RICE UNIVERSITY

A COMPARISON OF COMPUTER-ASSISTED DRILL & PRACTICE STRATEGIES IN A HIERARCHICAL TASK

by

Marc B. Barnett

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

Thesis Director's Signature:

[Signature]

Houston, Texas

August, 1975
ABSTRACT

A COMPARISON OF COMPUTER-ASSISTED DRILL & PRACTICE
STRATEGIES IN A HIERARCHICAL TASK

by

Marc B. Barnett

Computer-assisted drill & practice is investigated in the context of a short-term, novel mathematical task. A between groups design is used to statistically compare the efficiency of four methods (three response-sensitive and one response-insensitive) for structuring drill & practice in base eight arithmetic. In addition, a descriptive analysis of both task structure and suitability of the task for individualization is performed. While no statistical differences among the treatment groups were observed, substantial information concerning the properties of the task was obtained.
Acknowledgements

The author wishes to express appreciation to Drs. John Brelsford, David Schum, and Sarah Burnett who served patiently as members of his committee. I especially thank Dr. Brelsford for his thoughtful guidance and persistent encouragement.

The author would also like to thank John Crockett for the months of diligent computer programming.

Finally, I wish to express sincere gratitude to Anne Hoffacker whose support made the final preparation of this paper tolerable.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>DESCRIPTION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>1</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>ii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>METHOD</td>
<td>20</td>
</tr>
<tr>
<td>RESULTS AND DISCUSSION</td>
<td>25</td>
</tr>
<tr>
<td>APPENDIX 1</td>
<td>40</td>
</tr>
<tr>
<td>APPENDIX 2</td>
<td>46</td>
</tr>
<tr>
<td>APPENDIX 3</td>
<td>47</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>51</td>
</tr>
</tbody>
</table>
INTRODUCTION

Computer-assisted instruction (CAI), hailed initially like so many other novel instructional techniques as the "wave of the future for education," has generated in inconsistent record of successes in its applications (McKeachie, 1974). Understandable, the lack of strong empirical support for the hypothesis that CAI, conceived according to the principles of individualized instruction, should be superior to the conceptually less sophisticated textbook-lecture tradition is a matter of concern for supporters of applied learning theory. This concern, however, may be premature in the light of the atheoretical research orientation typically encountered in studies of educational techniques (Groen & Atkinson, 1966)- CAI not withstanding. Frequently it has consisted of large-scale, non-rigorous summative evaluations comparing groups receiving commercially available CAI to control groups not receiving CAI (Davis, 1972; Fletcher & Atkinson, 1972; Vinsonhaler & Bass, 1971). These studies, while providing useful information about the relative desirability of commercially available CAI programs, do not provide a sensitive assessment of many of the variables critical to the efficacy of CAI. A more reasonable approach, given the present state of the art, is to perform a series of theory-generated experimental manipulations in strictly controlled
environments for the purpose of observing "within-method" effects. With this notion in mind, the present study was undertaken as a preliminary investigation of some basic variables within CAI.

Although CAI includes variations that range from administrative recordkeeping to allowing students to manipulate large, complex data bases (cf. Atkinson & Wilson, 1969), perhaps its most important quality stems from its cybernetic potential (Stolurow, 1969) as an "adaptive teaching system" (Atkinson, 1972b). This is possible because CAI can achieve a unique, dynamic interaction with an individual student. Ideal dynamic interaction results in a flow of instruction tailored to the student because it has been derived from the combination of a tractable model of the learning process with an assessment of the student's state of learning or understanding of the material at any given point in time during the instructional session. The degree to which the interaction between student and computer is maximally efficient or "optimal" is a function of the adequacy of its components: 1) a prescriptive model or theory of the learning processes involved, 2) summarization of the student's observed learning progress as input for the model, 3) presentation of the actual instructional elements according to the most efficient algorithm implied by the interaction of the first two components (Atkinson, 1972b; Stolurow, 1969). This study is directed primarily
toward investigation of the interaction between the first and second components, i.e., the nature of the individualization used within a particular type of CAI format.

In developing a CAI program one of the first considerations is the type of instructional format to be used. The three generally acknowledged formats, dialogue, tutorial, and drill & practice represent a wide range of theoretical and operational complexity (Suppes, 1969). Dialogue, patterned after the one-to-one interaction between teacher and student, is the most elaborate of the three. Its ideal form consists of a non-structured conversation between the student and a computer programmed to perform all the functions of a human teacher. Presently this level of interaction is beyond our reach for two reasons. First, a "theory of instruction" (which prescribes the most desirable instructional policy given a set of contextual requirements) (Atkinson, 1972a) is not currently available for even the most simple instructional situations. And, second, voice-dependent or similar freely-structured man-computer interfaces have not yet been perfected.

Tutorial and drill & practice, the most widely used types of CAI (Atkinson & Wilson, 1969, p. 8), represent more reasonable levels of complexity. In tutorial programs the computer "take(s) over the main responsibility both for presenting a concept and developing skill in its use" (Suppes, 1969). While conceptually similar to the dialogue
format, tutorial is far less flexible in that it operates according to a relatively simple model or theory of the learning process involved and is restricted to a convenient student-computer interface, eg., a teletype. Drill & practice programs, composing the least complex level of CAI, are used primarily to give students an opportunity to practice material that has already been presented in a traditional classroom setting. Like the tutorial format, it is individualized based on a simple model or theory of the learning process and is accomplished through a similar man-computer interface. In situations where the teacher takes less responsibility for introducing new material, the distinction between tutorial and drill & practice may become academic.

For any of the formats just discussed the critical issue is not why, but how to achieve individualization of instruction (Atkinson, 1972a). Essentially, individualization is accomplished with a plan that prescribes what instructional elements should next be presented, based on some measure, or measures, of the student's past and present levels of learning. The most simple form of individualization, intrinsic programming (Crowder, 1962), is commonplace in programmed instruction. Typically this involves presenting a "frame" of material to the student, asking a question about the material, and then, depending only on that response, presenting more material using an
"if-then" decision rule. The theory that generates the decision rule is more often than not only a simple, intuitive statement regarding what frame to present in the event of a particular answer. Such a prescription implies that two students with dissimilar sets of responses to the first n-1 questions would receive the same material following question n if their responses to that question were identical. When, as is frequently the case, the only possible responses are coded as correct, incorrect, or A, B, C, or D (as in multiple choice), the probability of any two students giving the same response is quite high. Consequently, the probability of getting the same instruction as a result is quite high even though their relative "states of learning" (i.e., their present levels of learning taking into account their cumulative records of previous levels of learning) may be quite different. Therefore, the mildly individualized instruction that relies on intrinsic programming may yield only modest gains (cf. Briggs & Angell, 1964) because the theoretical base is too insensitive to the student's moment-to-moment states of learning.

More completely individualized instruction has resulted from elaboration on the concept of extrinsic programming introduced by Crowder (1962). As originally formulated, extrinsic programming utilizes "other data"
and "further computation to select the next material the student should see" (Crowder, 1962). The operational extension to extrinsic programming can be derived from a tractable model or theory of the learning process combined with some subset of the student's response history - the complete chronological record of his responses and their attributes. CAI individualized in this way can be characterized as being more or less "mathematically optimal" or subjectively optimal" contingent upon the model or theory from which it was generated.

Presently, the most sophisticated mathematically optimal CAI uses stochastic, N-stage (or approximately N-stage) mathematical models that consist of:

A) a set of learning states; B) a usually nondeterministic response rule which gives (for each state of learning) the probability of a correct response to a given stimulus; and C) an updating rule which provides a means of determining the new learning state (or distribution of states) that results from the presentation of a stimulus, the response the student makes, and the reinforcement he receives (Groen & Atkinson, 1966).

A model with these characteristics can be combined with the student's response history to yield a number of possible instructional strategies, or algorithms, for presenting the instructional elements. Mathematically, the best, i.e., "optimal" strategy can then be formally derived through the application of control theory to the instructional context (For a formal discussion of control theory
in CAI see Atkinson and Paulson, 1972; Chant & Atkinson, 1973; Karush & Dear, 1966.). Control theory uses the prediction of the student's state of learning at any given moment (made by the model of learning) to present the instructional element that maximizes his future state of learning with respect to some previously established measurable criteria (Atkinson, 1972b; Groen & Atkinson, 1966). Although, CAI generated in this way is mathematically optimal, it does not necessarily conform to what would intuitively be considered the best instruction. This can occur only if the model is both a good and an accurate representation of the learning process involved in the task. Since even the most sophisticated models used in CAI fail to take into account some potentially important variables (eg. motivation), CAI based on those models must necessarily fall short of ideal instruction. Mathematically optimal CAI is, however, much more in line with what intuition would suggest about the complexity of the instructional process than CAI based on intrinsic programming.

Unfortunately, availability of even partially adequate models of the type just described is generally restricted to models developed in the field of verbal learning to accommodate very simple tasks. For this reason, subjectively optimal CAI is used to generate instruction of more complex tasks for which there are no appropriate mathematical models. In lieu of a mathematical model,
a prescriptive theory of the learning process hypothesized is used to specify a set of rational statements from which an instructional algorithm can be inferred. The algorithm and response history are then combined in such a way that the presentation of instructional elements results in "subjectively optimized," or maximized, learning with respect to some criteria. Just as with mathematically optimal CAI, subjectively optimal CAI is only as adequate as its underlying theory. The following pair of studies was chosen to illustrate subjectively optimal drill & practice CAI and mathematically optimal tutorial CAI. Additionally, the first study provides an example of the type of task and context for which CAI is becoming popular, while the second introduces the experimental paradigm used in the present study.

The first study, conducted by Suppes & Morningstar (1972), took place in an applied, elementary school setting. During the 1966-67 and 1967-68 school years, data were recorded for nearly 2500 students in grades one through six in two states, California and Mississippi. The application involved providing the students with five to fifteen minutes per day of drill & practice in elementary mathematics. Over an estimated average of 150 days of a student's participation during a school year, this accounted for a substantial amount of practice in the mathematics curriculum. The purpose of the research was not to
test any specific experimental hypotheses (disregarding
the comparison against a non-CAI control group); but
rather, through a rigorous descriptive analysis of the
data, to develop formal learning models for mathematical
tasks.

The general organization of the drill & practice pro-
gram, designed as a long term aid to classroom instruc-
tion, had a hierarchical structure consisting of three
levels: 1) grade level, 2) concept blocks within the grade
level, and 3) material within the concept blocks.

The curriculum material for each of Grades 1 through
6 was arranged sequentially in blocks to coincide
approximately with the development of the mathemati-
cal concepts in several text series. There were 24
concept blocks in Grade 1 and 23 concept blocks in
each of Grades 2 through 6. Each concept block in-
cluded a pretest (on Day 1), five days of drill (on
Days 2-6), a posttest (on Day 7), and sets of review
drills (also on Days 2-5), and review posttests (one
on Day 6). (Suppes & Morningstar, 1972)

Since the sequence and content of concept blocks was fixed
for all students within a grade level, presentation of
material within concept blocks was optimized (subjective-
ly). This was done in two stages. For one stage it
required presenting a set of drill exercises at an optimal
level of difficulty. For the other it required using an
optimal presentation format for drill items within that
set.

The choice of an optimal level of difficulty for a
given day's drill was predicated on the assumption that
the level of difficulty assigned should be directly proportional to the student's comprehension of the material as indicated by his performance record. Similarly, review drill was assigned from a previously completed concept block on which the student performed most poorly. The resulting algorithm and use of response history can be illustrated by a brief narration of the seven-day schedule for a concept block. Depending on his performance on the pretest given on the first day in a new block, the student was presented with drill material from one of five possible levels of difficulty on the second day according to the branching criteria in Table 1. In addition to the drill material he also received review drill material from a previously encountered concept block with the lowest posttest score. The level of difficulty for the review drill material remained the same for the second through the fifth day and was also determined by the branching criteria in Table 1. The levels of difficulty for the drills given on days three through six were determined by percentage correct on the previous day's drill as per Table 1. On the sixth day a review posttest corresponding to the review drill material on the four preceding days was administered and that score replaced the original posttest score for the block from which the review drill
### TABLE 1

**BRANCHING CRITERIA**

<table>
<thead>
<tr>
<th>From Pretest to Drill and From Posttest to Review</th>
<th>From Drill to Drill</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage Correct</td>
<td>Level Assigned for Drill</td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>0 - 19</td>
<td>1</td>
</tr>
<tr>
<td>20 - 39</td>
<td>2</td>
</tr>
<tr>
<td>40 - 59</td>
<td>3</td>
</tr>
<tr>
<td>60 - 79</td>
<td>4</td>
</tr>
<tr>
<td>80 - 100</td>
<td>5</td>
</tr>
</tbody>
</table>

was taken. On the seventh day the student was required to take a posttest for the concept block.

Optimization of the item presentation format focused on the notion that a persistent error on an item could be eliminated using a feedback loop to provide successively stronger corrective cues to the student. The optimal algorithm specified that a correct response to an item would result in presentation of a new item while an incorrect response (either a wrong answer or time out) would result in feedback and another opportunity to solve the problem. After the first error the student received the message, "NO, TRY AGAIN;" after the second he received, "NO, THE ANSWER IS ______. TRY AGAIN;" and after the third he received, "NO, THE ANSWER IS ______" before being presented with a new item. This decision rule is illustrated in Figure 1.

A particularly good example of successful, well designed research on tutorial CAI is a study done by Atkinson (1972b) to assess the relative optimality of four instructional strategies (each using a different level of individualization) for teaching a second language vocabulary. The following excerpt provides a brief description of the experimental context.

The constraints imposed on the experimental task are those that typically apply to vocabulary learning in
Fig. 1. Flow chart of program logic for teletype drill program. From Computer-assisted Instruction at Stanford, 1966-68: Data Models, and Evaluation of the Arithmetic Programs, by P. Suppes and M. Morningstar, 1972. Copyright 1972 by Stanford University.
an instructional laboratory. A large set of German-English items are to be learned during an instructional session which involves a series of discrete trials. On each trial, one of the German words is presented and S attempts to give the English translation; the correct translation is then presented for a study period. A predetermined number of trials is allocated for the instructional session, and after some intervening period of time a test is administered over the entire vocabulary set. The problem is to specify a strategy for presenting items during the instructional session so that performance on the delayed test will be maximized. (Atkinson, 1972b)

Since learning German-English items is essentially a paired-associate task, Atkinson used different versions of a three state Markov model for all-or-none paired-associate learning as a basis for two of the instructional strategies. One strategy, OE (for equal parameter case), assumed all items to be of equal difficulty and required only five parameters be estimated for the model. The other strategy, OU (for unequal parameter case), made no assumption of equal item difficulty and therefore required five estimated parameters for each of the 84 items. The necessary response history for the strategies consisted of the subjects' success-failure record for the trials. (For a more complete discussion of these models see Atkinson, 1972b)

The two other treatment conditions were instructional strategies not generated from models of the hypothesized learning process. Because there has been some controversy regarding the student's ability to structure his own learning experience (Atkinson, 1972a), one of the
strategies, SS (for self selection), required the subject to choose a sequence of items for himself. The fourth strategy, RO (for random order), which provided that all items be presented an equal number of times in random order, was a response-insensitive (Atkinson, 1972a) control condition that made no use whatever of the student's response history in determining the instructional sequence. Each of the four strategies yielded an appropriate algorithm for selecting items for test and study during the instructional session.

Thirty Stanford undergraduates randomly assigned to each of the four treatments participated in a two hour instructional session during which they attempted 336 trials covering 84 German-English associates (See Figure 2 for an illustration of the procedure used during the instructional session.). One week later they were given a short test over all 84 items.

The results of the experiment (shown graphically in Figure 3) indicate that, on the delayed test, subjects given the most sophisticated mode-generated CAI performed best; self selection and equal parameter CAI subjects demonstrated similar performance at a lower level; and random
Fig. 2. Flow chart describing the trial sequence during the instructional session. The selection of a word for test on a given trial (box with heavy border) varied over experimental conditions. From "Ingredients for a Theory of Instruction" by R. Atkinson, American Psychologist, 1966, Oct., 926. Copyright 1966 by the American Psychological Association.
Fig. 3. Proportion of correct responses in successive trial blocks during the instructional session and on the delayed test administered 1 wk. later. From "optimizing the Learning of a Second Language Vocabulary" by R. Atkinson, Journal of Experimental Psychology, 1972, 96, 1, 126. Copyright 1972 by the American Psychological Association.
sequence subjects performed the worst. This ordering was an inversion of the groups' performance during the instructional session. According to Atkinson, this was an illustration of the principle used in the models: to learn the material efficiently, test only items that have not yet entered the permanently-learned state, i.e., discontinue testing learned items and test items not yet mastered. Intuitively, this process (apparently mimicked to some extent by subjects in the SS condition) would lead to a large number of errors during the instructional session but a maximal number of learned items at the end of the instructional session. Assuming the hypothesis to be correct, unequal parameter CAI was more sensitive and equal parameter CAI was less sensitive in determining which items were not fully mastered than were the students themselves. Random sequencing resulted in the best instructional session performance because items already mastered were tested more often than necessary, resulting in fewer errors. By the same token, fewer items entered the learning state and were retained at the long-term memory test.

Atkinson's study demonstrates that the level of individualization used to teach a relatively simple, non-hierarchical task can account for a substantial amount of variability in the relative long-term retention that occurs. Not suprisingly, the study also points out that:
1) instruction generated from a highly parameterized model can be superior to both instruction derived from a less complex model and instruction that a reasonably sophisticated subject could structure for himself; and 2) response-sensitive strategies can be more effective than response-insensitive strategies.

The CAI programs used in both of the preceding studies, developed at Stanford on a PDP-1 computer, were ventures that suggest considerable expense. Since the best individualization is most likely to occur when it is the product of a relatively sophisticated theoretical base (Atkinson, 1972a; Atkinson & Paulson, 1972; Groen & Atkinson, 1966), substantial dollar costs can be incurred in development and implementation. This being the case, it becomes necessary to determine what levels of complexity will result in judicious use of individualization. That is, given some context, how much individualization is adequate and cost effective? From the research conducted at Stanford over the past several years, Atkinson (1974) has concluded that the trend in CAI should be "... to develop low-cost computer assisted instruction that supplements classroom teaching and concentrates on those tasks in which individualization is critically important" not only because of high costs, but also because instruction of some tasks is better left to the classroom teacher. To follow Atkinson's suggestion appropriate combinations
of tasks and levels of individualization must be explored.

The present study attempts to investigate one such relationship by comparing various levels of individualization for the same experimental task using a paradigm similar to the one used in Atkinson, 1972b. Briefly, this involves using four different drill & practice programs to supplement simulated textbook instruction of a simple mathematical task -- adding, subtracting, multiplying, and dividing in the base eight number system. The experiment is designed to provide insight on the interaction between the type of task and the type of drill & practice by addressing two questions. First, does the task, hypothesized to have a hierarchical structure (as opposed to a linear structure exemplified by paired-associate tasks), suggest a need for individualization? And second, can very simple, informally optimized CAI be sensitive enough to provide a subject with more effective practice than either: a) practice that he could himself structure; or b) random practice that has no organized structure?

METHOD

Subjects. - Subjects for the study were 44 paid volunteers from the Rice University summer community. Each of the subjects was randomly assigned to one of the four experimental conditions resulting in a total of 11
subjects per group. The subjects, composed primarily of high school and college students, spanned a wide range of mathematical sophistication and age.

Experimental design. - A 1 x 4 between subjects design was used to compare four different CAI drill & practice programs for octal (base eight) arithmetic. For each subject the experimental session, which lasted approximately 80 minutes, consisted of a 15 minute (simulated) textbook course in octal arithmetic, an 8 minute octal pretest, a 40 minute drill & practice session, and an octal posttest, in that order. Presentation of the drill & practice programs was done using a PDP-8L computer equipped with a remote TV monitor and typewriter keyboard.

Materials. - An Octal Primer (See Appendix 1.), developed as a short, simulated textbook course in octal arithmetic, was used to provide subjects with all the information necessary to perform the experimental task. The Primer included illustrations and step-by-step instructions for performing addition, subtraction, multiplication, and division operations in base eight.

Four types of items, addition (A), subtraction (S), multiplication (M), and division (D), were presented during the drill & practice session and used on pretests and posttests. The assortment of items for each of the four types included a sampling of three general levels of difficulty: easy (1), intermediate (2), and difficult (3)
The 12 resulting sets of problems were stored as random access files in the computer. Two speeded paper-and-pencil tests, an octal pretest and an octal posttest, were designed to isolate the effects of the drill & practice conditions. Both the pre-test and the posttest had eight minute time limits and consisted of 24 and 48 items respectively. The tests employed repetitions of the following item sequence: easy items of all four types, intermediate items of all four types, and difficult items of all four types. Instructions to the subjects included the time limit and the cautions that items were to be worked in the order that they appeared on the test and that items were not to be skipped unless a solution could not be found after a minute or two.

Four CAI programs, two very simple response-sensitive programs and two control programs, were constructed to structure the sequencing of items during the drill & practice session.

The first, LH (for learning hierarchy), was based on the "learning sets" concept (Gagne's, 1962). The theory implies that subjects learn, to some arbitrary criterion, each of the levels in the hierarchy, beginning with the lowest level. The concept of hierarchy was approximated by having subjects practice problems in the following order: addition, subtraction, multiplication, and division.
according to the logic shown in Appendix 3a.

The second program, DO (for drop out), was based on a modification of the concept underlying the unequal-parameter drop out method used by Atkinson (1972b) to present vocabulary items. Used here, it implied presenting the type of item showing the poorest mastery, subject to the constraint that presentation of items within a type occur in ascending order of difficulty appropriate to the demonstrated mastery for that type. The program logic is shown in Appendix 3b.

The third program, SS (for self selection), was an adaptation of the SS condition used by Atkinson (1972b). The method, a control for both of the previous programs, is based on the theory that, given information concerning instructional goals, instructional units, and appropriate feedback, the subject can develop an efficient sequencing scheme for drill & practice. In this study, the subject was required to choose the practice problem he wanted by specifying a type (A, S, M, D) and a difficulty level (1, 2, 3) for each trial. This is illustrated in Appendix 3c.

The fourth program, RN (for random), was a response-insensitive control for all of the preceding conditions. Like Atkinson's RO condition, items were presented in no logical order as indicated in Appendix 3d.

Procedure -- After a brief introduction to the
experiment, each subject was seated in a sound-deadened room containing a TV screen, remote typewriter computer terminal, and a scratchpad. During the first part of the experiment each subject was asked to spend 15 minutes reading and studying the Octal Primer. In addition, he was informed that a base eight arithmetic test would follow the study period. After the octal pretest was completed, each subject was given one of the four drill & practice treatments for 40 minutes. Before the session began, however, each subject was made aware that a final octal arithmetic test would follow the drill & practice.

For the drill & practice each subject was seated in front of the TV screen and the keyboard and provided with the Primer and a scratchpad. A series of "practice trials," each divided into a response segment and a study segment, was then presented according to the algorithm (See Appendices 3a-d.) for the condition to which the subject was assigned. For the LH, D0, and RN conditions the experimenter initiated each trial by typing in a "type and level" code (A1 for easy addition, M2 for intermediate multiplication, etc.) from the teletype. In the SS condition, the subject, who had been previously instructed how to select his own sequence of practice problems with the type and level code, initiated each trial.

During the response segment of each trial an unsolved
octal arithmetic problem was displayed on the screen. In two minutes or less, the subject, using scratch paper and referring to the Primer if necessary, was required to type in an answer and hit the "return key" on the keyboard. After an answer was entered or the time had expired, the study segment of the trial began. The computer, in addition to displaying an indication of "correct" or "incorrect," also displayed the subject's original answer and the fully worked octal solution for a period of either two minutes or until the subject terminated the study period by hitting the return key. At the end of the study segment a new trial was initiated. This process continued for 40 minutes.

After the completion of the drill & practice session and administration of the posttest, each subject was paid and debriefed.

RESULTS AND DISCUSSION

Since the criterion for skill in an arithmetic task is often considered to be some combination of speed and accuracy, the dependent measures in this study were chosen to reflect those qualities. Speed was measured by the total number of correctly solved items for a given period of time (i.e., during the pretest, the drill & practice, or the posttest) while accuracy was measured by percent
correct of the total number of items attempted in a given period. Due to intentional biasing effects induced through the general instructions to the subjects (e.g., to work carefully and not to skip problems), it is likely that the variability of the accuracy measure was attenuated; and hence probably contributed less to the total variability of the speed-accuracy construct than did the speed variable. Although this artificial constraint is somewhat undesirable, the manipulation was necessary to guard against improper functioning of the LH and DO drill & practice programs (which used response histories composed of the record of correct and incorrect responses); and to reduce subjects' tendency to skip over difficult test items in order to work more simple ones. It is not suspected that the instructional biasing had differential effects on any systematic variability induced by the treatment conditions themselves.

The first two issues to be discussed are those bearing directly on the nature of the task itself, i.e., task structure and general appropriateness of individualizing drill & practice for the task. An analysis of the pooled response data (for all subjects across all treatments) from the drill & practice sessions provides some affirmative support for the hypotheses that the task was both hierarchical and suitable for individualization of drill & practice.

A task is hierarchical when it is composed of a set of functions increasing in complexity such that the least
complex are generally mastered before the most complex because the former are usually elements of the latter. Since the functions addition, subtraction, multiplication, and division were presented as a hierarchy in the Octal Primer, evidence of the same hierarchy was expected to occur during drill & practice on octal arithmetic items. If it can be assumed that the ordering in the hierarchy is reflected by the ordering of the mean difficulties of the functions, then the structure of the hierarchy can be inferred by a plot of the mean difficulties.

It will be remembered that the drill & practice items were constructed to conform to three levels of increasing difficulty for all four functions, and that items in each of the levels (across functions) were to be of the same relative difficulty. Assuming that response latency (the time needed to enter an answer) for correct items is a reasonable index of item difficulty, the plot of the mean response latencies for all 12 function-level combinations (See Figure 4) indicates that the expected hierarchy did

Insert Figure 4 about here

occur. Even though the levels were not equally spaced or parallel, the relative ordering of the functions was generally maintained at each of the three levels. A clearer demonstration of the hierarchy is illustrated by plots of the mean response latencies and the mean
Fig. 4. Mean response latencies (pooled for all subjects) for each of the 12 categories (function-level combinations) presented during the Drill & practice.
probabilities of error in Figures 5 and 6, respectively.

Insert Figures 5 and 6 about here

The following is offered as a probably explanation of the observed data. Addition and subtraction are at the bottom of the hierarchy because they involve the fewest operations. The drill & practice items were constructed such that the probability of a borrow in subtraction was greater than the probability of a carry in addition, thus accounting for the greater difficulty of subtraction. Multiplication occupies the next position in the hierarchy because it also includes subtraction operations. By the same logic, division is the top of the hierarchy because it also includes multiplication and subtraction operations.

Again turning to the drill & practice data, the issue of the task's suitability for individualization can be addressed. Although not confirming the actual need for individualization, a large amount of between subjects variability in the SS condition might at least provide modest support for the hypothesis of "suitability." Figures 7a-c

Insert Figures 7a-c about here

are plots of the frequency with which subjects in the SS condition chose one of the 12 possible function-level combinations on any of the first 27 trials (the lowest number of trials completed by any subject in the SS condition).
Fig. 5. Mean response latencies (pooled for all subjects) collapsed across levels for each of the function types presented during drill & practice.
Fig. 6. Mean probabilities of error (pooled for all subjects) collapsed across levels for each of the function types presented during drill & practice.
Fig. 7a. Frequency distribution of SS subject item choices during drill & practice plotted in ascending order of difficulty as observed in Figure 4, trials 1-10.
Fig. 7b. Frequency distribution of SS subject item choices during drill & practice plotted in ascending order of difficulty as observed in Figure 4, trials 11-20.
Fig. 7c. Frequency distribution of SS subject item choices during drill & practice plotted in ascending order of difficulty as observed in Figure 4, trials 21-27.
The frequency distributions, arranged according to increasing difficulty (as shown in Figure 4), do indicate a trend toward more difficult items on later trials, but a substantial amount of variability is still evident. In addition, Table 2 presents the mean and standard deviation for the

\begin{table}
\centering
\begin{tabular}{|c|c|}
\hline
    & \\
\hline
Row 1 & \text{Column 1} \\
Row 2 & \text{Column 2} \\
\hline
\end{tabular}
\caption{Table 2}
\end{table}

number of total trials, and for the total number of each of the four functions attempted by subjects in the SS condition. Here again, a considerable amount of variability among subjects is evident. The fact that subjects chose different practice strategies, even in the face of potential biasing toward a single strategy (learning first to add, then to subtract, etc.) as a result of both prior experience and the structure of the Octal Primer, does suggest that individualizing the practice may have beneficial effects.

To test the main effect of the different drill & practice treatments a one-way multivariate analysis of covariance was used (See Huck & McLean, 1975, for a discussion of analyses of pretest-posttest designs.). A linear combination of the total number of correctly solved items on the posttest and the associated percent correct on the posttest served as the multivariate analysis variate (the speed-accuracy variable), while the same measures from the pretest were used as covariates. The test yielded an (approximate) $F_{6,74} = .8845$ with $p \geq .5$. In addition,
TABLE 2

Means and standard deviations of the total number of trials attempted and of the types of trials attempted by subjects in the SS condition.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total trials</td>
<td>50.36</td>
<td>18.20</td>
</tr>
<tr>
<td>Add. trials</td>
<td>9.63</td>
<td>9.96</td>
</tr>
<tr>
<td>Sub. trials</td>
<td>7.63</td>
<td>6.32</td>
</tr>
<tr>
<td>Mult. trials</td>
<td>17.81</td>
<td>8.94</td>
</tr>
<tr>
<td>Div. trials</td>
<td>15.27</td>
<td>12.28</td>
</tr>
</tbody>
</table>
separate analyses of covariance were performed on each of the two variables used in the linear combination. For the total number of correct responses $F_{3,39} = .51$ with $p = .5$, while $F_{3,39} = .48$ with $p = .5$ for percent correct responses. Given that the assumptions required for proper application of the above analyses were met, these results do not contradict the null hypothesis of "no differences between treatment means" with any degree of certitude.

While it is never logically possible to pinpoint the reasons for experimental results that fail to reject the null hypothesis, future studies may often be improved as a consequence of judicious a-posteriori examination of the data. In the present study, the lack of observed differences between the response-sensitive strategies and the response-insensitive strategy (a result that has occurred reliably in previous studies (cf. Atkinson, 1972a, 1972b)) is probably due to a combination of noise factors. The most likely of these relates to the experimental population. Table 3 points out that not only did groups differ substantially on the premeasure, but that they also demonstrated a considerable range of performance on both pre-tests and posttests. Small sample sizes and large variances make detection of a weak signal difficult.

Another factor that may have obscured between
TABLE 3

Pretest and posttest means and standard deviations for dependent measures, number correct and % correct.

<table>
<thead>
<tr>
<th>Cond.</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># corr.</td>
<td>s.d.</td>
</tr>
<tr>
<td>SS</td>
<td>11.81</td>
<td>6.17</td>
</tr>
<tr>
<td>LH</td>
<td>9.36</td>
<td>5.32</td>
</tr>
<tr>
<td>DO</td>
<td>10.82</td>
<td>8.28</td>
</tr>
<tr>
<td>RN</td>
<td>8.0</td>
<td>7.59</td>
</tr>
</tbody>
</table>


treatment differences was the possibility of overlearning during the drill & practice sessions. If the sessions were too lengthy it is not inconceivable that any type of drill & practice may have resulted in improved performances on the posttest.

These factors suggest that a more sensitive test of the experimental hypothesis would be possible if 1) large samples were to be drawn from a relatively homogeneous population, and 2) drill & practice sessions were to be terminated as soon as the subject reached error-free performance levels.

In conclusion, it would seem that while there was evidence favoring the use of individualized drill & practice for the experimental task; the statistical comparison performed indicated no support for using relatively simple CAI drill & practice programs. The statistical results do not, however, preclude the possibility that those, or similar programs may be effective. That determination can only come from more precisely controlled replications of the study.
INSTRUCTIONS

This is an experimental investigation of some of the properties of computer-assisted instruction. As a subject in this experiment you will be taught how to add, subtract, multiply, and divide in the base eight number system.

The base eight (OCTAL) system has all the properties of the base ten (DECIMAL) system. The difference is that the base eight system has only eight characters, zero through seven. Counting proceeds in this way: 0, 1, 2, 3, 4, 5, 6, 7, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, ... Because of this change in counting you will need to revise the way you would normally add, subtract, multiply, and divide in decimal to perform the same operations in octal. The following set of instructions and examples explains how to do the operations in octal. You will have only 15 minutes to read and study these instructions. At the end of the 15 minute study period you will be given another OCTAL ARITHMETIC TEST, so please do your best to understand the instructions.

OCTAL ADDITION:

1) Do a decimal addition of the first column. If the sum is less than 8 for the column, write down the digit as the octal sum for the column.

EXAMPLE:

\[
\begin{array}{c}
2 \\
+2 \\
\hline
4 \\
\end{array}
\text{ (octal sum)}
\]
2) If the column sum is greater than or equal to 8, do a decimal subtraction of 8 from the column sum and write down the remainder as the octal sum for that column. Carry a 1 to the decimal addition of the next column.

**EXAMPLE:**

\[
\begin{array}{c}
25 \\
+16 \\
\hline
\
\end{array} \quad \begin{array}{c}
+1 \\
+2 \\
\hline
+5 \\
\hline
411 \text{ (decimal column sum)}
\end{array} \quad \begin{array}{c}
25 \\
+16 \\
\hline
\
\end{array} \quad \begin{array}{c}
43 \text{ (octal)}
\end{array}
\]

\[
\begin{array}{c}
25 \\
+16 \\
\hline
\
\end{array} \quad \begin{array}{c}
+1 \\
+2 \\
\hline
+5 \\
\hline
411 \text{ (decimal column sum)}
\end{array} \quad \begin{array}{c}
25 \\
+16 \\
\hline
\
\end{array} \quad \begin{array}{c}
43 \text{ (octal)}
\end{array}
\]

\[
\begin{array}{c}
4 \quad 11 \quad \frac{3}{4} \\
-8 \\
\hline
3 \quad 3 \quad 0
\end{array}
\]

**MORE EXAMPLES:**

\[
\begin{array}{c}
63 \\
+27 \\
\hline
\
\end{array} \quad \begin{array}{c}
+1 \\
+2 \\
\hline
+3 \\
\hline
9 \quad 10 \quad 2 \quad (\text{octal sum})
\end{array} \quad \begin{array}{c}
63 \\
+27 \\
\hline
\
\end{array} \quad \begin{array}{c}
112 \text{ (octal)}
\end{array}
\]

\[
\begin{array}{c}
747 \\
+155 \\
\hline
\
\end{array} \quad \begin{array}{c}
+1 \\
+1 \\
\hline
+4 \\
\hline
9 \quad 10 \quad 12
\end{array} \quad \begin{array}{c}
747 \\
+155 \\
\hline
\
\end{array} \quad \begin{array}{c}
1124 \text{ (octal)}
\end{array}
\]

\[
\begin{array}{c}
747 \\
+155 \\
\hline
\
\end{array} \quad \begin{array}{c}
+1 \\
+1 \\
\hline
+4 \\
\hline
9 \quad 10 \quad 12
\end{array} \quad \begin{array}{c}
747 \\
+155 \\
\hline
\
\end{array} \quad \begin{array}{c}
1124 \text{ (octal)}
\end{array}
\]

\[
\begin{array}{c}
-8 \\
-8 \\
\hline
1 \quad 1 \quad 2 \quad 4 \quad (\text{octal sum})
\end{array}
\]

**OCTAL SUBTRACTION:**

1) If the minuend is larger than or equal to the subtrahend, subtract just as you would in decimal and write down the result as the octal column difference.

**EXAMPLE:**

\[
\begin{array}{c}
6 \text{ (minuend)} \\
-4 \text{ (subtrahend)} \\
\hline
2 \text{ (octal difference)}
\end{array}
\]
2) If the subtrahend is larger than the minuend a borrow must be made from the next column of the minuend. Reduce that column of the minuend by 1 and do a decimal addition of 8 to the minuend column to the right. Do a decimal subtraction of the subtrahend and write down the result as the octal column difference.

**EXAMPLE:**

\[
\begin{array}{c}
22 \\
-6 \\
\hline
\end{array}
\begin{array}{c}
1 \\
6 \\
\hline
4
\end{array}
\]

(Octal difference) \(\frac{14}{14}\)

**MORE EXAMPLES:**

\[
\begin{array}{c}
134 \\
-61 \\
\hline
\end{array}
\begin{array}{c}
0 \\
6 \\
\hline
5
\end{array}
\begin{array}{c}
3 \\
1 \\
\hline
3
\end{array}
\]

(Octal difference) \(\frac{53}{53}\)

\[
\begin{array}{c}
1211 \\
-532 \\
\hline
\end{array}
\begin{array}{c}
0 \\
5 \\
\hline
4
\end{array}
\begin{array}{c}
1 \\
3 \\
\hline
7
\end{array}
\]

(Octal) \(\frac{457}{457}\)

**OCTAL MULTIPLICATION:**

1) Do a decimal multiplication of the first (rightmost) digit in the multiplicand by the first digit in the multiplier. If the result is less than 8 write it down as the octal product for that column.

**EXAMPLE:**

\[
\begin{array}{c}
3 \text{ (multiplicand)} \\
\hline
2 \text{ (multiplier)} \\
\hline
6 \text{ (octal product)}
\end{array}
\]

2) If the decimal multiplication of the first digit in the
multiplicand by the first digit in the multiplier is a product greater than or equal to 8, do a decimal subtraction of the largest multiple of 8 (8, 16, 24, 32, 40, 48, etc.) that is just less than or equal to that product. Write down the remainder as the octal column product and carry the order of the multiple of 8 (Carry 1 if 8 was the largest multiple; carry 2 if 16 was the largest multiple; carry 3 if 24 was the largest multiple, etc.).

EXAMPLE:

\[
\begin{array}{c}
6 \\
\times 5 \\
\hline
? \\
\end{array}
\]

\[
\begin{array}{c}
+3 \\
\times 5 \\
\hline
6 \\
\end{array}
\]

\[
\begin{array}{c}
6 \\
\times 5 \\
\hline
30 \\
-24 \text{ (carry 3)} \\
\hline
6 \text{ (octal product)} \\
\end{array}
\]

3) If there is another column in the multiplicand do another decimal multiplication and then a decimal addition of the carry if any. From this number subtract a multiple of 8 as in step 2.

EXAMPLE:

\[
\begin{array}{c}
25 \\
\times 4 \\
\hline
? \\
\end{array}
\]

\[
\begin{array}{c}
+1 \quad +2 \\
\times 2 \quad 5 \\
\hline
10 \quad 20 \\
-8 \quad -16 \\
\hline
124 \text{ (octal)} \\
\end{array}
\]

\[
\begin{array}{c}
25 \\
\times 4 \\
\hline
10 \quad 20 \\
-8 \quad -16 \\
\hline
124 \text{ (octal)} \\
\end{array}
\]

4) If there is more than one digit in the multiplier repeat steps 1 through 3 and do an octal addition of the partial products.
EXAMPLE:

\[
\begin{array}{cccccc}
47 & +2 & +3 & 7 & +3 & +5 \\
x & 4 & 4 & 6 & 0 \\
\hline
19 & 28 & 29 & 42 & 24 & 40 \\
\hline
2 & 3 & 4 & 3 & 5 & 2 \\
\end{array}
\]

Now do an octal addition of the first and second octal products.

\[
\begin{array}{cccccc}
47 & x64 \\
\hline
2 & 3 & 4 & 352 & 3754 \\
3 & 7 & 5 & 2 & 4 \\
\end{array}
\]

OCTAL DIVISION:

1) Estimate the first digit in the octal quotient by doing a decimal division of the dividend by the divisor. Then check the estimate by doing an octal multiplication of the estimated quotient by the divisor. If the result is less than or equal to the dividend, retain the estimated value as the first digit of the octal quotient. Otherwise, lower the estimate and again do an octal multiplication of the estimated quotient by the divisor. Continue lowering the estimate until the octal product of the estimate and the divisor is less than or equal to the dividend.

EXAMPLE:

\[
\begin{array}{cccc}
2) & 12 & 6 & (estimated octal quotient) \\
\hline
2) & 12 & (dividend) \\
\hline
2) & 5 & 5 & (octal) \\
\hline
\end{array}
\]

6 (octal mult.) 
14 (14 is too large estimate 5) 
12 (octal quotient) 
0 (remainder)
2) When the first digit of the octal quotient has been determined, subtract the octal product of the first digit of the quotient and the divisor from the dividend. The remainder will now form the new dividend. Find the second digit of the octal quotient by using decimal division to estimate as in steps 1 and 2.

EXAMPLE:

\[
\begin{array}{c}
5)1577 \\
\underline{3} \quad \text{(estimate) } 3 \\
5)1577 \\
\underline{15} \quad \text{(octal)} \\
\underline{12} \quad \text{Too large, try 2.} \\
\underline{2} \quad \text{(estimate) } 2 \\
5)1577 \\
\underline{10} \quad \text{(octal)} \\
\underline{12} \quad \text{Ok, now subtract and estimate the second digit.} \\
\underline{27} \quad \text{(estimate) } 7 \\
5)1577 \\
\underline{12} \quad \text{(octal)} \\
\underline{37} \quad \text{Too large, try 6.} \\
\underline{26} \quad \text{(estimate) } 6 \\
5)1577 \\
\underline{12} \quad \text{(octal)} \\
\underline{36} \quad \text{Ok, now subtract and estimate the third digit.} \\
\underline{263} \quad \text{(estimate) } 3 \\
5)1577 \\
\underline{12} \quad \text{(octal)} \\
\underline{37} \quad \text{The estimate is correct as it stands.} \\
\underline{36} \\
\underline{17} \\
\underline{17} \\
\underline{0}
\end{array}
\]
Appendix 2

Examples of problems for test and study in base eight:

level 1 addition: \[ \begin{array}{c} 5 \\ \hline +3 \\ \hline \end{array} \]
\[ \begin{array}{c} +3 \\ \hline 10 \end{array} \]
solution: \[ \begin{array}{c} \_ \\ \hline ? \end{array} \]

level 2 subtraction: \[ \begin{array}{c} 53 \\ \hline -24 \\ \hline \end{array} \]
\[ \begin{array}{c} -24 \\ \hline 27 \end{array} \]
solution: \[ \begin{array}{c} \_ \\ \hline ? \end{array} \]

level 3 multiplication: \[ \begin{array}{c} 1247 \\ \hline \times 305 \\ \hline \end{array} \]
\[ \begin{array}{c} 3765 \\ \hline 405203 \end{array} \]
solution: \[ \begin{array}{c} 6503 \\ \hline 0000 \end{array} \]

level 3 division: \[ \begin{array}{c} 22 \) 1714 \\ \hline \end{array} \]
\[ \begin{array}{c} 66 \\ \hline 154 \\ \hline 154 \\ \hline 0 \end{array} \]
solution: \[ \begin{array}{c} 22 \) 1714 \\ \hline 154 \\ \hline 154 \\ \hline 0 \end{array} \]
Appendix 3a.

Item presentation algorithm for the LH condition. Subjects receive appropriate items with type (A, S, M, or D) and level (1, 2, or 3) depending on values of X, i, and j. The initial item in the session is an Ai denoted as A_{1,0}.
Item presentation algorithm for the D0 condition. Subjects receive appropriate items with type (A, S, M, or D) and level (1, 2, or 3) depending on values of X, i, and j. The initial item is a randomly chosen $X_{1,0}$. 

Appendix 3b.

display item $X_{1j}$ where $X=A,S,M$, or D; $i=1,2, or 3$; and $j=0 - n$. Start with lowest $i$ and $j$ examining $i$ then $j$; use random choice for tie

is $\text{timer}_1<2\text{min.}$?

no

yes

is $i>1$?

no

yes

is answer corr.?

no

yes

is $j<2$?

no

yes

is $i<3$?

no

replace $i$ with $i-1$ set $j=0$

replace $j$ with $j+1$

replace $i$ with $i+1$ set $j=0$

Terminate practice session

Start practice session

reset $\text{timer}_1=0$

display feedback

is $\text{timer}_2<40\text{min.}$?

no

yes

set $j=0$

no
Appendix 3c.

Item presentation algorithm for the SS condition. Subjects select their own items by typing A, S, M, or D for the type and 1, 2, or 3 for the level.
Appendix 3d.

Item presentation algorithm for the RN condition. Subjects receive items from all 12 type-level categories in random order.
REFERENCES


Huck, S. W. & McLean, R. A. Using a repeated measure ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. Psychological Bulletin, 1975, 82, 4, 511-518.


