DOCUMENT RESUME

ED 082 135 CS 000 725

AUTHOR	McMullen, David W.
TITLE	Computer-Assisted Vocabulary Acquisition: A Terminological Approach. Final Report.
INSTITUTION	State Univ. of New York, Stony Brook.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C. Regional Research Program.
BUREAU NO	BR-1-B-076
PUB DATE	1 Apr 73
GRANT	OEG-2-71-0076
NOTE	41p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	College Students; *Computer Assisted Instruction; *Computer Programs; Course Content; Course Evaluation; Cues; Data Bases; Educational Research; Educational Technology; Instructional Innovation; *Program Development; Student Reaction; *Vocabulary Development

ABSTRACT

A terminological approach, constructed around key words in textual material, emphasizes vocabulary in the context of organized frames of reference. This research was directed toward stating a terminological approach in operational terms in order to develop and test procedures for acquiring vocabulary by computer. Three phases are described: (1) the construction of a computer program for comparing two acquisition models, (2) the development of a data base (course content) through the use of the above procedures, and (3) the evaluation of the approach and acquisition models in an experimental study. A blocked-presentation model was compared with spaced presentation, the former supported by concept formation studies and the latter by studies in school learning. Results favoring spaced presentation are discussed in terms of information encoding, and performance gains are described along with suggestions for improving effectiveness. (Author/LL)

FINAL REPORT

Project No. IB076
Grant No. OEG-2-710076

David W. McMullen
State University of New York at Stony Brook
Stony Brook, New York 11790

COMPUTER-ASSISTED VOCABULARY ACQUISITION: A TERMINOLOGICAL APPROACH

April, 1973

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

National Center for Educational Research and Development

(Regional Research Program)

ABSTRACT

Vocabulary acquisition is often approached atomistically using general-purpose words. A terminological approach, constructed around key words in textual material, emphasizes vocabulary in the context of organized frames of reference. This research was directed toward stating a terminological approach in operational terms in order to develop and test procedures for acquiring vocabulary by computer.

Three phases are described: (1) The construction of a computer program for comparing two acquisition models, (2) The development of a data base (course content) through the use of the above procedures, and (3) The evaluation of the approach and acquisition models in an experimental study. A blocked-presentation model was compared with spaced presentation, the former supported by concept formation studies and the latter by studies in school learning. Results favoring spaced presentation are discussed in terms of information encoding, and performance gains are described along with suggestions for improving effectiveness.

The project represents a step in the direction of a generative system for computer-assisted instruction. Subsystems are described that are built around procedures like those used in the terminological approach of this project. Generally positive student response suggests that generative C.A.I. should be extended into areas beyond vocabulary acquisition.

Final Report

Project No. IB076
Grant No. OEG-2-710076

Computer-Assisted Vocabulary Acquisition: A Terminological Approach

David W. McMullen

State University of New York
at Stony Brook

Stony Brook, New York

April 1, 1973

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

I. Background

While remedial reading programs generally recognize the importance of acquiring vocabulary, their methods often reflect an atomistic approach to the problem. Vocabulary items are regarded as discrete entities with little relation to one another, and they are presented as an aggregate of building blocks rather than as elements in a structure of thought.

The major objective of this research has been to use computer technology in teaching vocabulary items as terms. The assumption has been that terminology related to relevant content fields has more utility for students at the college level than general purpose vocabulary, which has only incidental relation to organized frames of reference. With the volume of reading required of undergraduates, their need is to concentrate on a "word in place," Adler's definition of a term (1957), and to see how it functions within a configuration of other terms to articulate a conceptual framework.

Acquisition models. With this guiding assumption, the research has dealt with both theoretical and practical problems from the standpoint of developing procedures for teaching terminology by computer. The theoretical problem was to compare two models upon which to base optimization strategies for vocabulary acquisition. While one model conceptualizes acquisition as a serial or single-channel process, in which one term is mastered at a time, the other model regards acquisition as a parallel-processing activity that operates concurrently along several channels, each engaged in bringing one or more terms to mastery.

Serial mastery is similar to concept formation tasks in which positive and negative instances are blocked by concept, i.e. instances of one concept are not interspersed among instances of another. Evidence has favored unmixed over mixed presentation (Kurtz & Hovland, 1963). Though parallel mastery is mixed presentation, it also can be regarded as a form of distributed or spaced practice, which has been found superior to massed practice in the learning of verbal concepts (Reynolds & Glaser, 1964). Since terms are verbal concepts, a parallel mastery model using spaced practice may provide a better theoretical base for vocabulary acquisition than a serial mastery model using massed practice.

Procedural problems. One of the practical problems was to develop a response-sensitive computer program by which to organize and present a set of terms. The program, or course model, required an instructional data base that could be conveniently accessed by an IBM 1500 system and an instructional decision procedure for guiding vocabulary acquisition. In addition, a generator was necessary to produce execution code for the 1500, and routines were needed to analyze performance data. Another practical problem was to define a terminological approach in operational terms. In order to make the approach available for general use, procedures were needed both to specify a set of words that qualify as terms and to develop around them a data base upon which the computer program could operate. While the acquisition models guided development of the decision procedure, an operational statement of a terminological approach was needed to guide construction of the data base.

The statement that emerged in the study defined the approach as having two stages: selection and amplification, corresponding to two basic principles, context and clarification. A terminological approach first considers a context in order to make a selection of terms. Atomistic vocabulary-building usually draws items from standard frequency counts (e.g. Lorge-Thorndike, 1944), but selecting terms depends more on context than on frequency. Whether a word is critical in illuminating a context may be unrelated to its frequency in general vocabulary, since the context is likely to confer upon the word a special significance.

Similarly, the amplification of a term into a series of student-program transactions, or prepared "script," requires a different principle from that which guides atomistic vocabulary-building. General-purpose dictionary definitions or synonyms rarely suffice to communicate the distinctive role of a term in its context. Clarification is needed by coming at the term from several angles and helping the learner to converge on the precise sense of the term.

Defining a terminological approach operationally, then, is a matter of describing two kinds of procedures: one for selecting terms and another for amplifying them into a structured data base suitable for a series of clarifying transactions.

In summary, three major problems were identified in this research: constructing a data base, constructing a course model (program), and comparing two theoretical models of vocabulary acquisitions. The objectives were to

develop a set of generalized procedures for teaching terminology, using a computer as the instructional vehicle; to generate a specific course by means of these procedures; and to use this course as an experimental setting both to evaluate these procedures and to investigate the acquisition models.

II. Method

Though work proceeded on all three problems concurrently, the project was divided into three phases. The first phase was primarily occupied with constructing a course model, the second with preparing a data base, and the third with experimental evaluation.

Phase I: The course model.

During the first phase, a course model was developed around two considerations: the form and the sequencing of student-program transactions. The form of the transaction evolved over several months of pilot experimentation involving two university courses. One, offered in the summer of 1971, supplemented instruction with a series of quizzes over assigned readings (McMullen, 1972). The transaction was similar to a test question, but with immediate knowledge of results that included both the correct answer and a cumulative score. Questions were either true-false or matching, but in either case the student's task was to link up a term with relevant cues while discarding irrelevant cues. Each series of transactions, or quiz, contained 20 items - four items per term - and required at least 80% mastery.

The second course, in the fall of 1971, used a similar model but added features to make it more informative to the student. Each transaction included a feedback comment to help distinguish between relevant and irrelevant cues. Since the additional feedback placed more emphasis on instruction than on testing, the sequence of transactions was called an exercise rather than a quiz. At the end of each exercise an attempt was made to diagnose which terms had been mastered on the basis of number of errors per term, and students were told which terms required more study.

The sequencing of transactions. The model became more response-sensitive as the second course progressed. Because each term was associated with eight items (see Phase II for a description of the data base), the decision about which item to present was made in two steps. First the particular term was chosen and then the item within the term. At the beginning of the course, the model employed a stratified random sampling procedure (without replacement), drawing a number to select a term and another number to select an item. For each exercise, a set of five terms was chosen, with four items per term. Presentation occurred in two modes:

blocked and spaced. In blocked mode, the four items for each term appeared successively (serial presentation), while in spaced mode the program cycled through all five terms one item at a time (round-robin or parallel presentation).

A later version included items from all eight terms but made the number contingent on student performance. If two consecutive items for a term were correct, that term was dropped from consideration during the exercise. The program therefore gradually converged on those terms most in need of practice. Again the presentation occurred in either blocked or spaced modes.

The vocabulary program for this project used the response-sensitive model just described but with a carefully structured data base.

Phase II: The data base.

Since a primary objective of the project was to make a terminological approach operational, procedures for developing the data base or informational content of the course required major attention. A system was needed that guided the selection and amplification of terms, just as frequency counts and dictionaries guide atomistic vocabulary-building programs.

The Selection of terms. Because of the computer system's design, a decision was made to fix the number of terms per exercise at eight. The primary storage locations, called counters, were limited both in number and in length. Efficiency and flexibility in other programming had been achieved by restricting a set of items to eight. In addition to being compatible with the system, this number conforms with limits on information-processing suggested by several lines of evidence, as summarized by Miller (1956). Eight terms may be about as many as students can efficiently process in one exercise.

Terms require a context, and if the context is common to all terms they will be related. In concentrating on a single overall context or frame of reference, the project constructed a term topology that showed each term "in place." To avoid an arbitrary arrangement of terms into a hierarchy or network unique to its creator, the topology was textually based. The context, in other words, was defined as a text and the topology sought to draw from the text those words most critical to appropriating (i.e. reading and comprehending) this overall context. A topology of this kind can also be defined more operationally than one that is built by specifying logical relations, such as a taxonomy or outline.

The operations involved in selecting terms were the following:

1. Select the textual material
2. Segment the material into units
3. Within each unit identify one or more key terms, depending on the number of exercises planned.

The material selected was a semi-programmed text, Conditioning and instrumental learning by W.T. Smith and J.W. Moore. The book deals with terminology relevant both to education students, for whom the vocabulary program was planned, and to students of the social sciences generally. The basic units of this text are chapter divisions, each containing several paragraphs followed by programmed frames for review.

In each unit at least one key word receives special emphasis by appearing either in a heading, in italics, or in a programmed frame as a word omitted in the frame but shown in the margin. Other key terms appear but their importance is implicit, not explicit, and they tend to be general-purpose vocabulary. Both kinds of terms were included in the topology by restating operation 3 as follows:

3. Within each unit select one term that receives explicit emphasis and one that is emphasized only implicitly.

With two terms per unit, four units provided a set of terms for one exercise. The first 16 units in the text concluded the discussion of classical conditioning and supplied enough terms for four exercises, which were sufficient for the project's objectives. The term topology for the course is shown in Appendix A.

The amplification of terms. Once identified, terms passed through a series of operations designed to clarify the particular sense in which the term is used in the text. Together with the selection operations, the following steps made up the data base generation procedure:

4. Assign half of the terms in the exercise for use in a true-false task, the rest for use in a matching task.

5. Develop a set of cues for each true-false term as follows:

   a. restate a defining attribute mentioned in the text.

   b. restate a defining example mentioned in the text.

   c. state an attribute and an example from some other source (e.g. another psychology text).

   d. restate a non-attribute mentioned in the text, i.e. some characteristic excluded from the sense of the term.

   e. restate a non-example from the text, i.e. an example of something different though related.

   f. state a non-attribute and a non-example from some other source.

6. Develop a set of cues for each matching-task term by carrying out steps a-c of step 5. (Distractors provide non-attributes and non-examples.)

7. For each cue, add a brief feedback comment to indicate why the cue is either correctly or incorrectly related to the term.

Assignment to item types was made by alternating between explicit and implicit terms across cells. Thus, true-false terms for the first exercise were "maturation," "rubric," "habituation," and "increment," and the remaining terms in the first four cells were assigned to a matching task (see Appendix A). In this way, half of the terms for each item type were explicit and half implicit.

The structure of the data base for one exercise is shown in Figure 1. The eight terms are assigned to six channels, two terms to each of the channels containing cues for the matching task and one term to each of the true-false channels. Within each channel are two groups of cues, one based on the text ("old") and one based on other sources ("new"). Each group contains two types of cues: statements of attributes or features and statements of concrete examples. Every positive cue is accompanied by a negative cue that does not apply to the same term. Negative cues are related to competing terms, or foils, that differ in some important respect from the target terms and test comprehension of critical distinctions. In the matching task, each of the distractors

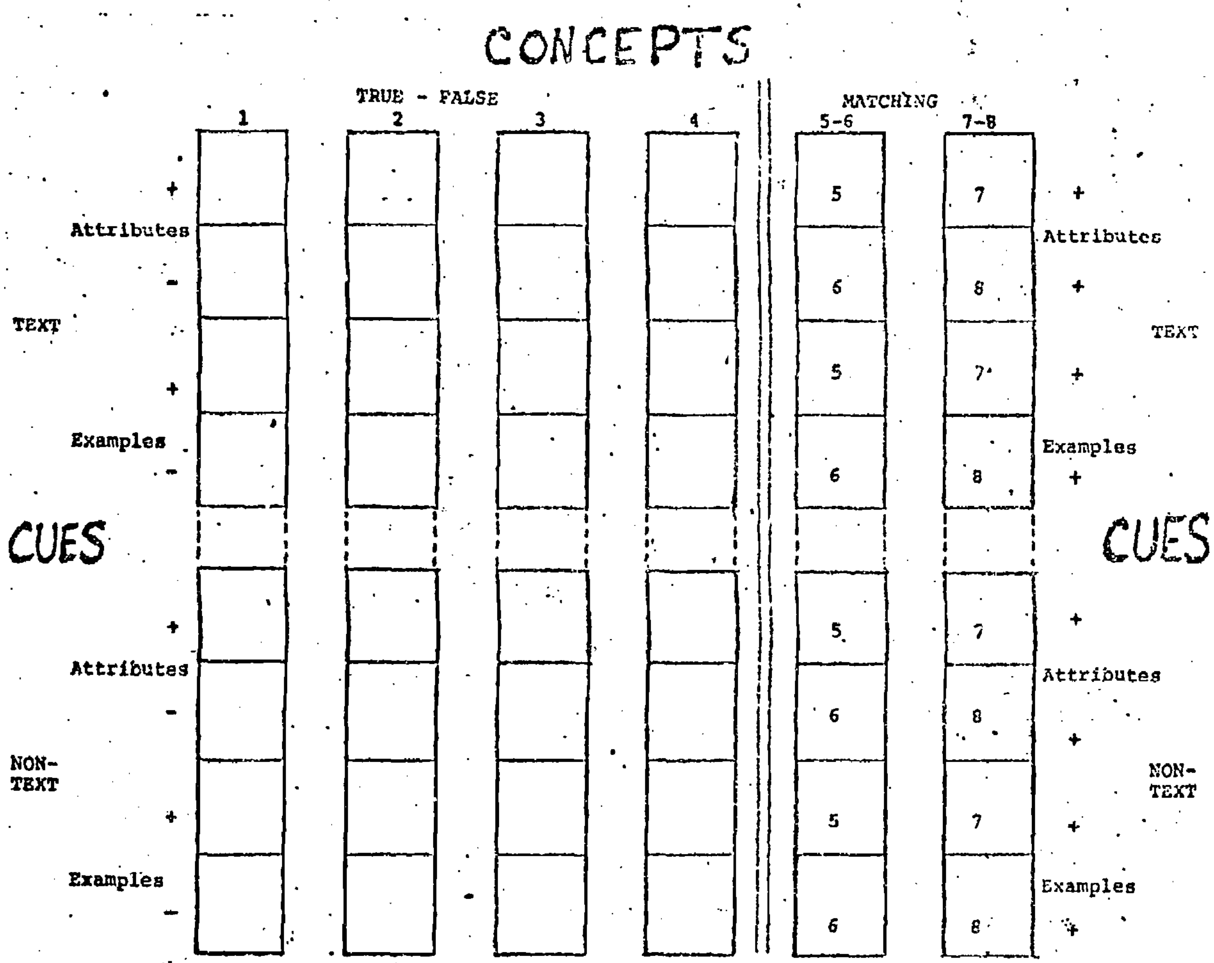

Figure 1. Exercise data base, showing cue types for eight terms.

acts as a foil; cues for one term are negative cues for the others. In true-false items, however, competing terms are not made explicit in the question, and additional cue-statements are required. A sample script for both item types is shown in Appendix B. A script showing the complete item bank for each exercise was printed at the same time execution code was generated by the IBM 370-155.

The structure of the data base reflects the close connection between teaching vocabulary terms and teaching concepts. A term can be viewed as the name of a concept, in which case the principal task of the learner is to classify linguistic experiences into categories called terms. It is as important to know what to exclude as what to include, since a discriminating use of a term neither under-generalizes nor over-generalizes (Markle and Tiemann, 1970). Non-instances, or negative cues, demarcate boundaries between related terms.

At the same time it is important to have both abstract and concrete representations of a term, just as concepts require both instantiation or denotation and a classifying rule covering a variety of instances in order to make clear how the concept is to be applied. Carroll (1964) has pointed out the need for both definitions and examples in the teaching of verbal concepts. Johnson and Stratton (1966) found that a mixed program led to better performance than either definitions or examples presented separately. The presence of attributes and accompanying examples in the data base of this project was intended to promote both deductive and inductive thinking during vocabulary acquisition.

Phase III. Experimental Evaluation

Subjects. To examine both the effect of the program on vocabulary acquisition and the relative effectiveness of the two acquisition models, the four exercises were presented to 50 undergraduates and 11 continuing education graduate students. The undergraduates participated in the experiment to fulfill a requirement to do a project for a course in social and psychological foundations of education. The graduate students took part during class time as an assignment in a course entitled "Contemporary Research on Teaching." All subjects were naive concerning the treatment variable except 16 undergraduates who participated late in the course. Their data is included because of the within-subjects design that enabled them to be their own control

Design. A pilot study indicated that no more than two exercises should be presented at a single session because of

fatigue produced by the intense concentration required. The need for two sessions provided an opportunity to compare the program's effectiveness in two ways: (1) by including terms from both sessions on a post-test following session one as well as on a pretest before session two, and (2) by allowing access to the text only between the first and final posttests. The effect of the program on immediate and delayed retention was examined by comparing first posttest and second pretest differences between terms studied the first week and those not studied until session two. Delayed access to the text provided some indication of the relative effectiveness of program and text by comparing gain during session one with gain between sessions, though no attempt was made to provide rigorous control for this comparison since no major objective was involved.

Primary attention was given to comparing the two acquisition models. One exercise during each session followed the blocked model and the other the spaced model. Subjects were randomly assigned to one of four groups, two of which took exercises one and two the first week and exercises three and four the second. The other two groups began with exercise three and ended with exercise two. One group in each of these two sets took the first and last exercises in blocked format and the middle two in spaced format, while the other had spaced before blocked. The design, then, was within-subjects with counterbalancing of both exercise content and type of model (blocked or spaced).

<u>The instructional sequence</u>. *S* participated in two sessions one week apart. Each session began with a pretest (T_1 and T_3) of 24 items, six from each exercise.(1) One item was randomly selected from each channel shown in Figure 1 so that six of the eight terms in each exercise were represented on the test. No feedback was provided except at the end when the possible score and *S*'s score were displayed. Posttests (T_2 and T_4) occurred at the conclusion of each session. Tests were identical in form but different in content for the most part since items were selected at random from the exercise channels.

Between T_1 and T_2 and again between T_3 and T_4 a teaching program occurred in the form of two exercises each containing at least 16 items followed by three kinds of feedback: the correct response (e.g. "true"), an updated score, and a brief sentence elaborating on why the response was correct. The score was calculated by adding ten points if the response was correct. If it was incorrect, the score remained the same unless *S* had previously received ten points for an item belonging to the same term. In that case, ten points were subtracted because the exercise criterion

required two consecutive correct responses for each term, or 160 as a final score. S was permitted 40 items in which to reach criterion.

At the end of the first session, S was handed a book and encouraged to read the first 50 pages before the next session for two reasons: to receive additional help in learning terminology and to observe an example of programmed instruction. The book was collected before S began T_4.

Differences between test and exercise decision procedures are shown in Figure 2. The test made six cycles through the four exercises selecting a different term at random on each cycle and one item at random per term. The order of exercises was rerandomized for each cycle. Each exercise made a random selection at only two levels: term and item. In blocked format the order of terms was random but selection remained at the item level until two consecutively correct responses occurred. In spaced format, both term and item were selected at random for each presentation, with a term removed from the round-robin whenever two of its items were answered correctly in a row.

A schematic representation of the overall system used in generating and teaching terms is shown in Figure 3. The process began with a context, defined as the source material or textual discussion in which the terms occur. The seven steps in Phase II were applied to the text to produce an instructional data base consisting of four exercises, each with the structure shown in Figure 1. The decision procedures, developed in Phase I and shown in Figure 2, operated upon the data base to produce the instructional sequence of Phase III. Data recorded on magnetic tape were then analyzed during the final phase of the project.

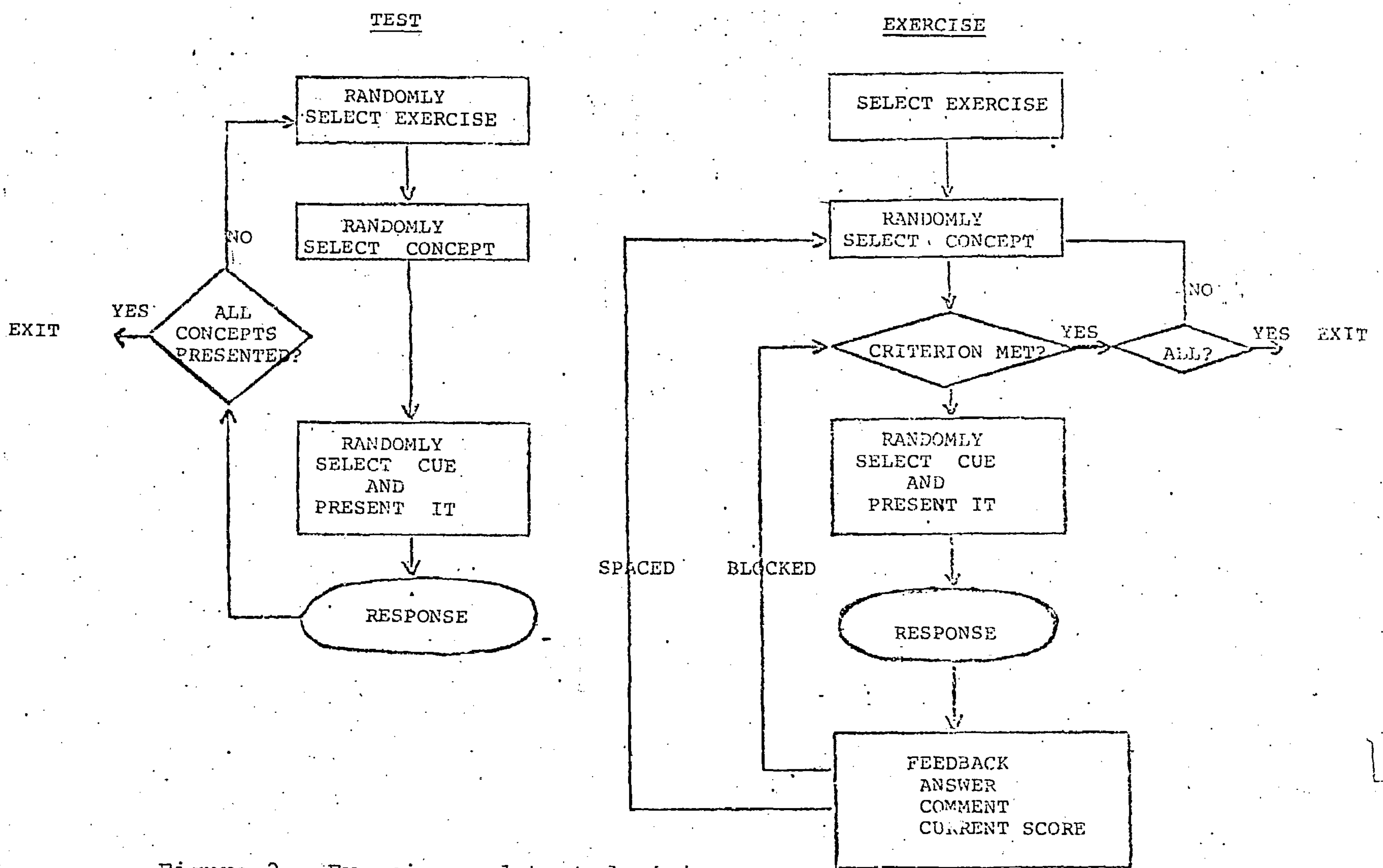

Figure 2. Exercise and test decision procedures, showing similarities and differences between components.

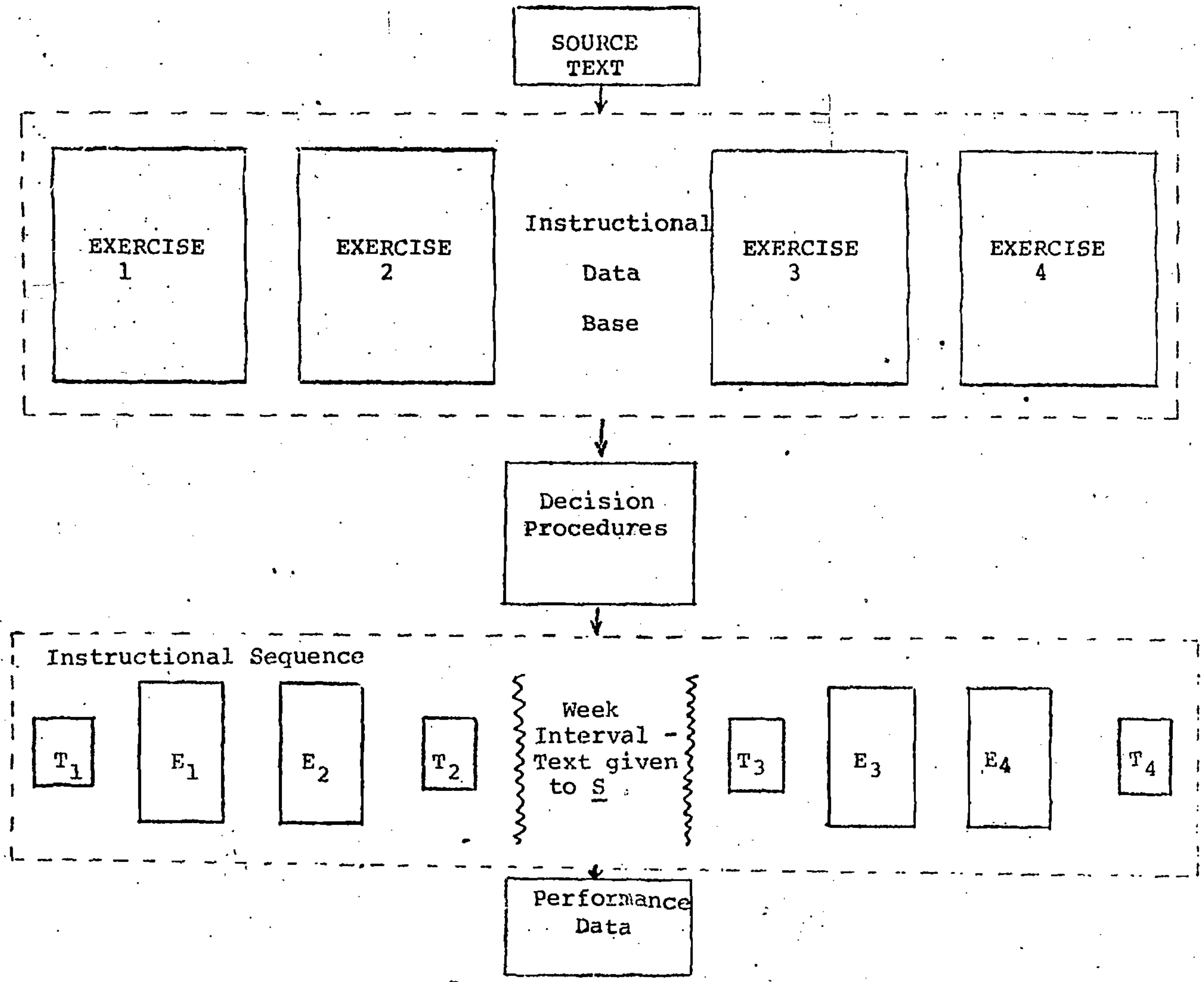

Figure 3. Schematic representation of generative C.A.I. system used in this study.

III. Results

Both formative and summative evaluation occurred during the project, the former primarily in pilot testing of the program during Phase I. Faculty colleagues as well as students monitored development, suggesting modifications, pointing out weaknesses, and acting as pilot subjects by which to study parameters of the model such as reliability, time required, fatigue effects, etc. Summative data were collected in two forms: response recording by the system and questionnaire reactions at the conclusion of the final posttest.

The major questions asked of the data were (1) did performance differ depending on whether terms were blocked or spaced, (2) did the teaching module (exercises) improve performance, and (3) how did students react affectively? Data relevant to the first two questions were proportion correct on the tests, number of items to criterion on the exercises, and time required to meet criterion on the exercises. Each of these three dependent variables was analyzed in terms of two independent variables, mode of sequencing and exposure to instruction, in a 2 x 2 repeated measures analysis of variance design. The first test (T1) was used as a covariate for separate analyses of the other three tests, as well as for the exercise data. The covariate adjusted scores where differences in initial performance existed either between blocked or spaced terms or between terms that were assigned or not assigned to instruction.

Blocked-spaced sequencing. Each of the four tests sampled equally from each of the four exercises. The mean proportion of correct responses for terms from the two blocked exercises is compared with proportion correct for terms from the two spaced exercises in Table 1, showing no significant differences on the four tests until T4. Spaced terms made nearly twice the gain of blocked terms over pretest levels by the time of the final test.

Half of each test consist of terms from session 1 exercises, one in each sequencing mode. The mean proportion correct in this half of the test revealed slight differences immediately after study (T2, Table 2) but an advantage for spacing over blocking on the final test.(2) Table 3 shows similar data for the other half of each test, i.e. mean proportion correct for terms studied in both modes during session 2. Again no differences occurred immediately after study (T4 in this case), but the design did not include a posttest for these terms. Blocked terms gained as much

Table 1

Mean Proportion Correct in Both Treatment Modes on Each Test (N=61)

	Blocked	Spaced	df	F
T_1	.66	.65	1,60	.15
T_2	.71	.73	1,59	1.58
T_3	.74	.72	1,59	.68
T_4	.76	.83	1,59	8.05**

** $p < .01$, ANOVA

Table 2

Mean Proportion Correct in Both Treatment Modes for Terms Studied Only in Session One (between T_1 and T_2)

	Blocked	Spaced	df	t
T_1	.67	.66	60	.17
T_2	.75	.76	60	-.21
T_3	.77	.75	60	.71
T_4	.75	.81	60	-2.16*

* $p < .05$, matched t-test

Table 3

Mean Proportion Correct in Both Treatment Modes for Terms Studied Only in Session Two (between T_3 and T_4)

	Blocked	Spaced	df	t
T_1	.65	.64	60	.40
T_2	.66	.70	60	-1.33
T_3	.72	.70	60	.42
T_4	.78	.84	60	-1.78

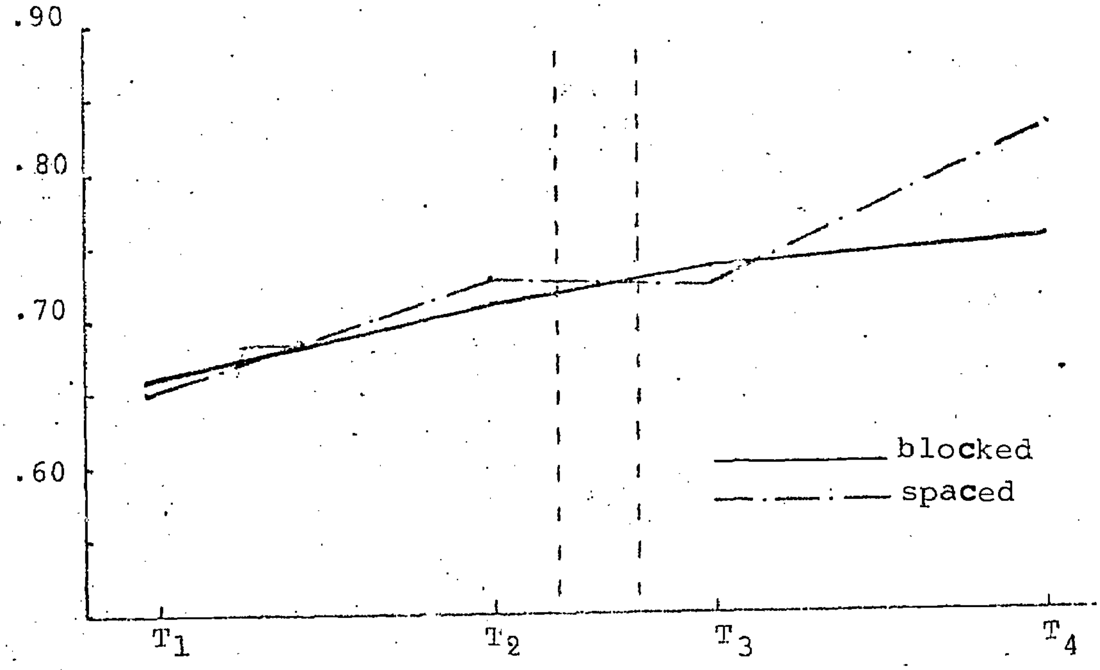

Figure 4. Overall proportion correct on the four tests, each sampling from all four exercises. (N=61.

between sessions as during session 2, while spaced terms gained dramatically during session 2, even making some improvement without study during session 1.

The learning curves in Figures 4, 5, and 6 contrast the steadier gains of blocking with the spurts of spacing after exposure to exercises. The curves suggest that an interaction occurred between mode of sequencing and the position of the tests. Tests that followed exercises (T2 and T4) tend to favor spacing, while any advantage on tests before exercises is in the direction of blocking. Another ANOVA was performed over all four tests with a third factor, type of test within session (before or after exercises). This analysis disclosed that a significant interaction did in fact exist between mode of sequencing and type of test ($F=7.46$, $df-1,60$, $p < .d$).

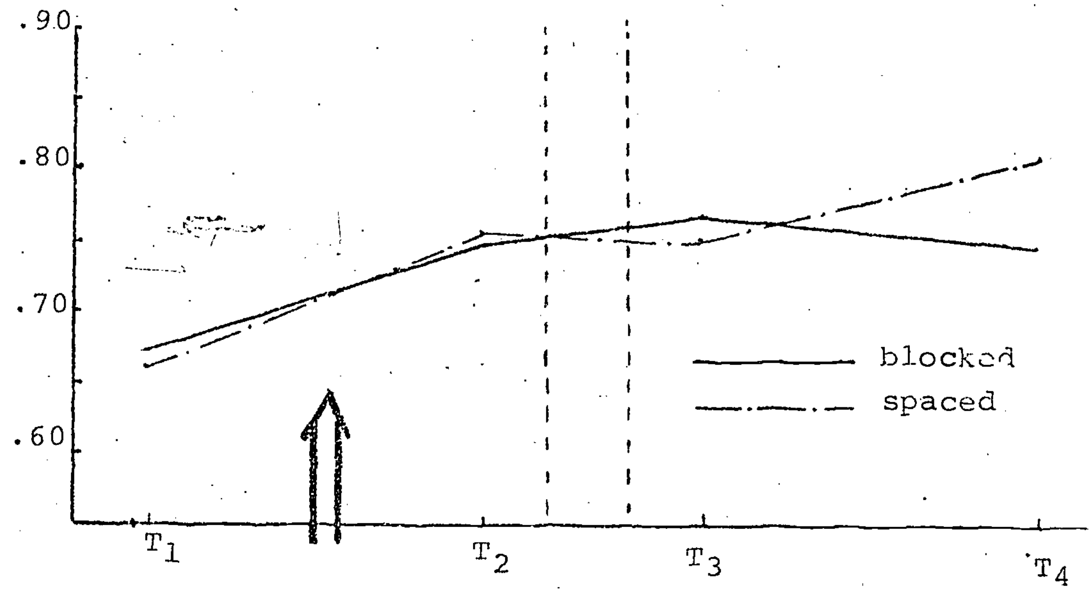

Figure 5. Proportion correct on the four tests for terms in first two exercises. (arrow)

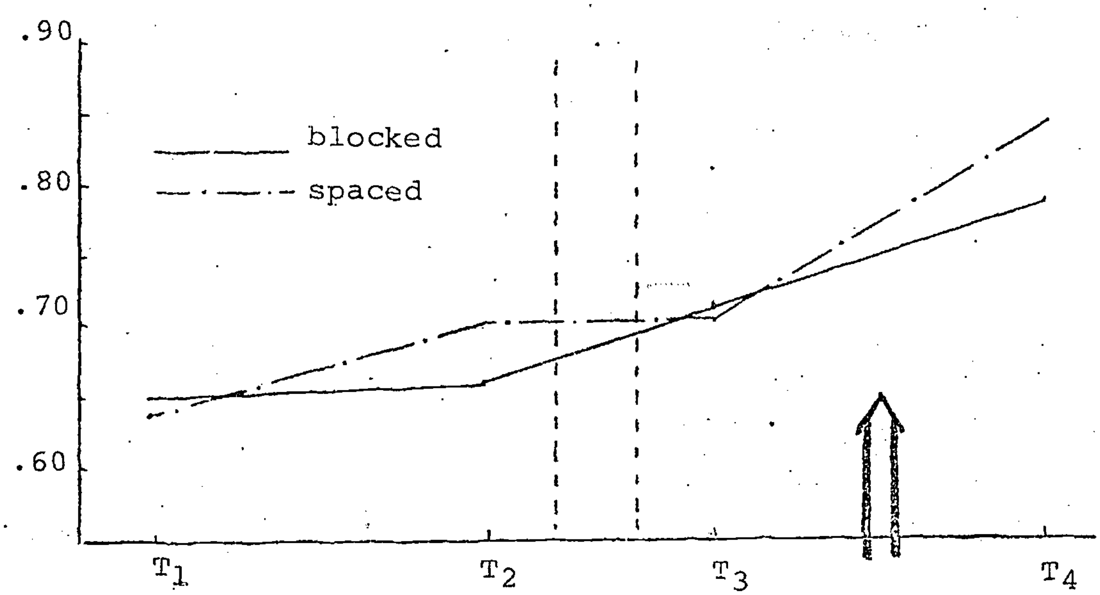

Figure 6. Proportion correct on the four tests for terms in last two exercises. (arrow)

Evidence from the exercises themselves did not disclose any differences between sequencing modes during instruction, though the direction favored spacing. The number of items required in spaced exercises during session 2 was 1.4 fewer than in blocked exercises, but not enough fewer to make overall differences significant (Table 4). The number of minutes to criterion in each mode was nearly identical (Table 4).

Performance gains. While the final test provides the critical data for comparing sequencing modes, data for assessing the effectiveness of the exercises comes from the last three tests. The immediate effect of session 1 exercises appears on T2 while T3 and T4 reveal two kinds of delayed effects: before and after exposure to other exercises. Thus Figure 5 is of primary interest, since it presents the learning curve for terms studied in session 1. Figure 6 is also of interest for purposes of comparison and for examining immediate retention effects for terms studied in session 2. Mean gains over pretest levels are presented in Table 5 for each mode and for combined terms in each session.

Terms studied in session 2 were not studied in session 1, i.e. they did not appear in exercises. A comparison of studied and non-studied terms on T2 reveals that the former made nearly twice the gain of the latter in proportion correct. After exposure to the textbook between T2 and T3, session 1 terms remained at about the same level but session 2 terms gained, perhaps because of measurement error in the case of blocked terms. (Other terms showed little or no change over T2.) The gain for session 2 terms on T4 over T3 parallels the gain for session 1 terms on T2 over T1. In both cases the gain is about ten points. The parallel is even stronger when non-studied terms are compared, i.e. session 2 terms at T2 and session 1 terms at T4. In both cases, spaced terms far surpass blocked terms, gaining about six points over the previous test.

The overall pattern, blocked and spaced terms combined, is a net gain for studied terms of about five or six points (studied minus non-studied), with little effect occurring because of exposure to the textbook, the retention interval, or exposure to other exercises. The pattern for spaced terms, however, suggests that exposure to other exercises had a facilitating effect on retention.

Exposure to the textbook may have had little effect because little time was spent studying it between sessions. On the questionnaire, however, Ss reported having spent an average of 34 minutes reading the book. Questionnaires were keyed to student numbers and false reports may have been given despite assurances that responses were confidential and in no way to be communicated to the instructor before grades were posted.

Table 4

Mean Number of Items and Minutes to Criterion in Each Session and Overall
(N=61)

	Blocked	Spaced	df	t	F
Items					
Session 1	26.9	26.4	60	.49	
Session 2	24.8	23.4	60	1.59	
Overall	25.9	24.9	1,59		1.94
Minutes					
Session 1	21.8	20.7	60	.41	
Session 2	16.5	15.4	60	.52	
Overall	19.2	18.0	1,59		.38

Table 5

Mean Gains in Performance over Pretest (T_1) in Proportion Correct

	Terms Studied in Session 1			Terms Studied in Session 2		
	Blocked	Spaced	Combined	Blocked	Spaced	Combined
T_2	.08	.10	.09	.01	.06	.04
T_3	.10	.09	.10	.07	.06	.07
T_4	.08	.12	.12	.13	.20	.17

Affective reactions. At the end of the experiment, *S* filled out a questionnaire consisting of ten Likert-type items and a request for the time estimate referred to above. Each of the ten items used a 5-point scale ranging from "strongly disagree" to "strongly agree." Mean values are reported in Table 6 for all but five *S*s, who were inadvertently overlooked. Two of the items were related to the sequencing variable to determine whether the preferred mode was also the mode associated with higher performance. Item 8 shows a tendency to favor blocked presentation ("one word at a time") and item 9 is consistent, indicating a reluctance to endorse spaced presentation ("switching from one word to another"). A matched t-test revealed a significant difference in response to the two items ($t=2.05$, $df=54$, $p<.05$), and the correlation between the two items was - .78. Higher performance, then, occurred in the least preferred mode (spacing). Table 7 also indicates that performance on the final test had little relationship to preference as expressed in these items. A modest but non-significant negative correlation (-.21) occurred between preference for blocking and proportion correct for blocked terms.

With the exception of item 7, which showed that true-false questions were considered harder than matching questions, the remaining items indicate global reactions. The most favorable aspect of the experience appears from item 1 to have been its novelty or lack of boredom. Item 3, next highest, agrees with the data that show performance gains; improvement was perceived as well as demonstrated. In general, the global items indicate that the vocabulary course was challenging, worthwhile, non-aversive, and supportive of learning. The degree of individualization (item 5) was limited, however, but not because the approach was too mechanical (item 6).

Additional reactions appeared in term papers submitted by the undergraduates about their experience. Comments ranged from extreme enthusiasm to moderate skepticism. An example of the former is given below.

> My feelings concerning C.A.I. are all positive. I can see it as a great potential godsend for mankind. It could be likened to the famed Seven League Boots....

This student was especially impressed by the reduced fear of answering incorrectly that allows the student to concentrate on the learning task itself rather than on grades and to reveal what he knows to the teacher. As a result,

> The teacher may now gauge the student's progress, seeing where he goes wrong and being able to pick up any weak spots...Best of all this could be done immediately.

Table 6

Mean Ratings and Standard Deviations on Questionnaire (Five-point scale: 1=strongly disagree; 3=neutral; 5=strongly agree; N-56)

$\overline{X}$			s.d.
4.16	1.	I found this approach to vocabulary study interesting.	.82
3.64	2.	I felt challenged to do my best work.	1.04
3.86	3.	The study helped to improve my vocabulary.	.90
2.68	4.	I found myself just trying to get through the material rather than trying to learn.	1.24
3.11	5.	There appeared to be efforts to suit the presentation of the material specifically to me.	1.20
2.60	6.	The whole approach to teaching vocabulary is too mechanical.	1.09
2.78	7.	True-false questions were easier than matching questions.	1.06
3.38	8.	I preferred working on only one word at a time.	1.00
2.87	9.	I was helped by switching from one word to another.	.94
3.40	10.	I would recommend this "short course" in vocabulary, even if my friend did not <u>have</u> to take it.	1.22

Table 7

Correlations between Questionnaire Items and Proportion Correct on Spaced and Blocked Terms at T_4

	Item 8 (Blocking Preference)	Item 9 (Spacing Preference)	Spaced Proportion Correct
Item 9	-.78		
Spaced P.C.	-.04	.04	
Blocked P.C.	-.21	.14	.23

Another student, "determined to meet the challenge of the machine," took extra time to inspect the feedback and to recall previous items.

> I was glad for the repeated drilling that the machine would do because I felt that I was getting a chance to learn the material.

One student voiced skepticism if such a program were used with small children because the "test format" would get boring. Stronger and more reinforcing stimuli would be needed.

> But if a teaching machine like this one were changed to show pictures, stories, a child would get more out of it.

On the other hand, students noted that the machine was patient, did not get bored like teachers, and made the task into a game by offering points.

Individualization and interaction were frequent themes in student comments. The self-pacing and diagnostic features were popular, and students liked the de-emphasis of competition. They also liked the immediate reinforcement and active involvement that the program provided.

While several students reported that some portions of the program were overcued, so that responses were too easily evident, they also reported that learning had transferred to other contexts. The word "rubric," for example, appeared in two other social science courses between the experiment and one student's paper. Before the experiment the word was unfamiliar to the student.

> Both instructors used this term and as soon as they said it I knew exactly what it meant. This is certainly a very good example of the tremendous effect which this particular machine and its content had on me.

Another student suggested that such a program would be a useful prerequisite for certain courses because of the familiarization training it provided.

One student summarized the program in the following way.

> I found this experiment with a computer to be most rewarding. I found the computer to be easier to work with than the book. I enjoyed competing with myself rather than with other

> individuals. I enjoyed working at my own pace not feeling pressured by anyone else progressing faster than I. In fact I don't think I even noticed anyone else in the room with me. The computer held my attention, although I don't know whether it was due to the novelty....Nevertheless I did learn from the computer, more in fact than any other two hour session I ever spent with conventional study methods....I truly enjoyed the sessions (it felt good to learn something without first being bored to death by either a bad lecturer or a dry book) and wish more classes were given with computers. I would surely enroll.

The student points out several ways in which the course was judged to be a viable alternative to conventional modes of instruction.

IV. Conclusions

The central objective of this research was to work out procedures for a terminological approach to teaching vocabulary by computer. In the process of carrying out the three phases described in Section II, this study formulated a structure in which to consider not only vocabulary instruction by computer, but other kinds of instruction as well, such as the teaching of concepts (McMullen, 1973). The structure may be described as a form of generative computer-assisted instruction. It is based on a distinction between data base and decision rule, and it is recursively organized into four sub-systems: authoring, production, transaction, and evaluation.

The authoring sub-system was discussed under Phase II above. As presented in Figure 7, its input is drawn from the cognitive structure of the initiator (author) of instruction, who applies selection and amplification procedures to the initial knowledge base, or text. The result is a script composed of information elements, e.g. terms, each tied to a cue-net that functions to locate the element in relationship to other key elements in the knowledge base. The script also contains specifications for the second and third sub-systems, i.e. directions for production and transaction.

Production specifications are values for a set of parameters that define options available to the author in transforming script elements into working code.

The production sub-system is built around generator programs that use the computer in a background mode to prepare the enormous amounts of execution code required for most C.A.I. applications. In this project, the production system was not given the prominence of the other sub-systems, but in many respects the others were only the tip of the iceberg. The underlying structure that made it possible to turn a script into performance data was a PL-I program, modified continuously over the project so that the other sub-systems could be implemented.

Phase I encompassed both the production and transaction sub-systems, i.e. the design and development of the course structure. Phase III activated the transaction sub-system, as well as the evaluation process, bringing the project full circle after the pilot run since additional authoring was found to be desirable in correcting and refining the data base. The project demonstrated the need for recursive organization of the generative structure because each sub-

SUB-SYSTEMS

	Authoring	Production	Transaction	Evaluation
INPUT	Output from Author's Cognitive Structure	Author Specifications	Output from Learner's Cognitive Structure	Author Questions
DATA BASE	Text	Script Elements	Coded Elements; History Files	Performance Data
DECISION PROCEDURE	Selection and Amplification Strategies	Generator Program	Study and Test Strategies	Reduction and Analysis Strategies
OUPUT	Script Elements and Specifications	Execution Code	Learner Behavior (Performance Data)	Decisions

Figure 7. Four sub-systems in generative C.A.I. and their major components.

system tended to call others in order to complete its task.

Mode of presentation. The data support the spaced-presentation model, although its advantage over blocked presentation does not become evident until additional concepts have been studied. Learning under a condition in which S must continuously deal with interference appears to aid retention after exposure to interfering material. The situation is not unlike school learning, when mastery of one concept (e.g. mean, median, or mode) also requires the ability to discriminate between concepts on the final examination. Linear study is less effective than cyclic-random study in combating the effects of long-term interference.

The interference explanation, however, suffers from an impreceision about what mechanisms may underlie observed performance. Interference theory is rooted in association learning assumptions such as the notion that associations have "strength" which decays over time or is diminished by competing associations. From the standpoint of structural learning, performance may be a function not of interference but of information. Concepts presented in spaced mode may be tagged with more information in the learner's cognitive structure than blocked concepts. If that structure is a network of nodes and each mode a configuration of cues, the learner may establish a better configuration and a more well-defined network when concepts (nodes) are studied concurrently rather than serially. With other concepts as background foils to provide a bas-relief for the target concept, the configuration and the network are continuously under development.

Besides using contrast, the spaced mode may overload short term memory with information. The consequence may be to require S to develop chunking or organizing rules for dealing with the overload. These rules may correspond to the concepts to which the terms refer. Whether the actual concepts or not, the rules would serve to assist encoding in memory. The questionnaire data, showing a lack of preference for spaced presentation, may buttress this interpretation, since information overload is likely to be regarded as taxing.

Course effectiveness. Both performance and questionnaire data support the general conclusion that the course was effective as instruction. Gains were steady, though not dramatic, and student interest was maintained. However, the level of performance on the final test indicates that instruction can be considerably improved. Four types of changes need to be explored in order to determine how the

overall system may function more effectively. First, the transaction data base should be validated by item analyses and a check for ambiguities. The massive amounts of performance data require a well-developed evaluation subsystem in order to reduce and examine data at the level of detail needed for validation. Such a sub-system was beyond the scope of this project.

The other three types of changes have to do with the transaction decision structure: (1) enriching components, (2) raising the criterion, and (3) altering the sequencing strategy. As shown in Figure 2 the exercise transaction consisted of three major components: presentation, response, and feedback. Each components appeared in the course in its simplest form, except that more than one kind of feedback was supplied. In a more highly developed transaction, the presentation might use audio-visual material or additional prompts at the very least. Response-processing might allow constructed responses and perhaps probing of student input to gain a more accurate assessment of how much is known. Confidence testing techniques, for example, might assist in detecting student weaknesses by encouraging him to reveal his subjective probabilities. Feedback might be made more helpful by permitting the student to view the effects of his decision in some simulated situation, perhaps one that was described in the presentation. He would thus have the sensation of manipulating the term or concept under study.

Changes in the criterion might be either local, global, or both. The local criterion, specifying acceptable performance on a given term might be set to four in a row correct or certain combinations of items correct (e.g. one positive, one negative). The global criterion might specify performance on both exercises and tests as well as requiring that the student meet local criterion for each term.

Most importantly, the sequencing strategy appears to be a promising variable. Both the blocked and spaced sequencing modes can be described as parameter-free because they do not depend on either the properties of the items in the data base or the characteristics of the students. Differences between these modes indicate that sequencing can improve effectiveness but even larger differences may appear with parameter-dependent strategies (Atkinson and Paulson, 1972). Key item parameters may be difficulty level, cue type (e.g. attribute or example), and relationship of cue to term (e.g. positive or negative, relevant or irrelevant). A "scientific rhetoric" is needed that specifies the conditions under which such parameters should be set to certain values (cf. Carroll, 1964).

Dissemination. The four courses which the project has used as an environment for program development and testing have already introduced numbers of students both to useful vocabulary and to the techniques of the project. Professionals have also been acquainted with the project and its results through reports at conferences (McMullen, 1973a; McMullen, 1973b).

What remains is to carry through the changes described above in order to make the program more effective. In addition, other faculty in other courses need to be acquainted with the techniques employed in this project so that more generalized procedures may be developed. A paper is now available to provide faculty and students with an operational description of generative C.A.I. in the context of this project (McMullen, 1973a).

Summary. Though productive in a number of ways, the project merely opened the way to dealing with vocabulary acquisition from a terminological standpoint, just as it only sketched the barest outline of what generative C.A.I. entails. The evolution of each of the four sub-systems resulted in useful experimental results, enthusiastic learning, and a set of tested procedures. What may prove most valuable, however, is the framework this project helped to build in which to model the teaching process. As work progresses in the context of generative C.A.I., the thinking that germinated during this project may contribute to problems far more comprehensive than vocabulary acquisition.

FOOTNOTES

(1) Initially each test consisted of 20 items, 5 per exercise, but the last 38 subjects took tests of 24 items, 6 per exercise, in order to include all true-false terms and to improve reliability.

(2) Matched-tests were used to detect the presence of simple effects and must therefore be interpreted with caution.

REFERENCES

Adler, M.J. How to read a book. New York. Simon and Schuster, 1940.

Atkinson, R.C. and Paulson, J.A. An approach to a psychology of instruction. Psych. Bulletin, Vol. 78, No. 1, 1972.

Carroll, J.B. Words, meanings and concepts. Harvard Ed. Review, Vol. 34, 1964.

Johnson, D.M. and Stratton, R.P. Evaluation of five methods of teaching concepts. Jour. of Ed. Psych., Vol. 57, 1966.

Kurtz, K.H. and Hovland, C.I. Concept learning with differing sequences of instances. Jour. of Exper. Psych., Vol. 51, No. 4, 1956.

Lorge, I. and Thorndike, E.L. The Teacher's word book of 30,000 words. New York: Teachers College, Columbia U., 1944.

Markle, S.M. and Tiemann, P.W. Conceptual learning and instructional design. Jour. of Ed. Technology, Vol. I, No. 1, 1970

McMullen, D.W. A concept-sampling procedure for quiz-oriented instruction. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, February, 1972 (ERIC ED D62 400).

McMullen, D.W. Generating concepts by computer. Paper presented at the Fourth Annual Invitational Inter-disciplinary Meeting on Structural Learning, Philadelphia, April, 1973.

McMullen, D.W. Spaced-blocked presentation and the learning of verbal concepts. Paper to be presented at the Annual Meeting of the Eastern Psychological Association, Washington, D.C., May, 1973.

Miller, G.A. The magical number seven; plus or minus two: Some limits on our capacity for processing information. Psychol. Rev., 63, 1956.

Reynolds, J.H. and Glaser, R. Effects of repetition and spaced review upon retention of a complex learning task. Jour. of Ed. Psych., Vol. 55, 1964

Smith, W.I. and Moore, J.W. Conditioning and instrumental learning. New York: McGraw Hill, 1966.

APPENDIX A.

Term Topology

Each cell is one chapter division (unit) in the text, with explicitly emphasized terms above the dashed line and implicitly important terms below. Each row of cells is one exercise (eight terms).

EXERCISES				
I	maturation	performance	habituation	hierarchy
	transitory	rubric	correlated	increment
II	reinforcer	elicit	classical conditioning	respondent behavior
	empirical	salivary	paradigm	repertoire
III	simultaneous	interstimulus	latency	extinction
	parameter	reflex	dependent	extraneous
IV	generalization	discrimination	unconditioned stimulus (US)	sensitization
	property	trial	primary stimulus	artifact

APPENDIX B

Sample Scripts For Both
Item-Types in the Data Base

(a) true-false term: maturation

Maturation...	(heading)
is unlearned, biological development.	(text-attribute-positive)
It is a growth process and cannot be learned.	(feedback comment)
produces temporary physiological changes in behavior	(text-attribute-negative)
Development implies a relatively permanent change.	(feedback)
enables a child to take his first step:	(text-example-positive)
"Readiness" is primarily responsible, i.e., muscular development.	(feedback)
enables a puppy to obey the command, "Stay."	(text-example-negative
Obeying an order is learned behavior.	(feedback)
is growth arising from internal genetic endowment.	(non-text-attribute-positive
Maturation "occurs;" it is gradual biological alteration.	(feedback)
is growth dependent upon experience or training.	(non-text-attribute-negativ
Learning is the product of experience, not maturation.	(feedback)
gradually enables an infant to organize his world of sensation.	(non-text-example-positive)
This depends on developing sense organs and nervous system.	(feedback)
enables the hungry baby to smile when he hears the mother's footsteps.	(non-text-example-negative)
He has learned to associate her approaching steps with food.	(feedback)

(b) Matching terms: primary stimulus and sensitization

discrimination property	primary stimulus sensitization	(heading)
always operates naturally on the organism	(primary stimulus)	(text-attribute)
It corresponds to a reflex - having an automatic character.		(feedback)
often described as pseudo-conditioning	(sensitization)	(text-attribute)
It appears to be learning (conditioning) but it is not permanent.		(feedback)
meat powder presented to a dog	(primary stimulus)	(text-example)
It leads almost automatically to salivation.		(feedback)
said by critics to be responsible for the "conditioning" of flatworms	(sensitization)	(text-example)
Their controversial responses may be the result of an increase in sensitivity.		(feedback)
a US that is not the product of training	(primary stimulus)	(non-text-attribute
It could not have formerly been a CS		(feedback)
leads an organism to respond to stimuli with which it has had no past experience.	(sensitization)	(non-text-attribute
The threshold of response to novel stimuli is lowered by this process.		(feedback)
a blow just below the knee	(primary stimulus)	(non-text-example)
It evokes a rather mechanical response.		(feedback)
Example: a cat jumps back when a bell sounds after shock was administered, yet bell and shock had not been associated.	(sensitization)	
He is temporarily sensitized to novel or unexpected stimuli.		(feedback)