Some thoughts on the psychology of mathematics learning

Research in PSYCHOMATHEMATICS
(Research in the Emerging Discipline of Psychomathematics)*

By JOSEPH M. SCANDURA
University of Pennsylvania
Philadelphia, Pennsylvania

DOING basic research on mathematics learning is a risky business. In the first place, it is hard to know whether one is asking the right questions, and, in the second place, it has been difficult to formulate significant questions involving mathematics learning in a researchable form. One of the major reasons for this state of affairs has been the lack of any suitable theoretical superstructure from which to work. The purpose of this paper is to outline some of the theoretical work under way at the University of Pennsylvania and to show how this work has helped to improve our understanding of how people learn mathematics.

Background

Theoretical development in educational psychology has been extremely slow. One major reason has been the typically imprecise definition of independent and dependent variables in research on meaningful learning and teaching. Such research can bear only an ambiguous relationship to theory. Similarities and essential differences often go undetected. Stating research objectives and defining variables in unambiguous terms, however, is not sufficient. The teaching-learning process has all too frequently been studied in terms of administrative variables such as class size, grade level, IQ, and teaching experience. To be theoretically relevant, the variables chosen must have broad explanatory potential. They should not merely be symptomatic of, and inextricably related to, the question at hand. Theory development depends on much more than mere fact-finding.

To provide a substantive base for their research, educational psychologists have frequently resorted to the languages, paradigms, and theories of the mother science of psychology. Mediation elaborations and operant conditioning paradigms of the S-R (stimulus-response) language and more general (but less well specified) cognitive theories have been popular.

Each approach has important limitations.1 From one point of view, parsimony

---

* The author thanks Joan Barksdale for her assistance in the preparation of this article. Her participation was made possible by a graduate-research training fellowship supported by the U. S. Office of Education.

suggests that the properties of overt S-R associations should also be attributed to mediational links. Yet practice has shown that mediational interpretations become increasingly cumbersome and less precise as situations become more complex. Similar difficulties have plagued researchers who have used operant techniques to study meaningful verbal learning. The results simply are nowhere nearly so clear in complex human learning as they are in the “Skinner Box.” It is increasingly recognized, for example, that knowledge of results is not directly analogous to feeding a pigeon and that in any case other factors, such as subject-matter structure, are probably of greater importance in promoting efficient learning. A general limitation of cognitive theories is their relative imprecision. Typically, “cognitions” are either not clearly specified in observable terms or are only partially defined. Under these conditions it has been impossible to construct a predictive theory—the sort of theory needed if practical implications are to be obtained.

In short, the choice to date has been between a precise, but seemingly inappropriate, S-R language and presumably more relevant cognitive formulations which leave much to be desired insofar as scientific cohesiveness and rigor are concerned.

In an attempt to overcome these difficulties, I have proposed a new scientific language for formulating research questions on meaningful learning. Because the language is framed in terms of the mathematical notions of sets and functions, the name “Set-Function Language” (SFL) was adopted.

Before outlining the SFL, let us briefly consider the S-R mediation language. In S-R psychology, the basic building block is the association, a construct which was abstracted directly from observed connections between overt stimuli and overt responses.

Learning a concept, presumably a more complex form of learning, involves the ability to give a common response to any one of a set of stimuli. To say that a subject has acquired the concept of “red,” for example, implies that he is able to give some common response, when shown any red-stimulus object, but will know not to give this response to any nonred stimulus. Similarly, a child may be said to have acquired the concept of “four” if he can say “four” when presented with any conglomeration of four objects but will not say “four” to any conglomeration not containing four objects—assuming, of course, that the child is operating under the same set of instructions in each case. In short, whereas an association pairs one stimulus with one response, a concept is a many-to-one relationship.

S-R theorists have felt obliged to represent the many-to-one concept relationship as a composite of one-to-one associations, as shown below.

\[
\begin{align*}
S_1 & \rightarrow M_{rs} \rightarrow R \\
S_2 & \\
S_3 &
\end{align*}
\]

Notice that the stimuli \( S_1, S_2, \) and \( S_3 \) are connected to the mediating response \( M_{rs} \) whose stimulus properties, in turn, elicit the observable response \( R \).

The Basic Unit in Meaningful Learning

Most subject-matter learning involves neither associations nor concepts but, as they have been variously called by different investigators, “rules,” “principles,” “schemata,” “heuristics,” and “TOTE units.” This is even more true of mathe-

\[1 \text{ See Footnote 1. The descriptor "set-function" should not be confused with set functions.} \]
matics learning than of learning in many other fields. To be more specific: Meaningful learning implies the ability to give the appropriate response in a class of functionally distinct responses to any stimulus in a class of functionally distinct stimuli. Unfortunately, this fact has often been overlooked because the term "concept" has been used so widely in discussing subject matters. When we say that a child "has the concept of addition," for example, what we frequently mean is that he can give the appropriate sum when presented with any pair of numbers. Put another way, the learning involved connects a large class of stimuli with a large class of responses. By definition, a concept connects a class of stimuli to exactly one response.

To see what is involved in meaningful learning, learn the following S-R pairs (i.e., overt inputs and outputs): (4 3 1) → 3, (8 1 6) → 2, (7 9 2) → 5, and (9 5 1) → 8. Now, on the basis of what you have just learned, give what you think should be the responses to the stimuli (7 2 1) and (4 7 2). Did you give the responses 6 and 2? If so, you probably acquired a unit of knowledge (i.e., rule) which might be stated, "Subtract the number in the third position from that in the first." If you did not give these responses, you presumably learned the original pairs as discrete entities—i.e., as distinct associations without noticing any relationships between them. (Of course, having no suggestion that there was any such relationship, you may not have been looking for one.) In this case, the rule-governed responses, 6 and 2, would be expected on the basis of chance alone.

Let us emphasize that this situation was not picked arbitrarily simply to embarrass S-R psychologists. Whereas a "patch job" can be done with certain special cases, to date no satisfactory way of representing rules et al. solely in terms of associations has been found. Stimulus properties which uniquely determine the responses (e.g., the first and third positions) and combining operations or transformations by which the responses are derived from the determining stimulus properties (e.g., subtraction) appear to be crucial aspects of all rules. While stimulus properties and derived responses play a central role in S-R mediation theory, there is no counterpart for the transform or combining operation.

The Set-Function Language (SFL)

Fortunately, all of these characteristics play a central role in the SFL. In fact, it is proposed that four characteristics are needed to specify a principle. Three of these characteristics specify a rule, and a fourth determines when the rule is to be applied. Those stimulus properties which determine (D) the corresponding responses constitute one such characteristic, the covert responses or derived-stimulus properties (R) are another, and the transform or combining operation (0) by which these covert responses are derived from the determining properties is the third. The fourth consists of those (usually higher order) contextual properties which identify (I) the rule to be applied. For example, the rule, $n^2$, for summing number series, where $n$ is the number of terms, applies only to those series which consist of the consecutive odd integers beginning with 1 (e.g., $1 + 3 + 5 + 7 = 4^2 = 16$).

Since certain simple principles, as well as associations and concepts, can be represented in the S-R mediation language, it is appropriate to ask whether these notions can also be represented in the SFL, in which rules and principles are taken to be basic. This is indeed the case. Furthermore, unlike the S-R language, explicit distinctions are made between (1) the observable S-R instances of a rule—the denotation, (2) the rule or principle itself,
that which underlies the behavior and whose presence can be inferred only indirectly, and (3) statements of the rule or principle in symbolic form.

The denotation is simply a function, a (mathematical) notion which may be defined as a set of ordered stimulus-response pairs such that to each stimulus there is one corresponding response. The denotation of a concept is simply represented as a constant function in which there is one functionally distinct response common to all stimuli. To represent an ordinary association, the set is further restricted so as to include only one S-R pair.

The rule construct is characterized as an ordered triple \((D, O, R)\), where \(D\), \(O\), and \(R\) are as defined above. Notice that \(D\) can be viewed as the domain of a function, \(R\) as the range, and \(O\) as a rule connecting them. Principles, of course, are ordered four-tuples \((I, D, O, R)\). In the case of concepts and associations, there are certain relationships among these characteristics\(^7\) but they need not concern us here. In stating rules or principles, primes may be used to distinguish signs (i.e., \(I', D', O', \text{and } R'\)) used to represent the constructs \(I\), \(D\), \(O\), and \(R\) from the constructs themselves.

The one point to emphasize is that even S-R theorists have been forced to adopt the idea of a transformation or combining operation in order to represent meaningful learning and thinking. Berlyne has done this, for example, in his fine book on thinking.\(^8\) In a recent article on the topic of principle learning, Gagné\(^9\) has adopted this point of view as well. Even many long-time mediation enthusiasts have recently become convinced that mediation theory is inappropriate for dealing with verbal learning.\(^10\)

In so doing, these theorists are in fact denying the primacy of the association as a construct. The idea of a transformation or combining operation takes over this primary role. We might note parenthetically that what here have been called response-determining properties of stimuli, S-R theorists have called functional stimuli. Similarly, our covert responses (or, derived stimulus properties) correspond to functional responses. What the combining operation does is to make explicit how the mediating responses are determined from the antecedent mediating stimuli. While this process may not be important in many forms of simple learning, we have just seen here how it becomes crucial in rule learning. The transformation, mapping, combining operation (whatever term is used) becomes of central concern, and to still call any suitable representation associationistic or even neoassociationistic is stretching the term beyond its reasonable limits.

In the final analysis, of course, the choice between scientific languages (and theories) involves efficiency and cohesiveness as well as the sheer ability to represent or account for observable phenomena. It is in this sense that the much-heralded adaptive quality of the S-R language too often has led psychologists to overlook the fact that it is always possible to patch up an existing formulation to meet new situations. Parsimony does not simply refer to the maintenance of existing concepts but to the formulation of emerging structures in the simplest possible way. What is presently being done with the S-R language (e.g., the inclusion of associative structures, reference mechanisms, etc.) is analogous to what pre-Copernican astronomers were doing when they invented ad hoc mechanisms to explain planetary motions in an attempt to salvage geocentric theory.

### Empirical Research Based on the SFL

Now consider two problems which have long plagued mathematics educators, rule

---

\(^7\) For more details see Mathematical Formulation (Footnote 1).


\(^10\) Personal communication.
generality and discovery learning: Let us attempt at least partial resolutions of these two problems and, in the process, indicate how the SFL helped to achieve these resolutions by making it possible to formulate the underlying questions in a precise way.

Rule generality

In instructional situations, the question often arises as to how general the presentation of material ought to be. Subject-matter specialists and most educators tend to emphasize that the more general the presentation, the more useful it will be. Learning-oriented experimental psychologists, on the other hand, are often inclined to point out that the more specific the presentation, the better the learning. The question of generality arises often in mathematics instruction. Should addition and subtraction be taught as two distinct, although related, operations, as has been done traditionally, or as one operation, as is done in more modern treatments? Should the three cases of percentage be taught separately or as variants of the rule, "Base \times rate = percentage"? Should pupils be taught the method of "casting out nines" or be taught the more general principles of modular arithmetic? How generally should theorems be stated? Proofs? The answer to these and related questions hinges, in part, on the learnability and utility or scope of the rules involved.

Mathematics educators have concerned themselves with such questions, but they have had to make judgments on largely intuitive grounds. There is a real need to better understand the psychological principles involved. Unfortunately, however, previous studies involving rule-learning have dealt only incidentally with this problem. Perhaps more important, even the best-designed studies in this area are subject to criticism for failure to distinguish between structure and behavior variables. The variables chosen (e.g., rule and example given, discovery, answer given) are often merely symptomatic of, rather than basic to, what is involved.

The fundamental assumption underlying our approach to the problem was that more rapid progress can be made by distinguishing clearly between structure and behavior variables and by identifying the important parameters of each. Indeed, we can never hope to understand mathematics and other subject-matter learning without making such a distinction.

At the time the generality study was designed, the SFL had developed only to the point where rules were defined in terms of their denotative sets of ordered stimulus-response pairs. No consideration was given to the nature of the underlying-rule construct. Even so, the fact that sets can be ordered as to their inclusiveness led naturally to the question of rule generality. Not only did this question have practical relevance, but, more important from the standpoint of theory, the SFL provided a basis for rigorously defining just what is meant by generality. One rule is said to be more general than another if the denotative set of the former includes that of the latter as a proper subset (i.e., the former set includes all of the instances of the latter plus some of its own).

Our primary motivation for the rule-generality study, then, was to "try out" this definition to see if it had the sort of straightforward behavioral implications we had hoped. In particular, notice that all S-R instances of a rule are treated equally. Any stimulus within the scope of a rule should provide an adequate test of its acquisition. Similarly, performance on extra-scope test stimuli should be uniformly nil. Once learned, a highly general rule would, of course, be expected to induce appropriate performance on a wide variety of tasks. At the same time, however, it is quite possible that the ease of

---

learning a rule, as judged by the ability to use it, may vary directly with its specificity.

Another facet of our rule-generality research concerned the consistency with which a learned rule is applied. In some earlier research, it was found that under certain conditions experimental subjects respond consistently in accordance with a derived rule. When told that their first response was correct, those subjects who used a rule as the basis for responding to a first test item also used the rule on a second test item. These pilot results were obtained in a discovery-learning situation with simple materials. There was a need to extend this finding to more complex subject matters which are presented by exposition. In general, we found strong support for this contention. Apparently, people tend to respond in a consistent manner, unless the context is changed or feedback otherwise indicates that the rules have changed.

For the purposes of this discussion, we shall be primarily concerned with only two hypotheses. First, performance on problems within the scope of a rule does not differ appreciably, and successful problem solving is limited almost exclusively to within-scope problems. Second, the ease of learning a rule statement so that it can be applied to within-scope problems varies directly with the rule’s specificity. Thus, the more general a rule-statement, the poorer the learning.

To test these hypotheses we conducted two experiments, only one of which is outlined here. The crux of the experimental design and the results can be seen in Table 1. The three groups of experimental subjects (S’s) were undergraduates enrolled in a mathematics-education course for elementary teachers. Each group was presented with one of three rule statements of varying generality. All of the S’s were then tested on the same three problems. Rule S was the most specific and was appropriate for solving any problem in a certain class—Problem 1 was one such problem. Rule SG was more general and was potentially applicable to a wider range of problems. In particular, it was logically sufficient to solve both Problem 1 and Problem 2, but not Problem 3. Rule G was the most general rule; it was applicable to all three test problems.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Pr. 1</th>
<th>Pr. 2</th>
<th>Pr. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>17</td>
<td>13</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>SG</td>
<td>17</td>
<td>5</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>17</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

In agreement with the first hypothesis, there was almost no extra-scope transfer, and performance on within-scope problems was essentially the same. Notice, in particular, that the performance obtained could be predicted on the basis of a strictly logical argument. If people learn an unfamiliar rule at all, then they should be able to apply it to any problem or stimulus within the scope of the rule but not to similar problems beyond the scope of the rule. To check this hypothesis, consider the performance of the three groups on Problem 1. Since this was the only problem within the scope of all three rules, we expect on the basis of our hypothesis that

---


---

Actually, the facts are not quite this simple, but the statement appears to be a good first approximation. John Dunnin and I have a study under way in the Penn Laboratory which we hope will add further clarification.

586 The Mathematics Teacher October 1968
Group S would do better than Group SG, which in turn would do better than Group G. As you can see, Group S did do much better than the others, but the performance of the subjects in Groups SG and G did not differ appreciably.

To explain these results, it was found necessary to define an underlying construct, a cognitive competency underlying the behavior we observed. This led to the four-tuple characterization previously outlined. When the competencies necessary for interpreting the rule-statements of varying generality were analyzed, it became apparent that the more general the rule, the more is required of the learner. Thus, to apply a highly general rule-statement, once memorized, requires that the learner be able to apply any rule of lesser generality, but not conversely. The ability to subtract any two numbers, for example, implies that 2 can be subtracted from 6, but being able to subtract 2 from 6 does not imply that the learner can subtract with any two numbers.

The fact that performance of the groups given the most general (G) and intermediate (SG) rule-statements did not differ can probably be attributed to prior learning. In effect, those additional abilities needed to interpret and hence to apply Rule G were not likely to have caused the college students involved any difficulty. The corresponding S-SG difference was more substantial.

Since postdiction is held in generally low regard in the psychological (but not scientific) community, let us add a brief word of defense. In this particular study, we began with a precise definition of rule generality based on a preliminary version of the SFL; we conducted a study, and, finally, we used the results to extend and otherwise improve the very foundations on which our a priori analysis had been based.

What educational implications can be drawn from this study? To begin with, the results of these experiments demonstrate, in a rather conclusive fashion, the behavioral relevance of rule generality. For the most part, successful performance was noted only on tasks within the scope of verbally stated rules. When rules are presented in an expository fashion, it is normally too much to expect generalization to problems to which the principle does not immediately apply.¹⁴

Potentially of even greater significance was the lack (there was one exception) of performance differences on the within-scope problems and the consistency results cited earlier. The former result demonstrates that, under certain specifiable conditions, any stimulus within the scope of a rule is equally as difficult to respond to correctly as any other. Furthermore, the obtained consistency results suggest that only one (new) test stimulus may be needed to determine whether, in fact, a given rule has been learned.

No more information is gained by using additional test instances. These results could have far-reaching implications for the development of highly efficient measuring instruments.

In addition, the pronounced tendency of the S's to attack all of the test problems in the same way, irrespective of whether the procedure used was appropriate, suggests that knowing how to solve problems and knowing when to use this knowledge are quite distinct abilities. Testing for the latter ability necessarily must involve the presentation of extra-scope problems.

Most important, in view of the problem posed originally, this study helps to reconcile the views of subject-matter specialists and learning theorists as to how general the presentation of material ought to be. While one may expect maximum transfer potential by introducing rules of the greatest possible generality, this transfer potential is bought at a price the teacher may or may not be willing to pay. The greater its generality, the harder a rule-statement is to interpret. Hence, before determining

¹⁴ See Footnote 13. I might add that some transfer does take place, and we are now attempting to pin down the source of this transfer.
how general a rule-statement should be, it is essential to first consider whether the students involved have the necessary requisite interpretative abilities.\textsuperscript{15}

Actually many teachers already do this, at least intuitively. All that the present study does is to make these intuitions more explicit. Frankly, I am always pleased when our results conform to what might be called "common sense." Too often psychological results, while perhaps relevant to animal learning or the memorization of nonsense-syllable lists, have very little to say about the learning of mathematics and other subject matters.

\textit{Discovery learning}\textsuperscript{14}

One of the fundamental assumptions underlying several of the new mathematics programs is that discovery methods of teaching and learning increase the student's ability to learn new mathematics. Indeed, this assumption has guided the development of many new curricula in all of the subject-matter fields. Attempts to demonstrate advantages or disadvantages of self-discovery, however, either have failed, have been open to criticism on scientific grounds, or are seemingly inconsistent even when apparently well controlled.

Research on discovery learning has been confounded by differences in terminology, the frequent use of multiple dependent measures, and vagueness as to what is being taught and discovered. While the difficulties due to the use of inconsistent terminology can often be minimized by a careful reading of research reports, the use of multiple dependent measures often makes it impossible to interpret experimental results unambiguously. Several investigators, for example, have found that groups which are given an expository statement of a rule perform better on transfer tests than groups which are required to discover this rule for themselves (from instances of the rule). The obtained differences in transfer ability, however, may well have been because the discovery groups simply did \textit{not} discover the rule!

Gagné and Brown\textsuperscript{17} overcame the dependent-measure problem by equating original learning and investigating only transfer differences on new problems. On the basis of an analysis of the learning programs used in the study by Gagné and Brown, Eldredge hypothesized that the obtained results could have been due to a number of flaws in the programs used.\textsuperscript{18} Eldredge proposed that exposition and discovery situations may be better characterized as differences in order of presentation. Exposition may be defined as giving rules and then examples of these rules, whereas discovery may be defined as giving the examples and then the rules. Contrary to his hypothesis, however, his discovery group did evidence more transfer than his exposition group. Unfortunately, there were a number of difficulties with the study that make the results hard to interpret.

The Set-Function Language was used as an aid in removing these difficulties. The resulting analysis of what is involved in discovering rules indicates that discovery learners may learn "something" by which they can derive solutions to an entire class of problems. Roughead and I called this "something" a \textit{derivation rule}. Thus, discovery learners who actually succeed in making a discovery should be expected to perform better than expository learners on tasks which are within the scope of


\textsuperscript{17} Garth M. Eldredge, "Discovery vs. Expository Sequencing in a Programmed Unit on Summing Number Series," in \textit{Sequence Characteristics of Text Materials and Transfer of Learning}, eds. Gabriel M. Della-Piana, Garth M. Eldredge, and Blane R. Worthen (Salt Lake City: Bureau of Education Research, 1965).
such a derivation rule. If the new problems presented have solutions beyond the scope of a discovered derivation rule, however, there would be no reason to expect discovery S's to have any special advantage.

This study was concerned with two basic questions. First, can "what is learned" by discovery be identified; and, if so, can that knowledge be taught by exposition with equivalent results? According to the SFL, all behavior is controlled by rules so that there might well be some identifiable rule which is equivalent to "what is learned" by discovery. Specifically, we hypothesized that "what was learned" by guided discovery in the study by Gagné and Brown could be identified and, hence, could be presented by exposition. Second, how is "what is learned" by discovery dependent on what the learner already knows and/or on the nature of the discovery treatment itself? More particularly, we hypothesized that the discovery of a derivation rule can actually be hindered by having too much prior information.

Assuming transfer depends only on whether the derivation rule is learned, sequence of presentation should have no effect on transfer so long as the subject is forced to learn the underlying derivation rule. That is, presenting the derivation rule by exposition or by guided discovery either before or after presenting the desired responses should have no effect on performance on transfer tasks. On the other hand, if a discovery program simply provides an opportunity to discover (with hints as to the solution) but does not guide the learner through the derivation procedure, sequence of presentation might well have a large effect on transfer. Assuming the learner is capable and motivated, he may well succeed in determining the appropriate responses and, in the process, discover a derivation rule. It is not likely, however, that a person would learn such a derivation rule if he already knew the correct responses.

We made three hypotheses: (1) what is learned by guided discovery can be presented by exposition with equivalent results; (2) presentation order is not critical when learners are effectively "forced" to learn derivation rules, either by exposition or by guided discovery; and (3) presentation order is critical when the discovery guidance provided is specific to the respective responses sought, rather than relevant to a general strategy or derivation rule.

The task we used was essentially identical to that used by Gagné and Brown and by Eldredge, and it involved finding formulas for summing the terms in number sequences. That is, the stimuli were number series, like 1 + 3 + 5 + 7, + . . . + (2n - 1), and the responses were formulas in n, the number of terms, for summing such series. For example, the appropriate formula for summing 1 + 3 + 5 + 7 + . . . + (2n - 1) is \( n^2 \).

Using the SFL as a guide (i.e., by identifying, in turn, D, O, and R) we were able to identify that derivation rule taught in the guided-discovery program used by Gagné and Brown. On the basis of this knowledge, four programs were constructed: (1) the formula-given program simply stated the correct summing formula for each problem series; (2) the guided-discovery program remained essentially as it was in the earlier studies; (3) the expository program consisted of a precise expository description of that derivation rule which was presumably equivalent to that learned by guided discovery—it consisted of a general procedure by which the desired formulas could be derived; and (4) the opportunity-to-discover program, in which the problem sequences were presented along with encouragement and hints as to what the desired formulas were. These hints involved such statements as, "The formula has a '2' in it." The same number sequences were used in each of these four programs.

Seven treatments were constructed by
combining these four basic programs. After going through a common introductory program, one group of subjects simply went through the formula program. The other six groups received the formula program together with one of the other three programs. Two of these six groups received the guided-discovery program together with the formula program; two additional groups received the expository and formula programs; and the final two groups received the opportunity-to-discover and formula programs. One group, in each of the resulting three pairs, received the programs in one order; the other group received them in the reverse order. Only the order of presentation was varied. After finishing their respective programs, all of the students were tested on new series to see how well they could determine the appropriate summing formulas.

The results were rather clear-cut. Essentially, the group given the formula program only and the group given the formula program followed by the opportunity-to-discover program performed at one level. The other five groups performed at a common and significantly higher level. Two points need to be emphasized.

First, "what is learned" during guided-discovery learning can, at least sometimes, be taught by exposition—with equivalent results. Of course, there are undoubtedly a large number of situations where, because of the complexity of the situation, "what is learned" during discovery can not be clearly identified. It is still an open question, for example, whether still higher order-derivation rules, which have a more general effect on the ability to learn, may be learned by discovery. If the answer to this question is affirmative, there may be no real alternative to learning by discovery, unless or until we can identify just what is involved. Nonetheless, intuition-based claims that learning by self-discovery produces superior ability to solve new problems, as opposed to learning by exposition, have not withstood experimental test. The value of some forms of discovery to transfer ability does not appear to exceed the value of some forms of exposition. Apparently, the discovery myth has come into being, not so much because teaching by exposition is a poor technique as such, but because what has typically been taught by exposition leaves much to be desired. As we identify just what it is that is learned by discovery in more and more situations, we shall be in an increasingly better position to impart that same knowledge by exposition.

The second point to be emphasized concerns the sequence effect. While the group that was given the opportunity-to-discover program and then the formula program performed as well on the transfer problems as those given the derivation principle in a more direct fashion, the group given these programs in the reverse order (i.e., the formula-opportunity group) did no better than those S's given the formula program alone. In effect, if a person already knows the desired responses, then he is likely not to discover a more general derivation rule.

An extrapolation of this result suggests that if $S$ knows a specific rule, then he may not learn one which is more general, even if he has all of the prerequisites and is given the opportunity to do so. The reverse order of presentation may enhance discovery without making it more difficult to learn more specific rules at a later time. In effect, prior knowledge may actually interfere in a very substantial way with later opportunities to discover. In spite of this fact, there may be some advantages inherent in learning more specific rules. Although the data are not entirely clear on this point, it is quite possible that specific rules may make it possible to determine responses more quickly than rules which are more general.

This sequencing result may have important practical and theoretical implications. The practical implications will be attested to by any junior high school
mathematics teacher who has attempted to teach the “meaning” underlying the various computational algorithms after the children have already learned to compute. The children, in effect, may say to themselves something like, “I already know how to get the answer. Why should I care why the procedure works?” Similarly, drilling students in their multiplication facts before they know what it means to multiply may interfere with their later learning what multiplication is.

This point should be clearly understood; it is an important one. It is not that we should teach meaning first simply out of some sort of dislike for *rote* learning—for certain purposes rote learning may be quite adequate and the most efficient procedure to follow. The point is that learning such things as how to multiply, without knowing what multiplication means, may actually make it *more* difficult to learn the underlying meaning later on.

In addition to the studies described above, we have conducted a number which are based on the same general theoretical orientation. One of them is designed to clarify the role of attribute and operation cueing in discovering mathematical rules. Another deals with the role symbolism plays in mathematics learning. We are also deeply involved in developing a completely new methods course in mathematics for elementary school teachers, closely tied to this point of view.

**Concluding Remarks**

In this paper, I have tried to share with you some of my thoughts on the psychology of mathematics learning—or what I like to think of as the *emerging* discipline of *psychomathematics*. This was done, not in a direct manner, but by pointing out certain inadequacies in existing behavioristic theory as it relates to mathematics learning and, more important, by describing an alternative scientific language and showing how it can be used to formulate research questions that involve mathematical learning and performance.

The mathematical notions of *sets* and *functions* were proposed as a basis for representing (1) the denotative or observable aspect of rules and principles, (2) the underlying knowledge itself, and (3) metalinguistic descriptions or representations of the underlying knowledge. Very recently, I have become intrigued with the idea of integrating these ideas by borrowing the very fundamental but more abstruse mathematical idea of a *functor*. The functor may also make it possible to distinguish in a very precise way between the sort of “ideal” competencies which have long been championed by linguists and competencies as they actually exist in human beings. In the present version of the SFL, it has only been possible to deal with idealized rules (i.e., competencies).

To ensure continued progress, it seems to me that a dual emphasis is needed in *psychomathematics*. On the one hand, there are a large number of unspecified, but crucial, “ideal” competencies which underlie mathematical behavior. These need to be identified. This is a problem area which in many ways is analogous to linguistics. There is also the urgent need to consider how the inherent capacities of learners and their previously acquired knowledge interact with new input to produce mathematical learning and performance. Again by analogy, we have a field much like psycholinguistics, a field which will seek to integrate knowledge concerning mathematical structures and modes of thought with psychology. This kind of distinction, I believe, will prove crucial to any deep understanding of how mathematics (and other subject matters) are learned.