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

ED 014 912 56 EM 006 015

PREDICTION OF EFFECTS WITH SELECTED CHARACTERISTICS OF LINEAR PROGRAMMED INSTRUCTION. FINAL REPORT.

BY- SMITH, MARTIN E. SEIBERT, WARREN F.
PURDUE UNIV., LAFAYETTE, IND.

REPORT NUMBER BR-5-0954 PUB DATE DEC 66
GRANT OEG-7-24-0280-273
EDRS PRICE MF-$0.75 HC-$5.36 132P.

DESCRIPTORS- *PROGRAMED INSTRUCTION, *LINEAR PROGRAMING, *VERBAL LEARNING, *RESPONSE MODE, *FACTOR ANALYSIS, STATISTICAL ANALYSIS, PURDUE UNIVERSITY, LAFAYETTE, INDIANA

ULTIMATE OBJECTIVE OF THIS STUDY IS A SYSTEM IN WHICH INSTRUCTORS MAY IDENTIFY RELEVANT AUDIENCE CHARACTERISTICS AND THEN DESIGN A PROGRAM TO OPTIMIZE LEARNING. MULTIPLE REGRESSION, FACTOR ANALYSIS, AND CROSS-VALIDATING PROCEDURES WERE USED TO ASSESS THE IMMEDIATE AND DELAYED LEARNING EFFECTS OF PROGRAMMED INSTRUCTION. INDEPENDENT VARIABLES STUDIED WERE IN THE AREAS OF VERBAL LEARNING, READABILITY, INCIDENTAL LEARNING, AND RESPONSE DIFFICULTY. 144 PAID STUDENT VOLUNTEERS, PARTICIPANTS IN A PRIOR STUDY, WERE RANDOMLY ASSIGNED TO 3 GROUPS--EXPERIMENTAL, CONTROL, AND CLOZE SCORE. THE 2200 FRAME LINEAR PROGRAM USED TAUGHT CLASSICAL AND OPERANT CONDITIONING. 183 CONSTRUCTED RESPONSE ITEMS FORMED THE CRITERION TEST, AND 6 GAIN SCORES WERE COMPUTED FOR EACH ITEM. COMPLETE STATISTICAL ANALYSIS WAS PREFORMED AND METHODOLOGICAL PROBLEMS WERE DISCUSSED FOR 37 INDEPENDENT VARIABLES. 9 FACTORS IDENTIFIED AND DISCUSSED WERE--QUANTITY OF INSTRUCTION (STEP SIZE), SENTENCE COMPLEXITY, CONCRETE VERSUS ABSTRACT INSTRUCTION, REVIEW, CLOZE SCORE, KEY TEAM, WORD COMPLEXITY, LARGEST FRAME, AND UNNAMED. NEW VARIABLES PROPOSED FOR FUTURE RESEARCH INCLUDE--CATEGORIES OF LEARNING, ESTIMATE OF THE EXTENT TO WHICH RE-CONDITIONS FOR MORE ADVANCED LEARNING ARE SATISFIED, AND INSTRUCTIONAL VARIABLES, SUCH AS FRAME ORDER. (LH)

PREDICTION OF EFFECTS WITH SELECTED CHARACTERISTICS OF LINEAR PROGRAMMED INSTRUCTION

FINAL REPORT

December 1966

Prepared by
Martin E. Smith and Warren F. Seibert

INSTRUCTIONAL MEDIA RESEARCH UNIT
Audio Visual Center
Purdue University
Lafayette, Indiana

Grant Number 7-24-0280-273
United States Office of Education
Department of Health, Education and Welfare

PREDICTION OF EFFECTS WITH SELECTED CHARACTERISTICS OF LINEAR PROGRAMMED INSTRUCTION

FINAL REPORT
December 1966

Prepared by

Martin E. Smith* and Warren F. Seibert

Instructional Media Research Unit, Audio Visual Center

Purdue University
Lafayette, Indiana

Grant Number 7-24-0280-273

United States Office of Education,

Department of Health, Education and Welfare

* This final report exists also as the Ph.D. dissertation of the senior author; the junior author was principal investigator of the project.

FOREWORD

The study reported here is one of several related studies recently completed or nearing completion within the Instructional Media Research Unit, Purdue. These recent studies include the M.S. thesis of the present senior author, the dissertation of Jon B. Myers, and the nearly completed dissertation of Robert W. Heckman. The present study, plus the senior author's thesis and the Myers dissertation, employ correlational and multiple regression analysis in searching for those stimulus variables within verbal instructional material which are predictive of observed instructional effects. Typically, their dependent variables are indexes of learning gain that are associated with individual criterion items.

The first of these studies, the Myers dissertation, was generally successful in its use of quantified stimulus variables to predict posttest performances, employing data from both an "immediately" administered criterion measure and from the same measure administered (to a different randomly constituted group) following about one month of post-instruction delay; multiple correlations ranged from the mid-fifties to the mid-sixties. The present senior author's thesis seemed even more successful in its search for stimulus variables which could be combined to predict instructional effects. In that study, the subject matter was spelling and the multiple correlations ranged from the seventies to the mid-eighties.

Without additional work, however, it would have been unwise, to say the least, to emphasize the substantive results of these earlier studies. Each had its share of procedural problems and these problems required early consideration and, where possible, remedy. It was known, for example, that the studies incorporated more stimulus (predictor) variables than was desirable. Because of this, impressive levels of prediction might be initially developed, but might also fail in cross-validation. An important need was therefore to provide immediately for cross-validation of results, if not also to restrict the number of variables included. In the present study, cross-validation and some reduction in the number of variables have been provided.

There was a further difficulty stemming from the possible identification of stimulus variables which, while not "causing" the effects with which they correlated, were nevertheless predictive of effects, perhaps because of their correlation with obscured causal variables. This problem suggests the use of factor analytic methods as one means of reducing the many stimulus variables to a smaller set of principal stimulus dimensions. This, too, has been provided for within the present study; the stimulus matrix was factor analyzed and the orthogonal factor scores employed to predict the learning criteria.

Still another interest and concern stems from the possible existence of "aptitude-by-treatment" interactions. If present, they could contribute to the finding of relationships which would change as other changes were made, either in relevant aptitude characteristics of the learner group or in characteristics of the stimulus material. Although aptitude-by-treatment interactions have not been directly considered in the present study, they are the concern of the Heckman dissertation and are being investigated with many of the same data employed in the present study.

There are continuing problems also in the identification of the stimulus material that constitute an instructional "sequence" (i.e., the set of sentences, typically, which contribute to development or shaping of each criterion behavior); when this cannot be resolved, even the otherwise simple matter of determining frame count and sentence count becomes unreliable. Once a sequence is determined, further problems remain. In sets of hundreds or even thousands of sentences, as in the present study, it is no simple matter to identify accurately every appearance of concrete referent nouns, affixed words, prepositional phrases, and sequence prompts, to name just a few variables which may be sources of difficulty. For the present study, it is suspected that problems in the correct identification of sequence sentences and of indexing their contents are more acute than in the related and earlier studies. The senior author is endeavoring to re-examine the instructional material used in the study and to improve its analysis-indexing.

Since much of the prediction initially derived from predictor variables in this study failed in cross-validation, it might be argued that difficulties inherent in the procedures are too great and that it is time to return to more customary study designs. We do not see such argument as convincing and believe still that there are reasons for continuation of the work. First, the present efforts at prediction and the identification of functioning stimulus variables were not totally a failure; some prediction remained in cross-validation. Second, alternative designs for the study of functioning variables within complex stimulus fields typically make their own compromises and have their own problems; to shift over would not provide pure gain. Finally, there is the continuing appearance of speed and economy in the present procedures, if these can be suitably refined. This view is partially quoted on pages 2 and 3 of the present report. More completely, the quotation (from Donald Campbell and Julian Stanley, as cited) is:

> Such correlational data are relevant to causal hypotheses inasmuch as they expose them to disconfirmation. If a zero correlation is obtained, the credibility of the hypothesis is lessened. If a high correlation occurs, the credibility of the hypothesis is strengthened in that it has survived a chance of disconfirmation. To put the matter another way, correlation does not necessarily indicate causation, but a causal law of the type producing mean differences in experiments does imply correlation...the

relatively inexpensive correlational approach can provide a preliminary survey of hypotheses, and those which survive this can be checked through the more expensive experimental manipulation.

The problem and the procedures now seem clearer, even though most of the answers are not.

MES and WFS

TABLE OF CONTENTS

INTRODUCTION

This study was designed to identify characteristics of linear programmed instruction which are predictive of immediate and delayed learning effects. The study was sponsored by the Instructional Media Research Unit as part of its efforts to explore the relationships between and interactions among characteristics of instructional material and learners. The objective is a system in which instructors may identify "relevant" characteristics of an audience, then provide an instructional program incorporating characteristics to optimize learning.

The present study was designed to overcome certain criticisms of the published research on programmed instruction. One disquieting trend is the inattention to the complexities of instructional stimuli. The typical experimental design entails manipulation of one or two independent variables while presumably holding constant the multitude of other variables which may influence learning. When used for instructional research, this strategy appears deficient in several respects. First, since no comprehensive theory yet exists to guide research, experimenters too often have little basis for selecting variables to be manipulated. Secondly assuming that important variables are chosen, the experimenter must still select "dosage" levels within a crucial range. Finally, in many studies, control of variables not under investigation is more apparent than real. In fact, the multitude of variables which affect learning may virtually preclude the degree of control needed when manipulating independent variables. Illustrations of inadequate controls are numerous. Lumsdaine (1963, p. 627), for example, says the following about two studies of step size (Coulson and Silberman, 1959; Evans, Glaser, and Homme, 1959): "smaller and more numerous steps gave better learning scores, but these results...are difficult to interpret because the small step program (1) provided more practice on varied examples, and (2) took considerably more time."

The point is not that designs which consider only a very few variables are inappropriate. Rather, it is contended that "the means should be developed for measuring, or otherwise more exactly specifying, the variables in programmed materials..." (Holland, 1965, p. 92). It is believed that a multiple regression design may provide some of these "means." Not only does regression analysis yield quantitative statements of variables and their relationships, but it also readily accommodates virtually an unlimited number of independent variables, which yields additional advantages. First, if one views instructional research as still in an exploratory stage, then "...the relatively inexpensive correlational approach can provide a preliminary survey of hypotheses, and those which survive can be checked through the more expensive experimental manipulation" (Campbell and Stanley, 1963, p. 234). Secondly, multiple regression analysis minimizes the immediate need to control stimulus variables experimentally. Statistical controls inherent in such analyses provide at least a partial substitute. Thirdly, regression analysis can be extended to include factor analysis as a means of arranging stimulus variables into a parsimonious set of underlying factors. Finally, cross-validation designs may be employed to provide safeguards against premature identification of insignificant stimulus-effects relationships.

Programmed instruction research can also be criticized for the degree to which a few program variables dominate the literature to the exclusion of other promising variables. For example, the field of verbal learning has lately produced findings which suggest the importance of syntactic (i.e., sentence) and semantic (i.e., word) variables in verbal learning. While these developments are based largely on artificial laboratory tasks, they nevertheless appear relevant to programmed instruction. Several investigators have shown grammatical structure to be a significant factor in rote learning (Epstein, 1961, 1962; Coleman, 1965; Marks and Miller, 1964) and in paired associate learning (Glanzer, 1962; Martin, Davidson, and Williams, 1965). In commenting on such results, Epstein (1961) and Miller (1962) believe that syntax organizes words into efficient chunks for memorization. As Miller (1965) later wrote: "the syntactic structure of a sentence imposes groupings that govern the interactions between the meanings of the words in that sentence" (p. 18). Other experimenters have dealt with the contextual constraint effects of syntax. The problem investigated in such research is: given a sentence with a word deleted, does grammatical structure aid a subject in guessing the missing word. Aborn, Rubenstein and Sterling (1957) found that "the length, distribution, and grammatical structure of...(a sentence)...are all independently effective sources of constraint on words in sentences." Within the field of programmed instruction Markle (1964, p. 57) has described syntax as a "formal" prompting device.

Additional evidence for the role of syntax in written verbal instruction comes from studies of specific grammatical structures. Gray and Leary (1935), for example, report data indicating that simple sentences may be easier to comprehend than either compound or complex sentences. In a study of grammatical transformations and their relation to rote learning of sentences, Coleman (1965) concluded that two short coordinate clauses tended to be retained better than an equivalent long clause. "Apparently a person can process content morphemes packaged into two clauses more easily than he can process content morphemes packaged into a single clause" (pp. 340-341). Within a programmed instruction setting, Smith (1965) analyzed the relation of learning effects with certain grammatical structures which he called "multi-word modifiers," e.g., prepositional phrases, verbal nouns and adjectives. For a sample of young children, these variables appeared detrimental to learning. Partial support for this finding comes from Gray and Leary (1935) who report a negative correlation between prepositional phrases and reading comprehension but a positive correlation between infinitive phrases and comprehension. Miller (1951, p. 134) prudently notes that effects of sentence complexity may be confused with sentence length effects.

Along with syntactic factors, Marks and Miller (1964) noted that semantic factors were related to errors in memorizing sentences. For example, Myers (1964) in a programmed learning study and O'Connor (1950) in a film learning study both observed a relationship between frequency of affixes and learning effects. Smith (1965) also found a positive correlation between suffixes and learning gains. Yngve (1962) speculated that affixes reduce sentence complexity by replacing other words. On the other hand, Miller (1951, p. 134) theorized that "counting affixes is another way of estimating the proportion of hard words." Since affixes increase the average of syllables per word, it is natural to suppose a relationship between syllables and learning. Smith (1965) and Gray and Leary (1935), however, report conflicting findings. The former obtained a slight positive correlation between syllables per word and learning while the latter observed a negative correlation. Lastly, Smith (1965) also analyzed the effect of modifiers (adjectives, adverbs and possessive pronouns) and concluded that they facilitate learning of written instruction.

The literature on readability furnishes additional support for the notion that syntactic and semantic variables influence learning. The Flesch and most other readability formulae consist of a word variable (e.g., number of syllables per 100 words) and a sentence variable (e.g., number of words per sentence). In reviewing studies of reading comprehension as a function of reading ease, Klare (1963, pp. 134-135) concluded that "...variation of one factor (word or sentence) was not sufficient to provide increased comprehension, but that a combination of the two was." Comparative studies of readability measures suggest that cloze procedure may be the most sensitive measure of reading ease. (See, for example, Taylor, 1953). Greene (1965) prefers to think of cloze procedure as "a means of measuring the degree of correspondence between the language of a message and the language system of a reader" (p. 213). Another characteristic which makes cloze procedure attractive for programmed instruction research is its similarity to the experimental task of textual constraint studies. Furthermore, Weaver and Kingston (1963) in their factor analytic study of reading and language ability have shown that cloze procedure is closely correlated with a factor they call redundancy retrieval. "The underlying ability here seems to be the recognition of redundancy characteristics of language" (p. 259), a property one might hypothesize as related to the comprehension of programmed materials.

In addition to syntactic and semantic variables, other factors which bear investigation include incidental learning and response difficulty. The incidental presentation of responses to be learned is so common, one wonders about its role in programmed instruction. Likewise, little attention has been given to differences in the difficulty of the responses to be learned. Prior studies of the Instructional Media Research Unit have assumed for convenience that the responses to be learned were of equivalent difficulty. It is here hypothesized that the length of the response to be learned may provide a quantitative measure of the difficulty of the learning task.

Finally, many pay lip-service to assessing retention effects of instruction, but few do anything about it. An instructional treatment may produce relatively inefficient learning but greater resistance to forgetting than other methods. A prominent example is the use of variable-ratio reinforcement schedules in animal studies. Within programmed instruction, Krumboltz and Weisman (1962) compared the effects of overt and covert responding and found that "written responses to programmed materials were found to aid in the retention of new learning when measured after two weeks interval although an immediate test of learning failed to show any differences."

To summarize, this study employed a multiple regression design to assess the immediate and delayed learning effects associated with a number of characteristics of programmed instruction. The characteristics selected for study include variables previously studied within the context of programmed instruction and variables recently investigated in the areas of verbal learning and readability. The effects of incidental learning and of response difficulty were also selected for study.

PROCEDURE

Subjects

One hundred forty-four Purdue students who had participated in a prior study of the Instructional Media Research Unit served as Ss; they were paid volunteers recruited originally from freshman English courses. The Ss were divided into three groups: the experimental group (65 Ss), the control group (60 Ss), and the cloze score group (19 Ss). The experimental and cloze score groups were randomly selected from the randomly constituted experimental group of the prior study. Similarly, control Ss were drawn from the earlier control group. Through the course of the study four students dropped from the experimental group and five from the cloze score group; the study's analyses are thus based on the remaining Ns of 61, 60, and 14.

The median age for all groups was 18 years, and the sexes were distributed as follows: 31 males and 30 females in the experimental group, 35 males and 25 females in the control group, and five males and nine females in the cloze score group. More complete descriptions of the Ss are reported by Fincke (1967).

Program

The instructional program chosen for this study was Holland and Skinner's Analysis of Behavior (1961). This linear program consists of about 2200 frames devoted to various aspects of classical and operant conditioning. It was selected for several reasons. First, its horizontal format provides reasonable assurance that each S will be exposed to the same stimulus material under conditions of adequate stimulus control. Secondly, it is sufficiently long to permit an extensive criterion test. Lastly, it teaches material which few, if any, of the participants have encountered previously.

Criterion Test

Instructional effects were assessed by means of 183 constructed response items adapted from the program's review sets (chapters).[1] Appendix A contains the review frames which served as test items.[2]

---

[1]A 183 item matching test, virtually identical in content to the constructed response test, was also administered but was not analyzed, since total test scores on the two tests were highly correlated ($r = .95$).

[2]Nineteen (19) review items were omitted from the criterion test because they appear to test the same learning as other items. Four review frames were combined to form two test items. Three review items were inadvertently omitted from the criterion test.

Administration

The task of the experimental Ss was to complete the programmed text within a five week span. A reading room was made available to them for twelve hours a day, Monday through Thursday, for eight hours on Fridays, and for parts of three weekends. It was felt that the task was of such size that Ss should be free to study whenever they chose, with few restrictions. The principal restriction imposed was that they complete the program within the specified five week period. To ensure that all Ss read the text, they were required to record responses to each frame on specially provided answer sheets. The experimental Ss also received immediate and delayed administrations of the criterion test. To minimize recency effects, the immediate posttesting was done in four parts, with each part assessing the instructional effects of about one-fourth of the text. Like the instruction, testing was conducted on an individual basis. Three weeks after termination of the program and immediate posttest period, experimental Ss were provided with a two week period during which each was to complete the criterion test a second time. In contrast to the immediate posttest administration, Ss generally completed the delayed posttest administration in one sitting. Consequently, the average interval between instruction and delayed posttest varied among the four blocks of instruction. Table 1 presents the

Table 1

Means, Standard Deviations, and Ranges of the Interval between Instruction and Test Administrations

	Immediate Posttest			Delayed Posttest		
Subtest	Mean	Stan. Dev.	Range	Mean	Stan. Dev.	Range
17	1.66*	2.73	0 to 12	47.52	6.12	37 to 60
29	0.62	1.45	0 to 7	36.80	4.96	24 to 49
41	0.38	1.08	0 to 7	31.82	5.21	21 to 44
53	0.21	0.66	0 to 4	26.28	5.52	16 to 40

* Data are expressed in terms of days.

means, standard deviations, and ranges of the time intervals between program termination and test administrations.

Control Ss served only to take the criterion test. For them, the test administrations were also on an individual basis, as in the experimental group. Cloze score Ss received a specially prepared version of the Holland and Skinner text, one in which every fifth word had been deleted. Cloze score Ss were required to read the test and to supply the missing words. The purpose of administering this version of the text was to provide an empirical measure of the reading difficulty associated with various segments of the program. Study facilities and conditions for the cloze score Ss were identical to those for experimental Ss.

Criterion Variables

For each of the 183 test items the following indexes were computed:

(1) Immediate absolute gain score: the immediate posttest item difficulty index minus the item difficulty index derived from control group data.

(2) Delayed absolute gain score: the delayed posttest item difficulty index minus the item difficulty index derived from control group data.

(3) Immediate relative gain score (May and Lumsdaine, 1958, p. 11; McGuigan and Peters, 1965): the

immediate absolute gain score divided by the quantity 1.00 minus the control group item difficulty index. Relative gain score is merely observed gain (1 above) divided by maximum possible gain. It controls for the negative bias effect of prior knowledge (as estimated from control group data) on absolute gain.

(4) Delayed relative gain score: the delayed absolute gain score divided by the quantity 1.00 minus the control group item difficulty index.

(5) Immediate residual gain score (Du Bois, 1962, pp. 77-80; Rankin and Tracey, 1965; Wodtke, 1966): the immediate posttest item difficulty index (in standard score form) minus the product of the control group item difficulty index (in standard score form) and the Pearson correlation coefficient between the control group item difficulty index and the immediate posttest item difficulty index.[3] Residual gain score partials out the effects of prior knowledge from the posttest. Correlations between the residual score and predictor variables are essentially part correlations between the predictors and posttest values with prior knowledge held constant.

---

[3]In algebraic terms, the residual score formula is: $z_{2.1} = z_2 - z_1 r_{12}$, were z_1 stands for the control group item difficulty, z_2 for the posttest item difficulty, and r_{12} for the correlation of control group and posttest item difficulty indexes.

(6) Delayed residual gain score: the delayed posttest item difficulty index (in standard score form) minus the product of the control group item difficulty index (in standard score form) and the Pearson correlation coefficient of the control group item difficulty index and the delayed posttest item difficulty index.

Program Characteristics (Predictor Variables)

To understand how the independent variables were derived, one should be aware of the assumptions underlying their derivation. First, it was assumed that the terminal behaviors to be taught by the program would not be uniformly achieved, i.e., posttest item difficulty indexes and gain scores would exhibit inter-item variability. Secondly, it was assumed that the instructional program could be divided into sequences of frames, with each sequence serving to develop (or "teach") a criterion behavior. Further, it was assumed that these frame sequences could be described in terms of quantifiable characteristics which exhibit inter-sequence variability. Finally, it was assumed that item difficulty indexes (and gain scores) vary, at least partially, as a function of the variability among the frame sequence characteristics.

The first step in the procedure was to identify the frames intended to teach the correct response to each criterion item. The frame sequence for test item 1702a, for example, consisted of the 25th and 28th frames of Set 8 and the 15th frame of Set 10. Correspondence with the senior author of the programmed text[4] revealed that there were no existing records which could serve in determining the instructional frames associated with each test frame response. Thus, it was the task of the present author to review the entire text and to decide which instructional frames would constitute each frame sequence. Reliance was placed upon such cues as the appearance of the key term in a frame (particularly as a response) and the set to which readers are referred for review of that item. The identification process was further complicated by teaching frames which appeared _after_ the test items to which they related. For example, the concept of "stimulus generalization" is taught in Set 22 and tested by item 2916a. Later, in Set 52, the concept of "transference" is taught as a special case of stimulus generalization. Quite naturally, stimulus generalization is briefly reviewed at this point in the program. Instances such as this could not help a _S_ answer 2916a on the immediate posttest but very possibly affected delayed posttest performance. Thus, for delayed posttest analyses, a number of frame sequences were augmented by such late appearing frames. Nevertheless, for

[4]Holland, J. G. Personal Communication, April, 1966.

immediate posttest analyses, frame sequences ranged from one to 34 frames with a median of eight frames; while in delayed posttest analyses, sequences ranged from one to 35, also with a median of eight frames. The frame sequence for each criterion item is listed in Appendix B.

Once the frame sequences were determined, the next step was to index the 37 program characteristics selected for investigation. Thirty-two (32) variables were characteristics of the frame sequence per se; three were characteristics of the test item; and two described the incidental presentation of the "key term," i.e., the word or words which constituted the answer to the criterion test item. Each independent variable is defined in Appendix D. Also included in Appendix D are the mean and standard deviation of each program characteristic obtained from the immediate posttest data.

Since cataloguing most of the predictors was highly judgmental, interrater reliability coefficients were obtained for 24 variables. The reliability study was based upon 25 frame sequences randomly selected. These coefficients appear in Appendix D.

Since the actual values of the program characteristics assigned to each frame sequence are not presented in this report[5], it may be informative to illustrate the form

---

[5]The actual values assigned to each test item on these variables may be obtained from the Instructional Media Research Unit, Purdue University, Lafayette, Indiana.

of the raw data by returning to the example of item 1702a and its frame sequence. Within this three-frame sequence, the typical sentence had 17.4 words, of which 1.67 were responses, 5.2 were modifiers, etc. On the average, each word in this sequence had .33 affixes and 1.77 syllables. There were no references to an exhibit in this sequence. The key term, "operant," was required twice as a response and appeared twice in the stimulus portion of the frames.

Statistical Analysis

The 183 test items were divided randomly into two item samples, one of 91 and the other of 92 items. The following operations were then carried out: multiple regression analyses, double cross-validations, and factor analysis.

Multiple Regressional Analyses. For each item sample, intercorrelations among the 37 program variables and the criterion gain scores were generated. Stepwise multiple regression analyses were computed using the Weighted Regression Analysis Program (WRAP). The WRAP performs regression analyses using a matrix inversion routine similar to the Dwyer square root method. This method first computes the multiple regression with all independent variables included, then proceeds to delete variables one at a time, removing the variable which contributes least to the multiple for the particular set of variables still retained. This technique is sometimes called the "tear down" method. All told, twelve regression analyses were computed: two item samples times three gain scores times two test administrations.

Cross-Validations. Double cross-validations were executed. That is, the raw score regression weights determined in Sample 1 and Sample 2 were used to predict the observed criterion index in the other sample. This procedure was repeated for each gain score and each administration, i.e., six double cross-validations were executed. Cross-validations were limited to those independent variables which appeared to be most predictive of the criterion indices in the original sample.

Factor Analysis. Using all 183 items, intercorrelations were generated among the 37 independent variables. A principal components factor analysis was executed using image covariance estimates of the communalities. The image covariance estimate is the square of the multiple correlation coefficient between each variable and all the remaining variables, with appropriate adjustment of the off-diagonal elements to maintain the positive semi-definiteness of the matrix (Guttman, 1953). Fourteen factors were extracted, and nine were rotated to simple structure according to the varimax criterion (Kaiser, 1958).

RESULTS

Original Analyses

Table 2 presents the means and standard deviations derived from the three criterion test administrations; these statistics are based on the original full set of 207 criterion items. (The principal analyses were based on the responses to 183 items, summed by item rather than by subject.) As Table 2 shows, The Analysis of Behavior text produced substantial learning gains which were retained virtually intact until the delayed posttest administration.

Table 3 presents the means and standard deviations for the control group item difficulty indexes, the immediate and the delayed posttest item difficulty indexes, the criterion gain indexes and the intercorrelations among these variables. Of the three gain indexes, the residual gain score is the most desirable for three reasons. First, in contrast with absolute gain, there is no regression effect associated with residual gain, i.e., no spurious negative correlation between residual gain and initial status. Secondly, among the three gain indexes, residual gain produced the highest intercorrelations with the other gain scores and the posttest item difficulty indexes (.88 versus .77 for absolute gain and .84 for relative gain). Finally, residual gain had

Table 2

Means and Standard Deviations of Subjects' Total Test Scores for the Control, Immediate Posttest and Delayed Posttest Administrations of the Criterion Test[1]

	Mean	Standard Deviation
Control	34.40	40.10
Immediate Posttest	133.31	36.13
Delayed Posttest	129.62	40.82

[1]Based upon 207 items.

Table 3

Intercorrelations, Means and Standard Deviations of the Control Group Item Difficulty Index, the Immediate and Delayed Posttest Item Difficulty Indexes, and the Criterion Gain Indexes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Mean	S.D.
(1) Control Group item difficulty index	1.00*	.44	.52	-.44	-.36	-.04	-.10	-.01	.00	.16	.19
(2) Immediate posttest item difficulty index		1.00	.84	.61	.52	.81	.72	.89	.72	.63	.21
(3) Delayed posttest item difficulty index			1.00	.39	.61	.61	.84	.68	.85	.61	.22
(4) Immediate posttest absolute gain index				1.00	.83	.84	.63	.90	.72	.47	.21
(5) Delayed posttest absolute gain index					1.00	.70	.82	.75	.93	.45	.20
(6) Immediate posttest relative gain index						1.00	.83	.91	.74	.56	.26
(7) Delayed posttest relative gain index							1.00	.75	.91	.55	.25
(8) Immediate posttest residual gain index								1.00	.80	0.00	.90
(9) Delayed posttest residual gain index									1.00	0.00	.85

*Correlations greater than .34 are significant at the .001 level of confidence.

larger standard deviations than either absolute or relative gain.

Inter-rater reliability coefficients were obtained for 24 of the 37 predictors. The reliabilities ranged from .18 to 1.00 with a median of .73. These coefficients appear in Appendix D. Appendix E lists the intercorrelations among the independent variables. Appendixes F and G present the validity coefficients obtained in the immediate and delayed posttest analyses.

Multiple Regression Analyses. Tables 4 through 7 present the multiple regression coefficients for the twelve regression analyses. Each table contains the *R* with all 37 predictors, the cumulative *R*s for the predictors retained latest in the "tear-down" and their associated validity coefficients and standard partial regression coefficients (beta weights). The predictors included in the tables were selected (a) because retaining another predictor increased the *R* by less than three correlation points or (b) because the next predictor to be retained was the random variable.

Using all 37 predictors, the mean *R*s for the immediate and delayed posttests were .69 and .70, respectively. Similarly, the average *R*s for the absolute gain index, the relative gain index, and the residual gain index were .70, .69, and .71, respectively. For Sample 1 the average *R* was .62, but for Sample 2 it was .76. These last two *R*s were significantly different at the .07 level.

Table 4

Cumulative Rs, Validity Coefficients, and Beta Weights for Independent Variables Related to Immediate Posttest Criterion Indexes (Sample 1)

	Number of Variables	Variable	Cumulative R	Validity	Beta Weight
Absolute gain	1	24	.27	-.27	-.224
	2	13	.36	.19	.357
	3	35	.49	-.21	-.331
	4	33	.52	.15	.276
	All 37		.66		
Relative gain	1	22	.24	.24	.367
	2	31	.31	.20	.230
	3	33	.36	-.23	-.214
	4	23	.39	.08	-.423
	5	34	.42	-.01	.261
	All 37		.62		
Residual gain	1	33	.21	-.21	-.257
	2	13	.31	.17	.295
	3	31	.38	.22	.240
	4	35	.41	-.08	-.173
	All 37		.60		

Table 5

Cumulative Rs, Validity Coefficients, and Beta Weights for Independent Variables Related to Immediate Posttest Criterion Indexes (Sample 2)

	Number of Variables	Variable	Cumulative R	Validity	Beta Weight
Absolute gain	1	7	.34	.34	.503
	2	10	.42	-.06	-.305
	3	8	.49	.24	.273
	4	37	.54	.19	.287
	5	28	.56	.16	.318
	6	17	.61	.06	.309
	All 37		.76		
Relative gain	1	7	.28	.28	.313
	2	37	.39	.28	.360
	3	17	.43	.06	.310
	4	28	.48	.12	.259
	All 37		.72		
Residual gain	1	7	.28	.28	.337
	2	37	.40	.27	.341
	3	28	.43	.17	.315
	4	17	.50	.03	.303
	All 37		.76		

Table 6

Cumulative Rs, Validity Coefficients, and Beta Weights for Independent Variables Related to Delayed Posttest Criterion Indexes (Sample 1)

	Number of Variables	Variable	Cumulative R	Validity	Beta Weight
Absolute gain	1	33	.28	-.28	-.302
	2	13	.35	.15	.300
	3	35	.41	-.16	-.256
	4	31	.47	.21	.243
	All 37		.60		
Relative gain	1	33	.32	-.32	-.335
	2	31	.42	.30	.290
	3	24	.47	.12	.202
	All 37		.65		
Residual gain	1	33	.30	-.30	-.321
	2	31	.40	.27	.265
	3	13	.44	.12	.196
	All 37		.61		

Table 7

Cumulative Rs, Validity Coefficients, and Beta Weights for Independent Variables Related to Delayed Posttest Criterion Indexes (Sample 2)

	Number of Variables	Variable	Cumulative R	Validity	Beta Weight
Absolute gain	1	33	.43	-.43	-.272
	2	37	.47	.19	.254
	3	28	.49	.21	.301
	4	17	.53	.08	.298
	5	7	.58	.29	.398
	6	10	.63	-.09	-.275
	All 37		.77		
Relative Gain	1	33	.34	-.34	-.311
	2	19	.37	-.09	-.551
	3	18	.42	.08	.534
	4	20	.49	-.06	-.343
	5	14	.53	-.30	-.214
	All 37		.74		
Residual gain	1	33	.45	-.45	-.394
	2	10	.49	-.14	-.352
	3	7	.54	.25	.302
	4	37	.59	.23	.291
	All 37		.81		

Cross-Validations. The variables selected for cross-validation are listed in Table 8. Each set of predictors was validated against the same criterion index obtained in the other sample. Thus, there were six double cross-validations, i.e., three gain indexes times two test administrations. Table 1 presents the number of variables cross-validated, the cross-validation Rs, and the Rs obtained in the multiple regression analyses with the predictors used in cross-validation. Four cross-validation Rs were significant. All four were associated with the delayed posttest and appeared to be largely the result of variable 33's correlation with the criterion indexes. Variable 33 is the rank order in which the initial learning frames first appear in the programmed text.

Factor Analysis. Twenty-one factors were extracted and nine factors, accounting for 82.99% of the variance, were rotated orthogonally. The factors were designated as: (1) quantity of instruction, (2) sentence complexity, (3) concrete versus abstract instruction, (4) review, (5) cloze score, (6) key term, (7) word complexity, (8) largest frame, and (9) unnamed. Tables 9 through 17 list the variables which load .40 or more on each of the factors. Appendix H presents all rotated factor loadings.

"Quantity of instruction" included such variables as total number of frames and number of initial learning frames which reflect the sheer size of frame sequences. Factor 2,

Table 8

Multiple Regression and Cross-Validation Rs for Selected Predictors

Criterion Index	Sample 1 Weights			Sample 2 Weights		
	Number of Variables	Multiple Regression R	Cross Validation R	Number of Variables	Multiple Regression R	Cross Validation R
Immediate Post-test Absolute Gain Index	4	.52**	.15	6	.61**	.32
Immediate Post-test Relative Gain Index	5	.42**	.17	4	.48**	.28
Immediate Post-test Residual Gain Index	4	.41**	.15	4	.50**	.28
Delayed Post-test Absolute Gain Index	4	.47**	.22	6	.63**	.35*
Delayed Post-test Relative Gain Index	3	.47**	.34**	5	.53**	.28
Delayed Post-test Residual Gain Index	3	.44**	.40**	4	.59**	.33*

* Significant at the 5% level.
** Significant at the 1% level.

Table 9

Variable Loadings on Factor 1: Quantity of Instruction

Variable	Loading
1. Total Frames	-.87
2. Initial Learning Frames	-.96
3. Review Frames	-.41
4. Exhibit Referrals	-.47
6. Key Term as Stimulus	-.71
7. Key Term as Response	-.43
10. Formal Prompts	-.53
11. Thematic and Sequence Prompts	-.84
19. Maximum Cloze Score	-.46

Table 10

Variable Loadings on Factor 2: Sentence Complexity

Variable	Loading
26. Sentences per Frame	-.71
30. Modifiers per Sentence	.72
32. Verbals per Sentence	.44
35. Prepositional Phrases per Sentence	.40
36. Clauses per Sentence	.74
37. Words per Sentence	.86

Table 11

Variable Loadings on Factor 3: Concrete versus Abstract Instruction

Variable	Loading
14. Affixed Word Ratio	-.41
17. Concrete Noun Ratio	.48
25. Syllables per Word	-.55
27. Applications per Frame	.88
28. Rules per Frame	-.89

Table 12

Variable Loadings on Factor 4: Review

Variable	Loading
1. Total Frames	-.41
3. Review Frames	-.79
7. Key Term as Response	-.56
22. Intentional Sets	-.68
23. Vehicular Sets	-.71
33. Complete Program Order	.43
34. Vehicular Uses of Key Term	-.49

Table 13

Variable Loadings on Factor 5: Cloze Score

Variable	Loading
18. Cloze Score	-.81
19. Maximum Cloze Score	-.58
20. Minimum Cloze Score	-.56

Table 14

Variable Loadings on Factor 6: Key Term

Variable	Loading
8. Letters in Key Term	-.87
9. Syllables in Key Term	-.88

Table 15

Variable Loadings on Factor 7: Word Complexity

Variable	Loading
14. Affixed Word Ratio	-.55
25. Syllables per Word	-.49
34. Vehicular Uses of Key Term	.44

Table 16

Variable Loadings on Factor 8: Largest Frame

Variable	Loading
12. Maximum Words per Frame	-.78
13. Maximum Syllables per Frame	-.63

Table 17

Variable Loadings on Factor 9: Unnamed Factor

Variable	Loading
16. Personal Word Ratio	-.49
35. Prepositional Phrases per Sentence	.65

"sentence complexity," consisted of various grammatical variables such as prepositional phrases and clauses. The highest loading variable, the number of words per sentence, merely indicates that the more complex sentences tend to be the longer sentences. Variable 26, the number of sentences per frame, had a high negative correlation with Factor 2. Apparently, the more sentences in a frame, the less complex these sentences tend to be. Factor 3 was labeled "concrete versus abstract instruction" because it represented the dominance of either abstract generalizations (variable 28) or concrete examples (variable 27) within a frame sequence. Factor 4 reflected the "review" dimension of instruction and included not only intentional review frames (variable 3) but also incidental review frames (variables 23 and 34). Variable 7, the number of appearances of the key term as a response, was included primarily because this variable was the principal cue for identifying review frames. The "cloze score" factor is defined by the three cloze score variables. Factor 6 consisted of only the two quantitative characteristics of the key term. Factor 7 contained two characteristics of words: affixes and syllables. The high loading of variable 34, the number of vehicular uses of the key term, on Factor 7 is difficult to interpret. Since vehicular appearances of the key term were negatively related to the order in which the key terms are taught, its correlation with Factor 7 may indicate that more complex words predominate in the frame sequences that appear later. The "largest frame" factor was defined by two characteristics of the largest frame within each frame sequence. Evidently this factor reflects the upper bound either of the quantity of information presented in one frame or of the complexity of instruction within a frame sequence. No interpretation of the ninth factor will be attempted. Considering the number and definition of variables loading on each, only the first four factors appear to be generalizable beyond the present study.

Post Hoc Analyses

Further analyses were undertaken to explore ways of increasing prediction. These analyses included (1) prediction using dichotomized independent variables, (2) prediction using factor scores, and (3) deviant case analyses.

Dichotomized Independent Variables. A number of predictor variables were sufficiently skewed to raise questions concerning assumed linear regression between these variables and the criterion variables. Consequently, every predictor with either (1) at least fifty per cent of the cases occurring within the upper or lower twenty per cent of the range, or (2) the largest frequency occurring in either of the extreme values, was dichotomized at approximately the median. The variables dichotomized were 2, 3, 4, 6, 7, 10, 11, 16, 17, 21, 22, 23, 24, 27, 28, 29, 31, and 34. Since these dichotomies are artificial, tetrachoric correlations were employed for pairs of the dichotomized variables, while biserial correlations were employed to correlate dichotomized variables with continous variables. The resulting tetrachoric and biserial correlations, plus the Pearson correlations among continuous variables, were merged into one matrix using the "Smerge" operation of the P-Stat System (Buhler, 1964). Since no computer program could be found to execute a stepwise multiple regression analysis from a correlation matrix, the Jenkins short-cut method (Lawshe and Balma, 1966, pp. 342-352) was used to select predictors for cross-validation. A more exact $\underline{R}$ and beta weights were then obtained using the "Multr" program of the P-Stat System (Buhler, 1964). This procedure was carried out for the immediate posttest data. In all, six multiple regression analyses were computed (i.e., two item samples times three gains scores). The obtained regression weights were then cross-validated.

Table 18 presents the predictors selected for the regression analyses, their validities and beta weights, and the $\underline{R}$ for Sample 1. Table 19 presents the corresponding information for Sample 2. The cross-validation $\underline{R}$s are shown in Table 20. While the multiple regression $\underline{R}$s tended to be slightly higher than those obtained in the original analyses, none of the cross-validations were significant.

Factor Scores. Thus far, this study has provided large

Table 18

Rs, Validity Coefficients, and Beta Weights Derived from 19 Continuous and 18 Dichotomized Independent Variables (Sample 1)

Criterion Index	Variables	Validity	Beta Weight	R
Absolute Gain	13	.19	.275	.46
	35	-.21	-.227	
	33	-.20	-.204	
	7	.36	.182	
Relative Gain	29	.31	.451	.56
	33	-.23	-.370	
	12	.18	.208	
	22	.19	.116	
	27	.19	.182	
Residual Gain	29	.25	.375	.50
	33	-.21	-.355	
	12	.21	.264	
	27	.16	.165	

Table 19

$\underline{R}$s, Validity Coefficients, and Beta Weights Derived from 19 Continuous and 18 Dichotomized Independent Variables (Sample 2)

Criterion Index	Variables	Validity	Beta Weight	R
Absolute Gain	2	-.34	-.070	.64
	7	.20	.416	
	16	.23	.234	
	12	.20	.372	
	11	-.20	-.741	
	10	.13	.225	
Relative Gain	37	.28	.327	.47
	7	.15	.390	
	2	-.23	-.347	
	36	.24	-.040	
Residual Gain	16	.28	.237	.49
	37	.27	.217	
	12	.24	.291	
	2	-.23	-.213	

Table 20

Multiple Regression and Cross-Validation *R*s for 19 Continuous and 18 Dichotomized Independent Variables

Criterion Index	Number of Variables	Multiple Regression R	Cross-Validation R	Number of Variables	Multiple Regression R	Cross-Validation R
Absolute Gain	4	.46**	.15	6	.64**	.28
Relative Gain	5	.56**	.18	4	.47**	.16
Residual Gain	4	.50**	.20	4	.49**	.15

** Significant at the .01 level.

initial multiple correlations but mostly non-significant Rs in cross-validation. When one considers the large number of predictors (37), relative to the number of observations (91 in one sample, 92 in the other), the following remarks of Horst (1965) seem particularly cogent:

> "...one of the chief difficulties encountered in the classical multiple regression technique is with reference to the problem of overfitting, or of degrees of freedom....the classical methods break down when, in a particular sample, we have as many or more attributes as entities.
>
> Even though the number of entities may be somewhat greater than the number of attributes, we may still get into difficulty because of not having enough degrees of freedom. A matrix of regression vectors for transforming a data matrix of predictor measures to an estimated matrix of criterion measures may give very poor results on another sample, simply because we have capitalized on random variance and covariance components of the predictor variables (p. 551).

These comments preface Horst's discussion of prediction based upon factor scores. The rationale for such a procedure "...is based on the assumption that with a small number of factors as compared to the number of predictor variables, one can reproduce the matrix of predictor measures for the sample with sufficient accuracy by the major product of the factor score and factor loadings matrices" (p. 551). Furthermore, it is assumed that the factor score matrix represents only the "reliable" variance, i.e., "that part of the predictor scores which is free from error or random variation."

> The traditional regression matrix may be regarded as consisting in part of this random variation. But

> suppose one uses the factor score matrix obtained from the predictor measures as a basis for estimating the criterion measures in the sample. The regression matrix calculated from the factor score matrix may be assumed to be free of the random error in the predictor part of the matrix, and therefore it should be more efficient when applied to other samples (Horst, 1965, p. 552).

In accord with Horst, a new set of predictions was undertaken and based on the nine rotated factors mentioned earlier. Nine standardized factor scores were computed for each frame sequence using the "short method" described by Harman (1960, pp. 349-356). "Tear-down" multiple regression analyses were then carried out for each of the immediate posttest criterion variables. The regression weights of the factors retained latest were cross-validated. Table 21 presents the factors retained latest in the tear-down, their validities and beta weights, and the cumulative *R*s. Table 22 compares the multiple regression *R*s and cross-validation *R*s for each set of factors.

The multiple regression *R*s are lower than those obtained in the original analyses. However, the cross-validations of Sample 1 weights are all significant.

Deviant Case Analyses. The frame sequences whose predicted criterion scores ($\hat{y}$) deviated most from their obtained criterion scores (y) were examined in an attempt to identify uncontrolled sources of variation. For this purpose, the two item samples were combined, a multiple regression analysis was executed against the immediate posttest residual

Table 21

Cumulative Rs, Validity Coefficients, and Beta Weights for Factors Related to Each Criterion Index

Criterion Index	Number of Factors	Factor	Cumulative R	Validity	Beta Weight
Absolute Gain (Sample 1)	1	8	.22	-.22	-.237
	2	9	.30	-.21	-.199
	4	4	.35	-.15	-.182
	all 9		.40		
Relative Gain (Sample 1)	1	4	.27	-.27	-.303
	2	8	.33	-.13	-.150
	3	2	.34	.10	.122
	all 9		.36		
Residual Gain (Sample 1)	1	4	.21	-.21	-.226
	2	8	.31	-.20	-.227
	all 9		.34		
Absolute Gain (Sample 2)	1	8	.31	-.31	-.302
	2	2	.38	.17	.206
	3	6	.42	-.20	-.203
	4	4	.47	-.25	-.202
	all 9		.49		
Relative Gain (Sample 2)	1	2	.26	.26	.287
	2	8	.34	-.23	-.266
	3	4	.40	-.18	-.114
	all 9		.43		
Residual Gain (Sample 2)	1	8	.30	-.30	-.325
	2	2	.42	.24	.282
	all 9		.46		

Table 22

Multiple Regression and Cross-Validation Rs for Selected Factors

Criterion Index	Number of Factors	Multiple Regression R	Cross-Validation R	Number of Factors	Multiple Regression R	Cross-Validation R
Absolute Gain	3	.35**	.33*	4	.47**	.26
Relative Gain	3	.34**	.32*	3	.40**	.22
Residual Gain	2	.31**	.36**	2	.42**	.24

* Significant at the .05 level.

** Significant at the .01 level.

gain index, and 23 cases with the largest $y-\hat{y}$ differences were selected for closer scrutiny. Of the 23, 18 cases were over-predictions; i.e., $\hat{y}$ was larger than y, and five were under-predictions.

Group comparisons were attempted first. After reviewing the test items and their corresponding frame sequences, several new variables were identified as potential sources of $y-\hat{y}$ differences: (A) number of other key terms tested in same test frame, (B) percent of key terms which are technical terms, (C) number of key terms taught at same time, and (D) number of appearances of principal test item prompt within frame sequence. Mean values on each new predictor were computed for the 18 over-predicted cases, the five under-predicted cases, and all 183 cases. These values appear in Table 23.

Only variables B and C seem to merit comment. Variable C was computed by first identifying the set to which readers are directed for review of the key term in question, then determining how many other key terms are also reviewed in the same set. The number of extreme cases is too small for the observed differences to be significant; nevertheless, the trend among the means suggests that the <u>greater</u> the number of key terms taught at approximately the same point in time, the <u>greater</u> the probability of the different frame sequences interfering with each other.

Variable B is a characteristic of the criterion test item per se. A "technical" term is here regarded as one whose meaning is specifically defined by the program's authors. "Non-technical" terms are those which are used by the program's authors as one would use them in normal or "everyday" writing and speaking. Such a variable may function as a crude index of Ss' familiarity with the key terms. The data in Table 23 suggest the faint but interesting possibility of a curvilinear relationship between variable B and the criterion index: learning gains are more accurately predicted for non-technical key terms than for technical key terms. Table 24, which presents the distribution of $y-\hat{y}$ differences for technical and for non-technical key terms, lends some additional support to this hypothesis. It should be emphasized though that the difference in predictive accuracy is slight. Hopefully, the differences would be greater with a less crudely indexed variable.

The next phase of the deviant case analyses involved examining the 23 deviant cases, individually. This examination is summarized in Table 25, which presents the standardized $y-\hat{y}$ difference score, key term, immediate posttest item difficulty index, most common responding errors and their frequency of occurrence, and comments for each of the 23 most deviant cases. It is impossible to identify all the extraneous sources of criterion variation, but three are immediately apparent. For instance, proactive inhibition effects are apparent from the error responses to items 1730b

Table 23

Mean Values on Five New Predictors Computed for the 18 Most Over-Predicted Cases, the Five Most Under-Predicted Cases, and All 183 Cases

Variable	18 Over-Predicted Cases	5 Under-Predicted Cases	All 183 Cases
(A) Number of other key terms tested in same test frame	1.11	1.20	1.13
(B) Percent of key terms which are technical terms	.83	.80	.64
(C) Number of key terms taught at same time	4.89	3.60	4.24
(D) Number of appearances of principal test item prompt within frame sequence	1.89	1.40	-

Table 24

Distribution of Y - Y Differences for Technical and Non-Technical Key Terms

	Y - Y Differences				
	+2 SD	+1 SD	M	-1 SD	-2 SD
Technical Key Terms (N=117)	0%	17%	63%	14%	6%
Non-Technical Key Terms (N=66)	0%	15%	75%	8%	2%

Table 25

Standardized $Y - \hat{Y}$ Difference Scores, Key Terms, Immediate Posttest Item Difficulty Indexes, Most Common Errors and Their Frequencies, and Comments for the 23 Most Deviant Cases

Test Item	$\frac{Y - \hat{Y}}{SD_{Y-\hat{Y}}}$	Key Term	Item Difficulty	Most Common Errors and Frequencies	Comments
1706c	1.78	Smooth Muscles	.95	-	Important information presented in exhibit and, hence, not adequately indexed.
1730b	-2.71	Response	.39	Respondent (.23) Operant (.11)	Discrimination between operant and response proactively inhibited by discrimination between operant and respondent.
1731a	-1.32	Fear	.39	Lying (.15)	
1733a	-1.54	Reinforcer	.25	Reinforcement (.26)	Discrimination between reinforcer and reinforcement not completely developed; ambiguous test item.
1736a	-1.67	Generalized Reinforcer	.26	Various types of stimuli (.30) Various types of reinforcers (.21)	Too many frames may have been included in the frame sequence.
1738a	-2.?2	Adaptation	.02	Forgetting (.70)	Discrimination between extinction and adaptation proactively inhibited by discrimination between extinction and forgetting.

Table 25 (continued)

Test Item	$\frac{Y - \hat{Y}}{SD_{y-y}}$	Key Term	Item Difficulty	Most Common Errors and Frequencies	Comments
1739a	1.76	Common Elements	.89	-	Recency effect: 1739a is last key term taught before first part of immediate posttest. Item difficulty drops to .44 on delayed posttest.
2914a	-1.86	Forms	.15	Other terms taught in same set (e.g., repertoire, fields) (.20)	Discrimination between form and other terms not completely developed.
2918b	1.76	Extinguish	.89	-	None.
2920a	1.92	(Water) deprivation	.89	-	None.
2922a	-1.43	Fixed-interval	.51	Variable-interval	Matching item involving graphs. Since graphs are presented in exhibits only, important information was not adequately indexed.
2922c	-1.56	Fixed-ratio	.41	Fixed-interval(.25) Variable-ratio(.23)	Same as 2922a.
2922d	-1.38	Variable-ratio	.24	Fixed-interval(.24)	Same as 2922a.

Table 25 (continued)

Test Item	$\frac{Y - \hat{Y}}{SD_{y-\hat{y}}}$	Key Term	Item Difficulty	Most Common Errors and Frequencies	Comments
2927a	-2.30	Differential (reinforcement)	.07	Conditioned (25) Discriminating (.15) Necessary (.13)	The reinforcement described in the test item is described as "conditioned differential" within the frame sequence; however, the conditioned aspect is not the critical property. The objective described in the text is discriminating subtle differences among stimuli to be imitated, and the means by which this objective is attained is differential reinforcement. Some Ss apparently confused the two terms.
4118a	-1.80	Aversive	.25	Punished (.15)	Confusion of terms, an aversive stimulus is presented in punishment.
4118b	-2.0[illegible]	Reinforcing	.16	Reinforced (.2[illegible])	Confusion of terms.
4121a	-1.93	Is	.56	Is not (cannot) (.41)	84% of the control group answered this item correctly. Item appears "too easy" to Ss who have read program.

Table 25 (continued)

Test Item	$\frac{Y - \hat{Y}}{SD_{y-\hat{y}}}$	Key Term	Item Difficulty	Most Common Errors and Frequencies	Comments
5302a	-2.32	Algebraic Summation	.12	Indecision (and synonyms) (.23) Avoidance (.15)	Key term always prompted by two blank answer spaces (indicating two words) within frame sequence. This prompt was omitted in the test item.
5303b	1.86	Incompatible	.87	-	None.
5308b	-1.98	Reinforcing	.16	Positive (.67)	Scoring problem. The word "positive" is presented as an acceptable synonym within the text.
5308c	-2.09	Aversive	.13	Negative (.67)	Similar to 5308b.
5316a	-2.03	Isolate	.33	Control (.25) Interpret (.18)	Purpose of experimental control confused with the goals of science.
5324d	-2.83	Law	.03	Respondent (.11)	The probability of answering 5324d correctly depends on answering 2324a, b, c correctly.

and 1738a. With regard to the latter item, learning to discriminate "forgetting" from "extinction" presumably interfered with learning to discriminate "adaptation" from "extinction." Similarly with item 1730b, learning the "operant"-"respondent" discrimination inhibited learning the "operant"-"response" discrimination. Error variance may also be attributed to identifying inaccurately the relevant instruction for each key term. A substantial portion of the instruction for such key terms as 2922a, c, and d, is presented in the exhibits which, unfortunately, were excluded from computing most of the frame sequence characteristics. The characteristics of test items per se constitute a third source of uncontrolled variation. Heretofore, it had been tacitly assumed, for convenience, that all test items were equivalent. Obviously, they are not. A few of the items are poorly written, e.g., 1733a, 4121a. A more complex problem involves criterion test items which are presented together in one frame. Among such items, the former items generally serve as prompts for the latter. Examine the following frame:

> We sometimes speak of cause and effect. A cause is a(n) (1)______ variable, and an effect is a(n) (2) ______ variable. The relation is a(n) (3)______ ______ or a(n) (4)______.
>
> Answers: (1) independent, (2) dependent, (3) functional relation, (4) law

The immediate posttest item difficulties of the four blanks were: .61, .61, .38, and .03. No S who missed 5324a answered 5324b correctly. Only one person who missed 5324b answered 5324c correctly. Only two Ss answered 5324d correctly, and both individuals had answered the other three items correctly; it is inferred that items which are sequentially dependent upon prior items are less likely to yield higher rates of correct responding upon posttest.

Both the group comparisons and the examination of individual cases suggest that criterion test item characteristics may moderate the relationships between frame sequence characteristics and criterion gains. To explore this possibility further, the test items were sorted into six rationally determined categories: (1) given an example, name the class to which it belongs; (2) given a set of conditions, complete the conclusion; (3) given a term, complete its definition; (4) given a definition, name the term defined; (5) given an effect, name the cause; and (6) rote fact or miscellaneous. Table 26 presents the number of items, mean residual gain score, and the distribution of $y-\hat{y}$ differences associated with each category. Although no tests of significance were performed, the data encourage further investigation of such item subgroups.

Table 26

Number of Items, Mean Residual Gain Score, and Distribution of $Y - \hat{Y}$ Difference Scores Associated with Six Categories of Criterion Test Items

Test Item Category	N	Mean Residual Score	Difference Scores				
			Under-prediction			Over-prediction	
			+2 SD	+1 SD	M	-1 SD	-2 SD
1	25	0.131	0%	20%	68%	4%	8%
2	35	0.197	0%	20%	71%	9%	0%
3	31	0.011	0%	16%	74%	10%	0%
4	26	-0.345	0%	12%	58%	19%	12%
5	33	-0.001	0%	21%	58%	15%	6%
6	33	-0.040	0%	6%	79%	12%	3%
All Items	183	0.00	0%	16%	68%	12%	4%

DISCUSSION

This study was undertaken with the intention that it contribute to understanding of programmed instruction characteristics and how they relate to learning. It was partially successful, not only with respect to the intended examination of stimulus-and-effects relationships, but also in the better identification of methodological problems. Within the present section, discussion of the principal results will be reserved until the last, with earlier consideration given to the methodological problems; to novel analysis procedures, including those employed in the study's post hoc analyses; to new variables which may merit consideration in future work; and to results of a factor analysis of the indexed stimulus variables. Recommendations will be made with respect to the following problems: (1) degrees of freedom, (2) skewness, (3) heterogeneous criterion tasks, and (4) allocation of frames to sequences.

Methodological Problems

Problem of Degrees of Freedom. Contrary to the cautions found in most basic statistics texts, the "degrees of freedom" problem was ignored with the consequence of dramatic shrinkages in cross-validations. The least squares solution capitalizes upon any chance deviations which favor high multiple correlation. Thus, the multiple correlation coefficient is an inflated estimate of the multiple correlation in the population (Guilford, 1956, p. 398). The multiple correlation model was originally chosen because it provides for the simultaneous investigation of many independent variables, yet it presents the problem of either recruiting a very large number of Ss or of reducing the number of predictors, which undercuts study purposes. Fortunately, factor analysis appears to provide for an acceptable compromise. It can reduce substantially the number of predictors, yet still reproduce the reliable variance among independent variables with sufficient accuracy. When contrasted with analyses based upon all 37 independent variables, factor scores yielded lower multiple regression Rs, but higher and statistically significant cross-validation Rs. The advantages of prediction based on factors rather than on the original variables are two-fold: an increase in the degrees of freedom and more reliable predictors. The latter is particularly important when one considers the disappointing inter-rater reliabilities of several of the original independent variables.

Several other issues should be mentioned, particularly for those who contemplate applying factor analysis to problems of prediction. First, by executing the present factor analysis on all 183 cases, an important problem was evaded.

In Horst's words:

> One difficulty with this approach is that one can not (sic) be sure how many factors he needs, (sic) to account for the predictor data matrices in the successive samples. He may need to account for more or less than in the original sample. Furthermore, certain problems of transformation of the factor score matrices on the new samples enter in, because one can not (sic) be sure that the arbitrary factor scores represent the same factors in the sample. (Horst, 1965, p. 552).

Secondly, to reproduce the predictor data matrices as accurately as possible, unities (1.00) are recommended for the diagonal elements of the matrix to be factored (Horst, 1965, p. 552; Harman, 1960, p. 349). Again, the present study deviated from usual practice by using image-covariance estimates of communalities in the diagonal. One reason for this decision was practical; the amount of variance accounted for (82,45%) was deemed acceptable. The second reason was theoretical; prediction based upon common factors was desired. Placing unities in the diagonal enables one to account for a greater amount of the variance in predictor data matrices, but it tends also to produce factors specific or unique to individual variables.

The final comment concerns another way in which factor analysis may be applied to prediction problems. Following examination of the variables loading on each factor, one could select one variable to represent each factor, then generate a prediction equation from these variables alone. This procedure was attempted but not reported because it yielded Rs generally comparable to those of the original analyses.

Problem of Skewness. Seriously skewed distributions, like several encountered in this study, violate the assumptions of rectilinear regression and homoscedasticity which underlie computation of the Pearson product-moment correlation coefficient. Faced with the task of expressing the data in a statistically acceptable form, the following alternatives were considered: (1) ignore skewness and compute product-moment correlations, (2) normalize the skewed distributions by some transformation procedure and compute product-moment correlations, (3) dichotomize variables exhibiting skewed distributions at the median and compute tetrachoric and biserial correlations, and (4) dichotomize the skewed distributions at the point of true dichotomy (i.e., zero versus all positive values) and compute phi and point-biserial correlations. The first alternative was selected for its computational convenience and justified on the grounds that validity coefficients based upon skewed variables would be conservative rather than inflated estimates (Carroll, 1961). The second alternative was not feasible since many of the skewed distributions were also truncated, thus preventing their normalization. The third alternative was chosen in preference to the fourth alternative because tetrachoric and biserial correlations tend to yield larger coefficients than phis and point-biserials (Guilford, 1956, pp. 303, 313). The results of cross-validation favor the use of only product-moment correlations. The reason is that the Pearson correlations were more invariant across the samples. If the product-moment correlation of any two variables in one sample is compared with the product-moment correlation between the same two variables in the other sample, then 8% of these comparisons are significantly different at the .10 level. Similar comparisons of the two integrated matrices of Pearson, tetrachoric and biserial correlations produced 22% of the comparisons which were significantly different at the .10 level. (See Guilford, 1956, pp. 300, 308-309 for discussion of this point.) In retrospect, it probably would have been better to split the skewed distributions at the point of true dichotomy and to use phi and point-biserial correlations, since these are more stable than tetrachoric and biserial correlations (Guilford and Lacey, 1947).

In summary, the issue of skewness has not been completely resolved. Ignoring the skewness problem and computing product-moment correlations produced slightly better cross-validations than did dichotomizing the skewed variables and computing tetrachoric and biserial correlations. However, in both instances, results were disappointing. The appropriateness of phi and point-biserial correlations is yet to be tested. In any event, a more desirable solution would be to rewrite the instructional program so that the independent variables would be normally distributed.

<u>Problem of Heterogeneous Criterion Behaviors</u>. Earlier,

it was suggested that heterogeneity among criterion test items, either qualitatively or quantitatively, may have obscured relationships between one or more program characteristics and the criterion indexes of learning. The implication is "that of drawing distinctions among the different classes of behavior to be established, as a basis for inferences concerning how modification of pre-existing behavior can be undertaken" (Gagné, 1965b, p. 25). If qualitative differences exist, discrete categories of post-learning behavior might be established empirically, such as by factor analyzing the test items, or rationally, such as by utilizing the taxonomies developed by Gagné (1965a, 1965b), Gilbert (1962), or Bloom (1956). Prediction could then be undertaken independently for each subgroup of criterion behaviors. If, on the other hand, quantitative differences along some dimension can be demonstrated among the criterion items, then this dimension might be incorporated into multiple regression analyses as a moderator variable, i.e., "a variable which is correlated with the <u>correlation</u> between two other variables" (Marks, 1964, p. 797). "The 'moderated multiple regression' provides a simple generalization to the case in which the basic parameter is not membership in some group, but score on some continuous variable" (Saunders, 1956, p. 209). Potential moderators include the item difficulty of an answer which prompts responses to the item in question, and some measure of the <u>Ss</u>' familiarity with the key terms such as the Thorndike - Lorge word count.

Problem of Determination of Frame Sequences. The term "frame sequence" has been used to refer to a set of instructional frames which are judged to teach a particular criterion behavior. Ideally, the determination of frame sequences would proceed from a detailed outline of the instructional program. Since such plans are unlikely to be retained by authors, investigators must rely upon sketchy rules and their own judgment to identify the frame sequences. Errors of inclusion or exclusion naturally distort the predictor scores derived for each frame sequence.

After completing the analyses of the present study it was apparent that a set of more rigorously stated decision rules was required, formulated in general for all frame sequences, then re-formulated as specific requirements for the frames of each particular sequence. The first decision involves determining the class of each criterion test item. This would require a set of operationally defined categories, e.g., as described by Gagné (1965a, 1965b) and a statement of the definition, functional relation, discrimination, etc., embodied in the test item. An additional set of decisions would be required in the identification of frame sequences. This should probably begin with the frame which most clearly states the rule pertinent to the test item. Working forward and backward, frames which develop or strengthen the criterion behavior are then identified. Usually these frames are contiguous with "rule" frames or with similar "shaping" frames, and usually, too, they conform to an obvious inductive or deductive sequence. Frequently, this task becomes one of first eliminating clearly non-relevant frames. Difficulty arises with frames containing instruction relevant to more than one criterion objective, such as transition links between two sequences. In the present study, such frames were included in both frame sequences. Additional problems arise when indexing delayed review frames. To date, no objectively stated identification rule has been devised. The third and last set of decisions concerns identification of incidental review of the criterion behavior. There are shades of "incidentalness", ranging from the mere presentation of a key word to a complete re-statement of the principle or concept. Again, too little is known about the incidental presentation of instruction to permit a statement of reliable rules for the identification of such material. However, a rather "loose" definition of what constitutes an "incidental" frame has been used in the past with disappointing results. One might, therefore, use a more narrow definition. Table 27 presents an illustration of how these decision rules might be phrased.

New Variables

The majority of the independent variables investigated

Table 27

Illustrations of Suggestions for Determination of Frame Sequences

Item	When the hold-up victim turns over his wallet, he <u>escapes</u> a threat.	
Criterion Behavior	Class Concept	Upon presentation of stimuli which differ widely in their physical appearance, makes a response which identifies them as instances of a class which distinguishes them from instances belonging to other Classes. (Gagné, 1965b, p. 50)
Item Description	Given an example of behavior which terminates an unconditioned stimulus, identifies example as "escape."	
Decision Rules for Identification of Frame Sequence	(1) Terminal frame must state formally the following elements: (a) operant behavior (b) which terminates (c) an unconditioned stimulus (d) is called escape behavior. (2) Other frames should be identified on the basis of such cues as (a) one or more of the above elements presented either as a rule or as an example, (b) the appearance of the word "escape" particularly as a response, (c) proximity to the terminal frame.	
Decision Rules for Identification of Incidental Frames	(1) Incidental frame must present the word "escape" as a response.	

in this study were drawn from among the more traditional program variables, e.g., number of review frames, and from the research dealing with semantic and syntactic determiners of readability. Since the major portion of criterion variation still remains unexplained, further utilization of the multiple correlation design in programed instruction will have to include variables not employed in this study.

Of the 37 predictors, variables included in this study, the one producing the highest validity coefficients in delayed posttest analyses was variable 33, the order in which the initial learning frames first appear in the program. Oversimplified, it could be claimed that this variable indexes instructional recency. Its negative correlation with the learning criteria indicates that greater gains tend to be associated with those sequences appearing earlier in the text. Among the hypotheses which can be proposed to account for this relationship, the most appealing concerns the organization of the instruction, i.e., the text is organized as a hierarchical sequence of increasingly complex instructional goals built on simpler goals. Certain elementary concepts, for example, must be mastered before progressing to more advanced principles. If the criterion responses taught earlier in the program are less complex, they are probably more readily learned and they thus exhibit the greater gains.

If the interpretation above is correct, then it should also be fruitful to investigate correlates of program order.

The following three implications of a hierarchical organization of instruction are proposed for additional study: (1) certain criterion responses are more complex and, consequently, more difficult to acquire; (2) certain knowledge is required as a pre-condition for more complex instruction; and (3) different learning tasks require different instructional strategies. The first implication has already been discussed in connection with the problem of heterogeneous criterion behavior. The second implication is illustrated when, for example, the concepts of differential reinforcement and of successive approximations must be learned before learning the principle of shaping. Some index of the mastery of the simpler concepts may predict the extent to which the more complex principle is learned by the Ss. One might choose the percent of correct responses to terminal frames for simpler concepts, i.e., the frames within the text which test the criterion behavior unaided by prompts. If the third implication is valid, then emphases in instructional variables should be adjusted to meet requirements of different learning tasks. Recent publications of Gagné (1965a, 1965b) have considered this topic at length. Since the position taken in this section coincides with Gagné's, a sampling of his recommendations may indicate how some of the above suggestions might be applied:

> Multiple discrimination: Present stimuli in a manner that emphasizes distinctiveness. Repetition may be needed to reduce interference among individual connections. Confirm correct responses.

> Concept learning: Present a suitable variety of stimuli to represent the concept class, each stimulus having a connection with a common response. Verify by presenting a novel stimulus that is also a member of the class.
>
> Principle learning: Inform the learner of the performance to be expected. Invoke recall of component concepts by verbal instructions. Make verbal statement of principle. Verify by the direction "Show me." (Gagné, 1965a, pp. 254-255).

Needless to say, this is not an exhaustive catalogue of the instructional techniques or strategies which might profitably be applied to the above tasks. But the examples serve to illustrate the shaping of instruction to meet demands placed upon learners.

Several "new" variables have been proposed for future research. These include: (1) categories of learning, (2) estimate of the extent to which pre-conditions for more advanced learning are satisfied, and (3) certain instructional variables.

The Factor Analysis

To review briefly, a principal components factor analysis of the 37 independent variables was executed with image covariance estimates of the communalities. Twenty-one factors were extracted and nine factors, accounting for 82.45% of the variance, were rotated orthogonally. Multiple regression and cross-validation analyses against the immediate posttest criteria were undertaken using factor scores derived by means of Harman's short method (1960). The nine sets of factor scores were uncorrelated. Of the six cross-validations, one was significant at the .01 level, two at the .05 level, and another missed significance by one correlation point. Validity coefficients for the nine factors are presented in Table 28. The factor analysis and rotation were later re-computed with the immediate posttest residual gain score included among the variables. The loadings of this criterion index on the nine factors are also presented in Table 28. The factors will now be discussed individually.

Factor 1: Quantity of Instruction. The First factor reflected the amount of instruction associated with each criterion behavior. This factor includes the number of frames (variable 1), of initial learning frames (variable 2), of presentations of the key term both as a stimulus (variable 6) and as a response (variable 7), of the thematic and sequence prompts (variable 11), etc. This factor had non-significant positive correlations with learning gain indexes in one item sample and slightly negative validities in the other sample. It is rather disconcerting to discover that one of the most basic of all programming variables and the factor which accounts for more variance than any other was unrelated to the learning criteria. Specialists in programmed instruction may know this factor as "step size"; "in experimental settings the number of items per concept is the most frequent definition...of...step size." (Holland, 1965, p. 88).

Table 28

Validity Coefficients of the Factors and the Loadings of Immediate Posttest Residual Gain Index on the Factors

	Sample 1			Sample 2			
Factor	Absolute Gain	Relative Gain	Residual Gain	Absolute Gain	Relative Gain	Residual Gain	Factor loading of immediate posttest residual gain index
1	-.14	-.08	-.09	.17	.10	.09	.04
2	.02	.10	.05	.17	.26*	.24*	.13
3	.06	.06	.03	-.02	.05	-.02	.01
4	-.15	-.27**	-.21*	-.25*	-.18	-.24*	-.17
5	-.01	-.02	.00	-.02	-.01	-.02	.00
6	-.11	.00	.01	-.20	-.07	-.05	.04
7	.06	.08	.08	-.02	.00	.00	-.04
8	-.22*	-.13	-.20*	-.31**	-.23*	-.30**	-.37**
9	-.21*	-.10	-.11	-.02	-.07	-.07	.05

* Significant at .05 level.

** Significant at .01 level.

A cursory search of the literature on step size reveals contradictory findings. For instance, Coulson and Silberman (1959) compared the learning effects of 56 and of 104 frames excerpted from The Analysis of Behavior. Similarly, Evans, Glaser, and Homme (1960) compared effects of mathematics programs with 30, 40, 51, and 68 frames. Both studies found significant differences favoring the longer programs. On the other hand, Smith and Moore (1961) and Smith (1965) found no relationship between number of steps and learning effects. The former study involved spelling programs of 1128, 830 and 546 frames. The latter study compared sixty sequences of spelling frames ranging in size from four to 123 frames. It appears that no generalization about step size and effects on learning can be stated with a high degree of confidence.

Factor 2: Sentence Complexity. Factor 2 was defined primarily by syntactic variables, e.g., modifiers per sentence (variable 30), verbal nouns and adjectives per sentence (variable 32), prepositional phrases per sentence (variable 35) and clauses per sentence (variable 36). The variable with the highest loading indexed the number of words per sentence (variable 37), which implies that sentence length varies as a function of sentence complexity. Factor 2 was significantly and positively related to learning gains in one sample of test items but not in the other.

Several of the studies reviewed in the

Introduction (see page 3) indicate syntactic variables facilitate certain types of learning, e.g., rote and paired associate learning. Smith (1965) and Gray and Leary (1935) reported negative correlations of syntactic complexity with learning. Closer scrutiny of the studies in question reveals that they differed both in terms of the experimental task and in the nature of their Ss. The latter point bears elaboration. The former experiments employed college student Ss almost exclusively, while the latter two studies were based on young children and on adults of limited reading ability, respectively. It is tempting to speculate about a hypothetical minimal level of reading ability. For people who have achieved this hypothetical level of reading ability, syntactical variables organize the content more efficiently for memorization (e.g., Miller, 1965) and/or provide textual constraint for key terms (e.g., Aborn, et al, 1957). Support for this hypothesis comes from Holland (1933) who found that "sentences which contain a conditional clause at the beginning are read at a greater rate of speed by subjects above the seventh grade and at a slower rate by pupils below this level." Thus, it appears that any generalization about sentence complexity and learning must consider the level of reading ability among Ss.

Factor 3: Concrete versus Abstract Instruction. Factor 3 reflects the relative presence within a frame sequence of instructional rules and examples (variables 27 and 28, respectively). In addition, a greater proportion of complex words, defined in terms of affixes (variable 14) and of syllables (variable 25), appeared in rule frames, while a larger percentage of concrete nouns (variable 17) was contained in example frames. Validities associated with this factor are consistently disappointing, particularly in view of the recommendations quoted earlier from Gagné, who considers the presentation of a variety of examples and the stating of principles to be important tactics for teaching "concepts" and "principles," respectively. In view of the criterion heterogeneity problem discussed earlier, it is surely best to reserve judgment concerning the effects of "abstract-versus-concrete" instruction.

<u>Factor 4: Review</u>. The review factor reflected the massed-versus-spaced dimension of instruction. A frame sequence with a high score on Factor 4 tended to have more intentional review frames (variable 3) and more incidental review frames (variable 34) distributed over a greater number of sets (variables 23 and 24). Factor 4 was significantly related to the criterion indexes; hence, it is inferred that spaced review facilitates programmed learning. The findings of Reynolds and Glaser (1964), Myers (1964), and Rothkopf (1965) support this inference. Conflicting evidence is reported by Smith (1965). Since Smith's findings are based upon grade school children, while the other studies employed college and high school Ss, some developmental

hypothesis might be invoked to account for the fact that certain age groups apparently profit more than others from such instructional variables as distributed review.

Variable 7, appearances of the key term as a response, correlated with Factor 4 because this variable was the principal cue used in identifying review frames. Perhaps this variable ought to be discussed on its own merits. It was the most effective predictor in the immediate posttest analyses and the second best in the delayed posttest analysis. The response requirement of programmed instruction has been a controversial topic since the earliest days of programmed instruction research. The majority of investigations have not found significant differences in learning, when contrasting active and covert response treatments. Kemp and Holland (1966) have recently described a procedure which produces a "blackout ratio," and this, they feel, will resolve the response controversy. The blackout ratio is the "percent of total words which could be removed without affecting error rate." Kemp and Holland consider this ratio to be a measure of the degree to which instruction is programmed or response contingent, i.e., the extent to which the written instruction must be read in order to supply the required responses. They applied the blackout ratio to twelve sets of materials used in previous studies of overt versus covert responding. Ratios ranged from 11.1% to 74.6%. The four lowest ratios were obtained f r programs which had previously yielded significant differences favoring overt responding. The remaining eight ratios were for programs which had previously yielded no significant differences between overt and covert responding. In other words, the response variable seems important only to the degree that the written instruction must be read in order to respond correctly.

Factor 5: Cloze Score. Factor 5 was limited to the three cloze score variables: mean cloze score (variable 18), maximum cloze score (variable 19), and minimum cloze score (variable 20). Correlations of this factor with the criterion indexes were essentially zero. Previous studies (reviewed by Rankin, 1965) had suggested that cloze score is a highly valid measure of comprehension. As Green (1965) has suggested, further efforts might be directed toward exploring the effects of deleting only certain types of words, e.g., nouns.

Factor 6: Key term. This factor was defined by the number of letters (variable 8) and the number of syllables (variable 9) in the key term. These variables had been hypothesized as measures of criterion task difficulty, but this factor was found to be unrelated to the criterion indexes. Implementation of the recommendations discussed earlier may lead to the identification of relationships between learning task difficulty and learning gains.

Factor 7: Word Complexity. The following variables correlated with Factor 7: affixed word ratio (variable 14), syllables per word (variable 25), and vehicular appearances of the key term (variable 34). The loading of variable 34 seems to imply that less complex words were associated with the frame sequences having the fewest vehicular appearances of the key term, viz., the sequences which appeared earlier in the program. It is certainly plausible that more complex vocabulary would be developed as instruction progresses. Word complexity, however, did not predict the criteria. Previous studies by O'Connor (1950) and by Gray and Leary (1935) had observed positive correlations between frequency of affixes and learning. Myers (1964) obtained the opposite relation when he correlated affixes ıd learning effects. In line with the present study, Smith (1965) found no significant relationships between word complexity variables and learning from a spelling program. In view of the conflicting findings reported about word complexity variables, evaluation of these variables should be withheld pending further investigation.

Factor 8: Largest Frame. The variables with the highest loadings on Factor 8 were the number of words in the frame with the most words (variable 12) and the number of syllables in the frame with the most syllables (variable 13.) Variable 26, the number of sentences per frame, had a moderate loading. Factor 8 produced the highest validities of any factor. A sequence containing a large frame tended to produce high learning gains. One might interpret this finding as support for a large "step" style of programming. Obviously, this is not a sufficient interpretation of this factor's functioning; one might examine the program for correlates of variables 12 and 13 which can more easily be interpreted.

Factor 9: Unnamed. Factor 9 appears to be a residual factor. It was unrelated to the criteria, with an exception that seems attributable to chance.

Other Factor Analytic Studies

Two factor analyses of written instruction have been published. Both were based upon data originally reported by Gray and Leary (1935). Using the principle axis method with a quartimax rotation, Stolurow and Newman (1959) factor analyzed 23 features of written prose. The two most significant factors, accounting for 51% of the variance of the rotated factors, were labeled semantic difficulty, or easy versus difficult word factor, and syntactical difficulty, or easy versus difficult sentence factor. Brinton and Danielson (1958) selected 20 language elements for their analyses. Using the centroid method with graphical rotation, they obtained four interpretable factors: vocabulary, sentence, stylistic or devices of content, and grammatical complexities. The first two factors correspond to those obtained by Stolurow and Newman.

The data from Smith's study (1965) was also factor analyzed using the principal components method and a varimax rotation. Smith had employed a multiple regression design to study the relationships between 38 program variables and learning effects. Many of the variables used in the present study had also been employed by Smith. Five interpretable factors were obtained: quantity of instruction, sentence length, number of responses, word frequency, and word complexity.

Inspection of the variable loadings obtained in the four analyses suggests the generality of the following factors: (1) quantity of instruction (present study, Smith), (2) sentence length (Stolurow and Newman, Brinton and Danielson, Smith) and/or sentence complexity (present study, Brinton and Danielson), (3) word frequency (Stolurow and Newman, Brinton and Danielson, Smith), (4) word complexity (present study, Smith).

SUMMARY AND CONCLUSIONS

The results of this study have been encouraging, in some respects, and disappointing, in other respects. Significant relationships with the learning criteria were found for distributed review and requiring the key term as a response. A less strong relationship was found for syntactic complexity. Correlates of program order and of large frames were recommended for future research. It was also noted that future research applications of multiple regression in programmed instruction depends upon resolving such issues as degrees of freedom, skewed independent variables, heterogeneity of criterion behaviors, and determination of frame sequences. The following solutions were tentatively proposed: prediction based on factor analysis, re-writing instructional materials to produce less skewed distributions, sub-grouping of criterion behaviors, and more rigorously defining decision rules.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aborn, M., Rubenstein, H., and Sterling, T. D. Contextual constraint in sentences. J. exp. Psychol., 1957, 57, 171-180.

Bloom, B. S. (Ed.) Taxonomy of educational objectives. N. Y.: Longmans, Green, 1956.

Brinton, J. E., and Danielson, W. A. A factor analysis of language elements affecting readability. Journalism Quarterly, 1958, 35, 420-426.

Buhler, R. P-stat; A system of statistical programs for the 7090/7094. Princeton, New Jersey: Princeton University Press, 1964.

Campbell, D. T., and Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Carroll, J. B. The nature of the data, or how to choose a correlation coefficient. Psychometrika, 1961, 26, 347-372.

Coleman, E. B. Learning of prose written in four grammatical transformation. J. appl. Psychol., 49, 332-341.

Coulson, J. E., and Silberman, H. F. Results of an initial experiment in automated teaching. Santa Monica, Calif.: System Development Corp., 1959.

DuBois, P. H. The design of correlational studies in training. In R. Glaser (Ed.), Training research and education. Pittsburgh: University of Pittsburgh Press, 1962.

Epstein, W. The influence of syntactical structure on learning. Amer. J. Psychol., 1961, 74, 80-85.

Epstein, W. A further study of the influence of syntactical structure on learning. Amer. J. Psychol., 1962, 75, 121-126.

Evans, J. L., Glaser, R., and Homme, L. E. A preliminary investigation of variation in properties of verbal learning sequences of the "Teaching Machine" types. In A. A. Lumsdaine and R. Glaser (Eds.), Teaching machines and programmed learning: A source book. Washington, D. C.: National Education Association, 1960, pp. 446-451.

Fincke, G. E. A factorial approach to the life history correlates of training task performance. Unpublished Ph.D. Dissertation. Purdue University, Lafayette, Indiana, 1967.

Gagné, R. M. The conditions of learning. New York: Holt, Rinehart and Winston. 1965 (a).

Gagné, R. M. The analysis of instructional objectives for the design of instruction. In R. Glaser (Ed.) Teaching machines and programed learning, II: Data and directions. Washington, D. C.: National Education Association, 1965, pp. 21-65. (b)

Gilbert, T. F. Mathetics: the technology of education. J. Mathetics, 1962, 1, 7-73.

Glanzer, M. Grammatical category: A rote learning and word association analysis. J. Verb. Learn. Verb. Behav., 1962, 1, 31-41.

Gray, W. S., and Leary, B. E. What makes a book readable: An initial study. Chicago: University of Chicago Press, 1935.

Green, F. P. Modification of the cloze procedure and changes in reading test performances. J. educ. Measmt., 1965, 2, 213-217.

Guilford, J. P. Fundamental statistics in psychology and education. (3rd ed.) New York: McGraw-Hill, 1956.

Guilford, J. P., and Lacey, J. I. (Eds.) Printed classification tests. Army Air Force Aviation psychology research program reports, Report 5. Washington, D. C.: U. S. Government Printing Office, 1947. Cited in J. P. Guilford. Fundamental statistics in psychology and education. (3rd ed.) New York: McGraw-Hill, 1956.

Guttman, L. Image theory for the structure of quantitative variates. Psychometrika, 1953, 18, 277-296.

Harman, H. H. Modern factor analysis. Chicago: University of Chicago Press, 1960.

Holland, B. F. The effect of length and structure of sentences on the silent reading process. Psychol. Bull., 1933, 30, 668-669.

Holland, J. G. Research on programing variables. In R. Glaser (Ed.), Teaching machines and programed learning, II: Data and directions. Washington, D. C.: National Education Association, 1965, pp. 66-117.

Holland, J. G., and Skinner, B. F. The analysis of behavior. New York: McGraw-Hill, 1961.

Horst, P. Factor analysis of data matrices. New York: Holt, Rinehart and Winston, 1965.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 1958, 23, 187-200.

Kemp, F. D., and Holland, J. G. Blackout ratio and overt responses in programmed instruction: Resolution of disparate results. J. educ. Psychol., 1966, 57, 109-114.

Klare, G. R. The measurement of readability. Ames, Iowa: Iowa State University Press, 1963.

Krumboltz, J. D., and Weisman, R. G. The effect of overt versus covert responding to programmed instruction on immediate and delayed retention. J. educ. Psychol., 1962, 53, 89-92.

Lawshe, C. H., and Balma, M. J. Principles of personnel testing. (2nd ed.) New York: McGraw-Hill, 1966.

Lumsdaine, A. A. Instruments and media instruction. In N. L. Gage (Ed.), Handbook on research in teaching. Chicago: Rand McNally, 1963, pp. 583-682.

Markle, Susan M. Good frames and bad. New York: Wiley, 1964.

Marks, L. E., and Miller, G. A. The role of semantic and syntactic constraints in the memorization of English sentences. J. Verb. Learn. Verb. Behav., 1964, 3, 1-5.

Marks, M. R. How to build better theories, tests, and theories: The off-quadrant approach. Amer. Psychologist, 1964, 19, 793-798.

Martin, J. G., Davidson, Judy, and Williams, W. L. Grammatical agreement and set in learning at two age levels. J. exp. Psychol., 1965, 70, 570-574.

May, M. A., and Lumsdaine, A. A. Learning from films. New Haven, Conn.: Yale University Press, 1958.

McGuigan, F. J., and Peters, R. J. Assessing the effectiveness of programed texts: Methodology and some findings. J. prgmd. Instrn., 1965, 3, 23-34.

Miller, G. A. Some psychological studies of grammar. Amer. Psychologist, 1962, 17, 748-761.

Miller, G. A. Some preliminaries to psycholinguistics. Amer. Psychologist, 1965, 20, 15-20.

Myers, J. B. The identification and effects of frame and frame sequence characteristics related to learning retention from programmed instruction. Unpublished Ph.D. Dissertation. Purdue University, Lafayette, Indiana, 1964.

O'Connor, V. J. An examination of instructional films for characteristics of an effective teaching presentation. Harv. educ. Rev., 1950, 20, 271-284.

Rankin, E. F. The cloze procedure - a survey of research. In E. Thurston and L. Hafner (Eds.), The philosophical and sociological bases of reading. Yearbook of the National Reading Conference, 1965, No. 14.

Rankin, E. F., and Tracey, R. J. Residual gain as a measure of individual differences in reading improvement. J. Reading, 1965, 8, 224-233.

Reynolds, J. H., and Glaser, R. Repetition and spaced review variables in programmed instruction. Investigations of learning variables in programmed instruction. Pittsburgh: Programmed Learning Laboratory, University of Pittsburgh, 1962.

Rothkopf, E. Z. Some theoretical and experimental approaches to problems in written instruction. In J. D. Krumboltz (Ed.) Learning and the educational process. Chicago: Rand-McNally, 1965, pp. 193-221.

Saunders, D. R. Moderator variables in prediction. Educ. psychol. Measmt., 1956, 16, 209-222.

Smith, M. E. The prediction of learning effects from linear program characteristics. Unpublished M. S. thesis. Purdue University, Lafayette, Indiana, 1965.

Smith, W., and Moore, J. W. Size-of-step and achievement in programmed spelling. Psychol. Reps., 1962, 10, 287-294.

Stolurow, L. M., and Newman, J. R. A factorial analysis of objective features of printed language presumably related to reading difficulty. J. educ. Res., 1959, 52, 243-251.

Taylor, W. L. Cloze procedure: A new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.

Weaver, W. W., and Kingston, A. J. A factor analysis of cloze procedure and other measures of reading and language ability. J. Communication, 1963, 13, 252-261.

Wodtke, K. H. On the assessment of retention effects in educational experiments. A paper presented to the American Educational Research Association, Chicago, February, 1966.

Yngve, V. Computer programs for translation. Scientific American, 1962, 206, 68-76.

APPENDIX A

Criterion Test

The following items constituted the criterion test. They were adapted from the program's review sets and are identified by the number they bear in the program. The first two digits, e.g., 17, indicate the review set, while the latter two denote the frame within the set. The words underlined are the answers.

1701. On a cumulative record, the slope of the line indicates (a) rate, and the hatch marks or pips usually indicate (b) reinforcements.

1702. (a) Operant behavior is strongly influenced by the consequences of previous similar responses, whereas (b) respondent behavior depends upon a preceding stimulus.

1703. In a conditioned reflex, when a conditioned stimulus is repeatedly presented alone, the magnitude of the conditioned response (a) decreases and the latency of the conditioned reflex (b) increases, until (c) extinction is complete.

1704. When a pigeon is reinforced for pecking a key, the reinforcing stimulus occurs (a) after a peck, and the (b) rate at which this response is (c) emitted (TT) increases.

1705. Turning off a television commercial is reinforced by termination of a(n) (a) negative reinforcer; turning on a very funny program is reinforced by the presentation of a(n) (b) positive reinforcer.

1706. Name the response systems involved in the following: walking to the table, putting food in the mouth and chewing it, (a) striated muscle; moistening food with saliva, (b) glands; passing food into stomach, (c) smooth muscle; and providing stomach with digestive juices, (d) glands.

1707. Many so-called traits ascribed to individuals (aggressiveness, persistence, friendliness, etc.) are simply alternate ways of indicating an individual's (a) rate of emitting certain types of behavior.

1708. In differential reinforcement, one form or magnitude of behavior is (a) reinforced and other, possibly rather similar forms or magnitudes, are not reinforced. (Most blanks were considered as separate items. In contrast, 1708 was considered as one item, and a student was credited with a correct answer only when he supplied both terms.)

1709. The experimenter deliberately arranges reinforcement for key pecking, but superstitious behavior is conditioned by (a) accidental reinforcement.

1710. In a reflex, the (a) threshold of a stimulus is the intensity which is barely sufficient to (b) elicit a(n) (c) response.

1711. An important aspect of respondent conditioning is the (a) temporal relation between presentations of the initially neutral stimulus and of the unconditioned stimulus.

1712. Not used.

1713. Conditioned operants are eliminated in two contrasting ways: the response is emitted without reinforcement in the process called (a) extinction, but is not emitted in the process called (b) forgetting.

1714. In conditioning a reflex, as the number of pairings of the conditioned and unconditioned stimuli increases, the latency of the conditioned reflex (a) decreases and the magnitude of the conditioned response (b) increases, until both reach a limit.

1715. A psychologist fed a baby when he emitted "coos," but not when he cried. We would expect that crying when hungry would be (a) extinguished (TT). (1715b not used.)

1716. Certain groups of responses, such as those elicited by a sudden loud noise, are characteristic of a state of (a) emotion.

1717. When we differentially reinforce successive approximations to a final form of behavior, we are (a) shaping behavior.

1718. Persistent head scratching, pencil chewing, table tapping, etc., while studying are frequently conditioned (a) superstitious operants. (1718b not used.)

1719. Two ways of effectively preventing unwanted conditioned behavior are: (a) to (a) extinguish it by withholding reinforcement, or (b) to condition some (b) incompatible behavior.

1720. A stimulus which elicits a response without previous conditioning is called a(n) (a) unconditioned stimulus; a stimulus which elicits a response only after conditioning is called a(n) (b) conditioned stimulus.

1721. If an airplane spotter never sees the kind of plane he is to spot, his frequency of scanning the sky (a) decreases. (1721b not used.)

1722. You will not continue to work if your pay checks "bounce" because the (a) conditioned generalized reinforcing effect of such a check disappears in (b) extinction.

1723. A simple operant can be conditioned very rapidly if the organism is (a) adapted to the situation and if a reinforcer follows the response (b) quickly.

1724. In shaping any given behavior, we gradually change the criterion for reinforced responses. The desired behavior is approached by (a) successive approximation.

1725. To condition a reflex, a neutral stimulus is paired with a(n) (b) unconditioned stimulus. (1725a not used.)

1726. In the usual experiment, when a peck operates the food magazine the (a) conditioned reinforcement is immediate, whereas the (b) unconditioned reinforcement is slightly delayed.

1727. In a reflex, the more intense the stimulus, the greater the (a) magnitude of the response and the shorter the (b) latency of the reflex.

1728. When a response is elicited by a stimulus without previous conditioning, the sequence is called a(n) (a) unconditioned reflex.

1729. The pairing of two stimuli is necessary for conditioning (a) respondent behavior; reinforcement is necessary for conditioning (b) operant behavior.

1730. Reaching for a glass of water or saying "Water, please" are examples of (a) operant behavior; any specific instance of such behavior, however, is called a(n) (b) response.

1731. Lying generates stimuli which have acquired the power to elicit the conditioned responses which occur in (a) fear.

1732. A particularly slow learner may require many reinforcements before developing a high rate of responding. He is (a) less likely to develop superstitious behavior than a faster learner.

1733. Operant behavior has direct consequences on the environment. A consequence which results in an increase in the subsequent rate of the operant response is called a(n) (a) reinforcer. (TT)

1734. Not used.

1735. Smooth muscles change the (a) dimensions of various internal organs. (The two blanks were scored as one item. Credit was given only when both answers were supplied.)

1736. A conditioned reinforcer can become a(n) (a) generalized reinforcer by being paired with several unconditioned reinforcers appropriate to various deprivations.

1737. The reinforcers used by animal trainers are (a) deliberately arranged, but a pigeon foraging for food among leaves in a park is working under (b) natural contingencies.

1738. When behavior decreases in frequency and when, so far as we know, no previous conditioning of the behavior has taken place, we call the process not extinction but (a) adaptation.

1739. Learning to say "ball" makes it easier for the child to learn to say "fall" because the two responses have (a) common elements.

2901. The professional winetaster can make very fine (a) discriminations. He shows little (b) generalization among various wines.

2902. Properly steering an automobile requires that the organism have a(n) (a) continuous repertoire.

2903. Availability of reinforcement depends on the passage of time in (a) interval schedules, and on the number of responses in (b) ratio schedules.

2904. The frequency with which a normal organism eats, given free access to food, changes in (a) cycles.

2905. Independent of the specific deprivation state present at the moment, a(n) (a) generalized reinforcer can be used to condition a new response.

2906. For most states of deprivation, there are alternative procedures which belong to the same (a) class since they have similar effects on a whole (b) class of responses.

2907. Each stimulus in a chain has the dual function of (a) reinforcing the response it follows and being a(n) (b) S^D for the response it precedes.

2908. Not used.

2909. A response occurring immediately after a reinforcement is never reinforced on a(n) (a) fixed-interval schedule. A response immediately after reinforcement is sometimes reinforced on a(n) (b) variable-interval schedule.

2910. The operant called "pressing a lever for food" is composed of many stages or parts integrated as a(n) (a) chain of stimuli and responses.

2911. A high rate of responding (a) is not in itself evidence for inferring a high level of deprivation without knowledge of other factors such as the reinforcement schedule.

2912. Responses reinforced by the generalized reinforcers of affection, approval, etc., are often extinguished very (a) slowly because reinforcement has occurred (b) intermittently due to the subtlety of the stimuli.

2913. Not used.

2914. Objects at two adjacent points in space usually control reaching responses of only slightly different (a) forms.

2915. Response magnitude varies closely with stimulus intensity in the case of (a) respondent behavior, but much less so in the case of (b) operant behavior.

2916. An organism may emit the same response to two fairly similar stimuli when only one of them has been present during reinforcement. The term for this phenomenon is (a) stimulus generalization.

2917. When paired with several unconditioned reinforcers appropriate to various deprivations, a conditioned reinforcer will be effective under several types of (a) deprivation. (2917b not used.)

2918. In teaching a child to copy a line, we do not wait for a perfect drawing. We (a) reinforce movements which produce lines fairly similar to the copy, but (b) extinguish movements which produce lines very different from the copy.

2919. When a response is under the control of a single property of a stimulus (which cannot exist alone), we call it a(n) (a) abstraction.

2920. Excessive sweating due to heavy work or emotion is similar to (a) water deprivation in its effect on the class of responses which have been (b) reinforced by water.

2921. Intermittently reinforcing temper tantrums makes them very (a) resistant to extinction.

2922. (Four graphs were presented. Ss were asked to match each graph with the reinforcement schedule most likely to have produced the graph.)

(a) fixed-interval
(b) variable-interval
(c) fixed-ratio
(d) variable-ratio

2923. The more food an animal has just eaten, the (a) less probable its ingestive behavior.

2924. Not used.

2925. In operant discrimination we speak of a three-term contingency. Events are arranged in this order: (a) present the S^D, (b) wait for the response, and (c) reinforce.

2926. In the laboratory, an animal's weight expressed as a percentage of its normal weight is used as a measure of its history of (a) food deprivation.

2927. Skill in "drawing from copy" will continue to be shaped without help from other persons once "likeness" has become a reinforcer, since this will automatically provide the (a) differential reinforcement needed for shaping.

2928. In establishing a discrimination, a response is (a) reinforced in the presence of one stimulus and (b) not reinforced in the presence of another stimulus.

4101. In the experiment demonstrating shock avoidance, the low initial rate suggests that anxiety (a) is necessary for adequate avoidance.

4102. The reflex activities comprising the activation syndrome occur together in emotions. We (a) do not define any one specific emotion (anger or fear) by listing the reflexes involved.

4103. When the hold-up victim turns over his wallet, he (a) avoids a threat and (b) escapes physical injury.

4104. In the experiment showing the ineffectiveness of punishment in removing a response from the repertoire, punishment was in effect for the first 10 minutes of (a) extinction of a food-reinforced response.

4105. Not used.

4106. Organisms in situations we call "emotional" frequently show the (a) activation syndrome.

4107. Punishment is effective in preventing a response when behavior which (a) avoids the punished behavior is established. But when this behavior has been extinguished, the (b) punished response will again be emitted.

4108. A pigeon reinforced with food for pecking a key during an S^D but not during an S will peck a second key which delays the appearance of (a) S.

4109. If a timid person forces himself to attend many group meetings, his timidity may undergo the process of (a) extinction or (b) adaptation.

4110. With continuous punishment of a response maintained by positive reinforcement, the greater the severity of the punishment (after some minimum intensity is surpassed), the (a) lower the response rate.

4111. In avoidance behavior, the longer the response postpones the aversive stimulus, the (a) lower the rate.

4112. Punishment does not (a) eliminate a response. Rate of responding temporarily (b) decreases, in part, because punishment and the resulting conditioned aversive stimuli generate (c) anxiety.

4113. In the activation syndrome, inhalation (a) quickens, bronchioles (b) dilate, and adrenalin secretion (c) begins.

4114. (Three graphs were presented. Ss were asked to match each graph with the reinforcement schedule most likely to have produced the graph.)

(a) Control session
(b) Stimulant injected
(c) Tranquilizer injected
(d) Buzzer onset
(e) Shock

4115. Reinforcement consists of (a) presenting a positive reinforcer or (b) terminating a negative reinforcer. Punishment consists of (c) terminating a positive reinforcer or (d) presenting a negative reinforcer.

4116. Operant behavior which postpones an unconditioned aversive stimulus is called (a) avoidance behavior. Behavior which terminates an unconditioned aversive stimulus is called (b) escape behavior.

4117. The activation syndrome would usually be biologically (a) useful to a cave man, and it would usually be biologically (b) useless to a public speaker.

4118. The exhilarating emotional reaction from riding a roller coaster or being in a speeding car is a reaction to stimuli which are often (a) aversive but in this instance are (b) reinforcing (TT).

4119. It is poor technique to shape skillful behavior with (a) negative reinforcers because the aversive stimuli, which must be presented in order to be terminated, elicit many respondents which (b) are incompatible with the behavior to be shaped.

4120. In the presence of a conditioned aversive stimulus, a food-reinforced response is emitted more (a) slowly, and a response with an avoidance history more (b) rapidly. These and similar changes define (c) anxiety.

4121. A given emotion (a) is defined by the events which serve as reinforcers or by the increased probability of a group of responses.

4122. The teacher who uses strong aversive control to get students to study may (a) decrease the probability that they will study after graduation.

4123. Not used.

5301. Not used.

5302. When a dog stands "not too close and not too far" from a strange object, its behavior shows (a) algebraic summation. When it approaches, jumps back, approaches, jumps back, approaches, etc., it shows (b) oscillation.

5303. Reaction formation may be interpreted as behavior which removes (a) S^Ds which make the punishable behavior more probable; it may also be interpreted as behavior which is (b) incompatible with punishable behavior.

5304. When a person cannot recall some previously punished behavior, we say he has (a) repressed it.

5305. Rationalization (b) is not the same as lying. (5304a not used.)

5306. Inadequate self-knowledge may result because we have not yet been (a) conditioned to notice some aspect of our behavior, or because we engage in behavior which is (b) incompatible with noticing this behavior.

5307. A word which "expresses multiple meanings" is controlled by (a) many stimuli.

5308. In self-control, the controll-(a) ed response usually has both (b) reinforcing and (c) aversive consequences.

5309. In the laboratory, it (a) is possible to isolate simple functional relations; but in interpreting most events outside of the laboratory, we must be alert to the possibility of (b) multiple effects in complex situations.

5310. In so-called self-control, the controlling response is established through (a) reinforcement. Usually this takes place through the reduction of the (b) aversive stimuli associated with the controlled response.

5311. In psychotherapy, a patient often manifests strong love or hate toward his therapist. This is sometimes called (a) transference.

5312. The effects of conditioned aversive stimuli are (a) decreased by drinking alcohol. This is shown when behavior which automatically generates aversive stimuli occurs (b) frequently under alcohol.

5313. Sneezing which clears the upper respiratory passages is (a) respondent behavior. However, the "imitative" sneezing by the little boy who "only does it to annoy" is (b) operant behavior.

5314. The goal of a science of behavior is to be able to (a) predict, (b) control, and (c) interpret the behavior of living organisms.

5315. The bodily malfunctions called "psychosomatic" result from prolonged elicitation of the responses characteristic of the (a) activation syndrome.

5316. By using simple organisms with controlled histories in simple environments, we hold many variables constant in order to (a) isolate one effective variable at a time.

5317. A therapist frequently tries to change his patient's behavior by acting as a(n) (a) non-punishing audience and thereby the aversive properties generated by previously punished behavior can become (b) extinguished.

5318. A man who strikes his adversary shows (a) counter-aggression.

5319. When we admit, even to ourselves, only the least punishable reasons for our behavior, we are (a) rationalizing.

5320. By tickling his throat with a feather, a man can regurgitate poisonous food. The controlled response is a(n) (a) respondent (TT) and the controlling response is a(n) (b) operant. (TT)

5321. Transference during psychotherapy can be interpreted as an example of (a) stimulus generalization.

5322. A man is said to be addicted to a drug if withholding the drug produces (a) withdrawal symptoms which furnish (b) aversive stimuli.

5323. To eliminate undesired behavior, such as a dog's scratching a door, it (a) is not necessary to use aversive control.

5324. We sometimes speak of cause and effect. A cause is an independent variable, and an effect is a(n) (b) dependent variable. The relation is a(n) (c) functional relation or a(n) (d) law. (5324a not used.)

5325. In self-control, the (a) controlling response affects variables in such a way as to change the probability of the (b) controlled response.

5326. A single aversive stimulus used in punishment elicits respondents, conditions other stimuli to elicit these respondents, and makes possible the conditioning of avoidance behavior. The single aversive stimulus has (a) multiple effects.

5327. Not used.

APPENDIX B

Frame Sequences Related to Each Criterion Test Item

The underlined numbers refer to review set frames which served as test items. The letter "a" refers to the first or only response in that frame; the letter "b" to the second response in that frame; etc.

1701A: set 12 - 5,6,7,8,9,10,11,12,13,21,22,27,28,29,30; set 13 - 18,19; set 18 - 20,23,24; set 21 - 6.

1701B: set 12 - 17,18,19,20; set 18 - 21,25,29; set 21 - 5.

1702A: set 8 - 21,29; set 10 - 15.

1702B: set 8 - 9,21; set 10 - 22,27.

1703A: set 4 - 25,28; set 10 - 27.

1703B: set 4 - 25,28.

1703C: set 2 - 18,19,20,21,23,29,30; set 3 - 21,22,23,24; set 4 - 26,27,28,29; set 5 - 6,16,17,33,44,46,47,49,57.

1704A: set 7 - 5,6; set 8 - 6,8,11,24,27; set 10 - 17,29; set 11 - 37; set 14 - 4; set 15 - 35.

1704B: set 7 - 8,16,17,18,19,29; set 8 - 5,6,10,18,24,25,26, 30; set 9 - 1,9,12,13,14,16,21,27,29; set 10 - 4,16,18; set 11 - 56; set 15 - 35; set 18 - 1; set 32 - 14,15; set 37 - 4, 5,11.

1704C: set 7 - 27; set 8 - 27; set 9 - 13; set 10 - 18,29;

set 14 - 40,42; set 21 - 51,52,57,59.

<u>1705A</u>: set 9 - 4,5,6,7,9,10,12,20,21; set 10 - 1,2,3; set 14 - 34,35,37; set 32 - 12,17; set 23 - 1; set 36 - 14.

<u>1705B</u>: set 9 - 3,4,5,8,9,11,12; set 10 - 5; set 32 - 16.

<u>1706A</u>: set 6 - 1,2,3,11,12,13,17,18,21,23,26,27,28; set 8 - 22,23.

<u>1706B</u>: set 6 - 8,9,11,16,17,23,24,25,26.

<u>1706C</u>: set 6 - 6,11,13,14,17,19,20,22,23,24.

<u>1706D</u>: set 6 - 8,14,16,17,19,22,23,25,29.

<u>1707A</u>: set 10 - 11,12,13,14.

<u>1708A</u>: set 15 - 1,2,3,4,5,6,10,11,12,13,14,15,16,17,18,22,24, 43; set 16 - 4,5,7,10,11,13,14,15,16,17,28,29.

<u>1709A</u>: set 14 - 2,7,8,9,10,11,12,13,15,36,37,41,43,44,45,46, 47,49; set 37 - 32; set 45 - 26.

<u>1710A</u>: set 1 - 13,15,21,23,29,35,40,52.

<u>1710B</u>: set 1 - 8,9,11,26,31,37,38,52; set 2 - 3,17; set 5 - 1,3,13,15; set 6 - 24; set 13 - 60,61; set 21 - 59,75; set 31 - 18; set 32 - 9: set 43 - 2.

<u>1710C</u>: set 1 - 2,3,10,14,19,20,28,34,38,41,52,53.

<u>1711A</u>: set 2 - 8,9; set 3 - 10,12,13; set 5 - 14.

<u>1712A</u>: set 11 - 44,45,48,49,50,51,53.

<u>1713A</u>: set 8 - 12,13,14,16,31; set 9 - 23,24,26,28; set 10 - 8,19,31; set 11 - 16,29,32,41,59; set 14 - 14,29,30,33; set 15 - 37; set 18 - 2,6; set 21 - 23.

<u>1713B</u>: set 8 - 15,16; set 10 = 30,32.

<u>1714A</u>: set 4 - 16,17,18,19,20.

1714B: set 4 - 12,13,15,16,20; set 10 - 27.

1715A: set 8 - 12,13,14,28,31; set 9 - 23,24,26,28; set 10 - 8,19,31; set 11 - 16,29,32,41,59; set 14 - 14,29,30,33; set 15 - 37.

1716A: set 5 - 19,24,25,26,27,28,29,30,31.

1717A: set 15 - 13,14,16,23,25,33,34,45,47; set 16 - 2,12,18, 19,20,21; set 48 - 12.

1718A: set 14 - 16,17,18,19,20,21,22,23,24,25,26,27,28,31,32, 38,41,42,45; set 18 - 39,40.

1719A: set 8 - 12,13,14,31; set 9 - 19,23,24,26,28; set 10 - 8,10,19,31; set 11 - 16, 29, 32,41,59; set 14 - 14,29,30,33; set 15 - 37; set 18 - 2,6; set 21 - 23.

1719B: set 5 - 54,58; set 9 - 17,18,19,22; set 10 - 9,10; set 13 - 1; set 15 - 21.

1720A: set 2 - 12,16,25; set 3 - 5,7,27; set 4 - 1; set 5 - 1,20,48.

1720B: set 2 - 11,12,15,28; set 3 - 15,16,17,18,20,27; set 4 - 4,5,6,7,8,9,23; set 5 - 3,5,9,15,34,40,41,43,45,51,52,53, 56,59.

1721A: set 7 - 10,29; set 9 - 24,25; set 10 - 19.

1722A: set 11 - 45,46,49,50,51,52.

1722B: set 11 - 29,32.

1723A: set 13 - 36,37,38,39,40,41,46.

1723B: set 13 - 37,38,39,48,50,51,52; set 15 - 27,30,31,40, 42,44; set 16 - 24,27.

1724A: set 15 - 7,8,9,12,13,14,16,23,46,48; set 16 - 3,13,14, 15,20; set 22 - 22; set 24 - 23; set 48 - 11.

1725B: set 2 - 22; set 3 - 11,29; set 5 - 2,42,58,59; set 10 - 57; set 21 - 55; set 31 - 24.

1726A: set 11 - 12,13,14,15,17,18,21,22,24,25,26,27,30,35, 36,37,38,39; set 13 - 8,10,11,12,14,29,51,52,58; set 15 - 27, 28,29; set 16 - 26; set 27 - 1,2,6,9.

1726B: set 11 - 5,6,7,17,18,19,20,21,22,31,33,36; set 13 - 11,12,14,51; set 15 - 39; set 16 - 25; set 27 - 1.

1727A: set 1 - 17,18,24,32,43,44,45,47,54; set 42 - 6,10,15,

1727B: set - 12,16,22,27,39,41,42,44,45,46,47,51; set 42 - 6.

1728A: set 2 - 16,24; set 3 - 7,8; set 31 - 16.

1729A: set 10 - 26; set 13 - 57,60.

1729B: set 8 - 29; set 10 - 26; set 13 - 56.

1730A: set 8 - 3,4; set 10 - 20.

1730B: set 8 - 3,4; set 10 - 20.

1731A: set 5 - 36,37,38,39.

1732A: set 14 - 50,51.

1733A: set 8 - 11; set 9 - 13,29; set 11 - 1,38.

1735A: set 6 - 4,5,6,7,10,11,15,19,20; set 8 - 22.

1736A: set 11 - 44,45,46,47,48,49,50,51,52,53,54,55,56,57,60; set 27 - 11,12,13.

1737A: set 7 - 4,11,14.

1737B: set 7 - 12,13,14; set 14 - 3.

1738A: set 13 - 1,2,3,4,5,8,27,28,30,31; set 31 - 17,20,22, 23; set 35 - 2.

1739A: set 15 - 32,36; set 16 - 1,6,8,9,22,23.

2901A: set 18 - 36; set 21 - 28,29,30,31,35,36,37,46,47,49, 50; set 22 - 35,37,38,41,44,51; set 24 - 36,38,39,40.

2901B: set 22 - 18,24,35,44,45,51.

2902A: set 24 - 3; set 25 - 1,22,23,24,25,26,28,29,33,34,35, 37.

2903A: set 18 - 14,15,16,18; set 20 - 3.

2903B: set 18 - 10,11,12,13,16; set 20 - 4.

2904A: set 28 - 6,7,8,9,10,11,12,13,14,21,23,25,26.

2905A: set 11 - 46,49,51; set 27 - 11,12,13,14,16,17,18,19, 20,21,22,31.

2906A: set 26 - 56,57,58,59,60,62,63,68; set 27 - 32.

2906B: set 26 - 43,44,45,46,47,48,49,50,56,68; set 31 - 2,4.

2907A: set 23 - 4,5,6,7,8,9,10,11,12,13,15,16,18,24,25,26,27, 28,29,30,31.

2907B: set 23 - 5,6,7,8,9,10,11,12,14,15,16,18,24,25,26,27, 28,29,30,31.

2909A: set 18 - 33,35,36,37; set 19 - 5,6.

2909B: set 19 - 5,6,7.

2910A: set 23 - 17,19,20,21,23,25,30,32; set 27 - 1,3,4.

2911A: set 27 - 25.

2912A: set 20 - 14.

2912B: set 20 - 12,14.

2914A: set 15 - 13,14,15,16.

2915A: set 21 - 62.

2916A: set 22 - 8,9,10,11,12,13,14,15,16,17,18,19,20,21,22, 23,24,25,26,29,30,31,32,33,34,36,40,42,52,57; set 52 - 21,22.

2917A: set 11 - 44,45,46,48; set 27 - 11,12,13,16,17,18,19, 20,31.

2918A: set 24 - 19,20,21,22,31.

2918B: set 24 - 19,20,31.

2919A: set 22 - 61,62,63,64,65,66,67,68,69,70,71; set 24 - 26.

2920A: set 26 - 16,17,18,56,57,58,59,60; set 27 - 32.

2920B: set 26 - 10,11,15,19,44,46,47,50,56,57,58,59,60.

2921A: set 18 - 3,7,8,9; set 19 - 40,41,42,43,44,45,46,47; set 20 - 9,10,13,29,30; set 36 - 25.

2922A: set 18 - 17,18,19,25,26,27,28,29,30,31,32,33,34,35,36, 37,39,40; set 19 - 27,31,33,34; set 20 - 21.

2922B: set 19 - 1,2,3,4,8,9,10,11,12,13,14,15,16,17,27,30; set 20 - 20,26,28; set 40 - 1,6.

2922C: set 18 - 13; set 19 - 18,19,20,22,23,24,25,26,28,30, 32,33,36; set 20 - 5,22,23,24,25; set 21 - 1,2; set 35 - 17.

2922D: set 18 - 13; set 19 - 26,27,28,29,35,36; set 20 - 5, 15,16,17,18,19,32; set 27 - 24.

2923A: set 26 - 7,8,9,13,14,23,27,30,37,38,39,42,43,49,53,61, 65; set 27 - 5,7,28,30.

2925A: set 21 - 12,51,52,53,54,56,57,59,60,61,65,66,69,70,71, 72,74,75; set 22 - 43,47,48,50,53,55; set 23 - 2,22; set 24 - 27; set 48 - 5.

2925B: set 21 - 51,53,54,56,57,59,66,69,72,74,75; set 22 - 53,55; set 48 - 5.

2925C: set 21 - 51,53,54,56,57,59,66,69,70,72,74,75; set 22 - 53,55; set 23 - 1; set 48 - 5.

2926A: set 28 - 16,17.

2927A: set 24 - 25,28,29,30,32,33,34,41.

2928A: set 18 - 36; set 21 - 9,10,11,12,13,14,15,16,25,26,27, 28,29,31,34,35,36,46,47,50,57,65,68,73; set 33 - 27; set 45 - 15,16,18,21,22,29.

2928B: set 18 - 36; set 21 - 9,10,11,13,14,17,18,19,20,21,22, 23,24,26,27,29,31,34,35,36,46,47,50,57,65,68,73; set 33 - 27; set 45 - 15,16,18,20,22,29.

4101A: set 34 - 11.

4102A: set 30 - 23,24,32; set 31 - 9,12.

4103A: set 33 - 18,19.

4103B: set 33 - 19.

4104A: set 38 - 2,3,4,5,16,17,18.

4106A: set 30 - 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18, 19,23,24,25,31,32; set 31 - 18,19,21,27; set 33 - 10,20; set 37 - 27; set 38 - 14; set 39 - 2; set 40 - 3.

4107A: set 38 - 30; set 39 - 13,16,17,18,19,20,22,29.

4107B: set 39 - 21,23,24.

4108A: set 33 - 19,20,21,26,27,28,29,30.

4109A: set 36 - 6.

4109B: set 36 - 6.

4110A: set 40 - 27,28,39.

4111A: set 32 - 31,32; set 34 - 14,25,26,27,28,30,31.

4112A: set 38 - 8,10,11,12,13,27; set 39 - 1.

4112B: set 38 - 5,6,7,9,12,26,28; set 39 - 1,4; set 40 - 9.

4112C: set 31 - 8; set 33 - 11; set 38 - 15,23,28; set 39 - 1; set 44 - 31,32,33.

4113A: set 30 - 4,7,8,9,10,19.

4113B: set 30 - 4,7,8,9,10,19.

4113C: set 30 - 5,6,7,8,9,10,19.

4114A: set 35 - 6,7,10,12.

4114B: set 35 - 9,10,11,12,16.

4114C: set 35 - 13,14.

4114D: set 35 - 2,3,5,6,7,8,21,22,24,28,29.

4114E: set 35 - 7.

4115A: set 32 - 1,15; set 33 - 23; set 37 - 1,3,4,5,11,15,18, 20,21; set 38 - 1.

4115B: set 31 - 14; set 32 - 1,2,14,15; set 36 - 14; set 37 - 2,3,4,5,11,14,18,20,21,25; set 38 - 29; set 39 - 12,14; set 45 - 32.

4115C: set 37 - 8,9,10,12,13,15,16,17,19,20,22.

4115D: set 37 - 6,7,10,12,13,14,16,17,19,20,26,31,32; set 39 - 11.

4116A: set 32 - 18,19,20,21,22,26,27,28,29,30,35; set 33 - 3, 7; set 34 - 1,2,8,9,10,12,13,15,16,17,18,19,20,21; set 35 - 19; set 45 - 12,28.

4116B: set 32 - 3,4,10,11,13,19,27,35; set 33 - 1,7,8; set 36 - 11; set 45 - 9,12.

4117A: set 30 - 20,27.

4117B: set 30 - 21,26,27.

4118A: set 32 - 5,12,13; set 36 - 3,4,9,10,12; set 39 - 7.

4118B: set 36 - 1,2,5,8,13,14.

4119A: set 32 - 37,38.

4119B: set 32 - 36,37,38; set 38 - 14; set 39 - 3.

4120A: set 35 - 3,7,8,12,15,24,27,30; set 37 - 29,30; set 38 - 28.

4120B: set 35 - 8,24,27,28,30; set 37 - 29,30; set 38 - 28.

4120C: set 31 - 8; set 33 - 11; set 35 - 4,12,15,25,26,27,30; set 37 - 30; set 38 - 28.

4121A: set 31 - 1,2,3,5,6,7,8,10,11,12,28,29,30,31,32.

4122A: set 33 - 25,26.

5302A: set 44 - 18,20,21,22,27.

5302B: set 43 - 29; set 44 - 23,24,25,26,27,28,30,31.

5303A: set 51 - 25,26,27,28.

5303B: set 51 - 23,24,27,28.

5304A: set 39 - 25,26,27,28; set 48 - 23,24,25,26,27,28,29, 30,33; set 52 - 4,9.

5305B: set 49 - 23,24.

5306A: set 48 - 1,2,13,14,15,19,31.

5306B: set 48 - 16,17,18,19,20,26,32.

5307A: set 44 - 7,8,9,10,11,12.

5308A: set 47 - 7,11,12,14.

5308B: set 47 - 7,8,9,11,12,13,14.

5308C: set 47 - 7,8,9,12,13,14.

5309A: set 43 - 9.

5309B: set 43 - 9,10.

5310A: set 47 - 9,10,14,25,28,31,32.

5310B: set 47 - 9,10,14,28,29,32.

5311A: set 22 - 49; set 52 - 23,24.

5312A: set 50 - 3,5,10,11,12,13; set 51 - 20,21,22.

5312B: set 50 - 9,10,12.

5313A: set 46 - 1,2,3,4,5,6,7,8,9,16.

5313B: set 46 - 10,11,12,13,14,15,16,17.

5314A: set 42 - 1,2,3,6,8,9,25,26,27.

5314B: set 42 - 1,2,3,4,5,8,9,25,27.

5314C: set 42 - 1,2,3,7,8,25,28; set 51 - 1.

5315A: set 30 - 28; set 51 - 16,17,18,19.

5316A: set 43 - 17,18,19.

5317A: set 52 - 7,8,10,11,25,26,28.

5317B: set 52 - 8,11,12,15,25.

5318A: set 51 - 8,9,10,11,12,13,14.

5319A: set 49 - 10,11,12,13,14,15,16,17,18,19,20,21,22,23,24, 25,26,27,28,29,30.

5320A: set 46 - 17,18,19,20,21,23,24,25,27,30; set 47 - 15, 16,17.

5320B: set 46 - 17,18,19,20,22,23,24,25,26,27,30; set 47 - 15,16,17.

5321A: set 22 - 49; set 52 - 23,24.

5322A: set 50 - 16,17,18,19,20,21,22,24,25,27; set 51 - 22.

5322B: set 50 - 15,16,17,19,21.

5323A: set 45 - 15,16,17,18,19,20,21,24,27,28,29,30.

5324B: set 42 - 10,11,12,13,14,15,16,17,18,23,24,26.

5324C: set 42 - 15,16,17,18,19,23,24,29,30.

5324D: set 42 - 20,21,22,23,29,30,31,32,33.

5325A: set 47 - 1,2,3,4,5,6,18,19,20,21,22,23,24,30,33.

5325B: set 47 - 1,2,3,4,5,6,17,18,19,20,21,22,23,24,30.

5326A: set 43 - 21; set 51 - 15,31.

APPENDIX C

Item Difficulty and Gain Score Indexes Associated with the Criterion Test Items

Item	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)*	(9)*
1701a	.017	.459	.672	.442	.655	.449	.667	2.545	3.659
1701b	.000	.459	.770	.459	.770	.459	.770	2.585	4.148
1702a	.000	.508	.475	.508	.475	.508	.475	2.811	2.819
1702b	.167	.459	.492	.292	.325	.350	.390	2.194	2.434
1703a	.217	.705	.918	.488	.701	.623	.895	3.213	4.215
1703b	.217	.705	.885	.488	.688	.623	.853	3.213	4.066
1703c	.050	.721	.885	.671	.835	.706	.879	3.678	4.527
1704a	.483	.951	.852	.468	.369	.905	.714	3.727	3.182
1704b	.500	.934	.820	.434	.320	.868	.640	3.609	2.992
1704c	.000	.672	.688	.672	.688	.672	.688	3.569	3.778
1705a	.150	.803	.934	.653	.784	.768	.922	3.823	4.472
1705b	.150	.770	.918	.620	.768	.729	.904	3.670	4.400
1706a	.017	.885	.885	.868	.868	.883	.883	4.513	4.619
1706b	.050	.836	.934	.786	.884	.827	.930	4.210	4.748
1706c	.017	.951	.902	.934	.885	.950	.885	4.818	4.695
1706d	.017	.885	.885	.868	.868	.883	.883	4.513	4.619
1707a	.067	.426	.574	.359	.507	.385	.543	2.275	3.079
1708a	.067	.623	.721	.556	.654	.596	.701	3.186	3.742
1709a	.067	.606	.721	.539	.654	.650	.701	3.107	3.742
1710a	.033	.524	.623	.491	.590	.508	.610	2.808	3.394
1710b	.033	.590	.721	.557	.688	.576	.711	3.113	3.836
1710c	.400	.902	.852	.502	.452	.837	.753	3.695	3.412
1711a	.017	.410	.393	.393	.376	.400	.382	2.318	2.402
1713a	.033	.820	.836	.787	.803	.814	.830	4.175	4.354
1713b	.000	.492	.524	.492	.524	.492	.524	2.737	3.039
1714a	.433	.885	.934	.452	.501	.797	.884	3.539	3.690
1714b	.550	.918	.934	.368	.384	.669	.853	3.417	3.366
1715a	.017	.852	.869	.835	.852	.849	.867	4.361	4.546
1716a	.300	.574	.902	.274	.602	.391	.860	2.414	3.913
1717a	.033	.639	.574	.606	.541	.627	.559	3.339	3.173
1718a	.000	.443	.426	.443	.426	.443	.426	2.511	2.598
1719a	.600	.984	.918	.384	.318	.960	.795	3.606	3.156
1719b	.000	.574	.606	.574	.606	.574	.606	3.116	3.409
1720a	.067	.688	.639	.621	.572	.642	.592	3.486	3.372
1720b	.167	.738	.705	.571	.538	.685	.646	3.483	3.393
1721a	.583	.999	.984	.417	.401	.999	.962	3.715	3.500
1722a	.033	.393	.426	.360	.393	.372	.406	2.202	2.507
1722b	.050	.705	.672	.655	.622	.689	.655	3.604	3.568
1723a	.050	.426	.590	.376	.540	.396	.568	2.315	3.198
1723b	.367	.918	.836	.551	.469	.870	.741	3.846	3.431
1724a	.000	.738	.656	.738	.656	.738	.656	3.874	3.634
1725b	.017	.688	.492	.671	.475	.683	.483	3.603	2.848
1726a	.100	.754	.541	.654	.441	.727	.490	3.714	2.840

Item	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)*	(9)*
1726b	.083	.705	.492	.622	.409	.678	.446	3.527	2.666
1727a	.233	.738	.738	.505	.505	.658	.658	3.328	3.359
1727b	.100	.885	.918	.785	.818	.872	.909	4.319	4.538
1728a	.033	.361	.148	.328	.115	.339	.119	2.046	1.254
1729a	.183	.574	.524	.391	.341	.478	.417	2.688	2.534
1729b	.050	.590	.508	.540	.458	.568	.482	3.729	2.829
1730a	.017	.475	.590	.458	.573	.466	.583	2.619	3.290
1730b	.200	.393	.393	.193	.193	.241	.241	.597	2.151
1731a	.000	.262	.410	.262	.410	.262	.410	1.674	2.526
1732a	.250	.852	.869	.602	.619	.803	.825	3.815	3.903
1733a	.000	.246	.361	.246	.361	.246	.361	1.600	2.305
1735a	.200	.836	.803	.636	.603	.795	.754	3.858	3.743
1736a	.000	.262	.328	.262	.328	.262	.328	1.674	2.156
1737a	.167	.705	.902	.538	.735	.646	.882	3.330	4.281
1737b	.317	.705	.738	.388	.421	.568	.616	2.979	3.127
1738a	.000	.016	.115	.016	.115	.016	.115	.538	1.197
1739a	.000	.885	.443	.885	.443	.885	.443	4.553	2.674
2901a	.000	.738	.475	.738	.475	.738	.475	3.874	2.819
2901b	.033	.672	.672	.639	.639	.661	.661	3.492	3.615
2902a	.000	.672	.475	.672	.475	.672	.475	3.569	2.819
2903a	.000	.688	.672	.688	.672	.688	.672	3.643	3.706
2903b	.000	.738	.787	.738	.787	.738	.787	3.874	4.224
2904a	.000	.606	.459	.606	.459	.606	.459	3.264	2.747
2905a	.000	.377	.328	.377	.328	.377	.328	2.206	2.156
2906a	.617	.590	.738	-.027	.121	-.070	.316	1.745	2.298
2906b	.433	.623	.738	.190	.305	.335	.538	2.329	2.807
2907a	.167	.918	.836	.751	.669	.902	.839	4.315	3.983
2907b	.300	.623	.557	.323	.257	.461	.367	2.640	2.359
2909a	.017	.705	.606	.688	.589	.700	.599	3.682	3.618
2909b	.017	.656	.688	.639	.671	.650	.683	3.455	3.731
2910a	.050	.738	.639	.688	.589	.724	.620	3.757	3.419
2911a	.100	.623	.639	.523	.539	.581	.599	3.108	3.281
2912a	.117	.492	.557	.375	.440	.425	.498	3.463	2.865
2912b	.000	.279	.377	.279	.377	.279	.377	1.253	2.377
2914a	.000	.148	.049	.148	.049	.148	.049	1.148	.900
2915a	.117	.475	.524	.358	.407	.405	.461	2.385	2.716
2915b	.017	.541	.590	.524	.573	.533	.583	2.924	3.290
2916a	.067	.721	.639	.654	.572	.785	.687	3.638	3.372
2917a	.083	.524	.492	.441	.409	.481	.446	2.691	2.666
2918a	.200	.885	.934	.685	.734	.856	.918	4.085	4.334
2918b	.100	.885	.803	.785	.703	.872	.781	4.319	4.020
2919a	.000	.361	.180	.361	.180	.361	.180	2.132	1.490
2920a	.017	.885	.721	.868	.704	.883	.716	4.514	3.880
2920b	.033	.639	.557	.606	.524	.627	.542	3.339	3.097
2921a	.167	.770	.787	.603	.620	.724	.744	3.631	3.763
2922a	.450	.508	.557	.058	.107	.105	.194	1.757	1.944
2922b	.100	.459	.393	.359	.293	.399	.326	2.351	2.173
2922c	.200	.410	.393	.210	.193	.262	.241	1.890	1.897

Item	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)*	(9)*
2922d	.150	.443	.443	.293	.293	.345	.345	2.160	2.260
2923a	.367	.820	.738	.453	.371	.716	.586	3.393	2.989
2925a	.567	.885	.918	.318	.351	.734	.811	3.225	3.247
2925b	.550	.869	.885	.319	.335	.709	.744	3.191	3.146
2925c	.183	.885	.902	.702	.719	.859	.880	4.125	4.236
2926a	.000	.770	.688	.770	.688	.770	.688	4.022	3.778
2927a	.000	.066	.164	.066	.164	.066	.164	.769	1.418
2928a	.150	.803	.852	.653	.702	.768	.826	3.823	4.102
2928b	.083	.787	.721	.704	.638	.768	.696	3.906	3.697
4101a	.750	.738	.885	-.012	.135	-.048	.540	2.118	2.593
4102a	.183	.820	.820	.637	.637	.780	.780	3.824	3.867
4103a	.000	.541	.557	.541	.557	.541	.557	2.964	3.188
4103b	.317	.738	.688	.421	.371	.616	.543	3.132	2.902
4104a	.000	.295	.230	.295	.230	.295	.230	1.827	1.715
4106a	.017	.623	.606	.606	.589	.616	.599	3.303	3.362
4107a	.000	.508	.459	.508	.459	.508	.459	2.811	2.747
4107b	.033	.639	.574	.606	.541	.627	.559	3.339	3.173
4108a	.317	.639	.738	.322	.421	.471	.616	2.674	3.127
4109a	.167	.852	.902	.685	.735	.822	.882	4.010	4.281
4109b	.033	.705	.475	.672	.442	.695	.457	3.644	2.727
4110a	.[illegible]50	.885	.754	.535	.404	.823	.622	3.734	3.108
4111a	.[illegible]0	.410	.443	.210	.243	.262	.304	1.890	2.122
4112a	.100	.656	.688	.556	.588	.618	.653	3.261	3.502
4112b	.483	.885	.902	.402	.419	.778	.810	3.422	3.407
4112c	.050	.721	.590	.671	.540	.706	.568	3.678	3.198
4113a	.400	.885	.902	.485	.502	.808	.837	3.616	3.637
4113b	.233	.705	.721	.472	.488	.615	.636	3.176	3.283
4113c	.533	.885	.685	.352	.352	.754	.754	3.305	3.193
4114a	.467	.688	.738	.221	.271	.415	.568	2.549	2.713
4114b	.300	.623	.574	.323	.274	.461	.391	2.640	2.436
4114c	.417	.705	.590	.288	.173	.494	.297	2.745	2.184
4114d	.117	.361	.393	.244	.276	.276	.312	1.858	2.126
4114e	.067	.426	.311	.359	.244	.385	.262	2.275	1.895
4115a	.417	.672	.672	.255	.255	.437	.437	2.592	2.554
4115b	.417	.574	.656	.157	.239	.269	.410	2.140	2.482
4115c	.400	.590	.688	.190	.288	.317	.480	2.253	2.673
4115d	.383	.623	.705	.240	.322	.389	.522	2.446	2.796
4116a	.117	.787	.705	.670	.588	.759	.666	3.826	3.531
4116b	.000	.508	.557	.508	.557	.508	.557	2.811	3.188
4117a	.050	.606	.606	.556	.556	.585	.585	3.147	3.271
4117b	.050	.574	.656	.524	.606	.552	.638	2.999	3.496
4118a	.017	.246	.295	.229	.278	.233	.283	1.561	1.961
4118b	.017	.164	.230	.147	.213	.150	.217	1.182	1.668
4119a	.333	.820	.852	.487	.519	.730	.778	3.473	3.597
4119b	.017	.508	.557	.491	.540	.499	.549	2.771	3.141
4120a	.083	.524	.557	.441	.474	.481	.517	2.691	2.959
4120b	.233	.721	.705	.488	.472	.636	.615	3.250	3.211

Item	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)*	(9)*
4120c	.000	.361	.213	.361	.213	.361	.213	2.132	1.638
4121a	.833	.557	.574	-.276	-.259	-.999	-.999	1.087	.963
4122a	.467	.934	.885	.467	.418	.876	.784	3.686	3.375
5302a	.000	.115	.148	.115	.148	.115	.148	.995	1.346
5302b	.000	.475	.443	.475	.443	.475	.443	2.659	2.674
5303a	.083	.590	.541	.507	.458	.553	.499	2.996	2.887
5303b	.033	.869	.738	.836	.705	.864	.729	4.402	3.912
5304a	.117	.902	.541	.785	.424	.889	.480	4.358	2.793
5305b	.317	.738	.803	.421	.486	.616	.712	3.132	3.420
5306a	.083	.770	.623	.687	.540	.749	.589	3.827	3.256
5306b	.000	.787	.770	.787	.770	.787	.770	4.100	4.148
5307a	.300	.492	.574	.192	.274	.274	.391	2.035	2.436
5308a	.133	.541	.426	.408	.293	.470	.338	2.652	2.230
5308b	.000	.164	.098	.164	.098	.164	.098	1.222	1.120
5308c	.000	.131	.164	.131	.164	.131	.164	1.069	1.418
5309a	.883	.967	.984	.084	.101	.718	.863	2.864	2.671
5309b	.283	.787	.787	.504	.504	.703	.703	3.438	3.442
5310a	.100	.738	.705	.638	.605	.709	.672	3.640	3.578
5310b	.000	.377	.180	.377	.180	.377	.180	2.206	1.490
5311a	.050	.721	.590	.671	.540	.706	.690	3.678	3.198
5312a	.217	.492	.590	.275	.373	.351	.476	2.229	2.737
5312b	.283	.557	.590	.274	.307	.382	.428	2.375	2.555
5313a	.183	.623	.541	.440	.358	.538	.438	2.914	2.610
5313b	.000	.656	.541	.656	.541	.656	.541	3.495	3.116
5314a	.217	.787	.557	.570	.340	.728	.434	3.592	2.588
5314b	.300	.918	.754	.618	.454	.883	.648	4.003	3.246
5314c	.467	.836	.656	.369	.189	.692	.354	3.233	2.343
5315a	.000	.590	.606	.590	.606	.590	.606	3.190	3.409
5316a	.367	.328	.492	-.039	.125	-.062	.197	1.120	1.881
5317a	.000	.820	.311	.820	.311	.820	.311	4.253	2.080
5317b	.067	.738	.475	.671	.408	.719	.437	3.717	2.633
5318a	.000	.410	.115	.410	.115	.410	.115	2.358	1.197
5319a	.600	.902	.902	.302	.302	.755	.755	3.227	3.084
5320a	.133	.410	.361	.277	.228	.319	.263	2.047	1.938
5320b	.000	.426	.328	.426	.328	.426	.328	2.432	2.156
5321a	.017	.459	-.079	.442	.262	.450	.266	2.545	1.889
5322a	.017	.721	.295	.704	.278	.716	.283	3.756	1.961
5322b	.017	.295	.164	.278	.147	.283	.150	1.787	1.371
5323a	.083	.820	.803	.737	.720	.804	.785	4.058	4.067
5324b	.300	.606	.574	.306	.274	.437	.391	2.561	2.436
5324c	.067	.377	.377	.310	.310	.332	.332	2.049	2.192
5324d	.000	.033	.016	.033	.016	.033	.016	.616	.751
5325a	.033	.524	.524	.491	.491	.508	.508	2.808	2.948
5325b	.050	.524	.541	.474	.491	.499	.517	2.768	2.978
5326a	.367	.885	.918	.518	.551	.818	.870	2.694	3.800

* A constant of 3.000 was added.

Legend for Appendix C

1. Control Group Item Difficulty (N = 60)
2. Immediate Posttest Item Difficulty (N = 61)
3. Delayed Posttest Item Difficulty (N = 61)
4. Immediate Posttest Absolute Gain Score
5. Delayed Posttest Absolute Gain Score
6. Immediate Posttest Relative Gain Score
7. Delayed Posttest Relative Gain Score
8. Immediate Posttest Residual Gain Score
9. Delayed Posttest Residual Gain Score

APPENDIX D

Definitions, Means, Standard Deviations, and Inter-rater Reliability Coefficients of the Program Characteristics

1. Total Frames: the total number of frames in the "frame sequence" associated with a criterion test item. A frame sequence consists of all the frames intended to instruct the answer to the test item. Identification of these "teaching" frames was necessarily subjective and was guided by such cues as the appearance of the key term in a frame, particularly as a response, and the set to which readers are referred for review.

   Mean: 9.70 Standard Deviation: 7.30 Reliability: not obtained

2. Initial Learning Frames: the total number of frames related to a particular test item (A) which present new information relevant to answering the test item and (B) which follow the "new information" frames by less than five consecutive frames unrelated to the test item. (The cut-off point of five consecutive frames was chosen arbitrarily.)

   Mean: 5.70 Standard Deviation: 4.62 Reliability: .89

3. Review Frames: the total number of frames within a frame sequence separated from initial learning frames by five or more consecutive frames unrelated to the test item.

Mean: 3.99 Standard Deviation: 4.84 Reliability: .56

4. Exhibit Referrals: the number of explicit references to exhibits contained within a frame sequence.

   Mean: 1.72 Standard Deviation: 3.20 Reliability: .98

5. Random Variable: A normal distribution of numbers was generated with a mean of 20 and a standard deviation of 5. These numbers were then randomly assigned to each frame sequence. Since 27 items were discarded before the final analysis, this variable actually had a mean and a standard deviation slightly different from the values originally specified.

   Mean: 20.06 Standard Deviation: 4.82 Reliability: assumed to be 0.00

6. Key Term as Stimulus: the number of times that the key term appears in the stimulus (printed) portion of the frame sequence. The key term is the word or words that make up the answer to the criterion test item.

   Mean: 3.48 Standard Deviation: 3.75 Reliability: .50

7. Key Term as Response: the number of times that the frame sequence requires the learner to supply the key term wholly or in large part.

   Mean: 4.30 Standard Deviation: 4.47 Reliability: .93

8. Letters in Key Term: the number of letters in the key term.

   Mean: 9.46 Standard Deviation: 4.27 Reliability: assumed to be 1.00

9. Syllables in Key Term: the number of syllables in the key

term.

Mean: 3.26 Standard Deviation: 1.58 Reliability: assumed to be 1.00

10. Formal Prompts: the total number of cues within a frame sequence regarding the form or appearance of a response. In this program, formal prompts were of three types: (a) the symbol (TT) indicating that the desired response is a "technical term," (b) two answer blanks indicating that the desired response consists of two words, and (c) a partial spelling of the response indicating that the learner should complete the spelling of the desired response.

Mean: 1.65 Standard Deviation: 2.71 Reliability: .18

11. Thematic and Sequence Prompts: the total number of cues regarding the meaning of the desired response. This type of prompt usually took the form of a word which was identical or highly similar to the desired response and which appeared in the same frame as the desired response or in one of the three frames preceding the frame containing the response in question. (The cut-off point of three preceding frames was arbitrarily chosen.)

Mean: 4.90 Standard Deviation: 4.70 Reliability: .82

12. Maximum Words per Frame: the number of words contained in the frame with the most words.

Mean: 34.64 Standard Deviation: 5.46 Reliability: .95

13. Maximum Syllables per Frame: the number of syllables

contained in the frame with the most syllables.

Mean: 61.29 Standard Deviation: 9.39 Reliability: .92

14. Affixed Word Ratio: ratio of suffixes and prefixes to the total number of words within a frame sequence.

Mean: .46 Standard Deviation: .10 Reliability: .52

15. Program Order Relative to Immediate Posttest: the rank order in which the initial learning frames appeared in the program relative to the immediate posttest. Thus, items 17, 29, 41, and 53 were instructed earliest among the items tested in their respective immediate posttests. These four items each received the rank of 01.

Mean: 24.14 Standard Deviation: 14.57 Reliability: not obtained

16. Personal Word Ratio: ratio of "personal" words to the total number of words within a frame sequence. A "personal" word refers to one or more persons and may be (a) a noun with natural gender, (b) a personal pronoun, (c) a collective noun such as "people," or (d) an adjective with gender.

Mean: .03 Standard Deviation: .04 Reliability: .40

17. Concrete Noun Ratio: ratio of concrete nouns to the total number of nouns within a frame sequence. A concrete noun is any noun referring to an object that can be photographed.

Mean: .19 Standard Deviation: .13 Reliability: .48

18. Cloze Score: a measure of reading ease. The procedure

consists of (1) deleting every nth word (in this case, every 5th word) from the text, (2) replacing each word with a blank of standard size, (3) asking 15 Ss, not involved in the main part of the study, to fill in the missing blanks, (4) counting up the correct insertions, and (5) comparing the percentage of correct replacements among the various frame sequences. Synonyms were scored as incorrect with the exception that the words "reflex" and "respondent" were considered interchangeable. Misspellings were accepted if it was apparent that the S was attempting to write the correct answer.

Mean: .57 Standard Deviation: .09 Reliability: not obtained

19. Maximum Cloze Score: the highest cloze score obtained by a frame in the frame sequence.

Mean: .78 Standard Deviation: .15 Reliability: not obtained

20. Minimum Cloze Score: the lowest cloze score obtained by a frame in the frame sequence.

Mean: .35 Standard Deviation: .13 Reliability: not obtained

21. Percent of Responses in First Third of a Frame: the percentage of responses in the frame sequence which are required within the first third of a frame's words.

Mean: .05 Standard Deviation: .11 Reliability: .97

22. Intentional Sets: the total number of sets containing frames intended to teach the key term.

Mean: 1.93 Standard Deviation: 1.32 Reliability: not obtained

23. Vehicular Sets: the number of sets containing frames which incidentally present the key term as part of the contextual material intended to instruct another key term.

Mean: 2.18 Standard Deviation: 2.41 Reliability: not obtained

24. Alternatives Accepted: the number of alternative answers accepted as correct for the criterion test item.

Mean: 2.20 Standard Deviation: 1.91 Reliability: not obtained

25. Syllables per Word: the number of syllables per word within a frame sequence.

Mean: 1.81 Standard Deviation: .17 Reliability: .44

26. Sentences per Frame: the number of sentences per frame within a frame sequence.

Mean: 1.44 Standard Deviation: .32 Reliability: .98

27. Applications per Frame: the ratio of applications, illustrations, examples within a frame sequence to the number of frames.

Mean: .60 Standard Deviation: .30 Reliability: .29

28. Rules per Frame: the ratio of rules, generalizations, conclusions within a frame sequence to the number of frames.

Mean: .47 Standard Deviation: .30 Reliability: .28

29. Shared Frames Ratio: the percentage of frames within a sequence which are also included within another frame sequence. Thus, a "shared" frame instructs more than

one key term.

Mean: .60 Standard Deviation: .78 Reliability: .44

30. Modifiers per Sentence: the number of adjectives, adverbs and possessive nouns per sentence within a frame sequence.

Mean: 5.82 Standard Deviation: 1.58 Reliability: .87

31. Responses per Frame: the number of responses per frame within a frame sequence. A multi-word response was considered as one response.

Mean: 1.44 Standard Deviation: .34 Reliability: .81

32. Verbals per Frame: the number of gerunds, infinitives and participles per sentence within a frame sequence.

Mean: 1.26 Standard Deviation: .63 Reliability: .65

33. Complete Program Order: the rank order in which the initial learning frames first appear in the program, regardless of their relation to the immediate posttest.

Mean: 92.07 Standard Deviation: 52.81 Reliability: not obtained

34. Vehicular Uses of Key Term: the number of frames which incidentally present the key term as part of the contextual material intended to instruct another key term.

Mean: 10.58 Standard Deviation: 21.94 Reliability: not obtained

35. Prepositional Phrases per Sentence: the number of prepositional phrases per sentence within a frame sequence.

Mean: 2.18 Standard Deviation: .84 Reliability: .91

36. Clauses per Sentence: the number of clauses per sentence

within a frame sequence.

Mean: 1.80   Standard Deviation: .46   Reliability: .65

37. Words per Sentence: the number of words per sentence within a frame sequence.

Mean: 18.13   Standard Deviation: 4.07   Reliability: 1.00

APPENDIX E

Intercorrelations among the Independent Variables*

	1**	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	-	76	78	48	07	71	66	14	10	61	87	40	34	-13	-01	-10	13	26	60
2		-	19	41	-01	61	33	14	13	43	77	35	38	-03	-03	-09	10	19	44
3			-	33	11	48	68	07	03	51	57	27	14	-17	01	-06	10	21	48
4				-	-05	36	14	09	10	27	33	20	18	-20	-01	-26	05	19	32
5					-	-10	12	06	07	10	05	03	03	11	07	02	-07	-02	02
6						-	30	18	19	38	59	32	37	-05	05	-16	04	17	41
7							-	11	08	51	61	29	22	-09	-03	-00	13	18	35
8								-	90	21	07	01	11	14	19	04	03	03	10
9									-	26	06	02	16	14	14	03	-02	-01	07
10										-	55	17	27	06	04	-07	08	10	28
11											-	39	35	-06	01	-05	08	21	49
12												-	70	-22	07	-05	11	35	41
13													-	21	01	-17	-08	13	25
14														-	10	-10	-21	-34	-22
15															-	08	18	05	-04
16																-	27	-21	-20
17																	-	03	15
18																		-	70
19																			-

	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
1	-38	-14	54	26	-10	-18	-03	23	-22	-09	-10	-07	-03	-20	19	-10	-03	-13
2	-25	-09	14	-06	-02	-14	-05	15	-17	-08	-11	-05	-13	-01	06	-02	-07	-11
3	-33	-12	69	44	-13	-14	-09	20	-17	-06	-05	-05	08	-29	23	-13	03	-09
4	-06	-05	23	12	-04	-15	-05	17	-18	-02	02	11	-09	-13	06	17	-06	06
5	-11	-03	09	-04	01	13	-03	10	-08	-04	-03	-10	01	-01	-07	-16	01	-02
6	-25	-11	34	11	-18	-06	04	11	-07	-02	00	-14	-09	-08	15	-01	-06	-09
7	-24	-11	53	32	-24	-14	-02	13	-11	-05	-13	02	07	-33	22	-22	00	-15
8	-03	-10	06	-06	-17	08	06	09	-11	08	-01	-18	12	06	-06	-13	02	-05
9	-01	-05	08	-06	-15	15	11	11	-10	03	01	-21	06	09	-05	-11	-01	-06
10	-19	-12	42	14	-12	03	07	15	-09	-06	-06	-08	02	-14	07	-13	-04	-12
11	-32	-10	33	12	-09	-15	03	16	-14	-13	-06	-13	-01	-14	12	-08	01	-09
12	-03	-13	16	-07	-03	-40	30	12	-02	-06	-12	-02	03	08	-05	11	07	08
13	-09	-06	13	-16	02	18	24	-10	18	-09	05	-13	12	18	-13	17	02	09
14	-19	08	-04	-21	09	71	-03	-29	27	-06	08	-29	24	23	-27	-07	04	-05
15	05	-06	-02	-08	-05	-08	19	-01	-00	-08	-06	-22	13	01	-12	-08	03	-04
16	-06	-15	-09	-04	-10	-16	16	22	-24	-00	-23	-05	-10	26	-02	-37	24	-09
17	-18	-10	-00	-07	10	-40	26	40	-40	-07	-33	14	-04	-12	01	-25	-13	-31
18	30	-25	23	12	-06	-27	02	09	-07	-04	-15	22	04	-17	04	00	-01	-10
19	-07	-26	36	22	01	-23	10	23	-22	-04	-29	03	-02	-17	18	-15	-20	-31
20	-	-04	-14	-04	-11	-07	-03	-16	16	02	14	19	07	-00	-04	17	16	20
21		-	-11	-01	05	10	-13	-22	21	-07	25	04	06	03	00	35	-07	20
22			-	55	-09	04	02	11	-07	-03	-16	01	04	-30	21	-16	01	-16
23				-	-11	-06	-05	02	-00	01	-16	18	-01	-33	73	-18	04	-17
24					-	08	01	-00	03	-11	02	-10	-10	06	-08	06	-07	-02
25						-	-08	-42	39	-03	17	-19	17	13	-15	-00	-05	-05
26							-	10	-01	-11	-58	-05	-34	18	01	-34	-37	-64
27								-	-93	06	-21	06	-11	15	06	-19	14	-06
28									-	-10	19	-10	07	-13	-06	20	-17	03
29										-	-20	18	00	09	-01	19	05	23
30											-	-09	33	-00	-12	44	32	71
31												-	-11	-11	19	14	07	06
32													-	-05	-11	05	31	34
33														-	-27	01	11	10
34															-	-11	00	-16
35																-	-05	59
36																	-	62
37																		

* Correlations .21 or greater are significant at the .05 level.
Correlations .27 or greater are significant at the .01 level.

** The numbers refer to the independent variables as listed in Appendix B.

Decimal points are omitted.

APPENDIX F

Validity Coefficients for Immediate Posttest Administration*

Variable	Absolute Gain Score		Relative Gain Score		Residual Gain Score	
	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
1	.20	.03	.20	.07	.19	.12
2	.14	-.16	.06	-.10	.10	-.08
3	.19	.19	.26	.19	.21	.24
4	.02	-.11	.01	-.08	.00	-.09
5	-.12	.21	-.10	.18	-.09	.27
6	.15	-.06	.07	.00	.08	-.02
7	.35	.34	.32	.26	.32	.28
8	.16	.24	.04	.13	.04	.11
9	.09	.13	-.01	.03	-.03	.01
10	.12	-.06	.14	-.04	.14	-.10
11	.19	-.04	.15	-.05	.15	.03
12	.24	.20	.18	.14	.21	.24
13	.19	.10	.07	.05	.17	.12
14	-.02	-.10	-.08	-.19	-.05	-.14
15	.04	.19	-.08	.15	-.05	.12
16	.13	.09	.11	.13	.09	.14
17	.26	.06	.20	.06	.21	.03
18	.06	.09	.10	.09	.05	.11
19	.11	-.05	.11	-.09	.08	.00
20	-.07	.10	-.02	.03	-.03	.05
21	-.08	-.02	.03	-.09	-.04	-.01
22	.14	.12	.24	.11	.19	.14
23	-.02	.03	.08	-.03	.04	.10
24	-.27	-.12	-.08	-.02	-.05	-.02
25	-.16	-.15	-.11	-.18	-.10	-.17
26	.12	-.06	-.05	-.16	.04	-.13
27	.06	-.12	.04	-.07	.02	-.11
28	-.06	.16	-.06	.12	-.02	.17
29	-.05	.08	.03	.10	.03	.07
30	-.03	.10	.00	.16	-.05	.13
31	.15	.01	.20	.07	.22	.04
32	.05	.16	.04	.09	.03	.11
33	-.20	-.13	-.23	-.09	-.21	-.10
34	.01	-.01	.10	-.06	.07	-.02
35	-.21	.07	-.08	.04	-.08	.05

Variable	Absolute Gain Score		Relative Gain Score		Residual Gain Score	
	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
36	.14	.09	.16	.24	.13	.18
37	-.07	.19	.04	.28	.01	.27

* Correlations .21 or greater are significant at the .05 level.

Correlations .27 or greater are significant at the .01 level.

APPENDIX G

Validity Coefficients for Delayed Posttest Administration*

Variable	Absolute Gain Score		Relative Gain Score		Residual Gain Score	
	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
1	.14	.05	.10	.06	.11	.10
2	.10	-.13	.02	-.07	.06	-.08
3	.12	.18	.13	.15	.12	.21
4	-.01	.01	-.04	.02	-.03	.03
5	-.11	.16	-.08	.15	-.09	.19
6	.07	-.06	-.02	.00	.01	-.04
7	.27	.29	.18	.20	.22	.25
8	.02	-.01	-.17	-.06	-.09	-.13
9	-.01	-.06	-.19	-.11	-.13	-.18
10	.08	-.09	.08	-.10	.09	-.14
11	.14	-.10	.06	-.11	.09	-.06
12	.16	.12	.08	.07	.12	.13
13	.15	-.06	.06	-.07	.12	-.07
14	-.04	-.25	-.07	-.30	-.06	-.31
15	-.02	.13	-.15	.09	-.10	.07
16	.11	-.06	.05	-.01	.07	-.05
17	.29	.08	.18	-.07	.23	.09
18	.08	.08	.05	.08	.08	.11
19	.08	-.05	.00	-.09	.05	-.03
20	-.03	-.02	.01	-.06	.00	-.05
21	-.04	.01	.06	-.06	-.02	.01
22	.08	.18	.11	.17	.09	.20
23	.05	.21	.10	.12	.08	.20
24	-.14	-.04	.12	.05	.05	.04
25	-.13	-.16	-.10	-.15	-.10	-.17
26	.11	-.18	-.03	-.22	.05	-.22
27	-.01	-.20	-.08	-.17	-.06	-.18
28	.00	.21	.07	.17	.06	.18
29	-.02	-.08	.06	-.01	.04	-.09
30	-.05	.12	-.01	.15	-.06	.13
31	.21	.13	.30	.17	.27	.19
32	.00	.09	.00	.02	-.03	.03
33	-.28	-.43	-.32	-.34	-.30	-.45
34	.04	.15	.12	.07	.08	.12
35	-.16	.14	.04	.09	-.04	.13

Variable	Absolute Gain Score		Relative Gain Score		Residual Gain Score	
	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
36	.12	.06	.11	.16	.10	.11
37	-.05	.19	.07	.25	.01	.23

* Correlations .21 or greater are significant at the .05 level.

Correlations .27 or greater are significant at the .01 level.

APPENDIX H

Rotated Factor Loadings

	Factors								
Variable	1	2	3	4	5	6	7	8	9
1	-.87	-.02	.16	-.41	-.05	-.02	-.08	-.12	.03
2	-.96	-.08	.03	.17	-.05	-.04	.13	-.01	-.07
3	-.41	.05	.21	-.79	-.02	.01	-.25	-.16	.10
4	-.47	.04	.16	-.12	-.08	-.06	.09	-.02	.30
5	-.02	.04	.08	-.06	.00	-.02	-.30	-.02	-.05
6	-.71	-.04	.01	-.17	-.02	-.12	-.02	-.11	.08
7	-.43	.00	.10	-.56	-.03	-.03	-.15	-.16	-.10
8	-.16	-.02	.04	.00	-.02	-.87	-.12	.03	.01
9	-.12	-.03	.02	.04	.00	-.88	-.14	.02	.02
10	-.53	-.03	.05	-.31	.01	-.18	-.20	-.08	-.01
11	-.84	.01	.06	-.21	-.03	.03	-.05	-.14	-.08
12	-.34	-.01	.10	-.01	-.19	.00	.10	-.78	.02
13	-.39	.02	-.23	.12	-.04	-.10	-.19	-.63	.09
14	.00	.01	-.41	.21	.21	-.10	-.55	.11	-.12
15	.04	-.02	.02	.04	-.05	-.20	-.06	-.14	-.11
16	.13	-.02	.28	.04	.21	-.06	.03	-.05	-.49
17	-.05	-.31	.48	-.01	.02	.02	.03	-.08	-.11
18	-.16	-.04	.07	-.14	-.81	-.01	.15	-.13	.07
19	-.47	-.23	.18	-.27	-.58	-.01	-.04	-.12	.09
20	.35	.21	-.16	.15	-.56	-.08	.28	-.03	-.01
21	.06	.12	-.20	.05	.26	.08	.10	.06	.25
22	-.26	-.03	.04	-.68	-.10	-.07	-.15	-.07	.02
23	-.01	-.05	-.05	-.71	-.02	.02	.23	.16	-.10
24	.03	-.06	-.01	.18	.00	.23	-.13	.05	.07
25	.06	.01	-.55	.05	.13	-.10	-.49	.25	.02
26	.05	-.71	.04	.09	.01	-.10	-.01	-.36	-.21
27	-.14	-.03	.88	.00	-.05	-.08	-.04	.04	-.09
28	.16	-.02	-.89	-.03	.04	.08	.04	-.15	.09
29	.12	.14	.16	-.02	.10	-.15	.16	-.03	.19
30	.00	.72	-.20	.11	.12	.02	-.05	.06	.27
31	.10	.03	.10	-.09	-.19	.15	.36	.07	.04
32	.06	.44	-.11	-.07	-.11	-.08	-.27	-.05	-.04
33	.05	.03	.12	.43	.13	-.09	-.15	-.17	-.10
34	-.09	-.09	-.02	-.49	-.07	.00	.44	.18	-.12
35	-.01	.40	-.12	.17	.05	.10	.17	-.15	.65
36	.05	.74	.15	-.02	.01	-.06	.00	-.12	-.34
37	.06	.86	-.01	.16	.08	.04	.11	-.11	.21