The Development and Empirical Validation of Programmed Instructional Materials to Improve Conceptual Mastery

Richard M. Woolfenden

Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Educational Psychology Commons

Recommended Citation

https://scholarworks.wmich.edu/masters_theses/2827

This Masters Thesis-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Master's Theses by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
THE DEVELOPMENT AND EMPIRICAL VALIDATION
OF PROGRAMMED INSTRUCTIONAL MATERIALS
TO IMPROVE CONCEPTUAL MASTERY

by

Richard M. Woolfenden

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment
of the
Degree of Master of Arts

Western Michigan University
Kalamazoo, Michigan
August 1972
ACKNOWLEDGEMENTS

I would like to express my appreciation to Dr. Richard Malott for his aid, guidance, and encouragement during the course of this study and preparation of this thesis. I would like to thank the members of my thesis committee, Dr. Howard Farris and Dr. Jack Michael, for their suggestions, comments and criticisms. For the financial aid and intellectual training received during my graduate studies, I express my gratitude to the Department of Psychology.

Richard M. Woolfenden
MASTERS THESIS

WOOLFENDEN, Richard Marshall
THE DEVELOPMENT AND EMPIRICAL VALIDATION OF
PROGRAMMED INSTRUCTIONAL MATERIALS TO IMPROVE
CONCEPTUAL MASTERY.

Western Michigan University, M.A., 1972
Education, psychology

University Microfilms, A XEROX Company, Ann Arbor, Michigan
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INDEX OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Improving Source Materials</td>
<td>2</td>
</tr>
<tr>
<td>Elements of Programmed Instruction</td>
<td>3</td>
</tr>
<tr>
<td>The Development of Programmed Instructional Materials</td>
<td>7</td>
</tr>
<tr>
<td>Programming Techniques</td>
<td>10</td>
</tr>
<tr>
<td>Levels of Learning</td>
<td>12</td>
</tr>
<tr>
<td>Teaching Concepts</td>
<td>13</td>
</tr>
<tr>
<td>Cost Analysis</td>
<td>15</td>
</tr>
<tr>
<td>The Problem</td>
<td>16</td>
</tr>
<tr>
<td>METHOD</td>
<td>23</td>
</tr>
<tr>
<td>Subjects</td>
<td>23</td>
</tr>
<tr>
<td>Procedure</td>
<td>23</td>
</tr>
<tr>
<td>Punishment program objectives</td>
<td>25</td>
</tr>
<tr>
<td>Initial development of punishment program</td>
<td>25</td>
</tr>
<tr>
<td>Ratio program objectives</td>
<td>28</td>
</tr>
<tr>
<td>Initial development of ratio program</td>
<td>28</td>
</tr>
<tr>
<td>One-to-one testing</td>
<td>29</td>
</tr>
<tr>
<td>Validation testing</td>
<td>31</td>
</tr>
<tr>
<td>RESULTS AND DISCUSSION</td>
<td>34</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>43</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>44</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>1</td>
<td>Results of punishment validation testing</td>
</tr>
<tr>
<td>2</td>
<td>Results of ratio schedules validation testing</td>
</tr>
<tr>
<td>3</td>
<td>Error analysis of punishment program for Group P</td>
</tr>
<tr>
<td>4</td>
<td>Error analysis of ratio program for Group P</td>
</tr>
</tbody>
</table>
The goals of mass education are not being achieved as effectively as possible. According to Malott (1972, 8-3), the objective of the educator is "to teach as much as possible to as many as possible." This task becomes increasingly more difficult to accomplish as the student-faculty ratio continually grows each year. Although this certainly is a problem confronting education, it is by no means its greatest problem.

There are two other major problems facing education today. One is concerned with the inadequacy of the source materials used in the classrooms; the other is the inadequate contingency management system, which is used to induce students to study these source materials (Michael, 1972). Source materials are the materials used by teachers to transmit information to the student (e.g., textbooks, lectures, films). This study is primarily concerned with the first problem—the inadequate source materials used to teach students.

Source materials can be deficient in any one of four ways. The material may be poorly sequenced, important materials may be missing, irrelevant material may be present, and there may be no mastery criterion for advancement to new material (Michael, 1972). The mastery criterion for advancement is necessary only when a student's mastery of new material is dependent upon his mastery of the prerequisite materials. If the source material is inadequate in any of these four ways, the student will not learn effectively. Learning may result from reinforcement contingencies. Stimuli are
presented, in the form of source materials. The student responds
to these materials in some manner, and his responses are then con­sequated. Consequation is the application of consequences, reinfor­cing, neutral, or punishing, following a behavior. The consequation
could be the acquisition of new knowledge or perhaps a test admin­istered sometime later. When these antecedent stimuli are inade­quate, the student will not be able to respond to them as effective­ly as possible. Perhaps the first step in improving the effective­ness of education should be to improve the source materials used by the students.

**Improving Source Materials**

Programmed instruction concerns itself with the improvement of the source materials. It focuses on the learning process by utilizing methods that actually teach, as opposed to current methods of education that may merely serve as selection devices. Education largely teaches on the "sink or swim" principle (Skinner, 1968, 118). Those students that manage to learn without being taught struggle to the surface, those less fortunate sink to the bottom. This method of teaching obviously will never achieve the objective of education. An alternative to this method of teaching would be one which taught all students equally well. Programmed materials are an attempt to approximate this goal.

Standard textbooks are not specifically designed to teach, but rather are designed to present information (Espich & Williams, 1967). The amount of learning that a student acquires from inter-
acting with a standard textbook is both a function of the student's study habits and his ability to discriminate important material from unimportant material (Espich & Williams, 1967). Those students able to make this critical discrimination will be more likely to learn. This problem must be resolved in the methods of teaching are to be improved. Teachers can greatly reduce the student's burden by utilizing adjunctive programming techniques (Espich & Williams, 1967). These could be in the form of study objectives that coincide with the text material, or materials designed to supplement the information provided by the text. By means of these two simple devices the teacher is able to specify what the student should learn from reading the text or is able to clarify difficult and ambiguous portions of the text. The student is then able to interact with the text material without the additional task of discovering what he is expected to learn or what point the author is attempting to convey.

Programmed materials, unlike standard textbooks, are specifically designed to teach. The program contains only that information essential to the development of the student's mastery of the material. No irrelevant information is included. Consequently, the student can acquire an excellent repertoire without having to discriminate important information from unimportant information.

**Elements of Programmed Instruction**

Pipe (1966) suggests that there are four basic elements involved in programmed instruction. (1) The material to be learned is
presented in small, well-sequenced steps, (2) the student actively responds to this material in each step, (3) the student receives immediate feedback concerning the correctness of his response, and (4) he proceeds at his own rate.

Most programmed instructional materials present information in small units or frames. By means of these small steps, the student is gradually led to mastery of the material. Initial material presented in the program is relatively easy. As the student progresses through the program, new material, based on previous units in the program, is presented. Thus the student's repertoire is gradually increased as he progresses from one unit to the next. New material is introduced gradually and in small steps, thereby enabling the student to readily master it before proceeding to subsequent units. This method of material presentation can be contrasted to that employed in a standard textbook, where large amounts of material are presented. The student who reads the text must then discover a method which enables him to discriminate the important information from the unimportant in order to have a chance to effectively master the material.

The second element of programmed instruction requires that the student actively respond to each unit of material presented. His responses are overt, and therefore can be observed and consequated. When a student interacts with a textbook, his responses are usually covert. Since these responses are not readily observable, there is no effective way to monitor his behavior. Overt
responding is not an essential requirement for learning. But by requiring the student to emit responses in a program, the probability that the student is attending to the material may be increased. Thus by increasing the probability of attending behavior, the probability that the student will learn the material is also increased.

After the student overtly responds to a program unit, he receives immediate feedback, in the form of the correct answer. When the student compares his answer to the answer provided in the program, correct responses are reinforced and incorrect responses are punished. The student's behavior is consequated (reinforced or punished) at each step of the learning process. This immediate and frequent consequation results in more effective control of behavior, and thereby increases the probability that the student will learn the material.

Programmed instruction provides differential consequation for correct responding and provides indirect consequation for participation (interacting with the material). Traditional methods of teaching do not provide clear-cut consequences for correct responses to the source material. The only reinforcement for correct responses is that the student 'understands' the point the author is making. These responses are reinforced when "his reading responses are controlled by other factors than the textual stimuli and to the extent that these other factors are the important ones relevant to the author's point" (Michael, 1972, 9). This form of consequation often may not be an effective means of controlling the student's
behavior. The only direct consequence presented is for participating responses. This is usually in the form of a test grade, in which the student is tested over text material he has read. However, these participating responses may be indirectly reinforced when they facilitate correct responding. The consequences involved in reading a textbook are relatively subtle or vague, and therefore may not be adequate in controlling a student's behavior.

The fourth essential feature of programmed instruction is that the student progresses at his own rate. He proceeds through the program at a pace which is maximally effective for him. He is not forced to progress at the same rate as the 'brighter' students, nor is he delayed by the 'slower' students.

The advantages of programmed instruction seem to greatly outweigh those of traditional teaching methods. One would think that this technology would be readily adopted by education. This has not been the case. Education, like many other established institutions, is resistant to change. Industry, however, has more readily designed more effective systems to meet its changing needs. Programmed instruction was rapidly adopted by industry to teach basic essential skills to employees. Here the objectives are clearly defined, procedures are easily changed, and the resulting gains lead to subsequent administrative action. This can be contrasted with education. The objectives of an educational system are often vague and unobservable, and the methods used are extremely resistant to change. The gains from any innovation are usually
too vague to be accessed by administrators, and therefore control little or none of their behavior (Skinner, 1968). Although education might profit tremendously by using programmed instruction, educators have been slow to adopt the technology.

The Development of Programmed Instructional Materials

The development of empirically validated programmed instructional materials is an extremely expensive process, both in terms of time and money. In order to effectively develop a program, one must proceed through the five phases of systems analysis. These are: (1) stating the behavioral objectives; (2) designing; (3) implementing; (4) and evaluating the system; (5) recycling through the phases of systems analysis (Malott, 1972).

The first step in preparing a program is to state the objectives of the program in behavioral terms. One has to determine what the terminal behavior will be; what behaviors should the student be able to emit after he has completed the program. The objectives must be stated in terms of observable behavior, so that the program can be evaluated as to whether the objectives were successfully accomplished. Specification of objectives in terms of observable behavior serves four purposes (Pipe, 1966). First, it provides a means of control to direct the programmer in the construction of the program. It allows the teacher to determine whether or not the program is applicable to the specific needs of his class. Third, it provides the student with a clear statement of the goal of the program, thereby enabling him to determine his level of
success after completing the program. A statement of the program objectives also enables administrators to assess the degree of the program's effectiveness.

These are three essential features of good behavioral objectives. The objectives must be attainable, observable, and functional (Malott, 1972). When writing behavioral objectives one must always "think small". Emphasis must be placed on making the objectives small enough that both the programmer can accomplish his own objectives and that the student using the program is able to attain them. If the objectives are unreasonably large, the student's behavior will not be adequately reinforced, and consequently it will extinguish. The objectives must also be stated in terms that involve observable behavior. The student's terminal behavior must be overt, so that it can be monitored and consequated. If the objectives are not specified in terms of observable behavior, it will be impossible to evaluate the effectiveness of the program. It would not be possible to determine if the program achieves its objectives. Finally, the objectives must be functional. The student who has completed the program should have learned behaviors that will be functional in his natural environment. If the objectives do not specify functional behavior, the behaviors which are taught by the program will rapidly extinguish when the student completes the program and returns to the natural environment. Such a program would be of little or no use to anyone. An effective program teaches behaviors that are functional. Those functional be-
haviors are reinforced in the natural environment and consequently are maintained.

After specification of the behavioral objectives, the programmer must design his behavioral system or program. There are numerous kinds of programming techniques available, each designed to accomplish different objectives. The programmer must take care to select a technique which is most appropriate for his program objectives.

Once the programmer has selected a programming technique and has written the program, he must then implement and evaluate it. There are three recommended methods for testing a given program (Espich & Williams, 1967). The first method involves a one-to-one testing situation. The programmer is with a student, while the student reads through the program one frame at a time. Whenever the student encounters difficulty with a particular frame, the programmer discusses the possible sources of difficulty. This one-to-one situation yields much information which can guide the programmer's revisions, and ultimately save him a great deal of unproductive revising. After numerous one-to-one testings the programmer should administer the program, which is still in the developmental stages, to a small number of students. The results of this testing will provide the programmer with some idea of how adequately the current version of the program teaches the student. It will also yield valuable information on the areas of the program that are still weak. The third test involves giving the pro-
gram to the population for which it is designed. This test is for the validation of the program.

The evaluation of the program's effectiveness as a teaching device must be relative to the specified behavioral objectives. Often the validation test ("field test") does not meet the overall objective's criteria. If the objectives are not accomplished, the program is not successful. It is therefore necessary to recycle through the phases of systems analysis to discover and correct the problem. Perhaps the objectives of the program were too large. This is one of the common mistakes made by beginning programmers.

Programming Techniques

The programmer interested in accomplishing the specific objectives has numerous programming techniques at his disposal. The most widely utilized technique has been the constructed response frame sequence (Espich & Williams, 1967). Although this programming technique is relatively easy to develop, there are some major drawbacks that limit its effectiveness. Since the program is broken into many very small steps, students find this technique to be boring. Because of the wide range of potential student responses, the program must proceed with very small steps, where the programmer has overly-determined the student's response, thereby reducing the probability of student error. This may result in the answer to a given frame being so obvious that getting the correct answer may serve only as a very weak reinforcer (Whaley & Malott, 1971). Also with this technique, new material is presented at a slow rate.
This over-determination and slow rate of new material presentation may be two of the major reasons for student boredom with this technique.

With discrimination frame sequences the student is provided with the criteria upon which the discrimination is based, followed by a series of items. He must then select those items that meet the specified criteria. Concept learning is essentially a problem of bringing the behavior under stimulus control. The student's behavior must be controlled by the essential stimulus characteristics of the concept. Because of this technique's effectiveness in discrimination training, it is often used to teach concepts.

Perhaps the most infrequently used form of programming is the retrogressive chaining technique. This method is most effective in the shaping of new behaviors and in teaching stimulus-response chains. The procedure used in retrogressive chaining is somewhat analogous to backward chaining in that the terminal behavior is often the first behavior that is taught.

The final programming technique to be discussed is adjunct programming. The most widespread use of adjunct programs is in conjunction with the standard textbook. After reading a paragraph in the text, the student is asked a series of diagnostic questions, which focus on the main points of the paragraph. A variation of this is that the student reads a section in the text and then completes a program that is designed to supplement the text. Adjunctive devices such as these may enable the student to more effectively
discriminate important text material from unimportant material.

There are more programming techniques than the ones described above, in addition to the different variations of the aforementioned methods. The programmer should spend considerable time analyzing the strengths and weaknesses of each technique in making his decision as to what technique might best aid in the accomplishment of his program objectives.

Levels of Learning

There is a prevalent misunderstanding concerning the teaching capacities of programmed instruction. Many people believe that programs are only effective in teaching simple motor skills and transmitting verbal knowledge (Skinner, 1968). However, programming techniques can teach at five different levels of learning (Espich & Williams, 1967).

The first is the exposure level. This is the least sophisticated level of learning. When a student learns material at the exposure level, he merely knows that the material exists. It serves a similar purpose as enrichment material. Unless such material is absolutely necessary for the student to achieve the program objectives, this exposure level material should not appear in the program.

The second level of learning is the recognition level. Basically this level of learning teaches the student to make broad discriminations.

Programming can also teach material at the recall level. Recall enables a student to define terms in his own words. Much of
current education involves learning at this level.

The subsequent level involves learning at the memory level. This enables the student to recite information exactly as he has learned it.

The most sophisticated level of learning is concept learning. This level enables the student to make extremely fine discriminations and generalizations and to apply what he has learned to novel situations. This type of learning is most beneficial to the student in his interactions in the natural environment.

**Teaching Concepts**

The objective of this study was to teach specific concepts to students enrolled in the introductory psychology course at Western Michigan University. The program objectives would be accomplished when students emitted the specified conceptual behavior. Whaley and Malott (1971, 187) define conceptual behavior as a "generalization within a concept or stimulus class and discrimination between concepts or stimulus classes." The same authors also define a concept as "a set of stimuli, all of which have some common property."

Englemann (1971, 241) defines a concept as "the set of characteristics shared by a set of instances in a given universe of concepts and not shared by other instances in that universe." In order to teach conceptual learning, one must teach a double discrimination: "(1) the discrimination of relevant characteristics of instances... from relevant characteristics of non-instances...;"
(2) the discrimination of relevant from irrelevant characteristics . . . within instances or not-instances" (Englemann, 1971, 241).

The essential stimulus characteristics must acquire stimulus control over the student's behavior. When teaching a concept, one must present a set of instances and not-instances of the concept, and vary the non-essential stimulus characteristics of the instances and not-instances within the set. It is imperative that a set of concept instances and not-instances is presented as it is impossible to teach a concept with only one instance and one not-instance (Englemann, 1971). The non-essential stimulus characteristics of a set of instances and not-instances must be varied so that the learner will not mistake a non-essential characteristic as an essential characteristic for either an instance or non-instance.

A student's mastery of a given concept can be depicted as occurring along a continuum. The learning of a concept is not an all-or-nothing process. A student may partially learn the concept, in which case he would be able to effectively respond to some instances and not-instances of that concept, yet his repertoire would be deficient in such a way that he would not be able to respond appropriately to other instances and not-instances of that same concept. Most traditional textbooks do not effectively teach mastery of concepts. Their standard approach to concept teaching is to provide the student with a definition and perhaps also provide a few examples. Surprisingly enough, students may acquire a limited range of conceptual behavior with this method of teaching. They may even
learn partial conceptual behavior by reading a text, while never being tested with conceptual quiz questions. The text that is used in the introductory psychology course at Western Michigan University presents both instances and not-instances of concepts. The students acquire a substantial amount of conceptual behavior, even though they receive recognition-level quizzes. Perhaps by providing more examples of concept instances and not-instances, a higher level of concept mastery could be achieved.

Cost Analysis

From the above discussion, it should be evident that the development and validation of a program is extremely costly. It is therefore adviseable that the potential programmer conduct a cost-benefit analysis before he begins to work on the program. He must be able to justify the expense and effort necessary to produce a program that will achieve his desired objectives. Currently only mass education and industry can afford to support the development of these expensive educational materials.

A cost analysis was performed concerning the two programs developed for this study. Each program could be considered as one assignment in an introductory psychology course. The students frequently have fifty assignments per semester and pay approximately fifteen dollars for materials (texts and workbooks), which is thirty cents per assignment. Therefore the two programs could be marketed for thirty cents each, or a total of sixty cents. The royalties received by the author (programmer) might be fifteen percent.
of the sale price, or in this case, nine cents. Approximately 250 hours were spent developing, testing and validating the programs. If a professional programmer, being paid a wage of four dollars an hour, were hired to develop these programs, the cost of development would have been one thousand dollars. The programmer, or the company that hired him, would then have to sell over 11,000 copies of the programs to regain the initial investment. These costs would be substantially greater if the programmer were an experienced Ph.D., who might receive between ten and fifteen dollars an hour. However, these programs were developed by a novice, who received academic credit that applied towards a master's degree. The total cost for development was minimal. Therefore, the number of programs that must be sold to balance the cost of development, if any, is reduced tremendously. By utilizing novice programmers who receive academic credit rather than monetary payoffs for their endeavors, it is possible to develop empirically validated instructional materials without spending large sums of money. If the developed programs are marketed commercially, the programmer might receive some royalties. If no programs were sold, he would still have a master's thesis, and no money would be lost. This could prove to be a very effective and economical means of producing high quality educational materials.

The Problem

Previous investigation indicated that many students enrolled in the introductory psychology course at Western Michigan University
did not have an adequate conceptual repertoire concerning specific psychological concepts. They were not able to correctly identify concept instances from not-instances and were not able to write examples that correctly illustrated the specific concepts. It was discovered that students had difficulty in achieving the desired level of mastery for such concepts as positive and negative reinforcement, positive and negative punishment, discriminated extinction, fixed ratio and variable ratio schedules of reinforcement, extinction of behavior maintained by positive and negative reinforcement, and others. It was not surprising that the students had not developed adequate conceptual repertoires concerning the specified concepts. Although they were quizzed daily over material that dealt with these concepts, the quizzes were not conceptually oriented. The primary emphasis of these quizzes was recognition learning.

This study was concerned with the development and empirical validation of instructional materials that would teach students specific concepts at a high level of mastery. The concepts to be taught were taken from Elementary Principles of Behavior (Whaley & Malott, 1971). Ideally, these instructional materials would be used by the student after he had completed specified chapters of the textbook, although they could be used independently of text materials.

The two sets of programmed instructional materials that were developed dealt with positive and negative punishment, and fixed
and variable ratio schedules of reinforcement.

In the past, introductory psychology students had not encountered a great deal of difficulty with the concept of positive punishment. Positive punishment is the response contingent presentation of a stimulus, resulting in a decreased frequency of responding. The relative ease with which students have mastered this concept may largely be due to its 'intuitive appeal'. That is, this type of punishment procedure has been utilized and experienced by students throughout their lives to control behavior. At one time or another they have all 'punished' their dog or cat for chewing on the furniture, or for other undesirable behaviors. This punishment had probably been in the form of physical energy, such as hitting or a slap. Since this form of punishment was familiar to most students, they usually had little difficulty recognizing the concept of positive punishment. The punishment procedure always results in a decrease in response rate. In order for a stimulus to be a punisher, it must produce a decrease in the frequency of the response that it follows (Whaley & Malott, 1971).

The essential stimulus characteristics of the concept positive punishment, relative to negative punishment, are (1) decrease in frequency of a response, and (2) due to response contingent presentation of a stimulus. In teaching this concept, the essential stimulus characteristics were contrasted to non-essential stimulus characteristics in both concept instances and not-instances. Below are two frames taken from the punishment program.
Sam always watched T.V. when he came home from school. One day he turned on the tube and received a shock that threw him across the room. Now Sam listens to the radio, instead of watching T.V., when he’s home from school.

Andy always drove 10 m.p.h. over the posted speed limit. One day a cop, who was hiding behind a fire hydrant, nailed him and gave him a speeding ticket. This didn't seem to bother Andy. He continued to drive 10 m.p.h. over the posted limit wherever he went.

The first example illustrated positive punishment. It contained both essential stimulus characteristics of the concept. The second example possessed only one of the two essential stimulus characteristics, and consequently was not a concept instance. The student must be taught to discriminate relevant stimulus characteristics of concept instances from relevant stimulus characteristics of not-instances, and relevant from irrelevant stimulus characteristics in both instances and not-instances. When a student's behavior is controlled by these stimulus characteristics, he has mastered the concept.

The concept of negative punishment was also taught with positive punishment. Negative punishment is the response contingent removal of a stimulus, resulting in a decrease in the frequency of responding. The essential stimulus characteristics of negative punishment, relative to positive punishment, are (1) decrease in the frequency of a response, and (2) response contingent removal of a stimulus. Students often had difficulty discriminating an example of negative punishment from an example of extinction of behavior maintained by positive reinforcement. Below are two frames from the
punishment program:

Curly plays with his hair a lot. Whenever his hands are free, he runs his fingers through his golden locks, pulling and twisting the beautiful strands. Curly's mother finds this habit to be very obnoxious. Whenever Curly starts to play with his hair while watching T.V., his mother immediately turns off the tube until he has not touched his hair for two minutes. As a result, Curly plays with his hair less.

Zeke jumped on his chopped Harley and kicked the starter. The engine didn't start. He kicked it ten more times -- still nothing. Dejected, Zeke stopped kicking, removed his helmet, and headed for the bus stop.

The first frame was an instance of negative punishment. There was a decrease in the frequency of "hair-playing behavior" due to the response contingent removal or termination of the television picture. It contained both essential stimulus characteristics of negative punishment and consequently was a concept instance. However, the second frame was not a concept instance. Although there was a decrease in the frequency of "kicking" responses, this decrease was due to the withholding of a stimulus, rather than the response contingent removal of a stimulus. This example illustrated extinction of behavior maintained by positive reinforcement. Again, the student must learn to discriminate between essential and non-essential stimulus characteristics.

A second instructional program was developed to teach the concepts of fixed ratio and variable ratio schedules of reinforcement. A fixed ratio schedule of reinforcement is one in which a fixed or specified number of responses must be emitted in order for reinforcement to be presented. The only essential stimulus char-
acteristic of a fixed ratio schedule, relative to a variable ratio schedule, is that a fixed number of responses must be emitted in order for reinforcement to be delivered. The following frame, taken from the program, illustrates this essential characteristic.

Bad Al is a member of the high School basketball team. The coach wants Al to practice his 'foul shot' shooting. So, every evening after practice, Al practices shooting foul shots. For every 35 foul shots that Al shoots, the coach gives him a cigarette. Bad Al then heads for the bench and immediately lights up.

On a variable ratio schedule of reinforcement, reinforcement is presented after a variable number of responses have been emitted. The essential stimulus characteristic of this concept, relative to a fixed-ratio schedule, is that the individual must emit a variable or unfixed number of responses in order to produce reinforcement. This characteristic is illustrated in the following frame.

Wendy loves to play roulette. She meets with the boys every Wednesday night for a few hours of fun and excitement with the little ball. She wins a little bit, but she loses more than she wins. Sometimes it may be 10 plays between wins, sometimes it may be 50 plays between wins.

In both fixed and variable ratio schedules reinforcement is dependent upon a number of responses being emitted. Reinforcement is not dependent upon the passage of time. Therefore, a situation in which reinforcement is dependent upon the passage of an interval of time would be a not-instance of the ratio concept. The following frame illustrates this characteristic.

Ronald dined at McDonald's every day. He really dug the burgers, but didn't like to clown around waiting in line to place his order. Some days he might have
to stand in line for as long as 10 minutes before being waited on; other times he might be able to order right away.

The instructional programs were developed to teach students psychology concepts at a conceptual level. By attaining a high level of mastery at the conceptual level, students would then be able to respond effectively and appropriately in novel situations. It is important that the student has a conceptual mastery of both positive and negative punishment, so that he may utilize these procedures to modify behavior in his natural environment. There are two types of punishment procedures; one involves the response contingent presentation of a stimulus, the other involves the response contingent removal of a stimulus. In order to become an effective behavior modifier, the student should master both of these procedures at the conceptual level.

It is also important to teach to the potential behavior modifier the concepts of schedules of reinforcement and their effects upon behavior. The student should know that behaviors maintained on intermittent schedules are more resistant to extinction than those maintained on a continuous reinforcement schedule. He should also know that ratio schedules produce higher response rates than interval schedules, and that fixed schedules produce pauses after reinforcement. Perhaps the first step in teaching these concepts, so that the student can use them in effective behavior modification, is to teach the procedures of ratio schedules of reinforcement at a conceptual level.
METHOD

Subjects

The subjects were students enrolled in the introductory psychology course at Western Michigan University. These students had read the textbooks for the course, *Contingency Management in Education* (Malott, 1972) and *Elementary Principles of Behavior* (Whaley & Malott, 1971) prior to their participation in the study.

Procedure

The objectives of the two instructional programs were to teach positive and negative punishment and fixed ratio and variable ratio schedules of reinforcement at a high level of conceptual mastery. This means that the student should be able to discriminate concept instances from not-instances, and write original non-laboratory examples that illustrated each concept.

Since the purpose of the programs was to teach conceptual behavior, the discrimination frame sequence was utilized. This particular programming technique had been found effective in teaching at the concept level (Espich & Williams, 1967). In the discrimination frame sequence the student was required to discriminate concept instances from not-instances. In order to make this discrimination, the criteria on which the discrimination is based were presented to the student. A series of concept instances and not-instances were then presented. The student had to examine the cri-
teria, analyze the items presented, and make the necessary discrimination. He indicated his answer by means of an overt response (Espich & Williams, 1967).

The discrimination frame sequence emphasized the importance of the covert response. The overt response was simply emitted so that the repertoire could be more readily observed and consequtated. Overt responding was not an integral feature of the program. Perhaps the main function of the overt responding was to increase the probability that the student attended to the material presented in the program.

The programs were also intended to serve as adjunctive material supplementing the textbook and enabling the student to learn more effectively from the text. Ideally, these programs would be utilized by the student after he had read the relevant chapters in the text. The student should acquire a factual repertoire from the text and a more conceptual repertoire from the programs.

One of the major shortcomings of many programs was that they were boring. The terminal behavior may not have been achieved because the students did not complete the program (Popham & Baker, 1971). Therefore, special emphasis was placed on making the programs both functional and interesting. It was attempted to make the frames of the program as intrinsically reinforcing as possible. Humor was an important element in maintaining the students' participation with the programs. Its purpose was to reinforce attending to the program, and thereby increase the probability of learning.
Humor did not reinforce correct responses per se.

**Objectives of Punishment Program**

After completing the program, the student should be able to correctly identify examples of positive punishment, negative punishment, or neither. He should also be able to write original non-laboratory examples correctly illustrating both types of punishment.

**Initial Development of Punishment Program**

There are many behavioral procedures used to decrease the probability of a response. Among these are positive punishment, negative punishment, extinction of behavior maintained by reinforcement, and the conditioning of incompatible behaviors. These seem to be the most frequently used techniques for decreasing behavior. These are also the techniques that students have had difficulty discriminating among. In addition to the aforementioned procedures, many students erroneously believe that negative reinforcement is a procedure used to decrease the frequency of a response. Actually, negative reinforcement increases the frequency of a response. Therefore these were the areas that were dealt with in the program.

Students were taught to discriminate examples of positive and negative punishment from examples illustrating extinction of behavior maintained by positive reinforcement, conditioning of incompatible responses, negative reinforcement, and examples that involved the response contingent presentation or removal of stimuli that
did not result in a decrease in behavior.

They were taught the basic steps used in the analysis of a situation to determine whether or not it correctly illustrated positive or negative punishment. Also, before classifying each example, the student was required to answer a short series of questions that guided his analysis and subsequent answer. After the student had written his answer, he then compared it with the answer and explanation provided in the program. Analytical feedback was provided so that the student could readily identify any errors he made in analyzing a given example.

Many versions of the program were written, tested, and revised before the final format was achieved. One of the most interesting versions of the punishment program involved the extensive use of positive (+) and negative (−) sign notation. The student was taught to analyze the examples in terms of the type of consequence, and whether the stimulus in that consequence was presented, removed, or withheld. If the stimulus involved in the consequence of the behavior was aversive (a punisher), it was represented by a −. If it was a 'pleasant' stimulus (positive reinforcer), it was represented by a +. If the stimulus was presented, removed, or withheld, the procedure was represented by a +, −, or 0 respectively. The student then wrote down the two appropriate symbols; one for the stimulus, and one for the procedure. Using the basic rules of multiplication, he multiplied the two symbols together. If the result of this multiplication was a −, then the example illustrated
punishment. If the result was a+ or a 0, punishment was not involved. For example, if a response was emitted that was followed by the removal of a positive reinforcer, the stimulus involved in the consequences would be symbolized as a+, the procedure (removal of the stimulus) would be symbolized as a -. Multiplication of the two signs, (+) x (-), results in a -. Therefore this example would be an illustration of punishment.

Although this was an interesting approach, it was eventually discontinued. It could not effectively teach the student to discriminate between positive and negative punishment, and it could not deal with situations other than punishment and extinction, nor could it deal with situations that involved the response contingent presentation or removal of stimuli but with no resultant decrease in behavior. Also, many students did not have adequate mathematical repertoires to multiply two symbols together and obtain a correct result.

In the initial stages of development, a considerable amount of time and energy was spent in testing the program on relatively large groups of students. Revisions were then made on the basis of these group tests. This method of large group testing proved to be an inefficient means of developing a program, both in terms of the subjects' and the programmer's time (Baker & Popham, 1971; Espich & Williams, 1967). Subsequent program revisions were made on the basis of data obtained from one-to-one testing situations.
**Ratio Program Objectives**

After completing the program the student should be able to correctly identify examples of fixed or variable ratio schedules of reinforcement or neither. He should also be able to write original non-laboratory examples of the two ratio schedules.

**Initial Development of Ratio Program**

There are many schedules of reinforcement used to establish and maintain behavior. Among these are fixed ratio, variable ratio, fixed-interval and variable-interval schedules. Students were taught to discriminate fixed and variable ratio schedules from fixed and variable-interval schedules. Continuous reinforcement schedules were included in initial versions of the program, but were eventually excluded because, by definition of a fixed ratio schedule, a continuous reinforcement schedule could be considered a fixed ratio, where one response results in reinforcement.

The first set of frames presented the criteria on which the discriminations were to be based. In addition, diagramatic representations of response patterns and reinforcement for both fixed and variable ratio schedules were presented. The purpose of these diagrams was to provide a textual stimulus that enabled the student to directly observe the temporal relationship between responding and reinforcement in these two schedules.

After the expository material, a series of examples were presented to the student. These examples were either instances of fixed ratio or variable ratio schedules or they were not-instances.
The not-instances consisted of approximations of fixed or variable-interval schedules. Since the program did not deal with laboratory situations, it was difficult to find examples of "pure" schedules operating in the natural environment. There are very few "pure" fixed-interval or variable-interval schedules controlling behavior in the natural environment (Whaley & Malott, 1971).

The first set of examples contained questions the student had to answer prior to classifying the given example. These questions provided specific steps for the student to follow when analyzing the example. As the student progressed through the program, these questions were gradually omitted.

Since the ratio program was being developed concurrently with the punishment program, considerable time was also spent with large group testings. Later program revisions were made on the basis of one-to-one test results.

One-to-One Testing

After the program was initially developed, the frames were put on 3" x 8" index cards and placed in a ring notebook. The program consisted of information frames and test frames. The information frames presented relevant information to the student. The test frames consisted of examples, either concept instances or not-instances, and a small series of questions after each example. On the back of the test frames were the correct answers, with a short explanation of why the answers were correct. The index cards were arranged so that only one frame could be viewed at a time.
The program was tested and developed with individual students on a one student to one programmer basis. Students who participated in this test-development were volunteers from the introductory psychology course. They received bonus points for participating, which were applied to their final grade in the course. Each student was informed that the purpose of this activity was to discover the weaknesses and inadequacies of the program being developed. The student was instructed to provide verbal feedback to the programmer whenever he encountered anything unclear or ambiguous.

After these instructions, the student took a ten-question pre-test. The definitions of the concepts being tested were provided in the pre-test. The first eight questions consisted of concept instances and not-instances. The student had to determine if the given example was a correct concept instance or a not-instance. The student read each example and vocalized his answer. The answers were recorded by the programmer. The last two questions on the pre-test required the student to write original non-laboratory examples of the concepts being tested. The student was given as much time as necessary to complete the pre-test.

When the student had completed the pre-test, he was instructed to go through the program one card at a time. He was informed that, after he had answered a test frame, he should compare his answer to the answer provided on the back of the card. If his answer did not correspond to the answer in the program, or if he did not understand why his answer was incorrect, he was told to stop and
discuss the problem with the programmer. The student's answers were recorded by the programmer. After the student completed the program, he then filled-out an evaluation of the program, indicating the helpful, unhelpful, and unclear frames.

He then took a post-test similar to the pre-test. The two tests were then graded, and the programmer discussed with the student the item he had answered incorrectly.

This one-to-one testing situation, with direct student feedback, proved to be a source of invaluable information. Poorly-written frames were immediately detected and either rewritten or replaced. Problem areas were spotted and supplementary frames added as the need arose. This method of program development was very effective. After numerous individual testings and revisions, the program was ready for validation testing.

**Validation Testing**

There were two test forms, 1 and 2, used as the pre-test and post-test. These were the same tests used in the development of the program. Six different sets of materials were assembled and put into large envelopes. The content of each set of envelopes was as follows:

- P1 - pre-test form 1, program, post-test form 2
- P2 - pre-test form 2, program, post-test form 1
- B1 - pre-test form 1, excerpts from *Elementary Principles of Behavior*, post-test form 2
- B2 - pre-test form 2, excerpts from *Elementary Principles of Behavior*, post-test form 1
N1 - pre-test form 1, post-test form 2
N2 - pre-test form 2, post-test form 1

The assembled envelopes were placed in random order and distributed to the students as they entered the testing room. Students with the same pre-test form sat in alternate seats. Students were required to participate in the validation testing, and received 1/2 of a quiz point for each correct answer on the pre and post-tests. These quiz points counted toward the students' grades in the introductory course.

Due to time limitations, students were not allowed to progress at their own rate. They were allowed to work on the pre-test for ten minutes. At the end of this time, they placed the pre-test back in the envelope and began working on the second section.

The second section consisted of the program for Group P. Group B read relevant portions from chapters in *Elementary Principles of Behavior* (Whaley & Malott, 1971), which dealt with the concepts being tested. These excerpts were approximately the same length as the program. Group N did not have a second section. Students in this group were required to remain in their seats, and were not allowed to read any psychology materials. Students had thirty minutes to complete the second section.

At the end of this thirty minute period, all students were required to place section two materials in the envelope. They were then allowed ten minutes to complete the post-test. At the end of this ten minute period, all materials were placed in the envelope.
and then collected.
RESULTS AND DISCUSSION

The program produced significant differences between pre and post-test scores for the eight objective questions, two subjective questions, and combined ten questions (overall). Neither of the other two groups showed significant differences between pre and post-test scores in any of three areas analyzed. Table 1 shows the summarized results for the three groups. Table 2 indicates that there were significant differences between pre and post-test scores in each of the three areas for the program group, while differences between pre and post-test scores for the group that read portions of the text were not significant. The group that took only pre and post-tests showed a significant difference on the objective portion, but did not show significant differences on the other two portions.

The significant difference obtained by the group that only took the tests may possibly have resulted because the students changed their method of 'guessing' on the post-test. Since they did not receive any feedback concerning the correctness of their pre-test answers, it was unlikely that they 'learned' a significant amount by taking the pre-test.

A detailed error analysis was performed for the pre and post-tests for each program. This was done by first categorizing the questions in each test according to the correct answer. When the correct answer was "neither", the questions were classified as to
## RESULTS OF PUNISHMENT VALIDATION TESTING

<table>
<thead>
<tr>
<th>Questions</th>
<th>Group</th>
<th>Pre-test Mean</th>
<th>Post-test Mean</th>
<th>D</th>
<th>t</th>
<th>S_d</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 Objective</td>
<td>P</td>
<td>47.1%</td>
<td>70.6%</td>
<td>23.5%</td>
<td>5.63</td>
<td>.334</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>B</td>
<td>56.3%</td>
<td>58.8%</td>
<td>2.5%</td>
<td>0.61</td>
<td>.330</td>
<td>p &gt; .05</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>49.3%</td>
<td>54.9%</td>
<td>5.6%</td>
<td>1.51</td>
<td>.299</td>
<td>p &gt; .05</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>38.5%</td>
<td>65.1%</td>
<td>26.6%</td>
<td>5.45</td>
<td>.099</td>
<td>p &lt; .05</td>
<td></td>
</tr>
<tr>
<td>2 Subjective</td>
<td>B</td>
<td>47.5%</td>
<td>47.5%</td>
<td>0.00%</td>
<td>0.00</td>
<td>.205</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td>N</td>
<td>41.7%</td>
<td>50.0%</td>
<td>8.3%</td>
<td>0.84</td>
<td>.202</td>
<td>p &gt; .05</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>45.4%</td>
<td>69.6%</td>
<td>24.2%</td>
<td>6.67</td>
<td>.363</td>
<td>p &lt; .05</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>B</td>
<td>54.5%</td>
<td>56.5%</td>
<td>2.0%</td>
<td>0.48</td>
<td>.414</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td>N</td>
<td>47.8%</td>
<td>53.9%</td>
<td>6.1%</td>
<td>1.31</td>
<td>.389</td>
<td>p &gt; .05</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 2

RESULTS OF RATIO SCHEDULES VALIDATION TESTING

<table>
<thead>
<tr>
<th>Questions</th>
<th>Group</th>
<th>Pre-test Mean</th>
<th>Post-test Mean</th>
<th>D</th>
<th>t</th>
<th>S&lt;sub&gt;D&lt;/sub&gt;</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>57.0%</td>
<td>73.5%</td>
<td>16.5%</td>
<td>2.56</td>
<td>.516</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>8 Objective</td>
<td>B</td>
<td>60.8%</td>
<td>58.3%</td>
<td>-2.5%</td>
<td>0.47</td>
<td>.431</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>44.5%</td>
<td>61.0%</td>
<td>16.5%</td>
<td>2.97</td>
<td>.444</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>44.0%</td>
<td>69.0%</td>
<td>25.0%</td>
<td>3.88</td>
<td>.129</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>2 Subjective</td>
<td>B</td>
<td>30.0%</td>
<td>43.5%</td>
<td>13.5%</td>
<td>1.02</td>
<td>.266</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>50.0%</td>
<td>47.0%</td>
<td>-3.0%</td>
<td>0.21</td>
<td>.281</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>54.4%</td>
<td>72.5%</td>
<td>18.1%</td>
<td>3.13</td>
<td>.579</td>
<td>p &lt; .05</td>
</tr>
<tr>
<td>Overall</td>
<td>B</td>
<td>54.7%</td>
<td>55.3%</td>
<td>0.6%</td>
<td>0.10</td>
<td>.597</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>45.6%</td>
<td>58.1%</td>
<td>12.5%</td>
<td>1.61</td>
<td>.714</td>
<td>p &gt; .05</td>
</tr>
</tbody>
</table>
what other concept they represented (see Tables 3 and 4). Then the probability of each of the possible incorrect responses was computed for each category of question. The probability was:

\[
\frac{\text{(number of errors of particular type for a given category of question)}}{\text{(number of questions in that category) x (number of students taking quiz)}}
\]

For both programs the pre-test results were essentially the same for all three groups. The two control groups (Groups B and N) showed no substantial improvements for any single type of error. Therefore only the data for the experimental groups (Group P) are shown in Tables 3 and 4.

In the punishment program, all groups had a high probability of mistaking negative punishment, negative reinforcement, and differential reinforcement of incompatible behavior for positive punishment; and positive punishment and extinction for negative punishment. However, Group P showed a decreased probability of mistaking positive punishment for negative punishment, and negative reinforcement for positive punishment. The program failed to produce significant decreases in the probability of mistaking negative punishment and differential reinforcement of incompatible behavior for positive punishment, and extinction for negative punishment.

In the ratio program, all groups had a high probability of mistaking 'fixed interval' schedules for fixed ratio schedules, and 'variable interval' schedules for variable ratio schedules. The program produced substantial decreases in the probability of these errors,
<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Positive Punishment</th>
<th>Negative Punishment</th>
<th>Neither</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
<td>Pre-test</td>
</tr>
<tr>
<td>Positive Punishment</td>
<td>.63</td>
<td>.96</td>
<td>.37</td>
</tr>
<tr>
<td>Negative Punishment</td>
<td>.30</td>
<td>.24</td>
<td>.65</td>
</tr>
<tr>
<td>Negative Reinforcement</td>
<td>.40</td>
<td>.13</td>
<td>.02</td>
</tr>
<tr>
<td>Extinction</td>
<td>.18</td>
<td>0.00</td>
<td>.74</td>
</tr>
<tr>
<td>Differential Reinforcement of</td>
<td>.67</td>
<td>.33</td>
<td>.15</td>
</tr>
<tr>
<td>Other Behavior</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'Punishment Procedure'</td>
<td>.15</td>
<td>.07</td>
<td>.07</td>
</tr>
<tr>
<td>But No Behavioral Decrease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stimulus</td>
<td>Fixed Ratio</td>
<td></td>
<td>Response</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------</td>
<td>--------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
<td>Pre-test</td>
</tr>
<tr>
<td>Fixed Ratio</td>
<td>.94</td>
<td>.88</td>
<td>0.00</td>
</tr>
<tr>
<td>Variable Ratio</td>
<td>.03</td>
<td>.03</td>
<td>.85</td>
</tr>
<tr>
<td>'Fixed Interval'</td>
<td>.47</td>
<td>.31</td>
<td>.19</td>
</tr>
<tr>
<td>'Variable Interval'</td>
<td>0.00</td>
<td>0.00</td>
<td>.81</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>.38</td>
<td>.13</td>
<td>.31</td>
</tr>
</tbody>
</table>
yet the probability of error was still high (greater than .20).
It is now evident that the groups did not have difficulty discriminating fixed ratio from variable ratio schedules of reinforcement. They did have difficulty discriminating 'fixed interval' from fixed ratio and 'variable interval' from variable ratio schedules. Perhaps it would be more effective to teach these four schedules simultaneously, rather than teaching them in two separate programs.

Another interpretation could be that the programs simply taught students to answer the type of questions that appeared on the post-tests. This is a reasonable interpretation. However, the desired terminal behavior the student was to acquire after completing the program was that he be able to correctly answer such questions, and write his own examples illustrating the concepts taught. In this sense, he was taught to answer such questions. But those were the program objectives.

The standard argument against the use of multiple choice questions, objectives, and tests is that the multiple choice format can not teach at a conceptual level. Many people feel that a multiple choice format is only effective in teaching at the exposure or recognition levels. However, by means of multiple-choice program formats, students were able to learn how to write original non-laboratory examples, even though they were not required to do this in the program.

As previously mentioned, the scores for students in Group P were not as high as were desired. The validation testing of the
punishment program was conducted on the day prior to the final examination in the introductory psychology course. Students were required to participate in this testing. The students who had 90% or more of these "in-class" points could receive an "A" in the course if they received an "A" on the final exam. Since many of these students already had 90% of the "in-class" points, the ten points that they could receive for scoring high on the pre-test and post-test may not have been an effective reinforcer for participating in the activity.

Because of time limitations, students were not allowed to proceed at their own rates during the validation testing. They were only allotted ten minutes to complete the pre-test, thirty minutes for the program or textbook materials, and ten minutes for the post-test. These restrictions may not have allowed some students to score as high as possible on either the pre or post-tests. Some students failed to answer all ten questions on the tests. This could have been due to lack of time, an inadequate repertoire, or inadequate motivation.

The eight objective questions on both pre and post-tests were machine scored. The machine was checked for scoring accuracy. A few errors were detected in the scoring of the ratio schedules program tests. These tests were rescored by hand. No errors were detected in the scoring of the punishment program tests. The two subjective questions on the test were scored by the programmer.
A reliability check was conducted by another graduate assistant.
assistant in the introductory course, to determine the accuracy of
the scoring of the tests for both programs. He rescored all of
the tests. The pre and post-test reliabilities for the punishment
program were 91.8% and 91.9% respectively. The reliabilities ob-
tained on the pre and post-tests of the ratio schedules program
were 95.2% and 87.1% respectively, or a mean reliability of 91.2%.
These reliability scores may be biased because the programmer had
initially marked on the test papers the answers that were incor-
rect. When the reliability check was conducted, these marks may
have biased the graduate student's grading.
CONCLUSIONS

Both the punishment and ratio schedules programs produced a significant difference between pre and post-test scores for students who read the programs. They scored significantly higher on the post-tests in each of the three areas; objective, subjective, and overall. Students who read portions of the course text did not demonstrate a significant difference between pre and post-test scores in any of the three areas on either punishment or ratio schedules. Similar results were obtained for the group that took both tests, without reading any materials. However, this group did show a significant difference between pre and post-test scores on the objective portion of the ratio schedules test. The two programs taught the specified concepts at a higher level of mastery than that obtained by the textbook alone.
REFERENCES


