R E P O R T   R E S U M E S

ED 010 540   56

THE INTERACTION OF WORDS AND GRAPHIC SYMBOLS, INVESTIGATED VIA A PROGRAMED SEQUENCE OF CONCEPT FORMATION EXPERIENCES RELATED TO VECTOR SPACES.

BY- SMITH, M. DANIEL   BOLYARD, A. JOYCE
EARLHAM COLL., RICHMOND, IND.

REPORT NUMBER BR-5-0888   PUB DATE   65
REPORT NUMBER NDEA-VIIA-1284
GRANT   OEG-7-24-0090-250
EDRS PRICE   MF-$0.09   HC-$1.28   32P.

DESCRIPTORS- *PROGRAMED INSTRUCTION, *CONCEPT FORMATION, TASK PERFORMANCE, *PROGRAMED UNITS, COMPARATIVE ANALYSIS, LEARNING ACTIVITIES, GRADE 5, GRADE 6, *SEQUENTIAL APPROACH, *NONVERBAL LEARNING, RICHMOND, INDIANA

A SEQUENCE OF LEARNING TASKS WHICH USED NONVERBAL STIMULI TO INTRODUCE CONCEPTS OF VECTOR SPACES WAS CONSTRUCTED. THE SAMPLE WAS 20 CHILDREN FROM GRADES 5 AND 6 WHO WERE MATCHED ON THE BASES OF INTELLIGENCE, READING, AND ARITHMETIC ACHIEVEMENT. EACH STAGE OF THE PROGRAMED SEQUENCE WAS PRESENTED TO EACH SUBJECT AS A CONCEPT-FORMATION PROBLEM. FOLLOWING EACH LEARNING TASK, THE SUBJECT DREW HIS ANSWER IN A SIMILAR FRAME. THE FIVE SEQUENCES AVERAGED 11 STAGES EACH. THE TOPICS TREATED WERE VECTORS AS OPERATORS, VECTOR ADDITION, SCALAR MULTIPLES, COMMUTATIVITY, AND ASSOCIATIVELY OF VECTORS. DATA RECORDED FOR EACH STAGE WERE--(1) NUMBER OF EXEMPLARS TO CRITERION (TWO CORRECT RESPONSES), (2) MEAN TIME PER RESPONSE, AND (3) MEAN ACCURACY OF RESPONSE. CONCLUSIONS WERE THAT (1) COMPARATIVE PENETRATIONS (TRIALS TO CRITERION) ON THE FIRST 3 STAGES SHOWED VERBAL TASKS WERE LOWER, (2) NONVERBAL TASKS REQUIRED FEWER TRIALS TO CRITERION, AND (3) VERBAL STIMULI WERE REGARDED AS HIGHLY REDUNDANT IN RELEVANT DIMENSIONS FROM CONTROLLING RESPONSES. FURTHER INDICATIONS WERE THAT NONVERBAL SEQUENCES COULD BE EMPLOYED IN TEACHING CERTAIN CONCEPTS. EVALUATION OF THE RESULTS BY THE CRITERIA OF TRIALS-TO-TIME, TIME, AND ACCURACY FAVORED USE OF THE NONVERBAL SEQUENCES. (RS)

5-0888
VII-A-1284

THE INTERACTION OF WORDS AND GRAPHIC SYMBOLS, INVESTIGATED VIA A PROGRAMED SEQUENCE OF CONCEPT FORMATION EXPERIENCES RELATED TO VECTOR SPACES

(Original title: An Investigation of Certain Variables Suggested by Programed Sequences Using Non-Wordal Stimuli: Response Dependent Redundancy Interaction of Words and Non-Wordal Stimuli, and Three Criterion Measures.)

M. Daniel Smith
Earlham College
Richmond, Indiana

Title VII Project Number 5-0688 A- 1284
National Defense Act of 1958
Grant No. 7-24-0090-250

The Research Reported Herein Was
Supported by a Grant from the
U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Office of Education

Abstract

The Interaction of Words and Graphic Symbols, Investigated via a Programed Sequence of Concept Formation Experiences Related to Vector Spaces.

M. Daniel Smith
A. Joyce Bolyard
Earlham College

Learning tasks were constructed in which no words were involved. Each task represented one stage in a programed sequence, and each stage was presented to each S as a concept-formation problem. S constructed (drew) his response at all stages of the initial sequence, and at some stages in later sequences. There were five sequences, averaging eleven stages each; they treated vectors as operators, vector addition, scalar multiples, commutativity, and associativity of vectors.

Data recorded for each stage were (1) number of exemplars to criterion (two correct responses), (2) mean time per response, and (3) mean accuracy of response. Performance of ten fifth grade Ss was compared with that of another ten who covered a similar sequence with word tasks added to each exemplar. Simultaneous and successive presentations were also involved.

For the first sequence of eleven stages, a trend analysis of data type (1) above indicated that the trends for the wordless and wordal versions were significantly different (.05). The wordal treatment differed from the non-wordal on data type (2) (.01), as did simultaneous and successive treatments (.001). Trend differences were found on data type (2) for non-wordal-wordal (.01) and simultaneous-successive (.05). On data type (3) the wordal-non-wordal treatment difference was significant (.001). All differences favored non-wordal stimuli and successive presentation.

On the second sequence of fifteen stages, similar differences were demonstrated for Type (1) and Type (2) (no accuracy scores). There was also a significant difference between the wordal and non-wordal treatments on Type (1) (.05), favoring non-wordal.

On the last three sequences differences tended to become non-significant, with the exception of a wordal-non-wordal trend difference on Type (1) (.01) in the last sequence. These data were clouded by the dropping out of one, then two Ss from one treatment (non-wordal successive).

A post-test was designed to yield data similar to those recorded during learning. It involved seven tasks with four exemplars each. Ss responded to the first exemplar of each task, with feedback after each response; they were given the second exemplars, in different order, again with feedback, and then the third exemplars and so on. This test was administered two weeks after completion of learning, and again two months later. On the "immediate" post-test there was a simultaneous-successive treatment difference on Type (1) (.10); no other differences appeared for this data type. For Type (2) there was a difference in trend for all treatments, and an interaction (all .01). All differences favored non-wordal stimuli and successive presentation.

On the delayed post-test, for Type (1) there was a weak trend difference for wordal-non-wordal (.25). There was also a Type (2) difference in trend for interaction of treatments (.25).

Scores on the delayed post-test were generally better than corresponding scores on the immediate post-test, and Ss generally took less time. On the two test tasks yielding accuracy scores, the accuracy of the simultaneous groups decreased while that of the successive groups increased. Accuracy for non-wordal was superior on the immediate, slightly inferior on the delayed post-test.

GENERAL INTRODUCTION

The inclusion of this general introduction is prompted by the somewhat unique nature of this investigation, and by difficulties experienced in the past in communicating its various dimensions to other investigators and interested persons. The investigation relates to three major areas; programed instruction, the teaching of mathematics and science, and basic research in concept formation.

In the area of programed instruction the investigation deals with step-size, frame format, and more basically the problem of small-sample trial-and-revision development (experiment analysis) versus large-sample comparisions-of-treatments (experimental design); it also deals with problems of pre-testing and post-testing.

In the area of research in mathematics teaching, it deals with the use of relatively sophisticated concepts in early grades. This involves considerations of concept heirarchies, the identification of means by which basic all-embracing concepts may be adapted for use early in learning so that later learning may be more efficient and effective. It also deals with optimal learning sequences and improved stimulus presentations, matters affecting conventional text construction and class-room teaching as well as programing.

In the area of research in concept formation, it adapts concept-attainment procedures and language to a problem in applied teaching, and at the same time collects data relevant to basic problems investigated in concept formation research (verbal vs. non-verbal stimuli, simultaneous vs. successive presentation).

<u>Further Details:</u> Originally a "linear" program, the sequence was turned into a series of concept-formation tasks. Then by setting a learning criterion at each stage, it became a "by-passing" procedure which returned a difficulty score (trials to criterion) for each student at each stage; this process adjusts the difficulty of the program to the individual. Time-to-respond and accuracy-of-response were also also recorded. A post-testing procedure was designed which gives a relearning score comparable to difficulty scores gathered during original learning. To analyze data, a trend analysis was employed which yields significant treatment differences over groups of stages, in spite of the small N typical of programing research.

The program itself is a sequence of learning tasks which introduces the subject of vector spaces. A unique feature of this investigation is its use of stimuli which have no words: such stimuli are termed "non-wordal". Initial investigations into the feasibility of such sequences showed that they enabled one to include a number of other features which conventional frames cannot encompass, among them the concept-attainment approach, the recording of a difficulty score at each stage, and others described above. Words were added to the non-wordal stimuli, thus treatment to gain some indication of the interaction of the two types of stimuli.

The learning tasks feature the use of a two dimensional arrow as a representation of a vector, and a dot placed somewhere in the vicinity representing a portion of the space operated on by the vector. To treat the vector as an operator, it was necessary to convey the idea of initial and final states (of the space, as represented by the dot). To do this, X would present S with a problem frame (a square of paper with the stimulus thereon), and then would present him with a "student answer frame"; here the point, which was originally solid, would be represented by a hollow dotted circle. After S had a chance to respond (ordinarily he could not do so on the first frame) he was presented with the "answer frame"; this gave him the new position of the point as well as the old position. The whole process conveyed the concept of initial and final states. The success of this approach was partially demonstrated at a later stage, where Ss were able to add vectors in the space of a few trials, by being given the problem of two vectors acting on one point. Thus forming an additional treatment.

REPORT OF RESEARCH

A sequence of learning tasks employing a relatively sophisticated mathematical approach to vectors and vector spaces was constructed without using words. The feasibility of teaching via such non-wordal stimuli was demonstrated, using fourth and fifth graders in an individual laboratory setting, and then field tested with modest success in a public school classroom (15). Subsequently, wordal tasks were designed to be added to the frames in order to investigate their interaction with the non-wordal stimuli: samples of the non-wordal and wordal tasks are shown in Figure 1.

The use of the term "non-wordal" is based on an analysis of verbal learning by Skinner (14), who treats as verbal all behavior whose reinforcement is contingent upon the resulting behavior of another. Using this analysis, graphical stimuli such as vectors and mathematical symbols are "verbal". Thus the stimuli used in this study must be classed as either "wordal" or "non-wordal", since both are verbal.

There is some research related to the interaction of wordal (verbal) and non-wordal (non-verbal) stimuli. Wittrock and Keislar (19) investigated the effect of verbal cues chosen from an intuitively established hierarchy, and found that levels within this hierarchy differed in effectiveness. They did not order their concepts sequentially, as was done in the present study, nor did they use non-wordal stimuli except in tests of transfer. Runquist and Hutt (12) and Johnson and O'Reilly (7) studied the comparative effects of pictorial and verbal stimuli in concept formation, and in each case the verbal (here "wordal") version was more effective; again there was no sequential development of concepts. On the other hand, developmental research in programed instruction by Holland (13) and Hively and Porter (6) evidence successful intuitive use of non-wordal sequences. Furthermore, non-wordal approaches to teaching have been used by such innovators in the teaching of mathematics as Beberman (1) and Suppes (16). There also seems to be a relationship between non-wordal (wordless) stimuli as used here and "meaningless" verbal learning stimuli used in research reported by Postman (11), where results implied that "meaningless" material is retained as well as "meaningful" material, given that they are learned to the same criterion (there was even a hint that meaningless material may be retained better under certain circumstances).

One objective of this research, in addition to investigating the interaction of words and non-wordal stimuli, was to form a sequence of tasks that would be "well-ordered" in that it would represent an optimal use of student energy in achieving understanding of the concepts and processes involved. It was assumed that to do this, a sequence would have to be both "challenging" and capable of reinforcing "discovery" behavior. To achieve this, the step-size" of the programed sequence was made (intuitively) fairly large, and then a number of similar tasks were developed at each stage (stage here corresponds to the conventional "step" in programing jargon). Each S was allowed to move to the next stage as soon as he reached criterion, which was two consecutive correct responses. Thus each stage became a concept-formation task, using constructed responses. Since S passed to the next stage after reaching criterion, "branching" or "bypassing" was also involved.

The intuitive process by which additional tasks were constructed for each stage can be described best in terms identified with concept identification research. Each of the tasks at a given stage had the same number of relevant and irrelevant dimensions but the irrelevant ones were varied from task to task. The desirability of this approach is suggested by the investigations of Kurtz and Hovland (9) and Peterson (10), which imply that exemplars of a concept should be presented in succession, without other concepts interspersed. The tasks at a given stage were not inherently ordered as far as we could determine: they assumedly could be reordered without changing the outcomes of the experiment, although some seemed more difficult than others as they were administered to Ss.

The ordering of stages may also be described in concept-identification terms, although the order was originally arrived at intuitively. It involves the addition of relevant and irrelevant dimensions, which according to Walker and Bourne (17) results in added difficulty. Both this study and that of BOurne and Haygood (3) imply that greater ease in learning could have been achieved by increasing the redundancy of either relevant or irrelevant dimensions or both; however, such "cueing" was deemed unnecessary since four or more exemplars were available at each stage.

Another aspect of the research reported here relates to mode of presentation. Comparisons between simultaneous and successive presentations by Bourne, Goldstein and Link (2), Cahill and Hovland (4) and Kates and Yudin (9) imply that availability of previous exemplars has a positive effect on learning, particularly if time for response is not limited. Thus availability (to S at each presentation) of the previous two exemplars, characteristic of the simultaneous treatment in this study, should be beneficial, especially since Ss were allowed as much time per frame as they needed.

Materials and Treatments: This report involves five sequences of learning tasks; the first introduces the concept of a vector (uncentered) as an operator, and the concept of vector addition, and has eleven stages. Each stage is made up of a number of exemplars each exemplar having the same number of relevant and irrelevant dimensions; most of the irrelevant dimensions vary from exemplar to exemplar. Each exemplar consists of a problem frame, a student's response frame, and an answer frame (Figure1). The sequences which follow are made up in a similar manner and cover vector notation, scalar multiples, commutativity, and associativity.Each stage, or set of several exemplars, is called a "P-series". The number of exemplars within the series is generally four, but on P1, P2, P3, and P6 in the first sequence it was necessary to provide many more exemplars to bring students to criterion. Criterion was two successive correct responses, after which S was presented the first exemplar of the next P-series. Correctness was judged subjectively by both the experimenter and the timer.

Each learning task (i.e. each exemplar) was cast in two formats. One involved only the graphical (or "symbolic") representation of the task; the other was exactly the same except that a relevant wordal task was added. The non-wordal treatment was abbreviated "N", the wordal, "W". In the first sequence P1, P2, and P3 could be considered collectively a "learning phase", with P4 - P11 the "applications": however, P6 was a new concept (vector addition),with its own applications, as well as a "transfer task". For full information on the first sequence, see Appendix B.

Subjects: Ss were fifth and sixth graders enrolled in local public schools, matched on the basis of intelligence (California Mental Maturity), reading achievement, and arithmetic achievement (California). Matching consisted of selecting groups of four with similar profiles, and placing one in each cell (NU, NI, WU, WI). In each cell there were two students with IQ greater than 115, and three with IQ less than 115. Comparisons between the above-115 and the below-115 groups (dubbed "A" and "B" respectively) were made on the basis of two in each group in each cell: the third "B" student in each cell had been given a slightly different program due to an error which occurred early in the investigation.

Learning Procedure: Each S, individually, was shown (by X) a frame which stated the learning task. Then S was given a similar frame on which to construct his response. No verbal instructions were given. Finally, S was shown another frame with the correct answer. If S asked a question concerning the response frame he was asked, "Would you like to see the answer?" If the question concerned the correctness of his answer, S was told, "I'll let you decide about that."

The correct answer was removed last, so that the correct response was always the last in view. S was then given another exemplar of the same task. This continued until the S had responded correctly to two tasks in a row: then he was presented with the first task in the second stage.

Each S was timed, beginning when the problem frame was presented and ending when the S completed his response. Most Ss covered the first book in one session of about three-quarters of an hour.

There were three scores: penetration, time, and accuracy. "Penetration" was the number of exemplars required to reach criterion at a given stage; "time" was the mean response time per frame in a given stage; "accuracy" was the total of two scores, one for length and one for direction (each had a maximum of two points). These data were plotted for each set of treatments and for each interaction; two graphs for the W-N dimension are given in Figure 2.

There were two procedural treatments. In one, the answer frames from the previous two tasks were retained within the view of the student; in the other, no previous material was left in view. This corresponds to treatments described as "simultaneous" and "successive" (abbreviated I and U respectively) in studies of concept identification.

<u>Testing Procedures</u>: A testing procedure was designed to yield data compatible with those recorded during learning. Each of seven test tasks was cast in four forms, comparable to exemplars at each stage in the learning sequences. S responded to the first exemplar of each task, one at a time, and was shown the correct response to each before going on to the next. After S responded to the second of each; in the second phase, the order was changed; as before, S was shown the correct answer to each task (after responding) before being given the next. This was carried on until S had responded to all four exemplars of each of the seven tasks. Since S was shown the answer in each case, this was a relearning experience as well as a conventional test. Scoring was as follows ("C" means "correct", "W" means "wrong"):

First exemplar:	W	W	W	W	W	W	W	W	C	C	C	C	C	C	C	C
Second exemplar:	W	W	W	W	C	C	C	C	C	C	C	C	W	W	W	W
Third exemplar:	W	W	C	C	C	C	W	W	C	C	W	W	W	C	W	W
Fourth exemplar:	W	C	W	C	C	W	C	W	C	W	C	W	C	W	C	W
Score:	5	4	4	3	2	4	3	4	1	3	2	4	2	3	2	4
			*			*	*	*		*	*	*	*	*	*	*

Thus, for example, a student who in answering the four exemplars for the first test task got the first wrong and all the rest right received a score of "2", whereas, if he got the first right, the second wrong, and the others right, he received a "2" also. In the latter case, the asterisk indicates that there was an inconsistency in his response pattern: at least one "W" is found <u>following</u> a "C". This scoring is similar to the scoring of "penetrations" on the learning stages. Inconsistencies were recorded to see whether they furnish an additional qualitative criterion for comparing learning among groups.

Results:

A. Sequence I: Tables 1, 2, and 3 represent the results of trend analyses patterned on examples in Edwards (18).

a. For penetration there is a significant interaction between wordal-non-wordal and stages: the curves are significantly different at the .05 level. These curves are shown in Figure 2; differences favor wordal for P1-P3, non-wordal thereafter.

b. For accuracy and time, the wordal and non-wordal treatments are significantly different at the .005 level (for accuracy curves see Figure 2); differences favor non-wordal.

c. For time, the successive and simultaneous treatments are also significantly different (.01). There is a significant interaction of wordal-non-wordal with stages, also of simultaneous-successive with stages (.05). Finally stages are significantly different for all three criteria, at the .005 level. Differences favor non-wordal stimuli and successive presentation.

While there were some hints of interactions with intelligence, it was not felt that the sampling procedure was well enough defined in this area to merit statistical analysis; an example of a potential interaction is given in Figure 3. Additional details regarding Sequence I are given in Appendix A.

B. Sequence II: Under exemplars-to-criterion, the trend analysis again shows the curves for wordal and non-wordal to be significantly different (.05); as before, differences favor the wordal treatment early in the sequence then non-wordal later. However, in this sequence the wordal and non-wordal treatments are also different (.05), favoring non-wordal. Under time, the non-wordal treatment is significantly quicker than the wordal (better than .001); there is a weak difference between simultaneous and successive, favoring the latter (.25). The trend analyses show time differences for all treatments and the interaction, (all .01); this favors non-wordal and successive respectively. The responses in Sequence II were not of a type to yield an accuracy score.

C. Sequence III: Errors in computation force postponement of reporting results for this sequence.

D. Sequence IV: Under exemplars-to-criterion, there were weak effects favoring the non-wordal treatment (.25) and the non-wordal-successive combination (.25). Time data replicated results in previous sequences.

E. Sequence V: Under exemplars-to-criterion, there was a weak effect favoring the non-wordal treatment (.25), and a slightly stronger trend difference favoring non-wordal (.1). Time data replicated previous results. There were two stages which yielded accuracy scores: no significant differences resulted.

F. Immediate Post-Test: Scores derived according to the process described previously, similar to exemplars-to-criterion scores in the learning sequences, showed no wordal-non-wordal differences either in treatments or in trends. There was, however, a surprising difference between simultaneous and successive treatments, (.1), favoring successive. There was also a difference in frequency of inconsistencies, favoring the successive treatment: this was not tested statistically. In respect to time required to respond to each task, there was a weak difference between simultaneous and successive treatments (.25) favoring successive; all trends showed significant differences, favoring non-wordal, successive, and their combination (all better than .01).

G. Delayed Post Test: The scores derived according to the process described previously, similar to exemplars-to-criterion in learning sequences, showed a weak trend favoring the non-wordal group (.25). For a time, a weak trend effect was found favoring the non-wordal successive combination (.25).

H. Correlations between Sequences and between Sequences and Tests: In a previous study, data suggested the possibility of an optimum trials-to-criterion level for post-test performances, i.e.; it seemed possible that students having an intermediate trial-to-criterion score might show better post-test performance than those having low or high scores. No evidence of this was found here: all correlations between sequences and between sequences and tests show simple positive linear-like scatter plots. Sequence-test correlations increase as one progresses from Sequence I to Sequence V.

I. Immediate vs. Delayed Post-Test:

(1) Scores on delayed post-test were generally better than corresponding scores on the immediate post-test, and Ss generally took less time.
(2) On the two test tasks yielding accuracy scores, the accuracy of the simultaneous groups decreased while that of the successive groups increased.

Discussion: While the addition of a wordal component obviously adds to the relevant and irrelevant dimensions of a task, thus making it more complex and time consuming, there is some reason to believe that the information the words contain should aid the student in solving the task, and thereby decrease the number of trials to criterion. That this has an element of truth is attested to by comparative penetrations (i.e. trials to criterion) on the first three stages of Sequence I where wordal Ss were lower. On subsequent stages the non-wordal Ss had the advantage. Moreover, at Stage 6, where a new concept was introduced (addition of vectors), the non-wordal Ss tended to require fewer trials to criterion. A similar early-late difference was noted in Sequence II.

These results indicate that previous studies comparing pictorial and verbal presentations must be interpreted conservatively. Since the concepts used in those studies (4,5) involved dimensions which have been thoroughly learned by Ss previous to the experiment (e.g; the color of wing, shape of beak), words may have recalled previous experience more effectively than in this study where previously learned concepts are not as clearly applicable to the subject matter. Words seemed to be of some help during early stages, where fundamental relationships were learned.

To compare non-wordal and wordal stimuli, one may regard the latter generally as being highly redundant in relevant dimensions (when well learned by S) and thus capable of preventing irrelevant dimensions from controlling responses.

One may challenge the effectiveness of the particular words used in this study, but results suggested that the superiority of any given wordal presentation can no longer be considered a foregone conclusion. Experience gained in developing the wordal and non-wordal components of these frames indicates that further improvement in both, as well as in their interaction, is to be expected.

In one sense these data support research indicating advantages of a "Rul-Eg" approach, since in this study words aided learning in early stages; but the non-wordal stimuli seemed more effective in later stages. On the other hand, the failure of words to assist in the learning of a new concept (Stage 6) implies that the matter is more complex than the simple "Rul-Eg" formulation admits, whether one takes the point of view described above or simply looks on the wordal version as a rul-eg process where the responses to the words first and then responds to the non-wordal task (this was most frequently the procedure, and many Ss failed to respond to the non-wordal component on the early tasks at a given stage).

One outcome from Sequence I which was not anticipated was the highly significant difference in accuracy of response found in both sets of treatments, W-N and I-U. Accuracy, like time, seems to be a function of the complexity of the stimulus field, rather than of any particular treatment variable. Incidently, where wordal Ss failed to respond to non-wordal components of a task, these tasks were not used in computing mean accuracy.

The overall absence of significant differences between simultaneous and successive treatments in trials to criterion (penetration) may be interpreted as a failure to replicate the results of Yudin and Kates (8) and of Bourne, Goldstein and Link (2). It is true that the number of subjects is rather small, but ample time was allowed for response. Time and accuracy data imply that stimulus complexity is a more powerful factor than reduced memory load, at least in this learning situation.

The immediate post-test results contain several simultaneous-successive comparisons which suggest that the successive treatment had some differential effect on the Ss: this may reflect the fact that two of the successive-wordal students had dropped from the experiment. On the other hand, the delayed-immediate comparisons for accuracy show a puzzling difference which would not seem to be biased by drop-outs: the successive students scores increased while the simultaneous group's scores decreased. No hypothesis is offered for this effect; it would seem to call for further investigation with more subjects. It may be that the strong time effect, seen all the way through, relates to it.

The appearance of even weak effects on the post-tests is worthy of some note. The difference in time required to respond, favoring non-wordal as well as successive, is interesting since there is of course no treatment difference for the groups on the post-test. This implies that Ss learned to take different amounts of time to respond as a result of treatments which during learning required more time; of course, there are no data to prove that the two groups did not differ in time-to-respond before beginning the experiment, and while unlikely this possibility should be considered in designing future experiments of this type. That this difference had largely disappeared by the delayed post-test except for a weak wordal-successive trend interaction would imply that the groups did not differ originally in this respect.

The improvement of scores from immediate to delayed post-test may only reflect familiarity with the test, since the nature of the test made it a fairly thorough learning situation as well. However, since response to the first exemplar of each task wasgiven the greatest weight, there may be some reminiscence-like effect as well.

The nature of feedback here requires some comment. Information regarding errors is implicit in the showing of the correct response following the response of S. This by no means assures strong feedback, however: it was observed that many Ss interpreted the repeated showing of the correct response to mean that they had the wrong hypothesis, rabher than that they were not sufficiently accurate in their response; this interferred with the formation of the correct concept. Such feedback is "weak" in the terminology of Wallace (18) and thus conducive to hypothesizing and testing. The investigators originally conceived of this feedback as a means of shaping "discovery" behavior, as well as of making pre-testing an integral part of the programing process (ordinarily one is never sure in programing whether one has made the task difficult enough). Further attention is being given the feedback dimension, since it seems particularly important in working with lower ability Ss. It may also be worthwhile in the future to record the time devoted by S to inspecting the correct answer, once he has responded: this datum may be or may not be highly correlated with time-to-respond, and it may be related to the "learning style" of S.

Summary: This study suggests that non-wordal sequences can be employed advantageously in teaching certain concepts; the criteria of trials-to-criterion, time, and accuracy all favor non-wordal sequences and in some cases, successive treatment as well.

This study indicates the feasibility of replicating aspects of basic concept formation research in applied subject matter areas, and suggests that terminology from such research can be applied to "practical" teaching to good advantage.

This study implies that small-group program-development studies which emphasize experimental analysis of behavior can be given an experimental-design emphasis without going to prohibitively high "N's"; however, results from such a study should be taken as suggestive for more rigorous investigations by others who concentrate on experimental design and comparison of treatments.

Finally, this study demonstrates that research in curriculum design involving the experimental use of new approaches to teaching, stemming from advanced sophisticated subject matter frames of reference, can be carried on in conjunction with research in learning; indeed, in the opinion of the investigator, it is impossible to separate this dimension of research from that of research in learning. In this study, the approach to teaching vectors (and in a larger sense to teaching algebra) which resulted from our desire to draft non-wordal sequences produced a learning sequence which has unusually powerful implications for the more effective teaching of algebra at all levels.

Tables I, II, III
Trend Analyses
Sequence One

(1) Trials to Criterion

Source of Variation	SUM OF SQUARES	d.f.	MEAN SQUARE	F
WN	.2228	1	.2228	
IU"	18.0410	1	18.0410	1.C694
WN x IU	4.9499	1	4.9499	
Error (a)	269.9273	16	16.8705	
Stages	1,027.7455	10	102.7746	14.9916 ***
WN x St.	141.7272	10	14.1727	2.0673 *
IU x St.	47.3090	10	4.7309	
WN x IU x St.	53.8001	10	5.3800	
Error (b)	1,096.8727	169	6.8555	

(2) Time

Source of Variation	SUM OF SQUARES	d.f.	MEAN SQUARE	F
WN	24,825.314	1	24,825.314	22.142 ***
IU	10,213.641	1	10,213.641	9.110 **
WN x IU	206.222	1	206.222	
Error (a)	17,939.237	16	1,121.202	
Stages	9,816.610	10	981.661	7.405 ***
WN x St.	3,499.935	10	349.994	2.640 **
IU x St.	3,110.009	10	311.001	2.346 *
WN x IU x St.	1,008.829	10	100.883	
Error (b)	21,212.163	160	132.576	

(3) Accuracy

Source of Variation	SUM OF SQUARES	d.f.	MEAN SQUARE	F
WN	10.5603	1	10.5603	10.6230 ***
IU	.1326	1	.1326	
WN x IU	.0115	1	.0115	
Error (a)	15.9049	16	.9941	
Stages	35.8053	10	3.5805	10.4145 ***
WN x St.	4.5517	10	.4552	1.3240
IU x St.	2.7274	10	.2727	
WN x IU x St.	3.8025	10	.3803	1.1059
Error (b)	55.0028	160	.3438	

CODE: *** .005 ** P .01 P .05

Tables IV, V
Trend Analyses

(1) Trials to Criterion

Source	df	S S	M S	F
WN	1	28.2134	28.2134	7.4279 *
IU	1	0.0000	0.0000	
WN x IU	1	5.8800	5.8800	1.6481 "
Error (a)	16	60.7733	3.7983	
St.	14	87.3667	6.2405	3.7664 **
WN x St.	14	47.0866	3.3633	2.0299 *
IU x St.	14	8.3000	0.5929	
WN x IU x St.	14	16.0200	1.1443	
Error (b)	134	222.0267	1.6569	
Total	209	475.6667		

(2) Time

Source	df	S S	M S	F
WN	1	74261.3334	74261.3334	27.1492 **
IU	1	5598.7201	5598.7201	2.0468 "
WN x IU	1	604.9192	604.9192	
Error (a)	16	43764.8270	2735.3017	
Stages	14	67603.6667	4828.8333	185. **
WN x Stages	14	75206.8666	5371.9190	206.5 **
IU x Stages	14	2103.4799	150.2486	5.78**
WN x IU x St.	14	71843.14[illegible]5	5131.6534	206.5 **
Error (b)	134	3484.7063	26.007	
Total	209	344471.6667		

CODE: * .05 ** .01 " .25

Table VI

		IPT	Penetration	
Source	d.f	SS	MS	F
WN	1	0.7404	0.7404	
IU	1	10.2912	10.2912	3.9701(.1))
WN X IU	1	4.4628	4.4628	1.7216
ERROR (a)	14	36.2913	2.5922	
ST	6	77.0000	12.8333	11.2949
WN X ST	6	5.1096	0.8516	
IU X ST	6	4.1838	0.6973	
WN X IU X ST	6	1.9789	0.3298	
ERROR (b)	84	95.4420	1.1362	
Total	125	235.5000		

Table VII

		IPT	Mean Time	
Source	d.f	SS	MS	F
WN	1	1225.1520	1225.1520	1.0277
IU	1	2528.7240	2528.7240	2.1212(.25)
WN X IU	1	3.1452	3.1452	
ERROR (a)	14	16689.9947	1192.1425	
ST	6	26569.9541	4428.3307	193.2360
WN X ST	6	1333.6369	222.2728	9.6992(.01)
IU X ST	6	825.2399	137.5400	6.0017(.01)
WN X IU X ST	6	893.9909	148.9985	6.5017(.01)
ERROR (b)	84	1925.0053	22.9167	
Total	125	51994.8730		

Table VIII

		DPT	Penetration		
Source	d.f.	SS	MS	F	
WN	1	2.909	2.909		
IU	1	3.7866	3.7866		
WN X IU	1	1.1689	1.1689		
ERROR (a)	13	69.7620	3.059		
ST	6	125.1429	20.8571	22.14	
WN X ST	6	8.0889	1.3481	1.431	(.25)
IU X ST	6	1.8218	.3036		
WN X IU X ST	6	0.910	.1520		
ERROR (b)	78	73.4650	.9418		
Total	118	287.0551			

Table IX

		DPT	Mean Time		
Source	d.f.	SS	MS	F	
WN	1	148.3695	148.3696		
IU	1	27.6895	27.6895		
WN X IU	1	291.7937	291.7931		
ERROR (a)	13	6306.1647	485.0896		
ST.	6	26847.3950	4474.5658	33.5372	
WN X ST	6	820.1523	136.6921	1.0245	
IU XST	6	383.7609	63.9602		
WN X IU X ST	6	1082.9994	180.4999	1.3529	(.25)
ERROR (b)	78	10406.8353	133.4210		
Total	118	46315.1597			

FIGURE I

Example of One Exemplar from STAGE 3

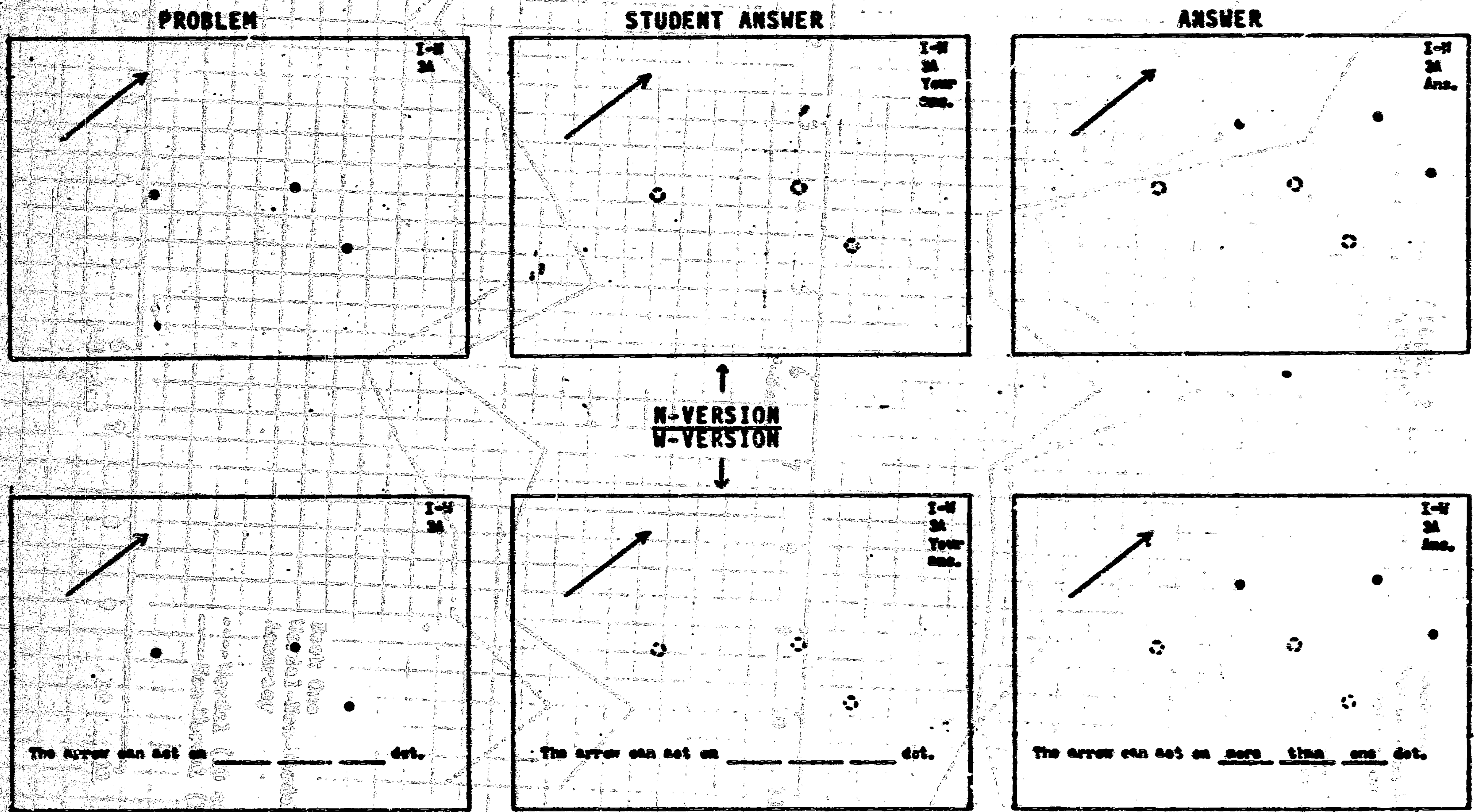

FIGURE 2
GRAPHS

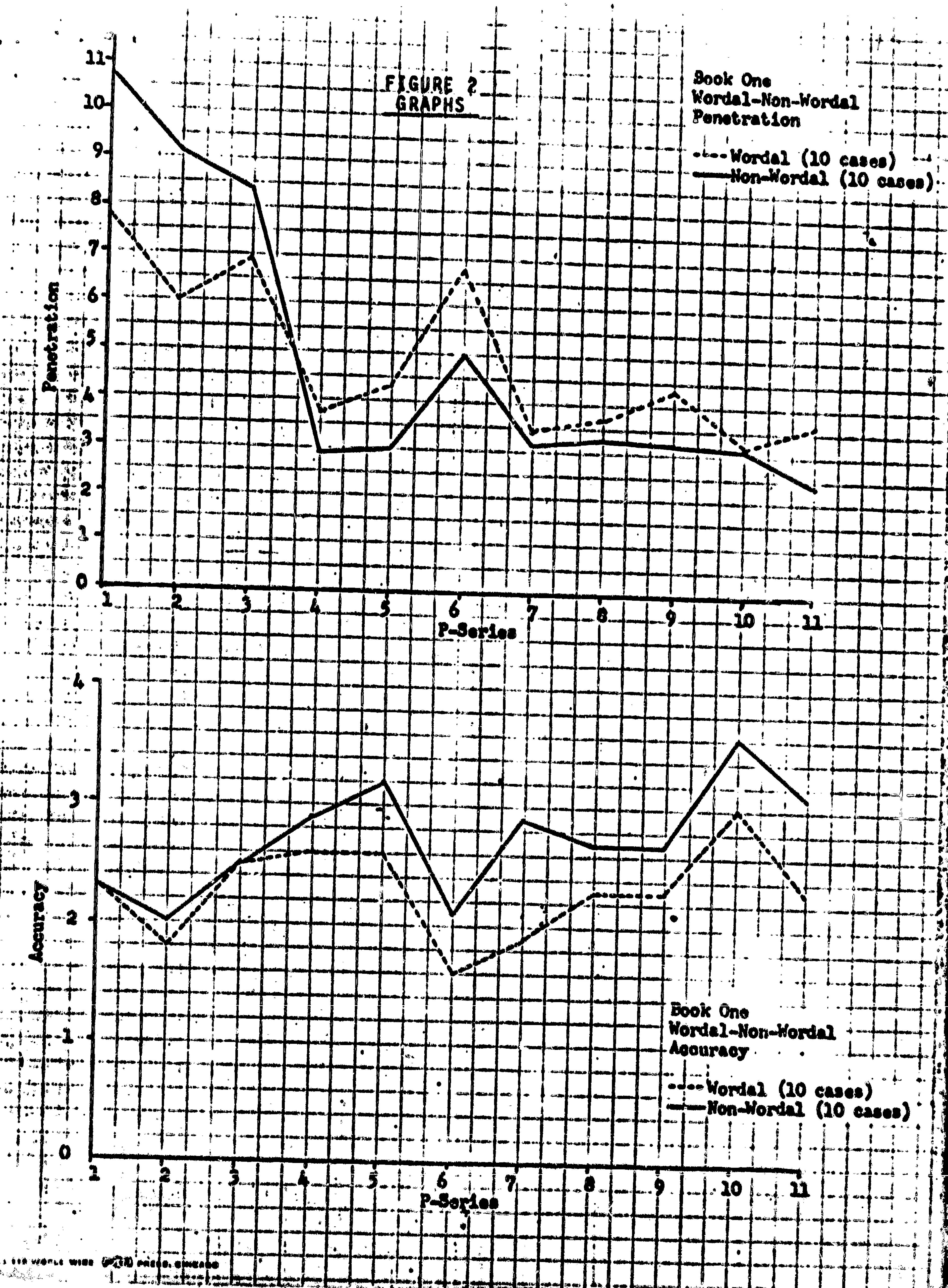

FIGURE 3

Example of Possible Interaction of I-U Dimension with Intelligence.

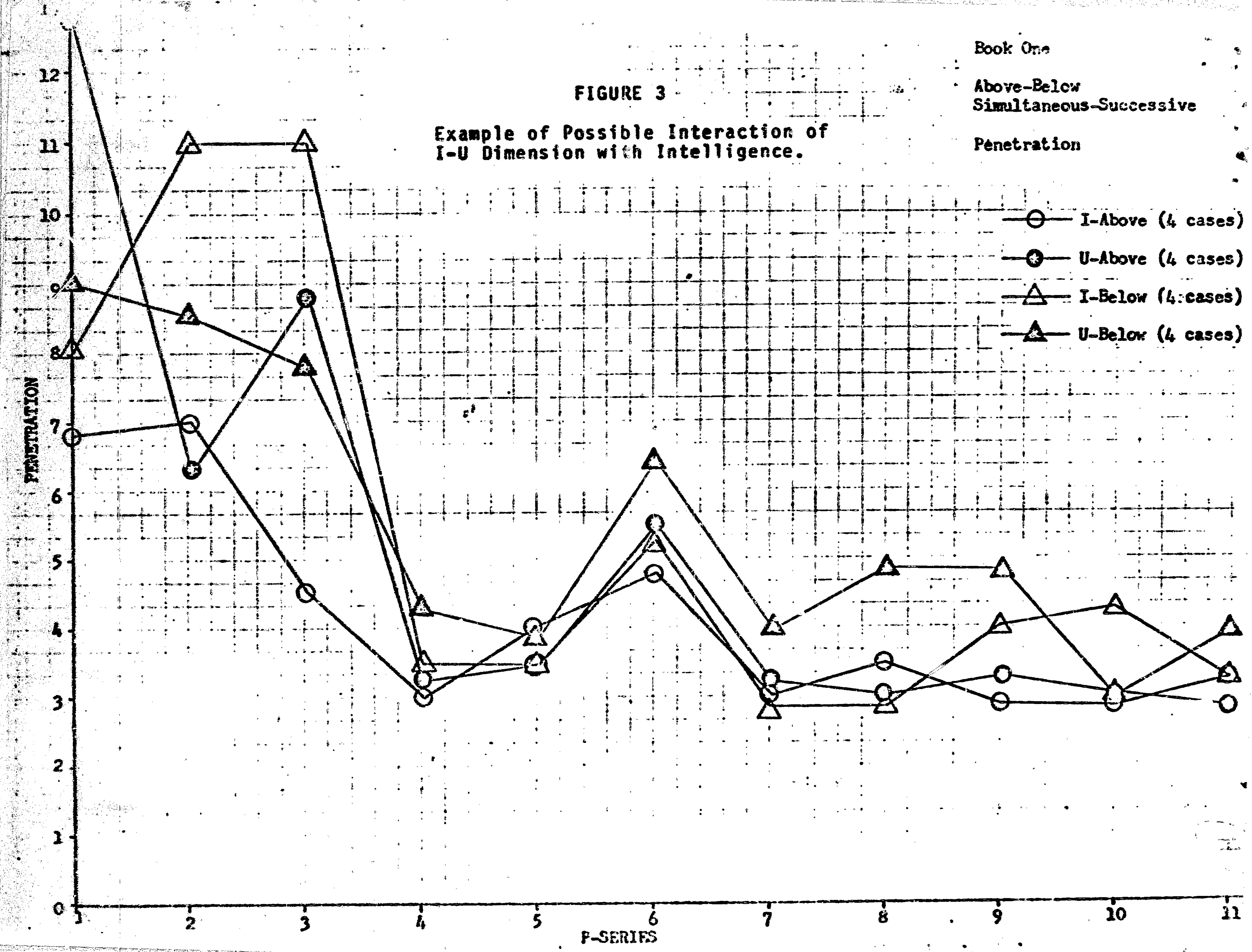

Figure IV

COMPARISON IPT - DPT

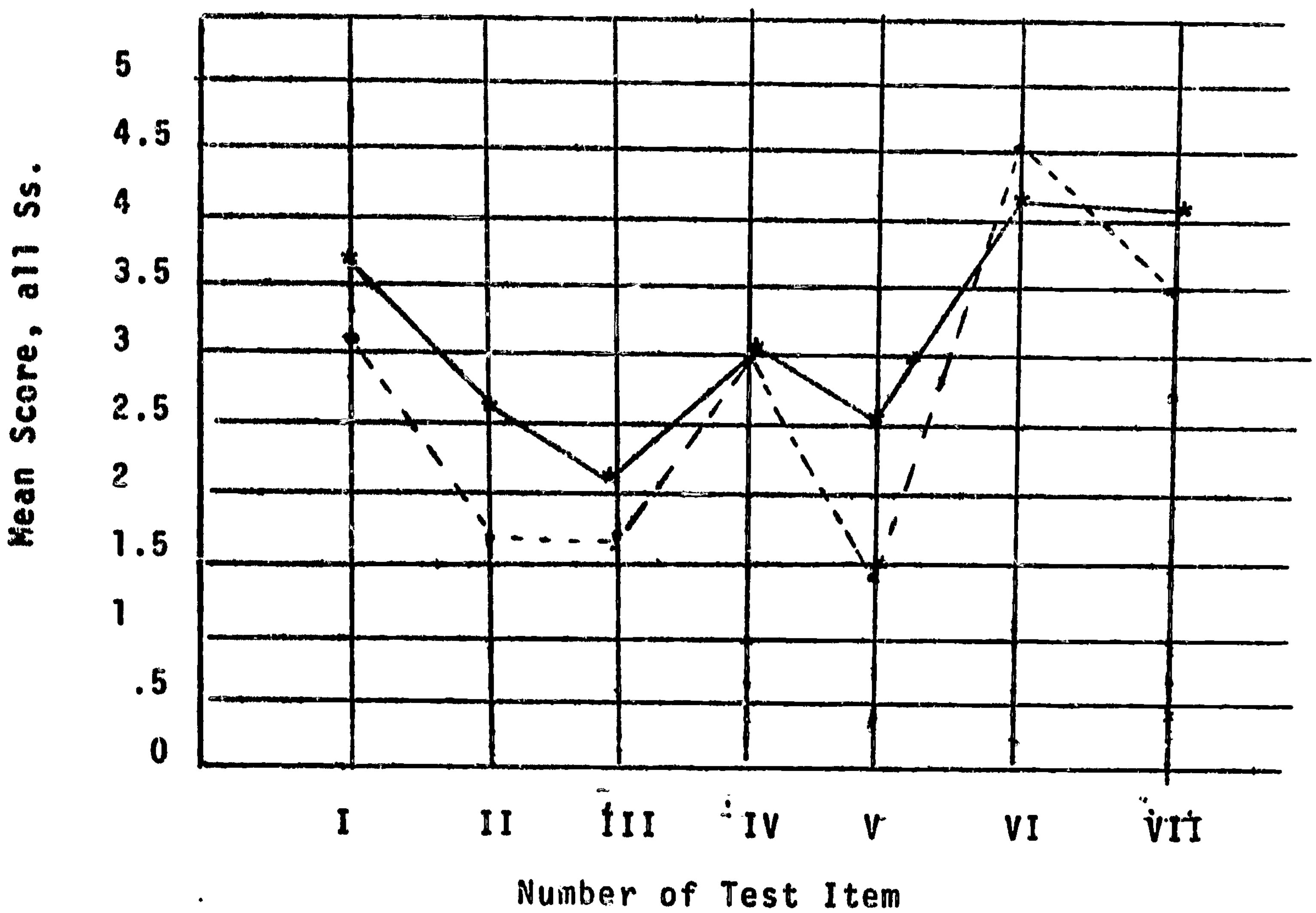

TABLE X

Post-Test Accuracy

Mean Accuracy on delayed Post-Test minus that on Immediate Post-Test.

WI	.11
WU	-.85
NI	.29
NU	-.70

i.e. accuracy <u>decreased</u> for I, increased for U.

REFERENCES

(1) Beberman, Max — Demonstration Class in Mathematics for Seventh Grade, held at Earlham College March, 1962.

(2) Bourne, L. E. and Haygood, R. C. — "The Role of Stimulus Redundancy in Concept Identification", Journal of Experimental Psychology, Vol. 58, No.3, pp.232-238, 1959.

(3) Bourne, L.E. and Goldstein, S. Link, W.E. — "Concept Learning as a Function of Availability of Previously Presented Information", Journal of Experimental Psychology, Vol.67, No.5, pp. 439-448.

(4) Cahill, H.E. and Hovland, C.I. — "The Role of Memory in the Acquisition of Concepts", Journal of Experimental Psychology, Vol.67, No.5, pp. 439-448, 1964.

(5) Edwards, A.L. — Experimental Design in Psychological Research, New York, Holt, Rinehart, and Winston, 1963.

(6) Hively, W. and Porter, D. — Research in pre-reading by matching-to-sample-tasks, Harvard University, 1960-1961, personal communication.

(7) Johnson, D.M. and O'Reilly, C. — "Concept Attainment in Children: Classifying and Defining", Journal of Educational Psychology, Vol.55, No.2, pp.71-74, 1964.

(8) Kates, S.L. and Yudin, L. — "Concept Attainment and Memory", Journal of Educational Psychology, Vol.55, No.2, pp. 103-109, 1964.

(9) Kurtz, K.H. and Hovland, E.I. — "Concept Learning with Different Sequences of Instance", Journal of Experimental Psychology, Vol.51, No.3, pp.239-243, 1956.

(10) Peterson, M.D. — "Some Effects of the Percentage of Relevant Cues and Presentation Method on Concept Attainment", Journal of Experimental Psychology, Vol.51, No.3, pp.239-243, 1956.

REFERENCES

(11) Postman, L. "The Present Status of Interference Theory", Verbal Learning and Verbal Behavior, McGraw-Hill Series in Phychology, Chap.7, pp. 152-195, 1961.

(12) Runquist, W.N. and Hutt, V.H. "Verbal Concept Learning in High School Students with Pictorial and Verbal Representation of Stimuli", Journal of Educational Psychology, Vol.52, No.2, pp. 104-111, 1961.

(13) Skinner, B.F. "Teaching Machines", Scientific American, November, 1961.

(14) Skinner, B.F. "Verbal Behavior", New York, Appleton-Century Crofts, 1957.

(15) Smith, M. Daniel "Non-Wordal Programing in Mathematics and Science", Journal of Research in Science Teaching, in press.

(16) Suppes, P. Sets and Numbers, Blaisdell Publishing Co., New York, 1962.

(17) Walker, C.M. and Bourne, L.E. "Concept Identification as a Function of Amounts of Relevant and Irrelevant Information", American Journal of Psychology, Vol. 74, pp.410-417, 1961.

(18) Wallace, J. "Concept Dominance, Type of Feedback as Related to Concept Attainment", Journal of Educational Psychology, Vol. 55, No.3, pp.59-66, 1964.

(19) Wittrock, M.C. and Keislar, E.R. "Verbal Cues in the Transfer of Concepts", Journal of Educational Psychology, Vol. 56, No.1, 16-21, 1965.

APPENDIX A

Particular Conclusions regarding Treatments in Sequence I:

N vs. W: The non-wordal treatment is generally preferable according to the three criteria, although the number of exemplars required is lower for the wordal sequence in the first three stages. The mean time is 21.3 seconds per frame less for the non-wordal.treatment (the mean time of the N treatment is 16.5 seconds.).

I vs. U: Successive presentation saves time: there is no consistent difference in penetration or accuracy, although the simultaneous presentation has the advantage in a majority of the series.

A vs. B: The above-115 group is superior on all criteria; the time difference is least.

Conclusions regarding Interactions:

NW X IU: NI and WI frequently appear as outer points in the penetration distribution (see P2, P5, P6, P9, P11). While relative mean times balance this part, the time graph does not show the same symmetry, nor is this reflected in accuracy data except on P9. If one considers the N group only, on penetration and accuracy, I is consistently superior to U this seems to reflect more effective use of additional (previously presented) information when words are not involved. Time still favors U, however.

NW X AB: Considering A alone, the N vs. W comparison is less consistently in favor of N than in the over-all situation. The A group, then, benefits more from words than does the over-all W group, In penetration, WA is superior to NA at several points: interestingly, however, this is balanced by a greater time differential, i.e.; WA shows the greatest mean time of any group at most stages, especially at P5 and P8. Thus while words (for the group) add information to the stimulus, .theyalso require more time for processing.

IU X AB: In penetration in the first stages there is an interesting contrast between IA (low) and IB (high); this is repeated less strikingly at P10. At these stages it seems that only the A-group was able to utilize the additional information in the simultaneous presentation, and that the B-group only became confused. In later stages, on the other hand, it is evident that the extra information aided the B group, since BI has lower penetration than BU on P6, P7, P8, P9, and P11. Thus it would seem that in the confusion of meeting a new situation, the simultaneous presentation added to the problems of adaptation, while subsequently it aided learning.

Of course, the time comparisons favor U, but less at certain stages (P6,P10,P11). One interesting feature of the accuracy graph is found in comparing early and late performance: IB for instance begins low in early frames but is high on P6, P7, P8, low again on P9 and P10, even on P11. Such erratic behavior is hard to account for. There are many instances in both accuracy and penetration data of such erratic behavior: some investigation is suggested into this matter.

Discussion: Inclusion of wordal stimuli is analagous to the addition of relevant and irrelevant dimensions; this interferes with learning, as demonstrated by Walker and Bourne (17), and confirmed incidentally, by others. The NW X AB interactions imply that these added dimensions are redundant for the A group, another way of saying that this group "understood" the words. The NW X AB interaction suggests also that the N approach is preferable for use with heterogeneous groups, since A vs. B differences are minimized.

It is interesting to compare the W and N modes with programing methods dubbed "Rul-Eg" and the like. N would assumedly be described as Eg-Eg-Eg- . . ., while the W would combine both Rul and Eg at each step (observation of Ss indicate that they were following a Rul-Eg-Eg-Eg-. . . process in fact, since they did not attend markedly to the wordal component after the first exemplar).

The intuitive non-wordal approaches of Beberman and Suppes seem to have been validated in part by this experiment. Subjective observations during this and previous projects indicate also that the students have little trouble taking up where they have left off in the sequence in spite of days or weeks between sessions: research reported by Postman (11) implies that relatively meaningless material (represented here by the N treatment) is retained well once learned.

APPENDIX B

Plans for a
Follow-Up
Investigation

Next Steps: A new investigation was designed and stimulus materials drafted to follow up various dimensions of the present one. This is described in some detail here, because it represents an important outcome of this investigation and defines operationally what we consider to be its' strong and weak points.

Subjects: Fifth grade students will be screened for previous learning of the criterion task through a pre-test. Several treatment groups will be matched in that their frequency distributions (using pre-testing data) are similar.

Materials, Learning, and Procedure: An ordered sequence of wordless concept formation tasks related to the subject of vector spaces will be used; this will lead to addition and subtraction of directed numbers, assumedly using addition and subtraction of vectors as the covert mediator for this. In some of the treatments, wordal tasks (designed to communicate concepts or processes similar to those in the non-wordal tasks) will be presented either before or after the corresponding non-wordal ones. Ss will be required to attempt all tasks regardless of their performance on the previous one (in contrast to the present study in which subjects passed on to the next stage after two correct responses). Both non-wordal and wordal tasks will require constructed responses. Feedback on non-wordal tasks will be obtained by having the student superimpose the correct response on his own response sheet; since the sheets are semi-transparent, the two can be compared. To facilitate this superposition process, students will proceed through the text from "back" to "front". Feedback on wordal tasks will consist of the response words, typed under the blank space; these also can be superimposed on the student answer frame.

Treatments and Experimental Procedure: Each concept formation task will require each S to construct a response to each of twelve exemplars, regardless of correctness of response. Within the twelve, there will be four modes. Letting N represent non-wordal tasks and W represent wordal tasks, the four modes will be as follows:

N N N N N N N N N N N N (symbolized "N12)

W W W W N N N N N N N N (symbolized "W4N8")

W W N N N N N N N N N N (symbolized "W2N10")

N N N N W W N N N N N N (symbolized "N4W2N6")

The order within the series or twelve of less non-wordal exemplars will vary from student to student, since exemplars have different difficulties and since we have not analyzed and classified these difficulties as yet. Considering each sequence of twelve as one stage of the program, and using Roman numerals for stages, one can present the overall sequence as follows (in the W4N8 mode):

I (W4N8), II (W4N8), III(W4N8), etc.

To give an idea of the progression qua subject matter, stage I will treat the vector as operating on a point, Stage II will present vector addition (two vectors operating on one point),etc.

The stages are ordered intuitively. It is not a trivial question whether Stage II benefits from the learning of Stage I. This question of proactive facilitation (or transfer) is one which should be investigated. To some extent, however, the procedure is "heirarchical", leading from the concept of a vector as an operator through vector addition, inverse, scalar multiples, and ordered pairs, to addition of directed numbers as an abstraction of the addition of scalar multiples of a single arbitrarily chosen vector, e.g.; $3x + {-2}x =$ _____ x leads to $3 + {-2} =$ _____.

Scores will be derived for each S for each stage as follows:

P equals penetration, the number of responses preceding the first correct response.

C equals consistency, the number of times an incorrect response follows a correct response (high C means low consistency).

T equals time, the total time on each stage.

A equals accuracy of response where applicable.

Ss will fill in a time sheet at the end of each stage to give T, C, and P will be recorded following the experimental session. Thus P,C, and T scores will be derived for each stage for each student.

A pre-test will be administered prior to the experiment; it will include pre-requisite tasks (matching to sample of geometric designs similar to the non-wordal tasks in the experimental sequence, also addition and subtraction of non-directed numbers) and target tasks (adding and subtracting vectors, adding and subtraction directed numbers, also transfer tasks relating to these two areas). Post-tests will be administered covering the same topics with less emphasis on prerequisite behaviors. One post-test will be delayed two days, the other two months. The tests will be concept-attainment tasks similar to those to be used in the program; four tasks at the same level of difficulty will constitute one set, and there will be at least five sets. The tasks within a set will be scattered randomly through the test. Thus for set A (e.g. adding two vectors) there will be four exemplars, a1, a2, a3, a4; similarly for B and C and others. A simplified version of the test consisting of only three sets will then be ordered as follows:

a1, b1, a2, c1, b2, c2, a3, c3, b3, c4, a4,....

From such a test one can take the first of each set and score performance as on a conventional test; one can also get a relearning score. One additional test will be used: this will consist of a series of one hundred problems in adding directed numbers; it will be designed to explore the possibility that covert mediation can be extinguished in such a situation if it is not explicitly reinforced. The explicit reinforcement group will be given problems in adding multiples of a vector graphically; these will be administered at intervals during the test.

To summarize the treatments, they will be as follows:

(a) N12, W4N8, W2N10, and N4W2N6 constitute four treatments.

(b) Explicit reinforcement of the mediator vs. no such reinforcement made up two other groups; these might be called "R" and "E", for "reinforcement" and "extinction".

Hypotheses: There will be several null-hypotheses available:

(1) Average penetrations for the non-wordal components of the four groups are not significantly different as compared through a trend analysis (analysis of variance applied by stages).

(2) The consistencies are not significantly different.

(3) The times are not significantly different.

(4) The accuracies are not significantly different.

(5) The reinforcement group is not significantly different from the extinction group in relation to errors on the test on adding directed numbers.

(6) The order of exemplars within each stage does not cause a significant difference in penetration, consistency, time, or accuracy.