ALGORITHMS EXIST for solving many types of problems. The step-by-step computational procedures used in arithmetic perhaps provide the most familiar examples, but algorithms are also used in dealing with all sorts of practical and theoretical problems — from “trouble shooting” to mathematical physics. A common feature of such procedures is that they can be applied mechanically without “understanding”.

Many educators and subject matter specialists would maintain that algorithms are limited in their usefulness to specific problem situations and that transfer to variants of the original problems requires an understanding of underlying principles. The fact that understanding is typically defined in terms of performance on some transfer task (8), however, poses a problem for researching such conjectures. To operationally define understanding in terms of problem solving transfer, would clearly lead to circularity. One way to overcome this difficulty is to operationally define “understanding” in terms of the amount of information presented.

This study was designed to determine whether given information about problem structure and/or subtasks is necessarily reflected in problem solving performance. It was hypothesized that such information facilitates algorithm learning only when the information is specifically needed to perform the algorithm. Other information about problems reflects itself only to the extent that the algorithm is inadequate to solve transfer problems and to the extent that the information provides a basis for additional positive transfer to such problems. In particular, definition of the problems themselves may have little effect on transfer whereas generalization of the algorithm taught may provide a profitable mode of attack.

METHOD

Materials and Subjects

The materials were based on some abstract materials (4) developed for use in problem solving research. The material was based on an extension of card tasks frequently used to study concept formation (1). Sixty-four cards were used. Each card had two objects on it, six two-valued attributes in all. Corresponding to each of the 12 attribute values was a familiar symbol. Each card was assigned one of several possible labels. A symbol set was said to be associated with a given labeled card if the symbols in the set were related to the card properties in a specified way. A card problem consisted of determining all symbol sets associated with each card comprising the problem. There was an algorithm for solving each of the many possible problem types. Problem types differed in the card labels used, and/or the algorithm pattern and number of solution sets.

For the purposes of this experiment, four levels of the material were identified. At level one, (Symbol) were the 12 one-to-one correspondences between the card properties and symbols which represented those properties. A common principle relating each property to the corresponding symbol. For example, S1 and T2 stood for the first object is Small and the second is a Triangle, respectively. Level two (Criterion) consisted of labeling the cards (either +all or +1) and defining corresponding (criterion) principles by which it was possible to decide, for each set of symbols, whether it was or was not associated with any given labeled card. A set of symbols was associated with a +all (+1) card if, and only if, all (exactly one) of the symbols corresponded to the card properties. Level three (Problem) consisted of defining the problems (i.e., finding all symbol sets associated with each of the three labeled cards comprising a problem) and pointing out some of the relationships between an illustrative problem (see R problem, following) and its solution. For example, every symbol used in writing the solution sets corresponded to some property of each of the +all cards and the symbol which corresponded to the +1 card property appeared in each solution set. In no case was information given as to how the solution was obtained. Level four (Algorithm) consisted of an efficient algorithm for solving the R problems.
There were four operations involved in each of the algorithms used: 1) writing those six symbols corresponding to the properties of the first (+all) card, 2) crossing out some of these symbols according to the second card, 3) encircling and coding one or more of the remaining symbols according to the third card, and 4) writing all solution sets on the basis of the algorithm pattern resulting from the first three operations.

In addition to the R problems, two other types were considered. The routine generalization (RG) problems also consisted of two +all cards and one +1 card, but successful RG problem solving required a relatively minor modification of the algorithm taught (i.e., the algorithm pattern was either \(-000\), \([1]\), or \(-000\), \([1]\), as opposed to the \(-000\), \([1]\), pattern used for the R problems\(^2\)). The novel (N) problems each consisted of a +all, -all, and +2 card. The algorithm patterns were \(-000\), \([2]\), and \(-000\), \([2]\). Successful RG performance depended primarily on the valid generalization of the third and fourth algorithm operations whereas N problem solving involved modification of the second operation as well.

A statement of the -all and +2 N criteria, each with one associated symbol set as an illustration, also was prepared.

The level one material was used in the description of all of the other levels, including the algorithm; the level two material was used only in describing the level three material; but, neither the level three nor level four material was prerequisite to any of the other levels. The R algorithm could be performed without knowledge of either level two or level three. The levels were related to each other and to the test problems as indicated in Figure 1.

Since the structural relationships indicated in Figure 1 played a central role in the present research, the bases for these analyses deserve further explanation. Two of the relationships indicated, use of the material in presenting new material and sufficiency of the algorithm, have a largely objective basis. The former relationship was indicated only in those cases where the terms (i.e., words and symbols), denoting the notions described or defined at one level, were actually used in the presentation of the higher order material. An example was provided by the use of the term “symbol-set” to describe and illustrate the +1 criterion. Sufficiency was indicated in that the algorithm provided a sequence of mechanical steps by which any R problem could be solved. Other than reading skills and other general abilities only knowledge of the symbol-property relationships, introduced at level one, was necessary to follow the procedure. Transfer potential was more subjectively based. Judgments were made as to whether the levels involved were related by a common principle(s). Learning the subtask above the arrowhead (see Figure 1) could be facilitated by prior acquisition of the common principle(s) during the learning of the other subtask. The principle relating the +1 and +2 criteria at level two and N criteria, respectively, might be stated, “the number tells how many symbols must agree.” In addition, knowledge of the +1 criterion (level two) could have facilitated modification of the third operation in the RG algorithm. This encircling and coding operation depended on the number of properties the +1 card had in common with the eligible (i.e., uncrossed) symbols. The N criteria (-all and +2) were essential to both the second and third operations in N problem solving (via algorithm). A final relationship also needs clarification, the possibility of solving the problems via non-algorithm (heuristic) modes of attack. Without the algorithm or the level three material there would be little possibility of even partially solving the problems. In the latter case, the problems would be undefined. Although it might be possible to infer what the problem is with less information, previous research (4, 5) with these materials suggests not only that this is not to be expected, but that heuristic modes of attack are far less efficient than those involving the algorithm.

The experimental materials consisted of a one-half to two page double spaced introduction to each level of the material, a number of practice problems (enough to keep even the fast Ss busy throughout the corresponding practice period) with answers on the following page, and some routine arithmetical and algebraic factoring problems which served as filler. The Ss were 84 11th grade college preparatory mathematics students at the Kenmore East High School in Kenmore, New York. None of these Ss was in an honors section.

**Design and Procedure**

The amount of prerequisite material was varied. After having been shown and having practiced different amounts of relevant prerequisite material, each of four groups was presented with the algorithm and given an opportunity to practice using it. The total time spent on the algorithm was 17 minutes. To control for possible massing and fatigue effects, all Ss had approximately 33 minutes of preliminary work (including the prerequisite instruction and/or filler material) before the Algorithm was presented. Group A was given 39 minutes of filler. Group SA had eight minutes to learn and practice the level one (Symbol) material and 31 minutes of filler. Group SCA had eight and 14 minutes to learn and practice the level one and level two (Criterion) material, respectively, and 17 minutes of filler. Group SCPA had eight, 14, and 17 minutes to learn and practice the level one, two, and three (Problem) material, respectively. A table of random numbers was used to assign the 84 Ss equally to the four treatment groups.

All of the materials were presented by means of dittoed booklets and the instructions were given by tape recorder for uniformity. The presentation at each level consisted of both a concise statement of
FIGURE 1

RELATIONSHIPS BETWEEN THE R, RG, AND N PROBLEMS AND THE PREREQUISITE MATERIAL

\[ R \leftarrow \text{LEVEL THREE} \]
\[ \text{ALGORITHM} \]
\[ \text{LEVEL TWO} \]
\[ \text{LEVEL ONE} \]

\[ RG \leftarrow \text{LEVEL THREE} \]
\[ \text{ALGORITHM} \]
\[ \text{LEVEL TWO} \]
\[ \text{LEVEL ONE} \]

\[ N \leftarrow \text{LEVEL THREE} \]
\[ \text{ALGORITHM} \]
\[ \text{LEVEL TWO} \]
\[ \text{LEVEL ONE} \]

--- INDICATES POSSIBILITY OF SOLVING PROBLEMS VIA NON-ALGORITHM (HEURISTIC) MODE OF ATTACK

---- INDICATES POTENTIAL POSITIVE TRANSFER BY MEANS OF STRUCTURAL SIMILARITIES OR RELEVANCY OF INFORMATION FOR SUCCESSFUL PROBLEM SOLVING

----- INDICATES USE OF MATERIAL IN PRESENTING NEW MATERIAL

--------- INDICATES SUFFICIENCY OF ALGORITHM FOR (R) PROBLEM SOLVING
and questions or problems based directly on the material. Filler material was used to replace entire levels, not merely the practice period within a level. All Ss were run at the same time. Three different classrooms were used, but each treatment group was equally represented in each room.

Immediately after the learning period, all Ss were given ten-minute R, RG, and N tests, each consisting of two problems. Reliability estimates of .90, .61, and .65, respectively, were obtained by computing the correlation coefficients between the correct scores on the two problems in each test. The problems were randomly designated "X" or "Y" so that half of the X(Y) scores were from problem one and half from problem two. Three testing orders were used in each treatment group: (R, RG, N), (RG, N, R), and (N, R, RG). The -all and +2 criteria were presented just prior to the N testing (3 minutes). Total learning and testing time (98 minutes) was the same for all treatment groups.

RESULTS AND DISCUSSION

Differences due to class and testing order were minimal, so that these factors were ignored in the subsequent analyses. The only tendency noted was that RG (and N) performance was slightly enhanced (not significantly) when the R (and RG) problems came first, possibly due to learning during the R (and RG) test period(s).

Due to the non-normality evident within the treatment groups on most measures and heterogeneity of some of the variances, the conservative Welch approximation test (9: 37-38) was used to test the hypothesized relationships.

R Problems

The number of correct solution sets and the total number of sets offered (including incorrect sets) ranged from zero to 16 (100 percent). Time spent on the R problems ranged from one and three-fourths minutes to the full ten minutes allowed. A summary of all means and standard deviations is given in Table 1.

The difference between the A and SA groups (number correct) on the immediate R test was highly reliable (t = 5.248, df = 20, p < .001) whereas none of the SA, SCA, and SCPA paired-comparisons approached significance (t < 1). There were 19, 1, 0, and 3 zero scores and 1, 7, 9, and 5 perfect -all -16 -scores in the four groups, respectively. These results conformed to expectations. According to Figure 1, the level one material was necessary for successful application of the algorithm; levels two and three were not.

Twelve A Ss offered no solution sets; all but one of the remaining Ss offered some. There were six, 11, 16, and 16 Ss, respectively, who offered 16 sets (the total number of spaces present on the answer sheet). These results are not surprising since the symbol meanings had not been presented to the A Ss. Even the SA Ss had had little practice in writing symbol sets, and none with labeled cards, prior to level A. Fifteen of the A Ss spent the full ten minutes on the R test whereas four, two, and two of the other Ss did, respectively.

As can be deduced from Table 1, there were some significant time differences. The A Ss consistently took the most time to complete the tests. The lack of significant differences between the other groups, even on the RG and N tests, suggests that the time scores may have reflected more than degree of learning. In particular, amount of information about the problems may have interacted with the success achieved on the algorithm level practice problems to produce differential test motivation levels.

The algorithm was relatively new to the SA (SCA) Ss so that extinction, with respect to the card material, may not have progressed as far, for example, as in the SCA (SPCA) group (for the relatively unsuccessful Ss). The time scores could have reflected such motivational effects, together with degree of learning, making interpretation difficult.

RG Problems

Overall, the Ss performed better on the R problems than on the RG problems. Whereas no S achieved a perfect RG score - 14, almost as many Ss wrote a symbol set in each of the 16 spaces provided as did on the R problems. As suggested in Figure 1, the A-SA difference (number correct) was significant (t = 4.033, df = 20, p < .001), but none of the paired-comparisons between the SA, SCA, and SCPA means were. The SA-SCA difference, however, was suspect (t = 1.209). According to the method-section-discussion, such an effect would have been attributed primarily to familiarization with the +1 criterion.

N Problems

A somewhat different pattern of results was obtained with the N problems. Whereas the A-SA difference (number correct) was not significant (t = 1.436), the SA-SCA difference was (t = 2.148, df = 20, p < .05). The SCA mean was slightly greater than the SCPA mean, but again the difference was not significant.

Although all of the hypothesized effects were not significant, Figure 1 adequately represents the magnitude and direction of the N mean differences. The SCA (and SCPA) Ss had three potential positive transfer routes whereas the SA Ss had only two. Apparently, learning the +1 criterion significantly facilitated +all transfer learning.

The number of sets offered and the time taken to complete the N tests paralleled the results on the other tests. The differences, however,
TABLE 1
SUMMARY OF MEANS AND (STANDARD DEVIATIONS)

<table>
<thead>
<tr>
<th></th>
<th>Correct</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>RG</td>
<td>N</td>
<td>R</td>
<td>RG</td>
</tr>
<tr>
<td>A</td>
<td>.81</td>
<td>.24</td>
<td>.14</td>
<td>5.19</td>
<td>5.67</td>
</tr>
<tr>
<td></td>
<td>(3.52)</td>
<td>(.54)</td>
<td>(.48)</td>
<td>(7.28)</td>
<td>(7.57)</td>
</tr>
<tr>
<td>SA</td>
<td>9.05</td>
<td>2.62</td>
<td>.67</td>
<td>13.19</td>
<td>11.67</td>
</tr>
<tr>
<td></td>
<td>(6.29)</td>
<td>(2.65)</td>
<td>(1.62)</td>
<td>(4.32)</td>
<td>(5.33)</td>
</tr>
<tr>
<td>SCA</td>
<td>10.24</td>
<td>3.67</td>
<td>2.29</td>
<td>14.62</td>
<td>14.05</td>
</tr>
<tr>
<td></td>
<td>(6.43)</td>
<td>(2.97)</td>
<td>(3.05)</td>
<td>(3.07)</td>
<td>(3.43)</td>
</tr>
<tr>
<td>SCPA</td>
<td>9.29</td>
<td>3.95</td>
<td>2.19</td>
<td>15.33</td>
<td>13.57</td>
</tr>
<tr>
<td></td>
<td>(5.87)</td>
<td>(3.22)</td>
<td>(3.06)</td>
<td>(1.50)</td>
<td>(3.31)</td>
</tr>
</tbody>
</table>

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>RG</td>
<td>N</td>
<td>R</td>
<td>RG</td>
<td>N</td>
</tr>
<tr>
<td>A</td>
<td>8.37</td>
<td>8.13</td>
<td>8.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.70)</td>
<td>(2.65)</td>
<td>(1.91)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SA</td>
<td>5.20</td>
<td>5.11</td>
<td>7.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.77)</td>
<td>(2.25)</td>
<td>(1.80)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCA</td>
<td>4.79</td>
<td>6.06</td>
<td>7.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.41)</td>
<td>(2.67)</td>
<td>(2.52)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCPA</td>
<td>5.60</td>
<td>5.14</td>
<td>7.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.17)</td>
<td>(2.14)</td>
<td>(2.08)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These results demonstrate that successful problem solving does not necessarily depend on an understanding of the problem involved. Only the SCPA Ss were presented with a definition of the problems, but this information was not reflected in problemsolving performance. Perhaps even more surprising was the finding that transfer also did not depend on “understanding”. Of course, it was conceivable that the successful non-SCPA Ss may have inferred what the problems were during the testing and/or the level four presentation. If so, none of the Ss were able to state any reasonable facsimile of the level three definition. Judging from the comments made, it was more likely that the successful Ss had detected relationships among the algorithms and were able to modify them according to the syntactic constraints present between the algorithm and the individual problem characteristics. For example, the -all label may have induced the Ss to cross out those symbols which corresponded to the card properties rather than to eliminate symbols as when the opposite -all label appeared. Similar constraints could have made it possible to deal effectively with the RG problems.

What is meant by solving a problem “withunderstanding” as opposed to “mechanically” has never been adequately specified. Whereas degree of understanding may often be defined in terms of a generalized sort of transfer, this was clearly not sufficient with the card problems. Should understanding reflect the degree to which problem elements and the goal are meaningful in terms of S’s prior knowledge (as indicated, for example, by the level three definition) or should it reflect the attainment of pattern constraints among the problem-solving procedures? These are different, yet equally valid, conceptions of understanding. The major implication, of course, is that the assessment of one or the other of these “kinds of understanding” depends critically on the transfer measure used. As suggested by Cronbach (2), global assessment involves making hazardous inferences and can only confound such issues, never clarify them. Subject matter, particularly language and mathematics, generally involves both semantic and syntactic constraints. A limitation of this study was that no attempt was made to explicitly separate these factors. Further clarification of the distinctions between semantic and syntactic constraints and the development of techniques for determining the acquisition of such abilities would be an important milestone in educational science.

Another facet of this research deserves mention. These results demonstrate the feasibility of predicting problem-solving performance on the basis of a subjective analysis of the structural relationships between the criterion task and the information presented. The general approach does not appear to be limited to problem solving. There is no intrinsic reason why similar analytical procedures could not be applied to the learning of new material as well. In addition, there is no reason why the S’s background knowledge could not equally well have been...
assessed. Intensive, rather than extensive, testing procedures could be used to determine what is known that might relate to each class of problems. Combined with structural analyses, such information could then be used to provide information as to how Ss might perform on the criterion tasks. Again, more precise prediction awaits a more complete understanding of just what is involved in such assessment and analyses.

The approach used in the present study to determine structural relationships was reminiscent of task analysis procedures which have been popularized, in education, by Gagné (3). Instead of asking, however, "What would S need to know in order to perform this task?" as Gagné did, the present form of analysis was concerned with the question, "Could this information facilitate problem solving, and, if so, how?" At present, both sorts of analyses can be evaluated only in terms of how well they predict the data. It would be highly desirable to have an independent measure.

The form of structural analysis used was at once more general and less precise than task analysis. It was more general in that it provided for prediction even though the information presented did not appear to be sufficient for successful criterion performance and less precise in that no attempt was made to identify molecular subtasks. Although task analysis is currently being employed in devising new science curricula (7), a more molar approach may be more feasible. It is obviously impossible to include in a curriculum everything S needs to know about a particular topic. Relatively simple structural analysis procedures may make it possible for the educational psychologist, in conjunction with subject matter specialists, to explore the problem-solving implications (assuming problem-solving ability is a desired objective) of a particular set of curricular materials before the costly tasks of formal evaluation and implementation are undertaken. Of course, this approach may lend itself more naturally to some areas, like mathematics, than to others.

SUMMARY

Four groups of 21 Ss were presented with an algorithm (A) and tested on routine (R), generalization (RG), and novel (N) problems. The A Ss had no prior training; the SA Ss were presented with information (level S) deemed necessary for learning the R algorithm; the SCA Ss also were presented with information (level C) deemed useful in modifying the algorithm (so as to solve the RG and N problems); the SCPA Ss, in addition, had the problems defined and were presented with relationships between an illustrative R problem and its solution (level P). The SA Ss performed better than the A Ss on the R and RG problems (p < .001); there were no other significant increases due to amount of information presented. Only the SA-SCA RG difference resulted in a greater than 1. On the N problems, only the SA-SCA difference was significant (p < .05), the A-SA difference resulted in a t of 1.44. Successful problem solving did not depend on an understanding of the problem involved. Transfer was attributed to learning syntactic constraints (principles) relating the algorithm and the problem characteristics. The results also demonstrated the feasibility of predicting problem solving performance by subjectively analyzing structural relationships between the criterion and the information presented. Implications were discussed.

FOOTNOTES

1. This research was supported by the Cooperative Research Program of the U. S. Office of Education, Project S-097 (6). Thanks are due Robert Magee for his help with both the preparation of the materials and the conduct of the experiment and Richard Eiswirth for help with the data analysis.

2. Dashes (—) represent symbols used in solution sets and encircled dashes (•) represent those symbols that correspond to +1 or +2 card properties. The numbers in brackets indicate the number of encircled dashes (i.e., symbols) in each solution set. For more details see the article by Scandura (4).

REFERENCES