R E P O R T   R E S U M E S

ED 013 377   56   AA 000 229

COMPARATIVE STUDIES OF PRINCIPLES FOR PROGRAMMING MATHEMATICS IN AUTOMATED INSTRUCTION. FINAL REPORT.

BY- STOLUROW, LAWRENCE M.   BEBERMAN, MAX
ILLINOIS UNIV., URBANA, BUR. OF EDUC. RESEARCH

REPORT NUMBER NDEA-VIIA-806   PUB DATE   JUL 64
EDRS PRICE   MF-$0.25   HC-$2.04   51P.

DESCRIPTORS- *EDUCATIONAL RESEARCH, *MATHEMATICS INSTRUCTION, COMPUTER ASSISTED INSTRUCTION, *PROGRAMED INSTRUCTION, TEACHING TECHNIQUES, INSTRUCTIONAL TECHNOLOGY, *LITERATURE REVIEWS, *RESEARCH PROJECTS, RESEARCH METHODOLOGY

AN ASSESSMENT WAS MADE OF PRINCIPLES USED FOR PROGRAMING MATHEMATICS IN AUTOMATED INSTRUCTION. THE FIRST CHAPTER PRESENTED A SUMMARY OF RESEARCH ON THE TEACHING AND LEARNING OF MATHEMATICS BY PROGRAMED INSTRUCTIONAL PROCEDURES. THE RESEARCH AND FINDINGS WERE CONSIDERED WITH RESPECT TO THEIR RELATIONSHIP TO THE DEVELOPING TECHNOLOGY OF EDUCATION. THE SECOND CHAPTER CONSISTED OF A SUMMARY OF RESEARCH PURSUED BY THE RESEARCH TRAINING LABORATORY ON PRINCIPLES IN PROGRAMING WITHIN THE CONTEXT OF MATHEMATICS. THE RESEARCH WHICH WAS REPORTED DEALT WITH LINEAR AS OPPOSED TO BRANCHING FORMS, PROGRAMING FOR DISCOVERY LEARNING, ORDERING OF FRAMES, STEP SIZE, COVERT AS OPPOSED TO OVERT RESPONDING, CONVENTIONAL AS OPPOSED TO PROGRAMED INSTRUCTION, ABILITY AND LEARNING, AND METHODOLOGICAL STUDIES. (GD)

VII A-806

TRAINING RESEARCH LABORATORY

Department of Psychology — Bureau of Educational Research
University of Illinois — 8 Lincoln Hall — Urbana, Illinois

University of Illinois
Urbana, Illinois

Lawrence M. Stolurow and Max Beberman

COMPARATIVE STUDIES OF PRINCIPLES FOR PROGRAMMING MATHEMATICS IN AUTOMATED INSTRUCTION

Final Report
July 1964

Co-Investigators:

Lawrence M. Stolurow
Professor, Dept. of Psychology
Training Research Laboratory

Max Beberman

Professor, College of Educ.
Univer. of Ill. Committee
on School Mathematics (UICSM)

Project Sponsor:

Educational Media Branch
U. S. Office of Education
Title VII

Project No. 711151.01

Table of Contents

Chapter 1

Programed Instruction and the Teaching of Mathematics[1]

This chapter presents a summary of research on the teaching and learning of mathematics by programed instructional procedures. The research and findings are considered with respect to their relationship to the developing technology of education.

Studies of Mathematics Teaching and Learning

The teaching of mathematics represents the largest of all subject matter areas in the research on the use of self-instructional programs (see Hendershot, 1964; Hanson, 1963). Although most of the programs and research are based on topics from secondary school level mathematics, the actual range is from elementary school through college. Topics covered include arithmetic, algebra (including Boolean), geometry, set and number theory, trigonometry, calculus, vectors, as well as areas of applied mathematics such as conventional statistics both descriptive and inferential, (e. g., Hickey, Autor and Robinson, 1962) and linear programing for management decision making (e. g., Glaser and Reynolds, 1962).

---

[1] Based upon a paper presented at the AAAS Symposium on Learning Research Pertinent to Educational Improvement, Cleveland, Ohio, December 29, 1963.

both descriptive and inferential, (e. g., Hickey, Autor and Robinson, 1962) and linear programing for management decision making (e. g., Glaser and Reynolds, 1962).

Two Types of Research

Two types of research, each with different objectives, are dominant in the studies of mathematics teaching using programed materials. One is the educational technology study in which mathematics is used simply as a convenient subject matter vehicle for the study of a basic problem relating to the technology of self-instruction (e. g., Alter, 1962; Alter, Eigen and King, 1962; Angell and Lumsdaine, 1962; Beane, 1962; Brown, 1962; Cartwright, 1962; Dick, 1963; Eigen, 1963a; Eigen, 1963b; Feldhusen and Eigen, 1963; Frye, 1963; Glaser and Reynolds, 1962; Heimer, 1962; Joos, 1961; Levin and Baker, 1963; Popham, 1962; Price, 1962; Roe, 1962; Roe, Case and Roe, 1962; Shay, 1961; Silberman, et. al., 1962; Smith, 1962; Smith and Moore, 1962; Stolurow and Walker, 1963; Suppes and Ginsberg, 1962a; Uttal, 1961, 1962; Weiss, et. al., 1961).

The other is the substantive study which consists of studying mathematics in its own right as a conceptual, intellectual, and skill domain to determine how it is learned by students (e. g., Gagné, 1962; Gagné, 1963; Gagné and Brown, 1961; Gagné and Dick, 1961; Gagné, et. al., 1962; Hendrix, 1947; 1960; 1961; Jacobs and Smith, 1960).

Some research involves both of these objectives as it is most efficient to study the specific problems of learning subject matter while also considering a technological problem (e. g., Gagné and Paradise, 1961; Keislar, 1959; Rigney and Budnoff, 1962; Smith and Quackenbush, 1960; Stolurow and Beberman, 1964; Suppes, 1960, 1963; Suppes and Hill, 1962; Suppes and Ginsberg, 1962b; Vicory and Corrigan, 1963; Wolfe, 1963).

These two types of research, the technological and the substantive, are by no means equivalent in their development nor is the total pattern which they reveal the most logical from all points of view. For example, it could be argued that the technological research which deals with problems of behavioral synthesis or "shaping" should follow the substantive research concerned with the analysis of mathematical concepts and skills (e. g., Gagné and Paradise, 1961). From this point of view, the argument would be that it is necessary to know what behavior is to be synthesized before doing research on techniques for accomplishing different types of syntheses. The fact that the organization of the student's behavior, its shaping, is accomplished indirectly by means of stimulus control (e. g., presenting sets of stimulus materials in a particular sequence, Mattson, 1963), and that stimulus control also is used, at a different level of stimulus complexity, to analyze processes makes it possible to work on problems of analysis while also working on synthesis. For example, molar units of behavior (e. g., the solution of linear equations) can be synthesized by means of stimulus manipulations and different levels of stimulus complexity. This makes it possible to work on problems of behavioral synthesis while also developing concepts and techniques of behavioral analysis studies at a more molecular level (e. g., prompting and confirmation). Any survey of research needs to keep the different levels of analysis clearly in mind if the research findings are related to either theory or practice.

The Problem of Behavioral Units

The use of the terms analysis and synthesis of behavior immediately suggests a need for some specification of the units used. Defining a behavioral unit is an important unsolved problem that at the moment, must be considered on intuitive grounds (see Sheffield, 1961). There is little doubt, however, that there are molecular and molar elements of behavior since certain intact behaviors at one time were known to consist of more elementary behaviors. An integrated set of molecular units can be called a molar unit, e.g., a sentence is more molar than a word; furthermore, a word can be thought of as an integrated collection of letters. Also, an analysis of behavior in terms of observable gross movements required to produce a sentence is more molar than that required to write individual letters. Similarly, gross movements of the torso are more molar than specific arm muscle movements. Similarly, "learning sets" (Harlow, 1949) such as those contributing to the solution of equations ("simplifying an equation by adding and subtracting terms to both sides," Gagné and Paradise, 1961, p. 6) are more molar than learning that two plus two equals four or that subtracting five from eight leaves three.

Important in the analysis of behavior repertoires is the unit of analysis employed. Most current psychological theories of learning deal with behavioral units much more molecular than those of concern to the educator (see Stolurow, 1964). This difference in the units used to describe behavior probably accounts for some of the failure in communication between educator and psychologist, and the problem of developing behavioral units is critical to the technology of education. Needless to say, this problem also arises in the research on the teaching of mathematics. Unfortunately it is not one of the active research problems in the teaching of mathematics at the present time. The purpose in raising it here is that it is basic to the present treatment of the research problems relating to programed instruction and the teaching of mathematics.

The Problem of Language

One way in which the problem of units enters into decisions about research rather obviously is in the application of the language of stimulus and response to behaviors more molar than than those to which these terms are traditionally applied. This is the problem of the psychological language to be used in formulating problems relating to the teaching of mathematics. There are many different theories of learning that use S-R language and not all of them use the terms stimulus and response to refer to environmental events and behaviors which are at the same level of description. Guthrie, for example, used s and r to refer to more molecular events than those to which Skinner applied these same verbal labels (see Hilgard, 1956). With the application of S-R language to the more molar educationally relevant events and behaviors, an extended use is being made of these terms. Furthermore, their use in this context does not mean that a particular S-R theory of learning also is being extended. Rather, these terms are used to achieve greater objectivity in both communication and description so as to minimize surplus meaning and to permit operational descriptions of material and procedures without also implying in the language of description the effects these materials are intended to produce. We use the word sentence to refer to one word and to ordered sets exceeding twenty words.

The Problem of Objectives

In addition to problems of units and of language, there are problems of objectives. It seems useful to distinguish two types of studies relating to different but important objectives in the technology of programed instruction. One is concerned with <u>analysis</u> and has implications for the psychological architecture of associative structures designed for particular educational purposes. Here the concern is with studies using programed materials in which the objective of the research is to analyze complex behaviors in a way that contributes to the design and test of hypotheses concerning the elements of complex mathematical concepts and skills. The other is concerned with synthesis and the problems involved in developing associative structures. The former, dealing with the analysis of complex behavior, will be referred to as architectural studies since their implications relate most directly to the design of programs; the latter, dealing with the principles of synthesis or the rules of efficient assembly of complex structures, will be referred to as an engineering studies.

Studies with Engineering Implications

The research on programed self-instruction in mathematics has concerned itself primarily with the problems of the technology of teaching (Stolurow, 1961). However, the implementation, or engineering, problems predominate and comprise the bulk of the research although not its more exciting developments.

Response Form

A behavioral engineering problem is the determination of the form of the response to be used in a learning situation. This is the problem, for example, of deciding to use an overt or a covert response, an actual response or a symbolic response, etc. The response that the program requires the student to use is determined as a result of several considerations, the most important one of which is the objective of the instruction. Another is the presumed effect of the form of practice upon learning, immediate retention, or transfer. In any extension of learning theory into educational practice, it is necessary to examine alternative forms of response to determine their implications for learning, retention, and transfer effects both with respect to the specific content and to the information processing skills. It is necessary to begin by considering the form of desired response in relation to the student's repertoire at the beginning of practice. A response that is in the student's repertoire in the exact form required by the new learning experience has different implications for programing than a response that is not in a form that can be used directly. While the objectives define the terminal behavior and the initial repertoire defines the beginning behavior, the intervening behavior also needs to be specified. In building a behavioral structure, it is necesary to specify the form of the behavior at each stage of learning. If behavior is already in the form that can be used, then the engineering problem is one of putting it under specific stimulus control. In this case, one question that arises is whether or not the student should practice the response overtly. There seems to be less need to practice overtly when the response is already established (e. g., Walker and Stolurow, 1962). However, if the student's responses are not already in the desired form then the engineering problem is to assemble or shape the behavior which is available; in this case, the program should require an overt form of response. For example, if students do not know how to do long division, then each of the component skills of estimating, multiplying, and subtracting may need to be taught. However, only one component skill may be missing, in which case only it must be taught. To make these decisions, the task must be analyzed into components and their relationships specified. Once this is done in the development of instructional programs, one of the basic psychological problems that remains is to determine whether the need is to shape responses or to develop stimulus control over existent responses. At different stages of learning each is required. Once the response is available, it can be put under appropriate stimulus control.

Once the psychological problems have been identified, their engineering implications must be considered. The decision may be to use either overt or covert responses, or it may be to use the actual or symbolic form of response. One engineering consideration that arises immediately is the visibility of the response. If shaping is required, then the behavior needs to be visible. By making the student respond visibly, errors are readily seen and these data can be used in revising the program. Thus, shaping may be the factor that determines whether or not the student's response is to be overt.

Visibility of response is particularly important in the early stages of the development of a program. It also is important in any new use of an established program; for example, the use of one developed for the fourth grade either with a younger group of students or with learners who have different cultural backgrounds from those with and for whom the program was originally developed and validated. In this type of situation responsibility is needed to reveal whether the presumed repertoire actually exists.

Overt vs. Covert Response

The research on overt and covert response in programed instruction has indicated that when the objective is to achieve stimulus control of established responses then the use of covert responses results in equivalent achievement in less time than the use of overt responses (e. g., Lambert, Miller and Wiley, 1962; Walker and Stolurow, 1962).

Some findings suggest (Krumboltz and Weisman, 1962) that the use of overt responses may enhance retention, but these results are contradicted by others showing no such relationship (Tobias and Weiner, 1963). This discrepancy suggests the need for further analyses of the situations to determine the possible presence of another variable.

One function of overt response is to focus the learner's attention on particular instructional materials. In this usage overt responses are employed to make cues more salient, or, in other words, as a step in achieving stimulus control of responses. This may be a very important step to take if the environment is distracting; however, it appears to be much less important when the student is in an experimental setting or working with a self-instructional program and his attention is already concentrated on the task than it is when the student must read a book or listen to a lecture to learn. Nevertheless, it sometimes is important even in programed instruction although, with a program, attention is narrowed to a frame at a time. For example, in a frame of a program intending to teach the form of a binomial, the pattern of the cue-stimulus may be difficult to detect without focusing the learner's attention on it in the absence of other potentially distracting stimuli. Therefore, by requiring the student to respond overtly to specific features of a complex stimulus, it is possible that he will acquire a more efficient discrimination than he would if the program omitted steps that required him to do this. Here the response involved is an instrument for sharpening a discrimination, consequently, an established form of behavior, one already in the student's repertoire, is used. The next question is whether or not it is important to have the response visible when its purpose is to make features of the stimulus salient. In other words, theproblem that arises at this point can be thought of as a trade-off that can be made between time and visibility when the objective is to focus the learner's attention on critical cues. It may be more important to have students learn rapidly than it is to know what errors they make during learning, particularly if the error rate is likely to be low, e. g, $< 10\%$.

Using an overt response in working with a program (e. g., Lambert, Miller and Wiley, 1962; Stolurow and Walker, 1962) may not add to performance on tests taken upon its completion. However, the data pertaining to this point probably were not obtained when it could be assumed that the primary problem was the shaping of responses. It might be assumed that different results would be obtained from the use of overt responses if they were not already in the student's repertoire. The number of such studies and, consequently, the number of different areas of mathematics that have been studied is very small, and does not justify generalizations to all levels and topics in mathematics. The advantages that can accrue from making the response visible seem sufficient to warrant continued use of overt behavior in a mathematics program at the present time (Williams, 1963). Furthermore, while the required responses may be present in the learner's repertoire, they may not occur at short enough latency or be sufficiently interconnected into molar segments to meet terminal performance standards, in which case, overt practice of the required combinations would be indicated.

Constructed vs. Multiple Choice Response

The psychological issues that arise when a decision is made to use either constructed or multiple choice response are comparable to those associated with the overt-covert problem. They relate first to the availability of the responses in the learner's repertoire. If the response does exist in his repertoire, then the use of a multiple choice format permits the student to make his responses visibly and also quickly. Consequently, delays that would occur if the responses were constructed are eliminated with the possible advantage that the motivation of the students is higher than it would be under the constructed response condition.

Data from mathematics, unfortunately, are meager. Price (1962) compared multiple choice and constructed response modes using mentally retarded students. His results could be interpreted in terms of response availability, but not unambiguously. He found that the multiple choice mode resulted in superior performance when the students were taught subtraction, but not when they were taught addition. This raises several questions: "What is the critical response?" Is it writing or selecting the numeral, or is it the behavior that the learner engages in to determine the visible numeral? Presumably it is the latter for, in Price's study, the students constructed or selected numerals in both the addition and subtraction programs. Granting this, there is the related question, "Why were the requisite processing responses more available for learning subtraction with multiple choice responses?" A possible hypothesis that could account for his results is that having learned the skills of addition, the students could use those skills to answer the subtraction problems when they were presented in multiple choice form since they could convert the problems to addition and try each alternative as an addend. However, they were required to recall many specific answers to the problems presented in the constructed response form and had neither learned these nor the required skills well enough to do the subtraction.

Stimulus Encoding and Support

There are several problems relating to the presentation of mathematical concepts for efficient teaching. These are encoding problems; unfortunately, they have been given little attention. One of them, for example, is the representation of a variable. The use of "frames," "boxes," or "empty" geometric forms such as squares and circles is one apparently effective form of encoding to convey the idea of a variable. It appears to be a more effective way to represent variables than letters of the alphabet (Page 1961; 1962). One hypothesis is that empty "boxes" suggest the idea of a container that could have a variety of things placed in it; therefore an empty geometric form symbolizes a box or container and is an efficient way of encoding to symbolize a variable. In contrast, letters are treated as fixed entities. This is a transfer problem which is amenable to research; although no studies have been conducted; consequently, the geometric form of encoding variables as boxes is not known to be superior either for all students or for students at the lower ages. By writing the program with different mathematical symbols to teach the same concepts, it would be possible to compare specific forms of encoding; e.g., letters with "boxes". It would be possible to determine whether the learning of either the concept of a variable or the specific operations were facilitated by the use of empty geometric forms to symbolize them.

Another problem concerns the choice between algebraic and geometric presentations of a problem. In some, as yet unpublished, studies[2] for example, students were required to "discover" a formula that applied to some, but not all, features of a set of displays. Each display contained two points located in different positions within the first quadrant of a field. The intent was that the students would use an algebraic procedure to solve the problem, particularly those in groups given partial or full information since the information was so encoded when it was given; however, it was found that some students gave geometric solutions. This finding raised a question about the assumption that students would use an algebraic algorithm. To determine the implications of this another study was done in which some groups were deliberately given a geometric solution principle for the problem, others were given the algebraic solution principle, and both groups had to discover how to apply the principle to the same

---

[2] Stolurow, L. M. and McHale, T. J. Study of the Transfer Effects of Written Instructions to Task Performance and of Task Performance to Task Performance. In L. M. Stolurow (Principal Investigator) Psychological and Educational Factors in Transfer of Training. Urbana, Ill.: Univer. of Ill. Trng. Res. Lab., USOE Title VII Contr. #2-20-003, Quarterly Reports No. 7, January - March, 1964, Chapter 2.

displays. On the whole the group given the algebraic principle performed better than the group given the geometric principle, even though the two solution principles were sufficient for the purpose and theoretically equivalent in their effectiveness. This difference in actual effectiveness was probably due to the fact that the algebraic principle could be applied more rapidly in solving the problem encoded in the frame.

The data on the effects of different forms of encoding in mathematics are meager. Some ambiguous data relating to the problem of efficient encoding of mathematics concepts comes from a study by Hickey, et al (1962). These investigations failed to find the significant differences they expected in favor of a graphic encoding as contrasted with an algebraic encoding in teaching Boolean algebra. They suggest that different symbolism can make a difference in the rate of learning but this remains an hypothesis since they do not provide sufficient data. Data are almost non-existent for retention and transfer, consequently, the implications of encoding differences for these objectives are unknown.

Related to the encoding problem is that of stimulus support in the presentation of mathematics materials for learning. Support for a correct response can be provided by various prompting devices. In contrast, students can be given no prompts but when they respond correctly their behavior can be confirmed by a stimulus that would match their production. Prompting and confirmation procedures can be used either separately (pure form) or together (mixed strategy) in a program. Rigney and Budnoff (1962) used both pure prompting, pure confirmation procedures, and combinations of them in teaching Boolean algebra. They found that pure confirmation, the condition with least stimulus support, led to lower error scores in learning than did the mixture, a vanishing procedure, that consisted of prompting followed by confirmation. This finding held for both upper and lower intelligence groups. However, the reverse held for the middle intelligence group. This most unusual finding is puzzling and clearly indicates the need for replication.

Theory asserts that prompting, which maximizes stimulus support, is the preferred initial learning procedure, for prompting quickly raises the probability of the correct response. Once the response occurs with sufficient probability, it is possible to withdraw stimulus support (the prompt) so that confirmation, which minimized stimulus support, can be used efficiently. No theory suggests the reason why the middle ability group in the Rigney and Budnoff study would not respond in the same way as the extremes of ability. Their data appear to be due to chance.

Angell and Lumsdaine (1962) studied vanishing and used a mixture of prompting and confirmation procedures. They found that the mixture resulted in performance scores equivalent to those of a group for whom stimulus support was not withdrawn. Both the pure and mixed procedures were equivalent for learning. However, two weeks later the 5th and 6th graders trained with the vanishing (mixed) procedure achieved higher retention scores. Their results are, therefore, consistent with those of Stolurow (1963) and Stolurow and Lippert (1964) and with the theory described above, although the retention test was required to reveal the difference.

An important problem in many learning situations is the development of discriminations. This type of problem occurs whenever there are similar stimuli requiring distinctive responses. A student who adds instead of multiplying, for example, is failing to discriminate an operator sign. Conventions in the use of mathematical symbols suggest that the need to differentiate them arises with very young students; however, this problem has not been studied, nor have programs been written which suggest a recognition of it by requiring that the student's initial practice consist in discrimination training. While it would seem that as a problem it would be most critical with young children to introduce discrimination training early in their practice, especially if the students are just learning to read, this may not exhaust the set of problems. A program for secondary students, for example, also may effectively include stimulus discrimination sequences early in practice whenever new forms of expression are introduced, e. g., quadratic equations.

Feedback Characteristics

The requirements for optimum feedback in all complex learning situations are poorly understood and mathematics programs are no exception. If the particular events which follow response are considered as feedback then several potential dimensions of effect upon the student can be hypothesized. The most salient of these is the reinforcement effect. This means that certain events which follow response increase the likelihood that the student will make that response when the stimulus is presented again. Typically, reinforcement is predictable from a variety of classes of stimuli; one of these is called rewards, another provides information, still another appears to have motivational effects. Unfortunately, the relative importance of these aspects of feedback is unknown even in mathematics instruction. If it is assumed that any event following a response can have one or more of these implications, then, since each is potentially capable of independent manipulation as a variable, its effect on behavior could be determined so as to permit comparison. Since teachers differ greatly in the way they repond to the student's performance, it is also meaningful from a practical point of view to determine these separate effects of different feedback. For example, teachers do not always tell the student that he is correct. This form of teacher behavior is tantamount to using a partial reinforcement schedule.

Also, the language used to inform the student of the correctness of his responses may be designed to evaluate the quality of the student's performance, in which case it may function as a reward. Explorations could or could not be used so that learning experiences which would be the same otherwise could differ in the information provided the learner. There is still another type of stimulus contingency, based upon the nature of the learner's response, which would be designed to motivate the learner to try harder or less hard.

It is conceivable that with computer-assisted instruction (CAI) a program could be presented in which selective use could be made of each of these types of feedback and each could be provided as appropriate and important for optimum results with different learners. Alter, Eigen and King (1962) studied the effects of certain rewards on learning using a program with five and six year olds that taught numerals and the concepts of "oneness" to "nineness." With some students, they added trinkets to verbal knowledge of results; however, the trinkets produced no differences in the average level of student performance. While this suggests that the concrete rewards are not critical, even at this early age, the study would have been more informative if the psychological reward value of the trinkets to the learners had been determined in advance and those differing appreciably had been used to see if they made a difference in rate of learning. We do not know how much reward value the trinkets used held for the students. Furthermore, we do not know whether or not their use made the students work more rapidly and it is possible that the trinkets had a motivational effect which revealed itself in the student's rate of work but not in his accuracy.

Review Conditions

There are many different ways in which review can be accomplished. One of these was studied by Dick (1963a, 1963b) who used an algebra program with college students who worked alone or in pairs. In the latter condition, each student was allowed to discuss different points in the program with another student. A test of achievement did not reveal a significant mean difference between the two groups, but when the groups were retested one year later there was a significant difference between them with better retention shown by the paired group. This suggests that selective review can be an aid to retention.

Studies with Architectural Implications

The development of the psychological plan for the educational engineering of a self-instructional program has two aspects. One is the delineation of the various associative structures that properly relate elements of knowledge of the subject matter to one another as cues and responses in a functional manner representing terminal performance. The other is the delineation of strategies.

Gagné and Paradise (1961) provide a key to the analysis necessary for the identification of hierarchical associative structures, or "learning sets." Their key is the question, "What would the individual have to know how to do in order to perform this task after being given only instructions?" By asking this question, each learning set can be specified; by repeating the question and addressing it to the previous response, every subordinate level is described down to the simplest, and lowest learning sets. A possible "hierarchy of knowledge" becomes explicit by this process.

Two problems emerge with this approach. One is that this is not a sufficient procedure for generating an instructional program, since crucial objectives other than those pertaining to the relationships among the elements of knowledge also are to be accomplished. For instance, "cognitive styles" or strategies of search, information processing, and cue selection also are sought as program objectives. To secure comparable information on the structure of the strategies, a different question must be asked. It is concerned with the procedures, methods, and techniques to be used to acquire the information or to accomplish the objective. For example, the student must learn to read English from left to right or to locate properly the product of two numbers in accomplishing long division. Consequently, in analyzing the task, the question is "What must the learner do in order to perform this task?" We can think of the answers to this question as a set of operations which the learner must perform. Computational skills in particular areas of mathematics represent the processing aspects of mathematics as distinct from the conceptual aspects. The other problem concerns the method of teaching to use when the material does have an hierarchical architecture. Arguing from the physical architecture it would seem necessary to consolidate the learning of the lower order elements before teaching the higher order elements. However, Merrill's (1964) data do not indicate that this is necessary. This finding raises questions about the analysis provided by Gagné (1962).

Associative Structures

The hierarchical structures of knowledge identified by Gagné and his collegues (Gagné, 1962; Gagné and Dick, 1961; Gagné and Dick, 1962; Gagné, Mayor, Garstens and Paradise, 1962; Gagné and Paradise, 1961) represent associative structures which presumably depend upon positive transfer for their efficient formation. The units from which the hierarchical structures are built are more molar than those typically studied in the learning laboratory. For example, at level V (Gagné and

Paradise, 1961) symbol recognition is a class of behaviors, not a single stimulus response connection. It is a "learning set" which means that there is a common principle involved in relating the student's responses to each of the exemplars of a class of stimuli. Gagné has suggested that the basic level to be identified in a hierarchy is specified by pure factor tests. These, then, are alternative ways of specifying the elements of the more molar associative structures involved in learning mathematics.

Chapter 2

Summary of Research

Principles Useful in Programing Mathematics

The primary objective was to conduct research on principles in programing within the context of mathematics. Individual studies were designed to determine answers to specific questions concerning the technology of programing. Separate but related hypotheses were tested, each of which dealt with a particular problem of programing relating to either step form or step sequence.

Linear vs. branching. The objective of the first series of studies was to compare the linear and branching step forms. Comparisons were made in terms of performance (errors and time scores) not only on the learning task itself but also upon a test of achievement and attitude about the forms (e. g., see Beane, 1965).

Programing for discovery learning. The second series of experiments had as its objective the study of techniques for implementing "discovery learning." Two approaches were used to study this problem. One approach used students who had been taught by the discovery method and allowed them to continue their studies of mathematics via a self-instructional program. Two groups were formed; one was taught by means of an expository programed sequence, the other by means of a discovery sequence. Measures were then taken on an achievement and transfer test. The second method involved three treatments. Subjects were introduced to the topic in one of three ways: (1) they were allowed to work with the stimulus elements; (2) they were introduced to the mediating principles; (3) they were allowed to work with the response terms. Measures were taken of performance on a test after the introductory material and after the task material.

Web of association. The third series of experiments had as its objective the study of techniques for building a "web of associations." The hypothesis was that learning is faster with a systematic presentation of material than with a nonsystemic ordering. Performance was measured via achievement tests following presentation of the materials.

Step size. A fourth series of experiments had as its objective the study of problems associated with step size. One aspect dealt with was an attempt to relate physical and relational properties of steps to empirical difficulty (e. g., percentage of errors). An attempt was made to find a judgmental *a priori* method of determining step size.

Covert vs. overt responding. A fifth series of experiments had as objective the study of covert and overt response sequences. The hypothesis was that the cover-overt response sequence is superior to the overt-covert. This was abandoned after pilot work suggested this was not so.

Program vs. teacher. A sixth series of experiments was concerned with the comparison of the different modes of presentation of materials (programed text or teacher) and the possible combinations of these two methods of presentation.

Ability correlates of mathematical achievement. A seventh series of experiments, and, in fact a variable controlled in all of these studies, was aimed at the relationships between mental abilities and performance and an attempt to find an underlying associational structure.

Methodological studies. An eighth series of experiments were aimed at various methodological problems such as (1) a most efficient method for sampling data; (2) a most efficient method for analyzing this sample; (3) the optimal sample size needed for accurate revision of a programed text.

Summary

Each of these series of experiments is described briefly in the following sections. The last section will attempt to summarize that which has been set forth into a coherent whole.

The First Series: Branching vs. Linear Form

To get at the question of the comparison of form, a series of programed plane geometry texts were developed. Two forms were prepared: (1) a linear form; and (2) a branching form. The design of the experiment (Beane, 1962, 1965) involved the use of repeated measures. One group was given conventional instruction. A second group was given the first half in the branch form and the second half in the linear program form. A third group had the first half in the linear program form and the second half of the branch program form. A fourth group had only the branch and a fifth group had only the linear program.

An achievement test was given before and after the experimental period and seven weeks after the end of the experiment. Prior to the experiment, a test of mental maturity was given. On the basis of this test, high and low IQ subgroups of the five groups were formed. The experimental groups were also given an attitude questionnaire at the halfway point, at the end of the experimental period, and seven weeks after the end of the experiment.

No differences were found between the groups on achievement at the end of the experiment or seven weeks thereafter. However, the high ability subgroups had significantly higher scores on these two tests than the low ability subgroups. There was a significant difference in time spent studying the material. The branch-branch and branch-linear groups spent less time than the other three groups.

Attitudes toward programed instruction. One of the most interesting outcomes of this study had to do with the attitudes toward programed instruction. At the first administration, all groups preferred programed instruction to conventional. This same attitude prevailed at the end of the experiment. But seven weeks later the students expressed a neutral attitude. The attitudes of the four experimental groups did not differ significantly nor did the attitudes of the ability subgroups differ significantly. Furthermore, the linear-branch and branch-linear groups both expressed a preference for the linear program. Both of the high ability subgroups of these groups expressed a similar preference.

Thus, it appears that the linear form is superior to the branch form in the sense that students prefer it. The branch form did not improve performance although it seems to have led to a savings in time. Furthermore, students apparently prefer programed instruction to classroom instruction only when they are actively involved in programed materials. This, however, may have been an artifact due to the brevity of this experiment (two weeks).

The Second Series: Programing for Discovery Learning

Non-Specific Transfer

The first approach to the use of the discovery method (Wolfe, 1963) attacked the problem of whether or not the student's ability to take advantage of the expository method is lessened later on. Two sets of programed materials were developed for this study. The two sets differed only in that one stressed the expository and one the discovery method of presentation. The students were matched on the basis of ability measures. One number of the matched pair was given the discovery sequence, the other the expository. All students had studied for at least a year previously using the discovery method. An achievement and transfer test was administered after all students had completed their sequence. No significant differences in learning were found to result from the discovery or expository treatment within the ability levels for either ninth or tenth graders. There was no interaction between ability level and treatment in either grade.

The second series (Stolurow, Rosen, Frincke, Batchelor, Himmel, 1964) compared methods used to implement discovery learning. Three sets of programed materials were developed for use in this study. The books covered the topic of grouping conventions. One group of subjects was introduced to the topic by means of presentation of stimulus elements, another by means of the response elements, and the third by means of the stimulus-response linkages or mediators (rules for grouping). Then all three groups were given another book which presented the whole task: the conventions for grouping and simplifying. Achievement tests were given following the first book and following the second one too. In a small pilot try-out no significant differences in performance as measured by the end tests or by amount of time spent in learning were found between the groups. These groups were equated for prior performance.

Later this study was run again using a larger sample size (Rosen, Frincke, Stolurow, 1964). Again, none of these treatments produced any difference in performance on the end tests. This time students were given the actual learning task in the conventional classroom manner, however. The amount of time spent on the programed segment again was not significantly different.

Thus, in conclusion, it appears that use of the discovery method in teaching does not impair the students' ability to benefit from the expository method. Also, prior practice with the mediating elements or principles does not aid the learning of the task any more than familiarization with the stimulus or response element of the task does. Any or all of these predifferentiation procedures can aid transfer.

The Third Series: Ordering of Frames

This series of experiments was aimed at the comparison of a systematic and a nonsystematic ordering of frames within a sequence.

Two forms of a segment of the beginning unit of the UICSM program were developed. One form, the systematic form, was the original segment unchanged. The revised form was a scrambled sequence of the same frames. In the summer pilot tryout (Stolurow, Rosen, Frincke, Batchelor & Himmel, 1964) neither group differed significantly in terms of end test score or average time spent per page. In the latter, larger sample size, replication (Rosen, Frincke, & Stolurow, 1964) the scrambled sequence produced significantly ($\alpha = .05$) better performance on the end test. However, there was no significant difference in mean time per page.

Another approach used to attack this problem employed a series of programed texts which covered the topic of fractions. Four groups were used. One group used a conventional textbook and three other groups used the programed materials. These three different sequences presented two different orderings of the materials. Sequence A maintained the sequence used in the textbook. Sequences B and C rearranged these segments so that the fractions were presented in numerical order. An end test was given to evaluate final student performance. There were two main types of questions asked: inferential and definitive types of questions. A pre-instruction performance test was also given. The residual scores between the first and second administration were then computed. In general, there appeared to be a difference between the testbook groups and the program sequence B group. There was also a difference between sequence B and sequence C, and between sequence A and B. Sequence C differed from B only in that the direction of the associations were reversed. Thus, the order of presentation of concepts does indeed effect final performance (Smith, 1962).

The Fourth Series: Step Size

There is a great deal of talk about large and small step size but little or no attempt has been made to quantify the concept. Even less work has been done to determine an _a priori_ measure. One possible _a posteriori_ measure of step size is empirical difficulty. A possible _a priori_ measure is achieved through judgemental ratings. These two techniques were employed and their interrelationships studied in this experimental sequence.

Two forms of a segment of the programed text series were developed. One had fewer frames than the other. This was designated the large step size version and the untouched version the small step size version. In the pilot tryout, students were put into groups such that the mean achievement of the two groups was about the same. An end-test was administered after the student completed his assigned book. The students did not differ in performance on the end-test but the group with the small step size took significantly longer to finisn the programed segment. However, the average time per page was approximately the same for both groups. The difficulties and judgemental ratings of complexity were intercorrelated for both; however, the correlations were not high and accounted for very little of the variance (Stolurow, Rosen, etc., 1964).

A similar study was conducted with a larger sample size. The design of the experiment crossed the size of step version with the mode of presentation in a 3 x 2 analysis of variance framework. The judgemental ratings used were the same as those obtained and used in the pilot tryout. Again, neither factor affected performance on the end test and the large step size group took less time to complete their version than the small step size group.

The Fifth Series: Covert vs. Overt Responding

The purpose of this study was to test the hypothesis that a covert response to a stimulus produces an S-R pairing which is more amenable to change than does an overt response. Thus, students who read and respond covertly to a program and then read and respond overtly should achieve better performance on a final end-test since the incorrect S-R pairs which are formed on the initial run through should extinguish more easily for these students than for students who cover the material in the reverse manner.

To test this hypothesis the first book of the UICSM Unit I programed series (the revised version) was used. This version covered the distinction between number and numeral. The presentation was linear in form and in what is commonly called the "zebra" format. Performance measures were taken by means of a test at the end of each run-through. There were no significant differences between the two groups on either test. Furthermore, there was no significant time savings on the overt runthrough after the covert reading. That is to say, the overt-covert group spent as much time on the overt runthrough as the covert-overt group spent on the overt run-through (Stolurow, Rosen, Frincke, Batchelor, & Himmel, 1964). After this pilot tryout of this experiment failed, this line of research was abandoned as unsuccessful.

Another approach to covert responding is to consider the conventional classroom technique. While the teacher is lecturing the responding that the student may do is generally covert. Thus, the conventional mode of teaching may be considered as covert responding on the part of the student. From this point of view, the comparison of covert and overt responding is involved in the comparison of programed instruction only and conventional instruction only. In general, in these studies programed and conventional instruction only are shown to lead to equivalent performance. This comparison is discussed further, with the presentation of results of studies bearing upon the comparison, in the sixth and eighth series of experiments.

The Sixth Series: Conventional vs. Programed Instruction

One of the underlying currents of this entire sequence of experiments has been the comparison of the efficacy of conventional and programed instructional modes. The first attempt at this comparison was made using a series of programed texts which covered Unit I of the UICSM series. Part tests were administered after each segment of the sequence. Although the two modes of presentation led to about the same mean performance on the tests, the programed instruction maintained a far more homogeneous level of performance throughout the course of instruction (Brown, 1962).

This series of programed texts was then revised and a new problem was attacked. The question under investigation was whether or not some teacher-program combination might lead to improved performance. Three modes of presentation were used:

1) "pure" condition: programed materials and no conventional classroom instruction;
2) "anticipating" condition: programed materials preceded the usual classroom development of topics;

3) "control" condition: no programed materials and the usual classroom development of topics.

The results of this experiment led to the conclusions that (1) the programed texts, when used as the sole agent of instruction, did not teach as effectively as the trained teacher alone; (2) the trained teachers using the UICSM programed texts to precede their lectures did not teach more effectively than the trained control teachers; and (3) the trained UICSM teachers using the programed texts teach more effectively than the programed text used as the sole agent of instruction. These comparisons were made on pairs of students matched for initial ability (Brown, 1964).

A similar type of study was performed using a segment of the total sequence as the material to be covered. A small pilot study was performed in which two conditions were used:

1) "follow" condition: program followed lecture; and

2) "lead" condition: program preceded lecture.

The two groups were equated for initial performance. The two groups did not differ significantly in terms of end-test performance on mean time spent on the program (Stolurow, Rosen, Frincke, Batchelor, & Himmel, 1964).

A follow-up series of experiments was performed (Rosen, Frincke, and Stolurow, 1964) on this same topic. In this larger study, a program only condition was used as well. Using the same materials, but different teachers, no effect was found for presentation mode in terms of end-test score or time spent on the program. However, on one other part, presentation mode had an effect upon performance; the lead x mediator practice resulted in the best performance. However, presentation mode led to no time savings in any conditions.

The Seventh Series: Ability and Learning

One way to get at the primary mental abilities necessary for successful learning in the educational situation is via the correlational method. Eight classes of students were given an introduction to fractions either via the conventional classroom approach or else via programed instruction. Prior to the educational experience the students were given a battery of tests to determine their "mental characteristics." They were also given a performance test (a) prior to the experience; (b) immediately after, and (c) six weeks later. The ability and achievement scores were then intercorrelated.

Common sense suggests that learning is a "snowballing process." That is, that later learning depends greatly on prior learning. If this is true, and if it is also true that things learned just recently are more important (or vice versa), then one ought to find that the correlational matrices of the end-tests form a simplex. This was not found in data gathered on the programed version of the UICSM Unit I texts (Rosen and Stolurow, 1964, Tech. Rep. #5). Since this hypothesis was disconfirmed, a further investigation was performed. A principal axis factor analysis was run on data gathered on an earlier version of the same program as well as on the data gathered from the revised form. The results of the analysis suggested that there are two factors underlying the learning which went on in this situation. One appeared to be a general reasoning factor and the other perhaps a rote memory factor (Rosen and Stolurow, 1964, Tech. Report #6).

The Eighth Series: Methodological Studies

In developing and analyzing programed materials, several methodological problems arise in the process of interpreting the data gathered. One problem is finding the number of students necessary to decide whether to revise, delete, or leave unchanged, any item in a programed text. To evaluate this problem the data from a portion of the UICSM programed sequence was employed. For the frames in this segment, error rates were determined for each of the items on the basis of several samples taken from the total student response population. The samples taken ranged in size from N = 1 to N = 100; there were 21 independent samples of seven different sizes (three per size). The results of the study demonstrated the hazards of making decisions on the basis of small sample sizes (up to N = 15). "Wide variations in efficiency among samples of a given size were observed both in terms of (a) rejection of acceptable frames, and (b) failing to reject unacceptable ones." Coupled with the inconsistency of small pretesting samples is the high frequency of erroneous rejections. It was recommended that pretest samples be both as large as practical and chosen so that the product of the desired rejection criterion and the sample size are integers, so as to maximize the power of the test. (Frincke and Stolurow, 1965).

Another problem often encountered is how to sample from the data collected in such a way as to produce a "happy balance" between accuracy and cost. What is needed is a means of estimating the population value from a smaller sample. The use of the upper and lower 27% provides such an estimate. This procedure and its rationale has been discussed elsewhere (Rosen and Stolurow, 1964).

A concommitant problem is how to process these responses most efficiently and economically. Three methods for processing were performed. The student worksheet data was directly punched into

IBM cards (wrong, right, omit) or recorded on a condensed tabulation sheet (a pictorial display of student errors and omits), or else the responses (wrong, right, omit) were recorded on SCRIBE answer sheets (Educational Testing Service). It was decided that the most efficient and productive method was to punch the IBM cards directly (Frincke and Stolurow, 1964).

Chapter 3

Administrative Information

Completed Technical Reports

Technical Report No. 1: A comparison of linear and branching techniques of programed instruction in plane geometry.
Donald G. Beane
July, 1963

Technical Report No. 2: Programed learning in elementary school: an experimental study of relationships between mental abilities and performance.
Leone Mary Smith
August, 1962

Technical Report No. 3: A comparison of test scores of students using programed instructional materials with those of students not using programed instructional materials.
O. Robert Brown, Jr.
July, 1963

Technical Report No. 4: Effects of expository instruction in mathematics on students accustomed to discovery methods.
Martin Sylvester Wolfe
January, 1963

Technical Report No. 5: In pursuit of a simplex.
Ellen F. Rosen
July, 1964

Technical Report No. 6: Analysis of structure underlying a self-instructional programed sequence in high school mathematics.
Ellen F. Rosen and Lawrence M. Stolurow
July, 1964

Completed Technical Reports (continued)

Technical Report No. 7:	Three methods of processing student response data in programed instruction. Gerald L. Frincke and Lawrence M. Stolurow July, 1964
Technical Report No. 8:	Description of variables and their implementation in studies of the principles of programing. Ellen F. Rosen and Lawrence M. Stolurow July, 1964
Technical Report No. 9:	Experimental studies of principles of programing. Lawrence M. Stolurow, Ellen F. Rosen, Gerald L. Frincke, David Batchelor, and Clark Hummel July, 1964
Technical Report No. 10:	A method for estimation of difficulty and discrimination indices in programed learning. Ellen F. Rosen July, 1964
Technical Report No. 11:	A sampling study. Gerald L. Frincke July, 1964
Technical Report No. 12:	Analysis of achievement by students using programed instruction or classroom teacher or both. O. Robert Brown, Jr. July, 1964
Technical Report No. 13:	An attempt to find an *a priori* measure of step size. Ellen F. Rosen and Lawrence M. Stolurow July, 1964
Technical Report No. 14:	Coding and flow charting as applied to UICSM programed texts (1962-1963 versions). Lawrence M. Stolurow, Ellen F. Rosen and Clark Himmel July, 1964

Summary of Project Personnel (June 1961-June 1964)

Almy, C. V.	June 1962 to Sept. 1962
Anderson, Bonnie	June 1963 to Sept. 1963
Anderson, Valerie	April 1962, May 1962 to July 1962
Attebury, F.	Oct. 1962 to Nov. 1962
Bakstansky, P.	Sept. 1961 to Dec. 1961
Baldwin, Elizabeth A.	Nov. 1961
Batchelor, David K.	June 1963 to Aug. 1963
Beberman, Dr. M.	July 1961 to Jan. 1964
Berkson, William K.	May 1962
Bird, William G.	Dec. 1962, Jan. 1963, Sept. 1963 to Nov. 1963
Blair, Gowin B.	Dec. 1962, Jan. 1963 to Feb. 1963
Blecher, Freda	Oct. 1962 to Sept. 1963
Bluhm, Eugene L.	July 1962 to Sept. 1962
Bonnell, Jane A.	Nov. 1962 to Dec. 1962, Feb. 1963
Brill, John B.	Sept. 1962 to Dec. 1962, Jan. 1963 to Nov. 1963
Brown, O. R., Jr.	June 1961 to Dec. 1964
Bucher, Mary A.	Dec. 1961 to Feb. 1962
Campbell, Wm. B.	Nov. 1961
Cutler, P.	Nov. 1962 to June 1963
Davidson, Allen L.	Sept. 1961 to Oct. 1961
Day, George H.	June 1962 to Sept. 1962
Day, Susan E.	May 1962 to Sept. 1962
Des Champs, Francois	Aug. 1962 to Sept. 1962
Driscoll, Ellen A.	Feb. 1963 to March 1963
Emami, Shahla	Feb. 1966 to Nov. 1966
Fejfar, J.	May 1962 to Aug. 1962
Frericks, Karen L.	Dec. 1962, Jan. 1963 to April 1963
Frincke, G.	Sept. 1962 to Aug. 1963
Fritz, Donald E.	Sept. 1961 to Oct. 1961 Feb. 1962 to May 1962 Sept. 1962 to Dec. 1962
Golden, W.	June 1961, Aug. 1961, June 1962 to Aug. 1962
Griffin, Ray A.	Sept. 1961 to Oct. 1961
Grubb, Gaila M.	Nov. 1961 to Feb. 1962
Hart, Bennett R.	Sept. 1961 to Oct. 1961
Hart, Robert S.	May 1962 to July 1962
Heins, Samuel O.	July 1962 to Aug. 1962

Summary of Project Personnel (continued)

Herman, T. L.	Dec. 1961
Holmes, A. H.	Aug. 1961 to Sept. 1961
Kansky, R. J.	Aug. 1962
May, R.	Sept. 1961 to April 1963
McCoy, E. M.	Sept. 1961 to June 1962
	Sept. 1962 to Oct. 1962
McHale, T. J.	Jan. 1962 to Feb. 1962
Pahel, Kenneth	June 1962
Perkins, Margot H.	April 1962
Podlogar, M.	Nov. 1961 to Aug. 1962
Rehwald, Rae A.	Oct. 1961 to Feb. 1962
Rosen, E. F.	Oct. 1962 to Aug. 1963
Schwartz, S.	Sept. 1961 to Oct. 1961
	Jan. 1962 to March 1962
Shah, C.	March 1963 to April 1963
Shah, S.	July 1961 to Aug. 1961
Sherman, E. R.	June 1962 to Aug. 1962
Skulte, Carol H.	June 1962 to Aug. 1962
Skulte, Robert J.	June 1962 to Aug. 1962
Slater, Donald C.	June 1962 to Aug. 1962
Snyder, John J.	Dec. 1962
Stone, Marcia R.	Nov. 1962 to March 1963
Stenson, H.	June 1962
Strachan, D.	May 1963 to Dec. 1963
Sturmthal, Susan M.	July 1962
Suh, D. K.	June 1962
Taylor, R.	May 1962 to Oct. 1962
Tremblay, C. W.	Aug. 1962 to Dec. 1963
Vander Linden, Richard	July 1961 to June 1963
Van Horn, C. J.	Oct. 1961 to Jan. 1962
Vaughn, V.	Oct. 1962 to Oct. 1963
Viemont, Ronald J.	Oct. 1961 to Dec. 1961
	Feb. 1962 to March 1963
Wadsworth, Adrian R.	Feb. 1962 to Aug. 1962
Weitzel, W. F.	Dec. 1961 to June 1962
Wills, H.	Sept. 1961 to Aug. 1963
Williamson, Alvia L.	July 1962 to Aug. 1962
	Nov. 1962 to March 1963
Wilson, M.	Aug. 1961 to Oct. 1962