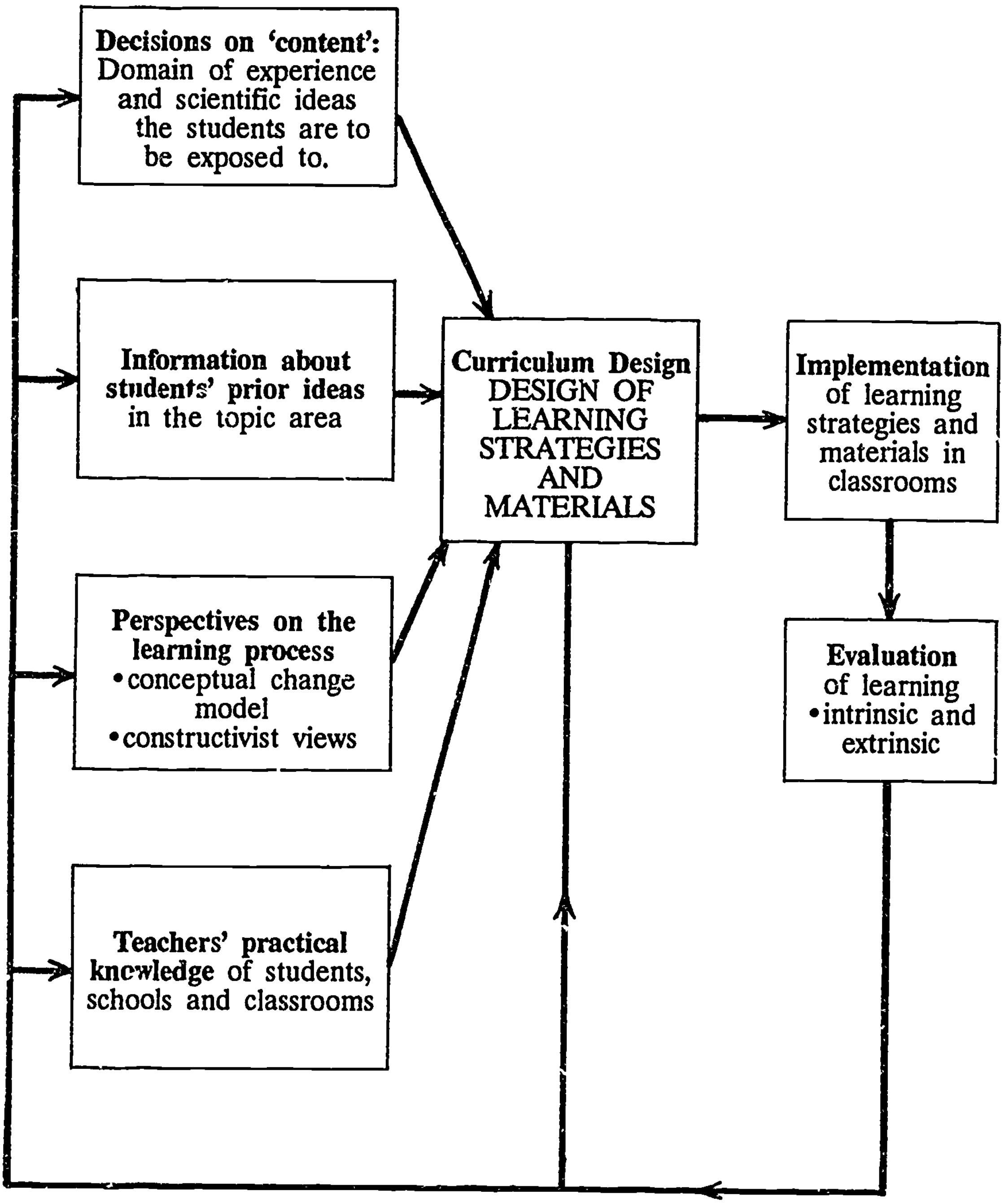



Figure 2. A constructivist model for curriculum development.
(Driver & Oldham, 1986, p.13)

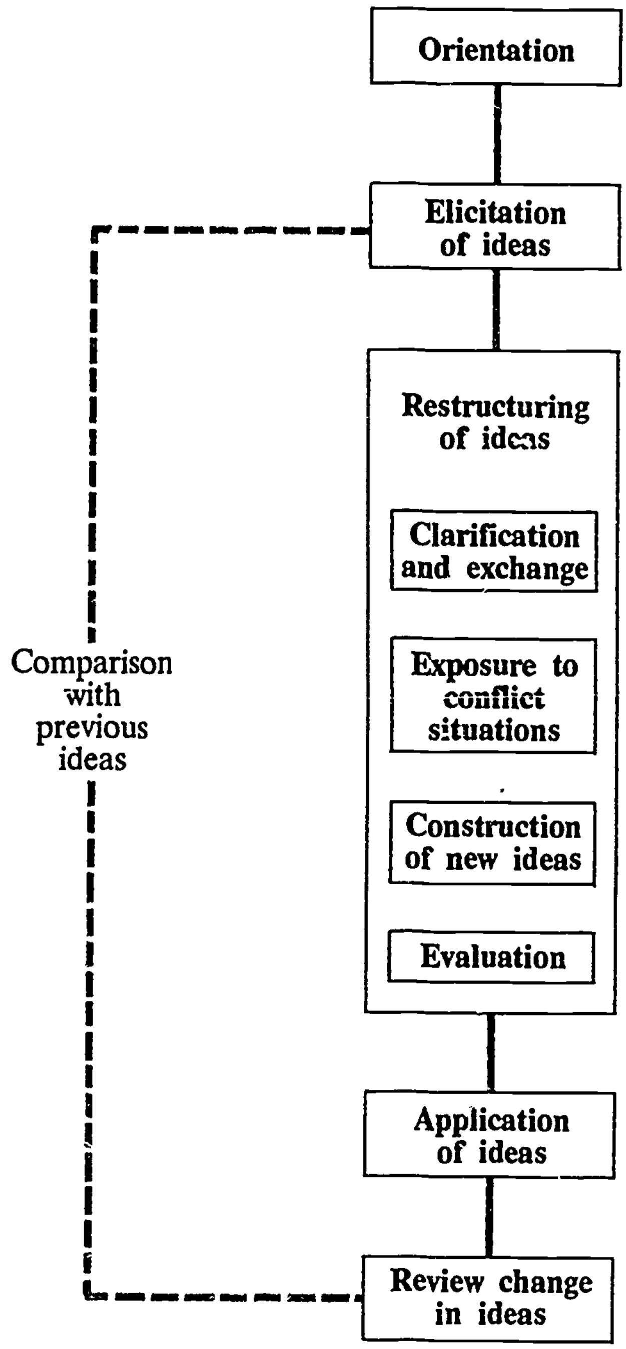



Figure 3. A constructivist teaching sequence.
(Driver & Oldham, 1986, p.18)

the use of a disconfirming or surprise demonstration. This type of discourse gives students an opportunity to develop an appreciation and tolerance for different notions used to explain or describe the same phenomenon.

From here pupils move into the evaluation of alternative ideas, possibly including the scientific one if they have suggested it. These ideas may be tested against experience, either experimentally or by thinking through their implications. Often, students can be given the chance to be imaginative in devising ways of testing these ideas (Nussbaum & Novick, 1982; Osborne, 1981). Different groups of students may test different ideas and report their findings to the whole class. As a result of this dialog and discussion, students may feel dissatisfied with their existing conceptions and, hence, receptive to change (Posner, Strike, Hewson, & Gertzog, 1982).

Some students may have constructed a reasonable scientific view from prior experiences; thus, the scientific view may have been presented and tested along with a range of alternative conceptions. Whether or not this has happened, the teacher must present and explain that particular view at some point and provide opportunities for pupils to construct meanings by empirical tests and language activities. The appearance or the discovery of a scientific view and the chance for students to begin to make sense of it occur at various points in the restructuring phase.

In the application phase students are given the opportunity to use their developed ideas in a variety of familiar and novel situations. In this manner new conceptions are consolidated and reinforced by extending the contexts within which they are seen to be useful.

In the final review phase of the sequence, students are invited to reflect on how their ideas have changed by comparing their thinking now with that at the start of the unit.

In summary, these two examples--story shell and constructivist teaching units--reflect the current attempts to rethink the curriculum engineering problem in light of current research from cognitive science.

## CONCLUSIONS AND PRINCIPLES FOR CURRICULUM ENGINEERING

The research results that have been discussed in this paper tend to support the following conclusions:

1. The use of generic schema, developed over long periods of time and by continual exposure to related events and exemplars, promotes both problem-solving skill and recall of textual material. These schema appear to guide, organize, and direct both the search for a problem solution and the retrieval of expository or story details.

2. The encoding, comprehension, and retrieval of information is aided when material is presented in a form that has structure and when the student is cognizant of that structure. In particular, these processes are facilitated when the information can be assimilated into an existing schema of the learner.

3. When information is presented in a story or expository text, the transitions and states leading directly to the goal or objective are remembered best. This critical path is probably related to a generic story schema which directs encoding. This story schema appears to be a part of a more general problem-solving schema, having as a primary component the cause and effect relation.

4. Although students appear to make use of cause and effect relations in encoding stories and text, and in solving problems not requiring specific content knowledge, they have difficulty with conditions regulating the use of specific mathematical properties. Failure to recognize these conditions often results in the development of buggy algorithms and the inappropriate application of mathematical theorems.

5. Problem-solving ability and encoding of information are enhanced when schemata are interrelated and form a hierarchical arrangement analogous to the way knowledge is used.

More specifically, just what do these conjectures have to say about the design of curriculum? We believe the following principles to be direct consequences of the research that has been summarized in this paper:

Principle 1 Conceptual strands should be specified.

The main generic schemata (i.e., measurement, mappings, proportionality) that we wish to develop in school children must be identified, and a spiral curriculum built around those conceptual strands (Vergnaud, 1983, calls these conceptual fields). These strands should be selected because of their generality and ability to subsume more specialized components of the curriculum deemed desirable for the development of problem-solving ability and quantitative reasoning.

Principle 2. The strands should be segmented into curriculum units that take two to four weeks to teach.

Students should be expected to construct meanings, interrelate concepts and skills, and use those meanings in a variety of problem situations. One cannot learn interrelationships by studying concepts, skills, and problems in isolation.

Principle 3. Students should be exposed to the major conceptual strands as they arise naturally in problem situations.

Ideas are best introduced when students see a need or a reason for their use. Promoting the development of integrated schemata requires an integrated curriculum. There can be little justification in maintaining a curriculum separated into, for example, arithmetic, algebra, geometry, and trigonometry. Also, a new look must be taken at what mathematics young children and adolescents can learn at various ages.

Principle 4. Each curriculum unit should tell a story.

Each curriculum unit should have a beginning and an end and culminate in some knowledge deemed beneficial by the student. The story setting should (a) review the background material necessary to comprehend the unit, (b) make clear to the student the goals of the unit, and (c) describe why the goals are important and worthwhile. The transition from the initial state to the goal state should be clear to the students, and the students should be actively involved in achieving the goal. Students should feel the excitement of investigation and the thrill of obtaining their own solutions. Ideally, the goal state should suggest further development of related conceptual strands.

Principle 5. The activities within each unit should be related to how students process information.

Each unit should provide review of prior concepts and skills and lay foundations for concepts and skills to be learned later. Activities used to teach algorithms should differ from those used to teach problem solving, and activities requiring assimilation should differ from those requiring accommodation. For example, students might be addressed as a large group when being exposed to information and work in small groups when inventing, proving, or applying. Assimilation may require exercises requiring little prior knowledge, while accommodation may demand a dissimilar array of problem situations involving varying cognitive structures. A higher degree of teacher-imposed structure and control may be desirable for lower-level cognitive outcomes, while a greater degree of group autonomy may aid higher-level cognitive outcomes.

Principle 6. Every unit should have students involved in inventing, abstracting, proving, and applying mathematics.

It is doing mathematics, analogous to a "hands-on" experience in the natural and physical sciences, that contributes to the formation of rich knowledge structures.

Principle 7. Students should be given ample opportunities to work with open-ended problems.

Situations requiring an action or a change in state might be presented students with the view toward soliciting varied student reaction. Evidence from the expert-novice investigations suggests that many students are more comfortable working with this bottom-up mode of processing. After all, most problems in the real-world are of this type.

Principle 8. Self-regulatory or metacognitive mechanisms should be continually stressed.

Good problem solvers appear to use such general heuristics as planning ahead, looking for qualitative or alternative representations, and monitoring problem-solving efforts. Development of a general problem-solving schema incorporating these and other general thinking skills should be of tremendous benefit in many fields of knowledge and in every day life.

If operations within knowledge structures resemble those within production systems, it is difficult to imagine learning not accompanied by active cognitive activity on the part of the learner. Learners construct their own cognitive structures. Only through a great deal of practice and reflection does organization of schemata become proficient. When a student generates a bit of knowledge on the way to a problem solution, the action taken and the conditions that made the action possible form an "indelible print" or production in memory, serving to expand, integrate or relate the associated schemata. If these same or similar conditions arise in the future, the action component of the production may bring to mind the associated bit of knowledge. And problem solving becomes more efficient.

Principle 9. Curriculum units should always be considered as problematic.

All curriculum sequences need to be adopted and modified in light of what knowledge the students bring to the unit and the context in which instruction takes place.

Principle 10. The teacher's role is not as a dispenser of information but as an instructional guide.

The role of the teacher and the nature of instruction differ radically as a result of these considerations. The implications of this principle are explored in more depth in the next chapter.

## References

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J. R., Greeno, J. G., Kline, P. J., & Neves, D. M. (1981). Acquisition of problem solving skill. In J. R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 191-230). Hillsdale, NJ: Erlbaum.

Anderson, R. C., Spiro, R. J., & Anderson, M. C. (1978). Schemata as scaffolding for the representation of information in connected discourse. American Educational Research Journal, 15, 433-440.

Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. Psychological Review, 86, 124-140.

Barnett, J. E. (1984). Facilitating retention through instruction about text structure. Journal of Reading Behavior, 16, 1-13.

Berliner, D. C. (1982). The executive functions of teaching. Paper presented at the meetings of the American Educational Research Association, New York City.

Bishop, A. J. (1982). Implications of research for mathematics teacher education. Journal of Education for Teaching, 8, 118-135.

Black, J. G., & Bower, G. H. (1980). Story understanding as problem solving. Poetics, 9, 176-193.

California Statewide Mathematics Committee. (1972). The second strands report, mathematics framework for California public schools: Kindergarten through grade eight. Sacramento, CA: California State Department of Education.

Carpenter, T. P., Bebout, H. C., & Moser, J. M. (1985, April). The representation of basic addition and subtraction word problems. Paper presented at the annual meeting of the American Educational Research Association. Chicago, Illinois.

Carpenter, T. P., Corbitt, K. M., Kepner, H. S., Lindquist, M. M., & Reys, R. E. (1981). Results from the Second Mathematics Assessment of the National Assessment of Educational Progress. Reston, VA: National Council of Teachers of Mathematics.

Carpenter, T. P., Fennema, E., & Peterson, P. (1985). Cognitively guided instruction: The application of cognitive and instructional science to mathematics curriculum development. Paper presented at the International Conference on Mathematics Education, University of Chicago.

Carpenter, T. P., & Moser, J. M. (1982). The development of addition and subtraction problem-solving skills. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), Addition and subtraction: A cognitive perspective (pp. 9-24). Hillsdale, NJ: Lawrence Erlbaum.

Case, R. (1978). Implications of developmental psychology for the design of effective instruction. In A. M. Lesgold, J. W. Pellegrino, S. D. Fokkema, & R. Glasser (Eds.), Cognitive psychology and instruction (pp. 441-463). New York: Plenum.

Chi, M. T. H., Claser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence, Vol. 1 (pp. 7-75). Hillsdale, NJ: Erlbaum.

Clement, J. (1979). Mapping a student's causal conceptions from a problem-solving protocol. In J. Lochead & J. Clement (Eds.), Cognitive process instruction. Research on teaching thinking skills (pp. 133-146). Philadelphia: Franklin Institute Press.

Curts, J. B. (1985). Teaching college biology students the simple linear regression model using an interactive microcomputer graphics software package. (Doctoral dissertation, University of Wisconsin-Madison, 1985). Dissertation Abstracts International, [need volume, page no.] (University Microfilms No. 85-00,000)

Driver, R., & Oldham, V. (1985, August). A constructivist approach to curriculum development in science. Paper presented at the meeting of the British Educational Research Association. Sheffield.

Fremont, H. (1967). New mathematics and old dilemmas. The Mathematics Teacher, 19, 91-94.

Gagné, R. M. (1965). The conditions of learning. New York: Holt,Rinehart and Winston.

Gall, M.D., & Gall, J.P. (1976). The discussion method. In N. L. Gage (Ed.), The psychology of teaching methods: Seventy-fifth yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press.

Gardner, H. (1985). The mind's new science. New York: Basic Books.

Glaser, R. (1984). Education and thinking, the role of knowledge. American Psychologist, 39, 93-104.

Greeno, J. C. (1980). Psychology of learning, 1960-1980. One participant's observations. American Psychologist, 35, 713-728.

Greeno, J. D., & Bjork, R. A. (1973). Mathematical learning theory and the new "mental forestry." In P. H. Mussen & M. R. Rozenzweig (Eds.), Annual Review of Psychology (Vol. 24). Palo Alto, CA: Annual Reviews Inc.

Harvey, J. G., McLeod, D. B., & Romberg, T. A. (1970). The task analysis for developing mathematical processes, arithmetic book 2: Writing mathematical sentences (Working Paper No. 49). Madison, WI: Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin.

Hewson, P. H., & Posner, G. J. (1984). The use of schema theory in the design of instructional materials: A physics example. Instructional Science, 13, 119-139.

Hornsey, M., & Horsfield, J. (1982). Pupils' discussion in science: A strategem to enhance quantity and quality. School Science Review, 63, 763-767.

Kantowski, E. L. (1975). Processes involved in mathematical problem solving (Doctoral dissertation, University of Georgia, 1975). Dissertation Abstracts International, 36, 2734A. (University Microfilms No. 75-23,764).

Larkin, J. H. (1979). Processing information for effective problem solving. Engineering Education, 70, 285-288.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Models of competence in solving physics problems. Cognitive Science, 4, 317-345.

Larkin, J. H., & Rainard, B. (1984). A research methodology for studying how people think. Journal of Research in Science Teaching, 21, 235-254.

Lindvall, C. M., & Bolven, J. O. (1967). Programmed instruction in the schools: An application of programming principles in Individually Prescribed Instruction. In P. C. Lange (Ed.), Programmed instruction: Sixty-sixth yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press.

Mandler, J. M. (1984). Stories, scripts, and scenes: Aspects of schema theory. Hillsdale, NJ: Lawrence Erlbaum.

Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. Reading Research Quarterly, 16, 72-103.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81-97.

National Assessment of Educational Progress. (1983). The Third National Mathematics Assessment: Results, trends and issues (Advance Copy). Denver: Education Commission of the States.

Nussbaum, J., & Novick, S. (1982). Alternative frameworks, conceptual conflict and accommodation: toward a principled teaching strategy. Instructional Science, 11, 183-200.

Osborne, R. J. (1981). Toward solutions: The work of the Physics Action Research Group. Waikato, N.Z.: Science Education Research Unit, University of Waikato.

Paige, J. M., & Simon, H. A. (1966). Cognitive processes in solving algebra word problems. In B. Kleinmuntz (Ed.), Problem solving. New York: Wiley.

Papert, S. (1980). Mindstorms: Children, computers, and powerful ideas. New York: Basic Books.

Phillips, D. C., & Soltis, J. F. (1985). Perspectives on learning. New York: Teachers College Press.

Posner, G., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. Science Education, 66, 211-227.

Rabinowitz, M., & Mandler, J. M. (1983). Organization and information retrieval. Journal of Experimental Psychology, 9, 430-439.

Romberg, T. A. (1970). Curriculum, development and research. In M. F. Rosskopf (Ed.). The teaching of secondary school mathematics. Washington, DC: National Council of Teachers of Mathematics.

Romberg, T. A. (1983). A common curriculum for mathematics. In G. D. Fenstermacher & J. I. Goodlad (Eds.), Individual Differences and the Common Curriculum: Eighty-second yearbook of the National Society for the Study of Education (pp. 121-158). Chicago: University of Chicago Press.

Romberg, T. A. (1984, August). Curricular reform in school mathematics: Past difficulties, future possibilities. Paper presented at the Fifth International Congress on Mathematical Education, Adelaide, Australia.

Romberg, T. A. (1985). School mathematics: Options for the 1990s. Chairman's report of a conference. Washington, DC: U.S. Government Printing Office.

Romberg, T. A., & Carpenter, T. P. (1985). Research on teaching and learning mathematics. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd Edition). New York: Macmillan.

Romberg, T. A., & Price, G. (1983). Curriculum implementation and staff development as cultural change. In G. Griffin (Ed.), Staff development: Eighty-second yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press.

Rowell, J. A., & Dawson, C. J. (1983). Laboratory counter examples and the growth of understanding in science. European Journal of Science Education, 5, 203-216.

Rumelhart, D. E. (1977). Understanding and summarizing brief stories. In D. LaBerge & J. Samuels (Eds.), Basic processes in reading: Perceptions and comprehension. Hillsdale, NJ: Erlbaum.

Shavelson, R. J. (1974). Methods for examining representations of a subject-matter structure in a student's memory. Journal of Research in Science Teaching, 11, 231-249.

Simon, H. A. (1978). Information-processing theory of human problem solving. In W. K. Estes (Ed.), Handbook of learning and cognitive processes (Vol. 5, pp. 271-295). Hillsdale, NJ: Erlbaum.

Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In R. S. Siegler (Ed.), Children's thinking: What develops? (pp. 325-348). Hillsdale, NJ: Lawrence Erlbaum.

Slater, W. H., Graves, M. F., & Piche, G. L. (1985). Effects of structural organizers on ninth-grade students' comprehension and recall of four patterns of expository text. Reading Research Quarterly, 20, 189-201.

Spilich, G. J., Vesonder, G. J., Chiesi, H. L., & Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain information. Journal of Verbal Learning and Verbal Behavior, 18, 275-290.

Stavy, R., & Berkovitz, B. (1980). Cognitive conflict as a basis for teaching qualitative aspects of the concept of temperature. Science Education, 64, 679-692.

Stein, N. L., & Trabasso, T. (1982). What's in a story: An approach to comprehension and instruction. In R. Glaser (Ed.), Advances in instructional psychology (Vol 2). Hillsdale, NJ: Erlbaum.

Stephens, W. M. (1983). Mathematical knowledge and school work: A case study of the teaching of developing mathematical processes. (Doctoral Dissertation. University of Wisconsin-Madison, 1982). Dissertation Abstracts International, 43, 1414A. (University Microfilms No. 82-16,267)

Stephens, W. M. (1984, April). Public policy, the social context of schools, and mathematics education. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Strike, K. A., & Posner, G. J. (1982). Conceptual change and science teaching. European Journal of Science Education, 4, 231-240.

Sweller, J. (in press). What do students learn while solving mathematics problems? Journal of Educational Psychology.

Tall, D. (1985). Using computer graphics as generic organizers for the concept image of differentiation. In L. Streefland (Ed.), Proceedings of the Ninth International Conference for the Psychology of Mathematics Education (pp. 105-110). Utrecht, The Netherlands.

Tukey, J. W. (1977). Exploratory data analysis. Reading, MA: Addison-Wesley.

Tyler, R. W. (1931). A generalized technique for constructing achievement tests. Educational Research Bulletin, 8, 199-208.

Vergnaud, G. (1983). Multiplicative structures. In R. Lesh & M. Landau (Eds.), Acquisition of mathematics concepts and processes. New York: Academic Press.

Wittmann, E. (1984). Teaching units as the integrating core of mathematics education. Educational Studies in Mathematics, 15, 25-36.

Woods, D. R., & Crowe, C. M. (1984). Characteristics of engineering students in their first two years. Engineering Education, 74, 289-295.

## Chapter 16

## PSYCHOLOGY IN THE MATH CLASS: COMMENTS ON CHAPTERS 12-15

Gary Glen Price

This is a reaction to the preceding four chapters, which all consider implications for school mathematics that can be drawn from psychology--particularly, cognitive science. Following a summary of the few differences and several similarities among the chapters, I provide an interpretation of each. I conclude with my own reflections.

The chapters differ in several distinct ways. The chapters by Greeno and by Hatano and Inagaki are primarily psychological and secondarily mathematical. Greeno's knowledge structure program and Hatano and Inagaki's cognitive Berlynean theory are psychological perspectives that are applied to, but not tied to, mathematics education. Fischbein's chapter is obviously well tied to recent developments in cognitive psychology, but its overall emphasis is squarely in the tradition of introspective psychologizing by mathematicians--the tradition of Jules Henri Poincaré (1913), George Polya (1954a, 1954b), and Seymour Papert (1980). Romberg and Tufte's chapter is similarly well tied to recent developments in cognitive psychology, but its overall emphasis is on curriculum development in mathematics education.

### SIMILARITIES

Despite the differences among these chapters, there are also several striking similarities.

Criticism of current pedagogy. The authors are nearly unanimous in their criticism of current pedagogy and in their optimism about the feasibility of improvement. Greeno surmised that children's lack of understanding may result from "a perverse method of instruction." Hatano and Inagaki wrote, "Teachers' conventional methods of motivating students, such as grading or reward . . . may prevent learners from understanding things deeply." Romberg and Tufte wrote, "The fragmentary nature of many existing mathematics programs leaves the student with an almost total inability to apply mathematics in any but routine situations and, in fact, with very little experience with mathematical thought itself." Fischbein is the exception: He did not criticize teaching; nor, however, did he mention teachers.

Positive appraisal of children's capabilities. These authors share optimism that most children can do considerably more mathematics than they presently achieve. Clearly, they emphasize what children can do, not what they cannot do. Gone is a former truism of the psychometric tradition--that greater-than-average intellectual ability is required to do tasks such as those that mathematicians do (McNemar, 1964; Osler & Fivel, 1961; Osler & Trautman, 1961). Gone also are claims that children's failure to attain a particular number concept (Piaget, 1952) or stage of cognitive development (Piaget, 1966) prevents them from "doing mathematics."

Prior knowledge affects acquisition of new knowledge. All four chapters pay heed to the enabling influence of well-conceived, well-organized prior knowledge and the disabling influence of misconceived, badly organized prior knowledge. Students' mental representations are very important. This perspective is a departure from the recent past, which treated lack of knowledge as a problem but underestimated the inertial impediment of misconceptions. One of Romberg and Tufte's central theses is that "new information is fitted or assimilated into existing cognitive structures." As they put it, "What we learn, and in fact what we are capable of learning, depends on the mental models each of us has developed." This new truism about the inertia of prior knowledge offers new insight into John Locke's (1699/1964) frequent reference to the tutor's task as being one of "laying foundations."

Changes in students' representations. A corollary of Romberg and Tufte's thesis is that a central purpose of educators is to induce changes in students' mental representations. Clearly Greeno (1976, chapter 12) has long been involved in the knowledge structure program, in which cognitive models have been treated as instructional objectives. The comprehension activity that Hatano and Inagaki seek to encourage is a process by which students "build an enriched and coherent representation." Fischbein describes analogic, intuitive representations as "the way we think," a set of processes that needs to be coordinated with, but not stifled by, formal meanings and formal implications. Romberg and Tufte take as their "major thesis . . . that the mathematics curriculum should reflect the way knowledge is optimally organized in the semantic and factual knowledge base." In support of that view, Greeno considered recent research to suggest strongly that "general principles and concepts play a significant role in organizing information and procedures that the child acquires."

The expert as a point of reference. One could accept that mental representations are important, and that educators should seek to provoke changes in them, but still not know which representations should be fostered. The authors of these chapters regard experts as a normative source: It is desirable to replace novices' representations with experts' representations. This could fairly be called the think-as-experts-think curriculum. When Greeno adapted Smith's (1983) framework to characterize procedures that are learned by students in mathematics instruction, he demonstrated the utility of the framework by using it to describe differences among children, expert mathematicians, and unschooled domain experts. The initial assumption of Hatano and Inagaki is that "one of the major goals of education is the acquisition of a well-organized body of knowledge." This heightened interest in the organization of knowledge stems from its identification as a distinguishing attribute of experts in a domain. Fischbein participates in the granddaddy of think-as-experts-think curricula--the previously mentioned tradition of introspective psychologizing by mathematicians. Romberg and Tufte justify the stressing of such general heuristics as planning ahead on the grounds that good problem solvers appear to use them.

Getting beliefs out into the open. Consonant with the purpose of changing students' mental representations, the authors have emphasized the desirability of eliciting explicit, representation-revealing statements from students. Greeno argues that we need to create environments in which students learn to ask meaningful questions and compose arguments. Hatano and Inagaki believe that situations in which a student must "make explicit what he/she knows only implicitly," are likely to induce discoordination, a type of cognitive incongruity, a precondition of comprehension activity. Fischbein claims that "special [presumably discursive] exercises should be devised to train students to analyze concepts and definitions in order to distinguish clearly the properties imposed by definitions and those suggested by intuitive components." Romberg and Tufte cite the elicitation phase used by Driver and Oldham (1985), in which students' ideas are "out in the open."

## JAMES GREENO

Greeno serves in the first part of his chapter as an intellectual historian of the knowledge structure program. Greeno is both an analyst of what the program has accomplished and a visionary who identifies paths by which the program can realize new promises. As one of the principal architects of the program, he is eminently well qualified to be both.

### Essence of the "Knowledge Structure Program"

The distinguishing feature of the knowledge structure program has been the way in which it frames instructional objectives. Like the behavioral objectives of other programs, the instructional objectives of the knowledge structure program concern individuals--multiple individuals, perhaps, but individuals nonetheless. Unlike behavioral objectives, its instructional objectives are models of cognition built from the theoretical constructs of cognitive psychology--constructs like production systems, schemata, and semantic networks. Greeno gives a brief history of these constructs, as well as some of the methods that have been associated with them (e.g., protocol analysis, simulation modeling). When models of cognition are used as instructional objectives, a double meaning is given to the term model, because exemplary structures and processes singled out as instructional objectives are model models.

This endeavor requires not only apt characterization of cognition, but judicious selection of exemplary cognition. As Greeno remarked, the dominant role of this research has been to understand knowledge that is required for successful performance of school tasks. A program centered on success in schools as they presently are may seem an unlikely means of sparking educational change. Greeno clearly understands that, because he devoted the concluding section of his chapter to the inculcation of "abilities to think mathematically and cognitive resources for reasoning in situations other than classrooms."

## Accomplishments of the "Knowledge Structure Program"

The first examples that Greeno cites to illustrate the accomplishments of the Knowledge Structure Program (Brown & Burton, 1980; Sleeman, 1984) fit a category that he previously termed models of knowledge (Greeno, 1978). They are models of knowledge because they detail knowledge that underlies performance. It is possible but not certain that such explications of important knowledge will make it available to some students who might not otherwise have discovered or constructed it. However, models of knowledge fail to fit another category that Greeno identified in 1978--models of learning. Models of learning detail the transitions through which novices pass en route to expertise. Thus, models of learning have more direct implications for educational practice than do models of knowledge. Considering that a decade has passed since Greeno discussed the need for models of learning, it is significant that he is now able to cite two examples (Anderson, 1983; Anderson, Boyle, Farrell, & Reiser, 1984).

An important development to which Greeno refers obliquely is the shift from general knowledge structures and strategies, such as those represented in Newell and Simon's 1972 work, to a concentration on domain-specific aspects, such as work concerning school mathematics. The knowledge structure program has contributed to a profound change in educational folklore--a change from the relative neglect of domain-specific knowledge to intense interest in it. Not so long ago, phenomena thought to be domain-general (e.g., Piagetian concepts) held center stage in mathematics education. In studies of experts, however, the weight of evidence has forced educators to reckon with the importance of domain-specific knowledge. Some of the recent research has shown that, even within a domain, some knowledge, like the low-level mathematics of unschooled experts, is context-bound (e.g., Acioly & Schliemann, 1986).

Greeno surveys evidence of children's ability to reason intelligently with mathematical ideas. He concludes that children are capable of much, and he is led thereby to blame current educational practice for children's failure to use school to blame current educational practice for children's failure to use school mathematics outside of school. Greeno is clearly optimistic, however, that more ambitious goals are feasible for mathematics education.

## New Directions for the Knowledge Structure Program

Greeno's chapter ends with a visionary call for the inclusion of two new quests in the knowledge structure program. The first is to better understand why mathematics learned in school so seldom transfers to individuals' reasoning and problem solving in practical, everyday situations. Clearly, for a fortunate few, mathematical knowledge is a valuable resource used in everyday reasoning. Why are there so few of these beneficiaries of mathematical resources? This is an important question whose answers should be rich with educational implications. In posing the question this way, Greeno has set aside the conventional question of whether students learn the mathematics that schools teach. Greeno initiates this quest with several conjectures that are both plausible and amenable to study, so we should know more after they have been tested.

The second new quest is to develop concepts of schooling and of mathematics education that are congenial to the kind of deep conceptual growth needed to transfer school mathematics to nonschool settings. This involves rethinking the goals of mathematics education. Referring to Kitcher's (1984) five components of a mathematical practice, Greeno shows that current instruction targets only the last two components. He acknowledges that going beyond this point will "take cognitive research into territory that is almost entirely uncharted."

Greeno concludes by identifying promising, innovative approaches. The features that guided Greeno's selection (placement of students into active, knowledge-constructing situations and collaborative mathematical work) are also prized by Hatano and Inagaki.

## GIYOO HATANO & KAYOKO INAGAKI

Hatano and Inagaki have addressed themselves to the question of why thoroughgoing comprehension (nattoku) is so rare. The Japanese word nattoku is translated as the achievement of having found satisfactory explanations of why a given rule is valid or why a given procedure works. Hatano and Inagaki attribute the rareness of nattoku to the rareness of comprehension activity. The question then shifts to the reasons why comprehension activity is so rare. To answer this question, Hatano and Inagaki have developed a theory of motivation for comprehension, which is rich with educational implications.

The basis for nattoku is "an enriched and coherent representation." To build this basis, children must engage in directed, persistent (time-consuming) comprehension activity, which includes activities like generating inferences, checking the plausibility of inferences, and coordinating pieces of old and new information. Unfortunately, these activities require effort that children are not commonly motivated to expend. If educators wish to increase students' engagement in comprehension activity, they should understand more about what motivates such activity when it does occur. They should also understand why certain circumstances --too often school circumstances--fail to motivate comprehension activity. One reason for lack of such motivation is that lack of nattoku is seldom a serious deficit. As Hatano and Inagaki note, "Lack of 'nattoku' becomes a serious deficit only when unusual, novel problems are posed." It thus seems to be the lot of the educator to devise situations in which the motivation for comprehension exceeds that which exists in everyday contexts.

Hatano and Inagaki advise against manipulating motivation for comprehension directly. Instead, they would have educators design situations in which intrinsic motivation for comprehension will come into play. To make this feasible, more needs to be known about motivation for comprehension, which, they argue, is different from achievement motivation. Neither the research literature on motivation nor the research literature on cognition has examined motivation for comprehension, and they seek to end that neglect.

## Hatano and Inagaki's Cognitive Berlynean Theory

Hatano and Inagaki have developed a theory of motivation for comprehension, which they describe as an elaboration and extension of Berlyne's (1960, 1965) theory of epistemic behavior. Their theory retains Berlyne's focus on intrinsic motivation for knowing, his description of the conditions under which motivation for knowing is aroused, and his prescriptive suggestions about how to motivate students.

Hatano and Inagaki have borrowed from Berlyne's construct of epistemic curiosity, which functions to motivate comprehension activity. To Berlyne, epistemic curiosity was an uncomfortable state from which one was driven to seek relief. Hatano and Inagaki eschew Berlyne's "discomfort drive state," but they do not elaborate on the reasons why epistemic curiosity produces sustained comprehension activity.

Cognitive incongruity is a state of awareness that provokes epistemic curiosity. It is akin to Berlyne's notion of conceptual conflict, but without his view that it is an uncomfortable state. Cognitive incongruity usually occurs when a person becomes aware that his or her comprehension is inadequate. Hatano and Inagaki identify three types of experience in which persons become aware of inadequacies in their comprehension--surprise, perplexity, and discoordination.

Students must monitor their own comprehension before they can recognize inadequacies in it. Therefore, comprehension monitoring activity is a prerequisite of cognitive incongruity. Recent research has shown comprehension monitoring activity to be a limited resource that educators cannot take for granted (e.g., Glenberg & Epstein, 1985; Markman, 1979). Drawing implications from that research, Hatano and Inagaki provide suggestions for fostering comprehension monitoring.

According to the theory, cognitive incongruity does not inevitably provoke epistemic curiosity. Whether it does so depends on two fundamental beliefs. First, a person must believe in his or her own capability to comprehend; comprehension activity appears futile to a person who lacks confidence. Second, a person must believe that the knowledge domain containing the cognitive incongruity is important enough to merit the effort of comprehension activity. Herein lies an educational problem identified by the theory, but not solved by it. Hatano and Inagaki do claim that some social milieus, such as that provided by dialogical interaction (Miyake, 1986), can raise students' interest. The theory fails to explain why this is so, but it does explain why some conventional methods of motivating students can be counterproductive.

## Strategies for Inducing Cognitive Incongruity

The theory's clearest educational implications derive from its elucidation of experiences that produce cognitive incongruity--surprise, perplexity, and discoordination. Teachers can induce surprise by having students make a prediction, then giving disconfirming evidence. Teachers may also induce surprise by having students encounter plausible predictions that differ from their own. Effective use of surprise requires that students already have acquired fairly rich and well-structured knowledge in a domain--knowledge that nonetheless includes misconceptions, false mental models, "bugs," etc. The surprise of having one's prediction disconfirmed may be strengthened by requiring that the predictions be expressed publicly. Teachers can induce perplexity by juxtaposing rival ideas. The presence of proponents of the rival ideas among peers amplifies the perplexity. Teachers can induce discoordination by having students explain or defend their ideas to others. To convince or teach, one must make explicit what was previously implicit. Persuasion requires orderly presentation, hence better internal organization. It also requires one to coordinate different points of view, and "one feels strong discoordination only when he or she struggles to coordinate."

Hatano and Inagaki have conducted a series of studies of the effectiveness of <u>Kasetsu-Jikken-Jugyo</u> (Hypothesis-Experiment-Instruction), a science education method developed in Japan (Itakura, 1962). They conclude their chapter with a description of the method and their findings. Although the method was not spawned by Hatano and Inagaki's theory, it is consonant with the theory.

Consequently, their example does demonstrate the plausibility and educational richness of the Hatano-Inagaki theory of motivation for comprehension. Another congenial example not cited by Hatano and Inagaki is the work of Hewson and Hewson (1984) on the role of conceptual conflict in helping to bring about conceptual change in science education.

## EFRAIM FISCHBEIN

Fischbein's chapter is squarely in the tradition of introspective psychologizing by mathematicians. He cites introspective accounts of several mathematicians (Hilbert, cited in Reid, 1970; Papert, 1980; Poincaré, 1913; Polya, 1954a, 1954b; and Tall, 1980).

Fischbein's chapter is concerned with the intuitive aspect of mathematical activity, which he distinguishes from formal and algorithmic aspects. The formal aspect of mathematical activity involves the deductive, logical structure of mathematics--axioms, definitions, theorems, and proofs. The algorithmic aspect involves standardized procedures--mathematical operations, formulas, and solution strategies. The intuitive aspect, with which Fischbein's chapter is concerned, involves subjective interpretations and connotations that individuals attach to mathematical truths as they make sense of them, assimilate them, and "integrate them in the fundamental schema of . . . mental behavior." This intuitive aspect, according to Fischbein, is often overlooked in mathematics instruction. He contrasts "cognitive components, deeply rooted in our adaptive behavior, like images, models and beliefs" with "propositional networks governed by logical rules." Intuition concerns constructs that synthesize these various aspects into unitary cognitive structures.

### The Influence of the Intuitive

Figural, intuitive representations, which most persons attach to abstract mathematical entities like point, line, and surface, "may influence the ways of reasoning even if the person is aware of the purely abstract nature of the respective entities." This influence of the intuitive on mathematical thinking--sometimes intrusion, sometimes inspiration--expresses itself in two ways during the problem-solving process. These are anticipatory intuition and affirmatory intuition. Anticipatory intuitions and affirmatory intuitions are intimately related, because they are both rooted in the intuitive meanings that persons attach to mathematical concepts. Also, a person's confidence in them exceeds what the evidence at hand merits.

Fischbein blends cognitive psychological constructs and mathematicians' introspections to describe what it is that effective mathematical problem solvers do. The particular attention Fischbein gives to the role of intuitions and analogies clearly distinguishes his chapter from those of Greeno and Hatano and Inagaki. The phenomena with which Fischbein is concerned--impingements of subconscious associations on conscious thought--are receiving newfound appreciation in cognitive science, as reported in Kihlstrom's (1987) recent article on the cognitive unconscious.

For both its benefits and its detriments, Fischbein argues that intuition will always be an important aspect of mathematical activity. Anticipatory intuitions may inspire a new direction for solution attempts. Affirmatory intuitions may help the student to construct a deeper, more personal, and more productive understanding of a concept. Both types of intuition provide the individual with the appearance of firm and reliable grounds, which is beneficial because confidence in one's mathematical grounds sustains mathematical effort--even if the confidence is unwarranted. This last benefit is paradoxical when placed into Hatano and Inagaki's theoretical framework: Illusory confidence is said to sustain comprehension activity, yet illusory confidence seems incompatible with comprehension monitoring activity, a prerequisite for comprehension activity.

In addition to its benefits, intuition also bedevils mathematical activity in ways that Fischbein seeks to illuminate cognitively. Fischbein provides an analysis of "the intervention of an intuitive meaning." The original, genuine meaning of a concept can be distorted by one's intuitive loading of that concept. And this initial, intuitive meaning can continue to color one's way of reasoning. Conflicts between intuitive meaning and formal constraints can escape the notice of both student and teacher. This analysis, which Fischbein has done in the context of mathematics, is consonant with recent findings on naive concepts in science education (Hewson & Posner, 1984; Posner, Strike, Hewson, & Gertzog, 1984).

Several of Fischbein's examples include frequently occurring mathematical problems in which students' intuitive interpretations of arithmetic operations affect their choice of solution and interfere with later understanding (e.g., multiplication as repeated addition). Other examples illustrate how the subjective certainty felt by students leads them to doubt the necessity of a proof and consequently to be suspicious of mathematics educators who insist on the importance of proofs. Fischbein's pedagogical recommendation in this case is that educators replace self-evident statements with counter-intuitive ones or that they frame statements in situations where they will lack intuitive meaning.

## Didactical Suggestions

Fischbein concludes his chapter by listing suggestions about how educators can foster fruitful coordination between intuition and other aspects of mathematical activity. The fact that his discussion of the role of intuition could lead to concrete suggestions deserves notice. His discussion of the role of intuition belongs to the literature on expert-novice differences in that it elucidates differences between novices and experts. In contrasting experts and novices, Fischbein has focused on a single dimension of difference, i.e., intuition. Also, an asymmetry should be noted in these novice-expert contrasts. Information about experts comes primarily from their own introspective accounts, whereas information about novices apparently comes from Fischbein's observations.

In elucidating what experts do well, Fischbein says little about the processes of acquisition and development through which the experts once passed. In Greeno's (1978) terms, Fischbein presents a model of knowledge, not a model of learning. Nonetheless, Fischbein suffers no shortage of recommendations about how educators should be able to affect processes of acquisition and development. This leap of logic assumes that effective pedagogical techniques are self-evident once one has a clear understanding of one's pedagogical goal, which in this case is the expert's ability to benefit from intuitions without falling victim to them. Although I have labeled this as an assumption, I do not criticize it. As I argue later, teachers' models of learning may suffice once teachers are given apt descriptions of novice-expert differences.

THOMAS A. ROMBERG & FREDRIC W. TUFTE

Romberg and Tufte have sought to apply recent cognitive science research to curriculum engineering in mathematics education. They describe curriculum engineering as an iterative process by which one invents and implements a curriculum, which is "an operational plan detailing what content is to be taught to students, how students are to acquire and use that content, and what teachers are to do in carrying out that curriculum." Romberg and Tufte's applications of cognitive science take the form of implied suggestions; they are practices suggested theoretically by the research, but they are not practices that have been tested in curriculum research.

Romberg and Tufte propose a constructivist approach to curriculum engineering in mathematics education. Their conception of constructivist epistemology is distinctly psychological: "We believe that information about how individuals personally construct knowledge and store it in memory should be the basis of curricula engineering." They have contrasted this approach to curriculum engineering with Ralph Tyler's (1931) approach, which they refer to as traditional curriculum engineering.

In their overview of cognitive science, Romberg and Tufte have emphasized the thesis that important forms of learning involve active processing in which the learner fits new information into his or her existing cognitive structures. This thesis, which is shared by other authors in this work, implies that prior knowledge makes the meaning of any new experience peculiar in some respects to each student.

Romberg and Tufte reviewed three areas of cognitive science: formal models of problem-solving protocols; qualitative differences between novices and experts; and recall of lists, stories, and prose. I would like to call particular attention to two ideas they emphasized in their review of problem-solving models. The first idea involves condition-action mechanisms and their educational importance. In the words of Romberg and Tufte, "As much care must be exercised in teaching the conditions under which mathematical properties and theorems can be applied as in the actual applications of these properties and theorems." This perspective, which is justified by the findings to date in cognitive science, can be regarded as direct instruction in transfer. It is based on the notion that failures of transfer often arise from a failure of pattern recognition. This notion has strong connections with the literature on activating schemas, and a review that coordinates these seldom-mingled literatures would be useful.

Second, I would like to call attention to Romberg and Tufte's characterization of the tension between formal systems of logic and idiosyncratic, informal knowledge structures developed by individuals. Romberg and Tufte, like Fischbein, regard informal knowledge structures as (1) powerfully affecting mathematical reasoning; (2) having an effect that is often, but not inevitably, detrimental; and (3) being omnipresently characteristic of human problem solving. The first two premises have been used heretofore as grounds for seeking to eradicate informal knowledge structures. Fischbein and Romberg and Tufte, heeding the third premise, believe that such efforts to eradicate are futile. Rapproachement between formal and informal thinking is advocated in both chapters, but the models of rapproachement are different. Fischbein's model of rapproachement is analogous to Freud's (1933/1965) model of sublimation, by which energy (libido) is channeled into constructive, creative activity. Informal intuition, the underlying engine, is not changed in Fischbein's model, but its force is kept within bounds and directed towards consciously pursued, formal purposes. Romberg and Tufte's model of rapproachement involves "molding or changing the informal structures of novices." This is a model of metaphor supplantation: Holistic metaphors[1] are acknowledged to be necessary in mathematical problem solving, but old metaphors can be supplanted by new ones. Consequently, mathematics educators should seek to supplant students' old detrimental metaphors with new ones that are more harmonious with formal mathematical systems.

Romberg and Tufte emphasize two qualitative differences that have been found between novices' and experts' approaches to problems. First, novices tend to use a means-end or working-

---

[1]What I have termed holistic metaphors are comparable to the organized wholes of Gestalt psychology (Kohler, 1929, pp. 187-223).

backward strategy, whereas experts tend to use a forward-looking strategy. Second, experts spend more time reading a problem and encoding it in an effort to make it conform better to an existing knowledge structure. Romberg and Tufte did not explicitly develop the pedagogical pertinence of these novice-expert contrasts.

Romberg and Tufte provide an overview of research on the recall of lists, stories, and prose. This research is the basis for some of the illustrative applications of cognitive science to curriculum engineering, which Romberg and Tufte describe at the end of their chapter. It is especially pertinent to Romberg's (1983) story-shell curriculum units.

In the final section of their chapter, Romberg and Tufte present ten principles of curriculum engineering that they believe to be "direct consequences of the research that has been summarized in this paper." The connection of some of these principles to the research reviewed is tenuous. But Romberg and Tufte would argue that it is overly constraining to draw from research literature only the causal, if-then assertions that have been tested directly. Instead, like Cronbach (1975) and Bishop (1982), they consider the shifts in perspective generated by researchers to be as important as specific causal assertions arising from research.

## REFLECTIONS

An emphasis on knowledge and its organization clearly distinguishes all of these "psychology in the math class" chapters from many of their counterparts of a few years ago. In the process-product studies of a recent era, instructional correlates of desirable mathematical performances were sought; if causal ambiguities in correlation could be nullified by experimental design, so much the better. But little attempt was made to peer into the black boxes of the cognitive processes underlying desirable mathematical performance. An expected dividend of the process-product research was practical knowledge of which instructional processes to use. Educators were hoping to capitalize on some practically useful empirical connections, whether or not the reasons for the existence of the connections were understood.

The process-product quest in education was analogous to the pragmatic procedure followed when aspirin was adopted as a drug. Appreciation for the analgesic benefits of aspirin long preceded any pharmacological understanding of the mechanisms by which it produces its benefits. The procedure sought to accumulate replicable practical lore, and it deferred questions about why things that worked did so. Now that advances have been made in understanding the physiological mechanisms by which aspirin acts, it is easy to appreciate the cliché that "nothing is as practical as a good theory." The authors of the foregoing chapters believe (and hope) that recent theoretical developments in understanding the cognitive processes of doing mathematics is analogous to recent theoretical developments in understanding the action of aspirin. To wit, they believe that cognitive theory is good enough to generate practically useful assertions not previously suggested by observation.

## The Novice-to-Expert Transition

Two hazards must be considered in using experts as a normative source of representations to be fostered in students. First, an inherent conservatism is implied by the acceptance of the notion that present experts have all the answers. Second, the transition from novice to expert is a gradual one that could involve many alternative routes. The end points--novice at one end, expert at the other--do not by themselves reveal the best path of transition. There remains, then, a new kind of information that is meager to date. As Lesgold, Pellegrino, Fokkema, and Glaser (1978) wrote a decade ago, "Work in modern cognitive psychology has focused primarily on the processes that underlie perceptual, memorial, and problem-solving performance and has only indirectly investigated how these process skills are learned and how broader competences and knowledge are acquired" (p. 1). What was lacking is what Greeno (1978) referred to as a theory of learning.

In the intervening decade, cognitive psychology has made headway on this problem, a noteworthy part of it in the domain of mathematics. A primary form of advance has been the illumination of the ways in which novices and experts differ, in some cases including the illumination of intervening, transitional states. We knew more today about what cognitive changes to watch for, but there is still more conjecture than understanding when we seek to explain the mechanisms that create those changes. There is much more to know, but I share the authors' optimism that new insights are accumulating rapidly. In addition to recognition of the importance of students' mental representations, experiments have been conducted to change them. Hewson's work in physics education illustrates the feasibility and importance of metaphor supplantation. We know the importance of being explicit, committed, public, and involved. We know that some common practices can disrupt motivation for understanding. The cognitive Berlynean theory of Hatano and Inagaki is to be applauded, because it seeks to illuminate the black box in which cognitive transitions occur.

Thanks to assorted studies of experts, we certainly know more about the goal state of being an expert. However, our strategy of working backward from the goal state reveals that we are still novices at the problem of moving a student from the initial state of novice to the goal state of expert

## Should Curriculum be Organized Ahistorically?

Romberg and Tufte's concern with the way in which the mathematics curriculum organizes mathematical knowledge is obviously merited. However, their goal that the knowledge be "optimally organized" assumes an optimum. Clearly, some ways of organizing mathematical knowledge are better than some other ways. But no single organization is, for all purposes, superior. Romberg and Tufte surely know that, so, in some respects, this criticism may seem to be semantic nitpicking. However, I believe the problem runs deeper than that, and it extends beyond Romberg and Tufte--certainly as far as Greeno, and possibly to the knowledge structure program in general. The problem is that a description of a desirable end state, no matter how apt it is as a description, is ahistorical; more precisely, it is abiographical. Consequently, it yields limited suggestions about the specific activities and sequences that will be effective in fostering the novice-to-expert transition.

In defense of the contributions of cognitive science, problem solving is not even possible if one does not know what the problem is. We now better understand what the problems are. Light shed on experts' cognition has brought problem statements within the ken of educators. There is now genuinely better insight than the obtuse recognition of success in performance with which we were previously saddled. Not so long ago, we were in the unenviable position of describing experts as those who could succeed at a task, and novices as those who could not. Clearly, cognitive science has cast some light into those black boxes, allowing us to know more about the thought processes--lacking in novices--that enable experts to succeed. Knowledge of experts' thought processes does help educators to clarify their goals.

Perhaps the ahistorical aspect is not so great a problem as my discussion thus far has suggested. Teachers may not need specific guidance on the relative efficacy of different ways to spur novice-to-expert movement. Perhaps cognitive science is already beginning to remedy its most glaring deficiency, the lack of a good problem statement. Apt descriptions of experts may constitute just the refinement of problem statement that teachers most need. Indeed, there is evidence that, once they understand a problem, teachers can invent effective solutions to it (Fennema, Carpenter, & Peterson, 1986).

## Motivation

Motivation, especially intrinsic motivation, often has been assigned a circular definition: Intrinsic motivation is what motivated persons have. Extrinsic motivation escaped this circularity, because the motivating conditions could be defined independently of the motivated state. Intrinsic motivation took the leftovers: Intrinsic motivation is what motivates persons when there is no extrinsic motivation. Recent advances, and I would number the Hatano and Inagaki work among them, improve this situation in that they illuminate the conditions that can disrupt intrinsic motivation.

## Distinguishing Expertise from Intelligence

As they plumb the contributions of knowledge structures, cognitive psychologists risk taking unnecessary blind alleys if they choose not to assimilate findings of earlier eras in psychology. Some practices now being identified as "expert practices" suspiciously resemble what for many years were called "intelligent practices." For example, Romberg and Tufte refer to Larkin's work (cited in Woods & Crowe, 1984) as evidence that experts who are successful in solving a problem "spend considerably more time than novices reading the problem statement before beginning to write equations." Was Larkin witnessing domain knowledge or something more general? Sternberg (1977), in his studies of intelligence, noticed a similar apportionment to "encoding" analogy problems. Labeling a practice as expert emphasizes its domain-specificity and its accessibility through sustained study. Labeling it as intelligent emphasizes its domain-generality and its relative recalcitrance to educators' efforts. It would seem important to distinguish the relatively plastic and relatively implastic ingredients of expert performance. To illustrate this point, consider the following problem.

> I have two coins.
> Together they make 55 cents.
> One of them is not a nickel.
> What are they?

This problem was used by David R. Olson (1986, p. 340) to illustrate sensitivity to subtleties of language. Olson attributed the item to Milton Rokeach, who reportedly used it to measure open-mindedness. I recently gave this problem to a class of 45 undergraduate students. After a wait of 30 seconds, only one hand signifying confidence in an answer was raised. On another occasion, when Alana, a five-year-old, heard the question posed to others, she quickly showed signs of insight. However, she was too unfamiliar with 50¢ pieces to solve it. In her case, lack of domain-specific knowledge prevented expert performance. When I reframed the problem by replacing "55 cents" with "15 cents," others present were still stumped. Alana replied, "You have a nickel and a dime. You said one of them is not a nickel, but the other one could be!" All problems require knowledge, and persons who lack essential knowledge will be unable to solve them. However, examples like this suggest that some kinds of problem solving involve knowledge that has already been acquired by some five-year-olds and has not been acquired by many collegians. Success on "nonentrenched" (Sternberg, 1981) tasks like this one probably involves knowledge structures that are relatively less plastic. The knowledge structure program would be more useful to educators if the relatively plastic, accessible structures could be distinguished from those that are relatively implastic.

As Greeno noted, unschooled domain experts link concrete objects of a particular domain to quantities, and they manipulate those quantities. They do not link these manipulations (operations) well with mathematical symbols. Greeno conjectures that the structures of unschooled domain experts, although they are abstract, are not as general as the structures used by experts in mathematics. Perhaps both the accomplishments and the weaknesses of unschooled domain experts should be interpreted in light of Olson's (1986) discussion of the linkage between intelligence and literacy. Mathematics would then be approached as one form of literate intelligence.

## Purposes of Mathematics Education

Which do we need, then--the stuff of mathematicians or the stuff of mathematics? Fields in which the application of mathematics is useful have long existed. For persons entering those fields, the desirability of extensive mathematics education was evident. But what about persons entering other fields? Defenses of mathematics education have, in general, followed one of two rhetorical tacks. The first tack emphasizes the "stuff of mathematicians"--habits of mind, disciplined inventiveness, perseverance, and the like. An illustration of this tack is quoted below.

> Would you have a man reason well, you must use him to it betimes, exercise his mind in observing the connexion of ideas and following them in train. Nothing does this better than mathematics which therefore I think should be taught all those who have the time and opportunity, not so much to make them mathematicians as to make them reasonable creatures. (Locke, 1699/1964, pp. 165-166)

The second tack for defending mathematics education emphasizes the stuff of mathematics--its concepts, symbols, and procedures. When this tack is taken, it is usually coupled with arguments about why the general public needs more of the stuff of mathematics. The authors of the foregoing chapters did not say so, but they have been spared the need to provide those arguments. Those arguments are not needed today because there is a spreading appreciation for the extent to which information-age societies have been suffused with mathematical concepts. The need to understand the stuff of mathematics is, at present, a truism.

Just as Resnick (1985) and others have argued that high literacy is both a necessary and a feasible goal for most students, so these authors are aspiring to a future for mathematics education in which all students do more of what mathematicians do. This is an estimably democratic idea, although some will say that it is a subtle way of legitimizing as the proper focus of schools those things in which the rich and powerful already excel.

References

Acioly, N. M., & Schliemann, A. D. (1986). Intuitive mathematics and schooling in understanding a lottery game. Recife, Brazil: Universidade Federal de Pernambuco.

Anderson, J. R. (1983). Acquisition of cognitive skill. Psychological Review, 89, 396-406.

Anderson, J. R., Boyle, C. F., Farrell, R., & Reiser, B. (1984). Cognitive principles in the design of computer tutors. Proceedings of the Cognitive Science Society, 2-11.

Berlyne, D. E. (1960). Conflict, arousal, and curiosity. New York: McGraw-Hill.

Berlyne, D. E. (1965). Structure and direction in thinking. New York: Wiley.

Bishop, A. J. (1982). Implications of research for mathematics teacher education. Journal of Education for Teaching, 8, 118-135.

Black, J. G., & Bower, G. H. (1980). Story understanding as problem solving. Poetics, 9, 176-193.

Brown, J. S., & Burton, R. R. (1980). Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 4, 379-426.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence (Vol. 1, pp. 7-75). Hillsdale, NJ: Erlbaum.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. American Psychologist, 30, 116-127.

Driver, R., & Oldham, V. (1985, August). A constructivist approach to curriculum development in science. Paper presented at the meeting of the British Educational Research Association, Sheffield.

Fennema, E., Carpenter, T. P., & Peterson, P. L. (1986). Teachers' decision making and cognitively guided instruction: A new paradigm for curriculum development. Paper presented at the Seventh Annual Psychology of Mathematics Education Conference, London, England.

Freud, S. (1965). New introductory lectures on psychoanalysis (J. Strachey, Ed. & Trans.). New York: Norton. (Original work published 1933)

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 702-718.

Greeno, J. G. (1976). Cognitive objectives of instruction: Theory of knowledge for solving problems and answering questions. In D. Klahr (Ed.), Cognition and instruction (pp. 123-160). Hillsdale, NJ: Erlbaum.

Greeno, J. G. (1978). A study of problem solving. In R. Glaser (Ed.), Advances in instructional psychology (Vol. 1). Hillsdale, NJ: Erlbaum.

Hewson, P. W., & Hewson, M. G. A. (1984). The role of conceptual conflict in conceptual change and the design of science education. Instructional Science, 13, 1-13.

Hewson, P. W., & Posner, G. J. (1984). The use of schema theory in the design of instructional materials: A physics example. Instructional Science, 13, 119-139.

Itakura, K. (1962). Instruction and learning of concept force in static based on Kasetsu-Jikken-Jugyo (Hypothesis-Experiment-Instruction): A new method of science teaching. Bulletin of National Institute for Educational Research, 52. (Published in Japanese.]

Kihlstrom, J. F. (1987). The cognitive unconscious. Science, 237, 1445-1452.

Kitcher, P. (1984). The nature of mathematical knowledge. New York: Oxford University Press.

Köhler, W. (1929). Gestalt psychology. New York: Liveright.

Lesgold, A. M., Pellegrino, J. W., Fokkema, S. D., & Glaser, R. (1978). Cognitive psychology and instruction. New York: Plenum.

Locke, J. (1964). Some thoughts concerning education (F. W. Garforth, Ed.). Woodbury, NY: Barron's Educational Series. (Abridged and edited version reprinted from Locke's 4th edition, 1699)

Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. Child Development, 50, 643-655.

McNemar, Q. (1964). Lost: Our intelligence? Why? American Psychologist, 19, 871-872.

Miyake, N. (1986). Constructive interaction and the iterative process of understanding. Cognitive Science, 10, 151-177.

Newell, A., & Simon, H. A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Olson, D. R. (1986). Intelligence and literacy: The relationships between intelligence and the technologies of representation and communication. In R. J. Sternberg & R. K. Wagner (Eds.), Practical intelligence: Nature and origins of competence in the everyday world. New York: Cambridge University Press.

Osler, S. F., & Fivel, M. W. (1961). Concept attainment, I: The role of age and intelligence in concept attainment by induction. Journal of Experimental Psychology, 62, 1-8.

Osler, S. F., & Trautman, G. E. (1961). Concept attainment, II: Effect of stimulus complexity upon concept attainment at two levels of intelligence. Journal of Experimental Psychology, 62, 9-13.

Papert, S. (1980). Mindstorms: Children, computers, and powerful ideas. New York: Basic Books.

Piaget, J. (1952). The child's conception of number. New York: Norton.

Piaget, J. (1966). Necessite et signification des recherces comparatives en psychologie genetique. Journal International de Psychologie, 1, 1-13.

Poincaré, J. H. (1913). The foundations of science. (Holsted, Trans.). New York: Science Press.

Polya, G. (1954a). How to solve it. Garden City, NY: Doubleday-Anchor.

Polya, G. (1954b). Induction and analogy in mathematics. Princeton: Princeton University Press.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1984). Accommodation of a scientific conception: Towards a theory of conceptual change. Science Education, 66, 211-227.

Reid, C. (1970). Hilbert. Heidelberg: Springer.

Resnick, L. B. (1985, April). Cognition and curriculum. Invited address presented at the meeting of the American Educational Research Association, Chicago.

Romberg, T. A. (1983). A common curriculum for mathematics. In G. D. Fenstermacher & J. I. Goodlad (Eds.), Individual differences and the common curriculum: Eighty-second yearbook of the National Society for the Study of Education (pp. 121-158). Chicago: University of Chicago Press.

Sleeman, D. (1984). An attempt to understand students' understanding of algebra. Cognitive Science 8, 387-412.

Smith, B. C. (1983). Reflection and semantics in a procedural language (MIT/LCS/TR-22). Cambridge, MA: Massachusetts Institute of Technology, Laboratory for Computer Science.

Sternberg, R. J. (1977). Intelligence, information processing, and analogical reasoning: The componential analysis of human ability. Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1981). Intelligence and nonentrenchment. Journal of Educational Psychology, 73, 1-16.

Tall, D. (1980). The anatomy of a discovery in mathematics research. For the Learning of Mathematics, 1, 25-34.

Tall, D. (1985). Using computer graphics as generic organizers for the concept image of differentiation. In L. Strefland (Ed.), Proceedings of the Ninth International Conference for the Psychology of Mathematics Education (pp. 105-110). Utrecht, The Netherlands: PME.

Tyler, R. W. (1931). A generalized technique for constructing achievement tests. Educational Research Bulletin, 8, 199-208.

## OUTCOMES OF SCHOOL AND THEIR ASSESSMENT

In this fourth set of background papers we begin to address the most critical problem in designing a monitoring system: namely, what is a reasonable approach to assessing the outcomes of instruction in mathematics given the shifts in emphasis due to the reforms. In chapters 17 and 18 members of the staff have summarized the past approaches to student assessment in light of the reform movement and found that there is a need to develop more valid procedures. In chapter 19 Kevin Collis presents an approach that relates methods of assessment to levels of reasoning. In chapter 20 Brian Donovan and Tom Romberg summarize the relationship between knowledge structures and assessment of understanding in mathematics. Normar Webb, in chapter 21, critically examines the arguments presented in chapters 17 to 20. The final two chapters examine a different aspect of instructional outcomes: namely, attitudes. In chapter 21 Gilah Leder summarizes the varied work on attitude assessment in mathematics. Doug McLeod provides a critique of that chapter in chapter 22 as well as an examination of recent approaches to attitude research.

Chapter 17

# MEASURES OF MATHEMATICAL ACHIEVEMENT[1]

Thomas A. Romberg

Information from students about their mathematical achievement is important. This is particularly true for the study of the effects of changes in what is being taught or how instruction is carried out. Only by repeatedly gathering achievement data over time can one reliably argue about actual effects.

In this chapter, I first briefly describe what is meant by the term achievement as it is applied to school mathematics. Then I give a short history of testing. In the third section a description of the three contemporary types of tests (standardized norm-referenced tests, profile achievement tests, and objective-referenced tests) is given with a discussion of the strengths and weaknesses of each. The chapter concludes with a rationale for the development of new tests that would be more valid indicators of mathematics achievement.

## MATHEMATICS ACHIEVEMENT

Achievement can be considered as reasonable pupil outcomes following a set of instructional experiences in school courses. Detailing those outcomes is, of necessity, quite complex. Minimally, however, the acquisition and maintenance of concepts and skills, preparation for new concepts and skills, acquisition of a positive attitude toward mathematics, and the use of concepts and skills to solve problems should be included. Although these outcomes are essential, they do not exhaust the list of pupil outcomes one might usefully observe in assessing the effect of the manner in which a particular content unit or course has been taught. In fact, in this period of change, mathematical concepts and skills have become more important, and emphasis has shifted from acquiring a large number of concepts and calculation routines to estimating, conjecturing, and developing strategies for solving problems.

Academic achievement is a subset of achievement associated with academic courses (as contrasted with vocational, technical,

---

[1]Sections of this paper are from a presentation given at The National Conference on the Influence of Testing on Mathematics Education, June 27-28, 1986, University of California at Los Angeles.

and physical education courses, for example). The concepts and skills of academic courses are associated with subject-matter disciplines (language arts, mathematics, physics). The goals of such courses not only emphasize acquisition and maintenance of concepts and skills, but, in particular, stress preparation for continued study in the subject area and subsequent use of that knowledge in various occupations.

In addition to the complex question of what outcomes should be examined, we must ask how to elicit the information needed. At least four aspects should be considered. First, the decisions about effects must be specified. Second, the implications of each decision to be made must be examined. This involves consideration of both the kind of statistical errors (Type I and II) one is willing to live with, and whether the decision is irrevocable. Next, the "unit" about which the decision is to be made must be determined (individuals, groups, classes, school, materials, etc.).

Finally, the question of measurement procedures and decision rules must be answered. This involves specifying the source, the scaling procedure, the reliability, and the validity of the measurement process.

The most common method of gathering information about mathematics achievement is paper-and-pencil tests given to groups of students. Although other procedures (for example, interviews, observations, and judgments about work samples) could be used, the ease of development, the convenience, and the low cost of such group testing has made it common in American schools. To understand how this has occurred, let us first examine how such tests have become so dominant.

## TESTING IN THE U.S.

The history of the measurement of human behavior, with primary reference to the capacities and educational attainments of school children, may be divided roughly into three periods. During the first period, from the beginning of historical records to about the 19th century, measurement in education was quite crude. During the second period, embracing approximately the 19th century, educational measurement began to assimilate from various sources the ideas and the scientific and statistical techniques which were later to result in the psychometric testing movement. The third period, dating from about 1900 to the present, can be characterized as the psychometric period.

### Early Examinations

The initiation ceremonies by which primitive tribes tested the knowledge of tribal customs, endurance, and bravery of young men prior to admission to the ranks of adult males may be among earliest examinations employed by human beings. Use of a crude oral test was reported in the Old Testament and Socrates is known to have employed searching types of oral quizzing. Elaborate and exhaustive written examinations were used by the Chinese as early as 2200 B.C. in the selection of their public officials. These illustrations may be classified as historical antecedents of performance tests, oral examinations, and essay tests.

## Educational Testing in the 19th Century

Three persons made outstanding contributions to 19th-century developments. The ideas of these men--Horace Mann, George Fisher, and J. M. Rice--appear to be forerunners of developments during the present century.

The first school examinations of note appear to be those instituted in the Boston schools of 1845 as substitutes for oral tests when enrollments became so large that the school committee could no longer examine all pupils orally. These written examinations, in arithmetic, astronomy, geography, grammar, history, and natural philosophy, impressed Horace Mann, then secretary of the Massachusetts Board of Education. As editor of the Common School Journal, he published extracts from them and concluded that the new written examination was superior to the old oral test in these respects.

1. It is impartial
2. It is just to the pupils.
3. It is more thorough than older forms of examination.
4. It prevents the "officious interference" of the teacher.
5. It "determines, beyond appeal or gainsaying, whether the pupils have been faithfully and competently taught."
6. It takes away "all possibility of favoritism."
7. It makes the information obtained available to all.
8. It enables all to appraise the ease or difficulty of the questions.

(Greene, Jorgenson, & Gerberich, 1953)

Although these ideas were apparently those represented by modern tests, the instruments themselves were inadequate. However, in successive issues of the Common School Journal, Mann suggested most of the elements in examinations that are found in the contemporary measurement.

To Reverend George Fisher, an English schoolmaster, goes the credit for devising and using what were probably the first objective measures of achievement. His "scale books," used in the Greenwich Hospital School as early as 1864, provided means for evaluating accomplishments in handwriting, spelling, mathematics, grammar and composition, and several other school subjects. Specimens of pupil work were compared with "standard specimens" to determine numerical ratings that, at least for spelling and a few other subjects, depended on errors in performance (Greene, Jorgenson, & Gerberich, 1953).

The inventor of the comparative test in America was J. M. Rice. In 1894, he developed a battery spelling test. Having administered a list of spelling words to pupils in many school systems and analyzed the results, Rice found that pupils who had studied spelling 30 minutes a day for eight years were not better spellers than children who had studied the subject 15 minutes a day for eight years. Rice was attacked and reviled for this "heresy," and some educators even attacked the use of a measure of how well pupils could spell for evaluating the efficiency of spelling instruction. They intended that spelling was taught to develop the pupils' minds and not to teach them to spell. It was more than ten years later that Rice's pioneering resulted in significant attention to objective models in educational testing (Ayres, 1918).

## The Psychometric Period

This era began shortly after the turn of the century. Although the historical antecedents sketched in the preceding paragraphs were essential prerequisites, developments first in mental testing and shortly after in achievement testing are at the roots of this era.

General intelligence tests. Attempts to measure general intelligence, or ability to learn or ability to adapt oneself to new situations, had been made both in America and in France. The first individual test was developed in France, and the first group test was developed some years later in America.

Individual intelligence scales were originated in 1905 by Binet and Simon. Their first scale was devised primarily for the purpose of selecting mentally retarded pupils who required special instruction. This pioneer individual-intelligence scale was based on interpreting the relative intelligence of different children at any given chronological age by the number of questions of varied types and increasing levels of difficulty they could answer. These characteristics were all re-embodied in the 1908 and 1911 revisions of the Binet-Simon Scale and remain basic to most individual intelligence scales today. The 1908 revision introduced the fundamentally important concept of mental age (MA) and provided means for obtaining it (Freeman, 1939). Several American adaptations of these pioneer scales appeared between 1911 and 1916. All make use of the intelligence quotient (IQ), based on the relationship between the subject's mental age and chronological age (Freeman, 1939).

The first group intelligence test was Army Alpha, used for the measurement and placement of army recruits and draftees during World War I. It was the product of the collaboration of various psychologists working on group intelligence tests when the United States entered the war. This test, widely used to test men who could read and understand English, was accompanied by Army Beta, a nonlanguage test for use with illiterates and men who, although perhaps literate in a foreign language, could not read English

(Freeman, 1939). Other group intelligence tests began to appear almost immediately following World War I, and the period from 1918 to the middle 1920s was marked both by the publication of many such tests and by an upsurge of interest in intelligence testing.

Aptitude Tests. The measurement of aptitudes, or those potentialities for success in an area of performance that exist prior to direct acquaintance with that area, was closely related to intelligence testing. Early attempts to measure general intelligence tested many specific traits and aptitudes, but that approach was abandoned after Binet showed that tests of more complex forms of behavior were superior. It was soon apparent, however, that general intelligence tests were not highly predictive of certain types of performance, especially in the trades and industries. Münsterberg's aptitude tests for telephone girls and streetcar motormen were followed by tests of mechanical aptitude, musical aptitude, art aptitude, clerical aptitude, and aptitude for various subjects of the high school and college curricula prior to 1930 (Watson, 1938). Spearman's (1904) splitting of total mental ability into a general factor and many specific factors had its influence on this movement.

Achievement Tests. Modern achievement testing was stimulated by Thorndike's 1904 book on mental, social, and educational measurements. Through his book and his influence on his students, Thorndike was predominantly responsible for the early development of standardized tests. Stone, a student of Thorndike's, published the first arithmetic reasoning test in 1908. Between 1909 and 1915, a series of arithmetic tests and five scales for measuring abilities in English composition, spelling, drawing, and handwriting were published (Odell, 1930). Literally thousands of standardized achievement tests have been published during the last half-century.

## Summary

The reasons for presenting this brief history are threefold. First, what is referred to as the modern testing movement began with a selection problem (Binet & Simon) and a placement problem (Army Alpha). It was assumed that a single measure (e.g., MA) or index (e.g., IQ) could be developed to compare individuals on what was assumed to be a general fixed unidimensional trait. In turn, the procedures that evolved in developing and administering these tests were used in aptitude and achievement tests. Second, the testing procedures we now consider typical were developed for group administration of early intelligence tests. An example from the Lorge-Thorndike Test (1954) is shown in Figure 1. Such tests are comprised of a set of questions (items), each having one unambiguous answer. In this sense, such tests are "objective" since subjective inferences are not necessary. All subjects are administered the same items under standard (nearly identical) situations with the same instructions, time, constraints, etc. Furthermore, subjects' answers (usually chosen from a set of alternatives as in Figure 1) could be easily scored as correct or

One word has been left out of each sentence on these two pages. Choose the word that will make the best, the truest, and the most sensible complete sentence. Look at sample sentence 0.

0. Hot weather comes in the ———.
A fall B night C summer D winter E snow — A B C D E

The best answer is summer. The letter before summer is C, so you should make a heavy black pencil mark in the C answer space for sentence 0.

Now look at sentence 00.

00. ——— bark at cats.
F Cows G Mice H Cats J Hens K Dogs — F G H J K

The best answer is Dogs, so you should make a heavy black pencil mark in the K answer space for sentence 00.

*Do all the sentences on these two pages in the same way. Try every sentence.*

1. Boys will become ———.
A infants B little C intelligent D stupid E men — A B C D E

2. We see ——— only at night.
F children G plants H stars J houses K trees — F G H J K

3. Fred was six years old. There were six ——— on his birthday cake.
L candles M boys N girls P parties Q children — L M N P Q

4. Not every cloud gives ———.
R weather S shade T sky U climate V rain — R S T U V

Figure 1. Excerpt from Lorge-Thorndike Intelligence Tests (Lorge & Thorndike, 1954).

not, the total number of correct answers tallied, tallies transformed, and transformed scores compared. Psychometrics involving the application of statistical procedures to such tests developed as a field of study in the 1920s.

Most importantly, it should be understood that the testing movement was a product of a historical era. It grew out of the machine-age thinking of the industrial revolution of the past century. The intellectual contents of the machine age rested on three fundamental ideas. The first was reductionism. For several centuries, our world view argued that everything we experience, perceive, touch, feel, or handle is comprised of parts. The machine age, preoccupied with taking things apart, was founded on the idea that, in order to deal with anything, you had to take it apart until you reached ultimate parts.

The second fundamental idea was that the most powerful mode in thinking was a process called analysis. Analysis is based in reductionism. It argues that, if you have something you want to explain or a problem you want to solve, you start by taking it apart. You break it into its components, you get down to simple components, then you build up again.

The third basic idea of the machine age has been called mechanism. Mechanism is based on the theory that all natural phenomena can be explained by cause-and-effect relationships. The primary effort of science was to break the world into parts that could be studied to determine cause-and-effect relationships. The world was conceived of as a machine operating in accordance with unchanging laws.

These ideas gave rise to what we now call the first industrial revolution. In this era, work was defined in physical terms; mechanization involved the use of machines to perform physical work. Man as an energy source was supplemented by machines. Man-machine systems were developed to do physical work in such a way that mechanization was facilitated.

This process is clearly reflected in what has happened in school mathematics during the last half-century. Mathematics was segmented into subjects and topics, and eventually reduced to its smallest parts: behavioral objectives. At this point, a network diagram was created (a hierarchy) to show how these components were related to produce a finished product.

Next, the steps through that hierarchy were mechanized via textbooks, worksheets, and tests. Teaching was dehumanized to the point that the teacher need do little but manage the production line.

Business, industry, and, in particular, schools have been conceived, modified, and operated based on this mechanical view of the world since before the turn of the century. Today, however, a new world view has emerged. It is a view we should use in our considerations of school mathematics and its assessment.

## ACHIEVEMENT TESTS

During the past three-quarters of a century, a variety of different achievement tests have been developed. In this section, the three most widely used types of tests are described, and their appropriateness for monitoring changes in school mathematics assessed.

### Standardized Tests

Norm-referenced standardized tests have become part of the yearly ritual in most schools. The purpose of such tests is to rank respondents with respect to a particular type of mental ability or achievement or to indicate a respondent's position in a population. A standardized test is comprised of a set of independent multiple-choice questions. The items have necessarily been subjected to a preliminary tryout with a representative pupil group, so that it is possible to arrange the items in a desired manner with respect to difficulty and the degree to which they discriminate among students. Also, each test is accompanied by an appropriate table for transforming resulting scores into meaningful characterizations of pupil mental ability or achievement (grade-equivalent scores, percentiles, stanines, etc.)

For example, millions of students each year take one of the major college admissions tests, the Scholastic Aptitude Test (SAT) or the American College Test (ACT). Both are standardized tests. Scores derived from these tests are used to make selection and placement decisions.

Four features of such tests require comment. First, although each test is designed to order individuals on a single (unidimensional) trait, such as quantitative aptitude, the derived score is not a direct measure of that trait. It is, for example, as if one were measuring Houston Rocket basketball star Ralph Sampson's height but not reporting that he is 7' 4"; rather, what is reported is that he is at the 99th percentile for American men. Furthermore, for mathematics achievement, there is no theoretically single trait (like height) that is being assessed.

Second, because individual scores are compared with those of a norm population, there will always be some high and some low scores. This is true even if the range of scores is small. Thus, high and low scores can not be judged as "good" or "bad" with respect to the underlying trait.

Third, test items are assumed to be equivalent to each other. They are selected on the basis of general level of difficulty (p value) and some index of discrimination (e.g., nonspurious biserial correlation). Furthermore, there is no claim that the items are representative of any well-defined domain. For example, in many subtraction computation standardized tests, items such as that shown in Figure 2 are common. Such an item, because of a zero

$$\begin{array}{r} 304 \\ -\ \underline{176} \end{array}$$

A) 272
B) 138
C) 238
D) 128
E) 232

Figure 2. A Typical Three-Digit Subtraction Test Item.

in the tens' place, requires successive regroupings and discriminates between good and average subtractors. However, if one were to randomly generate three-digit subtraction problems, few like this would ever appear.

Finally, such tests have only predictive validity. Scores on the SATs are useful only because they are reasonable predictors of how well students will do in college.

The strength of standardized tests is that they do what they were designed to do reasonably well. (Note that the SAT is an aptitude test, not an achievement test.) They are relatively easy to develop, inexpensive, convenient to administer, and provide comprehensible results. Their primary weakness is that they are often used as the basis of decisions they were not designed to address. For example, aggregating standardized scores for students in a class (school, district, etc.) to get a class profile of achievement (class mean) is a very inefficient method of profiling a class; standardized tests provide too little information for the cost involved. They are of little value for evaluation or research, since test items are not selected to be representative of the curriculum. Unfortunately, their common use appears to be more strongly related to political, rather than educational, issues. For example, it is claimed that elected officials and educational administrators increasingly use test scores comparatively to indicate which schools, school districts, and even individual teachers appear to be achieving better results (National Coalition of Advocates for Students, 1985). Such comparisons are misleading. One can only conclude that standardized tests are unwisely overused, and their derived scores are of little value as indicators of achievement which could be used in monitoring the health of the system.

Profile Achievement Tests. In contrast to standardized tests, profile achievement tests are designed to yield a variety of scores for groups of students. As early as 1931, Ralph Tyler outlined a procedure for test construction and validation which clearly pointed out the essential dependence of a program of achievement testing on the objectives of instruction and the recognition of forms of pupil behavior indicating attainment of the desired instructional outcomes. Tyler, more than any other single test specialist, was responsible for the extension of achievement testing to the outcomes of instruction. His contributions in the 1930s doubtless did much to replace the narrow concept of standardized testing with a broad, modern conception of evaluation.

The current approach to profile testing is to specify a content-by-behavior matrix. For example, the matrix used for profiling 8th-grade performance in the Second International Mathematics Study is shown in Table 1 (Crosswhite et al., 1986, pp. 80-81). Content topics are crossed with hypothesized cognitive levels. The content topics are judged to be appropriate for that grade, and the cognitive levels are usually based on some adaptation of those in Bloom's Taxonomy (1956). Items, similar to those in standardized tests, are prepared for each cell in the matrix. Item data then can be reported in several ways. First, data can be reported in terms of item means. Second, cell means can be calculated. For example, in Figure 3, the means are presented for six items on a topic (each given in a different instrument) for different students at different grades in Ontario (McLean, 1982b). Third, item scores can be aggregated by columns to yield cognitive level scores or by rows to yield topic scores (see Figure 4).

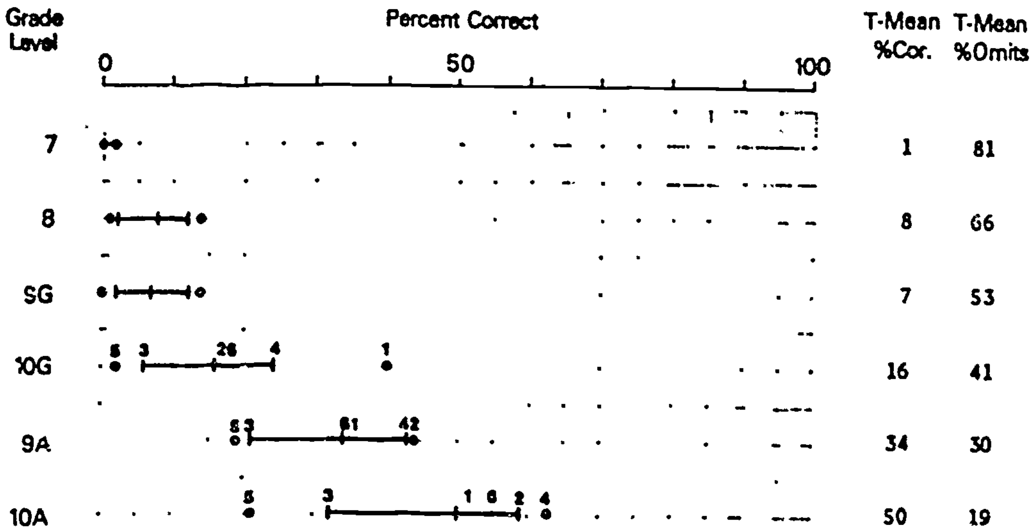
Table 1

Population A: Importance For Instrument Construction Of Content Topics And Behavioral Categories

| Content Topics | Behavioral Categories* Computation | Comprehension | Application | Analysis |
|---|---|---|---|---|
| **000 Arithmetic** | | | | |
| 001 Natural numbers and whole numbers | V | V | V | I |
| 002 Common fractions | V | V | I | I |
| 003 Decimal fractions | V | V | V | I |
| 004 Ratio, proportion, percentage | V | V | I | I |
| 005 Number theory | I | I | - | - |
| 006 Powers and exponents | I | I | - | - |
| 007 Other numeration systems | - | - | - | - |
| 008 Square roots | I | I | - | - |
| 009 Dimensional analysis | I | I | - | - |
| **100 Algebra** | | | | |
| 101 Integers | V | V | I | I |
| 102 Rationals | I | I | I | I |
| 103 Integer exponents | Is | - | - | - |
| 104 Formulas and algebraic expressions | I | I | I | I |
| 105 Polynomials and rational expressions | I | Is | - | - |
| 106 Equations and inequations (linear only) | V | I | I | Is |
| 107 Relations and functions | I | I | I | - |
| 108 Systems of linear equations | - | - | - | - |
| 109 Finite systems | - | - | - | - |
| 110 Finite sets | I | I | I | - |
| 111 Flowcharts and programming | - | - | - | - |
| 112 Real numbers | - | - | - | - |
| **200 Geometry** | | | | |
| 201 Classification of plane figures | I | V | I | Is |
| 202 Properties of plane figures | I | V | I | I |
| 203 Congruence of plane figures | I | I | I | Is |
| 204 Similarity of plane figures | I | I | I | Is |
| 205 Geometric constructions | Is | Is | Is | - |
| 206 Pythagorean triangles | Is | Is | Is | - |
| 207 Coordinates | I | I | I | Is |
| 208 Simple deductions | Is | I | I | I |
| 209 Informal transformations in geometry | I | I | I | - |
| 210 Relationships between lines and planes in space | - | - | - | - |
| 211 Solids (symmetry properties) | Is | Is | Is | - |
| 212 Spatial visualization and representation | - | Is | Is | - |
| 213 Orientation (spatial) | - | Is | - | - |
| 214 Decomposition of figures | - | - | - | - |
| 215 Transformational geometry | Is | Is | Is | |
| **300 Descriptive Statistics** | | | | |
| 301 Data collection | Is | I | I | - |
| 302 Organization of data | I | I | I | Is |
| 303 Representation of data | I | I | I | Is |
| 304 Interpretation of data (mean, median, mode) | I | I | I | - |
| 305 Combinatorics | - | - | - | - |
| 306 Outcomes, sample spaces and events | Is | - | - | - |
| 307 Counting of sets, (P(A B), P(A B), independent events | - | - | - | - |
| 308 Mutually exclusive events | - | - | - | - |
| 309 Complementary events | - | - | - | - |
| **400 Measurement** | | | | |
| 401 Standard units of measure | V | V | V | - |
| 402 Estimation | I | I | I | - |
| 403 Approximation | I | I | I | - |
| 404 Determination of measures: areas, volumes, etc. | V | V | I | I |

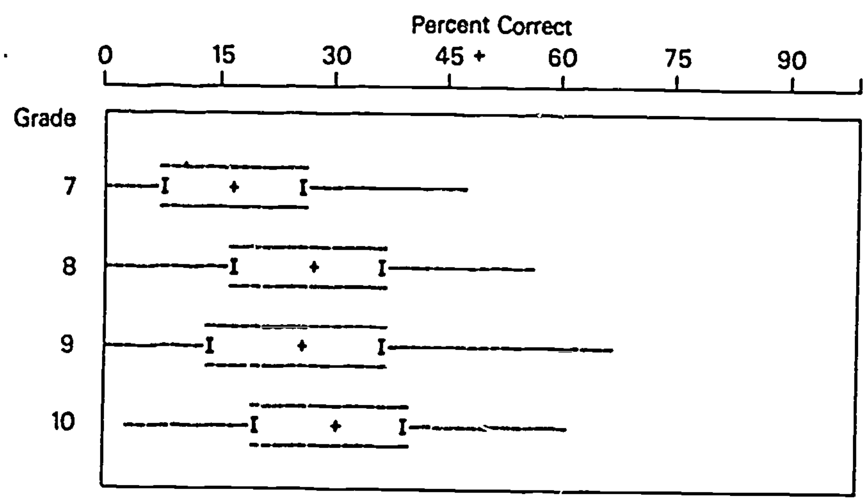
* The following rating scale has been used: V = very important; I = important; Is = important for some countries. A dash (-) indicates that the topic was not considered important enough to warrant trial items being found or constructed.

Source: Crosswhite, F. J., et al., 1986, pp. 80-81.




- this topic is not part of the Gr. 7 or Gr. 8 program.
-· a surprisingly large number of Grade 10 Advanced students omitted these items.
-. results indicate that where this skill is needed in Grade 11 and 12 it should be reviewed and practised then.

Figure 3. Algebra--Equations and Inequalities.
Range of Correct Responses to the Six Instruments, by Grade
(from McLean, 1982b, p. 207)




Statistical Summary

| Grade Level | No. of Classes | Grade Mean | Grade St. Dev. |
|---|---|---|---|
| 7 | 97 | 18.6 | 11.8 |
| 8 | 98 | 26.8 | 12.9 |
| 9 | 122 | 25.6 | 15.4 |
| 10 | 103 | 30.4 | 13.8 |

Figure 4. Percentages.
Range of Correct Responses to Topic Group, by Grade
(from McLean, 1982b, p. 138)

Profile tests have become popular alternatives to standardized tests. They have been developed for several major studies of mathematical performance, such as the National Longitudinal Study of Mathematical Abilities (NLSMA), National Assessment of Educational Progress (NAEP), First International Mathematics Study (FIMS), Second International Mathematics Study (SIMS), and several different state assessments.

There are four features of profile assessments that make them quite different from standardized tests. First, there is no assumption of an underlying single trait. Instead, instruction at any grade in mathematics is assumed to focus on several topics; the tests are designed to reflect the multidimensional nature of mathematical content. It must be noted that there is often a temptation to aggregate and derive a single total score, which would be very misleading. Second, the unit of investigation is a group, not an individual. Matrix sampling is usually used so that a wider variety of items can be given. Third, as in Figures 3 and 4, comparisons between groups are done graphically on actual scores. No transformations are needed. Finally, validity is determined in terms of content and/or curricula validity. Mathematicians and teachers are asked to judge whether individual items reflect a content behavior cell in the matrix and sometimes to judge whether or not the item represents something that was taught in the curriculum. The strength of profile achievement tests is that they provide useful information about groups. They are particularly useful for general evaluations of changed educational policy that directly affects classroom instruction. Thus, profile tests are very useful for monitoring purposes.

However, there are four weaknesses of these tests. First, because they are designed to reflect group performance, they are not useful for individual ranking or diagnosis. An individual student answers only a sample of items. Second, they are somewhat more costly to develop than are standardized tests, and they are harder to administer and score. Third, because they yield a profile of scores, they are difficult to interpret. In particular, comparisons between groups with different profiles often do not yield simple results.

However, their primary weakness is in the outdated assumptions underlying the two dimensions of content-by-behavior matrices. The content dimension (see Table 1) involves a classification of mathematical topics into "informational" categories. As Romberg (1983) has argued:

> "Informational knowledge" is material that can be fallen back upon as given, settled, established, assured in a doubtful situation. Clearly, the concepts and processes from some branches of mathematics should be known by all students. The emphasis of instruction, however, should be "knowing how" rather than "knowing what." (p. 122)

Furthermore, the items in any content category are independent of each other. For profiles, we should use content domains that reflect how that material is learned. Also, the items should reflect the interdependence (rather than independence) of ideas in that field. Gerard Vergnaud (1982) referred to such domains as "conceptual fields."

The behavior dimension of matrices has always posed problems. All agree Bloom's Taxonomy (1956) has proven useful for low-level behaviors (knowledge, comprehension and application) but difficult for the higher levels (analysis, synthesis, and evaluation). Single-answer, multiple-choice items are not reasonable for those levels. One problem is that this taxonomy suggests that the "lower" skills should be taught before the "higher" skills. As Resnick (1987) argued:

> This assumption--that there is a sequence from lower level activities that do not require much independent thinking or judgment to higher level ones that do--colors much educational theory and practice. Implicitly at least, it justifies long years of drill on the "basics" before thinking and problem solving are attended to or demanded. Cognitive research on the nature of basic skills such as reading and mathematics provides a fundamental challenge to this assumption. (p. 8)

The real problem is that Bloom's Taxonomy fails to reflect current psychological thinking. It is based on the naive psychological principle that simple individual behaviors become integrated to form a more complex behavior. In the past 30 years, our knowledge about learning and how information is processed has changed and expanded. Today, we should discard Bloom's Taxonomy and use a contemporary alternative that reflects current ideas from cognitive psychology.

Objective-referenced tests. These tests (often called criterion-referenced tests) are a product of the behavioral objectives movement in the U.S. during the 1960s. Statements of the following form are behavioral objectives: "the subject when exposed to the conditions described in the antecedent displays the action specified in the verb in the situation specified by the consequent to some specified criterion" (Romberg, 1976, p. 23). Items randomly selected from a pool designed to represent the antecedent conditions and the same action verb are given to students. From their responses, diagnosis of problems or judgments of mastery of objectives can be made.

Four features of these tests should be mentioned. First, they are usually designed as part of a curriculum and are to be administered to individuals at the end of some instructional topic. They often are administered individually and judgments are made quickly by teachers. For example, they are a part of such elementary mathematics programs as Individually Prescribed Instruction (Lipson, Koburt, & Thomas, 1967) and Developing

Mathematical Processes (Romberg, Harvey, Moser, & Montgomery, 1974, 1975, 1976). Second, they have occasionally been used in group settings. For example, the comprehensive achievement monitoring scheme (Gorth, Schriber, & O'Reilly, 1974) assesses student performance periodically on a set of objectives. Also, in 1974 Wisconsin used objective referencing for the construction of a state test (Wisconsin Mathematics Assessment Committee, 1974). Third, decisions on performance are made with respect to some a priori criterion. Often, a 75%-80% correct threshold has been used. For example, in Wisconsin's 1974 state test, variable criteria were used. First, objectives were defined by three priorities:

> Priority I:
> Objectives that deal with skills, concepts and applications which are essential for all students and/or are minimum prerequisites for continued study of mathematics.
>
> Priority II:
> Objectives that deal with skills, concepts and applications which are essential, but in-depth mastery is not expected at this level.
>
> Priority III:
> Objectives that expose students to new topics or challenging problems, provide motivation or create interest. (WMAC, 1974, p. 6)

Then, performance on the items in the priority level were evaluated using the scheme depicted in Figure 5.

*Evaluation*

| | Acceptable | Unacceptable | Understandable |
|---|---|---|---|
| **Priority I** | 75% or more of the students responded correctly to the item. | Less than acceptable student performance. | Unacceptable performance resulting from test item construction. |
| **Priority II** | 50% or more of the students responded correctly to the item. | Less than acceptable student performance. | Unacceptable performance but difficult test item. |
| **Priority III** | None were Assessed | | |

Figure 5. Interpretive Analysis Used in First Wisconsin State Mathematics Assessment (from WMAC, 1974, p. 7)

Finally, in some programs there have been a few attempts to aggregate performance across objectives. For example, the "80/80" criterion was used to describe whether a student had succeeded on a topic ("80/80" meaning that, for at least 80% of the objectives, the student had gotten at least 80% of the items correct).

The strength of objective-referenced tests lies in their instructional usefulness. As long as instruction on some topic focuses on the acquisition of some concept or skill, such tests can be used to indicate whether the concept has been learned or the skill mastered. Furthermore, such tests are scored easily and are readily interpretable.

Three weaknesses should be mentioned. First, such tests are costly to construct because there are often hundreds of objectives in any instructional program. Second, aggregation across objectives is not very reasonable. Third, and most importantly, these tests share some of the same conceptual problems that trouble profile tests. Objectives are assumed to be independent n c interdependent; items for higher level or complex problem-solving processes are hard to construct; and only correct answers (not strategies or processes) are scored.

Other Tests. In this brief review, other tests often used in mathematics education research have not been mentioned. For example, personality tests, ability tests (e.g., spatial ability), or even diagnostic tests are often administered. They simply do not fit the conception of assessment of mathematical achievement needed for the monitoring of school mathematics.

Summary. The purpose of this section was to reflect on current practice and to outline what tests now in common use can and cannot do. The main point is that, while these tests have been useful for some purposes and undoubtedly will continue to be used, they are products of an earlier era in educational thought. Like the Model T Ford assembly line, objective tests were considered as an example of the application of modern scientific techniques in the 1920s. Today, we ought to be able to develop better indices of achievement.

## NEED FOR ALTERNATIVES

Sometimes educational reform is directed toward making schooling more efficient. Under those conditions, expected outcomes do not change, and assessment procedures may remain the same if they reflect those expectations. However, if expectations change, new assessment procedures must be developed. This can only be done by comparing and contrasting the old and new expectations, using the assessment tools designed for both, discarding those no longer appropriate, and developing new procedures when needed. Today, schools should be planning to change the emphasis from drill on basic mathematical concepts and skills to explorations that teach students to think critically, to reason, to solve problems, to interpret, to refine their ideas, and to apply them in creative ways.

I base the need for new assessment procedures which reflect those changes on four assumptions.

Assumption 1. We are now in a new economic age--The Information Age--which will significantly alter the character of American schooling.

Zarinnia and Romberg in chapter 2 of this monograph argued:

> The most important single attribute of the Information Age economy is that it represents a profound switch from physical energy to brain power as its driving force, and from concrete products to abstractions as its primary products. Instead of training all but a few citizens so that they will be able to function smoothly in the mechanical systems of factories, adults must be able to think. . . . This is significantly different from the concept of an intellectual elite having responsibility for innovation while workers take care of production. (pp. 23-24)

Assumption 2. Thinking skills must be the focus of instruction in mathematics.

Lauren Resnick (1987) has argued:

> Thinking skills resist the precise forms of definition we have come to associate with the setting of specified objectives for schooling. Nevertheless, it is relatively easy to list some key features of higher order thinking. When we do this, we become aware that, although we cannot define it exactly, we can recognize higher order thinking when it occurs. Consider the following:
>
> - Higher order thinking is nonalgorithmic. That is, the path of action is not fully specified in advance.
> - Higher order thinking tends to be complex. The total path is not "visible" (mentally speaking) from any single vantage point.
> - Higher order thinking often yields multiple solutions, each with costs and benefits, rather than unique solutions.
> - Higher order thinking involves nuanced judgment and interpretation.
> - Higher order thinking involves the application of multiple criteria, which sometimes conflict with one another.

- Higher order thinking often involves uncertainty. Not everything that bears on the task at hand is known.

- Higher order thinking involves self-regulation of the thinking process. We do not recognize higher order thinking in an individual when someone else "calls the plays" at every step.

- Higher order thinking involves imposing meaning, finding structure in apparent disorder.

- Higher order thinking is effortful. There is considerable mental work involved in the kinds of elaborations and judgments required.

This broad characterization of higher order thinking points to a historical fact that is often overlooked when considering the school curriculum, a fact that helps to resolve the question of what is new about our current concerns. American schools, like public schools in other industrialized countries, have inherited two quite distinct educational traditions--one concerned with elite education, the other concerned with mass education. These traditions conceived of schooling differently, had different clienteles, and held different goals for their students. Only in the last sixty years or so have the two traditions merged, at least to the extent that most students now attend comprehensive schools in which several educational programs and student groups coexist. Yet a case can be made that the continuing and as yet unresolved tension between the goals and methods of elite and mass education produces our current concern regarding the teaching of higher order skills. (pp. 2-3)

Assumption 3. Higher order skills are not to be learned after other skills.

Again, Resnick (1987) has stated:

The most important single message of modern research on the nature of thinking is that the kinds of activities traditionally associated with thinking are not limited to advanced levels of development. Instead, these activities are an intimate part of even elementary levels of reading, mathematics, and other branches of learning--when learning is proceeding well. In fact, the term "higher order" skills is probably itself fundamentally misleading, for it suggests another set of skills, presumably called "lower order," needs to come first. (p. 8)

Assumption 4. The three contemporary approaches to achievement testing (standardized tests, profile achievement tests and objective referenced tests) are conservative inhibitors to needed reform.

Les McLean (1982a) has stated that "achievement tests as we have known them are obsolete and teachers should discontinue their use as soon as possible" (p. 1). Peter Hilton (1981) argued even more strongly:

> What should we do to improve the situation? The answer is simple and obvious: avoid these glaring blemishes in the standard pedagogy! But it is not so easy in practice. We have given many reasons for the inertia in the system, for the remarkable stability of those practices which militate against effective mathematics education. Let us be explicit about one further potent factor in the preservation of the status quo--the standardized tests. These tests, beloved of (some) educational psychologists and (many) educational administrators, superimpose a further degree of artificiality on that which is already present in the curriculum. They force students to answer artificial questions under artificial circumstances; they impose severe and artificial time constraints; they encourage the false view that mathematics can be separated out into tiny water-tight compartments; they teach the perverted doctrine that mathematical problems have a single right answer and that all other answers are equally wrong; they fail completely to take account of mathematical process, concentrating exclusively on the "answer." Particularly perverse and absurd is the multiple-choice format. I have been doing mathematics now as a professional for nearly 40 years and have never met a situation (outside finite group theory!) in which I was faced with a mathematical problem and knew that the answer was one of five possibilities. Moreover if faced, artificially, by such a situation, my approach would, and should, be quite different from that in which I simply had to solve the problem.
>
> Tests tyrannize us--they tyrannize teachers and children. They loom so large that they distort the teaching curriculum and the teacher's natural style; they occur so frequently, and with such dire consequences, that they appear to the child (and, perhaps, to the teacher) to be the very reason for learning mathematics. (p. 79)

Lauren Resnick (1987) stated the case against standardized tests differently:

> Many of the higher order training programs aspire to types and levels of cognitive functioning to which standardized reading tests are not likely to be adequately sensitive. . . .
>
> Clearly, a most important challenge facing the movement for increasing higher order skill learning in the schools is the development of appropriate evaluation strategies.

> Part of the problem is our penchant for testing. American pressures for standardized testing, especially at the elementary and secondary school levels, make it difficult for curriculum reforms that do not produce test score gains to survive. But most current tests favor students who have acquired lots of factual knowledge and do little to assess either the coherence and utility of that knowledge or students' ability to use it to reason, solve problems, and the like. (pp. 33-34)

## CONCLUSIONS

To conclude this chapter, I emphasize four points.

First, the educational system as a whole and the teaching and learning of mathematics in particular need to be changed. Current reform efforts must encompass more than simple reactions to current weaknesses. To remedy weaknesses, we cannot return to the same methods of curriculum development, teacher training, and pupil assessment used in the past. Unless these, too, are changed, the same difficulties of sterile lessons, further deskilling of teachers, and so on will have been created.

Second, information on student performance is important for educational decisionmaking and the monitoring of the effects of change. It is not clear how influential test data and other data on students actually are in educational decisionmaking. Most educators certainly believe that test data has a strong influence. Whether this is myth or reality is not clear. However, there is no question that valid data could and should influence decisions. Clearly, if the content of courses and methods of instruction change, the monitoring of student achievement is necessary if the effects of these changes are to be determined.

Third, current testing procedures are unlikely to provide valid information for decisions about the current reform movement. Current tests reflect the ideas and technology of a different era and world view. They can not assess how students think or reflect on tasks, nor can they measure interrelationships of ideas.

Finally, work needs to be started on new assessment procedures. Only by having new assessment tools can we provide educators with appropriate information about how students are performing with respect to the goals of the reform movement.

## References

Ayres, L. P. (1918). History and present status of educational measurements. In S. C. Parker (Ed.), The measurement of educational products: Seventeenth yearbook of the National Society for the Study of Education (Part II). Bloomington, IL: Public School Publishing Co.

Bloom, B. S. (Ed.) (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook 1. Cognitive Domain. New York: Longmans.

Crosswhite, F. J., et al. (1986). Second international mathematics study detailed report for the United States. Champaign: Stipes.

Freeman, F. N. (1939). Mental tests: Their history, principles and applications (rev. ed.). Boston: Houghton Mifflin.

Gorth, W. P., Schriber, P. E., & O'Reilly, R. P. (1974). Comprehensive achievement monitoring: A criterion-referenced evaluation system. New York: Educational Technology Publishers.

Greene, H. A., Jorgensen, A. N., & Gerberich, J. R. (1953). Measurement and evaluation in the elementary school (2nd ed.). New York: Longmans.

Hilton, P. Avoiding math avoidance (1981). In L. A. Steen (Ed.), Mathematics tomorrow (pp. 73-82). New York: Springer-Verlag.

Lipson, J. I., Koburt, E., & Thomas B. (1967). Individually prescribed instruction (IPI) mathematics. Pittsburg: Learning Research and Development Center.

Lorge, I., & Thorndike, R. (1954). The Lorge-Thorndike Intelligence Tests, Verbal Battery (Level 3, Form A). Boston: Houghton Mifflin.

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

McLean, L. D. (1982b). Report of the 1981 field trials in English and mathematics -- Intermediate Division. Toronto, Ontario: The Minister of Education.

National Coalition of Advocates for Students. (1985). Barriers to excellence: Our children at risk. Washington, DC: Author.

Odell, C. W. (1930). Educational measurements in high school. New York: Century.

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

Romberg, T.A. (1976). Individually guided mathematics. Reading, MA: Addison-Wesley.

Romberg, T. A. (1983). A common curriculum for mathematics. In G. D. Fenstermacher & J. J. Goodlad (Eds.), Individual differences and the common curriculum: Eighty-second yearbook of the National Society for the Study of Education (Part I). Chicago: University of Chicago Press.

Romberg, T. A., Harvey, J. G., Moser, J. M., & Montgomery, M. E. (1974, 1975, 1976). Developing mathematical processes. Chicago: Rand McNally.

Spearman, C. (1904). 'General intelligence' objectively determined and measured. American Journal of Psychology, 15, 201-293.

Thorndike, E. L. (1904). An introduction to the theory of mental and social measurements. New York: Teachers College, Columbia University.

Tyler, R. W. (1931). A generalized technique for constructing achievement tests. Educational Research Bulletin, 8, 199-208.

Vergnaud, G. (1982). Cognitive and developmental psychology and research in mathematics education: Some theoretical and methodological issues. For the Learning of Mathematics, 3(2), 31-41.

Watson, G. (1938). The specific techniques of investigation: Testing intelligence, aptitudes, and personality. In G. M. Whipple (Ed.), The scientific movement in education: Thirty-seventh yearbook of the National Society for the Study of Education (Part II, pp. 365-366). Bloomington, IL: Public School Publishing.

Wisconsin Mathematics Assessment Committee (1974). Interpretive report on Wisconsin state mathematics assessment. Madison, WI: Department of Public Instruction.

Zarinnia, E. A., & Romberg, T. A. A new world view and its impact on school mathematics. In T. A. Romberg & D. M. Stewart (Eds.), The monitoring of school mathematics: Background Papers. Vol. 1: The monitoring project and mathematics curriculum (pp. 21-61). Madison: Wisconsin Center for Education Research.

## Chapter 18

# CONSEQUENCES OF THE NEW WORLD VIEW TO ASSESSMENT OF STUDENTS' KNOWLEDGE OF MATHEMATICS

Thomas A. Romberg and Anne Zarinnia

In this paper we consider the consequences of the emerging world view on assessment of students' knowledge of mathematics and their ability to use that knowledge both creatively and routinely in solving the variety of problems encountered in the course of life. In chapter 2, we argued that metaphor and model cause the prevailing world view to exert enormously powerful forces over people's thoughts and activities. We pointed out the emerging characteristics of the new view of the world, including the fact that intellectual conflict between the old and the new is impeding any serious progress toward curricular improvement. The crucial point was that the world is changing so rapidly that, unless those involved in mathematics education adopt a proactive view and develop a new model for the twenty-first century, the mathematical understanding of children will remain permanently inadequate and a source of trauma.

An implicit premise of this project is that assessment, which has usually involved some testing procedure, has an impact on curriculum and instruction, if only by demanding and providing information. It is openly acknowledged that the content emphasis of assessment has a direct impact both on what is taught and how it is taught. The school outcomes sought determine curricular elements to be assessed and monitored, and that which is monitored is almost inevitably emphasized. Thus, the selection of indicators cannot be regarded as neutral (Oakes, 1986), and monitoring is an instrument of reform (National Science Board Commission on Precollege Education in Mathematics, Science, and Technology, 1983).

This chapter argues that the nature, forms, purposes and design of major models of assessment are dominated by the prevailing, old world view, helping to perpetuate it, and that there is an iterative relationship which inhibits change. In particular, as Romberg argues in chapter 17, this is true for Profile Achievement Tests, which are the type of assessment procedure most applicable for monitoring purposes. If assessment of progress toward a new curriculum is dominated by the forms and functions of the old-world view, progress toward a new curriculum will be impeded by the process of assessment itself. Consequently, it is essential to lay bare the ways in which contemporary assessment procedures, particularly group-profile testing procedures, are redolent of the old world view and to point to alternatives.

To develop this argument, let us begin by examining the current framework for the profile assessment of knowledge -- content by behavior matrices. This will be followed by our argument as to why this approach is no longer appropriate in the light of the new world view. In turn, we then summarize new directions and new partial models before drawing conclusions for this monitoring project.

## CONTENT-BY-BEHAVIOR MATRICES

### Introduction

As argued by Romberg in chapter 17, Profile Achievement tests are comprised of items that reflect the combination of two classifications. One is related to the content of the items, the other to the behavioral outcomes response. Classification is a fundamental intellectual activity that underlies most practical and theoretical activities. The role of classification in practical activities, such as sorting the laundry, is self-evident; objects, both concrete and intellectual, are sorted into convenient groupings. However, efforts to formulate laws of nature also involve stating the relationships between members of different classes. To pursue the laundry analogy, reds are washed separately from whites to avoid the anathema of pink undershirts. In other words, "a taxonomy not only classifies phenomena; it also orders them, and it must be a satisfactory enough tool to reveal significant relationships between the phenomena" (Romberg & Kilpatrick, 1969, p. 282). In science,

> formulation of laws presupposes classification. . . . While every theory presupposes a classificatory scheme, this scheme is in turn influenced by the content of the theory. . . . The investigator will frequently have to develop his own classificatory scheme rather than take over from a developed explicit theory. The place of the theory is taken by a provisional model or scheme of the whole situation in which the inquiry has taken place. Use of such a model suggests that a classificatory scheme is required that, when modified as a result of inquiry, will in turn suggest modifications of the model. (Korner, 1976, pp. 5-6)

Succinctly, a particular classification is a schematic model of its underlying theory; taxonomy reflects theory. Nomenclature, in turn, depends on taxonomy. Thus, an established nomenclature tends to preserve the principles of the taxonomy which it describes, and both combine to indirectly perpetuate the theory on which the taxonomy was based (Korner, 1976).

Classification as a logical process is essentially a matter of partitioning a domain into sets and subsets, culminating in a class containing a single, unique member. The process typically relies on the assumption that each set is extensionally definite, although in practice this causes problems. Division into subsets is appropriate only if no two subsets have anything in common, and all of the subsets together contain all the members of the partitioned set. In other words, the subsets are mutually exclusive and jointly exhaustive (Korner, 1976).

Prototype theory (Cohen & Murphy, 1984) highlights one of the problems with deterministic classification of concepts. Some examples of a concept are less typical than others. A thistle head is less obviously a flower than is a rose. Both have a delightful fragrance, form, and color; both have uncomfortable prickles for the unwary (a nonflower concept). One is cultivated, the other killed. In other words, any concept involving more than one case almost inevitably becomes a system of concepts and consequently a fuzzy set. Wohlwill (1973) pointed out, for example, that the behavioral classifications of the stage theory of cognitive development (Piaget, 1973) rely on the underlying assumption that there is synchronous passage from one stage to the next in the various facets of behavior. Temporary lags in one aspect or another suggest that (a) the model should be modified to encompass time-lagged relationships, or (b) the theory is based on unwarranted assumptions, or (c) basing the theory on the notion of extensionally definite sets is inappropriate.

In summary, classification of objects in a domain starts with the broadest, most inclusive categories and progressively subdivides. At any given level in the resultant subsets of subject matter, categories are theoretically mutually exclusive. Each subset may be subdivided according to some principle of internal coherence until a set containing only one object is reached. Equally, subsets may be recombined to reform the initial set. This process of ordered set division, the larger set being an aggregation of its own subsets, is the organizing principle of all hierarchies. It is a method of analysis that has been used on everything from land forms to library collections. In the process of outlining the work of students and teachers, the principles of classification have also been applied to both the organization and the sequencing of the content to be taught and learned (e.g., Thorndike, 1904; Tyler, 1931) and, with the behavioral objective movement, to the behaviors exhibiting orders of understanding (e.g., Bloom, 1956).

For example, precisely such a comprehensive coverage of the mathematics curriculum was a stated goal of the first IEA (Husén, 1967), the aim being not to pass judgment on an individual student but to survey cognitive achievement without using a predetermined standard. Therefore, in planning the battery of tests, the field of school mathematics was viewed as a whole, and traditional classifications of mathematics were used to ensure inclusion of all subfields.

As an organizing tool, classification of both content and behavior is well illustrated in the major mathematical evaluations around which the following discussion will be focused. These are:

1. Husén, T. (Ed.). (1967). International Project for the Evaluation of Educational Achievement (IEA). Phase I. International study of achievement in mathematics: A comparison of twelve countries. (Volumes I-II).

2. Weinzweig, A. I., & Wilson, J. W. (1977, January). Second IEA mathematics study. Suggested tables of specifications for the IEA mathematics tests. Working paper 1.

3. Romberg, T. A. & Wilson, J. W. (Eds.). (1969). The development of tests. NLSMA reports No. 7.

4. National Assessment of Educational Progress. (1981). Mathematics objectives: 1981-82 assessment.

5. Carpenter, T. P., Coburn, T. G., Reys, R. E. & Wilson, J. W. (1978). Results from the First Mathematics Assessment of the National Assessment of Educational Progress.

6. Carpenter, T. P, Corbitt, M. K., Kepner, H. S., Lindquist, M. M. & Reys, R. E. (1981). Results from the Second Mathematics Assessment of the National Assessment of Educational Progress.

## Content

Classification of mathematical content typically depends on the identification of mathematical objects and their attributes. At the broadest level, categories of mathematical content are a convenient way of dividing knowledge into such large chunks as semester courses, textbooks, and major examinations. At an intermediate level, the categories may be used to organize chapters in the textbook or weeks in the course. At an even more specific level, small independent categories of content are the organizing principle for parts of the daily lesson plan, a unit in the text, or a homework assignment. Such categories are advantageous in that they break work into manageable chunks and restrict teaching to the presentation of a clearly defined segment of the content. By structuring content into a hierarchy, it is possible to ensure comprehensive coverage of the subject, whether in teaching, testing, or learning.

Unfortunately, the classifications on which the sequencing of instruction and consequent assessment have been based are largely spurious, a means toward the linear ordering of work. Note that most instructional sequences have been constructed for purely practical reasons and are not true hierarchies. Often strands, and subjects within strands, are specified, but no conceptual or psychological dependence is apparent or assumed. If a strict partial ordering of the segments could be found, a content hierarchy might be constructed. However, if the structure of instruction and assessment is to have a positive influence, mathematical content must be arranged, where appropriate, in true hierarchies based on the interdependence of skills and concepts. Two approaches to this problem have emerged, facility hierarchies (Hart, 1980) and conceptual fields (Vergnaud, 1982, 1983a, 1983b).

## Behaviors

The power of classification as a logical organizer also appealed to college examiners looking for a theoretical framework to facilitate communication. Thus, at the 1948 American Psychological Association Convention in Boston, after considerable discussion, there was agreement that such a theoretical framework might best be obtained by classifying the objectives of the educational process. The resultant taxonomy and nomenclature were intended to improve communication in the community because the objectives provided the basis around which curricula and tests could be built (Bloom, 1956). The proposal rested on the premise that educational objectives stated in behavioral forms have their counterparts in the behavior of individuals, which can be observed, described and, therefore, classified. However, fear was expressed that:

> It might lead to fragmentation and atomization of educational purposes such that the parts and pieces finally placed into the classification might be very different from the more complete objective with which one started. (Bloom, 1956, pp. 5-6)

Nevertheless, it was felt that the structure of the hierarchy would enable users to understand clearly the place of objectives in relation to each other. Consequently, the taxonomy was formally presented at the Chicago meeting of the American Psychological Association (APA) in 1951. It was subsequently published (Bloom, 1956) and incorporated into the plan for a large-scale cross-national study of mathematics presented by Bloom at Eltham and Hamburg in 1958 (Husén, 1967). The taxonomy of behaviors complemented the classification of content as an organizing tool. As a result, the principles of taxonomy formed the basis for a matrix model of assessment which ensured comprehensive coverage of both behavior and content in the first IEA.

Bloom's (1956) taxonomy first divided educational objectives into three domains: the cognitive, the affective, and the psychomotor. Only the first two were incorporated into the IEA (Husén, 1967), and those with somewhat different strategies because they were the responsibility of different committees. The committee charged with the affective domain distributed questionnaires covering both attitudes and descriptions of the learning environment. That responsible for content viewed instructional objectives as having three dimensions: the behavior to be demonstrated (cognitive, affective, and psychomotor); the content; and a field of application. Because another committee was responsible for the affective aspects, the content committee eventually agreed on a primarily cognitive list.

1. Lower mental processes (use or repetition of learned intellectual activity)

    a. Knowledge and information: recall of definitions, notation, concepts

    b. Techniques and skills: solutions

2. Transitional processes (higher or lower, depending on the novelty of the context)

    c. Translation of data into symbols or schema and vice versa

3. Higher mental processes (demanding lines of thought not previously used)

    d. Comprehension: capacity to analyze problems, follow reasoning

    e. Inventiveness: reasoning creatively in mathematics

The committee's content-by-behavior matrix showing the number of test items for each category of content and each kind of behavior does not follow a breakdown identical to its short list of objectives. Despite this, it is clear that, while over 40% of items tested the lowest level in the taxonomy, fewer than 3% tested inventiveness.

Even more clear is that, as a theoretical framework for ensuring a comprehensive approach to both content coverage and range of behavioral objectives, the content-by-behavior matrix is a powerful organizing structure. It enables a rapid overview of the entire structure and of the relative emphases on one part or another. Consequently, despite modification of the specifics on each axis, the matrix approach persisted in subsequent evaluations. It was integral to the model of mathematics achievement in the National Longitudinal Study of Mathematical Abilities (NLSMA) (Romberg & Wilson, 1969, pp. 29-44) and the National Assessments of Educational Progress (NAEP, 1981).

NLSMA, for example, originally considered "an eleven-by-seven content-behavior matrix" (Romberg & Wilson, 1969, p. 35). However, content was combined and reduced to three categories: number systems, geometry, and algebra. The behavioral axis was consolidated from seven categories to four: computation, comprehension, application, and analysis. In the second IEA, Weinzweig and Wilson (1977) recommended a matrix identical to the NLSMA on the behavioral axis but subdivided into nine categories on the content axis. By comparison, the second NAEP, using a content-by-process matrix, divided the behavioral axis (process) into knowledge, skill, understanding, and application. The third assessment expanded application to include problem solving and added an attitude category (NAEP, 1981; Romberg, 1986). Nevertheless, specific modifications of the content or behavioral axes and change in nomenclatu_e from "behavior" to "process" (Romberg & Wilson, 1969, p. 38) are not important. Persistence of the matrix as a tool for organizing activity is important and probably reflects:

1. its power as an organizing tool;

2. its visual facility;

3. the strong continuity between assessment projects created by relying on those with the most relevant experience in the field when planning the next project.

## Items

A practical problem of testing is that any test attempting to be comprehensive in approach takes a long time for children to complete and a long time for teachers to grade. Consequently, those designing the first IEA (Husén, 1967) had to resolve the conflict between time and practicality. The European countries almost all used complex items with an open response format, while the United States typically used a collection of short tasks. Responses to the tasks were controlled by a multiple-choice technique. It was not claimed that the two approaches measured the same thing. However, the controlled, multiple-choice response offered advantages:

1. It made possible much more extensive and representative sampling of the content topics because it tested more topics less deeply.

2. It was easy to score.

3. It was cheaper and faster than scoring open-ended responses.

4. Questions could be designed to stand alone and test a specific objective.

5. Because the items were classified according to location in the matrix, a more detailed profile of groups of students became possible.

6. Item design was philosophically congruent with the

theoretical model for evaluation; they were both constructed around a matrix.

The matrix model is now so widespread that it is accepted virtually without question and is the framework within which item-banks of questions are compiled to test the concepts in a given cell of a matrix (Wisconsin Department of Public Instruction, 1986).

### Psychometrics

Items for assessment not only had to be judged appropriate for a particular cell of the matrix, they also had to have certain psychometric properties. Ideally, each was to be of moderate difficulty ($p$ values between .4 and .8) and related to other items in that cell (positive, nonspurious biserial correlation greater than .30). These criteria were adopted from those used to select items for standardized tests. They ensure variability and discriminability of scales derived from those sets of items (Romberg & Wilson, 1969). Unfortunately, the items which meet these criteria contribute to the questionable validity of profile achievement tests.

### Summary

The deep structure of the theoretical model implicit in major evaluations of mathematical education is based on a matrix of taxonomies of content and behaviors. The convenience and power of the model is reflected in its persistent use in the face of changing circumstances.

## DISCONTENT

### Introduction

Noting the failure of the mathematics reform efforts of the 1960s and early 1970s, Westbury (1980) argued that change involves the abandonment of practices, as well as their adoption. The deep structures of formal and informal institutional apparatus, procedures, forms, and rituals tend to preserve the status quo, frustrating efforts at curricular reform. However, just as students have difficulty in learning because they fail to modify old conceptions, so ingrained theoretical structures carry an intellectual baggage that impedes change.

A new cohesion between the goals of education, its practices, and the methods of assessment, which would promote educational change rather than stifle it, therefore depends on divestiture of old styles of thinking. For that to happen, there must first be a recognition of the ways in which the concepts of the old world view dominate the deep structure of present evaluation. Despite long-standing and growing concern, the values and forces that dominated mathematical education twenty years ago are embedded in the theoretical structures of prevailing methods of assessment.

## Behaviorism

Behaviorism reflects the application of the engineering approach of scientific management to the problems of education. Scientific management rested on three basic principles: specialization of work through the simplification of individual tasks, predetermined rules for coordinating the tasks, and detailed monitoring of performance (Reich, 1983). These microprinciples pervaded American education with the same thoroughness with which they were applied in the economy. They dominated the breakdown of knowledge, the roles of teacher and students, instructional and administrative processes, the building-block approach of Carnegie units, the content and structure of textbooks, belief in the textbook as an effective tool for transmitting content, the structure of university education, and monitoring and evaluation. Hence emerged the notion of progress through the mastery of simple steps, the development of learning hierarchies, explicit directions, daily lesson plans, frequent quizzes, objective testing of the smallest steps, scope and sequence curricula.

Unfortunately, these are only the more obvious aspects. One consequence of such meticulous planning is that it renders the unplanned unlikely. A second is that a system designed to eliminate human error and the element of risk also eliminates innovation. A third is that, like factory work, it is crashingly dull, uninspiring, and unmemorable except for its boredom, for personal involvement and the mnemonics of the unexpected are nonexistent.

Bloom's taxonomy of educational objectives epitomizes the domination of American education by scientific management, for it completed the process by which not only the content of learning but the proxies for its intelligent application were classified, organized in a linear sequence and, by definition, broken into a hierarchy of mutually exclusive cells. The consequences in the classroom were far reaching. Scope and sequence charts prescribed which parts of a subject were to be covered in what order; each cellular part of each subject was put into a matrix (e.g. Romberg & Kilpatrick, 1969, p. 285); behaviors suggesting desirable intellectual activity were also sequenced. However, given the multiplicity of subject cells to be covered, the easiest way to finish the prescribed course of study was to simply cover content without worrying too much about thought. Furthermore, matrices are difficult to construct effectively on paper in more than two dimensions. Consequently, few scope and sequence charts addressed both levels of thinking and specific aspects of content within an overall discipline in a very coherent manner. Thus, a focus of concern in documents addressing the quality of education has been the failure of students to reach the "higher order intellectual skills" (National Commission on Excellence in Education, 1983, p. 9).

Recognition of this failure is a reflection of the incorporation of Bloom's taxonomy into the fabric of national and international evaluation and, by implication, a tacit expression of the depth of penetration of scientific management. It also indirectly reflects perceived inadequacy of the stimulus-response philosophy as a model of human behavior.

The continued dominance of behaviorism and scientific management over the thinking of leading mathematics educators is reflected in the persistence of Bloomian content by behavior analyses in the second IEA and in both NLSMA (Figure 1) and the activities of NAEP (Figure 2). Continuation of this pattern would be catastrophic because it would suggest that those responsible for evaluation have failed to take cognizance of the power of the deep structures to constrain curriculum development (Westbury, 1980) through the implicit goals suggested by the form and content of evaluation.

Attacking behaviorism (e.g., Suppes, 1965) as the bane of school mathematics, Eisenberg (1975) criticized the dubious merit of a task-analysis, engineering approach to curricula because it essentially equates training with education, missing the heart and essence of mathematics. Expressing concern over the validity of learning hierarchies, he argued for a reevaluation of the objectives of school mathematics. The goal of school mathematics is to teach students to think, to feel comfortable with problem solving, to help students question and formulate hypotheses, investigate, and simply tinker with mathematics.

Persistence of Bloom's intel ctual model is also reflected in the continued use of associated nomenclature. Use of the term higher order thinking, for example, directly expresses reliance on Bloom's taxonomy for the theoretical model. This is of particular concern because of its associated intellectual baggage; it implies that lower order thinking precedes higher thinking processes. However, activities associated with higher order thinking are not limited to advanced levels of development. Failure to stress higher order features of thinking because of the belief that a lower order must be attended to first is a source of major learning difficulties. In reading, for example, cognitive science has suggested that "processes traditionally reserved for advanced students . . . might be taught to all . . . especially those who learn with difficulty" (Resnick, 1987). This approach is subliminally impeded by continued reliance on nomenclatures and models of assessment that have Bloom's taxonomy as their underlying construct.

| | Number Systems | Geometry | Algebra |
|---|---|---|---|
| Computation | | | |
| Comprehension | | | |
| Application | | | |
| Analysis | | | |

Figure 1. NLSMA Model for Mathematics Achievement (from Romberg and Wilson, 1969, p. 44).

CONTENT

| | A. Numbers and Numeration | B. Variables and Relationships | C. Shape, Size and Position | D. Measurement | E. Statistics/ Probability | F. Technology |
|---|---|---|---|---|---|---|
| I. Mathematical knowledge | | | | | | |
| II. Mathematical skill | | | | | | |
| III. Mathematical understanding | | | | | | |
| IV. Mathematical application and problem solving | | | | | | |
| V. Attitudes toward mathematics | | | | | | |

Figure 2. Objectives Framework for Third NAEP Assessment (from NAEP, 1981, p. 10).

Content

By definition, classifications of knowledge, whether for the purpose of organizing the curriculum or for monitoring curricula, make an implicit statement of theory. Statements of curricula that focus around knowledge broken into subjects for study, such as mathematics into algebra, geometry, etc., have the immediate impact of stating that:

1. Knowledge can be broken into clearly defined, independent, self-sustaining parts;

2. Such an approach is important, more important than any other approaches which might follow;

3. There is a logical sequence of development in which each part builds on a preceding foundation;

4. It is important to know about the divisions of knowledge enumerated.

Such implicit assumptions may be unwarranted if, for example, knowledge is regarded as unitary and emphasis is on knowing rather than knowing about. The approach may also be unsuitable if there is genuine concern with application and problem solving. Stated simply, purpose should suggest form, and form implies purpose; incoherence may be inferred from anything less.

For instance, for schools dominated by traditional curriculum engineering, the IEAs, NLSMA, and the NAEPs are tightly coherent. However, the model around which they were built is congruent neither with the efforts of those schools which construed the purpose of mathematics differently from those designing the evaluations, nor with the purpose of schools trying to reform the mathematics curriculum. For example:

> Criticisms of the content outline and thus of the international grid . . . were partly due to the fact that for some National Committees the order and grouping of the topics in the outline were thought to imply an underlying philosophy or instructional treatment different from that commonly espoused in their particular country. The proportion of content which was common to all countries was considerable, but wording or placement in the content outline caused some National Committees to express doubts about the validity of the proposed grid for their curricula. (Garden, in press)

Thus, discussing the validation of the cognitive instruments in the second IEA, Garden (in press) made explicitly clear the continued dominance of the Bloomian model and the difficulties experienced by the Belgians in relating the study content to their conception of mathematics as a field of inquiry.

Disagreement over the precise structure and arrangement of content in a grid is only part of the problem. Westbury (1980) pinpointed a more fundamental concern: the difference between the intellectual structure of a discipline and its institutional structure in schools, where it is an administrative framework for tasks. The consequence is that administrative stability impedes intellectual change. For similar reasons, Romberg (1985) described mathematics in schools as a stereotyped, static discipline, in which the pieces have become ends in themselves. A similar response to the impact of scientific management and behaviorism on mathematics as a school subject is Scheffler's (1975) denunciation of the traditional, mechanistic approach to basic skills and concepts:

> The oversimplified educational concept of a "subject" merges with the false public image of mathematics to form quite a misleading conception for the purposes of education: Since it is a subject, runs the myth, it must be homogeneous, and in what way homogeneous? Exact, mechanical, numerical, and precise--yielding for every question a decisive and unique answer in accordance with an effective routine. It is no wonder that this conception isolates mathematics from other subjects, since what is here described is not so much a form of thinking as a substitute for thinking. What is in point is the process of calculation or computation, the deployment of a set routine with no room for ingenuity or flair, no place for guesswork or surprise, no chance for discovery, no need for the human being, in fact. (p. 184)

Item Independence

The single most severe criticism of objective test questions designed to assess a specific item of content at a specific level of behavior is that they trivialize learning and knowledge (Berlak, 1985). This is almost inherent to such questions for several reasons. First, they are designed to test a single, specific objective, clearly defined in the matrix. Thus, elements in the multiple-choice format are designed so that the candidate can pick an answer which is sufficiently specific to unequivocally demonstrate the sought behavior. This tends to eliminate synthesis between content and behavior. Second, the very nature of objective tests, which ask the user to choose among alternatives, eliminates creativity in answering. Even the intent militates against creativity in answering because it is microanalytical rather than synthetic or creative.

Frederiksen (1984) observed that a multiple-choice format does not measure the same cognitive skills as a free-response form, and that "efficient tests tend to drive out less efficient tests, leaving many important abilities untested--and untaught" (p. 201). One example of a desirable outcome untested and untaught is the ability to cope with ill-structured problems, which are not found on standardized achievement tests.

A less obvious impact was observed by the Assessment of Performance Unit (Cambridge Institute of Education, 1985). The multiple-choice format is an interventional mode of questioning which appears to offer a greater chance for success in situations where the student is unfamiliar with the material. However, in other situations, students benefited from the opportunity to think, achieving greater success with the free-form response.

Another aspect of most objective tests is that, even though some questions may be designed to test lower level thinking and others are designed to evaluate higher thought processes, they are usually tested independently of each other, allowing little notion of a child's approach to a given problem.

In addition to their direct effects, such tests exert powerful indirect effects on both the style of teaching and the style of learning. When one studies for an essay exam, one progressively surveys and synthesizes, putting the parts together and developing a mental model of the structure of the subject. One also develops points of view, arguments to advance and support, for those are the expectations. By contrast, in studying for an objective, multiple-choice test, one learns to cover the parts and make fine distinctions between alternative ways of stating the same thing in order to distinguish a "right" answer from a "wrong" one, the implication being that there is a single right answer. In other words, the one requires that students create their own models of mathematics, the other reinforces the view of mathematics as a ground to be covered.

## Summary and Conclusion

The intent of the content-by-behavior matrix is in every respect hierarchical. It leads to ranking of those assessed by standardized tests according to their position on a normal curve, with the result that

> despite the lip service we pay to the myriad ways in which individuals differ. . . . [i]t is the performance on these tests--with their narrow and rigid definition both of when children should be able to perform particular skills and how they should be able to exhibit their knowledge--that determines whether we see children as "okay" or not. In the process we damage all children--we devalue the variety of strength they bring with them to school. All differences become handicaps. (National Coalition of Advocates for Students, 1985, p. 47)

It is easy to be dispassionate about a theoretical model. However, the accompanying objective testing invariably results in poor, minority and handicapped students placing at the low end of the curve. It stamps with failure the groups most dependent on the educational system for improvement and acts as a dangerous social filter.

Unfortunately, it is incredibly difficult to shrug off old habits. For example, the architecture of current evaluations, the two-dimensional content-by-behavior matrix, is a seductively convenient model for organizing information visually; occasionally a three-dimensional version expands the possibilities (e.g., Carpenter et al., 1978; Foxman, Cresswell, & Badger, 1981) but increases the conceptual load and so is used less (e.g., Carpenter et al., 1981). The intellectual consequences of using a two-dimensional matrix bear thought. It encourages a tendency to tacitly view successive cells in a row or column as entities having a sequential and linear relationship to each other. It also causes visual separation of nonadjacent cells, subliminally interrupting perception of relationships between them. If such relationships do exist, the visual patterns of the matrix have a powerful, often mnemonic, impact. If not, the framework is not inert, it suggests relationships which are simple, numerically restricted, and linear. Persistence of the matrix form is likely to continue as long as information is presented on paper. However, the potential of electronic data bases and computer-based modeling suggests that multiple viewpoints may be more revealing and less constraining.

In a stable situation in which there is coherence between purpose, curriculum, and evaluation, testing what has not been taught is ludicrous. However, in a situation where there is dissatisfaction with both what is taught and what is not taught, where change is sought, it is vital to consider the purpose of teaching mathematics and to test for what is sought regardless of whether it is taught. The purpose of teaching mathematics is no longer computation and routine algorithms; anything that can be reduced to a routine algorithm can now be done by computers. However, while purpose has changed, the content and structure of evaluation remain the same.

> As long as segmented structures and segmentalist attitudes make the very idea of innovation run against the cultural grain, there is a tension between the desire for innovation and the continued blocking of it by the organization itself. (Kanter, 1983, p. 75)

The key issue is that structures constrain, whether they are subject structures, behavioral structures, or theoretical frameworks for testing and assessment. Modifying the detail of a structure is merely an exercise in fine tuning. Fine tuning the theoretical structure of the content-by-behavior framework by modifying content, or by seeking ways of attaining higher levels on the behavioral axis, can only be ameliorative. Substantive progress will be accomplished only through a remodeling of the fundamental theoretical framework.

# NEW PURPOSE: MANAGING COMPLEXITY

## Introduction

One of the most notable features of the existing framework for assessment of mathematics achievement in the United States is its logical congruence with the world view of science and society that has existed during this century, and thus with the intellectual structures and purposes dominating the education it was designed to assess. That coherence contributed largely to the inherent intellectual power of the content-by-behavior matrix as a theoretical model. If alternative frameworks for assessment are to be equally powerful, they must be equally congruent with the forces requiring their construction, namely the changed views on science and society.

Traditionally, developments in mathematics and mathematical education have been coherent with the prevailing philosophy of science. At the root of scientific management and the matrix model of assessment were the concepts of classical dynamics. The classical view emphasized stability, order, uniformity, equilibrium, linear relationships, results proportional to input and lawfully predictable from the current state of the system, and the separation of theory and technology. Hence, the intellectual foundations of the content-by-behavior matrix were rooted in a view of processes as linear, stable, uniform, equilibrial, and proportional. Thus, analysis was acceptable as the dominant intellectual tool. It was, according to Prigogine and Stengers (1984), "a world in which the only events which could occur were those deducible from the instantaneous state of the system" (p. 225). From the classical standpoint, it was perfectly feasible to analyze, isolate, experiment, and deduce.

This stable, linear, hierarchical approach was also the dominant social philosophy. That fact was reflected in, for example, managerial, political, and ecclesiastical organization, and in the rank-ordering philosophy of assessment practices. However, these traditional views are now being regarded as too simple to account for complex reality, whether in science or social organization. For example, uncertainty and instability are part of reality but the bane of short-term economic forecasting (e.g., Clark, 1986). Alternative and radically different models of knowing (e.g., Bohm, 1983) and learning (e.g., Kuyk, 1982) have been proposed which incorporate the most recent views of physics and mathematics. In areas such as classification (e.g., Farradane, 1980a, 1980b) and truth (e.g., Rescher, 1979) and in practical problems facing mankind, such as global warming, a more coherent approach is sought to cope with complexity. The search for tools to handle complexity is clearly reflected in the intellectual trends of mathematics (National Research Council, 1984):

1. the concern with nonlinearity;

2. the increased role of discrete mathematics, essential to network node location and the distribution of information;

3. the increased role of probabilistic analysis;

4. the development of large-scale scientific computation.

Even more significant in the effort to handle complexity is a double movement toward a new coherence in the mathematical research community: internally, toward unifying ideas, blurring the boundaries so that diverse mathematicians again participate in a common enterprise; externally, toward interaction with science and technology (National Research Council, 1984). Jaffe (1984) pointed to the reunification of mathematics with theoretical physics and its revolutionary consequences. This reunification with other sciences is best illustrated by the concurrent but independent development by meteorology, genetics, and theoretical physics of nonlinear mathematical models (Hofstadter, 1986), which thereby illustrates Jaffe's (1984) view of the iterative relationship between excellent mathematics and practical application.

In chaos theory, for example, Prigogine and Stengers (1984) created a synthesis between linear and nonlinear causality; singular anomalies, normally ignored for the purpose of abstraction in the classical approach, under conditions far from equilibrium play a disproportionate role when part of a reactive loop. Under such conditions, there are no universally valid laws on which predictions can be based. Random and irreversible events reach a threshold at which bifurcation takes place and there is, unpredictably, either new order or further disorder. Thus the new view of science blends the linear and the circular; it emphasizes probability and stochastic processes, the importance of chance. Because apparently minor events can have disproportionate results, it renews the importance of practice as a source of theory. At a practical and an intellectual level, the individual is no longer doomed to insignificance.

Pask's (1984) characterization of an overused concept illustrates one problem of long-term stability in a theoretical model. When a concept is first learned and applied, the user can describe how it was conceived and used. However, as its application becomes automatic, the concept becomes ingrained and rigid because there is no longer a conscious transfer of information to link procedures. For the person who has learned to ride a bike, for example, only disturbance of the equilibrium, prompted by the need to learn something novel like riding a tandem, or by the desire to teach somebody else, will renew awareness of the application of the concept. This is essentially the case with the content-by-behavior framework of assessment, which persisted in the face of a changing world view. Even advocates of new assessment schemes (practical tests, problem solving, etc.) such as the Assessment of Performance Unit (APU) (Foxman, 1985) have used a content-by-behavior matrix (Romberg, 1986). Nevertheless, the real significance of the situation lies in the principle of complementarity:

> No single theoretical language articulating the variables to which a well-defined value can be attributed can exhaust the physical content of a system. Various possible languages and points of view about the system may be complementary. They all deal with the same reality but it is impossible to reduce them to one single description. (Prigogine & Stengers, 1984, p. 225)

In other words, the basic problem is not that the matrix model was wrong, but merely that it was inadequately simple and insufficiently flexible to accommodate new theoretical developments. It makes the assumption that items, cells, columns, and rows are independent.

In any model, cohesion comes from purpose. The tight cohesion of the content-by-behavior framework came from the intent to assess and, implicitly, sort children according to their knowledge about mathematics and, secondarily, by their ability to think. It was, and is, a quantitative, linear model of content, process, and people. The content-by-behavior matrix of evaluation does not question the purpose for teaching mathematics, it reflects purpose. It also derives from congruence between purpose and intellectual tools. The purpose was to sort people into linear rankings of extensionally definite sets, which is precisely what a matrix does.

Evaluation and assessment in a stable paradigm may take for granted the purpose for teaching and the philosophical foundations of the subject under evaluation, but in a period of major societal change, such nonchalance is unwise. Not only has stability been a relative matter in this century, but the new world view specifically rejects the consequences of old cohesion, of which the content-by-behavior matrix is a microcosm. Because the stated purpose is no longer to rank-order, but to cooperate in the creation of knowledge, that concept should become the cohesive force of any new theoretical model. It is, furthermore, a qualitative, rather than quantitative, concept.

As argued in chapter 2, the new world view is, above all, integrative; it sees everything as part of a larger whole, with each part sharing reciprocal relationships with other parts. It seeks a rational balance between education and training, between cooperation and individual effort, between the development of intelligence and its measurement, between the integration of intuitive and analytical thinking and an exclusive stress on the analytical, and between constant learning for the purpose of innovation and adaptability as opposed to one-time schooling for life. The new world view stresses the acquisition of understanding by all, including the traditionally underprivileged, to the highest extent of their capability, rather than the selection and promotion of an elite. It is a philosophy that simultaneously stresses erudition and common sense, integration through application, and innovation through creativity. Most importantly, it stresses the creation of knowledge. It is as tightly coherent as the old world view; to espouse the intent but retain the old model of assessment is to lose the integrity of the old without gaining that of the new.

To recap, the process of assessment affects the educational process it is designed to evaluate, and the power of the old model derived in large part from its congruence with the underlying and coherent philosophies of science and society. Cohesion is a matter of purpose. Logically, if it is to be a powerful tool for intervention, any new model should be as closely congruent with the purposes, philosophy, and methods of the new world view as the matrix model was with the old.

## Recent Statements of Purpose

The goal of mathematics as a domain is creation (or discovery) of new knowledge. Children are inventive (e.g., Moser, 1980). Thus, the primary objective of mathematical education should be not to perpetuate existing knowledge, but to foster a contemplative approach which will support the creation of new knowledge.

> The objective is to produce new mathematics, to create new theories, to help in the solution of new problems which are only now being identified and recognized. . . . We need all the creative power of youth, we need new forms of thought which we cannot envisage. The primary objective of mathematics education is not to perpetuate knowledge or to push existing knowledge a little further . . . but to foster the creation of new knowledge. (D'Ambrosio, 1979, p. 193)

On a practical level, the aim of mathematics education is to provide students with the understanding, processes, and language needed for communication and problem solving in adult life (Committee of Inquiry into the Teaching of Mathematics in Schools [CITMS], 1982). This stress on application and problem solving has engendered a move towards interdisciplinary efforts. In consequence, the stress on content in school mathematics is giving way to a stress on the processes of mathematics and learning. Succinctly, emphasis in mathematical education has changed from knowing about mathematics to knowing mathematics (Romberg, 1983).

This concern with the directions of mathematical education is not a parochial matter restricted to the United States, but a serious concern internationally. In the United Kingdom, for example, the aims of mathematics teaching have been described, in practical terms, as the use of mathematics in communicating information and ideas, its use as a powerful problem-solving tool (especially in analyzing relationships), and its fascination (DES, 1985). Equally important is the learning of attitudes and habits, the sense of mathematics as a creative process requiring imagination, initiative, and flexibility, and the habits of working systematically whether independently or cooperatively. Most of all, the learning of mathematics should be an experience from which students derive enjoyment and confidence (DES, 1985).

Processing/Strategies

Reaction against the long domination of the survey-the-domain approach to content has resulted in a strong stress on process in mathematics education, on the process of mathematics, on the process of learning mathematics, and, contributing to the learning process, on the context in which that learning takes place (D'Ambrosio, 1979; Freudenthal, 1983; Romberg, 1983). Both innovation and adaptation involve recognition of a problem or an opportunity, hypothesizing of a solution, and resolution of ensuing problems. Thus, modeling (Buck, 1965; D'Ambrosio, 1979), conjecture (Schwartz, 1985), and problem solving (CITMS, 1982; National Council of Teachers of Mathematics [NCTM], 1980) have been described as the heart of the mathematical process. More lucidly, the mathematical process was distilled to abstraction, invention, proof, and application (Romberg, 1983).

In essence, a new common thread has emerged. The making of conjectures is essentially the abstraction of concepts into a mental model, a process whereby certain qualities of actual events are internalized and others ignored. Similarly, the process of mathematical modeling is an extension of the process of concept formation (cf. Skemp, 1979), in that there is an iterative process of model abstraction, validation through simulation or actuality testing, and further reflection. "To do mathematics is to create and manipulate structures" (Lesh, 1985, p. 81). Thus, one makes conjectures and, having extracted a workable model or concept, uses it. Further problems may require fine tuning of the model, or may prompt the development of new models. There is, therefore, a cohesion of thinking between the methods of mathematics and the processes of the mind based on a commonality of purpose--the creation of new knowledge.

In many respects, this is similar to the Kuhnian (1962) notion of science. Hence, mathematics is the science of order, the identification, description, and understanding of complex situations (Jaffe, 1984). Mathematics codifies such situations with elegance and simplicity so that it is possible to prove or disprove abstractions and to evaluate predictions based on the model; the process then supports further abstraction. There is an iterative, if unpredictable, relationship between abstraction and application; abstraction leads to applications and hard problems lead to the invention of new mathematics (Jaffe, 1984). In other words, mathematics supports the processes of thinking, communication, and practical activity.

Freudenthal (1983) stressed that, while problems in mathematics may be isolated, those in mathematical education may not. While the history of mathematics has been one of progressive schematizing and formalization, in learning mathematics the psychological progression of understanding is more important than a historically sequenced development of content. Consequently, while stressing process to the virtual exclusion of content, Freudenthal's first emphasis was on the processes of learning. For mathematical education, these include:

1. paradigms of how children learn mathematics;

2. diagnosis of problems in learning mathematics and prescriptions for solutions; and

3. identification of levels of mathematical learning, which would facilitate cooperative activity between children at different levels.

This emphasis on process draws attention to the most fundamental distinction between mathematics and mathematical education: the focus of the first is the discipline; the focus of the second is the child (Skovsmose, 1985).

The psychological problems of mathematical education are integral to the process of creating new mathematical knowledge. For example, children must learn to reflect on and argue their intuitions in order to develop formalizations. Hence, one danger of training in algorithms is that it will block the pathways to intuition (Freudenthal, 1983). In addition, language and a sense of appropriate precision also are important, both in formalizing and codifying an argument and in application.

Because the problems and processes of mathematics education are interwoven, the context in which that education takes place is crucial to the development of learning processes and of mathematical attitudes. Mathematical attitudes are not synonymous with attitudes toward mathematics but are a reflection of a coherence (or lack thereof) between language and notational system, a feeling for mathematical structure and perspective. Thus, a major challenge is the creation of situations which will encourage the process of doing mathematics (Freudenthal, 1983).

The goal is for children to think for themselves (Bell, 1985). If the context in which children learn mathematics is regarded as a separate issue from the processes and content of mathematics, instruction in techniques replaces instruction in content. Children need to be able to identify and initiate their own problems, to express their own ideas, to make and test their own hypotheses, to rationally defend their own ideas, and to constructively criticize the ideas of others (Bell, 1983). The process of teaching children mathematics is therefore changing from exposition and drill in algorithms and skills to a combination of discovery and diagnosis (Bell, 1985). Exposition and algorithms are seen as more properly following experience with realistic problems (DES, 1985). The intuitions and naive source-ideas derived from basic "paradigmatical experiences" (Davis, 1972) are crucial to the development of understanding.

Content

Proposals abound for modification of the content of mathematics education for children. Some recommendations proceed from the assumption that there is a curriculum in place that needs modification; such recommendations suggest additions and deletions to existing courses (Usiskin, 1984). Others (e.g., California, 1985; Illinois, 1985) specify concepts and subjects to be taught. Project 2061's initial phase, for example, is content identification. In considering content, an early meeting of the mathematics panel (American Association for the Advancement of Science, 1985) brought up graphs, math as a language, attitudes, algorithmic computation, arithmetic. The NCTM (1980) advocated problem solving as the focus of school mathematics in the 1980s, which is less a recommendation of content than of process. Perhaps because of stress on process, more recent recommendations (DES, 1985) have emphasized the objectives of mathematics instruction and the consequent criteria for content. Thus, the widely stated aims of mathematics teaching ask that children acquire facts, skills, conceptual structures, general strategies, and personal qualities. From these aims, the following criteria for content may be derived.

1. Students are able to cover content successfully at their own appropriate level;

2. Content is not so extensive that it impose restrictions on the range of classroom approaches;

3. Content forms a coherent structure;

4. Students are exposed to a broad content;

5. Content meets the mathematical needs of the rest of the curriculum;

6. Content meets the basic mathematical needs of adult life, including employment;

7. Content includes elements which are intrinsically interesting and important;

8. Appropriate weighting is given to the essential and the desirable;

9. Content takes account of the potential of electronic calculators; and

10. Content is increasingly influenced by developments in microcomputing.

It is significant that the Department of Education and Science (DES) curriculum guide (1985), while recommending abilities, attitudes, classroom approaches and assessment strategies, completely abstained from any recommendation of specific content, such as algebra, geometry, discrete mathematics. This essentially follows the argument of the report from the National Advisory Committee on Mathematical Education (NACOME, 1975) that, in a rapidly changing world, no specific curriculum could ever be recommended. One possible reason, of course, is that content is changing so rapidly that specification of particular content or courses is futile. More to the point is a clearer perception of what it means to do mathematics.

In one respect, it can be argued that mathematics has no content. Its objects are all imaginary, belonging to the intellectual world, its content its own epistemology. Consequently, the traditional content of mathematics--arithmetic, geometry, algebra, or calculus--never was content but always process. If emphasis on process and the creation of mathematical knowledge means that the content of mathematics is its own epistemology, several things follow:

1. Context, content, and process are inextricably related. Some sense of this is emerging in, for example, the work of Vergnaud (1982, 1983a, 1983b, 1984) on conceptual fields, in which the context, the relational invariants, and the signifiers are all not merely related but are a tightly cohesive system. In essence, the move toward a coherent approach in science is already being reflected in some parts of the pedagogy of mathematics.

2. Interdisciplinary activity is a natural corollary once mathematics is seen as a process in search of content and context. Thus, it makes more sense for children trying to understand entirely abstract processes to root their understandings in concrete contexts from the real world, whether cake baking or stream flow.

3. A clear understanding of the significance of an epistemological emphasis is essential to the creation of a framework for assessing the mathematical progress of children.

An Epistemological Approach

Epistemology is concerned with the origin, nature, methods, and limits of knowledge. Therefore, emphasis on the creation of knowledge virtually requires an epistemological perspective. However, that carries with it a long-standing controversy over whether the knower and that which is to be known are separate entities (von Glaserfeld, 1983). When the knower and the known are seen as separate entities, knowing involves making cognitive structures match the reality which they are supposed to represent. However, because experience is the way to knowing, knowledge is necessarily subjective and constructive and cannot be separate from the knower. In this context, public knowledge structures ensue from communal agreement about private cognitive structures. If a usable coherence is to emerge around which to create a new framework for assessment, it is important to consider some implications of the two approaches.

> On the one hand is the view that declares that: The acquisition of a social knowledge like mathematics is not reducible to a process of spontaneous construction by children, adolescents and adults, even if one considers as essential a constructive approach to learning. (Vergnaud, 1983c, p. 2)

Vergnaud (1983b) has a very distinct view of the interrelationships between meaning and complexity, which holds that the meaning of mathematics comes from practical and theoretical problems to be solved. This view is crucial to his perception of mathematics as arising from contexts. He emphasized the theory of didactic situations--conceptualizations depend on the context in which they are formulated and are eventually modified in the face of new situations. In other words, knowledge emerges in situ, and there is a tight relationship between the context, the conceptual properties of the context, and the symbolic representation (cf. Kaput, 1983) which best represents both concept and context. If a student does not have a coherent system of concepts, relationships, and symbols appropriate to a given situation, the level of complexity is commensurately higher and obstructs understanding. Conceptual development is so slow that it is desirable to study the same field year after year, going deeper, meeting new contexts through different problems to be solved (Vergnaud, 1982). The problem of complexity, therefore, is not simply one of memory overload but of the difficulties inherent in conceptualizing tightly interrelated structures of concept, procedure, and representation. This constitutes a serious problem for the transfer of concepts from one context to another. It is a matter of cognitive dissonance:

> This source of resistance to change lies in the fact that an element is in relationship with a number of other elements. To the extent that the element is consonant with a large number of other elements and to the extent that changing it would replace these consonances by dissonances, the element will be resistant to change. (Festinger, 1957, p. 27)

Hence, Vergnaud (1983b) stressed the importance of identifying and classifying situations according to their conceptual fields, apparently emphasizing the inherent properties of the matter to be known. Good teaching therefore requires that a set of relations be learned in one context and then in another, so that the relational invariants and common structure can emerge. A gradual increase in complexity relies on controlled changes of structure in a fixed context and deliberate transfers of structure from one context to another (Bell, 1985). In other words, control over increases in complexity depends on a moderated introduction of cognitive dissonance.

By contrast, von Glaserfeld (1983) adopted the constructivist perspective to epistemology, stressing that knowledge is not necessarily a picture of reality but provides structure and organization to experience. Davis and Hersh (1981) took a similar position:

> The whole object of mathematics is to create order where previously chaos seemed to reign, to extract structure and invariance from the midst of disarray and turmoil. . . . To create order--particularly intellectual order--is one of the major human talents, and it has been suggested that mathematics is the science of total intellectual order. (pp. 172, 173)

The all-important function of such constructed knowledge is to enable the solution of problems. Knowledge is not a transferable commodity but a matter of the students' conceptual organization of their own experience. "Most of our heuristic knowledge of mathematical enquiry is tacit; built on our experience and our unconscious systematization of that experience" (Ruthven, 1985, p. 106). Rightness is not a matter for assessment against another's standards but of the "fit" of the internal order with the external problem. Understanding consists of fitting a concept to the language at hand, analogous to the process of matching knowledge to experience.

Thus, there is an apparent polarity of the epistemological approach, with strong pedagogical implications, which is likely to make the search for a new cohesion difficult. Vergnaud attends primarily to the matter known, von Glaserfeld to the knower (Kilpatrick, 1983); the former is domain-centered, the latter child-centered. The divergence, emphasized by use of such terms as "transmission" of knowledge (Vergnaud, 1983c, p. 2), seems to shatter any hope for a cohesion around which to build an assessment framework whose purpose centers around the creation, rather than acquisition, of knowledge.

It would be easy to misread Vergnaud and regard stress on conceptual fields as an updated version of the content-oriented, cover-the-ground philosophy. While it is domain-oriented, it is also constructivist and child-centered. Knowledge is built by children from problems they have solved and situations they have mastered; their conceptions, models, and theories are shaped by the situations they have met (Vergnaud, 1982). This point is crucially important because it obviates a philosophical conflict. The strategies-and-errors approach of diagnostic teaching (e.g. Bell, 1985; Hart, 1984), which is the essential groundwork for efforts to establish and use conceptual fields, is equally essential to the constructivist, coherentist approach (e.g., Rescher, 1979; Skemp, 1979; von Glaserfeld, 1983) to knowledge and learning.

It was argued in chapter 2 that the goal of mathematical education should be the ability to produce new knowledge, whether personally new, new in the sense of a new solution to a problem, or new to the domain. This cannot occur in a vacuum; it must be viewed in reference to the structure of personal knowledge, practical application and the structure of the domain, for that is the process by which new knowledge is validated.

The key that links the domain orientation of conceptual fields with the constructivist viewpoint is diagnostic teaching (e.g., Bell, 1982; Bell, Swan, Onslow, Pratt, Purdy, and others, 1986). This is based on critical tasks, designed to reveal students' strategies and errors. The tasks should be embedded as closely as possible in the context in which the student is likely to apply the principle being learned (Bell, 1985). In other words, practical, problem-solving activities are part and parcel of the diagnostic approach.

In one sense, strategies-and-errors research and diagnostic teaching have, by implication, an iconic model against which the child's knowledge is compared, with the intent of transforming the cognitive structure of the novice so that it matches that of the expert; the analytical framework of conceptual fields contributes significantly to the process (e.g., Bell, 1985; Bell et al., 1986; von Glaserfeld, 1983). If that is all it is, then both diagnostic teaching and conceptual fields remain a cover-the-ground approach, albeit of somebody else's cognitive structure rather than somebody else's factual knowledge of the domain.

From another perspective, diagnostic teaching monitors the success of the child's strategies against the problem attempted. From this view, a strategy is only in error if it fails to adequately solve the problem, even though more efficient strategies may exist. Most importantly, if diagnostic teaching regards anomalies as important and, in the process of diagnosis, not only errors but unique strategies and ways of looking at problems emerge, that amounts to the documentation of new knowledge production. It is, in essence, an approach to the identification and validation of new knowledge.

Our argument runs as follows:

1. An epistemological approach to mathematical education is required.

2. An epistemological approach invokes a fundamental conflict between the views of knowledge originating independently of the knower or inseparably from the knower.

3. Conflict is resolved for the purpose of mathematical education by diagnostic teaching and an approach to the validation of knowledge which relies on cognitive systematization and applicative adequacy (cf. Rescher, (1979).

Several common threads run through the conceptual-fields approach, diagnostic teaching, and the constructive, coherentist, cognitive systematization of knowledge; these common strands offer a source of congruence and cohesion.

1. Each regards knowledge structures as emerging from experience. In diagnostic teaching, it is the problem; in

conceptual fields, the situation; and in cognitive systematization, the phenomenon.

2. Each assumes a cohesive set of relationships: the relational invariants and related signifiers of conceptual fields; the developing concept structures of diagnostic teaching and cognitive systematization.

3. Each involves checking against a systematized model of conceptual structure, whether of the domain or of the picture in the individual's head.

4. Each has the goal of progressively developing conceptual structures, thereby creating order out of disorder.

5. Each relies on disequilibrium to precipitate progress in the development of structures, either deliberately created for the purpose of causing controlled cognitive conflict, or occurring spontaneously in the form of incoherent phenomena.

6. For each, predictive value is applied as a test of the adequacy of a theoretical model.

7. Each has the process of communication as its cohesive force.

Pursuit of an epistemological approach to mathematical education virtually ensures the development of a new coherence between mathematical education and the trends in science and society, because it ensures a close coherence between the pedagogy of mathematics and the science of mathematics. Coherence with the philosophical trends of mathematics as a science offers the greatest possibility of coherence with science in general because mathematics and the development of science are inextricably linked. Science, in turn, inevitably influences practical activities, the economy, and the way people think about the world. The resulting web of connections offers the greatest hope of congruence between didactics, science, and society.

## Alternative Intellectual Structures: The Theoretical Network Model

It was argued that the content-by-behavior matrix was originally powerful because it was congruent with the contemporary philosophies of science and society, losing its value as it became increasingly inadequate to cope with complexity. The intellectual structure most congruent with present trends in science and society is the theoretical model, which is endlessly versatile. Anything that can be diagramed can be modeled, because a diagram is simply a model of an idea (cf. Albarn & Smith, 1977) Consequently, theoretical models may be created for anything ranging from the economics of moving icebergs (Cross & Moscardini, 1985) to the role of intuition in the process of creating knowledge (Kuyk, 1982). Models may be steady-state or dynamic, continuous or discrete, statistical or stochastic (Cross & Moscardini, 1985). Mathematical descriptions of relationships in the model may rely on anything from network theory to catastrophe theory (e.g., Andrews & McLone, 1976), depending on the relationship that exists and the facility with which it may be described. The significant point is that models are far more capable than the matrix of describing complex relationships and, partly for that reason, are also intellectually consistent with the new world view.

The construction and use of a theoretical model includes problem identification, gestation through reflection, model building, simulation, and payoff (Cross & Moscardini, 1985). In other words, a model originates in a situation, after which a systematized and idealized form of knowledge is validated by testing its predictive power (McLone, 1976). This is remarkably congruent with the epistemological approach, for only reliance on disequilibrium as a means of progress is missing, although in one sense the only way to test a model is to make every effort to destabilize it. While theoretical models are essentially stable structures, it is possible to model instability (Chillingworth, 1976)--and, if disequilibrium is seen as essential to the development of knowledge, that may be important. This reflects the fact that theoretical models are versatile tools for the handling of complexity (Prigogine & Stengers, 1984).

Networks are an especially significant variety of theoretical model. They consist of nodes and arcs, which depict the direction of relationships between the nodes (Carre, 1976). Any node may be connected to any other node, regardless of size or other connections. There is no preordained sequence to the connections unless a strict partial ordering actually exists in the phenomenon modeled, in which case one small part of the overall network reflects the hierarchy.

Consequently, a network is capable of depicting both linear and iterative relationships and their directionality, the impact of singular anomalies as well as major subsystems, complex relationships in addition to linear and numerically restricted ones, and interrelationships rather than segmentations. It is a more powerful model than the hierarchy for handling complexity simply because a hierarchy is only one very limited form of network and so, by definition, is less versatile. By subsuming hierarchical models, the network model also reduces the danger of overreliance on a single theoretical model.

The power and congruence of the network model is suggested by its use as an intellectual framework in cognition (e.g., Lesh, Landau, & Hamilton, 1983), computer-based communications (e.g., Glossbrenner, 1985; Hiltz & Turoff, 1978), anthropology (Pelto & Pelto, 1978), ad hoc and formal organizational grouping (e.g., Hine, 1977), industrial organization (e.g., Kanter, 1983; Toffler, 1985), mathematics (e.g., Carre, 1976) and epistemology (e.g., Rescher, 1979). Of these, the most pertinent is the application of the network in mathematics and epistemology. Nevertheless, the other applications are important because they suggest that the network model has enormous potential for congruence and, therefore, for power.

The cohesive force of networks and other theoretical models is purpose. For this reason, networks are dynamic; as purpose changes, parts of the network atrophy and others grow. This applies to all networks, whether in transportation, electronic conferencing, or epistemology. A topic, for example, is a system of concepts which can only be defined by identifying its construction principle or purpose, usually arising in response to some problem (Jackson, 1984).

In epistemological networks, conversation theory (Pask, 1984) suggests that conversations take place between participants, which are coherent systems of concepts. When the incoming information is inadequately absorbed or rejected by the cohesive relationships among a system of concepts, there is the equivalent of Davis and Hersh's (1981) chaos out of order. Subsequently, a phase of schism results from the juxtaposition of incoherent or contradictory data. For structured knowledge to emerge (order out of chaos), a generalized and relevant analogy is essential to the development of new cohesion (Pask, 1976). This new order is new knowledge, and the resulting knowable public topics, together with their relationships, can be represented by the network of an entailment mesh (Pask, 1984). Projects based on the computerized application and testing of these ideas, with programs such as Caste and Thoughtsticker (e.g., Ferraris, Midoro, & Olimpo, 1984; Gregory, 1984), suggest an epistemological and constructive alternative to the kind of computer-adaptive testing based on standardized, objective testing methods such as Project Adapt (Frechtling, 1986).

In summary, theoretical models are intellectually powerful and versatile tools with which to replace the content-by-behavior matrix. This is especially the case with network models, partly because the network lends itself to representation of the communication process inherent to the epistemological/diagnostic approach. The application of network models in epistemolgy (e.g., Pask, 1984; Rescher, 1979) means that, if used in the assessment process, they may be especially useful in promoting an emphasis on the creation of knowledge. However, a network model appropriate to the needs of assessment and monitoring needs to be developed.

## Alternative Assessment Procedures

It was argued in the discussion about the existing framework for assessment that there was a strong congruence between the purpose for assessment, the model of assessment, and the tools for assessment. In other words, an intense cohesion unites the hierarchical purpose of ranking, the content-by-behavior matrix, and standardized, objective, group testing, epitomized by the multiple-choice format. For an equally intense cohesion to be developed, alternative methods of assessment must be designed which are congruent with the theoretical model and its purpose, which is to assess the ability and achievement of the educational system in teaching children to create knowledge. While any number of indirect proxies may be postulated, the only direct indicator is the kind of knowledge created by children in the system. Thus, tools are needed to assess children's progress in creating knowledge.

Work in artificial intelligence suggests that there are two basic facets to creating knowledge: first, a data base of facts and assertions; and, second, an inference engine. There are, therefore, several ways of adding to knowledge, whether individually or cooperatively: increasing the power of the inference engine; adding to the facts in the data base; and adding to the network of assertions in the data base. Significantly, power in knowledge creation is primarily a consequence of the knowledge base and only secondarily a consequence of the power of the inference method (Feigenbaum, 1984). Furthermore, the most important aspect of the knowledge base is the structure of assertions (Robinson, 1984). These facts reinforce the notion of knowledge creation as a matter of searching for new structures. It is essentially similar to the conclusions reached by Pask (1976, 1984) on the importance of analogic reasoning (cf. Pimm, 1980; Pelto & Pelto, 1978) in the creation of new knowledge and to the use of analogy in the mathematical modeling of complex systems (Cross & Moscardini, 1985, p.15). In summary, for policy purposes, it is important to have a framework for comparative evaluation of parts of the system that is congruent with the intent of the system. However, knowledge is created by individuals and groups, so for the purpose of intervention, it is equally important to have tools that monitor children's strategies, problems, and achievements. Theoretical models, especially network models, offer distinct advantages for both aspects of the monitoring process, the framework and the instruments.

## Principles of Construction for Tools of Assessment

Instruments for assessment should embody the commonalities among the epistemological approaches to mathematical education and diagnostic teaching, the character of theoretical models, and the insights of artificial intelligence, namely:

1. All knowledge is rooted in experience.

2. Knowledge entails the structural modelling of perceived regularities.

3. Cohesion of structure is integral and derived from purpose.

4. Quality is determined by predictive power.

5. Disequilibrium is essential to the process.

To these might be added a sixth:

6. Knowledge is both individual and communal.

Simply stated, there is a need for tools that document the production of knowledge and not merely the proxies that contribute to the process, such as time spent learning or the quality of the teaching staff. A sufficiently detailed view of the process is essential in order to have some idea of how to construct policies for intervention. However, if there is any lesson to be learned from the old paradigm, it is that parts of the process cannot be analyzed in isolation, and then aggregated, with the result regarded as an adequate indicator.

Because knowledge is derived from experience, it seems logical to monitor the quality of the experience in which children learn to create knowledge and to assess it in a practical and realistic context. Evidence suggests that this strategy would have a rapid and significant impact on the teaching and learning process (Frederiksen, 1984, 1986). At the moment there are only very indirect proxies for monitoring the quality of experience, such as the professional qualifications of teachers, quality of textbooks, and class size.

Although it seems desirable to use practical assessment techniques, the notion of assessing in a practical and realistic context is typically restricted to such areas as teacher education, medical school, flight training, and some of the Advanced Level General Certificate of Education exams. However, in England, the APU gave practical tests in topics that included measurement of mass and area and extended problem-solving situations (Joffe, 1985) as part of its program to assess secondary mathematics. The more usual avoidance of practical testing is largely because conventional, group testing has emphasised cost-efficient, standardized, objective testing, while practical testing is viewed as difficult, costly, and time consuming. There is also a more subtle reason. Standardized, objective, group tests are prepared by an external authority and merely administered locally, often by an official proctor. By comparison, practical tests require more local and internal knowledge and authority, which reduces their perceived validity. Such local authority is a particularly fraught question when the capabilities of teachers are under fire.

There is an additional consideration. The standardized objective testing approach lends itself readily to quantification when items are scored right or wrong, 1 or 0. In the context of evaluating collaborative effort and the quality, structure, and predictive power of knowledge, efforts can no longer be scored right or wrong; the exclusively quantitative nature of group testing is no longer tenable. The first step of many assessment procedures will almost inevitably be qualitative, even though means may be devised for subsequent quantification.

Several approaches offer some promise. One instrument that is a cost-effective tool for group assessment of intellectual structure in context is the Superitem (Collis, Romberg, & Jurdak, 1986), based on the SOLO taxonomy (Biggs & Collis, 1982). A second, which promises to offer the information needed for diagnostic teaching, is the constellation of innovative approaches being tried in Britain. These incorporate pencil-and-paper testing, practical testing, diagnostic interviewing for the identification of strategies and errors in problem solving, and the effort to develop graduated assessment in mathematics. A third approach that examines the cooperative nature of knowledge production, but is only a proposal, may be termed Coaker's Wild Idea.

Superitems

A superitem (Collis, Romberg, & Jurdak, 1986; see also Collis, chapter 19) consists of a paragraph describing a problem situation (stem) and a series of ensuing items that can be answered by reference to the information provided in the stem. The intent (Romberg, Collis, Donovan, Buchanan, & Romberg, 1982) is for a series of interdependent questions of increasing complexity to originate in a common, realistic context. Thus, a superitem consists of a problem situation containing considerable information and an accompanying set of open-ended questions carefully graduated according to the SOLO taxonomy (Biggs & Collis, 1982). This categorizes the child's response according to its capacity and structure, relating operation, consistency, and closure. The SOLO taxonomy addresses the structure of ideas derived from an experience, and superitems attempt to elicit that structure. One practical advantage of superitems is that they proffer an alternative to independent, multiple-choice items but may still be administered to large groups.

Assessment in Britain

Some of the innovative approaches to assessment in Britain may prove useful. The Assessment of Performance Unit in Britain, similar to the National Assessment of Educational Progress in the United States, was commissioned to prepare a national profile on the educational achievement of children The work of the APU is geared toward causing educational change by having assessment procedures precipitate curricular change (A. Clegg, personal communication, July 1985). The direction of change is essentially that outlined as desirable by the Cockroft Commission (CITMS, 1982) which advocated, among other things, links with other curricular areas, practical work, the importance of language, a diagnostic approach to testing (cf. Bell, 1985), mathematics for the majority, a graduated assessment, and records of progress. In the process, the APU gave completion tests to a large number of students. One facet consisted of a matrix-sampling approach organized around a content-by-behavior matrix to which had been added a third dimension that addressed understanding, practical application, problem solving, and attitudes. The third dimension, involving the more innovative efforts, was assessed separately by sending test booklets to small samples.

The APU's assessment methods for the practical and problem solving parts (Foxman, 1985; Foxman & Mitchell, 1983) are a combination of pencil-and-paper answers to complex and realistic situations and practical assessment with manipulatives in a diagnostic assessment interview (e.g., Denvir & Brown, 1985, 1986; Joffe, 1985). The situational questions are largely analogous to the superitem approach (Romberg et al., 1982) in that there is a problem stem with considerable information, followed by a series of increasingly complex questions. Answers can range from the simple to the complex. Diagnostic interviewing of a small sample of students engaged in a practical test is conducted according to a script, but with some flexibility for clarification, limited prompting, or amended answers. Responses are checked against a precoded list, but unanticipated answers are recorded in detail. The result is valuable insight into students' mathematical thinking (Burstall, 1986), a conclusion supported by other studies (e.g., Confrey, 1980).

While one-to-one interviewing in a practical test yields a wealth of valuable information, a disadvantage is that it is time consuming and costly to conduct and analyze. One alternative is content analysis, both global and propositional (Bell, Brook, & Driver, 1985). It is in some respects analogous to Pask's (1984) conversation theory. Comparison between responses in a written exam and answers to essentially similar questions derived in an interview situation showed that the same range of propositions was used in each format. However, the response level in interviews was higher, and students were more likely to suggest alternative responses and describe their thinking in more detail. While arguing the stability of concepts between written form and interview, a curious statement was made: "Over 50% of the students gave the same type of response in written form and in interview" (Bell, Brook & Driver, 1985, p. 210). The reciprocal inference is that almost 50% of the students changed their conceptions between one form and the other. Consequently, substitution of questionnaire for interview needs closer examination.

In chapter 2, it was argued that intelligence must now be regarded as multifaceted (Walters & Gardner, 1985) and susceptible to improvement. Therefore, methodologies and instruments are needed that do more than produce a crude terminal score purporting to summarize years of a child's achievement. A number of strategies have been tried in Britain which essentially link internally created portfolios with external assessment. The General Certificate of Secondary Education (GCSE) (Scruton, 1986) requires both external assessment by examination and an internal record of achievement. The internally assessed but externally moderated record is intended to promote many of the practices attempted in pilot efforts (Wharmby, 198_):

1. a modular approach;

2. practical work;

3. extended project work;

4. written assignments;

5. oral assessment;

6. written assessment;

7. assessment as an integral part of the learning process;

8. greater involvement of the teacher in the assessment process;

9. a cumulative profiling of students' mathematical achievement; and

10. an implicit intent to send all students, and not just the brightest and most mathematically able, into the adult world with some mathematical understanding and confidence.

Graded Assessment in Mathematics (GAIM) is one such project. The curriculum is divided into progressive levels (Brown, 1986) determined by the facility hierarchies identified in the Concepts in Secondary Mathematics and Science Project (Hart, 1980). A year's portfolio would contain at least 4 practical problems, 4 investigations and 1 extended project among the minimum of 10 required pieces (Graded Assessment in Mathematics, 1986).

A portfolio record of assessment in artistic learning is also being tried in the United States in a project jointly administered by Project Zero and Educational Testing Service (Zessoules, 1986) However, with respect to teachers' assessment of children becoming part of a permanent record of achievement, it is important that (Department of Education and Science, Welsh Office, 1984)

1. the picture be fair, reasonable, and confined to matters of direct knowledge and evidence;

2. assessment concentrate on the positive qualities;

3. the assessment include concrete examples;

4. the statement be written in sentences and not in the form of checks, numbers, or letter grades. As with practical testing, this approach to assessment places heavy reliance on the professional abilities of teachers.

It was argued in chapter 2 that self-direction and self-assessment are essential to life-long learning. An element of self-direction is implicit to extended project work, and thus self-assessment is also being considered as an essential element. The Emrys ap Iwan school in Abergele, North Wales, which takes all students between the ages of 11 and 18, adopted a scheduling strategy to encourage investigation and project work. Every afternoon for nine weeks, children are in the same two-hour block with the same teacher. Self-assessment, guided by a checksheet and monitored by the teacher, plays a large part. Experience with self-assessment by pupils and internal assessment by teachers showed that it was essential for teachers to monitor student self-assessment, because children tended to judge their own work too harshly, and that external monitoring of the entire process was essential for similar reasons (D. Newman, Principal, Personal communication, July 1985).

Both the new world view outlined in chapter 2 and the epistemological approach to mathematics education require cooperative effort in the creation and validation of new knowledge. Cooperative learning (e.g., Johnson, Johnson, Holubec, & Roy, 1984) has not been a matter for traditional assessment. However, the most recent initiative of the APU is development of an assessment framework that looks at four aspects of group behavior (Joffe & Foxman, 1986): social interaction, working on the task, mathematics used, and communication. Different contexts, sizes (2-4), and composition (friendship, gender, teacher recommendation) of groups are being tried. That this approach to mathematics is unfamiliar to most students participating highlights the proactive approach of the APU.

In summary, the thrust of the effort in Britain is toward . much wider variety of teaching and learning strategies, with the assessment process regarded as a catalyst. It is a multifaceted strategy that has the potential for providing a more flexible and much more detailed picture of children's achievement. It is a strategy we should consider.

## Coaker's Wild Idea

The traditional pattern of testing is to isolate the student from all sources of information and assistance. This is not realistic if the intent is to evaluate the production of knowledge, which may be initiated by the individual but is an inherently cooperative process. Coaker (1985), an industrial mathematician formerly with British Petroleum, argued that mathematics is a language, and so the need for communication is intrinsic. It is also a practical and cooperative activity:

> To solve our problems, if we hit a snag, we "cheat" as much as possible. We ask our colleagues, we look up what other people have done before, we search in libraries, we discuss the problem and work together as much as possible. This involves communication, but I suggest that, after primary school, such ideas are rather alien to most school work. (p. 169)

Coaker argued that whatever mathematics is taught should be used with confidence, applicable to a wide range of problems, transferable to other topics and subjects. One failure of the present system is that students are expected to find the "right" answer and find it in less than 30 minutes. He argued a need for employers to link with the schools and provide locations for topic work, assisting the management in the solution of technical or other problems. Assessment of such an approach would necessarily be school designed because practical applications and problem solving are less easily done in timed examinations. Coaker's wild idea is that assessment of such work entail a collaborative effort of teachers, parents, employers and students.

> A compulsory part of the final assessment system should include a special project. In this, pupils will be put into teams of four, of mixed ability, and given a cross-curricular task to perform. This would occupy a week, in which time they would be allowed access to all forms of information and calculation, workshops and laboratories, as required. At some stage during the week, an additional question would be asked, which they should be able to solve from their work so far, and to which an answer is required in a short time. Other information would be provided in a foreign language, not necessarily one they had met before.
>
> At the end of the week, each team would present its results, using whatever aids they required. All should take part. The assessors would be a mix of teachers, parents and local employers. Each member would also write a report on the project and their views of the contributions of the other members of their team. (p. 169)

Obviously, Coaker's wild idea is intended for the terminal assessment of secondary school children. Nevertheless, the principles embodied could apply at any level. They include:

1. knowledge grounded in practical, realistic, inter-disciplinary experience;
2. [illegible]ledge creation as a realistically collaborative ef[illegible] individuals of widely ranging ability;
3. the intrinsic and essential role of various kinds of communication in the process of creating and communicating knowledge;
4. realistic use of widely ranging information sources;
5. recognition of the inadequacy of traditional assessment tools;
6. recognition of two kinds of problem, the urgent and the important;

7. collaboration between school, home, and community in the process of teaching children to create knowledge and assessment of that process;

8. experience by students with the reality that, in the world beyond school, activity is usually collaborative and assessment involves both peers and superiors.

Coaker's wild idea seems a feasible strategy in the context of recommended by the Carnegie Foundation for the Advancement of Teaching (Boyer, 1983). However, the notion assumes that students have experience in such an approach. A proactive view would suggest that, whether they have or not, the stated intent of conducting such an assessment would have an impact.

## SUMMARY AND CONCLUSION

Traditional monitoring practices have consistently used a content-by-behavior matrix as their theoretical framework and relied heavily on independent, multiple-choice items. Cost efficiency almost eradicated other approaches to group testing. However, the mathematical, psychological, sociological, and pedagogical theories embedded in the model are, quite simply, inadequate. Consequently, it is important to replace the matrix model with one more capable of handling complexity and one that will stimulate change. Unfortunately, the cohesive power of the matrix model exerts a powerful influence which subliminally impedes change.

It is essential that the new model be powerful and have both tight internal coherence and congruence with the trends in mathematics, science, and society. It is also important that the key indicators and instruments for measuring be equally coherent and congruent, the cohesive force being purpose, namely the creation of knowledge.

It is argued that theoretical models, especially network models, are both widely used and consistent in philosophy with approaches to the creation of knowledge. They are also capable of modeling complex processes and, in consequence, likely to exert powerful pressure in stimulating change toward the new world view in mathematical education.

Because the intent is to assess the creation of knowledge and the processes involved rather than to measure the extent to which children have acquired a coverage of the field of mathematics, a much wider variety of measures, many of them qualitative, are needed. Relevance, for example, is crucial to the assessment of knowledge. Yet there is no single system for evaluating relevance, although there are some common considerations (Saracevic, 1976). These include knowledge and a knower, selection based on inference, mapping of structures, dynamic association, and redundancy. Considerable effort is needed to find instruments adequate for the purpose.

Only a few of a wide variety of approaches to assessment have been discussed. They were selected as representative of the range of instruments that might form a coherent repertoire. The urgent need is for a much greater variety of learning and assessment tasks (Ruthven, 1985), a coherent body of tools that will precipitate curricular change. No reference has been made to the need for longitudinal consistency in methodology with previous monitoring programs, because our purpose is not to see how far the education of children has progressed since World War I, or even since the Vietnam War. Our purpose is to ensure that children develop a mathematical understanding adequate to the twenty-first century and monitoring to promote that end.

A wild idea (Coaker, 1985) is a conjecture, which is the heart of the mathematical process (Schwartz, 1985). The intent of the assessment strategy is to intervene in an unproductively stable situation, to create awareness of disequilibrium and cognitive conflict in order to promote progress. More wild ideas need conceiving and testing, because destabilizing the present situation is like trying to rock an iceberg--and without destabilization, significant change is improbable.

## References


Albarn, K., & Smith, J. M. (1977). Diagram: The instrument of thought. London: Thames and Hudson.

American Association for the Advancement of Science. (1985, October). Project 2061: Phase I. What understanding of science and technology will be important for everyone in tomorrow's world? Washington, DC: Author.

Andrews, J. G., & McLone, R. R. (Eds.). (1976). Mathematical modeling. London: Butterworths.

Bell, A. W. (1982). Diagnosing students' misconceptions. The Australian Mathematics Teacher, 38(1), 6-10.

Bell, A. W. (1983). Teaching for combined process and content objectives. In M. Zweng, T. Green, J. Kilpatrick, H. Pollak, & M. Suydam, (Eds.), Proceedings of the Fourth International Congress on Mathematical Education (pp. 587- 590). Boston: Birkhauser.

Bell, A. W. (1985). Some implications of research on the teaching of mathematics. In A. Bell, B. Low, & J. Kilpatrick, (Eds.), Theory, research & practice in mathematical education (pp. 61-79, 5th International Congress on Mathematical Education, Adelaide, South Australia, 1984). Nottingham: Shell Center for Mathematical Education, University of Nottingham.

Bell, A. W., Swan, M., Onslow, B., Pratt, K., Purdy, D., and others. (1986). Diagnostic teaching: Teaching for long-term learning (Report of the ESRC Project HR8491/1). Nottingham: Shell Centre for Mathematical Education, University of Nottingham.

Bell, B., Brook, A., & Driver, R. (1985). An approach to the documentation of alternative conceptions in school students written conceptions. British Educational Research Journal, 11(3), 201-213.

Berlak, H. (1985, October). Testing in a democracy. Educational Leadership, 43(2), 16-17.

Biggs, J. B., & Collis, K. F. (1982). Evaluating the quality of learning: The SOLO taxonomy (structure of the observed learning outcomes). New York: Academic Press.

Bloom, B. S. (Ed.). (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain. New York: Longman.

Bohm, D. (1983). Wholeness and the implicate order. London: Ark Paperbacks.

Boyer, E. L. (1983). High school. A report on secondary education in America. New York: Harper & Row.

Brown, M. (1986). Developing a model to describe the mathematical programs of secondary school students (11-16 years): Findings of the Graded Assessment in Mathematics Project. In Psychology of mathematics education. Proceedings of the Tenth International Conference. (pp. 135-140). London: The University of London Institute of Education.

Buck, R. C. (1965). Goals for mathematics instruction. American Mathematical Monthly, 72(9), 949-956.

Burstall, C. (1986). Alternative forms of assessment: A United Kingdom perspective. Educational Measurement: Issues and Practice, 5(1), 17-22.

California State Department of Public Instruction. (1985, March). Mathematics framework for California public schools: Kindergarten through grade twelve (Preliminary edition). Sacramento: Author.

Cambridge Institute of Education. (1985). New perspectives on the mathematics curriculum: An independent appraisal of the outcomes of APU mathematics testing 1978-1982. London: Her Majesty's Stationery Office.

Carpenter, T. P., Coburn, T. G., Reys, R. E., & Wilson, J. W. (1978). Results from the First Mathematics Assessment of the National Assessment of Educational Progress. Reston, VA: National Council of Teachers of Mathematics.

Carpenter, T. P, Corbitt, M. K., Kepner, H. S., Lindquist, M. M., & Reys, R. E. (1981). Results from the Second Mathematics Assessment of the National Assessment of Educational Progress. Reston, VA: The National Council of Teachers of Mathematics.

Carre, B. A. (1976). Network flow models. In J.G Andrews & R.R. McLone (Eds.), Mathematical modeling (pp. 199-216). London: Butterworths.

Chillingworth, D. R. J. (1976). Structural stability of mathematical models. In J. G. Andrews & R. R. McLone (Eds.), Mathematical modeling (pp. 231-258). London: Butterworths.

Clark, L. H. (1986, Monday July 14). The outlook: Forecasts are flimsy basis for policy. Wall Street Journal, p. 1.

Coaker, P. (1985). Why teach mathematics? 1985 presidential address [to the Mathematical Association]. Mathematical Gazette, 69(449), 161-174.

Cohen, B., & Murphy, G. L. (1984). Models of concepts. Cognitive Science, 8, 27-58.

Collis, K. F., Romberg, T. A., & Jurdak, M. E. (1986). A technique for assessing mathematical problem-solving ability. Journal for Research in Mathematics Education, 17(3), 206-221.

Committee of Inquiry into the Teaching of Mathematics in Schools. (1982). Mathematics counts (The Cockroft Report). London: Her Majesty's Stationery Office.

Confrey, J. (1980). Clinical interviewing: Its potential to reveal insights in mathematics education. In R. Karplus (Ed.), Proceedings of the Fourth International Conference for the Psychology of Mathematics Education. (pp. 400-408). Berkeley: Lawrence Hall of Science, University of California.

Cross, M., & Moscardini, A. O. (1985). Learning the art of mathematical modeling. Chichester: Ellis Horwood.

D'Ambrosio, U. (1979). Overall goals and objectives for mathematical education. In UNESCO International Commission on Mathematical Instruction (ICMI), New trends in mathematics teaching (pp. 180-198). Paris: UNESCO.

D'Ambrosio, U. (1983). Successes and failures of mathematics curricula in the past two decades: A developing society viewpoint in a holistic framework. In M. Zweng, T. Green, J. Kilpatrick, H. Pollak, & M. Suydam, (Eds.), Proceedings of the

Fourth International Congress on Mathematical Education) pp. 362-364). Boston: Birkhauser.

Davis, P. J., & Hersh, R. (1981). The mathematical experience. Boston: Houghton Mifflin.

Davis, R. B. (1972). Naive foundations for a theory of mathematics learning. In W. E. Lamon (Ed.), Learning and the nature of mathematics (pp. 69-83). Chicago: Science Research Associates.

Denvir, B., & Brown, M. (1985). Understanding of number concepts in low attaining 7-9 year olds: Part I. Development of descriptive framework and diagnostic instrument. Educational Studies in Mathematics, 17(1), 1-22.

Denvir, B., & Brown, M. (1986). Understanding of number concepts in low attaining 7-9 year olds: Part II. The teaching studies. Educational Studies in Mathematics, 17(2), 51-72.

Department of Education and Science. (1985). Mathematics from 5 to 16 (Curriculum Matters 3). London: Her Majesty's Stationery Office.

Department of Education and Science, Welsh Office. (1984). Records of achievement: A statement of policy. London: Author.

Eisenberg, T. (1975). Behaviourism: The bane of school mathematics. Journal of Mathematical Education, Science and Technology, 6(2), 163-171.

Farradane, J. (1980a). Relational indexing. Part I. Journal of Information Science, 1, 267-276.

Farradane, J. (1980b). Relational indexing. Part II. Journal of Information Science, 1, 313-324.

Feigenbaum, E.A. (1984). Knowledge engineering: The applied side. In J. E. Hayes & D. Michie (Eds.), Intelligent systems: The unprecedented opportunity (pp. 37-55). Chichester: Ellis Horwood.

Ferraris, M., Midoro, V., & Olimpo, G. (1984). Petri nets as a modeling tool in the development of CAL courseware. Computers and Education, 8(1), 41-49.

Festinger, L. (1957). A theory of cognitive dissonance. Stanford, CA: Stanford University Press.

Foxman, D. (1985). Mathematics assessment framework. In Mathematics report for teachers: 1 (DRAFT). London: Assessment of Performance Unit, Department of Education and Science.

Foxman, D. D., Cresswell, M. J., & Badger, M. E. (1981). Mathematical development. Primary survey report no. 2: Report on the 1979 primary survey from the National Foundation for Educational Research in England and Wales to the Department of Education and Science, The Department of Education for Northern Ireland and the Welsh Office. London: Her Majesty's Stationery Office.

Foxman, D., & Mitchell, P. (1983, November). Assessing mathematics: 1. APU framework and modes of assessment. Mathematics in Schools, 12(5), 1-5.

Frechtling, J. (1986). PROJECT ADAPT: A state and local cooperative effort in computerized adaptive testing. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Frederiksen, N. (1984). The real test bias. American Psychologist, 39(3), 193-202.

Frederiksen, N. (1986). Toward a broader conception of human intelligence. American Psychologist, 41(4), 445-452.

Freudenthal, H. (1983). Major problems of mathematical education. In M. Zweng, T. Green, J. Kilpatrick, H. Pollak, & M. Suydam, (Eds.), Proceedings of the Fourth International Congress on Mathematical Education (pp. 1-7). Boston: Birkhauser.

Garden, R. A. (1981). Validation of the cognitive instruments. Manuscript submitted for publication.

Glossbrenner, A. (1985). The complete handbook of personal computer communications: Everything you need to know to go on-line with the world. London: St. Martins.

Graded Assessment in Mathematics. (1986, Spring). Newsletter 4. London: King's College/University of London/ILEA.

Gregory, D. (1984). Caste (Course Assembly System and Tutorial Environment) and CVI (Combat Vehicle Identification): A first application of an intelligent tutorial system to combat vehicle identification (Report No. AMTE(S)/R83403).

Hart, K. (1980). Secondary school children's understanding of mathematics. A report of the mathematics component of the Concepts in Secondary Mathematics and Science Programme. London: Mathematics Education/Centre for Science Education, Chelsea College, University of London.

Hart, K. (1984). Ratio: Children's strategies and errors. A report of the Strategies and Errors in Secondary Mathematics Project. Windsor, Berks.: NFER-Nelson.

Hiltz, S. R., & Turoff, M. (1978). Human communication via computer. Reading, MA: Addison-Wesley.

Hine, V. (1977). The basic paradigm of a future socio-cultural system. World Issues, 11(2), 19-22.

Hofstadter, D. (1986). Metamagical themas: Questing for the essence of mind and pattern. New York: Bantam Books.

Husén, T. (Ed.). (1967). International study of achievement in mathematics: A comparison of twelve countries. International Project for the evaluation of educational achievement in education (IEA). Phase I. Volumes I and II. New York: John Wiley.

Illinois State Board of Education Mathematics Committee. (1985, February). New thrusts in mathematics in Illinois. Springfield: Illinois State Board of Education. (Workshop draft)

Jackson, P. (1984). Towards a theory of topics. Computers and Education, 8(1), 21-26.

Jaffe, A. (1984). Ordering the universe. Appendix C in National Research Council, The Commission on Physical Sciences, Mathematics, and Resources, Renewing U.S. mathematics: Critical resource for the future. Report of the Ad Hoc Committee on Resources for the Mathematical Sciences (pp. 117-162). Washington, DC: National Academy Press.

Joffe, L. (1985). Practical testing in mathematics at age 15. London: Assessment of Performance Unit, Department of Education and Science.

Joffe, L., & Foxman, D. (1986). Factors affecting small group performance in problem solving. Psychology of Mathematics Education. Proceedings of the Tenth International Conference. London: University of London Institute of Education.

Johnson, D. W., Johnson, R. T., Holubec, E. J., & Roy, P. (1984). Circles of learning: Cooperation in the classroom. Alexandria, VA: Association for Supervision and Curriculum Development.

Kanter, R. (1983). The change masters: Innovation and entrepreneurship in the American corporation. New York: Simon and Schuster.

Kaput, J. J. (1983). Representation systems and mathematics. In J. C. Bergeron & N. Herscovics (Eds.), Proceedings of the fifth annual meeting of PME-NA, The North American Chapter of the International Group for the Psychology of Mathematics Education (Vol. 2, pp. 57-66). Montreal: Université de Montréal, Faculté des Sciences de l'Education.

Kilpatrick, J. (1979). Methods and results of evaluation with respect to mathematics education. In International Commission on Mathematics Instruction (ICMI), New trends in mathematics teaching (pp. 162-179). Paris: UNESCO.

Kilpatrick, J. (1983). Knowing mathematics and knowing research. In J. C. Bergeron, & N. Herscovics (Eds.), Proceedings of the fifth annual meeting of PME-NA, The North American Chapter of the International Group for the Psychology of Mathematics Education (Vol. 1, pp. 118-130). Montreal: Université de Montréal, Faculté des Sciences de l'Education.

Korner, S. (1976). Classification theory. International Classification, 3(1), 3-6.

Kuhn, T. S. (1962). The structure of scientific revolutions. Chicago: University of Chicago Press.

Kuyk, W. (1982, July). A neuropsychodynamical theory of mathematics learning. In Proceedings of the Sixth International Conference for the Psychology of Mathematical Education (pp. 48-62). Antwerp: Universitaire Instelling.

Lesh, R. (1985). Conceptual analyses of mathematical ideas and problem solving processes. In L. Streefland (Ed.), Proceedings of the Ninth International Conference for the Psychology of Mathematics Education. Volume 2: Plenary addresses and invited papers (pp. 73-96). Utrecht, The Netherlands: Research group on Mathematics Education and Educational Computer Centre (D. W. & O. C.), Subfaculty of Mathematics, State University of Utrecht.

Lesh, R., Landau, M., & Hamilton, E. (1983). Conceptual models and applied mathematical problem solving research. In R. Lesh & M. Landau (Eds.), Acquisition of mathematics concepts and processes (pp. 263-343). New York: Academic Press.

McLone, R. R. (1976). Mathematical modelling--the art of applying mathematics. In J. G. Andrews & R. R. McLone (Eds.), Mathematical modeling (pp. 1-11). London: Butterworths.

Moser, J. M. (1980). What is the evidence that children invent problem solving strategies? In R. Karplus (Ed.), Proceedings of the Fourth International Conference for the Psychology of Mathematics Education (pp. 312-316). Berkeley, CA: Lawrence Hall of Science, University of California.

National Advisory Committee on Mathematical Education. (1975). Overview and analysis of school mathematics, K-12. Washington, DC: Conference Board of the Mathematical Sciences.

National Assessment of Educational Progress. (1981). Mathematics objectives: 1981-82 assessment (No. 13-MA-10). Princeton, NJ: Educational Testing Service.

National Coalition of Advocates for Students. (1985). Barriers to excellence: Our children at risk. Boston: Author.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. Washington, DC: U.S. Government Printing Office.

National Council of Teachers of Mathematics. (1980). An agenda for action: Recommendations for school mathematics of the 1980s. Reston, VA: Author.

National Research Council, Commission on Physical Sciences, Mathematics, and Resources. (1984). Renewing U.S. mathematics: Critical resource for the future. Report of the Ad Hoc Committee on Resources for the Mathematical Sciences. Washington, DC: National Academy Press.

National Science Board Commission on Precollege Education in Mathematics, Science and Technology. (1983). Educating Americans for the 21st century: A plan of action for improving the mathematics, science and technology education for all American elementary and secondary students so that their achievement is the best in the world by 1995. Washington, DC: National Science Foundation.

Oakes, J. (1986). What mathematics, science, and technology education? Indicators of curriculum (Paper prepared for the National Science Foundation project "Monitoring National Progress in Mathematics, Science and Technology Education"). Santa Monica, CA: Rand Corporation.

Pask, G. (1976). Conversation theory: Applications in education and epistemology. Amsterdam: Elsevier.

Pask, G. (1984). Review of conversation theory and a protologic (or protolanguage) Lp. Educational Communications and Technology Journal, 32(1), 3-40.

Pelto, P. J. & Pelto, G. H. (1978). Anthropological research: The structure of an inquiry. Cambridge: Cambridge University Press.

Piaget, J. (1973. The child and reality: Problems of genetic psychology. New York: Grossman.

Pimm, D. (1980). Metaphor and analogy in mathematics. In R. Karplus (Ed.), Proceedings of the Fourth International Conference for the Psychology of Mathematics Education (pp. 157-162). Berkeley, CA: Lawrence Hall of Science, University of California.

Prigogine, I., & Stengers, I. (1984). Order out of chaos: Man's new dialogue with nature. New York: Bantam Books.

Project 2061. (in press). Report of the National Forum of school Science. Washington, DC: American Association for the Advancement of Science.

Reich, R. B. (1983). The next American frontier. Harmondsworth, Middlesex: Penguin Books.

Rescher, N. (1979). Cognitive systematization: A system theoretic approach to a coherentist theory of knowledge. Totowa, NJ: Rowman and Littlefield.

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

Resnick, D. P., & Resnick, L. B. (1985, April). Standards, curriculum and performance: A historical and comparative perspective. Educational Researcher, 14(4), 5-20.

Robinson, J. A. (1984). Logical reasoning in machines. In J. E. Hayes & D. Michie (Eds.), Intelligent systems: The unprecedented opportunity (pp. 19-36). Chichester: Ellis Horwood.

Romberg, T. A. (1983). A common curriculum for mathematics. In G. D. Fenstermacher & J. T. Goodlad (Eds.), Individual differences and the common curriculum: Eighty-second yearbook of the National Society for the Study of Education (Part I, pp. 121-159). Chicago: University of Chicago Press.

Romberg, T. A. (1985, December). The content validity for school mathematics in the U.S. of the mathematics subscores and items for the Second International Mathematics Study (Paper prepared for the Committee on National Statistics, National Research Council of the National Academy of Sciences). Madison: Wisconsin Center for Educational Research.

Romberg, T. A. (1986). National assessment of mathematical performance (Paper prepared for the Study Group on National Assessment). Washington, DC: U.S. Department of Education.

Romberg, T. A., Collis, K. F., Donovan, B. F., Buchanan, A. E., & Romberg, M. N. (1982). The development of mathematical problem-solving superitems. A report on the NIE/ECS Item Development Project. Madison: Wisconsin Center for Educational Research.

Romberg, T. A., & Kilpatrick, J. (1969). Appendix D. Preliminary study on evaluation in mathematics education. In T. A. Romberg & J. W. Wilson (Eds.), The development of tests. NLSMA reports No. 7 (pp. 281-298). Stanford, CA: School Mathematics Study Group.

Romberg, T. A., & Wilson, J. W. (1969). The development of tests. NLSMA reports. No. 7. Stanford, CA: School Mathematics Study Group.

Ruthven, K. (1985). A-level mathematics in an information age. The Mathematical Gazette, 69(448), 103-109.

Saracevic, T. (1976). Relevance: A review of the literature and a framework for thinking on the notion in information science. Advances in Librarianship, 6, 79-138.

Scheffler, I. (1975, October). Basic mathematical skills: Some philosophical and practical remarks. In The NIE Conference on basic mathematical skills and learning, Euclid, Ohio, 1975. Volume I: Contributed position papers (pp. 182-189). Los Alamitos, CA: SWRL Educational Research and Development.

Schwartz, J. (1985). The geometric supposer (Computer software). Pleasantville, NY: Sunburst Communications.

Scruton, P. (1986). Mathematics: GCSE, a guide for teachers. Milton Keynes: Open University Press.

Skemp, R. (1979). Intelligence, learning and action: A foundation for theory and practice in education. Chichester: Wiley.

Skemp, R. (1983). The functioning of intelligence and the understanding of mathematics. In M. Zweng, T. Green, J. Kilpatrick, H. Pollak, & M. Suydam (Eds.), Proceedings of the Fourth International Congress on Mathematical Education (pp. 533-540). Boston: Birkhauser.

Skovsmose, O. (1985). Mathematical education versus critical education. Educational Studies in Mathematics, 16, 337-354.

Suppes, P. (1965). On the behavioral foundations of mathematical concepts. Monographs of the Society for Research in Child Development, 30(1), 61-96.

Thorndike, E. L. (1904). An introduction to the theory of mental and social measurements. New York: Teachers College, Columbia.

Toffler, A. (1985). The adaptive corporation. New York: McGraw-Hill.

Tyler, R. W. (1931). A generalized technique for constructing achievement tests. Educational Research Bulletin, 8, 199-208.

Usiskin, Z. (1984). Needed changes in mathematics curricula. In T. A. Romberg & D. M. Stewart (Eds.), School mathematics: Options for the 1990s. Proceedings of the Conference (pp. 53-62). Madison: Wisconsin Center for Education Research.

Vergnaud, G. (1982, November). Cognitive and developmental psychology and research in mathematics education: Some theoretical and methodological issues. For the Learning of Mathematics, 3(2), 31-41.

Vergnaud, G. (1983a). Multiplicative structures. In R. Lesh & M. Landau (Eds.), Acquisition of mathematics concepts and processes. New York: Academic Press.

Vergnaud, G. (1983b). Pourquoi une perspective epistemologique est-elle necessaire pour la recherche sur l'enseignement des mathematiques? In J. C.Bergeron & N. Herscovics (Eds.), Proceedings of the Fifth Annual Meeting of PME-NA, The North American Chapter of the International Group for the Psychology of Mathematics Education (Vol. 1, pp. 21-40).

Vergnaud, G. (1983c). Why is an epistemological perspective a necessity for research in mathematics education? In J. C. Bergeron & N. Herscovics (Eds.), Proceedings of the fifth annual meeting of PME-NA, The North American Chapter of the International Group for the Psychology of Mathematics Education (Vol. 1, pp. 2-20). Montreal: Université de Montréal, Faculté des Sciences de l'Education.

Vergnaud, G. (1984). Problem-solving and symbolism in the development of mathematical concepts. Proceedings of the Eighth International Conference for the Psychology of Mathematics Education (pp. 27-37). Darlinghurst, NSW: Mathematical Association of New South Wales.

von Glaserfeld, E. (1983). Learning as a constructive activity. In J. C. Bergeron, & N. Herscovics (Eds.), Proceedings of the fifth annual meeting of PME-NA, The North American Chapter of the International Group for the Psychology of Mathematics Education (Vol. 1, pp.41-69). Montreal: Université de Montréal, Faculté des Sciences de l'Education.

Walters, J. M., & Gardner, H. (1985). The development and education of intelligences. In F. Link (Ed.), Essays on the intellect (pp. 1-21). Alexandria, VA: Association for Supervision and Curriculum Development.

Weinzweig, A. I., & Wilson, J. W. (1977, January). Second IEA mathematics study. Suggested tables of specifications for the IEA mathematics tests (Working paper 1). Wellington: IEA.

Westbury, I. (1980, January 7-11). Change and stability in the curriculum: An overview of the questions. In Comparative studies of mathematics curricula: Change and stability 1960-1980 (pp. 12-36). Proceedings of a conference jointly organized by the Institute for the Didactics of Mathematics (IDM) and the International Mathematics Committee of the Second International Mathematics Study of the International

Association for the Evaluation of Educational Achievement (IEA). Bielefeld, FRG: Institut fur Didaktik der Mathematik der Universitat Bielefeld.

Wharmby, D. (1986). SEC-sponsored Development Project on Practical Mathematics: A review of research projects in mathematics. Manchester: Joint Matriculation Board.

Wisconsin Department of Public Instruction. (1986, May). 1985 Wisconsin assessment report (Bulletin No. 6378). Madison, WI: Author.

Wohlwill, J.F. (1973). The study of behavioral development. New York: Academic Press.

Zessoules, R. (1986). Description of the Rockefeller Project in Artistic Learning and Assessment. (Available from Rieneke Zessoules, Harvard Project Zero, Longfellow Hall, Appian Way, Cambridge, MA 02138.)

Chapter 19

# LEVELS OF REASONING AND THE ASSESSMENT OF MATHEMATICAL PERFORMANCE

K. F. Collis

Modern methods of assessing achievement in the classroom have been influenced by three very different research traditions. The first of these is the psychometric approach; the second, Bloom's Taxonomy; and the third, Piaget's theory of cognitive development. Each of these will be discussed briefly to indicate their strengths and limitations and to provide a background against which one might evaluate current theories and practices, particularly in the assessment of mathematical performance.

## The Psychometric Approach

Modern psychometric testing has its origins in research conducted at the turn of the century by such pioneers as Galton in England, Wundt in Germany, and Cattell in America. In 1908 Stone, a student of Thorndike, published the first standardized arithmetic achievement test, and by 1917 more than 200 achievement tests were available for school use, including 11 in arithmetic (Resnick, 1982). Binet, working in France with Simon, in 1905 published the first individually administered intelligence test. The items were arranged in order of increasing difficulty and so constituted the first scale for measuring an individual's level of mental development.

Since that time, enormous numbers of standardized tests of intelligence and achievement have been published, and statistical techniques have become progressively more sophisticated. All standardized tests share certain characteristics. These include a fixed set of items carefully designed and pretested to measure a clearly defined sample of behavior, explicit procedures for administering and objectively scoring the test, and normative data, derived from administering the test to carefully selected groups (often based on age or grade), as an aid to interpreting test scores.

The psychometric model has also guided recommendations for measurement in the classroom. Textbooks on educational measurement and teacher-made tests over the last 40 years typically have required teachers to list instructional objectives in terms of learning outcomes as the first step in evaluating performance. To this end, teachers were instructed to subdivide their curriculum into the separate skills or areas of knowledge they hoped to teach and to select objectives from each such area. Both existing standardized tests and teacher-constructed tests were recommended for different purposes. Teacher-constructed tests ranged from the essay or short answer to objective tests including multiple choice, matching, and true-false items. To improve the quality of the tests they constructed, teachers were taught concepts of validity and reliability, methods of assigning grades, and statistical treatment of the test data.

The extensive use of standardized tests of ability has met with widespread criticism. A multidisciplinary committee established in America to examine testing practices (Wigdor & Garner, 1982) attended primarily to the social and legal implications of ability testing. However, it also concluded that, while the strength of modern mental measurement has been its mathematical and statistical foundations, similar progress has not been made in understanding what is being measured. That is, test construction has not been guided by any powerful psychological theory of the behavior under examination. Rather, there have been two separate approaches to the study of abilities that have not tended to draw strength from one another. The first has focused on internal processes and their ontogenesis, using a variety of clinical techniques; Binet and Piaget are examples of this type of research. The second has concentrated on the external correlates of test scores; pioneers of testing such as Cattell, Galton, Thorndike, and Thurstone worked in this mode. The latter work has generated the advanced psychometric methodology under discussion.

The psychometric model has also proved inadequate for guiding the teaching-learning situation in the classroom. Its lack of integration with a coherent theory of learning has meant that test results provide teachers with little insight into what to do next with their students, or how to overcome problems. Teachers have tended to ignore their psychometric training and to rely on past experience and intuition when selecting test items. In mathematics, especially in the area of elementary applications of mathematics, the emphasis has been placed on mechanical features, such as setting items that range from easy to hard by increasing the number of steps in a problem or making the numbers bigger. The aim has been to obtain a quantitative measure that ranks students and gives an acceptable range and spread of scores, rather than to provide a qualitative account of the students' understanding of content.

## Bloom's Taxonomy

The Taxonomy of Educational Objectives for the cognitive domain (Bloom, 1956) attempted to rectify a situation in the late 1940s in which the methodology of measurement was becoming increasingly sophisticated but notions about what was being measured, particularly in the educational field, remained disorganized. Bloom and his colleagues gathered a large number of educational objectives from institutions, from the literature on curriculum development and evaluation, and from the unpublished material of examiners and curriculum specialists to develop a comprehensive classification system of educational objectives.

In the absence of an existing theoretical base to guide the structuring of the taxonomy, the committee decided to use the naive psychological principle that individual simple behaviors become integrated to form a more complex behavior. Accordingly, the behaviors specified by the cognitive objectives were organized from the simplest to the most complex and placed into six major classes: 1.00 Knowledge; 2.00 Comprehension; 3.00 Application; 4.00 Analysis; 5.00 Synthesis; 6.00 Evaluation.

This was a somewhat tentative ordering of classes, and Bloom (1956) himself expressed some reservations about it:

> Our evidence on this is not entirely satisfactory, but there is an unmistakable trend pointing toward a hierarchy of classes of behavior which is in accordance with our present tentative classification of these behaviors. (p. 19)

The taxonomy has subsequently been widely used to generate techniques for evaluating students' progress toward educational objectives. The Handbook on Formative and Summative Evaluation of Student Learning (Bloom et al., 1971) is an example of the sort of concepts and materials made available to teachers using this framework. It provides models for the evaluation of particular areas of schooling, including an evaluation of learning in secondary school mathematics (Wilson, 1971). Wilson developed a classification matrix that sets levels of behavior against content areas in mathematics. The four main levels of behavior--computation, comprehension, application and analysis--were a modification of Bloom's taxonomy.

There are several problems with Bloom's Taxonomy and the models derived from it which stem from their lack of coherent theoretical base. The taxonomy was developed in the early 1950s, before Piaget's theories had revolutionized educational thinking. Piaget emphasized the qualitatively different nature of the child's thinking from that of the adult and the way in which "knowledge" was actively constructed by the child. Bloom's categories, on the other hand, were established using an entirely different point of departure. His starting point was not children's behaviors at different stages of the learning process (as was Piaget's), but lists of educational objectives, devised by adults who presumably had already mastered the curriculum material, and who were not sensitized to the qualitative changes that occur in cognitive development.

## Piaget

Piaget (1929), picking up the Binet and Simon thread in test construction, developed the clinical interview technique and used it throughout his subsequent research. Piaget's clinical technique was designed to investigate cognitive processes rather than their end products. It involved careful observation and questioning, usually in a one-to-one interview. In essence, it consisted of presenting the same task to children across a range of ages and allowing the examiner to vary the line of questioning, or modify the task, with a view to clarifying the nature of the child's reasoning.

Piaget's methodology has been criticized on several grounds, but his analysis of children's thought processes nevertheless provided a major advance in our understanding of children's logical reasoning at various age levels. He found that there were qualitative differences in the operational structures available to children at different ages. This led to his proposal that there are four main stages of intellectual development from birth to adolescence, each with its characteristic form of logical functioning.

The importance of Piaget's theory in this paper is twofold. First, his theory, when combined with information-processing concepts as they have been applied to human cognition, leads directly to present-day, post-Piagetian structuralist notions of both cognitive development itself and of the learning of specific intellectual skills such as those involved in learning mathematics. Second, his clinical method of investigation has opened the way for significant insights into techniques of evaluation that allow us to assess the level of understanding a student has of a particular content area.

Let us examine very briefly four recent post-Piagetian theories that emphasize structure in the development of intellectual functioning. The theorists to be considered are Fischer (1980), Case (1985), Halford (1986), and Biggs and Collis (1982).

## Cognitive Development Theory in the 1980s

Fischer (1980) integrated behavioral and developmental concepts with a view to providing a method of predicting developmental sequences and synchronies for various domains of human functioning.

He listed 10 clearly distinguishable levels that form a hierarchical sequence and that can be applied *inter alia* to cognitive development. These levels are grouped into three tiers (or stages) according to the level of abstraction of the attributes ascribed to the objects, events, or people involved in the processing. Progression through a tier follows a cycle of four levels represented by specific structures, each of which is defined in terms of set theory. The highest level of one tier becomes the lowest level of the next. Movement from one level to the next occurs according to certain transformation rules and can occur only when the individual controls a skill at a particular level and thus has available a structure that allows for one or more sources of variation.

Case (1985) put forward a theory that, while heavily mediated by information-processing constructs, can be traced back to the original Piagetian formulations. He proposed four major stages of intellectual development from birth to adulthood and identified a universal sequence of three substages that occur within each stage; the highest substage at one level is the lowest substage of the next. Each stage is associated with a particular type of mental element, and each substage is associated with the number and organization of these elements. The latter in turn are related to the short-term memory space available. Case postulated that integration of existing structures is a key notion in considering the acquisition of new processes, both within and between stages. In the latter case, however, he proposed that the transition occurs via hierarchical integration of executive structures that were put together in the earlier stage but whose shape and purpose at that stage were considerably different than at the higher stage. He also listed the processes by which transformation to a higher stage takes place and suggested typical life situations that facilitate this development.

Halford (1986) described cognitive development as a hierarchy of increasingly powerful organizations, where higher level structures combine and integrate lower level ones. His theory argues that higher level organizations make greater information-processing demands than do lower levels, that the amount of information that can be utilized in a single decision increases with age, and that there are minimum ages below which particular mental processes cannot be attained. He described four levels of thought that are hierarchically ordered such that a representational system at one level is a composition of two or more systems at the previous level. Halford holds that there are two kinds of elements involved in thinking: environmental, where the objects and events are actually in the individual's environment; and symbolic, where the elements are the individual's internal representations of objects and events from the environment.

Halford's sole criterion for assigning tasks to a particular level of thought is the minimum amount of information, or number of relationships, required to make a decision. Subjects who operate successfully at a particular level, therefore, must not only have the requisite processing capacity but must be well trained in the individual aspects of the task and operate with maximum efficiency. Within levels, training will therefore affect performance. However, transition between levels is dependent on increased processing capacity, which, according to Halford, is not influenced by training. Although he saw increased processing capacity as a necessary condition for transition across levels, he did not regard it as a sufficient one. Halford (1986) proposed three subsequent processes that facilitate transition. The first is the composition or integration of lower level systems. Once this is achieved, the second process, the detection of consistency and inconsistency, is put into action. This determines whether the new system, at the higher level, is a valid or consistent representation of the corresponding environmental system. Once consistency is established, the third step, the discovery of new rules and applications, may be taken. This allows the new system to be extended to a variety of problem situations within the new level.

Biggs and Collis (1982) put forward a set of proposals compatible with those already described but more thoroughly worked out in terms of evaluating students' responses. As this is the main focus of this paper, their proposal will be described more fully.

A detailed analysis of children's responses to questions asked in a variety of school content areas, of observations recorded in a range of developmental research data, and of observations of skills development in various contexts suggested that there were two phenomena involved in determining the level of an individual's response to an environmental cue. The first was what Biggs and Collis chose to call the Hypothetical Cognitive Structure (HCS) and the second the Structure of the Observed Learning Outcome or Response (SOLO).

The former was closely related to the existing notion of Piaget's stages of cognitive development--sensorimotor (birth to 2 years), intuitive/preoperation or iconic (2 to 6 years), concrete symbolic (7 to 15 years), formal operational (16+ years)--in which each stage has its idiosyncratic mode of functioning and, as far as intellectual development is concerned, its own set of developmental tasks. The latter, on the other hand, was concerned with describing the structure of any given response as a phenomenon in its own right, that is, without the response necessarily representing a particular stage of intellectual development.

The Structure of the Observed Learning Outcome or Response that occurs within each stage becomes increasingly complex as the cycle within that stage develops. Prestructural responses represent no use of relevant aspects of the mode in question; unistructural responses represent the use of only one relevant aspect of the mode; multistructural responses represent several disjoint aspects, usually in a sequence; relational responses involve several aspects related into an integrated whole; and an extended abstract response takes the whole process into a new mode of functioning. These notions may be best summarized by considering the diagram in Table 1.

The first column indicates the various "stages of development" or typical modes of functioning at the various age ranges indicated; the second represents the cycle of learning that recurs at each stage of development; and the third illustrates the model's implications for the psychological concept of conservation as it applies to the extended abstract level of each mode.

Table 1

STAGES OF COGNITIVE DEVELOPMENT AND RESPONSE DESCRIPTION

| 1<br>Mode<br>(Developmental Stage) | 2<br>Response Structure<br>(Learning Cycle) | 3<br>Example,<br>Conservation |
|---|---|---|
| Sensorimotor (infancy) | Unistructural<br>Multistructural<br>Relational = Prestructural | |
| Intuitive/Preoperational or Iconic (early childhood to preschool) | Extended Abstract = Unistructural<br>Multistructural<br>Prestructural = Relational | Objects |
| Concrete Symbolic (childhood to adolescence) | Unistructural = Extended Abstract<br>Multistructural<br>Relational = Prestructural | Classes |
| Formal--1st order (early adult) | Extended Abstract = Unistructural<br>Multistructural<br>Prestructural = Relational | Systems |
| Formal--2nd order and higher order (adult) | Unistructural = Extended Abstract<br>Multistructural<br>etc. | Theories (of increasingly higher order) |

Perhaps the most outstanding feature of the above model is the marriage between the cyclical nature of learning and the hierarchical nature of cognitive development. Each level of functioning within a cycle has its own integrity, its own idiosyncratic selection and use of data, and yet each provides the building blocks for the next higher level. The movement from relational to extended abstract within a cycle marks the transition to a new mode of functioning, a new stage of development. The next higher mode subsumes the earlier one and then proceeds through a similar structural reorganization until it eventually is itself subsumed. Such absorption is not entire, however, as the learner always has the option of operating at a lower level than the one attained. This last fact is of considerable importance when we come to assessing student responses.

There is little theoretical difficulty with the question of learning within modes. Basically, for a given task or skill, this can be related to general (nonstructuralist) variables, such as simultaneous and successive processing and working memory capacity (M-space). The latter concept is of particular importance, as the M-space available to complete the necessary operations involved in the task and to monitor the processes involved is directly related to the complexity of the task that can be handled successfully at that stage. Indeed, progress through a mode can be seen in terms of an increasing degree of automaticity and familiarity that the individual achieves with the task elements and operations involved. The more familiar the individual becomes with these variables, the more M-space is cleared for processing the data.

The question of transition across modes, however, is more intractable. It is possible that there are fundamental endogenous processes at work that we have not considered to date. Epstein (1978), for example, pointed out that certain periods of rapid growth in brain-associated areas coincide fairly well with the periods of cognitive change noted by the Piagetians. It would be premature, however, to elaborate on how such physiological growth phases may affect cognitive functioning.

Instead, let us look more closely at the question of transition itself. Within each mode of functioning, there is an increasing development of power to organize and control the individual's interactions with the environment. Paradoxically, this increasing power, represented by higher-level responses within the current mode of functioning, sows the seeds for the individual to recognize the inadequacies of that mode and thus causes a striving to raise the level of functioning (Halford, 1970).

For example, the individual responding at the relational level in the concrete symbolic mode is able to use all the data and their interrelationships to come to a generalization. This represents a considerable increase in power over the previous multistructural response in the same mode, where decisions were reached by a selection of unrelated data from those given. However, the person responding at the relational level is likely to make hasty overgeneralizations that will cause inconsistent judgments. If the area of inconsistency is significant to the individual, an attempt will be made to resolve it (Halford, 1970), because consistency leads to increasing control over the environment. However, resolution of inconsistency only comes about by upward movement to the next level of functioning.

The most remarkable aspect of the four theories outlined above is the common threads running through them despite their different points of departure and the different methods by which they came to their conclusions. It is true that there are clearly distinguishable theoretical differences among them. Case, Halford, and Biggs and Collis give much more emphasis to the importance of working memory capacity than Fischer, for example; Halford seems more wedded to the developmental stage notion than the others; the Case and Halford views of the point at which formal operations normally begin differs by some years from the views of Fischer and of Biggs and Collis; the transformation rules differ somewhat from theory to theory and indeed are worked out more systematically in some than in others. Their differences are vitally important to the science of psychology, but their common elements are highly significant for planning teaching strategies, curriculum content, and assessment techniques.

All four theories regard cognitive development as a series of hierarchical skill structures that can be grouped into sets of levels (for convenience, a set of levels may be called a stage of development). These sets of levels incorporate skills of gradually increasing complexity, with a skill at a higher level developing directly from specific skills at the preceding level. The processes of development within each stage are parallel from stage to stage and involve the capacity to cope with increasingly abstract concepts.

While all normal human beings appear to attain a form of logical functioning by adolescence, specific intellectual skills involving mathematics, for example, are only developed by careful and lengthy attention to their attainment. That is, the general level of cognitive skill achieved by average 4- to 6-year-old children enables them to begin work on the development of the specific intellectual skills involved in mathematics (or other academic subjects such as reading and writing), but these skills will reach a high level only with careful attention to skill development and practice. Moreover, specific skills seem to feed into and enhance the individual's general level of cognition. Each of the theories can handle this specific skill development in a variety of academic content areas, as well as the development of more general logical functioning.

## The SOLO Taxonomy

This paper is most concerned with the insights that these recent theories provide for assessment of school learning. Of the four theories outlined above, the Biggs and Collis formulation has concentrated on the evaluation of the quality of school learning. Biggs and Collis took the cycle of learning associated with their concrete symbolic mode of functioning and applied it to a wide range of academic content areas from early elementary school to senior college and university. They found that a student's response could be analyzed and evaluated according to its structure and categorized according to the level it reached in the learning cycle. Independent support for this approach has been supplied by Marton (1981).

Marton's qualitative categories were devised in a way similar to that described by Biggs and Collis when they set up a particular cycle of learning; that is, the structure of a particular response is regarded as a phenomenon in its own right. Marton, like Biggs and Collis' SOLO taxonomy approach, is concerned with providing practitioners, researchers, and teachers the tools to analyze and react to student responses.

As originally developed by Collis and Biggs (1979), and Biggs and Collis (1982), the SOLO taxonomy used an open response format in which student responses were examined for structural organization by an assessor. A later development (Collis & Romberg, 1981) enabled the technique to be used in a closed format. Let us look at some examples of these two formats.

SOLO Taxonomy: Open Format. In this form the student is either given information and asked a question requiring a response, or given a task that requires the student to draw on his or her long-term memory store for suitable data to complete the task. An example of the first type of task, taken from the history content area, is presented in Figure 1, with comments indicating the SOLO analysis of a selection of responses. The comments after each example of a response at a particular level indicate both the criteria used for the categorization and the typical *modus operandi* of students responding at that level.

The study of ancient history in particular often requires an interpretation of a display when some crucial evidence is missing. Lodwick (reported in Peel, 1959) presented children aged 7:6 to 15 years with the passage in Figure 1 and a picture of Stonehenge.

The type of task in Figure 2 would apply to creative writing tasks, where the student is expected to recall the relevant facts as well as to organize them into an argument. The open technique presents particular difficulty with categorizing responses in mathematics in that the student's response does not always indicate how the material was manipulated to obtain the result. Thus, in the absence of the student's actual "working," the assessor must interview the student; the examples in Figure 2 were identified by interview. Of course, having found the responses that represent a

The Function of Stonehenge

Stonehenge is in the South of England, on the flat plain of Salisbury. There is a ring of very big stones which the picture shows. Some of the stones have fallen down and some have disappeared from the place. The people who lived in England in those days we call Bronze Age Men. Long before there were any towns, Stonehenge was a temple for worship and sacrifice. Some of the stones were brought from the nearby hills but others, which we call Blue Stones, we think came from the mountains of Wales.

Question: Do you think Stonehenge might have been a fort and not a temple? Why do you think that?

Prestructural

"A temple because people live in it."

"It can't be a fort or a temple because those big stones have fallen over."

Comment: The first response shows a lack of understanding of the material presented and of the implication of the question. The student is vaguely aware of "temple," "people," and "living," and he uses these disconnected data from the story, picture, and questions to form his response. In the second response, the pupil has focused on an irrelevant aspect of the picture.

Unistructural

"It looks more like a temple because they are all in circles."

"It could have been a fort because some of those big stones have been pushed over."

Comment: These students have focused on one aspect of the data and have used it to support their answer to the question.

Multistructural

"It might have been a fort because it looks like it would stand up to it. They used to build castles out of stone in those days. It looks like you could defend it too."

"It is more likely that Stonehenge was a temple because it looks like a kind of design all in circles and they have gone to a lot of trouble."

Comment: These students have chosen an answer to the question (i.e., they have required a closed result) by considering a few features that stand out for them in the data, and have treated those features as independent and unrelated. They have not weighed the pros and cons of each alternative and come to a balanced conclusion on the probabilities.

Relational

"I think it would be a temple because it has a round formation with an altar at the top end. I think it was used for worship of the sun god. There was no roof on it so that the sun shines right into the temple. There is a lot of hard work and labor in it for a god and the fact they brought the blue stone from Wales. Anyway it's unlikely they'd build a fort in the middle of a plain."

Comment: This is a more thoughtful response than the previous ones; it incorporates most of the data, considers the alternatives, and interrelates the facts.

Extended Abstract

"Stonehenge is one of the many monuments from the past about which there are a number of theories. It may have been a fort but the evidence suggests it was more likely to have been a temple. Archaeologists think that there were three different periods in its construction so it seems unlikely to have been a fort. The circular design and the Blue Stones from Wales make it seem reasonable that Stonehenge was built as a place of worship. It has been suggested that it was for the worship of the sun god because at a certain time of the year the sun shines along a path to the altar stone. There is a theory that its construction has astrological significance or that the outside ring of pits was used to record time. There are many explanations about Stonehenge but nobody really knows."

Comment: This response reveals the student's ability to hold the result unclosed while he considers evidence from both points of view. The student has introduced information from outside the data, and the structure of his response reveals his ability to reason deductively.

Figure 1. Open history item: Constructing a plausible interpretation from incomplete data. (Biggs & Collis, 1982, pp. 47-49)

Find the value of $\triangle$ in the following statement:

$(72 \div 36) \times 9 = (72 \times 9) \div (\triangle \times 9)$

Prestructural responses

"Have not done ones like that before, so I can't do it."

"Don't want to do it."

Comment: Both respondents indicate that they are unwilling to engage in the task.

Unistructural responses

"36 - because there is no 36 on the other side."

"2 - because $72 \div 36 = 2$."

Comment: Both responses take only one part of the data into account. The first response shows a low level "pattern completion" strategy. The second response shows one closure and then an ignoring of the remainder of the item. Both of these strategies give "correct" responses to certain items; for example, the correct answer to the item $3 + 4 = 4 +$   is readily obtained by the first strategy or a slight variation of the second.

Multistructural response

$2 \times 9 = 18$, and $648 \div (\triangle \times 9)$

$648 \div ? = 2$ that is, 324 (looking for 18 $(2 \times 9)$)

Hence 324

Comment: This response incorporates a series of arithmetical enclosures to reduce the complexity and to focus on "$\triangle$." However, the students appear unable to keep the overall relationship in mind throughout the closure sequences and lost in a "maze" of their own creation.

Relational response

$2 \times 9 = 18$, and $648 \div (\triangle \times 9)$

$648 \div 9 = 72$, then $72 \div 4 = 18$

Hence 4

Comment: This response also involves a sequence of arithmetical closures, but the students are able to keep the relationships within the statement in mind and thus successfully solve the problem.

Extended abstract response

First step involves obtaining an overview of the relationships between the numbers and operations involved, for example:

$(72 \div 36) \times 9 = (72 \times 9) \div (\triangle \times 9)$

The pattern suggests something akin to the "distributive" property--this hypothesis is tested out thus:

$$\frac{a}{b} \times y = \frac{a \times y}{b}$$

This immediately solves the problem (without necessity for closure) as follows:

$(72 \div 36) \times 9 = (72 \times 9) \div 36 = (72 \times 9) \div (4 \times 9)$

Hence 4

Comment: This response shows the following characteristics:

1. Focusing on the relationships between the operations and the numbers rather than regarding the operations as instructions to close;

2. a hypothesis suggested by the data;

3. avoiding closures wherever possible as these change the form of the statement and "hide" the original relationship.

Figure 2. Open mathematics item.
(Biggs & Collis, 1982, pp. 83-84)

particular category of functioning, assessors could use this knowledge to set multiple-choice questions so long as they avoid obvious pitfalls, such as those mentioned in Figure 2 with respect to the item, "find the value of $\triangle$, in the following statement, $3 + 4 = 4 + \triangle$."

SOLO Taxonomy: Closed Format. This form was developed initially for use in testing mathematical problem solving (Collis, 1982; Collis, Romberg, & Jurdak, in press) by combining the superitem technique devised by Cureton (1965) with the cycle of learning notion from the SOLO taxonomy. It requires the writing of an item stem that provides data for four questions devised in such a way that each requires an ability to respond at one of the SOLO levels: unistructural, multistructural, relational, or extended abstract. The basic criteria for designing the questions are as follows:

Unistructural: use of one obvious piece of information coming directly from the stem;

Multistructural: use of two or more discrete closures directly related to separate pieces of information in the stem;

Relational: use of two or more closures directly related to an integrated understanding of the information in the stem;

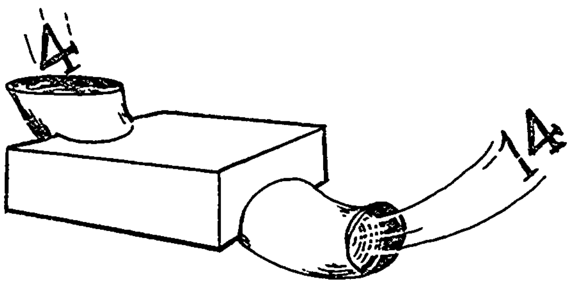
Extended abstract: use of an abstract general principle or hypothesis which is derived from (or suggested by) the information in the stem.

The method of construction and certain psychometric analysis on the data gathered are in press for both mathematical problem solving (Collis, Romberg & Jurdak, in press) and school science (Collis & Davey, in press); work is also in progress on this type of format for the social science area. Examples from mathematics and science, with some explanatory comment, are set out in Figures 3 and 4.

SOLO Taxonomy and Psychometric Analysis. The psychometric analyses carried out so far on the data generated by SOLO items, both open and closed formats, seem to indicate the usefulness of the technique. Although the SOLO procedures are of relatively recent origin, some results are available for both open and closed versions. There is insufficient space in this paper to develop the details in full, but a summary of some of the results for both versions seems appropriate at this point.

Studies (Biggs & Collis, 1982) using the open format have shown this technique to have good reliability (interjudge agreement: correlation coefficients between .71 and .95) and validity (teacher rating of response vs. SOLO level independently rated: correlation coefficients between .65 and .75).

This is a machine that changes numbers. It adds the number you put in three times and then adds 2 more. So, if you put in 4, it puts out 14.




U. If 14 is put out, what number was put in?
Answer..............

Answer: 4

Comment: Students have to understand the problem well enough to be able to close on the correct response which is displayed in the stem.

M. If we put in a 5, what number will the machine put out?
Answer..............

Answer: 17

Comment: Students need to comprehend the set problem sufficiently to be able to use the given statements as a recipe and thus perform a sequence of closures which they do not necessarily relate to one another.

R. If we got out a 41, what number was put in?
Answer..............

Answer: 13

Comment: An integrated understanding of the statements in the problem is necessary to carry out a successful solution strategy in this case. Correct solutions may involve working backwards or carrying out a series of approximation trials. It should be noted that the solution requires only data-constrained reasoning in that no abstract principles need to be invoked.

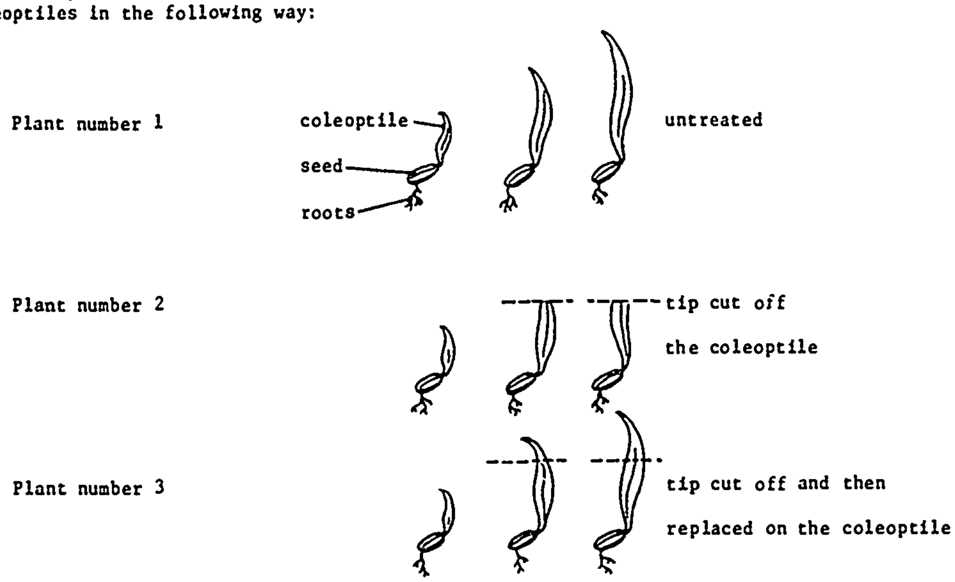
E. If "X" is the number that comes out of the machine when the number "Y" is put in, write down a formula which will give us the value of "Y" whatever the value of "X."
Answer..............

Answer: $Y = \frac{X - 2}{3}$

Comment: A correct response involves extracting the relationships from the problem and setting them down in an abstract formula. It involves using the information given in a way quite different from that of the lower levels.

Figure 3. Closed mathematics item.
(Adapted from Collis, Romberg, & Jurdak, in press)

A student performed an experiment in which he germinated three oat seeds and treated the coleoptiles in the following way:




| Plant number | Height in cms at start | 1 week | 2 weeks |
|---|---|---|---|
| Plant number 1 | 1 | 2 | 2.5 |
| Plant number 2 | 1 | 1.4 | 1.4 |
| Plant number 3 | 1 | 2 | 2.5 |

U. Which oat seedling had the tip cut off its coleoptile and not replaced?
Answer...............

Answer: Plant number 2

<u>Comment</u>: Students must understand the problem well enough to select (or close on) the correct piece of information clearly displayed in the stem.

M. What is the height difference after two weeks between the seedling which had its tip removed but not replaced and the seedling which had its tip removed then replaced?
Answer...............

Answer: 1.1 cms

<u>Comment</u>: Students must understand the problem well enough to make a sequence of appropriate selections from the data displayed in the stem and use them to come to a conclusion.

R. How does the coleoptile tip affect the growth of a seedling?
Answer...............

Answer: Growth takes place at the tip; no tip, no growth.

<u>Comment</u>: An integrated understanding of the various data displayed in the stem is necessary to extract this general principle. It should be noted that the principle is still data bound.

E. Develop a general theory that could have been tested by the above experiment, and list three other factors that would need to be controlled.
Answer...............

<u>Comment</u>: Several responses would be acceptable so long as the student showed familiarity with the bases of scientific experimentation, as well as some knowledge of plant biology. A response at this level requires the student to go outside the given data to hypothesis formulation and abstract principles, and to then use the data given as specific information on which to test the abstractions. The use of data at this level of response is quite different from its use at the lower levels.

Figure 4. Closed science item.
(Adapted from Collis & Davey, in press)

Factor analysis confirmed that two aspects of achievement were measured: one relies on a pinpointing ability to identify the correct answer; the other on a relating ability to take aspects of a situation and integrate them. In addition, canonical analysis suggested that SOLO is closely involved with school achievement, and that high SOLO levels are obtained by highly intrinsically motivated students who search for meaning and who avoid rote learning facts and details. The studies indicate that high quality of learning, as indexed by high SOLO levels, is different from high quantity learning that involves the reception and retention of facts.

Results for the closed format data are similarly supportive, (Collis, Romberg, & Jurdak, 1982; Collis & Davey, in press). Items were examined for their scaleability in the Guttman (1941) sense. The indices used for this purpose were the coefficient of reproducibility (Guttman 1941) supported by the goodness of fit procedures derived by Proctor (1970). The results were highly significant and positive. In addition, cluster analysis revealed that students at various age levels could be assigned to interpretable groups that reflected the sequence of SOLO levels. The results indicated the utility of the SOLO responses categories for evaluation purposes. Finally, Wilson (1985) examined closed SOLO science items from the perspective of a family of Rasch measurement models and found that, in general, these analyses confirmed the hypothesized patterns of the learner's responses.

## SOLO Taxonomy in Instructional Design

Although the taxonomy is applicable to a wide range of skills, the discussion in this paper is focused on academic skills--those associated with school subjects. Academic subjects are taught with two main effects in mind: the facts and concepts that constitute knowledge of the subject, and the cognitive processes that are induced by a proper understanding and application of the subject, the way of thinking for that subject. Moreover, it is a reason·ble hypothesis that development of skills in this latter domain interact with and enhance the general level of intellectual functioning. Leaving aside this last speculation, however, it is clear that learning an academic subject has dimensions of both content and process. Bruner (1960) emphasized the interaction between content and process and put forward the notion of "the spiral curriculum" on the basis that the content/process dimensions of a content area are assimilated and understood on a cumulative basis. In his view, understanding increased with appropriate experience and cognitive maturity. It is in measuring the quality of assimilation in terms of progressive structural complexity that the SOLO taxonomy has its main strength. It is concerned with specifying "how well" (qualitative) something is learned, rather than "how much" (quantitative). This distinction is important, especially in school mathematics where the demand in recent times has begun to place a premium on understanding applications and developing problem-solving skills.

With this background, it is clear that there are three basic areas in which the technique described would be very useful: curriculum analysis, teaching procedures, and assessment of performance. Let us take these one at a time and examine them briefly from the SOLO perspective.

Curriculum Analysis. Two aspects of curriculum analysis are particularly amenable to explication using the technique under discussion, namely, task analysis and specifying adequate performance levels.

Keeping in mind the cycle of learning structure and the various associated levels, the critical variables in analyzing a task become clear. It is not possible to go into details in this paper, but some general principles can be suggested. First, it is necessary to determine the basic elements involved in the task; these minimal features must be defined so that achievement of one would indicate a unistructural response level, and achievement of several, a multistructural level. The next step would be to establish a relating concept that would identify the movement to a relational response. The step up to an extended abstract response is marked by hypothesis testing and the use of the information in a new way. Instead of using the elements provided to determine the response, the individual goes outside the "givens" to formulate a relevant hypothesis and then examines the hypothesis in relation to what is given. These steps are clearly discernible in the mathematics examples given earlier. With respect to the open example, successful performance of the closure of a simple binary arithmetical operation is the basic element; the ability to keep all the relationships in mind while closing a sequence of these operations is the relating factor. The closed item shows another example of the same phenomenon.

Specification of adequate performance is a critical issue. In the past, it has been difficult to make specifications because of lack of suitable objective criteria. The SOLO technique can be used to define realistic levels of expected performance and then to monitor their achievement. For example, analysis of the reading material bought by the vast majority of the community at large shows that it would be classified at a multistructural level in the concrete symbolic mode. It is equally clear that most people, even those in many professions, do not need to be able to respond to mathematical tasks at better than relational level, concrete symbolic mode. In respect to this, the credibility of many academic subjects, including mathematics, has been damaged over the years by setting up higher than necessary achievement as a basic goal for entry to various occupations and professions. It can be readily demonstrated that many of even the most prestigious professions have no need for mathematics beyond the relational level, concrete symbolic mode. Current research indicates that the move from relational level to extended abstract level functioning is much bigger than the moves within a level and involves a high level of commitment from the person concerned. First of all, the individual must recognize that inconsistencies in judgment arise as a result of relational level, concrete symbolic mode, responding; the individual must then work hard at resolving them. Working hard in this context means spending many hours at tasks in which negative feedback is the norm in the early stages. Motivation must be extremely high, and eventual success at the task must offer significant personal reward.

It can be seen that, particularly in school-related activities, there is an alternative to making the effort required to raise the level of functioning, and that is to drop out of the activity involved. This alternative will be familiar to teachers in the middle ranges of high school. Many students recognize implicitly that it is possible to cope with the demands of everyday living, including holding a lower-level but technically skilled job and raising a family, without responding above the level represented by a relational structure in the concrete symbolic mode in many academic activities.

A survey of community achievement levels as they relate to expectations in various content areas would be of enormous benefit to curriculum workers setting up course programs. Once these programs have been set, evaluation and monitoring of individual student performance can take place with an eye on the achievement of a particular level of performance appropriate to the student's interest, ability, and ambitions.

Teaching Procedures. There are several ways in which the SOLO procedures can assist in thinking about the most effective instructional methods. Perhaps the most obvious is the assistance it can afford in adjusting the level of exposition to the level of the students' current performance. As the SOLO level is a measure of the complexity of the content, and the teacher can determine the SOLO level at which students are responding, it is possible to make a reasoned judgment about the level at which to set instruction. It may be appropriate to set it so that the levels match, or it may be appropriate to use the "plus one" strategy (Rest, Turiel, & Kohlberg, 1969) whereby the instruction is pitched at one level above the average response level of the class. It would appear to be nonproductive, for example, to attempt to present content at the extended abstract level to a group whose responses indicated unistructural or multistructural levels of functioning.

In the instructional context, the importance of a student's prior knowledge for likely current performance is highlighted by the SOLO approach. It is obvious from the examples given earlier that one of the determinants of higher level responding is how much and how well the student has grasped the information and concepts taught previously. If the student has not thoroughly automated the basic elements, he or she will be unable to use the concepts, skills, and discriminations necessary for relational and extended-abstract responses. All of this is well known, and teachers usually are careful to design instruction to fit what they believe the students already know. The particular contribution of SOLO here is that component analysis can define the target concepts or skills that are the key to performing the set task.

Assessment of Performance. In an important sense, this entire paper has addressed the value of the SOLO technique in evaluation, but it appears useful at this point to indicate several of its specific features. The technique would seem to fulfill in a meaningful way the fundamental requirements of an evaluation procedure in that it has immediate and direct relevance to curriculum content and teaching procedures and allows for end-of-course gradings. Moreover, it can provide both a diagnostic and a monitoring function in all three contexts. There are several features that make this possible:

1. It provides a vocabulary for describing the levels of attainment.
2. Target levels of achievement can be set with easily understood criteria.
3. Students can be assessed on an individual skill, and the teacher can know what is required to arrive at the next level.
4. It is oriented towards finding out the level of functioning rather than ranking and classifying.

If the more formal terms of measurement theory are evoked, it can be said (a) SOLO is useful for both formative and summative evaluation, although its major use would be in the former mode, and (b) it is suitable for both norm-referenced and criterion-referenced evaluation, although it has most to contribute to the latter.

## Conclusion

The SOLO taxonomy has been designed within the framework of cognitive development theory. It has sought to extract and amalgamate what is most useful from the statistical techniques of psychometric testing and the clinical procedures of Piaget to provide a structure that will help the educator make judgments about the quality of classroom learning. Its use in this context presupposes that the teachers, curriculum workers, or evaluation experts have clear-cut definite intentions concerning the amount and quality of learning that is to take place, and that they can analyze the skills to be taught into their component parts in terms of basic elements and relating factors. While the model is ideal for assessing mastery of academic material and problem solving, both fundamental aims of education, it is not meant to apply to other important but open-ended aspects of the child's educational experience such as learning social skills and attitudes. Nor does it apply to straight fact learning, which has its place in certain parts of the curriculum.

## References

Biggs, J. B., & Collis, K. F. (1982). Evaluating the quality of learning: The SOLO Taxonomy. New York: Academic Press.

Bloom, B. S. (Ed.). (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook 1. Cognitive domain. New York: Longmans, Green and Co.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill.

Bruner, J. S. The process of education. (1960). Cambridge, MA: Harvard University Press.

Case, Robbie. (1985). Intellectual development: Birth to adulthood. New York: Academic Press.

Collis, K. F. (1982). The SOLO Taxonomy as a basis of assessing levels of reasoning in mathematical problem solving. In Proceedings of the Sixth International Conference for the Psychology of Mathematical Education. Antwerp, Belgium: University of Antwerp.

Collis, K. F., & Biggs, J. B. (1979). Classroom examples of cognitive development phenomena: The SOLO Taxonomy. University of Tasmania: ERDC Report.

Collis, K. F., & Davey, H. A. (in press). A technique for evaluating skills in high school science. Journal of Research in Science Teaching.

Collis, K. F., & Romberg, T. A. (1981). Cognitive functioning, classroom learning and evaluation: Two projects. In Proceedings of the Fifth International Conference for the Psychology of Mathematical Education (pp. 64-77). Grenoble, France: University of Grenoble.

Collis, K. F., Romberg, T. A., & Jurdak, M. E. (1986). A technique for assessing mathematical problem solving ability. Journal for Research in Mathematics Education, 17(3), 206-221.

Cureton, E. E. (1965). Reliability and validity: Basic assumptions and experimental designs. Educational and Psychological Measurement, 25, 326-346.

Epstein, H. T. (1978). Growth spurts during brain development: Implications for educational theory and practice. In J. S. Chall & A. F. Masky (Eds.), Education and the brain: Seventy-seventh yearbook of the National Society for the Study of Education (Part II). Chicago: University of Chicago Press.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. Psychological Review, 87, 477-531.

Guttman, L. (1941). The quantification of a class of attributes: A theory and method for scale construction. In P. Horst (Ed.), The prediction of personal adjustment (pp. 319-348). New York: Social Science Research Council.

Halford, G. S. (1970). A theory of the acquisition of conservation. Psychological Review, 77, 302-316.

Halford, G. S. (in press). Children's utilization of information: A basic factor in cognitive development. New York: Academic Press.

Marton, Ference. (1981). Phenomenography--describing conceptions of the world around us. Instructional Science, 10, 177-200.

Peel, E. A. (1959). Experimental examination of some of Piaget's schemata concerning children's perception and thinking, and a discussion of their educational significance. British Journal of Educational Psychology, 29, 59-103.

Piaget, J. (1929). The child's conception of the world. New York: Harcourt.

Proctor, C. H. (1970). A probablistic formulation and statistical analysis of Guttman scaling. Psychometrika, 35, 73-78.

Resnick, D. (1982). History of educational testing. In A. K. Wigdor & W. R. Garner (Eds.), Ability testing: Uses, consequences and controversies, Part II (pp. 173-194). Washington, DC: National Academy Press.

Rest, J., Turiel, E., & Kohlberg, L. (1969). Relations between level of moral judgment and preference and comprehension of the moral judgment of others. Journal of Personality, 37, 225-252.

Stone, C. W. (1908). Arithmetical abilities and some factors determining them (Contributions to Education, No. 19). New York: Teachers College, Columbia University.

Widgor, A. K., & Garner, W. R. (Eds.). (1982). Ability testing: Uses, consequences and controversies, Part I. Washington, DC: National Academy Press.

Wilson, J. W. (1971). Evaluation of learning in secondary school mathematics. In B. S. Bloom, J. T. Hastings, & G. F. Madaus (Eds.). Handbook on formative and summative evaluation of student learning (pp. 643-696). New York: McGraw-Hill.

Wilson, M. (1985). Modelling SOLO science items. In Proceedings of the Annual Conference of the Australian Association for Research in Education (pp. 266-267). Hobart: AARE.

Chapter 20

# KNOWLEDGE STRUCTURES AND ASSESSMENT OF MATHEMATICAL UNDERSTANDING

Brian F. Donovan and Thomas A. Romberg

## Introduction

The purpose of this paper is to describe the need for a fundamental reappraisal of the content of school mathematics, to propose an alternative view of how knowledge is structured, and to outline an assessment strategy related to that alternative.

Toffler (1980) regarded today's social and economic changes as interdependent and argued that to view them as largely isolated was to miss their larger significance. Such a view also prevents design of a coherent and effective response.

> So profoundly revolutionary is this new civilization that it challenges all our old assumptions. Old ways of thinking, old formulas, dogmas and ideologies, no matter how cherished or useful in the past, no longer fit the facts. The world that is fast emerging from the clash of new values and technologies, new geographical relationships, new lifestyles and modes of communication demands wholly new ideas and analogies, classifications and concepts (Toffler, 1980, p. 18).

Toffler (1980) used the clash of waves as a metaphor for charting the history of civilization. Until now, the human race had experienced two great waves of change, each of which largely obliterated earlier cultures or civilizations and replaced them with ways of life inconceivable to those that went before. The first, the agricultural revolution, lasted thousands of years before playing itself out. The second, the rise of industrial civilization, lasted a few hundred years. Toffler suggested that the third wave has already arrived and is likely to complete itself in just tens of years.

First-wave societies drew energy from human and animal power, or "living batteries," as Toffler (1980) described them, as well as from sun, wind, and water. In the second wave, the mechanical engine provided energy. The third wave has substituted some works of the human brain with the intelligent machine. The second wave might be characterized as providing artificial arms; the third wave is providing artificial brains which produce artificial intelligence.

Papy (1982) associated particular mathematical knowledge with each of the three waves. He saw the mathematics of the first wave exemplified by the geometry of idealized physical space. The mathematics of the second wave broke the Euclidean hold and, in this industrial period, gave rise to calculus, matrices, various spaces and the structures emphasized by Bourbaki (e.g. 1968). Papy noted that the need for a well-defined spatial territory was evident in the regeometrization of "modern mathematics," marked by the creation of a collection of spaces, including, for example, vector spaces, topological spaces, Hilbert spaces, and Banach spaces. Papy described the mathematics of the third wave as being that of the most conceptual aspects of the great abstract structures of Bourbaki (1968). Papy (1982) concluded that the fundamental importance of a conceptual approach surpasses the possibilities offered by the artificial brains of the third wave:

> As all the very important computational aspects of the second wave can be performed by comput rs, the conceptual aspect becomes more and more important and fundamental. Because of the hand-calculators, it is not anymore important to teach a child to compute long numerical calculations, but the pupil has to know more than before the meaning of the operations and of the other concepts of mathematics. (p. 39)

The meaning and consequences of a conceptual approach for the third wave, and the critical deficiency of a second wave view, is more fully analyzed by Romberg (1984), who characterized the second wave perspective of schooling as a mechanical view growing out of the machine-age thinking of the industrial revolution. The intellectual contents of the machine age, according to Romberg, rest on three fundamental ideas: reductionism, analytical processes, and mechanism. Reductionism refers to a preoccupation with taking things apart. Under such a perspective, perceptions and experiences are viewed as the sum of parts; the fragmenting of mathematics into pieces is a natural product of this approach. The second idea, analytical processes, is based on reductionism. It emphasizes that problem solving is most facilitated by a process of breaking into components, then rebuilding the whole. Mechanism, the third fundamental idea, is based on the theory that all phenomena can be explained in terms of cause-and-effect relationships.

## Manufacturing Versus Revealing

Romberg's (1984) description of the fundamental characteristics of the second wave have been contrasted with third-wave characteristics taken from Toffler (1980). In this section, Heidegger's (1977) interrogation of technology is examined to disclose second wave practices and thinking that are helpful in considering a third wave alternative for knowing school mathematics.

Heidegger (1977) argued that technology cannot be understood as a means to an end. As such, this apparently value-free view is correct, but limited. It is an instrumentalist conception, insufficient to disclose the essence of technology. While it focuses on a pertinent element in technology, it can only condition attempts to recognize human agency in its proper relation to technology. Heidegger insisted that the two statements, "Technology is a means to an end," and "Technology is a human activity," reveal the true nature of technology. He opposed a reductionist approach to the definition of technology; just as he proposed that the instrumental and anthropological aspects of technology be considered in their dynamic and mutual relationship, a similar requirement is necessary for an examination of the fundamentals of school mathematics appropriate to the third wave.

Instruction and learning in the third wave require a consolidation of content matter and curricular form and process, with a view to revealing human interactions with others and with the environment. In this sense, content does not stand apart. Nor can it stand as a curriculum object, even when related to aims, methods and evaluation, as in Tylerian rationalism. This rationality does not sufficiently disclose the human agency and interest bases in curriculum, including content. Content is transformed by teachers and students acting within a complex of ends and bounds as they develop definitions of mathematics knowledge within the dynamics of their particular social setting. This points to knowing mathematics as problematic. It does not mean that content is of little significance, nor that content structures are out of place. Rather, it is a recognition that knowing and doing mathematics surpass manufacturing products to reveal human possibility. Presently, procedural knowledge, which is a manufactured product, is dominant in the content and practices of school mathematics.

Procedural knowledge, that is, skills development, has a life of its own. Where it was once a means to an end, it has become the goal. Popkewitz, Tabachnick and Wehlage (1982) point to pedagogical, ideological and sociocultural interests that seem to perpetuate such instrumental approaches. Diagnosed as skills development, procedural knowledge is associated with the will to master and manage learning more efficiently and effectively, but it has perverted what it means to know mathematics. As an instrumental conception of knowing mathematics, it seems to have conditioned attempts for people to have a right relation to knowledge. In formulating aims and objectives, in defining basics, we must keep in mind that the proposing of ends and means is a human activity. That which is known is integrally related to the "knower."

Manufacturing content in school mathematics has as its industrial wave equivalent mass production on factory assembly lines. It is evident in the prominence of procedural knowledge, in the dominance of skill over critical and conceptual development, and in the fragmentation of content that is consistent with this packaging of knowledge. In manufacturing, analysis processes have been employed to break down the desired production of learned outcomes into component parts to facilitate production, and to make it more efficient and effective. Students are to work in such production but usually lack ownership of its processes. Not surprisingly, therefore, students demonstrate various forms of resistance observed by researchers in relation to social class, gender and race (cf. Anyon, 1981). In manufacturing terms, this wastes resources. From a perspective of learning as revealing, it degenerates human possibility and will not stimulate new questions or disclose new approaches to new problems in a changed and ever-changing world.

Alternative content should be built upon a recognition that knowledge is socially constructed. In particular, it should acknowledge that students construct their own knowledge and that learning should be directed towards the development of general principles and critical awareness. The industrial wave characteristics of fragmentation, analysis, and mechanism disguise such recognition and limit more creative and critical human possibility. Indicators of third wave school mathematics will include context and holism, synthesis, and acknowledgement of the problematic nature of knowledge. Conceptual fields offer possibilities for such new fundamentals in school mathematics.

## Conceptual Fields

Gerard Vergnaud, of the Centre d'Etude des Processes Cognitifs et du Langage in Paris, has developed a framework he terms conceptual fields, which emphasizes contexts, relationships, and wholes in mathematics education. Where Piaget focused on cognitive development and the logical structure of tasks, Vergnaud has taken an epistemological approach (Vergnaud, 1982). He has synthesized psychogenesis and learning by applying cognitive developmental theories to the study of specific mathematics content. Mathematical knowledge is seen to emerge from working with problems. The word 'emerge' has special significance here; it indicates that students' concepts, models, and theories are shaped by situations and problems. Vergnaud envisages students' concepts as changing only in response to problems they are unable to solve. In this way, students come to accommodate their views and procedures to new relationships. Such constructions certainly do not occur spontaneously but develop over long periods of time. In this section, conceptual fields are defined, and examples are discussed to illuminate their third wave character.

Vergnaud (1983a) defined a conceptual field as "a set of problems and situations for the treatment of which concepts, procedures and representations of different but narrowly interconnected types are necessary" (p. 127). Important elements in conceptual fields include problems and situations, operations of thought, and symbolic representations (Vergnaud, 1982). A field is not described solely in terms of content; it is described as the interrelationships between problems and situations, and students' procedures and operations of thought in addressing them. A student's construction of symbolic representations, such as diagrams, algebra, graphs, equations and tables, is integrally related to situations and operations of thought.

Additive structures are one example of a conceptual field. They incorporate problems, operations of thinking, and symbolic representations relating to measurement, addition, subtraction, time transformations, comparison relationships, displacement and abscissa on an axis, and natural and directed numbers (Vergnaud, 1981). Another conceptual field is multiplicative structures, involving problems, operations of thought, and symbolic representations of multiplication, division, fractions, ratio, proportion, linear function, similarity, vector space, and dimensional analysis (Vergnaud, 1983a). These two fields are not mutually exclusive. Developing understanding of multiplicative structures requires some reliance on relationships within the field of additive structures. Also, there are other fields that to some extent intersect additive and multiplicative structures, yet cover diverse situations and levels of operational thinking. Examples of these include spatial measures, dynamics, and classes, classifications and Boolean operations (Vergnaud, 1982).

Conceptual fields are systems that involve integrative ways of looking at the learning of mathematics. On its own, any given problem will not involve all the properties of a concept. The concept of addition, for example, is shown in the following situations to involve complex operations of thought that vary between situations, progressive understanding that students build over a long period of time, and relationships for which the set of natural numbers is inadequate:

Situation 1 — There are four boys and seven girls around the table. How many children are there?

Situation 2 — John just spent \$4. He now has \$7 in his pocket. How much did he have before?

Situation 3 — Robert played two games of marbles. In the first game, he lost four marbles. He then played a second game. In total, he now has won seven marbles. What happened in the second game? (Vergnaud, 1981)

The first situation exemplifies a measure-measure-measure relationship, in which the measure of children is a composite of the more elementary measures of boys and girls. The second situation illustrates a different relationship, one involving measure-transformation-measure. The "spending" transformation gives a temporal aspect to this situation, which also distinguishes it from the first, a static relationship. The third situation is an example of a transformation-transformation-transformation relationship. Robert's overall winning transformation with seven marbles is a composition of two transformations, only one of which is given. Vergnaud (1981) pointed out that Situation 2 is generally solved by students one or two years older than those in Situation 1, and Situation 3 is not solved by about 75 percent of 11-year-old students. Furthermore, the transformations in the second and third situations are inadequately represented by natural numbers, nor are these situations adequately represented by equations in N. The use of natural numbers is appropriate in the first situation for measures of discrete sets. However, in the second and third situations, transformations should be represented by directed numbers. But students usually work with situations similar to the second well before they learn of directed numbers; these tend to be taught as a separate topic at a later stage and in a manner that highlights abstract mathematical properties rather than building from problem bases. It is not surprising, therefore, that the discrepancies between the structure of problems students meet and the mathematical concepts they are taught, mean that much of the learning of mathematics is carried on at an instrumental level.

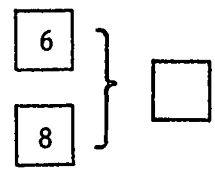
The building of concepts, in Vergnaud's view (1983b), will most effectively occur in settings in which students confront with integrity a range of problems over time. Integrity refers to students working on problems that have not been so condensed in their different relationships that they provide little opportunity for building operational knowledge. Such knowledge requires attention to relationships that remain the same over broad sets of problems. Vergnaud (1983b) referred to these relationships as <u>relational invariants</u> and notes that they are the very core of operational knowledge. He identified broad categories of relationships within conceptual fields, such as addition and subtraction problems and situations.

Vergnaud (1981) identified the main categories of relationships in addition and subtraction problems involving time: (a) composition of two measures, (b) a static relationship linking two measures, (c) composition of two transformations, (d) a transformation linking two static relationships, and (e) composition of two static relationships. Each category is described below in structural terms, an example is cited and a diagram used to represent the relationships. All too frequently in school mathematics equations derived from procedural knowledge are accepted as adequate expressions of thought, although they do not reveal underlying relationships in a situation. This point is elaborated later in the paper when distinctions are made between relational and numerical calculus.

<u>Category 1</u>: <u>Composition of two measures.</u>

This refers to situations with a static relationship in which two measures are combined, under addition, into a third measure. The vertical format of the measures in the diagram is meant to convey their static relationship.

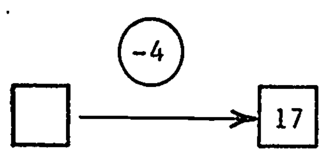
Problem: Peter has six marbles in his right-hand pocket and eight marbles in his left-hand pocket. How many marbles does he have altogether?




Category 2: Transformation linking two measures.

This class of situation is identified by a state-transformation-state arrangement.

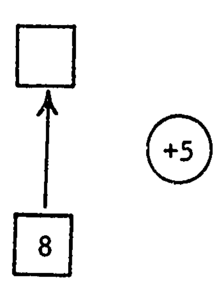
Problem: Peter had 17 marbles after playing. He had lost 4 marbles. How may marbles did he have to start with?




Category 3: A static relationship linking two measures.

This category differs from Category 2 in the form of the relationship. Where Category 2 refers to dynamic relationships, this classification is distinguished by the static nature of the relationship. In the diagram, the vertical arrow is meant to symbolize the static relationship.

Problem: Peter has 8 marbles. He has 5 more than John. How many does John have?



**Category 4:** Composition of two transformations.

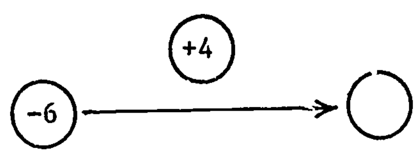
In this category, two transformations are viewed as equivalent to a third transformation.

Problem: Peter won 6 marbles in the morning. He lost 9 marbles in the afternoon. What happened overall?

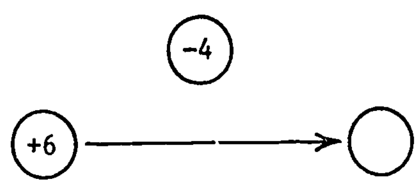

**Category 5:** Transformation linking two static relationships.

This class of problem involves static relationship-transformation-static relationships structures.

Problem: Peter owed Henry 6 marbles. He gave him 4 marbles. How many marbles does he still owe Henry?




or

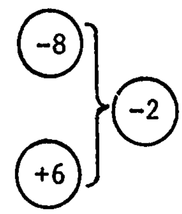



The first diagram might be interpreted as representing what is owed from Peter's point of view and the latter Henry's view.

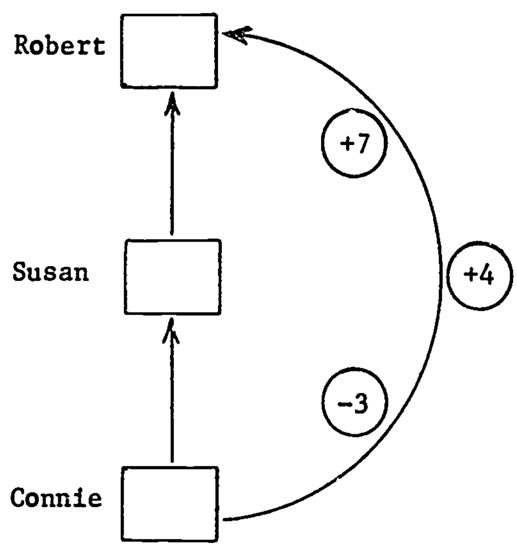
**Category 6:** Composition of two static relationships.

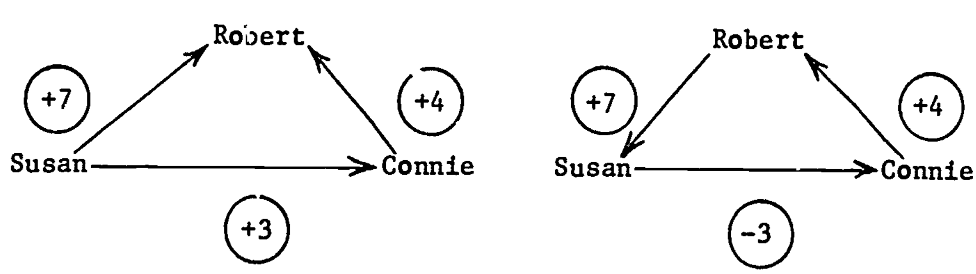
In this class, two static relationships are combined to produce a third relationship. Both of the following situations are examples of this structure.

Situation A: Peter owes 8 marbles to Henry, but Henry owes 6 to Peter. So Peter owes 2 marbles to Henry.




Situation B: Robert has 7 marbles more than Susan. Susan has 3 marbles fewer than Connie. Robert has 4 marbles more than Connie. This situation is represented by each of the following diagrams.




In these categories, the operation of addition and subtraction remain the same even though the type of relationship changes. Aspects of such change might involve static or dynamic relationships, presence of a unary positive or negative operation, or the presence of a part-whole relationship between the initial and final states. Knowing addition and subtraction goes beyond the mechanics of computation to recognition of invariant relationships over very different problems and situations. This recognition is unlikely to be articulated by students, particularly younger students, but will be observable as theorems in action over a broad range of contexts.

Theorems in action are operations which students use to solve or process problems and situations. For example, the Category 1 problem might involve operations such as counting all, counting on from the smaller quantity, or counting on from the larger quantity (cf. Carpenter, Moser, and Romberg, 1981). Where such operations are recognized as appropriate across a large variety of problems and situations, they become theorems in action. In particular, the discovery that the relation * is an additive relationship where $m(a * b) = m(a) + m(b) \ \forall \ a, b$ in many varied contexts is a theorem in action. Theorems in action, however, are not taught as such but are syntheses of operations students have in dealing with a broad range of contexts. Recognition of relational invariants and development of theorems in action within particular conceptual fields require a focus on relationships rather than procedures.

Vergnaud (year) employs the term relational calculus to describe students' operational knowledge which directs their theorems in action. He used numerical calculus in reference to the ordinary operations of addition, subtraction, multiplication and division. In the Category 2 problem described above, the relational calculus is the inverse of a negative transformation, −4, applied to the final state, 17. The numerical calculus is the addition in $17 + 4 = \square$ But recognizing the numerical calculus as leading to the solution, while widely accepted as demonstrating mathematical knowledge, in fact merely demonstrates the mechanism for arriving at the product. It misses the process in that it neither simulates the problem nor the operational thinking of the student. The problem would be simulated by either

$$\square - 4 = 17 \quad \text{or} \quad \square \xrightarrow{-4} 17$$

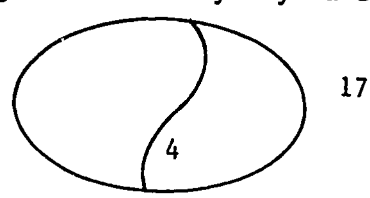
The emphasis in school mathematics on procedural knowledge and numerical calculus impedes a functional approach to mathematical symbolism. It separates signifiers or symbolic systems from the signified, failing to recognize their duality. Signifiers are made functional, however, when they assist students in the process of solutions which might otherwise not be found. Also, signifiers are made functional by enabling students to discriminate between situations, relationships and operations they might otherwise confuse. In this sense, the symbolic representation $17 + 4 = \square$ is a poor signifier of the situation, relationship and operation involved in the Category 2 problem. It misstates the situation, since young children associate addition with increase, and 17 + 4 does not convey the meaning of decrease in the context of the problem. The meaning of the statement $17 + 4 = \square$, if expressed in terms such as "I started from 17, then I added 4 and I got $\square$, is a unary, not a binary operation." In the procedure, +4 is an external operation on 17. Also, the equality sign is interpreted as producing the outcome and would therefore not express a symmetric relationship. It would be meaningless to write $\square = 4 + 17$. As noted above, the statements

$$\square - 4 = 17 \quad \text{and} \quad \square \xrightarrow{-4} 17$$

are likely to be functional as simulations of the problem, each representing the negative transformation involved.

Relationships in problems and situations are not equally well signified by various symbolic systems. A Euler-Venn diagram for example, is not capable of representing negative transformations, although this symbolic system is appropriate for representing a composition of measures as in the following problem:

> There are 17 children around the table for Joan's birthday. Four of them are girls. How may boys are there?




A Euler-Venn diagram would not be adequate, however, to represent problems involving relationships, such as: Tony has 17 marbles. He has 4 more than Robert. How many marbles does Robert have? Arrow diagrams adequately represent such relationships; for example:

$$\square \xrightarrow{-4} 17$$

Symbolic representations, as generally used, are vehicles for the efficient manipulation of data. Some, however, ar_ unable to represent problems that imply certain relationships. Some are unlikely to assist students to distinguish between representations of problems and representations of solutions. Also, some symbolic systems carry meanings that fall short of adequately conveying the mathematical relationships embedded in the situational context. The importance of symbolic representation to the construction and synthesis of different meanings is generally unrealized in school mathematics of the industrial wave.

## Implications for Assessment

One goal of any assessment procedure is to provide evidence about the level of understanding any student has with respect to a particular domain. If a conceptual field is a means of describing the interrelationship of ideas in a domain (what constitutes knowledge about that domain), then a new assessment perspective is called for. Specifically the assessment should reveal both the aspects of the domain the student has constructed and how the student reasons about those aspects and their relationships. Conceptual fields provide one framework for specifying knowledge structures for mathematics. The task then is to identify an assessment methodology which can be used to identify the extent of any student's knowledge about that domain. We believe that the methodology should be based on the notions about network models as a means of representing levels of knowledge, within a conceptual field, and the notions from cognitive psychology about how information is constructed.

## Network Models

A conceptual domain as described by Vergnaud (1981) may be considered as an example of a network model. In the past in most educational disciplines the concepts and skills upon which curricula and instructional procedures were based were considered as independent aspects to be mastered by students one at a time. Furthermore, in assessing understanding, student responses to each test item were considered to be independent of responses to other items.

Networks describe the interdependence of the aspects of a domain. Curricula, instruction, and assessment must reflect those interdependent relationships. Thus, assessment should begin with a set of exercises to be presented to students that reflect the important aspects of a conceptual field. Then from responses to those exercises a map of what the student knows about that domain would need to be constructed. However, the responses should not simply be a tally of the number of items that the student answered correctly. Instead the responses should be coded in terms of how the student reasons about the relationships.

## An Example

To illustrate how a conceptual field could be assessed we refer to the work of Carpenter and Moser (1983) who studied how students reason about addition and subtraction problems in a manner similar to that of Vergnaud (1981).

The domain includes learning to symbolically represent a variety of problem situations (often via word problems), operate on the symbols, and interpret the results. For example, to solve a typical addition and subtraction word problem, one first must understand its implied semantic meaning. Quantifying the element of the problem comes next (e.g., choosing a unit and counting how many). Then, the implied semantics of the problem must be expressed in the syntax of addition and subtraction. Next one must carry out the procedural (algorithmic) steps of adding and subtracting. Finally, the results of these operations must be expressed. Children bring to such problems well-developed counting procedures, some knowledge of numbers, and some understanding of physical operations, such as "joining" and "separating," on sets of objects.

Not all word problems involving addition and subtraction have the same semantic structure. In fact, most current work uses four broad classes of addition and subtraction problems: Change, Combine, Compare, and Equalize (Carpenter & Moser, 1983). There are two basic types of Change problems, both of which involve action. In Change-Join problems, there is an initial quantity and a direct or implied action that causes an increase in that quantity. For Change-Separate problems, a subset is removed from a given set. In both classes of problems, the change occurs over time. Within both the Join and Separate classes, there are three distinct types of problems depending on which quantity is unknown (see Table 1). Both Combine and Compare problems involve static relationships for which there is no action. Combine problems involve the relationship existing among a particular set and its two, disjoint subsets. Two problem types exist: the two subsets are given and one is asked to find the size of their union, or one of the subsets and the union are given and the solver is asked to find the size of the other subset. Compare problems involve the comparison of two distinct, disjoint sets. Because one set is compared to the other, it is possible to label one set the referent set and the other the compared set. The third entity in these problems is the difference, or the amount by which the larger set exceeds the other. In this class of problems, any one of the three entities could be the unknown--the difference, the referent set, or the compared set. The larger set can be either the referent set or the compared set. Thus, there exist six different types of Compare problems.

The final class of problems, Equalize problems, are a hybrid of Compare and Change problems. There is the same sort of action as found in the Change problems, but it is based on the comparison of two disjoint sets. The question is posed, "What could be done to one of the sets to make it equal to the other?" If the action to be performed is on the smaller of the two sets, then it becomes an Equalize-Join problem. On the other hand, if the action to be performed is on the larger set, then an Equalize-Separate problem results. As with Compare problems, the unknown can be varied to produce three distinct Equalize problems of each type.

To build the connection between semantic forms and relevant symbolism, one form is usually used as a model to introduce the symbolism. Given that there are many semantic forms for which the same symbolic sentence is appropriate, the pedagogical problem is how to relate the symbolism to all the semantic problems. Traditionally, the symbolism has been taught independently of word problems; that is, the symbolic procedures were taught, and some word problems were assigned so that students could apply their symbolic procedures. No serious consideration was given to the semantic structure of the problems. In fact, it is now clear that in many texts only a few of the semantic forms are ever included (see DeCorte, Verschaffel, Janssens & Joillet, 1984). It is no

TABLE 1
Semantic Classification of Word Problems
(Carpenter & Moser, 1983)

| Join | Separate |
| --- | --- |
| **Change** | |
| 1. Connie had 5 marbles. Jim gave her 8 more marbles. How many marbles does Connie have altogther? | 2. Connie had 13 marbles. She gave 5 marbles to Jim. How many marbles does she have left? |
| 3. Connie has 5 marbles. How many more marbles does she need to have 13 marbles altogether? | 4. Connie had 13 marbles. She gave some to Jim. Now she has 8 marbles left. How many marbles did Connie give to Jim? |
| 5. Connie had some marbles. Jim gave her 5 more marbles. Now she has 13 marbles. How many marbles did Connie have to start with? | 6. Connie had some marbles. gave 5 to Jim. Now she has 8 marbles left. How many marbles did Connie have to start with? |
| **Combine** | |
| 7. Connie has 5 red marbles and 8 blue marbles. How many marbles does she have? | 8. Connie has 13 marbles. Five are red and the rest are blue. How many blue marbles does Connie have? |
| **Compare** | |
| 9. Connie has 13 marbles. Jim has 5 marbles. How many more marbles does Connie have than Jim? | 10. Connie has 13 marbles. Jim has 5 marbles. How many fewer marbles does Jim have than Connie? |
| 11. Jim has 5 marbles. Connie has 8 more than Jim. How many marbles does Connie have? | 12. Jim has five marbles. He has 8 fewer marbles than Connie. How many marbles does Connie have? |
| 13. Connie has 13 marbles. She has 5 more marbles than Jim. How many marbles does Jim have? | 14. Connie has 13 marbles. Jim has 5 fewer marbles than Connie. How many marbles does Jim have? |
| **Equalize** | |
| 15. Connie has 13 marbles. Jim has 5 marbles. How many marbles does Jim have to win to have as many marbles as Connie? | 16. Connie has 13 marbles. Jim has 5 marbles. How many marbles does Connie have to lose to have as many marbles as Jim? |
| 17. Jim has 5 marbles. If he wins 8 marbles, he will have the same number of marbles as Connie. How many marbles does Connie have? | 18. Jim has five marbles. If Connie loses 8 marbles, she will have the same number of marbles as Jim. How many marbles does Connie have? |
| 19. Connie has 13 marbles. If Jim wins 5 marbles, he will have the same number of marbles as Connie. How many marbles does Jim have? | 20. Connie has 13 marbles. If she loses 5 marbles she will have the same number of marbles as Jim. How many marbles does Jim have? |

surprise, then, that for different types of problems students have found little connection between the problems and the symbolic procedures they had been taught (e.g., Vergnaud, 1982).

To assess a child's understanding of addition and subtraction, Carpenter & Moser (1983) administered six problem types (tasks) given under six conditions. The six types included two problems solvable by addition of the two given numbers and four problems solvable by subtraction of the two given numbers. The types differed in terms of their semantic structure. The semantic characterization for these six problem types is detailed in Carpenter and Moser (1983).

Table 2 presents representative problems. The six semantic problem types used were presented under six conditions, although

TABLE 2
Problem Types

| Task | Sample Problem |
|---|---|
| 1. Change/Join (Addition) | Pam had 3 shells. Her brother gave her 6 more shells. How many shells did Pam have altogether? |
| 2. Change/Separate (Subtraction) | Jenny had 7 erasers. She gave 5 erasers to Ben. How many erasers did Jenny have left? |
| 3. Combine/Part Unknown (Subtraction) | There are 5 fish in a bowl. 3 are striped and the rest are spotted. How many spotted fish are in the bowl? |
| 4. Combine/Whole Unknown (Addition) | Matt has 2 baseball cards. He also has 4 football cards. How many cards does Matt have altogether? |
| 5. Compare (Subtraction) | Angie has 4 lady bugs. Her brother Todd has 7 lady bugs. How many more lady bugs does Todd have than Angie? |
| 6. Change/Join, Change set Unknown (Subtraction) | Gene has 5 marshmallows. How many more marshmallows does he have to put with them so he has 8 marshmallows altogether? |

not all children responded to all conditions. Four conditions resulted from crossing smaller number (SN) problems and larger number (LN) problems with presence and absence of manipulative materials. The last two conditions involved two-digit numbers. In one set no regrouping (borrowing or carrying) was required to determine a difference or sum when a computational algorithm was used. In the second subdomain regrouping was required.

Trained interviewers have administered the tasks to children in several studies. The interviewers coded the responses for each child. (See Martin & Moser, 1980, for details of interviewer-training procedures and reliability.)

Children use a variety of strategies to solve the variety of additional subtraction word problems. For addition and subtraction three basic levels of operating have been identified: strategies based on direct modeling with fingers or physical objects, strategies based on the use of counting sequences, and strategies based on recalled number facts. For example, in addition problems, the most basic strategy is "Counting All With Models." Here physical objects or fingers are used to represent each of the addends, and then the union of the two sets is counted (see Carpenter & Moser, 1983).

From such a carefully constructed set of tasks it has been possible to construct a map of what a child knows about a domain at a point in time. Also, as was done by Carpenter & Moser (1983), by repeatedly administering the set of tasks one can portray changes in strategies used over time. Finally, although this example amply demonstrates the power of this assessment for understanding what a particular child knows and how he/she reasons, is the same strategy appropriate for monitoring group performance? The answer to this important question is yes. For example, Romberg & Collis (1987) used the tasks and coding procedures in a cross-sectional study to compare groups of children. Data were aggregated by class and cognitive level. Thus, both within and between group comparisons can be provided.

## Conclusions

The characteristics of the industrial wave, principally fragmentation, analysis, and mechanism, continue to permeate approaches to school mathematics. They underlie a manufacturing basis that objectifies school mathematics for supposed efficient and effective delivery to student consumers of the product. However, insufficient attention has been directed to mathematics as a social development, as a human enterprise in which student construction and creativity are valued. Outcomes of significance in this latter orientation are process rather than product, problem posing more than problem solving, questioning as well as responding, skills built within the context of problems, and reflective and operational thinking with less procedural thinking.

Vergnaud's development of conceptual fields offers possibilities for students of school mathematics in the postindustrial wave. It is problem- and situation-based. From this base, symbolic systems and the contextual meanings they signify are viewed as a duality. Operational thinking, at the core of which lies recognition of relational invariants, links problems and situations to symbolic representations and solution paths. As students distinguish between classes of problems they seem to employ theorems in action, syntheses of operations they have constructed and appropriated.

Vergnaud's approach challenges the very fundamentals of school mathematics that have so characterized it in the industrial wave. Conceptual fields as an approach makes clear the limitations of dominant industrial wave thinking and provides possibilities for working at school mathematics in ways that stress content and holism, are based on synthesis, and acknowledge the problematic nature of knowledge.

## References


Anyon, J. (1981). Social class and school knowledge. Curriculum Inquiry, 11(1), 3-42.

Bourbaki, N. (1968). Elements of mathematics: Theory of sets. Paris: Hermann.

Carpenter, T. P., & Moser, J. M. (1983). The acquisition of addition and subtraction concepts. In R. Lesh & M. Landau (Eds.), Acquisition of mathematics concepts and processes, (pp. 7-44). New York: Academic Press.

Carpenter, T. P., Moser, J. M., & Romberg, T. A. (1981). Addition and subtraction: A cognitive perspective. Hillsdale, NJ: Lawrence Erlbaum.

DeCorte, E., Verschaffel, L., Janssens, V., & Joillet, L. (1984). Teaching word problems in the first grade: A confrontation of educational practice with results of recent research. Leuven, Belgium: University of Leuven.

Heidegger, M. (1977). The question concerning technology and other essays. New York: Harper & Row.

Martin, C., & Moser, J. M. (1980). Interview training procedure. (Working Paper No. 300). Madison: Wisconsin Research and Development Center for Individualized Schooling.

Papy, G. (1982, July). The need for an epistemological understanding of mathematics. In A. Vermandel (Ed.), Proceedings of the Sixth International Conference for the Psychology of Mathematical Education, 2, 36-47. Antwerp.

Popkewitz, T. S., Tabachnick, B. R., & Wehlage, G. (1982). The myth of educational reform. Madison, WI: The University of Wisconsin Press.

Romberg, T. A. (1984, August). Curricular reform in school mathematics: Past difficulties, future possibilities. Paper prepared for the Fifth International Congress on Mathematical Education (ICME 5). Adelaide, Australia.

Romberg, T. A., & Collis, K. F. (1987). Different ways children learn to add and subtract. Journal for Research in Mathematics Education. Monograph 2.

Toffler, A. (1980). The third wave. New York: William Morrow.

Vergnaud, G. (1981). A classification of cognitive tasks and operations of thought involved in addition and subtraction problems. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), Addition and subtraction: A cognitive perspective (pp. 39-66). Hillsdale, NJ: Lawrence Erlbaum.

Vergnaud, G. (1982). Cognitive and developmental psychology and research in mathematics education: Some theoretical and methodological issues. For the Learning of Mathematics, 3(2), 31-41.

Vergnaud, G. (1983). Multiplicative structures. In R. Lesh & M. Landau (Eds.), Acquisition of mathematics concepts and processes (pp. 127-174). New York: Academic Press. (a)

Vergnaud, G. (1983, Sept., Oct.) Why is an epistemological perspective a necessity for research in mathematics education? In J. C. Bergeron & N. Herscovics (Eds.), Proceedings of the Fifth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, 1, 2-20. Montreal. (b)

## Chapter 21

# ANOTHER LOOK AT ASSESSMENT: A REACTION TO CHAPTERS 17-20

Norman L. Webb

Mathematics achievement and its assessment are the central topics of chapters 17-20. Together, the four chapters argue for new procedures for assessing mathematics achievement and indicate what needs to be considered in the development of such tests. The general argument posed by the four chapters goes like this: A new age is upon us, resulting in a need for reform in the mathematics curriculum and, consequently, for reform of the procedures we use to assess mathematics achievement. The focus of mathematics education and our understanding of the mental structure of knowledge are changing. Assessment procedures also must change to better reflect our current understanding about how knowledge is constructed and the mathematics that students should know. In addition, continued use of current assessment procedures will inhibit needed reform in the mathematics curriculum.

Assessment of mathematics achievement generally refers to some measure of a student's or group's command of mathematics. Three fundamental factors must be considered in decision making about the appropriateness of a particular measure. These factors are:

1. What is the purpose for the assessment?

2. Does the assessment procedure measure what it is intended to measure?

3. Is the assessment procedure reliable?

These three factors are not based on any assumptions related to historical or economic era, content, or school of psychology but are fundamental factors that must be considered for deciding the appropriateness of any measure.

Using a balance scale to measure the weight of a block of wood would provide very little useful information if the purpose of the measuring was to determine whether the block would fit into a box. The appropriateness of a measure and the procedure used to obtain it can only be judged relative to the purpose for obtaining the measure. This is true for all measurement procedures, including testing. As noted by Cronbach (1970), "Tests must be selected for the purpose and situation for which they are to be used" (p. 115).

A procedure also must provide an appropriate measure of that which is to be assessed; that is, the procedure must be valid.

This is true for any form of measurement for any purpose. A ruler graduated by inches is a valid measure for estimating the lengths of several new pencils to the nearest inch, but it is not a valid measure of the differences in pencils' lengths, which will vary by only a small fraction of an inch.

The third factor is based on an assumption that all measures have errors associated with them. The error of measurement must be small enough so that the measure consistently provides the information needed. For example, if a ruler used to measure a pencil were made of string that would stretch when pressure was applied, the measures would be inconsistent or unreliable.

Other considerations related to most measurement situations include the unit of measure, the precision of measure, the frequency of measure, the sampling for measure, and the generalizability of measure. However, these can all be subsumed into one or more of the three basic factors--purpose, validity, and reliability.

These fundamental factors comprise the model used in reviewing each chapter. The chapter will be discussed individually. Then a brief summary of content will be followed by a reaction to the chapter's main issues. I conclude my comments with observations about the content in the four chapters that is related to monitoring school mathematics.

Chapter 17 makes the case that new assessment procedures are needed to monitor educational reform. It can be assumed that such reform will produce new or different educational results; for example, the reformed curriculum emphasizes higher order thinking skills while it deemphasizes mastery of algorithms. If policy decisions are to be relevant to national reform, assessment procedures must be sensitive to the goals and purposes of the reforming curriculum.

Romberg, in chapter 17, comments on current tests. "While these tests have been useful for some purposes and undoubtedly will continue to be used, they are products of an earlier era in educational thought. . . . Today we ought to be able to develop better indices of achievement." In selecting or developing assessment procedures, it is important that the purpose be clearly understood and that a procedure, test, or other instrument be evaluated on its appropriateness to that purpose. The major reason to accept or reject the use of standardized tests is not so much based on the tests' historical roots as it is on the purpose for which the tests were designed. The knowledge of a test's historical roots, or the era of which it is a product, is useful in explaining why the procedure was used and in offering a deeper understanding of its purpose. A historical analysis also contributes to an understanding of the context in which a procedure was used, which is helpful in identifying factors important to test development.

A historical analysis reveals that norm-referenced tests were "based on the psychology of individual differences rather than upon the psychology of learning" (Tyler & Wolf, 1974) and were the product of an era in which the prevalent societal view espoused the survival of the fittest, a view that encouraged the selection of the nation's best and brightest to be officers in the army, to attend college, and to work in select professions. This is interesting, but a judgment about whether a norm-referenced test is an appropriate tool for assessing educational reform must be based on considerations of the purpose of the assessment and how well a norm-referenced test meets that purpose. If there is a need to order individuals on a single trait, to use items that are assumed to be equivalent, or to predict future achievement or success, then norm-referenced tests may be appropriate.

The importance of knowing the purpose for assessment is also true of other test forms. Profile achievement tests have been designed to evaluate educational programs or to assess a population's command of knowledge in a content area, such as mathematics. For example, the purposes of some of the programs listed by Romberg to illustrate the use of profile testing follow.

- The National Longitudinal Study of Mathematical Abilities (NLSMA) was organized by the School Mathematics Study Group (SMSG) as a long-term study of the effects on students of various kinds of mathematics programs. (Romberg & Wilson, 1969, p. vii)

- Three main purposes for the Second International Mathematics Study can be summarized as follows:

    1. to investigate the ways in which mathematics is taught;

    2. to describe student attainment in terms of both attitude and achievement; and

    3. to relate these outcome variables to the curriculum studied and the way it was taught.

    (Crosswhite et al., 1986, p. 3)

- The purpose of the First International Mathematics Study is ". . . to evaluate uniformly the educational practices (including 'standards') of different countries" (Husén, 1969, p. 338).

- The purpose of the National Assessment of Educational Progress is to gather information which will help answer the question, "How much good is the expenditure [for the testing] doing, in terms of what young Americans know and can do?" (Finley, 1974, pp. 95-96)

- The National Assessment has been designed to sample the things which children and youth are expected to learn in school, and to find out what proportion of our people are learning these things (Tyler, 1974, p. 94).

- The purposes [of the Wisconsin statewide assessment] are to provide:

  . measures of student performance in selected academic areas;

  . comparisons of student performance to a national average in mathematics, reading, and language;

  . descriptions of changes in student performance over time; and

  . technical assistance in the area of testing and evaluation. (Wisconsin Department of Public Instruction, 1986, p. 1)

The purpose, then, for using profile tests is to evaluate achievement or the effects of programs over a large group of students. In evaluating the programs, some scheme is needed to ensure that the range in content--both that which is included in the programs being evaluated and that which students being tested have taken--is represented. The content-by-behavior scheme was judged to be appropriate for use in NLSMA and the international studies mentioned above, but it was not appropriate for use in the Wisconsin State Assessment. In some cases, where the content-by-behavior matrix is not as appropriate, the results were reported by curriculum areas or objectives, while in other cases, such as for the National Assessment, the results were reported by item. "The results are reported in terms of percent of each population group that was able to perform the exercise. These exercises show the public both what our children are learning and how many are learning each thing" (Tyler, 1974, p. 94).

The content-by-behavior matrices have been used to fit the purpose of the study and to help in providing content and curriculum validity. Considering the historical context, using content-by-behavior matrices as a framework for constructing assessment instruments made sense; curriculum and instruction were greatly influenced by such matrices, based on the work of Tyler (1970) and others. Whether the use of the content-by-behavior matrix would be appropriate to monitor reform depends on how the reform curriculum will be structured, whether or not concurrent assessment is needed of students using curriculum materials based on content-by-behavior matrices, and whether such a model meets the assessment purpose.

The behavior dimension of the matrix frequently has been based on Bloom's Taxonomy (1956). As Romberg noted, this taxonomy fails to reflect current psychological thinking. Nonetheless, the matrix model may serve a purpose in reform assessment if its behavior dimension is replaced by a more contemporary notion of psychology applied to learning. Problems with the basic structure of the matrix model for guiding instruction and assessment are that the dimensions are considered orthonogal and that all of the cells created should be filled. A decision to use the matrix model should be based on the assessment's purpose and a judgment about the model's validity.

Similarly, the use of objective-referenced tests for assessment must be evaluated on the basis of the purpose for the assessment. Even though objective-referenced tests are related to criterion-referenced measurements, objective-referenced measurements are interpreted by referencing the specific behavioral objective(s) for which a test item was written (Sanders & Murray, 1976). As Swezey (1981) noted, "These [objective-referenced] test items are considered to be operational definitions of the behavioral objectives" (p. 4). Objective-referenced tests can be used in many different ways, including providing measurement of individual performance and evaluating an instructional program given to a group. If an acceptable criterion is associated with the objective-referenced test, it becomes a criterion-referenced test.

Romberg discusses in chapter 17 some of the major drawbacks to objective-referenced tests, such as the meaning of aggregated results across objectives, the assumption of independence of items, and the cost. To develop an appropriate objective-referenced test is both costly and time consuming. However, the appropriateness of this form of testing for monitoring school mathematics must be judged on the assessment's purpose and on the validity of the test in meeting this purpose. If the performance of specific tasks are part of the reform curriculum and there is a need to establish absolute measures of performance, then some form or related model of objective-referenced instrument may be appropriate.

An issue arises here regarding measurement. A measure is a quantification of some object, entity, or behavior. It is an abstraction. Any measure will be inadequate to describe the object in its entirety. When a student responds to a test, task, exercise, question, situation, or any other stimulus, and his or her performance, response, answer, description, or writing is recorded and used as a measure, behaviors are involved. A weakness in the construction of objective-referenced tests has been that traditional test items measured one small part of all behaviors related to the objective. It is assumed that, if items are randomly chosen from a pool, the aggregated score of the items will be a measure of the objective. In practice, objectives and items are articulated very specifically. However, this is not so much the fault of the procedure as it is a weakness in the manner in which the procedure has been put into practice. If the reform is guided in any way by goals, outcomes, and/or objectives--which I suspect it will be--then some form of objective-referenced testing will be appropriate in the assessment of the reform.

In summary, Romberg argues that as reform in mathematics education takes hold, new indices of achievement will be needed. New indices will be needed because what students will know and how students will gain this knowledge will be different from previous eras. But if the true impact of the reform is to be recorded, some comparative evidence will be needed between what students know as the result of reform versus what students know as a result of education from earlier periods. The case for the effectiveness of a reform is strengthened if the new knowledge students possess is shown to be the direct result of the reform movement. For example, there may be a need to show that the reform did, or did not, depending on the purpose of the reform, produce an elite class of students who benefit disproportionately from other groups and achieve much higher than all other students. Some form of standardized norm-referenced test would be useful in doing this. There may be a need to judge the impact of the reform material in comparison to the impact of the more traditional curriculum materials. If the current materials have been developed using a content-by-behavior matrix, then a form of a profile achievement test may have some validity with respect to the older curriculum. If the reform is guided by mandates that all students are to have certain knowledge, then some form of criterion-referenced test will be needed. The forms of assessment needed to measure the impact of reform must not be discarded simply because they were developed in a previous era, but must be judged on their own merits and how appropriate they are to the purpose for assessment and the expectations for the reform.

Now let us consider four assumptions Romberg (in chapter 17) uses as evidence of the need for new assessment procedures; we will evaluate each assumption using the general criteria for selecting an assessment procedure: Is the procedure alid and reliable, and does it provide the information needed to meet the purpose?

Assumption 1. The character of American schooling will be significantly altered in the new age. This suggests that the outcomes of schooling probably will need to be different to meet the demands of the new age. If we are to achieve outcomes different from those of the current system, we will need to teach new content in a different way. If reform is to be monitored adequately, assessment instruments must be sensitive enough to denote changes in student outcomes that result from a change in instruction designed to better prepare students for the new age. If the purpose for assessment is to determine how well students are learning the new content or have changed in light of new instruction, then the assessment procedures must include new-age content validity and some form of content reference. If the purpose for the assessment is to determine whether students are prepared for the new age and will perform well in light of its demands, then the assessment procedures will need to have predictive validity. If the purpose for the assessment is to determine whether the students are functioning at some predetermined level that has been deemed necessary for the new age, then the assessment procedures will need to have criterion validity.

Assumption 2. Mathematics instruction must focus on thinking skills. Thus, if the reform has as one of its goals the learning of higher order thinking, monitoring the reform will require assessment of this type of thinking and the ability to detect changes in the use of higher order thinking by different groups. Again this issue is one of assessment content and procedural validity; the measurement used must be sensitive to higher order thinking and able to detect changes in its use. If the nature of higher order thinking prohibits it from being precisely defined, as alluded to by the quote from Resnick (on pp. 146-147), then some of the more structured forms of measurement, such as objective-referenced and profile-achievement testing, may be inappropriate and may, in fact, discourage higher order thinking. But this is a question of validity and of the use of procedures that will allow higher order thinking to be observed and measured.

Assumption 3. Higher order skills are not to be learned after the mastery of other skills. In short, instruction of higher order thinking skills needs to be a part of the curriculum and its assessment at all age levels. This issue relates to the purpose of monitoring: What age levels are to be included, and how diagnostic or descriptive is the monitoring to be? Is the purpose to assess the higher order thinking of students at particular times during their schooling experience? Or is the purpose to assess how schools facilitate the development of higher order thinking throughout a student's school career, assuming this is one of the goals or possible outcomes of the reform? Assumption 3 implies that, if development is an issue, then monitoring must be conducted at all ages. This assumption has implications for defining the purposes of the monitoring and using procedures that are sensitive to providing the needed information for the purpose. This assumption is intrinsically related to instruction. Is the monitoring to look at outcomes or instruction? Is it reasonable to assume the existence of a hierarchy so that a student's inability to do a routine task does not imply his or her inability to do higher order thinking? This assumption is related to generalizability with regard to skills and their relations to each other; evidence of one form does not deny or confirm another form. This implies the need for a rethinking of prerequisite knowledge and skills and the structure of reform curriculum. It also means, for example, that what is currently considered an eighth-grade level of thinking may need to be considered a skill for students at all levels. A corollary issue is the question of how higher order thinking manifests itself. How do you know when higher order thinking is being used or has been used? How do higher order thinking skills relate to mathematics achievement? Is doing more advanced mathematics doing higher order thinking? Or, is mathematical higher order thinking an independent educational area in which students are expected to achieve? Is this related to creating mathematics and solving problems?

Assumption 4. Current approaches to achievement testing inhibit needed reform. This assumption is based on two issues. One is the issue of validity and that current tests are not aligned with the existing or reform goals for education. The second is the issue of the degree of influence that tests have on the curriculum and the advancement or retardation of any reform. The three references Romberg quotes in chapter 17 (McLean, 1982; Hilton, 1981; and Resnick, 1987) all raise concerns about current tests not reflecting what mathematics is being taught or what mathematics should be taught. They raise a good point and suggest that any test used should be aligned with the desired outcomes. If higher order thinking is an intended outcome, then the tests being used to evaluate students or a program should include some measure of higher order thinking. If the tests do not, then the validity of the evaluation procedures is in question. If a test is to measure reform, then the test needs to be aligned with the intended outcomes of the reform movement.

To make the assumption that the approach to testing inhibits reform is more difficult to do than to make the assumption that the content being tested is what has the real influence. Depending on what the reform curriculum is, existing approaches to testing can be valid for measuring reform. What is more important is that any one approach to testing will be insufficient to measure the depth and breadth of the reform curriculum. Standardized tests can measure some levels of cognitive functioning of higher order training programs but will not, necessarily, be sensitive to all levels of cognitive functioning. Some form of open ended question or interview will probably be needed. Rather than discarding tests outright because of the approach taken, we need to judge tests individually on how well they are aligned with the intended outcomes of the curriculum.

The second issue related to Assumption 4 is the degree of influence tests have on the curriculum and reform. The argument here is that teachers and school administrators make curriculum decisions based on the tests being used. That is, the curriculum is test driven, or, to use the term coined by Popham, Cruse, Rankin, Sandifer, and Williams (1985), instruction is measurement driven. Hilton (p. 148, chapter 17) is quoted as saying, "[Tests] loom so large that they distort the teaching curriculum and the teacher's natural style." This overstates the case. The influence of tests on the curriculum will vary according to the payoff placed on the test results. High-stake tests, such as those needed to graduate from high school, gain admission into college, or receive a license for a profession, will have more influence on what is taught than tests used to group students for learning. Tests that are used to evaluate and rate the effectiveness of teachers will have more influence on what is taught than tests used to evaluate programs.

The test or approach to testing achievement has the potential of inhibiting reform to the degree that results from the tests are used to make decisions with high payoffs. Currently, the use of high-stake testing is localized and varies from district to district and from state to state. There is some question about exactly how influential tests are on what is taught. Stiggins and Bridgeford (1985) noted that their findings from administrating an extensive questionnaire to teachers in a range of grades, subjects, and school districts showed that only from 8 to 19 percent of the eighth- and eleventh-grade teachers reported using published tests for any of the five purposes (diagnosing, grouping, grading, evaluating, and reporting). The form of assessment used by the highest percentage was teacher-made objective tests. In another study, the scores of students whose teachers taught to specific objectives on standardized tests did not seem to differ greatly from the scores of students whose teachers did not attend to the objectives (Mehrens & Phillips, 1986).

Tests are a part of the infrastructure of education and in that sense interact with the curriculum, instruction, and outcomes. The assumption that current approaches to achievement testing will inhibit needed reform, however, needs to be considered in context of how valid the individual test and its approach is to measuring the intended outcomes of the curriculum and how much weight is placed on the test results.

The four conclusions Romberg provides at the end of chapter 17 support reform and recognize the role of testing. His point that curriculum change must be accompanied by a concurrent change in evaluation, including assessment procedures, is an astute observation and is well taken. Tests are ingrained in our educational system, and as that system changes, the mode of testing needs to change as well.

Chapter 18 by Romberg and Zarinnia offers additional support for changing assessment means; their historical analysis takes into consideration the economic, social, and psychological environments. In a field whose beginnings are founded in this century, this chapter verges on a philosophical study of mathematics education which looks at the reasons, explanations, and meanings of certain happenings based on broader contexts. Such analysis, possible only if there is some history to a field, is an important factor in establishing the uniqueness of an area of study. The ability to draw meaning from a context that provides information about how a field of study was developed helps in understanding the dynamics of the field and offers insight to make predictions for the future. It must be noted, however, that predictions are no more than mere speculation, and there is no way to know what the world will be like 20 years from now or, much less, what kinds of mathematics people will be using. In planning for the future, and in making decisions about what should be done now, it is important to use all the information we have available, including considerations about how the world situation has effected educational trends in general and mathematics education in particular; about the most reasonable projections of what the world will be like in the near and extended future; and, based on this projected world view, about mathematics education and its assessment.

Education in this country has two fundamental purposes. It prepares students for the future by teaching the mathematics they will need for work and further education in 10 to 20 years; this is a utilitarian purpose. Education is also designed to transmit our culture from one generation to the next. For this purpose, educators must consider what has come before and what it means to be a part of a culture. In mathematics education, this means learning about what mathematics is, why it is important, and what tools it requires. It is important to keep these two purposes in mind as we review chapter 18.

Romberg and Zarinnia state their purpose as the consideration of the consequences of the emerging world view, called the Information Age, on assessment of students' knowledge of mathematics and their ability to use this knowledge creatively and routinely to solve a variety of problems encountered in life. Their argument holds that the nature, forms, purpose, and design of major models of assessment are dominated by the prevailing Old World views. They also argue that the "old" forms of assessment will impede the progress of reform. The discussion rests on their description of a dominant structure for creating achievement tests, the content-by-behavior matrices; the authors note that use of this structure relies on assumptions that a taxonomy exists, that the matrix is the product of a behaviorism tradition, that items selected to fit into this framework are frequently multiple choice, and that the psychometric characteristics of items used along with the matrix were selected based on assumptions used to select items for standardized tests. Romberg and Zarinnia express dissatisfaction with current testing because of the content-by-behavior matrix structure reflected in most of these forms. The problem with this structure, according to the authors, is that it reflects an engineering approach to education, it inhibits change in the curriculum, and tests developed according to its tenets misrepresent learning and knowledge. The chapter concludes with a description of mathematics education as it should be in view of existing and emerging psychological theories, epistemologies, and organizations of content; the authors offer suggestions about some forms of assessment that come close to reflecting the new view of the curriculum. A network approach is proposed as an alternative to the content-by-behavior matrix.

This chapter needs to be reviewed in light of its underlying assumptions and its purpose. It is intended to examine assessment from a new world view. It does not focus on the issue of monitoring, except to suggest that procedures used to monitor reform should reflect, in part, forms of assessment appropriate to the new world view.

There is no question that society is in the process of change, as suggested by Zarinnia and Romberg in chapter 2, Volume I, of this work. The level of productivity per individual has increased in industry and agriculture; fewer people are producing more goods--one indication of a prospering economy. With the development of technology and computers, employment in service and information jobs is on the rise. Such professions as consulting, nonexistent 20 years ago, have created new jobs. Change is the status quo. In fact, Zarinnia and Romberg may be understating the magnitude of social change in limiting their discussion to the transformation from an industrial age to an information age. Some observers have suggested that society is moving from the Modern Age into a whole new age of human history--a process which has occurred only three times in the last 2,000 years: the fall of the Roman Empire marked the first such transition; the Middle Ages was the second; and the evolution of the Middle Ages into the Modern Age was the third (Gust, 1986). The period we are in now has been labeled the post-Modern age.

The issue here, then, is not whether change is occurring; the real questions involve the ways in which the educational system will react to change in terms of the curriculum in general and mathematics in particular. The view expressed by Romberg and Zarinnia in chapter 18 is based on an underlying assumption that education prepares students to function in society. They examine the consequences of the emerging world view as it relates to students' "ability to use [their knowledge of mathematics] both creatively and routinely in solving the variety of problems encountered in the course of life" (p. 1, chapter 18). The authors take a utilitarian view of the role of education. An alternative view of the role of education sees it as the transmission of culture. This view holds that schools are to transmit to students the accumulation of knowledge to date. Students are prepared for work and further education only to the extent that work and education are seen as components of the culture. A response to a rapidly changing world based on this latter view of education would be very different from that which would emerge from the former. This thermostatic view suggests that "in a culture of high volatility and casual regard for its past such a responsibility [the conserving function of school] becomes the school's most essential service. The school stands as the only mass medium capable of putting forward the case for what is not happening in the culture" (Postman, 1979, pp. 21-22).

From a thermostatic point of view, curriculum reform in the face of volatile change would stress the nature of the content area, its history, its structure, and its place in society. The curriculum would not be taught as a series of skills in isolation, but as an integrated body of knowledge inherent to our society.

The thermostatic view of curriculum is presented here to suggest an alternative view to the utilitarian role of education that may offer some different directions for reform and, consequently, assessment. For example, the utilitarian view may suggest that statistics and probability are important topics for all students to master because these topics will be increasingly important in the work force and in describing our world. The thermostatic view would argue that it is uncertain what mathematics will be commonly used in 20 years; it is possible to guess, but no one can be sure. Statistics and probability are important mathematical topics which have evolved over time to describe and model chance events. The topics have important applications in the world today and students should know what role these topics play and how they relate to other mathematical topics. Response to change from a thermostatic view provides a firm foundation for students to build on. This approach would emphasize more the concepts and nature than procedures.

Chapter 18 talks about assessment of students' knowledge of mathematics as a consequence of the new world view. The prevailing view of the content-by-behavior matrix used to construct tests is that it is deficient because of the underlying theory it represents; because it can be used to separate things into distinct cells; because the classification of content is spurious; because it reflects behaviorism and scientific management, which misrepresent the thinking process; because its use trivializes learning and knowledge; and because it encourages the use of multiple-choice items that, by their very nature, must be independent.

The content-by-behavior matrix is described as providing the framework for profile achievement tests. As noted in the discussion of chapter 17, the appropriateness of using a content-by-behavior matrix as a framework for constructing tests is really a question of whether the matrix is valid for the intended purpose. If the curriculum was based on such a matrix, it is appropriate to use the matrix to guide the development of the assessment. Romberg and Zarinnia have noted that the content-by-behavior matrix has provided a powerful organization scheme for many assessment programs. The idea of a matrix, as advanced by Tyler (1970), was a guide for planning curriculum. In his rationale, he cautions against being too specific or too general; effort should be made to include a workable number of objectives, from 10 to 30. The ordering of categories came later, with the introduction of the behavior taxonomy. Romberg and Zarinnia's argument that the "intent of the content-by-behavior matrix is in every respect hierarchical" (chapter 18, p. 166) refers to a common use of the matrix but does not reflect the only way it can be used or how Tyler viewed its applicability in the early stages of its development.

The critical question is not so much what is wrong with the content-by-behavior matrix as a framework for designing assessments, but what is the best framework for the intended purpose. That the matrix may be associated with an "old world" view, in which an engineering approach to scientific management dominated, is not a sufficient reason to dismiss its use. That the application of the matrix was pushed in some cases to its extreme to partition objectives into very small cells, resulting in inattention to educational outcomes that required the combined applications of skills covered by all the objectives, does not indicate that the matrix has to be inappropriately used in the future. It is necessary to identify in some way that which is to be measured in light of the changing curriculum. Once the purpose for the assessment is adequately defined, a framework that will satisfy the purpose can be specified, selected, or derived.

Creation of knowledge is to be one of the main purposes of education in the future. In identifying a framework for assessing mathematics achievement, creation of knowledge serves a function similar to that which behavior serves for the content-by-behavior matrix. Another factor is that the "new view of science blends the linear and the circular; it emphasizes probability and stochastic processes" (chapter 18, p. 169). In their section "New Purpose: Managing Complexity," Romberg and Zarinnia have provided a very thorough conceptualization, well grounded in the literature and current recommendations, of the direction mathematics education is likely to take. They argue for an epistemological approach to mathematical education based on conceptual fields. Current trends in science and society are most congruent with the theoretical model, which is depicted by diagramming networks, as compared with the scientific model, which is depicted by forming matrices.

The appropriateness of a network model for structuring the assessment of outcomes from mathematics instruction must be considered in light of the purpose for the assessment and the validity of the network model in meeting this purpose. The network model, in theory, appears to have some validity to projected epistemological approaches to mathematics education. It can depict relationships among many different factors; it is flexible, so that factors and links can be added or deleted without affecting others in the model. This makes the network model more relevant to a constructive notion of knowledge formation than the matrix model, which assumes that behaviors cross all content areas and that it is more difficult to delete just one or two cells. The network model is very suitable for depicting the relation among situations, mathematical ideas, and possible representations.

Before the actual validity of the network model can be determined, it must be tested to determine whether it works in practice. Several questions must be answered before putting the model into use. First, in creating a network in which nodes are interlinked, what will the nodes represent and what will the links represent? Will the nodes be concepts, ideas, processes, or some combination of the three? Will the links represent the same type of relationships in the network, such as a subclassification of a broader category, or will the links represent different kinds of relationships, depending on which nodes are connected? In addition, the network model does not preclude falling into some of the same traps that plagued the content-by-behavior matrix. For example, what is the level of specificity needed to adequately develop an assessment framework? In using the network model, how refined do the nodes have to be? It is possible to become very specific, which could result in the reductionism that has occurred when very refined behavioral objectives were written on the basis of a matrix framework. The matrix provided a powerful organization scheme that depicted a plan for a large-scale assessment in a relatively small space. For example, the matrix used for NLSMA (Romberg & Wilson, 1969) fits on a single page. This matrix guided the use of a number of testing instruments and a variety of items. Do levels of networks exist so that one can depict the general approach to a large assessment? How will different content divisions, as we currently know them, be depicted? Will mathematics be divided into general areas or fields such as multiplicative field, additive field, geometric field, etc.? If so, will these fields be linked by a network, or will they be depicted disjointedly, each with its own network? Theoretically, the network approach is appealing and appears to be more reflective of current thinking on cognition, but many practical issues must be resolved before the approach has the functional power of the content-by-behavior matrix.

In chapter 19, Collis described an approach that is actually an intermediate model between the content-by-behavior matrix and the network model. The SOLO Taxonomy, in structure, is a hierarchy similar to Bloom's Taxonomy (1956). However, the SOLO Taxonomy is based on recent cognitive development theories and on Piaget's stages of cognitive development. In this sense, the model takes into account some of the new notions of knowledge. The taxonomy has been developed for use in evaluating students' responses. An added benefit to the system is that the taxonomy has been found to have validity for developing both an open format and a closed format of tasks. Because the approach has been tested, it is possible to specify the steps needed to analyze tasks in preparing an evaluation instrument. The system also has some applicability to analyzing the level of mathematics functioning required in particular professions.

The SOLO Taxonomy seems especially suited to the evaluation of individuals and to making instructional decisions regarding individuals. Concurrently, the system could be used to analyze curriculum and plans for teaching. The system is not as related to the constructive notion of knowledge, where the meaning of mathematics is drawn from the situation; it does not model as closely Resnick's description in chapter 17 of lower and higher order thinking skills that are not necessarily learned or experienced in a hierarchy. The superitem technique, which uses the closed format, may have possibilities for applications in other systems. What is important in developing assessment procedures for monitoring school mathematics is that the structure of the SOLO Taxonomy be considered in regard to the purposes of the assessment and its validity for assessment over groups of students.

Donovan, in chapter 20, described a "fundamental reappraisal" of the content of school mathematics based on Vergnaud's (1983) notion of conceptual fields. The approach is described in the context of the assumption that knowledge is socially constructed. This is an epistemological approach, different from the cognitive development theory espoused by Piaget or an approach that focuses on the logical structure of tasks. Students' concepts, models, and theories are shaped by situations and problems. The three important elements of conceptual fields are problems and situations, operations of thought, and symbolic representations. Examples of conceptual fields are additive structures, multiplicative structures, spatial measures, and dynamics.

Conceptual fields provide a new way of organizing content and of thinking about the assessment of mathematics, consistent with a constructive notion of the structure of knowledge that is applied to mathematics. Network models, as described above, appear to be directly applicable to this form of thinking about mathematics. Donovan illustrated the application of a conceptual field to an assessment of addition and subtraction (Carpenter & Moser, 1983). Using a carefully constructed set of tasks and interviews with individual students, it was possible to construct a map of what a child knows about the additive conceptual field. Donovan noted that a similar procedure was used by Romberg and Collis to collect data that were aggregated by class and cognitive level.

Progressing from conceptual fields to a well-developed plan for monitoring school mathematics is not a trivial matter. Major conceptual fields would need to be identified and defined; as Vergnaud (1983) noted, these fields would not be disjoint. Within each field, three major elements would have to be defined with enough specificity so that variations in situations as they relate to the conceptual field could be identified. For addition and subtraction, Carpenter and Moser defined semantic structure for six different types of addition and subtraction problems. For other fields, such as the multiplicative conceptual field, the number of types that could be identified and are applicable is unknown. The network model, as described in the discussion of chapter 18, may have the potential of providing a framework useful for describing and depicting conceptual fields. As is often the case, the process of specifying content or conceptual field to the degree necessary for assessment can make a real contribution toward advancing the use of conceptual fields in guiding instruction. The issue of the validity of assessment procedures based on conceptual fields to existing curriculum and their future evolutions must be resolved.

The major issue addressed in chapters 17-20 involves the need for some system to monitor school mathematics on a national level. Such a system is needed because the time is right. Rumblings of reform in mathematics education are in evidence in the number of standards being issued by blue ribbon committees and commissions; in the emergence of technology into everyday use in schools and work; and in the serious issues facing educators, such as teacher qualifications and shortages, student dropouts, school financing, and student achievement. As changes occur in the curriculum, the effects of these changes should be measured in such a way that measurement results could be used by policymakers and, in fact, could influence the direction or the acceleration of the changes. Current large-scale assessments are more sensitive to the status quo and therefore are insensitive to changes that may occur in reform curriculum. It is clear that a new monitoring system is needed.

In conclusion, several issues must be raised. The term assessment has been used to describe a variety of processes, such as the assessment of individual abilities, the assessment of student learning, the assessment of schooling, state assessment, and national assessment. Romberg articulates the need to identify the unit being assessed; such a specification is very relevant in judging whether or not a particular procedure is appropriate. In the historical analysis of testing procedures, the application of a test has gone beyond its intended purpose, particularly regarding the unit of testing. Standardized tests, which were developed to sequence individuals on a line based on the scores of a norm group, are frequently used to group students (Stiggins & Bridgeford, 1985). However, the tests that were developed to make decisions regarding individuals are generally not appropriate to make decisions regarding school programs. Variation in group results can fluctuate considerably depending upon the technique used to compute the group score (Baglin, 1986). At the same time, results from NAEP (which is an example of a profile form of assessment) cannot be used to make decisions regarding individuals because the sampling technique used requires an individual to take only a small sample of the exercises; the sampling technique does provide information for the nation. In short, the assessment procedure must be appropriate for the intended purpose. Mathematics education reform will be monitored to make decisions regarding large groups; the form of assessment to be used must be selected with this in mind.

Another consideration, the precision of measurement, was not discussed by Romberg but is relevant to determining an appropriate means for assessment and in evaluating whether an "old" form will suffice. How precise does the measurement have to be to provide the needed information to make decisions? This will depend on the decisions to be made and the costs or results of making a wrong decision. If the process of monitoring reform and the information to be derived from the measurement affects the allocation of large sums of money, the assessment instruments need to be precise. The importance of the measurement's precision also depends on the amount of change expected; if the results of reform are to be grand, affecting a large number of students, the form of assessment can be more coarse. If the results of the reform are to be subtle or gradual, the form of assessment must be able to detect minute changes. For example, if the reform is to require that students learn a totally different topic than that which is currently being taught (e.g., probability and statistics in the eighth grade), then a limited number of tasks can be used to show change. However, if evidence of the reform is to involve more students performing better on what is currently being taught, and the change is to be very gradual, a much larger number of tasks (items) are needed with a very refined calibration strategy.

Another relevant factor to be considered involves the assumptions of how the reform will take place. This issue is explicitly relevant to the assessment strategy, but also to the form of assessment. If it is anticipated that reform will be evenly distributed across a population and will occur continuously over time, some form of a standardized test may be appropriate. More likely, however, the effects of reform will be localized and will occur in steps and stages. This suggests that the form of assessment must be adaptable, flexible, and fluid so that local changes can be observed, while being built on solid conceptual foundation to measure different forms of a central idea.

It is difficult to conclude that current forms of assessment will be inappropriate for monitoring reform without specifying and researching more about what shape the reform will take. Without a refined notion of the anticipated changes, one appropriate strategy may be the shotgun approach where a battery of assessment instruments are used based on a number of different forms. Chapters 17-20 offer guidance in the process of developing some useful assessment procedures for monitoring school mathematics. What is needed now are the details.

## References


Baglin, R. F. (1986). A problem in calculating group scores on norm-referenced tests. Journal of Educational Measurement, 23(1), 57-68.

Bloom, B. S. (Ed.). (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain. New York: Longman.

Carpenter, T. P., & Moser, J. M. (1983). The acquisition of addition and subtraction concepts. In R. Lesh & M. Landau (Eds.), The acquisition of mathematical concepts and processes (pp. 7-44). ew York: Academic Press.

Cronbach, L. J. (1970). Essentials of psychological testing. New York: Harper & Row.

Crosswhite, F. J., Dossey, J. A., Swafford, J. O., McKnight, C. C., Cooney, T. J., Downs, F. L., Grouws, D. A., & Weinzweig, A. I. (1986). Second international mathematics study detailed report for the United States. Champaign, IL: Stipes.

Finley, C. J. (1974). Not just another standardized test. In R. W. Tyler & R. M. Wolf (Eds.), Crucial issues in testing (pp. 95-102). Berkeley, CA: McCutchan.

Gust, D. (1986, August 23). Coping with changing times. The Arizona Daily Star, p. B9.

Hilton, P. (1981). Avoiding math avoidance. In L. A. Steen (Ed.), Mathematics tomorrow (pp. 73-82). New York: Springer-Verlag.

Husén, T. (1969). Educational evaluation: New roles, new means. In R. Tyler (Ed.), Educational evaluation: New roles, new means. The 68th yearbook of the National Society for the Study of Education, Part II (pp. 335-350). Chicago: University of Chicago Press.

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

Mehrens, W. A., & Phillips, S. E. (1986). Detecting impacts of curricular differences in achievement test data. Journal of Educational Measurement, 23(3), 185-196.

Popham, J. W., Cruse, K. L., Rankin, S. C., Sandifer, P. D., & Williams, P. L. (1985). Measurement-driven instruction: It's on the road. Phi Delta Kappan, 66, 628-634.

Postman, N. (1979). Teaching as a conserving activity. New York: Delta.

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

Romberg, T. A., & Wilson, J. W. (Eds.). (1969). The development of tests, National Longitudinal Study of Mathematical Abilities. Stanford, CA: School Mathematics Study Group.

Sanders, R. J., & Murray, S. L. (1976). Alternatives for achievement testing. Educational Technology, 16(3), 17-23.

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. Journal of Educational Measurement 22(4), 271-286.

Swezey, R. W. (1981). Individual Performance Assessment: An approach to criterion-referenced test development, Englewood Cliffs, NJ: Reston.

Tyler, R. W. (1970). Basic principles of curriculum and instruction. Chicago: University of Chicago Press.

Tyler, R. W. (1974). Why evaluate education? In R. W. Tyler & R. M. Wolf (Eds.), Crucial issues in testing (pp. 91-95). Berkeley, CA: McCutchan.

Tyler, R. W., & Wolf, R. M. (Eds.). (1974). Crucial issues in testing. Berkeley, CA: McCutchan.

Vergnaud, G. (1983). Multiplicative structures. In R. Lesh & M. Landau (Eds.), Acquisition of mathematics concepts and processes (p. . 27). New York: Academic Press.

Wisconsin Department of Public Instruction. (1986). 1985 Wisconsin Pupil Assessment Report. Madison: Wisconsin Department of Public Instruction, Bulletin No. 6378.

Chapter 22

# ATTITUDES TOWARD MATHEMATICS

Gilah C. Leder

In recent years there has been a growing recognition that understanding the nature of mathematics learning requires exploration of affective as well as cognitive factors. Large scale surveys of students' performance in mathematics, such as the National Assessment of Educational Progress (NAEP, 1983), the Second International Mathematics Study (SIMS) (Crosswhite, Dossey, Swaffort, McKnight, & Cooney, 1985), and the National Assessment of Participation and Achievement of Women in Mathematics (Armstrong, 1985), have in fact included items designed to produce a measure of students' attitudes toward mathematics. The generally comprehensive Handbook of research on teaching (Wittrock, 1986), on the other hand, does not explore in any depth the interaction between attitudes and school learning. The authors of one of its chapters (White & Tisher, 1986) explain this omission as follows:

> Research has been handicapped by absence of a mature theory encompassing the nature of attitudes and their relation to other constructs. The external boundaries of attitudes with personality attributes and with abilities are blurred, and so are the internal ones between interests, feelings, values, and appreciations. (p. 892)

To help set in context the difficulties that face those concerned with attitudes toward mathematics, a brief overview of issues related to the definition and measurement of attitude in a broader context is essential.

## Attitude: The Problem of Definition

Consensus about the central position of attitude research in social psychology is not mirrored in agreement about the definition of attitude. Many investigators seem to select for their definition a measurement procedure that is convenient for the purpose of their study. Until recently, those concerned with measurement typically defined attitude as unidimensional, while those concerned with theory building have tended to use a broad multistructural definition.

The difficulty of equating the operational definition of attitude with its theoretical construct was highlighted by Fishbein and Ajzen (1975) who identified more than 500 different methods of measuring attitude in their review of research published between 1968 and 1970. Nevertheless, as can be seen from the sample of definitions of attitude summarized in Table 1 that span approximately five decades, there is much more overlap.

Table 1

Some Definitions of Attitude

| Author(s) | Year | Main Features of the Definition |
|---|---|---|
| Thurstone | 1928 | The sum total of the individual's inclinations and feelings, prejudice or bias, preconceived notions, ideas, fear, threats, and convictions about any specific topic. |
| Allport | 1935 | A mental and neutral state of readiness, organized through experience. It exerts a directive and dynamic influence upon the individual's response to all objects and situations with which it is related. |
| English & English | 1958 | An enduring learned predisposition to behave in a consistent way toward a given class of objects. |
| Shaw & Wright | 1967 | A relatively enduring system of Wright evaluative, affective reactions, reflective of the beliefs which have been learned about the characteristics of a social object (or class of social objects). |
| Fishbein & Ajzen | 1975 | A learned predisposition to respond in a consistently favorable or unfavorable manner to a given object. |

Several important components emerge from these definitions: attitude is learned; it predisposes to action that may be either favorable or unfavorable; and there is response consistency. The consensus implied by these commonalities is illusory, however. There is disagreement among theorists about the degree of interrelationships among the three components and whether or not they should be examined as separate entities. Furthermore, there are differences in the ways the key components are interpreted. For example, the notion of response consistency is interpreted by some to imply that an individual will perform consistently, given the same stimulus. Others concentrate on the notion that different responses elicited by any one object should be consistent with each other. Still others are more concerned with evaluative consistency, i.e., <u>overall</u> favorability or unfavorability expressed toward an object by a set of behaviors. Since these different interpretations are reflected in the way attitude is measured, i.e., inferred from observable behavior, these distinctions are not merely academic but have considerable practical implications for attitude research in general, and research on attitude toward mathematics in particular.

The notion of attitude as a predisposition is equally ambiguous. Predispositions must be inferred from consistencies in behavior, a requirement open to at least three different interpretations, as discussed above, and hence at least three quite distinct measurement approaches. "These problems are compounded when the level of dispositional specificity fails to correspond to the interpretation of response consistency. In a typical example, an investigator may infer attitude by observing overall evaluative consistency but assume a predisposition to perform a specific behavior" (Fishbein & Ajzen, 1975, p. 9).

Exactly how attitude is learned, which of the individual's previous experiences determine consistently favorable or unfavorable behavior toward an object, also continues to be an area of controversy and disagreement over optimal operational definitions. Some of the relevant nuances are captured well in the following excerpt:

> Attitudes involve what people think about, feel about, and how they would like to behave toward an attitude object. Behavior is not only determined by what people would like to do but also by what they think they should do, that is, social norms, by what they have usually done, that is, habits, and the expected consequences of the behavior. (Triandis, 1971, p. 14)

Thus, more specifically, attitude toward mathematics should not be treated as a unitary concept, nor can a simple link be assumed between attitudes toward mathematics and student outcome measures pertaining to mathematics.

The perspective from which attitudes are investigated depends largely on the theoretical orientation of the investigator. More prosaically, practical constraints will also affect measurement techniques. Fishbein and Ajzen (1975) discussed a number of theoretical approaches and indicated consistencies and differences between these and their own preferred conceptualization of attitude. Despite the risk of oversimplification, some central themes and concerns are summarized in Table 2.

Table 2

Theories of Attitude

| Approacl. | Key Concepts |
| --- | --- |
| Learning theories attitudes | Typically concerned with the ways in which attitudes are acquired. Explanations are given in terms of both classical and instrumental conditioning. Relations between attitudes are explored. Conflicting evaluations are considered to be resolved according to the congruity principle, i.e., with a shift in the differing evaluations toward equilibrium or congruity. |
| Expantancy-value theories | A causal relationship is postulated between behavior and the expected value of the outcome. The individual's attitudes toward an object depend on whether it is perceived as being instrumental in obtaining a positively valued goal or avoiding a negatively valued goal. Thus attitudes are determined by beliefs and associated evaluations. |
| Balance theory | Concerned with the qualitative relations between elements. If there are inconsistencies in an individual's perceptions of these relations, then there will be stress toward change and a balanced state (through, e.g., a change in attitude, attribution, or behavior). Failure to achieve balance results in tension. |
| The congruity principle | While the balance model and the congruity principle are both concerned with the qualitative relations between elements, the former focuses on perceived relations, while the latter treats these relations as assertions, i.e., as given. A state of congruence is said to exist when evaluations of two objects are equally intense and in consistent directions. When a state of incongruity exists, the extent to which the assertion is believed determines the degree of attitude change. |

| Approach | Key Concepts |
|---|---|
| Cognitive dissonance theory | According to Festinger (1957) there are four main sources that contribute to cognitive dissonance: the discrepancy between the cognitive elements, the importance of these elements, forced compliance, and the individual's commitment. Maximum dissonance is hypothesized to occur when the discrepancy is large, the elements are important, the individual has selected a particular behavior without coercion and is committed to the outcome of that behavior. Dissonance reduction can be achieved through changing one's opinion, attempting to influence others, or by devaluing their importance. Fishbein and Ajzen (1975) point out that at least some of the conflicting findings obtained in dissonance theory research can be attributed to a conceptual blurring between attitudes and beliefs. |
| Attribution theory | Examines how the effects produced by an action are attributed. Such attributions may be internal (i.e., ability or motivation) or external (i.e., difficulty of the task or luck) and may be shaped by the presence or absence of a specific factor in the presence or absence of the effect of interest. Attributions are hypothesized to be influenced by consistency (Is the same behavior exhibited to that object on different occasions?), by distinctiveness (Is the behavior shown to that object different from behavior shown toward other objects?), and consensus (Do other individuals behave in the same way toward that object?). The degree to which attribution theory helps to explain the formation of beliefs about one's self is still a matter of some debate. |

While all too brief, the summaries in Table 2 illustrate that a range of theoretical perspectives, with consequent differences in the variables selected as central to the various theories, is brought to attitude research. To foreshadow the later section on the measurement of attitudes toward mathematics, traces of the different approaches are embedded, to varying degrees, in the instruments used to tap attitudes to mathematics. For example, the attribution to mathematics scales used by Wolleat, Pedro, Becker, and Fennema (1980) and Leder (1981, 1984) and the enrollment in mathematics course data used by Armstrong (19[illegible]), as well as in numerous other studies, can be linked to the attribution theory and the expectancy value theory of attitude respectively. The different conceptualizations of attitude lead to differences in operational definitions of attitude and ultimately to differences in the interpretation of observed outcomes or behavior.

Many of the approaches used to measure attitude in fact rely on self-report paper-and-pencil instruments. These, as noted by Kiesler, Collins, and Miller (1969), do not make use of overt behavior. Other approaches to attitude measurement include drawing inferences from observing overt behavior in a natural setting, from considering an individual's reaction to or interpretation of partially structured stimuli, from an individual's performance on "objective" tasks, and from the physiological reaction of respondents to the attitudinal object or representation of it.

Before discussing measurement approaches used to assess attitude toward mathematics per se, it is useful to consider which variables are most frequently examined in conjunction with attitude toward mathematics.

In a then timely review of the literature, Aiken (1970) summarized the findings of a large number of journal articles, doctoral dissertations, and other reports concerned with attitude toward mathematics. This review included an overview of techniques used to measure attitude toward arithmetic and mathematics, the distribution and stability of attitude toward mathematics, interaction between attitude toward and achievement in mathematics, and the effects of different mathematics curriculum and practices on attitude toward mathematics. Also discussed were the effect of student variables such as anxiety, general ability, and gender, and the importance for student attitude toward mathematics of parents' attitude and teachers' attitude as well as selected other teacher characteristics. Aiken concluded, "Of all the factors affecting student attitude toward mathematics, teacher attitudes are viewed of particular importance" (Aiken, 1970, p. 592).

Many of the variables reviewed by Aiken were also examined in subsequent reviews of research on mathematics by Kulm (1980) and Bell, Costello, and Kuchemann (1983). All three reviews concluded that, though the correlation between attitude and achievement in mathematics was positive, its magnitude was small. "Broadly speaking, the set of people who like mathematics has only a relatively small overlap with the set of those who are good at it" (Bell et al., 1983, p. 255). The parallel between this and the typically weak relationship between an individual's attitude and behavior is inescapable. Triandis' (1971) warning, quoted earlier in this paper, that many variables confoun the relationship between attitude and behavior can be translated to the mathematics setting. Mathematics related outcomes are influenced by attitudes which in turn are affected by the individual's thoughts, feelings, preferred model of behavior (e.g., level of achievement), habits, expected consequences (of the level of achievement, say), and the social norms of the society within which the individual functions. In recent years particular attention has been paid to the effect of gender and race on attitudes toward mathematics.

The list of variables linked with attitude toward mathematics in reviews such as those cited above is reflected in the multidimensional approach of the more recent measures of attitude toward mathematics. It is appropriate to turn now to the important characteristics of distinctly different methods used to tap attitudes toward mathematics.

## Attitude Toward Mathematics: The Problem of Measurement

The techniques selected for discussion and the approach used in this section rely heavily on an earlier article by Leder (1985). For maximum clarity, each method discussed is illustrated by relevant examples, taken from recent large scale testings.

The following techniques are discussed: Thurstone scales, summated rating scales exemplified by (the most common) Likert-type scales, semantic differential scales, interest inventories and checklists, preference ranking, projective techniques, enrollment data, other forms of data gathering such as clinical and anthropological methods, and psychological responses. While the majority of these techniques are self-report paper-and-pencil measures, examples of instruments in other categories are also included.

## Thurstone (Equal-Appearing Interval) Scales

Possible item: I will do more mathematics because my mother thinks that mathematics is really important.

Development of a Thurstone scale requires a number of steps. In the first instance a pool of items, reflecting a continuum of attitude to arithmetic, say, is written. A group of "judges" is then asked to place these items in one of (typically) 11 piles, with the items considered most favorable to be put into the first pile, the least favorable into the last pile, and the other items in between, as deemed appropriate. A scale value (the mean or median of the ratings assigned by the judges) can thus be calculated for each statement. Items to which the judges assign widely differing ratings are omitted from the final scale. Respondents to whom the scale is administered are asked to identify those items with which they agree. The mean or median of the scale value of the items selected represents each respondent's attitude score.

Critics of Thurstone's approach have questioned his assumption that the judges' own biases would not influence their ratings. The alternate scaling procedure suggested by Likert overcomes this problem.

Recent testings have typically not used a Thurstone scale to assess student attitude. Yet it is interesting to consider one of the approaches described by Armstrong (1985) to tap student attitude toward mathematics. Students were asked to order nine factors to indicate the influence of each on their decision to take further mathematics courses. The mean value for each of the items could be computed. The responses of different groups (boys and girls, students of different ages) could thus be compared by examining the mean values assigned to each item by the different groups. This procedure shared some of the features used in the development of a conventional Thurstone scale:

1. A pool of items is selected (in this example, presumably on the basis of earlier research findings)

2. "Judges" are asked to rank order the items

3. However, instead of the Thurstone procedure of asking students to respond to the derived scale, the judgments are examined for group similarities and differences.

Likert Scale

Typical item: Mathematics is useful in solving everyday problems

SD D U A SA

Collecting a large pool of items reflecting either a positive or a negative attitude toward mathematics is the first step in constructing a Likert scale. While items indicating a neutral attitude are appropriate for a Thurstone scale, they are eliminated from a Likert scale. Subjects to whom the scale is administered are asked to indicate their response to each item, typically on a five-point scale ranging from Strongly Agree to Strongly Disagree. Strong agreement and disagreement with favorable items are scored as 5 and 1 respectively. Appropriate ratings are given to the intermediate responses. Scoring is reversed for unfavorable items. On the assumption of unidimensionality, i.e., that all the items measure the same construct, attitude is defined as the sum of the item scores. Items that do not correlate significantly with the overall attitude score are not retained. After trial, the 20 or so items with the highest correlations form the Likert scale.

The Fennema and Sherman (1976) Mathematics Attitudes scales are a widely used example of Likert scales. These researchers conceptualized attitude toward mathematics as comprised of a number of components, most meaningfully reported separately. Their scales consist of eight distinct clusters of items designed to measure confidence in learning mathematics, effectance motivation in mathematics, attitude toward success in mathematics, mathematics anxiety, mathematics as a male domain, and father's, mother's and teacher's perceptions of the student as a learner of mathematics.

There is much overlap in the approach used by Fennema and Sherman (1976) and that found in the NAEP, SIMS, and the Assessment of Performance Unit (APU) studies. For example, separate scales were used to assess attitudes toward mathematics and society, mathematics and myself, mathematics as a process, mathematics and gender (SIMS), and mathematics as an emotive subject, mathematics as a useful subject, confidence in doing mathematics, enjoyment in doing mathematics, and perceived difficulty of mathematics (APU, see Joffe & Foxman, 1984).

## (Osgood's) Semantic Differential Scale

Possible item: Mathematics

Worthwhile . . . . . . . Trivial

The semantic differential technique was originally developed by Osgood, Suci, and Tannenbaum (1957) to measure meaning. It consists of a number of stimulus words or concepts; subjects indicate the position on a line between pairs of bipolar adjectives (such as good/bad or masculine/feminine) that best reflects their feeling about that stimulus. A seven-point rating scale is commonly used. The ratings are combined and analyzed in various ways to describe the respondent's attitude. Factor analysis typically reveals that three basic dimensions underlie the common explainable variance: evaluation, potency, and activity.

The value of the technique depends to a large extent on the suitability of the stimulus words or concepts chosen, as well as on the relevance to them of the bipolar adjectives selected.

The semantic differential is often regarded as a less transparent, more indirect measure of attitude than the other measures discussed so far.

One example of its use is in a study conducted by Nimier (1976) of attitudes toward mathematics in 24 high school classes in France. His choice of bipolar adjectives--useful/useless, repulsive/attractive, easy/difficult, voluntary/compulsory, not feasible/feasible, unrealistic/realistic--overlaps with the components selected in studies using Likert scales to assess attitudes to mathematics. Tapped again are the usefulness, difficulty and enjoyment to be derived from doing mathematics. Given the different cultural setting of this study, it is worth noting that students concentrating on the sciences rated mathematics as more positive (or closer to the positive pole) than did students concentrating on the humanities. Furthermore, for each of the seven adjectives cited, boys' mean ratings were more favorable than those of the girls.

Inventories and Checklists

Typical items: A list of occupations
A list of words (adjectives, verbs)

Inventories and checklists are two other examples of subjective rating scales. The former typically consists of a list of careers, activities, hobbies, or adjectives. The respondent is asked to indicate items of particular interest.

Checklists are used to obtain descriptions or self-descriptions or to elicit stereotypes about groups of people. Respondents are asked to indicate those words they consider most applicable to themselves or to the target group, as appropriate.

Asking students to choose, out of a sample of eight careers, the career they expect to follow (Armstrong, 1985) is an example of the inventory approach. In Armstrong's study the link to mathematics was made more explicit by the follow-up question which required an indication of the amount of mathematics thought to be necessary for that career.

A checklist was used by Nimier (1976) as one of his instruments to gauge students' attitudes toward mathematics. Students selected three verbs, out of a list of 42, to indicate how they felt when doing mathematics. The range of verbs used was varied and included pervert, struggle, destroy, worry, discover, conquer, arrange, and assimilate.

Preference Rankings

Typical item: A list of school subjects, to be ranked in order of preference; asking students to specify their favorite school subject.

Preference ranking requires students to list the subjects they study at school in order of preference. The rank assigned to mathematics is thus obtained. However, the relative nature of the measure imposes limitations. A student with a very favorable attitude to school could put mathematics last and yet have a more positive attitude toward mathematics than another student who ranked mathematics first.

Asking students to indicate their favorite school subject as well as a selection of questions about that subject was part of the attitude toward mathematics data gathering approach used by the APU.

Projective Technique

Typical item: A request "to write about" a cue figure or to complete a partially formed sentence.

Projective techniques represent an indirect approach to the measurement of attitudes. They therefore rely less on the honesty and cooperation of respondents than do more explicit methods. Projective techniques may involve sentence completion (A good mathematics lesson . . . ), a word association test, a picture preference test, or a request to tell a story in response to a given cue. Because of the difficulty of ensuring satisfactory validity, reliability, and particularly consistent scoring of projective measures, they are not used often to tap mathematics attitude. Nevertheless, responses to partially structured stimuli can provide powerful insights into respondents' attitudes.

An interesting example of a projective technique is the use of repertory grids by Walden and Walkerdine (1985) in their study of students' progress, particularly in mathematics, as they moved from the primary to the secondary school. Students were asked to write about people they liked and people they disliked. Various themes emerged from the analysis of these stories.

> When the grids were compared the most interesting data were concerned with the relationship of the construct clever/not clever to the subjects which the children did in class. For boys, cleverness and being good at mathematics were close together. Girls linked cleverness and being good at mathematics with being good in English and being popular. (Walden & Walkerdine, 1985, p. 67)

Interviews conducted with the students typically supported the repertory grid data.

Enrollments

Typical item: Statistics on enrollment in mathematics courses.

A number of factors, including a positive attitude to mathematics, are generally assumed to influence students' decision to continue with mathematics courses once they are no longer compulsory. Haladyna, Shaughnessy, and Shaughnessy (1983), for instance, argued that "a positive attitude toward mathematics may increase one's tendency to elect mathematics courses in high school and college" (p. 20). Their interpretation rests on a willingness to accept a decision to continue with a course, say mathematics, as a measure of attitude to mathematics. A similar interpretation is prevalent in studies that consider gender-linked differences in mathematics learning. However, because of the widely recognized role of mathematics' prerequisites as a critical filter into other courses, apprenticeships, and occupations, the importance of other variables is likely to confound the attitude to mathematics component as a determinant of mathematics course taking.

The NAEP, SIMS, and Armstrong (1985) surveys all reported enrollment in mathematics course data and used these statistics as one measure of attitudes toward mathematics.

Other Forms of Data Gathering: Clinical and Anthropological Observations

Typical item: Observation of overt behavior in a natural setting.

In the study referred to earlier, Walden and Walkerdine (1985) video-taped regular classroom sessions and inferred individuals' attitudes toward classmates and curriculum areas from an analysis of the tapes. The difficulty of extracting attitudes to mathematics from the many other factors that determine behavior has already been discussed.

Physiological Measures

Possible item: Measures of heart rate and/or electrical skin resistance.

Physiological ratings (electrical skin resistance, breathing rate, blood pressure, heart rate) of attitude toward mathematics have been found in a number of research studies.

Most recently McLeod (1986) and Mandler (1986) have talked of changes such as increased muscle tension and rapid heart beat as physiological adjuncts to problem solving in mathematics. Because of the difficulties associated with obtaining such physiological measures per se, their use as indicators of attitudes toward mathematics is likely to remain limited. Some of the relevant information could, however, be captured through self-report measures.

The review of measures of attitudes toward mathematics served a threefold purpose. It allowed a range of different techniques to be discussed; it alluded to the findings of consistent differences in attitudes to mathematics of certain groups, specifically boys and girls; and it revealed that contemporary large scale surveys concerned with assessing attitudes to mathematics have used a multifaceted approach. Thus there is a clear recognition that attitude should not be represented by a single score, representative of an overall, general predisposition to the subject of mathematics. Instead, attitude is best regarded as a complex construct, influenced by a host of variables that cannot be measured adequately by a conventional unidimensional scale.

The data in Table 3 show the variety of attitude measures used in the large scale surveys referred to throughout the review.

Table 3

Summary of Attitude Measures Used in Selected Large Surveys

| Survey | Country | Main Measures Used |
|---|---|---|
| SIMS | 20 different countries | Enrollment data. Likert scales to assess attitudes about the usefulness of math and the importance of math to society, gender stereotyping of math, mathematics as a process, and students' views of themselves as learners of mathematics. |
| NAEP (1983) | USA | Enrollment data. Likert scales (similar to those used in SIMS) |
| National Assessment of Women in Mathematics (Armstrong, 1985) | USA | Enrollment data, actual and intended Likert scales (similar to SIMS, though not as comprehensive). Preference ranking. Interest inventory. |
| APU (Jaffe & Foxman, 1984) | UK | Preference ranking. Likert scales to assess attitude toward mathematics as an emotive subject and as a useful subject, confidence in doing mathematics, enjoyment in doing math, and perceived difficulty of math. |
| Walden & Walkerdine (1985) | UK | Projective measure. Anthropological observations. Interview. |
| Nimier (1976) | France | Semantic differential. Likert scales to assess attitude toward various aspects of mathematics. Checklist (verbs). |

## Attitude Toward Mathematics: Group Differences

As noted by Leder (1986) the issue of gender-linked differences in mathematics is extremely complex. Despite the inroads made by females into mathematics and related careers, students and teachers continue to perceive mathematics as a male domain. Society continues to highlight the difficulties faced by successful females, the price they need to pay to achieve success in traditional male areas. Such stereotyping is reflected in students' attitudes toward mathematics, e.g., in their attitudes to the usefulness of mathematics and themselves as learners of mathematics.

The conclusions of the APU survey (Joffe & Foxman, 1984) are illustrative of commonly found gender-related differences.

- When asked to rate statements and indicate the perceived difficulty and usefulness of mathematical topics and items, girls tend to make more moderate assessments; they use extremely positive and extremely negative positions on the rating scales far less than boys do.

- Girls express greater uncertainty about their mathematical performance. Boys express a greater expectation of success.

- Boys overrate their performance in mathematics in relation to written test results; they do not do as well as they expect to do. Girls underrate their performance and do better on tests than they expect. (p. 25)

The NAEP data have also highlighted race-related differences in achievement in mathematics and participation in mathematics courses once they are no longer compulsory. These differences are accompanied by and reinforce race-linked differences in attitudes toward mathematics, with differences in the perceived usefulness of mathematics and in the way students perceive themselves as learners of mathematics again being two notable areas. Future investigations should be sensitive to subtle but consistent group differences in attitudes toward mathematics.

## Concluding Comments

The definition and measurement of attitudes are interdependent--both in the broader context and in the area of mathematics. There is agreement that attitude toward mathematics should be conceptualized as a multidimensional construct, with the varying components most effectively assessed separately using several quite distinct techniques, if possible. When interpreting the results obtained, due attention should be paid to the restrictions imposed by the operational definition selected.

The continuing concern of social psychology with attitude research serves as testimony to the complexity of the area. Attitudes involve individuals' thoughts, feelings, and preferred behavior. They are also affected by the social norms and standards of behavior prevalent in the society within which the individuals function. Attitudes toward mathematics are similarly complex and multifaceted. Instruments used to measure attitudes toward mathematics should reflect these various dimensions.

The summary of attitude toward mathematics measures used in recent large-scale testings in a number of different countries has illustrated the heavy reliance placed on readily quantifiable outcomes such as enrollment data, as well as on self-report paper-and-pencil measures. Both approaches need to be interpreted with caution. As pointed out earlier, behavior is determined not only by the attitude being studied but as well by a host of other variables--both situational and psychosocial. Distortion of overt responses cannot be ruled out with self-report measures, particularly those whose purpose is obvious to the respondent.

Practical constraints suggest that self-report paper-and-pencil techniques will continue to be popular methods for assessing attitudes toward mathematics. Ways of improving their efficacy thus seem well worth exploring. Suggestions made in the general literature include adding items that focus on a different component from the one being studied or adding other somewhat irrelevant items, on the assumption that such inclusions would help mask the purpose of the instrument. Ensuring anonymity of reply is also believed to lower the distortion rate. Whatever the eventual approach selected, the aim should be to quantify attitudes rather than attitude toward mathematics.

## References

Aiken, L. R., Jr. (1970). Attitude towards mathematics. Review of Educational Research, 40, 551-596.

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), A handbook of social psychology. Worcester, MA: Clark University Press.

Armstrong, J. M. (1985). A national assessment of participation and achievement of women in mathematics. In S. Chipman, L. R. Brush, & D M. Wilson (Eds.), Women and mathematics: Balancing the equation. Hillsdale, NJ: Erlbaum.

Bell, A. W., Costello, J., & Kuchemann, D. (1983). A review of research in mathematical education. Windsor, Berks.: NFER, Nelson.

Crosswhite, F. J., Dossey, J. A., Swafford, J. O., McKnight, C. C., & Cooney, T. J. (1985). Second international mathematics study: Summary report for the United States. Washington, DC: National Center for Educational Statistics.

English, A. C., & English, H. B. (1958). A comprehensive dictionary of psychological and psychoanalytical terms: A guide to usage. London: Longmans.

Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitudes scales: Instruments designed to measure attitudes toward the learning of mathematics by male and female. Abstracted in the JSAS Catalog of Selected Documents in Psychology, 6(2), 31.

Festinger, L. (1957). A theory of cognitive dissonance. Stanford, CA: Stanford University Press.

Fishbein, M., & Ajzen, J. (1975). Belief, attitude, intention and behaviour: An introduction to theory and research. Menlo Park, CA: Addison-Wesley.

Haladyna, T., Shaughnessy, J., & Shaughnessy, J. M. (1983). A causal analysis of attitude toward mathematics. Journal for Research in Mathematics Education, 14, 19-29.

Joffe, L., & Foxman, D. (1984. September). Attitudes and sex differences. Mathematics in School, 22-26.

Kiesler, C. A., Collins, B. E., & Miller, N. (1969). Attitude change. New York: Wiley.

Kulm, G. (1980). Research on mathematics attitude. In R. J. Shumway (Ed.), Research in mathematics education. Reston, VA: National Council of Teachers of Mathematics.

Leder, G. C. (1981). Learned helplessness in the classroom? Research in Mathematics Education in Australia, 2, 192-203.

Leder, G. C. (1984). Sex differences in attribution of success and failure. Psychological Reports, 54, 57-58.

Leder, G. C. (1985). Measurement of attitude to mathematics. For the Learning of Mathematics, 5, 18-21, 34.

Leder, G. C. (1986, April). Gender-linked differences in mathematics learning: Further explorations. Paper presented at the research presession of the annual meeting of the National Council of Teachers of Mathematics, Washington, DC.

McLeod, D. B. (1986, May). Affective influences on mathematical problem solving. Paper presented at the conference Affective Influences in Mathematical Problem Solving, San Diego.

Mandler, G. (1986, May). Affect and learning: Causes and consequences of emotional interactions. Paper presented at the conference Affective influences on mathematical problem solving, San Diego.

National Assessment of Educational Progress. (1983, April). The third national mathematics assessment: Results, trends and issues. Denver: Education Commission of the States.

Nimier, J. (1976). Mathematiqué et affectivité. Paris: Collection Pernoud, Ed. Stock.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). The measurement of meaning. Urbana: University of Illinois Press.

Shaw, M. E., & Wright, J. M. (1967). Scales for the measurement of attitudes. New York: McGraw-Hill.

Thurstone, L. L. (1928). Attitudes can be measured. American Journal of Sociology, 33, 529-554.

Triandis, H. C. (1971). Attitude and attitude change. New York: Wiley.

Walden, R., & Walkerdine, V. (1985). Girls and mathematics: From primary to secondary schooling (Bedford Way Papers 24.) London: Institute of Education, University of London.

White, R. T., & Tisher, R. P. (1986). Research on natural sciences. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd edition). New York: Macmillan.

Wittrock, M. C. (Ed.). (1986). Handbook of research on teaching (3rd edition). New York: Macmillan.

Wolleat, P. L., Pedro, J. D., Becker, A. D., & Fennema, E. (1980). Sex differences in high school students' causal attributions of performance in mathematics. Journal for Research in Mathematics Education, 11, 356-366.

Chapter 23

# NEW APPROACHES TO RESEARCH ON ATTITUDE

Douglas B. McLeod

In chapter 22 Gilah Leder presents a state-of-the-art report on attitudes toward mathematics. Her review of related work presents the complexities of research on attitudes, including both the strengths and weaknesses of investigations on this topic. For example, she presents a clear picture of the importance of the affective domain in the monitoring of school mathematics and suggests a variety of strategies that are effective in measuring attitudes. She also notes the problems involved in defining attitudes and measuring attitudes, difficulties that plague all research in this area. For further discussion of the practical problems of measuring attitudes, see Henerson, Morris, and Fitz-Gibbon (1978).

In spite of these difficulties in research on attitudes, the last decade has been a period of substantial progress in our knowledge of attitudes toward mathematics. Reyes (1984) documented the progress that has been made in this area, especially in research on gender differences in mathematics education. Much of this progress has come about through the extensive use of the Fennema-Sherman scales (Fennema & Sherman, 1976) and similar instruments. It is reassuring to find, as Leder did in an earlier paper (Leder, 1986), that research on gender differences in mathematics education is producing relatively consistent results in terms of attitudes. This consistency is found not only among studies conducted in North America but also in the research conducted in Australia and the United Kingdom. Moreover, these confirmatory results often come from relatively large assessment projects (McLean, 1982; Foxman, Martini, & Mitchell, 1982), not just from small-scale research studies. So we have evidence that reasonably good data can be obtained on attitudes, even in large-scale efforts to monitor school mathematics.

Research on attitudes has made progress not only in the consistency of the results but also in the development of more sophisticated models to guide the research. This line of research has expanded to include investigations of gender differences in attributions of success and failure in mathematics (Reyes, 1984). The connection between research on attitudes and on attributions (Weiner, 1979) has been particularly useful in mathematics education and promises to make further contributions to our understanding of the relationships among attitudes, achievement, and gender (Fennema & Peterson, 1985).

Although research on attitudes toward mathematics has made substantial progress, there is general agreement that much more needs to be done. The purpose of this chapter is to argue that a new approach to the affective domain could yield substantially more progress, especially in developing better theories for affective factors, in making connections to contemporary theories of learning, and in monitoring higher-order thinking and problem solving in mathematics. To implement this new approach, conceptions of affect need to be broadened to include more than the usual attitude dimensions of liking mathematics, seeing the usefulness of mathematics, and feeling confident about mathematics. This broadened perspective requires a new theoretical framework; this chapter discusses the relevance of such a framework to the problems involved in measuring attitudes and other affective factors. Finally, this chapter discusses the implications of these ideas for the monitoring of school mathematics.

## Conceptions of the Affective Domain

Leder presents a thorough discussion of definitions of attitude as traditionally employed in both psychological research and mathematics education. In this section, however, I would like to expand the discussion to a broader view of affect. Attitudes are a part of the affective domain, but not all of it. For this chapter, affect will be used as a general term to represent all the feelings that seem to be related to mathematics learning and teaching--the attitudes, beliefs, moods, and emotions that may have an influence on mathematical performance. Emotion will be used to signify a more visceral kind of affect, a response that is quite intense but of relatively short duration. In Simon's (1982) terms, emotion is used to refer to affect that is sufficiently powerful to redirect attention. Moods (again from Simon, 1982) provide a context within which cognitive processes are carried out; moods are not so intense that they redirect attention. Beliefs (Silver, 1985) fall in the intersection of the sets of student knowledge and feelings; beliefs about the usefulness of mathematics, for example, are often treated as an attitude variable. Finally, attitudes will refer to affective responses that are relatively consistent, but not especially intense; this view is consistent with Leder's position.

In this chapter the focus is on the two extremes of the affective domain, emotions and attitudes. Sometimes the distinction is made between "hot" and "cold" affect, where emotions like joy, frustration, and fear are considered hot, and attitudes (liking mathematics, seeing mathematics as useful) are considered cold. For further clarification of terminology for the affective domain, see Simon (1982) and Reyes (1987).

This expansion of the affective domain to include more visceral, emotional responses to mathematics is related to new views of what it means to learn mathematics. If mathematics education is viewed as the teacher pouring a set of facts into the minds of the students, then perhaps student attitudes are the most important part of the affective domain. But if students are actively engaged in constructing their knowledge of mathematics, rather than just absorbing it, their affective responses will be more intense. If students are active rather than passive learners, their emotions as well as their attitudes will influence their learning. This new view of the learner is already having a substantial impact on paradigms for research on cognitive issues in mathematics learning and teaching (Romberg & Carpenter, 1986). Now it is time for this new view to influence how we approach research on affective issues related to mathematics education.

The need to expand the view of the affective domain is justified by more than current constructivist views of learning. It also results in part from a renewed emphasis on higher-order thinking and problem solving in mathematics. The recommendation from the National Council of Teachers of Mathematics (1980) to make problem solving the central goal of the mathematics curriculum also has implications for affect. Instruction in problem solving generates more intense reactions from students than instruction on more traditional topics. Trying to solve nonroutine problems is often frustrating; drill and practice exercises are generally more boring than frustrating. Posing problems and making conjectures (Brown & Walter, 1983) can provide a sense of joy and accomplishment that is much more intense than what we normally consider to be an attitude toward mathematics.

Further evidence that learning mathematics involves rather intense emotions comes from a variety of research studies. The clinical methodology of these studies provides a rich set of data on student responses to mathematics. These sources suggest that students' affective responses are often more emotional in tone than attitudinal. For example, Buxton (1981) presented a careful analysis of adults' affective responses to mathematics and used the term <u>panic</u> to describe what occurs in the minds of many. This panic is manifested both in chaotic reactions to mathematical tasks and in the tendency of some people to freeze--to be immobilized when asked to solve a problem. Ginsburg and Allardice (1984) noted similar intense reactions to mathematics among elementary school children, even when the mathematics appears to be relatively simple from an adult perspective. At the secondary level, Wagner, Rachlin, and Jensen (1984) reported further evidence along these lines in their study of algebra students; some of these students seemed to lose control of their cognitive processes and grope wildly for an answer, whether or not the answer made sense in terms of the problem they were trying to solve.

It is important to remember that students have positive as well as negative experiences with mathematics; good teachers of problem solving work hard to present students with opportunities for insight and illumination, and students report these experiences as extremely satisfying and even joyous (McLeod, 1985). Although research has tended to concentrate more on the negative emotions (such as frustration and anxiety) rather than the positive, teachers of problem solving know the importance of emphasizing the positive emotions (Mason, Burton, & Stacey, 1982).

In summary, there are at least three reasons to expand our view of the affective domain to include emotions as well as attitudes. The new view of the learner as an active processor of information suggests that the learner will also have more intense affective responses. Changes in the curriculum that emphasize higher-order thinking will result in more intense student reactions. Finally, data from clinical studies suggest that affective responses are more intense than traditional attitude instruments would indicate.

## The Need for Better Theory

As Leder (chapter 23) indicated, there are a variety of theoretical positions that have been used as the basis for research on attitudes. Most of these positions come from a foundation in behavioral psychology or social psychology. They do not, in general, represent positions that are consistent with the dominant paradigm for current research on learning, generally referred to as cognitive psychology or information-processing psychology (Mandler, 1985).

Research on attitudes has in fact often seemed to proceed in rather an atheoretical fashion. A typical approach would be to specify certain factors (e.g., liking, utility, confidence) that are hypothesized to be important in the affective domain and then devise a questionnaire that measures those factors. The researcher would then gather some data, examine the characteristics of the instrument, and apply the appropriate statistical analysis package. The results would then be interpreted and implications drawn for practice, but little thought would be given to the development of a sound theoretical framework. The driving force in much of this research seems to be the statistical methodology rather than the theory.

The researcher in this case seems to assume that the affective domain can be modeled by a vector space and that the questionnaire will span the space and produce factors that describe the space adequately. Current research on cognitive psychology suggests that an alternate mathematical model might build on the notion of a topological space, rather than a vector space, and that the major aspects of interest in this space would involve concepts like connectedness, networks, and other topological properties.

Difficulties with current research on affect have been discussed by many authors. In psychology, Abelson (1976) noted that theories about attitudes are confused and contradictory. Mandler (1972) observed that research on anxiety is generally not cumulative and that researchers have been preoccupied with measurement issues to the neglect of theory. In mathematics education, Kulm (1980) called for better theory to guide research on attitudes toward mathematics, and numerous authors (Begle, 1979; Suydam & Osborne, 1977) have noted the relatively weak relationship between attitudes and achievement in mathematics.

In summary, research on affect in mathematics education lacks a strong theoretical base, and results so far have been relatively weak. If we are to monitor affective factors in school mathematics, we need to establish a stronger theoretical framework that can guide the development of a suitable evaluation system. For a new approach to research on affect, we turn to the work of George Mandler (1975, 1984).

## A New Perspective on Affect

In 1972, after many years of working on both cognitive and affective issues, Mandler expressed his concern with the lack of an acceptable theory for research on anxiety and went on to write a book called Mind and Emotion (Mandler, 1975). His position (refined in Mandler, 1984) is an extension of the theory and methods of cognitive psychology to the affective domain. Although it is not possible to do justice to his theory here, let me briefly describe its essence. Mandler's view is that affective responses result mainly from interruptions of the student's plans or planned actions. Using the terminology of cognitive psychology, the plans come from the activation of schemas, and the schemas induce actions. If these actions are blocked or interrupted, the individual's autonomic nervous system responds with some sign of arousal, such as an increase in heartbeat or a tensing of the muscles. The individual then interprets this reaction of the autonomic nervous system as frustration, surprise, or some other emotion.

The notion of blockages or interruptions is also at the center of what it means to solve a mathematical problem. If there is no block to a student's first attempt at a problem, then there is really no problem for that student, only a routine exercise. Thus it seems that instruction in higher-order thinking and problem solving will be intrinsically more emotional than more traditional kinds of mathematics education.

When a student is interrupted, the interpretation of that interruption is based on the student's knowledge, beliefs, and previous experiences. The interpretation may result in either a positive or a negative emotion. For example, some students interpret the blockage as a challenge and enjoy the opportunity to work on a nontrivial problem. Other students interpret the blockage as a sign that they should get help from the teacher. The students' interpretations reveal a great deal about what they have learned to value in mathematics and about what they believe about their role as mathematics students.

If interruptions generate emotions, then I suggest that repeated interruptions generate attitudes. If a student is regularly faced with interruptions in the same context, then the student's response will become automatic. The role of automaticity is the same in the affective domain as in the cognitive: Human information processing allows certain responses to become more and more automatic, thus freeing the individual's limited processing capacity for action on unfamiliar problems or situations (Resnick & Ford, 1981). These automatic responses seem to be a crucial part of the consistency of attitudes toward mathematics. For a more extensive discussion of automaticity in affective responses to mathematical tasks, see McLeod (1986). For a more detailed exposition of how attitudes develop, see Abelson (1976). Although he uses the terminology of script processing in his definition of attitudes, Abelson's ideas carry over into Mandler's theory quite well.

Now let's move on from attitudes to beliefs. If interruptions generate emotions, then, just as with attitudes, repeated interruptions generate affect-laden beliefs. The development of belief systems from a cognitive perspective has recently been receiving more attention. D'Andrade (1981), for example, discussed how individuals learn about their culture through what is essentially guided discovery. In the case of mathematics education, the responses that the student receives from the surrounding cultural environment provide the guidance in the development of the student's belief system about mathematics. For some concrete examples of how this occurs in mathematics, see Schoenfeld (1985).

This brief discussion suggests that Mandler's (1984) theory could provide the kind of framework that is needed to guide research on affect. Mandler's view is comprehensive and could be used to explain attitudes and beliefs as well as more intense emotional responses to mathematics.

## Implications for Monitoring School Mathematics

If we want to monitor school mathematics, then clearly we need to monitor the affective domain. Leder (chapter 23) has analyzed the issues involved in monitoring attitudes toward mathematics. In this section, I want to suggest some ways to go beyond attitudes and monitor other affective influences on learning.

Although this section will emphasize the monitoring of students' affective responses, the health of school mathematics depends on the affective responses of other groups as well. Teachers, parents, and administrators will all have an influence on students' affective responses to mathematics. Development of indicators for all of these groups seems appropriate.

Although this section will emphasize affective reactions to problem solving in mathematics, there is no intention of slighting other areas of the mathematics curriculum such as the teaching of concepts and procedures. In addition, the increasing importance of the computer in mathematics classrooms suggests that we should pay special attention to technology issues. For a contemporary approach to evaluating affective reactions to computers, see Turkle (1984).

## What Do We Monitor?

If we agree that problem solving is a major goal of the mathematics curriculum, then we should monitor students' affective responses to nonroutine problems as well as to more routine tasks. If students are working on a problem, and their progress on the problem is blocked, what are their reactions? Do they quit? Do they become frustrated and repeat the same unsuccessful attempt at a solution many times? Or do they continue to work and develop more information about the problem, even when frustrated? Can they even see a challenging problem as a positive experience?

How long will students work on a nonroutine problem before giving up? Wertime (1979) suggested the notion of courage span--analogous to attention span--as a way of measuring student willingness to address nonroutine problems. The courage span is the time that a student spends trying to find a way to solve a problem that is unfamiliar to them.

Perhaps more important than the amount of time spent on the problem is the reason for stopping. Do students quit because they have gotten in a rut and want to return to the problem later with fresh ideas? Or do they quit because they assume automatically that any nonroutine problem will be beyond their ability? Do they quit because they are feeling so much emotional stress that they cannot think clearly? Is their limited cognitive capacity totally absorbed in dealing with their emotional reactions, leaving no room in their working memory to deal with the problem?

Along with courage span, one could monitor the "heuristic" span of a problem solver. Students who are unable to manage their emotional reaction to the blockages that are involved in problem solving often use only one heuristic, or one strategy for solving the problem. If the students are given a problem where the goal is clearly specified, and if their strategy is to compute using the numbers in the problem, they often compute over and over until they quit in frustration. If their strategy is to draw a picture, they draw and redraw, waiting for outine solution to appear, but don't think to try investigating a simpler version of the problem or some other strategy. The lack of use of alternate strategies may also be a measure of emotional overload on students' limited processing capacity. The relationship of emotions to the students' metacognitive processing should be a particularly interesting aspect of the monitoring of school mathematics (Garofalo & Lester, 1985).

Another area that requires monitoring is student beliefs. A substantial amount of work of this type has been done, and national assessment data on student beliefs (Carpenter, Corbitt, Kepner, Lindquist, & Reys, 1981) suggest that many of these beliefs have an affective component.

In addition to monitoring students, teachers, and other audiences separately, it would be useful to gather data that involve the interactions of students and teachers in the classroom and the school more generally. The availability of a variety of measures of classroom and school climate, as well as large amounts of extant data gathered from an ethnographic perspective, as in the National Science Foundation case studies (Suydam & Osborne, 1977), should make it possible to assess affective influences in those arenas.

## How Do We Monitor?

Monitoring attitudes has long been a part of school mathematics evaluation and will continue to be appropriate, as Leder (chapter 23) has indicated. However, monitoring affective factors from an information-processing perspective requires a change of methods from the usual assessment of attitudes. Ericsson and Simon (1980) gave elaborate justifications for the usefulness of interview data and for the importance of the density of observations of individuals. Clearly, simple adaptations of attitude questionnaires will not be sufficient. Even transcriptions of verbal data are not enough; students can insist that they hate mathematical problem solving, even when we have just observed them work a nonroutine problem with considerable deftness and obvious enjoyment. So the monitoring should include not only interviews but also observational data on student performance in problem-solving settings that are as realistic as possible.

A variety of researchers have used interviews and observations to obtain data that go beyond the usual measures of cognitive performance. For example, Cobb (1985) reported data related to affective influences on the development of early number concepts. Similarly, Confrey (1984) reported data on beliefs and affect among secondary school students. Related data on teachers were reported by Thompson (1984).

Observations of students should include not only what they say and do but also their physical reactions. Muscle tension and facial expression can tell a great deal about the emotional state of the individual. Many teachers are quite adept at assessing the individual student's emotional condition; it would be interesting to investigate the basis on which those teachers make their assessments.

Although interviews and intensive observations are important, practical considerations suggest that less costly methods be developed that could provide reasonable data on affect. I suggest a modification of the superitem format used in assessing problem-solving ability (Collis, Romberg, & Jurdak, 1986).

Superitems include a stem (a paragraph that specifies the problem situation) and a series of questions about the information in the stem. These questions would normally range over a set of taxonomic levels from lower to higher cognitive levels. For our purposes, the questions should range from cognitive to affective dimensions. Within the affective domain, the questions could range from simple responses regarding attitudes and beliefs to more extensive questioning on the student's emotional states.

If students were assigned a nonroutine problem, their response could include not only attempts to solve the problem, but also thei: emotional reactions at various points in the solution process. Presumably the assessment of their emotional state could be accomplished with minimal disruption to their problem-solving performance. More detailed questions about their emotional reactions to the problem could be attempted at the conclusion of the problem-solving episode. Field tests ~f this procedure would indicate how disruptive it might be to ask students about their emotional state at regular intervals during the solution process. It seems likely that procedures could be developed for large-group administration of superitems that would include assessment of a broad range of affective responses.

We need to develop a variety of ways to assess affective responses of varying intensity. The development of superitems that incorporate questions about the affective domain appears to be a useful strategy. It could be used to assess attitudes and beliefs as well as more intense emotional reactions to mathematical problem solving.

## Summary

Affect plays an important role in the health of school mathematics. Any realistic effort to monitor school mathematics needs to include indicators from the affective domain. The importance of assessing attitudes toward mathematics is well established, in spite of what is currently a relatively weak theoretical foundation for that work. This chapter has suggested that it is possible to develop a stronger theo~etical foundation for the measurement of attitude, and that such a foundation could also support measures of the affective domain that go beyond the usual attitude factors. In particular, the monitoring of school mathematics should pay substantial attention to the emotions that are an integral part of solving nonroutine problems.

## References


Abelson, R. P. (1976). Script processing in attitude formation and decision making. In J. S. Carroll & J. W. Payne (Eds.), Cognition and social behavior (pp. 33-45). Hillsdale, NJ: Erlbaum.

Begle, E. G. (1979). Critical variables in mathematics education: Findings from a survey of the empirical literature. Washington, DC: Mathematical Association of America.

Brown, S. I., & Walter, M. (1983). The art of problem posing. Philadelphia: Franklin Institute Press.

Buxton, L. (1981). Do you panic about maths? London: Heinemann.

Carpenter, T. P., Corbitt, M. A., Kepner, H. S., Lindquist, M. M., & Reys, R. E. (1981). Results from the Second Mathematics Assessment of the National Assessment of Educational Progress. Reston, VA: National Council of Teachers of Mathematics.

Cobb, P. (1985). Two children's anticipations, beliefs, and motivations. Educational Studies in Mathematics, 16, 111-126.

Collis, K. F., Romberg, T. A., & Jurdak, M. E. (1986). A technique for assessing mathematical problem-solving ability. Journal for Research in Mathematics Education, 17, 206-221.

Confrey, J. (1984, April). An examination of the conceptions of mathematics of young women in high school. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

D'Andrade, R. G. (1981). The cultural part of cognition. Cognitive Science, 5, 179-195.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. Psychological Review, 87, 215-251.

Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman Mathematics Attitude Scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. Journal for Research in Mathematics Education, 7, 324-326.

Fennema, E., & Peterson, P. (1985). Autonomous learning behavior: A possible explanation of gender-related differences in mathematics. In L. C. Wilkinson & C. Marrett (Eds.), Gender influences in classroom interaction (pp. 17-35). Orlando: Academic Press.

Foxman, D. D., Martini, R. M., & Mitchell, P. (1982). Mathematical development: Secondary Survey Report No. 3. London: Her Majesty's Stationery Office.

Garofalo, J., & Lester, F. K., Jr. (1985). Metacognition, cognitive monitoring, and mathematical performance. Journal for Research in Mathematics Education, 16, 163-176.

Ginsburg, H. P., & Allardice, B. S. (1984). Children's difficulties with school mathematics. In B. Rogoff & J. Lave (Eds.), Everyday cognition: Its development in social context (pp. 194-219). Cambridge, MA: Harvard University Press.

Henerson, M. E., Morris, L. L., & Fitz-Gibbon, C. T. (1978). How to measure attitudes. Beverly Hills, CA: Sage.

Kulm, G. (1980). Research on mathematics attitude. In R. J. Shumway (Ed.), Research in mathematics education (pp. 356-387). Reston, VA: National Council of Teachers of Mathematics.

Leder, G. C. (1986, April). Gender-linked differences in mathematics learning: Further explorations. Paper presented at the research presession of the annual meeting of the National Council of Teachers of Mathematics, Washington, DC.

Mandler, G. (1972). Helplessness: Theory and research in anxiety. In C. D. Spielberger (Ed.), Anxiety: Current trends in theory and research (pp. 359-374). New York: Academic Press.

Mandler, G. (1975). Mind and emotion. New York: Wiley.

Mandler, G. (1984). Mind and body: Psychology of emotion and stress. New York: Norton.

Mandler, G. (1985). Cognitive psychology: An essay in cognitive science. Hillsdale, NJ: Erlbaum.

Mason, J., Burton, L., & Stacey, K. (1982). Thinking mathematically. London: Addison-Wesley.

McLean, L. D. (1982). Willing but not enthusiastic. Toronto: Ontario Institute for Studies in Education.

McLeod, D. B. (1985). Affective issues in research on teaching mathematical problem solving. In E. A. Silver (Ed.), Teaching and learning mathematical problem solving: Multiple research perspectives (pp. 267-279). Hillsdale, NJ: Erlbaum.

McLeod, D. B. (1986, May). Affective influences on mathematical problem solving. Paper presented at the conference on Affective Influences in Mathematical Problem Solving, San Diego.

National Council of Teachers of Mathematics. (1980). An agenda for action. Reston, VA: Author.

Resnick, L. B., & Ford, W. W. (1981). The psychology of mathematics for instruction. Hillsdale, NJ: Erlbaum.

Reyes, L. H. (1984). Affective variables and mathematics education. Elementary School Journal, 84, 558-581.

Reyes, L. H. (1987, April). Describing the affective domain: Saying what we mean. Paper presented at the research presession of the annual meeting of the National Council of Teachers of Mathematics, Anaheim.

Romberg, T. A., & Carpenter, T. P. (1986). Research on teaching and learning mathematics: Two disciplines of scientific inquiry. In M. Wittrock (Ed.), Third handbook of research on teaching. New York: Macmillan.

Schoenfeld, A. H. (1985). Mathematical problem solving. Orlando: Academic Press.

Silver, E. A. (1985). Research on teaching mathematical problem solving: Some underrepresented themes and needed directions In E. A. Silver (Ed.), Teaching and learning mathematical problem solving: Multiple research perspectives (pp. 247-266). Hillsdale, NJ: Erlbaum.

Simon, H. A. (1982). Comments. In M. S. Clark & S. T. Fiske (Eds.), Affect and cognition: The Seventeenth Annual Carnegie Symposium on Cognition. Hillsdale, NJ: Erlbaum.

Suydam, M. N., & Osborne, A. (1977). The status of pre-college science, mathematics, and social science education: 1955-1975. Vol. 2. Mathematics education. Columbus: Ohio State University.

Thompson, A. G. (1984). The relationships of teachers' conceptions of mathematics and mathematics teaching to instructional practice. Educational Studies in Mathematics, 15, 105-127.

Turkle, S. (1984). The second self: Computers and the human spirit. New York: Simon and Schuster.

Wagner, S., Rachlin, S. L., & Jensen, R. J. (1984). Algebra learning project: Final Report. Athens, GA: University of Georgia.

Weiner, B. (1979). A theory of motivation for some classroom experiences. Journal of Educational Psychology, 71, 3-25.

Wertime, R. (1979). Students, problems, and "courage spans." In J. Lochhead & J. Clement (Eds.), Cognitive process instruction (pp. 191-199). Philadelphia: Franklin Institute Press.