

1986

# Teaching complex skills in a PSI psychology course : a thesis ...

Robert Alan Kutner
*University of the Pacific*

# Teaching Complex Skills in a PSI Psychology Course

A Thesis

Presented to

the Faculty of the Graduate School

University of the Pacific

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

by

Robert Alan Kutner

June 1986

This thesis, written and submitted by

Robert Alan Kutner

is approved for recommendation to the Committee on Graduate Studies, University of the Pacific.

Department Chairman or Dean:

Thesis Committee:

M L Davis Chairman

Esther Cohen

Dated May 12, 1986

Acknowledgements

I am grateful and indebted to the members of my thesis committee consisting of Michael Davis (Chairperson), Kenneth Beauchamp, and Esther Cohen for their valuable contributions and continued interest in this arduous research project. I am especially indebted to Kenneth Beauchamp for allowing me to use his Self-Control course for this project and his patience in both designing and grading the course materials. I would also like to thank Paul Vinciguerra, Esther Cohen, and Michael Davis for serving as item-raters. Your availability and expertise were invaluable. This thesis project is the culmination of several years of correspondence between Michael Davis and myself, and Michael deserves special thanks for his never ending support and encouragement. Thank you all very much for your contributions.

## Abstract

The Personalized System of Instruction (PSI) is designed to individualize instruction based on traditional learning theories. Students are required to demonstrate mastery before advancing to new material. A self-pacing feature allows students to dictate their rate of progress. Compared to lecture-discussion instruction, PSI courses have demonstrated superior examination performance as well as increased ratings of course quality. However, studies have been criticized for testing only basic skills while ignoring more complex processes.

In this research project, the PSI study guides were designed to emphasize complex processes and mastery test and review examination questions reflected increased item-level complexity.

Results showed that students were able to master these complex items at the required 90% criterion. Performance on the comprehensive review examinations was slightly lower for complex items.

Expected differences relating to the three group sequence requirements were not obtained. Nevertheless, mastery performance on the complex items was achieved by all students regardless of experimental group.

# Teaching Complex Skills in a PSI Psychology Course

The Personalized System of Instruction (PSI) was developed by Keller (1968) as a method of individualizing instruction on the basis of traditional reinforcement and learning theories (Skinner, 1954). Skinner's theory of learning suggested that, in order to teach effectively, one must present the material to be learned in small, sequential amounts with all terminal behavioral objectives well defined. The learning situation must also be structured such that every learner will receive immediate feedback on performance, and that the learning environment be relatively free of punishing circumstances which would inhibit student achievement.

Using these principles, Keller (1968) identified some fundamental characteristics which became the basis of PSI. First, he emphasized the written word for effective communication between the teacher and the learner. Second, he created the concept of unit perfection which required that a student demonstrate mastery before advancing to new material.

In this context, tests are taken by students as many times as necessary on a given unit until a threshold level of mastery is achieved (usually 85 to 100%). Third, and most important, was the self-pacing feature which allowed a student to dictate the rate at which units in the PSI course were attempted. Students moved through the course at a rate which was most comfortable for them while simultaneously mastering the course materials. Other important features of PSI were the use of lectures as a vehicle of motivation rather than as a source of critical information, and the reliance on students to serve as peer-tutors or proctors in the course. The proctor served the role of test-giver and test-grader, tutor and peer advisor, and as a provider of critical feedback related to a student's performance in the course.

Since its inception, PSI has generated considerable empirical research. In comparison to lecture-discussion instruction, the PSI courses have been reported to result in superior examination performance (e.g., Born, Gledhill, & Davis, 1972; McMichael & Corey, 1969; Sheppard & MacDermot, 1970). Kulik, Kulik, and Cohen (1980) examined 312 reports on

instructional technology and college teaching. They reported that on measures of student achievement, ratings of course quality, course completion, and the correlation between aptitude and achievement, students clearly favored teaching which employed instructional technologies such as PSI, competency-based instruction and programmed instruction. Of these technologies, PSI studies reported stronger results than the other studies in the category of student achievement which specifically measured examination performance. In summary, Kulik, et al., suggested that not all technologies are equal in their results and that of the technological approaches studied, Keller's PSI had the most pronounced effect on student ratings of instruction.

Testing for Complex Learning in PSI Courses

Studies of PSI have often been criticized for testing only simpler skills; that is, critics have argued that testing has been limited to direct recall of factual information and that more complex processes such as comprehension, application, analysis, synthesis, and evaluation have been ignored (Austin & Gilbert, 1974; Keller & Sherman, 1974). These

criticisms may appear valid for specific courses taught in the PSI format. However, as Semb, Conyers, Spencer, and Sanchez-Sosa (1975) point out, a review of traditionally taught courses may warrant the identical criticism.

The complexity of learning which occurs in any course is not inherently a function of the type of instructional approach utilized. In order to test a student's performance on more complex skills, Watts (1973) has suggested that the focus be shifted to the type of test questions the student is required to answer. He suggested that using questions which force the student to go beyond the literal content of instruction will facilitate learning by promoting rehearsal and application of the instructional materials, all of which result in a greater degree of information processing by the student. Questions that require a student to apply concepts or principles can have a similar effect.

Research examining the effects of different types of test questions on achievement has been rare (Andre, 1979). The manner in which questions influence achievement remains unspecified. Studies previously

published have attempted to equate research on depth-of-processing with support for a level-of-questions effect (e.g., Anderson, 1970). In depth-of-processing research, perceptions of stimuli are analyzed at a number of levels or stages. Preliminary stages focus on physical or sensory features while later stages are concerned with matching the stimuli against stored abstractions from past learning. Later stages emphasize pattern recognition and extraction of meanings (Craik & Lockhart, 1972). This concept of processing stages is often referred to as "depth of processing" where greater depth implies a greater degree of semantic or cognitive analysis by the subject. Since retention of information is related to depth-of-processing, various factors such as amount of attention paid to the stimulus and the processing time available will both affect the depth at which information is processed. In a classroom demonstration of depth of processing, Chaffin and Herrmann (1983) found that simply repeating information did not improve long-term retention but that thinking about the meaning of a word did. After a five-word list was read to the

class, one of two interpolated tasks were performed. One group repeated the list rapidly over and over while the other group was told to say "hello" rapidly for 15 sec. Results demonstrated that recall was greater for students who received the "hello" requirement. After instructing the students to devise a way to do well on this immediate recall task and suggesting they use their ingenuity or some mneumonic devise the students in the "hello" list group were more successful in transferring much of this information from short-term to long-term memory while those in the repetition group were not nearly as successful.

The level-of-questions effect states that as the level of complexity of an item increases, the greater the depth at which the information is processed. The differentiation between depth-of-processing and level-of-question effects is basically a matter of semantical differences as opposed to a real theorectical difference. The hypothesized effect is that information may be processed along a hierarchy of depth levels ranging from superficial processing of perceptual features to processing for the meaning of

information. The greater the depth of processing, the higher the probability the material will be retained and later recalled (Anderson, 1970; Craik & Lockhart, 1972).

Andre (1979) reviewed studies which examined the effects of requiring both grammar and high school students to answer questions which prompted greater depths or levels of processing. Of interest was whether answering higher-level questions while reading would facilitate productive learning. Specifically, what effect would different levels of questions asked during instruction have on recall and test peformance. Research suggested that posttest performance was greatly enhanced when these questions were inserted either in, before, or immediately after reading passages. Several other researchers (e.g., Duell, 1974; Frase, 1968; Rickards & DiVesta, 1974; Watts, 1973) have also reported that asking higher-level questions generally facilitated posttest performance. In the Rickards and DiVesta (1974) study, grammar school students read a passage and were asked either rote factual questions, rote idea questions, or meaningful learning questions which required the use

of supporting sentences in the passage to justify the general assertion. When questions were inserted after every two paragraphs, meaningful learning questions lead to greater recall than did the rote factual or rote idea questions. These questions appeared to have this effect by directing attention to more specific information. Having attended to more information, students receiving higher-level or meaningful learning questions tended to recall more.

Several studies have employed Bloom's (1956) taxonomy of educational objectives in selecting a desired level of questioning. According to Bloom's taxonomy, educational objectives or tasks can be arranged in classes from simple to complex in a hierarchical manner. Within this hierarchy, six distinct classes of objectives have been defined. The first, and most basic educational task, involves the recall of specifics and facts in a given situation. This class has been defined as knowledge by Bloom. The second class involves an ability to communicate or interpret factual information and has been labeled as comprehension. Application, the third class, involves the use of abstract concepts in concrete situations.

In this task, a student must correctly apply an appropriate abstraction to a new problem without having to be prompted as to its selection or useage. The fourth class, analysis, involves the breakdown of a task into its elements or parts such that detection of the relationships of the parts can be discovered. Synthesis, the fifth class in the hierarchy, involves arranging elements or parts as to form a pattern or structure not initially identifiable. The final, and most complex educational task is evaluation. This task involves making judgments about the value of some idea, work, solution, or method. For example, it might require the use of criteria such as standards for appraising whether the idea or work is accurate, effective, or satisfying. In a study by Hunkins (1969), sixth-grade students learned social study materials containing "knowledge-level" or "evaluative-level" questions. At the end of the period they took a posttest containing questions at all six levels in the Bloom taxonomy. Students who received evaluative-level questions during instruction did better on new evaluative-level questions on the posttest. In another study using high school seniors,

Watts and Anderson (1971) examined the effects of requiring students to apply what they had read to some new situation. Using three types of inserted questions (repeated examples, new examples, or recall of factual information) they found that students in their "application-level" questions group were better able to transfer their knowledge of the concepts and principles to new examples than those receiving only repeated examples or factual information type questions.

PSI Study Guide Questions as Independent Variables

A major component of PSI is the study guide including study questions and instructions. The study guide questions can be classified in terms of their levels of complexity using Bloom's (1956) taxonomy. As an independent variable, the study guide questions can be selected to prepare the student to answer either primarily recall or higher level questions. Although study guides have been found to facilitate learning and retention (Hinton, 1978), if the questions are too difficult the student will not perform well. On the other hand, if the questions prove to be too simple, a student may have difficulty

using the information to help answer more complex questions based on the identical information. The usage of Bloom's taxonomy in structuring study guide questions has not previously been reported in the literature (Appendix A).

Since the conditions under which higher-level questions produce better transfer to new situations and greater recall are not yet totally understood, a secondary question was raised. Will the order in which a student is asked to study for and answer more complex questions make a difference in his/her ability to answer such questions? That is, will the student who experiences more complex material earlier in the course outperform those students having only attended to simplier questions?

Using Bloom's (1956) taxonomy, students in the present study were required to answer questions coinciding with various levels of the educational hierarchy as judged by independent raters. Three different learning sequences were followed by each of the three experimental groups in this study, each representing a different degree of item-level complexity. Students were expected to be able to

correctly answer higher-level questions similar to those on the study guides. Each student was assigned study questions, included in the study guides, that were at the knowledge-comprehension (KC) level, knowledge-comprehension and application-analysis (KC & AA) levels, or knowledge-comprehension, application-analysis, and synthesis-evaluation (KC, AA, & SE) levels. The students were assigned to three experimental groups each of which was required to answer questions at the various levels on a predetermined schedule. Thus, during the initial phase of the course, some students were required to answer primarily KC level questions, some KC and AA questions, and others KC, AA, and SE level questions. The dependent variables included the number of attempts needed to pass each unit test, the level of mastery on each unit test (percentage correct), the amount of time required to complete each test, and performance on the major comprehensive examinations given throughout the course.

# Method

## Subjects

The participants in this study were 55 freshman and sophomore students registered for an introductory, general education course entitled Self-Control (APY 73, Fall, 1980). These students had no prior knowledge that a research project was being initiated at this time. Of the 55 students who initially registered for the course, 38 completed all course requirements. Included in this number were 21 females and 17 males. Each student was randomly assigned to one of three experimental groups. The number of withdrawals by experimental group were 5, 6, and 6 respectively.

## Staff

The course instructor, a full professor of psychology, presented all lecture material, discussion groups, and graded the major examinations. Two graduate assistants in psychology served as proctors and PSI test administrators. Each graduate student scheduled 20 hours per week to work in the testing center administering and grading unit mastery tests throughout the semester.

## Setting

Lectures were given twice weekly for 1 hour in a large classroom adjacent to the Psychology Department. All PSI quizzes were given in a smaller classroom designated the PSI Testing Center. Course materials, including unit quizzes and study guides were located in the testing center. Operating hours were posted weekly on the testing center bulletin board with the center being open approximately 40 hours per week.

## Introductory Class Meeting

During the first class meeting, students were given basic information on the research project by the course instructor. Students were told why this particular class was selected for the experiment and were informed that the experiment would have no effect on course grades. Students were told that they would be asked to study for and answer both simple and complex questions, all of which were included in the study guides for each unit of material. Students were also informed that the class would be divided into three experimental groups with differing study questions and unit mastery tests. At this time students were given a research participation consent form (Appendix B) which all agreed to sign.

During the first week of classes, the graduate student proctors informed each student of their group assignments. The students were also told that an optional follow-up examination would be given after the course was completed. If they chose to particpate, each student would receive a $5.00 payment for their effort.

PSI Procedures

Students were given study guide materials which instructed them to study a particular set of questions taken from the course readings. After the material was studied, the student would report to the PSI Testing Center for a unit mastery test. Immediately following the completion of the test, one of the proctors would score it and inform the student of the outcome. If the student answered 90% of the questions correctly, a "pass" grade was assigned for the unit. Otherwise, the student was required to re-study, then re-take an alternative test covering the same material. Upon completion of each unit test at 90% correct, the next study guide in the sequence became available.

The graduate student proctors graded approximately equal percentages of PSI tests. Proctor

A graded 281 tests (40%) while Proctor B graded 263 tests (38%). The course instructor graded 154 tests (22%).

Course Materials

The two textbooks used in the course were divided into 15 distinct units, each with approximately 25-30 pages of reading material. Units 1 through 9 were taken from the 12 chapters in the text by Watson & Tharp (1977). Units 10 through 15 were taken from the 5 chapters in the text by Chance (1979). A student needed to pass an average of at least one and one-half units per week throughout the 10 week semester in order to finish on time. Critical dates or deadlines were imposed to help curb student procrastination. In most cases, a student needed to pass two or three units between those specified dates or the instructor suggested that the student withdraw from the course.

Study Guides

Three different study guides were written for each unit of material, and students received a study guide designed for their particular experimental group. The study guide and questions for each unit were designed to emphasize the level of questions

asked on the unit mastery tests. Students in Group One received study guides which emphasized only basic knowledge and comprehension for the first 10 units of material or Phases One and Two of the course (see Table 1). The remaining five units (Phase Three) covered basic knowledge and comprehension information plus questions pertaining to more complex educational objectives such as application, analysis, synthesis, and evaluation.

The second experimental group, Group Two, received a different set of study guides for each five unit phase. This group first received study guides at the knowledge-comprehension level. Study guides covering the next five units emphasized questions which required students to use both knowledge-comprehension and application-analysis skills. Study guides for the remaining five units emphasized all of the above skills plus synthesis and evaluation. The last phase, units 11-15, was identical to that of the first experimental group.

The third experimental group, Group Three, received study guide material emphasizing the most complex level of study questions for all 15 units of material (see Table 2).

Table 1

Mastery Test Item Format and Taxonomy Sequences

| Experimental Group | Phase 1 | Phase 2 | Phase 3 |
|---|---|---|---|
| One (Knowledge-Comprehension) | 15 KC | 15 KC | 5 KC |
| | *[3 AA | *[3 AA | 5 AA |
| | 3 SE] | 3 SE] | 5 SE |
| n=12 | | | *[6 UA] |
| Two (Application-Analysis) | 15 KC | 8 KC | 5 KC |
| | *[3 AA | 7 AA | 5 AA |
| | 3 SE] | *[3 SE | 5 SE |
| n=12 | | 3 UA] | *[6 UA] |
| Three (Synthesis-Evaluation) | 5 KC | 5 KC | 5 KC |
| | 5 AA | 5 AA | 5 AA |
| | 5 SE | 5 SE | 5 SE |
| n=14 | | | |

Note. - Abbreviations: KC = knowledge-comprehension, AA = application-analysis, SE = synthesis-evaluation, UA = un-answerable or nonsense items. Phase 1 = Units 1-5, Phase 2 = Units 6-10, Phase 3 = Units 11-15.
*- For the items in brackets, students were not graded on their answers.

Table 2

Emphasis on Study Guide Materials

| Experimental Group | Phase One | Phase Two | Phase Three |
|---|---|---|---|
| One | Knowledge-Comprehend | Knowledge-Comprehend | Synthesis-Evaluation |
| Two | Knowledge-Comprehend | Application Analysis | Synthesis-Evaluation |
| Three | Synthesis-Evaluation | Synthesis-Evaluation | Synthesis-Evaluation |

Mastery Test Items

A panel of raters, two psychology professors and one graduate student, categorized each test item as fitting one of the aggregated groups; knowledge-comprehension, analysis-application, or synthesis-evaluation (see Appendix A). All test items were taken directly from the reading material and were written by the course instructor and the research assistant. Two out of three raters had to agree on the item's categorization before it was accepted into the item pool for a particular unit mastery test. Items about which the raters could not agree were rewritten and re-rated by the panel until agreement was reached. A few items were discarded because of a lack of agreement. All test items were written in either multiple-choice, fill-in, or short-answer essay format. The majority of knowledge-comprehension items were written in either multiple-choice or fill-in format. The application-analysis items were primarily written as short-answer essay type with occasional usage of multiple-choice items. The synthesis-evaluation items were, for the most part, written exclusively in short-answer essay format.

Mastery Tests

Throughout the first 10 units of material (Phases 1 and 2), mastery tests given Groups One and Two contained 21 items each. Only the first 15 items were graded. Group Three was given a mastery test containing 15 items (see Table 1, pg. 19). In each case, a student was required to correctly answer 13.5 out of the first 15 questions presented in order to pass the test. The remaining six items on the tests given Groups One and Two served an experimental purpose, related to the original research questions, and had no effect on the outcome of the test grading for the student. On certain tests, these six items were used to assess a student's performance on items of greater complexity than were required in the study guide instructions. For example, with a study guide emphasizing simple recall (knowledge-comprehension), the additional questions assessed a student's performance at a more complex level. For those students already receiving more complex questions, these extra items served to equalize the number of items on each test across Groups One and Two. An extra item might cover a topic unrelated to the

material presently being studied. Students were asked to answer these items, but were told they would not be counted toward the 90% mastery criterion. These items were categorized as unanswerable (see "UA" in Table 1, pg. 19).

Comprehensive Examinations

Each phase of the course was followed by a 70-item comprehensive review examination which covered the previous five units of material. Each examination contained a proportion of items drawn from the item pool for each unit. Fifty-percent (approximately seven from each unit) of the items were at the knowledge-comprehension (KC) level; thirty-percent (approximately four per unit) were at the application-analysis (AA) level; and twenty-percent (approximately three per unit) were at the synthesis-evaluation (SE) level. All three review examinations were given to the class as a whole with no time limitation imposed. The first review examination was given 5 weeks into the semester with the two remaining review examinations spaced about 3 weeks apart. Immediately following the completion of the exam, the course instructor and the graduate

research assistant began scoring the examination. Each grader scored the same question on each examination until all questions were graded. This procedure enhanced scorer reliability on the more complex items. Partial credit was available for the short-answer essay questions.

Final and Follow-up Examinations

A comprehensive final examination covering all material presented throughout the course was administered to all students during the last week of the semester. This examination contained questions previously used on the three review examinations. With 107 total items, 140 points were possible on this examination. The majority of items counted one-point apiece although a few of the more complex items were worth two or three points apiece. The examination contained 81 knowledge-comprehension (KC) questions at one-point each, 15 application-analysis (AA) questions with a few two-point items, and 11 synthesis-evaluation (SE) questions with approximately one-third of the questions worth either two or three points apiece.

Following a 2 month break, students were contacted by phone and asked to return to the psychology department to take a follow-up examination. This examination was given to students at their convenience and each student received $5.00 for completing the exam. The format of this examination was almost identical to final examination. On the follow-up examination, with a total of 104 items, 142 points were possible. Of these, 80 items worth one-point apiece were knowledge-comprehension (KC), 15 items at two-points each were application-analysis (AA), and the 12 synthesis-evaluation (SE) items included eight items worth three-points apiece. Approximately 80% of the items on the follow-up examination had appeared on a previous midterm or final examination. Both examinations were graded using the same procedure described in the comprehensive review examination section.

Experimental Design

The experimental design employed in the study was a split-plot factorial, denoted as an SPF 3.35 using Kirk's (1968) system, containing three sequences of question level complexity, three testing phases with

five unit tests per phase. Table One (pg. 19) displays the type of questions being emphasized for each experimental group through the three phases of the course (units 1-15). In the first sequence which was defined by the types of questions answered by each student, those students in Group One were required to answer and were only graded on the most basic skills in Bloom's taxonomy for the first two phases of the course (10 units) and ended with grading on all three levels of complexity during the last phase (5 units). Experimental Group Two followed a progressively more difficult question answering and grading sequence which began with grading on only basic skills (5 units) and gradually required terminal graded performance on the most complex items in a package similar to that of Group One. Students in the third experimental group, Group Three, were graded on all three levels of item complexity throughout the three phases of the course (units 1-15).

## Results

Prior to an analysis of the primary dependent measures, an analysis of variance (ANOVA) was

performed to determine whether the apriori assignment procedures resulted in differences in the GPA of the three experimental groups. Student grade point averages were obtained upon completion of the course and compared. On a 4.0 scale, students assigned to Group One (n=12) averaged 2.94 with a range of 1.90 to 3.93. Students assigned to Group Two (n=12) averaged 3.05 with a range of 2.5 to 3.75. The remaining students who comprised Group Three (n=14) averaged 3.20 with a range of 1.77 to 3.92. As expected, no statistically significant group differences were evident, $F$ (2,17) = .60, $p > .05$, although Group Three had a slightly higher overall mean GPA than the other two groups.

PSI Performance

A number of measures were taken as a student attempted to pass each unit mastery test. First, the amount of time required to complete each test was recorded. Second, the percentage of correct responses by unit and by experimental group was recorded (mean mastery performance). Third, the number of students with errorless performance on any given unit mastery test (15 out of 15 correct) was recorded, and finally,

the percentage of students who passed each test on their first attempt (13.5 out of 15 correct). Each measure was examined within the three phases of material presentation and between experimental groups.

The mean amount of time needed to complete a unit mastery test was 50 min and ranged from 10 min to more than 2 hours. Each experimental group differed on mean completion times across the 15 unit tests. Group One averaged 46.7 min, Group Two averaged 49.2 min, and Group Three needed an average of 53.2 min per test. These group means differed significantly, $\underline{F}(2,36) = 8.54$, $\underline{p} < .002$.

The mean mastery performance by the class as a whole averaged slightly above the required mastery criterion of 90% correct. The class averaged 13.8 correctly answered questions per unit out of a possible 15 questions. This resulted in a mean mastery performance of 92%. Table 3 displays the mean number of items answered correctly across all 15 units of material. There were no significant differences in performance between experimental groups.

Table 3

Mean Number of Correct Items on PSI Tests

| Phase One | | Phase Two | | Phase Three | |
|---|---|---|---|---|---|
| Units | $\bar{X}$ | Units | $\bar{X}$ | Units | $\bar{X}$ |
| 1 | 13.2 | 6 | 14.4 | 11 | 13.2 |
| 2 | 13.7 | 7 | 14.0 | 12 | 13.6 |
| 3 | 14.1 | 8 | 13.9 | 13 | 14.1 |
| 4 | 13.4 | 9 | 14.1 | 14 | 14.0 |
| 5 | 13.9 | 10 | 13.7 | 15 | 14.1 |
| $\bar{X}$ Total | 13.7 | | 14.0 | | 13.8 |

Note. A score of 13.5 is equivalent to 90% correct. Group One mean correct = 13.8; Group Two mean correct = 13.9; Group Three mean correct = 13.8.

The number of students who performed errorlessly varied from unit to unit. An average of 9 students (23.6%), SD=4, had errorless performance on any given unit mastery test. The range was from 3 to 17 perfect scores per unit. Group differences were negligible except during the second phase of the course (units 6-10). During this phase, students in Group One totalled 22 perfect scores compared to 14 and 15 for Groups Two and Three (see Table 4).

Table 4

Errorless Performance on PSI Tests

| Unit | Group One | Group Two | Group Three | Total |
|---|---|---|---|---|
| 1 | 3 | 2 | 0 | 5 |
| 2 | 5 | 3 | 0 | 8 |
| 3 | 2 | 5 | 4 | 11 |
| 4 | 2 | 2 | 5 | 9 |
| 5 | 5 | 3 | 5 | 13 |
| Phase Total | 17 | 15 | 14 | 46 |
| 6 | 6 | 5 | 6 | 17 |
| 7 | 4 | 2 | 1 | 7 |
| 8 | 3 | 2 | 1 | 6 |
| 9 | 7 | 2 | 6 | 10 |
| 10 | 2 | 3 | 6 | 11 |
| Phase Total | 22 | 14 | 15 | 51 |
| 11 | 1 | 0 | 2 | 3 |
| 12 | 2 | 1 | 1 | 4 |
| 13 | 3 | 3 | 4 | 10 |
| 14 | 2 | 2 | 5 | 9 |
| 15 | 5 | 5 | 2 | 12 |
| Phase Total | 13 | 11 | 14 | 38 |
| Group Mean | 3.5 | 2.6 | 2.9 | 9.0 |

The percentage of students who passed a unit mastery test on the first attempt was 83.7%; 14.5% successfully passed on their second attempt. On 13 occasions a third attempt was needed (1.8% of the students) (see Figure 1, Appendix E for first-attempt performance).

Comprehensive Review Examinations

Table 5 displays group statistics on scores from the three comprehensive review examinations. Performance on the three examinations was similar across experimental groups. On the first exam, following Phase One, the class as a whole performed at a mastery level of 87% ($\bar{X}$=58 out of a possible 67 points). Forty-five percent of the students answered greater than 90% of the questions correctly. Performance on the three item levels (KC, AA, SE) did not differ significantly across experimental groups. Overall, students correctly answered an average of 88% of the knowledge-comprehensive items, 75% of the application-analysis items, and 83% of the synthesis-evaluation items.

Student performance on the second review examination following Phase Two was comparable to scores on the first exam with the overall scores slightly lower than the first review examination. The mean mastery for the class as a whole was 84%. On the other hand, performance on the knowledge-comprehension and application-analysis items was improved. Students correctly answered an

average of 90% of the knowledge-comprehension items, 82% of the application-analysis items, and 71% of the synthesis-evaluation items. The increase in performance on the application-analysis items may reflect the application objectives inherent in much of the Watson & Tharp textbook practice exercises.

Table 5

Performance on Comprehensive Review Examinations

| Group | n | Mean Score | Range | Median | Item Level KC Mean Scores | AA | SE |
|---|---|---|---|---|---|---|---|
| Review Exam Number One | | | | | | | |
| 1 | 12 | 57 | 47-66 | 58.5 | 48 | 6 | 3 |
| 2 | 12 | 61 | 57-65 | 60.5 | 51 | 6 | 4 |
| 3 | 14 | 56 | 45-63 | 58.3 | 47 | 6 | 3 |
| Review Exam Number Two | | | | | | | |
| 1 | 12 | 58 | 52-63 | 58.0 | 32 | 16 | 10 |
| 2 | 12 | 58 | 47-65 | 58.6 | 31 | 17 | 10 |
| 3 | 14 | 60 | 53-65 | 59.0 | 32 | 16 | 12 |
| Review Exam Number Three | | | | | | | |
| 1 | 12 | 51 | 37-67 | 50.5 | 33 | 14 | 4 |
| 2 | 12 | 55 | 32-65 | 57.5 | 35 | 15 | 5 |
| 3 | 14 | 53 | 33-64 | 56.5 | 33 | 15 | 5 |

Note. Maximum score on Review Exam Number One = 67. Maximum scores on exams Two and Three = 70. Possible scores; Review One: KC=55, AA=8, SE=4. Review Two: KC=35, AA=20, SE=15. Review Three: KC=40, AA=20, SE=10. Review Exam One contained more KC items since two groups were receiving a majority of these types of items.

Student performance on the third review examination was somewhat lower than the previous two exams. This may reflect the fact that both the course materials as well as the mastery test requirements became more difficult. Mean mastery performance on this examination was 76%. Scores on the aggregated item types (KC, AA, & SE) also showed a decrease from the previous two examinations. On the third review examination, students correctly answered an average of 84% of the knowledge-comprehension items, 73% of the application-amalysis items, and 46% of the synthesis-evaluation items.

Table 1 in Appendix D summarizes three one-way (Score x Group) ANOVAs on the midterm comprehensive review examination data. No significant differences were found between the three experimental groups. The group differences in performance at each of the level-of-questions were also examined by a series of ANOVAs. These ANOVAs (one for each exam) were computed on the scores for each type of item (level of complexity). The results of these ANOVAs are summarized in Table 2 of Appendix D. Only one ANOVA yielded a significant ($p < .03$) effect, for the

knowledge-comprehension items on the first examination. Because the pattern of mean differences did not make sense in the present experimental design, post-hoc analyses were not calculated.

Final Examination

The comprehensive final examination was taken by 38 students. A total of 140 points were possible on this examination. The mean score achieved for the class as a whole was 110. This score represents a class mastery level of 79%. Table 6 displays group differences in performance on the final examination. Of the three experimental groups, Group Three had slightly higher scores on the synthesis-evaluation type items. A summary of a series of one-way analyses of variance (ANOVA) on group differences is displayed in Table 3, Appendix D. There were no significant group differences for each level of question complexity in the final examination scores.

Table 6

Group Score Comparisons on the Final Examination

| Group | n | $\bar{X}$ Score | Range | Median | KC | AA | SE |
|---|---|---|---|---|---|---|---|
| Total | 38 | 110 | 84.5 - 124 | 111.75 | 73 | 20 | 16 |
| 1(KC) | 12 | 108 | 84.5 - 124 | 110.00 | 72 | 19 | 16 |
| 2(AA) | 12 | 110 | 88.5 - 122 | 112.50 | 74 | 21 | 15 |
| 3(SE) | 14 | 112 | 93.5 - 123 | 112.75 | 74 | 20 | 18 |

Follow-up Examination

Results of the follow-up examination were compared to student scores on the final examination. Data from the 20 students included in the follow-up sample showed a slight decrease in performance following a 2 month interval between examinations. The mean score on each examination was 110 and 106.5 points, respectively. These group score comparisons are summarized in Table 7. The analyses of variance on the level-of-questions effect for this examination are summarized in Appendix D, Table 4. Performance on the higher-level items was maintained at approximately the same level by all three experimental groups. The ANOVA on group differences in performance on the level-of-questions was not found to be significant.

Table 7

Group Score Comparisons on the Follow-up Examination

| Group | n | $\bar{X}$ Score | Range | Median | KC | AA | SE |
|---|---|---|---|---|---|---|---|
| Total | 20 | 106.5 | 85 - 128 | 104 | 67 | 23 | 17 |
| 1(KC) | 7 | 108.2 | 95 - 127 | 105 | 66 | 23 | 19 |
| 2(AA) | 6 | 104.5 | 91 - 128 | 99 | 67 | 23 | 14 |
| 3(SE) | 7 | 106.9 | 85 - 123 | 103 | 66 | 23 | 18 |

A substantial decrease in the amount of time needed to complete the follow-up examination, from a mean of 107 min to 76 min was observed. Students in Group Three displayed the greatest decrease in amount of time needed, from 124 min down to 75 min.

The follow-up examination included approximately equal numbers of students from each of the three experimental groups. These students were also compared in terms of grade point averages and no signficant differences were evident. Students in Group One had an average GPA of 3.11 on a 4.0 scale. Students in Group Two averaged 3.18, and students in Group Three averaged 3.45. An analysis of variance on these group differences, $F$ (2,17) = 0.60, $p > .05$ was not significant.

## Discussion

Performance on the unit mastery tests and on the major review and final examinations was studied as a function of the level of questions incorporated in the study guide materials presented to the students. On the average, students performed better than expected on the unit mastery tests with approximately 92% of the questions answered correctly on any given unit mastery test. No significant differences between experimental groups were evident. Approximately 84% of the students passed each unit mastery test on their first attempt.

Student performance on the three comprehensive review examinations was also similar across experimental groups. Following Phase One, the class as a whole performed at a mastery level of 87% on the first comprehensive review examination. Performance on the three item level types did not differ signficantly across experimental groups even though exposure to the higher-level questions was limited for Groups One and Two. On the second and third review examinations, class performance was slightly lower than expected with mastery levels of 84% and 76%,

respectively. The decrease in performance can be partially attributed to a higher ratio of the more complex synthesis-evaluation type items and to the introduction of a new, more theoretically oriented textbook (Chance, 1979) during the last phase of the course.

Student performance on the comprehensive final examination was somewhat lower than expected. The class as a whole achieved a mastery level of approximately 80% correct. Experimental Group Three performed slightly better on the more complex synthesis-evaluation type items, but the difference between groups was not significant.

Results of the follow-up examination were compared to scores obtained on the final examination since both tests were similar in composition and level-of-questions presented. Of the students (n=20) who returned following a 2 month interval to take the follow-up examination, overall performance was similar to scores obtained on the final examination. The mean score on the final examination was 110 points and on the follow-up 106.5 points. Performance on the higher-level items was maintained at approximately the

same level, however, on those items requiring more rote memorization, scores decreased slightly.

In general, expected differences relating to the three sequence requirements were not obtained. The order in which a student was required to study for and answer more complex questions did not make a difference in his/her ability to answer such questions. It was hypothesized that the student who experienced more complex questions earlier in the course would outperform those students having only attended to simpler questions. The data failed to support this hypothesis. In fact, no differences between student groups were obtained in comparing the unit mastery test, comprehensive review examination scores and the final examination scores. There were no significant group differences in performance on any of the item-type questions. The students in Group Three did not show any consistent superiority on the more complex synthesis-evaluation type items even though they received these items from the initial study guide materials. These results might suggest that within the structured PSI system, specific hierarchical sequencing of higher-level questions is

not essential for sufficient learning to take place. That is, a student need not master basic knowledge-level questions before experiencing and correctly answering higher-level questions.

Watts and Anderson (1971) suggested that answering higher-level questions would facilitate later performance on these types of items by encouraging students to process the content of the information more thoroughly. According to this hypothesis, behavior subsequent to the receipt of study guide questions is modified and forces the student to adjust his study activities in preparation for these more complex items. Results of the present study relating specifically to the initial focus on the effects of study question levels on subsequent review examination performance did not support this hypothesis. Students in the present study who received study guide questions emphasizing only knowledge and comprehension skills were able to perform adequately on higher-level items without prior exposure to those more complex items types.

Student performance on the follow-up examination was better than expected. On the higher-level

questions, particularly the application-analysis items, students performed better than they did on the final examination. Students tended to correctly answer the basic knowledge-comprehension type items less consistently than the higher-level items. The mean difference in performance (total scores) on the follow-up examination reflected only a slight decrease in retention following a 2 month interval. On the average, students also used approximatley 30% less time to complete the follow-up over the final examination.

Several factors may have contributed to these last results. First, a self-selection factor may have been operating with the more motivated, over-achieving students completing the follow-up examination. A comparison of grade point averages of students who returned for the follow-up exam was made and showed that each experimental group had similar GPA's. However, this analysis tells us little about the differences between the follow-up students and those who failed to return for the examination. Another factor which may have contributed to higher scores and a decrease in the amount of time needed may have been

the students' familiarity with these higher-level items. Approximately 80% of the items were retained from the final examination.

A few suggestions for future courses may be made. First, a means to control for practice effects on the higher-level questions is essential to ensure proper measurement of actual achievement on these types of items. In the present study, some students answered higher-level questions even though they were not being graded on them at that particular point in time. It was possible that this exposure made these types of items easier to answer later in the course when they were being graded. Originally, these extra items were to be used in assessing a students' performance on higher-level items before the study guide materials were received. This analysis proved too laborious and futile given that some of these items included "unanswerable" or "irrelevant" items used as fillers. Second, a method needs to be devised to write higher-level questions in a multiple-choice or fill-in answer format. This type of format would reduce the amount of time needed to answer and score these items. One suggestion is that mastery test items be written

in multiple-choice format with two correct choices. One choice being a basic knowledge-comprehension answer, and another at a higher-level. With a possible added explanation, this would enable the experimenter to determine at what level the student is actually processing the information given that both answers are correct. Third, although students in the present study achieved mastery of the PSI materials, several changes in the actual level of course materials would be recommended. The course was highly content defined as an application-analysis course. That is, the self-control content in Watson & Tharp (1977) included mostly application-analysis objectives. Extrapolating higher-level items often reflected many hours of staff time before suitable items were written. This difficulty in creating suitable items increased the work load on both the course instructor and item-raters and also increased test completion time for the students.

In conclusion, it was hypothesized that students who initially received study guide materials emphasizing higher-level questioning would outperform other students on these types of items on the

comprehensive review examinations and final examinations. Also, it was hypothesized that the order in which a student was asked to study for and answer more complex questions would make a difference in his/her ability to answer such questions. The data failed to support both hypotheses. There was no consistent superiority of those students who received the more complex items earlier in the course over those who received primarily knowledge-comprehension type items. The order in which a student was required to study for and answer more complex questions made little difference in his/her ability to answer such questions.

Finally, it is worth noting that higher-level performance clearly resulted from the presentation of material and study guides in this course. Testing was not limited to direct recall of factual information and students demonstrated mastery level performance on items which included application-analysis as well as synthesis-evaluation processes. Future research in this area must demonstrate greater control of the assessment procedures which take into account the nature of the course materials and their possible

limitations for developing and testing performance on higher-level questioning specifically within the Personalized System of Instruction format.

# References

Anderson, R. C. (1970). Control of student mediating processes during verbal learning and instruction. Review of Educational Research, 40, 349-369.

Andre, T. (1979). Does answering higher-level questions while reading facilitate productive learning? Review of Educational Research, 49, 280-318.

Austin, S. M. & Gilbert, K. E. (1974). Student performance in a Keller course in introductory electricity and magnetism. In J. G. Sherman (Ed.), Personalized system of instruction: 41 germinal papers. Menlo Park, CA: W. A. Benjamin.

Bloom, B. S. (Ed.) (1956). Taxonomy of educational objectives. New York: David McKay.

Born, D. G., Gledhill, S. N., & Davis, M. L. (1972). Examination performance in lecture-discussion and personalized instruction courses. Journal of Applied Behavior Analysis, 5, 33-43.

Chance, P. (1979). Learning and behavior. Belmont, CA: Wadsworth.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11, 671-684.

Duell, O. K. (1974). Effect of type of objective, level of test question, and the judged importance of tested materials upon posttest performance. Journal of Educational Psychology, 66, 225-232.

Frase, L. T. (1968). Some unpredicted effects of different questions upon learning from connected discourse. Journal of Educational Psychology, 59, 197-200.

Hunkins, F. P. (1969). Effects of analysis and evaluation questions on various levels of achievement. Journal of Experimental Education, 38, 45-58.

Keller, F. S. (1968). "Good-bye teacher..." Journal of Applied Behavior Analysis, 1, 79-89.

Keller, F. S., & Sherman, J. G. (1974). PSI: The Keller plan handbook. Menlo Park, CA: W. A. Benjamin.

Kirk, R. E. (1968). Experimental design: Procedures for the social sciences. CA: Brook-Cole.

Kulik, C. C., Kulik, J. A., & Cohen, P. A. (1980). Instructional technology and college teaching. Teaching of Psychology, 7, 199-205.

McMichael, J. S., & Corey, J. R. (1969). Contingency management in an introductory course produces better learning. Journal of Applied Behavior Analysis, 2, 79-84.

Richards, J. P., & DiVesta, F. J. (1974). Type and frequency of questions in processing textual material. Journal of Educational Psychology, 66, 354-362.

Semb, G., Conyers, D., Spencer, R., & Sanchez-Sosa, J. J. (1975). An experimental comparison of four pacing contingencies. In J. M. Johnston (Ed.), Behavior Research and Technology in Higher Education. Springfield, IL: Charles C. Thomas.

Sheppard, W. C., & MacDermot, H. G. (1970). Design and evaluation of a programmed course in introductory psychology. Journal of Applied Behavior Analysis, 3, 5-11.

Skinner, B. F. (1954). The technology of learning and the art of teaching. Harvard Educational Review, 24, 86-97.

Watson, D. L., & Tharp, R. G. (1977). Self-directed behavior: Self-modification for personal adjustment, (2nd ed.). Belmont, CA: Brooks-Cole.

Watts, G. H. (1973). Effects of prequestions on control of attention in written instruction. The Australian Journal of Education, 18, No. 1.

Watts, G. H. Anderson, R. C. (1971). Effects of three types of inserted questions on learning from prose. Journal of Educational Psychology, 62, 387-394.

# Appendix A

Item Rater Instructions: Using the following criteria, indicate which category (if any) each mastery test item is most appropriate. Check the blank which corresponds for the selection of each item. If you feel that an item may fit into more than one category then check more than one blank. If you feel any question is ambiguous or diffiuclt to understand, please circle the question or note on which page it appeared. Thank you for your assistance.

Knowledge-Comprehension: Involves the recall of specifics and universals, the recall of methods and processes, or the recall of a pattern, structure, or setting. The recall situation involves little more than bringing to mind the appropriate material. This includes knowledge of specific terminology, specific facts, trends and sequences, and knowledge of criteria, methodology, and knowledge of theories and structures. Comprehension involves understanding such that the individual knows what is being communicated and can make use of the material or

idea being communicated without necessarily relating it to other material.

Application-Analysis: Is the use of abstractions in particular and concrete situations. The abstractions may be in the form of general ideas, rules of procedures, or generalized mehtods. The abstractions may also be technical principles, ideas, and theories which must be remembered and applied. Analysis is the breakdown of a communication into its elements or parts such that the relative hierarchy of ideas is made clear and/or the relations between the ideas expressed are made explicit. This includes the ability to recognize unstated assumptions, the connections and interactions between elements and parts of a communication and the ability to recognize the general techniques used in the communication.

Synthesis-Evaluation: Synthesis is the putting together of elements and parts so as to form a whole. This involves the process of working with pieces, parts, elements and arranging them in such a way as to constitute a pattern or structure not clearly there before. This includes the development of a communication in which the writer or speaker attempts

to convey ideas, feelings, and/or experiences to others. Evaluation is the ability to make judgments about the value of material and methods for given purposes. Quantitative and qualitative judgments about the extent to which material and methods satisfy criteria. It also includes the ability to indicate logical fallicies in arguments and evaluation of material with reference to selected or remembered criteria.

Key Words for Taxonomy

| | Infinitives | Examples of Direct Objects |
|---|---|---|
| K-C: | to define, recall, recognize, identify, determine, extend, distinguish, fill in | criteria, basics, uses, methods, elements, procedures, meanings, definitions, factors |
| A-A: | to apply, relate, choose, use, employ, classify, identify, deduce, analyze, compare, contrast, detect | principles, laws, ideas, conclusions, effects, fact, intent, biases, generalizations, cause-effect, point of view |
| S-E: | to write, tell, relate, modify, propose, plan, design, derive, develop, combine, judge, argue, consider, compare, contrast, appraise | structure, patterns, design, efforts, plans, solutions, concepts, theories, hypotheses, accuracy, flaws, consistent, utility, errors |

Note: To simplify this procedure, most of the K-C items are straight recall items from sentences in each unit. The A-A items often relate to

everyday examples of the material in each unit and may be short-answer types. The S-E items are more complex asking for student evaluation and communication of each important concept in the unit. These items may involve putting together many concepts and making a final decision in a written format.

Definitions of Action Words

a) Identifying: The student selects the correct object of a class name. This class of performances also includes identifying object properties (rough, smooth) and, in addition, kinds of changes such as an increase or decrease in size.

b) Distinguishing: Identifying objects or events which are potentially confusible (e.g., square or rectangle), or when two contrasting identifications (such as right or left) are involved.

c) Constructing: Generating a construction or drawing which identifies a designated object or set of conditions. For example: Beginning with a line segment, the request is made, "complete this figure so that it represents a triangle."

d) Naming: Supplying the correct name for a class of objects or events.

e) Ordering: Arranging two or more objects or events in proper order in accordance with a stated category.

f) Describing: generating and naming all of the necessary categories of objects, object properties, or event properties that are relevant to the description of a designated situation.

g) Stating a Rule: Makes a verbal or written statement which conveys a rule or principle including the names of the proper classes of objects in their correct order.

h) Applying a Rule: Using a learned principle or rule to derive an answer to a question. The answer may be a correct identification, the supplying of a name, or some other similar kind of response.

i) Demonstrating: Peforming the operatins necessary to the application of a rule or principle.

j) Interpreting: Ability to identify objects or events in terms of their consequences.

---

Source: Partial reprint from Bloom, B. S. (Ed.) (1956) Taxonomy of educational objectives. New York: David McKay.

While all results of this study cannot be anticipated as of now, all participants will have the opportunity to hear or read a summary description of the study and its major results during the Winter or Spring, 1980 terms.

I understand that records of my test-taking performance in this course will be used by a graduate student in psychology for research purposes and my signature authorizes athe use of my test data.

_____ _____

Signed Date

Course Instructor: Kenneth Beauchamp, Ph.D.
Graduate Student in Psychology: Robert Kutner

Appendix C

Student GPA and Follow-up Request Consent Form

Self-Control
Fall 1980

In order that we may be able to reach you to provide further information about the results of the experiment which involved the Self-Control class and to notify you of the opportunity to earn money by taking another final exam (follow-up) in February, please indicate your campus address and phone number during February 1980:

Address: Phone Number:

In order to assess the equivalence in ability of the randomly assigned groups 1, 2, and 3, we are requesting permission for Robert Kutner to see your cumulative collegiate GPA (not your grades in any class, just your overall GPA at the end of the Fall semester). If you grant permission, we will ask a clerk in the Registrars office to provide, with your permission, your cumulative GPA from high school as calculated by the Admissions office. These figures

will never be published in any way that an individual student's name. Again, our only purpose is to examine group similarities and differences as reflected in achieved grades.

If you do grant permission to Robert Kutner to secure your two GPA figures for the above reason with the assurance that the information will remain anonymous, please sign below:

Name: _____

# Appendix D

## Table 1

### One-way Analyses of Variance:

### Group Scores on Comprehensive Exams

| Effect | Source | df | MS | F |
|---|---|---|---|---|
| Review Examination Number One | Group | 2 | 40.2 | 1.85 |
| | Error | 35 | 21.6 | |
| | Total | 37 | 22.6 | |
| Review Examination Number Two | Group | 2 | 4.5 | 2.91 |
| | Error | 35 | 15.4 | |
| | Total | 37 | 14.9 | |
| Review Examination Number Three | Group | 2 | 106.0 | 1.56 |
| | Error | 35 | 67.8 | |
| | Total | 37 | 69.8 | |

# Appendix D

## Table 2

One-way Analyses of Variance:
Level-of-Questions Effect between Experimental Groups

| Effect | Source | df | MS | F |
|---|---|---|---|---|
| Review Examination One | | | | |
| Item Level | | | | |
| KC | Group | 2 | 149.9 | 3.82* |
| | Error | 32 | 39.3 | |
| | Total | 34 | 45.8 | |
| AA | Group | 2 | 24.7 | 0.12 |
| | Error | 32 | 199.9 | |
| | Total | 34 | 189.6 | |
| SE | Group | 2 | 419.0 | 1.31 |
| | Error | 32 | 318.0 | |
| | Total | 34 | 323.9 | |
| Review Examination Two | | | | |
| Item Level | | | | |
| KC | Group | 2 | 3.9 | 0.14 |
| | Error | 35 | 28.1 | |
| | Total | 37 | 26.8 | |
| AA | Group | 2 | 4.2 | 0.06 |
| | Error | 35 | 64.6 | |
| | Total | 37 | 61.3 | |
| SE | Group | 2 | 366.3 | 2.60 |
| | Error | 35 | 140.7 | |
| | Total | 37 | 152.9 | |

Appendix D

Table 2
(continued)

| Effect | Source | df | MS | F |
|---|---|---|---|---|
| Review Examination Three Item Level | | | | |
| KC | Group | 2 | 235.7 | 1.77 |
| | Error | 35 | 133.1 | |
| | Total | 37 | 138.7 | |
| AA | Group | 2 | 227.1 | 1.21 |
| | Error | 35 | 186.9 | |
| | Total | 37 | 189.0 | |
| SE | Group | 2 | 549.7 | 1.22 |
| | Error | 35 | 449.7 | |
| | Total | 37 | 455.0 | |

* $p < .03$. KC=knowledge-comprehension, AA=application-analysis, SE=synthesis-evaluation.

# Appendix D

## Table 3

### One-way Analyses of Variance:

### Level-of-Questions Effect on the Final Examination

| Effect: Item Level | Source | df | MS | F |
|---|---|---|---|---|
| KC | Group | 2 | 23.87 | 0.63 |
| | Error | 17 | 37.53 | |
| | Total | 19 | 36.09 | |
| AA | Group | 2 | 0.18 | 0.02 |
| | Error | 17 | 9.14 | |
| | Total | 19 | 8.20 | |
| SE | Group | 2 | 33.11 | 3.09 |
| | Error | 17 | 10.70 | |
| | Total | 19 | 13.06 | |

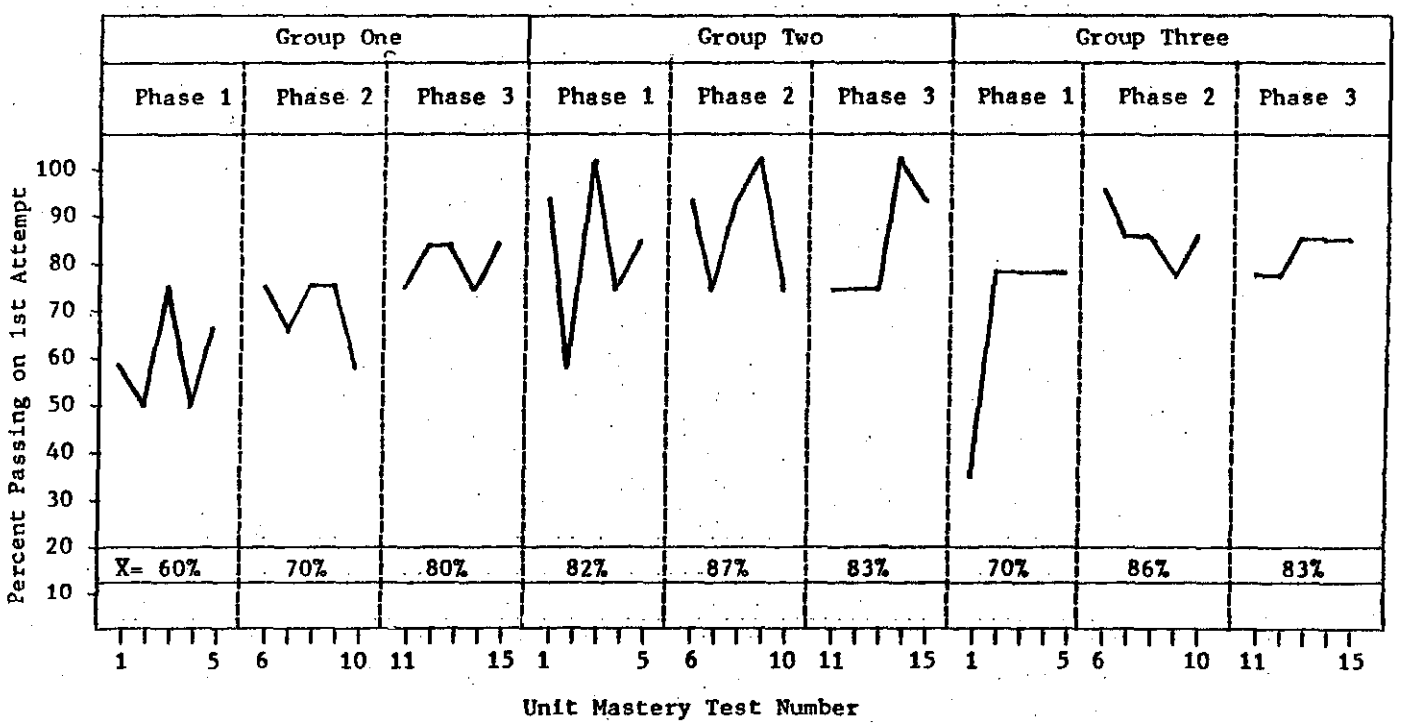
# Appendix D

## Table 4

One-way Analyses of Variance:

Level-of-Questions Effect on the Follow-up Exam

| Effect: Item Level | Source | df | MS | F |
|---|---|---|---|---|
| KC | Group | 2 | 1.77 | 0.04 |
| | Error | 17 | 38.20 | |
| | Total | 19 | 34.36 | |
| AA | Group | 2 | 0.66 | 0.02 |
| | Error | 17 | 25.39 | |
| | Total | 19 | 22.79 | |
| SE | Group | 2 | 38.93 | 1.15 |
| | Error | 17 | 33.75 | |
| | Total | 19 | 34.30 | |




Appendix E

Figure 1. First-attempt performance on all PSI unit tests for all three experimental groups.