DOCUMENT RESUME

ED 040 967 24 SP 004 017

AUTHOR Natkin, Gerald L.
TITLE Research and Theory on the Effects of Instructional Sequencing.
INSTITUTION Bucknell Univ., Lewisburg, Pa.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
BUREAU NO BR-8-B-100
PUB DATE Jun 70
GRANT OEG-2-9-480100-1005(010)
NOTE 27p.

EDRS PRICE EDRS Price MF-$0.25 HC-$1.45
DESCRIPTORS *Achievement, High School Students, *Learning Processes, *Programed Instruction, *Sequential Learning, *Sequential Programs

ABSTRACT

Although previous research has indicated no significant relationship between sequencing of programed instructional materials and terminal achievement, it is felt that certain conditions might produce such a relationship. In order to test this idea, a hierarchy of word meanings was constructed, in which each definition was dependent on the previous one. The hierarchy was divided into several different linear and scrambled versions according to the amount of memory and information reorganization required to learn the material. The linear and scrambled versions were then administered to two groups of 80 high school students. Factoral analysis of results showed a significant difference in achievement between students using the linear version and those using a scrambled version for the lower degrees of memory requirement. However, there was no difference in achievement at the higher levels of memory requirement. Conclusions were that sequence does have an effect in programed instruction when a certain optimal memory load is not exceeded, and that further research should be done on the learning mechanisms involved in programed instruction. (RT)

RESEARCH AND THEORY ON THE EFFECTS OF INSTRUCTIONAL SEQUENCING

Project No. 8-B-100

Grant No. OEG-2-9-480100-1005-(010)

by

Gerald L. Natkin

June, 1970

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare.

Bucknell University

Lewisburg, Pennsylvania

CONTENTS

RESEARCH AND THEORY ON INSTRUCTIONAL SEQUENCING

Introduction

The development of optimally effective instructional sequences is one of the most complicated and difficult of educational problems. Decisions about the order in which concepts and principles are to be taguht, and about the order in which skills are to be developed occur in most instructional planning efforts, at all levels of education. Intuitively they seem to be among the most important of educational decisions.

Sequencing decisions are also quite important theoretically. From a Skinnerian point of view, for example, if the instructional sequence is not correct the learner will make many errors during learning, and be reinforced too infrequently for effective shaping. Or, the somewhat different model of Gagné, (1962) predicts that incorrect sequencing would impede the development of prerequisite knowledge, thus preventing achievement of the terminal objective.

Research evidence on the importance of sequencing is rather sparse, and is largely concerned with sequencing in programmed or computer-assisted instruction. With only one exception, however, the results are negative regarding its effect on terminal achievement level. In box-score terms, sequencing appears not to be as important as intuition as some learning theories suggest.

This conclusion should probably not be taken at face value, even though at least one author (Niedemeyer, 1968) appears to believe it. Some of the studies are flawed methodologically, and none have been based on a detailed analysis of how sequencing could operate to effect terminal achievement. The latter failing is especially critical: without specifying how sequencing should produce an effect it is quite probable that irrelevant factors could be operating to obscure it.

Review of Related Research

This section will be concerned with reviewing several of the previous studies, and summarizing their findings, followed by an attempt to formulate a preliminary model of the psychological basis for sequence effects, and a re-examination of these studies in the light of this model.

Studies of sequencing in PI, while varying widely in the content areas used and the learner populations sampled, have used very similar methods in one respect: all have used two different types of presentation sequences: (1) "Linear": ("Ordered," "Logical," "Organized") in which the frames were presented in their original order, and (2) "Scrambled," in which the frames are presented in random order. In one study (Niedemeyer, et al., 1968), the material was also presented backward. Ss are then usually compared on the number of errors they make on in-text questions, the amount of time required to complete the program, and post-test performance.

Roe, Case, and Roe (1962) presented 36 college students with a 71-frame program on probability theory. Ss were given either linear or scrambled versions. The groups were stratified on mathematical ability. After completing the program Ss took a ten-item post-test. The results indicated, contrary to expectation, that there were no differences due to sequencing on the post-test, in-text errors, or study time. They concluded that sequencing may not be as important as had previously been supposed.

This study suffers from two main difficulties. First, a 71-frame program is extremely short, and cannot have covered much probability theory. It seems likely that the portion of the topic this program could have covered would tap Ss' pre-knowledge of the content, which would tend to obscure any treatment effect. Second, their criterion test, being so short, might well have been very unreliable, and hence insensitive to any effect.

Levin and Baker (1963) ran a similar experiment. In this study 36 second grade students were given PI in informal geometry. Although the program was quite long, only a specific sequence of 60 frames was studied experimentally. The 60 frames were broken into three blocks of 20 frames each. The L group received the original program. The S group received the blocks of 20 in their original order, but within each block, frames were randomly ordered. A 52-item post-test (reliability = .82) was administered. No significant differences were found on in-text errors or on criterion performance.

Pre-familiarity was probably not a problem here, nor was test unreliability. It may be, however that unscrambling a 20-item sequence does not represent a great difficulty. This possibility is supported by Niedemeyer's (1968) suggestion that the sequence effect may be a function of program length. There does not, however, seem to be any evidence on this point.

More recently, Payne, Krathwohl, and Gordon (1967) hypothesized that sequence effects would be a function of the degree of interrelatedness in a program, such that scrambling a loosely-interrelated program would have little effect, while scrambling a tightly-interrelated one would. Three programs on various measurements and statistical concepts, varying in degree of judged interrelatedness were given to 195 Ss in either S or L sequences. All of the permutations, SSL, SLS, etc. were used.

Ss were given a 56-item immediate post-test and a 53-item delayed post-test two weeks later. There were no significant differences in either immediate or delayed post-test scores. Nor were there any significant differences on in-text errors, though there was a trend toward greater differences as interrelatedness increased.

This study suffered from one major flaw. The Ss were given the material to study on their own time for a 48-hour period. During this time they could have unscrambled the program, taken as much time as they wanted, or even studied a friend's copy! Such a degree of uncertainty renders the results of little value.

The studies discussed above are all based on the implicit hypothesis that a "logical" or "ordered" sequence should be optimal for terminal achievement. None of the authors, however, have specified any operational criteria for deciding when these adjectives apply. It seems, rather, to have been assumed that no sequence could be more "logical" than an instructional program as originally written. This is a very questionable assumption, in the absence of some set of rules for constructing sequences, and may, in part, account for the negative findings; in order to detect effects due to destroying a sequence, it is necessary to have begun with an effective one. What, then, is an "ordered" or "logical" sequence?

It seems fair to assume that the authors mentioned above intended reference to something like Gagné's (1962) notion of a "hierarchy of knowledge." In Gagné's procedure a terminal objective is stated, and the subordinate capabilities ("learning sets") directly effecting achievement of the objective are hypothesized. The same procedure is then applied recursively to the resulting learning sets, until some lowest desired level is attained.

When completed, this task analysis yields a diagram (knowledge hierarchy) which specifies precisely what a learner must master in order to achieve the terminal objective or any of the sub-objectives. Gagné (1962) gave some evidence that the sequence of learning sets resulting from a hierarchical analysis has a high degree of sequential dependency.

Assuming that component learning sets are correctly identified, we may then define a logical sequence as one in which the components are taught in their hierarchical order.

(Care is needed in distinguishing "logical" from "instructionally optimal" order. A logical sequence is one in which we can be sure, at every point, <u>S</u>s have been presented with all <u>necessary</u> information to master the next stage. It is not necessarily the most effective sequence.)

Using this definition, it can be seen that it implies that scrambling a hierarchy should have the effect of increasing the number of errors that <u>S</u>s make while studying the program (Wodtke, et al., 1967). When studying a scrambled program, Ss will not always have been presented with all the information necessary to answer in-text questions. The failure to obtain significant differences on this measure can, therefore, be taken as evidence that none of the studies discussed above began with logical sequences. Their results probably tell little about sequence effects. We must, therefore, look elsewhere.

Wodtke, Brown, Sands, and Fredericks (1967) gave 74 college students a CAI sequence on conversion of base 10 numbers to base N, which was reportedly a hierarchy. Ss with pre-knowledge of the topic were screened out. The post-test was a set of 23 base-conversion items. Ss were given either linear or scrambled versions of the program.

No significant differences were found on the post-test, but there were significantly more in-text errors in the scrambled group. Further, the error rate in the scrambled group decreased almost linearly over time, until it was slightly lower than that of the linear group, which remained almost constant throughout. These data are at least consistent with the presence of a hierarchy. In addition, the decreasing error rate suggests that Ss may have managed to "unscramble" the program for themselves.

Niedemeyer, Brown, and Sulzen (1969) presented Gagné and Brown's (1961) "Number Series" program to 64 ninth grade algebra students. This program, which was thought to be based on Gagné's (1962) analysis, teaches Ss to find formulas for sums of series. The program was given in three sequences: linear, scrambled, and backward.

Comparisons were made on in-text errors, time, a 10-item test of introductory concepts, and a 10-item transfer test. Considering only the linear and scrambled groups, they found significant differences in errors, but none on time, introductory concepts, or transfer. The curve of error rates over time is similar to the one found by Wodtke, et al., (1967), again indicating the possibility of unscrambling. The transfer test results may have been suppressed by a floor effect, all means being quite low.

Brown (1970) used linear and scrambled versions of the same program on 44 eleventh grade trigonometry students. He found significant differences, favoring the linear groups, on all measures. This is, to the author's knowledge, the only clear-cut sequence effect on post-test achievement reported in the literature. The discrepancy between these findings and those of Niedemeyer, et al., can perhaps be attributed to the fact that older Ss were used, thus removing the floor effect. The discrepancy with the Wodtke, et al., results may (according to Brown) be due to the relatively low level of conceptualization with which the base conversions program terminated.

To summarize, sequence effects on terminal achievement have been found in only one study, that of Brown (1970), in which a hierarchy was used. All such studies have found that scrambling increases errors, and in two cases (Wodtke, et al., 1967; Niedemeyer, et al., 1969) these errors have decreased throughout the program. The last point may indicate that scrambling has an effect initially, due to the absence of critical information, but that Ss are able to "unscramble" much of the material.

Redefinition of the Problem

Enough evidence is now available to show that, in the case of auto-instructional programs, sequence is ordinarily not a very potent factor; the point must be considered reasonably well established. Further studies comparing scrambled and linear sequences are unlikely to yield radically different results.

In the author's opinion, more is to be gained from approaching the sequencing problem in a somewhat different manner. Given that sequencing usually doesn't have an effect on terminal achievement, are there any conditions under which it can be expected to have such an effect?

The question can be answered quite generally: sequence will have an effect on terminal achievement whenever it interferes with the process of unscrambling. We must, therefore, try to conceptualize this process, in order to determine where sequence could have such an effect.

Suppose it was necessary to teach the meaning of a word, and that acquiring this meaning depended on knowledge of the meaning of seven other words and six qualifiers, as shown in figure 1.

Figure 1. A hierarchy of word meanings.

A ⟶ B ⟶ C ⟶ D ⟶ E ⟶ F ⟶ G

Assume that all the attributes of A are known, and that learning B requires only qualifying A, etc. What advantage would there be in presenting these definitions in the logical, linear order? The obvious answer is that doing so assures that all of the information logically necessary for comprehension at any point has been presented. Informationally, it would always be possible to comprehend the Nth word after having been presented with the first N-1. From the same point of view, however, it is also clear that the first N-1 words need not have been presented in any particular order; as long as the learner can sort them out, any order would be logically sufficient to comprehend word N+1.

A concrete example will help to make this point clear. Suppose we had three concepts, "A," "B," and "C," and a "program" to teach them. Let "A" signify "dogs," "B" signify "large dogs," and "C" signify "large brown dogs." Suppose these were presented in the following logical sequence:

(1) "A" means "dog"

(2) "B" means "large A"

(3) "C" means "brown B"

We would know at each point that enough information had been presented to render the upcoming concept meaningful. However, if a sequence such as the following were given,

(1) "A" means "dog"

(2) "C" means "brown B"

(3) "B" means "large A"

we might expect the learner to experience more difficulty: some of the information is literally meaningless for him until the end. Logically, however, it is possible, given certain capabilities, for the learner to achieve the same terminal state as an S studying a linear program. If a non-linear sequence is to lower the terminal achievement, then, it must be because these capabilities are not present in sufficient degree, or that the sequence interferes with their functioning, which amounts to the same thing. The identification of these capabilities should then provide a conceptual basis for further research on sequencing.

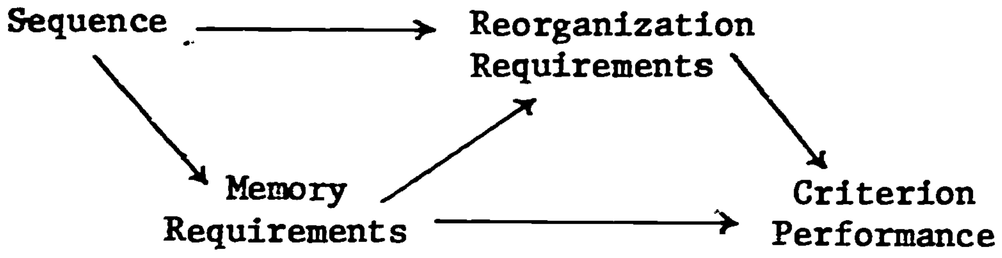
Two important capabilities stand out when the problem is viewed this way: reorganization and memory.

Reorganization refers to the behaviors involved in unscrambling. In the type of task described above, this would require that Ss be able to transfer meaning acquired later in the sequence into previously meaningless terms. Thus, given "C" means brown "B," Ss would, then given "B" means large "A," be required to reconstruct "C means large, brown dog."

The ability to reorganize may depend on being able to keep track of information. If memory is rendered relatively unimportant, as by keeping all information readily available to the learner, the task should be purely a test of inferential capabilities. Should the learner be required to recall everything, however, without any aids, his ability to reorganize is limited by his ability to retrieve necessary information. In the second sequence shown above, for instance, if the learner could not recall that "A" means "dog," he simply could not extract much useful information from "B" means "large A."

It is hypothesized, then, that the ways in which varying the sequence of presentation operates on learning from hierarchies are (1) through changing the amount of reorganization necessary to achieve success on the task, and (2) changing the amount of meaningless information which must be retrieved in order to reorganize. This analysis is summarized in figure 2.

Figure 2. Summary of hypothesized sequence effects.



In the studies described below an attempt has been made to vary these two aspects of sequences independently, and to determine their effects on terminal achievement.

Operational Definitions:

1. Knowledge Hierarchies. Initially it was planned to use existing units of programmed instruction which conformed to Gagné's definition of a hierarchy. Several problems with that method became apparent after attempts to implement it. First, few if any programs, aside from those developed by Gagné, have been constructed by hierarchical analysis. Hence it was unlikely that any would conform very well to the hierarchical model. Second, even though a program was developed by hierarchical analysis, there would be no guarantee that there is a unique logical structure to the task. Repeated attempts at hierarchical analysis by the author and some of his associates on the same task often led to quite different hierarchies. (Niedemeyer, et al., noted that this problem occurred with the "Number Series" program.) Third, existing programs contain devices, particularly repetition and adjunct questions, to assure mastery of sub-tasks. These serve useful functions in the programs, but could easily mask any effect that sequence has on recall or reorganization.

For these reasons, it was decided to develop a hierarchical learning task which could avoid all these difficulties. Such a task would have (1) a known, unique hierarchical structure, and (2) would not be restricted by usual programming methods.

The task designed for these purposes is of the type discussed in the previous section. The objective of the task is to learn the definition of the last item in a sequence of seven terms which form a hierarchy. The linear version of the task is shown in table I in the appendix.

2. Memory Requirements (MR). Initially it had been planned to use various types of material to interrupt the sequence, hence controlling memory requirements of the task. This was attempted in pilot work with college students, using various amounts of poetry to be read, irrelevant programmed instruction, and some arithmetic tasks. These methods proved ineffective in producing interference. In later pilot work it was found that an effective way to vary memory requirements was simply to lengthen the sequence, adding an extra one, two, or three concepts. This procedure was used in the experiments.

3. Reorganization Requirements (RR). The notion of reorganization requirements can be operationalized by scrambling. That is, a hierarchical program which has been randomly re-ordered poses greater reorganization requirements than does a linear ordering of the same program. The scrambling notion, however, is qualitative, and recognizes too few distinct levels of reorganization requirements; there is either the minimal level (linear) or the maximal level (scrambled). Scrambling, however, could produce a linear program, a backward program, or anything in between the two, so the procedure appears too imprecise to yield much information about the effects of RR.

The procedure used in these studies to vary RR is based on the following hypothesis: the greater the sum of the linear distances between all pairs of adjacent concepts, the greater the RR in a sequence (by linear distance we mean the distance in terms of numbers of concepts, between a pair of concepts in the original linear sequence). This definition incorporates the essential idea of RR: the greater the separation between adjacent concepts, the greater the extent to which the learner will have to rely on his own capability to integrate meaningful and meaningless labels during learning. Thus, degree of RR is varied by varying the sum of linear distances between adjacent concepts.

Experiment I

The purpose of this experiment was to determine the effects of varying the following two sequence parameters: (1) reorganization requirements, and (2) memory requirements, on the highest level of achievement in a knowledge hierarchy.

Method

Materials. The learning task was a hierarchy of sequential definitions of labels. Each label was a nonsense syllable given meaning by associating it with a meaningful English noun or adjective, and/or with other labels. Thus, Ss might read "Glux" means "House," then "Baf" means "red Glux."

Sixteen versions of the task were prepared, varying in level of RR (linear, 11, 15, 20) and MR (7, 8, 9, or 10 concepts).

The materials were assembled into stapled booklets, with one "frame" per page. Each page contained a single statement of the type shown above.

The post-test was simply a list of the nonsense syllables, presented on a single page. Ss were requested to write out, in English, all the characteristics associated with the label.

Subjects. Ss were 160 male and female tenth and eleventh grade students of Hamilton High School East, Trenton, New Jersey. The experiment was conducted early in June, 1969. All Ss volunteered for the study.

Design. The design was a 4 x 4 factorial plan, including all combinations of RR and MR. There were 10 Ss per cell.

Procedure. Ss were randomly assigned to the sixteen groups. Ss were run in groups of 40 during three successive class periods. They were told that they were participating in a psychological experiment, and that the results would be kept confidential, not being disclosed to the principal, teachers, counselors, or parents.

They were seated in a large, well-lighted room, and the booklets, which had been randomly ordered, were distributed. The instructions (see appendix) were read, Ss were asked if they had any questions, and were told to begin. At one minute intervals the experimenter told them to turn the page. When the last page was completed, Ss were told to do the post-test.

There was one possible biasing factor in the execution of the study. Since the various MR conditions required different amounts of time to complete the material, it was thought best to run them at different times. Thus, MR was confounded with sessions.

Results

Each S received three scores on each item: (1) the number of correct attributes recalled, (2) the number of intrusions, and (3) the number of correct attributes excluded.

Individual protocols were then scored for the highest item in the sequence which the S recalled without either intrusions or exclusions. This scoring was based on the assumption that in a hierarchical task most, if not all of the information learned is contained in the highest achievement level. It also gives a numerical measure of progress toward the objective. This score will be referred to as the total score. Table 1 gives a 4 x 4 ANOVA of these scores.

Table 1. ANOVA of the total scores, experiment I.

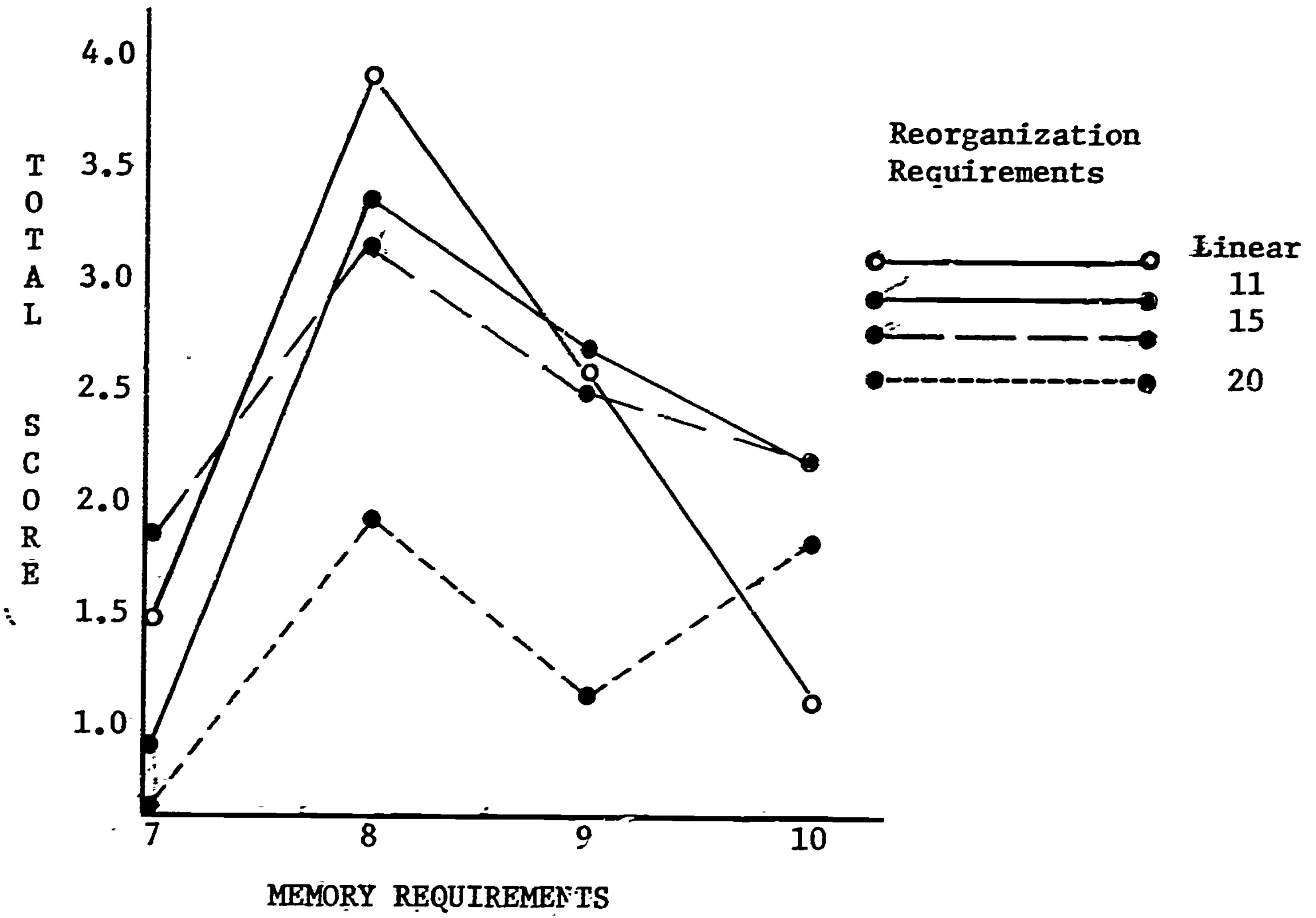
Sources	SS	df	MS	F	P
RR	20.87	3	6.96	4.64	<.01
MR	77.90	3	25.98	3.04	n.s.
RR x MR	76.83	9	8.55	5.70	<.01
Ss: RR x MR	215.90	144	1.50		

The analysis was for a mixed model, since the RR levels, except for the linear case, were sampled.

There were two significant effects, one for RR ($F=4.64$, $P<.01$), and one for RR x MR ($F=8.55$, $P<.01$).

Figure 3 shows the curves of the various RR groups plotted as a function of MR. The differences are concentrated at the 8 and 9 MR levels, and appear to be

Figure 3. Results of experiment I.



between the (linear, 11, 15) and 20 RR groups.

The curves, except for the 20 RR group show definite curvilinearity with marked increases in performance from 7 to 8 MR, and decreases thereafter.

The extremely poor performance of the 7 MR groups was unexpected. It is difficult to understand how the lightest memory load could result in poorest performance, particularly for the linear group.

It was suspected that the results may have been biased by some extraneous environmental factor. Although no such factor was evident during the study, such a bias was possible, since all the groups run at a single sitting were homogeneous on MR. Such an extraneous factor could not have made a difference in the RR main effect, but could have obscured an MR effect, or inflated the MR x RR interaction, since these were both nested under sessions.

It was therefore decided to repeat the experiment, with more adequate control over the possible extraneous factors.

Experiment II

Method

Materials. The materials of experiment II were identical to those of experiment I, except that two versions of each of the non-linear sequences were prepared. This was done in order to obtain a more adequate sampling on RR.

Subjects. Ss were 196 eleventh grade male and female students of Downington High School, Downington, Pennsylvania. Participation was non-voluntary. The experiment took place on January 8, 1970.

Procedure. Ss were randomly assigned to the 28 treatment groups. They were run in a well-lighted, quite room during the first two periods of the day. All treatment combinations were run during both periods. This was accomplished by starting the MR groups successively. It was felt that this procedure offered less opportunity for bias than the one used in experiment I.

All other procedural details were identical with experiment I.

Results

A preliminary 3 x 4 x 2 ANOVA of the non-linear sequences, in which there were three levels of RR, four levels of MR, and two groups of sequences (A and B) indicated no significant effects of any type. Therefore, the A and B sequences for each MR-RR combination were pooled for the analysis.

Table 2 shows a 4 x 4 unweighted means ANOVA, in which the A and B sequences

Table 2. ANOVA of the total scores, experiment II.

Source	SS	df	MS	F	P
RR	55.608	3	18.526	5.02	<.01
MR	29.915	3	9.971	3.19	n.s.
RR x MR	28.129	9	3.125	<1.00	n.s.
Ss: RR x MR	663.997	180	3.688		

are pooled. There was a significant effect for RR ($F=5.02$, $P<.01$), but no effect for either MR or RR x MR interaction. Figure 4 shows the results.

Figure 4. Results of experiment II: total score.

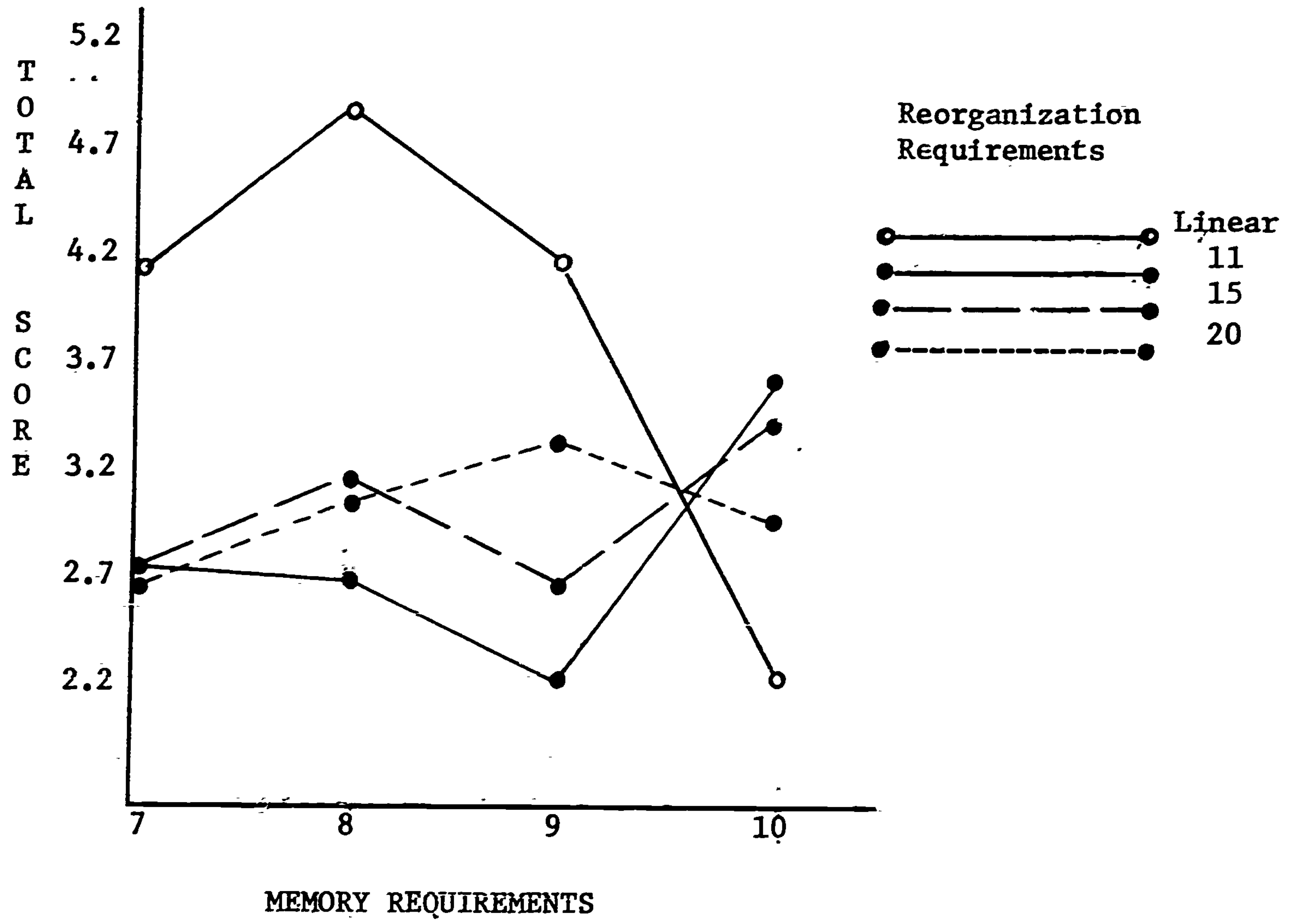


Figure 4 reflects the fact, mentioned previously, that there were no significant differences among the various non-linear treatment conditions due to either MR or RR, and shows the RR effect to be located in the 7, 8, and 9 MR groups.

Analysis of Reversals

In examining the protocols for a hierarchical response pattern, it became obvious that correct responses sometimes occurred after errors, and that these reversals were not distributed at random among the groups. Each protocol was therefore scored for the number of reversals, by the unweighted means analysis described above. Table 3 and figure 5 show the results.

Table 3. ANOVA of reversals: experiment II.

Source	SS	df	MS	F	P
RR	3.287	3	1.095	4.90	<.01
MR	2.570	3	.857	3.90	<.05
MR x RR	1.977	9	.219	.98	n.s.
Ss: MR x RR	37.966	170	.223		

Figure 5. Results of experiment II: reversals.

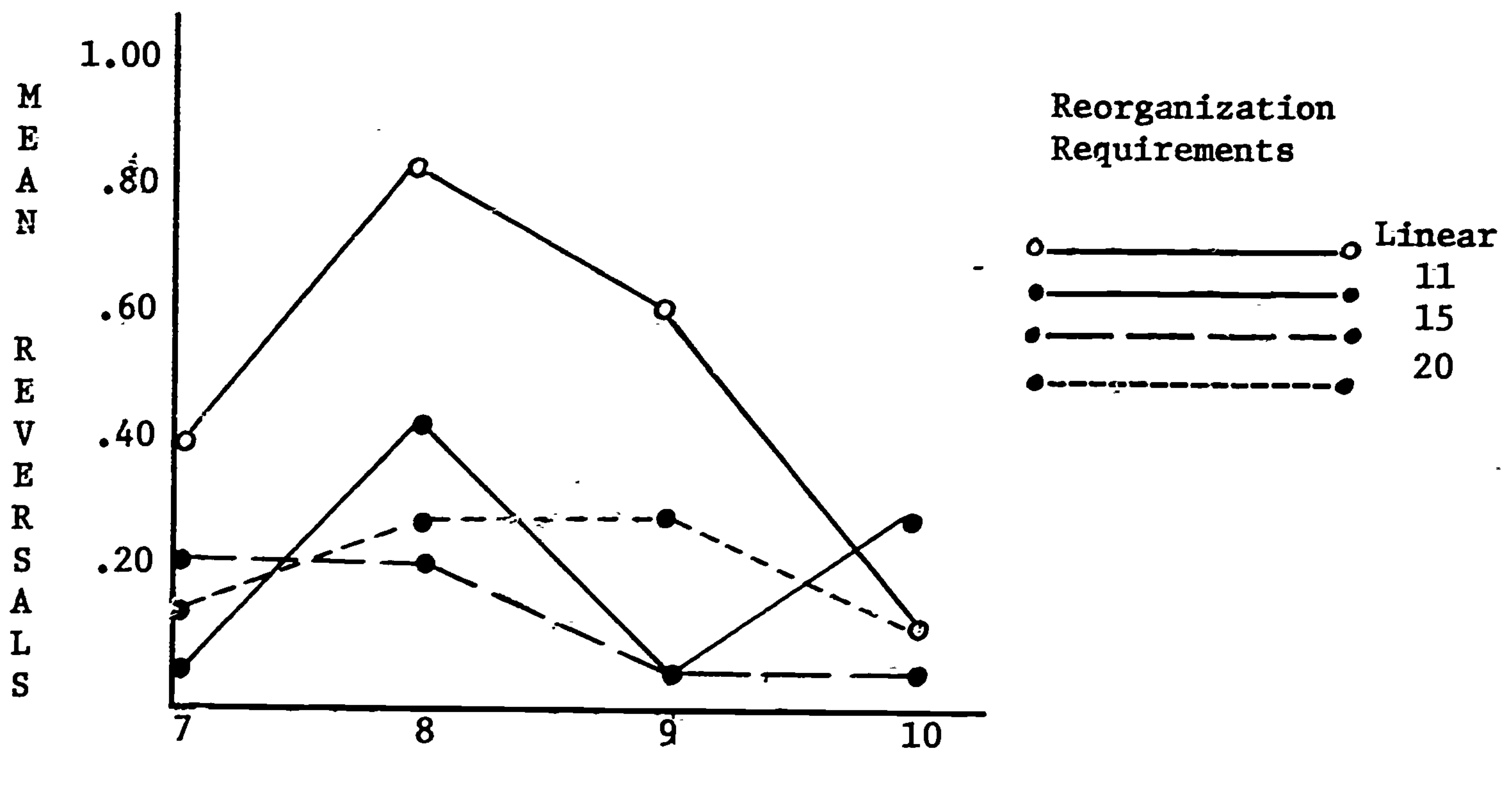


Figure 5 reveals a strong resemblance between the pattern of results on the total score and reversals, which suggests that they should be correlated variables. Since most Ss made zero reversals, this correlation could not be adequately assessed on individuals. Instead, mean reversals were correlated with mean total scores. The resulting value was $r=.80$, $P<.01$.

Re-analysis of Total Scores

It was thought that the total score differences may have been mediated by the reversals, i.e., that total score differences were attributable to points earned after errors had occurred. In order to test this hypothesis, the protocols were re-scored in the following way: each S was given a total score equal to the highest item he performed perfectly on prior to an error. That is, no correct responses were counted after an error.

These corrected scores were then analyzed by the unweighted means procedure. Figure 6 and table 4 give the results of this re-analysis.

Figure 6. Total scores corrected for reversals: experiment II.

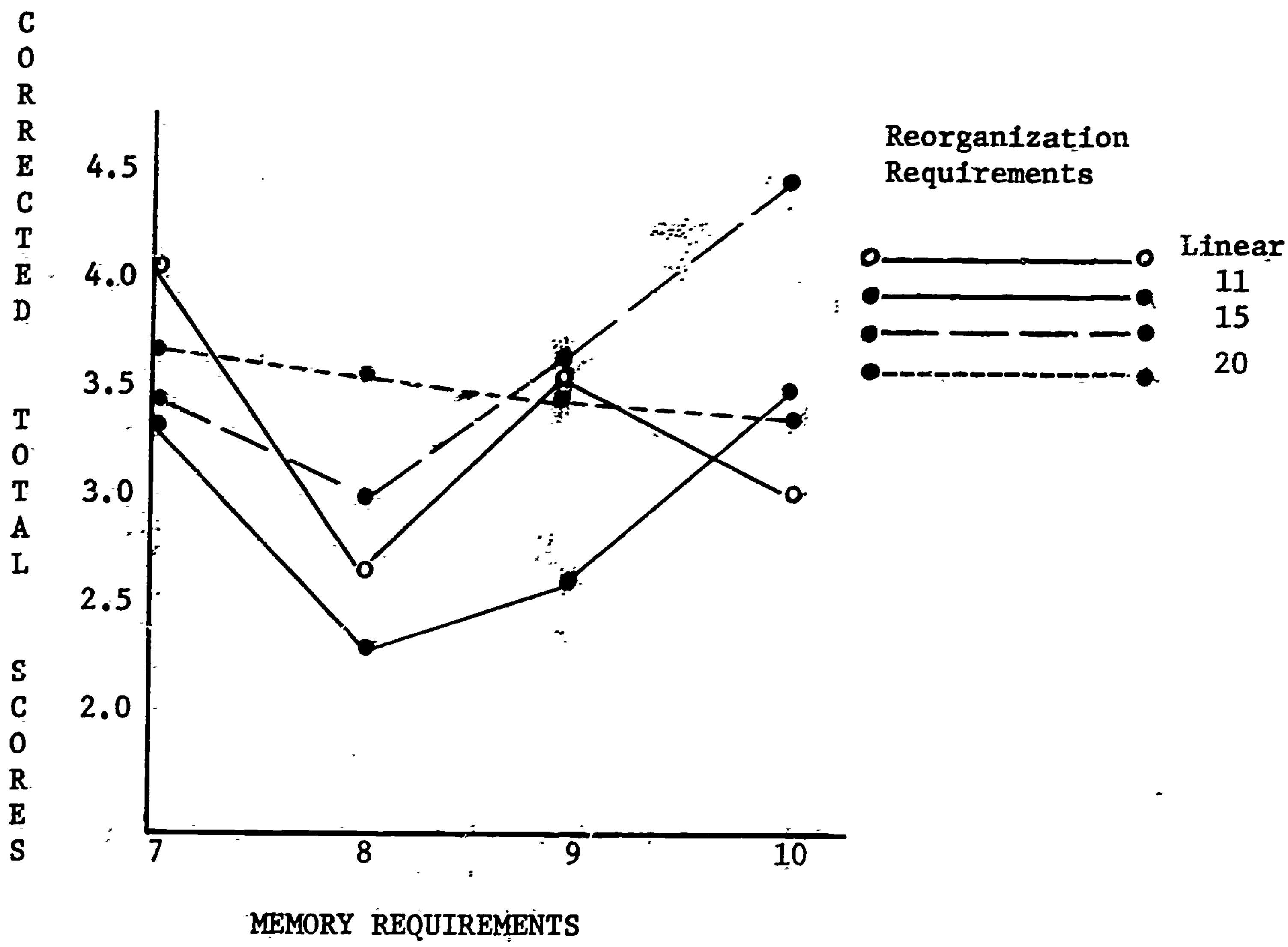


Table 4. ANOVA of total scores corrected for reversals: experiment II.

Source	SS	df	MS	F	P
RR	6.449	3	2.15	.80	n.s.
MR	6.786	3	2.26	1.57	n.s.
RR x MR	12.974	9	1.44	.54	n.s.
Ss: RR x MR	482.143	180	2.68		

Discussion

The results of experiments I and II are consistent regarding the effect of RR on terminal achievement. In both cases RR was found to have a significant effect on the highest level of achievement in the knowledge hierarchy. They are inconsistent regarding the interaction of MR and RR, however; such an effect was found in the experiment I, but not in experiment II.

It is the author's opinion that this discrepancy is almost completely explainable in terms of the nesting of MR x RR combinations within sessions in experiment I, since the MR or MR x RR effects, or both, were the ones confounded by nesting. It is assumed, therefore, that the results of experiment II are the more representative.

Several points concerning experiment II are of interest.

First, RR did have a significant effect (in both experiments). This effect was confined to the contrast between linear and non-linear groups at the 7, 8, and 9 MR levels in experiment II. In light of the accummulated evidence on the ineffectiveness of sequencing, this stands out as an important finding. The interpretation of the difference between this finding and those of previous studies is not entirely clear, since our methods and materials differ in many ways from those used in earlier work. It seems likely, however, that the difference is due in part to the presence of a knowledge hierarchy in the present study. As mentioned earlier, unless there is some evidence of a clear sequential dependency in a knowledge structure, there is no basis for predicting sequence effects. In the present studies there is no question regarding the presence of a hierarchy, and there is complete information regarding its structure.

Second, it seems that, at least for the range of values included in the present study, variations in RR among the non-linear sequences had no effect. All the non-linear sequences performed quite poorly.

Third, the effectiveness of the linear sequence dropped off significantly when the memory load reached ten concepts. (Comparing the first three linear conditions (MR=7, 8, 9) with the fourth (MR=10), we obtained $F=5.74$, $df=1, 26$, $P<.05$). This finding may help explain earlier results. If the memory requirements of the programs used were particularly large, it might be expected that Ss in the linear groups would perform as poorly as Ss in the scrambled group.

Finally, we come to the curious relationship between total score and reversals. In part, this relationship may simply reflect a simple probabilistic phenomenon. It may be, for example, that there is a constant probability of a reversal for each item, and that people who recall more items simply have a higher probability of reversing at one or another point.

The probability explanation may be correct, in part, though it seems unlikely to account for an r of .80. Too, the results of removing correct responses after an error indicate that the reversals play a more fundamental role in the criterion performance than this essentially artifactual argument would suggest. One plausible interpretation seems to fit all of the data, the pattern of total score means, reversal means, and their correlation, so it will be presented conjecturally: It is hypothesized that the results reflect differences in the learning strategies available to, and possibly used by, the linear and non-linear groups. As mentioned above, at each point in the linear sequence the S has been presented with enough information so that it is possible for him to translate the new concept into his natural language. Given that fact, it is also possible, with a certain amount of rehearsal time, for him to store each in a fully expanded English version. In an ideal form of this type of strategy the recall of any item would be completely independent of every other item. Thus, if this were the dominant strategy in the linear groups, it would be expected that a recall or storage failure at some point in the sequence would not necessarily be fatal; the information contained in the missing point could serve its mediating role, then be forgotten, all without hampering later performance. Reversals would be expected here, then, since Ss can theoretically recall later items even after a retrieval failure on earlier ones.

The situation is quite different with the non-linear sequences. Here an S cannot put each item into English as it is presented, since components of several concepts will be undefined. Thus, the independent storage strategy could not be used here in the same way. At best, the S could store several of the definitions in the form in which they are presented, independently. But he cannot use them independently on the post-test, since they depend on one another for meaning. Should he, then, fail to recall any of the definitions, he cannot go further. The effects of a complete recall failure here are, so to speak, irreversible. (The few reversals found in the non-linear conditions could have occurred in cases where Ss only recalled an item after they had failed to answer it correctly; in any case we are not arguing that reversals should never occur there, but only that they should be infrequent.)

This model qualitatively explains the correlation: Ss who store the definitions independently can overcome an error on a single item, and hence get some points after the error; Ss who can only store items in coded form are essentially limited to recalling those items prior to an error, since the error signifies, in effect, the inaccessability of some part of the language necessary to correctly recall later items.

Conclusions

The results point to several tentative conclusions.

First, sequence effects have been unequivocally found in the type of task used in these studies. The salient characteristics of this task are: (1) It embodies a hierarchy of knowledge having a known, unique structure, (2) no repetition was included, and (3) no in-text questions were used. This task was intended to be a prototypical hierarchy, in which standard programming methods could not mask sequence effects.

Second, the sequence effect (in both studies) was dependent on memory load. When the latter exceeded nine concepts, the linear sequence was not superior to the non-linear ones It may be that sequence effects have not occurred in most previous studies because the optimal memory load has been exceeded in the linear sequence.

Third, the superiority of the linear sequences stems from the fact that they allow the learner to forget some intermediate items while still having access to later ones. Intermediate items must be stored temporarily in both linear and non-linear conditions to attain later points, but failure to retrieve the intermediate items is fatal only in the non-linear conditions.

It was hypothesized that this result is due to a difference in strategies available to and used by the linear and non-linear groups. The linear group, due to the order of presentation, could store each item independently while the non-linear group could not. Thus a retrieval failure would be less detrimental in the former group. The sequencing problem, then, may profit from a detailed analysis of the learning strategies available for a given type of task.

The evidence presented here, in the light of the predominantly negative results of previous studies, indicates that studies of sequencing cannot prove very informative unless the materials used are well controlled. It could be argued that the task used in the present study was artificial, and perhaps unrepresentative of programmed instruction. That is doubtless true, but the results do shed some light on the learning mechanisms involved in sequence effects, and the methods can be extended to more complex and representative tasks. The method of scrambling an existing program may be more representative, in one sense, but can lead to uninterpretable results even when effective.

APPENDIX

Table I. Means and standard deviations of total score: experiment I.

		MR			
		7	8	9	10
	L	1.9000	4.0000	2.9000	1.6000
		.9944	2.1081	1.7289	1.2838
	11	2.1000	3.3000	2.9000	2.5000
RR		1.1972	1.7029	2.2335	1.6499
	15	2.3000	3.5000	2.8000	2.4000
		1.1004	1.9003	2.4854	1.3904
	20	1.3000	2.1000	1.0000	1.9000
		.6749	1.7920	1.6465	.8756

Table II. Means and standard deviations of Reversals: experiment I.

		MR			
		7	8	9	10
	L	0	.3000	.2000	0
		0	.4830	.4216	0
	11	0	.1000	.3000	.2000
RR		0	.3162	.4830	.4216
	15	0	.1000	.1000	0
		0	.3162	.3167	0
	20	0	.2000	.1000	.2000
		0	.6325	.3162	.4216

Table III. Means and standard deviations of total scores: experiment II.

RR		MR 7	MR 8	MR 9	MR 10
	L	4.1441	4.8586	4.1441	2.2864
		1.4638	2.7342	2.4494	1.1262
	11	2.7846	2.7132	2.2134	3.3700
		1.7619	1.7228	.8018	2.1881
	15	2.7132	3.1416	2.6418	3.3558
		1.9779	2.1070	1.8634	1.6919
	20	2.6418	2.9988	3.2844	2.9274
		1.9259	2.0000	2.1636	2.0555

Table IV. Means and standard deviations of reversals: experiment II.

RR		MR 7	MR 8	MR 9	MR 10
	L	.4287	.8574	.5716	.1429
		.7071	.8997	.7868	.3779
	11	.0715	.4290	.0715	.2860
		.1506	.6462	.1906	.4688
	15	.2145	.2145	.0715	.0715
		.4688	.5789	.1506	.1506
	20	.1430	.2860	.2860	.1430
		.3631	.4688	.4688	.3631

INSTRUCTIONS

This is a test of how people learn the meanings of a series of words. You are going to be given a series of nonsense words one at a time. Your task is to learn the meaning of each of them. One nonsense word and its definition appear on each page of the booklet. For example:

> Page 1 - quay means book.
> Page 2 - dab means large quay.

Your task is to learn the definition of each word in the booklet and you will be tested after you have completed reading the booklet.

On the test that follows you will be asked to write out the meaning of each of the nonsense words IN ENGLISH. For example:

> Given quay - your answer should be book.
> Given dab - your answer should be large book.

You will have one minute to look at each page. Do not turn the page until you are told to do so. Once you have turned the page do not turn back. Remember you will be tested when you have finished the booklet.

Table V. List of items used in the experiments.

1. Glux - house
2. Baf - red house
3. Dex - 2 red houses
4. Wuf - 2 small red houses
5. Lep - 2 small pretty red houses
6. Nus - 2 small pretty good red houses
7. Zas - 2 small pretty good long red houses
8. Acb - 2 small pretty good long tall red houses
9. Cet - 2 small pretty good long tall square red houses
10. Nah - 2 small pretty good long tall square hard red houses

References

Brown, J. L. Effects of logical and scrambled sequences in mathematical materials on learning with programmed instruction materials. Journal of Educational Psychology, 1970, 61, No. 1, 41-45.

Gagne, R. M. The acquisition of knowledge. Psychological Review, 1962, 69, 355-365.

Gagne, R. M., & Brown, L. T. Some factors in the programing of conceptual learning. Journal of Experimental Psychology, 1961, 62, 313-321.

Levin, G. R., & Baker, B. L. Item scrambling in a self-instructional program. Journal of Educational Psychology, 1963, 54, 138-143.

Niedermeyer, F. C. The relevance of frame sequence in programmed instruction: An addition to the dialogue. AV Communication Review, 1968, 16, 301-317.

Niedermeyer, F. C., Brown, J., & Sulzen, R. Learning and varying sequences of ninth-grade mathematics materials. Journal of Experimental Education, 1969, 37, 61-66.

Payne, D. A., Krathwohl, D. R., & Gordon, J. The effect of sequence on programed instruction. American Educational Research Journal, 1967, 4, 125-132.

Roe, K. V., Case, H. W., & Roe, A. Scrambled versus ordered sequence in auto-instructional programs. Journal of Educational Psychology, 1962, 53, 101-104.

Wodtke, K. H., Brown, B. R., Sands, H. R., and Fredericks, P. Scrambled versus ordered sequencing in computer-assisted instruction. Monograph of the Pennsylvania State Computer-Assisted Instruction Laboratory, 1967.