NEW DIRECTIONS FOR THEORY AND RESEARCH
ON RULE LEARNING

II. EMPIRICAL RESEARCH

JOSEPH M. SCANDURA

Graduate School of Education, University of Pennsylvania,
Philadelphia, Pa., USA

In Part I of the present monograph on 'New directions for theory
and research on rule learning', a new scientific language, called
the Set-Function Language (SFL), was introduced and contrasted with
the S–R language. In this (Part II), the second of three articles in the
series, some related empirical research is reported.

Much of the to-be-reported research was formulated and completed
prior to the development of the SFL in its present (but still preliminary)
form. Indeed, in the beginning, the research was based largely on a
vague feeling that a new approach to research on meaningful learning
was needed (Scandura, 1964a, b; Scandura, 1966a, b, c; Scandura
and Behr, 1966). Both the language and the experiments evolved
simultaneously, with the language helping to give impetus to the experi-
ments, and the experimental results pointing to theoretical inadequacies
in need of clarification. I have tried to indicate both the chronology of
developments to date and some of the pitfalls which have helped to
shape my thinking.

PILOT RESEARCH ON RESPONSE CONSISTENCY

During the summer of 1962, Greeno and Scandura (1966) found,
in a verbal concept learning situation, that essentially S either gives the
correct response the first time he sees a transfer stimulus or the transfer
item is learned at the same rate as its control.¹

The thought later occurred to me that if transfer obtains on the first
trial (if at all), then responses to additional transfer items, at least
under certain conditions, should be contingent on the response given

¹ The learning (list 1) and transfer and control (list 2) stimuli were
Underwood and Richardson (1956) nouns, the responses were nonsense
syllables, and the lists were learned by a self-paced anticipation method. The
transfer stimuli belonged to the same concept categories as did the learning
stimuli.
to the first transfer stimulus. In effect, a first transfer stimulus could serve as a test to determine what had been learned during the original learning, thereby making it possible to predict what response $S$ would give to a second transfer stimulus. To test this assumption, I had a total of about fifteen (highly educated) $S$s overlearn the list shown in fig. 1.

Prior to learning the list, both the $S$s and the experimenter agreed on the relevant dimensions and values — size (large—small), color (black—white), and shape (circle—triangle). The $S$s were told to learn the pairs as efficiently as they could since this might make it possible for them to respond appropriately to the transfer stimuli. After learning, the Test One stimuli were presented and the $S$s were instructed to respond on the basis of what they had just learned. Positive reinforcement was given no matter what the response. Then, the Test Two stimuli were presented in the same manner.

The results were clear-cut. All but three of these $S$s gave the responses 'black' and 'large' respectively to the two Test One Stimuli (see fig. 1) and, also, responded with ‘white’ and ‘small’ to the Test Two Stimuli. On what basis could this happen? It was surely not a simple case of stimulus generalization; the responses did not depend solely on common stimulus properties. The first Test One Stimulus, for example, is as much like the fourth learning stimulus as the first (see fig. 1).

Perhaps the simplest interpretation of the obtained results, is that most of the $S$s discovered the two underlying principles during list one learning and later applied them to the test stimuli. These principles might be stated, 'If (the stimulus is a) triangle, then (the response is the name of the) color' and 'If circle, then size'. The former principle may
be characterized by letting I involve the property of being a triangle, D = the set of color properties, R = the set of color names, and O = the transform which maps color properties onto their names. The denotation of this principle would consist of the set of S-R pairs, \{(S_i, R_i) | i \in I \} where S_i is a colored triangle and R_i is the name of the corresponding color).

The results obtained in this miniature pilot experiment (which I have repeated a number of times) provide support for the contention that principle learning is an all-or-none affair (Scandura, 1965; Greeno and Scandura, 1966). The Ss either learned the principles or they didn’t; in almost every case, Test Two performance was compatible with that obtained on Test One. This is not to say, of course, that all of the Ss learned the two principles indicated above. Apparently, some of the Ss ignored the similarities between the original pairs and learned them in rote fashion – i.e., as four distinct principles, each involving one stimulus (actually one equivalence class of stimuli) and the corresponding word response. Under such circumstances, random test performance would be anticipated.

The results of another pilot study, conducted at the University of Michigan during the summer of 1963 and reported by Scandura (1966d), were also revealing. In this case, high dominance nouns taken from the Underwood and Richardson (1956) list were used. Each of these nouns elicited a single adjective associate with a frequency greater than 50%. Eleven college Ss overlearned a list of eight such nouns associated with a total of four adjective categories; two nouns were associated with each adjective category. Both stimuli in a given category were assigned a common response. The four Test One and four Test Two stimuli were also high dominance nouns associated with the same four adjective categories. The task was put in the context of a game in which an explorer is lost and must determine which direction (the four responses) he should go given various hints (the transfer nouns). S had the option of responding to the transfer nouns with, ‘I don’t know’, when none of the four learned responses seemed appropriate. Without this control, appropriate responding to a transfer noun would have occurred by chance in about one out of four cases. Again, positive reinforcement was given for all choices.2

The results with these concept materials were equally revealing. In

---

2 The predictive value of the methodology described is dependent on knowing when S will continue to employ the same rule or responding set.
those cases where transfer potential was indicated, the responses to the second set of stimuli conformed to prediction in 47 of 52 cases. Furthermore, when asked, all but two of these Ss correctly identified the common adjective as the basis for their test responses. Consistency of response to test-stimuli may be influenced by feedback as well as by instruction variables operating between the first and second test responses. The effects of positive, negative, and neutral reinforcement of the first test response may be crucial. Telling S how he should respond or indicating that the 'rules have changed,' by hint or choice of test stimuli, may also affect response consistency. On the other hand, suggesting that the first response is appropriate apparently encourages use of the same responding set on the second test. As indicated above, S was told in our pilot experiments that he was correct regardless of how he responded to the test stimuli. He also was encouraged to respond on the basis of his prior learning. In effect, the experimental situation was designed to both control and capitalize on Einstellung for assessment and predictive purposes. Under these conditions, response consistency was near perfect. Nonetheless, these results cannot be interpreted as unambiguously as in the first pilot study. With verbal materials it is almost impossible to identify all of the important stimulus dimensions. Nonetheless, in concept learning experiments, using the Underwood and Richardson (1956) materials, it does appear that the dominant adjective associate plays the predominant role.

With actual subject matters, assessment sometimes presents additional problems. In the first place, it is not always easy to specify uniquely the basis for an overt response. There is usually more than one path to the goal. Consider, for example, a situation in which S is asked to compute 35-449+35-551 as rapidly as possible. S can laboriously multiply 35 times 449 and 35 times 551 and then add the products or he can recognize this is an instance where the distributive rule would allow him to compute 35 (449 + 551) = 35 • 1000 = 35,000. Clearly, it is not the sum alone which determines what is learned (i.e., the 'way' in which the problem is solved), but the time it takes to respond. If the correct answer is given in a short time, the distributive rule or some equivalent was probably used. Giving the answer in a relatively long time would likely indicate use of the usual computational algorithm. In both cases, the physical response alone would not be adequate to specify the rule. If S gives an incorrect answer or if the problem is so easy that there would be little time differential no matter how S does the computations, further complications would be introduced.

In short, the careful selection of test stimuli and responses is essential in order to assess knowledge. Ideally, these elements should be chosen so as to eliminate all modi operandi but the one in question. Although probably not attainable, this ideal can be approached in many cases.

Another problem involved in work with actual subject matters is that of complexity. More than one principle may, and usually does, enter into a single test response. To determine the knowledge underlying the response, it is often
Principle Learning

The question of relationships between S–R pairs seems so basic, and so obvious, that one wonders why it has not been studied extensively. Because it provides a simple context in which to contrast mediation and set-function formulations, this problem is discussed in some detail.\(^5\)

Consider a paired associate (PA) context in which the relationships between four pairs are varied while the other factors are held constant. In fig. 2, such a manipulation is accomplished by selecting the two principles indicated by, 'If black, then shape' and 'If white, then size'.

<table>
<thead>
<tr>
<th>EXPERIMENTAL</th>
<th>S-R MEDIATION</th>
<th>SFL</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAPEZOID</td>
<td>S(<em>{1}) - R(</em>{1})</td>
<td>({S_1, R_1})</td>
</tr>
<tr>
<td>CIRCLE</td>
<td>R(_{1})</td>
<td>I = (BLACK)</td>
</tr>
<tr>
<td>CIRCLE</td>
<td>S(_{2})</td>
<td>D = R = {SHAPE}</td>
</tr>
<tr>
<td>SMALL</td>
<td>R(_{2})</td>
<td>R = {SMALL}</td>
</tr>
<tr>
<td>LARGE</td>
<td>S(_{3})</td>
<td>({S_3, R_3})</td>
</tr>
<tr>
<td>LARGE</td>
<td>R(_{3})</td>
<td>I = D = {LARGE, BLACK, TRIANGLE}</td>
</tr>
<tr>
<td>SMALL</td>
<td>S(_{4})</td>
<td>({S_4, R_4})</td>
</tr>
<tr>
<td>TRAPEZOID</td>
<td>R(_{4})</td>
<td>I = D = {LARGE, WHITE, TRIANGLE}</td>
</tr>
</tbody>
</table>

Fig. 2. Sample paired-associate lists, together with S–R mediation and SFL representations of these lists. In the experimental list the pairs are interrelated; in the control list, they are not. Two principles are involved in the experimental list: (1) If black, then shape, (2) If white, then size.

Broken lines indicate associations to be learned; solid lines indicate previous associations. The symbols, \(rs\), refer to both the mediating response and the response produced stimulus. In the S-R mediation representation of the experimental lists, \(rs_1\) corresponds to 'black,' \(rs_2\) to 'shape,' \(rs_3\) to 'white,' \(rs_4\) to 'size,' \(rs_1\) to 'triangle,' \(rs_2\) to 'circle,' \(rs_3\) to 'large,' and \(rs_4\) to 'small.' In the control list representation, \(rs_1'\) corresponds to 'circle,' \(rs_2'\) to 'small,' \(rs_3'\) to 'large,' and \(rs_4'\) to 'triangle.'

In the SFL representation both the principles and denotations are characterized. \(O\), in each case, is the mapping between D and R. Necessary to assess each principle individually, as in diagnostic work with school children.

In many test situations, there are few available responses from which to choose (as in True-False and Multiple Choice tests). Under these conditions,
In the experimental list, two pairs correspond to each of the two principles involved. The stimulus properties (sizes, colors, and shapes) and the (internal) responses (names of shapes and sizes) in the control list are identical with those in the experimental list. Thus, any differences in the learnability of these lists would be hard to attribute to anything but the presence of relationships between pairs in the experimental list. Assuming S and E agree on the relevant stimulus dimensions, S's task in learning the experimental list can be viewed as that of discovering the principle identifying (I) and response determining (D) attributes since O is essentially an identity map (i.e. naming) between D and R. On the other hand, S could learn the experimental list without noting any relationship between the pairs. Only one real alternative is available in learning the control list; the list was constructed so that no principle exists which involves more than one pair. To the extent that relationships between pairs are noted, the experimental list should be easier to learn.

There are additional problems of assessment since there is a high probability of giving any particular response (by guessing) irrespective of what is learned. A similar problem obtains in assessing concept learning. There are at least three ways of minimizing this problem: (1) present more than one test stimulus, (2) include appropriate controls for comparison (e.g., GREENO and SCANDURA, 1966), and (3) provide an alternative to guessing as was done in the pilot study described above.

The assessment methodology employed in this research may be used in conjunction with two types of variable: (1) those which affect the probability of rule learning and (2) those which affect response consistency. Giving directions and presenting cues, hints, or other attention-getting devices provide examples of the former type of variable. The consistency with which S responds according to a learned rule may be influenced by feedback, as well as instruction variables operating between the first and second test responses.

When this research was conducted, I, in the four-tuple characterization of a principle, was viewed as a set of stimulus properties which determine when a particular rule, (D, O, R) is to be applied. It now seems more reasonable to view the process by which rules are selected for use as higher order rules defined on the entire contextual situation, including the desired goal as well as the stimulus context.

It may appear that an appropriate control list could be constructed by pairing the experimental stimuli and responses in random fashion. Alas, this turns out not to be a critical control. The responses in the experimental list would all be names of properties of the experimental stimuli but this would not necessarily be the case in the control list. In effect, any differences between the groups could be attributed to prelearned associations between the respective stimuli and responses in the experimental list rather than to relationships between the pairs.
The S–R mediation description of the list contingencies in fig. 2 leaves much to be desired. The representation of principle learning is relatively complex and would have been even more so had we not let ‘rs’ represent both mediating responses and their assumed stimulus properties. No single chain, for example, can adequately represent rule or principle learning in which more than one pair is involved. The one-to-one pairing between the \( S_i \) and \( R_i \) \( (i = 1, \ldots, 4) \) does not follow from a simple analysis of the S–R links in the longer three-stage chain. This chain does not make clear, for example, why \( R_1 \) is the response to \( S_1 \) rather than \( R_2 \). The more direct two-link chains involving the \( rs_i \) \( (i = 1, \ldots, 4) \) serve this purpose. In effect, corresponding stimuli and responses are connected by two chains; those which do not correspond are connected by only one.

In view of this complexity, perhaps the most crucial limitation of S–R formulations may prove to be their inability to lead to practically important questions concerning meaningful learning. The S–R representations that would seem to be called for bear more than a passing resemblance to Copernican epicycles and related attempts to salvage concentric theory.

**Learning principles in paired-associate lists**

Paired-associate learning (PA) has been studied as a function of many variables (e.g., meaningfulness, association value, pronounceability) but little attention has been given to relationships among different S–R pairs. The purpose of this exploratory effort was to determine relationships between (1) the number of S–R pairs related by a common principle, (2) learning rate, and (3) transfer (Scandura, 1967a).

The materials to be learned consisted of 12 pair lists. Each stimulus had a property relating to shape, border, shading, outline, and color. Four colors and eight values of each of the other four attributes were used. The responses were descriptive labels attached to the non-color stimulus properties (e.g., circle). Of the 12 pairs in each list, six were instances of one principle (P6), three were instances of another (P3), two were instances of a third (P2), and one was an instance of a fourth (P1). The principles were constructed so that the same principle applied to all stimuli having a particular color. The response determining cue was either a shape, a border, a shading or an outline. The four colors and the determining attribute dimensions (e.g., shape) were randomly
paired to form four principles (e.g., If black, then shape) which appeared equally often under each condition. The PA list was learned by the anticipation method to a criterion of three consecutive errorless trials.

**Fig. 3. Sample stimulus-response pair.**

To determine whether the principles were acquired sometime during the list learning, each S was shown two transfer lists of four new stimuli each, eight in all. Each transfer list included one stimulus associated with each of the four learning principles. Responding according to one of the principles was presumed to indicate that that principle had been learned.

Prior to learning the original list, each of the 20 college Ss was pre-trained so that he was familiar with the stimulus dimensions and could name each stimulus property. These responses were typed on a card and were always available to S. In addition, S was told that a pattern was involved which might facilitate his learning and guide his responses to the transfer stimuli.

The dependent variables were the average number of errors per instance (i.e., an S–R pair associated with a principle) for each S (on each of the four principles) and the number of appropriate responses to the transfer stimuli.

Except for a very small reversal between treatment P3 and P2, learning rate (i.e., the average number of errors per instance) decreased with the number of instances per principle: 5.0, 3.4, 3.5 and 2.7, respectively ($F = 8.76, df = 3/76, p < .001$). The difference between P1 and P2 was significant ($F = 11.50, df = 1/76, p < .01$), but none of the other adjacent means differed significantly. Under the experimental conditions, the rate of learning an S–R pair increased with the addition of a second S–R instance but increasing the number of instances still further apparently had little effect.
The number of appropriate responses to the transfer stimuli was also affected by the number of instances per principle. There were 27, 8, 15, and 9 appropriate responses (as indicated by the experimental principles) given to the P6, P3, P2, and P1 transfer stimuli, respectively. Although the trend was not entirely regular, a sign test indicated that the degree of principle learning was higher in treatment P6 than in the average of treatments P3, P2, and P1 ($z = 2.6, p < .005$).^7

Another analysis demonstrated that P6 transfer was related to learning rate. Of those 9 Ss who responded appropriately to both P6 transfer stimuli, 7 had below median (2.61) error scores, indicating more rapid learning; of those 11 Ss who responded appropriately to at most one test stimulus, 8 had above median error scores, indicating slower learning. An exact probability test (Finney, 1948) on the resulting 2 × 2 contingency table indicated a significant relationship between P6 transfer and learning rate ($p < .035$). The small number of Ss who gave two appropriate (transfer) responses with respect to the other principles precluded the possibility of obtaining significant relationships. Only 3, 5, and 2 Ss gave both desired responses to the P3, P2, and P1 test stimuli, respectively.

The list learning and transfer results were not entirely consistent. The inclusion of more than two instances did not affect learning rate, but it may have affected transfer. These results could reflect real differences or be simply artifacts of the situation. In either case, the overall pattern of results was sufficiently clear to make any interpretation in terms of stimulus or response generalization extremely difficult, if not impossible. Some resort to S–R generalization (Hull, 1943; Berlyne, 1965) or rule learning would appear necessary. For reason primarily of parsimony, a rule interpretation would seem preferable.

**RULE GENERALITY AND CONSISTENCY IN MATHEMATICS LEARNING**

In many instructional situations, the question often arises as to how general the presentation of material ought to be. Some proponents

---

^7 It might be argued that the difference in the number of appropriate responses was due to there being more responses per category in treatment P6. When in doubt, the Ss may have tended to give a response from the most frequently experienced category. A comparison however, of the average number of P6 responses given to the P3, P2 and P1, transfer stimuli (16) was not significantly higher than the ten P3, P2, and P1 responses given to the P6 stimuli ($p > .10$).
emphasize that the more general a rule the more useful it will be; others, that the more specific the rule, the better the learning. There is a real need to better understand the psychological principles involved, but previous studies dealing with rule (or principle)\(^8\) learning (e.g., Craig, 1956; Gagne and Brown, 1961; Haselrud and Meyers, 1958; Kersh, 1958, 1962; Kittle, 1957; Scandura, 1964a, 1966a; Wittrock, 1963) have dealt only indirectly with this question.

Assuming that the answers hinge, at least in part, on learnability as well as general utility, and armed with the denotative characterization of a rule as a function, we (Scandura, Woodward and Lee, 1967) set out to explore this question. In particular, we were concerned with the effects of rule generality on learnability and transfer. We also explored the response consistency hypothesis with more complex materials. Two experiments were conducted, the independent variable in both cases being the scope (i.e., generality) of a rule statement. Scope was defined in terms of the corresponding denotation, one statement being more general than another if the denotation of the former included the latter.\(^9\)

Our original hypotheses were that: (1) the scope of a rule would be fully reflected in performance, there would be no success on extra-scope problems and little difference in performance on within-scope problems, (2) the learnability of a statement, as determined by within-scope performance, would vary inversely with scope, and (3) the rule taught would be used on all problems under conditions of nonreinforcement.

**Experiment one**

In the first experiment, each group of 17 college Ss (majors in elementary education) was presented with one of three ordered rules dealing with a variant of the number game called NIM. In the game, two players alternately select numbers from a specified set of consecutive integers, beginning with one, and keep a running sum. The winner

---

\(^8\) Although a legitimate distinction may be made between *rules* and *principles*, the distinction is fine and was not recognized until after the study was completed. The terms have been used synonymously throughout the paper except in the concluding section where the distinction is outlined.

\(^9\) Notice that defining a rule as a function makes it possible to consider a variety of other relationships between different rules. In particular, two functions (i.e., sets) may be disjoint (i.e., have no instances in common), overlap, or be identical in addition to being ordered (i.e., one being more general than the other). Such relationships might provide a basis for formalizing questions concerning the effects of prior learning on later learning.
is the one who picks the last number \((j)\) in a series with a predetermined sum. If, for example, this sum is 31 and the set consists of the integers 1–6, the players alternately select numbers from 1–6 until the cumulative sum is either 31 or above (in which case no one wins). There are rules which allow the player who goes first to always win.

Any such game can be characterized by an ordered pair of integers. The application of each winning rule was illustrated with a common \((6, 31)\) game. The least general rule (S), adequate for winning only \((6, 31)\) games, was stated, '... make 3 your first selection. Then ... make selections so that the sums corresponding to your selections differ by 7'. Rule SG was adequate for solving \((6, j)\) games \(j = 1, 2, \ldots, n\) and was stated, 'the first selection is determined by dividing the desired sum by 7 and making the remainder your first selection ... Then ... make selections so that the sums corresponding to your selections differ by 7'. The most general rule (G) was adequate for solving \((i, j)\) games \(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n\) and was stated, 'the first selection is determined by dividing the desired sum by one more than the largest integer in the set from which the selections must come and making the remainder your first selection ... Then ... make selections so that the sums corresponding to your selections differ by one more than the largest integer in the set'.

All Ss, including two control groups, were tested on three problems. The first was within the scope of each rule, the second within the scope of all but rule S, and the third only within the scope of rule G.

The results were straightforward. Of the 13 Ss in group S who solved problem one, none solved problem two, and one solved problem three. The corresponding numbers for groups SG and G were, respectively, 5, 4, 0 and 5, 5, 4. Within the scope of each rule there were only chance differences in performance on the problems. On the other hand, only one S solved an extra-scope problem.

The relative interpretability of the three rule statements was determined by comparing group performance on problem one which was within the scope of each. Rule S proved to be easier to learn, under the self-paced conditions, than were rules SG and G \(p < .005\) in both cases). There was, however, no discernable difference in the interpretability of rules SG and G.

10 Exact probability tests on \(2 \times 2\) contingency tables were used to test the various hypotheses.
The third facet of this research was concerned with the consistency with which presented rules are applied. We wanted to determine whether the S and SG Ss would use the rule taught even when it was inappropriate (on the second and third problems). To make this possible, no information was given as to when the rules were and were not appropriate.

Of the 17 S Ss, 13, 9, and 8 used the rule taught on problems one, two, and three, respectively. The corresponding numbers in groups SG and G were 7, 7, and 5 and 6, 6, and 6. Although there was a slight tendency to not use the rules taught on problems two and/or three, where they were inappropriate, there were no significant differences in frequency of use.

These results certainly provided strong support for our original hypotheses: (1) performance on within-scope problems did not differ appreciably, even though the common illustration was more similar to problem one than the others, and successful problem solving was limited almost exclusively to within-scope problems, (2) rule S proved easier to interpret than rules SG and G, and (3) the rules taught tended to be used consistently on all problems whether they were appropriate or not.11

About the only major unanticipated result in experiment one was that rule G proved as easy to interpret as rule SG. In view of the rather low proportion of successes in these groups, we were originally tempted to attribute the lack of such an effect to scale insensitivity near its lower extreme.

11 The first mentioned result has particular relevance for the psychologist since it crystallizes the fact that no generalization gradient is to be expected when the stimulus values are discrete rather than based on a continuous physical dimension. If there were such a gradient, performance on the first test problem which was more similar to the example, should have been superior to that on the other problems. Even S–R associationists are generally agreed that the lack of such an effect provides indirect support for a rule interpretation. To the extent that the variables involved in meaningful learning are discrete, a rule interpretation may prove more useful.

Furthermore, when the underlying stimulus dimension(s) are continuous, S–R theorists will need to consider the possibility that generalization gradients are simply artifacts of averaging individual differences in perceptual discrimination (and, hence, what rule is learned) over continuous dimensions (e.g., Lykken, Rose, Luther and Maley, 1966).
Experiment two

To determine the generality of these findings, a second experiment, dealing with arithmetic series, was conducted simultaneously with junior high school Ss. In this experiment, both scope (S, SG, G) and example (present, absent) were varied independently. Since most of the Ss were already familiar with arithmetic operations introduced and, to some extent, with number series generally (i.e., as in adding lists), it was felt that examples might provide a basis for generalization, via discovery, to extra-scope problems. Another difference between this experiment and the first was that rule S, $50 \times 50 = 2500$, was effectively an answer given treatment and applied to only one series. This series was used both as the common example and as problem one. In experiment one, rule S applied to a number of different game sequences.

Although the pattern of results shown in Table I paralleled those of experiment one in most respects, there were several important differences. First, the presence of the example (problem one) along with rule S resulted in significantly better performance on problem two than when rule S was shown alone, the only case in either experiment where nonnegligible success was noted on an extra-scope problem. This effect may have been due to the form of the combining operation, '50 $\times$ 50', in the rule S statement. '50 $\times$ 50' is clearly an instance of the more general SG combining rule, 'n $\times$ n = $n^2$'. Presumably, the statement of rule S, together with the common illustrative series, 1 + 3 + 5 + ... + 97 + 99, provided the successful S Ss with enough cues to generalize. In particular, they may have discovered that this series had 50 terms. Hindsight suggests that this difficulty might have been overcome by simply stating the sum, 2500, of the illustrative series rather than '50 $\times$ 50'.

Second, table 1 indicates that only three of the 19 G-with-example Ss

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary: Number of correct sums (uses of rule taught) on problems one, two, and three.</td>
</tr>
<tr>
<td>Rule</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Group S</td>
</tr>
<tr>
<td>Group SG</td>
</tr>
<tr>
<td>Group G</td>
</tr>
</tbody>
</table>
solved problem three whereas 18 solved problem one and 14 solved problem two. The decrement between problems two and three was significant (p < .003). The reason for this difference was not immediately apparent especially since 15 of these Ss applied rule G to the third problem. A more intensive post hoc analysis of the situation, however, suggested that the result may have been due to a difference in ease of determining N, the number of terms, for use in the G combining rule, [(A + L)/2] N. N could be determined from problem series one and two by taking the average of the first and last terms. A careful examination of the test papers suggested that this led to the incorrect value (25, rather than 24) for N in the third series, 2 + 4 + 6 + . . . + 46 + 48. In short, the difficulty was not in the rule but in finding the correct value of N. Such difficulties may be circumvented in future experimentation by controlling for such unwanted differences.

Third, although the results of experiment two were in the hypothesized direction, only the overall effect of scope on interpretability was significant. This led us to wonder whether interpretability of the rule statements depended solely on generality. Could the rule statements also have differed as to the difficulty of interpreting the actual terms or symbols used? After consideration of this possibility, interpretability was rejected as an important factor in experiment two since a recheck convinced us that we had succeeded reasonably well in stating each rule as clearly as possible. Perhaps a more likely interpretation is that the Ss' familiarity with arithmetic interacted with the materials used so as to reduce the effects of statement generality.

Fourth, only one of the Ss who was shown the rule, 50 × 50, applied it to problems two and three. This result can probably be attributed to an interfering effect due to prior familiarity with addition problems. The Ss may simply have mistrusted rule S. How could a rule like 50 × 50, having only one answer, be the sum of all three problem series? Most junior high school Ss would find it unreasonable that the series 1 + 3 + . . . + 99 (problem one) and 1 + 3 + . . . + 79 (problem two) have the same sum (50 × 50). Some such reluctance may also have obtained on problem one with group S-without-example.

12 It may be desirable to think of properties, such as N, as being derived from lower order (i.e., more easily discernible) stimulus properties. Thus, the rule, (A + L)/2, worked for problems one and two whereas L/2 was required for problem three.
Nonetheless, we were surprised that only 8 of those 20 Ss, not presented with the illustrative series, gave the correct sum (2500 or 50 × 50) for problem one.

**Implications and theoretical comment**

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 rule does not immediately apply (however, see Scandura and Durnin, 1968).

Of perhaps even greater practical significance were the lack (there was one exception) of performance differences on within scope problems and the consistency results. The former result demonstrates that (almost) any stimulus within the scope of a rule is equally as difficult to respond to correctly as any other. Furthermore, coupled with the consistency data cited above, the obtained consistency results suggest that only one (new) test stimulus is 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 Ss to attack all of the test problems in the same way, irrespective of whether the procedure used was appropriate, suggests that the ability (i.e., knowing how) to solve problems and knowing when to make use of this ability to solve problems are quite distinct. Testing for the latter ability necessarily must involve the presentation of extra-scope problems.

Perhaps even more important than the results of these exploratory experiments were the post hoc analyses they made both necessary and possible. In particular, the results of these experiments indicated that the roles played by various aspects of a principle statement need to be more clearly specified than the form, 'If I', then R" indicates. It does not detail all that appears relevant. For one thing, it was not possible in the rule generality study to distinguish between the roles played by A, L, and N (where A, L, and N have a particular meaning) and the algebraic expression [(X + Y)/2] Z (where the variables have general relevance). The former variables relate to properties (D) of the series
Stimuli, while the latter is a ternary operation (O) by which another such property (e.g., sums) may be derived. \( I' \), of course, although it played no role in the rule generality study is also critical. It tells when, in fact, a rule can and cannot be applied. Thus, the rule, \( N^2 \), is appropriate whenever an arithmetic series consists of consecutive odd integers beginning with 1 while \( [(A + L)/2] \) \( N \) works whenever there is a common difference between adjacent terms.

These observations suggest that a principle statement might be represented more appropriately by the form, 'If \( I' \), then \( O'(D') = R' \), where \( I' \) refers to the set of stimulus properties which indicate when the rule denoted \( O'(D') \) should be applied, \( D' \) refers to the set of those properties which determine the responses, and \( O' \), the operation from which the responses, denoted by \( R' \), may be derived from the properties referred to by \( D' \). This representation led naturally to the four-tuple characterization of a principle \((I, D, O, R)\) and to the previously mentioned distinction between a rule and a principle.

Although the actual symbols used in a statement may be an important factor as has been suggested above and as will be demonstrated in the next section, the hypothesis advanced in the rule generality study to the effect that scope and learnability are inversely related finds a formal rationale in the nature of the characterizing elements. Making operational use, for example, of the arithmetic series property (i.e., dimension), 'the difference between adjacent terms is some common value', necessarily presumes that 'the difference between adjacent terms is two', 'three', 'etc.', can all be correctly interpreted. The converse does not necessarily follow. A similar relationship exists with respect to the rules, \( 50 \times 50 \) and \( N \times N \). To correctly apply the latter, more general, rule to any particular series requires the ability to determine any value of the dimension \( N \), including 50. Being able to apply \( 50 \times 50 \), however, does not.

It would appear that the more general the rule the more is expected of the learner. Whether such differences will be reflected in behavior, however, may depend not only on rule generality but the population involved, particularly on whether the Ss have the necessary requisite abilities.

In effect, differences in generality appear, on analysis, to be equivalent to differences in abstraction level. Thus, the number \textit{two} is more abstract than the property \textit{two oranges} because the former applies to a collection of sets only one (subcollection) of which has the latter
property. For the same reason, the property represented by the place holder \( X \) is more abstract than the number \textit{two} since it refers to a still higher order collection. Unfortunately, we have not yet conducted a study designed to provide definitive information on these points. For the present, this analysis remains hypothetical.

\textit{Interpretability and symbolism}

To help clarify the role symbolism plays in mathematics learning, I recently completed a study (Scandura, 1967b) with the help of John Davis in which we varied both the symbols actually used to construct statements of a principle and the ability of an \( S \) to interpret these symbols. It seems almost axiomatic that the ability to interpret a statement of principle depends critically on the ability to interpret the symbols of which it is composed, be they mathematical symbols or elements of the native language (e.g., English). Nonetheless, in mathematics learning the use of mathematical symbolism is frequently, if not always, preferred to ordinary English. The reason why, however, is never made explicit.

The purpose of this study was to determine whether: (1) principles are more easily memorized when stated symbolically or when stated verbally and (2) the ability to correctly use constituent symbols and the required (grammatical) combining rules is a necessary and/or sufficient condition for applying a learned (i.e., memorized) principle statement.

\textit{Method}

Four artificial principles, each unfamiliar to the 24 \( Ss \) (college majors in elementary education), were selected for study. Each principle was based on one of the following notions: greatest integer, sigma notation for sequential addition, vector, and partial derivative. Two statements of each principle were prepared; one was composed of unfamiliar mathematical symbolism and the other of carefully, yet succinctly, worded English. For example, the greatest integer rule was stated (in English).

(1) Take the greatest integer in \( X \).
(2) Take the greatest integer in \( Y \).
(3) Divide the result of step one by the result of step two.
(4) Take the greatest integer in the quotient obtained in step three.

The symbolic form of this rule was \([([X] \div [Y])]\).

Tasks were designed to train the \( Ss \) to interpret the constituent
symbols. For example, one set of tasks involved the greatest integer function (i.e., \((W, [W], | \text{all real } W)\)). In addition, all of the Ss were required to demonstrate proficiency in the use of parentheses as a means of signifying the order in which binary operations are to be taken. These conventional rules of grammar were involved in all four principles. Of course, neutral materials were used to teach and assess proficiency in the use of parentheses. In no case did the pretraining or assessment include either a complete rule or one of its instances.

A 2 \(\times\) 2 factorial design, with repeated measures, was used. Each S effectively served as his own control. One factor was the form in which a given principle was stated, symbolic or English. The other factor involved the presence or absence of training on the constituent symbols. Of course, the principles were counterbalanced over treatments so that each was used equally often under each of the four treatments. All other unwanted factors were randomized, including presentation order.

Separate measures of learning rate and interpretability were obtained. Learning rate was determined by presenting each principle statement for a fixed period of time for study and testing to see if the Ss could completely reproduce them in written form. All four principle statements were shown once before the next go-through (trial) began. Ease of learning was determined by the number of trials it took to learn each principle to a criterion of two perfect reproductions in a row.

Interpretability was measured immediately after all of the statements had been well-learned. To demonstrate his 'understanding' of the statements, S was required to apply each of the corresponding (underlying) principles to two stimulus instances. For example, one of the problems used to determine whether S could apply the integer rule was stated simply, 'If \(x = 8.64\) and \(y = 3.24\) then . . .' All four principles were tested once before the second set of test problems was given.

Results

The results demonstrated quite clearly that: (1) symbolic rules are learned more rapidly, whether the constituent symbols are familiar or not \((p < .01)\) – there were only 2 exceptions (out of 24) to this generalization, (2) rules, stated in symbolic form, are applied successfully if and only if the Ss have been taught how to apply the constituent symbols (and the necessary grammatical rules) – there were only 4 exceptions to the sufficiency part of this generalization and none as regards necessity, (3) rules, stated in the native English language, are
applied equally well whether or not training in the use of the corresponding mathematical symbols is given, and (4) English statements, once learned, are applied equally as well (in this study, somewhat better) as symbolic statements in which use of the constituent symbols has previously been mastered.

These results are not entirely surprising but they do, nonetheless, make explicit at least one aspect of the role symbolism plays in mathematics learning. The use of symbolism makes mathematics learning more efficient when the constituent symbols and grammatical combining rules have previously been mastered. Symbols, of course, also serve the practical function of requiring less space in printing. Nonetheless, these results suggest that under certain conditions it may be well to remember that ordinary English can be used to teach mathematical ideas.\(^{13}\)

**Extra-scope transfer in learning mathematical rules**\(^{14}\)

The main purpose of this research was to help identify the underlying causes of generalization from one instance of a rule or strategy to another. Our more immediate aim was to provide evidence relevant to the interpretation offered in the Rule Generality Study to explain the one case of extra-scope transfer obtained (i.e., from the rule statement ‘50 \(\times\) 50’ to the statement ‘\(n \times n\)’). In essence, it was hypothesized that if a given rule or strategy is a restriction of a more widely applicable rule then a statement of this restricted rule may very well provide a basis for generalization to extra-scope items, which are instances of the more general rule or strategy. A restricted rule statement may be viewed as one obtained by replacing the variables entering into the statement of

\(^{13}\) These results provide a rational basis for making one type of branching decision that, while intuitively obvious, needs to be made explicit in computer-assisted instruction. Given the objective of learning a particular principle and an expository mode of instruction, one might proceed as follows: (1) test to see if \(S\) can make use of the constituent symbols; (2) if so, present the principle in the more efficient symbolic form; (3) if not, present the principle in English.

Although learning feedback has long been recognized as an important factor in promoting efficient learning, it has been unclear as what sort of feedback to measure. The present results suggest that specific sorts of feedback are needed in order to make specific kinds of decisions.

\(^{14}\) This section is based on a paper presented by John Durnin and the author at the American Educational Research Association Convention in Chicago on February 10, 1968.
a more general rule with the specific values of a particular instance. For example, the restricted rule statement, ‘50 X 50’, can be obtained from the statement, ‘n X n’, by replacing ‘n’ with the specific value, ‘50’.

A secondary purpose of the study was to obtain further information on the consistency hypothesis. That is, we wanted to find out whether a correct response to one instance of a generalized rule would imply success with other instances of the more general rule. Since the scope of a rule was directly related to the number of variables entering into the rule (i.e., the number of dimensions which were allowed to vary), it was hypothesized that if transfer to one instance indicates that a particular rule (e.g., ‘50 X 50’) has been generalized along one or more dimensions (e.g., to ‘n X n’), then transfer should be expected to additional instances which differ from the original only along the same dimension(s).

**Method**

The task used was again the number game ‘NIM’. Restricted statements of the three game winning rules, S, SG, and G, used in Experiment One of the Rule Generality Study, were constructed. These restrictions were applicable only to (6, 31) games. Statement S′ was essentially identical to the rule S Statement. Statement SG′ was a restriction of statement SG in the sense that the rule was restricted to one value (i.e., 31) of the ‘desired sum (m)’ dimension. It was stated,

‘The appropriate first selection is determined by dividing 31 by 7. The remainder 3 should be your first selection...’

Statement G′ was a restriction of statement G in that the rule was restricted to one value (i.e., 31 and 6, respectively) of the ‘desired sum (m)’ and ‘size of selection set (n)’ dimensions. It was stated,

‘The appropriate first selection is determined by adding one to six (1 + 6), and dividing 31 by this result. The remainder 3 of this division is the selection which should be made first... It is important to notice that 7 = 6 + 1.’

Three of the four treatment booklets included one of the restricted statements (S′, SG′, and G′) together with a common (6, 31) game which was provided for practice. The fourth booklet served as a control. The first two test problems, 1A and 1B, were (6, 31) games. Test problems 2A and 2B were (6, m) games, which differed along the
'desired sum' dimension, with \( m = 25 \) and \( m = 29 \), respectively. Test problems 3A and 3B were \((n, m)\) games, which differed along both the 'desired sum' and 'size of the selection set' dimensions, with \( n = 5 \), \( m = 26 \) and \( n = 7 \), \( m = 33 \), respectively.

The Ss were 88 West Philadelphia High School students enrolled in an academic mathematics program. They were randomly assigned to three experimental groups (S', SG', G') and a control (C) so that each group included 22 Ss.

Each S completed one of the four treatment booklets and the test in that order. The experiment was self-paced and with only a few exceptions the Ss completed the experiment well within the time limit of 40 minutes.

The criterion measure was use of the appropriate pattern (AP). S was given credit for using the AP if he won the game and employed an appropriate game winning strategy. All of the tests conducted were applied to \( 2 \times 2 \) contingency tables.

### Table 2

<table>
<thead>
<tr>
<th>N</th>
<th>Problem</th>
<th>Problem</th>
<th>Problem</th>
<th>Problem</th>
<th>Problem</th>
<th>Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1A ((6, 31))</td>
<td>1B ((6, 31))</td>
<td>2A ((6, 25))</td>
<td>2B ((6, 29))</td>
<td>3A ((5, 26))</td>
<td>3B ((7, 35))</td>
</tr>
<tr>
<td>Group C</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Group S'</td>
<td>22</td>
<td>16</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Group SG'</td>
<td>22</td>
<td>19</td>
<td>20</td>
<td>7</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Group G'</td>
<td>22</td>
<td>18</td>
<td>18</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

**Results and discussion**

Table 2 shows that restricted rule statements may provide an adequate basis for generalization. The three experimental groups performed at essentially the same level on the \((6, 31)\) games (Problems 1A and 1B), but there were 12 Ss in groups SG' and G', as compared to none in group S', who were successful on problems 2A and 2B. This difference was significant at the .01 level.

A cursory review of the literature suggests that the transfer observed in a number of other studies may also have involved generalizing a restricted rule statement. MAIER (1945), for example, found that
providing S with a problem solving procedure, as it applied to one
problem (i.e., with a restricted statement), improved the level of per-
formance on a second problem (which was presumably within the scope
of a more general rule). Some such generalization mechanism may also
be involved in what some investigators have called ‘remote transfer’.
Thus, in a recent study, Wittrock’s (1967) non-replacement strategy
group was presented with a restriction of a general strategy which was
applicable to his remote transfer items. Apparently, what these Ss
actually learned (i.e., discovered) was the more general strategy.

The performance of the G’ Ss, however, suggests that transfer can
not necessarily be expected to all problems within the scope of the rule
from which the restricted statement is derived. Of the five Ss in group G’
who were successful on problems 2A and 2B, none was successful on
problem 3A and only two, on problem 3B.

These differences between problems 2A and 2B and problems 3A
and 3B suggest that the level of performance on transfer problems may
depend on the particular dimension(s) involved. Problems 2A and 2B
required that the G’ statement be generalized only along the ‘desired
sum’ dimension whereas problems 3A and 3B required generalization
along the ‘size of selection set’ dimension as well. Apparently, the G’
Ss were more capable of making the former generalization than the
latter.

To test the consistency hypothesis, those Ss who used the AP on
problems 1A, 2A, and 3A and those who did not (non AP users) were
compared as to AP use on problems 1B, 2B, and 3B, respectively.
There were significantly more AP users on problem 1A who were
AP users on problem 1B than was the case for non AP users on
problem 1A (p < .001). The same relationship held for problems 2A
and 2B (p < .001) and problems 3A and 3B (p < .001), respectively.
There were only four cases out of a total of 131 in which a non AP user
(in groups S’, SG’, and G’) on an ‘A’ problem became an AP user on
the corresponding ‘B’ problem. There was only one case (out of 67)
where an AP user on an ‘A’ problem was not an AP user on the
corresponding ‘B’ problem.

These results suggest that if transfer obtains on one new problem,
which differs (from the training problem) along one or more dimensions,
then transfer may be expected to other problems which differ along
these same dimensions.

Of course, the ease with which a correspondence can be determined
between statement cues and the determining properties of an illustrative problem undoubtedly depends heavily on individual differences as well as on the nature of the cue. A major task of future research will be to determine what the important individual differences are.

ATTRIBUTE AND OPERATION CUEING IN THE DISCOVERY OF MATHEMATICAL RULES

So far we have limited our illustration and discussion to reception learning; that is, learning which takes place via the interpretation of symbolic statements. Much, however, is learned by discovery. The typical instructional procedure involves presenting, one at a time, either stimuli or stimulus-response pairs corresponding to the to-be-discovered rule. To determine whether learning has taken place, the learner is usually asked to give the appropriate response to new stimulus instances. The former type of situation, in which S is tested repeatedly, is exemplified by:

The sum of the first 2 odd integers, $1 + 3 = ?$
The sum of the first 3 odd integers, $1 + 3 + 5 = ?$

...  

The learner would be expected to discover that the correct sum may be obtained by simply squaring the number of terms in the (odd integer) series.

Ostensibly to speed the discovery process, the teacher or auto-instructor might cue the critical aspects of the stimuli. Thus, italicizing the number of odd integers (as above) or printing them in red would presumably attract attention by making the cues more salient. Of course, it would be equally possible to identify the appropriate operation (i.e., squaring) or, to introduce various combinations of both determining (D) and operation (O) cues.

Although a good deal of verbal and non-verbal cueing goes on during the contemporary discovery lesson in mathematics, there have been very few attempts to uncover the basic mechanisms involved. Whereas, recently, there have been a few such studies concerned with concept

---

15 Other variants of reception learning are possible, such as learning from diagrams or pictures (icons), as in geometry, and from concrete objects, as with DIENES' (1967) multiple embodiments.
learning (Haygood and Bourne, 1965; Wittrock and Keisler, 1965) and, earlier, with problem solving (Maier, 1930), no consideration has been given to rules. This is indeed, unfortunate since rules seem to underlie so much of mathematics learning.

With this in mind, the members of my research seminar on mathematics learning 16 conducted a pilot study to determine the effects of verbal attribute and operation cueing on the rate of discovering mathematical rules (Scandura, 1966e). In particular, the study extended that of Haygood and Bourne (1965) in two ways: (1) rules were used instead of concepts and (2) the operations involved were arithmetic, rather than logical. The study was designed simply to determine whether identifying the determining attributes (i.e., the class D of stimulus attributes which determine the responses) or the appropriate combining operation (i.e., O) does, in fact, increase the rate at which arithmetical rules are discovered.17

**Method**

To minimize the effects of individual differences, we again used artificial materials. The stimuli were four-tuples of numbers (e.g., (4, 8, 9, 3)) and the responses were simply new integers that could be derived uniquely from exactly three of the four original integers by some combination of two (of the four) elementary arithmetic operations.

Three rules were used. The determining characteristics and operations, respectively, were (1) $A_1, A_3, A_4$; (2) $A_1, A_2, A_4$; (3) $A_2, A_3, A_4$; and (1) $X + Y - Z$, (2) $X \cdot Y \div Z$, (3) $X \cdot Y + Z$ where the subscripts, $i = 1, 2, 3, 4$, in $A_i$ refer to position in the four-tuple and $X, Y,$ and $Z$ to place holders.

A $3 \times 3$ design, with repeated measures on the second factor, was used. Factor one involved the type of cue given (none, determining attribute (D), operation (O)). The second factor was a composite of the rule in question and the order of presentation 18 but gave some

---

16 Mike Bundrick, John Davis, Rosalie Jensen, Bob King, Frank Pavlick, and Larry Smith.

17 Although the order in which stimuli are introduced is also used extensively to promote discovery, this factor was not considered in this preliminary effort.

18 Since the study was designed as much as a learning experience (for my graduate students) as one of advancing knowledge, no attempt was made to remove this confounding, by counterbalancing principles over presentation order.
indication of the effects of one discovery on the next. The 36 elementary education majors were randomly assigned to one of the three cue groups so that a given S was exposed to only one type of cue. Each S completed three discovery episodes.

The Ss were told that their job was to write that number which they thought corresponded to the four-tuple shown. They were also instructed, 'There is a procedure by which you can always determine the corresponding number when I show you the set'. Then, the control group was told, 'To help you discover this procedure as rapidly as possible, you should try to determine the three specific positions in the four-tuple and a rule which combines the numbers in these positions to yield the corresponding number'. The attribute group was told, '... determine a rule which combines the numbers in the (proper positions inserted) to yield the corresponding number'. The operation group was told, '... determine the three specific positions in the four-tuple from which the numbers X, Y, and Z are always taken where (proper rule inserted) yields the correct number'.

After S responded to a given four-tuple, either by writing a number or by indicating he 'didn’t know', the card on which the four-tuple appeared was turned over exposing the same four-tuple together with the correct number response. The Ss were given approximately 12 seconds to compare the four-tuple with its solution before the next four-tuple was shown. Ten such four-tuples comprised one problem and S’s score was the total number of correct responses made.

The probability of giving a correct response, by chance, without discovering the corresponding rule was relatively small. This was evidenced by the fact that once a correct response was given, S almost invariably gave the correct response thereafter.

Results

The results are summarized in table 3.

Both attribute and operation cueing induced significantly ($p < .01$) earlier discovery of all three rules. But, whereas a significant improvement, presumably due to practice, was noted between problems two and
three \( p < .01 \), there was essentially no difference in performance on the first two problems. This is surprising since so-called ‘warm up’ effects typically have their greatest effect in the beginning.

<table>
<thead>
<tr>
<th>Problem</th>
<th>One</th>
<th>Two</th>
<th>Three</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute</td>
<td>6.6</td>
<td>4.2</td>
<td>8.1</td>
</tr>
<tr>
<td>Operation</td>
<td>3.5</td>
<td>7.2</td>
<td>8.7</td>
</tr>
<tr>
<td>Control</td>
<td>1.1</td>
<td>0.6</td>
<td>4.0</td>
</tr>
</tbody>
</table>

More intensive comparisons, however, indicated that problem one practice significantly \( p < .01 \) improved operation group performance on problem two but actually hindered (significantly, \( p < .01 \)) attribute group performance – classic cases of positive and negative transfer. At least two possible interpretations may be given for the latter finding. First, having discovered that some combination of addition and subtraction \((A_1 + A_3 - A_4)\) worked on problem one, many of the attribute-cue Ss may have spent too much time trying various combinations of addition and subtraction on problem two. Second, the rule, \( A_1 \cdot A_2 \div A_4 \), needed to solve problem two, since it involved multiplication and division as opposed to addition and subtraction, may have been intrinsically harder than that needed on problem one. Of course, both interpretations may have some degree of truth. Having been given the appropriate rules, in both cases, the operation-cue Ss were not subject to such effects. Furthermore, whatever response set developed on the basis of discovering that positions 1, 3, and 4 were relevant on problem one was more than compensated for by the practice afforded.

Clearly, research aimed specifically at such questions is needed to provide definitive information. It is impossible to say, at this time, that we fully understand exactly what is involved or, even more important, what the boundary conditions for these findings are.

Since specifying boundary conditions has all too frequently been passed over or, at most, been paid ambiguous lip service in educational research, I should like to emphasize one point about these results. It is doubtful that attribute cueing will ever be shown to be unconditionally
better than operation cueing or vice versa. What future research may be expected to do, however, is to specify the conditions under which each will be superior.

**Discovery learning**

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, have either failed, 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 (Scandura, 1964b). 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 unambiguously interpret experimental results. 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 not discover the rule.

Gagne and Brown (1961) 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 Gagne and Brown study, Eldredge (1965) hypothesized that the obtained results could have been due to a number of flaws in the programs used. 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 difficult to interpret.

The Set-Function Language was used as an aid in removing these
difficulties (Scandura, 1966a; Scandura and Roughead, 1968). The resulting analysis of what is involved in discovering rules indicates that discovery learners learn 'something' by which they can derive solutions to an entire class of problems. Roughead and I called this 'something' a 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 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 Ss 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 is learned' by guided discovery in the Gagne and Brown study could be identified and, hence, could be presented by exposition. The second question was, how is 'what is learned' by discovery dependent on what the learner already knows and/or 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 or not 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.

**Method**

The task was essentially identical to that used by Gagne and Brown and Eldredge and involved finding formulas for summing the terms in number sequences. That is, the stimuli were number series, like $1 + 3 + 5 + 7$, and the responses were formulas in $n$, the number of terms, for summing such series. For example, the appropriate formula for summing $1 + 3 + 5 + 7 + \ldots + 2n - 1$ 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 Gagne 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 confronted in the learning program; (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) in the *opportunity-to-discover* program, 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 subjects were tested on new series to see how well they could determine the appropriate summing formulas.

Results and discussion

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 we believe that the answer to this question is in the affirmative, then there is 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, has 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 an opportunity to discover and then the formula program performed as well on the transfer problems as those given the derivation rule 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 Ss 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 available 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, must 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. Let me make this point clear, because it is an important one. It is not that meaning should be taught 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. The theoretical implications are even more interesting for the researcher and, in fact, may be crucial to any theory based on the rule construct. More will be said on this topic in the final article in this series on theoretical direction.

References


