

*A BEHAVIORAL TECHNOLOGY FOR PRODUCING CONCEPT  
FORMATION IN UNIVERSITY STUDENTS*

L. KEITH MILLER<sup>1</sup> AND F. HAL WEAVER<sup>2</sup>

UNIVERSITY OF KANSAS

Three experiments analyzed the effectiveness of a textbook incorporating "concept programming" in producing concept formation in university students. The concept programming portion of each lesson requires students to determine which concept is illustrated by each of 20 short fictional stories about everyday behavioral situations. The stories are selected to illustrate and contrast the concepts of that lesson. Student responses are heavily prompted during the initial stories of each lesson. The first experiment demonstrated that students generalize to entirely novel examples from the examples in the textbook. The second experiment demonstrated that the concept programming portion of the textbook is a critical component in producing generalization. The third experiment demonstrated that the amount of concept formation produced by the concept programmed textbook is greater than that produced by a widely used standard textbook.

DESCRIPTORS: academic behavior, college instruction, concept programming, teaching concepts, individualized instruction, multiple baseline, language, university students

---

The "Personalized System of Instruction", widely known as PSI, is the first comprehensive approach to higher education based on behavioral principles (Keller, 1968). Numerous studies have shown that this method produces greater academic achievement than traditional methods (Alba and Pennypacker, 1972; Born, Gledhill, and Davis, 1972; Bushell, 1965; Cooper and Greiner, 1971; Johnson and Pennypacker, 1971; Kulik, Kulik, and Carmichael, 1974; McMichael and Corey, 1969; Sapp, Edwards, and Thomas, 1972; Sheppard and MacDermott, 1970; Stalling, 1971; Witters and Kent, 1972).

Study guides are one of the major innovations of the PSI method. A study guide is a set of

questions that assists in studying a textbook. These study questions often comprise the pool of quiz items used to construct in-class quizzes to test mastery of the reading assignment. The effectiveness of study guides has been demonstrated by a recent experiment showing that with them, students attain quiz scores about 30% higher than scores attained by students without them (Semb, Hopkins, and Hursh, 1973).

The use of identical questions in quizzes and study guides emphasizes the simple recall of textual material, what Bloom (1956) has termed "knowledge". The use of similar, rather than identical questions in both study guides and quizzes, emphasizes what Bloom has termed "comprehension". Knowledge and comprehension have been classified by Bloom as the two most elementary educational objectives and there is considerable evidence in the literature that both of these objectives may be effectively attained by PSI methods (*e.g.*, Miller and Weaver, 1972; Semb, 1974).

Recently, "concept programming", which involves the use of discrimination training and

---

<sup>1</sup>We would like to thank Don Baer, Jim Sherman, Dave Born, George Semb, John Wright, and Steve Fawcett for their helpful suggestions on the design, interpretation, and writing of this work. Reprints may be obtained from L. Keith Miller, Department of Human Development, University of Kansas, Lawrence, Kansas 66045.

<sup>2</sup>Now at Southwest Indian Youth Center, 615 Adam Street, Tucson, Arizona.

programming principles, has been suggested (Miller and Weaver, 1975) as a method for producing more complex forms of learning. The stimulus material is arranged so that students learn to discriminate situations that exemplify a particular concept from those that do not. Also, students are heavily prompted to make the correct discriminations early in the programmed sequence, and the prompts are gradually withdrawn. The concept programming method is based on the finding that conceptual behavior can be produced by training the discrimination between a stimulus belonging to a conceptual class and a stimulus not belonging to that class, and by repeating this for many pairs of similarly selected stimuli (*e.g.*, Millenson, 1967).

Concept programming might be used to teach students to recognize examples of reinforcement. Such a program would start by presenting the student with a definition of "reinforcement". Next, he would be presented with a series of fictional stories of everyday behavioral situations, some of which illustrated reinforcement and others that did not. By using appropriate questions for the first few examples, the student can be prompted to analyze those critical aspects of the fictional stories necessary to determine whether they conform to the definition of reinforcement. Such questions might include: "Did the event precede or follow the response?", "Did the rate of the response increase?". After the student has learned to ask these questions as they analyze the stories, the questions can be faded from the program. By carefully including situations that meet one but not both criteria of reinforcement, students can be taught to make fine discriminations. Because the student can respond to many such examples, varying along many dimensions, the probability of generalizing to additional novel examples can be increased. If the student continues to discriminate examples that are not reinforcement, then the program should cause the student to form the concept of "reinforcement".

The strategy of testing for concept formation in the present experiments involved showing

students a variety of short fictional stories that they had not seen in the programmed material and then asking them to identify the behavioral concept involved. Since they had not been taught the answers in reference to these specific stories, a correct answer constitutes evidence of generalization (Markle and Tieman, 1970). Furthermore, by illustrating different concepts with different examples, each answer also constitutes evidence of discrimination—the wrong label was not applied to the situation. The more stories that are correctly analyzed by the student, the stronger the evidence that he would correctly answer additional novel stories from that same stimulus class. The stimulus control over a response that involves generalization to an entire stimulus class, such as all stories exemplifying "nonreinforcement", has been taken as the definition of concept formation in this paper (Ferster and Perrot, 1968; Skinner, 1953; Whaley and Mallott, 1971).

This paper reports three experiments. The first was designed to analyze the effect on concept formation of a textbook-plus-PSI teaching package that uses concept programming. The second experiment was designed to analyze the effect of the concept programming component of the text in comparison to the other components of the text. The third experiment was designed to compare the overall effectiveness in producing concept formation of the concept programmed textbook with another widely used but more traditionally written textbook.

## EXPERIMENT I

### METHOD

#### *Subjects*

All nine students enrolled in the 1974 summer session in a personalized, introductory undergraduate course covering the basic principles of behavior analysis in the Department of Human Development at the University of Kansas, served as subjects.

### *Textbook Description*

The textbook *Principles of Everyday Behavior Analysis* (Miller, 1975) was designed to use the principles of discrimination and generalization to develop abstract conceptual behavior. The book contains 26 lessons. Twenty-two lessons stress one major behavioral concept or several related major concepts, and four lessons are review. There are chapters on such topics as "methods of behavior observation", "reinforcement", "discrimination training", and "review of reinforcement".

Each nonreview lesson was divided into three parts: an "introduction", a "concept program", and a "self-quiz". The introduction typically consisted of one to four pages introducing the student to the concept of that lesson. This introduction would include a definition of each concept, several illustrations of that concept occurring in everyday behavior, and instructions on how to identify whether or not a situation exemplified that concept. Thus, this part of the lesson was similar in structure and intent to the major portion of most typical textbooks.

The "concept program" consisted of 20 fictional examples of everyday situations familiar to most university students. About 65% of the examples were chosen to illustrate the concept under study; the remainder were chosen to illustrate similar concepts from other chapters. Several examples contrary to common sense were also included, such as an example of spanking that increased the rate of behavior. All irrelevant features of the examples, such as names, settings, behaviors, and stimuli were varied. The principle was to design examples that required students to focus on the essential elements that defined the concept (Tennyson, Woolley, and Merrill, 1972).

The examples dealt with social interactions between adults, groups, professor-student interaction in university classrooms, students engaged in political behaviors, and parent-child interactions with university-age children. About 80% of the examples dealt with the behaviors of

normal young adults in everyday social settings. None of the examples dealt with animal behavior or the behavior of abnormal populations such as retardates, schizophrenics, or adults with phobias.

The task of the student in the concept program portion of the book was to identify the concept exemplified by the example. To assist the student in doing this, the first few examples were accompanied by several questions that pointed out how the example did or did not conform to the definition contained in the introduction to the lesson. These questions were included to prompt the correct identification of the example. These prompts were gradually faded out during the first 10 examples, so that there were no prompts in the last 10 examples. Answers to all the questions relating to the first 10 examples were provided in the back of the book, allowing for immediate feedback. All questions called for one-word fill-in answers.

The third part of each lesson consisted of a "self-quiz" containing 10 questions related to the introduction and 20 questions related to the examples contained in the concept program portion of the text. The questions related to the examples included short, four- to six-line summaries of the examples and the student was asked to identify the behavioral concept involved. All self-quiz questions required one word answers. No answers were supplied for any of these questions because they were also used on in-class quizzes. Use of the self-quiz was optional unless a student failed a quiz (see below). All questions were analyzed for frequency of errors during five semesters with a total of approximately 500 students. Every question missed by more than 10% of the students was revised and rechecked the next semester. Error rates for all questions were reduced to 10% or below.

Each review lesson provided an overview of three to seven lessons, including a glossary of terms introduced in those lessons. Forty to 70 questions over basic concepts and definitions were contained on each self-quiz; another 20 questions were based on fictional examples con-

structed for the review lesson. These questions focused primarily on the concepts in the unit under review, but also provided some review of concepts in earlier units.

### *Course Description*

The course was conducted along the lines of a standard PSI (Keller, 1968) course. The self-quiz portion of the textbook served as a study guide and students were required to take a quiz based on a sample of 10 of these self-quiz items (30 on review lessons) to demonstrate mastery of each lesson. If they did not pass the quiz with a score of 90% or better, they were required to fill in the answers to the "concept program" portion of the text before retaking the quiz. When they handed in those answers they were eligible to take an alternative form of the quiz (with totally different self-quiz questions), until they attained a score of 90%. Their grade in the course depended on their average score on the first form of the quiz for each lesson—a "B" being earned by an average of 90% or more, a "C" by an 80% to 89% average, and so on. They could earn one letter grade increment by completing a special project assigned by the instructor (usually involving serving as an experimental subject). Student proctors were available to answer questions on the introduction and the examples (but not on the self-quiz items, since they were the same as the questions on the in-class quiz) and to grade their quizzes. Proctors spent most of their time giving and grading quizzes, with relatively little time spent answering student questions. Proctors did not initiate questions to probe student understanding of the material. The proctors were trained by students who served as their proctors during the prior semester using a standardized training package that included role playing (Weaver and Miller, 1975). Students had to progress through the course by keeping up with a series of target dates assigned by the instructor (Miller, Weaver, and Semb, 1974). Thus, the course operation involved the use of study guides, frequent quizzing, mastery criteria, and student proctors.

### *Generalization Test*

The generalization of student responses from the examples in the textbook to novel examples was measured by means of a Generalization Test. This test consisted of 48 fill-in questions, similar to those appearing in the concept program, that had never been seen before by the students. The questions consisted of examples of 16 basic concepts drawn from three different content areas (see Table 1). To increase the novelty of the items, the content areas of the questions had not been used in the textbook: 16 examples involved animal experiments, 16 involved the research behavior of a psychologist, and 16 involved the behavior of welfare clients and caseworkers. This test was administered 10 times during the semester; students never received feedback on their answers.

The 48 questions were scored in four separate categories (12 items each) based on the four major units of the textbook: Methods, Reinforcement Control, Stimulus Control, and Aversive Control. The percentage of correct responses was then computed separately for each unit, and these scores were used as indices of student abil-

Table 1  
The 16 Concepts Used in the Generalization Test

|                               |
|-------------------------------|
| <i>Methods of Research:</i>   |
| Reliability                   |
| Interval recording            |
| Reversal Design               |
| Response Definition           |
| <i>Reinforcement Control:</i> |
| Extinction                    |
| Reinforcement                 |
| Shaping                       |
| Differential Reinforcement    |
| <i>Stimulus Control:</i>      |
| Discrimination                |
| Generalization                |
| SD                            |
| Concept                       |
| <i>Aversive Control:</i>      |
| Punishment                    |
| Escape                        |
| Avoidance                     |
| Negative Reinforcement        |

ity to analyze correctly examples of the concepts in each unit.

### *Experimental Design*

This experiment examined the effect of the teaching package on the students' ability to answer correctly questions on the Generalization Test. For this purpose, each of the four major units of the textbook, along with the associated course procedures (e.g., proctors and quizzes), was treated as a separate treatment package designed to teach the student how to answer questions in that content area. Scores on questions from each of the four units were examined before, during, and after the units were assigned to be read. Specifically, each student was given the Generalization Test in class once before any lessons and then again after completion of every third lesson, for a total of 10 administrations. Each student was required to complete the generalization test before they could progress to further lessons. A teaching assistant graded all generalization tests by comparing the one-word answers with an answer key that listed all acceptable answers. The senior author then spot-checked the scoring. No scoring discrepancies were noted.

This procedure defined a multiple-baseline design across the four question-answering baselines, where the treatment was the relevant part of the textbook and the associated course procedures. This defined five experimental conditions:

*Baseline.* During the first test, question-answering behaviors for all four units of the text were in pretreatment or baseline condition because none of the students had yet read any of the textbook.

*Treatment for methods.* During the second and third test administrations, students completed the six lessons dealing with methods of research. Responses to questions in the other three topical areas were in baseline condition during this period.

*Treatment for reinforcement control.* During the fourth through sixth test administration,

students completed the lessons dealing with reinforcement control; responses to methods questions were in a posttreatment condition during this time; responses to stimulus control and aversive control questions were still in baseline condition.

*Treatment for stimulus control.* During the seventh and eighth test administrations, students completed the lessons relating to stimulus control; responses to methods and reinforcement questions were simultaneously in posttreatment condition, while responses to aversive control questions were still in baseline.

*Treatment for aversive control.* During the ninth and tenth test administrations, students completed the lessons relating to aversive control. Responses to questions in the other three areas were in a posttreatment condition.

## RESULTS

Figure 1 shows the effect of the treatment packages on the mean per cent of correct responses on each generalization subtest. During baseline, the mean per cent of correct answers ranged from 11% to 50% correct for the four subtests. During treatment for the Methods subtest, this figure jumped to 78% and then gradually increased to 95%; very little change occurred in the other three subtests. During treatment for the Reinforcement subtest, the mean per cent increased to 76%, and then gradually to 87%, with little change in the other subtests still in baseline. During treatment for the Stimulus Control subtest, the mean per cent increased to 67% and then gradually to 75%, with little change in the Aversive Control subtest. When the treatment was applied to the Aversive Control subtest, the mean per cent of correct responding increased to 67%. By the final test, the mean per cent correct responding on the four subtests ranged from 67% correct to 95% correct. Thus, an increase was noted in each of the four subtests, and that increase was clearly associated with introduction of the textbook for that specific topical area.

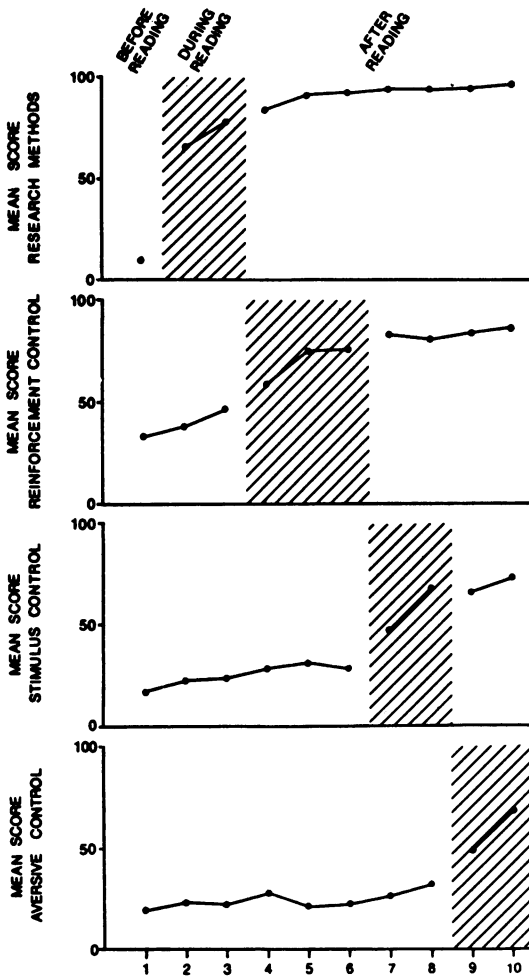


Fig. 1. Mean scores on each of four unit generalization tests administered to students following completion of three lessons in the textbook during Experiment I. The hatched bars indicate which scores were obtained when a portion of the text pertaining to a specific generalization test was being read by the students.

An analysis of the data for individuals revealed that all students scored higher in each of the four areas on the final test compared to their initial test, and that this increase was associated with introduction of the treatment packages.<sup>3</sup> Thus, the individual data are generally consistent with the group effect described

<sup>3</sup>The probabilities of the observed mean differences for each subtest between the first and last administration arising by chance are  $P < 0.0005$  by a correlated t-test.

by Figure 1. Examination of the number of quiz retakes indicated that students averaged about one retake every 10 quizzes.

## DISCUSSION

The results of this experiment showed that the percentage of correct responses to novel questions increased after the students read and passed quizzes over the relevant part of the textbook. The fact that most of the increase in correct responding for each unit occurred after introduction of the treatment package provides evidence that the package produced these increases. Further, the individual data indicate that this effect was reliably observed in the behavior of the individual students.

## EXPERIMENT II

The textbook used in Experiment I is similar to most texts except for one feature—the concept program. This part of the text is designed to help the student generalize to many examples of a concept while still discriminating them from examples of other concepts. However, it may be that the inclusion of definitions in the text is sufficient to produce concept formation (Anderson and Kulhavy, 1972).

Experiment II was designed to analyze the importance of the concept program in producing increases in the correct responses to the Generalization Test. Basically, the experiment undertook a component analysis of the textbook. First, students were taught to identify formal definitions of the concepts taught in the course; second, they were taught to answer questions about the “introduction” to each lesson, thereby learning not only definitions but also textual illustrations of those concepts and instructions about use of the concept; third, students were required to work through the concept program portion of the text. Each of these components was analyzed by means of a multiple-baseline design across the four generalization subtests associated with each major unit of the textbook. Our purpose was to analyze the relative effect of teaching students

formal definitions, requiring them to read the lesson introductions, and of requiring them to answer questions in the concept programs.

### *Subjects*

Nine students enrolled in Spring Semester, 1973, participated in Experiment II. They were selected from among a larger number of students who volunteered by eliminating all the students who had taken a prior course in Human Development. The students were freshmen and sophomores representing a variety of majors.

### *Procedure*

The experiment was administered in a small classroom separated from that used to teach the regular course. Subjects were not provided with the textbook used in the regular course and they were instructed not to borrow the textbook used in the course. Three types of instructional material were provided to them at different times during the experiment.

The first type of material consisted of all the terms defined in the textbook, their definitions, and a study guide consisting of the definitions with one word omitted from each question. The definitional material was divided into four sets correlated with the four major units of the textbook. The definitions for the first unit (Methods) were presented to the students, and after they had studied it they were given a quiz based on the study guide. When they passed a quiz at the 90% level, the generalization test was administered. If they did not pass at the 90% level, they were instructed to restudy the material and take alternative forms of the quiz until they reached 90%. When they completed the generalization test, they were given the definitional materials associated with the second unit in the textbook (Reinforcement Control), and the same procedure followed. This was repeated for the definitional material from the third and fourth units of the text.

The second type of material was the lesson introductions appearing in the textbook, along with the 10-question study guide based on the

introduction. This material was also divided into the four major units of the textbook. After students passed a quiz at the 90% level over the introductory material for the first unit, the generalization test was administered. Then they were given the introductory materials for the second unit in the textbook, and the same procedure followed. This was repeated for the introduction materials correlated with the third and fourth units.

The third type of material consisted of the concept programming section of the text and related study questions. Thus, at this point, students were using the entire text: the lesson introductions, the concept program, and the self-quizzes. Again, this was divided into four sets associated with the textbook units. Initially, only the material correlated with the first unit was given to the students; after they passed a quiz over this unit at the 90% level, the generalization test was administered to them. The materials associated with the second, third, and fourth units were administered in the same fashion.

The materials were controlled so that students could not work ahead. A pilot study in which this precaution was not observed revealed upward trends in many of the baselines just before the presumed introduction of the packages. This pattern suggested students were studying ahead, contrary to instructions.

The subjects were administered the course by trained proctors who followed the same procedures followed by proctors in the first experiment. Students were permitted to work at their own pace on the experimental materials. The same generalization test was used in this as in the first experiment. It was graded by comparing the one-word answers with an answer key that listed all acceptable answers. The senior author spot-checked for discrepancies; none were noted.

### *Experimental Design*

The experimental design might be characterized as a three-stage sequential multiple-baseline design. Each of the three stages was used to

determine experimentally the successive contribution of each one of the three types of textbook components: definitions, lesson introductions, and the concept program. During each of the stages, the additional effect produced by the material associated with each of those three treatment packages was analyzed in a multiple-baseline design across the same four generalization subtests used in the first experiment. Each stage established a baseline level of performance against which the effects of the treatment package introduced in the subsequent stage could be evaluated.

Before the first stage, baseline performance on each of the four generalization subtests was established. Then, the effect of teaching the subjects only the definitions of the concepts was examined in a multiple-baseline design. This required four administrations of the generalization test. Next, the effect of teaching the students the material in the lesson introductions was examined in a multiple-baseline design. The initial baseline for this second stage was the level of performance established by the first stage (definitions). Thus, this stage permitted determination of the added effect of introducing the lesson over the definitions. This also required four administrations of the generalization test. Finally, the third stage of the design involved requiring the students to work through the concept programs. This stage permitted determination of the added effect of the concept program over the level of performance produced by the definitions and lesson introductions. This, too, required four administrations of the generalization test. Altogether, the three-stage sequential multiple-baseline design involved 13 administrations of the Generalization Test. This means that the level of performance on each unit was determined four times before any change in conditions (except during the initial baseline).

RESULTS

Figure 2 shows the effect of each additional component of the textbook on the four sections of the Generalization Test. The initial baseline

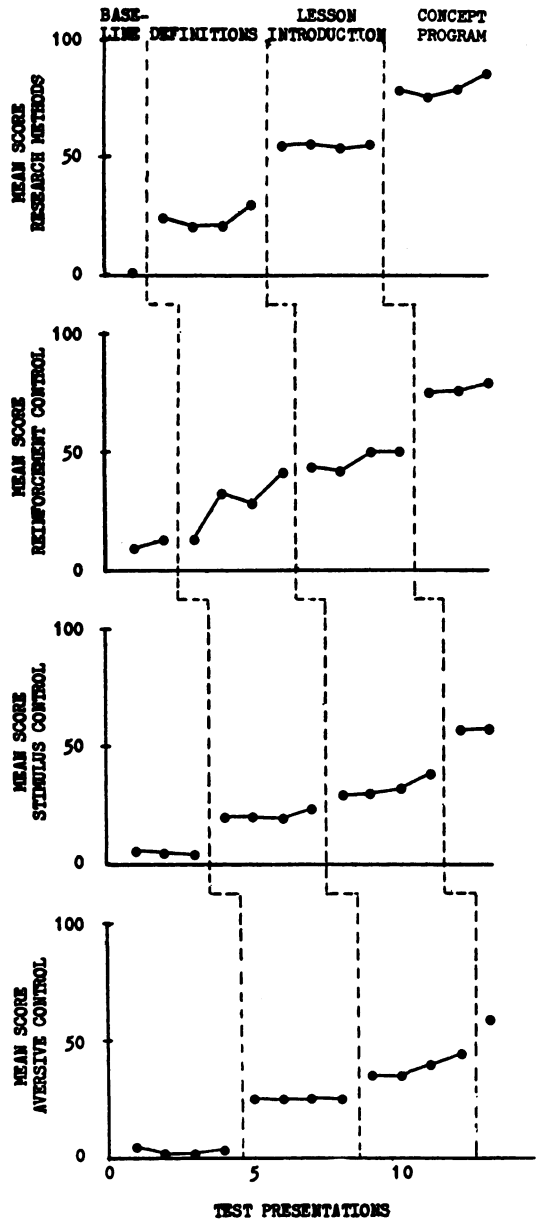


Fig. 2. The mean scores on each of four generalization subtests administered every time the student completed a set of materials associated with a major unit of the textbook during Experiment II. The materials consisted of (1) definitions of all terms in a unit; (2) introductions to all lessons in the unit; (3) concept programs for all lessons in a unit.

determination for the four subtests averaged about 4%. The second section of the figure shows the effect of teaching students formal definitions of the basic terms. Mean test score



increased from the baseline level of 4% to a mean of about 25%, ranging from about 15% to 45%. Correct responding did not increase significantly during baseline conditions for any of the subtests. Thus, the increase above baseline is tightly correlated with introduction of the definitions.

The third part of the figure shows that teaching students the lesson introductions increased the mean test score to about 42% correct. While this increase usually occurred after initiation of the appropriate part of the textbook, there was some upward movement of the behavior before that. This was particularly true with respect to the Reinforcement subtest, which showed little change after the new condition. The other three behaviors were relatively stable throughout this stage. The use of the lesson introductions was associated with an increase in subtest performance from 25% (after teaching the definitions) to 42%. In three of four cases, this increase was tightly correlated with introduction of the package.

The fourth part of the figure shows the effect of the concept program on each subtest. With the concept program, the mean per cent of correct answers increased to about 72%. Again, the four measurements during the lesson introductions condition were reasonably stable, so that this increase was correlated with introduction of the concept program.

#### DISCUSSION

This experiment suggests that the concept program was crucial to the students' performance on the Generalization Test. The first part of the experiment showed that students correctly answered about 25% of the generalization questions after exposure to the definitions. The second part of the experiment showed that scores could be further increased by adding textual discussion of the definitions for each unit. Finally, the average score was still further increased for each unit after introduction of the concept program. The multiple-baseline design provides evidence that these separate increases

can be attributed to introduction of the different components of the textbook. Since the concept program produced a large increase in Generalization Test scores over that produced by either the definitions or the definitions plus strictly expository textual materials, we can conclude that the concept program contributed importantly to final scores on the Generalization Test.

#### EXPERIMENT III

Experiment I demonstrated that the textbook and associated course procedures produced concept formation. Experiment II demonstrated that the concept programming portion of the textbook contributed importantly to the concepts that were developed. However, we still do not know whether this text is an improvement over typical textbooks. The third experiment was designed to compare the effectiveness of the concept programmed textbook with a more traditionally written, and widely used alternative textbook.

#### METHOD

##### *Subjects and Setting*

Twenty undergraduate students enrolled in an introductory behavior analysis course at the University of Oregon<sup>4</sup> and who scored lowest on the pretest administration of the Concept Formation Test described below were selected as subjects. The experiment was conducted in the classroom during the hours assigned to the course.

##### *Concept Formation Test*

To determine student ability to identify examples of "reinforcement", 20 fictional stories, each about a paragraph long, were presented to the subjects. Four stories were written by each of five faculty members at the University of Kansas for a total of 20 stories. This was done

<sup>4</sup>Thanks go to George Semb who administered this experiment during his summer teaching at the University of Oregon.

to avoid having the test items and the textbook written by the same author. Each story described the rate of a particular behavior before and after introduction of a new event into the environment. In 10 of the stories, the event followed the behavior and increased the rate of the behavior. These stories were, therefore, examples of reinforcing events. In five of the stories, the event followed the behavior but did not lead to an increase in rate. In the remaining five stories, the event led to an increase in the rate, but it preceded instances of the behavior. Therefore, these stories were not examples of a reinforcing event. Students were asked to indicate whether or not each story was an example of a reinforcing event. Scores on the test were computed as the per cent of correct answers.

All tests were scored from a standardized answer key listing the correct answer to each of the 20 questions. The senior author spot-checked this scoring and found no scoring discrepancies.

#### *Procedure and Experimental Design*

The 20 subjects were assigned to two groups matched for initial test scores. The two groups were then required to read the chapters on reinforcement in two textbooks, a traditional text and the concept programmed text.

The first group read the concept programmed textbook (Textbook B) followed by the traditional textbook (Textbook A). The second group reversed the order. The textbooks were considered "read" only after students handed in a study guide for these chapters published with the books. Then, the students were required to take the concept formation test a second time. They then studied the second textbook; after handing in a completed study guide they were given a third and final administration of the concept formation test. Thus, the experimental design consisted of a repeated-measurements group design in which the textbooks were presented to each of two groups in a counterbalanced fashion.

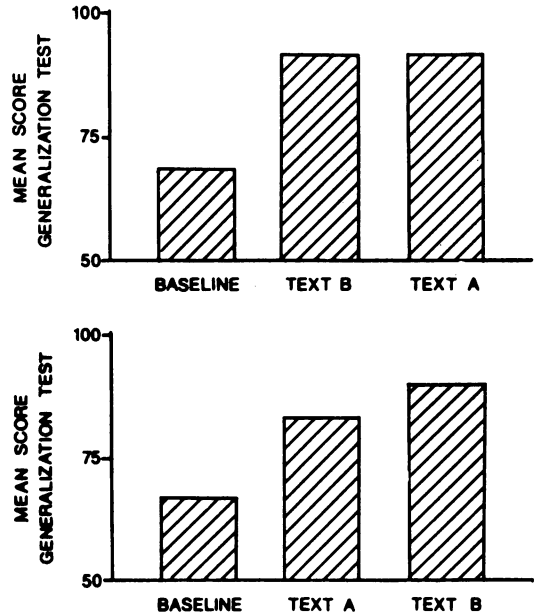


Fig. 3. The mean scores of a Reinforcement concept formation test for two groups of students during Experiment III. The first group was tested before reading Textbook B, after Reading Textbook B, and after reading Textbook A. The second group was tested before reading Textbook A, after reading Textbook A, and after reading Textbook B. Textbook A is a traditionally written textbook; Textbook B is a concept programmed textbook.

#### RESULTS AND DISCUSSION

Figure 3 shows the effect of the different textbooks on the concept formation test scores. Group I had a pretest score of about 68%, which increased to 92% after reading Textbook B, the one designed according to the principles of concept programming. No further gains were observed after reading the traditional textbook. Group II had a pretest score of 66%, which increased to 81% after reading Textbook A. This increased still further to 89% after reading Textbook B. Considering only the gain occurring after reading the first book, it can be seen that students gained an average of 15% after reading the traditional textbook, and an average of 24% after reading the text based on concept programming.<sup>5</sup> Thus, the data show that gains

<sup>5</sup>The probability of this difference occurring by chance is  $p < 0.10$  computed by a correlated t-test.

were made from baseline levels to about 90% mastery by the students after reading the text that included a concept program, but that smaller gains were made after reading the traditional textbook.

## GENERAL DISCUSSION

The present series of experiments demonstrate that a textbook incorporating concept programming produced a high score on a generalization test, that concept programming is a crucial component of that textbook, and that the resulting generalization test score is greater than that produced by at least one comparable traditional textbook.

A crucial point of concern in evaluating these findings involves the similarity of text items and Generalization Test items. On the one hand, they must be similar if they are to provide an example of the concept. On the other hand, they must be sufficiently different to be more than a trivial variation on examples already learned. As stated earlier, the examples were designed to be clearly different from the items used in the teaching materials. The most dramatic case of this was the use of animal examples on the generalization test, while all the textual examples referred to everyday behavioral situations involving normal humans. Although not as dramatic, the use of research behavior and welfare examples were also entirely different in content from the examples used in the textual materials. Furthermore, an attempt was made to eliminate terms or other word cues that might influence Generalization Test scores.

Thus, the results may be interpreted as evidence that the students were taught concepts. A widely used behavioral definition of "concept" is that it is a class of stimulus situations that have a defined similarity (Whaley and Mallot, 1971). By this definition, the unlimited set of verbal examples of "reinforcement" that all conform to the definition of that term would constitute a concept. Behavioral evidence of concept forma-

tion might involve a demonstration that a student applied the term "reinforcement" correctly to representative examples of reinforcement, and that it was not applied incorrectly to examples of other phenomena. The results of the last administration of the Generalization Test in all experiments provided just such evidence. Thus, the present experiments may be interpreted as support for the contention that concept programming is a behavioral procedure for producing concept formation with respect to academic material.

It should be pointed out that the present series of experiments do not support the conclusion that all components of the concept program are crucial. It may be that active responding, fading of prompts, and the inclusion of illustrative and contrasting examples are not all necessary. Resolution of this issue must await a component analysis of the concept program. However, the present experiments do support the conclusion that the package of components we have labelled a "concept program" does produce a reliable and significant effect on concept formation, even though we cannot conclude that each component of that package is necessary.

One interesting implication is that concept programming can teach students to generalize appropriately from a limited set of questions to a virtually infinite class of questions. This skill is classified as "application" within Bloom's (1965) taxonomy and is regarded as a more complex and advanced educational objective.

The results of the three experiments indicate that concept programming may be a useful addition to any textbook designed to teach concepts. However, the experiments should not be misunderstood to indicate that students do not learn from traditional textbooks. Students clearly learn the recall of knowledge and some conceptual behavior from such texts, particularly when the texts are used within a personalized system of instruction. Therefore, we view concept programming as a behavioral procedure than can be used to augment the nature and range of educational objectives attained by textual materials.

The concept programming procedure may also be useful in constructing supplementary material for use with a normal textbook. Such material would be similar to that used in a study guide and, in fact, might even be included in the same handout. With the addition of such materials, contingency management courses may take another step toward the development of a behaviorally-based program of higher education. The present research suggests that such an approach may well assist students in forming concepts related to their particular program of study.

### REFERENCES

- Alba, E. and Pennypacker, H. S. A multiple change score comparison of traditional and behavioral college teaching procedures. *Journal of Applied Behavior Analysis*, 1972, **5**, 121-124.
- Anderson, R. L. and Kulheuy, R. W. Learning concepts from definitions. *American Educational Research Journal*, 1972, **9**, 335-390.
- Bloom, B. S. *Taxonomy of educational objectives*. New York: Longmans, Green, 1956.
- Born, D. G., Gledhill, S. M., and Davis, M. L. Examination performance in lecture-discussion and personalized instruction courses. *Journal of Applied Behavior Analysis*, 1972, **5**, 33-43.
- Bushell, D. Jr. Textbook and programs: antithesis or synthesis. *National Society for Programmed Instruction Journal*, 1965, **4**, 3-5.
- Cooper, J. L. and Greiner, J. M. Contingency management in an introductory psychology course produces better retention. *Psychological Record*, 1971, **21**, 391-400.
- Ferster, C. B. and Perrott, M. C. *Behavioral principles*. New York: Appleton-Century-Crofts, 1968.
- Johnston, J. M. and Pennypacker, H. S. A behavioral approach to college teaching. *American Psychologist*, 1971, **26**, 219-244.
- Keller, F. S. Goodbye teacher . . . *Journal of Applied Behavior Analysis*, 1968, **1**, 79-89.
- Kulik, J. A., Kulik, C., and Carmichael, K. The Keller plan in science teaching. *Science*, 1974, **183**, 379-383.
- Markle, S. M. and Tieman, P. W. "Behavioral analysis of 'cognitive' content". *Educational Technology*, 1970, **10**, 41-45.
- McMichael, J. S. and Corey, J. R. Contingency management in an introductory psychology course produces better learning. *Journal of Applied Behavior Analysis*, 1969, **2**, 79-83.
- Millenson, J. R. *Principles of behavior analysis*. New York: Macmillan, 1967.
- Miller, L. K. *Principles of everyday behavior analysis*. Monterey, California: Brooks/Cole, 1975.
- Miller, L. K. and Weaver, F. H. A multiple baseline achievement test. In G. Semb (Ed.), *Behavior analysis and education—1972*. Lawrence: Support and Development Center for Follow Through, Department of Human Development, University of Kansas, 1972. Pp. 393-399.
- Miller, L. K. and Weaver, F. H. The use of "concept programming" to teach behavioral concepts to university students. In J. Johnston (Ed.), *Behavior research and technology in higher education*. Springfield, Illinois: Charles C Thomas, 1975. Pp. 44-57.
- Miller, L. K., Weaver, F. H., and Semb, G. A procedure for maintaining student progress in a personalized university course. *Journal of Applied Behavior Analysis*, 1974, **7**, 87-91.
- Sapp, G. L., Edwards, B. C., and Thomas, J. D. Reinforcement principles in an introductory educational psychology course. *Journal of Educational Research*, 1972, **66**, 72-75.
- Semb, G. The effects of mastery criteria and assignment length on college student test performance. *Journal of Applied Behavior Analysis* 1974, **7**, 61-70.
- Semb, G., Hopkins, B. L., and Hursh, D. E. The effects of study questions and grades on student test performance in a college course. *Journal of Applied Behavior Analysis*, 1973, **6**, 631-642.
- Sheppard, W. C. and MacDermott, H. C. Design and evaluation of a programmed course in introductory psychology. *Journal of Applied Behavior Analysis*, 1970, **3**, 5-11.
- Skinner, B. F. *Science and human behavior*. New York: Macmillan, 1953.
- Stallings, R. B. A one-proctor programmed course procedure for introductory psychology. *Psychologica Record*, 1971, **21**, 501-505.
- Tenneyson, R. D., Woolley, F. R., and Merrill, R. D. Exemplar and nonexemplar which produce concept classification behavior and specified classification errors. *Journal of Educational Psychology*, 1972, **63**, 144-152.
- Weaver, F. H. and Miller, L. K. Teaching students how to proctor in a PSI course by means of a role-playing procedure. In J. Johnston (Ed.), *Behavior research and technology in higher education*. Springfield, Illinois: Charles C Thomas, 1975. Pp. 168-182.
- Whaley, D. L. and Malott, R. W. *Elementary principles of behavior*. New York: Appleton-Century-Crofts, 1971.
- Witters, D. R. and Kent, G. W. Teaching without lecturing: evidence in the case for individualized instruction. *Psychological Record*, 1972, **22**, 169-175.

Received 24 October 1974.

(Final acceptance 24 December 1975.)