DOCUMENT RESUME

ED 035 545 SE 007 724

AUTHOR Scandura, Joseph M.
TITLE Research in Mathematics Education.
INSTITUTION National Council of Teachers of Mathematics, Inc., Washington, D.C.
PUB DATE 67
NOTE 132p.
AVAILABLE FROM National Council of Teachers of Mathematics, 1201 Sixteenth Street, N.W., Washington, D.C. 20036

EDRS PRICE EDRS Price MF-$0.75 HC Not Available from EDRS.
DESCRIPTORS *Educational Research, *Mathematics Education, *Research, *Research Methodology

ABSTRACT
This publication is a collection of eleven manuscripts which deal with research in mathematics education. Each of the first ten papers is concerned with one or more of the following areas: (1) significant research efforts in mathematics education; (2) the need for both "information-oriented" (basic) and "product-oriented" (applied) research; (3) the complementary nature of basic and applied research; (4) the potential impact of research on the teaching of mathematics; and (5) the implementation of research in the teaching of mathematics. The final paper, written by the editor, focuses on the more important aspects of the preceeding papers and offers a perspective for viewing them. (FL)

# RESEARCH

# MATHEMATICS EDUCATION

# RESEARCH IN MATHEMATICS EDUCATION

National Council of Teachers of Mathematics

1201 Sixteenth Street, N.W.
Washington, D.C. 20036

*Second Printing 1968*

Library of Congress Catalog Card Number: 67-23562

Printed in the United States of America

---

Editor: Joseph M. Scandura, Graduate School of Education
University of Pennsylvania

Editorial Consultants: Kenneth B. Henderson, College of Education
University of Illinois
Harold C. Trimble, Department of Education
The Ohio State University

# Preface

KENNETH B. HENDERSON
JOSEPH M. SCANDURA
HAROLD C. TRIMBLE
*Research Publication Committee*

QUESTIONS about research in education and its special case of research in mathematics education are timely.

For one thing, leading citizens believe in research as never before. They have long since noted that research in the physical and biological sciences has paid off in technological and medical advances that are everywhere evident. Why not give more support to research in education? In answer to this query, Congress has voted more financial assistance for education than ever before in history.

For another thing, many teachers are asking for evidence to support or deny the current crop of claims demanding changes in curriculum and pedagogy. There is a growing feeling that change for the sake of change is suspect. Recommendations for change should be based on research.

Yet many thoughtful people are critical of the quality of research in mathematics education. They look at tables of statistical data and they say "So what!" They feel that vital questions go unanswered while means, standard deviations, and *t*-tests pile up.

What should the *National Council of Teachers of Mathematics* do? Should it help identify questions on which research is needed? Should it serve as a critic of current research? Should it assist in the dissemination of results? Should it sponsor research projects? Should it encourage programs for training research workers to meet the demands that seem to be emerging? These are some of the questions the Research Advisory Committee is discussing.

The purposes of this special publication are conceived as follows: (1) to provide a rationale for both basic and applied research in mathematics education, (2) to exhibit significant research efforts, (3) to clarify the complementary nature of "information-oriented" (basic) and "product-oriented" (applied) research, (4) to demonstrate the potential impact of research and the implementation of research on the teaching of mathematics, and (5) to sample the reactions of members of the profession to a research-oriented journal in mathematics education.

The Curriculum Committee and the Board of Directors of the NCTM approved these purposes as proposed by the Research Advisory Committee. Then, the Board created the Research Publication Committee to get the job done. The task of collecting and editing manuscripts fell to Dr. Joseph M. Scandura.

Surely no two persons, nor even two committees, would come up with the same set of manuscripts. The Research Publication Committee made its own selections, and it does not apologize for its choices. But it wants the reader to think of these papers as samples. In fact, it hopes the Council may want to sponsor further research publications and, perhaps, to create a journal for those of its members who have a special interest in research.

In Paper I, Suppes makes a case for basic research in mathematics education. He views theory construction as an essential guide to data collection. In Papers II, III, and IV, the authors report studies designed to increase understanding about the teaching and learning of mathematics. Gagné is concerned with "The Acquisition of Knowledge" and the importance of prior learning in its acquisition. Dienes introduces "Some Basic Processes Involved in Mathematics Learning" and outlines the results of some of his recent collaborative research with Jeeves. Suppes and Groen describe "Some Counting Models for First-Grade Performance Data on Simple Addition Facts." The phrase "information-oriented" is used to describe these studies, studies which seek information leading to the development of theory about mathematics learning, teaching, and/or curriculum.

Paper V, "A Comparison of Discovery and Expository Sequencing in Elementary Mathematics Instruction," by Worthen, provides an example of basic information-oriented research which also has rather direct implications for classroom practice. In the latter sense, it is "product-oriented." The next paper (VI), "Evaluation of *Experiences in Mathematical Discovery*," by Berger and Howitz, is illustrative of the many problems confronted by the researcher in evaluating a new instructional product, *Experiences in Mathematical Discovery*.

In Papers VII and VIII, the authors describe new technologies, based partially on the analysis described by Gagné, for constructing instructional material and curricula along with the evaluation of sample curricula which were devised using these technologies. Lipson describes his group's efforts to individualize instruction and presents some very interesting results. Kersh's title, "Engineering Instructional Sequences for the Mathematics Classroom," adequately reflects his accomplishment and intent. The phrase "product-oriented" is used to describe these studies. Such research may utilize theory or technology to devise a new process or product and then, almost necessarily, evaluates the process or product with an eye towards its improvement.

In Paper IX, Becker and McLeod summarize the research over the past 75 years on "Teaching, Discovery, and the Problems of Transfer of Training in Mathematics." Then, Holtan reports a sampling of current activities in, and concerns about, mathematics education research in Paper X. In the last paper (XI), the editor points up some of the highlights of the earlier papers while attempting to provide a perspective in which they might be viewed. Finally, we wish to acknowledge the efforts of several other authors whose excellent manuscripts could not be printed due to space limitations.

The Research Advisory Committee hopes you will read this publication and find in it some helpful ideas.

# Contents

# I

# The Case for Information-oriented (Basic) Research in Mathematics Education

PATRICK SUPPES
*Stanford University*
*Stanford, California*

THE marvelously clear and definite structure that is characteristic of most parts of modern mathematics can be misleading when problems of mathematical instruction are considered. The very clarity of the structure of mathematics itself can lead to the mistaken view that nothing beyond this structure need be considered in analyzing and deciding how mathematics should be taught.

Yet anybody who has taught mathematics knows how far from the truth this claim is. It is not a straightforward or simple matter for the average student to learn mathematics! And there is no doubt that the ordinary student finds that he has to think harder in learning mathematics than in learning just about any other subject in the curriculum.

The case for basic research in mathematics education can be stated quite simply in terms of these well-known difficulties of students. It is the ultimate objective of basic research in mathematics education to understand how students learn mathematics, and to use this understanding to outline more effective ways of organizing the curriculum. It is probably also agreed, on all sides, that we are still very far from realizing this objective. Without question, we do not yet understand in any reasonable degree of scientific detail what goes on when a student learns a piece of mathematics, whether the mathematics in question be first-grade arithmetic, undergraduate calculus, or graduate-school algebraic topology.

In this brief article I want to survey some of the more important reasons for having a vigorous program in basic research in mathematics education.

## DEFECTS OF INTUITION

Many teachers, who would admit that the logical structure of mathematics alone is not sufficient to determine the mathematics curriculum and how it is to be presented to students, would still maintain that the remaining gaps can be closed by appropriate use of intuition.

The first puzzling thing about this claim for intuition is that most of us have only a vague idea of what another person means when he talks about knowing something by intuition. What is intuition? We all recognize the role of experience in the training of teachers. As a rule, the teacher who has taught several years is able to do a better job than the beginner. Intuition is involved—intuition as the acquisition of knowledge and information in an inexplicit and nonformalized way on the basis of teaching experience. No one faced with the complex problems of teaching mathematics or any other part of the curriculum would want to belittle the importance of experience and practice in the training of good teachers.

Yet many examples exist in the mathematics curriculum to show that it is not sufficient to leave the curriculum to the intuition of curriculum writers and the experience of teachers. The extensive research by Brownell and others on methods of subtraction has made everyone dealing with the curriculum in arithmetic sensitive to the analysis of the actual steps that must be taught children in learning the subtraction algorithm. Another example is the evidence that in the learning of a sequence of mathematical concepts, the important problem is often to minimize negative transfer rather than to facilitate positive transfer. The existence of negative transfer in passing from one concept to another is the sort of thing that is noticed by the very good teacher; it is also the kind of phenomenon that needs to be pinned down, in terms of research, and made part of the objective evidence presented to all teachers in telling them about learning difficulties. Another example that goes contrary to the formal structure of our standard teaching of geometry is found in the clear results concerning children's perceptions of rotations and stretches of standard geometrical figures in the plane. Although Euclidean geometry uses the fundamental notion of congruence that is invariant under rotations of figures, but not under stretches in their size, at the perceptual level this notion of congruence is more difficult for young children than perceiving the relation of similarity between figures that have the same orientation and shape but different sizes. Because teachers have themselves been taught Euclidean geometry and are familiar with the concept of congruence, it is all too easy for them to infer that this is the more natural concept for children. Without supporting research, it would be difficult to convince many teachers of the true state of affairs.

## Defects of Sheer Empiricism

It is also important to emphasize, in discussing the role of basic research in mathematics education, that simple applied empirical research will not answer all the many questions that confront us. For example, if we hope to determine by experimental research the optimum sequence of topics in the first two grades of elementary school (or, with equal pertinence, in the first two years of university mathematics), it is easy enough to show for either of these cases that the mathematical constraints that are placed on the possible sequences of topics are not sufficient to reduce the number of *possible* sequences of concepts to a manageable number of experiments. The number would be greater than all persons now working in mathematics education could perform in the next ten or fifteen years, even if they devoted themselves wholly to this question. The sort of mathematical constraint I have in mind is that the introduction of multiplication would, from a mathematical standpoint, have to be preceded by the introduction of addition, if multiplication is initially to be talked about in terms of repeated addition. On the other hand, there is no real reason why we could not experiment with the introduction of subtraction before addition.

Examples of a more practical nature center around questions of the following sort. Should addition and subtraction be introduced simultaneously? If not, should addition be carried to sums not greater than five, not greater than six, not greater than seven, etc., before subtraction (or at least the notation for subtraction) is introduced? Such purely empirical questions are endless in number, and I emphasize once again, there is no purely mathematical answer to them. Because there is no purely mathematical answer, the importance of a psychological theory of mathematics-learning is crucial, in order ultimately to provide appropriate answers to problems of curriculum organization.

Another way of putting the matter is that purely empirical research lacks conceptual power, because the absence of any theory prohibits us from making extensive generalizations to other situations and broader classes of problems.

From this standpoint, I would emphasize that the demands for a psychological theory of mathematics-learning, and thus for theoretical basic research as well as empirical basic research, are practical demands. Without such theory it is impossible for us to answer in any scientific way many substantive questions of curriculum organization. The vast literature on readiness, drills, practice, and overlearning in arithmetic and other subjects has made all of us aware of the complex and subtle nature of the empirical problems. Anyone who thinks that he can answer these problems either by intuition or by any simple experimental program, without facing the theoretical problems of weaving into one coherent theoretical pattern the many kinds of results already obtained, is surely daydreaming.

In this discussion of empirical problems I have emphasized the kind of questions that have arisen in elementary-school mathematics. The reason for this is simply that a greater body of research already exists in this area. The problems of mathematics-learning at the university level are certainly more complex and difficult, and may demand even more of an effort in basic research in order to begin to understand them.

### Analysis of Learning Difficulties

Given a particular organization of the curriculum in terms of the concepts to be taught and the sequence in which these concepts will be presented, it is still a major task of basic research to analyze and provide a theory for the kind of learning difficulties students encounter as they progress through this curriculum. It is again important to emphasize that the learning difficulties students encounter cannot be predicted by a nonpsychological mathematical analysis of the mathematical content of the curriculum itself—at least no one has proposed such a theory, and there are good reasons for thinking that no such theory shall be proposed.

It is not a part of arithmetic proper or of geometry proper to make psychological predictions about the difficulties students will have with the different concepts in these disciplines. It is the task of a psychological theory of mathematics-learning to predict and to offer an analysis of the kinds of difficulties that are encountered. The success of mathematics teaching depends upon understanding and providing successful practical remedies for the difficulties that students do encounter. In our increasingly technological age it is of greater importance than ever before that we, as educators, recognize the need for clear analysis of students' learning difficulties and the pressing need to develop theories that adequately deal with these difficulties. I have tried to emphasize in this brief discussion that neither intuition nor sheer empiricism is able to provide adequate answers to our problems. I have rested the case for basic research on the overwhelming practical importance of the solutions one hopes to find. I would like to conclude with some remarks in a somewhat different direction.

## Psychology of Learning and the Nature of Mathematics

It is my own conjecture that as we are able to dig deeper into the development of an adequate psychological theory of mathematics-learning, the results will have an impact on our conception of the nature of mathematics itself.

It is not possible here to defend this conjecture in a detailed way, but there is reason to think that concentration on mathematical thinking and the difficulties students have in learning to think mathematically will lead to a new conception of *invariance,* a conception that goes beyond that now encountered in the various parts of mathematics. Historically, the standard philosophies of mathematics have emphasized differing attitudes toward the nature of mathematical objects, but it is perfectly obvious that in most domains of mathematics the exact nature of the mathematical objects studied is not essential. What is of more central concern are the patterns of thought applied by mathematicians in reaching new results, or by students in finding for themselves solutions of problems or proofs of known theorems.

As yet, theories of learning have little to offer in providing insight into how one learns to think mathematically. The nature of abstraction, or the processes of imagery and association that are surely essential to thinking in any domain of mathematics, have as yet scarcely been studied from a scientific standpoint.

Like mathematics itself, research in mathematics education will necessarily have both basic and applied components. Research that is concerned with particular pieces of curriculum and particular learning difficulties of students will continue to occupy a major portion of research efforts, but it is also to be hoped that the kind of problems I have just been mentioning, problems that represent fundamental puzzles about the nature of human thinking, will come to occupy a larger place in research about mathematics learning.

# II

# The Acquisition of Knowledge*[1]

ROBERT M. GAGNÉ
*University of California*
*Berkeley, California*

THE growing interest in autoinstructional devices and their component learning programs has had the effect of focusing attention on what may be called "productive learning." By this phrase is meant the kind of change in human behavior which permits the individual to perform successfully on an entire *class* of specific tasks, rather than simply on one member of the class. Self-instructional programs are designed to ensure the acquisition of capabilities of performing classes of tasks implied by names like "binary numbers," "musical notation," and "solving linear equations," rather than tasks requiring the reproduction of particular responses.

When viewed in this manner, learning-programming is not seen simply as a technological development incorporating previously established learning principles, but rather as one particular form of the ordering of stimulus-and-response events designed to bring about productive learning. It should be possible to study such learning, and the conditions which affect it, by the use of any of a variety of teaching machines, although there are few studies of this sort in the current literature (cf. Lumsdaine and Glaser, 1960). In the laboratory, the usual form taken by studies of productive learning has been primarily that of the effects of instructions and pretraining on problem solving (e.g., Hilgard, Irvine, and Whipple, 1953; Katona, 1940; Maltzman *et al.*, 1956).

When an individual is subjected to the situation represented by a learning program, his performance may change, and the experimenter

* This article originally appeared in *Psychological Review*, LIX (1962), 355-65, and is reprinted here with the kind permission of the author and the American Psychological Association.

[1] This study was made possible in part by funds granted by the Carnegie Corporation of New York. The opinions expressed are those of the author, and do not necessarily reflect the views of that corporation.

then infers that he has acquired a new capability. It would not be adequate to say merely that he has acquired new "responses," since one cannot identify the specific responses involved. (Adding fractions, for example, could be represented by any of an infinite number of distinguishable stimulus situations, and an equal number of responses.) Since we need to have a term by means of which to refer to what is acquired as a result of responding correctly to a learning program, we may as well use the term "knowledge." By definition, "knowledge" is that inferred capability which makes possible the successful performance of a *class of tasks* that could not be performed before the learning was undertaken.

*Some initial observations*

In a previous study of programmed learning (Gagné and Brown, 1961) several kinds of learning programs were used in the attempt to establish the performance, in high school boys, of deriving formulas for the sum of *n* terms in a number series. Additional observations with this material led us to the following formulation: In productive learning, we are dealing with two major categories of variables. The first of these is knowledge, that is, the capabilities the individual possesses at any given stage in the learning; while the second is instructions, the content of the communications presented within the frames of a learning program.

In considering further the knowledge category, it has been found possible to identify this class of variable more comprehensively in the following way: Beginning with the final task, the question is asked, what kind of capability would an individual have to possess if he were able to perform this task successfully, were we to give him only instructions? The answer to this question, it turns out, identifies a new class of task which appears to have several important characteristics. Although it is conceived as an internal "disposition," it is directly measurable as a performance. Yet it is *not the same* performance as the final task from which it was derived. It is in some sense *simpler*, and it is also more *general*. In other words, it appears that what we have defined by this procedure is an entity of "subordinate knowledge" which is essential to the performance of the more specific final task.

Having done this, it was natural to think next of repeating the procedure with this newly defined entity (task). What would the individual have to know in order to be capable of doing *this* task without undertaking any learning, but given only some instructions? This time it seemed evident that there were two entities of subordinate knowledge which combined in support of the task. Continuing to follow this procedure,

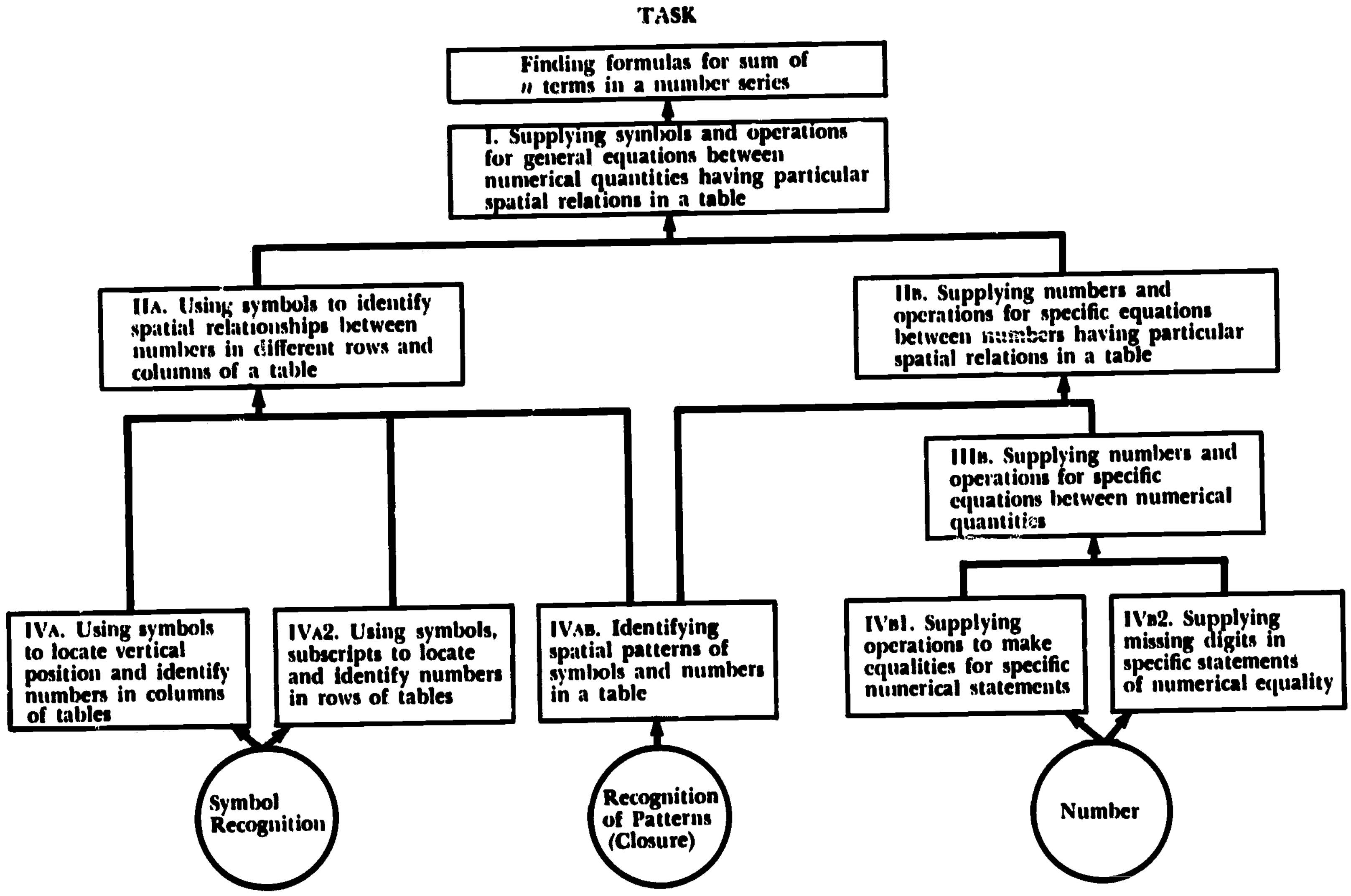


FIGURE 1.—HIERARCHY OF KNOWLEDGE FOR THE TASK OF FINDING FORMULAS FOR THE SUM OF $n$ TERMS IN A NUMBER SERIES

we found that what we were defining was a *hierarchy* of subordinate knowledges, growing increasingly "simple," and at the same time increasingly general as the defining process continued.

By means of this systematic analysis, it was possible to identify nine separate entities of subordinate knowledge, arranged in hierarchical fashion (see Fig. 1). Generally stated, our hypothesis was that (*a*) no individual could perform the final task without having these subordinate capabilities (i.e., without being able to perform these simpler and more general tasks); and (*b*) that any superordinate task in the hierarchy could be performed by an individual provided suitable instructions were given, and provided the relevant subordinate knowledges could be recalled by him.

It may be noted that there are some possible resemblances between the entities of such a knowledge hierarchy and the hypothetical constructs described by three other writers. First are the habit-family hierarchies of Maltzman (1955), which are conceived to mediate problem solving, and are aroused by instructions (Maltzman *et al.*, 1956). The second are the "organizations" proposed by Katona (1940), which are considered to be combined by the learner into new knowledge after receiving certain kinds of instructions, without repetitive practice. The third is Harlow's (1949) concept of learning set. Harlow's monkeys acquired a general capability of successfully performing a class of tasks, such as oddity problems, and accordingly are said to have acquired a learning set. There is also the suggestion in one of Harlow's (Harlow and Harlow, 1949) reports that there may be a hierarchical arrangement of tasks more complex than oddity problems which monkeys can successfully perform. Since we think it important to imply a continuity between the relatively complex performances described here and the simpler ones performed by monkeys, we are inclined to refer to these subordinate capabilities as "learning sets."

## Requirements of Theory

If there is to be a theory of productive learning, it evidently must deal with the independent variables that can be identified in the two major categories of instructions and subordinate capabilities, as well as with their interactions, in bringing about changes in human performance.

### *Instructions*

Within a learning program, instructions generally take the form of sentences which communicate something to the learner. It seems possible to think of such "communication" as being carried out with animals lower than man, by means of quite a different set of experimental operations. Because of these communications, the human learner progresses from a point in the learning sequence at which he can perform one set of tasks to a point at which he achieves, for the first time, a higher level learning set (class of tasks). What functions must a theory of knowledge acquisition account for, if it is to encompass the effects of instructions? The following paragraphs will attempt to describe these functions, not necessarily in order of importance.

First, instructions make it possible for the learner to *identify the required terminal performance* (for any given learning set). In educational terms, it might be said that they "define the goal." For example, if the task is adding fractions, it may be necessary for the learner to identify $15\frac{3}{4}$ as an adequate answer, and $\frac{63}{4}$ as an inadequate one.

Second, instructions bring about proper *identifications of the elements* of the stimulus situation. For example, suppose that problems are to be presented using the word "fraction." The learner must be able to identify $\frac{3}{8}$ as a fraction and .4 as not a fraction. Or, he may have to identify $\Sigma$ as "sum of," and $n$ as "number." Usually, instructions establish such identifications in a very few repetitions, and sometimes in a single trial. If there are many of them, differentiation may require several repetitions involving contrasting feedback for right and wrong responses.

A third function of instructions is to establish *high recallability* of learning sets. The most obviously manipulable way to do this is by repetition. However, it should be noted that repetition has a particular meaning in this context. It is not exact repetition of a stimulus situation (as in reproductive learning), but rather the presentation of additional examples of a class of tasks. Typically, within a learning program, a task representing a particular learning set is achieved once, for the first time. This may then be followed by instructions which present one or more additional examples of this same class of task. "Variety" in such repetition (meaning variety in the stimulus context) may be an important subvariable in affecting recallability. Instructions having the function of establishing high recallability for learning sets may demand "recall," as in the instances cited, or they may on other occasions attempt to achieve this effect by "recognition" (i.e., not requiring the learner to produce an answer).

The fourth function of instructions is perhaps the most interesting from the standpoint of the questions it raises for research. This is the "guidance of thinking," concerning whose operation there is only a small amount of evidence (cf. Duncan, 1959). Once the subordinate learning sets have been recalled, instructions are used to promote their application to (or perhaps "integration into") the performance of a task that is entirely new so far as the learner is concerned. At a minimum, this function of instructions may be provided by a statement like "Now put these ideas together to solve this problem"; possibly this amounts to an attempt to establish a *set*. Beyond this, thinking may be guided by suggestions which progressively limit the range of hypotheses entertained by the learner, in such a way as to decrease the number of incorrect solutions he considers (cf. Gagné and Brown, 1961; Katona, 1940). Within a typical learning program, guidance of thinking is employed after identification of terminal performance and of stimulus elements have been completed, and after high recallability of relevant learning sets has been ensured. In common sense terms, the purpose of these instructions is to suggest to the learner "how to approach the solution of a new task" without, however, "telling him the answer."

Obviously, much more is needed to be known about the effects of this variable, if indeed it is a single variable. Initially, it might be noted that guidance of thinking can vary in *amount;* that is, one can design a set of instructions which say no more than "now do this new task" (a minimal amount); or, at the other end of the scale, a set of instructions which in effect suggest a step-by-step procedure for using previously acquired learning sets in a new situation.

### *Subordinate capabilities: learning sets*

When one begins with the performance of a particular class of tasks as a criterion of terminal behavior, it is possible to identify the subordinate learning sets required by means of the procedure previously described. The question may be stated more exactly as, "What would the individual have to be able to do in order that he can attain successful performance on this task, provided he is given only instructions?" This question is then applied successively to the subordinate classes of tasks identified by the answer. "What he would have to be able to do" is in each case one or more performances which constitute the denotative definitions of learning sets for particular classes of tasks, and totally for the entire knowledge hierarchy.

A theory of knowledge acquisition must propose some manner of functioning for the learning sets in a hierarchy. A good possibility seems to be that they are mediators of positive transfer from lower-level learning sets to higher-level tasks. The hypothesis is proposed that specific transfer from one learning set to another standing above it in the heirarchy will be zero if the lower one cannot be recalled, and will range up to 100 percent if it can be.

In narrative form, the action of the two classes of variables in the acquisition of knowledge is conceived in the following way. A human learner begins the acquisition of the capability of performing a particular class of tasks with an individual array of relevant learning sets, previously acquired. He then acquires new learning sets at progressively higher levels of the knowledge heirarchy until the final class of tasks is achieved. Attaining each new learning set depends upon a process of positive transfer, which is dependent upon (*a*) the recall of relevant subordinate learning sets, and upon (*b*) the effects of instructions.

### Experimental Predictions and Results [2]

Using the procedure described, we derived the knowledge heirarchy depicted in Figure 1 for the final task of "deriving formulas for the sum of *n* terms in number series."

As mentioned previously, it contained nine hypothesized learning sets. (The final row of circled entities will be discussed later.) Each of these subordinate knowledges can be represented as a class of task to be performed.

*Measuring initial patterns of learning sets*

It is predicted that the presence of different patterns of learning sets can be determined for individuals who are unable to perform a final task such as the one under consideration. To test this, we administered a series of test items to a number of ninth-grade boys. These items were presented on 4"-by-6" cards, and the answers were written on specially prepared answer sheets. This particular method was used in order to make testing continuous with the administration of a learning program to be described hereafter. Each test item was carefully prepared to include instructions having the function of identification of terminal performance and of elements of the stimulus situation.

Beginning with the final task, the items were arranged to be presented in the order I, IIA, IVA, IVA2, IVAB, IIB, IIIB, IVB1, and IVB2. For any given subject, the sequence of testing temporarily stopped at the level at which successful performance was first reached, and a learning program designed to foster achievement at the next higher level (previously failed) was administered. This program and its results will shortly be described. Following this, testing on the remaining learning set tasks was undertaken in the order given. The possibility of effects of the learning program on the performance of these lower-level learning sets

[2] The author is grateful to Bert Zippel, Jr., for assistance in the preparation of learning program materials and in the collection of a portion of the data.

TABLE 1
PATTERN OF SUCCESS ON LEARNING SET TASKS RELATED TO THE FINAL NUMBER SERIES TASK FOR SEVEN NINTH-GRADE BOYS

| SUBJECT | TASK | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Final | I | IIA | IIB | IIIB | IVA1 | IVA2 | IVAB | IVB1 | IVB2 |
| WW | − | + | + | + | + | + | + | + | + | + |
| WC | − | + | + | + | + | + | + | + | + | + |
| PM | − | − | − | + | + | + | + | + | + | + |
| GR | − | − | − | + | + | + | + | + | + | + |
| DJ | − | − | − | − | + | + | + | + | + | + |
| JR | − | − | − | − | − | + | + | + | + | + |
| RH | − | − | − | − | − | + | + | + | + | + |

Note that + = Pass; — = Fail.

(not specifically practiced in the learning) is of course recognized, but not further considered in the present discussion.

A particular time limit was set for each test item, at the expiration of which the item was scored as failed. If a wrong answer was given before this time limit, the subject was told it was wrong, and encouraged to try again; if the correct answer was supplied within the time limit, the item was scored as passed. It is emphasized that these time limits, which were based on preliminary observations on other subjects with these tasks, were *not* designed to put "time pressure" on the subjects, nor did they appear to do so.

The patterns of success achieved on the final task and all subordinate learning set tasks, by all seven subjects, are shown in Table 1. The subjects have been arranged in accordance with their degree of success with all tasks, beginning with one who failed the final task but succeeded at all the rest. Several things are apparent from these data. First of all, it is quite evident that there are quite different "patterns of capability" with which individuals approach the task set by the study. Some are unable to do a task like IIA (see Fig. 1), others to do a task like IIB, which is of course quite different. Still others are unable to do either of these, and in fact cannot perform successfully a task like IIIB. All seven of these subjects were able to perform IV-level tasks successfully, although in preliminary observations on similar tasks we found some ninth-grade boys who could not.

Second, the patterns of pass and fail on these tasks have the relationships predicted by the previous discussion. There are no instances, for example, of an individual who is able to perform what has been identified as a "higher-level" learning set, and who then shows himself to be unable to perform a "lower-level" learning set related to it.

If learning sets are indeed essential for positive transfer, the following consequences should ensue:

1. If a higher-level learning set is passed (+), *all* related lower-level tasks must have been passed (+).

2. If *one or more* lower-level tasks have been failed (—), the related higher-level tasks must be failed (—).

3. If a higher-level task is passed (+), *no* related lower-level tasks must have been failed (—).

4. If a higher-level task has been failed (—), related lower-level tasks may have been passed (+). The absence of positive transfer in this case would be attributable to a deficiency in instructions, and does not contradict the notion that lower-level sets are essential to the achievement of higher-level ones.

The relationships found to exist in these seven subjects are summarized in Table 2, where each possible higher-lower-level task relationship possible of testing is listed in the left-hand column. It will be noted that there are several relationships of the type higher (—), lower (+), as listed in Column 5. These provided no test of the hypothesis regarding hierarchical relations among learning sets. The instances in the remaining columns do, however. The ++ and —— instances are verifying, whereas +— instances would be nonverifying. As the final column indicates, the percentage of verifying instances is in all cases 100 percent.

TABLE 2
PASS-FAIL RELATIONSHIP BETWEEN RELATED ADJACENT HIGHER- AND LOWER-LEVEL LEARNING SETS FOR A GROUP OF SEVEN NINTH-GRADE BOYS

| RELATIONSHIP EXAMINED | NUMBER OF CASES WITH RELATIONSHIP | | | | TEST OF RELATIONSHIPS | |
|---|---|---|---|---|---|---|
| | Higher + Lower + | Higher — Lower — | Higher + Lower — | Higher — Lower + | N (1 + 2 + 3) | Proportion $\frac{(1+2)}{(1+2+3)}$ |
| Final Task: I | 0 | 6 | 0 | 1 | 6 | 1.00 |
| I: IIA, IIB | 1 | 5 | 0 | 1 | 6 | 1.00 |
| IIA: IVA1, IVA2, IVAB | 2 | 0 | 0 | 5 | 2 | 1.00 |
| IIB: IIIB | 3 | 2 | 0 | 2 | 5 | 1.00 |
| IIIB: IVAB, IVB1, IVB2 | 5 | 0 | 0 | 2 | 5 | 1.00 |

Note that + = Pass; — = Fail.

*Effects of learning program administration*

If the characteristics of instructions as previously described are correct, it should be possible to construct a learning program which can be begun for each individual at the point of his lowest successful learning set achievement, and bring him to successful achievement of the final class of tasks. Briefly, its method should be to include frames which have the functions of (*a*) insuring high recallability of relevant learning sets on which achievement has been demonstrated; (*b*) making possible identifications of expected performance and of new stimuli, for each newly presented task; and (*c*) guiding thinking so as to suggest proper directions for hypotheses associating subordinate learning sets with each new one.

A program of this sort was administered to each of the seven ninth-grade boys, beginning at the level at which he first attained success on learning set tasks (Table 1). This was done by means of a simple teaching machine consisting of a visible card file clipped to a board mounted at a 40° angle to the learner's table, and containing material typed on 4″-by-6″ cards. He wrote his answer to successive frames on a numbered answer sheet, then flipped over the card to see the correct answer on the back. He was instructed that if his answer was wrong, he should flip the card back, and read the frame again until he could "see" what the right answer was.

After completing the instructional portion of the program for each learning set, the learner was again presented with the identical test-item problem he had tried previously and failed. If he was now able to do it correctly, he was given five additional items of the same sort to perform, and then taken on by instructions to another learning set in either a coordinate or higher-level position in the hierarchy. This process was continued through the performance of the final task.

The data collected in this way yield pass-fail scores on each test item (representing a particular member of a class of tasks) *before* the administration of the learning program, and similar scores on the same item *after* learning. It is recognized that for certain experimental purposes, one would wish to have a different, matched, task for the test given after learning, to control for the effects of "acquaintance" during the first test. Since this study had an exploratory character, such a control was not used this time. However, it should be clearly understood that the first experience with these test items in question, for these subjects, involved only activity terminating in failure to achieve solution. No information about the correct solutions was given.

A striking number of instances of success in achieving correct solutions

TABLE 3
NUMBER OF INSTANCES OF PASSING AND FAILING FINAL TASK AND SUBORDINATE LEARNING SET TASKS BEFORE AND AFTER ADMINISTRATION OF AN ADAPTIVE LEARNING PROGRAM, IN A GROUP OF SEVEN NINTH-GRADE BOYS

| TASK | NUMBER FAILING BEFORE LEARNING | NUMBER OF THESE PASSING AFTER LEARNING | PERCENTAGE SUCCESS |
|---|---|---|---|
| Final Task | 7 | 6 | 86 |
| I | 5 | 4 | 80 |
| II | 5 | 5 | 100 |
| IIB | 2 | 1 | 50 |
| IIIB | 2 | 2 | 100 |
| Total | 21 | 18 | 86 |

to learning set tasks was found following learning as compared with before. These results are summarized in Table 3. Although for learning set IIB the percentage of success was only 50 percent (with two cases), there were two learning sets for which 100 percent success was achieved, and the percentage for all instances combined was 86 percent. These results provide additional evidence compatible with the idea of the knowledge hierarchy.

The learner in such a program does not "practice the final task"; he acquires specifically identified capabilities in a specified order. In as many as six out of seven cases, we were able by this means to bring learners from various levels of competence all the way to final task achievement. (It is perhaps important that the exception was JR, one of two who had most to learn). Of course, it must be recognized that two separable causes contribute to the effects of the learning program in this study: (*a*) the correctness of the learning set analysis; and (*b*) the specific effectiveness of the instructions contained in the learning program.

### IMPLICATIONS FOR INDIVIDUAL DIFFERENCES MEASUREMENT

It is evident that learning sets, as conceived in this paper, operate as "individual differences" variables, which, when suitably manipulated, also become "experimental" variables. There are some additional implications which need to be pointed out regarding the functioning of learning sets in the determination of measured individual differences.

As the process of identification of subordinate learning sets is progressively continued, one arrives at some learning sets which are very simple and general, and likely to be widespread within the population of learners for which the task is designed. Consider, for example, learning set IVB

(Fig. 1), which is represented by a task such as $4 \times 2 = 5 + ?$ If one makes a further analysis to identify a subordinate learning set for this task, the answer appears to be "adding, subtracting, multiplying, and dividing one- and two-place numbers." It is interesting to note that this is exactly the task provided by a set of factor reference tests (French, 1954) called Number. In a similar manner, the other two circled entities in the last row of Figure 1 were identified. One is Symbol Recognition (called Associative Memory by the factor researchers), and another is Recognition of Patterns (called Flexibility of Closure). The implication is, therefore, that these simplest tasks, identified by factor analysis techniques as common to a great many human performances, also function as learning sets.

The hypothesis has been proposed that learning sets mediate positive transfer to higher-level tasks. Very often, if not usually, the measurement of transfer of training implies that a second task is learned *more* rapidly when preceded by the learning of an initial task than when not so preceded. Accordingly, it seems necessary to distinguish between expected correlations of these basic factors (at the bottom of the heirarchy) with rate of attainment of higher-level learning sets on the one hand, and correlations of these same factors with achievement of higher-level learning sets on the other.

The implications of this line of reasoning would seem to be somewhat as follows: Factors which are found by the kind of psychological analysis previously described to lie at the bottom of the knowledge hierarchy should exhibit certain predictable patterns of correlation with higher-level learning sets. They should correlate most highly with rate of attainment of the learning sets in the next higher level to which they are related, and progressively less as one progresses upwards in the hierarchy. The reason for this is simply that the rate of attainment of learning sets in a hierarchy comes to depend to an increasing extent on the learning sets which have just previously been acquired and accordingly to a decreasing extent upon a basic factor or ability. Some analogy may be drawn here with the findings of Fleishman and Hempel (1954) on motor tasks.

The expected relationships between factor test scores and achievement scores (passing or failing learning sets) throughout such hierarchies seem to require a somewhat more complex derivation. First of all, such relationships will depend upon the effectiveness of a learning program, or perhaps on the effectiveness of previous learning. If the learning program is perfectly effective, for example, and if differences in rate of attainment are ignored, everyone will pass all the learning set tasks, and the variance will accordingly be reduced to zero. Under these circumstances, then, one may expect all correlations with basic factors to be zero. However, one must consider the case in which the learning program is not perfectly effective. In such a case, the probability that an individual will acquire a new learning set, as opposed to not acquiring it, will presumably be increased to the extent that he scores high on tests of related basic abilities. If one continues to collect scores on learning set tasks of both successful achievers and those who fail, the result will presumably be an increasing degree of correlation between basic ability scores and learning set tasks as one progresses upwards in the hierarchy. The reason for this is that the size of the correlation comes to depend more and more upon variance contributed by those individuals who are successful, and less and less on that contributed by those who effectively "drop out."

The difference in expectation between the increasing pattern of correlation with achievement scores, and the decreasing pattern with measures of rate of attainment, is considered to be of rather general importance for the area of individual differences measurement. Confirmatory results have been obtained in a recent study (Gagné and Paradise, 1961) concerned with the class of tasks "solving linear algebraic equations."

## Discussion

The general view of productive learning implied in this paper is that it is a matter of transfer of training from component learning sets to a new activity which incorporates these previously acquired capabilities. This new activity so produced is qualitatively different from the tasks which correspond to the "old" learning sets; that is, it must be described by a different set of operations, rather than simply being "more difficult." The characteristics of tasks which make achievement of one class of task the required precursor of achievement in another, and not vice versa, are yet to be discovered. Sufficient examples exist of this phenomenon to convince one of its reality (Gagné, *et al.*, 1962; Gagné and Paradise, 1961). What remains to be done, presumably, is to begin with extremely simple levels of task, such as discriminations, and investigate transfer of training to tasks of greater and greater degrees of complexity, or perhaps abstractness, thus determining the dimensions which make transfer possible.

The path to research on the characteristics of instructions appears more straightforward, at least at first glance. The establishment of identifications is a matter which has been investigated extensively with the use of paired associates. The employment of instructions for this purpose may need to take into consideration the necessity for learning differentiations among the stimulus items to be identified, as well as other variables suggested by verbal learning studies. The function of inducing high recallability would seem to be a matter related to repetition of learning set tasks, and may in addition be related to time variables such as those involved in distribution of practice. As for guidance of thinking, the distinguishing of this function from others performed by instructions should at least make possible the design of more highly analytical studies than have been possible in the past.

In the meantime, the approach employed in the experiment reported here, of proceeding backwards by analysis of an already existing task, has much to recommend it as a way of understanding the learning of school subjects like mathematics and science, and perhaps others also. Naturally, every human task yields a different hierarchy of learning sets when this method of analysis is applied. Often, the relationship of higher to lower learning sets is more complex than that exhibited in Figure 1. It should be possible, beginning with any existing class of tasks, to investigate the effects of various instruction variables within the framework of suitably designed learning programs.

The major methodological implication of this paper is to the effect that investigations of productive learning must deal intensively with the kinds of variables usually classified as "individual differences." One cannot depend upon a measurement of *general* proficiency or aptitude to reveal much of the important variability in the capabilities people bring with them to a given task. Consider, for example, the seven ninth-grade boys in our study. Each of them had "had" algebra, and each of them had "had" arithmetic. There was no particularly striking relationship between their ultimate performance and their previous grades in algebra (although there is no doubt some correlation), nor between this performance and "general intelligence." But the measurement of their learning sets, as illustrated in Table 1, revealed a great deal about how they would behave when confronted with the learning program and the final task. For some, instructions had to begin, in effect, "lower down" than for others. Some could do Task I right away, while others could not, but could do it equally well provided they learned other things first. The methodological point is simply this: if one wants to investigate the effects of an experimental treatment on the behavior of individuals or groups who start from the same point, he would be well advised to measure and map out for each individual the learning sets relevant to the experimental task. In this way he can have some assurance of the extent to which his subjects are equivalent.

## REFERENCES

DUNCAN, C. P. "Recent Research on Human Problem Solving," *Psychological Bulletin*, LVI (1959), 397-429.

FLEISHMAN, E. A., and HEMPEL, W. E. "Changes in Factor Structure of a Complex Psychomotor Test as a Function of Practice," *Psychometrika*, XIX (1954), 239-52.

FRENCH, J. W. (ed.). *Kit of Selected Tests for Reference Aptitude and Achievement Factors*. Princeton, N. J.: Educational Testing Service, 1954.

GAGNÉ, R. M., and BROWN, L. T. "Some Factors in the Programming of Conceptual Learning." *Journal of Experimental Psychology*, LXII (1961), 313-21.

GAGNÉ, R. M., and PARADISE, N. E. "Abilities and Learning Sets in Knowledge Acquisition," *Psychological Monographs*, LXXV, Whole No. 518 (1961).

GAGNÉ, R. M., *et al.* "Factors in Acquiring Knowledge of a Mathematical Task," *Psychological Monographs*, LXXVI, Whole No. 526 (1962).

HARLOW, H. F. "The Formation of Learning Sets," *Psychological Review*, LVI (1949), 51-65.

HARLOW, H. F., and HARLOW, M. K. "Learning to Think," *Scientific American*, CLXXXI (1949), 36-39.

HILGARD, E. R.; IRVINE, R. P.; and WHIPPLE, J. E. "Rote Memorization, Understanding, and Transfer: An Extension of Katona's Card-Trick Experiments," *Journal of Experimental Psychology*, XLVI (1953), 288-92.

KATONA, G. *Organizing and Memorizing*. New York: Columbia University Press, 1940.

LUMSDAINE, A. A., and GLASER, R. *Teaching Machines and Programmed Learning*. Washington, D. C.: National Education Association, 1960.

MALTZMAN, I. "Thinking: From a Behavioristic Point of View," *Psychological Review*, LXII (1955), 275-86.

MALTZMAN, I., *et al.* "Task Instructions for Anagrams Following Different Task Instructions and Training," *Journal of Experimental Psychology*, LI (1956), 418-20.

# III

# Some Basic Processes Involved in Mathematics Learning*

ZOLTAN P. DIENES†
*University of Sherbrooke*
*Sherbrooke, Quebec*

WHENEVER an organism is put in a new environment the initial interaction between organism and environment seems to be some kind of tentative exploring activity. The organism seems to wish to explore and manipulate the environment. It does this, presumably, for the purpose of being able to predict how the environment is going to respond. Mathematics learning would probably be no exception to this, but a preliminary groping period is notably lacking in most mathematics lessons. Children are not usually thrown into a mass of mathematical stimuli and encouraged to sort them out and make sense of them.

Such activity can probably best be described as play. Why does a kitten play with a ball of wool? The biological purpose is, no doubt, to become skilled at using its paws and to orient itself in space generally so that it can later catch food. A child plays for much the same reason. He moves his body around. He uses his mind in various acts of play in order to be able to meet requirements the environment is likely to pose later on. So play, it seems, should be regarded as an integral part of any learning cycle.

Mathematical play can be generated simply by providing children with a large variety of constructed mathematical materials. Suppose materials such as multibase arithmetic blocks, Cuisenaire rods, or various kinds of geometric materials that might induce them to learn about vector spaces, matrices, etc., have been made available.

* Inquiries should be addressed to International Study Group for Mathematics Learning, c/o Professor Zoltan P. Dienes, University of Sherbrooke, Sherbrooke, Quebec, Canada.

† Previously, Dr. Dienes has been affiliated with Teachers College, Columbia University, New York, New York, and with the University of Adelaide, Adelaide, South Australia.

The *first* thing to do with these materials, of course, is to leave them around for the children to play with! Then one of two kinds of play usually takes place. In one, which might be called purely manipulative, the child tries to find out, almost consciously, how the material handles. He wants to know what kind of a tool he has. In the other, which might be called representational play, the child adds his imagination to the manipulation—he makes up all sorts of "cover stories" and uses the material to represent these ideas.

Of course, the child who uses the materials inappropriately and tries by hook or by crook to fit them into his imaginings is not adapting to the environment as efficiently as the child who is making up suitable stories and devising appropriate uses for the material. The manipulative and representational kinds of play coalesce, really, into one stream of organic inquiry. The child, however, is not aware that he is inquiring into anything. He is merely having a good time playing with the materials.

Eventually, certain properties of the situation and other constraints will begin to make themselves clear to the exploring child. For example, the child may discover that some blocks do not stand up, that others do not fit alongside each other, that a triangle cannot be made out of squares or squares out of (some) triangles, and so on. The child, having realized the restrictions under which he is working and the possibilities that are open to him, will begin to ask questions concerning the conditions under which certain possibilities may be realized. For instance, can he build a certain kind of structure with a certain number of certain kinds of pieces? Is it possible to make windows in a certain part of the wall without causing the wall to collapse? Answers to such questions should be obtained rather easily by means of manipulating the material at hand.

## Abstraction

Abstraction is the gathering together of a number of different events or situations into a class, using certain criteria that must be applicable to all these events and situations. When we abstract we draw out from many different situations that which is common to them, and we disregard those things which are irrelevant to this common core.

If children are provided with a sufficient variety of mathematical materials, it will be more likely that the mathematical relationships determined during the course of their play with these materials will be abstract, rather than tied to certain particular situations. In effect, it is hypothesized that in mathematical learning abstraction will be more likely to take place if a multiple embodiment of a mathematical idea is provided, rather than a single embodiment such as Cuisenaire rods by themselves. Providing a number of embodiments enables the child to progress toward abstraction on a broad front. The more broadly based the abstraction, the more widely applicable will it be. In other words, if the abstraction is a result of gathering together the common properties of a large variety of situations, it is more likely that the final abstract concept will be applicable to a large variety of applications and situations.

The formation of abstract concepts seems to take place in cycles. The end point of each cycle can act as at least a partial beginning point of the next cycle. For example, the idea of natural number (or certainly of the cardinal aspect of natural number) is obtained partially by manipulating sets of objects, comparing them, and realizing that if two sets are in one-to-one correspondence then these sets are equivalent, i.e., have the same number of elements. When the order property is joined to this concept, the idea of natural number is made operational. This is the end point of a very long set of experiences in which the child finally realizes the irrelevance of various other properties of the sets and only the number property is retained. The child is probably unaware of this process, at least in the traditional educational setup.

Sets can undergo a large number of transformations, and still the number of elements might remain unaltered. Thus, any element of a set could be replaced by some other element without altering the property that the set has a certain number of elements. On the other hand, if one piece of a jigsaw puzzle is replaced by another piece, it is highly unlikely that the puzzle could still be completed.

## Higher Level Abstractions

The "natural number" property of sets can, in turn, be used as a starting point for further abstraction processes. The ideas of "even" and "odd" can be generated by getting children to arrange a set of objects into pairs. They will find that sometimes all of the objects can be arranged in pairs, and at other times there is one left over. The end point of a wide variety of such experiences will be the ideas of even and odd.

After this idea is achieved, it is possible to gain some appreciation of the relationship between them through making unions of disjoint sets with even and odd numbers of objects in them. By constructing such unions children can come to realize that the union of a set with an even number of objects with another set with an odd number of objects results in a set with an odd number of objects. At the end of this cycle the children will have realized the addition table of even and odd numbers —that is, that "even plus even equals even," "even plus odd equals odd," "odd plus even equals odd," and "odd plus odd equals even."

The resulting 2-by-2 table may later be recognized as isomorphic to the multiplication table of positive and negative numbers. In fact, it is an instance of the multiplication table, so-called, of the abstract mathematical group with two elements. It would also be possible for children to extend this notion by making up other tables of a similar kind (or of a different kind) that have more than two elements in the table, maybe three or four or eventually, perhaps, an infinite number.

In summary: the abstractions resulting from one cycle may provide a basis for the next, higher order, cycles. The experience cycle leading to natural number could, as we have said, lead to the ideas of even and odd. These, in turn, could lead to the connection between even and odd that might, then, be recognized as isomorphic to multiplication with equivalence classes of positive and negative numbers. The mathematic entity (two-group) invariant under this isomorphism could lead to cyclic groups of orders 3, 4, 5, . . . . Noncyclic groups, for instance the Klein group, could then be introduced by a variety of constructive experiences such as folding pieces of paper.

## Generalization

By generating abstractions out of previously formed ideas we are moving along an abstraction dimension.

There are, of course, many other dimensions of mathematical thinking. One of the more important dimensions is generalization—something hinted at, but not made explicit, in the preceding section. Whereas an abstraction is created from elements by virtue of realizing some common property of the elements, generalization is the extension of an abstract class to a wider class of elements that possess the same properties as the original class, or, possibly, properties only similar to them.

One might, for example, generalize from the even-and-odd situation to the rules for adding numbers that are divisible by three, those that when divided by three leave a remainder of one and those that when divided by three leave a remainder of two. The resulting addition rules result in what is known as a modulo-three arithmetic. This table does not have the same properties as the other table (modulo two), but it has some similar properties. For instance, any two kinds of numbers, when added, result in one of the three kinds. In other words, the situation is "closed." Another feature common to the two tables is that in each kind of table there is a neutral element. The neutral element in the even-and-odd table is the class of even numbers and in the modulo-three table it is the class of those numbers that are divisible by three.

Naturally, one can generalize to more than 3-by-3 tables. One could take a general $n$-by-$n$ table and invent rules for its operation. Or one might impose a restriction to certain kinds of rules. For example, it might be required that the rules result in an associative table or one with a neutral element, as in our two examples.

In all mathematics of generalization, the set of entities to which our operations are applicable is extended. Sometimes the process of generalization results in a new domain that not only is more extensive than the previous one but also includes an isomorphic image of the previous one. This situation might be referred to as "embeddedness." An example of this might be extending the group with two elements to a cyclic group of four elements—the modulo-four arithmetic. In the modulo-four arithmetic, the natural numbers are divided into four classes: those divisible by four and those that when divided by four leave remainders of one, two, or three. These four classes provide a system that is also an extension of the system involving three classes of numbers. But, in addition, the properties of the numbers that are divisible by four and those that leave a remainder of two have exactly the same properties as the even and odd numbers, respectively. The number classes are not identical, but the relationships involved are the same. In effect, a 2-by-2 table has been extended to a 4-by-4 table and in this 4-by-4 table there is a subtable that has the same properties as the original 2-by-2 table. This form of generalization is very common in mathematics.[1]

A similar situation exists in passing from natural numbers to integers. The positive and negative integers comprise a much wider class of entities than the natural numbers, and yet the properties of the positive integers are isomorphic to the properties of the natural numbers in reference to addition, subtraction, multiplication, and division. Still, a "positive two" is a very different concept than a "natural two." A positive two means that we are thinking of a "two-moreness" situation. Natural two means that we are thinking of a situation in which there "are two." Confusion between these two situations gives rise to much mathematical headache in the classroom.

In view of the foregoing, a question arises. Is it better to generalize on a narrow front and then abstract, or to abstract on a broad front and then generalize? In other words, is it better to restrict oneself to one or two situations to which the mathematical structure being learned is applicable, and at the same time pursue its mathematical generality as far as possible; or to look at a wide number of situations in which the

[1] In "On Abstraction and Generalization," *Harvard Educational Review* (Summer 1961), the term "generalization" is used only in the case where it occurs in conjunction with embeddedness.

structure is applicable, to encourage a broad abstraction of mathematics before extension of the mathematical structure itself is contemplated? Probably no easy answer is forthcoming. There might even be individual differences, and certainly the answer will depend at least in part on the type of mathematical situation being envisaged.

## PARTICULARIZATION

Abstraction and generalization are fundamentally different psychological processes. Abstraction, the creation of a class out of its elements, is an irreversible process. Once a class has been created, it is inconceivable that it can be uncreated. The generalization process, however, *can* be reversed. It is equally possible to pass from a more to a less extensive class and to pass from a less to a more extensive class. The former process might be called "particularization." Consider, for example, a two-dimensional vector space. The most general vector is an arbitrary ordered pair of real numbers. Suppose, however, the restriction is imposed that the sum of these two real numbers must be zero. This results in particularization from the entire vector space to a subset of this vector space. Further restrictions can be made; suppose the choice of vectors is now restricted to those in which the second component is a number that is two more than the first component. If both restrictions are to be satisfied, then only the vector $(-1, 1)$ will do. By two steps of particularization, coupled in each case by embeddedness, one particular vector has been identified in the two-dimensional vector space.

The fact that the restrictions are not applied in this systematic order in our mathematical learning does not necessarily indicate that this is not how it should be done. In fact, the results of obeying the first restriction but disobeying the second, obeying the second and disobeying the first, or disobeying both should also be considered by the learner. Considering these four possibilities would give a fuller mathematical context to the particular situation in which both restrictions are obeyed.

## SYMBOLISM

So far, emphasis has been given to conceptual structure, as it arises out of play, and particularly to the two dimensions of abstraction and generalization.

The role played by language has not been considered. This role is very important in the general scheme of mathematical learning, but exactly what it is is not yet clear. Very little research has been done on the role of language, either mathematical or metamathematical, in the learning of mathematics itself.

Some preliminary investigations were made at Harvard simply by observing children generate and play with symbols they had introduced themselves. It seemed that at times the introduction of symbols impeded the concept formation, while at other times the generation of symbols, particularly by the child himself, led to a considerable amount of exciting and creative thinking.

It is unclear what general laws govern the use of symbols in mathematical thinking. Until quite recently it had been taken for granted that the only way to learn mathematics is through symbols. Since it is now known that mathematics, even mathematics of quite sophisticated kinds (e.g., complex algebra, affine geometry, matrices, eigenvalues, etc.), can be learned via manipulative experiences with concrete objects, a question arises concerning the optimal use of symbols.

There seem to be some indications that a certain degree of abstraction is necessary before a symbol can effectively be used and applied in a wide range of situations. If symbolization takes place after only one embodiment has been introduced and children are asked certain questions to which the obvious answers (from the adult point of view) would be through the use of symbols, the children will almost invariably go back and manipulate the materials to provide the answers. On the other hand, if many different embodiments have been introduced and the symbols are beginning to mean for the children the common mathematical properties of these embodiments, then it becomes more likely that the children will use the symbols to provide the answer to a problem.

When a symbol to represent a certain situation has been either invented by a child or presented to him, it is always a problem to know what that symbol does, in fact, symbolize for that child. To what extent does the symbol denote that activity with those very things with which he is engaged, or to what extent does it denote a class of activities that he might engage in? It seems to be the hallmark of an intelligent child to think more in terms of classes of events than in terms of individual events.

Classifying events enables one to predict future events more accurately than regarding events simply as isolated individual occurrences. It seems a priori more probable that symbolization would be more effective after a high degree of abstraction has been achieved than if symbols are introduced at the very beginning.

Some people might argue that, on the other hand, introduction of the symbol would save a good deal of unnecessary work with concrete embodiments. Symbols are more easily transformable and manipulatable than concrete materials, so it is in a sense labor-saving to manipulate a symbol rather than an event. My reply to this criticism is that although it may be labor-saving, if the result of symbol manipulation is a knowledge only of how to manipulate further symbols it is of little use. At best, if the symbols denote one kind of activity then predictions regarding only that kind of activity will be possible as a result of manipulating the symbols. So it seems probable that the introduction of symbols *after* a variety of concrete experiences would be more effective than their introduction earlier—but just what variables are involved, only future research will determine.

### Interpretation

Once a language with which to talk about mathematical events has been constructed, the problem of decoding the language arises. Any nonmathematician, on looking inside an advanced mathematics textbook, will be horrified and will shut the book at once. This horror of alien symbols is due simply to the fact that most present-day adults were never taught, during their school days, how to decode mathematical symbolism. So symbolization and its concise interpretation should, perhaps, be emphasized in schools and experimented with in psychological laboratories.

The problem of decoding (interpretation) is the reverse of symbolization, just as particularization is the reverse of generalization. It is impossible to introduce an abstraction directly; so, to explain what a certain mathematical symbol conveys, it is necessary to choose a particular instance or representation and describe it, or to invent a language (e.g., an English metalanguage) with which it might be possible to explain what the strict mathematical language conveys.

One difficulty with mathematical language is its almost total lack of redundancy. Ordinary language is extremely redundant. English prose has been measured, I believe, to be on the average about 70 percent redundant. On the other hand, if any single part of a mathematical formula were to be left out, either the meaning of the formula would be altered or the statement would be reduced to nonsense.

Possibly a certain amount of redundancy should be allowed for, in introducing mathematical symbols. When we pass from ordinary communication to nonredundant mathematical communication there should, possibly, be some intermediate stages in which redundancy is gradually reduced to almost zero.

### Summary

To sum up, the following observations are offered.

One, to encourage children to abstract (that is, to determine the elements common to a large number of different situations), a large number of different situations must be provided. This leads to the principle of *multiple embodiment* of mathematical concepts.

Two, to encourage children to generalize, one must try to vary the values of the mathematical variables that make up the mathematical concepts to be taught. An illustration is that of varying the mod value in modular arithmetic. This leads to the principle of *mathematical variability*.

Three, if children are to symbolize and use their symbols effectively, it is probably better to let them have a hand in the process itself. Children might want to change their symbols as they change their breadth of abstraction. They might not want to use the same symbols when two or three situations are pulled together into one. If they originally used the symbols to represent only one of these situations, they might want a different and possibly more concise symbolism when they realize that there could be literally hundreds of similar situations. The principle involved might be referred to as the principle of *dynamic symbolization*. Normally, symbols are static; but in this conception symbols take on a dynamic role and become an integral part, indeed, of the abstraction and generalization cycles.

Four, to encourage children to interpret, they might first be given some practice in making up imaginative stories to which their structures are applicable. Soon they will realize the kind of stories that are applicable and those that are not. So if children are allowed to take a hand in the process of interpretation they are more likely to understand the interpretation they have abstracted than if the teacher does all the interpreting *for* them. This leads to a principle that might be called the principle of *image construction*. Children should be encouraged to construct their own images.

## Controlled Research

The preceding discussion has been rather general. The initial research of my collaborators and myself had to content itself with a type of research that might be described as "naturalistic" or "observational." This research was followed by some tentative theorizing to account for the facts observed. In a later series of experiments, conducted at the University of Adelaide, Dr. Jeeves and I looked into some of the detailed problems of how structures are built, and how the learning of one structure affects the learning of another.

For various reasons—partly because our subjects were unlikely to have come across them—mathematical groups were chosen as the structures to be learned. It could almost be guaranteed that each subject would start off at essentially the same position, that is, zero. Also, it was relatively easy to control the degree of task complexity.

We devised an experimental situation in which the responses of the subjects to the problems would be externalized. Many steps in the thinking process could be observed directly. This procedure gave us invaluable information as to how the thinking might have taken place from stage zero to criterion.

Several questions were investigated in detail. One dealt with the effects of conceptual symmetry or asymmetry of a task on performance. Another involved the effects of starting with a more complex task followed by a simple one, as against starting with a simple task followed by a more complex one. A third question we tried to answer is "What kind of strategies did the subjects use to solve the task that confronted them?" And, having isolated a number of different strategies, we then asked how and to what extent those strategies were related to the way in which the subjects themselves interpreted the nature of the task. We also asked questions about such things as the ability to extrapolate and the connections between performance, extrapolation, and intelligence. Extrapolation was operationally defined in terms of the number of times a certain combination had to be tried before it was, afterwards, faultlessly handled. Of course, a smaller score gave higher extrapolating ability.

This extrapolation measure was applied only to those subjects who did the complex task after the simple task, because some of the properties were unique to the complex task (but not vice versa). These subjects were required to guess what these properties were. Some of them guessed correctly from the very start, in which case the extrapolation score was zero, the best possible score for that ability.

Each task was administered approximately in the following way: The subject was given some cards. The experimenter had an identical set of cards, which he was able to put in the window of a piece of apparatus in front of the subject. The experimenter sat behind this apparatus. To begin, a certain card was placed in the window and the subject had to place one of his cards on the table. Which card was next placed in the window was determined by the card on the table, the card then in the window, and a set of rules. The subject had to predict what the next card in the window would be. After a certain number of successive correct predictions the subject was examined on the remaining possibilities; if he achieved a criterion of 90 percent correct, the task was discontinued. If not, the process was continued until he again reached the criterion number of successive correct predictions, after which he was once more examined on all the remaining possibilities.

The results were, roughly, these:

The extent of symmetry of the structure had a profound effect on behavior. The subjects tended to predict that symmetrical structures would be presented even when asymmetrical structures were being presented. For example, consider the Klein group as opposed to the cyclic-four group. In the Klein group, the three elements that are not the unit element are in a symmetrical relation to one another. That is, if the nonunit elements are $X$, $Y$, and $Z$, then $X$ and $Y$ produce $Z$, $Y$ and $Z$ produce $X$, and $X$ and $Z$ produce $Y$. In the cyclic-four group, this is not so. If $X$, $Y$, and $Z$ are the three nonunit elements, then although $X$ and $Y$ produce $Z$, and $Y$ and $Z$ produce $X$, $X$ and $Z$ do not produce $Y$. Instead, they produce the unit element—so there is an annoying kind of asymmetry about the situation that the subjects learning the asymmetrical (cyclic-four) group did not seem to like at all. They seemed to predict in the direction of the more symmetrical structure.

Another interesting finding was that those who received the four-group first appeared to do better than those who received the two-group followed by the four-group. This was especially true when the performance was measured in terms of the subject's verbal interpretation of what the tasks were about. This finding provided a hypothesis that was tested later on with more complex group structures. It seemed that there was little to choose between introducing a group of order three before a group of order five or vice versa, in the case of adults; but children did significantly better when they started on the five-group than when they started on the three-group. It also seemed that those who did the three-group before a six-group did considerably better than those who did the six-group before the three-group. Thus, the two-group followed by four-group subjects did not do as well as the four-group followed by two-group; but the three followed by the six did better than the six followed by the three. The conclusion is that while it is possible to throw people into a structure too deeply, it is also possible to allow them to get in too gingerly. It seems that there is no particular immediate rationale why the optimal level should be at any particular place rather than another. The construction of models predicting such things remains a task for the future.

The strategies used appeared to fall quite neatly into three distinct categories. The first we termed the *operational strategy;* the second, *pattern;* and the third, *memory.* The *operational strategy* was presumably the result of the subject's regarding the card he played as an operator acting on the card in the window. This strategy may have encouraged the subjects to play the same card several times against different cards in the window to find out what kind of operator it was. The *pattern strategy* involved playing with the combinations in particular areas of the 4-by-4 matrix constituting the associated triads to be learned in the group operation. For instance, a subject might investigate what would follow whenever the card in the window and the card on the table involved the same symbol; or he might investigate how the neutral card worked. The *memory* strategy appeared to be a random strategy in which the subjects restricted themselves to learning the combinations in a random order until the whole table had been memorized.

These strategies were related to the ways in which the subjects evaluated the tasks. Those who used an operational strategy indicated that the card played had a role in the tasks, different from that of the card in the window, and would affect in a particular way what card would next appear in the window. The pattern strategists viewed the tasks as depending on how the combination of particular kinds of cards would affect the result. The memory strategists simply felt that there were some combinations to memorize.

The relationships between the evaluations and the strategies were, in every case, statistically significant. The relationship between the evaluations and the number of instances required to complete the tasks was also tested. It appeared that the operational evaluators completed the task with the smallest number of instances, the pattern evaluators came next, and (as might have been expected) the memory evaluators came last. These differences also were significant.

With regard to extrapolation, it was found that ability to extrapolate correlated very highly with general performance on the task, as measured by the number of errors made. Extrapolation also correlated highly with intelligence, measured by ordinary group tests used for school purposes; yet there were no significant relationships found between performance scores and intelligence test scores.

These results led to further questions about the interrelations between tasks and performance. One question that was investigated is the differential effect of generalization and inclusion of one structure in another on performance. For example, in the three-group, the five-group, and the seven-group, there are no equivalent parts except the neutral element, and yet each structure is very clearly a generalization of the preceding ones. The way in which the tasks become more complex is by way of generalization. Now, if instead of taking the three-, five-, and seven-groups, we take the three-, six-, and nine-groups and take either the direct product of the three-group itself or the cyclic nine-group, we will have *embeddedness* at every stage. That is, the three-group is embedded into the six-group and the three-group is embedded into any nine-group. Of course, we also have overlapping between the six-group and the nine-group because they both contain the three-group as a subgroup.

Some subjects were run on the three-group followed by the five-group followed by the seven-group, and others on the three-group followed by the six-group followed by the nine-group. Control groups were run in the reverse orders to determine the differential effects of doing the complex tasks first. The criteria were performance on the third task, and the sum total of the errors on the first two tasks.[2]

There are strong indications that children can cope with *embeddedness* much more easily than with generalization.

Some other interesting problems are being opened up in the field of children's learning of logic. Up until quite recently it had been thought that children pick up logic incidentally as they mature. From experiments of William Hull[3] in Cambridge, of my own in different parts of the world, particularly in Hawaii, New Guinea, and South Australia, and of others, it is becoming obvious that young children are able to engage in quite sophisticated logical thinking if the stimulus situations are of a concrete character.

The Vigotsky blocks, adapted for this use by Hull, involve different sizes, shapes, colors, and thicknesses with each possible combination of attributes occurring exactly once. Conjunctions, disjunctions, and negations can be, so to speak, "played with" by children in this way. Children might, for example, collect all blocks which are both red and square, all blocks which are not circles, etc. Disjunctive combinations may, in addition, lead to implications. Thus, in a set where all blocks are either red or not square, all squares will be red. That is, from the disjunctive attribute "either red or not square" we can deduce the implication attribute, "if square then red." Further, all valid deductions, such as that above, can be shown to be equivalent to an inclusion relationship between sets. If an object is a member of *Set A,* which is included in *Set B,* then it is also a member of *Set B*. More precisely, if $A$ is included in $B$ and $x$ is a member of $A$, then $x$ is a member of $B$.

In South Australia, and later in New York, we experimented with the introduction of the quantifying operators for "all" and "there exists" and their relationship to negation. In fact, the children in a preparatory

[2] The results have now been collected and are to be published in the second volume of *Psychological Monographs*, Adelaide University, Adelaide, South Australia.

[3] See his "Concept Work with Young Children," *Bulletin of the International Study Group for Mathematics Learning*, 1962.

class, five-year-olds, recently compelled the kindergarten teacher to introduce the ideas for "all" and "there exists" because of their questioning attitude on how the various properties were to be represented by sets of blocks. Were all of the blocks in a set red, or were some of them red, or were all of the red ones in the set? Such inquiries coming from five-year-olds after a few months of experience in logical thinking are very encouraging. Controlled experiments on this kind of thinking have not yet been carried out, but some are being planned.

It will be appreciated that the research described here represents only a beginning. A great deal more dovetailing of laboratory and classroom research, and of mathematics-learning, logic-learning, and language-learning research will need to be done before we can consider the study of complex learning as truly undertaken.

# IV

# Some Counting Models for First-Grade Performance Data on Simple Addition Facts*

PATRICK SUPPES
GUY GROEN
*Stanford University*
*Stanford, California*

To MATHEMATICIANS or educators who have not thought very much about the matter it usually comes as a surprise, and occasionally even as a shock, to find out how little we know about the learning of mathematics. It is not uncommon to hear mathematicians say that because mathematics is a systematic subject with an inherent order imposed on the development of topics, it should be relatively straightforward to give quite an adequate account of how students learn mathematics. Because students do learn mathematics and because many of the mathematicians who make this sort of statement have themselves been successful teachers, it is not always evident what is the best way to bring out the gross inadequacies in our present knowledge of mathematics learning.

Perhaps the most effective way—at least we have found that it sometimes works—is to rely heavily on computer analogies. First challenge: If you understand so well how mathematics is learned, please program my computer to learn it. It does not take much discussion to bring out the difficulties of this task, and one can then move on to a second challenge: Predict the points at which students will have learning difficulties, and make explicit the principles used to make the predictions. The requirement of explicitness is needed to make the challenge a scientific and theoretical one that cannot be answered by the nonverbalized and intui-

* This research was supported by the U.S. Office of Education, The National Science Foundation, and the Carnegie Corporation of New York. The authors would like to thank Mrs. Sue Matheson for running the experiments in the school.

tive experience of a good teacher. While one's opponent is struggling with this second challenge, a third challenge on performance data can be put ready at hand: Predict systematic variations in performance data involving mathematical concepts and skills already taught, and again make the principles of prediction explicit. The view that mathematical colleagues may have difficulty giving a serious constructive response to these three challenges is not meant as a criticism of their scientific prowess. The only criticism implied is of the opinion that we already know how to meet these three challenges in any serious way.

The present article is meant to be a small step toward a positive response to the third challenge. From a mathematical standpoint the performance task we have selected is ridiculously simple, that of handling correctly the simple addition facts, with the sums being no greater than 5. From a psychological standpoint, however, this task is not as simple as most of those that lie at the heart of the classical experiments in learning theory. Moreover, attempts to develop mathematically well-defined performance models for even this simple task do not seem to exist in the literature.

We reserve more detailed comments until after we have presented in the next section goodness-of-fit results, i.e., the extent of correspondence between the theoretical predictions and the experimental outcomes for five closely related models. A broader conceptual framework for the viewpoint expressed here is to be found in Suppes.[1]

## An Experimental Test of Five Models

The results we will discuss are from an experiment in which a group of first-grade children in the first half of the school year were asked to solve a set of simple addition problems. Each problem was of the form

$$m + n = \_\_,$$

where $m + n \leq 5$. The line was colored red and the rest of the problem was printed in black. The task of each child was to provide the missing number.

Thirty subjects were used, randomly selected from two different homerooms. Each subject was run individually. Subjects were seated in front of a panel with six buttons marked 0, 1, 2, 3, 4, and 5. A sample problem was then projected on a screen in front of them. They were told that the red line meant that a number was missing and were instructed to

[1] Patrick Suppes, *The Psychological Foundation of Mathematics*, Technical Report No. 80 (Stanford, Calif.: Stanford University, Institute for Mathematical Studies in the Social Sciences, 1965). Pp. 33.

push the correct button for the missing number. When a subject had responded, he was shown a new slide with the correct answer (printed in red) replacing the red line. Each child was then presented with a sequence of twenty-one problems consisting of all possible combinations of integers $m$ and $n$, subject to the constraints

$$\begin{aligned} m + n &\leq 5, \\ m &\geq 0, \\ n &\geq 0. \end{aligned}$$

These problems were presented in a random order, the same sequence being used for each child. After each presentation of a problem, the child made a response and was shown the correct answer. Both the actual response and the response latency (the time between the onset [presentation] of the stimulus and the elicitation [occurrence] of the response) were recorded. This procedure was repeated for two more days. However, on the last two days, no preliminary instructions were given, and the child was asked to respond as quickly as possible. The order of presentation of items was different on each of the three days.

In this discussion, we will concentrate on the data obtained on the third day. It can be assumed that by then the children had become fully familiar with the experimental situation. The initial problem we proposed to consider was whether it is possible to formulate a simple model that will account in an approximate fashion for the children's responses.

Unfortunately, the error rate was too low for any systematic analysis to be based on this aspect of the response data. Although at least one subject made an error on each problem, seven subjects out of the thirty made errors on $1 + 3 =$ __ and $1 + 2 =$ __, and five subjects made errors on $4 + 1 =$ __, $3 + 2 =$ __ and $1 + 1 =$ __. On most other problems one or two subjects made an error.

As a result of these low error rates, it seemed more promising to consider the response latencies. The most reasonable basic assumption to make is that the variations in response latencies between problems are the reflection of some kind of counting process that the child is using. For a problem of the form $m + n =$ __, it is possible to distinguish between five different kinds of counting processes. In order to make this distinction, it is convenient to consider a counter on which two operations are possible: setting the value of the counter to a certain value (while clearing the previous value) and adding a number to the current value of the counter. The addition operation is performed by successively increasing the initial value of the counter by one until the second value has been added on. The operation of this counter is illustrated in Figure 1, as shown on the following page. Using this counter, an addition problem of the form $m + n =$ ___ can be solved in the following ways:

1. *The counter is initially set to 0, m is added and then n.*

2. *The counter is set to m (i.e., the left-most number) and n is then added.*

3. *The counter is set to n, and m is then added.*

4. *The counter is set to the minimum of m and n. The maximum is then added.*

5. *The counter is set to the maximum of m and n. The minimum is then added.*

The setting operation is assumed to take a constant time, independent of the value to which it is set. The addition time, on the other hand, is proportional to the number of times the counter must be increased. Suppose a counter takes time $\alpha$ to be set and time $\beta$ to be increased by 1. If a counter is to be set to a certain value and then increased $x$ times by 1 (which is equivalent to having $x$ added to it) the total time $T$ taken by the counter to perform these operations is

$$T = \alpha + \beta x. \tag{1}$$

Thus, Equation (1) gives the time taken to perform an addition problem of the form $m + n =$ ___. It will give differential predictions depending on the type of solution because, corresponding to the classification of solution types we have just proposed, $x$ is determined as follows:

*Type 1.* $x = m + n.$
*Type 2.* $x = n.$


FIGURE 1.—EXAMPLE OF A DEVICE WHICH SETS A COUNTER TO $a$ AND ADDS $x$ TO $a$

*Type 3.* $x = m$.
*Type 4.* $x = \max (m, n)$.
*Type 5.* $x = \min. (m, n)$.

If we wish to apply this model to the latencies of our experimental subjects it cannot be assumed that the values of $\alpha$ and $\beta$ are constant. Rather, it is correct to assume that $\alpha$ and $\beta$ are random variables with two different distributions. However, we can eliminate this problem by taking the mean latencies, $E(\alpha)$ and $E(\beta)$, over all subjects. We then have, for a particular problem $i$,

$$E(T_i) = E(\alpha) + x_i E(\beta), \tag{2}$$

where $x_i$ is computed according to the rules given above. For Equation (2) to hold, it is necessary, of course, that $x_i$ be constant for all subjects on a given problem. In other words, it is necessary to assume that all subjects use the same type of solution. If this assumption is incorrect, then the goodness of fit of observed-to-predicted data will be affected.

In order to evaluate the goodness-of-fit of these five models, it is necessary to estimate the expected values $E(\alpha)$ and $E(\beta)$. These estimates will be denoted by $\hat{\alpha}$ and $\hat{\beta}$. For each problem, it is possible to compute a value of $x_i$ under each of the five assumptions. Since Equation (2) is linear, $\hat{\alpha}$ and $\hat{\beta}$ can be computed for each model by means of a simple regression analysis, using $x_i$ as the independent variable and the observed average-success latency on each problem as the dependent variable, with the index $i$ ranging over all twenty-one problems. It is necessary to use the success latency rather than the overall latency for the dependent variable because it is reasonable to assume only that Equation (2) holds for correct solutions.

An analysis of this type was performed on the data obtained on the third day of the experiment. Two problems ($3 + 0 =$ __ and $2 + 3 =$ __) were omitted from the analysis. On both these problems, many individual response latencies were excessively high. The former was always the first problem to be presented. The high latencies on the latter can also be accounted for on the basis of sequential ordering effects. From the data obtained from the remaining nineteen problems, $\hat{\alpha}$ and $\hat{\beta}$ were evaluated for each of the five models, and two indexes of goodness of fit were computed. The first was the mean squared deviation between predicted and observed values:

$$s^2 = \frac{1}{17} \sum_{i=1}^{19} (T_i - \hat{\alpha} - \hat{\beta} x_i)^2,$$

where $T_i$ denotes the observed success latency for problem $i$. Also computed was the ratio of $\hat{\beta}$ to the standard error of $\hat{\beta}$. If $T_i$ is normally

TABLE 1
REGRESSION ESTIMATES FOR THE DIFFERENT SOLUTION TYPES
($\hat{\alpha}$ AND $\hat{\beta}$ MEASURED IN SECONDS)

| MODEL | $\hat{\alpha}$ | $\hat{\beta}$ | $s^2$ |
|---|---|---|---|
| 1. $x = m + n$. | 2.96 | .216 | .369 |
| 2. $x = n$. | 3.50 | .098 | .465 |
| 3. $x = m$. | 3.48 | .119 | .404 |
| 4. $x = \max(m, n)$. | 3.43 | .092 | .471 |
| 5. $x = \min(m, n)$. | 3.26 | .710 | .233 |

TABLE 2

| MODEL 1: $x = m + n$. | | | | MODEL 5: $x = \min(m, n)$. | | | |
|---|---|---|---|---|---|---|---|
| Problem | $x$ | Mean Success Latency (in seconds) | | Problem | $x$ | Mean Success Latency (in seconds) | |
| | | Pred. | Obs. | | | Pred. | Obs. |
| 0 + 0 | 0 | 2.96 | 2.98 | 0 + 0 | 0 | 3.26 | 2.98 |
| 0 + 1 | 1 | 3.18 | 3.36 | 0 + 1 | 0 | 3.26 | 3.36 |
| 1 + 0 | 1 | 3.18 | 3.27 | 1 + 0 | 0 | 3.26 | 3.27 |
| 0 + 2 | 2 | 3.40 | 3.57 | 0 + 2 | 0 | 3.26 | 3.57 |
| 1 + 1 | 2 | 3.40 | 2.67 | 2 + 0 | 0 | 3.26 | 2.88 |
| 2 + 0 | 2 | 3.40 | 2.88 | 0 + 3 | 0 | 3.26 | 3.45 |
| 0 + 3 | 3 | 3.61 | 3.45 | 0 + 4 | 0 | 3.26 | 3.48 |
| 1 + 2 | 3 | 3.61 | 4.20 | 4 + 0 | 0 | 3.26 | 3.40 |
| 2 + 1 | 3 | 3.61 | 4.28 | 0 + 5 | 0 | 3.26 | 2.85 |
| 0 + 4 | 4 | 3.83 | 3.48 | 5 + 0 | 0 | 3.26 | 3.03 |
| 1 + 3 | 4 | 3.83 | 4.18 | 1 + 1 | 1 | 3.97 | 2.67 |
| 2 + 2 | 4 | 3.83 | 3.90 | 1 + 2 | 1 | 3.97 | 4.20 |
| 3 + 1 | 4 | 3.83 | 4.04 | 2 + 1 | 1 | 3.97 | 4.28 |
| 4 + 0 | 4 | 3.83 | 3.40 | 1 + 3 | 1 | 3.97 | 4.18 |
| 0 + 5 | 5 | 4.05 | 2.85 | 3 + 1 | 1 | 3.97 | 4.04 |
| 1 + 4 | 5 | 4.05 | 4.49 | 1 + 4 | 1 | 3.97 | 4.49 |
| 3 + 2 | 5 | 4.05 | 5.15 | 4 + 1 | 1 | 3.97 | 4.53 |
| 4 + 1 | 5 | 4.05 | 4.53 | 2 + 2 | 2 | 4.68 | 3.90 |
| 5 + 0 | 5 | 4.05 | 3.03 | 3 + 2 | 2 | 4.68 | 5.15 |

distributed, then this has a $t$ distribution with $n - 2$ degrees of freedom. (In the present case, $n = 19$ [problems], so that $n - 2 = 17$. Although the summation is over subjects, as well as over problems, the details have been omitted so as not to obscure the basic ideas.) While it is not entirely clear whether or not the assumptions of the test are satisfied in the present experiment, its application does provide a rough index of whether or not the fit is satisfactory. The values of $\hat{\alpha}$, $\hat{\beta}$, and $s^2$ resulting from the analyses of the various models are shown in Table 1. Model 1 and Model 5 provided the best fits (i.e., $s^2$ was smallest for these models). The second goodness-of-fit computation resulted in levels of significance beyond .05 for all but these models. The predicted success latencies obtained on the basis of Model 1 and Model 5, together with the corresponding observed mean success latencies, are shown in Table 2. Notice that, especially in Model 5, each value of $x$ involves a number of data points. As a result, a clearer notion of the fit can be obtained by comparing the predicted latency with the observed latency averaged over all problems that contribute to a given value of $x$. This is done in Figure 2. It is clear that the best fit is provided by Model 5. Although there are

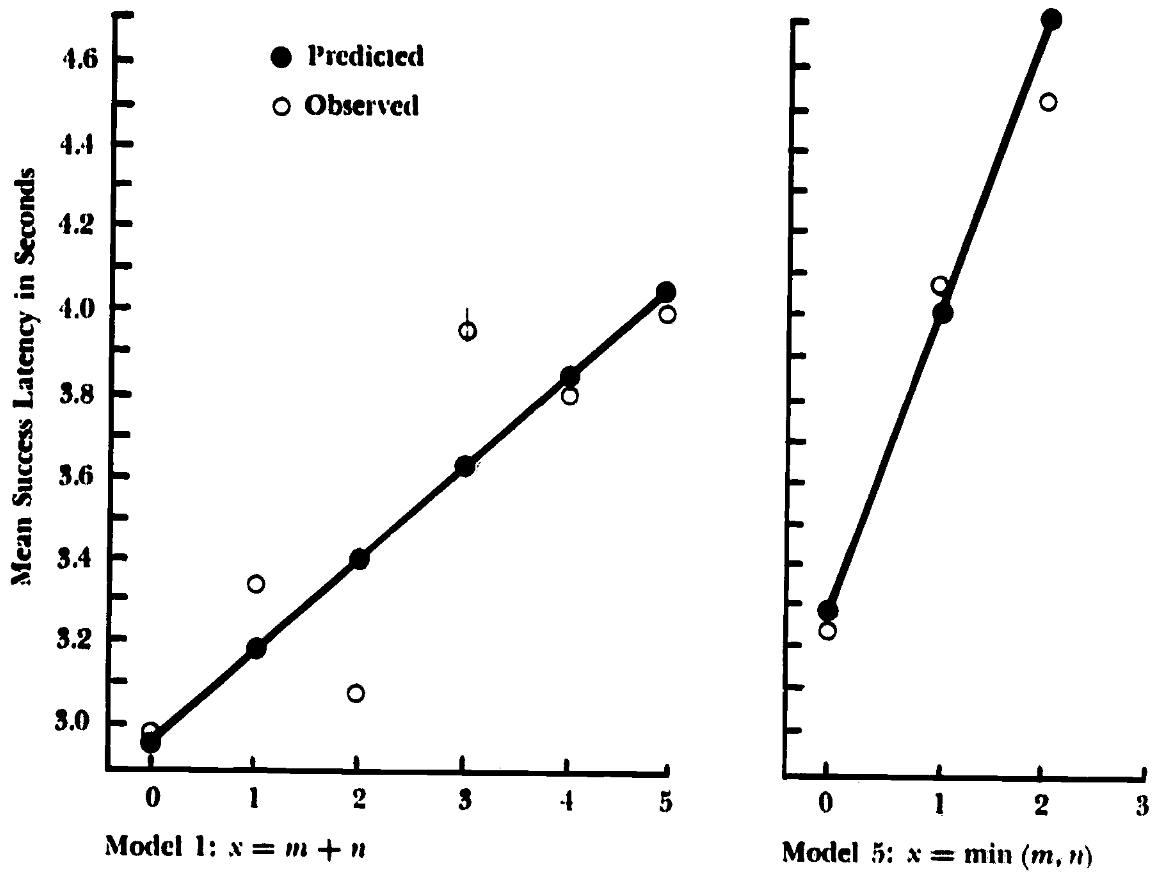


FIGURE 2.—PREDICTED AND OBSERVED MEAN SUCCESS LATENCIES PLOTTED AS A FUNCTION OF $x$ FOR THE TWO BEST-FITTING MODELS

less values of $x$ in Model 5, the better fit cannot be ascribed to this circumstance, since the value of $s^2$ is lower for Model 5, despite the fact that $s^2$ was computed for each model on the basis of deviations between predicted and observed latencies for individual problems. Further evidence in favor of Model 5 is the fact that, with the exception of $1 + 1$, *all* problems with $x = 1$ have larger latencies than those with $x = 0$.

While it can safely be concluded that Model 5 fits better than Model 1, this result can only be considered to be a first step. There is no guarantee that no other model exists that would fit the data in a more satisfactory fashion. Moreover, it cannot be inferred that the good fit of Model 5 implies that subjects tend to add two numbers according to the mechanism suggested by the model. For this model, $x$ ranges from 0 to 2. It is only when $x = 2$ that neither a 0 nor a 1 appears in the problem. Hence the data might be accounted for by a model which assumes specific algorithms for solving problems involving a 0 or 1 rather than the general algorithms used by the models we have proposed in this paper. Finally, there is, of course, the possibility that different individuals use different algorithms. Subsequent research that deals with these matters is now planned.

## Some Concluding Remarks

It would be good if we could report that the algorithm represented by Model 5 was the one explicitly taught the children by their teachers. This does not seem to have been the case. At the present time most first-grade teachers do not teach their students an explicit counting algorithm for handling the simple addition facts ordinarily taught in the first grade. As would be expected there is usually some mention, and often even a fair amount of discussion, of counting and its relation to the first introduction of addition. But—and this is the important point—an explicit algorithm is not developed and taught as is done later for addition of multi-digit numbers.

The results of the present paper suggest that more attention might profitably be devoted to these first algorithms, and that the algorithm of Model 5, which seems more sophisticated than that of Model 1, might well receive more explicit emphasis in the teaching of first-grade arithmetic.

It has not been our intention in this short paper to present any definitive research, but only to illustrate how even so simple a thing as learning the addition facts presents an interesting challenge to learning theorists and affords an opportunity to test some alternative mathematical models, each of which rests on a clear intuition of how a simple addition problem may be solved. The central idea of a counting model seems so natural that it seems difficult to think of other possible approaches, but this is not really the case—for example, a table-look-up model with parameters appropriately introduced for scanning the table can be formulated in such a way that it is identical in all behavioral predictions with Model 5. Moreover, simple counting ideas are not sufficient to account for all the significant variations in the observed data of Table 2, and as a larger body of data is accumulated, more complex and subtle ideas will be needed in constructing an adequate model of the observed phenomena. On the other hand, it seems to us that the learning of elementary mathematics affords a natural testing ground for mathematical models of learning or performance, and there is some reason to hope that in a first approximation, at least, models of a reasonable degree of simplicity will suffice.

It should be apparent that as such models are developed and the range and depth of their success is increased they will have increasing significance in suggesting and guiding curriculum modifications, particularly as regards the fundamental problem of finding out how students can on the *average* best learn mathematical concepts and skills.

# V

# A Comparison of Discovery and Expository Sequencing in Elementary Mathematics Instruction*

BLAINE R. WORTHEN†
*University of Utah*
*Salt Lake City, Utah*

CONSIDERABLE research which relates to the discovery-expository dimension of the task presentation has been conducted (for example, research conducted by Katona, 1940; Stacey, 1949; Craig, 1953; Sobel, 1954; Kittell, 1957; Corman, 1957; Haslerud and Meyers, 1958; Kersh, 1958; Gagné and Brown, 1961; Wittrock, 1963; and Scandura, 1964). To date these studies have failed to clarify many of the questions pertaining to discovery and expository instruction; rather, the findings of the various studies, when taken at face value, often seem to be contradictory. Perhaps the greatest factor which contributes to such equivocal research evidence is the differing specification among researchers as to what they mean by such terms as "discovery," "guided discovery," and "exposition." Since these terms have not yet been reduced to generally accepted operational definitions, it is highly probable that researchers working in what is nominally the same domain are not actually investigating the same phenomena at all. Even within the broadest framework

* This investigation was supported by the Cooperative Research Program of the Office of Education, U.S. Department of Health, Education and Welfare (Project 2277) and constitutes part of the final report of that project (Della-Piana, Eldredge, and Worthen, 1965). The data collected for this investigation also served as an essential portion of a master's thesis (Worthen, 1965) submitted to the Department of Education, University of Utah. For a more complete review of related research and a detailed description of all methods, analyses, instrument validation, results, etc., the reader is referred to the above sources.

† Now at The Ohio State University.

of agreement concerning elements which are generally characteristic of discovery and expository teaching, there is still another divisive factor at work. A review of the research literature shows that many of the relevant variables have been explored to a marked degree, while others have received relatively little attention. Such wide divergence in the variables controlled in various studies has led to investigation of widely differing facets of the discovery and expository processes, a too-general specification of task parameters, and a consequent noncomparability of the results.

Many of the investigators have been primarily concerned with the amount and type of external guidance to which the learner is subjected. Others have been concerned chiefly with the role of verbalization in the discovery-expository processes. One facet of investigation which has received somewhat less attention is that of the sequence characteristics of the learning tasks. In fact, many previous "discovery" studies have failed to consider or specify such task parameters as sequence. It could be argued that the type or amount of external guidance or verbalization is no more important in concept formation than the timing of such guidance or verbalization. Certainly this aspect of discovery teaching deserves investigation in its own right.

In addition to the lack of clarity of research evidence pertaining to the discovery-expository dilemma, there is another factor which often disturbs the practitioner who depends on research to determine the best instructional techniques for classroom use. Most "discovery" studies have been conducted in a laboratory setting and consequently have dealt with small time samples, small numbers of subjects, and very discrete and often manipulative learning tasks. One might argue that such sampling of time, subjects, and tasks is so restrictive and limited in scope that any attempt to generalize the results to classroom learning or instruction would be subject to serious question. It would seem that the results of a *carefully controlled* classroom experiment where both time sample and learning task are representative of typical school behavior and curriculum could be generalized to classroom practice with more confidence than could the results of the typical short-term laboratory experiment.[1]

The primary purpose of the present study was to describe and compare two instructional methods in a naturalistic setting where the learning tasks and time sample approximated normal classroom conditions. The methods compared were a discovery method and an expository method

[1] The difficulty of controlling research in a naturalistic classroom setting has been documented (Bellack, *et al.*, 1963, pp. 165-68; and McDonald, 1964, p. 542) and is acknowledged by the investigator. It would seem, however, that difficulty does not of itself preclude the possibility of finding productive ways to utilize the classroom as a research setting.

which differed primarily in terms of the sequence characteristics of the presentations, and secondarily in terms of teacher guidance necessary to maintain these sequence characteristics. No attempt was made to define *the* discovery method or *the* expository method.[2] Instead, attention was given to describing two methods which may be somewhat typical of the characteristics that normally serve to differentiate discovery techniques from expository techniques.

Specifically, the present study assessed the effects of two methods of teaching selected mathematical concepts to fifth- and sixth-grade subjects. The two sets of experimental sequences were presented to the subjects through quasi-textual instructional programs and were introduced by classroom teachers trained in both techniques of presentation. The criteria used to measure the outcomes of instruction included the following: tests of initial learning, retention, and transfer of the selected mathematical concepts; tests for transfer of heuristics; and measures of attitude toward the subject content. A complete listing and description of these criterion measures appears later in the section "Tests and Measures."

Secondary purposes of this study were the following: (1) to test the criticism that teaching by a discovery method is inherently more time-consuming than teaching by exposition; and (2) to point out fruitful directions that more focused research might take.

Brief definitions of the experimental methods appear below.

*Discovery method (Treatment D).* Treatment D is a method in which verbalization of each concept or generalization is delayed until the end of the instructional sequence by which the concept or generalization is to be taught.

*Expository method (Treatment E).* Treatment E is a method in which verbalization of each concept or generalization is the initial step in the instructional sequence by which the concept or generalization is to be taught.

It was hypothesized that Treatment D would produce superior results to Treatment E on each of the criterion measures.

---

[2] The discovery and expository techniques used in this study are not directly comparable to other techniques using the same descriptive titles. Indeed, the use of the term "discovery" in the literature is vague. As Wittrock and Keisler (1965, p. 20) note, the term is used to describe stimuli, intervening variables, and responses. The same is true, in a lesser degree, of the term "expository." The emphasis of this study is on the sequence characteristics of the stimulus materials.

## METHOD

### *Subjects*

The subjects were 538 fifth- and sixth-grade pupils in the Salt Lake City School District, Salt Lake City, Utah. The experimental sample was comprised of 432 of these pupils, who were equally divided among 16 classes. These classes were equally divided among eight elementary schools which were judged by district central office personnel to be representative of the elementary schools in the district, in terms of socioeconomic and geographical characteristics.[3]

The teachers were selected on the basis of the following criteria: (1) mathematical and general teaching competence, as judged by supervisors, (2) minimum of three years of teaching experience, and (3) willingness to participate in this research project. The selection of the teachers determined the selection of the sample; subjects used in this study were pupils in established classes of the selected teachers.

### *Experimental design and controls*

Two classes in each of eight schools served as experimental groups. In each school, both classes were taught arithmetic by the same teacher, one class by Treatment D and one class by Treatment E. This was done in order to control the dimensions of teacher personality and other teacher characteristics. Seven of the teachers taught two sixth-grade classes each, while the eighth teacher taught two fifth-grade classes.

Seven of the eight experimental teachers taught their own homeroom class as one of the experimental groups. In an attempt to control possible differential in pupil-teacher interaction between homeroom and nonhomeroom classes, the number of homerooms receiving each experimental treatment was balanced as nearly as possible. The assignment procedures also balanced as nearly as possible the number of classes receiving each treatment during any particular segment of the school day. Although there was no reason to believe that the selection and assignment procedures would bias the sample, a preliminary inspection of the mean values

[3] A control group, comprised of 106 pupils in 3 sixth-grade classes, received both the pre- and posttests but received no special instruction during the intervening six-week period. This group was included in the study in order to provide normal baseline data against which to assess effects of the two experimental treatments. Results of the intertreatment comparisons between the experimental groups and the control group appear in detail in previous reports of this research (see the introductory footnote) but are omitted here in the interest of brevity. It should be noted, however, that the results of these comparisons support the findings and conclusions reported herein.

for each treatment group was conducted on several pre-treatment measures including IQ, arithmetic computation skill, arithmetic problem-solving ability, prior knowledge of the selected mathematical concepts, prior attitude toward arithmetic, and pupil perception of teaching behavior. The only significant differences found between the Treatment D and the Treatment E groups were on the attitude measures. Pupils in Treatment E entered the experimental period with significantly better attitudes toward arithmetic than pupils in Treatment D.

The major nonexperimental variables controlled in this study are presented below.

1. The pupils in Treatments D and E received the same length of time to work on the learning tasks.

2. Although the type of verbal behavior varied to fit the two teaching models, the amount of verbalization in the teachers' oral presentation and in the written instructional materials was held constant in both treatments. Verbalization of the mathematical generalizations varied in sequence between the two treatments but was present in both.

3. In order to obviate the criticism that the instruction received by the two treatment groups was not actually different or did not match the experimental models, three techniques were used in this study in an attempt to assess the extent to which the teachers did, in fact, teach by the specified methods. These techniques (utilizing instruments described hereafter) included the following: (*a*) live rating by observer-raters of a 10 percent sample of the total teaching behavior of each teacher in each treatment; (*b*) rating of a 10 percent sample of total teaching behavior of each teacher in each treatment from lessons recorded on audio-tape; and (*c*) rating by pupils of teaching behavior on the discovery-expository dimension.

4. The research design and all of the various procedures and methods utilized were designed to negate any differential "Hawthorne Effect" between the two experimental groups.

5. An attempt was made to equalize the pre-experimental mathematical experiences of all subjects in Treatments D and E by presentation, during a two-month period immediately preceding the pretests, of a unit which included both specific and general mathematical concepts judged to provide necessary background for the experimental materials. In addition, pollution of the experimental results by nonexperimental arithmetic experiences was minimized by a request that no homework or out-of-school arithmetic assignments be given to the pupils. District personnel complied with this request and also elicited parental cooperation.

The experimental period consisted of three days of pretest administration, a six-week instructional period, and five days of posttest administration.

*Training program*

All raters and teachers attended a training class which met a minimum of two hours weekly for 20 weeks, 13 weeks prior to and 7 weeks during the experimental period. Extra training sessions were frequently inserted as they proved necessary. Training was given in four areas: (1) general mathematical concepts necessary as background; (2) all selected mathematical principles used in the instructional materials and criterion measures; (3) procedures for administering and scoring the various tests, scales, and questionnaires; and (4) use of the two specific methods of instruction. Training procedures included the following: (1) demonstrations by the investigator of all instructional units in each treatment; (2) practice teaching and critiques, during the training class, of portions of the instructional units; and (3) practice of instructional techniques in a third class set up in each school specifically for that purpose.

*Instructional materials*

The instructional materials were unique to each treatment and consisted of mimeographed textual materials for each subject. These materials presented several mathematical concepts selected on the basis of suitability for both discovery and expository teaching and probable unfamiliarity to subjects at the inception of the study. The mathematical concepts selected were the following: (1) notation, addition, and multiplication of integers (positive, negative, and zero); (2) the distributive principle of multiplication over addition; and (3) exponential notation and multiplication and division of numbers expressed in exponential notation.

The materials were equated in terms of the mathematical concepts, diagrams of physical models, number and type of examples, and degree of verbal presentation used in each treatment. The two sets of materials differed primarily in terms of sequence characteristics.

*Instructional procedures*

The instructional procedures in each treatment were largely determined by the requirement that the teachers follow the predetermined sequences of the instructional materials. However, a significant portion of total teaching behavior was judged to be independent of task sequence characteristics but still influential in affecting the impact of the instructional sequences on the subjects. The characteristics of teaching behavior which seemed most operative in this regard include the following: (1) interjection of teacher knowledge, (2) introduction of generalizations, (3) method of answering questions, (4) control of pupil interaction, and (5) method of eliminating false concepts. Model "discovery" teaching behavior and model "expository" teaching behavior on each of these five characteristics was specified, and a paradigm of teaching techniques for each characteristic was established in each treatment. Adherence to the model techniques of teaching specified for each of the treatments and to the sequence of presentation determined by the instructional materials was assessed by observer- and pupil-rating scales (described hereafter). Scores on these scales were used as an index of teacher fidelity in the presentation of the experimental treatments.

Because of the wide range of ability among classes, teachers were allowed to vary their *rate* of instruction in order to fit the needs of their particular class. (This in no way affected the total time consumed by each treatment, which was held equal, but merely dictated how far each class progressed in the instructional materials.) Teachers were required, however, to cover each concept and principle in the materials carefully, using the prescribed teaching techniques, following the sequence dictated by the materials, and making every attempt to make both treatments equally meaningful. In order to insure adequate presentation of the concepts to both treatment groups, the criterion was established that a minimum of 85 percent of each class must attain a specified minimum level of understanding of each concept before the teacher was allowed to proceed to the next concept.

*Tests and measures*

Ten instruments were developed for this study, nine of which were administered to all subjects while the tenth was used to rate teacher behavior.

Prior knowledge of the selected mathematical concepts was measured by a test (Concept Knowledge Test) administered to both treatment groups in the pretest series. Initial learning was measured by the four subsections of this test administered at the completion of the corresponding subsection of the instructional materials. A parallel form of this test (Concept Retention Test) was administered twice to both treatment groups, once five weeks after instruction and once eleven weeks after instruction, in order to measure retention.[4]

A concept transfer test (Concept Transfer Test) was administered to both treatment groups in the posttest series and was used to evaluate the subjects' ability to recognize and apply mathematical principles in situations unlike those in which they were originally presented. A negative concept transfer test (Negative Concept Transfer Test) was added to the Concept Transfer Test in order to assess the subjects' tendency to overgeneralize the principles to inappropriate situations.

Transfer of heuristics was measured by two tests. The first of these was a paper and pencil discovery test (Written Heuristic Transfer). The second consisted of a sequence of problems presented orally by the teacher, each of which could be solved easily if the subject discovered the "shortcut." On the second test, the final criterion behavior was determined by performance on a six-problem exercise (Oral Heuristic Transfer). Both of these tests were administered in the posttest series to subjects in both of the experimental treatments.

Pupil attitude toward arithmetic was assessed by two attitude scales (Statement Attitude Scale and Semantic Differential Attitude Scale) administered in the pretest series, and again in the posttest series, to the subjects in both treatment groups. The scores from these two scales were summed into a total attitude score (Total Attitude Scale).

In addition to these criterion measures, a questionnaire (Pupil Perception of Teaching Behavior) was administered to subjects in both treatment groups in both the pre- and posttest series on which they recorded, by responding to statements about teaching behavior characteristics of their teacher, their perception of their teacher's behavior along the discovery-expository continuum. This instrument, along with a rating scale (Observer Rating of Teaching Behavior) devised and used to rate teaching behavior through classroom observation and rating from audiotape recordings, was used to assess the degree to which teachers adhered to the prescribed teaching models in each experimental treatment.

The Pintner Intermediate Test, Form A (IQ) and the Metropolitan Achievement Test, Tests 5 and 6 (arithmetic computation and arithmetic problem solving) were used as measures of group comparability.

---

[4] The Concept Knowledge Test represents the summation of four discrete subtests, each of which was administered immediately upon completion of the corresponding subsection of instructional materials. This resulted in a series of four staggered posttests given approximately eight, six, four, and three weeks prior to the first administration of the Concept Retention Test. The four subscores were summed to yield a Total Concept Knowledge Test score. The average delay between administration of the subtests and the first Concept Retention Test was slightly over five weeks. The second administration of the Concept Retention Test came six weeks after the first. Thus, the average time between the subtests and the second retention test was slightly over eleven weeks.

γ

## RESULTS

### *Summary of analyses of teaching behavior*

As indicated in the method section, two instruments were used to gather data on teaching behavior which might be characterized as "discovery" or "expository" in nature. The data thus obtained were analyzed by use of the standard analysis of variance.

The results of analyses of the data obtained with these instruments were interpreted as measures of the degree to which the teachers were actually able to vary their teaching behavior and present both teaching models adequately.

*Observer rating of teaching behavior.*—There were no differences found between teachers in Treatment D or between teachers in Treatment E on their mean ratings on this instrument, nor were there any significant differences between the mean teacher ratings in each treatment and the maximum rating possible if teachers adhered to the prescribed models in each treatment. A significant difference was found between treatments on the mean teacher ratings on the discovery-expository continuum, further validating the proposition that pupils in the two treatments received instruction by two consistently different methods. Table 1 summarizes these four analyses of variance of the data yielded by observer ratings.

TABLE 1
SUMMARY OF ANALYSES OF VARIANCE OF TEACHER RATINGS ON OBSERVER RATING OF TEACHING BEHAVIOR

| COMPARISON | $df_1$ | $df_2$ | $F$ | $P$ |
|---|---|---|---|---|
| 1. Between Treatments D and E | 1 | 71 | 1,061.18 | $< .001$ |
| 2. Between actual ratings and "ideal" ratings for teaching models in D and E | 1 | 71 | .59 | n.s. |
| 3. Between teachers in Treatment D | 7 | 26 | 1.00 | n.s. |
| 4. Between teachers in Treatment E | 7 | 31 | .90 | n.s. |

*Pupil perception of teaching behavior.*—This instrument was used in an attempt to assess pupil perception of teaching behavior on the discovery-expository dimension, both before and after the experimental instructional period. This rating device was scaled so that the pre- to posttest gain score for each teacher in each treatment could be used as an index of the teacher's adherence to the teaching model. In the discovery treatment, high fidelity to the Treatment D model of teaching should have resulted in a positive pre- to posttest gain score. In the expository treatment, high fidelity to the Treatment E model of teaching should have resulted in a negative gain score.

Inspection of the mean pre- to posttest gain score for each treatment revealed changes for each treatment in the predicted direction. An analysis of variance which compared mean teacher gain scores in the two treatments revealed a highly significant difference between the treatments. These data were interpreted as further evidence that the teachers varied their behavior sufficiently to effect a real test of the two teaching models.

No significant differences were found between teacher mean pre- to posttest gain scores within either of the experimental treatments.

Analyses of these data are shown in Table 2.

TABLE 2
SUMMARY OF ANALYSES OF VARIANCE OF TEACHER PRE- TO POSTTEST GAIN SCORES ON PUPIL PERCEPTION OF TEACHING BEHAVIOR

| COMPARISON | $df_1$ | $df_2$ | $F$ | $P$ |
|---|---|---|---|---|
| 1. Between Treatments D and E | 1 | 398 | 25.59 | $<.001$ |
| 2. Between teachers in Treatment D | 7 | 192 | 1.48 | n.s. |
| 3. Between teachers in Treatment E | 7 | 192 | 2.12 | n.s. |

*Summary of tests of hypotheses*

Because of the noncomparability of the treatment groups on several pretreatment measures, statistical controls were imposed in all intertreatment data analyses (except analyses of teaching-behavior data discussed previously) by use of a two-way teacher-by-treatment analysis of covariance.

The choice of covariates was determined by an examination of the intercorrelations on all measures and variables. On this basis, IQ, arithmetic computation, and arithmetic problem solving were used as constant covariates in the analysis of each dependent variable. Pretest scores were used as additional covariates in analysis of the posttest of each instrument administered in both the pre- and posttest series. Posttest scores on the Concept Knowledge Test were used as an additional covariate in the analysis of the Concept Retention and Concept Transfer tests.[5]

This analysis yielded significant $F$ ratios for between-teacher effects and teacher-by-treatment interaction on all of the criterion measures. No attempt to explain these findings is given here; several plausible explanations are included in previous reports of this research (see the introductory footnote). Only the results yielded by direct comparisons between Treatments D and E are presented here.

*Initial learning.*—The data yielded by the Concept Knowledge Test did not support the hypothesis that Treatment D would produce superior results on an initial learning test. On the contrary, these data showed Treatment E to produce significantly better results than Treatment D on the initial learning criterion test.

*Retention.*—The hypothesis that Treatment D would produce superior results to Treatment E on a retention test given five and eleven weeks after instruction was supported by the evidence yielded by an analysis of the Concept Retention Test scores ($p < .05$ on the first administration and $p < .025$ on the second administration).

*Concept transfer.*—The data yielded by the Concept Transfer Test lent tenuous support to the hypothesis that pupils in Treatment D would show greater ability to transfer the concepts learned during instruction than would pupils in Treatment E.

*Negative concept transfer.*—There was no support in the data yielded by the Negative Concept Transfer Test for the hypothesis that Treatment D would produce less negative transfer than Treatment E. Rather, it was found that there were no differences in negative transfer between Treatment D and Treatment E.

*Attitude.*—Of the three possible comparisons between Treatments D and E on measures of attitude, none reached significance at a minimum acceptable level of significance. The hypothesis that Treatment D would

---

[5] It is questionable whether a legitimate test of transfer potential could have been obtained in this study without equating original learning, as indicated by performance on the Concept Knowledge Test, for the two treatments. Therefore, this covariate was included in order to obtain an estimate of what performance on the Concept Retention and Concept Transfer tests would have been if performance of the E and D groups had been equivalent on the Concept Knowledge Test.

Statisticians, however, are divided on the use of this technique. Some argue that it is not legitimate to use a covariant which has been effected by the treatments. These statisticians would prefer to use absolute measures of performance rather than a treatment-effected covariant or any variation of difference-score techniques. At present, this methodological issue seems to remain largely unsolved.

produce superior results to Treatment E on attitude measures was rejected.

*Transfer of heuristics.*—The hypothesis that Treatment D would produce superior results to Treatment E on tests of pupil ability to transfer heuristics was supported by the evidence yielded by analyses of both the Written Heuristic Transfer and the Oral Heuristic Transfer test scores.

Table 3 summarizes the analyses of covariance which yielded the above results.

TABLE 3
SUMMARY OF ANALYSES OF COVERAGE OF CRITERION MEASURE POSTTEST SCORES: BETWEEN TREATMENTS D AND E

| MEASURE | $df_1$ | $df_2$ | $F$ | $P$ | DIRECTION |
|---|---|---|---|---|---|
| Concept Knowledge Test | 1 | 412 | 7.435 | $<.01$ | D < E |
| Concept Retention Test 1 | 1 | 412 | 3.918 | $<.05$ | D > E |
| Concept Retention Test 2 | 1 | 412 | 5.868 | $<.025$ | D > E |
| Concept Transfer Test | 1 | 412 | 3.089 | $<.10$ | D > E |
| Neg. Concept Trans. Test | 1 | 413 | .098 | n.s. | |
| Sem. Diff. Attitude Scale | 1 | 412 | .161 | n.s. | |
| Statement Attitude Scale | 1 | 412 | 1.173 | n.s. | |
| Total Attitude Scale | 1 | 412 | 2.057 | n.s. | |
| Written Heuristic Trans. | 1 | 413 | 5.004 | $<.05$ | D > E |
| Oral Heuristic Trans. | 1 | 413 | 5.720 | $<.025$ | D > E |

## DISCUSSION AND CONCLUSIONS

### *Teaching behavior*

Of most importance for the interpretation of the results of this study was the clear-cut evidence that the subjects in the two experimental treatments received instruction by two consistently different methods of teaching, each of which closely paralleled the particular model prescribed. It can be concluded that both treatments were fairly presented and that no factors operated which would tend to give either method an unfair advantage. Although the necessity of experimental controls may have precluded either method from reaching its optimum power, this factor, if present, was equally operative in both experimental treatments.

*Tests of hypotheses*

In general, the findings of this study support many of the claims made by proponents of discovery methods. The most dramatic finding was the rather startling reversal in rank of Treatments D and E between the administration of the Concept Knowledge posttest and the first administration of the Concept Retention Test five weeks later. Although Treatment E was significantly superior to Treatment D on the tests of initial learning ($p < .01$), the retention test given after an average five-week delay showed Treatment E not only to have lost this initial superiority but also to have been surpassed by Treatment D. The pupils taught by the discovery method were able to retain significantly more material ($p < .05$) over the intervening period, notwithstanding the fact that they had evidenced knowledge of significantly less material than the Treatment E group on the test of initial learning. Analysis of the scores from the second administration of the Concept Retention Test eleven weeks after instruction showed pupils in Treatment D to have maintained this advantage over pupils in Treatment E ($p < .025$). This finding strongly suggests that presentation of mathematical concepts to sixth-grade pupils by techniques of discovery teaching causes the learner to conceptually integrate the content in such a manner that he can retain it more readily than if the concepts had been presented to him by an expository teaching method.

Another finding which clearly favors Treatment D is that dealing with subject acquisition of a problem-solving set. In light of the evidence yielded by both the Written Heuristic Transfer and the Oral Heuristic Transfer tests, it seems reasonable to conclude that learning by discovery techniques significantly increases pupil ability to use discovery problem-solving approaches in new situations, both those which require paper and pencil application and those which involve verbal presentation by the teacher. Treatment D was shown to be significantly superior to Treatment E on both of these dimensions in the present study.

Treatment D also seems superior to Treatment E in terms of transfer of mathematical concepts, although this finding is somewhat tenuous. It was the experimenter's opinion that the Concept Transfer Test was much too difficult for the subjects involved and that this factor resulted in random errors of measurement which reduced the possibility of finding more significant differences between the treatments. The obtained between-treatment $F$ ratio in the teacher-by-treatment analysis of covariance favored Treatment D over Treatment E at a minimum acceptable level of significance ($p < .10$) and the experimenter would speculate that modifications of the instrument to reduce the random error of measurement would result in more highly significant differences in favor of Treatment D.

The results yielded by the attitude measures were somewhat equivocal. None of the comparisons between Treatments D and E reached the .10 level of significance, although the differences were all in the direction predicted. A postexperimental evaluation of the research project also yielded provocative, although subjective, results related to the above. Among other questions, the eight experimental teachers were asked which of their two classes seemed to like the "new math" better. Six of the eight teachers responded that their Treatment D group gave considerably greater expressions of liking the new arithmetic program than did their Treatment E class. The remaining two teachers indicated that both of their classes seemed to like the arithmetic content equally well. This overall judgment was corroborated by the three rater-observers. In addition, several factors existed during the experiment which, if operative, would tend to negatively affect the attitudes of pupils in Treatment D toward arithmetic while not affecting the attitudes of pupils in Treatment E.[6] While not offered as conclusive evidence, these opinions were judged by the experimenter to be sufficiently perturbing to point to the need for future research specifically designed to test further the relative effects of discovery and expository methods on pupil attitude.

Although not a specific hypothesis, the question of relative practicality of discovery and expository teaching in terms of time consumption was of particular interest in this study, and controls were established to enable this question to be answered. The results indicate that the discovery method need not be more time consuming than the expository method of instruction. When given an equal amount of time to work on the learning task, pupils in Treatment D proved superior to pupils taught by Treatment E, in the majority of intertreatment comparisons. No support was found in this study for the notion that discovery is inherently more time consuming than expository instruction.

*Implications*

Implications which have been drawn from both experience in this study and an analysis of its results are of two types, implications for future research and implications for educational practice.

*Implications for future research.*—Replications of this study should be conducted (1) at other grade levels to test the generalizability of the results

[6] These factors are discussed in detail in previous reports of this research, listed in the introductory footnote.

and conclusions to other age groups; and (2) with more discriminating attitude measures and experimental controls specifically designed to test the effects of the two methods upon attitude toward the selected subject-matter content.

Programmatic research dealing with various discovery-expository variables of task presentation should be initiated. In addition to a continuation of research in which sequence characteristics of the learning task are manipulated, the present research design and instructional materials might be modified to provide tests of the relative effectiveness of various types and amounts of guidance along the discovery-expository dimension. Studies could be designed in which the present instructional materials are used to compare guided discovery with independent discovery. Further modifications of the present design and learning task could serve to compare discovery methods in which the verbal factor is varied from verbal to nonverbal discovery. Interrelationships among these relevant variables might then be explored.

*Implications for educational practice.*—Any generalizations based on the findings of this study must take into account the particular teachers, experimental population, instructional procedures, instructional materials, and criterion measures used. In addition, without the programmatic research suggested above, any conclusions drawn on the basis of this single study must be tentative at best. Furthermore, while many of the results of this study are statistically significant, the question of practical significance remains largely unanswered.

Conversely, this study was conducted under carefully controlled conditions which were judged to approximate normal classroom conditions with respect to all dimensions except those specifically varied for experimental purposes. Because of the relatively large time sample, the nature of the learning task, and the large number of subjects used, it would seem that the results can be generalized, at least to innovative teaching with similar subjects and subject-matter content, with a relatively high degree of confidence. Within this context, it is the experimenter's opinion that, pending further programmatic research, this study holds the following implications for educational practice:

1. To the extent that pupil ability to retain mathematical concepts and pupil ability to transfer heuristics of problem solving are valued outcomes of education, discovery techniques of teaching should be an integral part of the methodology used in presenting mathematics in the elementary classroom.

2. To the extent that immediate recall is a valued outcome of educa-

tion, expository instruction should be continued as the typical instructional practice used in the elementary classroom.

3. The present study also suggests that pupils' ability to transfer concepts will likely be increased in proportion to the degree to which discovery techniques are used in the classroom.

## REFERENCES

BELLACK, A. A., *et al. The Language of the Classroom: Meanings Communicated in High School Teaching.* New York: Bureau of Publications, Teachers College, Columbia University, 1963.

CORMAN, B. R. "The Effect of Varying Amounts and Kinds of Information as Guidance in Problem Solving," *Psychological Monographs,* LXXI, No. 2 (1957).

CRAIG, R. C. *The Transfer Value of Guided Learning.* New York: Bureau of Publications, Teachers College, Columbia University, 1953.

DELLA-PIANA, G. M.; ELDREDGE, G; and WORTHEN, B. R. *Sequence Characteristics of Text Materials and Transfer of Learning, Part I: Experiments in Discovery Learning.* Salt Lake City: Bureau of Educational Research, University of Utah, 1965.

GAGNÉ, R. M., and BROWN, L. T. "Some Factors in the Programming of Conceptual Learning," *Journal of Experimental Psychology,* LXII (1961), 313-21.

HASLERUD, G. M., and MEYERS, SHIRLEY. "The Transfer Value of Given and Individually Derived Principles," *Journal of Educational Psychology,* December 1958, pp. 293-98.

KATONA, G. *Organizing and Memorizing.* New York: Columbia University Press, 1940.

KERSH, B. Y. "The Adequacy of 'Meaning' as an Explanation for the Superiority of Learning by Independent Discovery," *Journal of Educational Psychology,* XLIX, No. 5 (1958), 282-92.

KITTELL, J. E. "An Experimental Study of the Effect of External Direction During Learning on Transfer and Retention of Principles," *Journal of Educational Psychology,* XLVIII (1957), 391-405.

McDONALD, F. J. "Meaningful Learning and Retention: Task and Method Variables," *Review of Educational Research,* XXXIV, No. 5 (December 1964), 530-44.

SCANDURA, JOSEPH M. "An Analysis of Exposition and Discovery Modes of Problem Solving Instruction," *Journal of Experimental Education,* XXXIII, No. 2 (1964), 149-59.

SOBEL, M. A. "A Comparison of Two Methods of Teaching Certain Topics in Ninth Grade Algebra," *Dissertation Abstracts,* XIV (1954), 1647.

STACEY, C. L. "The Law of Effect in the Retained Situation with Meaningful Materials," *Learning Theory in School Situations.* (University of Minnesota Studies in Education, No. 2.) Minneapolis: The University of Minnesota Press (1949), pp. 74-103.

WITTROCK, M. C. "Verbal Stimuli in Concept Formation: Learning by Discovery," *Journal of Educational Psychology,* LIV (1963), 183-90.

WITTROCK, M. C., and KEISLER, E. R. "Verbal Cues in the Transfer of Concepts," *Journal of Educational Psychology,* LVI, No. 1 (February 1965), 16-21.

WORTHEN, B. R. "Discovery vs. Expository Classroom Instruction: An Investigation of Teaching Mathematics in the Elementary School." Unpublished Master's thesis, University of Utah, 1965.

# VI

# Evaluation of *Experiences in Mathematical Discovery**

EMIL J. BERGER
*St. Paul Public Schools*
*St. Paul, Minnesota*

THOMAS A. HOWITZ
*University of British Columbia*
*Vancouver, British Columbia*

In December 1962 the Board of Directors of the National Council of Teachers of Mathematics authorized an expenditure of $40,600 to finance a General Mathematics Writing Project to produce text materials for non-college-bound ninth-grade students in the 25th to 50th percentile range in mathematical achievement. The following summer twelve writers, working under the direction of Dr. Oscar Schaaf, completed the preliminary edition of a text entitled *Experiences in Mathematical Discovery (EMD).* The preliminary edition of *EMD* was multilithed and bound in two volumes. A *Teacher's Commentary* accompanied the text.

The preliminary edition of *EMD* contains nine chapters having the following titles:

1. Patterns, Formulas, and Graphing Data
2. Arrangements and Selections
3. Intuitive Geometry
4. A New Look at Whole Numbers
5. Ratio, Proportion, and Per Cent
6. Learning to Use Directed Numbers
7. Measurement

---

* Dr. David R. Giese, Director of Research, General College, University of Minnesota, served as statistical consultant in analyzing the data collected during the course of this study.

8. Mathematical Thinking in Geometry
9. Fraction Numbers

As the titles indicate, each chapter involves significant mathematical ideas. The applied aspects of mathematics are stressed, and there is much new material (not just a review of old topics).

The style of exposition is based on the discovery approach. Also, the presentation in each chapter proceeds in such a way that the student is not compelled to give prolonged attention to long systematic developments. Another important characteristic of *EMD* is that practice work is incorporated as an integral part of the content development.

To determine the effectiveness of the preliminary edition of *EMD* an experimental evaluation was carried out during the 1963/64 school year. Comparisons were made between ninth-grade general mathematics classes using *EMD* and comparable classes using conventional ninth-grade general mathematics textbooks. Particular attention was given to comparisons involving student achievement in mathematics, and to student change of attitude toward mathematics. The reason for carrying out the evaluation with classes of ninth-grade general mathematics students is that students normally registered in such classes provided the best available approximation of the population for which *EMD* was written (i.e., 25th to 50th percentile range in mathematical achievement).

## METHOD

### *Sample*

The sample used in the study consisted of 86 ninth-grade general mathematics classes located in various parts of the United States.[1] The 86 classes were taught by 43 teachers, each teaching two of the classes in the sample. Selecting the sample involved finding schools such that each school had a teacher who was scheduled to teach *two* classes in ninth-grade general mathematics during the school year 1963/64. Thus, one class for each teacher served as an experimental class and the other as a conventional control class.

During the course of the study 14 pairs of classes (one pair for each of 14 teachers) were eliminated from the study for reasons that are explained in appropriate paragraphs of this report. Data from the remaining 29 pairs of classes, taught by 29 teachers, were analyzed in accordance

[1] The evaluation of the preliminary edition of *EMD* was also carried out with several tenth-grade general mathematics classes, but the present report is limited to the evaluation conducted with the ninth-grade general mathematics classes.

with the purposes of the study. This means that data for 29 experimental classes and 29 conventional control classes were analyzed.

*Instructional materials*

Each experimental class was provided with a class set of the preliminary edition of *EMD*, and the teacher was provided with a copy of the accompanying *Teacher's Commentary*. Each conventional control class had available the ninth-grade general mathematics textbook that was in normal use in the school in which the class was located. Although the participating schools used nine different conventional textbooks, the majority of the conventional control classes used either Stein's *Refresher Arithmetic* or Hart's *Mathematics in Daily Use.*

*Measuring instruments*

The School and College Ability Test (Form 3A) was administered as a pretest to all students, in order to obtain a measure of initial scholastic ability and also to determine whether or not the experimental and conventional control classes taught by each teacher were comparable in scholastic ability.

The Sequential Test of Educational Progress (Mathematics—Form 3A) was used as a pretest and the Sequential Test of Educational Progress (Mathematics—Form 3B) was used as a posttest for all students in both the experimental and the conventional control classes. The two different forms were used to obtain a measure of gains in mathematical knowledge resulting from participation in one of the two kinds of classes.

The School and College Ability Test (SCAT) and the Sequential Test of Educational Progress (STEP) were selected as measuring instruments for the following reasons: (1) SCAT provides a measure of both verbal ability and quantitative ability; (2) STEP, while considered a test of mathematical achievement, measures mastery in most of the broad mathematical concepts; (3) STEP and SCAT are widely used and are readily available; (4) national norms for both instruments are available and many schools have established their own local norms; and (5) both instruments have been used in mathematics curriculum studies and have been accepted by many researchers as valid and reliable instruments. Although STEP was published in 1957 by the Educational Testing Service, it was developed in the few years prior to that date. In view of the present trend in mathematics curriculum development the items in STEP would have to be classified as conventional (or traditional). Hence, some of the newer concepts presented in the preliminary edition of *EMD* could not be tested by STEP. Therefore, it is possible that students using the conventional textbooks had a slight advantage on STEP.

Since formation of favorable attitudes toward mathematics is generally considered to be a desirable outcome of instruction in mathematics, an effort was made to measure student attitude toward mathematics. The Mathematics Inventory, a test developed by Cyril J. Hoyt and Donald G. MacEachern at the University of Minnesota in 1958, was selected as the most appropriate instrument available for measuring student attitude toward mathematics. This test was administered both as a pretest and as a posttest to all experimental and conventional control classes to determine student attitude change toward mathematics.

The Mathematics Inventory was designed for use with junior high school students. Reliability and validity coefficients have been computed. The test consists of 110 statements to each of which the student is asked to respond in one of three ways: "agree," "uncertain," or "disagree." The test is machine-scorable.

If the Mathematics Inventory is used both as a pretest and as a posttest, the difference scores that are obtained can be interpreted as a measure of attitude change for an instructional period. High attitude scores have been shown to be indicative of the likelihood that students will elect further courses in mathematics and science. An acceptable attitude toward mathematics is, in itself, an important aspect of achievement.

To obtain a measure of student mastery of topics included in *EMD* a General Mathematics Achievement Test (GMAT—unpublished) was constructed and administered as a posttest. Test items were submitted to the investigators by members of the Advisory Committee of the General Mathematics Writing Project. Four mathematics educators rated the items that were collected, and the fifty considered to be the most suitable were incorporated in GMAT. In producing GMAT four criteria were established:

1. The test had to be objective.

2. The test had to have content validity.

3. Chapter sampling had to be representative; and there had to be a balance among items involving problem solving or interpretation and those involving recall of factual information.

4. Language usage peculiar to either conventional textbooks or to *EMD* had to be neutralized. This meant including definitions of words that were not common to both treatments.

Basically, the purpose of GMAT was to determine whether or not the use of *EMD* enabled students to learn what the writers of *EMD* had intended that students should learn.

*Experimental procedure and method of analysis*

During the early summer of 1963, as already indicated, 43 teachers, each scheduled to teach two ninth-grade general mathematics classes, were selected to participate in the evaluation of the preliminary edition of *EMD*. Instructions sent to each teacher emphasized two things: (1) that it was desirable for a teacher's experimental class and his conventional control class to be as much alike as possible, and (2) that the two classes were to be taught separately—that is, *EMD* was to be used only with the experimental class and the conventional textbook in normal use in the teacher's school was to be used only with the conventional control class.

In August 1963 each participating teacher was furnished with one class set of each of the following tests: SCAT (Form 3A), STEP (Mathematics—Form 3A), and the Mathematics Inventory. Also furnished were enough answer sheets and electrographic pencils for both of the teacher's classes. Detailed instructions were provided to insure uniformity of administration of the tests. Thirty-eight teachers administered SCAT (Form 3A), STEP (Mathematics—Form 3A), and the Mathematics Inventory to participating classes and returned the completed testing materials to the investigators.

Five teachers did not return results for the fall testing. Besides this, it was learned that in five other cases the same teacher had not been assigned to teach both an experimental class and a conventional control class. The classes involved in the two kinds of situations described were therefore dropped from the evaluation. This meant a reduction of ten pairs of classes in the anticipated sample size.

In April 1964 testing materials were again sent out, this time to each participating teacher who had correctly followed directions up to this point, and was therefore assumed to be actively participating in the evaluation. During the year the investigators had been notified that one pair of classes had been disbanded because of a school reorganization. This pair of classes was dropped from the evaluation. In all, testing materials were sent to 32 teachers. As in the fall, each teacher was provided with all materials and complete instructions for administering the tests. Thirty-one teachers administered STEP (Mathematics—Form 3B), the Mathematics Inventory, and the General Mathematics Achievement Test (GMAT) in accordance with instructions, and returned the completed materials. Testing materials for one pair of classes were never returned. After the test results were machine-corrected, all students who had *not* completed all parts of the testing program were eliminated from the evaluation. Elimination of students for this reason meant elimination of one more pair of classes from the evaluation because too few students remained in one of the two classes in the pair. When the spring testing was completed, there were 30 pairs of classes for which sufficient data were available for the analysis.

As a preliminary step in carrying out the analysis of the experiment, frequency distributions of the STEP (Mathematics—Form 3A) pretest scores and the SCAT (Form 3A) scores were developed. Upon examination of the means of the distributions it was realized that "low ability" mathematics students are not homogeneous with respect to SCAT and STEP scores. Some students who were classified as "low" in one school would have been classified as "very good" in another school. For example, the means for the experimental class of one particular teacher were 34 on SCAT (Form 3A) and 18 on STEP (Mathematics—Form 3A), while the means for the experimental class of a second teacher were 59 on SCAT (Form 3A) and 28 on STEP (Mathematics—Form 3A). Even more unusual than the differences of these means was the fact that the lowest student in the second teacher's class was above the highest student in the first teacher's class. Although not as extreme, there were large differences between the experimental and conventional control classes for several other teachers.

In view of the foregoing information it was decided that if the SCAT (Form 3A) mean scores for a teacher's two classes differed by more than ten points or if the mean scores on STEP (Mathematics—Form 3A) differed by more than five points, this teacher's classes would not be included in the primary analysis. This decision was made to insure comparability of the classes in each pair for which data were to be analyzed in the primary analysis. On the basis of this preliminary assessment of the data, six pairs of classes were excluded from the primary analysis and were placed in a special group. The data for each of these six pairs of classes were analyzed separately. Finally, one pair of classes was dropped from all analyses because the mean scores for one of the two classes in the pair were far below those of all other classes.[2] However, even after separating out the six pairs of classes described above and dropping one pair of classes entirely, it was still necessary to cope with rather large differences between schools for the remaining 23 pairs of classes.

Because of these differences it was felt that it would be impossible to

[2] This completes the accounting of the 14 pairs of classes that were reported as being eliminated from the original sample.

analyze classes simultaneously. Blocking on either the SCAT (Form 3A) scores or the STEP (Mathematics—Form 3A) pretest scores or both was considered, but rejected because of the inability to find suitable blocking scores which would not result in empty cells. Multidimensional covariance analysis was considered; however, the effect of nonhomogeneity of regression coefficients, which must have existed but which was not tested, was unknown. Instead, it was decided to group the 23 pairs of classes into five groups of approximately equal size, based on their SCAT (Form 3A) mean scores. In this way five groups, three of which contained five pairs of classes and two of which contained four pairs of classes, were constructed. The range of SCAT means for each group is given in the table below. The primary analysis was carried out separately for each of these five groups. To increase the precision of the analysis the students in the experimental and conventional control classes in each group were divided into two initial knowledge levels (low and high) using their STEP (Mathematics—Form 3A) pretest scores. The dividing scores for each of the five groups are shown in Table 1.

TABLE 1

| GROUP | NUMBER OF PAIRS OF CLASSES | SCAT SCORES RANGE OF MEANS | DIVISION POINT ON STEP PRETEST SCORES | |
|---|---|---|---|---|
| | | | LOW | HIGH |
| I | 5 | 31 – 37 | 17 | 18 |
| II | 4 | 38 – 44 | 19 | 20 |
| III | 5 | 45 – 48 | 21 | 22 |
| IV | 4 | 49 – 52 | 23 | 24 |
| V | 5 | 53 – 60 | 25 | 26 |

The scores on two of the posttests (STEP Mathematics—Form 3B and GMAT) and the Mathematics Inventory difference scores (gains) for each of the five groups were analyzed separately. In each case a three-way unweighted means analysis of variance was used to determine the effects of each of the following factors:

1. Treatment (two types, experimental and conventional)
2. Initial mathematical knowledge (two levels for each of the five groups identified in the table above)
3. Teacher (four or five depending on the group)

Besides determining the effects of the three factors described above, all interactions of the three factors were also tested. The test scores of the students in the six pairs of classes that did not fit into the primary analysis as described above were analyzed separately, using analysis of covariance.

Because the amount of variation among schools and between two classes within a school cannot always be anticipated, the final analysis was quite different from that which was originally planned.

## Summary of Results

*Conclusions pertinent to attitude*

1. *Treatment effects.*—Change of student attitude toward mathematics due to the experimental treatment was not significantly different from the change of student attitude toward mathematics due to the conventional control treatment.

2. *Initial knowledge effects.*—Examination of the $F$ ratios and the mean scores indicated that students with more initial mathematical knowledge not only received significantly higher attitude scores on the pretest but also raised their attitude scores significantly more during the year.

3. *Teacher effects.*—The change in attitude during the year was related to the teacher.

4. *Interactions.*—Only isolated significant interactions were identified in the analysis.

*Conclusions pertinent to mathematical knowledge as measured by the STEP posttest*

1. *Treatment effects.*—The treatment posttest results as determined by STEP (Mathematics—Form 3B) were not significantly different within groups, thereby indicating that both treatments were about equally effective in teaching what STEP (Mathematics—Form 3B) measures. The actual differences among groups were as expected, the groups with the higher SCAT scores getting the higher scores on the STEP posttest.

2. *Initial knowledge effects.*—There were large significant differences between levels on the STEP posttest. Students who knew more at the beginning of the experiment as measured by the STEP (Mathematics—Form 3A) pretest also made the higher scores on the STEP (Mathematics—Form 3B) posttest.

3. *Teacher effects.*—There were no consistent differences among teachers within any group; however, there were large differences among all teachers in the study.

4. *Interactions.*—There were no consistent interactions among the factors.

*Conclusions pertinent to the experimental material as measured by the General Mathematics Achievement Test (GMAT)*

1. *Treatment effects.*—There was a significant difference between the two treatments, with the experimental classes getting the higher scores.

2. *Initial knowledge effects.*—The students who knew more in the beginning of the experiment, as measured by STEP (Mathematics—Form 3A), earned significantly higher scores on GMAT.

3. *Teacher effects.*—There were significant differences among the teachers. The differences were complicated by a significant teacher-treatment interaction. This indicates that the treatment difference was not consistent with teachers.

4. *Interactions.*—Except for the treatment-teacher interaction discussed above, no interactions were consistently significant.

## DISCUSSION

The analysis indicated that the use of *EMD* had little, if any, differential effect either on attitude, as measured by the Mathematics Inventory, or on mathematical knowledge, as measured by STEP (Mathematics—Form 3B). However, it was apparent that students in the classes using *EMD* learned something that was not taught in the classes using conventional textbooks. The nonsignificant interaction of treatment and initial knowledge on the General Mathematics Achievement Test (GMAT) indicates that the better students in each group learned more than the poorer students, regardless of the teacher or the text materials that were used.

During the experimental tryout each participating teacher was asked to submit reports on each chapter of *EMD*. These reports included questions concerning mathematical content, difficulty of the material, time spent on each section of the chapter, and general opinion. Reports received indicated that the teachers were favorably disposed toward *EMD*. However, they were relatively noncommittal about the choice of topics and the mathematics contained in the text.

Copies of the preliminary edition of *EMD* were sent to 15 mathematicians and educators who were asked to give detailed chapter-by-chapter appraisals of the text. The appraisals suggested that the preliminary edition of *EMD* placed too much emphasis on the discovery approach and that more formalization of those mathematical concepts that students are expected to discover might be needed. The reviewing group also indicated that the preliminary edition of *EMD* might need strengthening in the development of basic mathematical concepts before it is published in final form.

Information on the readability of the preliminary edition of *EMD* was obtained by using the Flesch reading-ease formula adapted for mathematical materials. Samples selected showed that the reading levels of the chapters ranged from a grade level of 8.0 to a grade level of 11.5. It was estimated that the reading level of the students for whom this text was intended should probably be between a grade level of 7.0 and 8.0.

As a result of the statistical evaluation, the information obtained from the chapter reports submitted by participating teachers, the reviews obtained from mathematicians and educators, and the reading level study, the preliminary edition of *EMD* is now being revised.[3]

Evaluation of text materials, although long and involved, is necessary if the material produced is to be of value to the intended student population.

[3] The completed revision will consist of ten independent units. Each will be separately bound. Five of these units have already been completed and are now available from the National Council of Teachers of Mathematics, Washington, D.C.

## REFERENCES

BERGER, EMIL J. "*Experiences in Mathematical Discovery:* Preliminary Evaluation," in *The Low Achiever in Mathematics.* Washington, D.C.: U.S. Office of Education, 1965.

HOWITZ, THOMAS A. *The Discovery Approach: A Study of Its Relative Effectiveness in Mathematics.* Ph.D. thesis, University of Minnesota, Minneapolis, 1965.

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS. *Experiences in Mathematical Discovery* (preliminary edition). Washington, D.C.: The Council, 1963.

SCHAAF, OSCAR F. "Preparing the MNCB Text *Experiences in Mathematical Discovery*," in *The Low Achiever in Mathematics.* Washington, D.C.: U.S. Office of Education, 1965.

# VII

# Individualized Instruction in Elementary Mathematics*

JOSEPH I. LIPSON
*Learning Research and Development Center*
*University of Pittsburgh*
*Pittsburgh, Pennsylvania*

For the past two years, the Learning Research and Development Center has been involved in the development of an innovative system of mathematics instruction for the elementary school, Grades K–6. The purpose of the program is to allow each child to progress through the curriculum at his own rate and to reach objectives by means of tasks assigned on the basis of his unique abilities (Bolvin, 1966). The basic components of the system are (1) a sequential curriculum stated in terms of what the student is expected to do at each stage, (2) placement and diagnostic tests to determine what instruction shall take place, and (3) lessons (e.g., work-page assignments or teacher-directed activities).

## Method

### *Population*

The program has been in operation in 1964/65, 1965/66, and 1966/67 in the Baldwin-Whitehall school district near Pittsburgh. Approximately 220 children, who live in the immediate neighborhood, are enrolled in this school. The neighborhood is characterized by sociologists as lower-middle class, although the area consists exclusively of one- or two-bedroom single-family dwellings. There are only three managerial-class families and one truly poor family with children in the school.

---

* The research and development reported herein was performed pursuant to a contract with the U.S. Office of Education, under the provisions of the Cooperative Research Program. More information about this project, and the detailed curriculum, may be obtained by writing the author.

*Objectives*

The list of objectives is categorized by topic, such as addition or multiplication, and sequenced according to difficulty and prerequisite conditions. In total there are about 385 objectives, grouped into 85 units by topic and by level of difficulty. Many objectives are not "terminal objectives," in the sense that one would like all elementary school graduates to be able to display mastery of them. They are placed in the curriculum as "subordinate objectives," because it is believed that eventual mastery of these intermediate tasks is prerequisite to later mastery of other important mathematical concepts. For example, the children are expected to be able to say different names for the same number (e.g., $8 + 5 = 8 + (2 + 3) = (8 + 2) + 3 = 10 + 3 = 13$), in order to prepare them for such things as the associative law, rather than as an end in itself.

The objectives conform to what we call "classical new math." Once the cardinal and ordinal properties of number are abstracted from counting and matching operations with real objects, the laws of arithmetic are developed and then used to make the more complex operations and algorithms reasonable, and retraceable to the basic counting operations. Many programs in current use are built along the same lines.

*Tests*

Once the objectives were agreed upon, the next step was to evolve a set of tests and a set of instructional materials. Given the objectives, the test-writing was a fairly straightforward matter. Three kinds of tests have been developed under the direction of Dr. Richard Cox: (1) broad scale placement tests, (2) detailed diagnostic achievement pre- and posttests, and (3) curriculum-embedded tests. Since these tests are critical to the individualization procedures, let us consider each briefly.

Each *placement test* covers an entire topic in arithmetic, e.g., addition. At each level of difficulty in a given topical area (there are eight levels of difficulty in the program), test items were written in sufficient number to test general capabilities at that level. The tests were kept short enough so that the entire battery of twelve placement tests could be given in one week. In this way a placement profile for each child in the entire school can be completed within one week after classes start.

After the placement profile is completed for a student, he is given the *diagnostic pretest* for the lowest hierarchical unit in which his placement test indicates lack of competence. For example, if a student tests at the D-level of difficulty in all areas except multiplication, and if he indicated inadequacy in multiplication at the C-level, he would be given a diagnostic pretest in the C-Multiplication unit. If he shows lack of mastery, instruction will begin through individually assigned instructional tasks. After he has completed instruction in the objectives of C-Multiplication, the student would take a *post*test in C-Multiplication which is simply an alternate form of the pretest.

These pre- and posttests for each unit are called diagnostic achievement tests because each objective in the unit is tested by a sufficient number of items not only to indicate general mastery, but also to determine the specific operations which the child cannot perform. Thus, performance on these tests forms the basis for the individual instructional assignments.

Finally, in order to keep an up-to-date record of each student's progress, there are *curriculum-embedded tests.* These tests are given periodically as the student works through a unit to determine whether the ongoing instruction is effective and whether the student is able to apply prerequisite skills to new instructional tasks.

*Materials*

The materials were originally obtained from those commercial programs which seemed to most closely follow our objectives, e.g., GCMP. However, on the basis of information on student performance the curriculum-development staff, in cooperation with the teachers, has continually revised and added to this material. Today, 30 percent of the roughly 4,000 pages in use have been written by the teachers and center staff.

In preparing and revising the materials, the following sequence of operations has been followed. Initially, six sets of commercial workbook materials were bought. Each page was identified with one of the objectives in our program. Then, another set of commercial workbooks was cut up and all the pages identified with a given objective were assembled. These pages constitute the material which can be assigned for instruction on that objective.

In many cases, there were few or no pages from commercial sources for a given objective. In such instances, instructional materials were prepared by the staff. Once children are entered into the program, the pre- and posttest results provide continual information as to which materials are not providing adequate instruction. Whenever this happens, the teachers and students are interviewed, and the test results are examined by item in order to decide what the instructional problem is. After a decision is made, suggested new approaches are prepared and tried. Upon successful trial, the new approach is written up and installed as new additional material, or as replacement material. In this way thousands of new pages have been written in response to ongoing instructional problems. More important, as the most obvious problems are solved, we are able to turn our attention to other dimensions of the instructional materials. Thus, a system of ongoing revision, in which it may take as little as a month to go from an identified problem to the installation of new materials, promises to provide a wide variety of instructional approaches that can be used differentially so that an effective learning path can be found for each student.

### *Instructional procedures*

The instructional procedure revolves around diagnostic testing and daily assessment and assignment of work for each student. The idea is to make sure that no student ever receives instruction on an objective which he has already mastered while, at the same time, his instruction is always *based* on skills which he has mastered.

The student is first *placement*-tested to find the *general level* in each area at which he begins to show difficulty. More detailed *pretests* are then administered, starting from the lowest unit in the hierarchy until a unit is encountered in which the student shows lack of mastery of the objectives. The pretest is then examined to show which operations the student is unable to perform.

The student at this point goes into his daily work pattern. There is a large folder for each student with information on both his past performance and his current work assignment. The information contained in the folder is (1) his placement profile, (2) the record of all his work from the beginning of the year—units mastered, pre- and posttest scores, dates and days to complete each unit, and (3) his assignment sheet for the current unit of work.

His assignment sheet for the current unit of work includes (*a*) his pretest scores broken down with a score for each objective in the unit, (*b*) the teacher's decision as to which objectives need work, and (*c*) the list of assigned workpages and curriculum-embedded tests along with the student's score on each assigned page. Those pages the student is currently working on are also to be found in the folder.

Let us go through a cycle of evaluation, work assignment, and actual work by the pupil. At the end of a class each child puts his folder in a box in the classroom. The teacher then evaluates each folder before the next class. The folders are separated according to whether the student (1) needs a test for the next period, (2) needs additional workpages assigned, or (3) has sufficient work for the next period. If the student needs a test, it is a simple matter to have this in the folder before the next class period. If the student needs a new assignment, the student's immediate past work and his entire record are examined. On the basis of this information and the teacher's general assessment of the student's ability, a new work assignment (usually of the order of five pages of work) is made. If the student needs no additional work, the teacher need only decide whether the student is making sufficient progress or whether personal attention is required. Ideally, the student's progress is evaluated daily, and new assignments are made on the basis of past performance.

At the beginning of the next hour of arithmetic the children get their folders. For young children (first and second graders), the pages will have already been put in their folders by a clerk. The older children will note the pages assigned and get the pages themselves from a storage area immediately outside their classrooms. Each child then begins to work on his individual assignment. During the period he may (1) need help, (2) need work pages scored, (3) need a new assignment. If a student needs help, the teacher comes to that student and helps him personally; if a new assignment is needed, the teacher makes the new assignment on the spot, but this is not a preferred procedure. If the student needs work pages scored and he is in Grades 1–3, he takes the pages to a clerk to be scored. If an instructional problem is indicated in the scoring, the student is referred to a teacher. If the student is in Grades 4–6, he will normally score his own work from keys which are kept in loose-leaf notebooks with each page in a plastic protector. The older student must then exercise judgement as to when the teacher's attention is needed. The size of the group during individualized instruction is quite flexible—it has varied from 1 to 80.

The teachers have planning time for arithmetic at least once a week. At this time, the teachers, together with a specialist connected with the center, discuss the progress of the class as a whole and, in turn, the progress of each child.

The children work on their individual work for four of the five school days. The fifth day is called "math seminar day." The entire class meets as a group. The purpose of math seminar day is to (1) discuss topics of general interest to the entire class, (2) to promote a conversation between children of different abilities and at different levels of work, and (3) to cover broad areas of the curriculum in a discussion lecture. In other words, the math seminar day should give the student perspective on where he has been and where he is going, as well as a sense of the relation between arithmetic and his world of outside school interests.

To evaluate the program properly, the pupil-teacher ratio and the number of clerk assistants should be considered. For 7 classrooms with about 220 children there are 10 teachers. One of the teachers, not having a homeroom, is primarily a science teacher, another is primarily a librarian (who has duties in the individualized reading program), while the third is a "travelling teacher" who takes on a variety of classroom duties so as to enable the seven homeroom teachers to attend planning sessions. In addition, the usual quota of special teachers comes into the school to conduct classes in art, music, and physical education.

In order to handle the record-keeping and to score the tests, six local housewives assist as clerks under the direction of a Learning Center staff member. Some of the work done by the clerks is required by the Learning Center solely for experimental purposes. Perhaps three or four non-teacher clerks might otherwise be sufficient to carry the extra load involved in an individualized program in a school the size of Oakleaf. This, of course, does not take into account the many extra services provided by the center staff or the preparation and revision of instructional materials and tests.

## RESULTS

### *Achievement*

One of the commonest questions asked is "What do you do about the student who just can't learn something, for example, how to multiply fractions?" In this case, either the teacher provides tutorial assistance or, if enough pupils have difficulty, the instructional materials are revised. The argument is that the student must first master all of the prerequisite units and that the current work must then build on this foundation. It is our problem to find the instructional approach which will be successful for any particular learning problem. As a result, we can point to a floor of achievement for each class (Cox, 1965). For example, all of the children in the sixth grade will have mastered addition with carrying and simple multiplication with carrying, as well as subtraction with borrowing and simple division. This floor is much higher for each grade this year than in the first year. During the first year there were sixth-grade students who, at the beginning of the year, had not mastered the addition of single digit numbers. At the end of the year these students had mastered about one and one-half years' work by normal standards, but the floor was still very low. This year each class is advanced almost a year over the corresponding class last year.

Another question is "How well do your students do as compared to other students?" Since there is no control group as such, the question is answered in two ways. The first way is to point out that, on entrance into the individualized program, the students did very poorly on the placement and pretests—they had not mastered our objectives. In noting this, of course, one must be aware that since our objectives may differ from those of other programs, it would not be unexpected to find gaps in student performance when students from another program are inducted into our program. This has been borne out this year. Whenever new students have come into the school they tend to begin work near the bottom of the class distribution, regardless of their previous grades.

The companion question is "How well do Oakleaf students do on standard achievement tests?" The answer is interesting. At the end of the first year of the program, the first and second grade looked outstanding with almost every student ranking above the 80th percentile. The third and fourth grades looked average while the fifth and sixth grades had large numbers of students ranking below the 40th percentile.

Many of the upper class students had to go below grade level to make up deficiencies in their mastery level. For this reason, they were not seeing the material normally presented to students in the upper grades. Thus, on the standardized tests they did poorly while at the same time they were shoring up their understanding of earlier work.

The next question is "How well do last year's sixth graders do in the seventh grade of the junior high school?" Does the mastery of earlier objectives compensate for not encountering certain topics? The seventh-grade mathematics teacher reports that the Oakleaf students seem no different from his other students from other schools in the district.

While almost every student has been faced with a large remedial load, the mastery of earlier levels of the individualized curriculum seems to allow the student to perform satisfactorily when he goes back into a self-contained classroom. Of course, it remains to be seen how succeeding classes of students from Oakleaf will perform on various measures of scholastic ability. The final test will be to try to evaluate how well the students turn out as adults in a complex and demanding world. Certainly, ten years or more is a long time to wait for the results of an experiment. But, again, perhaps we had better begin.

*Number of units mastered*

The average number of units mastered in the first year was about 12 units per student. You will recall that there are a total of 85 units of varying length in the program, which encompasses Grades K–7 (we wrote objectives for the seventh grade in case some of the better students needed the additional work).

Some students (not counting the first grade) completed as few as 5 units, while others completed more than 20 units.

*Range of achievement*

Placement tests administered in September 1964 have been compared to those administered in September 1965. They show that the spread of achievement increased for the second and third grades and decreased for the fourth, fifth, and sixth grades. This decrease for Grades 4–6 may be due to the relatively rapid growth of the slower students, who had been hopelessly lost in the regular syllabus, along with the heavy remedial load faced by even the better students (Bolvin, 1966).

*Summer retention*

In view of the encouraging progress made by the children during the school year 1964/65, the question of retention over the summer became important. As is typically the case, there was some loss of skill. Some children did not perform satisfactorily on tests of objectives previously mastered. On the other hand, these losses tended to disappear by the end of the placement-testing period, i.e., the placement tests themselves served as a warm-up for the students, and the students usually passed on the pretest the objectives that they had mastered during the previous year. Furthermore, retesting after a three-month period during the school year showed generally higher scores on the retests than on the original post-tests given prior to the summer vacation. Our conclusion is that summer retention is very high in this program, and we attribute this to the mastery criterion for progression.

*Rate and IQ*

A very interesting result is the almost complete lack of correlation of rate of progression in the program with IQ (Yeager and Lindvall, 1966). Only on units for which we had independent evidence of instructional difficulty was there a correlation of time to complete a unit with IQ. Perhaps the simplest interpretation is that IQ is related primarily to the ability to leap over deficiencies in the instructional process.

*Transfer*

Since we have a program in which students are pretested before they receive formal instruction in a unit, we can look at instances where the students have shown mastery of objectives before they have been formally introduced (i.e., taught). The three major conclusions are these:

1. The probability of transfer of old abilities to new objectives is greater as the student acquires more knowledge of arithmetic.

2. There is a slight correlation of ability to show this type of transfer with IQ.

3. The objectives which are most difficult to master without formal instruction are those which involve the algorithms of multiplication with multi-digit numbers, and long division. It seems that as the progression of objectives becomes more logical and less dependent on memorized procedures the probability that the student will infer the rules of the algorithms is greatly increased. We made an attempt to capitalize on this observation in revising our materials during the summer of 1965, and preliminary results are encouraging.

### *Motivation*

There are two ways to report on motivation of the students. The first is that in the individualized program the students leave the classroom for many reasons and they do not need to ask permission to leave. If the work were punishing rather than interesting to them, they could avoid the work very easily. Independent observers have noted some students who not only do not try to escape, but who hurry back to the classroom. Second, students who are behavior problems in regular classrooms often are not as disruptive in an individualized setting where each student works on his own assignment. Anecdotally, it can be reported both by the staff and visitors that for most of the children motivation does not seem to be a problem, especially in the lower grades.

However, some students who seem to be progressing slowly and some students who perform well in the self-contained classroom may be missing something in a program in which they spend much of the time working alone. The trouble is that it is very difficult to tell the difference between a student who is working slowly for legitimate reasons and a student who is experiencing difficulty because of a mismatch between the student and the program. Certainly, in an individualized program we must pay attention to this problem.

A major problem is that certain units of instruction, especially at the earlier levels, require concrete materials for effective instruction. We are attempting at this time to prepare lessons which involve such materials. A special problem occurs in providing directions for the use of such materials to students who cannot read. In order to deal with this problem we are attempting to use two *audiotape devices.* One is a tape cartridge repeater similar to those being sold for automobiles, and the other is a device in which the tape is attached to a heavy card which can present graphic stimuli along with a sound message of about 20 seconds.

### *Implications*

In conclusion, let me state what I feel are two important implications. First, the greater the variability of student achievement in the classroom or school, the greater the potential of an individualized system. Thus, the general approach may be most useful in school districts which are undergoing integration or which, for other reasons, have large spreads in student ability. Second, a system of continuous revision of curricular materials, based on student performance, is a highly desirable way to avoid obsolescence of instructional materials and to arrive at effective working materials.

## REFERENCES

BOLVIN, JOHN O. *Variability of Pupil Achievement in Mathematics*. Symposium on Individually Prescribed Instruction at the American Educational Research Association Meeting, Chicago, February 1966.

COX, RICHARD. *Report on Progress of Students at the Oakleaf School*. Learning Research and Development Center Report, University of Pittsburgh, 1965.

GLASER, ROBERT. *The Program for Individually Prescribed Instruction*. Symposium at the American Educational Research Association Meeting, Chicago, February 1966.

LIPSON, JOSEPH. *Transfer and Generalization in Individually Prescribed Instruction*. Symposium at the American Educational Research Association Meeting, Chicago, February 1966.

YEAGER, JOHN L., and LINDVALL, C. M. *An Exploratory Investigation of Selected Measures of Rate of Learning*. Symposium on Individually Prescribed Instruction at the American Educational Research Association Meeting, Chicago, February 1966.

# VIII

# Engineering Instructional Sequences for the Mathematics Classroom*

BERT Y. KERSH
*Teaching Research Division*
*Oregon State System of Higher Education*
*Monmouth, Oregon*

THIS report reflects my concern for the classroom teacher and the difficult instructional problems he must resolve. The teacher needs procedures by which he can systematically arrange learning experiences so that the learner will attain prescribed instructional objectives efficiently, economically, and within practical limitations. However, this report is written from the point of view of an educational researcher concerned with the development of a science of instructional engineering. Engineers are people trained to design and to develop structures such as bridges, and man-machine systems such as computers. They draw heavily upon principles of physical science and mathematics, but they also have developed a body of knowledge through research which may be properly called an engineering science. As educators we too are concerned with designing and building man-machine systems, and we too are having to rely increasingly on our own research efforts because the information we need is not to be found in textbooks of social science.

It is frequently said that the classroom teacher will never be replaced by programs of self-instruction. Rather, he will be freed to guide the learning of his students in ways that only a human being can. Implicit in this statement is the assumption that some learning processes cannot be "automated" or learned independently. Learning processes which many say cannot be automated include such complex intellectual proc-

* Portions of this report were selected from a previous publication in R. Glaser, *Teaching Machines and Programmed Learning II, Data and Directions* (Washington, D.C.: National Education Association, 1965).

esses as reasoning, problem solving, and "learning how to learn." The behavioral components of such complex processes are elusive, so it is reasonable to believe they are best learned by interacting with another person who has mastered them or by wrestling with difficult problems under supervision.

The thesis of this paper is not that such complex learnings are adaptable to self-instructional programming techniques, but rather that the principles and techniques which underlie self-instructional programming can be employed equally as well in the development of suitable classroom instructional materials and procedures. The result may be very similar in appearance to classroom procedures which are presently employed by teachers, but the resemblance may end there. There is no greater similarity between conventional classroom techniques and programmed classroom techniques than there is between conventional self-study materials and programmed self-instructional materials.

## Instructional Design Requirements

Hereafter, instructional objectives will be classified in two categories: (1) as being amenable to "automatic" or self-instruction and (2) as being most readily attained through "human" instruction.

Instructional objectives which are most readily attained through human instruction may be distinguished from those which are amenable to automatic or self-instruction by identifying their instructional requirements. For example, assistance from another person may be required in the attainment of an instructional objective for any one or more of the following reasons:

1. The required behavior cannot be identified by machine processes presently available, or by the learner himself without previous instruction.

2. The required behavior cannot be reliably elicited except through direct or indirect intercommunication with another person who is capable of identifying the required behavior once it has been elicited.

3. The learner cannot determine that he is making progress toward the instructional objective by independently comparing his own behavior against a behavioral standard or model.

Usually instructional objectives which involve the attainment of factual knowledge, concepts, principles, or even some psychomotor skills will be amenable to automatic or self-instruction. Objectives which are most readily attained through human instruction will usually involve patterns of behavior occurring at unpredictable intervals and reflecting "media-tional" processes. This second class of objectives probably includes what Duncker (1945) calls formulating or restructuring problems during the problem-solving process, and what might be identified as hypothesis formation or "retroductive reasoning" (see Hanson, 1958, p. 85). Of course, involved in such complex behaviors as reformulating problems and forming hypotheses are many other behaviors (or behavioral tendencies) which have been variously described as "shifting," "searching for patterns," and "being flexible."

However, public school teachers today do not often limit themselves to teaching one thing at a time. If they wish to teach some computational skill in arithmetic, for instance, they also concern themselves with such by-products of learning as the attitudes of their students toward arithmetic; if they wish to teach theory of combustion, they are also concerned with "understanding scientific method" and "skill in problem solving." Even if teachers were satisfied to deal with a single objective at a time, psychologists would remind them that they must not only consider the objective from the standpoint of immediate learning, but that they should give consideration also to the maintenance and subsequent use (transfer) of the new learning. It is one thing to predict that the learner will be able to say something or do something that he is presently unable to do after completing the instruction, and something else to say that the learner will want to continue using it and will use it to good advantage in a great variety of appropriate situations. A single unit of instruction may include some objectives which can be taught through automatic or self-instructional techniques, and other objectives which may call for human instruction. When this is the case the instruction will be said to involve multiple or compounded objectives.

While existing procedures may be adequate for programming objectives one at a time, in the experience of the present writer they have not been adaptable to programming multiple or compounded objectives. In dealing with compounded objectives, the programmer must concern himself with two or more processes which will be operating at once, in about the same fashion that a composer of music must in developing a symphonic score.

A new procedure for planning classroom instruction is needed which will incorporate the techniques employed in developing self-instructional programs in the design and development of procedures for attaining compounded instructional objectives. It should be possible for an expert in human learning and a subject matter specialist to prepare, in advance, an outline of the learning process, just as an engineer does in designing, on paper, the structures and systems he builds. Then, from these "instructional designs," it should be possible for programmers and materials development specialists in our schools and colleges to actually "build" the instructional systems, try them out, and, if necessary, send them "back to the drawing board" to be modified.

The essential characteristics of such a procedure would appear to be the following:

1. It should provide a notation and charting technique with which the instructor can prepare in advance a detailed outline of the learning experience in terms of practice and reinforcement schedules, branching criteria, and related characteristics, without attending to the specifics of frame writing.

2. It should outline a precedure for preparing a basic instructional program aimed at objectives which are amenable to automated instruction, and then for "weaving in" programs involving human instruction, or vice versa. In this way, different processes of learning could be employed simultaneously in a single program, or a single program could be systematically altered for purposes of research and development.

3. The methodology should enable the instructor to deal with problems of program design separately from frame writing and materials development so that the latter can be accomplished by different individuals concurrently.

### An Example: The TRAC Procedure

One example of a methodology which meets the requirements specified above is referred to as the TRAC procedure simply because it was developed in connection with the Teaching Research Automated Classroom, called "TRAC," located on the campus of Oregon College of Education. The instructional procedure was designed for use in the TRAC facility but it might also be adapted for use in other semi-automatic instructional facilities.

By way of illustrating this procedure, consider a particular instructional sequence which was designed for research on discovery learning (Kersh, 1964). We wanted to test an hypothesis concerning learning by a process of discovery, defined as a specific instructional method. The design of the instructional sequence was complicated because we had more than one objective. Our "subject matter" objective was to teach the distributive law of arithmetic to capable fifth graders. In addition, our objectives were to teach the fifth graders to "discover" principles from concrete examples and to stimulate their interest in what they were learning.

As might be expected, a number of different criterion measures were employed. One consisted of a set of six "open sentences" (using the Illinois Program terminology), some of which correctly represented the distributive pattern and others which did not. The learner was asked to mark each example which would always produce a true statement when the "frames" were replaced with numerals. This test was used as a standard of learning for purposes of instruction. Instruction was continued, in other words, until each learner could complete the test with no more than one error.

A second standard was developed to determine whether or not each learner could employ certain prescribed behaviors which we called "searching behaviors." The instructors were trained to use an observation schedule to identify specific student behaviors identified as "searching for patterns," "checking for exceptions to a possible pattern," "checking to see if statements are true or false," and "employing frames." The procedure was to test each learner individually within 24 hours after he had attained the first instructional objective. The test consisted of three questions, each of which was designed to elicit a specific class of searching behaviors. In the first question, for example, the learner was asked to examine a set of four examples of a general law of arithmetic and to determine whether or not they were all examples of the same or of different laws. If his answer was yes, he was asked to write the general law using the notation of the Illinois Program. As each learner attempted to answer the questions, he was instructed to "think out loud" or to indicate what he was thinking by his scratch work.

As a test of "interest" an attempt was made to ascertain whether or not each learner spontaneously practiced with his new knowledge outside of class without instructions to do so. It was reasoned that the learner who actually put into practice what he was learning without being told to do so was manifesting interest in the task. We did not have a very precise measure of such "interest behavior," and had to rely on information obtained from each learner through interviews.

Finally, as a test of recall, a paper-pencil test consisting of problems similar to the ones used during instruction was administered to each subject within eight weeks following instruction.

Our task was to develop an instructional sequence for the classroom that would accomplish all of these objectives. The product of our labors was to be evaluated in part by simply teaching several groups of fifth graders and determining that their performance was acceptable by the standards we have previously established. We began our efforts at the drawing board, just as you would expect an engineering scientist to do.

In a very real sense of the word, we "designed" the entire instructional sequence, calling on our knowledge of the psychology of learning, on findings from previous research efforts, and on techniques of instruction which had been developed by others. The procedures we employed are described very briefly below.

*Preparation of a hierarchy of subordinate facts and processes*

According to Gagné (1962), tasks to be learned in the acquisition of knowledge may be identified by working backwards from the final task. The question is asked, "What would an individual have to know in order to perform this task successfully?" The answer to this question reveals subordinate knowledge which the individual must know in order to obtain the ultimate objective. The subordinate knowledge is presumably simpler and probably more general. This subordinate knowledge is again subjected to the question, "What does one have to know in order to achieve this?" And still more subordinate knowledge is revealed in the answer.

By continuing this questioning procedure and working backwards from the ultimate objective, a hierarchy of subordinate knowledge is established. In the end, the final content objective is seen to rest on a framework of subordinate knowledge which becomes increasingly simpler and more general.

The TRAC procedure differs somewhat in that *both* knowledge (the subject matter objective) and the complex behavioral objectives (e.g., how to "search for patterns" and to "check for exceptions") are treated as "ultimate" objectives. The hierarchy of knowledge to be acquired is identified by asking what the learner must *know* (after Gagné), and the hierarchy of complex behaviors is identified by asking what the learner must *do* in order to acquire both the knowledge and the complex behavior. [Although Gagné does use the word "know" rather than "do," he uses this term to indicate what a learner must be able to *do* in order to be able to *do*. The concern here is with *how* the learner acquires higher order forms of behavior.]

In the present example, the hierarchy of *knowledge* that was actually developed contained seventeen separate subordinate facts, called subfacts, to be learned. These subfacts were arranged in a logical sequence and diagrammed so that a programmer could readily determine the sequence of learning experiences for the lesson.

The complex behavioral objectives, on the other hand, were considered separately. Also, they usually were diagrammed separately. Figure 1, for example, illustrates the "hierarchy" (actually a set of programming speci-

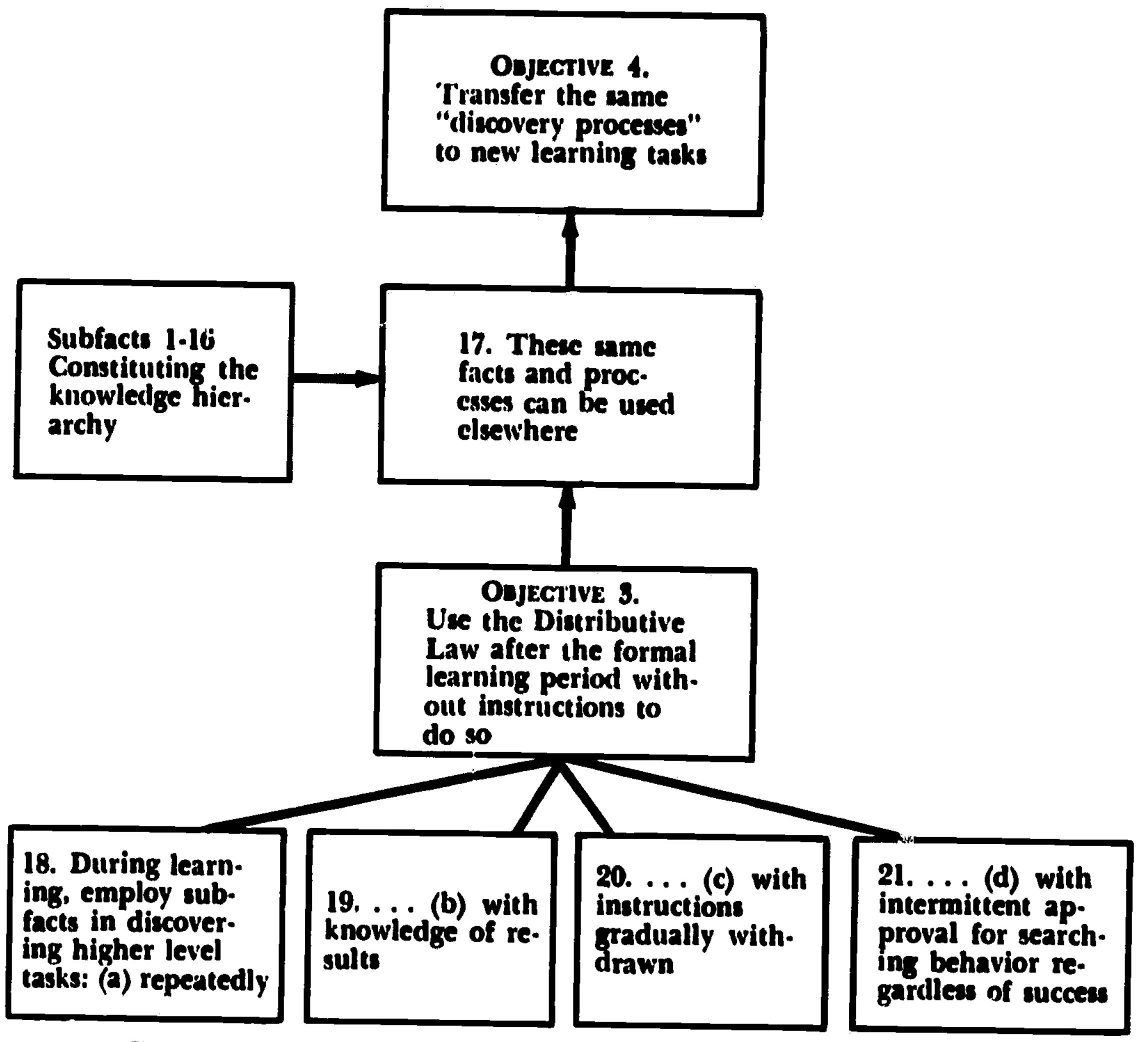


FIGURE 1.—EXAMPLE HIERARCHY OF COMPLEX BEHAVIORAL OBJECTIVES

fications) which specified what the learner should "do" in order to learn to use the distributive principle and to transfer those techniques called "searching behaviors" in learning other mathematical principles. In Figure 1, the box labeled "Objective 3" reveals that one requirement of the instructional unit is that the learner generate enough interest in the distributive law to use it after the formal learning period, without instructions to do so. What must the learner do to develop this interest? The answer is written in the four smaller boxes labeled 18, 19, 20, and 21. These subordinate process statements specify that the learner should employ subordinate knowledge in discovering higher-level tasks repeatedly; with the knowledge of results; on a schedule in which the teacher's instructions are gradually withdrawn; and with approval provided intermittently, regardless of the learner's success or failure.

Now locate Objective 4, the "transfer of discovery process" objective in the same figure. It is combined with the "knowledge" objectives (not shown). Clearly it is a higher level of learning than the knowledge objectives, and probably should be classed as a "learning set," but it rests on a framework of specific knowledge as indicated. The "discovery processes" employed by the learners are not precisely stated in Figure 1, primarily because the behaviors involved cannot be adequately described in general terms. Instead, the instructors learned to identify examples of the complex behavior in the context of standardized instructional and test situations. For example, "frames" (e.g., □, △, ○) were used in the notation instead of more conventional algebraic symbols (e.g., $x$, $y$, $z$) in writing abstract mathematical expressions. All learners were taught how to use frames. However, "using frames" also referred to a specific and somewhat complex behavior which was classified as a "discovery process." When learning by discovery, a student might have been given a set of mathematical statements such as the following:

$$3 + 3 = 2 \times 3.$$
$$5 + 5 = 2 \times 5.$$
$$8 + 8 = 2 \times 8.$$

Then the student might be asked to determine whether or not each is an example of the same general law. If while trying to determine the correct answer the learner was observed to use frames in an effort to reduce the three examples to a single abstract expression (e.g., $\triangle + \triangle = 2 \times \triangle$) he was said to be "using frames" as a discovery process.

*Preparation of flow charts for each subordinate fact and process*

Next, the instructional program was designed, using flow charts. The flow charts were prepared for each of the subordinate facts and processes.

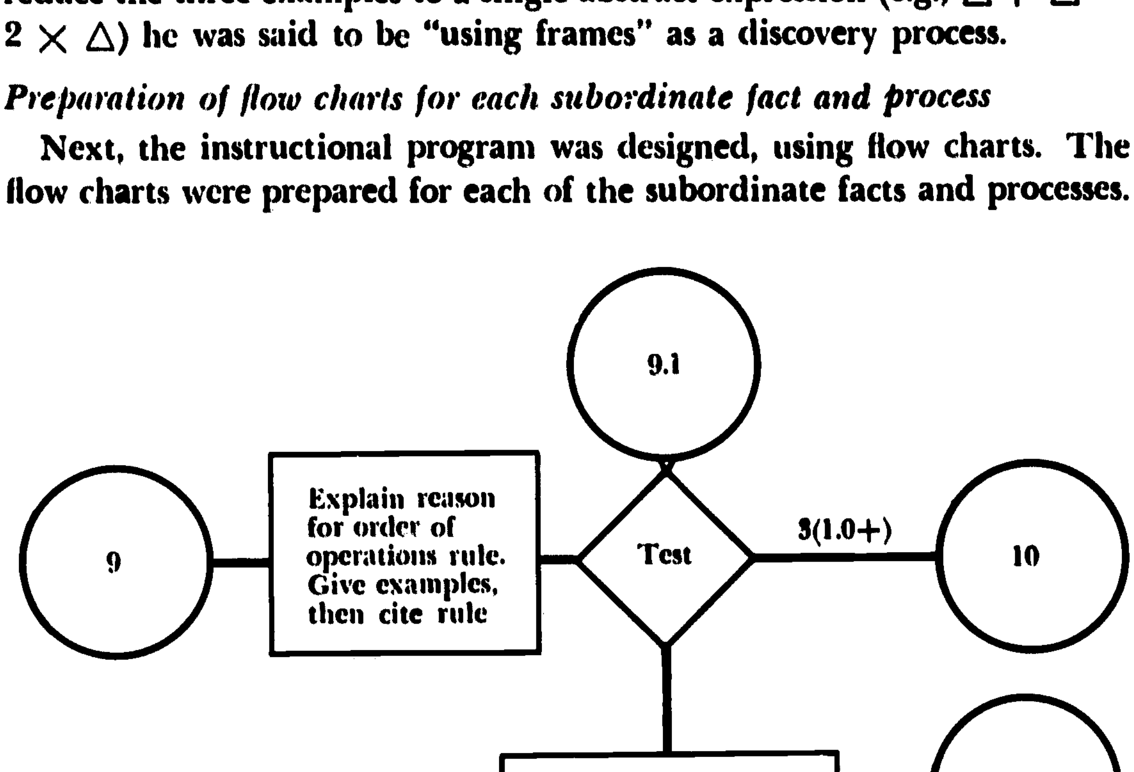


FIGURE 2.—PLAN A FOR SUBFACT 9, "CONVENTION: MULTIPLY FIRST, THEN ADD"

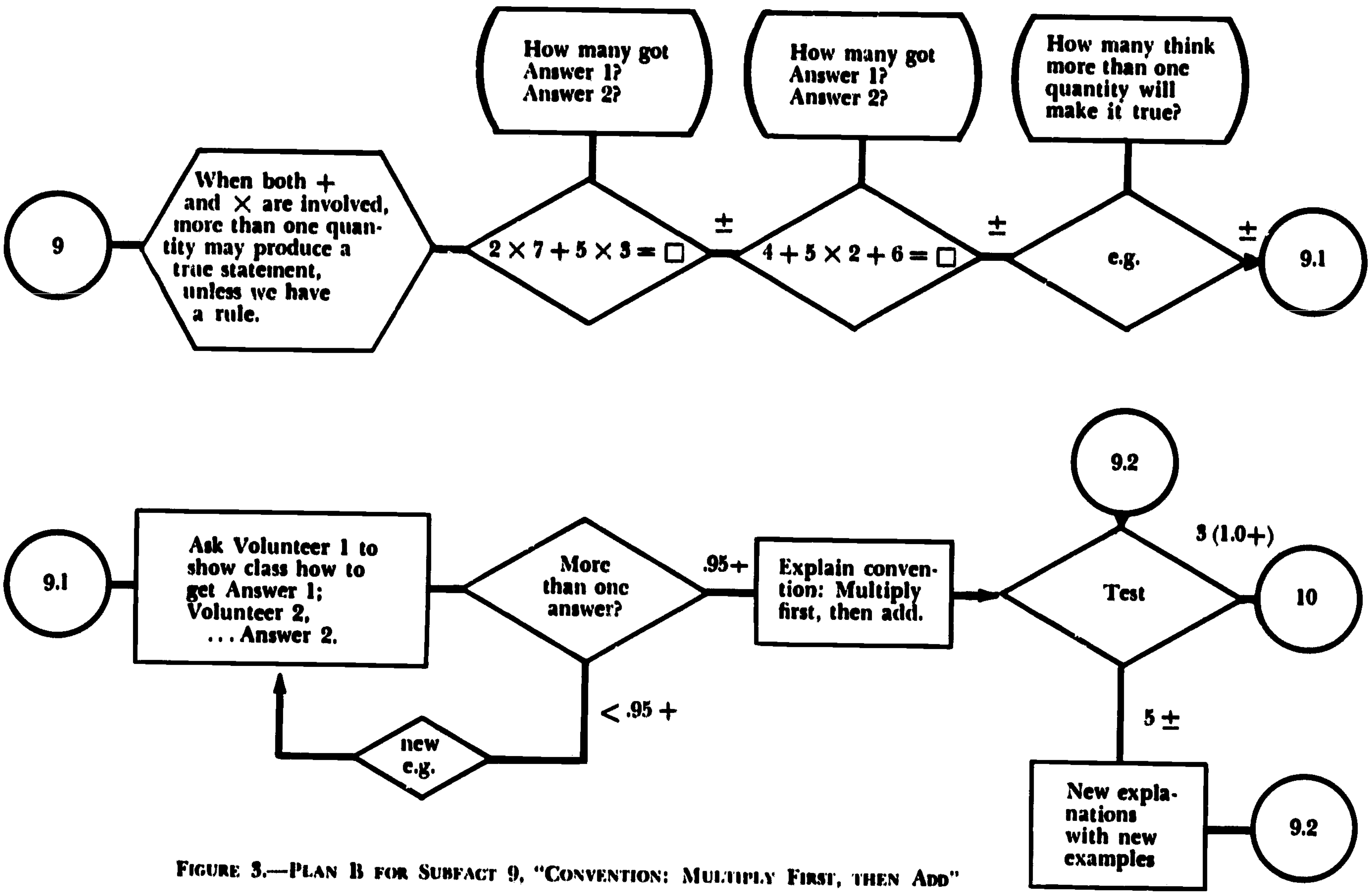


FIGURE 3.—PLAN B FOR SUBFACT 9, "CONVENTION: MULTIPLY FIRST, THEN ADD"

A special notation was developed to indicate specific teaching operations so that the detailed instructions and materials could be completed by another person independently, without consultation with the person preparing the flow charts. For example, instructions to the learner (as if from a teacher) were abbreviated and written in square boxes. Whenever there were problems or examples to be worked by the learners, they were indicated in diamond-shaped boxes. Additional notation such as "3(1.0+)" was used to indicate that the problem-solving exercise was to be continued until every member of the class achieved three problems in succession correctly. A less stringent criterion would be indicated by the notation, "3(.75+)."

Using the special notation, it was possible to outline for the programmer the essential characteristics of the instructional program in sufficient detail for him to carry on independently. The person doing the flow charting operated with the knowledge that he could alter the flow chart quite simply according to the subordinate process requirements after he had prepared the outlines for each of the subordinate facts.

As example, Figures 2 and 3 illustrate two different plans for teaching a subordinate objective which can be learned either by discovery or by some other method. Subfact 9 is the conventional order of operations (multiply first, then add). Typically, students were not required to "discover" a convention; however, it was decided in this case that the learners should have the experience of "discovering" the need for such a convention before being told the convention.

Plan A for Subfact 9 (which is not the discovery plan) is outlined in Figure 2. The first box in the flow chart indicates that the teacher should first explain to the student the reason for the order of operations rule, then give examples, and finally cite the rule. Next, the diamond indicates that a test should be given which continues until all learners answer three problems in a row correctly. Having reached the criterion, the flow chart indicates that the program should continue to the next step in learning, designated as Subfact 10. If, during the test, the criterion is not achieved after five problems, the program branches to a new explanation of the rule, using new examples, followed by a retest.

The outline for Plan B in Figure 3 (the discovery plan) appears very much more complicated. Starting with the hexagon after the circle numbered 9, the flow chart indicates that the program should be written in such a way that the students discover the need for a rule. The diamonds following the hexagon indicate in more detail how this is to be done. As is indicated, the learners are asked to complete the open sentence involving both multiplication and addition, with the expectation that two correct answers are possible without a rule regarding the order of operations. The fact that either one of the two answers is considered correct is communicated to the learners until the learners become aware that "something is wrong." At about this point (Circle 9.1 in the flow chart), the flow chart indicates that the teacher should ask for a volunteer or two to explain to the class how they obtained their particular "correct" answer. Then the class is asked if they believe that there is more than one possible answer. If more than 95 percent answer correctly (.95 +), the teacher explains that mathematicians have agreed to multiply first and then to add. If less than 95 percent of the class answer correctly, the procedure of giving examples and asking for volunteers to explain their procedure is continued until such time as the criterion .95 + is achieved. Finally, the flow chart ends with a test of the ability of the students to use the rule. The test has a criterion of "3(1.)" after which the program continues to the next step, Subfact 10.

Presumably, the flow chart writer could modify the learning experience in yet other ways. The appealing feature of this methodology is that it indicates rather precisely what changes are to be made. This is a happy feature from the experimental standpoint. It also indicates how changes can be made with relative ease after the fashion of an "executive routine" in a computer program.

This flow-charting procedure was continued until, finally, at the drawing board level, we had a very detailed notion of the instructional procedure and also regarding what might be called the "scope and sequence" of the instructional unit.

### *Development of specific instructions and materials*

The next step was to develop the actual instructional materials required in the teaching of the so-called subordinate objectives. We treated these separately in the beginning, without worrying too much about the ultimate objectives. This is an important point, because long-term instructional sequences are too complex to be treated as a whole. As instructional materials were completed for each subordinate objective, they were tried out with small groups of learners and were revised according to the results. Essentially we were evaluating each segment of the instructional sequence separately. Often, we found out that it was necessary to take the preplanned sequence back to the drawing board and to revise it in accordance with our tryout data.

In the end, we had a set of instructional materials together with detailed instructions for employing them which was developed on the basis of preliminary trials with several groups of fifth graders, the last of which had attained the ultimate objectives to our satisfaction.

## Final Evaluation

Normally, we have taken steps to ensure that the procedure could be used successfully in the classroom without special laboratory media. However, this particular sequence was designed for research purposes. So our next concern was to build a parallel instructional sequence, identical in every respect to the first, except that the particular instructional variable with which we were concerned was eliminated. When the second course sequence was completed, we had the ingredients for an experiment designed to test our theoretical "hunch" concerning the discovery method of teaching. By comparing one specially designed instructional sequence against the other, under controlled laboratory conditions, we were able to provide evidence to support (actually, in this case, to refute) our hypothesis. There is no way that I know of to do this except by making such comparisons. This is the method by which we solve puzzles of science. However, the hypothesis-testing paradigm is not necessary, or even very often appropriate, for evaluating the effectiveness of courses or course sequences which constitute the curriculum.

The particular version of the instructional sequence which we had designed and put together took from 20 to 23 class sessions lasting approximately 30 to 50 minutes each, which is admittedly not particularly long. However, the procedures we had followed in instructional design, construction, and evaluation are just as applicable to sequences lasting over periods of months and years. There is the added implication for long-range instructional sequences that even the most carefully planned sequences may fall short of expectations in the final analysis. It may be quite possible to evaluate small segments of a curriculum, for example, from the standpoint of student achievement, but even with instructional units as short as the one just described, it is very difficult to ascertain precisely where the program fell short when considered in its entirety. It is with such long-range instructional sequences that it may become increasingly important that we employ what may be called "second order criteria" involving logical and psychological considerations of the instructional design itself. To determine the effectiveness of an instructional sequence in meeting these second order criteria it is necessary either for experimental subjects to complete the entire instructional sequence or to make use of existing records of student groups who may have completed the entire instructional sequence in the past.

## REFERENCES

DUNCKER, K. "On Problem Solving." *Psychological Monographs,* LVIII, Whole No. 270 (1945), 1-113.

GAGNÉ, ROBERT M. "The Acquisition of Knowledge." This publication, reprinted from *Psychological Review,* LXIX (1962), 355-65.

HANSON, N. R. *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science.* London: Cambridge University Press, 1958.

KERSH, BERT Y. *Directed Discovery vs. Programed Instruction: A Test of a Theoretical Position Involving Educational Technology* (U.S.O.E. Res. Proj. No. 907, Title VII, NDEA of 1958). Monmouth, Oreg.: Teaching Research Division, Oregon State System of Higher Education, 1964.

# IX

# Teaching, Discovery, and the Problems of Transfer of Training in Mathematics

JERRY P. BECKER*
GORDON K. McLEOD
*Stanford University*
*Stanford, California*

LEARNING transfer (or transfer of training) is an important topic in many branches of psychology. The topic receives extensive coverage in texts of general experimental psychology and in treatments of learning. As one may expect, it is especially important in the psychology of human learning and in educational psychology.

In its broader sense, something like transfer of learning is basic to the whole notion of schooling. Those who support schools, like those who conduct them, must assume that the thing being taught at this particular moment will have some value at a later moment and in a somewhat different situation. For example, we assume that today's lesson in geometry will surely help in tomorrow's lesson in the same subject, that it may be of use in later study of analytic geometry, and, more ambitiously, that it may induce an appreciation of logic so profoundly that it affects the student's entire way of life. Clearly, without some degree of reliance on transfer, teaching would be hopelessly specific. It would be necessary to train each student in every specific situation he might ever encounter.

We believe most teachers of mathematics make the assumption that the skills and understanding which they endeavor to impart to their students will influence the behavior of the students beyond the classroom

* Will be at Rutgers—The State University, New Brunswick, New Jersey, beginning September 1, 1967.

setting in which the learning takes place. We expect specific learning in mathematics to transfer to ensuing situations both inside and outside school. When one takes account of the evidence, however, our assumption is not necessarily borne out in practice. This is, indeed, discouraging to teachers of mathematics. But what is more discouraging is the fact that students seem to have difficulty in effecting learning transfer from one situation to another even within the mathematics curriculum itself.

It seems reasonable to inquire into the degree of validity of the conjecture that there is a broad transfer power in the study of mathematics. For example, it is commonly stated that a significant outcome of the study of mathematics is the ability to think more logically. What we propose to ask as educators in mathematics is whether psychological theory can give us a basis for a hopeful view of the problem of learning transfer. This is, in fact, the objective of this paper. With psychological theory as our guide, we propose to consider the problem of structuring the learning situation in mathematics so that maximum transfer of learning can occur.

## Definitions and Model of Transfer

It seems appropriate to inquire about a definition of transfer at this point. It turns out that few people have actually defined the term. Consequently, we have concluded that transfer of learning can be thought of as a broad, inclusive phenomenon. Let us consider a few examples.

"Learning how to learn" to solve a class of problems is considered to involve an important type of transfer. Mathematics teachers consider the application of logical processes of analysis learned in geometry to nonmathematical situations to be a very desirable example of transfer. Experimenters in psychology consider as evidence of transfer the application of a principle in a test situation, where the test situation may differ only slightly from the training session in which the principle was learned. We submit that every learning situation involves transfer to some extent, since a learner brings his past learning experiences and attitudes to any new learning situation.

We think it would be useful to examine a model suggested by Ferguson (1956) in order to bring into focus the consideration of the problem of transfer. His transfer model, in its simplest form, is a mathematical function of three variables. If $y$ is the dependent variable representing a measure of performance on some particular task, then $y = f(x, t_x, t_y)$, where $x$ is a measure of performance on another task, while $t_y$ and $t_x$ represent the amount of practice on each of the two tasks. Here $x$ is also a function of $t_x$; that is, $x = \emptyset(t_x)$, so that $y = f(\emptyset(t_x), t_x, t_y)$. Ferguson (1956) used this model to describe a formulation of the concept of transfer and we propose to consider it in more detail.

When two tasks are the same, so that the measures of performance are identical, the expression for $y$ reduces to a function of one variable, since $x \equiv y$ implies that $t_x \equiv t_y$. Therefore we find $y = g(t_y)$. Clearly, this expression relates a measure of performance on a task to a measure of the amount of practice on the task and the result is a representation of the traditional learning curve. Thus, Ferguson's model suggests that learning is a special case of the more general phenomenon of transfer.

Looking at it another way, if no practice is allowed on the task represented by $y$, then $y$ reduces to a function of two variables so that $y = h(x, t_x)$. This case represents a transfer experiment where measurement is made of the effect of practicing one task upon the performance of another nonpracticed task.

Consideration of this model enables one to obtain a broad, general view of the problem of transfer. Further, it suggests the following definition: "Transfer of learning occurs whenever the existence of a previously established habit has an influence on the acquisition, performance, or relearning of a second habit" (McGeoch and Irion, 1952, p. 299).

There are many phenomena which are consequences of learning; among them are skills and understandings. In light of Ferguson's model, we will focus attention on these in this paper. Therefore, the term "habits" as used in the definition above will refer to skills and understandings in subsequent pages. It seems clear that an implication of the definition and the model is that *transfer can be positive or negative.*

## Theories of Transfer

Before proceeding to a consideration of transfer of learning in the educational setting, we think it is appropriate to examine briefly some general theories which deal with the mechanism of transfer. Man's first theory of transfer proclaimed that formal study in school subjects was the best way to secure the ability to apply sound judgment and logical reasoning to problems outside of school. It held that the more difficult the formal study, the more exercise for the mind and the better its training for transfer. For example, this theory held that the development of logical thinking in geometry would transfer automatically to sound logical reasoning in social studies.

The investigations of Thorndike and Woodworth (1901) at the turn of the century proved this theory inaccurate. In a series of experiments, the influence of special training in estimating magnitudes (lengths, areas, etc.) on the ability to estimate magnitudes of a more general nature was tested. The conclusion was that performance on the more general tests was not significantly influenced by the special training.

Later Thorndike (Thorndike and Woodworth, 1901) formulated his doctrine of identical elements to explain the phenomenon of transfer. It stated that transfer occurs only when identical elements are involved in the influencing and influenced function. McGeoch and Irion (1952, p. 343) claimed that by two identical elements. Thorndike seemed to mean any clearly discriminable aspect of two activities which is the same in each. It was further suggested by McGeoch and Irion (1952) that Thorndike wrote as if he intended the theory to cover more than strict identity. In the light of Ferguson's model, Thorndike's view would claim that performance on any task is largely reduced to the case $y = f(t_y)$. In words, practice must be specific to the performance being sought. Other writers have concluded that Thorndike's view on transfer was an extremely pessimistic one.

Travers (1963, p. 193) states the opinion that Thorndike's theory is thought of today as an oversimplification of the phenomenon of transfer. The famous experiment of Judd suggested the theory of generalization which has come to supplement Thorndike's theory. Modern day Gestalt psychologists talk about essentially the same phenomenon in terms of meaningful organization of learning or the reorganization of experience.

It has been demonstrated that this kind of learning leads to transfer power. Bruner states ". . . massive general transfer can be achieved by appropriate learning, even to the degree that learning properly under optimum conditions leads one to 'learn how to learn' " (1962, p. 6). We propose to devote much of the remainder of this paper to dealing with the following two questions: What is *appropriate learning* for transfer? What might be considered *optimum conditions* for such learning? We will not confine our discussion to the area of mathematics, although what is discussed is certainly relevant to learning transfer in mathematics.

## The Role of Principles

Judd (1908) conducted an experiment on maximizing transfer of learning. This experiment consisted of throwing darts at a submerged target. Judd reached the conclusion that the best way to guarantee transfer is to teach principles. However, he believed that a principle must be exercised in practice while it is being learned, since he found his experimental group, which had been supplied with the principle of refraction, to be not significantly better than the control group in the first test; therefore, he contended that knowing the principle was not a substitute for direct experience. However, having organized their experiences using the principle as a frame, the subjects in the experimental group readily worked out necessary adjustments in succeeding tests with the target at different depths. Judd (1908) also found that experiences *alone* led to confusion on succeeding tests. The control group was not able to adjust readily to changes in depth.

It is not possible to critically evaluate the research design of Judd's experiment since many details are not available. We do know that the groups of boys were equated on the basis of the teacher's judgments of their brightness; however, such things as the number of subjects, the apparatus details, the procedure used in teaching the principle to the experimental group, and the quantitative results are not reported. For these reasons, it is significant to mention that Hendrickson and Schroeder (1941) conducted an experiment in which they modified Judd's experiment so that the skill being tested was shooting an air gun at a submerged target. Their conclusions confirmed the main result of Judd, although the differences between the three groups in the study were not large.

The transfer measured in Judd's experiment can be represented in terms of Ferguson's model. The performance of the control group in throwing darts at the target, submerged to a particular depth, may have been dependent only on the group's practice at that depth. If we let this performance be represented by $y_c$ and let the amount of practice at this depth be $t_y$, then $y_c = f_1(t_y)$. Thus, this situation reduces to the usual learning curve. However, the performance of the experimental group was dependent not only on practice at a particular depth, but also on knowledge of the principle of refraction and on practice in its application at a previous depth. Thus, for the experimental group, if we let $x$ represent a measure of knowledge of the principle and let $t_x$ represent the amount of practice in applying this knowledge, then $y_e = f_2(x, t_x, t_y)$.

There is another way of looking at the transfer involved in Judd's experiment, and that is to attempt to provide an explanation for the poor performance of the control group in terms of negative transfer. We could conjecture that training at the first depth interfered with performance at the second depth. If we let $w$ represent a measure of performance at the first depth, then, for the control group, $y_c = g_1(w, t_w, t_y)$. Now in order to represent the performance of the experimental group, it is necessary to extend the model so that it is a function of five variables instead of three. We could conjecture that knowledge of the principle and practice with it in some way mediated the performance of the experimental group at the first depth so that the transfer effect of that experience is positive. Thus, we get that $y_e = g_2(x, t_x, w, t_w, t_y)$, where, as before, $x$ represents a measure of knowledge of the principle.

## THE ROLE OF DISCOVERY

Let us again refer to the study done by Hendrickson and Schroeder (1941). A significant observation reported in that study was the apparent importance of *discovery* of the solution by individual subjects. Knowledge of the refraction principle seemed to hasten this discovery for the subjects in the experimental group. Therefore, we see that discovery enters the picture in transfer of learning.

Ervin (1960) used third- and fourth-grade pupils to investigate transfer effects of learning a verbal generalization. She led pupils to discover the principle of reflection by means of experiments in ejecting a marble from a tube against a barrier. One experimental group worked out the verbal principle from its observations while the other was given non-verbal aid in observing relevant facts. All instruction was individual. While there were no overall differences between the two experimental groups and a control group in performance on the transfer criteria, one test item was a key one. Here a flashlight was to be aimed upwards towards a mirror so that it would reflect on a target. The mirror was tipped sharply, and the target was low, near the flashlight. The usual error is to aim the flashlight too high, thus sending the beam up to the ceiling (Ervin, 1960, p. 547). On other test items, subjects could achieve success by aiming at a point somewhere between the vertical projections of the target and flashlight. But this doesn't work when the mirror is tipped steeply; only subjects who adjusted the incidence angle could be correct. Striking differences were found on this transfer item, with superior performance for those subjects who arrived at the correct verbal rule during training. Finally, it should be noted that both groups in the study had been guided toward discovery.

In another study of discovery, Gagné and Brown (1961) prepared programs to instruct ninth- and tenth-grade boys in deriving formulas for summing various number series [e.g., $1 + 3 + 5 + \dots + (2n - 1)$]. Then, instead of testing transfer by summing series of the same type, they tested ability to develop new formulas for summing new series (e.g., $1 + 3 + 9 + \dots + 3^{n-1}$). They constructed three programs: The first (*R* and *E*) gave the *rule* (formula) for finding the sum of $n$ terms of each training series and taught subjects to apply it to *examples;* a second (*GD*) divided the task into forty steps of *guided discovery,* each step requiring an analysis of a small part of the series; finally, a third (*D*) demanded *discovery* of the formula and provided hints as needed. All groups showed improvement from one training series to another. The transfer test required subjects to find rules for new series utilizing a few hints as needed. Guided discovery was found to be superior to each of the other groups. It should be mentioned that the tasks selected by Gagné and Brown appear to be well chosen. Not only are they representative of series problems, but, insofar as one task can be, they are representative of all mathematics (Cronbach, 1965b, p. 4).

Gagné (1959) and Cronbach (1965a) report that claims for discovery, as a method of learning, have had widespread influence on mathematics educators. At the same time, they state that the answer to the question "What kind of training will make a student capable of discovery?" has not been given. Consequently, Gagné and Cronbach and others have called for more research in this area.

Even so, mathematics educators should be aware of the attention that has been given to the effect that "discovery" of principles has upon transfer of learning. In a study of the effect of external direction during learning on the transfer of principles, Kittell (1957) used 132 sixth-grade students, divided into three experimental groups, who were trained by different methods to select one word that did not belong in a set of five given words. During the training process, the subjects in the "minimum" treatment group were told when correct responses were made, but they were required to discover principles independent of other help. The appropriate principle was briefly stated in general terms for the "intermediate" treatment group for each task, but they had to discover how to apply it in each case. The "maximum" treatment group was given not only the principles but also correct responses. The design of the research was of the following type:

$$\begin{matrix} O_1 & T_1 & O_1 & O_2 & O_3 \\ O_1 & T_2 & O_1 & O_2 & O_3 \\ O_1 & T_3 & O_1 & O_2 & O_3 \end{matrix}$$

where $T_i$ $(i = 1, 2, 3)$ represent the treatments and $O_i$ $(i = 1, 2, 3)$ represent the observations. (In this experiment, the observations which preceded and immediately followed the treatments were made with the same test instrument.)

The second observation measured the application of principles, learned during the training period, to new items. The third observation measured the ability to discover and use new unpracticed principles. Kittell (1957) concluded that superiority of the "intermediate" group, which received a certain amount of direction in discovering the principles, was established at a statistically significant level for both observations. At the same time, the "maximum" help group was also significantly superior to the group which derived principles independently.

The technique utilized by Kittell (1957) to train his "intermediate" group could be thought of as a type of learning in which principles are taught by examples. Katona (1940) in several interesting transfer experiments compared the effectiveness of learning by means of examples with learning by rote. He thought of the former as meaningful learning and the latter as senseless learning. His conclusions indicated superior results for the method of meaningful learning when transfer of learning was tested. Also, there was substantial transfer for the groups that learned by examples and practically none for the groups that memorized.

Although most educators would not find Katona's conclusions surprising, his experiments were weak in several respects. For example, he used a very small number of tasks and questionable statistical controls. According to Melton (1941), Katona's major results were unreliable. He observed that "understanding" and "transfer" were not independently defined words; hence, the hypothesis that learning by understanding leads to greater transfer was not actually tested. Melton further suggests that a more defensible explanation of the results might be to attribute the difference in performance to a *shift from a rote-learning attitude to a problem-solving attitude.*

Melton's conjecture is supported by the results of an experiment by Kersh (1958) in which the effects of independent discovery, as compared to directed discovery, of a generalization were tested. He concluded that "the superiority of the independent discovery procedure may be better explained in terms of motivation than in terms of understanding" (Kersh, p. 290). He goes on to say that the independent learner is more likely to become motivated to continue the learning process or to continue practising a task after the learning period. However, in a later study, Kersh (1964) found that neither of the discovery groups employed the learned material more frequently after instruction than did the third group in the experiment. This suggests that his previous findings may be unique to the particular instructional setting or to the learning materials used in the earlier study.

The same contrast in approaches to the learning of mathematics is emphasized in a book by Bruner (1960). He points out that an overly passive approach to learning creates a situation in which the learner expects order to come from the outside, that is, from the material which is presented. Mathematical reasoning, however, requires unmasking, simplification, reordering, etc. Therefore, the role of attitudes is recognized here as important in learning and hence to transfer of learning.

Hilgard, Irvine, and Whipple (1953) repeated and extended Katona's card trick experiment using sixty high school students in an attempt to counter Melton's (1941) criticisms of poor research design. The conclusions supported the hypothesis that transfer to new related tasks is greater after learning by understanding than after learning by rote. However, these authors felt that "the failures of the understanding group were more impressive than their successes, in view of the logical advantages inherent in the methods they were taught" (1953, p. 290). Consequently, a second study was undertaken in an attempt to reduce the number of errors (Hilgard, Irvine, and Whipple, 1954). Subjects in the understanding group were taught by five different methods, but the overall differences in success among the methods were slight. Hence the complex nature of transfer was brought into focus.

Wittrock (1963) used college students to study the effect of different schedules of help and statement of rules in learning on the following criteria: initial learning, retention, transfer to new examples. Wittrock's results indicate that explicit and detailed direction appear to be most effective and efficient when the criterion is initial learning. An "intermediate" amount of direction, however, appears to produce the best results when retention and transfer are the criteria.

Craig (1956) also used college students to test the effect of giving the rule and providing help on the criteria of initial learning, retention, and discovery of new principles. The group which was given the principle was superior in the number of rules learned initially and retained many more items after thirty-one days. A test for discovery of new principles, however, did not reveal reliable differences.

A study by Haslerud and Meyers (1958) also compared the transfer power of a principle which was derived by the subject with the transfer power of a principle presented by the experimenter in the form of a statement and an example. The researchers concluded that independently derived principles transferred more readily than given principles. However, other researchers have questioned the interpretation of the results and the conclusions drawn by Haslerud and Meyers (see Cronbach, 1965b, pp. 6-7; Wittrock, 1965, p. 41).

### The Role of Verbalization

As suggested earlier, another important consideration in the transfer of learning is the question: "What role does verbalization play in transfer?" In a study previously cited, Katona concluded that "the ability to solve the tasks can be acquired without verbal formulation of what has been learned and successfully performed" (1940, p. 101). Several people have pursued this observation in research.

In one of these experiments, Hendrix (1947) tested three hypotheses. They were (1) the nonverbalized awareness method of learning a generalization is superior to the method in which an authoritative statement of the generalization comes first; (2) verbalizing a generalization immediately after discovery does not increase transfer power; and (3) the possibility exists that transfer power may decrease as a result of verbalization. We found no trace of statistical controls in the study and the type of transfer tested was somewhat limited in scope. This is borne out by the fact that only one principle was considered for the three methods of training. Hendrix suggests, in conclusion, that the "flash" of *nonverbalized awareness* is the phenomenon that accounts for transfer power. This conjecture, we believe, should be tested under an improved design.

The University of Illinois Committee on School Mathematics (UICSM) also has something to say on the question of verbalization. This group believes that the student should become aware of a concept before a name is assigned to the concept. Many mathematics educators share this view.

### Transfer in Geometry

In all of the research studies we have examined, the tasks performed in the experiments were not unlike the analysis of relationships encountered in mathematical problem solving. Thus, we accept the conclusions as being relevant to learning in mathematics. Under careful scrutiny, however, it will be realized that the tasks to which the learning was transferred were only slightly different from the training tasks. Mathematics teachers have long felt that there might be a more general type of transfer to be gained from the study of mathematics, namely, an improvement in reasoning ability outside of mathematics.

Several studies we have examined have dealt with the hypothesis that training to think logically in geometry can transfer to nongeometric situations. Parker (1924), Perry (1925), Fawcett (1938), and Ulmer (1939) conducted such studies. The study of Ulmer virtually entailed the others, and hence we will consider it alone.

Ulmer's (1939) experiment was designed to evaluate the results achieved by a number of high school geometry teachers in different communities who utilized a method of teaching in which emphasis was placed on the cultivation of critical thinking. Ten teachers and 1,239 students in seven high schools were used. The subjects were divided into three groups: the experimental group with 638 students, the nongeometry control group with 575 students, and the geometry group (traditional courses) with 416 students. The nongeometry control group was composed of sophomores from schools having geometry as a junior course. Only the most capable teachers were used for both the experimental and traditional geometry courses. In the experimental group, definite emphasis was placed on concise, logical thought and application of critical thinking to nongeometric situations.

The evaluation instruments were reasoning tests prepared at The Ohio State University. The results indicated significant gains in critical thinking at all levels of intelligence for the experimental group at no loss in the learning of geometry content. The geometry control group showed a slight gain and the nongeometry group displayed no gain. We agree that the study illustrated very vividly that even highly competent geometry teaching offers little hope for the transfer of critical thinking *unless* definite provision is made for it in the teaching act. On the other hand, if such provision is made, the results can be rewarding indeed.

## Discussion of the Research

The preceding review of studies dealing with various teaching methods reveals the lack of consistent empirical evidence on the relative efficacy of these methods and points to the need for more carefully controlled research. The hypotheses which precede these studies frequently focus on the extent to which discovery activity should be guided.

We submit that this may not be the critical variable and that possibly these studies can be better understood if we separate *what* happened from *why* it happened. In the experiments in which the subjects who were given the principle performed best, these subjects comprised the *group that had the most practice in using the principle*. They were practicing the principle on trials when the others were trying, sometimes unsuccessfully, to discover it.

Particularly in the instances when the transfer task was recognition of *new examples of a learned principle,* practice in using the principle may be the most important variable. Of the groups which were tested for ability to discover *new principles,* only the discovery groups in Gagné and Brown's (1961) study were more successful than the nondiscovery group. In the studies by Wittrock (1963), Craig (1956), and Kittell (1957), the subjects in the principle-given groups had the higher scores in discovering new principles. It is difficult to equate these studies, but the weight of this evidence does not appear to give an advantage to learning by discovery.

It is more difficult to attribute differences to practice in those experiments in which the guided discovery group performed best. We would hypothesize that it was a combination of practice, increased attention, and reflection upon what was learned that was responsible for the differences in results in these cases.

In regard to transfer, the argument appears to be that learning by discovery helps a student to organize knowledge and the knowledge therefore is more susceptible to transfer (Baskin, 1962; Bruner, 1961). Hilgard states, "Transfer to new tasks will be better if, in learning, the learner can discover relationships for himself, and if he has experience during learning of applying the principles within a variety of tasks" (1956, p. 487). However, Travers (1963) sees no advantage to learning by discovery and prefers the learning of principles and *overlearning* as the superior preparation for transfer.

In the experiments by Wittrock (1963), Craig (1956), and Kittell (1957) described above, the superior group had more opportunity for overlearning than any other groups in the same experiment. Mandler (1962, p. 425) cites evidence to the effect that "there is an initial negative transfer effect followed by a reversal to a positive direction after the organism has had longer experience with the original task." Thus Mandler's results would appear to argue for overlearning on specific tasks. But in an experiment by Duncan (1958), where one series of groups had different schedules of overlearning on a single problem task and another series of groups learned the responses to varied stimuli, the group with experience in "learning to learn" was superior on transfer tasks. Hence, the role of overlearning in transfer remains unclear.

## Summary

It is acknowledged that some aspects of the problem of transfer of learning have not been discussed in this paper. Much of the paper has been devoted to the best way to learn principles in order to maximize transfer. The conclusions of Haslerud and Meyers (1958) and of Kersh (1958, 1964) contradicted those of Kittell (1957) so that it is not clear whether principles should be derived independently by the learner or learned through a certain amount of direction from the teacher. Kersh (1958) is of the opinion that this is exactly the teacher's dilemma. The teacher has to decide whether the most important outcome of a learning experience should be maximum understanding or maximum motivation to continue learning. In our judgment, both outcomes are essential to maximum transfer. Thus, the teacher is confronted with the task of striking the proper balance.

Ausubel (1961) claims: "Learning by discovery has its proper place among the repertoire of accepted pedagogic techniques available to teachers. For certain designated purposes and for certain carefully specified learning situations, its rationale is clear and defensible." ( 1961, p. 53.) On the other hand, he argues that discovery methods are not unique in their ability to generate self-confidence, intellectual excitement, and sustained motivation for learning. Finally, he states his position that available research does not provide a basis for generalizing to any one position.

We have concluded from this investigation what other writers have concluded in the past; namely, that transfer of learning is not automatic. The objectives of the methodology must be carefully formulated with transfer as a primary goal and with provision for various learning experiences as a means to the goal. Also, we believe that the learning of principles increases positive transfer in most situations and that principles *discovered* by the learner are more susceptible to transfer than those learned by *rote*. Finally, it is not completely clear whether principles should be discovered relatively independently by the learner or through close direction from the teacher in order to increase transfer. A crucial question that needs to be answered here is whether the increased expenditure of time required for independent discovery warrants its use. Similarly, the role which verbalization plays in transfer of mathematics learning remains unclear. Consequently, specific additional research is needed in these areas.

Ausubel (1961), in reviewing a sample of research studies, states that such relevant learning variables as rote-meaningful, inductive-deductive, verbal-nonverbal, and intramaterial organization were not controlled. Thus, the generalizability of such studies is limited. Ausubel's observations should be considered in future research in teaching, discovery, and the problems of transfer of training in mathematics.

## BIBLIOGRAPHY

AUSUBEL, D. P. "Learning by Discovery: Rationale and Mystique," *Bulletin of the National Association of Secondary School Principals*, XLV (December 1961), 18-58.

BASKIN, S. "Experiment in Independent Study," *Journal of Experimental Education*, XXXI (1962), 183-86.

BRUNER, J. S. "The Act of Discovery," *Harvard Educational Review*, XXXI (1961), 21-32.

———. "On Learning Mathematics," THE MATHEMATICS TEACHER, LIII (December 1960), 610-19.

———. *The Process of Education*. Cambridge, Mass.: Harvard University Press, 1960.

CAMPBELL, D. T., and STANLEY, J. C. "Experimental and Quasi-Experimental Design for Research on Teaching," in *Handbook of Research on Teaching*, ed. N. L. GAGE. Chicago: Rand McNally, 1963.

CRAIG, R. C. "Directed Versus Independent Discovery of Established Relations," *Journal of Educational Psychology*, XLVII (1956), 223-34.

CRONBACH, L. J. *Educational Psychology*. New York: Harcourt, Brace & Co., 1954, pp. 245-68.

———. "Issues Current in Educational Psychology," in *Mathematical Learning*, eds. L. N. MORRISETT and J. VINSONHALER. (Monographs of the Society for Research in Child Development.) Chicago: University Press, 1965 (a).

———. "The Logic of Experiments on Discovery." Paper prepared for The Conference on Learning by Discovery, New York, N.Y., January 28-29, 1965 (b).

DUNCAN, C. P. "Transfer After Training with Single Versus Multiple Tasks," *Journal of Experimental Psychology,* LV (1958), 63-73.

ERVIN, S. M. "Transfer Effects of Learning a Generalization," *Child Development,* XXXI (1960), 537-54.

FAWCETT, H. P. *The Nature of Proof,* Thirteenth Yearbook, National Council of Teachers of Mathematics. New York: Columbia University, 1938.

FERGUSON, G. A. "On Learning and Human Ability," *Canadian Journal of Psychology,* VIII (1954), 95-112.

———. "On Transfer and the Abilities of Man," *Canadian Journal of Psychology,* X (1956), 121-31.

GAGNÉ, R. M. "Implications of Some Doctrines of Mathematics Teaching for Research in Human Learning," in *Psychological Problems and Research Methods in Mathematics Training,* eds. R. L. FEIERABEND and P. H. DUBOIS. U.S. Office of Education. St. Louis: Washington University, 1959.

GAGNÉ, R. M., and BROWN, L. T. "Some Factors in the Programing of Conceptual Learning." *Journal of Experimental Psychology,* LXII (1961), 313-21.

GAGNÉ, R. M., *et al.* "Factors in Acquiring Knowledge of a Mathematical Task," *Psychological Monographs,* LXXVI (7 [Whole No. 526], 1962).

HASLERUD, G. M., and MEYERS, S. "The Transfer Value of Given and Individually Derived Principles," *Journal of Educational Psychology,* XLIX (1958), 293-98.

HENDERSON, K. B. "Research on Teaching Secondary School Mathematics," in *Handbook of Research on Teaching,* ed. N. L. GAGE. Chicago: Rand McNally, 1963.

HENDRICKSON, G., and SCHROEDER, W. H. "Transfer of Training in Learning to Hit a Submerged Target," *Journal of Educational Psychology,* XXXII (1941), 205-13.

HENDRIX, G. "A New Clue to Transfer of Training," *Elementary School Journal,* XLVIII (1947), 197-208.

HILGARD, E. R. *Theories of Learning.* New York: Appleton-Century-Crofts, 1956.

HILGARD, E. R.; IRVINE, R. P.; and WHIPPLE, J. E. "Errors in Transfer Learning with Understanding: Further Studies with Katona's Card Trick Experiment." *Journal of Experimental Psychology,* XLVII (1954), 457-64.

———. "Rote Memorization, Understanding, and Transfer; An Extension of Katona's Card Trick Experiment." *Journal of Experimental Psychology,* XLVI (1953), 288-91.

JUDD, C. H. "The Relation of Special Training to General Intelligence," *Educational Review,* XXXV (1908), 28-42.

KATONA, G. *Organizing and Memorizing.* New York: Columbia University Press, 1940.

KENDLER, H. H., and VINEBERG, R. "The Acquisition of Compound Concepts as a Function of Previous Training." *Journal of Experimental Psychology,* XLVIII (1954), 252-58.

KERSH, B. Y. "The Adequacy of 'Meaning' as an Explanation for the Superiority of Learning by Independent Discovery," *Journal of Educational Psychology,* XLIX (1958), 282-92.

———. "Directed Discovery Versus Programmed Instruction: A Test of a Theoretical Position Involving Educational Technology." Final report. Title VII, Project No. 907. Monmouth, Oreg.: Oregon State System of Higher Education, 1964.

KITTELL, J. E. "The Effect of External Direction During Learning on Transfer and Retention of Principles," *Journal of Educational Psychology,* XLVIII (1957), 391-405.

KOLESNIK, W. B. "Note on Transfer of Training Research," *Peabody Journal of Education,* XXXIV (1957), 215-19.

MANDLER, G. "From Association to Structure," *Psychological Review,* LXIX (1962), 415-27.

MCDONALD, F. J. *Educational Psychology.* Belmont, Calif.: Wadsworth Publishing Co., 1959.

McGeoch, J. A., and Irion, A. L. *The Psychology of Human Learning*. New York: Longmans, Green, 1952.

Melton, A. W. "Review of Katona, Organizing and Memorizing," *American Journal of Psychology*, LIV (1941), 455-57.

Murdock, B. B., Jr. "Transfer Designs and Formulas," *Psychological Bulletin*, LIV (1957), 313-26.

Parker, E. "Teaching Pupils the Conscious Use of a Technique of Thinking," The Mathematics Teacher, XVII (1924), 191-201.

Perry, W. *A Study in the Psychology of Learning Geometry*. New York: Columbia University, 1925.

Thorndike, E. L., and Woodworth, R. S. "I. The Influence of Improvement in One Mental Function upon the Efficiency of Other Functions; II. The Estimation of Magnitudes," *Psychological Review*, VIII (1901), 247-61; 384-95.

Travers, R. M. W. *Essentials of Learning*. New York: Macmillan, 1963, pp. 187-216.

Ulmer, G. "Teaching Geometry to Cultivate Reflective Thinking: An Experimental Study with 1,239 High School Pupils," *Journal of Experimental Psychology*, VIII (1939), 18-25.

Wittrock, M. C. "The Learning by Discovery Hypothesis." Paper prepared for The Conference on Learning by Discovery, New York, N.Y., January 28-29, 1965.

———. "Verbal Stimuli in Concept Formation: Learning by Discovery," *Journal of Educational Psychology*, LIV (1963), 183-90.

Young, R. K., and Underwood, B. J. "Transfer in Verbal Materials with Dissimilar Stimuli and Response Similarity Varied," *Journal of Experimental Psychology*, XLVII (1943), 153-59.

# X

# Some Ongoing Research and Suggested Research Problems in Mathematics Education

BOYD HOLTAN
*University of Florida*
*Gainesville, Florida*

KNOWLEDGE of ongoing research activities is helpful both in planning one's own research and in improving existing programs in mathematics education. Because of poor communication, there is not only a good deal of unnecessary duplication, but also a lack of needed replication. The communication channel between research workers and classroom teachers must also be open. This is particularly serious since information and products that do not get to the practitioner can obviously have no practical value. There are research projects in mathematics education, both large and small, which are being conducted and are completely unknown to many mathematics educators. Therefore, in planning this publication, the Research Advisory Committee of the NCTM felt it would be appropriate to list some activities which would give indications of what is happening in mathematics education research.

A short questionnaire was sent to a sample of mathematics educators asking for a response to two questions.

The first question was, "What research is being conducted at your institution which is related to some aspect of mathematics education?"

The second question was, "What do you feel is the most pressing problem in mathematics education to which research might aid in contributing a solution?"

The answers received for each of the two questions are reported below under one of three categories: (1) Developmental Activity, (2) Product-Oriented Research, and (3) Information-Oriented Research.

## Ongoing Research [1]

*Developmental activity*

A Writing Project for Developing Text Materials for Elementary Teacher Training in Mathematics

The Development of Ways of Presenting Arithmetic to Elementary Teachers—Relating It Very Closely to the Real World

The Development of a Statewide Continuing In-Service Program for Secondary School Mathematics Teachers

The Development of a Graduate Level Course on the History of Mathematics

"The Development of Facility in Exposition" in a Methods Course for Students with Extensive Prior Training in Mathematics (Essentially a Mathematics Major)

The Development of a Mathematics Institute Program

Development of an Instructional System Involving Television, Text, and Teacher, for Teaching Mathematics to In-Service Elementary School Teachers

The Development of New Materials for High School Geometry

The Development of a System for Teaching Mathematics Through the Use of a Time-Shared Computer

A First Step Towards the Implementation of the Cambridge Mathematics Curriculum in a K–12 Ungraded School

The Development of Discovery Units

*Product-oriented research*

A Study of Textbooks Versus Lectures in the Preparation of Elementary Teachers

The Design and Evaluation of an Individualized Program in Elementary School Mathematics

The Development and Evaluation of Minnemast: Elementary School Science and Mathematics Programs

An Evaluation of the Presentation of First-Year Algebra in Two National Experimental Programs Based on Selected Criteria from the Theory of Learning

A Comparative Study of Two Methods of Teaching Mathematical Analysis at the College Level

---

[1] If the reported research obviously had both developmental and product-oriented (usually evaluation) aspects, it was listed under product-oriented research; if it had both product and information aspects, it was listed under information-oriented research.

The Development and Evaluation of a New Mathematics Curriculum, Grades 7–12

An Evaluation of the Effectiveness of Closed-Circuit Television in Teaching Mathematics to Prospective Elementary Teachers

An Evaluation of the Effectiveness of Teaching by Induction, via the Use of a Computer-Based Teaching Machine

The Standardization of a Number Systems Test for Elementary Majors

An Experimental Study of the Effectiveness of Computer-Mediated Instruction in Mathematics

A Study of the Effectiveness of Minnemast Materials on Groups, Vectors, and Transformations

The Development and Evaluation of Test Items for Elementary and Secondary School Mathematics Curricula at Each of Bloom's Taxonomic Levels

The Development of a Collection of Film Loops Which Depict Certain Well-Defined Teaching Strategies, and a Study of Their Effectiveness for Teacher Training

A Teach-Test Procedure for Obtaining Measures of Mathematical Aptitude

A Comparison of Two Methods of Presenting an Axiom System Using a Computer-Assisted Instructional Unit Designed to Teach Deductive Proof

The Effects of Team Teaching in Junior High School Mathematics

The Development and Evaluation of Procedures for Measuring Understandings in Arithmetic

The Effectiveness of Programmed Instruction in Teaching Plane Geometry

The Effects of Teaching a Unit on Logic as a Part of a College Course in Calculus

An Experimental Investigation of the Effectiveness of the Kansas Demonstrations of Mathematical Concepts in the Teaching of Mathematics in the Elementary Grades

An Investigation of the Effect of Types of Exercises in Teaching Mathematical Concepts to Prospective Elementary School Teachers

The Effects of Different Kinds of High School Experience with the Limit Concept on the Study of Calculus in College

A Comparison of Methods of Teaching Abstract Algebra in College

The Identification of the Algebraic Concepts Needed for the Instruction of Mathematics in the Elementary School and the Designing of a Related Course of Study

The Identification of Concepts from Probability and Statistics Needed for

Instruction of Secondary School Mathematics Teachers and the Designing of a Related Course of Study

The Difference Between Large and Small Sections in Calculus

The Development and Evaluation of a Test of Understanding of Selected Properties of a Number System: Primary Form

The Development and Evaluation of a Test of Arithmetic Principles: Elementary Form

*Information-oriented research*

An Analysis of the Learning Problems Involved in Teaching the First Grade

The Interrelationships Among Selected Personality Traits, Levels of Cognitive Structure, and Teaching Strategies

Mathematical Models as Mediators in Facilitating or Inhibiting Growth in Problem-Solving Ability

The Effect of Teaching Certain Concepts of Logic on the Verbalization of Discovered Mathematical Generalizations

The Influence of Discovery Teaching on the Ability to Solve Mathematical Problems

The Relationship Between "Strategy of Search Training" in Non-Mathematical Fields and the Learning of Mathematics

A Characterization of Provers and Nonprovers in an Axiomatic Geometry Course for Elementary Education Majors: A Discriminate Analysis

A Study of the Role of Symbolism in Learning Mathematical Principles

A Study of the Relationships Between Problem Solving and Prior Learning

A Study of the Effectiveness of Using Conceptual Organizers in Learning Abstract Mathematics

The Development of a Scientific (Theoretical) Language for the Precise Formulation of Basic Research on Mathematics Learning

The Role of Inductive Strategies in the Teaching of Mathematical Concepts and Generalizations

The Relationship Between Student Interest in the Instructional Materials and Mathematics Achievement

The Determination of How Children Solve Novel Mathematical Problems

The Identification of Factors Contributing to the Understanding of Selected Basic Arithmetical Principles and Generalization

The Relationship Between Teachers' Knowledge of Arithmetic and Pupil Gain

The Measurement of Teacher Attitude in Relation to Contemporary Mathematics Programs

The Relationships Between Underachievement and Low Achievement and Mathematics Learning

A Study of the Relative Importance of Certain Factors in the Prediction of Successful Performance in Seventh-Grade Mathematics

A Comparative Study of Selected Factors of Mathematics Achievement in Homogeneous Groups of Fifth-Grade Pupils Taught by a Discovery Approach

Success in Mathematical Statistics as a Function of Mathematical Background

The Measurement of Affective Changes Among Elementary Majors During Their Undergraduate Careers: A Longitudinal Study

PROBLEMS

*Developmental problems*

The Development of Additional Materials and Courses for Teachers of Prospective Elementary School Teachers

Procedures for Developing a Desire to Learn Mathematics, Especially for Students at About Eighth- or Ninth-Grade Level Who Have Been in the Lower Achievement Group

The Development of Better Diagnostic and Remedial Procedures for Use with Individuals

The Development of Improved Teacher-Behavior Training Programs

*Product-oriented problems*

An Evaluation of the Effectiveness of "Modern" Math Programs and Instructional Methods Related to "Modern" Topics

Using the "Best" Texts that Can Be Constructed Today, What Can the "Best Possible" Present-Day Teaching Accomplish with Various Levels of Students?

Teacher Training—What Kind of Programs of Teacher Training Can Best Perform the Function of Preparing Teachers to Do Justice to New Programs?

The Determination of Effectiveness of "Discovery Teaching"

To What Degree Do Modern Elementary Math Textbooks and Programs Which Are Almost Completely Dependent upon Diagrams, Games, Puzzles, Tricks, etc., Contribute to the Learning and Use of Basic Mathematics?

The Development of Valid Measuring Devices Which Will Not Only Measure Skills but Also Concepts and Applications in New Situations

At What Degree of Rigor Do High School and College Freshmen Best Learn Mathematics?

An Evaluation of Various Techniques for Keeping Mathematics Teachers Up-to-Date on Recent Developments in Mathematics and in the Teaching of Mathematics

The Development of Tests to Measure Concept Development and Problem-Solving Ability

The Development and Evaluation of Procedures for Content Selection and Placement in Relation to Objectives of Mathematics in the Elementary School

The Development of Procedures to Aid the Low Achiever in Mathematics

*Information-oriented problems*

The Role of Intuition in the Learning of Mathematics

Using Clearly Defined Criteria, Is It True that "Any Subject (Topic) Can Be Taught to Any Child of Any Grade Level in Some Intellectually Honest Manner?"

Acquiring More Knowledge About the Relationships Between Teaching and Learning (This Might Be Called "Methods Research" Which in the Past Few Years Has Taken a Back Seat to Curriculum Research)

How Are Mathematical Concepts Formed?

What Is the Ability to Read Mathematical Material?

An *Intensive* Study of Outstanding Teachers' Behavior in Relation to Students' Learning

Determine Optimum Levels for Introducing Specific Skills and Ideas

Determine Methods Which Contribute Most to Retention

How Do Individuals Learn Mathematics?

How Is Mathematics Learned at Various Levels?

The Need to Improve Our Understanding of How to Teach Mathematics

How Do Elementary School Children Develop Concepts in Mathematics?

THESE indications of research activity trends were reported by about two dozen mathematics educators. The research is not necessarily being conducted by them, but is being done at their institutions. Twenty-eight of the research projects were listed under product-oriented research, eleven under developmental activities, and twenty-two under information-oriented research. The problems posed were also about equally divided between product-oriented research and information-oriented research. The relatively small number of developmental projects and problems listed, however, may not truly represent the current situation since the contributors were *not* asked to list developmental activities and since many leaders in mathematics education do not classify such activities, though highly significant, as research. In effect, they make a sharp distinction between scientific research and artistic development. The former has been emphasized in this publication.

On the basis of this sample listing of ongoing research and research problems, it appears that there certainly is activity in mathematics education research which could be usefully shared by all who are interested in the field. Furthermore, a careful perusal of the projects and problems listed strongly suggests that what is a research problem at one institution may be an ongoing research activity at another. In addition to making mathematics educators aware of present-day research activity and concern, it is hoped that this compilation may also have a motivational effect on future research activities.

# XI

# Research in Mathematics Education—An Overview and a Perspective

JOSEPH M. SCANDURA*
*Graduate School of Education*
*University of Pennsylvania*
*Philadelphia, Pennsylvania*

WITH so many different kinds of research and development presently underway in mathematics education, it seems desirable, in this final article, to provide a perspective in which these activities might be viewed. Particular attention is given to the nature of and the relationships between information-oriented (basic) and product-oriented (applied) research. In the process, some of the major points made in the preceding articles are highlighted and some of the interrelationships between them are pointed out. The points raised, however, should be taken as selective rather than exhaustive.

Let me begin by making a distinction between scientific research and developmental activity, or, as it is frequently called, "action research." In the present context, development refers primarily to those innovative classroom activities which have had so great an effect on mathematics education in recent years. The term "development," rather than "research," is used because most, although not all, of the resulting materials and procedures were obtained not by applying any existing theory or technology, but simply on the basis of the perspicuitive intuition or artistry of mathematicians who were also master teachers. Many of the innovators, themselves, are quick to point out that neither the scientific method nor scientific results were used in any way.

This relatively informal and intuitive approach was sufficient in the

* The author would like to thank Drs. E. E. Boe, C. E. Dwyer, and J. P. Williams for their helpful comments on a draft of this article.

immediate past because the gap between mathematics, as practiced by twentieth-century mathematicians, and mathematics, as it then existed in the schools, had become an abyss. Bridges had to be built, almost any kind of bridges.

Now that the revolutionary period is giving way to a more thoughtful evolution, the situation is changing. Mathematics educators and others concerned with the new mathematics programs are beginning to demand "hard facts" to support the claims made by proponents of the various programs. If for no other reason, evaluation has been felt necessary to justify the funds spent on development. Since many of the innovators had neither the training nor the inclination to pursue this part of the task themselves, they have enlisted the aid of psychologists and specialists in educational research.

Originally, the concern was with the question, "Does this new program (set of materials, etc.) work as well as what we have been doing (using, etc.)?" [1] Berger and Howitz have reported the results of a comparative evaluation study designed to answer just this sort of question. More important, they have shown how some of the problems confronted in evaluating a new program can be handled. Anyone who has conducted such research knows how frustrating it can be when pupils get sick and are forced to miss a crucial test, when teachers unwittingly contaminate the experimental treatments, when administrative difficulties make the random assignment of pupils and teachers to treatments impossible, etc. It is satisfying to know that a variety of statistical procedures is available to partially compensate for such factors.

For the most part (there have been exceptions), the new materials and curricula (e.g., *Experiences in Mathematical Discovery*) that have been evaluated have proved to be more effective than the materials and curricula they were designed to replace, insofar as the newer topics are concerned, and equally as effective with regard to more traditional topics.

Once having demonstrated that a new set of instructional materials or a curriculum does no harm, and, indeed, seems promising, the next step is to improve it. For this purpose, a rather simple research strategy or methodology has been found useful. Determine the learning outcomes of the new materials or curriculum in question and, by comparison with certain predetermined and objective standards, determine where the materials and/or instructional procedures are adequate and where they are lacking. Such information, of course, is then used in revision, possibly followed by another evaluation cycle. During the course of such a

[1] Notice that this same question can be asked of any new product—whether it be a new light bulb, pill, or automobile.

development and evaluation cycle, material developers and research workers are often forced to reconsider their objectives and to translate these objectives into a form that can be measured. The result is almost always an improved product.

That part of the cycle referred to as materials development, since it is based almost entirely on intuition, is perhaps best viewed as an art and not research. The research phase of the cycle consists of the evaluation itself. This kind of comparison with absolute standards has long been used, in a slightly modified form, by teachers (in the course of periodic testing), was used somewhat later by program writers, and more recently is gaining favor as an alternative method of curriculum evaluation.[2]

Both approaches to evaluation, *comparative* and *predetermined standards,* since they deal with products, rather naturally fall into the category of "product-oriented" research. It must be apparent, however, that without formal guidelines to be used in the development of instructional materials, the materials produced depend almost entirely on the ability of the writers. In order to capitalize on the skills of specialists in a variety of related disciplines in developing materials, an increasing number of research and development centers have found it desirable either to apply existing technologies (i.e., systematic developmental procedures) or to devise new ones. Because of the difficult problems of integration and the like, there often is simply no other way to get the job done in an efficient manner.

The procedures described by Kersh and Lipson provide two excellent examples of such technologies. Although both procedures make general use of the task analysis technology described by Gagné, Kersh dealt with engineering instructional sequences for use in the classroom and Lipson with the development of materials for use with individual students. Although intuitive judgments are always involved to some extent in the development of any product, these articles make it clear that the purely artistic approach of the materials producer can be replaced by a clearly specified technology, one which is subject to review, criticism, and (hopefully) continued improvement.

The mathematics educator, of course, must play the key role in determining what the objectives are to be and in actually writing the material —these tasks require an intimate familiarity with the subject matter. The psychologist plays his major role in helping to translate these objec-

[2] A more complete description of the development-evaluation methodology described above may be obtained from Dr. Wells Hively, Department of Educational Psychology, University of Minnesota. An example of an evaluation study with predetermined standards may be obtained from Dr. Wai-Ching Ho, Greater Cleveland Mathematics Program.

tives into terms that can be measured and in devising effective procedures for achieving these objectives.

Nonetheless, the serious question remains as to whether present-day instructional technologies can improve on, or even equal, what the skilled mathematical artist has been able to accomplish. One answer to such a challenge is that as technologies continue to improve, the improvements become available not only to the technology developers, themselves, but to anyone else who wants to use them and who is willing to take the time to learn how. On the other hand, when the artist improves his style with practice, the benefit is only to the artist himself and to those who have direct access to him as a teacher or to his products (e.g., texts, etc.). Kersh's reference to "second-order" objectives and Lipson's mention of attempts to capitalize on the "learning how to learn" evidenced by the students at the Oakleaf School both suggest basic changes in the respective technologies originally proposed. It is quite possible that one of the major reasons why a number of prominent curriculum developers in mathematics have had a generally negative attitude towards stating objectives in behavioral (i.e., observable) terms is that, in its preliminary form, the approach paid too little attention to secondary objectives and learning how to learn. The innovator almost always has several objectives in mind when he introduces a topic, even if only at the intuitive level. It is to be expected that, as still further improvements cumulate with time, technologies will play an ever increasing role in mathematics education.[3]

In view of the above discussion, the case for *product-oriented* research is quite direct.[4] Whenever research (e.g., evaluation) demonstrates the value of one product over another or that a product meets certain standards, or, whenever a technology makes it possible to produce more and better materials in an efficient manner, both the practitioner and the student benefit rather directly.

When it comes to basic *information-oriented* research in mathematics education, however, the payoff is not always so immediate. Nonetheless, Suppes has made an excellent case for an active program of basic research in mathematics education. Since he has stated his arguments so clearly, it is unnecessary to elaborate here. Let me simply summarize what appear to be his key points: (1) intuition alone provides an insufficient base for

---

[3] While both of the technologies described in this monograph are based in varying degrees on task analysis, there are many other kinds of technological development underway. These activities range from programming a computer so that it will be able to provide almost immediate answers to an author's questions about the effectiveness of his material (UICSM) to devising efficient procedures for assessing mathematical knowledge (University of Pennsylvania).

[4] It is for this reason that the project committee did not feel that an article paralleling that of Suppes on basic research was necessary.

devising new curricula (or instructional procedures)—intuitive judgments and objective facts are too often at opposite ends of the pole, (2) the number of sheerly empirical studies is certainly large in number, if not uncountable—achieving order out of chaos will depend on the development of a sound theory of mathematics learning, based on carefully thought out information-oriented studies, (3) there is a need to analyze and provide a theory for students' learning difficulties, and (4) a better understanding of how mathematics is learned and how mathematicians think may lead to a revised conception of the nature of mathematics itself —in particular, a more central emphasis may be given to the patterns of thought found useful in dealing with mathematics. I find it hard to disagree with Professor Suppes, for agreeing with what he has said. Nonetheless, in order to provide a perspective from which to view the four reports of information-oriented research, let me make a few additional comments.

The time-honored purpose of basic research is theory development. To be classified as basic, the research must deal with (1) the identification of and relationships between (2) well-defined variables which are (3) theoretically relevant. Whereas different variations on this theme may be found, most scientists and philosophers of science would probably not find too much quarrel with this definition, particularly in the present context where it is being used primarily to specify one of two admittedly highly overlapping categories (i.e., information- and product-oriented research).

It is important to notice from the beginning that this definition makes no mention of experimental or statistical methodology—something which is often mistakenly taken as evidence of *basic* research in education. The position taken here is that any approach which furthers the goal of basic research deserves to be classified as such. In the *experimental* approach, for example, one or more variables are systematically varied, and the effects of this manipulation on other (dependent) variables are determined. The article by Worthen serves as an example of basic experimental research which also has rather direct practical implications. Perhaps the most noteworthy feature of this research is that it provides support for two major contentions of discovery enthusiasts. The discovery group not only performed better than the expository group on tests designed to measure the transfer of heuristics but they better retained the material that had been originally taught. While this is not the first time such results have been found,[5] Worthen's experiment certainly represents one of the best-

[5] See, for example, R. M. Gagné and L. T. Brown, "Some Factors in the Programming of Conceptual Learning," *Journal of Experimental Psychology*, LXII (1961), 313-21.

controlled comparisons of expository and discovery methods in mathematics which extended over a period of weeks. It is particularly encouraging to find that laboratory results and field trials often coincide.

Another common approach to information-oriented research, often called the *correlational* approach, involves uncovering relationships between two or more dependent measures. The strong relationship Dienes found between the way a mathematical task is perceived by a learner and the learning strategy followed illustrates the utility of this approach.

A third type of information-orieinted research involves setting up a single well-defined situation, determining the outcomes in an objective fashion, and, then, comparing the obtained outcomes with predictions made on the basis of one or more theories or analyses. The studies reported by Suppes and Groen and Gagné well exemplify this third approach. Suppes and Groen compared predictions, based on five alternative algorithms for finding the sum of two numbers, $m$, $n$, where $m + n \leqslant 5$, with the latencies (i.e., time between presenting a problem and the occurrence of the correct answer) actually obtained. The best fit was obtained by an algorithm, in which the largest of the two given numbers is stored and successively incremented by one until the smaller value (number) has been added on. In effect, characteristics of the group data (i.e., statistics of the obtained score distribution) could be best predicted by assuming that *all* of the experimental subjects used this algorithm to add. As the authors suggested, they do *not* necessarily believe that this is true, only that the group's mean performance could be predicted best by making this assumption. Gagné's rationale was based on the assumption that a learner's *existing state of knowledge* is equally as important in determining future learning as the instructions (or information) given. His results appear to provide strong support for this position. Furthermore, the relationship between learning and prerequisite performance, as determined during the learning sequence, and aptitude, as measured by standardized instruments, became stronger and weaker, respectively, as learning progressed toward the hierarchical apex.

On the surface, these findings of Gagné and those of Suppes and Groen appear to clash head on. To Gagné, the prior state of the learner appears to be critical in determining what will (or can) be learned. A rapid reading of the Suppes and Groen article, on the other hand, might lead one to think that individual differences have been ignored.

Rather than being contradictory, I feel that the differences exemplified by these studies have deep roots and, in fact, are suggestive of two critically important, but fundamentally different, aspects of mathematical learning and performance. Gagné was concerned largely with the logically determined prerequisites for successful performance on a mathematical task. His experimental data simply provided empirical support for the validity of his analysis. Had the results not conformed to prediction, the difficulty would have been due more to the logical inadequacy of the task analysis than to a lack in any theory of behavior. In the Suppes and Groen study it seems reasonable to suppose that most of the subjects had at their command the logical prerequisites for *all* five algorithms proposed, particularly since the five sets of prerequisites undoubtedly overlap. The reported results were obtained on the third day of the experiment, after the experimental subjects had attained a high level of mastery on the tasks, so that the experimental data probably reflected a *preference* for one of the algorithms rather than any additional learning. The basis for such a preference might well involve some sort of complex interaction between certain basic psychological capacities of learners (presumably reflecting underlying physiological capacities such as the amount which can be stored in short-term memory) and what is already learned. In short, Gagné was largely concerned with determining prerequisites for successful performance, while Suppes and Groen, implicitly assuming a common level of prior knowledge, sought to determine what knowledge *would* be used. The relative power of each approach depends on what kinds of predictions one wants to make.

*I would propose that both kinds of research are badly needed. Any reasonably complete understanding of mathematical learning and performance will depend on (1) the identification of those "ideal" competencies underlying various kinds of mathematical behavior (e.g., what are the prerequisites for syllogistic reasoning?) and (2) an understanding of how inherent psychological capacities and subject matter competencies already had by a learner interact with external stimulation to produce mathematical learning and performance.*

Before passing on, one further point deserves mention. Assessing a learner's state of knowledge cannot always be determined in a direct manner. Suppose, for example, an experimental subject has learned to give the integers, 8, 11, and 5, as responses to the four-tuples (stimuli) (3, 8, 9, 4), (9, 7, 8, 6), and (6, 5, 8, 9), respectively. The question remains as to *just what he has learned.* Has he learned the three four-tuple-integer pairs as distinct entities, noticing no relationships between them? Or has he learned (discovered) that the response integers can be determined from the stimuli by adding the numbers in the first and third positions of the corresponding four-tuple and subtracting from this sum the number in the fourth position?

Some of our [6] recent research suggests that presenting a new four-tuple, such as (4, 8, 9, 3), may provide a sufficient test for deciding between these alternatives. If, under certain conditions, the learner gives the response, 10, one can be quite certain that he has learned the rule stated above. If not, he has probably failed to notice the essential similarity between the three original four-tuple-integer pairs. Furthermore, having once used the rule, the learner will almost invariably use the same rule again when confronted with a second four-tuple—unless he either has conflicting knowledge at his command or has been led to believe that the rule is no longer appropriate or that his response to the first test stimulus was wrong (e.g., by telling him). This assessment procedure is quite general and can be used with any principle that can be stated in the form, "If *A*, then *B*." [7]

Still a fourth approach to information-oriented research involves the careful and often painstaking *naturalistic* observation for which Piaget is so famous. On the basis of intuition and detailed observations of how young children learn mathematics, Dienes has identified those kinds of activity which he feels are fundamental to all mathematics learning. He has singled out for special emphasis *play*, informal exploratory behavior; *abstraction*, the identification of that which is common to a number of situations; *generalization*, the extension of an abstract class to a broader class; *particularization*, the passage from a broader class to one more restrictive; *symbolization*, the symbolic representation of mathematical ideas; and *interpretation*, the determination of meanings underlying symbols. To this list may be added *deduction*, the (logical) derivation of new relationships, and *axiomatization*, the determination of a (small) basic set of relationships from which all others may be derived. Taxonomic activity of this sort is a general characteristic of any new science, in this case *"psycho-mathematics"* or the psychology of mathematics learning. Until the basic kinds of phenomena with which the new science must deal have been adequately determined, the variables chosen for study may lead to relationships which are merely symptomatic of, rather than fundamental to, an underlying theory.

Review articles, such as that by Becker and McLeod, also play a vital role in information-oriented research. This is particularly true when the authors provide a rationale both for classifying existing research and for placing proposed research into a perspective. While a few excellent examples exist in the mathematics education literature, there have been far

[6] Scandura, J. M., "Precision in Research on Mathematics Learning: The Emerging Field of Psycho-Mathematics," *Journal of Research in Science Teaching*, V (1967).

[7] The good teacher may notice the similarity between this test procedure and what is typically referred to in the classroom as the *Aha!* experience.

too few.[8] For many purposes, a simple listing will not suffice. Becker and McLeod have provided a valuable service not only in reviewing, but in providing a framework for viewing transfer of training, a topic of great concern to mathematics educators.

In order to dispel any remaining doubt, let me emphasize that, as defined herein, experimental research is not synonymous with basic information-oriented research. The typical comparative evaluation study, for example, would not meet the proposed criteria. In effect, finding relationships between variables is not a sufficient condition. Not only must variables be specified, but they must be well-defined in a mathematical sense. When one talks about one curriculum being better than another, the question remains as to just what makes it better. What goes into a curriculum, when presented by one teacher, may be quite different when presented by another. In short, equivalence classes of mathematical curricula typically are not behaviorally invariant, even in a probabilistic sense.

Even finding relationships between unambiguously defined variables, however, may not be sufficient. To have a *direct* effect on theory development, research should be aimed at determining fundamental variables and relationships. In many cases it is hard to determine just when this requirement is met since which variables are deemed basic and which theoretically superficial (although perhaps of immense practical concern) depends, in large part, on the stage of development of the science in question. An example may not only help to clarify this distinction but help to locate the present rapidly changing state of knowledge about the teaching-learning process. Consider grade level, a variable which is frequently included in educational experimentation. This variable is well-defined, but not basic according to the present definition. While it has been observed many times that certain topics are learned better when taught at one grade level than at another, it has more recently been established, by a number of investigators,[9] that prior learning may be the crucial factor involved. That is, the reason grade level has so often been related to teachability is probably that the necessary prerequisites have tended to covary with grade level. Obtained relationships between grade level and learning, then, should be deducible from a knowledge of the abilities had at the various grade levels involved. The facts that it might be difficult to measure all of the necessary pre-

[8] Examples are provided by K. E. Henderson, "Research on Teaching Secondary School Mathematics," in *Handbook of Research on Teaching*, ed. N. L. Gage (Chicago: Rand McNally, 1963), pp. 1007–30; *The Learning of Mathematics, Its Theory and Practice*, the Twenty-first Yearbook of the NCTM; and some of the U.S. Office of Education pamphlets edited by K. E. Brown.

[9] The study by Gagné provides a case in point.

requisites and that knowledge of these prerequisites is crucial to any complete understanding of mathematics learning do not alter the situation fundamentally. A study designed to determine relationships between grade level and teachability, while it might provide a great deal of practical information (information which might be put to use in preparing instructional materials) would not add to our store of fundamental knowledge. Such information-oriented research is typically referred to as being empirical in nature.

Nonetheless, empirical research frequently results in information which can *not* be derived from other findings. In such cases, the information so attained sometimes serves as an impetus for theory development. Too often, however, this is not the case. Facts, even discrepant facts, frequently pile up with little resulting attempt at theoretical explication. For these reasons, a strong case can be made for distinguishing between information-oriented research which is directed specifically at theory development and (empirical) information-oriented research in which the variables chosen for study neither have explanatory power themselves nor are explained in terms of more generic variables (having such explanatory power). The term "basic (or theory-oriented) research" might well be reserved for the former type, in which the concern is either with the identification of, or relationships between, fundamental variables or with research which, while derivable from more basic findings, makes these derivations explicit, whether in the form of highly elaborate theories or relatively imprecise rationales.

To avoid needless dispute, let me emphasize that it is often difficult to distinguish between information-oriented research and product-oriented research, let alone between information-oriented research which is explicitly theory-oriented and information-oriented research which is not. Furthermore, even developmental activity frequently provides valuable information (or at least raises important theoretical questions) while the results of information-oriented resetarch may find rather direct application. The many-faceted nature of much research is well exemplified by the Kersh and Worthen articles and by several of the listings of ongoing and needed research which were solicited and compiled by Holtan. Perhaps the ultimate basis for categorizing a study is the researcher's motivation—to find out why or to improve an existing situation.

The major purpose of this article has not been to favor information- or product-oriented research over artistic development but simply to help clarify some of the interrelationships between them. It has been suggested, however, that if mathematics education is to improve fundamentally beyond its present state more will be required than simply teaching more mathematics at an earlier age. We, as mathematics educators, will have to turn our attention more and more towards the development of improved technologies for preparing materials and for instructing students. Such advanced technologies, in turn, may be expected to depend increasingly on a more complete understanding of how mathematical knowledge is organized, learned, taught, measured, and created.

Information-oriented research, product-oriented research, and development are all necessary. Information-oriented research, without related product development, is of no use to mankind while product-oriented research and development, without supporting basic research, may too easily become tradition-bound—or, what is equally bad, revolution-bound.