THE POTENTIAL OF THE FREE RECALL TASK AS A MEANS OF ASSESSING STUDENT ACQUISITION OF COURSE RELATED ORGANIZATION

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
WALTER SHAWVER BROWN
1972

LIBRARY
Michigan State
University

This is to certify that the

thesis entitled

THE POTENTIAL OF THE FREE RECALL TASK AS A MEANS OF ASSESSING STUDENT ACQUISITION OF COURSE RELATED ORGANIZATION

presented by

WALTER S. BROWN

has been accepted towards fulfillment
of the requirements for

PH.D. degree in MEASUREMENT, EVALUATION, AND RESEARCH DESIGN

Joe E Byers
Major professor

Date 8-25-72

O-7639

ABSTRACT

THE POTENTIAL OF THE FREE RECALL TASK AS A MEANS OF ASSESSING STUDENT ACQUISITION OF COURSE RELATED ORGANIZATION

By

Walter Shawver Brown

Rationale of the Inquiry

This field study builds on Johnson's research (1967), in which a traditional verbal learning technique is used to examine the organization personalized by students, and this organization's relationship to formal curricular structure. Rather than using free association as Johnson has done, however, this study uses free recall (FR).

Primary Objectives

1. To examine the generalizability of laboratory findings vis a vis the free recall paradigm to the less constrained environment of an ongoing instructional program.

2. To extend the conceptualization of the associative cluster to include not only segmental clustering (taxonomic), but also suprasegmental clustering (subject-imposed coalition of within-unit segmental clusters, as opposed to between-unit clustering).

3. To develop the basic data to support the use of free recall as a mode of formative evaluation, both for individual students and for instructional material refinement.

Materials

A free recall task was constructed using 49 key terms drawn from four units of material presented in individualized carrel programs in basic teacher education course. All subjects were tested in groups of approximately fifteen subjects each. The free recall task was presented in booklet form, with four study trials and four recall trials. The students received 50 seconds for each study trial and 150 seconds for each recall trial. All study lists were composed of the 49 key terms in random order.

Design and Analysis

Different groups of subjects were randomly assigned to one of five testing times. The first testing time occurred before any students had studied any of the four units in the carrels. The second testing time occurred at the close of instruction for unit I. Testing time 3 occurred at the close of instruction for unit II, etc. A total of 350 subjects with 70 per testing time comprised the final samples.

Four dependent measures were available for each student. From the free recall task, a recall measure, a segmental clustering measure and a suprasegmental clustering measure were extracted for each subject for each of the four trials and four units nested within each of the trials. The fourth measure for each subject was the mastery test score achieved for each of the four units studied during the quarter.

Three major analyses were undertaken. The first or theoretical analysis examined the learning curves over trials in dependent measure performance. This analysis consisted of a series of multivariate analyses of variance using orthonormalized transformations and orthogonal polynomials.

The second analysis was an instructional analysis examining predicted changes over testing times in within unit dependent measure performance. This analysis consisted of a series of multivariate analyses of variance using simple contrasts.

The third analysis was also instructional and examined the nature of the relationship between unit dependent measure performance and unit mastery test performance. This analysis used Pearson product moment correlation coefficients.

Hypotheses

Theoretical: The free recall dependent measures will increase over trials.

Instructional-MANOVA: Within-unit dependent measures will peak at that testing time in which the unit in question was studied.

Instructional-Correlations: Before unit instruction there will be no significant correlations and after unit instruction there will be significant positive correlations between unit dependent measure performance and unit mastery test performance.

Results

Theoretical

The dependent measures do increase over trials, though the suprasegmental clustering measure is redundant to the segmental clustering measure. The nature of the recall and segmental clustering curves is complex with significant constant, linear, quadratic and cubic trends. The recall curves decelerate (negative quadratic components) and the segmental clustering curves accelerate (positive quadratic components) over time. The major difference between the curves at different testing times was the constant term or y-intercept, indicating an increase in trial 1 performance. The curves were, otherwise, highly similar within measures.

Instructional-MANOVA

The hypothesis was only partially supported for two out of the four units. In general the measures did not peak as predicted.

Instructional-Correlations

The correlational data evidenced primarily insignificant correlations. There was some evidence that at preinstruction, student differences in entry behavior did correlate with later mastery test performances. The post-instructional correlations were generally not significant and positive as predicted.

Conclusions

On the basis of this field study, it could be concluded that the practical application of the free recall task to the classroom setting is not possible. The loss of control that is necessitated by "in vivo" research destroys the viability of the free recall task. It is encouraging, however, from a theoretical point of view, that the learning curves observed were consistent within-measures and were not unlike those previously encountered in laboratory research. Finally, it would appear that the suprasegmental clustering measure, while redundant to the segmental measure in this study, might be of value in research with materials or designs that make segmental clustering difficult to use.

---

[1]P. E. Johnson, Some psychological aspects of subject matter structure, J. Ed. Psychol., 1967, 58:2, 75-83.

THE POTENTIAL OF THE FREE RECALL TASK AS A MEANS OF ASSESSING STUDENT ACQUISITION OF COURSE RELATED ORGANIZATION

By

Walter Shawver Brown

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Educational Psychology

1972

DEDICATION

My thanks to those who have supported, guided and nurtured my growth during my doctoral studies. Uppermost among these are the following:

My wife Janet and my parents Jack and Wanda

Dr. Joe Byers

Dr. Judy Henderson

Dr. Maryellen McSweeney

Dr. Gordon Wood

Perhaps the best way to repay all of them is to try to teach as I have been taught--with patience, understanding and caring.

TABLE OF CONTENTS

		Page
DEDICATION		ii
LIST OF TABLES		v
LIST OF FIGURES		vii
Chapter		
I.	THE PROBLEM	1
	Introduction	1
	Commentary	4
	Purpose of the Study	5
	Theory	6
	Commentary	10
II.	REVIEW OF THE LITERATURE	11
	Overview	11
	Historical Antecedents	11
	Commentary	15
	The Multi-Trial Free Recall Task	15
	Commentary	20
	Verbal Organization and Academic Performance	20
	Commentary	24
III.	DESIGN AND METHODOLOGY	26
	Overview	26
	Sample	26
	Commentary	27
	Materials	27
	Word List Characteristics	27
	Instrument	33
	Design	34
	Introduction	34
	Testing Times	35
	Units	38
	Trials	38
	Unit of Analysis	38
	Completion Index	39
	Analyses	40
	Research Hypotheses	41
	Theoretical Hypothesis	41
	Instructional Hypothesis	41

Chapter Page

IV. PRESENTATION OF THE DATA 43

Introduction. 43
The Completion Index 43
Theoretical Analysis 45
Instructional Analysis 57
Correlations Between Dependent Measures and Mastery Test Performance 70
Overview and Hypotheses 70
Process of Analysis 70
Correlations Between Dependent Variables Performance and Mastery Test Performance 72

V. DISCUSSION AND IMPLICATIONS 76

Overview 76
The Theoretical Analysis: Change Over Trials. 76
The Instructional Analyses: Change Within Units Over Time 81
The Instructional Hypotheses: Correlations Between Unit Dependent Measure Performance and Unit Mastery Test Scores. . . 84
General Implications and Recommendations for Future Research 86

APPENDICES 89

BIBLIOGRAPHY 104

LIST OF TABLES

Table		Page
1.	Unit I segmental clusters.	19
2.	Maximum possible mastery test score by unit. .	24
3.	Report of segmental and suprasegmental cluster scores obtained for each of the units by four different raters	29
4.	Word list for each unit broken into segmental clusters	31
5.	Maximum possible values for the dependent measures (recall, segmental clustering and suprasegmental clustering) for individual units and the total across units	32
6.	Summary of independent and dependent variables.	34
7.	A time-line illustration and explanation of the independent variable of testing times (fall quarter)	37
8.	"Ideal" and "obtained" completion indices by testing time	44
9.	Mean proportions of possible recall achieved over four units and standard deviations by trials and testing time	45
10.	Mean proportion of possible segmental clustering achieved over four units and standard deivations by trials and testing times . .	46
11.	Mean proportion of possible suprasegmental clustering achieved over four units and standard deviations by trials and testing times	46
12.	Step-down F tests and P values for three dependent measures and orthogonal polynomials for trials over each testing time.	49
13.	Independent tests for significance of linear, quadratic and cubic trends within the three dependent measures	51

Table Page

14. Estimated coefficients for the learning curves over trials, based on the recall data. 53

15. Estimated coefficients for the learning curves over trials, based on the segmental clustering data 55

16. The mean proportion of possible recall achieved over four trials by testing times and units. 58

17. The mean porportion of possible segmental clustering achieved over four trials by testing times and units 59

18. The mean proportion of possible suprasegmental clustering achieved over four trials by testing time and units. 59

19. The step down "F" ratios and "p" values for the pair-wise contrasts within the three dependent variables resulting from the interaction test of the overall MANOVA 62

20. The step down "F" ratios and "p" values for the dependent measures in units 3 and 4 respectively, resulting from two independent significant multivariate tests of equivalence of mean vectors in a one factor MANOVA. 64

21. Summary of the multivariate tests of equality of mean vectors for the unit 3 ordered contrasts for the dependent measure of recall 67

22. Summary of the multivariate tests of equality of mean vectors for the unit 4 ordered contrasts for the dependent measure of recall 67

23. The completion statuses, completion status codes, sample size and critical values for the correlation coefficients at $p \leq .10$, to be used in the insturctional correlational analysis 72

24. Matrices of correlations between unit dependent variable performance and unit mastery test performance by completion status 73

LIST OF FIGURES

Figure		Page
1.	Data matrix--overall	36
2.	Proportion of possible recall achieved over four units by trials and testing times . .	54
3.	Proportion of possible segmental clustering achieved over 4 units by testing times and trials.	56
4.	Proportion of possible unit recall achieved over four trials by testing times	68
5.	Proportion of possible suprasegmental clustering achiever over four units by trials and testing times	100
6.	Proportion of possible unit segmental clustering achiever over four trials by testing times.	102
7.	Proportion of possible suprasegmental clustering over four trials by testing times	103

CHAPTER I

THE PROBLEM

Introduction

As evidenced by the publication of books such as Neisser's, Cognitive Psychology (1967), Miller, Galanter and Pribram's, Plans and the Structure of Behavior (1960), and Saltz, The Cognitive Bases of Human Learning (1971), psychologists are becoming more interested in human cognitive functioning and processing. The modes of cognitive organization available to man are of particular interest. Bruner (1956, 1960), Gagne (1970), Bloom (1956) and others have described methods of elucidating the structure inherent to the subject matter areas. The findings and hypotheses being promulgated by such men as those mentioned are being translated by educators into numerous specific instructional programs designed to maximize student learning of verbal materials. Examples of such programs would be the SCIS and the AAAS science programs (Belanger, 1969; Smith, 1969), the SMSG and IPI mathematics programs (Romberg, 1969; Heimer, 1969), and the Bereiter, Engleman beginning reading program (Bereiter and Engleman, 1966). More generally, models for the nurturance of optimal learning, such as Bloom's mastery

learning conceptualization (Bloom, 1968) demand explicit analysis of the content and objectives, and flexible implementation of the curriculum, to the end that the majority of the students may be educationally successful. The success or failure of such attempts to optimize the students' personalization of structure can be assessed only by examining the final outcomes of instruction--student learning and cognitive organization.

The organization imposed on verbal input has become a central concern of yet another academic population within the past fifty years, specifically, the verbal learning psychologists. In ever increasing numbers, they have produced a voluminous literature base in a relatively short period of time. While this tremendous growth has produced a number of experimental paradigms and variables of import, the overwhelming majority of the work has been at a very basic level. Such research, while exciting, and most sensible perhaps, given the relative infancy of the field (Kuhn, 1970), has been challenged as irrelevant to the "real world." Indeed, in a recent review (Tulving and Madigan, 1970), of the state of the art, only ten percent (10%) of the research reviewed was seen as being truly relevant to the field itself, in terms of advancing the boundaries of understanding. It is apparent that if much of the inquiry seems irrelevant to those intimately involved in the

field, that the potential value of the research would be particularly difficult for those outside the microcosm of verbal learning to discern.

This irrelevance may, however, be more apparent than real or necessary. To clarify, let us take an example from both fields in relation to organization. The educator may directly attack the structure of knowledge and extract or make concrete that structure, through task analysis and behavioral objective development. All of the curricula mentioned earlier rely heavily on such direct procedures. The verbal learning psychologist, however, must examine indirectly the ways that individuals process the verbal information that they collect every day from peers, parents, teachers, tachistoscopes, etc. The necessity of indirectly studying complex behavior has created a number of ingenious tasks that allow one to gain insight into cognitive processing. Free association, paired-associate learning, serial recall and free recall each approach the problem of how man structures his verbal world and each does so in a slightly different way. The free recall task, for instance, presents subjects with a set of words or terms that are semantically related in some way, e.g., associatively or taxonomically (Cofer, 1965). Most often, several conceptual clusters or groups of words are identifiable within the total list. When the words

are presented in random order, people invariably begin to impose an organization on the list. They report words in chunks or clusters that give the researcher information about the associations and superordinate concepts that the individuals have internalized. This task has repeatedly shown itself valuable in exposing pieces and parts of the effective organization a person brings to the task and to the terms within it (Adams, 1967).

It is a basic assumption of this inquiry that the dilemma of the curriculum developer/teacher, that of finding new ways of examining student learning and organization, and that of the verbal learning psychologist, that of proving the relevance of his research, may be concomitantly served and at least partially solved by the extension of some of the traditional verbal learning techniques and understandings to the tasks set before the educator. Further, it seems that the free recall task would be an especially fruitful technique with which to begin this foray, since it does address itself uniquely to the problem of organization of verbal materials--a primary concern of the educator.

Commentary

Obviously, this movement from the laboratory to the classroom will not be done without some costs. A major trade-off is the manipulability of the research

environment. Traditional experimental designs often demand a degree of control which is impossible to obtain in contexts other than the laboratory. This does not make research in the field any less necessary or attractive, it simply requires developing and trying new research techniques and models. Such then is the context of this investigation. The next section will discuss in a more detailed manner the purpose of the study.

Purpose of the Study

The purpose of this field study is to determine the efficacy of the free recall task as a means of analyzing the structural outcomes of a course of study. The structure of particular interest is that imposed before and after instruction, on a list of key terms drawn from the materials studied in a basic course in teacher education. The words have been chosen because they form clusters based on the course objectives and because they occur contiguously within the textual materials (Henderson et al., 1971, and unpublished carrel materials). An instructor would hope that the organization that he "teaches" would display a high match with that organization that the student would learn, i.e., impose on key terminology upon exiting the course.

Of further interest is the free recall tasks's relationship to the more traditional testing modes used

in student evaluation. The four unit exams consist of multiple choice questions and short answer essay questions based on the unit objectives. Since both the test items and the free recall task involve the same key terms, concepts and principles, it is expected that a positive relationship will obtain between the level of performance achieved on the two tasks.

Should this study indicate a positive relationship between the dependent measures and more traditional testing modes, it is possible that the free recall task could provide valuable evaluative input to a developing program. Its utility would be in giving feedback to material producers about the success of their materials in teaching desired organizational structure.

Theory

Time has traditionally been treated as an independent variable in our educational system. Also, the normal curve has been used as a model for the expected outcomes of student learning. Bloom, in developing his model for mastery learning (1968), has taken both of these desiderata to task. First, he insists that time should be seen as a dependent variable, the length of time allowed on learning a task being decided not by administrative exigencies, but rather by the amount of time needed by the student to reach

mastery. Second, in regard to the "arbitrary" imposition of the normal curve on student outcomes, he opines that 95% of the student population can and should achieve mastery of the material under consideration if given the appropriate modes of instruction and sufficient time. Indeed, it is suggested that brandishing records of student achievement which approximate the normal curve is more proof of poor teaching than of the existence of poor learners (Bloom, 1968).

Of particular interest in this investigation is the concern voiced for new tools of formative evaluation. Formative evaluation has two purposes (Bloom, 1968, 1971): first it is designed to check the student's progress through a unit of material. If mastery is evident, then another unit may be entered. If, on the other hand, the unit objectives are not mastered, the evaluation must serve as a diagnostic tool, aiding the instructor in prescribing appropriate additional experiences. The second purpose of formative evaluation is only secondarily concerned with an individual student's progress. It, rather, emphasizes the further refinement of the program of instruction. If, for example, an individual or a group of students does not reach mastery, the evaluation tool should help identify those concepts and principles that were inadequately taught. Assuming that the objectives are specified, the

problem is then how best to change instruction to inculcate the desired cognitive organization. From a practical point of view, then, this thesis is concerned with testing the efficacy of the free recall task as an additional formative evaluation tool for the latter purpose--that of identifying weak areas in the instructional materials.

From a theoretical point of view, this represents an extension of free recall, a traditional verbal learning paradigm, from the laboratory to the world of the classroom. While research emphasis has been placed for some time on the effects of the relative meaningfulness of verbal inputs, meaningfulness generally remains restricted to a word's socio-cultural linguistic frequency or familiarity and its associative relationships to other words, i.e., the "Minnesota Norms" (Russell and Jenkins, 1954), and Deese's indices of "Interitem Associative Strength" and "Associative Meaning" (Deese, 1959, 1962). Further, while the psychological mechanisms involved in prose learning have been examined by many from varying points of view and with varying techniques (Rothkopf, 1968; Rothkopf and Coke, 1968, 1970; Frase, 1969, 1971; Ausubel, Stager and Gaite, 1968, 1969), each has carefully manipulated the stimulus materials to maximize the likelihood of producing information gain in a laboratory setting.

In this study, while the materials have been carefully constructed to maximize the clarity and the learning of the materials, it has, nonetheless, been done from an instructional point of view, not for the purposes of isolating and manipulating learning processes for experimentation. Thus, if we find that the laboratory results are replicable in the instructional setting and applicable to instructional materials, the techniques and theory involved may be seen to have great potency.

In this investigation, ordinary English words are being used, making them relatively high on a traditional scale of meaningfulness. However, rather than being concerned solely with the examination of pre-existing language habits or verbal organization, we are concerned with the changes in language habits that obtain as a result of instruction. Many of the terms may be familiar to the students in contexts other than that imposed by the study of educational psychology. Thus in some cases we may be extinguishing pre-established associations that are extraneous to the conceptual matrix that the course itself develops for the words. Particularly germane to this thesis is the work of Paul Johnson (1964, 1965, 1967) in free association. He used free association (FA) tasks in the context of regular classes, i.e., physics, and found that the constructs of the

material were reflected in FA performance as the course progressed. Further, performance on the FA task also correlated with degree of course involvement and general course performance as measured by more traditional means. His work will be discussed in greater detail in the next chapter.

Commentary

It should now be clear that this thesis has hypotheses and subsequent analyses of two basic types: instructional and theoretical. The instructional concerns are based upon the ability of the free recall task to expose subject organization of verbal material. The theoretical questions will center on the replicability of laboratory findings in the instructional setting and in the extension and/or refinement of the present methodologies.

CHAPTER II

REVIEW OF THE LITERATURE

Overview

The intent of this chapter is to review the literature considered relevant to this thesis as briefly described in Chapter I. The review is broken into three subtopics, each followed by a commentary which will summarize the major contributions of the preceding section, and/or discuss its implications and introduce the next sub-section. The three subtopics are, in order of discussion: historical antecedents, the multi-trial free recall task, and verbal organization and academic performance.

Historical Antecedents

The free association and the free recall tasks are both primary tools used by verbal learning psychologists to study learning processes. Of these tasks, the first to attract research interest was free association. It relies totally on the previous verbal experience of the subject, in that he is given a stimulus word (i.e., white) and is asked to write or say the first word or words that come to mind (i.e., black). It has served

as the basis for an inquiry into the nature of the relationships that words have to each other and has spurred a debate still on-going as to whether the association can effectively operationally define the kernel upon which all higher processes rest. Norms have been established which identify the associates of common or high frequency words in the English language (Russell and Jenkins, 1954; Kent-Rosanoff, 1910; Thorndike and Lorge, 1944), which reflect the responses of native speakers. Other norms have been established for special populations such as schizophrenics (Bleuler, 1951). Beyond the common associations that individuals of a given identifiable population will produce, of course, the uniqueness of individual experiences are reflected by largely ideosyncratic associations that are common, possibly to many individuals, but hardly common in the sense of the Minnesota norms.

While simple pairwise associations are the most obvious relationships studied with the free association paradigm a great deal more has been done. For example, if one asks for more than the first word that comes to mind, any number of levels of associations are possible. Much theorizing about the nature of such structure of associative hierarchies (Deese, 1965; Cofer, 1965; Cofer and Foley, 1942) has been undertaken. Research in concept formation, for example, is basically a search

for the mechanisms that allow words or objects to become related and labeled categorically. Some psychologists have added intervening variables such as the mediating link or "verbal mediation" to explain more complex behavior (Osgood, 1956, etc.). Others have called upon imagery (Paivio, 1969), and still others prefer straight SR generalization or chaining (Skinner, 1957; Staats and Staats, 1963). Regardless of the descriptive preference of the psychologist involved, the search is the same--How does one discover or better understand the processes by which human beings form relationships, simple and complex; how do they achieve the ability to conceptualize, impose and use subordinate and superordinate structuring to communicate and solve problems?

Given the obvious presence of more complex behavior patterns such as associative hierarchies and general concept formation capabilities, the free recall task has become popular as an investigative tool in the past twenty years. The task is again relatively simple in nature, yet it taps more complex and sophisticated processing. A subject is given a list of words to study either visually or aurally and is then required to recall as many words as possible in whatever order they occur to him. Much work is done with lists which have experimenter-selected conceptual categories; i.e., categories that would be known to a native speaker (colors,

furniture, vehicles, etc.)--words and word classes that are high frequency in the English language. If the list contains, for example, twenty words; four concepts with five exemplars each, and is presented to the subject with the words in random order, the subjects reliably begin to impose previous language habit structure on the words. In fact, if given sufficient trials most individuals when asked to recall the randomly ordered words will be able to report the words in clusters which reflect the conceptual constraints that the experimenter has imposed--i.e., all words from concept A would be recalled together, all words from B together, etc.

Such grouping or clustering in accordance with experimenter defined constraints or concepts has been joined by the study of subjective organization wherein the experimenter need not be concerned with imposing organization with the list prior to presentation (Tulving, 1962). Rather, the subject is given several trials to learn the list and the experimenter then examines the resultant protocols to discover any consistent patterns of recall over trials. Invariably, the S's do establish groups or clusters of words that seem to facilitate the memory and recall of the items on the list. It can be seen that this task also elucidates or makes concrete verbal organization patterns than an individual has acquired through experience.

Commentary

It is hoped that through this rather cursory review, the point is established that the free association and free recall tasks are intimately related, historically and conceptually. Each can help expose verbal structure or organization and each is amenable to diverse variable manipulations; a necessity for a fruitful research paradigm. It is not surprising, therefore, that while some task variables and outcomes are unique to each, there are marked similarities between them. The next section of this review will concentrate on the particular task of primary importance to this research--the multi-trial free recall task.

The Multi-trial Free Recall Task

Bousfield and Cohen (1953) first reported examination of subject categorical organization or clustering in a multi-trial free recall paradigm. They presented their subjects with a list of 60 nouns which contained four different categories and 15 exemplars each. The categories and examples were equated in terms of the Lorge-Thorndike word count (Thorndike and Lorge, 1944), which as mentioned earlier, is an estimate of the frequency per million words in ordinary English text. The four categories which they chose were: animals, names, professions and vegetables. The 60 words were presented

at a three-second rate to five independent groups of subjects. Each group received a different number of trials or reinforcements on the random stimulus list, ranging from one to five presentations each followed by an immediate recall. It was noted that, in terms of recall only (strictly defined as the total number of words present in the stimulus material recalled after a trial), the mean total recall increased from 23.9 words for one reinforcement or trial, to 37.9 words after five. This learning over trials (learning here referring to the increase in mean total of words recalled) has been consistently replicated in the literature (Murdock, 1960; Tulving, 1962, 1964). Total recall, then, will be a significant response measure or dependent variable in this thesis. It is expected that recall will increase over trials.

As in his work with single trial free recall (1953), Bousfield defined clustering for this experiment in terms of repetitions, a repetition being a sequence in recall of two or more items from one of the four categories of the stimulus list. The number of repetitions was defined as the number of items from a category recalled together, minus one. Bousfield then calculated the number of repetitions that could be expected on the basis of chance alone, and found that his data suggested that a significant tendency exists to

reorganize the randomly presented words into their four conceptual categories. In the multi-trial experiment, then, he found that not only did mean recall increase over trials, but that clustering also increased steadily from one to five reinforcements, with five trials having about twice the level of repetitions as one.

Generally speaking, then, this measure can be obtained in conditions wherein the total set of "X" words or items is constructed by the experimenter such that it consists of two or more mutually exlcusive subsets of items. Each item within a given subset must be assumed on some logical or empirical grounds to be more "similar" to other items within the subset than it is to any other items in other subsets. Such subsets have been defined in terms of belongingness of the words or items to a conceptual category (Bousfield, 1953), associative relations among words (Jenkins and Russell, 1952), parts of speech (Cofer and Bruce, 1959), etc. The basic finding that the degree of clustering increases over trials has been consistently replicated by the latter researchers and numerous others over the past two decades.

One major drawback of this conceptualization of subject organization of verbal material is that it reflects and examines only that organization that the experimenter imposes (Tulving, 1962, 1964, 1968). In

other words, the subject may impose organization in terms of trial by trial contiguity of word or item output, yet if it is not one of the repetitions defined as "correct" by the experimenter, it is completely missed in the analysis of the subject output protocols. A second type of organization, therefore, has been labeled by Tulving (1962) as "subjective organization" (SO). Since it is defined in terms of the consistency of output order over trials for each subject, it does not necessarily require that the experimenter know in advance of the experiment which items are to be grouped together. It can, therefore, be used for any set of items.

Given the existence of two types of secondary organization (clustering and subjective organization), this thesis is designed to examine the clustering that occurs in response to a set of 49 words drawn from four units of material presented in a basic Teacher Education course. Two measures of organization are proposed, segmental clustering (C_s) and supra-segmental clustering (C_{ss}). The segmental clustering is traditional and directly related to research cited under clustering measures. Within each unit of material, a series of two to five word clusters (C_s) have been identified. Each word within a segmental cluster is deemed more similar to other words within the cluster than it is to words in other segmental clusters within the same or different units.

The supra-segmental cluster, however, is a departure from traditional measures and can be thought of as one piece of that organization heretofore generally referred to as subjective organization. More specifically, the existence of a supra-segmental repetition or cluster (C_{ss}) is defined by the contiguity of two or more words from within the same unit of material, but not from the same segmental clusters. Thus, while much subjective organization will still be missed, a more concrete delineation of at least one type of organization previously typed as subjective organization is now possible, trial by trial rather than only over trials.

To further clarify the distinction between C_s and C_{ss}, let us examine some of the content of the Unit I list (Table 1).

TABLE 1.--Unit I segmental clusters.

A	B	C	D
teaching learning	design instruction	means givens ends	variables human environmental curricular

Within the list of Unit I words, there are four segmental clusters--A, B, C, D. If a subject in a test trial or output phase lists the words, means, givens,

and ends consecutively, two repetitions are evident: means-givens and givens-ends. However, a further organizational mode is possible, that of grouping words together within the unit but across segmental clusters. For example, a subject might report teaching, instruction and curricular together. From this investigator's point of view, this grouping would be referred to as a supra-segmental cluster and would also contain two repetitions: teaching-instruction and instruction-curricular. Note, therefore, that we have at our disposal two measures of clustering, each describing very distinct types of organization. Theoretically, then one would expect that both clustering measures will increase over trials.

Commentary

At this point, three dependent measures have been identified: recall, segmental clustering and supra-segmental clustering. Each of these will be an integral part of the research to be done and each will serve as dependent measures in both the theoretical and instructional analyses.

The following section will discuss the work of Paul Johnson and its relationship to the thesis.

Verbal Organization and Academic Performance

As mentioned earlier, one of the prime goals of an instructional sequence is to change the behavior of

the students involved. It is clear that students cannot memorize verbatim the entire content of a course of study in any permanent fashion. What must occur, then, is storage of the substantive essence of what is being taught; associations, concepts, principles, etc. To this end, the majority of what is read or heard by the student must, of necessity, be forgotten, because so much of communicative verbiage, by the very nature of our language's syntactic system, is redundant--it does not distinguish one thought from another, one concept from another, etc. When Ausubel, et al. (1969) speak of meaningful verbal learning and subsumptive processing, they are in fact discussing the process of dropping the redundant features of instructional communication and saving and organizing those features which are distinctive into previously existing hierarchies or into newly acquired hierarchies and relationships. Using such stored distinctive features, one can retain in long-term memory the essence of what has been taught and learned for future reference--i.e., for test taking as well as for problem solving in life situations.

If one were to rationally define what kinds of words in instructional materials would become "distinctive features" in cognitive storage, the result would probably be the key terminology defining the concepts and principles presented in a course of study. Moving to physics as a

specific subject matter, Johnson (1964) presented four groups of high school students a one response, free association task consisting of a list of 18 concepts in physics (i.e., volume, density, mass). The stimulus words were presented aurally one at a time. The groups were (a) students presently taking physics; (b) students having taken physics; (c) students planning to take physics; and (d) students not planning to take physics. He hypothesized that degree of present involvement in physics (group "A" being most involved, group "D" being the least) would be positively related to performance on the dependent variables: the frequency with which the physics stimulus words were given as free associates of one another and Deese's (1959) inter-item associative strength measure. He found that the associative measures were, indeed, related to the degree of involvement in physics. The mean number of appropriate responses to the stimulus words ranged from 8.07 for those taking physics to 2.11 for those not planning to take physics. The inter-item associative strength measures ranged from 44.3 for group "A" to 11.3 for group "D." In both cases, then, the ranking in group performance was A>B>C>D.

Having found that performance on the free association task exposes differences in verbal organization which are related to degree of course involvement, Johnson went a step further in his research in 1965.

In this study (Johnson, 1965), he examined the possibility that success in taking an actual physics problem-solving test should correlate with performance on the free association task. In other words, if a student has a good grasp of the organization or associative relationships between the physics words, he should be able to solve physics problems better than the student who has not internalized the appropriate relationships. He found, upon analysis of the data, that the performance of subjects in the problem-solving test was related to their performance on the free association test; subjects who produced a relatively large number of relevant associations solved more problems than subjects who produced relatively few such associations.

While the relationship between performance on a verbal learning task and performance on a traditional testing mode has not been investigated much beyond Johnson's work, the ideas which he investigated are enticing. Given the relationships discussed earlier between free association and free recall, there is no reason to expect that the cognitive organization tapped by free association cannot also be profitably tapped by the free recall task. Further, since the free recall task allows more complex systems of relationships to be exposed than does the free association task, it seems important to assess the ability of the free recall task to

elucidate student-acquired organization and cognitive structure in an instructional setting.

Commentary

On the basis of the above, a set of instructional analyses will be done which will allow examination of the growth of cognitive organization over the span of an entire course of instruction, as well as examination of the relationship between performance on a free recall task embodying key terminology from the course and performance on traditional tests. To this end, each student's performance on the mastery test has been recorded. The tests contain both multiple choice and short answer questions. Four such scores have been recorded for each student with maximum values as follows:

TABLE 2.--Maximum possible mastery test score by unit.

Unit	I	II	III	IV
Maximum score possible	27	33	29	22

The basic research relevant to this thesis has been presented. Specific research hypotheses will not be presented until the end of Chapter III, when the design will be more clearly delineated. The following general hypotheses are apropos at this time, however.

1. The total (R, C_S, C_{SS}) will increase over trials and testing times.

2. Performance on the dependent measures will correlate positively with performance on traditional tests of knowledge of course materials.

CHAPTER III

DESIGN AND METHODOLOGY

Overview

The intent of this chapter is to discuss in greater detail the methodology and design of this thesis. The five sections are as follows: the sample, the materials, the design, the analyses, and the research hypotheses.

Before going further it should be stressed that this research is not experimental in nature. Rather it is a developmental field study designed to observe free recall performance "in vivo." The ultimate intent, of course, is to identify new arenas for later experimental investigation.

Sample

Subjects were drawn from the first required course in MSU's teacher education sequence. It is a sophomore course and thus contained a majority of sophomores. Both males and females were included, though females predominated. The size of the potential sample was approximately 700, however, since testing was done only once in each small group section over the quarter, anyone who was absent on the day of testing was excluded from the

study. Given the large number of students involved and the length of time (a full quarter) over which testing occurred, it is doubtful that there have been any systematic effects of absenteeism.

After all the data were collected and the quarter was completed, a total sample of 382 tested subjects remained who had completed all course requirements by the end of the quarter. From this number, a total of 32 subjects were randomly deleted from various testing times so as to provide 70 subjects per testing time. This imposition of equal cell sizes was necessitated by the mode of analysis which will be discussed later.

Commentary

It is apparent that because of the requirement imposed on subjects that they complete the course in order to be included in the study that the results can generalized most comfortably only to those students who do finish "on time."

Materials

Word List Characteristics

As has been briefly alluded to in previous chapters, the course in which this research is being undertaken is a basic course in teacher education. The instruction is broken into four major units: introductory, assessment, objectives and strategies. A word

list was constructed by drawing key terminology from each of the four units. A total of 49 such words were chosen in the following fashion. The course's behavioral objectives were perused and the materials were then studied to identify those words which could be seen as being of primary importance. One further restriction on the choices was that they had to be members of a cluster of at least two words and that the words chosen could not easily be placed in more than one segmental cluster. The words chosen, then, form clusters ranging from two to four items.

Admittedly, this process involved some subjective judgments by the author. For instance, some words may appear in more than one unit, with one unit giving greater stress and detail in regard to their meaning and relationships. In such cases, that unit providing the greatest degree of stress and information in regard to the words in question was considered the "home base" for the concepts. It is possible then that different people extracting lists of key terminology could come up with variations in the unit placement of clusters. To check on the degree to which this concern may be considered dangerous to the validity of the findings, three other people intimately involved in material and general course development were asked to break down the random list into unit and within unit cluster groupings. Each of the

three have been involved in the writing and teaching of the course materials and in the production of mastery test items.

The individuals were independently given the trial one randomized list from the subject testing materials (see Appendix A). They were asked first to break the words into four groups defining the four units and then to cluster the words within each of the units.

The output of these people is summarized in the table below (Table 3). Each letter stands for a separate rater, "A" being the author.

TABLE 3.--Report of segmental and suprasegmental cluster scores obtained for each of the units by four different raters.

Rater	Segmental Clustering				Suprasegmental Clustering			
	Unit				Unit			
	I	II	III	IV	I	II	III	IV
A	7	5	10	9	10	7	16	12
B	7	4	10	8	10	6	12	12
C	7	5	10	9	10	7	12	12
D	7	5	9	9	10	7	12	12

The above data (Table 3) was then analyzed using Kendall's coefficient of concordance (Hays, 1963), or W. The resultant coefficients were W = .9562 for segmental

clustering and W = .9062 for suprasegmental clustering. It is clear that with an upper limit for W of 1.0, both of the measures fared relatively well.

Looking at the output summary table above (Table 3), the major area of disagreement is in suprasegmental clusters for Unit III. This, on reviewing the organization imposed by the other persons involved, is due to the placement of two two-word clusters in units different than that chosen by rater "A." In retrospect, their placement of these two clusters is perhaps the most defensible. (The two clusters in question are asterisked in the word list chart below.) The discrepancy will be discussed further in Chapter IV.

The following table (Table 4) will clarify which words were chosen from each unit, and what segmental clusters are identifiable within each unit.

The three dependent measures, recall (R), segmental clustering (C_s) and suprasegmental clustering (C_{ss}), were each defined on the basis of this list. For example, for Unit I there are 11 words that a subject could recall. The maximum recall score is therefore 11. For segmental clustering, the number of words listed consecutively by the subject in recall from any one experimenter-defined cluster is used as a base for computation of a total C_s score. From this total of "n" words, 1 is subtracted, producing the clustering score which summarizes the number of distinct pairs found within a

TABLE 4.--Word list for each unit broken into segmental clusters.

Unit I C_s	Unit II C_s	Unit III C_s	Unit IV C_s
teaching	I.Q.	behavioral*	strategies
learning	achievement	non-behavioral	respondent
design	pretest	complete	operant
instruction	Piaget	incomplete	modeling
means	preopera-	beyond-school	shaping
givens	tional	in-class	reinforcers
ends	formal	socio-emotional*	positive
variables	Kohlberg	intellectual	negative
human	moral	conditions	attitude
environ-		terminal	feelings
mental		criteria	classical
curricular		recognition	stimulus
		recall	response
		convergent	
		attending	
		valuing	
		committing	

given cluster. Looking once again at Unit I then, if a subject wrote learning and teaching consecutively in recall, he would receive a C_s score of 1 (2 words minus 1); if he wrote human, variables, environmental and curricular, he would receive a C_s score for that cluster of 3 (4-1). The total C_s score for Unit I would be 1 for the first cluster, plus 1 for the second cluster, plus 2 for the third cluster, plus 3 for the last cluster--a total possible segmental clustering score of 7 for Unit I.

Suprasegmental clustering assumes that words are recalled consecutively from within a unit but that they are

not recalled pairwise from within the segmental clusters identified within Table 4. For example, if a subject recalls teaching and learning together, that constitutes a segmental cluster, but if he recalls learning and design together, that forms suprasegmental cluster of two words. Again applying the formula (n-1) to this cluster, it has a score value of 1. The maximum C_{ss} score would be achieved if a subject recalled all 11 words from Unit I consecutively, and produced no segmental clusters. The score for this large chunk would then be 10 (11-1), which is the maximum possible value for C_{ss} in Unit I.

The table below summarizes the possible values, given the above word list, of recall (R), segmental clustering (C_s) and suprasegmental clustering (C_{ss}) for each of the four units and total.

TABLE 5.--Maximum possible values for the dependent measures (recall, segmental clustering and suprasegmental clustering) for individual units and the total across units.

	I	II	III	IV	Total
R	11	8	17	13	49
C_s	7	5	10	9	31
C_{ss}	10	7	16	12	45

For presentation to the subjects the word list was randomized once for each of the four study trials involved in the study, producing random lists A, B, C, and D. See Appendix A.

Instrument

The course has large numbers of students, so to facilitate data collection, test booklets were used and the students were tested in groups. Four free recall trials were deemed sufficient to guarantee exposure of the existence of course-related organization of the key terms and yet avoid problems of ceiling effects on any of the measures.

The booklet (see sample Appendix A) has a cover sheet requesting the student's name, small group section number, section leader, date, and information regarding which units had been finished or started at the time of testing. This sheet is followed by four randomized word lists for study trials, each followed by a blank sheet on which the students record the words they remember during the test trials.

The students are allowed one second per word for study trials (total of 49 seconds) and three seconds per word for recall or test trials (a total of 147 seconds).

All students receive the same booklet, regardless of testing times. As mentioned in the previous section,

each study trial is composed of a different randomly ordered word list. The following chart summarizes the student's study trial stimulus materials.

Trial	1	2	3	4
Random Order	A	B	C	D

Design

Introduction

This study includes the following independent and dependent variables:

TABLE 6.--Summary of independent and dependent variables.

Variables	Levels
Independent	
Testing times	5
Units	4
Trials	4
Subjects	350
Completion index	3
Dependent	
Recall (R)	
Segmental clustering (C_S)	
Suprasegmental clustering (C_{SS})	
Mastery test scores (MS)	

The dependent measures (R, C_S, and C_{SS}) were introduced in Chapter II and are described more fully on Page 27. The dependent variable of mastery test scores

was described on page 24. The independent or design variables will be discussed in this section.

The complete design is shown in Figure 1. Using this design as a base, two separate analyses will be undertaken. The first analysis is theoretical and will involve collapsing across units leaving only the dependent measures (R, C_s, and C_{ss}) nested within trials. The second analysis is instructional and will involve collapsing across the variable of trials leaving the dependent measures nested only within units. Both then will be 5 x 4 designs with three dependent variables. Trials and units respectively will be considered as repeated measures. In succeeding subsections, the independent variables of interest will be discussed in more detail.

Testing Times

In order to study the growth of organization over the entire course sequence, the testing instrument was given at five times during the quarter. The probes or testing times were spaced over the quarter such that as a unit reached completion in the carrels, a testing time was planned.

While officially the emphasis in the carrels may have shifted from Unit II to Unit III, some students may still be working on Unit I or Unit II. This possibility, while perhaps not ideal from a strict

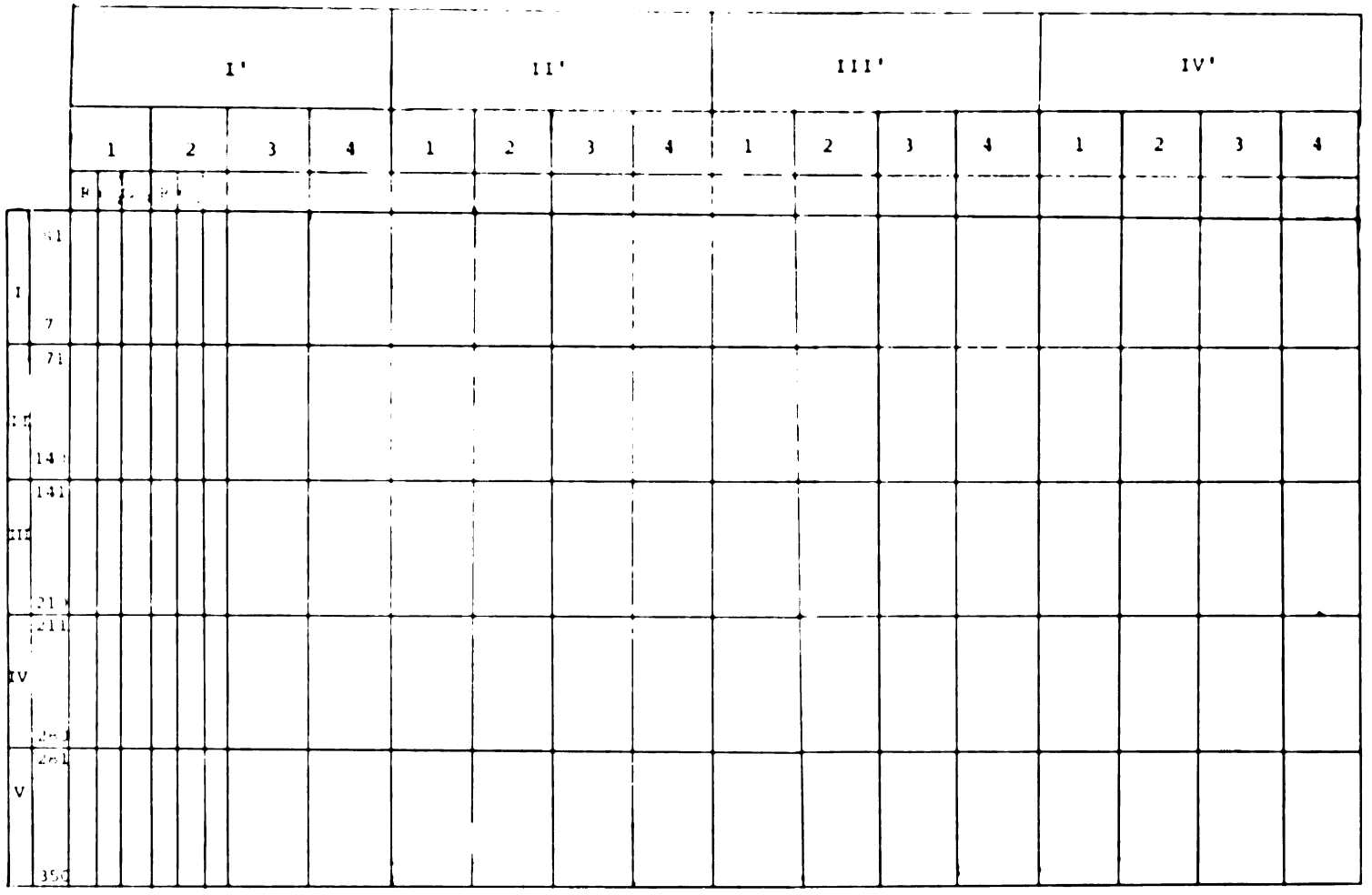

Figure 1.--Data matrix--overall.

experimental point of view, is necessitated because the course operates on a mastery model which allows students to move through the carrel programs at their own rate. Thus at any testing time, students will be spread in terms of which units are untouched, started or finished. The following table describes concretely the independent variables of testing times.

TABLE 7.--A time-line illustration and explanation of the independent variable of testing times (fall quarter).

Testing Times	Units
9/29- 9/30	I. Pre-instruction
10/11-10/14	II. Unit I nearing completion (Introductory)
10/18-10/21	III. Unit II nearing completion (Assessment)
10/25-10/27	IV. Unit III nearing completion (Objectives)
11/22-11/23	V. Unit IV nearing completion (Strategies)

At each testing time, exactly the same instrument was administered to 9 small group sections. A total of 45 small groups were randomly assigned to five testing times, thus allowing 9 independent groups per testing time. Each small group contained approximately 15 students, though absences on the testing day resulted in loss of subjects.

Units

As described earlier in Materials section (page 27), there are four units of subject matter. Performance data on the dependent measures of recall, segmental clustering and suprasegmental clustering for each unit will be gathered at each testing time, thus allowing examination of changes in within-unit performance over the quarter. As the research is to be done in an ongoing, instructional program, counterbalancing of the units is not possible and effects of order will not be testable.

Trials

As described earlier (page 33), four recall trials are given to each subject.

Unit of Analysis

Subjects are considered the unit of analysis in this thesis. While subjects are tested in preestablished small groups in Education 200, no threats to independence are in evidence. The small groups referred to are established for the purposes of personal growth and as such do not deal directly with the subject matter of intimate concern in this study. The substantive material on which this research relies is studied by all subjects in Education 200 in individualized study carrels via tapes, slides, film clips and programmed work books. There is no reason, therefore, to expect systematic

influences on carrel performance to accrue as a result of involvement in any particular small group. Further, the assignment of the 45 small groups to a particular testing time was done through the use of a table of random numbers.

Completion Index

Each subject at testing must indicate for each of the four units whether they are untouched (0), started (1), or completed (2). Thus, a sample student may have completed Unit I (2), started Unit II (1) and not yet begun Units III and IV (0,0). Since each of the four units contains a different number of words to be "learned," each unit contributes a different weight to the total task. Thus, to further increase our precision, each of the values just discussed will be multiplied by the constant weight on percentage of total words contained in the unit in question. The weights for the units are as follows: Unit I (11/49 = .2245), Unit II (8/49 = .1633), Unit IV (13/49 = .2653), Unit III (17/49 = .3469). To exemplify how these weights will be used, let us assume that a student reports his completion status as 2,1,0,0; he has completed Unit I, started Unit II and has not yet started Units III and IV. If we sum the weighted unit completion values, in this case, 2 (.2245) + 1 (.1633) + 0 (.2653) + 0 (.3469), we

produce .6123, an index of completion which describes in a relatively direct sense how far along in the substantive materials the subjects have progressed. These indices can range from 0 to 2. If this index is determined for all subjects within a testing time and the mean completion index for the group produced, we have an overall indication of how far the group, as a whole, has progressed toward course completion at each testing time. Though no attempt to align or transform the data on the basis of the completion index will be made, they will be used to provide a more realistic graphical portrayal of the variable of testing time. For example, if one were to graph the simple variable of testing times, the first reaction would be to space them equidistantly over the abscissa. It could be, however, that the rate of progress toward completion varies over the quarter such that the distance between testing time I and II should be much larger than the distance between time III, IV and V. If such absence of equidistance is the case, a more accurate graphical representation of the data will obtain through the use of the completion index.

Analyses

Since units have varying numbers of words and clusters, the data must be transformed such that the dependent measures (R, C_s, C_{ss}) are on an equivalent

scale--they have been transformed, therefore, into percentages by dividing the obtained score for the dependent measure by the maximum possible score for the unit and dependent measure in question.

The two basic analyses (instructional and theoretical) will be repeated measures, multivariate analyses of variance. The instructional analysis will be followed by post hoc trend analyses and by computations of correlation coefficients between performance on the dependent measures and mastery test scores.

Research Hypotheses

Theoretical Hypothesis

1. Recall, segmental clustering and suprasegmental clustering will increase over trials.

Instructional Hypotheses

2. a. Within unit, dependent measure totals will increase over testing times.

   b. (1) Unit 1 dependent measures will be significantly higher at testing time II than at testing times I, III, IV, and V.

   (2) Unit 2 dependent measures will be significantly higher at testing time III than at testing times I, II, IV and V.

   (3) Unit 3 dependent measures will be significantly higher at testing time IV than at testing times I, II, III and V.

   (4) Unit 4 dependent measures will be significantly higher at testing time V than at testing times I, II, III and IV.

3. a. At testing time I, or preinstruction, there will be no correlation between performance on the dependent measures and the mastery test scores.

   b. If a unit has been completed at testing times II, III, IV or V, there will be a positive correlation between performance on the dependent measures and the mastery test scores.

CHAPTER IV

PRESENTATION OF THE DATA

Introduction

This chapter will summarize the outcomes of the research outlined in the previous chapters. This presentation will attend closely to describing these outcomes, not discussing their implications, as that is reserved for Chapter V. Four major presentations will ensue: The completion index, the multivariate analyses of the theoretical hypothesis, the multivariate analyses of the instructional hypotheses and the instructional correlational analysis.

The Completion Index

As described on page 39, a weighted completion index was computed for each subject. The mean completion indices for each testing time have been computed and are reported below in Table 8. Note that the "ideal" values are those weights for each testing time that would have been obtained if students had moved in a totally rigid fashion through the course material: the "obtained" values reflect the population's freedom to move through the materials at their own pace.

TABLE 8.--"Ideal" and "obtained" completion indices by testing time.

	Testing Times				
	I	II	III	IV	V
Ideal Values	0	.4490	.7756	1.4694	2.0000
Obtained Values	.1032	.6198	.9239	1.1207	1.7630

Both ideal and obtained completion indices were plotted on the same graph. The curves were little different from each other. Further, the only reason for computing the completion indices was to more accurately describe the variable of testing times. It was felt that when the dependent measures were graphed as a function of time, the curves resulting from the use of testing times on the abscissa might look quite different than those obtained using the weighted completion indices. Such was not the case. Several sets of data were graphed using both time variables, and the appearance of the curves was virtually identical. Therefore, since no advantage accrued through the use of the more complex, weighted completion indices, the more parsimonious variable of equidistantly placed testing times will be used in graphing most of the data.

Theoretical Hypothesis

The hypothesis of interest in this analysis is: Recall, segmental clustering and suprasegmental clustering will increase over trials. Before detailing the series of analyses that were undertaken to test the above hypothesis, it would be useful to peruse the means and standard deviations that were obtained from the samples. The means and standard deviations are presented in Tables 9 (recall), 10 (segmental clustering) and 11 (suprasegmental clustering).

TABLE 9.--Mean proportions of possible recall achieved over four units and standard deviations by trials and testing time.

Trials	Testing Times					Overall
	I	II	III	IV	V	
1. Mean	.226	.246	.266	.269	.275	.256
SD*	.063	.081	.072	.077	.079	.075
2. Mean	.290	.307	.313	.331	.327	.314
SD	.080	.084	.082	.086	.088	.084
3. Mean	.345	.354	.365	.400	.397	.372
SD	.100	.094	.092	.106	.099	.098
4. Mean	.364	.390	.399	.401	.419	.395
SD	.107	.113	.096	.111	.117	.109

*SD represents standard deviation.

TABLE 10.--Mean proportion of possible segmental clustering achieved over four units and standard deviations by trials and testing times.

Trials	Testing Times					Overall
	I	II	III	IV	V	
1. Mean	.042	.050	.057	.059	.051	.052
SD*	.038	.045	.044	.050	.043	.044
2. Mean	.050	.055	.051	.068	.060	.057
SD	.044	.055	.046	.050	.045	.048
3. Mean	.063	.074	.071	.096	.086	.080
SD	.057	.055	.053	.076	.066	.062
4. Mean	.068	.096	.093	.108	.116	.100
SD	.054	.060	.058	.070	.069	.063

*SD represents standard deviation.

TABLE 11.--Mean proportion of possible suprasegmental clustering achieved over four units and standard deviations by trials and testing times.

Trials	Testing Times					Overall
	I	II	III	IV	V	
1. Mean	.040	.044	.057	.056	.058	.051
SD*	.034	.039	.041	.042	.043	.040
2. Mean	.069	.064	.061	.076	.086	.071
SD	.045	.049	.038	.047	.056	.047
3. Mean	.091	.101	.088	.110	.116	.101
SD	.065	.061	.044	.055	.057	.057
4. Mean	.089	.102	.099	.116	.104	.102
	.066	.070	.064	.068	.067	.067

*SD represents standard deviation.

Without concern at this time for significance, it would appear that the means generally increase over trials both within and across testing times for all dependent measures. Also in most cases, the individual trial means tend to increase over testing times.

If the above observations were shown to be significant, they would suggest that not only did the dependent measures increase significantly over trials (the hypothesis of interest), but also that the dependent measures increased significantly within trials over testing times; i.e., that trial 1 dependent measure performance at testing time V was significantly greater than trial 1 performance at testing time I.

The analysis of these data consisted of a series of multivariate analyses of variance (MANOVAS), using orthonormalized transformations and orthogonal polynomial contrasts to examine the nature of the learning curves over the four trials. The polynomial contrasts provide tests for four components: the constant term (the "y" intercept), the linear term (the slope of each curve), the quadratic term (first degree curvature or "x^2") and the cubic term (second degree curvature or "x^3"). The first analysis of this series was an overall test of equality of mean vectors with 70 subjects per cell. It was a two factor MANOVA; factor 1, testing times, had five levels and factor 2, trials, had four levels treated

as repeated measures. The three dependent measures (recall, segmental clustering and suprasegmental clustering), were nested within trials.

The overall multivariate F for equality of mean vectors was 634.199 for 12 and 344 degrees of freedom with $p < .0000$. The interpretation of the main effects of trials was obscured, however, because of a significant groups (testing times) by repeated measures (trials) interaction. The F ratio for this multivariate test of equality of mean vectors was 1.803 for 48 and 1288.6 degrees of freedom with $p < .0008$. This interaction suggests that change over the trials is differentially dependent upon the testing time in questions.

To attempt further explication of the changes over trials, a series of one factor MANOVAS was undertaken. Each analysis examined the trends over the factor of trials (four levels) within a particular testing time. The data from these analyses are summarized in Tables 12 and 13.

Table 12 presents the first segment of the relevant results of these analyses. This table summarizes the overall tests within each testing time allowing examination of significance of the contributions of each of the dependent measures considered in the following order: recall, segmental clustering and suprasegmental clustering. Since there were 12 variables for each testing time, an

TABLE 12.--Step-down F tests and P values for three dependent measures and orthogonal polynomials for trials over each testing time.

Dependent Measure and Trend Component	Testing Time I		Testing Time II		Testing Time III		Testing Time IV		Testing Time V	
	Step down F	p less than	Step down F	p less than	Step down F	p less than	Step down F	p less than	Step down F	p less than
Recall										
Constant	4.23	0.0435	1.40	0.2402	10.89	0.0016	15.26	0.0003	13.36	0.0005
Linear	309.25	0.0000	277.40	0.0000	408.21	0.0000	335.94	0.0000	306.71	0.0000
Quadratic	44.61	0.0001	68.02	0.0001	52.65	0.0001	36.03	0.0001	50.62	0.0001
Cubic	58.70	0.0001	45.77	0.0001	32.83	0.0001	49.17	0.0001	54.08	0.0001
Segmental Clustering										
Constant	0.04	0.8304	0.44	0.5108	1.98	0.1638	0.88	0.3228	6.05	0.0166
Linear	0.34	0.5626	0.44	0.5099	1.19	0.2798	0.93	0.3387	0.38	0.5391
Quadratic	10.48	0.0020	5.56	0.0216	.74	0.3940	8.01	0.0063	5.16	0.0266
Cubic	0.65	0.4225	13.00	0.0007	11.56	0.0012	3.93	0.0520	4.67	0.0345
Supra-Segmental Clustering										
Constant	2.22	0.1418	0.27	0.6036	0.66	0.4201	0.05	0.8192	2.02	0.1598
Linear	0.80	0.3741	0.24	0.6267	1.29	0.2601	0.91	0.3426	0.01	0.9273
Quadratic	5.94	0.0179	5.09	0.0278	0.13	0.7232	0.21	0.6483	0.96	0.3310
Cubic	1.76	0.1900	1.24	0.2703	0.02	0.8903	1.41	0.2392	2.41	0.1263
Multivariate F-test for Equality of Mean Vectors, df = 12, 58	121.64	0.0000	122.81	0.0000	131.84	0.0000	116.23	0.0000	137.23	0.0000

alpha level of 0.005 was used to make decisions regarding significance of the variables' contributions. This would result in an "analysis-wise" error rate of approximately 0.06.

Using the established alpha (0.005) and scanning the p-values associated with each F test from the bottom of Table 12 upwards, the conclusion was reached that at none of the testing times was the suprasegmental clustering measure contributing any unique information to the analysis of the data. In other words, once the variance explained by the recall and segmental clustering data had been taken into account, suprasegmental clustering added no further explanation of variance. Therefore, the suprasegmental clustering measure was dropped in all subsequent steps in the theoretical analysis.

Table 13 presents the second relevant segment of data obtained through the one factor MANOVAS. This table summarizes the within-measure tests of significance for the linear, quadratic and cubic trend components for recall and segmental clustering. The constant term was eliminated in these analyses as we were concerned here only with the shape of the learning curves, not with the "y" intercept of the various curves.

Scanning Table 13 led to the conclusion that for both dependent measures there was a reliable cubic, quadratic and linear trend component at each testing time.

TABLE 13.--Independent tests for significance of linear, quadratic and cubic trends within the three dependent measures.

Dependent Measure and Trend Component	Testing Time I		Testing Time II		Testing Time III		Testing Time IV		Testing Time V	
	Step down F	p less than	Step down F	p less than	Step down F	p less than	Step down F	p less than	Step down F	p less than
Recall										
Linear	310.8808	.0000	286.8421	.0000	490.3838	.0000	423.9125	.0000	368.9790	.0000
Quadratic	52.5823	.0001	69.0324	.0001	53.3418	.0001	34.7000	.0001	48.1780	.0001
Cubic	58.6332	.0001	46.9990	.0001	33.3878	.0001	53.6421	.0001	60.1036	.0001
Segmental Clustering										
Linear	39.4256	.0001	86.1514	.0001	81.7600	.0001	68.9062	.0001	65.4191	.0001
Quadratic	25.6994	.0001	31.8723	.0001	30.8160	.0001	37.1353	.0001	43.4347	.0001
Cubic	35.2648	.0001	12.3007	.0009	36.4104	.0001	18.1060	.0001	38.8391	.0001

Degrees of Freedom for each within measure hypothesis = 1

Degrees of Freedom for error for each = 69

This outcome suggests that the learning curves over trials for all testing times have multiple components that each significantly increase the accuracy of the prediction of the line of best fit for the data. In other words, looking at the means over trials, within testing times, in Tables 9, 10, and 11 we can now safety state that the means do tend to increase significantly over trials as was originally hypothesized, but that this increase is not strictly linear since the significant quadratic and cubic components suggest a degree of curvature identifiable within each dependent measure over trials by testing times.

In order to further clarify the differences in the slopes of the learning curves within dependent measures and testing times, the estimated coefficients for each component were computed. Tables 14 and 15 present estimates of the constant, linear, quadratic and cubic components of the curves; for recall and C_s respectively. Each table of estimated coefficients is followed by the graph of the means of the data for that dependent measure; recall, Figure 2, segmental clustering, Figure 3. For C_{ss} table and Figure, see Appendix C.

The striking feature of the tables of estimated coefficients is the general similarity within each dependent measure. Although the various components are statistically significant, the differences are relatively minor. The major differences seem to occur in the estimated coefficients for the constant component, which represents

TABLE 14.--Estimated coefficients for the learning curves over trials, based on the recall data.*

Testing Times	Constant	Linear	Quadratic	Cubic
I	.306	.105	-.022	-.006
II	.324	.107	-.013	.001
III	.336	.101	-.006	-.005
IV	.350	.104	-.030	-.017
V	.353	.109	-.016	-.008

*Supraseamental coefficients and graph are in Appendix C, pages 99 and 100.

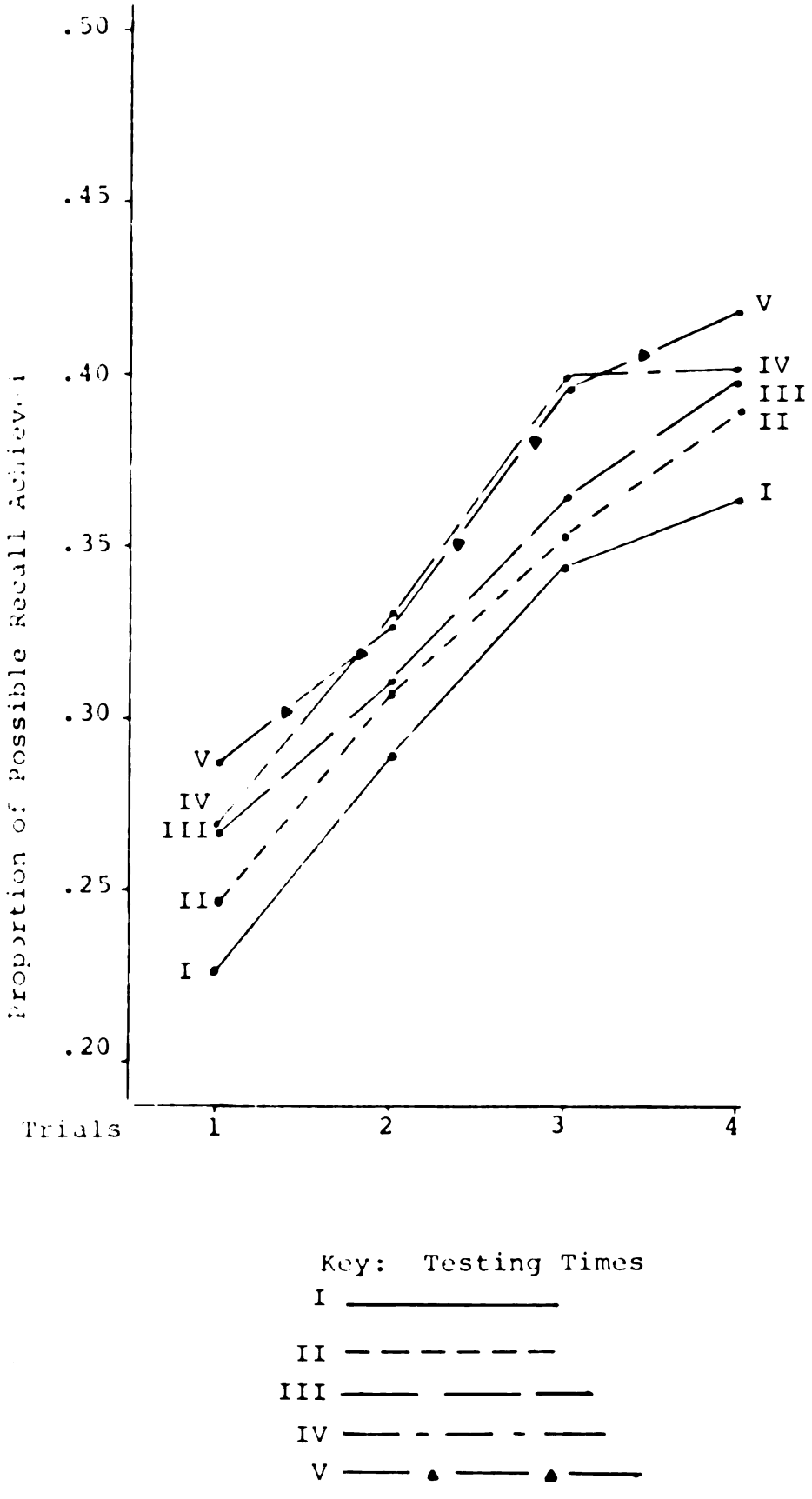

Figure 2.--Proportion of possible recall achieved over four units by trials and testing times.

TABLE 15.--Estimated coefficients for the learning curves over trials, based on the segmental clustering data.

Testing Times	Constant	Linear	Quadratic	Cubic
I	.055	.020	-.002	-.003
II	.069	.035	.008	-.002
III	.068	.029	.014	-.006
IV	.083	.039	.001	-.008
V	.078	.047	.011	-.001

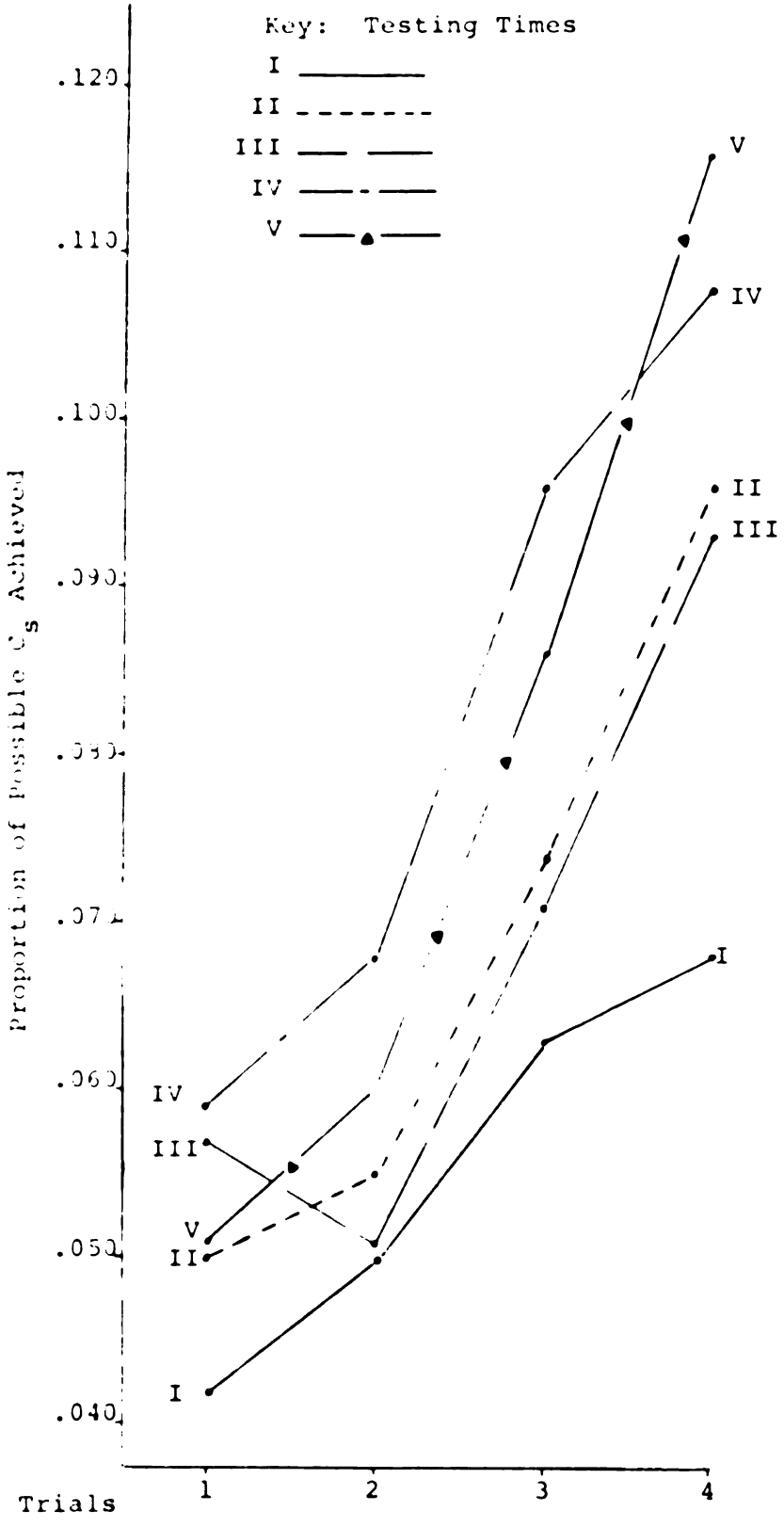

Figure 3.--Proportion of possible segmental clustering achieved over 4 units by testing times and trials.

the "y" axis intercept. In effect, it appears that the major difference between the curves obtained at the various testing times within measures is representative of a consistent increase in performance on the dependent measures at trial 1 over testing times. The minor differences in the linear, quadratic and cubic components within measures might suggest thta the curves themselves are highly similar regardless of testing time, once the absolute differences at trial 1 have been equalized. Also of note is the difference between the recall and segmental clustering measures in the nature of their non-linear components. All of the quadratic coefficients for the recall data are negative, suggesting decelerating curves or curves that flatten out over trials. All of the quadratic components for the segmental clustering data, except for testing time I, are positive, suggesting accelerating curves. Glancing at the graphs of the recall and segmental clustering data with the coefficients (Figures 2 and 3 respectively), the tendencies described are noticeable.

Instructional analysis

The hypotheses of interest in this analysis are:

a. Within unit, dependent measure totals will increase over testing times.

b. (1) Unit 1 dependent measures will be significantly higher at testing time II than at testing times I, III, IV, and V.

(2) Unit 2 dependent measures will be significantly higher at testing time III than at testing times I, II, IV, and V.

(3) Unit 3 dependent measures will be significantly higher at testing time IV than at testing times I, II, III, and V.

(4) Unit 4 dependent measures will be significantly higher at testing time V than at testing times I, II, III, and IV.

Before detailing the series of analyses that were undertaken to test the above hypotheses, it would be useful to peruse the within-unit and overall means that were obtained from the samples. The means are presented in Tables 16 (recall), 17 (segmental clustering), and 18 (suprasegmental clustering).

TABLE 16.--The mean proportion of possible recall achieved over four trials by testing times and units.

Testing Times	Units			
	1	2	3	4
I	.348	.403	.238	.300
II	.395*	.443	.250	.290
III	.375	.470*	.275	.295
IV	.388	.465	.303*	.310
V	.350	.445	.308	.363*

*The "b" hypotheses suggest that these means should be significantly higher than any other within-unit means.

TABLE 17.--The mean proportion of possible segmental clustering achieved over four trials by testing times and units.

Testing Times	Units			
	1	2	3	4
I	.090	.017	.048	.059
II	.126*	.028	.060	.058
III	.121	.029*	.062	.057
IV	.145	.029	.083*	.066
V	.094	.024	.083	.092*

*The "b" hypotheses suggest that these means should be significantly higher than any other within-unit means.

TABLE 18.--The mean proportion of possible suprasegmental clustering achieved over four trials by testing time and units.

Testing Times	Units			
	1	2	3	4
I	.075	.099	.057	.077
II	.101*	.115	.062	.064
III	.079	.132*	.071	.052
IV	.094	.137	.080*	.074
V	.075	.135	.085	.089*

*The "b" hypotheses suggest that these means should be significantly higher than any other within-unit means.

In regards to hypothesis "a" above, the means, if ranked by testing times should appear as follows: I<II<III<IV<V. The recall data (Table 16) might tend to support the hypothesis of interest as the rank order of overall means by testing times is I<II<III<IV=V. Testing times IV and V means are equivalent, however. The clustering measures both exhibit the same pattern (I<III<II<V<IV), which would tend not to support a simple increase over testing times.

Perusing the within-unit means with hypotheses "b," 1 through 4, in mind, some consistency is notable. First, note that on each table of means, the mean for each unit is asterisked that should be the maximum one according to the hypothesis being discussed. Recall for units 1, 2 and 4 do maximize as expected. Unit 3 recall maximizes at testing time V rather than at IV as hypothesized. For segmental clustering, units 2, 3 and 4 maximize as hypothesized, though unit 1 maximizes at testing time IV. Finally, for suprasegmental clustering, the mean for unit 4 does maximize as hypothesized but the other three units do not. In general, the raw means for recall and segmental clustering do tend to support the hypothesized relationships and suprasegmental clustering does not.

If all of the above discussions vìs a vís the hypothesized relationships among the raw means were found to be significant, the instructional hypotheses would

tend to be supported, though not strongly and not similarly across all dependent measures.

The actual analyses of these data consisted of a series of multivariate analyses of variance (MANOVAS) using simple contrasts to examine the hypothesized relationships. The first analyses of this series was an overall test of equality of mean vectors with 70 subjects per cell. It was a two factor MANOVA (5 x 4) with the five level factor being testing times and the four level factor being units, which were treated as repeated measures. The three dependent measures (recall, segmental clustering and suprasegmental clustering), were nested within trials. This overall test contrasted dependent variable means at testing times I, II, III and IV with the means at testing time V in a pair-wise fashion. The overall multivariate F for equality of mean vectors was 680.861 for 12 and 334 degrees of freedom with $p < .0000$. The interpretation of the main effects of units was obscured, however, because of a significant groups (testing times) by repeated measures (units) interaction. The F ratio for this multivariate test of equality of mean vectors was 2.1763 for 48 and 1288.6 degrees of freedom with $p < .0001$. Table 19 presents the step down F table for this overall test for the interaction. To ensure an "analysis wide" alpha level of approximately 0.06, an alpha level of 0.005 was used

to make decisions regarding significance of the "p" values for the pair-wise contrasts.

TABLE 19.--The step down "F" ratios and "p" values for the pair-wise contrasts within the three dependent variables resulting from the interaction test of the overall MANOVA ("p" values significant at 0.005 or less).

Variable	Contrast	Step down "F"	"p" less than
Recall	I versus V	3.7694	.0052
	II vs. V	5.5917	.0003
	III vs. V	5.2088	.0005
	IV vs. V	2.0211	.0911
Segmental Clustering	I vs. V	1.8700	.1153
	II vs. V	2.6777	.0318
	III vs. V	0.4282	.7883
	IV vs. V	0.6596	.6205
Suprasegmental Clustering	I vs. V	1.2858	.2753
	II vs. V	0.2190	.9278
	III vs. V	1.6506	.1612
	IV vs. V	0.7847	.5358

degrees of freedom for the hypothesis = 4
degrees of freedom for error = 345

Scanning the "p" values from the bottom of Table 19 upwards, it appears that neither clustering measure evidences any significant variance in the means. In essence then, the differences noted by the overall analysis fall within the recall measure and not within the clustering measures. Therefore, in subsequent analyses, only the recall data will be considered when discussing the results.

To attempt clarification of the significant, overall groups by repeated measures interaction, a series of four, one factor MANOVAS was undertaken. The only design factor in these MANOVAS was testing times, which had five levels. Again, the pairwise contrasts used pitted testing times I, II, III and IV against time V. Each of these analyses dealt with the performance on a single instructional unit and addressed itself to the question of whether within a given unit there are some significant differences among the means of the dependent measures over testing time.

The unit 1 analysis resulted in an F ratio for equality of the mean vectors of 1.7553 for 12 and 907.8 degrees of freedom with $p < .0514$. It was therefore not significant and suggested that the dependent measure means in unit 1 were equivalent over testing times. The unit 2 analysis resulted in an F ratio for equality of mean vectors of 1.6920 for 12 and 907.8 degrees of freedom with $p < .0636$, and also was not significant. In summary then, units 1 and 2 do not evidence significant differences over time and therefore were excluded from further analyses.

Units 3 and 4, however, did evidence significant differences over testing times. The unit 3 F ratio for equality of mean vectors was 2.4295 for 12 and 907.8 degrees of freedom with $p < .0042$. Analysis of unit 4 exposed an F ratio for equality of mean vectors of 3.1358 for 12 and 907.8 degrees of freedom with $p < .0003$.

Table 20 summarizes the step down F tables for the independent tests of units 3 and 4 respectively. To assure a low "analysis-wise" alpha level, an alpha level of .005 was used to make decisions regarding significance of the variables' descriptive contributions.

TABLE 20.--The step down "F" ratios and "p" values for the dependent measures in units 3 and 4 respectively, result-from two independent significant multivariate tests of equivalence of mean vectors in a one factor MANOVA (the factor was that of testing times, "p" values significant at 0.005 or less).

Unit	Dependent Variable	Step down "F"	"p" less than
3	Recall	6.7378	.0001
	Segmental Clustering	0.6292	.6420
	Suprasegmental Clustering	0.0966	.9836
4	Recall	5.1573	.0005
	Segmental Clustering	1.5350	.1916
	Suprasegmental Clustering	2.7613	.0277

Note that examination of the "p" values in Table 20 reaffirms the previous decision to exclude the clustering measures. The data presented to this point allow us to suggest that at least for units 3 and 4, the dependent measure of recall does tend to vary over testing times.

The final analysis of this series attends to the "b" hypotheses, or the relationship between the time of finishing the study of a unit and the dependent measure performance. Since there were no significant differences

found within units 1 and 2, these units will not be analyzed in terms of this set of hypotheses. This deletion leaves us with units 3 and 4, which did evidence significant differences. The next step then is to relate the outcomes of the two, one factor MANOVAS in which the pair-wise contrasts were differentially defined for each unit, according to the demands of the hypothesis in question. For unit 3, the fourth testing time is hypothesized as the maximum performance time. Therefore all other testing time dependent measures are contrasted with performance at testing time IV, i.e., I versus IV, II versus IV, etc. Furthermore, since a multivariate test of equivalence of mean vectors examines the contrasts non-independently, the contrasts are ordered such that the contrast between the means showing the smallest difference is tested first, the next largest difference next, etc. The order of the testing of the contrasts for unit 3, then, is IV versus V, IV versus III, IV versus II, and IV versus I. This ordering reflects the rank order of the recall means only, since it is the only one of the three dependent measures shown to be clearly significant. The ordering of the contrasts in the described manner, avoids the seemingly illogical possibility that because of the non-independence of contrasts that a difference tested at the beginning could appear significant, while a larger difference tested later could appear non-significant.

For unit 4 the fifth testing time was the location of the hypothesized maximum means, therefore the means were all contrasted with those obtained at testing time V. Further, the order of the contrasts was IV versus V, I versus V, III versus V, and II versus V. Again these contrasts were ordered on the basis of the decreasing recall means only, since recall alone has clearly been shown to be significant.

Tables 21 and 22 summarize the tests of significance for the separate contrasts within units 3 and 4 respectively. They are followed by the graphs of the recall data (Figure 4). For graphs of the C_s and C_{ss} data, see Appendix D. Using an alpha level of .05 and scanning the data in Table 21 from the bottom upwards the results suggest that for unit 3, the means at testing times I and II are significantly different from those at testing time IV. Also, the means for the recall data are equivalent at testing times III, IV and V. This result suggests that unit three performance did increase over the testing times, but that the data did not peak at testing time IV as hypothesized, since the means at testing times III, IV, and V are statistically not different.

The data in Table 22 for unit 4 suggests that all of four contrasts are significant. The "p" value for the II versus V contrast does not look significant, but it must be remembered that the non-independent natuee of multivariate testing allows room for such a paradox. It

TABLE 21.--Summary of the multivariate tests of equality of mean vectors for the unit 3 ordered contrasts for the dependent measure of recall ("p" values significant at "p" of .05 or less.

Contrast	"F" ratio	"p" less than
I versus IV	4.1681	.0065
II vs. IV	3.6242	.0134
III vs. IV	1.9749	.1175
V vs. IV	.1440	.9335

degrees of freedom 3, 343

TABLE 22.--Summary of the multivariate tests of equality of mean vectors for the unit 4 ordered contrasts for the dependent measure of recall ("p" values significant at "p" of .05 or less.

Contrast	"F" ratio	"p" less than
II versus V	1.3408	.2610
III vs. V	3.8732	.0096
I vs. V	3.6166	.0135
IV vs. V	3.8488	.0099

degrees of freedom 3, 343

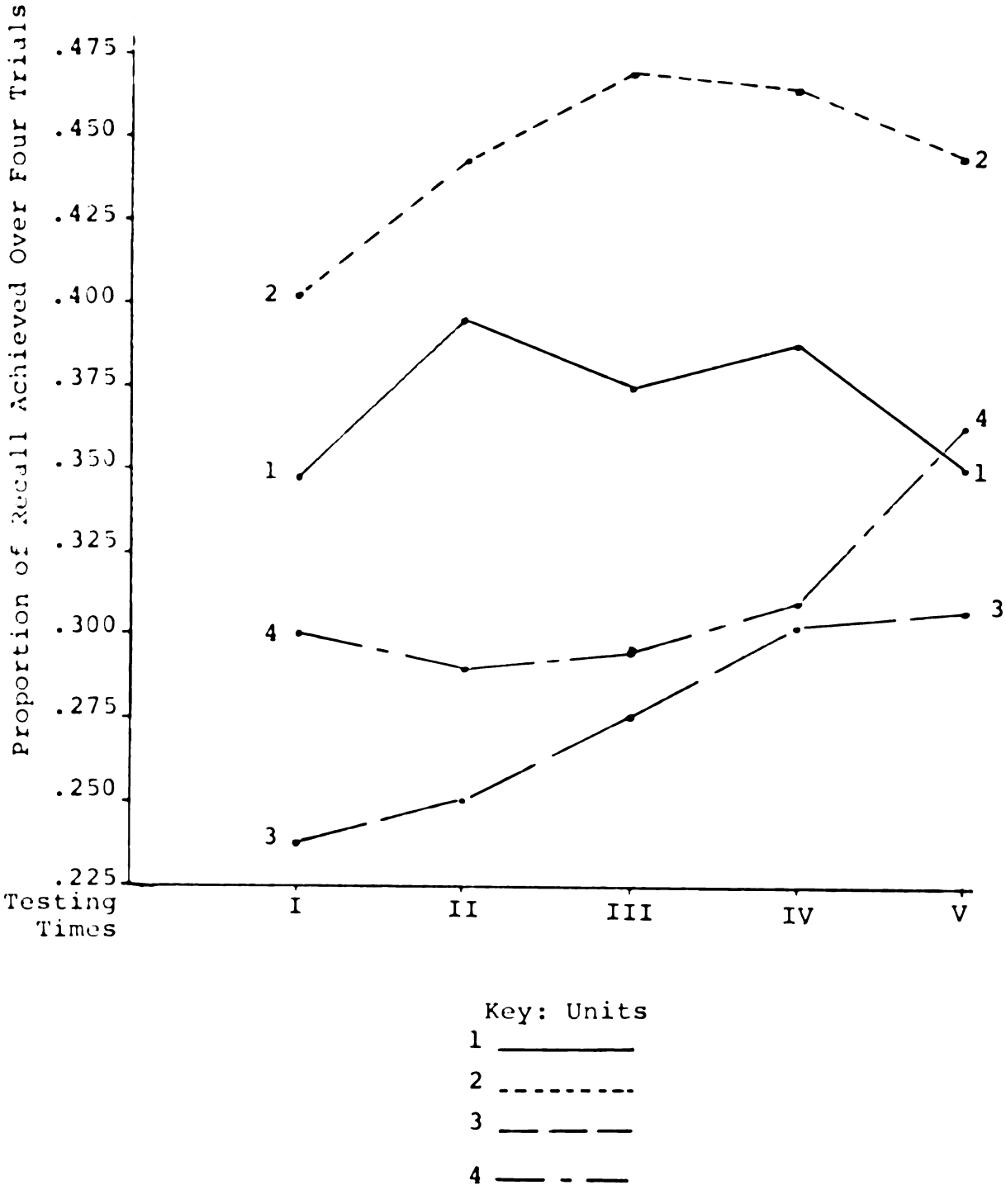

Figure 4.--Proportion of possible unit recall achieved over four trials by testing times.

*Graphs for segmental and suprasegmental clustering data are in appendix D.

is at this point that the ordering becomes useful, because we know that the absolute difference between the recall means at testing times II and V is greater than the absolute difference between the means at testing times IV and V. Therefore it may be concluded that for at least the recall data in unit 4, the performance peaked at testing time V and was significantly higher than that at any previous testing time, as was hypothesized.

The data and hypothesis tests relevant to the instructional multivariate hypotheses have been summarized. It can now be stated that the recall measure was the only one that included significant differences, and further, that only the recall data for units 3 and 4 evidence significant differences. For unit 3 and unit 4 recall data then, there is a significant increase over testing times, and the unit 4 recall data does peak as expected at testing time V when unit 4 has just been studied. From a general point of view, while some parts of the data did perform as expected, it could only be said that the multivariate instructional hypotheses were only weakly supported. Further discussion of this set of analyses will be presented in Chapter V.

Correlations Between Dependent Variables and Mastery Test Performance

Overview and Hypotheses

The following section first outlines the process of analysis used to examine the hypothesized relationships between unit dependent measure performance (Recall - R; Segmental clustering - C_s); Suprasegmental clustering - C_{ss}) and the unit mastery test performance. Once the process of analysis is clarified, the data itself will be presented and described.

3. a. At testing time I, or preinstruction, there will be no correlations between performance on the dependent measures and the mastery test scores.

   b. If a unit has been compelted at testing times II, III, IV or V, there will be a positive correlation between performance on the dependent measures and the mastery test scores.

Process of Analysis

A series of Pearson product moment correlation matrices were first computed using all 350 subjects divided by testing time. Therefore each matrix was based on a population of 70 subjects. The resultant correlations were low, with only a small percentage evidencing significance.

It was then decided that the above mentioned outcome might be an artifact of the wide range of completion statuses possible at any given testing time in a course wherein students may advance at their own rate (see page 35).

To reduce the possibility of this artifact's effect the five pure completion statuses (0000, 2000, 2200, 2220, 2222) were used to extract five new populations across testing times. Recall that 0 represents a unit not yet started and 2 signifies a unit that has been finished. In effect then, all people with one status code 2200 (units 1 and 2 finished, units 3 and 4 untouched) were grouped together, regardless of the testing time. These five new samples were then used as a basis for a second analysis of the relationship between dependent measures and the unit mastery test scores.

This second analysis produced more differentiated correlation coefficients. That is, a greater number of significant correlations were produced, and the range of the coefficients increased. The data to be presented and discussed then will be that data produced by the second analysis. The following table will summarize the five completion statuses used in the analysis, the completion status code that will be used for identification purposes in the matrices to be presented, the number of subjects (N) in each of the five samples and the critical values for the correlation coefficients at $p \leq .10$ for each of the five samples. Had reliability information been available, the coefficients reported would have been corrected for attenuation. Assuming a relatively low

TABLE 23.--The completion statuses, completion status codes, sample size and critical values for the correlation coefficients at $p \leq .10$, to be used in the instructional correlational analysis.

Completion Status	Completion Status Code	N	$p \leq .10$
0000	0	57	.21
2000	2	20	.36
2200	4	20	.36
2220	6	19	.37
2222	8	22	.36

reliability for the instruments, such correction would have increased the size of the correlations reported. Since reliability information was not available, the less conservative alpha level of $p \leq .10$ has been used. This less conservative alpha is used with the awareness that the usual interpretation of "significance" of the correlations may be spurious. Therefore, the interpretation of the significance of the relationships will be purely descriptive in nature.

Correlations Between Dependent Variable Performance and Mastery Test Performance

Two general comments are applicable to the data in Table 24. First note that out of 60 correlations, only 15 are significant at $p \leq .10$. This would suggest

TABLE 24.--Matrices of correlations between unit dependent variable performance and unit mastery test performance by completion status.

	Status Completion Code				
	0	2	4	6	8
			Recall (R)		
R_1 X MTS_1	.20	.11	.19	-.16	.14
R_2 X MTS_2	.35*	.22	.16	-.28	.56*
R_3 X MTS_3	.27*	.34	.32	-.23	.49*
R_4 X MTS_4	-.01	.36*	.22	.26	.14
			Segmental Clustering (C_s)		
C_{s1} X MTS_1	.22*	.05	.35	-.07	.06
C_{s2} X MTS_2	.03	.21	-.42*	-.09	.00
C_{s3} X MTS_3	.04	.12	.12	-.04	.17
C_{s4} X MTS_4	-.01	.39*	.16	.15	.12
			Suprasegmental Clustering (C_{ss})		
C_{ss1} X MTS_1	.19	.35	-.02	-.10	-.10
C_{ss2} X MTS_2	.22*	.48*	-.05	-.23	.37*
C_{ss3} X MTS_3	.36*	.40*	.32	-.38*	.48*
C_{ss4} X MTS_4	.05	.33	.30	.04	.18

*Indicates significance at $p \leq .10$

Key: MTS - Mastery Test Score
Subscripts represent the unit in question

that the relationship between dependent variable performance and mastery test performance, in an overall sense, is a tenuous one. Second, units 2 and 3 in the recall and suprasegmental clustering matrices contain eleven out of the fifteen significant correlations. This suggests that unit 2 and 3 recall and suprasegmental clustering performances do, from among all units and measures, have the greatest tendency to be related to the appropriate mastery test performances. For this reason, only the unit 2 and unit 3 correlations for recall and suprasegmental clustering will be discussed further.

In regard to the first hypotheses (no correlation at status 0), it is apparent for both units (2 and 3) that preinstructional recall and suprasegmental clustering correlate significantly and positively with later mastery test performance. In effect, there are individual differences among the students before instruction in units 2 and 3 that are not entirely destroyed by instruction, i.e., those students who perform well on the unit 2 and 3 dependent variables of interest tend to do well on later master tests on those units, and vice versa.

In regard to the second hypothesis, after instruction, again for units 2 and 3 in the recall and suprasegmental clustering data, the only consistent relationship is that at status code 8. In these cases there is a positive and significant correlation. In general, however,

the hypothesized post-instructional correlations are not in evidence. It would appear then that those individual differences which are exposed by free recall performance are not in any strong, systematic way related to individual differences in mastery test performance as was hypothesized.

In summary, neither of the hypotheses for the correlational data are clearly supported. Because of the small sample sizes used and other design weaknesses to be discussed in Chapter V, no definitive conclusions should be drawn without replication.

CHAPTER V

DISCUSSION AND IMPLICATIONS

Overview

This chapter will discuss and consider the implications of the data presented in Chapter IV. There are four major divisions within the chapter: the theoretical analysis, the instructional hypothesis; multivariate analyses, the instructional hypotheses; correlational analysis; and the general implications and recommendations for future research.

The Theoretical Analysis: Change over Trials

The hypothesis that the dependent measures would increase over trials was supported by the data. It does appear that the students' recall and clustering performance increases over trials within each testing time. The nature of the curves is complex, each reflecting significant cubic, quadratic, linear and constant components. In general, the large number of observations may have provided us with a test of great strength that found significance in some of the components that appear little different, both numerically and graphically. This is particularly true within the linear and cubic components.

The major differences in the curves appeared to have been within the constant and quadratic components.

The significant constant component is of particular interest because it would tend to support the conclusion that trial one dependent measure performance increases over testing times in a significant fashion. From an instructional point of view, this is heartening, as it implies that the students at each succeeding testing time have become more familiar with the terminology. Remember, however, that these findings relate primarily to the recall measure, secondarily to segmental clustering and little, if at all, to suprasegmental clustering.

The quadratic component differences between recall and segmental clustering are also of interest. The negative quadratic coefficients for recall suggest a decelerating curve. This may imply that the recall would, after another few trials, flatten out and never rise again. There is a second possibility that will be discussed after the succeeding statements about segmental clustering.

The segmental clustering, quadratic coefficients are primarily positive (excluding that for testing time I which was negative), which implies an accelerating curve, one that does not appear to be peaking or flattening out. This is important because it suggests that the reporting of the clustering organization was accelerating over trials. Perhaps, if more trials had been given, the clustering

would have increased in an absolute sense, to a sufficient degree to change the nature of the recall curves. For instance, it is possible that if eight or ten trials were given, the recall would, instead of flattening out from trial three or four on, merely reach a plateau for a few trials before resuming a second increase in performance or period of acceleration. One obvious suggestion then would be the possibility of studying performance over a longer series of trials.

A second feature of interest is the redundancy of the suprasegmental clustering measure. Once the effects of the recall and segmental clustering measures have been examined, the suprasegmental clustering measure adds no new information in regard to change over trials.

Though the suprasegmental clustering measure is redundant and unnecessary when the segmental clustering measure is obtainable for a set of words, or data, it is possible that by nature of the difference in definition of the clustering measures that there may be experimental questions based on materials that are not amenable to the use of the segmental clustering measures. These materials may be more amenable to the use of the suprasegmental clustering measure and analysis. This question in regard to the relative viability of the two clustering measures is one that appears worth further examination.

A final interesting feature is the low level of performance evidenced throughout the data. For example, looking at the overall means for the four trials in the recall measure (Table 9), the recall increases from 26% to 40% over four trials. This would indicate an approximate increase from 12.7 (26%) words on trial 1 to 19.6 (40%) words on trial 4. Bousfield and Cohen (1953) found an increase over five trials with 60 words from a mean of 23.9 (40%) at trial 1 to a mean of 37.9 (63%) at trial 5. It was expected that since the words used in this research had been so recently studied in relative depth, that the performance would at least equal and perhaps be higher, that the performance on words which had not necessarily been under recent examination as those in the research of Bousfield and others. These findings suggest for simple recall purposes, instructional terminology does not become equivalent to "high frequency" words such as colors, animals, professors, etc., with such short exposure.

Bousfield's clustering (1953) measure about doubled over the five trials, as did those in this study. Again it is not the percentage of increase that is disappointing, but rather the absolute level of performance observed. In general, it appears that the words and their relationships studied in the materials for this research did not behave as high frequency words might have been expected to. This would suggest that long term exposure

to the terminology, jargon, etc., inherent to various subject matter areas, would be necessary for the words to become "high frequency" for the individual. In retrospect, this makes a good deal of common sense. It would be interesting to experiment with doctoral students or scholars who are concentrating on a particular field of study over several years to see what effects could be noted in key terminology drawn from their fields, or in other words, to examine how long it takes people to acquire the often talked about "structure of knowledge" for a given subject area.

Of course, alternative hypotheses to that above are possible for explaining the generally low level of performance. Perhaps the students were poorly motivated. This is possible, but most verbal learning research is done with similar kinds of students, so it does not appear to be a strong possibility. Also, it may be related to the strangeness of the task. This possibility could be examined by the manipulation of the number of trials, as mentioned earlier. The analysis of the hypothesized change in dependent measure performance over trials has been presented. The ensuing section will discuss the hypotheses of within unit increase in dependent measure performance over testing times.

The Instructional Analyses: Change Within Units Over Time

The means themselves, in the absence of statistical analysis, seemed to support the hypotheses of interest. True, the interactions were complex, with different units behaving differently at different testing times on different dependent measures, but the trends of means themselves were seemingly supportive. The statistical analysis, however, suggested that neither clustering measure evidenced anything of particular significance and that the recall measures did so only in units 3 and 4.

The significance in units 3 and 4 makes one wonder why the lack of significance for the recall in units 1 and 2. It is possible that the words chosen from 1 and 2 were not as "crucial" in absolute terms as were those in units 3 and 4. On the other hand, if all could be defined as relatively equivalent in importance, perhaps the instruction for units 3 and 4 was more "successful" than that for the first two units. Either hypothesis, or both, could be true, and both would be difficult to "test." I might add, however, that from the judgment of some of the material developers, the last two units are more differentiated and well developed than the first two, and this feeling is supported to some extent by the fact that the first two units are presently undergoing major redevelopment by those involved in the course.

The lack of any evidence that the clustering measures were non-redundant with the recall measure was also unexpected. In effect, the clustering measures are of no significance in showing change within units over time. This suggests that the students did not pick up any significant degree of organization over the quarter that could be exposed by the free recall task. Certainly these findings would negate the possibility of using the free recall task to measure acquisition of organization that was treated in an incidental fashion by the course materials, as was the case here. To clarify, the students were not asked to memorize the key terminology, nor were they asked to form particular associations. The words were related contextually throughout the materials, but the way that the student stored the information he studied was not directly manipulated. If the structure of interest in this study was thought to be of crucial importance by the material developers, then the inclusion of specific objectives and experiences designed to teach that structure might make a difference. The results do imply, however, that students on the average do finish the course (recall that all subjects that did not finish the course were deleted from the samples), and appear not to have concerned themselves, on their own, with the relationships between the key terminology identified by the author. Also, since all students who finish the course must pass

a series of mastery tests, the organization tapped by the free recall task does not seem to be a direct outcome of successfully studying the course materials to pass the tests given in the course. This relationship between the dependent measure performance and mastery test performance is examined further in the next section of Chapter V, when the correlational analysis is discussed.

The possibility of the need for more direct tuition of the desired relationships is partially supported by re-examination of Johnson's work in free association (1967, 1965, 1964). He did his research in the context of physics. The words he chose were related to each other in the form of equations; which also incidentally must be known to solve physics problems which would comprise the evaluation instruments. The relationships among the various words chosen then were much more directly taught and mastery of them was demanded of the students to successfully complete the course.

The discussion of the analysis of the hypotheses of change in within-unit dependent measure performance has been presented. The next section discusses the outcomes of the analysis of the hypothesized correlations between dependent measure performance and mastery test performance.

The Instructional Hypotheses: Correlations Between Unit Dependent Measure Performance and Unit Mastery Test Scores

First, each of the hypotheses was discussed only on the basis of the significant correlations to be found in the data. In general, units 1 and 4 supported the first hypothesis (no relationship between entry dependent variable performance and later mastery test performance) and did not support the second hypothesis (positive correlations after unit completion). Units 2 and 3, on the other hand, did not support the first, and did support the second, particularly within the recall and suprasegmental clustering measures. In effect then, unit recall and suprasegmental clustering which evidence significant positive correlations before instruction also tend to show significant positive correlations after instruction and vice versa. The explication of this phenomenon is difficult. Perhaps some students had previous experience with the terminology, etc., in units 2 and 3, thus starting the quarter with an advantage over others who had no such previous experience. If this was the case, the instruction tended not to mitigate against or wash out these early advantages, as those who started out ahead, stayed ahead. It is possible that the post-instructional significant positives describe the same sort of individual differences in retrospect. This explanation would leave

two possibilities for the explanation of units 1 and 4. The first possibility is that the students had uniformly little advantage at the beginning and the instruction did not help or hinder students in any systematic way. The second possibility is that the students entered the units with the same variance of previous experience with the terminology, etc., but that the instructional materials helped the students without prior experience to overcome their initial disadvantage. One other possibility that has not yet been discussed is that for units 2 and 3, the mastery tests more directly test the organization exposed by the free recall task while unit 1 and 4 tests do not.

Which of the above explanations is the most accurate, if any of them are, is impossible to say. Further research testing the effects of entry behavior, the effects of differently composed mastery examinations and the effects of differently organized textual materials would perhaps help unravel the relatively obscure data outcomes. The use of larger sample sizes would also lend credibility to the findings, as the need to regroup subjects. The correlational analysis may have produced spurious results.

The general lack of significance and clear trends may be suggesting that the organization or learning needed to perform well on the free recall task is not closely related to or reflected by that necessary to do well on the mastery tests. Also, it is possible that the low

correlations are an artifact of limited differences in performance on the mastery tests. This is possible, because mastery tests by definition and design are not usually of high difficulty and discrimination in classical measurement terms.

General Implications and Recommendations for Further Research

The free recall task, at this time, would not appear to be of any practical use in an instructional setting. It seems to be subject to confounding influences and complex interaction patterns which provide anyone needing relatively clean information with results of little practical significance. The conceptualization for this research appeared logical enough, but it seems that, at the very least, we tried to look at too much too soon. Laboratory research is done under closely controlled conditions and it is perhaps naive to expect that so many controls could be successfully dropped at once. It would be wise in the future for researchers attempting to apply such laboratory techniques to an instructional setting, to limit the scope of the research much more severely at first.

In this study, for instance, group testing supplanted the more frequently used individual testing; multiple units of material, testing times and dependent variables were used. Perhaps one or two of these

divergencies at a time is more than any paradigm could withstand. Add to this, the fact that students move at their own pace through the course and the nature of the study materials and mastery tests is out of the control of the experimenter, and it is not at all surprising, in retrospect, that the outcomes were somewhat disappointing and unclear.

The new dependent variable of suprasegmental clustering still seems like it may be worthy of further investigation, if not in the instructional setting, then back in the laboratory. Certainly in the correlational analysis, suprasegmental clustering produced far more significant correlations than did segmental clustering. This suggests the possibility that examination of student subjective organization, a piece of which suprasegmental clustering represents as defined in this study (page 19), might account for more variance than did the segmental clustering and suprasegmental clustering together. Also, suprasegmental clustering may be able to supplant segmental clustering in materials where that type of clustering is inappropriate.

While the nature of classroom learning was not greatly illuminated in this study, it did make patently clear the complexity of the instructional setting and the probable need to develop new means of assessing and studying this type of learning. More such attempts at

transfer from laboratory to classroom may be necessary before we know for sure whether laboratory techniques such as the free recall task have any practical significance outside of the laboratory setting, but this research would suggest that they do not.

APPENDICES

APPENDIX A

TESTING BOOKLET

TESTING BOOKLET

Section #	Section Leader	Date
	Name	Student #

Place an "X" in the box beside those units you have completely finished.

I have completed the following units:

☐ Introductory

☐ Assessment

☐ Objectives

☐ Strategies

I completed the last of the above units on the following date: _____

I have started but not completed the following units:

☐ Introductory

☐ Assessment

☐ Objectives

☐ Strategies

Random Order A

Kohlberg	criteria
teaching	operant
negative	intellectual
stimulus	socio-emotional
behavioral	instruction
design	variables
terminal	Piaget
conditions	givens
I.Q.	convergent
recognition	response
curricular	complete
learning	recall
non-behavioral	incomplete
modeling	shaping
environmental	in-class
ends	attitude
moral	formal
valuing	feelings
beyond-school	attending
means	committing
pre-test	preoperational
positive	reinforcers
strategies	classical
respondent	achievement
human	

Random Order B

respondent

attending

beyond-school

recognition

positive

design

Piaget

criteria

intellectual

ends

convergent

Kohlberg

operant

shaping

non-behavioral

formal

givens

terminal

human

response

recall

committing

valuing

variables

attitude

stimulus

incomplete

teaching

classical

preoperational

moral

environmental

reinforcers

learning

feelings

complete

I.Q.

conditions

in-class

socio-emotional

means

pre-test

behavioral

strategies

achievement

modeling

instruction

negative

curricular

Random Order C

learning	pre-test
means	negative
I.Q.	response
operant	formal
Piaget	curricular
terminal	environmental
criteria	attitude
reinforcers	preoperational
instruction	stimulus
recall	committing
intellectual	achievement
valuing	recognition
moral	givens
behavioral	attending
variables	complete
shaping	in-class
Kohlberg	positive
design	modeling
classical	beyond-school
feelings	conditions
strategies	teaching
convergent	respondent
socio-emotional	human
non-behavioral	incomplete
ends	

Random Order D

means	environmental
classical	formal
pre-test	negative
response	ends
givens	positive
incomplete	human
valuing	attending
committing	stimulus
non-behavioral	feelings
attitude	conditions
socio-emotional	learning
I.Q.	achievement
modeling	in-class
beyond-school	terminal
reinforcers	shaping
intellectual	strategies
teaching	behavioral
operant	complete
recall	variables
curricular	convergent
respondent	Piaget
instruction	criteria
preoperational	Kohlberg
moral	design
recognition	

APPENDIX B

INSTRUCTIONS TO THE SUBJECTS

INSTRUCTIONS TO THE SUBJECTS

The following instructions were given verbatim to all subjects at all testing times.

> I am involved in research with Education 200. The task I am going to ask you to perform is designed to give us feedback to help improve the course. Your performance will in no way affect your grade or your progress in the course.
>
> Would you please fill out the cover sheet of the booklet I am passing out? Please do not open the booklet until I ask you to.

--Pause while the students fill out the cover sheet.

> The task may seem strange to you but I hope you will find it enjoyable. Please don't open the booklet until I finish giving the instructions. When you do open the booklet you will note that you have four pages that look like this (I hold up a sample word list). Each of these pages will have two columns of words. After each page like this, you will find a blank page. Okay. Now the task will go as follows. When I tell you to begin, you will have about one minute to study all of the words on the first page. When I tell you to stop, you should immediately turn to the following blank page and write down all of the words you remember from the study trial. You should write them down in whatever order they come to you. You will have about three minutes for each of these test trials.
>
> Finally, we will do four such trials, one right after the other. Now are there any questions?

--Pause for questions.

APPENDIX C

C_{ss} COEFFICIENTS AND GRAPH FOR THE THEORETICAL ANALYSIS

APPENDIX C.--Estimated coefficients for the learning curves over trials, based on the suprasegmental clustering data.

Testing Times	Constant	Linear	Quadratic	Cubic
I	.072	.039	-.016	-.004
II	.078	.048	-.009	-.012
III	.076	.034	.004	-.009
IV	.090	.048	-.007	-.101
V	.091	.036	-.022	-.011

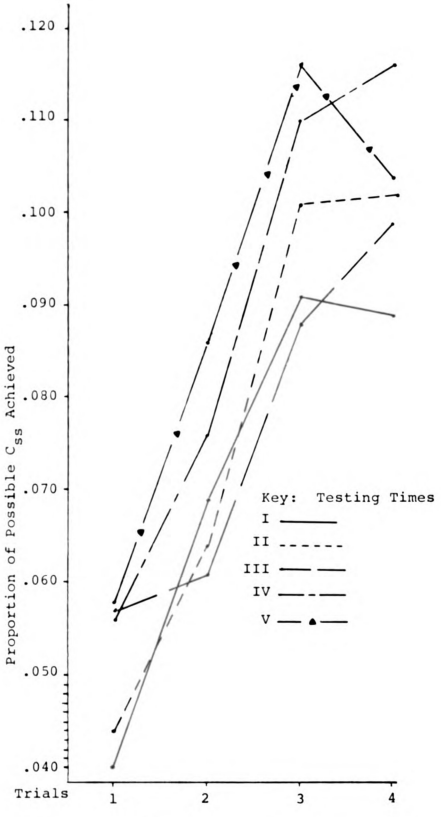

Figure 5.--Proportion of possible suprasegmental clustering achieved over four units by trials and testing times.

APPENDIX D

GRAPHS OF C_s AND C_{ss} DATA FOR THE INSTRUCTIONAL ANALYSIS

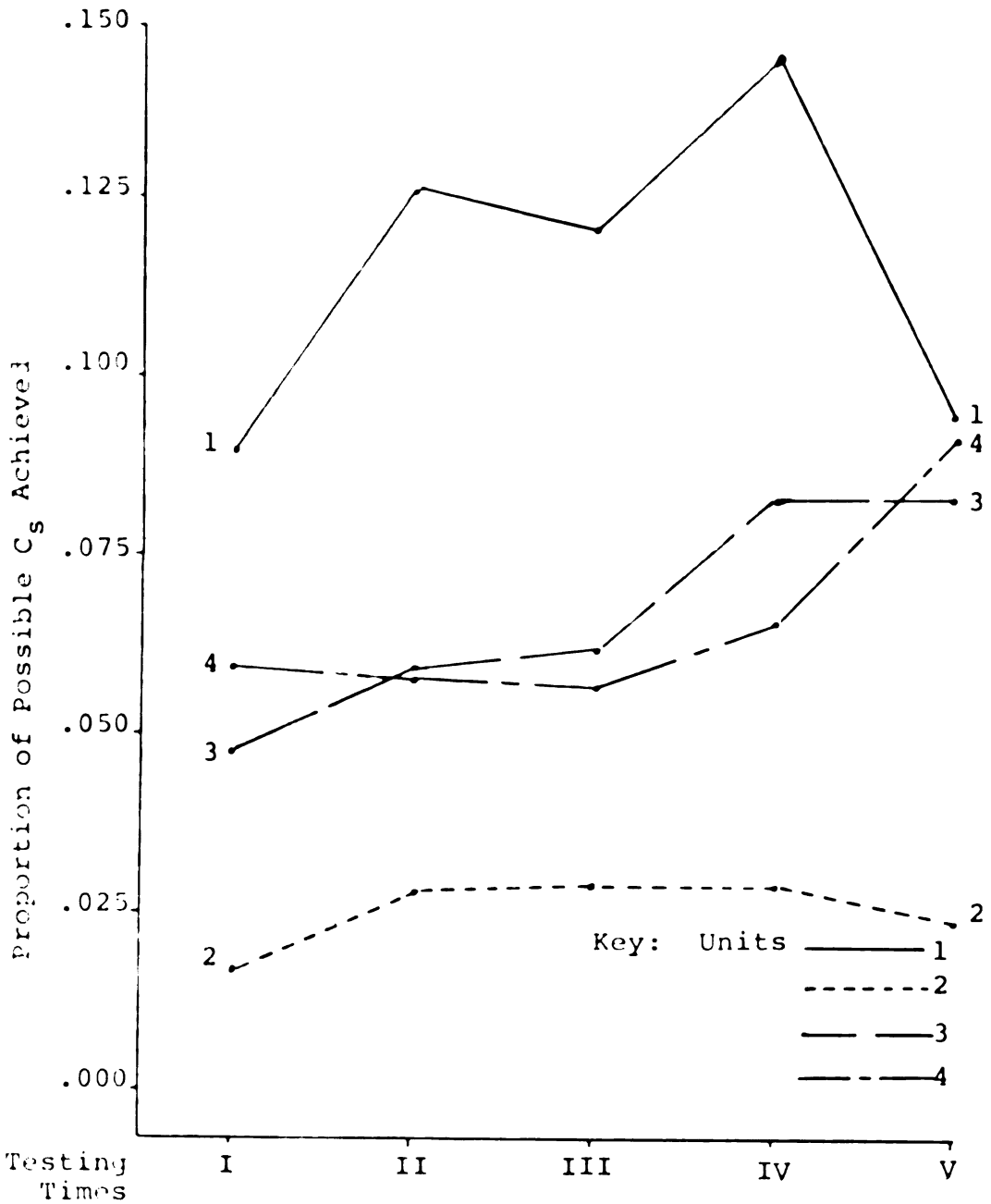

Figure 6.--Proportion of possible unit segmental clustering achieved over four trials by testing times.

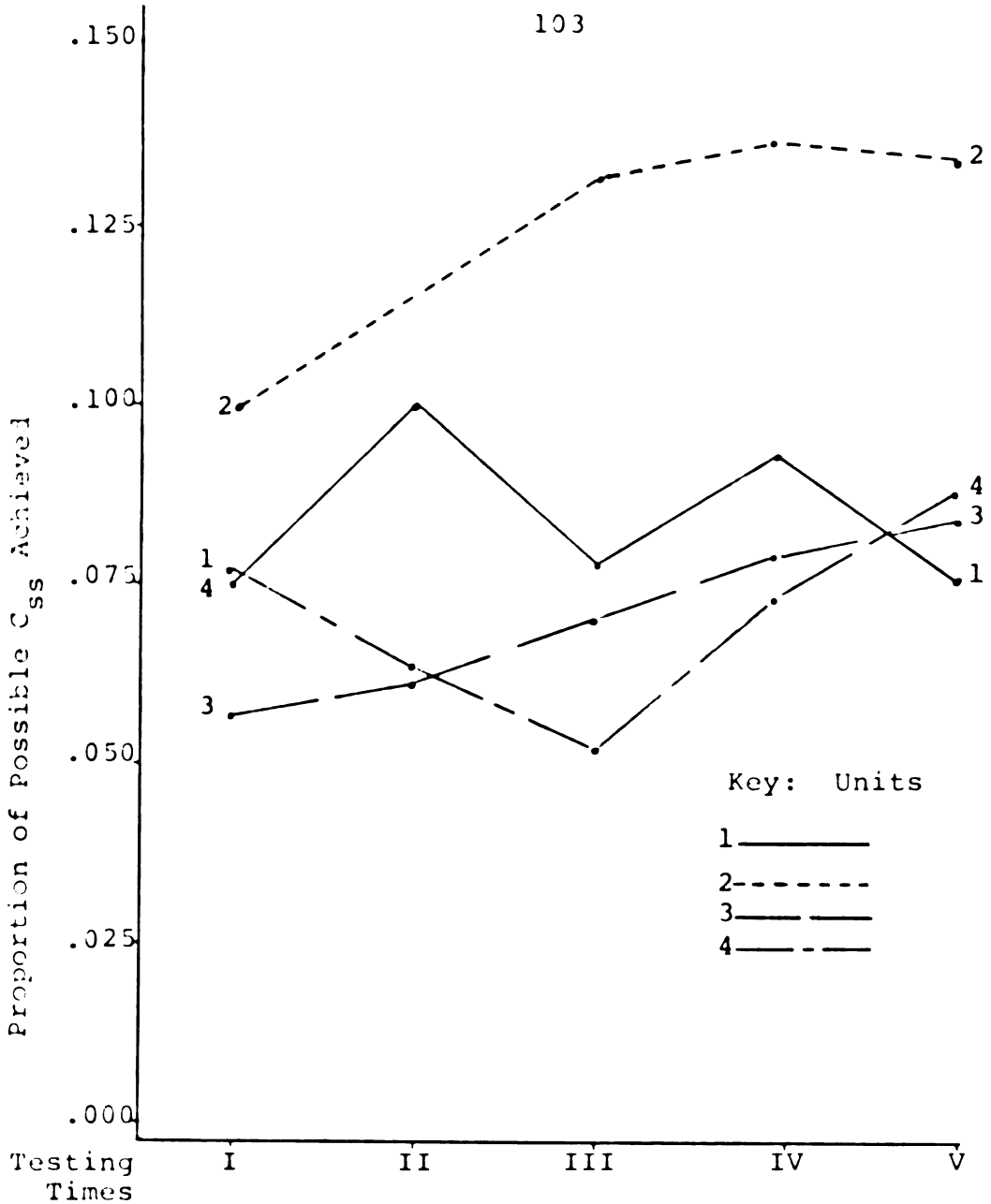

Figure 7.--Proportion of possible suprasegmental clustering over four trials by testing times.

BIBLIOGRAPHY

BIBLIOGRAPHY

Adams, J. A. Human memory. New York: McGraw-Hill, 1967.

Ausubel, D. P. A subsumption theory of meaningful verbal learning and retention. J. Gen. Psychol., 1962, 66, 213-214.

Ausubel, D. P., Stager, M., and Gaite, A. J. H. Proactive effects in meaningful verbal learning. J. Ed. Psychol., 1969, 60, 59-64.

Ausubel, D. P., Stager, M., and Gaite, A. J. H. Retroactive facilitation in meaningful verbal learning. J. Ed. Psychol., 1968, 59, 250-255.

Belanger, M. Learning studied in science education. Rev. Ed. Research, 1969, 39:4, 377-395.

Bereiter, C., and Engleman, S. Teaching disadvantaged children in the preschool. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

Bleuler, E. Organization and pathology of thought. (D. Rappaport, ed.). New York: Columbia University Press, 1951.

Bloom, B. S. Mastery learning. Evaluation comment I, no. 2. Los Angeles: Center for the Study of Instructional Programs, University of California, 1968.

Bloom, B. S. (ed.). Taxonomy of educational objectives: The classification of educational goals. Handbook I. Cognitive domain. New York: McKay, 1956.

Bloom, B. S., Hasting, J. T., and Madaus, G. F. Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill Book Company, 1971.

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. J. Gen. Psychol., 1953, 49, 229-240.

Bousfield, W. A., and Cohen, B. H. The effects of reinforcement on the occurrence of clustering in the recall of randomly arranged associates. J. Psychol., 1953, 36, 67-81.

Briggs, G. E. Retroactive inhibition as a function of the degree of original and interpolated learning. J. Exp. Psychol., 1957, 53, 60-67.

Bruner, J. S., Goodnow, J. J., and Austin, G. A. A study of thinking. New York: John Wiley & Sons, Inc., 1956.

Bruner, Jerome S. The process of education. Cambridge, Mass.: Harvard University Press, 1960.

Cofer, C. N. On some factors in the organizational characteristics in free recall. Am. Psychologist, 1965, 20, 261-272.

Cofer, C. N., and Bruce, D. R. Form-class as the basis for clustering in the recall of non-associated words. J. Ver. Learning and Ver. Behavior, 1965, 4, 386-389.

Cofer, C. N., and Foley, J. P., Jr. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. Psychol. Rev., 1942, 49:6, 513-540.

Deese, J. Influence of inter-item associative strength upon immediate free recall. Psychol. Reports, 1959, 5, 305-312.

Deese, J. On the structure of associative meaning. Psychol. Rev., 1962, 69, 161-175.

Frase, L. T. Cybernetic control of memory while reading connected discourse. J. Ed. Psychol., 1969, 60, 49-55.

Frase, L. T. Effect of incentive variables and type of adjunct question upon text learning. J. Ed. Psychol., 1971, 62, 371-375.

Gagne, R. M. The conditions of learning. (2nd ed.) New York: Holt, Rinehart and Winston, Inc., 1970.

Hays, W. L. Statistics. New York: Holt, Rinehart and Winston, Inc., 1968.

Heimer, R. T. Conditions of learning in mathematics: Sequence theory development. Rev. Ed. Research, 1969, 39:4, 493-508.

Henderson, J. E., Willard, S. M., Barnes, H. L., and Prawat, R. S. The individual and the school. East Lansing, Mich.: Michigan State University Printing, 1971.

Jenkins, J. J., and Russell, W. A. Associative clustering during recall. J. Abnorm. and Soc. Psychol., 1952, 47, 818-821.

Johnson, P. E. Associative meaning of concepts in physics. J. Ed. Psychol., 1964, 55:2, 84-88.

Johnson, P. E. Word relatedness and problem solving in high school physics. J. Ed. Psychol., 1965, 56:4, 217-224.

Johnson, P. E. Some psychological aspects of subject matter structure. J. Ed. Psychol., 1967, 58:2, 75-83.

Kent, G. H., and Rosanoff, A. J. A study of association in insanity. Am. J. Insanity, 1910, 67, 37-96.

Kuhn, T. S. The structure of scientific revolution. Chicago: University of Chicago Press, 1970.

Miller, G. A., Galanter, E., and Pribram, K. H. Plans and the structure of behavior. New York: Holt, Rinehart and Winston, Inc., 1960.

Murdock, B. B., Jr. The immediate retention of unrelated words. J. Exp. Psychol., 1960, 60, 222-234.

Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.

Osgood, C. E. Behavior theory and the social sciences. Behav. Sci., 1956, 1:3, 167-185.

Paivio, A. Mental imagery in associative learning and memory. Psychol. Rev., 1969, 76, 241-263.

Romberg, T. A. Current research in mathematics education. Rev. Ed. Research, 1969, 39:4, 473-491.

Rothkopf, E. Z. Textual constraints as a function of repeated inspection. J. Ed. Psychol., 1968, 59, 20-25.

Rothkopf, E. Z., and Coke, E. U. The concept of mathemagenic activities. Rev. Ed. Research, 1970. 40, 325-336.

Rothkopf, E. Z., and Coke, E. U. Learning about added sentence fragments following repeated inspection of written discourse. J. Exp. Psychol., 1968, 78, 191-199.

Russell, W. A., and Jenkins, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test. Technical Report No. 11, University of Minnesota, Contrast N8onr-66216, Office of Naval Research, 1954.

Saltz, E. The cognitive bases of human memory. Homewood, Ill.: Dorsey, 1971.

Skinner, B. F. Verbal behavior. New York: Appleton Century Crofts, Inc., 1957.

Smith, H. A. Curriculum development and instructional materials. Rev. Ed. Research, 1969, 39:4, 397-413.

Staats, A. W., and Staats, C. K. Complex human behavior. New York: Holt, Rinehart and Winston, Inc., 1963.

Thorndike, E. L., and Lorge, I. The teacher's word book of 30,000 words. New York: Cambridge University Press, 1944.

Tulving, E. Intratrial and intertrial retention: Notes toward a theory of free recall verbal learning. Psychol. Rev., 1964, 71, 219-237.

Tulving, E. Subjective organization in free recall of "unrelated words." Psychol. Rev., 1962, 69, 344-354.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon and D. L. Horton (eds.), Verbal behavior and general behavior theory. New York: Prentice Hall, Inc., 1968.

Tulving, E., and Madigan, S. Memory and verbal learning. Annual Rev. Psychol., 1971, 21-84.

MICHIGAN STATE UNIVERSITY LIBRARIES
3 1293 03082 1577