AN ANALYSIS OF EXPOSITION AND DISCOVERY MODES OF PROBLEM SOLVING INSTRUCTION

JOSEPH M. SCANDURA
State University at Buffalo

IN RECENT YEARS, discovery methods of instruction have gained increasingly widespread support (2, 11). As justification, it is pointed out that direct verbal presentation frequently leads to a sort of verbal glibness by the learner without true understanding. Many good teachers, nonetheless, still use some form of exposition. By drawing the attention of the student to basic principles, they feel that the educational process is made more efficient. Gagne (7: 50) supports this point of view when he asks, 'Is it [verbalization] bad for problem solving because it is expressed in words, or is it bad because it does not have enough words?'

The literature contains studies which appear to support each of these views. Haselrud and Meyers (9) and Hendrix (10) have found that individually derived (i.e. discovered) principles were better retained and led to more transfer than was the case when the Ss were told the principles directly. The findings of Craig (5), Kittle (13), and Sassenrath (14), on the other hand, indicate that superior learning results when information is given directly. Such apparently contradictory findings can be explained, at least partially, in terms of the inconsistent use of the terms direct and indirect (12). Variables other than "directness" also may be involved. The results of a study by Corman (4), for example, indicate that kind of information and ability both interact with directness.

The situation in the classroom is still more complex. As is becoming increasingly evident to educators (11, 1), exposition (E) and discovery (D) refer to classes of methodology — not uniquely defined methods. In addition, psychologists (8, 2, 17) recognize that known learning principles, by themselves, are inadequate to explain many educational situations. Nonetheless, failure to identify many of the basic variables and interrelationships operating in the classroom has made it difficult, if not impossible, to study the teaching-learning process in a systematic fashion. Furthermore, any such relationships can be of only passing interest to the teacher unless the variables involved are largely under his control. At the current stage of development, it would seem that more educational research should be exploratory in nature. Too much precision, too soon, is at best unrealistic and at worst misleading.

The dual purpose of this research was to help determine some variables and interrelationships which complicate experimental comparisons of exposition and discovery modes of instruction and to provide a framework for future more precise experimentation. The teaching and learning of problem solving in the classroom was of particular concern.

APPROACH

Three studies were conducted in sequence; one form of exposition (E) and one form of discovery (D) were compared in each. In addition to an objective reporting of results, using statistical techniques where appropriate, post-analysis played an important role in the research. The results of completed studies were used to suggest a format for the next.

Tape transcriptions of the instruction were made and measures of specific and non-specific transfer were obtained from specially constructed tests, called routine (R) and novel (N), respectively. The problems on the R-tests were of the same type as those taught while those on the N-tests were based on similar principles, but required modified modes of attack for successful solution. Student attitudes and opinions also were obtained by means of unstructured questionnaires.

The transcriptions together with the test and questionnaire results were analyzed to obtain more insight into the relationships operating between the teaching and learning variables. In each study, the information given by the experimenter was classified as to directness, amount and meaning and its effectiveness was assessed by subjectively analyzing...
S-reactions to that information. Thus, an attempt was made to characterize E and D and to collate judged informational effectiveness with the more objective findings (i.e., test and questionnaire results). In this way it was hoped that specific outcomes might be tentatively attributed to particular aspects of the instruction.

The use of abstract card material (16) facilitated the type of analysis used. Three aspects of this material were identified. The first, called the prerequisite or subordinant learnings, refers to the information necessary to make the problems meaningful. The second refers to the problems themselves and the third to an algorithm for solving the problems. Due to the novel nature of the material, it was feasible to assume the same initial level of relevant learning for all Ss - essentially zero. This assumption, together with the above categorization of the material, made it easier to analyze the growth of knowledge during training as a function of the instruction given. Apparently, its novelty also made it easier to maintain an almost uniformly high level of motivation.

Each method conformed to certain predetermined criterion. In the E-classroom, all of the information necessary for solving the training (T) problems, including the algorithm, was presented directly. Typically, the information was first presented verbally, and then illustrated before the Ss were given an opportunity to practice it. Questions were answered directly in a manner consistent with the form of E under study. In the D-classroom, learning was induced by simple directive statements, questions, and hints. The D-Ss were not told or shown how to solve the problems, nor were any of their questions answered directly. They had to go beyond the information given them. The algorithm was introduced incidentally, in a non-card problem context, after the D-Ss had reached a prescribed level of performance on the problems.

In spite of the restrictions imposed on the E and D instruction, a certain degree of method manipulation was still possible; the first two studies give some indication of the degree to which learning may be affected by such manipulation. To begin with positive, rather than inconclusive findings, Study One was designed so as to favor the D-method.

STUDY ONE

Procedure

Two essentially equivalent and above average sixth-grade classes, each having 23 students, were taught the card material, one by exposition (E) and the other by discovery (D). Each class met for instruction on May 16, 18, and 23, 1961. Testing took place two days later on May 25. Both classes were taught the same material and instruction continued until problem solving performance was judged approximately equal.

Instructional Analysis

The prerequisite aspects of the card material were covered rapidly in the E-class; card problems were introduced before the E-Ss had had much time to practice this subordinate material. As the instruction progressed, the E-Ss became more and more dependent on the teacher's help to solve the increasingly complex R-problems. Although all questions were answered, few, if any, were directed at the underlying principles, especially in the latter phases of the instruction. The experimenter made only brief references to the meaning underlying the problem solving algorithm and the E-Ss seemed to be primarily preoccupied with using the "technique" correctly.

In contrast, the D-Ss were given extensive experience with the prerequisite material before new "ideas" were introduced. Meaning was stressed in the sense that the questions asked and the hints given were directed at the underlying principles. After the D-Ss generally were able to solve the R-problems with good facility and the symbol algorithm was introduced in a non-card problem context, some of these Ss were able to apply the technique to the problems. Apparently, they understood the critical relationships between the algorithm and the problems because they were not shown how to apply it. However, it took the D-class longer to reach the desired level of R-problem facility - 153 minutes, versus 108 for the E-class.

Results

The mean scores for each class on the R and N portions of the test are presented in Table 1. The results favored the D-class, but only the N-score difference was statistically significant (t-test, .01 level).

<table>
<thead>
<tr>
<th></th>
<th>E (n = 23)</th>
<th>D (n = 23)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>12.2</td>
<td>14.9</td>
</tr>
<tr>
<td>N</td>
<td>6.4</td>
<td>12.2</td>
</tr>
</tbody>
</table>

* A perfect score was 36 (no S scored higher than 30).
Relationships between ability, as measured by the CTMM, performance on the R-test, and performance on the N-test were also explored (using scatter diagrams). In the E-class, a positive generally linear relationship was observed between the ability and R-scores. On the other hand, the N-scores were generally low except at the highest levels of ability — here the N-scores varied more widely. A comparison of the R and N-scores indicated similar, but slightly stronger relationships. The N-scores were almost uniformly low except for a few high N-scores in the upper R-score range. The corresponding relationships in the D-class were all positive and generally linear although wide variability was evident. Again, the R-N relationships appeared to be the strongest.

Since using the algorithm involved the use of paper and pencil techniques, class differences in mode of attack were readily detected. The data in Table 2 gives some indication of the relationships observed between methods, mode of attack and problem solving success.

Only two Ss switched from the algorithm category to the non-algorithm group although the Ss generally made less use of paper and pencil in attacking the N-problems, especially in the D-class. With the exception of some unsuccessful Ss who apparently guessed blindly, the E-Ss typically employed the algorithm taught. Nonetheless, only one E-S was able to modify the technique so as to solve an N-problem completely. Slightly less than half (ten) of the D-Ss used some form of paper and pencil technique. Those that did, however, were relatively more successful.

Student comments, obtained after the testing, also were indicative of method. Several E-Ss complained that they “forgot” on the test. One E-S summed up much of the intent, “...the classes were easy because you helped us, but on the test you weren’t (helping us).” The D-Ss indicated more self-reliance and although most felt the tests were harder and covered more (N-problems) than the classwork, many considered them a challenge. The E and D Ss apparently were highly-motivated; only on the N-problems, and more frequently in the E-class, did any of the Ss appear to “give up”.

Discussion

The differential performance of the E and D Ss on the R and N tests may be attributed to anyone or a combination of the factors identified: 1) directness of presentation, 2) emphasis on meaning, 3) amount of problem solving practice, 4) time at which the algorithm was introduced.

Although apparently more inefficient than that used in the E-class, the indirect D-instruction may have promoted more incidental learning — learning how to attack the problems. If so, the D-Ss should have been better able to solve the N-problems. Greater emphasis on learning structure (i.e., meaning) in the D-class might reasonably be expected to lead to similar results. It also can be argued that problem structure can be discovered while applying (i.e., practicing) an algorithm and that lack of sufficient practice in the E-class not only led to greater “forgetting” on the R-test but also contributed substantially to N-performance. Another possibility is that the degree of familiarity with prerequisite material before problems and/or algorithms are introduced (i.e., timing) may be an important factor in determining the transferability of what is taught.

The relationships indicated between the R and N scores may be interpreted in terms of a distinction between associative and conceptual learning (6). Since the E-Ss were shown directly how to apply the algorithm, meaning was not stressed, and “timing” was poor; it is entirely conceivable that many Ss may have simply associated the algorithm with the problems (i.e., memorized them). Thus, they could perform reasonably well on the R-problems without correspondingly good N-performance. Not having been shown how to use the algorithm, success of the D-Ss on the R-problems might reasonably be attributed to more conceptual learning — of which N-performance was the behavioral correlate. This view is generally supported by the mode of attack analysis. All of the E-Ss who had some success on the N-problems were algorithm users. If they didn’t at least learn to apply the algorithm, they just didn’t learn. On the other hand, those D-Ss who did discover and use the algorithm did relatively better than the non-algorithm users on the N-problems. Apparently, such discovery could not be expected without some grasp of the underlying structure.

It was apparent that the obtained results could have accrued from many factors. Rather than begin systematic and more closely controlled experimentation at that time it seemed more appropriate to conduct a second exploratory study differing from the first in several ways. It was hoped that further insights would be obtained with a relatively small expenditure of time.

STUDY TWO

Some of the conditions favoring the D-method were relaxed in Study Two. The prerequisite information given the D-class not only was indirect, but less to the point than that given in Study One and required somewhat more interpretation by the D-Ss. In addition, an attempt was made to reduce instructional time in the D-class by introducing the symbol algorithm earlier in the instructional sequence (relative to Study One). In the E-class, some effort was made to establish a meaningful relationship between the card problems and the symbol algorithm.
Progressively more difficult problems and algorithms were introduced only after the E-Ss had been given an opportunity to become familiar with the prerequisite material.

**Procedure**

The Ss were 46 summer school students who had just completed grades four and five. The available ability measures, although not strictly comparable, seemed adequate to classify the rather heterogeneous group as to high or low ability. A randomized block design was used in selecting the 23 E and D Ss (the grade and ability categories constituted the four blocks).

Each class met five times within a period of 13 days. The first three meetings were devoted to instruction and the other two to evaluation. The first test, consisting of four R and four N problems, was given on the same day that the instruction concluded and seven days later a short retention test (one R and one N problem) and questionnaire was given.

Both classes covered the same material. Identical prerequisite material and training problems were used wherever possible. As in Study One, the instruction continued until the consensus of the experimenter and an observer was that both classes had attained the same training (R) problem facility.

**Instructional Analysis**

Greater emphasis on the prerequisite aspects of the material and on the relationships between the problems and algorithms seemed helpful in the E-class, especially with the older, more intelligent, Ss. These Ss asked questions, directed not only at how the algorithm worked, but why it worked. Even most of those E-Ss who clearly had only a marginal understanding of what was going on apparently kept trying. Giving less specific direction and increasing the rate at which new information was presented, had less desirable effects on the D-Ss. They reacted hesitantly, although sometimes cleverly, almost from the beginning. Apparently the D-Ss took a broader view of the card material than was intended and it was more difficult to maintain the desired direction than it was in the Study One D-class. Insufficient direction, coupled with the introduction of new aspects of the card material before many of the Ss felt comfortable with the prerequisite material, left many D-Ss groping. Symbol algorithms were introduced at various times during each of the three class meetings: but none of the D-Ss apparently related the algorithm to the problems. Problem solving proficiency did improve progressively, but slowly, and a long third lesson (80 minutes) was needed to reach criterion. Although motivation was good, for the most part, interest diminished as the instruction progressed. Many of the D-Ss, apparently, had given up trying to understand and were content to use unsystematic trial and check methods.

It took longer than in Study One to teach the card material. In the E-class, 153 minutes were required while the D-class needed 199 minutes. The greater time required can probably be attributed largely to the fact that the Ss were younger and class heterogeneity was greater.

**Results**

Whereas both the E and D classes had a mean score of approximately 51 on the R-test, the E-class mean of 33 on the N-test was higher than the D-class mean of 21. Perfect scores were 100. No significant differences were found except between ability levels on the R-test. Grade differences were small (5) and no grade or ability differences were found that could be differentially attributed to methods. On the retention test, the E-class R-mean score was 11 percent higher and the N-mean score was 86 percent higher than the corresponding D-class mean scores. The tendency of the score distributions to be either rectilinear or bimodal suggested the inapplicability of an analysis of variance. Relationships between the R and N test scores were generally linear and positive. Tetrachoric estimates of .8 were obtained in both the E and D classes (see footnote 5).

Table 3 indicates that the successful E-Ss, as in Study One, all used algorithms to attack both the R and N problems. The D-Ss, successful or otherwise, all used intuitive (i.e., nonalgorithm) modes of attack. A few D-Ss seemed to find the solutions systematically, but haphazard modes were the rule, particularly on the more complex R-problems and the N-problems. It is possible to obtain a measure of the relative efficacy of the algorithms in attacking problems of different complexity. In order to equate problem solving success only those Ss who scored higher than 75 percent on a given problem were considered. There was a tendency for the algorithm users to perform relatively better than the nonalgorithm users on the high complexity problems and vice versa.

Most of the comments were positive, but several D-Ss indicated that they lost interest after getting behind—"I didn't know what was going on" being a typical statement. Only two Ss from each class solved one particularly novel problem, but their reactions to it were different. The E-Ss found it hard to adapt the algorithm whereas the D-Ss found it "different—fun and a challenge".

**Discussion**

The result with perhaps the greatest implication in the present context, is that the E-Ss performed somewhat better than the D-Ss on the N-problems although R-performance was almost identical. The
corresponding result in Study One had been attributed to some combination of the directness, meaning and timing variables. Although more abstract interaction effects are possible, this result in Study Two suggests that “directness” was not a critical variable. Apparently, solution of the N-problems was only incidentally determined by “learning how to learn” phenomena since the D-Ss in Study Two were forced to learn even more by themselves. The effects of learning to go beyond given information probably cannot be determined from such short term research although the differential S-reactions to the particularly novel problem, previously cited, may provide a clue. In addition, since meaning was given similar emphasis in both classes, the more difficult to specify variable, timing, was perhaps most decisive in promoting transfer.

Presumably, the premature introduction of new information may induce E-Ss merely to memorize relationships between problems and algorithms. Such associations are not likely in the D-class. After their early failure to recognize its potential usefulness, the Study Two D-Ss never did discover how to apply the algorithms even after problem solving facility had increased. Not having been reinforced for trying to relate the algorithm in the beginning, they may have stopped trying. Too large a “gap” apparently results in “no results” – the information goes unused and may actually lead to confusion, inattentiveness, and lowered motivation.

The preceding results also suggest that not only amount of practice but what is practiced must considered by the teacher – attacking a problem from first principles or applying an algorithm. If anything, the D-class solved more problems than the E-class, but the retention results tend to favor the E-class. Perhaps applying the algorithm was a simpler task (and this seems likely) than learning to solve the problem in a more heuristic fashion so that the same amount of practice went further. One last point, suggested by personal interviews with some of the better E-Ss, is that the algorithm concept, itself, was what led to N-problem transfer. That is, the Ss may have discovered an algorithm pattern which was then generalized so as to encompass the new problems. To see that this phenomenon is not specific to the algorithm used, one has only to consider the patterning evident in our ordinary arithmetical algorithms.

Due to conditions of class availability and other practical considerations, Studies One and Two are not directly comparable. This was not deemed crucial, however, since the purpose of these studies was solely to explore. Nonetheless, it was felt desirable to conduct a third, better controlled and more intensive study to determine the feasibility and some of the weak points of the preceding rationale.

STUDY THREE

The preceding results suggest strongly that, if equivalent proficiency is desired, instruction by D generally takes longer than by E. The question arises as to what effect increasing the instructional time in the E-classroom would have on specific and non-specific transfer. Instead of merely serving to increase R-facility and retention, for example, it might be argued that transfer also would increase.

The third study was shorter and involved fewer Ss. This made possible the elimination of some of the more obvious and unwanted confounding factors involved in the previous studies and made practicable a more intensive individual analysis (via the tapes). In addition to differences in instructional time, three other sources of variation were minimized: 1) the effects of outside study and other means of S-interchange of information outside the classroom were largely eliminated by having all of the instruction and testing during one session, 2) the range of individual differences was reduced by utilizing a select and relatively homogeneous population, and 3) the information necessary for making the N-problems meaningful was presented directly just before the N-test was given and in exactly the same way to both classes.

Although the study was not specifically designed to study “timing”, it did serve to demonstrate its role in two relatively well-defined and often used methods of instruction. Another aim was to explore further the effects of problem difficulty and to investigate more closely the modes of attack used.

Procedure

A simplified version of the card material3 was taught to two small classes of seven (E) and eight (D) gifted Ss who had already completed grades four and five. The selection procedure was designed to ensure that each class had its share of the very best and the poorest (not poor) Ss as judged by the university teacher of their special summer school class – otherwise selection was random. The E and D classes met once during which all of the instruction and short two-problem R and N tests were given. In order to explore the role of problem variation, the two R-problems were varied as to complexity and the N-problems as to novelty.

In order to equalize instructional time, the D-class met first. The instruction was designed to aid the Ss find systematic modes of attacking the R-problems (much as was done in Study One) and it continued until the D-Ss discovered how to apply an algorithm to a problem. The E-class met the next day and the same amount of instruction was given. The prerequisite material was presented only briefly before the R-problems and the algorithms were introduced. Application of the algorithm was mechanically illustrated and individual difficulties
were cleared up. Additional practice was planned for the remaining time.

**Instructional Analysis**

In the D-class, S-attention was directed toward the underlying fundamentals and new ideas were introduced only when the Ss were judged able to deal with them effectively. A considerable amount of time, 26 minutes, was spent building readiness before the R-problems were first presented. Then, the experimenter confronted the D-Ss with progressively more difficult R-problems, and questions, directed towards the underlying principles, were asked. After all but two of the eight Ss were solving the problems systematically and with good facility, the algorithm was incidentally introduced and several of the successful D-Ss were able to apply this technique to the next R-problem shown them. The instruction was concluded after the presentation of a second problem.

In the E-class, the prerequisite material was covered in five minutes. Next, the first problem was presented and the desired goal stated. Ten minutes were spent explaining and illustrating the algorithm but with little or no direct reference as to why the algorithm worked. Then, the Ss worked two problems at the blackboard and the teacher helped those two Ss who had some difficulty with the first. Meanwhile, two other Ss had discovered how to perform the algorithm in one step. Finally, the experimenter reviewed the algorithm once again and the last twenty-seven minutes were spent in practicing what was by then a well-learned technique. This practice was motivated, apparently with good success, by making it a race against time—boys versus girls. The total instructional time in both classes was sixty-seven minutes.

**Results**

Due to the small numbers involved and the J-shaped distributions obtained no statistical tests were applied. Mean scores on the R and N problems are given in Table 5. Although the E-Ss did appreciably better on the R-tests and only slightly more poorly in the N-test, other comparisons are revealing 1) Whereas the E-Ss did considerably better on the R-test, the D-Ss performed equally well on both; 2) Even though the second R-problem was judged to be more complex than the first, the E-Ss did almost as well on both while the D-Ss found the second more difficult; 3) Whereas the E-Ss did relatively poorly on the more novel N-problem, the D-Ss did just about as well on both.

A more intensive individual analysis (see Table 6) indicated the following. First, those D-Ss who were successful on the R-problems also were successful on the N-problems. This was not necessarily the case with the E-Ss. Second, the E-Ss were consistent in their use of formal (F), pencil and paper, techniques, and two were successful in abbreviating the algorithm (one step). Six of the D-Ss used formal (F) procedures, all but two of which appeared systematic. Procedures, typically, seemed less formal on the N-problems. Third, the sources of difficulty and error patterns were different in the two classes. The formal procedures, more typically used by the E-Ss, apparently served to reduce the incidence of “careless mistakes”, but three E-Ss were unable to modify the algorithm so as to solve the more novel N-problem. The modes of attack used by the six successful D-Ss apparently were adaptable to the N-problems and indeed led to fewer mistakes than they had when applied to the more familiar R-problems. As previously noted, those D-Ss who did discover and use the algorithm, were among the most successful.

**Discussion**

As suggested in the preceding discussions, transfer is not a simple function of instructional time—it depends on what goes on during that time. Applying an algorithm over and over to more and more problems of the same type may promote efficiency, but it does not necessarily increase transfer potential. That four (out of seven) E-Ss were successful on the N-problems may be attributed to their high intelligence and the simplified material used. It is not unlikely that these Ss were well able to maintain the rapid instructional pace set for them. The high R, low N, pattern was not evident in the D-class. Those D-Ss who attacked the R-problems systematically had relatively little difficulty with the N-problems. Presumably, they were able to do so because they understood the basic principles involved.

The effects of R-complexity and N-novelty on E and D performance add another complication. Although alternate explanations may be devised, parsimony suggests that mode of attack is critical. Use of a well mastered algorithm may serve to reduce what Bruner, Goodnow, and Austin (3) have called “cognitive strain”, thus, making it relatively easier to solve problems with more of the same sort of difficulties. Nonetheless, it is not surprising that the degree to which an algorithm must be modified to solve an N-problem was related to the performance level of the algorithm users. When basic principles remain the same, however, there is less reason to expect more heuristic problem solvers to perform differentially.

Although learning by drill has often been associated with poor motivation, this did not seem to be a factor during the instruction. It may be significant, however, that the unsuccessful E-Ss tended to give up on the N-test, while the two poor D-Ss persevered throughout the ten minute test period. Whether these differences were due to methods, of course,
can not be determined. Nonetheless, it is desirable that a person be able to develop his own means for solving new problems and it is unlikely that he will develop either the ability or the predisposition to do so unless given the opportunity. Long term research is clearly indicated.

Conclusions, Collations and Conjectures

The results of these studies make it difficult to conceive of any explanation of the teaching-learning process that does not consider S-feedback and the nature of the materials used. Perhaps the most striking conjecture to be drawn from this research involved the use of "timing" as a coordinating variable. Although evaluated subjectively in the present context, by analyzing tape recordings, it is felt that as more is learned about its nature more objective measures may become available. The author currently is investigating this possibility in the laboratory. A promising methodology, which makes it possible to infer the current cognitive state of an organism on the basis of S-feedback, has been developed for this purpose. It also is quite possible that timing eventually may be characterized in terms of other variables such as the massing (rate of presentation) of information, the complexity of the material, S-ability and prior relevant knowledge. Nonetheless, such reduction is not necessarily desirable. On the contrary, timing allows us to deal, at once, with the cumulative effects of these variables. In this manner the number of variables effectively operating is reduced considerably. Further, a recent study (18) indicates that teachers frequently fail to give information when S-feedback indicates its necessity and often give information when it is not needed. A better understanding of the role of timing and its effects on learning, retention and transfer may be of practical import in helping make the teaching-learning process more efficient.

Of course, timing is far from the whole answer. Its effects are tempered by other variables. The distinction between associative and conceptual learning, as proposed by Dienes (6), provides a framework for the relationships involved. Generally, it is assumed that the better the timing, the more conceptual the learning and the more transfer will obtain. When problem solving algorithms are represented directly and meaningfully at a point in time when S-feedback indicates good prerequisite comprehension, conceptual learning may be expected. When meaning is not stressed and timing is poor, the algorithm may be simply associated with the problems. In such a situation, practice in applying the algorithm apparently tends to facilitate specific transfer (R-problems) and perhaps retention, but it is not necessarily associated with non-specific transfer. There is a clear tendency, however, for direct presentation, coupled with moderate practice, to induce Ss to apply the algorithm to the N-problems as well as the R. Performance may be affected only slightly by R-problem complexity, but N-problem novelty may have considerable impact when the learning is associative.

Information given indirectly, acts as a catalyst. Instances of the desired concepts are presented and attention is directed so that the learner may abstract for himself. Although the discovery methods employed in this research generally were designed to exhibit the basic structure of the material, the Ss, on occasion, inadvertently were led off the main stream. Apparently, when prerequisite learning is inadequate, indirect information is of little value. Indeed, too early presentation may actually inhibit later discovery when the appropriate learning sets have been developed.

Relatively few of the D-Ss had time or were able to discover how to apply the algorithm, but those that did were generally successful not only on the R-problems but on the N-problems as well. Relatively speaking, those D-Ss who used systematic, yet heuristic, modes of attack seemed to be hindered more by problem complexity than novelty. It is possible that complex problems may overtax the heuristic problem solver's information processing capacity so that he makes mistakes. Novelty may not have this effect when the same well-learned principles are involved.

Any interpretation of the foregoing must take into account: 1) the exploratory nature of this research; 2) the almost uniformly high motivation involved; 3) the artificial material used and 4) the lack of relevant knowledge prior to the instruction.

Nonetheless, this research clearly demonstrates the potentially strong affect of within method differences on behavior as it relates to various educational objectives - specific skills, transfer, and/or attitudes. Comparisons between exposition and discovery necessarily involve the simultaneous variation of several variables and as such cannot be taken as definitive no matter what the results. How, for example, does one equate the timing of information based on S-feedback without also changing the amount of instructional time. Surely the latter variables can be predetermined and the first, at least, subjectively assessed, but to control the first means not to control the others. It is just this difficulty which besets most methodological comparisons. In this sense, global comparisons are design-wise unsound - more intensive research dealing with component method variables is needed.

Methods research has never attained much esteem from behavioral scientists. The facts are, however, that too few of them have concerned themselves with the problems of education. In order to further the scientific aspects of education, we must begin with a prudent selection (or identification) of central and meaningful educational variables - variables which
serve to integrate rather than complicate. If this involves the consideration of difficult to specify variables, then let us learn how to specify them better — this is one of the difficulties besetting any young science. Whenever possible, of course, such variables should be characterized in terms of meaningful laboratory variables. Reductionism is the means by which we relate all science. Both levels of research are needed. Without the former, the latter tends to become sterile and laboratory bound while the latter tends to make the former more precise. It is hoped that the recent educational fervor may lead to more cooperative ventures along these lines.

FOOTNOTES

1. This research is based in part on a dissertation completed in 1962 while the author was an instructor of mathematics at Syracuse University.
2. The author taught all of the experimental classes.
3. In studies one and two the following subset-card relationships (see Scandura, pp. 1-2) were introduced during class time: -all, -all, +1, -1, +2, ..., +6, -6. In study one, twelve three and four card test problems were used. The six routine (R) problems consisted of all and +1 cards; the novel (N) problems consisted of these plus at least two other card types. In study two, eight three-card test problems were used on the immediate test and two on the retention test. Otherwise, their makeup was comparable to those used in study one. In study three, only +all and -all cards were used (they were denoted type A and B, respectively). Only the type A cards were introduced during the instruction. The two routine (R) test problems each consisted of three +all cards with seven and fifteen solution subsets, respectively. The type B subset-card relationship was introduced immediately (in the same way to all Ss) preceding the novel (N) test — which included one problem consisting of three +all cards and another consisting of two -all and one +all cards. Both problems had seven solution subsets.
4. The class means on the California Test of Mental Maturity (CTMM) and the Metropolitan Achievement Test in Arithmetic (MAT) were:

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTMM</td>
<td>118</td>
<td>120</td>
</tr>
<tr>
<td>MAT</td>
<td>8.45</td>
<td>8.44</td>
</tr>
</tbody>
</table>

5. Since the data were only considered suggestive, correlation coefficients were not computed. Tetrachoric estimates (Peters, C.C. and Van-Voorhis, W.R., Statistical Procedures and Their Mathematical Bases, New York, McGraw, 1946, p. 370) of the relatively strong R-N relationships are given for comparative purposes: E-class, r = .5; D-class, r = .7. These estimates may be spuriously high since both R and N scores were generally positively related to ability.

6. Other variables such as total instructional time and amount of information have frequently been used in similar comparisons. The position taken here is that both are derived variables; that is, these variables may be characterized in terms of other more fundamental variables. Total time, for example, may be apportioned in many ways — time spent on meaningful prerequisite material, and practice appear to be more critical in the present context. “Amount”, as used by other authors (12), is similar to “directness” as used here. The latter term is felt to be more descriptive since from a logical point of view the same amount of information may be given directly or indirectly. It may merely be more difficult to interpret or encode information when presented indirectly.

7. Although CTMM scores were available for almost all of the Ss, they were of differing vintage and a “home made” abstract reasoning test was used to supplement them.

8. The fifth graders averaged eight months younger than their counterparts in Study One and the fourth graders averaged still another year younger.

9. In the first two studies this information was introduced as a natural adjunct of all of the other prerequisite material. For details, see Scandura (15, 16).

10. The author is indebted to Roy Callahan for his assistance in this regard (See Table 6).

11. Two adjectives, complex and novel, have been used to distinguish between novel problems. In Study Three, the second problem was judged to be of the same complexity but greater novelty than the first. Complexity was defined in terms of the attention to detail required to solve a problem and novelty in terms of the degree to which the R-algorithm had to be modified. In Study Two, the N-problem singled out in the discussion was of only moderate complexity, but high novelty.

REFERENCES

3. Bruner, J. S., Goodnow, J. J., and Austin,G.
A. A Study of Thinking (New York: Wiley and Son, 1956).

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER OF STUDENTS IN EACH METHOD, MODE AND SUCCESS CATEGORY</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Algorithm</th>
<th>Non-Algorithm</th>
<th>Algorithm</th>
<th>Non-Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above Median*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>12 (12)</td>
<td>0 (0)</td>
<td>7 (7)</td>
<td>5 (5)</td>
</tr>
<tr>
<td>Below Median*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>6 (5)</td>
<td>5 (6)</td>
<td>3 (2)</td>
<td>8 (9)</td>
</tr>
</tbody>
</table>

*Above or below class median
### TABLE 3
NUMBER OF STUDENTS IN EACH METHOD, MODE AND SUCCESS CATEGORY

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th></th>
<th>D</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Algorithm</td>
<td>Non-Algorithm</td>
<td>Algorithm</td>
<td>Non-Algorithm</td>
</tr>
<tr>
<td>Above Median*</td>
<td>12 (12)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>12 (12)</td>
</tr>
<tr>
<td>Below Median*</td>
<td>6 (6)</td>
<td>5 (5)</td>
<td>0 (0)</td>
<td>11 (11)</td>
</tr>
</tbody>
</table>

*Above or below class median.

### TABLE 4
NUMBER OF SUBJECTS IN EACH ALGORITHM-COMPLEXITY CATEGORY

<table>
<thead>
<tr>
<th></th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Algorithm</td>
<td></td>
</tr>
<tr>
<td>R (N)</td>
<td>20 (13)</td>
</tr>
<tr>
<td>Non-Algorithm</td>
<td></td>
</tr>
<tr>
<td>R (N)</td>
<td>27 (14)</td>
</tr>
</tbody>
</table>

### TABLE 5
MEAN SCORES OF EXPOSITION AND DISCOVERY CLASSES ON ROUTINE AND NOVEL TEST PROBLEMS*

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th></th>
<th>D</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>(N)</td>
<td>R</td>
<td>(N)</td>
</tr>
<tr>
<td>Low Complex (Novel) Problem</td>
<td>100</td>
<td>(86)</td>
<td>82</td>
<td>(77)</td>
</tr>
<tr>
<td>Medium Complex (Novel) Problem</td>
<td>96</td>
<td>(59)</td>
<td>71</td>
<td>(75)</td>
</tr>
<tr>
<td>Overall</td>
<td>98</td>
<td>(72)</td>
<td>76</td>
<td>(76)</td>
</tr>
</tbody>
</table>

*A perfect score was 100.
## Table 6
### Summary of Individual Results

<table>
<thead>
<tr>
<th>Student</th>
<th>Ability Rating</th>
<th>Student Comment</th>
<th>Routine Low Complex</th>
<th>Routine Medium Complex</th>
<th>Novel Low</th>
<th>Novel Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>E 1</td>
<td>Exceptional</td>
<td>liked what and how taught</td>
<td>100 F (1 step)</td>
<td>100 F (1 step)</td>
<td>100 F (1 step)</td>
<td>100 F (1 step)</td>
</tr>
<tr>
<td>E 2</td>
<td>Exceptional</td>
<td>Class at first difficult, N-problems can be solved with thought</td>
<td>100 F</td>
<td>100 F</td>
<td>100 F</td>
<td>100 F</td>
</tr>
<tr>
<td>E 3</td>
<td>High</td>
<td>“I liked everything!”</td>
<td>100 F (1 step)</td>
<td>93 F (1 step) Omission</td>
<td>100 F (1 step)</td>
<td>100 F (1 step)</td>
</tr>
<tr>
<td>E 4</td>
<td>High</td>
<td>class interesting</td>
<td>100 F</td>
<td>87 F Omission</td>
<td>100 F</td>
<td>100 F</td>
</tr>
<tr>
<td>E 5</td>
<td>High</td>
<td>“I soaked up a lot”</td>
<td>100 F</td>
<td>100 F</td>
<td>100 F</td>
<td></td>
</tr>
<tr>
<td>E 6</td>
<td>Moderate (Young)</td>
<td>liked class and material, R-test easy, N-test hard</td>
<td>100 F</td>
<td>100 F</td>
<td>100 F</td>
<td>O F - Started procedure, couldn't finish</td>
</tr>
<tr>
<td>E 7</td>
<td>Moderate</td>
<td>liked class, R-test easy, N-test hard</td>
<td>100 F</td>
<td>93 F Omission</td>
<td>O F - Started procedure, couldn't finish</td>
<td>O F - Started procedure, couldn't finish</td>
</tr>
<tr>
<td>D 1</td>
<td>Exceptional</td>
<td>novel N-problem challenging</td>
<td>100 F (1 step)</td>
<td>100 F (1 step)</td>
<td>100 F (1 step)</td>
<td>100 I</td>
</tr>
<tr>
<td>D 2</td>
<td>Exceptional</td>
<td>just needed to write answers (no algorithm)</td>
<td>100 F (1 step)</td>
<td>100 I (systematic)</td>
<td>100 I (systematic)</td>
<td>100 I (systematic)</td>
</tr>
<tr>
<td>D 3</td>
<td>High</td>
<td>“could” use algorithm on N (“easily”)</td>
<td>100 F (1 step)</td>
<td>93 F (1 step) Omission</td>
<td>100 F (1 step)</td>
<td>100 F (1 step)</td>
</tr>
<tr>
<td>D 4</td>
<td>High</td>
<td>none</td>
<td>100 I (systematic)</td>
<td>93 I (systematic) Omission</td>
<td>100 I (systematic)</td>
<td>100 I (systematic)</td>
</tr>
<tr>
<td>D 5</td>
<td>High</td>
<td>“fun”</td>
<td>86 I (systematic) Omission</td>
<td>87 I (systematic) Double Omission</td>
<td>100 I (systematic)</td>
<td>100 I (systematic)</td>
</tr>
<tr>
<td>D 6</td>
<td>High</td>
<td>did problems in head</td>
<td>100 I (systematic)</td>
<td>47 I (systematic) Many omissions</td>
<td>100 I (systematic)</td>
<td>100 I (systematic)</td>
</tr>
<tr>
<td>D 7</td>
<td>Moderate (Young)</td>
<td>liked; thought first, then checked</td>
<td>42 I (unsystematic) multiple omissions</td>
<td>33 I (unsystematic) multiple omissions</td>
<td>28 I (unsystematic) guessing, inadequate checking</td>
<td>0 I (unsystematic) guessing, inadequate checking</td>
</tr>
<tr>
<td>D 8</td>
<td>Moderate</td>
<td>“liked”</td>
<td>28 I (unsystematic) multiple omissions</td>
<td>13 I (unsystematic) multiple omissions</td>
<td>0 I (unsystematic) guessing, inadequate checking</td>
<td>0 I (unsystematic) guessing, inadequate checking</td>
</tr>
</tbody>
</table>