Journal of Educational Psychology
1978, Vol. 70, No. 3, 363–371

Broader Transfer Produced by Guided Discovery of Number Concepts with Preschool Children

Aletha Solter and Richard E. Mayer
University of California, Santa Barbara

In two experiments, 43 preschoolers learned the concept of one-to-one correspondence to identical behavioral criteria by matched discovery, expository, or observation methods of instruction. Performance on a subsequent post-test, administered by a "blind" experimenter, revealed a pattern in which the groups did not differ on short-term recall and near transfer, but the discovery group excelled on far transfer (conservation) and delayed recall. The effect of guided discovery on the acquisition of broader learning outcomes was discussed.

Discovery methods of instruction received much attention during the 1960s. The promises of "meaningful" learning outcomes, superior transfer, and longer retention were particularly attractive to educators (Bruner, 1961; Shulman & Keislar, 1966). However, consistent empirical support for the claims failed to materialize (Wittrock, 1966), and there has been a lack of agreement on how to define the concept of discovery or relate it to a useful theory of instruction (Strike, 1975).

More recently, Mayer (1975) has suggested a theory of instruction based on the idea that meaningful learning depends on the satisfaction of at least three conditions: (1) reception—the learner must be presented with and pay attention to the to-be-learned material, (2) availability—a set of related experiences must be available in the learner's long-term memory to serve as an assimilative set, and (3) activation—the assimilative set must be actively processed during learning. According to this view, discovery methods of instruction might be supposed to have their main effect on Condition 3 by encouraging subjects to actively search and process their existing meaningful knowledge and relate it to ongoing learning.

This research was supported by a Faculty Research Grant from the Academic Senate of the University of California, Santa Barbara, and by Grant SED-77-19875 from the National Science Foundation.

Requests for reprints should be sent to Richard E. Mayer, Department of Psychology, University of California, Santa Barbara, California 93106.

There are several situations in which discovery methods of instruction might fail to achieve the goal of broader learning outcomes. Even though the subject may actively search his existing knowledge and try to relate it to what is presented (Condition 3), if the subject fails to discover the to-be-learned principle (Condition 1), no learning can occur. This situation provides one interpretation of the many instances in which "guided discovery" results in more learning or better transfer than pure discovery (Forgus & Schwartz, 1957; Gagné & Brown, 1961; Wittrock, 1963). The present study attempted to overcome this problem by providing that all subjects (in discovery and in expository groups) learned to the same behavioral criterion of mastery, that is, by providing that all subjects achieved Condition 1.

A second situation in which discovery techniques might fail to deliver broader transfer and retention occurs if the discovery subjects do not have a meaningful set of related past experiences (Condition 2). Discovery methods that presumably encourage active search of existing knowledge (Condition 3) will be of little value if no useful related knowledge exists in memory (Condition 2). For example, Mayer, Stiehl, and Greeno (1975, Experiments 3 and 4) found that although all subjects could learn to solve simple binomial probability problems by discovery, only the group that received pretraining in basic underlying concepts performed better on a subsequent transfer test involving integrative problems. In addition, Egan and Greeno (1973) obtained Attitude × Treatment interactions in which the test performance of rule-method subjects was not affected by their scores on pretests for prerequisite concepts, while the test performance of discovery learners was positively related to their level of prerequisite concepts. These results suggest that the availability of prerequisite concepts (Condition 2) is essential for discovery learning but not for rule methods. In order to overcome this problem in the present experiment, a task (one-to-one correspondence) was chosen for which children in the sample were likely to have had some prerequisite concepts (such as previous playing with concrete objects and one-to-one matching in everyday life).

Finally, even if Conditions 1 and 2 are met (as noted above), discovery techniques will not produce different learning outcomes than expository methods if expository subjects are able to actively search and integrate old knowledge with new (Condition 3). For example, this third situation is consistent with Ausubel's (1968) claim that there may be many cases in which expository instruction can lead to meaningful learning. There is some evidence that instructional methods that serve to activate a learning set (Condition 3) may be particularly important for children, since they may not have developed assimilative learning strategies for expository instruction. In the present experiment, preschoolers are used as subjects in an attempt to assure that subjects will not generally have developed strategies for actively integrating information that is presented by expository methods.

Several investigators have studied the effects of discovery methods for instructing children in mathematical concepts (Anastasiov, Sibley, Leonhardt, & Borisch, 1970; Olander & Robertson, 1973; Peters, 1970). The results, as with those cited above, are contradictory. Much of the discrepancy may be accounted for by the non-uniform manner in which the learning outcomes have been evaluated. Also, the concept of discovery itself is rather vague and has included a wide range of very diverse instructional strategies.

Besides discovery training, another manner in which the assimilative set might be activated during learning is to allow the learner to manipulate concrete materials. Piaget (1965) has claimed that mathematical concepts can only be learned if the subject has an opportunity to manipulate real "concrete" objects. A major problem confronting the discovery issue is to clearly separate "discovery" from "active manipulation" in order to ascertain the contributions of each to learning.

In order to investigate the effect of discovery training and concrete manipulation, preschool children were given training in one-to-one correspondence. In the first study, a discovery method was compared to a matched expository method, both of which involved active manipulation of objects. The second study replicated the first study using different materials and included a third group (observation training) in which subjects did not manipulate objects. The learning outcomes were evaluated by tests of short-term recall, near transfer, long-term recall, and far transfer. This study attempts to reconcile some inconsistencies in the discovery learning literature by carefully measuring the learning outcome and by using a situation in which discovery subjects learn to a mastery criterion (Condition 1), are likely to possess some prerequisite concepts (Condition 2), and are not likely to otherwise actively integrate old and new material (Condition 3).

The assimilation theory cited above suggests several predictions that were tested in the present study. Discovery subjects should connect the new skill (one-to-one correspondence) with existing concepts, while expository subjects might simply add the new behavior without integrating it. Since both groups learned to the same criterion, performance on short-term retention should be similar for both groups; however, the broader learning outcome of the discovery group should result in superior performance on problems requiring far transfer of the learned material to novel situations and on long-term retention. An alternative hypothesis is that since both groups reach the same level of learning, they have learned the same thing and should perform similarly on all tests. In addition, the present experiments will provide information on whether training with manipulation (discovery and expository training) results in broader learning than training involving no manipulation (observation training).

Experiment 1

Method

Subjects. The subjects were 19 children between the ages of 3 years 6 months and 5 years 1 month, who attended a private nursery school near Santa Barbara, California and who failed a pretest for one-to-one correspondence (out of a larger group of 62 pretested children). They came primarily from white, middle-class, and upper middle-class homes; written parental permission was obtained for all subjects.

Design. Subjects were divided into two groups, with 10 subjects in the discovery training group and 9 subjects in the expository training group. All subjects were tested on the same four posttests, so that comparisons by type of posttest are within-subjects comparisons.

Procedure. Each subject participated individually in four sessions, sitting opposite the experimenter at a small table in the nursery school. Following each session, a small colored star was given as a reward for participating.

In Session 1, a pretest for one-to-one correspondence was administered to all subjects. Those who passed were eliminated from the study. In Session 2 (2 days later), those who failed the pretest were given one of two training programs for one-to-one correspondence. Subjects were grouped in pairs with an attempt to equate pairs for age and sex. Members of each pair were then randomly assigned to each of the two training methods. In Session 3 (5 to 7 days after Session 2), five recall tests (short-term recall) and five near-transfer tests were administered. In Session 4 (14 days after Session 3), three recall tests (long-term recall) and a test for conservation of number (far transfer) were administered. The experimenter who administered the tests did not know which training the subjects had received.

Materials. The materials consisted of 12 red and 12 blue poker chips (3.75 cm in diameter), 12 small red poker chips (approximately 1.75-cm diameter), 12 wooden blocks (2.5-cm sides), and a 25 × 35 cm piece of white cardboard separated lengthwise by a 1-cm ridge. This board was always placed in such a manner that the ridge was parallel to the edge of the table. In addition, standard data forms and a package of colored gummed stars were used.

Pretest. For the pretest, 6 blue poker chips were placed on the experimenter's side of the board (the side farthest from the subject), next to the ridge, and equally spaced. The subject was given the red chips with the instructions, "Now you put just as many red poker chips on your side. Make it so there are just as many red ones as blue ones." Following the subject's response, a second trial with 7 poker chips was given. The criterion for passing the pretest was a correct response on either or both of the trials.

Expository training. One blue poker chip was placed on the experimenter's side of the board next to the ridge, with instructions for the subject to watch carefully. One red poker chip was then placed directly opposite the blue one, and the experimenter said, "See, I'm putting just as many poker chips on your side as there are on my side." The red one was then removed, and the subject was asked to repeat what the experimenter had done. If the subject did not correctly place a red chip opposite the blue one, the experimenter demonstrated again. This continued until the subject responded correctly. The procedure was then repeated with 2, 3, 4, 5, 6, and 7 poker chips. Each error was corrected by the experimenter demonstrating again how to place the red chips. The training ended after the subject had correctly matched 7 blue poker chips with 7 red ones.

Discovery training. One blue poker chip was placed on the experimenter's side of the board next to the ridge, with the instructions, "Now you put just as many red poker chips on your side." If the subject did not do it correctly, then the experimenter demonstrated, as in the expository training. After one chip was correctly matched, two chips were presented.

If the subject did not correctly match 2 chips, the experimenter did not demonstrate immediately but returned to 1 chip, letting the subject match 1 chip again. If the subject still did not match 2 chips correctly the second time, then the experimenter demonstrated how to do it. This same procedure was then repeated with 3, 4, 5, 6, and 7 poker chips. Each time the subject made an error, the experimenter returned to the previous number, with demonstrations only as needed the second time around. The training ended after the subject had correctly matched 7 blue poker chips with 7 red ones.

Short-term recall tests. Each subject was tested on the numbers 5, 6, 7, 8, and 9 with a procedure identical to that of the pretest. For the numbers 8 and 9, the board was lengthened by adding an extra part to it.

Near-transfer tests. Five tests involving one-to-one correspondence were administered in which the conditions were slightly different from the training. The instructions to the subjects were always to put "just as many" objects on their side as there were on the experimenter's side of the board. The tests were as follows:

1. *Different objects.* Tan, 2.5-cm wooden blocks were used on both sides of the board instead of poker chips. One trial was given using 6 blocks.
2. *Smaller chips.* The subject was given small (1.75-cm diameter), red chips, while the experimenter's blue ones remained the same large size. One trial was given with 7 chips.
3. *Chips close together.* The experimenter placed the chips in a row next to the ridge with no spaces between them. The regular large-sized chips were used. One trial was given using 6 chips.
4. *Two rows of chips.* The experimenter placed 7 chips on the board in two rows. One row was next to the ridge and contained 4 chips, while the other was farther back and contained 3 chips opposite the spaces between

Table 1
Mean Percentage of Correct Responses by the Two Training Groups on the Four Types of Tests in Experiment 1

Instructional method	1 week after training: Short-term recall	1 week after training: Near transfer	3 weeks after training: Long-term recall	3 weeks after training: Conservation (far transfer)
Discovery group (n = 10)	96	72	100	56
Expository group (n = 9)	80	64	71	19
t test (df = 17)	*ns*	*ns*	$p < .08$	$p < .01$

Note. The Group × Delay of Test interaction was significant at $p < .05$. Two delay scores were short-term recall and near transfer versus long-term recall and conservation.

the first four and making a zigzag pattern. One trial was given.

5. *Piles of chips.* The experimenter's chips were placed in 3 piles of 2 chips each next to the ridge, with the top chips covering about half of the bottom chips. One trial was given.

Long-term recall tests. The subjects were tested on the numbers 5, 7, and 9 with a procedure identical to that of the pretest.

Conservation test (far transfer). The experimenter first presented 6 blue chips with instructions identical to those of the pretest. Any errors were corrected. Once there were 2 rows of chips in correct one-to-one correspondence, the experimenter lengthened the row of red chips by spreading them out, so that they surpassed the row of blue chips by one at each end. The subjects were then asked, "Are there still just as many red ones as blue ones, or does one of us have more?" The subjects' answers were written down. The same procedure was repeated with 7 chips, except that the red chips were pushed together instead of spread out, so that they were surpassed by one blue chip at each end.

Results

Thirty-four subjects passed the pretest and were eliminated from the study. Of the 28 subjects who failed, 19 completed the training and tests. Therefore, there were 9 dropouts, all of whom quit at some point on their own accord (refused to play). Two of the children dropped out after the pretest, 2 during discovery training, 3 during expository training, and 2 after completing the expository training but before the tests.

Of the 19 subjects who completed the training and tests, 10 were in the discovery group and 9 in the expository group. The average age of the discovery group subjects was 4 years 1 month (48.8 months), while the expository group subjects averaged 4 years 0 months in age (47.9 months). There was no significant difference between the mean ages of the two groups, $t(17) = .32, p > .20$. The discovery group contained 7 girls and 3 boys, while the expository group contained 3 girls and 6 boys. Only three of the discovery group subjects required demonstrations during training, with an average of three demonstrations per subject. The other 7 subjects either made no errors or corrected themselves.

The tests were scored as follows: Each subject received four different scores between 0 and 100, reflecting the percentage of correct answers in each of the four types of tests (short-term recall, near transfer, long-term recall, and conservation test). The means on each posttest for the two groups appear in Table 1.

Two separate analyses of variance were computed on the data. The first had one between-subjects factor (the type of training received) and one within-subjects factor (the type of test). The effect of the type of training was significant, $F(1, 17) = 4.66, p < .05$, with the discovery group performing better than the expository group overall. The effect of the type of test was also significant, $F(3, 51) = 28.7, p < .001$. However, the main prediction of the assimilation theory is that there should be a pattern of interaction in which the two groups perform at similar levels for short-term recall and near transfer, but the discovery group outperforms the expository group on long-term recall and far transfer to novel situations. The predicted Group × Test interaction was obtained in the general direction outlined above but did not quite reach significance at the .05 level, $F(3, 51) = 2.28, p < .1$.

The second analysis of variance also had one between-subjects factor (type of training), which is identical to the first analysis, but two within-subjects factors: delay of test (1-week delay or 3-weeks delay) and type of test (recall or nonrecall). For purposes of this analysis, the conservation and near-transfer tests were both considered as nonrecall tests. As expected, the effects due to delay of test and type of test were both significant, $F(1, 17) = 134.3$, $p < .001$, and $F(1, 17) = 84.2$, $p < .001$, respectively. However, there was also an interesting pattern of Group × Delay interaction, $F(1, 17) = 56.99$, $p < .001$, with the groups performing at similar levels on the two tests given at 1 week (short-term recall and near transfer), but the discovery group performing better on the tests given at 3 weeks (long-term recall and conservation). In addition, *t* tests revealed that the two groups did not differ in overall performance on the combined short-term recall and near-transfer tests, $t(17) = 1.02$, *ns*, but the difference was significant for performance on the combined long-term transfer and conservation tests, $t(17) = 2.60$, $p < .02$.

In order to more carefully analyze this trend, *t* tests were conducted to determine whether the difference in performance for the two groups was significant for each of the four tests. The two groups did not differ in performance on short-term recall ($t < 1$) nor near transfer ($t < 1$). These tests measured behaviors identical or very similar to those taught during instruction; apparently, as predicted, the groups did not differ in their levels of mastery of the presented material. In addition, the discovery group outperformed the expository group at only a marginally significant level for the long-term recall test, $t(17) = 1.94$, $p < .08$. The only statistically significant difference was obtained for the test of far transfer in which the discovery group performed better on the conservation task than the expository subjects, $t(17) = 3.03$, $p < .01$.

Experiment 2

Experiment 2 was conducted in order to provide a replicatory test of the interesting results of Experiment 1. In particular, Experiment 2 provided an additional test of the prediction that discovery and expository instructions should lead to similar levels of performance on tests of mastery of the presented material (short-term recall and near transfer), but that discovery should lead to broader outcomes capable of superior long-term retention and far transfer to novel situations (long-term recall and conservation). Experiment 2 used different materials and a slightly different set of tests and included a third group (observation group) in which the subjects merely watched other people move the pieces. In Experiment 2, the conservation test was given at both 1 week and 3 weeks in order to overcome the problem that delay and transfer were confounded in Experiment 1. If discovery subjects outperform other groups on the conservation test, even when it is presented after only 1 week, that would be support for the idea that learning outcomes differed in breadth rather than solely in terms of long-term retention.

Method

Subjects. The subjects were 24 children between the ages of 3 years 3 months and 5 years 0 months, who attended two private schools near Santa Barbara, California and who failed a pretest for one-to-one correspondence (out of a larger group of 53 pretested children). The children from School A (a half-day nursery school) came primarily from white, middle-class, and upper middle-class homes; while the children from School B (a day-care center) came from both white and black, middle-class homes. Written parental permission was obtained for all subjects.

Design. There were three between-subjects groups: 9 subjects served in the discovery group, 8 subjects served in the expository group, and 7 subjects served in the observation group. Since all subjects took the same five posttests, comparisons by type of posttest are within-subjects comparisons.

Procedure. Each subject who failed the pretest participated individually in four sessions sitting opposite the experimenter at a small table in the school. No rewards were given for participating.

In Session 1, a pretest for one-to-one correspondence was administered to all subjects. Those who passed were eliminated from the study as in Experiment 1. In Session 2 (1 to 14 days later), those who failed the pretest were given one of three training programs for one-to-one correspondence. Subjects were grouped in trios, equated as much as possible for age, sex, and school attended. Members of each trio were then randomly assigned to each of the three training methods. In Session 3 (5 to 7 days after Session 2), the recall test (short-term recall), near-transfer test, and conservation test (far transfer) were given. In Session 4 (14 days after Session 3), the same recall (long-term recall) and conservation tests (far transfer) were administered as in Session 3.

Materials. Materials consisted of 12 small, plastic, black-and-white panda bears; 12 small, plastic, partly eaten apples; 12 red and 12 blue standard poker chips as used in Experiment 1; and the board and data sheets used in Experiment 1.

Pretest. The children were first familiarized with the materials, learned to name each object, and were shown the bear "eating" one of the apples. Six bears were placed on the experimenter's side of the board next to the ridge, equally spaced, and facing the subject, with the instructions, "Now you put just as many apples on your side. Make it so there are just as many apples as bears." Following the subject's response, a second trial with 7 bears was given. The criterion for passing the pretest was a correct response on either or both of the two trials.

Discovery and expository training. The discovery training and the expository training were identical to Experiment 1, with the exception that bears and apples were used instead of poker chips.

Observation training. Observation training was done in the form of a game in which the experimenter placed various numbers of bears on one side of the board and apples on the other side, and the subject had to judge each time whether it was done "right" or "wrong." "Right" was the appropriate answer for a correct one-to-one correspondence, that is, the same number of apples as bears. Objects were placed in the following manner, each pair beginning with the number of bears on the experimenter's side of the board: 1 and 1, 2 and 2, 2 and 3, 3 and 2, 3 and 3, 4 and 5, 4 and 4, 5 and 4, 5 and 5, 6 and 5, 6 and 6, 7 and 6, and 7 and 7. The apples (with the exception of the extra one or the missing one) were each placed directly opposite a bear. In order to familiarize the subjects with the same vocabulary as was used in the other two training methods and the tests, during each trial the experimenter said, "I'm going to put just as many apples as bears here. Look, are there just as many apples as bears? Did I do it right or wrong?" If the subject made an incorrect judgment, the experimenter simply corrected by saying, "No, I did it right (or wrong) this time," and proceeded to the next problem. The last six problems were repeated over again in the same order as many times as needed until six successive correct judgments occurred.

Recall tests. The recall tests, both the short term and long term, were identical to the pretest.

Near-transfer tests. These were administered only at the first test session. There were four different tests involving one-to-one correspondence in which the conditions were slightly different from the training. The instructions to the subjects were always to put "just as many" objects on their side as there were on the experimenter's side of the board. The tests were as follows:

1. *Different objects.* Blue and red poker chips were used instead of bears and apples. One trial was given using 7 poker chips.
2. *Two rows of bears.* The experimenter placed 7 bears on the board in 2 rows of 4 and 3, forming a zigzag pattern. One trial was given.
3. *Bears close together.* The experimenter placed 6 bears in a row next to the ridge with no spaces between them. One trial was given.
4. *Piles of bears.* Six bears were placed next to the ridge in 3 piles of 2 bears each. One trial was given.

These four tests were equivalent to Transfer Tests 1, 4, 3, and 5, respectively, of Experiment 1.

Conservation test (far transfer). The conservation test was identical to the conservation test of Experiment 1, except that bears and apples were used instead of poker chips. Two trials were given at both test sessions.

Results

Sixteen subjects passed the pretest and were eliminated from the study. Of the 37 subjects who failed, 24 completed the training and tests. Therefore, there were 13 dropouts. Two of the children were dropped after the pretest because of inability to understand the instructions, 2 were dropped because of inability to learn the observation training, and 2 left school before completion of the study. The others dropped out of their own accord (refused to play): 3 after the pretest, 1 after discovery training, 1 after expository training, and 2 after observation training.

Of the 24 subjects who completed the training and tests, 9 were in the discovery group, 8 in the expository group, and 7 in the observation group. The mean ages of the subjects were 4 years 2 months for the discovery group, 4 years 1 month for the expository group, and 3 years 11 months for the observation group (49.8 months, 49.0 months, and 46.7 months, respectively); t tests revealed no significant differences between these means. The discovery group contained 3 girls and 6 boys, while the expository group contained 4 girls and 4 boys, and the observation group 4 girls and 3 boys. Four of the discovery group subjects required demonstrations during training, with an average of three demonstrations per subject.

The tests were scored as follows: Each subject received five different scores between 0 and 100, corresponding to the percentage of correct answers for each of the five types of tests (recall, transfer and conservation tests 1 week after training, and recall and conservation tests 3 weeks after training). The group means for the five tests are shown in Table 2.

Table 2
Mean Percentage of Correct Responses by the Three Training Groups on the Five Types of Tests in Experiment 2

Instructional method	1 week after training			3 weeks after training	
	Short-term recall	Near transfer	Conservation (far transfer)	Long-term recall	Conservation (far transfer)
Discovery group ($n = 9$)	78	53	44	100	28
Expository group ($n = 8$)	63	50	6	75	6
Observation group ($n = 7$)	71	40	0	57	0
F test ($df = 2, 21$)	*ns*	*ns*	$p < .025$	$p < .08$	$p < .06$

Note: The Group × Type of Test interaction was significant at $p < .05$. Two types of tests were short-term recall and near transfer versus long-term recall and the two conservation tests.

Two separate analyses of variance were computed on the data. The first had one between-subjects factor (the type of training) and one within-subjects factor (the type of test). The main effect of type of training did not reach significance at the .05 level, $F(2, 21) = 2.88, p < .10$, but the effect of type of test was significant, $F(4, 84) = 29.63, p < .001$. However, the Group × Test interaction was not significant, although it did appear to be in the same direction as in Experiment 1, $F(8, 84) = 1.03$, *ns*. Based on the Group × Delay interaction in Experiment 1, the main prediction of the assimilation theory for Experiment 2 is a pattern of interaction in which all groups perform at similar levels for short-term recall and near transfer, but the discovery group shows an advantage for far transfer and long-term recall. A second analysis of variance was therefore conducted, which grouped the short-term recall and near-transfer scores together and the long-term recall and two conservation test scores together. There was one between-subjects factor (whether the training was expository, discovery, or observation) and one within-subjects factor (type of test). As in the previous analysis, the groups did not differ in overall performance, but there was a significant difference in the difficulty of the two types of tests, $F(1, 21) = 17.41, p < .001$. Furthermore, the predicted Group × Test interaction was significant, $F(2, 21) = 3.60, p < .05$, and showed the same trend as in Experiment 1.

In order to more closely investigate these differences, individual one-factor analyses of variance were conducted for differences among the three groups on each of the five tests. As in Experiment 1, there were no significant differences among the groups on the short-term recall test given 1 week after training ($F < 1$) nor on the near-transfer test given 1 week after training ($F < 1$). However, there were significant differences for the conservation test given 1 week after training, $F(2, 21) = 5.42, p < .025$. A Newman–Keuls multiple-range test performed on these data revealed that the discovery group performed significantly better than the expository ($p < .05$) and that the discovery group performed significantly better than the observation group ($p < .05$), but as expected, there were no differences between the expository and observation groups on the 1-week conservation test. This result helps extend the results of Experiment 1 in that the discovery group performed best on a test of far transfer even when it was presented closer in time to original learning. On the long-term recall test given 3 weeks after training, there was only a marginally significant effect, $F(2, 21) = 2.96, p < .08$, and the only pairwise difference that reached statistical significance ($p < .05$) based on a Newman–Keuls test was the difference between the discovery and observation groups.

The conservation test given at 3 weeks produced only a marginally significant effect for differences among the three groups, $F(2, 21) = 3.31, p < .06$, and only the difference between the discovery and observation groups reached statistical significance at the .05 level based on a Newman–Keuls test. It should be noted, however, that unlike Experiment 1, the conservation test given at 3 weeks is a repeat of an earlier test. Finally, a one-factor analysis of variance was performed on the combined performance on both conservation tests for the three groups. The three groups differed significantly in performance on the conservation tests, $F(2, 21) = 5.26, p < .025$. A Newman–Keuls test revealed that the differences between the discovery and expository groups and between the discovery and observation groups were both significant ($p < .05$), while the difference between expository and observation groups was not. These results complement those of Experiment 1: Although there were no differences between discovery and expository groups for short-term recall nor near transfer tests, there were statistically significant differences on tests for far transfer (conservation).

Conclusions

These results provide some evidence concerning the effects of discovery and active manipulation of objects on children's learning of number concepts. If this study had used a posttest based only on mastery of the presented information (i.e., only a recall or near-transfer test), there would have been no evidence of differences among the training groups in either Experiment 1 or Experiment 2. However, when posttests are given that include far transfer, such as conservation tests and long-term recall tests, important differences emerge. In particular, the discovery group performed relatively better than the expository group on far transfer, while both groups performed at similar levels for short-term recall and near transfer. Another interesting piece of information is that in Experiment 2, there were no reliable differences in the performances of the expository and observation groups. Hence, without discovery, the active manipulation of objects (expository group) seemed to have little positive effect.

These findings are consistent with the idea that the discovery procedure encouraged subjects to activate their existing cognitive structures concerning number concepts and to assimilate the new information to form a broader learning outcome. Subjects in the expository and observation treatments apparently were more likely to add the new behaviors to memory without connecting them to related ideas. It should be noted that the discovery procedure developed for these studies was such that subjects continued on the problem until they reached a level of mastery.

These results also encourage further work in determining the optimal conditions for and limits of discovery techniques. The present experiment was successful in demonstrating differences in the breadth of transfer due to discovery; the situation was such that discovery subjects learned to criterion (Condition 1), discovery subjects were likely to possess prerequisite concepts (Condition 2), and nondiscovery subjects were not likely to normally use assimilative learning strategies (Condition 3). In order to more fully understand the effects of discovery instruction on retention and transfer, attention must also be paid to these additional cognitive variables for any particular situation.

An important pedagogic implication of these findings is that equivalent mastery on a behavioral level such as was displayed by each of the treatment groups does not guarantee equivalency in "what is learned." There may be situations in which mastery of the presented information (in this case, performance on tests of one-to-one correspondence) is a sufficient instructional objective. However, in the case of number concepts, the broader learning outcomes produced by discovery seem to be desirable "cognitive objectives" (Greeno, 1976), especially since it is unlikely that all the necessary objectives for a lifetime of mathematics learning could be individually taught. These results suggest that there are definable situations in which discovery can lead to subjects acquiring competencies for objectives that were not specifically taught.

References

Anastasiov, N. J., Sibley, S. A., Leonhardt, T. M., & Borisch, G. D. A comparison of guided discovery, discovery and didactic teaching of mathematics to kindergarten poverty children. *American Educational Research Journal,* 1970, *7,* 493–510.

Ausubel, P. P. *Educational psychology: A cognitive view.* New York: Holt, Rinehart & Winston, 1968.

Bruner, J. S. The act of discovery. *Harvard Educational Review,* 1961, *31,* 21–32.

Egan, D. E., & Greeno, J. G. Acquiring cognitive structure by discovery and rule learning. *Journal of Educational Psychology,* 1973, *64,* 85–97.

Forgus, R. H., & Schwartz, R. J. Efficient retention and transfer as effected by learning method. *Journal of Psychology,* 1957, *43,* 135–139.

Gagné, R. M., & Brown, L. T. Some factors in the programming of conceptual learning. *Journal of Experimental Psychology,* 1961, *62,* 313–321.

Greeno, J. G. Cognitive objectives of instruction: Theory of knowledge for solving problems and answering questions. In D. Klahr (Ed.), *Cognition and instruction.* Hillsdale, N.J.: Erlbaum, 1976.

Mayer, R. E. Information processing variables in learning to solve problems. *Review of Educational Research,* 1975, *45,* 525–541.

Mayer, R. E., Stiehl, C. C., & Greeno, J. G. Acquisition of understanding and skill in relation to subjects' preparation and meaningfulness of instruction. *Journal of Educational Psychology,* 1975, *67,* 331–350.

Olander, H. T., & Robertson, H. C. The effectiveness of discovery and expository methods in the teaching of fourth-grade mathematics. *Journal for Research in Mathematics Education,* 1973, *4*(1), 33–34.

Peters, D. L. Discovery learning in kindergarten education. *Journal for Research in Mathematics Education,* 1970, *2*(3), 33–34.

Piaget, J. *The child's conception of number.* New York: Norton, 1965.

Shulman, L. S., & Keislar, E. R. (Eds.). *Learning by discovery.* Chicago: Rand McNally, 1966.

Strike, K. A. The logic of learning by discovery. *Review of Educational Research,* 1975, *45,* 461–483.

Wittrock, M. C. Verbal stimuli in concept formation: Learning by discovery. *Journal of Educational Psychology,* 1963, *54,* 183–190.

Wittrock, M. C. The learning by discovery hypothesis. In L. S. Shulman & E. R. Keislar (Eds.), *Learning by discovery.* Chicago: Rand McNally, 1966.

Received June 23, 1977 ■