1980

A methodology for evaluating verbal classification schemes and verbal task variables.

Philip N. Chase
University of Massachusetts Amherst

A METHODOLOGY FOR EVALUATING VERBAL CLASSIFICATION SCHEMES AND VERBAL TASK VARIABLES

A Thesis Presented

By

Philip N. Chase

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

May 1980

Psychology

A METHODOLOGY FOR EVALUATING VERBAL CLASSIFICATION SCHEMES AND VERBAL TASK VARIABLES

A Thesis Presented

By

Philip N. Chase

Approved as to style and content by:

Beth Sulzer-Azaroff, Chairperson of Committee

John W. Donohoe, Member

Arnold D. Well, Member

James M. Royer, Member

Bonnie Strickland, Department Head
Department of Psychology

ACKNOWLEDGEMENTS

At long last, the first step in a long chain of scientific endeavors is completed. I am excited. I believe that I have been trained well. And so, I would like to thank everyone that has made this possible. But how does my training facilitate acknowledging all of those who have had a major impact on my behavior? First, it gives me a historical perspective. From this perspective I remember the warmth and love that was provided by my parents John and June and the love and competition provided by my brothers, Jack, Mike and Tim. Second, my training provides me with a social, environmental perspective. From this I recall the consistent, immediate support supplied by my friends. Especially, I thank Karen. Third, my training has provided me with a professional perspective. From late night planning, scheming and scamming with Kent Johnson to Tuesday organizing with Beth Sulzer-Azaroff, I have been fortunate to experience the best in behavioral training. Of course, all these categories overlap. There is no simple way to separate the historical from the social or the social from the professional. All that one can do is key in on certain ways that others have been of assistance. Therefore, the last perspective that my training has provided is to be able to isolate the key factors in this specific project. From this, I know to thank each of my committee members, John Donahoe, J.M. Royer and Arnold Well, a superb typist, Carol Vreeland, and especially, again, my advisors and friends, Kent Johnson and Beth Sulzer-Azaroff.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

Chapter
I. INTRODUCTION. 1

The Problem 3
Methodological Issues 19
The Classification of Verbal Tasks. 33

II. EXPERIMENT I. 50

Purpose 50
Methods 50
Subjects. 50
Personnel 51
Setting 51
Materials and Apparatus 51
General Procedure 51
Reliability (Interscorer Agreement) 53
Results 53

III. EXPERIMENT II 63

Purpose 63
Method. 63
Subjects. 63
Personnel 63
Setting 64
Materials 64
General Procedures. 64
Reliability (Interscorer Agreement) 66
Results 68
Discussion. 70

IV. EXPERIMENT III. 72

Purpose 72
Methods 73
Subjects. 73
Personnel 73
Setting 74
Materials and Apparatus 74
General Procedures. 75
Reliability 78
Results 81
Accuracy Data 81
Reading Rate Data 92
Relations Between Accuracy and Reading Rate Performance 96

Discussion. 100

V. CONCLUSIONS 111

Validation of the Typology. 111
Materials Assessment. 116
Summary 121

. . . . . .

REFERENCES 123

APPENDICES 131

LIST OF TABLES

1. Subject Classification Performance on All Twenty Tasks. 55
2. Subject Classification Performance on Tasks Written by Experimenter 56
3. Subject Classification Performance on Tasks From Commercial Materials. 57
4. Number of Minutes to Completion of Handout and Classification Test For Each Subject 58
5. Decision Rules for Accepting Concepts and Subject Answers to Exemplify Tasks. 67
6. Number of Correct Original Examples Written for Each Concept 69
7. Order of Conditions for Each Subject by Session 79
8. Matrix of Arcsine Transformations of Proportions. 82
9. Source Table for ANOVA of Accuracy. 83
10. Source Table for Scheffé's Post Hoc Comparison Between Means. . 85
11. Proportion of Correct Performances for Four Sample Subjects Across the Three Concepts 89
12. Proportion of Correct Performances for Four Sample Subjects on Four Different Sequences of Concept Presentation 90
13. Rank Order of Tasks by Proportion Correct 91
14. Source Table for ANOVA of Reading Rates 93
15. Single-Subject Mean Reading Rate Performances and Standard Deviations For Each Concept 97
16. Mean Reading Rates For All Subjects Across Sessions 98
17. Rank Order of Mean Reading Rates for Example Identification Tasks 99
18. Rank Order of Tasks on the Basis of Accuracy and Reading Rate . 107
19. A Typology of Verbal Behavior 115

LIST OF FIGURES

1. Single-subject Responses to Each Task for Each Concept. 86
2. Absolute Frequency of Pooled Reading Rates for Each Subject . . 94
3a. Scatter Plot of % Correct and $\bar{X}$ Reading Rate for Each Task, Constructional Approach 101
3b. Scatter Plot of % Correct and $\bar{X}$ Reading Rate for Each Task, Abulia. 102
3c. Scatter Plot of % Correct and $\bar{X}$ Reading Rate for Each Task, Tau Effect. 103

C H A P T E R I

INTRODUCTION

Stating that transfer of learning is quintessential to developing a systematic understanding of human behavior is, by this point, a truism. It is clear that few other domains of human interest have captured the imaginations and precipitated the efforts of more investigators. Countless philosophers, psychologists, linguists and educators have tackled the problems generated by a simple observation: that previous experiences affect behavior in subsequent situations. This general definition of transfer subsumes so much of human behavior, it is of little wonder that so many reviewers and researchers have expressed the importance of the topic. Ellis (1965) stated that there is perhaps no more important topic in psychology than transfer of learning. Deese (1958) concluded that there is no more important topic. More recent reviewers (Perkins, 1976; Royer, 1978) have expressed similar views. Perhaps the most cogent statement of this topic's importance is: "It is difficult to think of any human learning that is not affected by earlier learning" (Ellis, 1965, p. 5). Given such kudos, it is not surprising to find a vast depth and breadth of literature devoted to transfer of learning or training. Ellis (1965) referenced 197 publications in his frequently quoted review of the literature. Postman (1971) cited close to 300 references in his review chapter on transfer of verbal learning. Royer (1978) referenced

forty-four studies in a review limited to theoretical developments of learning transfer. Other recent sources of review in areas subsumed by the concept of transfer reveal similar interest in and concern for the effect of previous learning on behavior in subsequent environments (i.e. Anderson and Bower, 1973 - information processing; Clark, 1971 - concept formation; White, 1973 - cognitive hierarchies; Andre, 1979 - levels of questioning; Rickards, 1979 - adjunct questions in text).

This vast amount of literature and the variety of topics covered by the transfer phenomenon require any reviewer attempting an analytic and evaluative picture of the literature to limit his or her scope. Without restricting a review to certain areas of transfer research, to certain eras of investigation, to certain methodologies used to investigate transfer or to certain theories used to organize the results of transfer research, a reviewer would suffocate in a quagmire of exposition. To this end, this thesis is restricted in a number of ways. First, it is restricted to transfer of verbal or cognitive learning. More specifically, it is restricted to the practical consideration of the verbal transfer that occurs from training provided in secondary and post-secondary schools. How do the cognitive behaviors that are taught within these environments transfer to the situations a student faces outside of these specific environments? What practical contingencies can be derived from the hundreds of studies that have investigated verbal transfer? What are the most efficient means of effecting the greatest amount of behavior change? These general questions define the first restriction. Second, it is restricted to a particular conceptual stance: the development of a theory of verbal learning that is consistent with the experimental

analysis of behavior. Inherent in this restriction is a necessity for conducting research from a basically inductive methodological perspective and an emphasis on single-subject data. Finally, this thesis is restricted to isolating a few variables that are prerequisites to testing a typology of verbal tasks. The design of this typology is based on conclusions drawn from literature that are consistent with the first two restrictions. It is a typology that is designed to meet the practical concerns of secondary and post-secondary instruction and to meet the functional demands of an experimental analysis of behavior.

These are the practical, conceptual and experimental restrictions placed on this thesis. The rest of this introduction develops, in detail, each of these. Therefore, it is divided into three sections. In the first section, previous literature devoted to developing efficient and effective cognitive learning is analyzed in terms of its practical input. The second section explicates the need for a fusion of inductive and deductive methods of inquiry and single-subject methodologies to further investigate the problems of verbal transfer. Finally, the third section presents a typology of verbal tasks and delineates the questions that need to be answered with respect to this typology.

The Problem

One problem inherent in assessing transfer of cognitive training from one educational setting to other settings is defining the verbal behavior of interest. Bostow (1976) stated a general educational goal that may prove useful for tackling this problem. Bostow (1976) indicated the need to teach students to engage in the behavior of professionals or advanced students in a discipline. Content experts engage in various

kinds of discourse with respect to the materials within their discipline. Experts can state the facts and figures of their discipline. They can relate seemingly obscure similarities between concepts. They can identify real world instances of the concepts developed in their disciplines. They can provide students with intriguing examples of these concepts. Finally, they can ask questions and determine methods for answering these questions, or when faced with a problem can determine ways of solving the problem; in short, experts can problem-solve. The sum of these various classes of verbal behavior constitutes the complex verbal repertoire needed in advanced educational and professional environments. The particular components of this sum constitute what is necessary to teach students in secondary and post secondary instructional settings.

Explicit in Bostow's educational goals is the classification system for verbal behavior that was developed by Skinner (1957). Skinner categorized verbal behavior on the basis of functional distinctions. For example, an intraverbal is any verbal behavior, spoken, written or gestural, that is controlled by a spoken, written or gestural verbal stimulus that does not have point-to-point or one-to-one correspondence with the behavior. A tact, on the other hand, is verbal behavior that is controlled by a physical or non-verbal stimulus. Transcriptive behavior is a written response that is controlled by a written stimulus that has point-to-point correspondence with the behavior. The intraverbal, the tact and the transcriptive define functionally different relations between stimuli and responses. For the sake of isolating the different kinds of control exerted over verbal behavior, these classes are presented as discrete units. However, in the natural environment, most complex verbal

interactions are composed of combinations of these classes of behavior. Bostow (1976) claimed that instruction must be aimed at teaching these complex combinations of behavior.

In addition to proposing educational goals on the basis of Skinner's analysis of verbal behavior, Bostow (1976) also critiqued current teaching strategies. First, he claimed that most instruction is aimed at one class of behavior, the intraverbal. Second, he warned that teaching one class of behavior was insufficient to teach students the kinds of interactions in which professionals engage. Both of these assumptions are supported by data.

First, two recent surveys empirically validate Bostow's statement that teachers are directing their instruction to one class of responding. Semb and Spencer (1976) randomly selected 17 university instructors for their survey. They provided these instructors with definitions and examples of both recall and complex tasks. Recall tasks were those that required a specific answer to an item drawn directly from course material. Other tasks, including asking the student to identify or exemplify novel instances of a concept, were defined as complex. Then, they asked the instructors to estimate what percentage of their tasks met the definitions of recall and complex. The instructors estimated that 32.6% of their exam items were complex and 67.4% were of the recall variety. However, upon examination of these instructor's tests, it was found that only 9% were complex, whereas 91% were recall. These data indicate not only that instructors teach and test for predominantly rote recall, but also that they report that they teach a higher proportion of complex skills than is the case. Even though based on a limited sample, these results are consistent with earlier studies using Bloom's (1956) taxonomy (Hirshfield,

1969; Pfeiffer, 1967).

Chase, Johnson and Keenan (1977) conducted a survey of 20 commercially available introductory and educational psychology study guides. The definitions used for classifying recall and complex tasks were the same as those used in Semb and Spencer (1976). This survey found that only 10 of these study guides included complex cognitive tasks. 87% of the 1231 items randomly sampled from these 10 study guides were of the recall or definition variety.

At first blush, the results of these surveys indicate an astonishing trend: that teachers are not requiring complex verbal behavior from their students. These data seem to suggest that either educators assume that transfer will occur from these recall tasks to other classes of behavior or that they do not know what kind of behavior they are teaching. The disagreement between teacher estimates of the proportion of complex tasks and the actual proportion of complex tasks administered, suggest that teachers do not know whether they are teaching complex behavior or not (Semb and Spencer, 1976). Another survey lends support to this argument. Gallup and Warranch (1976) compare the accessory materials (i.e. study guides, instructors manuals, test item files) of 14 introductory psychology textbooks and found a great deal of inconsistency in their construction. They found that 12 of the accessory materials included fill-in and multiple choice questions, 3 had true-false items, and 3 provided matching or other items. This variety of tasks <u>across</u> instructional materials implies the lack of systematic rules for determining the kinds of tasks needed to teach students.

Although these studies lend empirical support to Bostow's (1976)

claim that teachers are requiring only one class of verbal behavior from their students, they indicated that Bostow should probably modify his conclusion that teachers are employing intraverbal tasks. Both Semb and Spencer (1976) and Chase, Johnson and Keenan (1977) demonstrated that the most prevalent tasks were transcriptive rather than intraverbal behavior. The definition of recall tasks used in both of these studies was similar to the definition of transcriptive behavior (Skinner, 1957). The stimulus (i.e. the text and instructional materials) has point-to-point correspondence with the response that the student has to make. The student needs only to memorize this point-to-point relation. This clarification is critical to validating Bostow's second assumption: that teachers need to teach their students more than one class of verbal behavior. If that class of verbal behavior is transcriptive, as demonstrated by the surveys, then there is research to support Bostow's claim. Most of the transfer studies reviewed employed memorization or definitional tasks that may be loosely defined as transcriptive behavior.

A number of studies have been conducted on memorization or transcriptive tasks. Whereas most of the studies indicate some facilitative effect of memorization, comparisons to other tasks, varied tasks and such variables as group participation during learning sessions have shown that memorization is not a superior form of instruction (Ellis, 1965).

As early as 1890, James demonstrated that memorization of one poem had no effect on the efficiency of memorizing a second poem. More recent studies have found little transfer from memorization tasks to complex problems. Katona (1940) found that memorization of the rules for performing a card trick led to superior performance on performing a new card

trick than no training at all. However, both watching a model perform each step of the trick and learning an orally stated principle about card tricks were superior to memorization. Hilgard, Irvine and Whipple (1953) found that subjects who simply memorized solutions to card tricks performed significantly worse than subjects who were taught to "reason out" the card trick from a formula on tasks that required learning new but related problems. Morrisett and Hovland (1959) found significantly better performance on problem-solving tasks when subjects practiced a variety of tasks as opposed to memorizing one particular task. Finally, Gagné and Brown (1961) found that instructional programs that were broken down into small sequential steps and required systematic use of previously learned concepts facilitated problem-solving performance better than programs that were also broken into small sequential steps, but required only the identification of a specific rule that could be memorized.

A number of studies have demonstrated that learning definitions (a transcriptive task) can facilitate transfer. Johnson and O'Reilly (1964) found that subjects could learn to discriminate between positive and negative instances of a concept (concrete descriptions that either exemplify the concept or exemplify another concept) when trained to supply the definitions of these concepts. Anderson and Kulhavy (1972) obtained similar results. However, other experiments have demonstrated that other kinds of tasks lead to superior performance identifying instances of a concept. Johnson and Stratton (1966) found that a variety of learning tasks (i.e. defining, using a concept in a sentence, giving synonyms, and classifying) resulted in better performance on novel instances of all these tasks, than training on any one task alone. Watts

and Anderson (1971) found that subjects asked to perform application questions (identifying examples and nonexamples of a concept) while reading a passage defining the concept, perform better on new application questions and as well on name (naming the concept read about and naming a person) and repeated examples (repeating an example of the concept given in the passage defining the concept) questions as did subjects who were asked either to answer name or repeated examples questions while reading. Both the name questions and the repeated examples questions can be categorized as transcriptive behavior. Miller and Weaver (1976) found improvement on identifying instances of concepts when subjects were given examples to identify during instruction. Subjects had previously been supplied only with definition tasks. Finally, Keenan and Grant (1979) discovered that subjects answered quiz and exam questions more accurately when study materials required responses to both positive and negative instance identification tasks, and definition tasks than when subjects were required to answer definition tasks alone.

To summarize, it appears that criticizing the current trend of providing students with memorization, definition or transcriptive tasks is justified. Research has demonstrated that these kinds of tasks are insufficient to teach students to engage in other kinds of verbal or cognitive behavior. Unfortunately, the answer to the question of what kinds of tasks are sufficient to teach a variety of classes of verbal behavior is not at all clear. The studies reported above and other research conducted to determine the components of effective cognitive learning have been criticized on a number of levels. Therefore, instructional strategies still need to be developed that teach students to engage in a variety of classes of cognitive behavior.

As indicated earlier, much transfer of learning research has been devoted to determining the component features of effective cognitive learning. In addition, other strategies for developing effective cognitive instruction (i.e. cognitive learning hierarchies) have been investigated. It is appropriate, therefore, to glean as much as possible from this previous research. Unfortunately, as several educational technologists have indicated, the development of classroom methods for promoting a variety of complex cognitive learning from research on cognitive behavior is limited (Bloom, Engelhart, Furst, Hill, & Kratwohl, 1956; Engelmann, 1969; Gagne, 1970; Markel and Tiemann, 1970; Clark, 1971; Royer, 1978). The reasons postulated for this lack of practical input from previous basic research do, however, suggest alternative strategies that can be investigated.

Three reviews of basic cognitive research have suggested that perhaps the vast differences between research and classroom contexts for teaching conceptual behavior contribute to the paucity of classroom applications of experimentally derived principles. Clark (1971) analyzed both research and classroom contexts and found five major differences. First, most research utilized conjunctive concepts, whereas classroom instruction requires a majority of conjunctive and relational concepts. Second, research subjects were required to attain already familiar concepts and several of these simultaneously. In contrast, the classroom student is required to learn new, unfamiliar and successive concepts. Third, the concept instances presented in research tasks were almost exclusively concrete, had more than one value, the values were discrete (e.g. black or white), the number of values was finite and the values were absolute (i.e. they were either there or not there and did not change

over time). On the other hand, the concept instances presented in classrooms are often abstract (verbal), have any range of values, the values are continuous (light blue-blue-dark blue), the number of values can be infinite and the values, properties and dimensions may change over time. Fourth, the research strategy for teaching concepts was most often discovery; the subject discovered the critical features, values etc. However, in the classroom, the most frequently used strategy is expository; the student is told the critical features to be learned. Finally, concept attainment in concept research was most often evaluated by a sorting task (either verbal or physical manipulation) while in the classroom students are asked to define or use the concept (Clark, 1971, pp. 254-255).

Gagne and Brown (1961) also speculated that the nature of research conducted on transfer of learning has prevented clear-cut application of experimentally derived principles to the classroom. The concepts acquired in the course of an experiment are not further used as in the solution of a problem, but are simply measured as being established. This point is similar to the final point made by Clark discussed above that the tasks used to evaluate subjects in basic research are different from the tasks used to evaluate students in classrooms.

In the same vain of criticism, Rickards (1979) analyzed the results of recent research conducted on inserting questions within prose passages. He claimed that all of these studies employed somewhat artificial settings in which subjects were not allowed to check the text passage when answering the inserted questions. Therefore, it is difficult to generalize to the natural environment in which students are reading textual materials and can always flip back through the pages to find the passage

or concept in question (Rickards, 1979, pg. 193).

Such analyses provide potential strategies to follow in designing research to determine the kinds of tasks that are necessary and sufficient to teach students in classroom settings to engage in a variety of verbal behavior. The use of laboratory concepts and settings that are more contrived than the concepts and learning environments of the real world is justified when it is necessary to control as many features of the environment as possible to determine the nature of simple learning. However, it is apparent that the nature of highly controlled, synthetic conceptual learning is different from learning concepts in the real world. Thus, two recommendations can be made for further investigations. First, real concepts can be investigated in the context in which concepts usually appear, within prose passages (cf. Ausubel, 1963). Second, cognitive learning laboratories can utilize a closer approximation to real learning environments. During study trials, subjects are allowed access to the text passages. However, during test trials, subjects are not allowed access to the passages or the study questions that they have answered. In these ways, controlled research can be conducted with concepts and environments that more closely approximate the situations in which real concepts are taught.

Two reviews have claimed that the lack of explicit rules for designing instructional tasks contributes to the problem of application. Gagné and Brown (1961) speculated that the rules that have been written for constructing learning programs (Skinner, 1958; Galanter, 1959) are too general to be used as guidelines. Gagné and Brown (1961) demonstrated that it was possible to generate three functionally different kinds of programs that adhered to the specified rules. This finding

implies that many kinds of programs, some effective and some not effective, can be written from these rules. In addition, Andre (1979) pointed out that it would be impossible to duplicate the kinds of questions that some researchers have found to be effective because the exact characteristics of the questions types are not explicated.

These criticisms identify the second recommendation for improving the methods used to investigate cognitive behavior. The definitions of the classes of behavior, and the methods for incorporating the classes of behavior into an instructional system must be made in exact functional terms. The analyses should take into account the following properties: the general, observable topography of the student's response, the conditions that are relevant to the occurrence of the response, and the dynamic or changing features of the student's response. By categorizing different classes of conceptual behavior according to these properties, new rules should avoid the trap of ambiguity.

Royer (1978) implied that the problem of applying principles taken from laboratory research on transfer of learning is a specific theoretical problem. The history of transfer of learning is replete with research conducted from a single perspective, environmental or stimulus-response theory. Briefly, environmental transfer theory assumes that events which share stimulus properties will control the behavior of the learner similarly, so that a response learned to the first event will also occur in the presence of the second event. This emphasis on requiring <u>shared environmental features</u> in order for transfer to occur is where the most difficulty is encountered. Royer (1978) presented examples of real-world problems that do <u>not</u> share features with the classroom tasks that one needs to learn in order to solve the problem. For

example, a child learns to compute the area of a rectangle in a classroom. After instruction the child is faced with the problem of determining the amount of carpet needed to cover a living room floor. On the basis of shared environmental features, one would not be able to predict whether or not the child can recognize the application of the mathematical skill learned in the class to the problem of carpeting the floor. If the child is able to transfer the skill of measuring areas of rectangles to measuring carpets and has not received examples of measuring carpets, the transfer must occur because of something other than shared environmental features. This example points out the limitations of the shared environmental features model. It is not clear that in all cases of transfer, that sufficiently similar features are available in both the learning and testing environment to predict transfer. Many researchers have turned to an internal model in attempts to solve this problem. In these analyses, the problem of transfer becomes one of retrieval of relevant knowledge that has been stored in internal structures. These models emphasize the behavior of the learners (if it is appropriate to call the internal processes of a living organism, behavior) as they interact with the environment. The learners develop schema or ways in which to organize the impinging environment. What seems necessary to investigate is how instruction can be organized so that students learn methods for structuring the environment of prose passages.

Although many interesting and potential explanations of this problem have been offered by those postulating the structure of internal mechanisms and the heuristic of schema, it may not be necessary to refer to the unobservable. It is possible to teach students to engage in a set

of observable responses that can be used to structure their learning from a text. Each of these responses can be defined according to the above guidelines (i.e. clear specification of the topography and the function of each relation between stimuli and verbal behavior) and the effect of making these responses on other tasks can be measured. Thus, the prediction of whether a response will occur or not during a testing session is not based solely on the similar features of the environment, but rather on the kinds of behaviors the subject has previously learned. The environment may have few similarities other than the general similarities of written passages.

The three basic criticisms stated above (i.e. the differences between real-world and laboratory cognitive learning, the ambiguous definitions of instructional tasks and experimental methods, and a theoretical concentration on the stimulus aspects of the environment) all emphasize the problem of deriving practical solutions from basic research. Unfortunately, in addition to these well documented problems, other prescriptive strategies have failed to produce valid procedures for teaching a range of cognitive behaviors. Various theoretical analyses have attempted to classify verbal or cognitive behavior in terms of complexity. The most notable of these analyses are by Bloom et al. (1966) and Gagné (1965). Both of these analyses have arranged cognitive behavior into hierarchial levels. The concept of level implies that behavior or tasks that are at the bottom of the hierarchy are prerequisites for tasks that are at the top of the hierarchy. For example, under Gagné's hierarchy it is assumed that students must learn to verbalize definitions before they can sort positive and negative instances of a concept. The concept of hierarchy implies that behavior at the top of the hierarchy subsume those at the

bottom of the hierarchy. For instance, Bloom's analysis level objectives subsume comprehension level objectives. A transfer of learning analysis of these assumptions would suggest that transfer occurs from higher level to lower level behavior or tasks, but not from lower to higher level tasks. Since the lower level tasks are prerequisites they would be subsumed by the higher level tasks. The instructional implications of this kind of categorization are that if a student is observed engaging in higher level tasks, then it can be assumed that his behavior will transfer to lower level tasks. However, if a student fails a lower level task, it will be necessary to teach both lower level and higher level tasks. Unfortunately, neither of these instructional implications is supported by conclusive empirical evidence. Whereas some studies have shown that it is necessary to engage in knowledge level behavior (memorizing a definition of a concept) which is a lower level task, before successfully engaging in sorting behavior, a high level task (Reed, 1946; Wolff, 1967), other studies have shown that subjects can perform sorting and application tasks (applying the concept in new situations) without having to verbalize the definition of the concept (Furth, 1961; deLuna, 1972). Gagné and his colleagues (Gagné and Paradise, 1961; Gagné, Mayor, Garstens, and Paradise, 1962; Gagné, 1962; Gagné and staff, 1965) demonstrated that in most cases, it was necessary to engage in the lower elements of different math hierarchies, before one can engage in the higher elements of the math hierarchies. The exceptions in Gagné's studies were attributed to methodological flaws. For example, White (1973) speculated that since pretesting was not used universally, some subjects not trained in lower level behaviors could already perform lower level tasks. However, Kolb (1967, 1968) used a similar methodology and the same hierarchy

as Gagné and colleagues and found that many subjects succeeded at tests of higher elements while failing to learn relevant lower level tasks. The conclusions that can be drawn from such discrepant data are limited. As White (1973) has indicated, both the studies that support the notion of hierarchies and those that refute this notion have had methodological problems. Therefore, it is impossible to construct a complete cognitive hierarchy at this point in time.

The fact that methodological problems have been cited as the cause of these inconclusive results is to be expected. Methodology arguments are one of the signs of a young and not very exact science. It may be true that particular hierarchies do exist, but that the method for testing the assumptions related to the hierarchy has not been implemented. This problem can be solved by utilizing alternative methodologies to test current hierarchies. However, there are other criticisms of current hierarchies that are more damaging from a practical standpoint. Discussions with instructors, curriculum designers and educational researchers suggest that the levels of these hierarchies are complex to learn and use (Sullivan, 1969; Williams, 1977). In addition, these two papers report that definitions of the hierarchy levels produce low agreement among content experts on the classification of objectives, or test items, into the categories of the hierarchies. This suggests that even if one could develop a study to demonstrate that one of these hierarchies does exist, the hierarchy would have to be rewritten in order to be of any practical use.

In sum, the inconclusive results, the methodological problems and the ambiguous definitions demonstrate that current hierarchies are of little use to the practitioner. However, as with the transfer and concept

research something can be gained from these studies. First, it has been shown that alternative methodologies need to be devised to test the hierarchial relations among cognitive behavior. The most important methodological rule gleaned from these studies is that exhaustive pretesting is necessary. White (1973) suggested that when testing the hypothesis that behavior I is higher than behavior II, it is essential to pretest all subjects and eliminate those who already possess behavior II. From a transfer of learning perspective this can be translated into pretesting for all classes of behavior for which transfer is to be evaluated. Therefore, if one is testing transfer from I to II, it would be essential to determine whether II is already part of the subject's repertoire.

The hierarchy research presented here also reinforces the previously stated notion that the categories of conceptual or cognitive behavior must be defined in more exact terms. Practitioners should be able to agree on what differentiates one class of behavior from another. Again, following the guidelines specified for functional definitions should help to accomplish this goal. Descriptive labels for each class of behavior should also be beneficial. Of course, these assumptions need to be tested. Williams (1977) has devised an empirical validation procedure for classification systems. All definitions need to be evaluated through this test (see Methods, Experiment I).

To summarize, the problems that others have found with previous strategies for teaching conceptual behavior have been helpful in generating rules for developing and testing alternative classification systems of conceptual behavior. First, it is important to use research concepts that are similar to real-life concepts. Second, it is beneficial to

define functionally the classes of conceptual behavior that are to be investigated. Each of these classes should be given a descriptive label and tested on practitioners. Third, it is necessary to concentrate on the behavior of the learner as well as the features of the environment. Finally, it is necessary to use methodologies that test all the assumptions of the classification scheme. This includes adhering to these rules while testing for various kinds of transfer and for testing assumptions about possible levels of difficulty across classes of behavior (hierarchies). A synthesis of these rules with consideration for Bostow's (1976) goals for higher education defines the practical restrictions placed on this research. These practical restrictions have been defined in order to facilitate solving the general problem under investigation: to determine a classification scheme of cognitive tasks that covers the range of behaviors in which professionals and advanced students of a discipline engage.

Methodological Issues

In order to tackle such a complex problem, a problem that has been investigated from many different perspectives without conclusive results, it is essential to take a conceptual stance. This thesis and the subsequent related research are concerned with the development of a theory of verbal learning that is consistent with the experimental analysis of behavior. As such, the conceptual framework springs from an analysis of verbal behavior that is a synthesis of the empirically determined methods, concepts and relations that are currently expounded by behavior analysts. The methodological components of the experimental analysis of behavior that differentiate it from other psychological approaches need to be

defined here to eliminate any possible confusion.

First, it is necessary to define the method of reasoning that is employed by experimental analysts. Many equate the inductive method of inquiry in psychology with behavior analytic approaches. This presumption is well founded when behavioral approaches are compared with other psychological approaches which adhere to a hypothetico-deductive method of reasoning. Skinner (1969) defined hypothetico-deductive methods in the following manner:

> Whenever a subject matter is inaccessible, when the variables of interest can not be manipulated or the effects observed, then tentative or hypothetical statements are made about them, theorems are deduced which refer to accessible data and the theorems are checked to confirm or refute the hypothesis.

Since behavior analytic research is concerned with subject matter that is observable, it has been concluded that the hypothetico-deductive method is unnecessary for behavioral research.

Sidman (1960) claimed that the explanation of behavior is not a logical process, thus, logical analyses (i.e. hypothetico-deductive) have failed to account for behavior. In support of this statement, Sidman suggests that beliefs about behavioral relations always come from the scientist's experiences. Since induction conforms to this assumption, then the inductive method of inquiry must be used to investigate behavior. These two investigators have been widely read and have influenced over two decades of behavioral research and reasoning. However, it can be shown that a simple dichotomy does not exist between inductive and deductive reasoning. The differences between their use in psychological experimentation are matters of degree. It is this difference in degree of induction and deduction that differentiates behavioral reasoning from

other psychological reasoning.

Inductive reasoning is generally defined as extending a relation(s) between two or more phenomena from specific cases to a general case. Thus, as Johnston and Pennypacker (1979) stated, induction involves accumulating facts, checking the similarities and differences between facts, organizing the facts according to similarities and describing these similarities. This statement can be augmented by saying that the inductive investigator conducts experiments to look for relations between and among two or more phenomena. Thus, the question that the inductive scientist asks is: what is the relation between one event and another. If no orderly relation is found, the procedures, methods, equipment and/or variables are manipulated until a relation is determined. Through these processes, inductive reasoning promotes attending to any relation that exists between phenomena, regardless of whether the data agrees or disagrees with previously stated assumptions. In fact, Polya (1954) claimed that the inductive process promotes looking for facts that tend to disagree with previously generated relations. This property is consistent with the scientific rule that no assumption can ever be absolutely proven, however, one anomaly, if reliably obtained, can disprove an assumption.

On the other hand, deductive reasoning generally involves finding particular cases which support a general relation that is believed to exist between two or more phenomena. Thus, as Skinner (1969) noted, tentative or hypothetical statements are made about the belief, theorems are deduced which refer to the possible results of the experiment and the theorems are checked to confirm or refute the hypothesis. This statement

can be augmented by the fact that in their most rigorous form, hypotheses are deduced by formal (mathematical) rules and are made in terms of relatively explicit predictions. The hypothesis takes the form: if event A occurs, then event B will occur. The deductive investigator conducts research to verify the hypothesis. The extent to which the experimental outcomes correspond to the hypothesis, reinforces the validity of the hypothesis. If the outcomes do not confirm the hypothetical statement, then the hypothesis is revised until a set of results are predicted.

Close examination of both methods of inquiry reveals problems with each. Three major difficulties have been found with inductive methods. First, pure induction is always impossible for it does not explain the means by which facts are organized. It is not enough to say that findings are organized according to similarities, because it may always be asked why particular events (with possible similarities) were observed. Usually, some form of theoretical, conceptual, or logical analysis guides the investigation of similarities among a number of phenomena. Second, induction alone is an inefficient process. If conceptual or theoretical analyses do not guide an investigation, then "all possible irrelevant categories would have to be explored and evaluated" (Johnston and Pennypacker, 1979, p. 50). This is not a likely practice. What is more likely is the willy-nilly collection of data on this relation and that relation without an attempt to incorporate these findings into some categorical scheme. This, of course, results in reams of facts that are difficult to compare, are seldom compared and therefore leads to much unnecessary repetition.

The third problem with inductive methods of inquiry is the logical fallacy known as Affirming the Consequent; one decides that event A is related to event B when event B is observed in the presence of event A. This is a logical problem in that it does not conform to the rules of prediction established by logicians. The logic of induction is faulty because the investigator does not predict the occurrence of event B before it occurs. However, the relevance of this argument to conducting research on human behavior is questionable. This problem will be analyzed at length after examining the problems of pure deductive reasoning.

The two major difficulties encountered with deductive methods of reasoning are related to two assumptions: that deductive reasoning is useful only when a discipline has matured sufficiently such that its basic facts are reliable and when its facts can be formally systematized (Johnston and Pennypacker, 1979). The field of psychology may conform to the first assumption, but because it does not conform to the second, two problems arise. First, it has been stated that the rules by which hypotheses are deduced are usually formal and generate relatively specific predictions. These predictions lead to a restricted set of questions of the form: when X occurs, does Y occur? These restricted predictions are justified when enough is known about the subject matter to both measure the relation accurately and state the hypothesis precisely. In addition, these kinds of predictions turn out to be justified when Y does in fact occur. The problem arises when Y does not occur. When Y fails to occur under deductive rules, then the statement of X is revised until Y is obtained. This emphasis on revising the hypothesis not only interferes with noting what happened instead of Y, but is also conducted without knowing whether the problem lies in the precision of

the hypothesis or the accuracy of the measurement system. Since psychology is a young science and its measurement systems are often inaccurate, this approach can be very misleading, and inefficient. In addition, this emphasis "generates an advocacy style of inquiry wherein experimental results are marshalled in support of hypotheses" (Johnston and Pennypacker, 1979, p. 53). It is assumed that such "marshalling" inhibits the synthesis of related findings that happen to occur in the context of two or more antagonistic hypotheses.

The second problem of employing deduction in psychological investigations is that the relations between the formal statement of a hypothesis and the procedures and operations of the laboratory are often ambiguous. When the deductive method has been used in other sciences, such as physics, there has been a great deal of correspondence between the mathematical elements of the hypotheses and empirical elements of interest. In other words, matter and motion can be measured at the same mathematical level of sophistication as the formal statement of an hypothesis. Since this is an uncommon relation in psychological experiments, psychologists have been forced to use less precise hypothetical statements and metaphorical models. The question always remains, therefore, as to what the relation is between various models and the real phenomena of interest. This question is never answered at the level of experiments involving the models. The unequivocal answers are obtained when the methological technology is developed to the degree where there is unambiguous correspondence between measurement and the phenomena being measured.

A third problem with advocating a pure deductive method of inquiry, regardless of the maturity of a discipline, is that deductive methods

fail to account for the means by which logical statements (i.e. the hypotheses, theorems, etc.) are obtained. Certainly, a verbal classification scheme is not divorced from the experiences that the classifier has had. General rules are obtained from specific experiences. Thus, all deductive reasoning has implicit inductive components. This realization is, of course, similar to an earlier conclusion that all inductive reasoning has implicit deductive components.

The conclusion that neither pure deduction nor pure induction exists suggests that both methods are necessarily employed in the investigation of psychological phenomena. Most experimental psychological approaches are, in fact, controlled by both sets of rules (cf. Hayes, 1973). However, with the advent of sophisticated inferential statistical techniques which eliminate the necessity for precise experimental control in order to determine differences between manipulations and which provide some semblance of a formal relation between measurement and hypothesis stating, most psychological approaches favor deductive rules. Induction is referred to in passing from experiences to hypotheses. The behavior analytic approach does not favor deduction. The reasons that this approach has followed a more inductive course are many and complex. Some of these reasons may even be superstitious! However, the basically inductive method of inquiry employed by behavior analysts can be justified on the basis of the features of induction and deduction presented here.

The reader is referred back to the three problems of induction. It will be recalled that the first problem negated the notion of pure induction. All reasoning must be guided by some a priori verbal classification scheme. This deductive aspect of all inductive approaches is shared by behavior analytic approaches. There are certain materialistic,

philosophical assumptions and previously recorded psychological facts that control the organization of behavioral data. Behavior analysis is not a form of pure empiricism. Behavioral research is tempered by previous assumptions and by a large body of literature suggesting strong relations between environmental events and behavior. Therefore, behavioral findings are organized according to these basic relations.

The second problem, inefficiency, is also eliminated by incorporating deductive properties. Behavior analytic research is not an inefficient, purely empiricist undertaking. Behavioral research is guided by a conceptual foundation and data base that helps to integrate data and direct investigations. Behavioral research is deductive in this sense. The steady, systematic progress of behavior analytic research is one of its strongest features. This progress has been made because attention is paid to previously obtained data.

The third problem that the inductive scientist encounters is the logical fallacy of Affirming the Consequent. Regardless of its logical problems, it seems that affirming consequences is the job of science. All experimental methods involve manipulating antecedent conditions (independent variables) and observing the consequent effect on the phenomenon of interest (dependent variable), (Johnston and Pennypacker, 1979). The question that needs to be addressed is whether predicting the occurrence of the consequent before an experiment is conducted eliminates the fallacy? According to the rules of logic, it does. But, does this type of prediction produce a different effect than simply affirming the consequent? It is clear from previous discussion that even when experimental predictions are made, they are made in the context of the experimenters' previous experiences. These experiences usually are in the form of

reading related literature or conducting experiments with related variables. In the course of these experiences, the experimenter affirms a number of consequences. The data are examined and conclusions are drawn on the basis of consequences of other experimental behavior. Then a new prediction is made and another experiment is conducted. This process is repeated, continuously. The question becomes, what is the difference between affirming other experimental consequences and those which occur when one has actually carried out the experiment? The difference seems to be a matter of reference. It all depends on where one leaps into the experimental stream. The experimenter's behavior at the point at which previously collected data are examined and future results of research are predicted is similar to the inductive scientist's behavior. The inductive scientist affirms a consequence and then predicts what is likely to occur in the future under similar conditions. Both are affirming consequences and both are making predictions.

Certainly, making a prediction before any data have been collected seems to be a more interesting intellectual pursuit. There is something about the construction of a correct hypothesis that suggests mysterious intellectual abilities. There are those who claim that this display of knowledge, which appears to have been acquired through super-ordinary channels, is what controls so many deductive investigations (Skinner, 1950). This need not be the case. If all researchers accurately and completely referenced the antecedent sources of control, the mystery would not exist and the prestige would be better distributed. However, since the notion of scientist as wizard is as old as alchemy, it is unlikely that this will occur for quite some time.

The conclusion of this verbiage is that whether a prediction is made about a particular experiment or about future behavior in general, scientists are always affirming consequences. If this is true, then the question remains as to when affirming consequences is truly a fallacy. Johnston and Pennypacker (1979) suggest that the only time this fallacy has real significance is when a conclusion is made in terms of the necessary and sufficient conditions for the consequent to occur. It would be faulty to conclude that because B occurred in the presence of A, that A is the necessary and sufficient condition for B to occur. However, this conclusion would always be faulty on the basis of one experiment. The more rigorous a formal hypothesis is, and the more accurate the measurement system is, the closer an experiment can come to providing an explanation at the level of necessary and sufficient conditions. Yet, even in the most extreme example there is still a possibility of error and replications should be conducted. When the hypotheses and measurement systems are as inaccurate and informal as those used in most of psychology, then replication and affirmation of many consequences are always needed. The more consequences that are affirmed the greater the likelihood of accepting a relation between two or more phenomena.

It should be clear from the above analyses, that the traditional traps of inductive reasoning are avoided by behavioral research. Behavioral inquiry is a fusion of induction and deduction. The features of induction that behavioral research shares are the data collection processes and the question asking procedure. Behavioral research begins with a question about the relation between two or more phenomena. Then data are collected and direct descriptive statements about the data are made. If the data are highly variable further manipulations are made on the

procedures, equipment and/or the variables under investigation. If a possible relation emerges, it is measured repeatedly. When all conditions have been examined and alternative explanations eliminated, then the relation is stated as a principle. The features of deduction that behavioral research shares are the assumptions and general relations that control the scope and analyses of behavioral investigations. The inductive properties of behavioral research are adopted because of the problems specified for deductive question asking and problem solving procedures. The deductive properties of behavioral research are adopted because of the inherent necessity of some form of deduction.

The difference in the kind of deduction that behaviorists use in comparison to other psychologists is that the general beliefs that guide behavioral research are all based on observable data. As Skinner specified, behavioral theory is not a theory in the traditional sense. It is not an explanation of an observed fact which appeals to events taking place somewhere else, at some other level of observation, described in different terms, and measured, if at all, in different dimensions (Skinner, 1950, p. 193). In other words, behavioral research does not measure one event and discuss another event that has been, at best, indirectly measured. Behavioral research is controlled by an integration of observable facts (Keller, 1973).

The integration of observable facts that controls the experimental analysis of behavior includes the specific experimental methods for investigating phenomena as well as the general inductive strategy employed. An integral component of this model has been the use of single-subject methodologies. The exploratory nature of inductive research requires the control of subject variability and the increased data obtained by

single-subject methodologies. Michael (1974) suggested a number of reasons that support this assumption. First, prolonged and intensive interaction between subject and experimenter undertaken to control irrelevant sources of variation constitutes a rich source of data that may control further investigations of the subject matter. This source of ideas is critical to exploratory research. The use of statistical control, instead of experimental control, deprives the experimenter of these data and the experimenter becomes more dependent upon deductive theory and other experimenters' comments.

Second, the knowledge developed in order to control sources of variation is useful to other experimenters and practitioners. Balancing, randomizing and running hundreds of subjects to increase power and statistically control variability does not give others the necessary information for progressing from a statistically significant result to a practical application that is guaranteed effective. The practitioner needs to take into consideration control procedures as well as experimental procedures in order to make an educated decision as to whether or not to use a particular strategy.

Third, since single subject designs emphasize experimental control of irrelevant variables, prolonged study of a small number of subjects and relatively simple methodological procedures, it is possible to change the experiment while it is underway. If some previously unrecognized source of variability is apparently affecting the dependent variables, the experiment can be changed to control this source of variation or manipulate it in some way. Moment to moment changes can be made without interfering with the interpretation of the results. However, studies employing inferential statistics can not be changed so readily. Their

hypothetico-deductive assumptions, the complexity of the designs and results that can not be interpreted until all the data are collected and significance tests carried out prohibit the kinds of on-going changes that must be made to gain experimental control over irrelevant variables. Exploratory research needs to use experimental designs that can easily be changed to investigate new leads that surface through intensive interactions with the subject matter.

Baer (1977) and others (Kratochwill, 1978; Hersen and Barlow, 1976) have pointed out another practical advantage of single subject research. In research conducted to solve practical and social problems, it is necessary to see that the solution effects change in many individuals, not just an average individual. If one must test for statistical significance on data collapsed across many individuals, one does not necessarily have a solution. A statistical significance should never be assumed to be practically significant. If one accepts the assumption that practical significance is equated with substantial changes in individual behavior, then group statistical significance can never indicate practical significance. Fisher (1956) stated that while one can make inferences about a population of individuals from properly obtained significant effects, one can never make inferences about an individual from group data. However, the use of single-subject methodologies allows both sets of inferences to be made. First, the data are collected on individual behavior. Second, generality of results can always be inferred if a result has been replicated across individuals. Thus, if the results are true for each individual, then it can be concluded that the result will be true for any individual and that the result is true for groups of individuals. If the results are not replicated across individuals, then

a general principle has not been found. If the result occurs with some individuals, but not others, then it will be necessary to determine those individual characteristics that interact with the phenomena of interest to produce the effect. These variables can then be used by the practitioner to make an educated decision on the use of the procedure, relation or manipulation.

To summarize, the conceptual position that directs this research and subsequent research on cognitive behavior is the experimental analysis of behavior. This position is both a methodological perspective and a data based, theoretical perspective. The methodological components have been stressed in the above discussion because behavioral methods substantially depart from those used traditionally to investigate cognitive behavior. It is hoped that the use of behavioral methodologies has been sufficiently justified in general. For this particular problem area, the need for inductive, single-subject methodologies should be obvious from the conclusions drawn about previous research in this area. The fact that few practical solutions have evolved out of this research suggests that investigators need to take a giant step backwards and begin exploring different alternatives. This is exactly what is proposed in this thesis. The exact specification of this alternative will be developed in the next section. Thus, the lack of specific predictions that can be made about this approach to studying cognitive behavior indicate the need for inductive research. The increased, individual data that is needed suggests single-subject methodologies. Therefore, these are the approaches that are used in this research.

The Classification of Verbal Tasks

In each of the previous sections, it was concluded that an alternative approach to investigating cognitive behavior is needed. In the first section, the problems with previous research were discussed and it was found that the most prevalent criticism of previous approaches was the lack of precision in defining the cognitive tasks. Therefore, the purpose of this section is to develop a classification system for verbal or cognitive behavior that is precise and testable.

Johnson and Chase (1978, 1979) have designed a typology of verbal behavior that conforms to the rules for defining cognitive behavior derived in the first section. This functional typology was originally based on Skinner's (1957) analysis of verbal behavior. Skinner's functional classification system is a viable theoretical model for integrating language in general. However, strict adherence to this model may cause some practical problems. All of Skinner's terms (i.e. mand, intraverbal, tact, transcriptives) are neologisms. Although Skinner's justification for adopting these terms was well founded (i.e. to eliminate confusion arising from the mentalistic use of traditional labels for cognitive behavior) these terms are new to most practitioners and are not readily assimilated into everyday use. In addition, these terms are not always descriptive of what they represent. In fact, often their descriptive quality begets errors. For example, the mand was derived from such words as demand, command and mandatory. However, demands and commands are not always considered mands. They are often tacts or intraverbals. Such limitations, when coupled with the fact that Skinner's analysis has never been experimentally validated, warrant the changes that have been made here. Thus, the following definitions of cognitive or

verbal behavior are derived from Johnson and Chase (1978, 1979) and Skinner (1957), but the labels have been changed to avoid the confusion inherent in Skinner's terminology. As each class of behavior is introduced the descriptive label is followed by the class of verbal behavior from which it is derived.

The first class of behavior defined for this typology is the copy task (transcriptive behavior). The student reads a passage that defines, describes or exemplifies a phenomenon. Then, the teacher presents a task that either explicitly asks the student to copy, reproduce or reconstruct the passage (or some subset of the passage) or implicitly increases the likelihood of copying by including sequences of words taken directly from the passage. The student's response duplicates the passage or portions of the passage.

For example:

The student reads the following passage:

Abulia is a term used to describe low rates of behavior that are caused by an abrupt change from frequent to infrequent reinforcement. Sometimes the number of times that we must perform a behavior before reinforcement occurs is too large to keep us behaving. The result is a decrease in the frequency that we engage in the behavior. Freud called the resulting low frequency of behavior abulia.

Then, the teacher presents the following task:

Complete the following sentences by finding the sentences in the passage and copying the appropriate words in the blanks.

1. _____ is a term used __ _____ ____ _____

___ _____ _____ ___ _____ ____ __ _____

_____ from _____ to _____ _____.

2. Freud called ____ _____ ___ _____ __ _____ _____.

The student responds by filling in all the blanks from the text above.

Notice that the student and the teacher have used words and sequences of words that are identical to those used in the prose passage.

One nonexample of a copy task is presented below:

The student reads the same passage defining abulia. The teacher asks the student to fill in the blanks in the following task:

1. Abulia occurs when a _____ or dense _____ __ reinforcement _____ _____ to a more _____ or _____ schedule ___ _____.

Notice the words presented by the teacher in this fill in task are completely different from those used in the prose passage. Therefore, this task makes it very unlikely that the student can copy from the passage.

Another example of a copy task is exemplified below:

The student reads the same passage presented previously. The teacher asks: "What term would you use to describe low rates of behavior that are caused by an abrupt change from frequent to infrequent reinforcement?" The student answers: "Abulia."

This is a copy task because the teacher's presentation includes words or sequences of words taken directly from the passage. The student needs only to look through the passage until he comes across these words and copy the related term from the sentence.

A nonexample that is similar in structure, but different in function to the copy task just presented is:

After the student reads the passage defining abulia, the teacher asks: "What term would you use to describe the following situation?":

> Julia found that every time she kissed Allen on the neck he would respond by taking her into his arms and kiss her passionately. As a result she would often kiss Allen on the neck. One dark Thursday, Julia kissed Allen on the neck 6 times and he continued to read the paper until the sixth kiss. Then he turned and gave her a big kiss. The next day Julia kissed him 10 times before he responded. After this, Julia rarely kissed Allen on the neck.

The student responds: "Abulia."

In this case, the student could not merely look back at the passage and find the answer. He has to apply the words presented in the passage to determine if the situation exemplifies abulia or some other concept.

A second nonexample that is structurally similar to the copy task presented above is:

After reading the passage defining abulia, the student is asked: "What psychological concept refers to decrements in behavior that result from rapid decreases in the schedule or amount of reinforcement that occurs for that response?" The student responds: "Abulia."

Again, the student can not simply reread the prose passage and copy the term abulia. The student must be affected by the defining features of the term and affected similarly by the synonymous sequence of words used in the task.

To summarize, a copy task is characterized by either the necessity for or the possibility that a student will copy sentences or phrases directly from instructional materials. If a task sets the occasion for copying from instructional materials and a copy response could be a correct response, then the task is a copy task.

The second class of behavior that is studied in this project is the define task (intraverbal behavior). The students read or hear a prose passage that defines, describes and/or exemplifies a phenomenon. Then, the teacher presents one or more terms, definitions, rules or partial definitions concerning the passage. If a term is presented, the teacher requests a novel (not previously described) verbal response. Students define or describe the term in their own words. If a definition, rule or partial definition is presented, the teacher uses words or sequences of words that are not used in the passage. The student identifies the definition with a term.

A task category that is easy to confuse with define tasks is the copy task. Thus, a nonexample of a define task is:

After reading that glabrous skin is defined as skin devoid of hair, the student is presented with the following task: "Define glabrous skin." The student responds: "Skin devoid of hair." Notice that the student's response is not novel. He has repeated the exact sequence of words provided by the instruction. Notice, also, that the task did not ask for a novel response. This is an example of a copy task, not a define task.

This particular illustration could be altered to make it a define task by simply inserting the phrase "in your own words" at the end of the request. If the student responded: "Glabrous skin is skin that has no hair," then a define task has been completed.

Another example of a define task is:

After reading the definition of glabrous skin, the student is asked the following question: "What term is used for skin that is hairless?" The student responds: "Glabrous."

This is an example of a define task because it requires the synonymous effect of "devoid of hair" and "hairless." In other words, the task uses words that are different from those used in the definition.

Another example of a define task is:

After reading two passages, one defining glabrous skin and one that defines pubescent skin as that covered with short or downy hair, the student is asked: "Compare and contrast the terms pubescent and glabrous skin. Do not repeat the definitions given in the text." The student answers: "Both terms are used to describe different kinds of skin. Glabrous refers to skin or sections of skin that are hairless. Pubescent refers to skin that has small, soft, often hard to see hair."

Again, this is an example of a define task because it explicitly asks for sequences of words different from those provided by the teacher and the student has responded as directed.

Another task category that is easily mistaken for define tasks is illustrated below:

After hearing a prose passage that defines glabrous skin, the student is asked to say which of the following is an example of glabrous skin:

a. the palms of the hand

b. the forearm

c. the soles of the feet

d. lips

e. dorsal side of toes

The student identifies a., c., and d., as glabrous skin.

Notice that whereas the words used in this task are different from those used in the definition, they are not terms, definitions, rules or partial definitions of the phenomenon. Rather they are concrete instances of the phenomenon. Thus, this is a nonexample of a define task.

An example of a define task that is similar to the nonexample presented above is:

After hearing the definition of glabrous skin, the student is presented the following question:

Say which of the following defines glabrous skin:

a. skin that has soft, downy hair

b. skin that is completely hairless

c. skin that has a hair, no hair pattern

d. skin that is covered with course hair

The student identifies b. as the definition.

Again, the words used in the task are different from those presented in the passage (pg. 38) and the choices are all rules or general descriptions of skin type. They are not concrete instances of skin. Therefore, this is an example of a define task.

In sum, the define task is defined as the presentation of words, terms or definitions to which the student must respond in his own words. If the student's response is made in general terms or the statement of a general rule, and this response is novel, then the task is a define task.

The third class of verbal behavior categorized by this typology is the exemplify task (intraverbal behavior). The student reads or hears a prose passage that defines, describes and/or exemplifies a phenomenon. Then, the teacher asks the student to give an original

example of the phenomenon or some subset of the phenomenon. The student's response is a concrete narration of a novel (not previously described) instance of the phenomenon. The student's narration includes properties of the environment that are irrelevant to the definition of the phenomenon.

This task category is relatively easy to identify. In all cases, it requires that the teacher explicitly request some original description of a concrete instance of a general rule, prediction or definition. For example:

The student reads or listens to a passage that discusses hygrometers. A hygrometer is defined as a device for measuring moisture in the air. An example is given that describes a wet and dry bulb hygrometer. The end of one thermometer is wrapped in cloth, the end of the cloth is extended down into a bottle of water. This hygrometer measures the relative humidity by measuring evaporation. The more water that evaporates, the less moisture that is already in the air. The teacher asks the student to give an original example of a hygrometer. The student writes: "At home we have a carved, wooden Swiss Chalet that houses a boy with an umbrella and a girl in a bathing suit. Each figure is standing on opposite ends of a swivel post that is attached to a tautly stretched human hair. Whenever there is a lot of moisture in the air, the boy swings out of the Chalet. Whenever there is little moisture in the air, the girl swings out." As the prose passage does not contain such a description of a hygrometer, this is an example of an exemplify task. The task explicitly asks for an original example. The student's response is a description of one instance of a hygrometer. It is certainly an original description in relation to the instruction.

One nonexample of an exemplify task is presented below:

> The student reads a chapter that describes hygrometers and barometers as instruments for measuring weather change. The teacher asks the student to give an original description of the similarities and differences between hygrometers and barometers. The student answers: "Both barometers and hygrometers are devices for observing changes in climate. The barometer measures changes in air pressure and the hygrometer measures changes in the amount of moisture in the air."

In this case, the teacher has not explicitly asked for an original example of an original description of one instance of the terms hygrometer and barometer. The student has answered the request in terms of general descriptions. She has stated three rules that relate and distinguish between hygrometers and barometers. Thus, this is a define task.

Of course the teacher could change the task above into an exemplify task by substituting the word "example" for "description" or by adding the words "by juxtaposing instances of each" at the end of the request. These changes require that the student answer the request with a concrete narration of instances of both a hygrometer and a barometer.

At this point it seems necessary to specify one critical difference between define tasks and exemplify tasks. Often instructional tasks do not explicitly state the kind of behavior in which the student needs to engage. In these cases, it is necessary to observe the student's response in order to determine the task category. For instance, in the illustration of hygrometer and barometer above, the student could have answered

the original question by juxtaposing novel examples of both hygrometers and barometers. He could have related these two instances by making two examples similar except for the specification that one measured air pressure and the other, water content in the air. The student would have correctly answered the teacher's task, however, the answer may not have been what the teacher expected. This original question was ambiguous. It seems clear that in such cases the teacher should rewrite the task, making explicit what is expected from the student.

One last nonexample of exemplify tasks should be sufficient. In this illustration, the student reads the passage defining and exemplifying hygrometers. Then, the teacher asks for an example of a hygrometer. The student repeats the example given by the teacher by stating:

> Take two thermometers, wrap one end of one of the thermometers in cloth and let one end of the cloth extend into a jar of water. This will measure the amount of evaporation and thus, the amount of water in the air.

The problem with this task is that the teacher did not ask for an original example. Therefore, the student has answered the task correctly by repeating the example given in the passage. Again, if the teacher wants an exemplify response, the task must be explicit. In this case, the student's behavior constituted copy behavior.

In conclusion, the exemplify task is characterized by a request to give an original example of a concept. If such a request is answered by a concrete description of one or more instances of a general rule, term or definition, then the task is an exemplify task.

The fourth class of behavior of the functional typology is the example identification task (tact behavior). The student reads or hears a prose passage that defines, describes and/or exemplifies a phenomenon. Then, the teacher presents one or more concrete narrations of novel instances of the phenomenon and/or novel instances of other phenomenon and asks the student to identify those instances that illustrate the phenomenon. The student identifies those descriptions that conform to the general rule of definition of the phenomenon given in the passage.

For example:

A student reads the following passage:

> The constructional approach is a relatively new way by which we can change the problem behavior of an individual. Currently, most methods for dealing with problem behavior focus on eliminating or alleviating the distressing behavior. An alternative, the constructional approach, focuses on teaching new behaviors that are followed by desirable outcomes. This is accomplished by determining the desirable outcomes that maintain the problem or distressing behavior. Then, the constructional therapist helps construct alternative behaviors that are maintained by the same critical or desirable outcomes, but that are not distressing to the client.

Then, the teacher presents the following illustrations with the request: "Circle the letter corresponding to each of the following that is an example of the constructional approach.

a. Bob enjoyed jogging because of the exhilarating feeling and because it kept him in shape. However, when he jogged in the winter, he got a sore throat, cold feet, and the chills. Since he didn't like these effects, but liked staying in shape, Bob was faced with a dilemma. After hearing Bob complain about this predicament for the hundreth time, his friend Mara suggested that he join a health club and work out

on the Nautilus machine. Bob decided to give this a try. Bob no longer jogs every day; he works out on the Nautilus instead. This keeps him trim and he no longer complains about his winter colds.

b. Gerry was a music major participating in a research study conducted by one of the music theory graduate students. Every five seconds, a note was played and Gerry was required to indicate the name of the note. The first note was a "B," which Gerry correctly identified. He incorrectly identified the next note as an "F"; it was really another "B." The third note was an "F," and he got that right. The next two notes were "C's"; he identified each of them correctly. Finally, he was presented another "B," which he identified correctly.

The student identified a. as an example of the constructional approach.

Notice that each of these descriptions is concrete and can be considered an instance of some phenomenon. Illustration a. does conform to the definition of constructional approach given in the passage.

A nonexample of an example identification task is:

After reading the same passage defining constructional approach, the student is presented the above illustrations, but is asked to say why he identified the illustrations as examples and nonexamples of the constructional approach. The student answers: "Example a. does illustrate constructional approach because a new behavior is learned that is reinforced by the same events, exhilaration and trimness, as jogging, but does not have the same bad effects as jogging in the winter. Jogging has also decreased."

In this case, the student does have to identify which is an example of the phenomenon, but he also has to justify his answer in terms of a general rule. Thus, this task is a define task as well as an example identification task.

Another nonexample is:

The teacher asks the student to define constructional approach in his own words. The student responds: "The constructional approach is a model of therapy that attempts to construct new behavior rather than get rid of old, unwanted behavior. The therapist tries to determine what are the positive and the negative consequences of a problem behavior. Then he teaches the client another behavior that gets the positive consequences, but does not get the negative consequences."

Of course, this is an example of a straight define task. The student hasn't had to identify any concrete narrations. He has provided a definition in his own words.

We could also ask the student to give an original example of the constructional approach. This would not be an example of an example identification task, it would be an example of an exemplify task. Here, the student is providing the novel narration; in the example identification task the teacher is providing the novel narration.

One last point about example identification tasks. The narrations that the teacher presents need to be different from those that are presented in instruction. If an illustration is presented in a passage and then presented as a task, the student need only to look back at the passage, find the narration and copy the term that is related to it. This would be an example of a copy task. Therefore, a task is identified as

an example identification task only when the student is presented with a novel description of an instance of a concept and must identify it as such.

The final class of behavior that is of interest to this project, the combination task, is not a single class of behavior at all. Rather, it is a class of tasks that is composed of various combinations of the previously defined classes. The student reads or hears a prose passage that defines, describes and/or exemplifies a phenomenon. Then, the teacher asks the student to engage in two or more kinds of tasks with respect to the phenomenon. The student's response is any combination of copying, defining, exemplifying and/or example identifying that is called for with respect to the phenomenon.

For example, the student reads a passage that describes the procedure of effective imitation training. She is then presented with the following task:

> Ada was a little-league baseball coach. To help the children learn to field groundballs, she had them hit balls to her while she demonstrated the essential elements of fielding. These elements included getting her body in front of the ball, kneeling on one knee, keeping her eye on the ball, and so on. Then she hit some balls to the children and had them try to field the balls. When they fielded the ball correctly she praised them.
>
> If the above is an example of effective imitation training procedures, key the components of the illustration to the components of your definition of imitation procedures.

The student answers by writing that the illustration does conform to the definition of effective imitation procedures. Ada has modeled the appropriate response, has had the players attempt the response and has given the players feedback on whether or not they are correct.

This task clearly exemplifies a combination task. First, the student has

to say whether or not the narration describes an instance of effective imitation procedures. This is the example identification component of the task. Second, the student has to generate a definition of effective imitation procedures. This, of course, is the define component. The task is a define/example identification combination task.

This example could be changed to become an example of a single task category very easily. We could make it a define task by deleting the description of "Ada" and instead stating the general rule for effective imitation procedures. We could make it an example identification task by simply asking the student to say whether or not the "Ada" illustration exemplified effective imitation procedures.

Another example of a combination task is illustrated below:

The student studies the relations among the demographics of the people in a voting precinct, the kinds of political candidates that have been elected in different communities and the probabilities of new candidates being elected given certain perspectives on issues. Then, the student is given the following task:

> Below are biographical sketches of the residents in three precincts. Accompanying each sketch is a detailed description of successful and unsuccessful candidates in prior elections. Describe the candidate who you would predict would be the most likely to win an upcoming election. Say why you have chosen such a candidate.

Notice that if the sketches and descriptions had been provided then the student would have to engage in three classes of behavior in order to answer this question completely. The student has to identify political trends from concrete narrations, an example identification task; the student has to describe the characteristics of a candidate most likely to win, an exemplify task; and the student has to justify his descriptions in terms of the general rules of political relations discussed in the

text, a define task.

Again, a combination task is any task that asks the student to engage in two or more classes of behavior.

This typology of cognitive or verbal behavior is not an empirically derived classification system. It is based on a set of logical assumptions that are extensions of an experimental analysis of behavior. Skinner (1957) discussed the need to develop a functional or "causal" analysis of verbal behavior. He speculated that regardless of the differences between verbal behavior and other kinds of behavior (i.e. that verbal behavior does not have a direct effect on the physical environment, but rather its effect is mediated by another organism), that verbal behavior is controlled by the same kinds of general principles as other behavior. This assumption has been validated by a number of studies that have observed the effect of reinforcement (Skinner, 1961), stimulus control (Sidman, 1974), fading (Hively, 1962; Sidman and Stoddard, 1967; Stoddard and Sidman, 1967), and other kinds of instructional control over verbal behavior. However, the development of an instructional classification system based on the classes of verbal behavior developed by Skinner (1957) has not been investigated. Therefore, a number of studies are needed to determine the utility of the typology presented here. First, it is necessary to determine whether the system is a reliable and valid system for classifying cognitive or verbal tasks. Second, it is necessary to develop materials that are stable and consistent enough to eliminate task confounding variables in subsequent research on the typology. Third, it is necessary to determine the components of the typology that are necessary and sufficient for transfer of verbal learning to occur with a variety of different populations (e.g. college students, high

school students, industrial employees, institutional staff).

This thesis is concerned with the first two problems, to test the validity of the typology presented here, and to design the materials that will be used in subsequent single-subject experiments on the different components of the typology.

C H A P T E R I I

EXPERIMENT I

Purpose

The purpose of the first study was to test the validity of the functional typology presented in the introduction. Two questions were asked. First, can content experts reliably agree on the categorization of tasks into the classes specified by the typology? Second, does the classification system cover the range of verbal instructional tasks that it purports to cover?

Methods

Subjects

Eight graduate and two undergraduate students in the Psychology Department at the University of Massachusetts at Amherst served as subjects. The graduate students were selected from the fields of Educational, Cognitive, and Developmental Psychology. Two of the eight had not used and were not familiar with other cognitive classification systems. The six remaining graduate subjects had used at least the taxonomy of Bloom et al. (1956) to classify test items and objectives. The undergraduate subjects were research assistants working in the area of Educational Psychology. Both of these subjects had learned to classify and write objectives according to Bloom et al. (1956) previous to the study. None of the ten subjects had had previous contact with the typology used here.

Personnel

A graduate student in Educational Psychology coordinated the study. This experimenter had programmed a number of study and test materials for a range of psychology courses. An undergraduate psychology major served as a research assistant. The research assistant scored subjects' answers and assisted in designing the validation test.

Setting

The study was conducted in the Psychology Department at the University of Massachusetts at Amherst. All experimental tasks were completed in two similar sound insulated carrels. Each carrel was supplied with a desk, two chairs, storage space for research materials and one-way mirrors.

Materials and Apparatus

Each carrel was supplied with a complete set of experiment materials. These included a photo-copied, 16 page handout defining and exemplifying the categories of the typology (pages 33-49 of Introduction), a set of twenty tasks for each subject to classify (Appendix A) and an answer sheet (Appendix B). An electric timer was also provided in each carrel.

General Procedures

Williams (1977) stated two important goals for designing typologies of instructional tasks. First, the categories must be clearly defined so that two or more individuals can agree on the correct classification of instructional tasks. Second, the typology must provide sufficient categories to cover the range of cognitive abilities used by learners

in instructional tasks. The following empirical test of these goals was carried out on the typology developed here.

Twenty tasks were selected by a research assistant from a range of instructional materials. Ten of the tasks were taken from materials that the experimenter had written and ten tasks were taken from accessory materials for commercially available Educational and Introductory Psychology text books. The research assistant was instructed to select two of each of the task categories defined in the typology from those materials written by the experimenter. She was also asked to select as wide a range of tasks as possible from the commercially available materials. These tasks were examined by the experimenter and were classified according to the definitions provided in the handouts. A master sheet was prepared from the experimenter's classification of the tasks. Five were copy tasks. Three were define tasks. Five were example identification tasks, two were exemplify tasks and five were combination tasks.

Subjects were then given the handout that defined and exemplified each task to study. The subjects were instructed to read the handout carefully and to ask any questions before the twenty tasks were given to them to classify. Subjects were told that they would be able to use the handout while classifying the tasks and therefore they need not memorize the definitions. Subjects were timed from the moment they started reading the handout to when they indicated they were finished.

Finally, the subjects were given the twenty tasks to classify. Each subject was told to read the sample tasks carefully, to record the kind of task they thought each task exemplified and to comment on any tasks that they did not think could be categorized according to the

typology presented. Subjects were asked to think critically about the typology and comment on any objectives, test items or questions that could not be classified by this typology. In addition, it was emphasized that any task that did not call for a novel or original response and could therefore be answered by copying from a text or other instructional materials, should be categorized as a copy task. The subjects were timed from the moment they began reading the first task to when they indicated that they were finished.

As soon as each subject was finished, the experimenter scored the subject's answers by comparing them to the master answer sheet. Subjects were asked to clarify any disagreement and were given feedback on their agreement with the experimenter.

Reliability (Interscorer Agreement)

Interscorer agreement was assessed by a research assistant. The assistant rescored five (50%) randomly selected subject answer sheets. The research assistant compared each of the subject's answers with the experimenter's master answer sheet. An interscorer agreement index was calculated by dividing the number of agreements by the total number of answers and multiplying by 100. Interscorer agreement was 98%. In both cases of disagreement, a subject had changed the classification of a task orally to the experimenter, but had not written the change on the answer sheet. Therefore, the experimenter scored the classification differently from the reserach assistant.

Results

Agreement between experimenter and subjects on the classification of the twenty tasks was calculated in two ways. Both agreement indices

were obtained by dividing the number of agreements by the total number of tasks and multiplying by 100. First, percent agreement was calculated for each subject on all twenty tasks. Table 1 presents each subject's performance on each task, each subject's percent agreement index, the mean percent agreement for both graduate subjects and undergraduate subjects (subtotals) and the total mean percent agreement for all subjects. Pluses indicate agreement and minuses indicate disagreement with the experimenter's master classification sheet. Mean percent agreement for graduate subjects was 89.37. Mean percent agreement for undergraduate subjects was 82.50. Total mean percent agreement was 88.00.

Second, percent agreement was calculated separately for the ten tasks written by the experimenter and for the ten tasks obtained from Introductory and Educational Psychology study guides. Table 2 presents each subject's performance on each task written by the experimenter, each subject's percent agreement index, the subtotal means for graduate subjects and for undergraduate subjects and a total mean percent agreement for all subjects. Mean percent agreement for graduate subjects was 95.00; for undergraduate subjects, it was 95.00. Total mean percent agreement was 95.00. Table 3 presents each subject's performance on tasks obtained from Introductory and Educational Psychology study guides. The data are broken down in the same fashion as Tables 1 and 2. Mean percent agreement for graduate subjects was 84.00. Mean percent agreement for undergraduate subjects was 70.00. Total mean percent agreement was 81.00.

Duration data for each subject were also recorded. Table 4 presents the duration that each subject spent on both reading the handout defining

Table 1

Subject Classification Performance on All Twenty Tasks

Plus signs indicate agreement with experimenter
Minus signs indicate disagreement with experimenter

Task #	Graduate Subjects								Undergraduate Subjects	
	1	2	3	4	5	6	7	8	9	10
1	+	+	+	+	+	+	+	-	+	-
2	+	+	+	+	-	-	+	-	+	+
3	+	+	+	+	+	+	+	-	+	-
4	+	-	-	+	+	-	+	+	+	+
5	+	+	+	+	+	+	+	+	+	+
6	-	+	+	+	+	+	+	+	+	+
7	+	+	+	-	+	+	+	+	-	+
8	+	+	+	+	+	+	+	+	+	-
9	+	+	-	+	+	+	-	-	-	+
10	+	+	+	+	-	+	-	+	+	+
11	+	+	+	+	+	+	+	+	+	+
12	+	+	+	+	+	+	+	+	+	+
13	+	+	+	+	+	+	+	+	+	+
14	+	+	+	+	+	+	+	+	+	-
15	+	+	+	+	+	+	+	+	+	+
16	+	+	+	+	+	+	+	+	+	+
17	+	+	+	+	+	+	+	+	+	+
18	-	+	+	+	+	+	+	+	+	-
19	+	+	+	-	+	+	+	+	+	+
20	+	+	+	+	+	+	+	+	+	+
Individual Totals	90%	95%	90%	90%	90%	90%	90%	80%	90%	75%
Sub-Totals	89.37%								82.50%	
Total	88%									

Table 2

Subject Classification Performance on Tasks
Written by Experimenter

Task #	Graduate Subjects								Undergraduate Subjects	
	1	2	3	4	5	6	7	8	9	10
4	+	-	-	+	+	-	+	+	+	+
5	+	+	+	+	+	+	+	+	+	+
6	-	+	+	+	+	+	+	+	+	+
11	+	+	+	+	+	+	+	+	+	+
12	+	+	+	+	+	+	+	+	+	+
13	+	+	+	+	+	+	+	+	+	+
14	+	+	+	+	+	+	+	+	+	-
15	+	+	+	+	+	+	+	+	+	+
16	+	+	+	+	+	+	+	+	+	+
20	+	+	+	+	+	+	+	+	+	+
Individual Totals	90%	90%	90%	100%	100%	90%	100%	100%	100%	90%
Sub-Totals	95.00%								95%	
Total	95%									

Table 3

Subject Classification Performance on Tasks
From Commercial Materials

Task #	Graduate Subjects								Undergraduate Subjects	
	1	2	3	4	5	6	7	8	9	10
1	+	+	+	+	+	+	+	-	+	-
2	+	+	+	+	-	-	+	-	+	+
3	+	+	+	+	+	+	+	-	+	-
7	+	+	+	-	+	+	+	+	-	+
8	+	+	+	+	+	+	+	+	+	-
9	+	+	-	+	+	+	-	-	-	+
10	+	+	+	+	-	+	-	+	+	+
17	+	+	+	+	+	+	+	+	+	+
18	-	+	+	+	+	+	+	+	+	-
19	+	+	+	-	+	+	+	+	+	+
Individual Totals	90%	100%	90%	80%	80%	90%	80%	60%	80%	60%
Sub-Totals	84%								70%	
Total	81%									

Table 4

Number of Minutes to Completion of
Handout and Classification Test
For Each Subject

Subject	Handout	Classification Test
1	23	-
2	18	25
3	26	-
4	27	26
5	21	20
6	15	24
7	16	43
8	25	26
9	25	32
10	23	23
Mean Duration	21.9	27.37

and exemplifying the task categories, and on classifying all twenty tasks. The dotted line separates the graduate subjects from the undergraduate subjects. Mean duration for reading the handout was 21.9 minutes. Mean duration for completing the classification test was 27.37 minutes.

Discussion

The results of the first study were fairly unambiguous. First, all agreement analyses revealed substantial differences between graduate and undergraduate subjects. This result was expected, as it indicated that population differences may affect performance on a typology classification test. These differences suggest that comparisons across validation studies may be inappropriate unless more detailed descriptions of the subjects are given. In this case, both undergraduate subjects were considered content experts and experts in the use of test items and objectives. If these criteria and only these criteria define the population, then different results could be expected with a sample consisting of less graduate students. The graduate students sampled probably comprise a subset of the total population of potential users of the typology. Therefore, the results found here can not be generalized to this population as a whole.

At this point in the development of the typology, the conclusion stated above is of little concern. The purpose of this study was to determine whether any subjects could reliably agree with the experimenter on the classification of tasks. This objective seems to have been met. First, eight of the ten subjects classified at least 90% of the twenty tasks the same as the experimenter. Second, all of the subjects

were in agreement on 90% or more of those tasks written by the experimenter. Third, even on the more stringent test of tasks taken from other sources, four of the subjects agreed at least 90% of the time and four subjects were in agreement 80% of the time.

The lower agreement indices obtained for those tasks taken from other sources may be attributed to one of two reasons. It is possible that the definitions do not really cover the questions, objectives or tasks that others might write. This conclusion is somewhat disturbing in that it limits the use of the typology and questions its validity. However, the second possible conclusion is that this result is the product of ambiguous questions. Three tasks account for 55% of the disagreements in this analysis (Tasks 2, 9 and 10). Upon further investigation of these tasks, ambiguities were found. In each case it is not clear whether the information given in the task is the same or different from the information given in the text. For example, in task 9, the description of the passage does not clearly specify that one of the examples in the passage was identical to the example given in the task. Therefore, some of the subjects classified the task as a copy task and others classified it as an example identification task. If the original passage had been available to the subjects, instead of a summary of the passage, then perhaps fewer subjects would have identified the task differently. In sum, it appears that even on the most stringent test of interrater agreement, the results can be interpreted as a high level of agreement on the classification of objectives and test items according to this typology.

The second question asked in this experiment (i.e. does the typology cover the range of tasks purported) is more difficult to analyze.

No subject was able to specify a task type that was not covered by the five kinds of tasks defined by the typology. However, three subjects suggested that further divisions could be made within the define class of tasks. For example, one subject suggested that define tasks should be further analyzed in terms of difficulty. Define tasks that ask for an original description of the relation between concepts, that is not explicitly stated in a passage, seemed to this subject to be different from define tasks that asked students to define a simple concept in their own words. Certainly, these two kinds of tasks are different. However, it may be argued that these differences are relative to what is presented in the instructional passage. An instructor, faced with the objective that students relate two concepts that are not presented as a relation in the text, might construct the relation for the students and then ask the students to state the relation in different words. Whether this kind of programming leads to results different from a program in which the relation is never specified for the student is an empirical question. Therefore, further research needs to be conducted to determine these differences.

To summarize, although there is no quantifiable evidence that the typology covers all the types of tasks that it purports to cover, subjects were not able to specify questions, tasks or objectives that were not part of the typology. Therefore, it can be concluded that the second objective of the experiment was met.

Time data were collected as a further test of the practicality of the typology. Williams (1977) suggested that study time on the definitions of tasks of a typology should not be excessive. Reported times of typologies tested by Williams (1977) ranged from 20 to 60 minutes. The time

data collected here indicate that the definitions of the tasks for this typology do not take any longer to study. Therefore, this analysis of the typology suggests that it fits within established guidelines for practicality.

C H A P T E R I I I

EXPERIMENT II

Purpose

The second study was conducted for two reasons. First, to select three concepts that were similar in difficulty and that were sufficiently esoteric. Similarity was necessary because repeated measures, single-subject experimental designs will be used in all subsequent experiments. If the concepts are similar, then the confound of changing concepts concurrent with changing conditions will be minimized. Esoteric concepts were needed in order to assure that subjects would not pretest out of subsequent experiments. Second, the study was conducted to obtain feedback on the clarity with which the prose passages that defined each concept were written.

Method

Subjects

Seventeen staff members of a State School for the Mentally Retarded served as subjects. All subjects selected had bachelors degrees in psychology or a related field. Their job classifications ranged from Mental Health Assistants to Psychologists. At the time of the experiment, all subjects were being trained to apply behavior analysis procedures.

Personnel

A graduate student in Educational Psychology conducted the experiment. Assistance was provided by a fellow Educational Psychologist serving as a Chief Psychologist at the state school.

Setting

The study was conducted in a classroom in one of the buildings on the school grounds. The classroom was supplied with five long tables, each with four chairs. All experimental tasks were completed at these tables.

Materials

Nine esoteric psychological concepts were selected to be tested. They were: intraverbal, correlative subsumption, conditional suppression, abulia, constructional approach, D prime, retroactive inhibition, canalization, and tau effect. These concepts were selected from a range of advanced textbooks in the areas of Experimental Psychology, Sensation and Perception, Concept Formation, Theories of Personality, and the Experimental Analysis of Behavior. A set of tasks was written for each concept. Each set consisted of a prose passage that defined the concept, a copy task and three exemplify tasks. These tasks were written by both the experimenter and the assisting Educational Psychologist, each of whom edited the work of the other. All tasks were typed on 8-1/2 x 11 ditto masters and copied for each subject. A master data sheet was designed and implemented by the experimenter (Appendix C).

General Procedures

The experiment took place during two, one hour sessions. During each session subjects were asked to sit at a table and were given scrap paper and pens. At the start of the first session, the study was described to the subjects as follows:

> The Psychology Department training staff are devoted to providing you with the best training possible. One aspect of this training is the individualized, written materials that we

> will provide for you. We would like to see that all of you learn as many different ways of talking about psychology as can be efficiently arranged. Unfortunately, although we know how to teach this verbal curriculum effectively, we do not know the most efficient process. Therefore, we are conducting a number of studies to try to evaluate our materials so that you can be trained in as little time as possible. What we would like you to do today is study text passages defining psychological concepts. After you have studied these passages, we will ask you to complete a number of questions related to them. The first type of question will ask you to copy words directly from the passage. You must use the exact words that are given in the passage in order for these to be marked correct. The second type of question will ask you to give an original example of the concept. For these questions, think of how the concept applies to your own life and write a story that illustrates or exemplifies the concept. You can use examples from work, from home or what ever other experiences you can relate the concept to, however each example must be different from any others that you have written. Are there any questions?

After answering all questions, the experimenters distributed one passage to each subject and asked them to begin. Since order of concept presentation was randomized, subjects began the experiment with the first concept of an individual sequence of concepts.

Subjects were instructed to read the passage and complete the accompanying copy task. The copy task served as an "observing response," assuring that the subjects had read the passage. Subjects were asked to raise their hands when they had completed the copy tasks. Then, one of the experimenters corrected the copy tasks and any errors were corrected by the subject. Once the copy task was completed correctly, the experimenter asked the subject to attempt each of the three exemplify tasks. Subjects were instructed to raise their hands again when three original examples were written or if they thought it was impossible to finish these tasks. Upon finishing the exemplify tasks, subjects were given the next prose passage. This sequence was repeated until each subject had completed the tasks for six concepts. The concepts tested were

intraverbal, correlative subsumption, conditioned suppression, abulia, tau effect and constructional approach. At the end of the session subjects were asked individually to comment on the materials. They were prompted to point out any words, phrases or sentences that were confusing or created difficulty in understanding the definitions. Comments were recorded by the experimenters.

After the first session was completed, it was decided that four of the concepts needed to be tested with more subjects. In addition, the three remaining concepts needed to be tested. Therefore, a second session was conducted with eight additional subjects. Seven concepts were tested: abulia, tau effect, constructional approach, conditioned suppression, D prime, canalization, and retroactive inhibition. All procedures were exactly the same as those used in the first session.

All exemplify tasks were corrected by the experimenter outside of the experimental sessions. A priori rules were established for correcting the exemplify tasks. Table 5 presents the rules for accepting and rejecting concepts, examples and the originality of examples. The exact decision rules for concepts were arbitrary. However, they are conservative (i.e. the concepts must be difficult in order to be accepted) and they do assure that the definitions communicate what they purport to communicate.

Reliability (Interscorer Agreement)

The second experimenter rescored 20% of all tasks completed by the subjects. The interscorer agreement index was 99% as the experimenters disagreed only once on a total of fifteen tasks.

Table 5

Decision Rules for Accepting Concepts
and Subject Answers to Exemplify Tasks

	Accept	Reject
Concepts	3 original examples written across S	< 3 original examples written by any S
Examples	1. original and 2. includes all critical features of concept	1. not original and/or 2. does not include all critical features of concept
Originality of Examples	1. behavior not used in previous examples and 2. setting not used in previous examples	1. behavior used in previous example and/or 2. setting used in previous example

Results

The results were analyzed after each of the two sessions. Table 6a illustrates the number of original examples written by each of the nine subjects for each concept in the first session. Table 6a demonstrates that two of these concepts, intraverbal and correlative subsumption, failed to meet the criterion for difficulty. The concept intraverbal was exemplified three times by two subjects. The concept correlative subsumption was exemplified correctly three times by one subject. Therefore, both of these concepts were eliminated from the sample of concepts.

Table 6b illustrates the number of correct original examples written by each of the eight subjects for each concept in the second session. The concept canalization was eliminated from the sample because it failed to meet the criterion for difficulty. The concepts D prime and retroactive inhibition were eliminated from the sample because they failed to meet the criterion of acceptability. Three concepts, abulia, tau effect, and constructional approach were chosen from the remaining five concepts because of similarity in the number of correct examples written for each during both sessions. Five subjects wrote one correct original example for both abulia and constructional approach. Five subjects wrote at least one correct original example of tau effect. Even though no single subject wrote three original examples of conditioned suppression, it was rejected. The total of 18 correct original examples that were written across the two sessions indicated a substantial difference between this concept and the three concepts that were accepted.

Tables 6a, 6b

Number of Correct Original Examples
Written for Each Concept

Table 6a

Subjects

Concepts	1	2	3	4	5	6	7	8	9	Total
intraverbal	3	0	0	3	0	0	0	0	0	6
correlative subsumption	1	0	0	3	0	0	0	0	0	4
conditional suppression	2	1	1	2	1	0	1	1	0	9
abulia	1	0	0	1	0	0	0	1	0	3
construc-tional approach	1	0	0	0	0	0	0	0	0	1
tau effect	1	0	1	0	0	0	0	1	0	3

Table 6b

Subjects

Concepts	1	2	3	4	5	6	7	8	Total
conditional suppression	1	2	0	0	1	1	2	2	9
abulia	0	1	0	0	0	0	1	0	2
construc-tional approach	1	1	0	0	0	0	1	1	4
tau effect	0	2	0	0	0	0	0	2	4
canaliza-tion	2	1	1	0	0	0	3	1	8
retro-active inhibition	0	1	0	0	0	0	0	0	1
D prime	0	0	0	0	0	0	0	0	0

Discussion

The three concepts selected all met the prespecified criteria for selection. In addition, a similar number of correct original examples were written for each of these concepts. Therefore, they appear to be similar in difficulty. However, this conclusion must be examined in light of the tasks used to test the similarity between concepts.

The exemplify tasks were selected for this study because previous experience with these tasks suggested that they were challenging to answer. Therefore, it was decided that the exemplify tasks would be sensitive to the differences between concepts. However, this is purely an assumption. It is not known how sensitive to differences exemplify tasks are. It is also not known whether performance on exemplify tasks is related to performance on other kinds of tasks.

Because the relation between exemplify tasks and other kinds of tasks is not known, it can not be concluded that the concepts selected are similar under all conditions. The definition of a concept includes the behavior or verbal practices of the community as well as the stimuli which set the occasion for the behavior (Keller and Schoenfeld, 1950; Skinner, 1957; Rosch, 1978). If a range of behavior can occur in the presence of a stimulus configuration, the concept must be defined in terms of all these classes of behavior. Of course, these definitions get unwieldy. Therefore, it is often necessary to divide the definition of a concept into separate definitions of conceptual behavior. What has been measured in this study is one class of conceptual behavior, exemplifying. Thus, one can conclude only that the conceptual behavior tested resulted in similarities between the three accepted concepts. In order to claim that the concepts are similar, other classes of behavior

must be measured.

The second objective of this experiment had to be analyzed qualitatively. Therefore, conclusions drawn about the clarity of the prose passages for the three selected concepts were subjective and speculative. One of the prespecified criteria, that at least three original examples be written across subjects, did assure that at least some of the subjects were able to understand the prose passage. Each subject was also asked to comment on the materials before they left the session. A number of comments were made and resulted in subsequent editing of the prose passages. However, the real test of the clarity of these passages will have to wait for further research in which subjects are exposed to the materials individually. During such exposure each subject's responses can be closely monitored and feedback on the materials can be gained.

C H A P T E R I V

EXPERIMENT III

Purpose

The third study was conducted for three reasons. First, to further test the similarity between the three prose passages that define the concepts constructional approach, abulia and tau effect (see Appendix F). Experiment II demonstrated that performances on exemplify tasks were similar across these three concepts. However, it was concluded that other kinds of tasks needed to be tested to determine whether the concepts were similar across all conditions. In this study it was decided to test example identification tasks.

The use of example identification tasks as a test for concept similarity also permitted testing the difficulty of the vignettes or illustrations that had been written for each concept. Since these examples and nonexamples varied in length and varied the kinds of situations in which the concepts were illustrated, it was assumed that some illustrations were more difficult than others. In addition, because these illustrations would be used for both study tasks and testing tasks in subsequent experiments, and because they would be used in both example identification tasks and combination tasks, it appeared necessary to determine the level of difficulty for each illustration. The easier tasks would then be used during initial learning trials and more difficult tasks presented later in the learning sequence. The most difficult tasks would be administered during the testing sequences. Thus, the second reason for conducting this experiment was to assess a rank order

of difficulty for all examples and nonexamples written for each concept.

Finally, the third reason for conducting Experiment III was to determine the relation between reading rate and accuracy as measures of subject performance. A number of writers have indicated that rate of task completion is positively correlated with both accuracy and retention performance (Johnston and Pennypacker, 1971; Houghton, 1979; Lindsley, 1979). In addition, Kintsch (1974) indicated that reading rates are a sensitive measure of text complexity. These studies suggested that reading rates need to be investigated.

Methods

Subjects

Twelve undergraduates at the University of Massachusetts at Amherst served as subjects. Subjects were recruited from two courses; Educational Psychology and a special Introductory Psychology course for transfer students. All subjects were upperclassmen majoring in either Psychology or Education.

Personnel

The graduate student who conducted Experiments I and II coordinated this study as well. A fellow instructional designer with much experience in writing study materials and evaluation instruments assisted in writing materials, editing materials and running subjects.

The research assistants were three advanced undergraduates from the Psychology Department at the University of Massachusetts at Amherst. Their primary task was to conduct the experimental sessions. This involved presenting all experimental tasks to the subjects, correcting the subjects' written responses and delivering feedback to subjects

during the experiment. In addition, one of these assistants transcribed audio-tapes of all sessions, checked the reliability and accuracy of the data collection process, and helped compute and analyze the data. A second assistant helped write examples and nonexamples of each concept.

All research assistants were trained by the experimenter before the study was implemented. Experimental training consisted of: 1) a detailed written description of correct procedures for each session; 2) modeling; 3) role-playing; and 4) feedback. All assistants met the criterion of no more than one procedural mistake during the role-playing session before they were permitted to participate in their assigned roles. All subjects were trained for content expertise as well. Training consisted of: 1) studying the prose passages for each concept; 2) answering all tasks; 3) feedback on performance; 4) answering all tasks that were incorrect; and 5) final feedback. All assistants met the criterion of 95% correct performance on these tasks before being permitted to correct subject responses.

Setting

The study was conducted at the Mastery Learning Center, located on the seventh floor of the University of Massachusetts Library. All experimental tasks were completed in four, similar sound insulated carrels. Each carrel consisted of a desk, two chairs and shelves for research materials.

Materials and Apparatus

The experimental materials consisted of three prose passages that defined the three esoteric psychological concepts (*abulia*, *constructional approach*, and *tau effect*) and a domain (Hively, 1970) of copy

tasks and example identification tasks for each concept. Tasks were written by the experimenter, the assisting instructional designer and one of the research assistants. Once a week for four weeks these three met to discuss and edit the illustrations that had been written. All examples and nonexamples were completed before the study began.

Other materials included a study behavior questionnaire (Johnston, O'Neill, Walters & Rasheed, 1975) (Appendix D), a pretest (Appendix E) and a comment and scoring sheet (Appendix C). All materials were typed on 8-1/2 x 11 ditto masters and copies were made for each subject.

Cassette tape recorders were used to record all interactions between research assistants and subjects. At the beginning of each session, the research assistants recorded their names, the subject's name, the date, the session number, and the concept being used.

An electric timer was also provided for each carrel. Duration on each task was recorded by the research assistants. In addition, a watch with a second hand was used by each subject to self-record duration of study behavior during the first session.

General Procedures

This study was conducted during four, one hour sessions. The first session was devoted to assessing the subjects' typical study behavior and to determine whether they had any previous experience with the three concepts. Thus, the study behavior questionnaire and the pretest were administered during the first session.

After assuring that all materials were ready, that the timer was set, and that the tape-recorder was functioning, the research assistant called each subject into the carrel. After a brief informal discussion

the research assistant described the study in general and the first session in detail.

The first task presented to each subject was the study behavior questionnaire. The assistant explained that the purpose of the questionnaire was to find out how the subjects typically studied for a quiz or test. Then, the questionnaire was given to the subjects to read. Then, the experimenter went over each item and asked the subjects if they had any questions. Finally, a 900 word passage (Robinson, 1976, pp. 351-355) was given to the subjects to study. The subjects were asked to study as they "typically" or "normally" did for a quiz and to fill out the questionnaire accordingly. Since most of the items on the questionnaire require an estimate of time spent on each kind of study behavior, subjects were given a watch so that they could record time on the questionnaire. Subjects were asked to indicate when they had finished studying and were ready to take a test on the 900 word passage. The total study duration was also timed by the assistant.

As soon as the subjects indicated that they were finished, the timer was turned off and the duration of total study behavior was recorded on the scoring sheet. Then each item was checked and the subjects were questioned to assure that they had filled out every relevant item. At this time, there was a 5 minute break. After the break, the assistant described the rest of the session to the subjects. The subjects were told that they were to complete a test. The test consisted of pretest questions on the three esoteric concepts and questions related to the passage the subjects had just studied. The assistant briefly defined the different kinds of questions that were on the pretest. The first question type asked the subject to define each concept

in their own words. The second type asked them to identify examples of each concept. The third type asked them to give their own novel examples of each concept and the fourth type asked them to solve problems that require defining, identifying, and exemplifying each concept within the same task. Subjects were asked to read each question carefully, to answer each question as completely as possible and to write "D.K." or "Don't Know" next to each question that they could not answer. Each subject was encouraged to guess on any question except the example identification tasks.

If the subject had no questions, the assistant presented the pretest, started the timer and asked the subject to indicate when they were finished.

As soon as the subjects indicated that they had finished, the timer was turned off, and the elapsed time was recorded on the scoring sheet. Then, the assistant checked to make sure that subjects had answered each question at least with a D.K. If not, subjects were asked to do so and the timer was turned back on until every question had been answered. When the subjects had answered each question, the assistant thanked the subjects and made an appointment for the next session.

The second session began with a detailed description of the rest of the experiment. The subjects were told that each of the remaining sessions would be devoted to learning one concept. They would study a prose passage defining that concept, complete copy tasks that referred back to the passage and finally they would be asked to discriminate examples from nonexamples of the concept without referring to the passage. After all questions were answered, the experiment began.

Each subject had been assigned to one of four groups defined by

concept presentation order (Table 7). Thus, the assistant presented the subjects with the prose passage defining the first concept for their group. As soon as the passage was placed on the table in front of the subject, the assistant started the timer. When the subjects indicated that they had finished studying the prose passage, the timer was stopped and the elapsed time was recorded.

After the duration of studying the prose passage was recorded, the copy tasks were presented. Again, as soon as the task was placed on the table in front of the subject, the timer was started. When the subjects finished the copy tasks, the timer was stopped and the assistant recorded the total time spent on copy tasks. Since the copy tasks were used as an "observing response" to assure that the subjects had read the prose passage, the assistant immediately corrected the copy tasks and asked that any incorrectly copied answer be corrected.

Upon completing the copy tasks, the subjects were presented with a randomly ordered sequence of example identification tasks. Each example identification task was presented separately. The subjects wrote "yes" for those illustrations that exemplified the concept and "no" for those illustrations that did not exemplify the concept. The assistant recorded the time that elapsed between presenting the illustration to the subjects and a finished written answer. Finally, the subjects were asked to justify their answers to 50% of the tasks. Justifications were recorded on the tape-recorder. This sequence of procedures was followed for sessions two through four for all subjects.

Reliability

In order to determine if the procedures were reliably implemented,

Table 7

Order of Conditions for Each
Subject by Session

	Session I	Session II	Session III	Session IV
S1	study questionnaire pretest	abulia	con. app.	tau effect
S2	study questionnaire pretest	abulia	con. app.	tau effect
S3	study questionnaire pretest	abulia	con. app.	tau effect
S4	study questionnaire pretest	tau effect	abulia	con. app.
S5	study questionnaire pretest	tau effect	abulia	con. app.
S6	study questionnaire pretest	tau effect	abulia	con. app.
S7	study questionnaire pretest	con. app.	abulia	tau effect
S8	study questionnaire pretest	con. app.	abulia	tau effect
S9	study questionnaire pretest	con. app.	abulia	tau effect
S10	study questionnaire pretest	abulia	tau effect	con. app.
S11	study questionnaire pretest	abulia	tau effect	con. app.
S12	study questionnaire pretest	abulia	tau effect	con. app.

ten of the 48 hours of taped interactions between experimenters and subjects were transcribed. Although no quantitative measure was taken, only six procedural or instruction mistakes were found by the experimenter. Each of these was minor and corrected by the assistant.

Interscorer agreement indices were calculated for each of the dependent measures. For accuracy scoring, a research assistant rescored ten tasks for each subject for each concept (approximately 20% of all tasks). The scores for these 320 tasks were compared to the original scores given by the experimenter. Agreement was calculated by dividing the number of agreements by 320 and multiplying by 100. The interscorer agreement index for accuracy was 98%.

Interscorer agreement on duration measures was determined in a slightly more complicated and less exact fashion. Since the tape-recorders used in the study were found to be slightly slower than real time and because the length of time that it takes to depress the stop switch varies, exact agreement was impossible. However, an estimate of agreement was obtained in the following fashion. First, depressing the start and stop switches of the timers made a distinct click that could be heard on the tape recorder. Therefore, the assistant transcribing the tapes recorded the length of time between timer clicks. The duration of 200 tasks were rescored (11% of all durations timed). Second, these times were compared to the times recorded by the original scorer. Third, scorers were considered in "agreement" if the original duration was within ± 2 seconds of the rescored duration. An agreement index was calculated by dividing the number of agreements by the total number of intervals scored and multiplying by 100. The interscorer agreement index for durations was 87.5%.

Results

The data for experiment III were analyzed in several ways in order to answer the three questions. First, both dependent measures, accuracy and reading rate were analyzed separately. Both group and single-subject comparisons were made for each. Then, the relations between accuracy and reading rate were analyzed. Study-behavior data and justification data were not analyzed. Justifications for these decisions appear in the discussion.

Accuracy Data

In order to isolate any differences that might exist between concepts, a concept by order of presentation, repeated measures Analysis of Variance was conducted. Table 8 presents the accuracy data for the ANOVA. These data are the arcsine transformations of the raw proportion of correct responses. The transformations were conducted because the variances obtained from proportions are always systematically related to the mean in the form: $\sigma^2 = \mu(1-\mu)$ (Myers, 1979). Therefore, the arcsine transformations are used to stabilize the variances sufficiently to assume homogeneity of variance. The letters C, A, and T at the top of the table represent the three concepts, constructional approach, abulia, and tau effect, respectively. The marginal totals are augmented by the marginal means in parentheses. The group means are presented below each group number on the ordinate of the table. One df was lost for each F test because the data for the cell $X_{1,2,1}$ was estimated. Table 9 shows that the main effect of concept was significant, $F(2,15)$ of 6.93 $p<05$ and the effect of the concept x order of presentation interaction was also significant, $F(6,15)$ of 3.14 $p<05$. The main effects of order of

Table 8

Matrix of Arcsine Transformations of Proportions

		C	A	T		
Group 1 (57.82)	1	47.01	70.54	63.15	180.7	(60.23)
	2	51.06	60.94	56.79	168.79	(56.26)
	3	61.14	57.86	51.94	170.94	(56.98)
Group 2 (59.59)	1	49.43*	63.44	60.67	173.54	(57.85)
	2	53.79	64.75	64.90	183.44	(61.15)
	3	49.66	64.75	64.90	179.31	(59.77)
Group 3 (57.11)	1	53.79	46.55	50.30	150.64	(50.21)
	2	61.14	59.67	59.34	180.15	(60.05)
	3	61.14	59.67	62.87	183.68	(61.22)
Group 4 (58.10)	1	53.79	56.23	56.79	166.81	(55.60)
	2	55.18	59.67	58.05	172.9	(57.63)
	3	59.60	61.62	62.03	183.25	(61.08)
		656.73	725.69	711.73	2094.15	
		(54.73)	(60.47)	(59.31)		

*X_{ijk} was estimated by the following: $\frac{\hat{X}_{ij} = nT_{i} + aT_{j} - T_{..}}{(n-1)(a-1)}$ (Myers, 1979)

Table 9

Source Table for ANOVA of Accuracy

Source	SS	df	MS	F	
A (concepts)	221.53	2	110.76	6.93	(2,15)*
B (order)	28.3	3	9.43	.244	(3,7)
C (subjects)	308.83	8	38.60		
AB	301.43	6	50.24	3.14	(6,15)*
AC	255.94	16	15.99		

* $p < .05$

presentation was not significant. Two further group tests were conducted to isolate the relations between the differences between concepts. First, the proportion of variance accounted for by the main effect of concept was estimated by the formula:

$$\text{est. } w_x^2 = \frac{(MS_x + MS_e)df_x}{SS\ total + MS_e}.$$

This analysis revealed that est. $w_A^2 = .16$ of the variance. Thus only 16% of the total variance can be attributed to the effect of concepts. The second test conducted was Scheffé's (1959) Post Hoc Test of Comparisons, for the null hypotheses that the difference between each of the concept means was zero. Table 10 shows that the difference between the mean obtained for constructional approach (C) and the means obtained for the other concepts was significant at the .05 level. The difference between the mean for abulia and the mean for tau effect was not significant.

Single-subject data for the effect of concepts were also analyzed. Between-subject and within-subject analyses, however, reveal considerable individual differences. Figure 1 presents the distribution of correct and incorrect responses made by four representative subjects. Sample subjects were randomly selected from each of the four groups. A within-subject analysis revealed considerable individual variability across concepts. The pattern of responses made by each subject demonstrated substantial differences between constructional approach and the other concepts. However, the patterns found for both abulia and tau effect are similar for three of the four subjects presented. Comparisons made between subjects on each concept indicate substantial variability also. Patterns of responses to each concept vary from subject to subject. On

Table 10

Source Table for Scheffé's Post Hoc
Comparison Between Means

		Group	
	Mean	A	T
Mean		60.47	59.31
Group			
C	54.73	-5.74	-4.58
A	60.43	----	1.12

F (2,15) = ±4.42 p<.05

Figure 1. Single subject responses to each task for each concept.

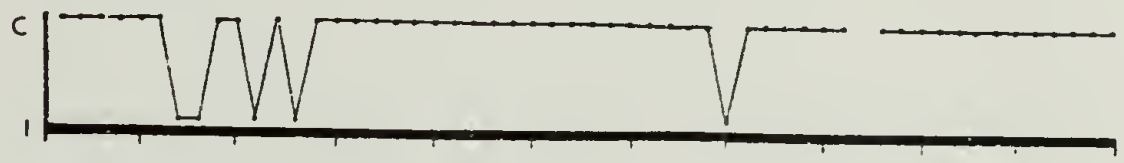
C : CORRECT

I : INCORRECT

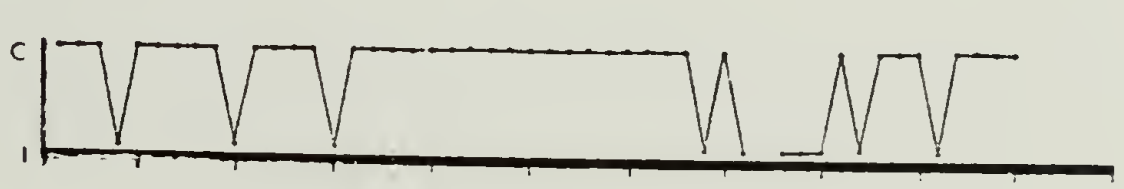
SUBJECT ONE

CONSTRUCTIONAL APPROACH

ABULIA

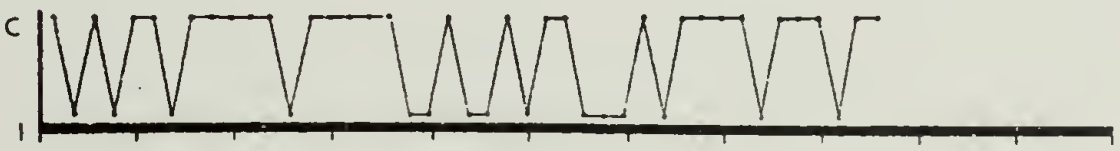


TAU EFFECT



SUBJECT FOUR

CONSTRUCTIONAL APPROACH



ABULIA

TAU EFFECT

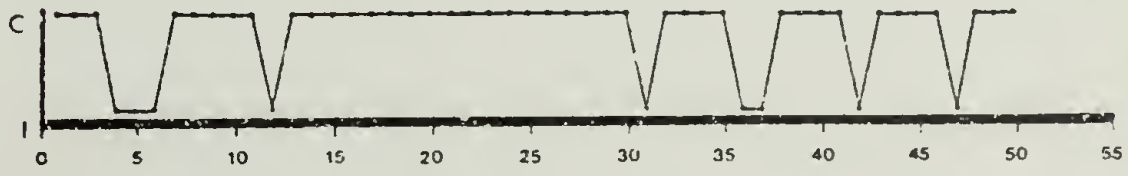


Fig. 1 (continued)

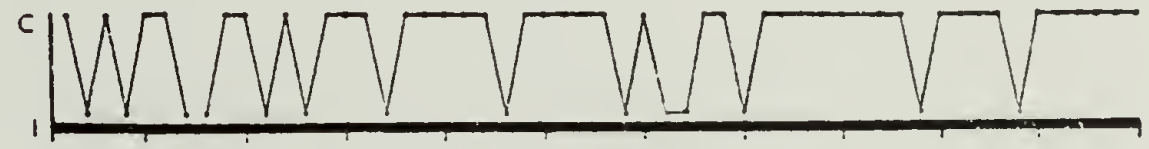
C : CORRECT

I : INCORRECT

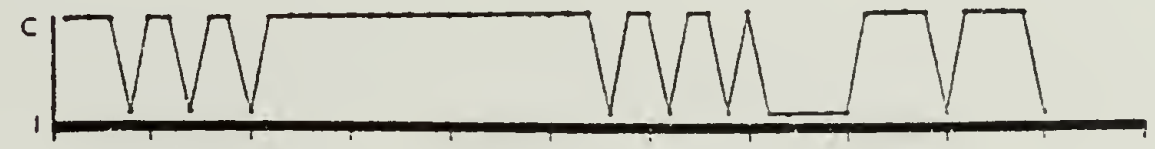
SUBJECT SEVEN

CONSTRUCTIONAL APPROACH

ABULIA

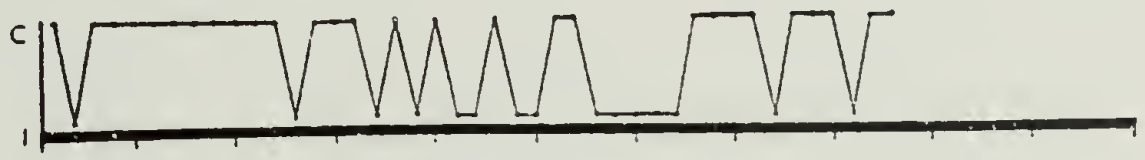


TAU EFFECT

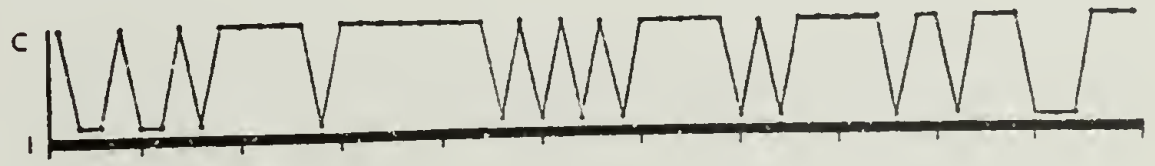


SUBJECT NINE

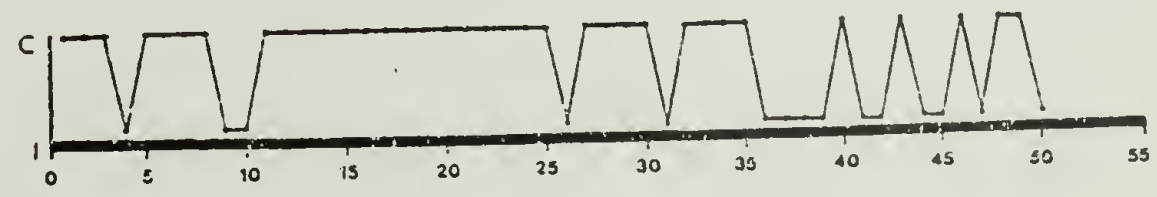
CONSTRUCTIONAL APPROACH



ABULIA



TAU EFFECT



the constructional approach, the patterns are extremely variable. On both abulia and tau effect, the patterns of responding are similar for three of the four subjects.

These differences are summarized in Table 11. Table 11 data are presented in raw proportions of correct responses. These data illustrate that each subject responded differently to each of the three concepts. It illustrates substantial differences between the constructional approach and the other concepts for individuals. Three of the four subjects scored substantially lower on the constructional approach than on the other two. However, there appears to be little difference in performance between abulia and tau effect across subjects. The one percentage point difference found between these two concepts for three of the four subjects is negligible. Only subject 1 performed substantially better on the tasks for abulia.

Table 12 is presented to illustrate that order of presentation of the concepts had no effect on the accuracy performance of these individuals. Again, the data are presented as proportions of correct performance. Subject 1 scored the highest on tasks for abulia, which appeared first in the sequence. Subject 4 performed best on abulia tasks as well, yet abulia appeared second in the sequence. Subject 7's performance was approximately equivalent across all concepts. Finally, subject 9 scored the highest on the tau effect, which appeared second in the sequence. This lack of presentation order effects were found for all 12 subjects.

The final analysis conducted on the accuracy data was a rank order of all tasks for each concept to indicate a difficulty continuum for all example identification tasks. Table 13 presents the distribution of tasks for each concept. The continuum ranges from tasks that had the

Table 11

Proportion of Correct Performances
for Four Sample Subjects
Across the Three Concepts

	Constructional Approach	Abulia	Tau Effect
S1	.53	.89	.78
S4	.65	.83	.82
S7	.77	.75	.74
S9	.65	.69	.70
Total	.65	.79	.76

Table 12

Proportion of Correct Performances
for Four Sample Subjects on
Four Different Sequences of
Concept Presentation

	Session 1	Session 2	Session 3
S1	Abulia .89	Con. App. .53	Tau Effect .78
S4	Tau Effect .82	Abulia .83	Con. App. .65
S7	Con. App. .77	Abulia .75	Tau Effect .74
S9	Abulia .69	Tau Effect .70	Con. App. .65

Table 13

Rank Order of Tasks
by Proportion Correct

CA	A	T.E.
37. 0.00	7. .08	54. .91	36. .08	16. 1.00
34. .09	8. .16	1. 1.00	38. .16	17. 1.00
47. .09	11. .16	16. 1.00	31. .16	20. 1.00
2. .18	35. .25	21. 1.00	10. .25	24. 1.00
51. .18	23. .33	33. 1.00	4. .33	27. 1.00
41. .27	31. .41	38. 1.00	42. .33	
39. .27	29. .50	40. 1.00	39. .33	
31. .45	2. .58	41. 1.00	34. .33	
38. .45	15. .58	42. 1.00	37. .41	
14. .45	52. .66	47. 1.00	26. .41	
5. .54	3. .66		40. .41	
33. .54	25. .66		45. .41	
21. .54	27. .66		50. .41	
24. .54	28. .66		44. .58	
32. .54	48. .66		35. .66	
8. .63	44. .66		12. .75	
49. .63	43. .66		47. .75	
30. .63	6. .66		46. .75	
10. .63	17. .66		48. .75	
7. .63	26. .75		30. .75	
35. .72	45. .75		21. .75	
46. .72	22. .75		2. .83	
44. .72	24. .75		5. .83	
13. .81	49. .75		7. .83	
16. .81	4. .83		25. .83	
17. .81	5. .83		29. .83	
19. .81	9. .83		32. .83	
26. .81	10. .83		33. .83	
40. .81	13. .83		41. .83	
1. .90	18. .83		43. .83	
6. .90	32. .83		49. .83	
9. .90	50. .83		3. .91	
53. .90	51. .83		6. .91	
52. .90	53. .83		9. .91	
43. .90	55. .83		15. .91	
4. 1.00	12. .91		18. .91	
11. 1.00	14. .91		19. .91	
12. 1.00	19. .91		22. .91	
36. 1.00	20. .91		23. .91	
44. 1.00	30. .91		28. .91	
45. 1.00	34. .91		1. 1.00	
50. 1.00	36. .91		8. 1.00	
23. 1.00	37. .91		11. 1.00	
	39. .91		13. 1.00	
	46. .91		14. 1.00	

lowest proportion of correct answers to tasks that had the highest proportion of correct answers. Task numbers appear on the left of each column.

A single-subject, rank order comparison of accuracy was impossible to calculate. Each task was either correct or incorrect for each subject.

Reading Rate Data

Reading rates were obtained for each subject on each task by the following formula: $60/X \cdot Y$ where X equals the duration spent on task and Y equals the number of words in the task.

A concept by order of concept presentation, repeated measures Analysis of Variance was conducted to determine the effect of concepts on reading rate. Table 14 is the source table for the ANOVA of reading rates. The data indicate that there were no significant main effects of concept $F(2,15)$ of 3.33 $p>.05$ or order of presentation $F(3,7)$ of .089 $p>.05$. There are also no interaction effects between concept and order of presentation $F(2,15)$ of 2.48 $p>.05$. One df was lost from each F test because the data for one cell, $X_{1,2,1}$ was estimated.

Single-subject data for the effect of concepts were also analyzed. Between-subject and within-subject analyses of reading rates reveal individual differences similar to those found on accuracy performance. Figure 2 presents absolute frequency distributions of reading rates pooled on the basis of 50 word per minute intervals. Two representative subjects' reading rates are presented for each concept. The differences in central tendency and variability are noticeable for both within subject, across concepts and within concept, across subjects comparisons. Similar differences were found with each of the twelve subjects. These

Table 14

Source Table for ANOVA of Reading Rates

Source	SS	df	MS	F	
A (concepts)	2494.2	2	1246.2	3.33	(2,15)
B (order)	1895.4	3	631.8	.089	(3,7)
C (subjects)	56498.4	8	7062.3		
AB	5575.8	6	929.3	2.48	(6,15)
AC	5990.8	16	374.42		

Figure 2. Absolute frequency of pooled reading rates.

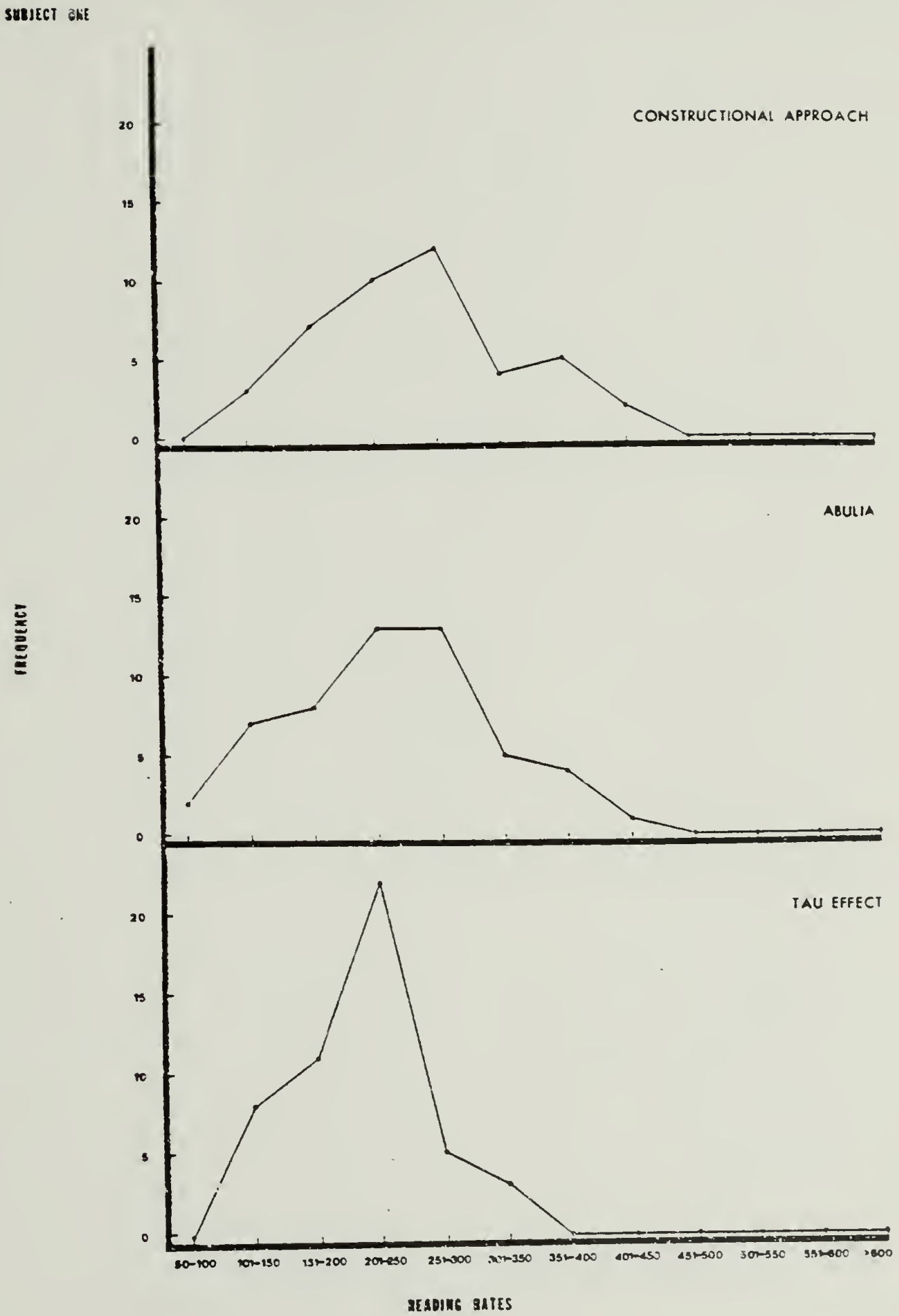
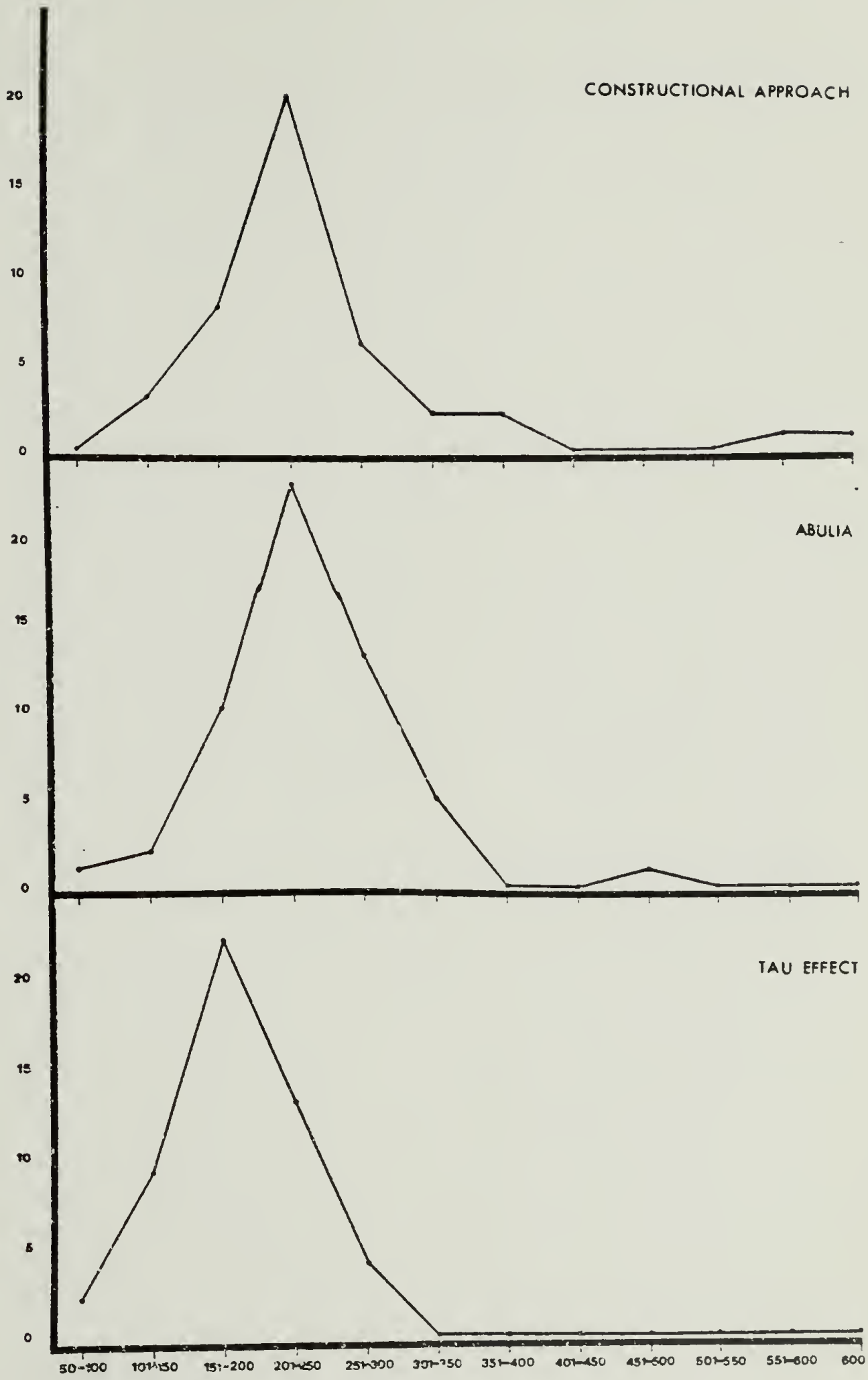


Fig. 2 (continued)

SUBJECT FIVE



data are summarized in Table 15 for all subjects. Table 15 shows that nine of the twelve subjects read tasks for constructional approach faster than the other two concepts. Of these nine, six had greater variability in reading rate on the constructional approach than on the other two concepts.

Table 16 is presented to illustrate whether order of presentation of the concepts had any noticeable effect on reading rate. Four subjects had their fastest reading performances during the first session, two with abulia and two with the constructional approach. One subject had his fastest performance during the second session, with the constructional approach. Finally, seven subjects had their fastest performance during the third session; however, six of these were with the constructional approach. These data indicate a possible interaction between constructional approach and order of presentation for individuals.

The final analysis conducted for reading rate data was a rank order of all tasks for each concept to indicate a difficulty continuum for all example identification tasks. Table 17 presents the distribution of tasks for each concept. The continuum ranges from tasks that had the lowest mean reading rates across subjects to tasks that had the highest mean reading rates across subjects. The task numbers appear on the left of each column.

Single-subject comparisons of the reading rate continuum are not included. Each subject obtained a completely individual rank order of tasks.

Relations Between Accuracy and Reading Rate Performance

A direct test of the relations between accuracy and reading rate

Table 15

Single-Subject
Mean Reading Rate Performances and Standard Deviations
For Each Concept

	Constructional Approach		Abulia		Tau Effect	
	$\overline{X}$	σ	$\overline{X}$	σ	$\overline{X}$	σ
S1	260.09	81.45	237.39	79.64	207.83	49.78
S2	116.06	56.17	180.00	66.27	170.19	64.89
S3	179.25	54.28	203.74	70.70	182.12	60.56
S4	218.92*	---	192.30	54.71	165.02	45.28
S5	241.53	78.58	235.66	59.67	184.24	49.17
S6	198.58	56.99	151.91	43.90	147.34	45.02
S7	165.39	53.70	138.23	42.18	151.04	47.92
S8	208.90	71.63	200.92	65.71	206.95	67.81
S9	207.42	71.35	199.13	95.13	228.65	53.51
S10	209.89	46.48	190.63	59.86	176.66	51.52
S11	330.15	72.74	249.38	66.36	297.29	64.37
S12	150.64	48.53	137.71	57.59	132.15	51.36

*estimate of X_{ij}

Table 16

Mean Reading Rates
For All Subjects Across Sessions

	Session 1	Session 3	Session 4
	Abulia	Con. App.	Tau Effect
S1	237.39	260.09	207.83
S2	180.00	116.06	170.19
S3	203.74	179.25	182.12
	Tau Effect	Abulia	Con. App.
S4	165.02	192.30	218.92*
S5	184.24	235.66	241.53
S6	147.34	151.91	198.58
	Con. App.	Abulia	Tau Effect
S7	165.39	138.23	151.04
S8	208.90	200.92	200.95
S9	207.42	199.13	228.65
	Abulia	Tau Effect	Con. App.
S10	176.66	190.63	209.89
S11	249.29	297.29	330.15
S12	137.71	132.15	150.64

*X_{ij} is estimated

Table 17

Rank Order of Mean Reading Rates
for Example Identification Tasks

Constructional Approach		Abulia				Tau Effect			
38.	136.94	28.	115.57	20.	216.07	47.	128.30	18.	215.04
46.	142.01	6.	118.78	10.	219.64	43.	130.49	8.	215.73
32.	156.94	49.	124.62	21.	220.83	31.	131.58	28.	226.15
31.	167.23	35.	130.07	22.	222.80	42.	134.14	27.	227.43
33.	167.26	4.	131.97	18.	225.06	46.	136.06	25.	235.29
49.	174.17	54.	134.15	9.	226.60	37.	154.13	15.	255.72
52.	174.90	16.	140.57	2.	227.10	35.	154.75	11.	266.69
40.	174.93	29.	141.96	39.	263.71	44.	157.57		
26.	175.26	50.	143.56	23.	264.46	50.	162.54		
41.	180.11	47.	144.01	38.	271.57	41.	166.32		
53.	181.43	33.	144.94	52.	282.81	34.	166.52		
7.	182.37	48.	151.08	14.	285.30	9.	170.52		
45.	185.96	53.	151.84			14.	170.90		
36.	189.70	15.	158.65			45.	173.59		
14.	190.60	43.	161.45			5.	175.51		
37.	191.60	7.	162.49			7.	176.58		
10.	192.01	17.	164.81			6.	176.60		
11.	197.44	26.	165.50			24.	177.03		
47.	198.44	25.	172.64			33.	177.05		
19.	200.42	12.	179.63			29.	178.57		
39.	202.37	42.	182.53			4.	180.80		
51.	205.52	55.	182.62			39.	181.48		
48.	214.49	37.	183.93			32.	184.49		
13.	215.15	24.	186.60			1.	185.79		
35.	218.99	46.	187.75			17.	188.26		
9.	220.01	8.	190.30			40.	189.52		
44.	221.54	32.	190.73			30.	191.03		
17.	226.19	45.	191.76			38.	191.26		
24.	226.92	51.	192.32			48.	194.40		
1.	228.61	1.	194.59			13.	194.58		
4.	229.15	41.	196.59			19.	197.77		
23.	231.52	27.	196.67			12.	197.97		
34.	231.84	11.	196.95			20.	198.16		
43.	236.26	30.	198.03			49.	198.27		
30.	242.38	34.	198.23			3.	201.62		
2.	243.25	5.	198.31			36.	207.76		
16.	244.47	31.	200.31			21.	208.69		
5.	246.93	36.	201.00			10.	209.50		
6.	273.18	3.	203.32			23.	210.14		
12.	285.55	19.	203.66			26.	210.58		
50.	285.08	40.	208.54			16.	213.03		
8.	288.24	13.	209.22			2.	213.86		
21.	337.96	44.	211.82			22.	214.65		

was conducted in order to determine whether these data could be combined for a single difficulty index of the tasks. Figures 3a-3c present scatter plots of the proportion of subjects who correctly answered each task by the $\overline{X}$ reading rate for each task. A visual analysis of these data was sufficient to conclude that a positive correlation did not exist between these two measures for any of the concepts. Therefore, transformations of these data into a single measure of difficulty (i.e. Z score transformations) could not be conducted.

Discussion

As stated above, three questions were asked in this experiment. First, are the three concepts of equal difficulty? Second, what is the rank order of example identification tasks for each concept? Third, are accuracy and reading rate measures related sufficiently to combine into a single index of performance? Of these three questions, only the latter was answered conclusively. All three kinds of analyses, single-subject, group, and correlational, indicated that there was little relation between accuracy and reading rate performance on the example identification tasks.

Single-subject analyses of accuracy and reading rate yielded similar results. Both revealed considerable inter and intra subject differences. Both indicated that constructional approach was different than the other two concepts. However, the direction of this difference appeared to be antithetical. When analyzing the accuracy data, it appeared that tasks for constructional approach were more difficult than those for abulia and tau effect. There were more errors for the constructional approach overall and eight of the twelve subjects performed worse on tasks for

Figure 3. Scatter plot of proportion correct and mean reading rates for each task.

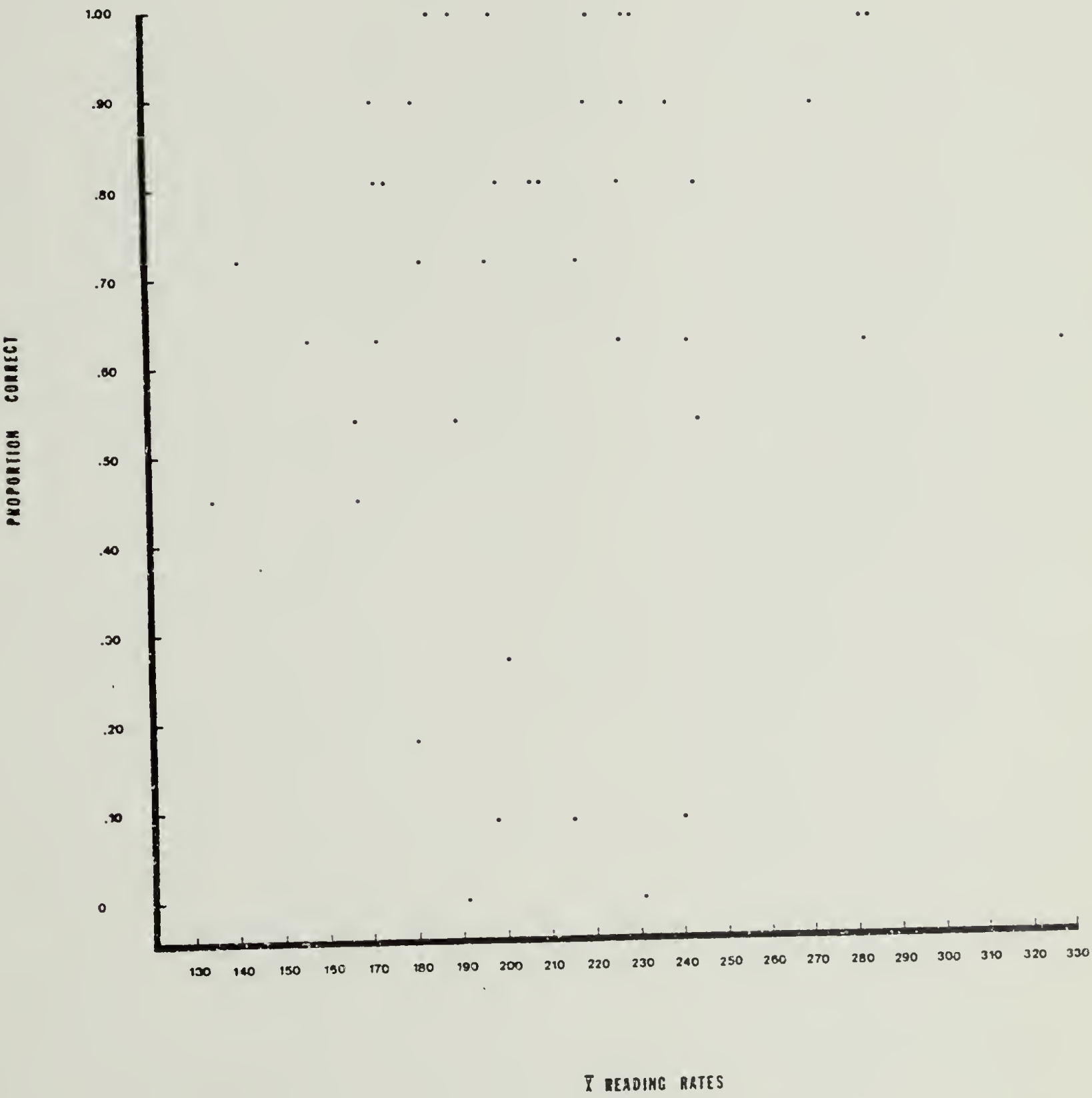


Fig. 3 (continued)

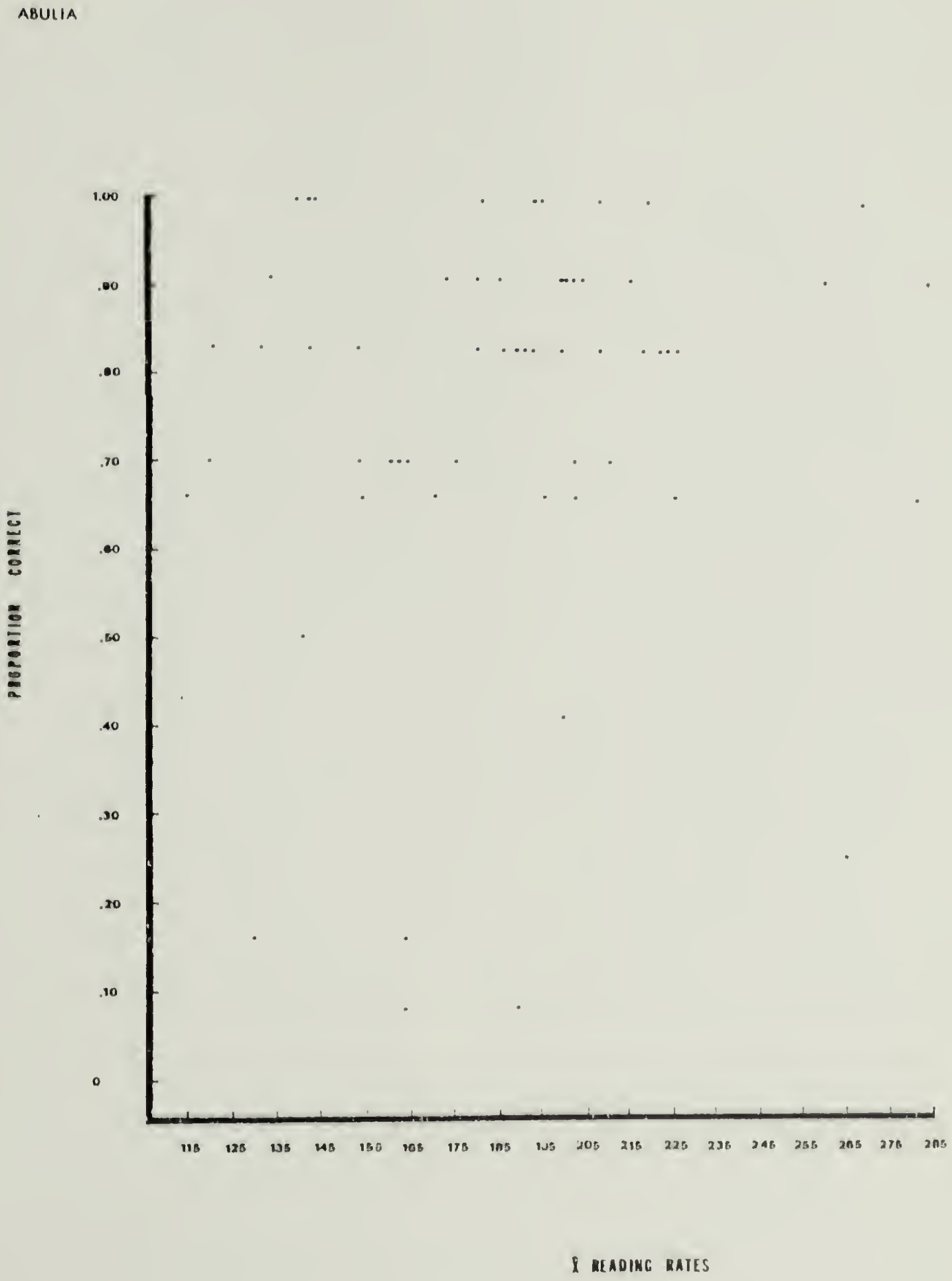
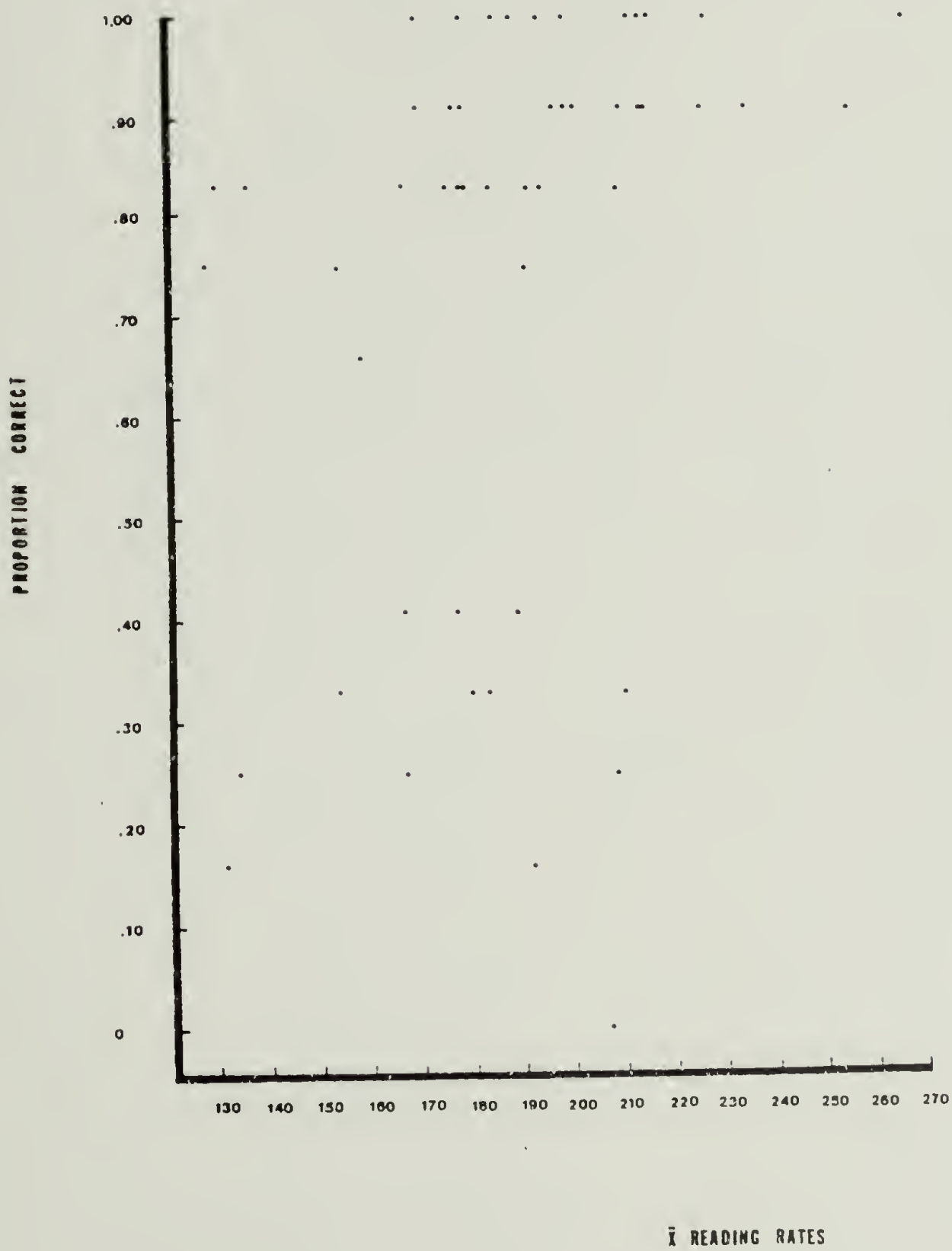


Fig. 3 (continued)

TAU EFFECT



constructional approach than they did on tasks for the other concepts. It was assumed that more difficult tasks would take longer to complete.

Of course, reading rates could be faster on very difficult tasks if subjects decided that they were unable to answer the tasks and therefore simply went through the motions of reading and answering the questions. This would also result in less accurate performance. However, two aspects of the experiment appear to make this conclusion unlikely. First, on over 50% of the tasks for each concept, subjects were asked to justify their answers vocally. The exact tasks that they were to justify were chosen individually and on a random basis by the experimenter. Therefore, there was little chance that the subjects could predict which tasks they would have to justify. There was no indication from these vocal responses that subjects knew less about the irrelevant aspects of the examples for constructional approach than for the other two concepts. This would be expected if subjects were not attending to constructional approach tasks. Second, even when accuracy was held constant and only those tasks that were answered correctly were analyzed, subjects still had faster reading rates for the constructional approach than for the other two concepts. These single-subject analyses fail to provide a conceptualization of the exact relation between task difficulty and reading rate. However, it appears that accuracy and reading rate are not related sufficiently to use them together as an index of performance.

The ANOVA's also revealed this apparent lack of relation between accuracy and reading rate measures. From the ANOVA in which accuracy data were used, it was found that there was a significant main effect for concepts and a significant interaction effect between concepts and order of presentation. Analyzing the same effects using reading

rate data resulted in no significant effects. If these two measures were related in some useful manner, it would be expected that both could be used to determine the relations between variables. The Post Hoc comparison of concept means led to the same conclusions. Post Hoc comparisons of the accuracy data resulted in significant differences between the mean for constructional approach and the other two concepts. Post Hoc comparisons of the reading rate means failed to reject the null hypothesis. Again, these results would be unexpected if the two measures were related.

A fourth comparison between accuracy and reading rate can be obtained by the juxtaposition of Tables 13 and 17. These tables present the rank order of tasks for each concept on the basis of accuracy performance and reading rate performance. A perusal of the task numbers of the left of each column reveals little similarity between accuracy and reading rate.

Finally, a direct test of the relationship between accuracy and reading rate failed to indicate any positive or negative correlation. The distributions in the scatter diagrams appear to be random, indicating a zero correlation. Given the strength of the data across all of these analyses, it did not seem necessary to conduct any further tests for a correlation.

It appears, then, that accuracy and reading rate are not related for the example identification tasks used in this study. The question remains as to how to use these data to answer the first two questions. The ambiguities surrounding the answers to these questions center on the different conclusions that can be made depending on whether accuracy or reading rate data are used.

The answer to the first question is fairly straight forward. Does one conduct the next experiment with the materials for these three concepts? After all, no significant main effect was found for the concept variable when reading rate was used as the data. Or does one rewrite or throw out constructional approach because it was found to be significantly different from the other two concepts when accuracy data were used? The somewhat arbitrary decision made by this investigator is to follow the latter. With only 12 subjects, the power of this experiment to detect a real difference when a real difference did in fact exist was weak. Yet, by using accuracy data a difference was found. More importantly, the single-subject analyses did reveal a systematic difference between the reading rates for constructional approach and the other concepts. Therefore, in order to be conservative, the materials for constructional approach will be rewritten and further tests conducted on them.

Unfortunately, there is not a straight forward answer to the second question. However, there are three possible solutions. The first solution is to use the accuracy data to guide the selection of tasks when one's primary interest is to observe the effects of experimental manipulations on accuracy. Similarly, one would use the reading rate data to guide the selection of tasks for those studies in which one is trying to manipulate reading rate. A second solution is to conduct more experiments to test the sensitivity of these measures to detect differences in materials. It appears from this study that accuracy data is more sensitive to these differences. This is witnessed by the significant differences found in the ANOVA and the Post Hoc comparisons. A third solution is to combine the accuracy and reading rate data in a somewhat arbitrary fashion. For example, Table 18 presents a rank order

Table 18

Rank Order of Tasks on the Basis of Accuracy and Reading Rate

Constructional Approach				Abulia				Tau Effect			
	.00		.90		.08				.08		
37.	191.60	56.	174.90	7.	162.49	32.	190.73	36.	207.76	25.	213.86
		53.	181.43			51.	192.32				
	.09	9.	220.01		.16	5.	198.31		.16		.91
47.	198.44	1.	228.61	8.	190.30	13.	209.22	31.	131.58	9.	170.52
34.	231.84	43.	236.26	11.	196.95	10.	219.64	38.	191.26	6.	176.60
		6.	273.18			18.	225.06			19.	197.77
	.18				.25	9.	226.60		.25	3.	201.62
51.	205.52		1.00	35.	130.07			10.	209.50	23.	210.14
2.	243.25	45.	185.96				.91			22.	214.65
		36.	189.70		.33	54.	134.15		.33	18.	215.04
	.27	11.	197.44	23.	264.46	12.	179.63	42.	134.14	28.	226.15
41.	180.11	44.	221.54			37.	183.93	34.	166.52	15.	255.72
39.	202.37	23.	231.52		.41	46.	187.75	4.	180.80		
		50.	285.08	31.	200.31	30.	198.03	39.	181.48		1.00
	.45					36.	201.00			14.	170.90
38.	136.94				.50	19.	203.66		.41	24.	177.03
31.	167.23			29.	141.96	20.	216.07	37.	154.13	1.	185.79
14.	190.60					39.	263.71	50.	162.54	17.	188.26
					.58	14.	285.30	45.	173.59	13.	194.58
	.54			15.	158.65			40.	189.52	20.	198.16
32.	156.94			2.	227.10		1.00	26.	210.58	16.	213.03
33.	167.26					16.	140.57		.58	8.	215.73
24.	226.92				.66	47.	144.01	44.	157.57	27.	227.43
5.	246.93			28.	115.57	33.	144.94			11.	266.69
21.	337.96			6.	118.78	42.	182.53		.66		
				48.	151.08	1.	194.75	35.	154.75		
	.63			43.	161.45	41.	196.59				
49.	174.17			17.	164.81	40.	208.54		.75		
7.	182.37			25.	172.64	21.	220.83	47.	128.30		
10.	192.01			27.	196.67	38.	271.57	46.	136.06		
30.	242.38			3.	203.32			30.	191.03		
8.	288.24			44.	211.82			48.	194.40		
				52.	282.81			12.	197.97		
	.72							21.	208.69		
46.	142.01				.75						
48.	214.49			49.	124.62				.83		
35.	218.99			26.	165.60			43.	130.49		
				24.	186.60			41.	166.32		
	.81			45.	191.76			5.	175.51		
40.	174.93			22.	222.80			7.	176.58		
26.	175.26							33.	177.05		
19.	200.42				.83			29.	178.57		
13.	215.15			4.	131.97			32.	184.49		
17.	226.92			50.	143.56			49.	198.27		
16.	244.47			53.	151.84			2.	213.86		
				55.	182.62						

of tasks that used accuracy as the primary index of difficulty. The reading rates are used to further differentiate those tasks that share the same accuracy index. The advantage of this solution is that one might be able to make a more fine-grain decision as to where tasks will be used in the process of teaching and testing. Unfortunately, this requires knowing more about the relation between reading rate and difficulty. Further research is needed to determine this relation.

Although the data that were analyzed led to some interpretation of the relations among concepts, among tasks for each concept and between accuracy and reading rate, functional conclusions were prohibited by the amount of intersubject variability. Single-subject analyses were inconclusive because subjects responded so differently. Group analyses did reveal some relations and these were presented. However, these group analyses were not satisfactory for the reasons argued in the introduction. The experimental plan had been that if individual differences were obtained, the other dependent measures (i.e. the study-behavior questionnaire and the vocal justification data) would be analyzed to determine possible reasons for these differences. Unfortunately, neither of these variables could be analyzed.

First, the study-behavior questionnaire was not sensitive enough to individual subject's study behavior to indicate any relations between example identification performance and study habits. The small sample size prohibited collapsing this data for group comparisons. Thus, these data were not analyzed formally. This result suggests three alternative future plans. First, to redesign the study-behavior questionnaire. Second, to administer the questionnaire repeatedly to the same subjects. Third, to wait until a sufficient number of subjects have completed the

questionnaire before analyzing the results. The first and second alternatives are more in line with behavioral methodologies.

The justification data were not analyzed because too many subject responses could not be transcribed. The ten hours of analyzed tapes reported in the reliability section were the extent of the interactions that could be understood from the tapes. Insensitive microphones and cassette tapes were the cause of this data loss. These results indicate that if vocal behavior is to be used as a dependent measure in future studies, better recording equipment is necessary.

In conclusion, this experiment provided direction for a number of different issues related to materials development. First, the results indicated that accuracy and reading rate are not correlated measures of performance on example identification tasks. Therefore, they can not be combined to determine a single index of difficulty. In addition, decisions made from each measure, alone, will not be similar. Thus, investigators, instructional designers and teachers need to decide whether either or both measures are important to their goals before using them to test materials and/or individuals. Second, the results showed that one concept, constructional approach, was significantly different from the other two concepts. Therefore, the materials for this concept need to be rewritten and tested further before they can be used in subsequent research. This result also suggests that a small number of subjects (12) is sufficient to determine differences between materials when a study is conducted under similar, controlled conditions. Third, the results demonstrated considerable variability across subjects. Although these intersubject differences can be controlled statistically, they suggest that further experimental control is needed before single-subject analyses

can be used. Research needs to be conducted to determine more sensitive tests of subject entering behavior and other relevant variables so that single-subject results can be interpreted.

C H A P T E R V

CONCLUSIONS

The purpose of this thesis was twofold. First, a proposed typology of instructional tasks was evaluated to determine whether others could reliably classify tasks according to the categories of the typology. Second, the concepts and tasks that were needed for subsequent research on the typology were developed and tested. In this chapter, both of these goals are discussed in detail. Specifically, the research conducted for this thesis is evaluated, conceptual and experimental issues are raised and future directions and studies are explicated.

Validation of the Typology

The first experiment provided evidence that other psychologists could classify tasks according to the typology. An overall agreement index of 88% between experimenter and subjects was as high or higher than those reported for other typologies (Sullivan, 1969; Williams, 1977). Although subjects scored lower on tasks obtained from commercially available materials, it was tentatively concluded that this was due to unclearly specified tasks. The amount of time that subjects needed to study the definitions of the task classes ranged from 15 to 26 minutes. These are as low or lower than those reported by Williams (1977). In addition, subjects could not specify verbal tasks that were not covered by the typology. A number of subjects indicated that the typology could be further subdivided. These suggestions

were noted and extensions of the typology have been started. Overall, though, it appeared that the typology, albeit simple, is a viable method for classifying cognitive or verbal behavior.

A number of issues have been raised by this study. First, it appears that comparisons between typologies are difficult and may be misleading. As indicated by the data presented, better descriptions of subject characteristics are necessary if comparisons are to be made across studies. Substantial differences were found between graduate and undergraduate subjects, yet both sets of subjects conform to the general criterion of expertise in writing behavioral objectives and test items (Williams, 1977). Although it was concluded that the differences were not important for the purposes of this study, these differences did prohibit comparisons to other studies.

The problem of comparison is further confounded by the fact that different strategies are used to teach the typologies to subjects. The definitions, text, examples, etc. that subjects study are inextricable from the typology itself. In this study, specific programming strategies were used to define the classes of tasks. It is not clear how other classification schemes have been taught. However, even if this was clarified, it is doubtful that the teaching programs would be similar enough to compare.

There are two possible solutions to these problems of comparison. First, the typologies could be compared on a logical basis. Any classification schemes that have been tested for reliability could be compared on the basis of the number of conditions that they cover.

Does the typology, taxonomy or hierarchy cover both study and evaluation tasks? Are the distinctions between instruction and evaluation specified? Can other tasks be named that do not fit into the classification scheme? On the basis of such questions, one could justify the use of one scheme over another. The problem with this solution is that it is not clear that such logical analyses are effective. A typology could be designed that covers every possible condition, but its complexity or generality might make it virtually unusable except by a handful of people. This, in fact, is one of the criticisms of Bloom's taxonomy (Sullivan, 1969).

The second solution might be to directly compare classification schemes through experimental procedures. Two or more classification schemes that have been previously tested individually might be taught to a group of subjects. The teaching programs and the classification tests could be designed similarly. Then, the subjects' differential performances on these classification tests could be compared. The simplicity of this solution is offset only by the length of time it would take to program the definitions of each classification scheme and to conduct the studies. Such expenditure of time and effort may be worthwhile in the long run. Thus, one way that the importance of this particular typology may be extended in the future is by such a direct comparison to the more notable classification schemes.

Another issue raised by Experiment I was whether the typology could be further subdivided. At least one subject indicated the need to distinguish between different levels within each class of task.

This break down is being conducted currently. Johnson and Chase (1979) presented a complex typology that included a 16 x 11 matrix of teacher and student behavior. This matrix involved 7 classes of tasks. Although this typology covered the kinds of questions asked by the subjects during Experiment I, it has proved too cumbersome to learn and use. Therefore, this typology is being simplified. Currently, it involves the five classes of tasks presented in this thesis with a general discrimination made at the level of Elementary tasks and Complex tasks (Table 19). Elementary tasks are those that are rigid, inflexible and are not extended to novel instances. For example, a formula, a technical definition and a one-of-a-kind object would all be elementary tasks when the relation between teacher behavior (instructional stimuli) and student behavior do not change from instance to instance. Complex tasks are those tasks that are extended to new situations or instances. Complex tasks are flexible, the behavior changes and the stimuli changes from instance to instance. Under the typology presented here, all elementary tasks are labeled as copy tasks. In the new typology, copy tasks are subdivided to include simple or fixed kinds of define, example identification and exemplify tasks.

Certainly, other changes and extensions can be made to this typology. However, all of these changes need to be evaluated. Future behavior with respect to this typology of verbal tasks will include reanalyzing the specific tasks, reprogramming the descriptions of the tasks and further testing of the validity and practicality of the typology.

Table 19

A Typology of Verbal Behavior

ELEMENTARY TASKS	COMPLEX TASKS
Copy (Transcriptive)	Flexible Define (Intraverbal)
Fixed Define (Intraverbal)	Flexible Example Identification (Intraverbal and Tact)
Fixed Example Identification (Intraverbal and Tact)	Flexible Exemplify (Intraverbal)
Fixed Exemplify (Intraverbal)	Combinations (Multiply Caused)

Materials Assessment

Both Experiments II and III were designed primarily to find at least three equivalent concepts that could be used in subsequent single-subject research. In order to accomplish this goal, these studies attempted to determine differences in difficulty of the concepts. The behaviors measured in Experiment II were sensitive enough to conclude that three concepts were approximately equal (constructional approach, abulia, tau effect). All other concepts were rejected as being too easy or too difficult. The behaviors measured in Experiment III indicated that one of the seemingly equivalent concepts, constructional approach, was, in fact, different. Analyses of both individual performance and group performance indicated significant differences between constructional approach and the other concepts. However, no differences were found between abulia and tau effect. Therefore, it was concluded that the concepts, abulia and tau effect were similar enough to be used in subsequent research on the tasks of the typology.

Although the results of Experiment II and Experiment III indicated that abulia and tau effect were similar, two analyses suggest that further research is needed to conclude that constructional approach was different. First, both the w^2 test and the single-subject data revealed that factors other than concepts account for most of the variability. The w^2 test demonstrated that only 16% of the variance was attributed to the effect of concepts. The single-subject analyses revealed substantial intrasubject variability within each concept

condition. Both of these analyses suggest that the measure of concept difficulty in this study was not stable. Without a stable measure, it would be erroneous to conclude that the resultant relations are reliable, practical relations.

A conceptual analysis of the differences between concepts further confounds these findings. A number of authors have indicated that it is impossible to define a concept without referring to the behavior that occurs with respect to that concept (Keller and Schoenfeld, 1950; Skinner, 1957; Rosch, 1978). This notion suggests that a concept is not a single entity, but rather it is a synthesis of different classes of behavior that occur in the presence of particular stimulus properties of the environment. For example, the concept dog is not simply the dictionary definition of a living creature. Rather it is a combination of the stimulus features that make up this creature and all the behaviors that the community emits in the presence of dogs. This notion is critical to discussing concept difficulty. For instance, the concept dog is a straightforward concept for most adults. However, identifying a dog is quite different from defining dog. In fact, if the concept dog was compared to the concept abulia, it is possible that more people could define abulia than could define dog. However, it is also likely that more people could identify dogs than could identify instances of abuiia. Whether this specific conclusion is correct or not, the point is that concept difficulty can not be judged on the basis of one class of behavior. Although this assumption can not be proven by the data analyzed here, the different results obtained in

Experiment II and Experiment III are consistent with this assumption. Specifically, the exemplify tasks used in Experiment II failed to find consistent differences between abulia and constructional approach. Both were exemplified correctly five times. In addition, the exemplify tasks suggest a slight difference between tau effect and the other concepts, as 7 exemplify tasks were answered correctly for tau effect. However, the example identification tasks used in Experiment III demonstrated that 8 of the 12 subjects performed less accurately on the constructional approach than they did on abulia. This study also showed that 6 of the 12 subjects scored almost identically on tasks for abulia and tau effect. The only relation between concepts that was the same for both studies was the relation between tau effect and constructional approach. Ten subjects scored less accurately on constructional approach than on tau effect in Experiment III.

The conclusions that can be drawn from these data at this time are limited. The data do suggest that other classes of behavior should be tested with respect to these concepts before concluding that constructional approach is different from the other concepts.

Another incidental and unexpected result of these two studies was the lack of a positive relation between accuracy and reading rate measures. A number of authors have discussed the sensitivity of reading rate data and the positive relation between reading rate and accuracy (Houghton, 1979; Lindsley, 1979; Kintsch, 1974). However, in Experiment III no relation was found. In addition, accuracy data appeared more sensitive to differences between materials than reading rate. These

results have been discussed in detail earlier and therefore, need not be discussed here. However, in defense of the possibility that reading rate is an important measure of verbal behavior, it must be stated that the timing procedure used was inaccurate. The reaction time of experimenters to starting and stopping the timer varied both across experimenters and across tasks for the same experimenter. Although it is expected that such differences would be randomly distributed across tasks and concepts, this procedure needs to be brought under stricter control before definite conclusions can be made.

One solution to this problem is to train research assistants on accuracy and speed of engaging the timers. A second solution is to use a more direct measure of rate. Subjects would be given a certain number of tasks to complete in a set amount of time. Rate would be the measure of the number of tasks completed in this interval.

In conclusion, the exploratory nature of these materials investigations prevented any clear-cut quantifiable conclusions. However, what does emerge from this morass is the need for a methodology to assess task and concept variables. Such a methodology is essential to determining the kinds of verbal learning that are most likely to facilitate transfer to a wide variety of behaviors. If the relations between tasks and the concepts that are being learned are not known, than it is likely that the relations between learning and transfer cannot be determined. Results obtained could be an artifact of the prose passages defining the concept, an artifact of the behaviors measured, or a combination of both. Statistical designs may be useful

to some extent in controlling for these problems. However, for the reasons stated earlier, there is a need for the increased data and individual decisions that are facilitated by single-subject methodologies. In order to carry out single-subject studies, it is essential to determine the relation between the concepts used.

Part of the methodology for testing the concept relations has been determined through this investigation. Other components are needed to flush out some of the problems found here. The following is an example of how this methodology would be carried out. First, it is necessary to measure each concept that is to be used. In order to measure a concept, it is necessary to develop a range of tasks for the concept. These tasks are edited and a small number of subjects are asked to answer the tasks. Explicit comments are obtained from these subjects. The comments should concentrate on three problems: a) communication problems (e.g. diction, grammar, sentence structure), b) motivation problems (e.g. boredom, lousy humor), and c) step size (e.g. too many questions, too few questions). As Markle (1967) indicated, no formula exists for predicting what is right and what is wrong with a task. Therefore, the more information that is considered, the more likely the production of stable tasks. All tasks should be modified on the basis of these comments. Finally, a large number of subjects should be tested on a range of task types for each concept. If the subject population prohibits repeated sessions with the same subjects, then a number of separate studies with different subjects can be conducted. Each study should measure the effect of a different type

of task on all the concepts.

When all of these steps have been carried out, then it is possible to conduct a transfer study and interpret the results. If a certain concept is answered more accurately on a particular type of task, then greater relative transfer to this task type can be expected with this concept. On the other hand, if a concept is answered inaccurately on a particular task type, then less relative transfer can be expected. With the kinds of data afforded by all the above steps, many different kinds of transfer studies can be conducted.

Summary

In conclusion, the research conducted here has determined some of the prerequisites for conducting single-subject transfer of learning studies. The range of tasks and the classification scheme for this range has been developed. Two concepts appear to be of similar difficulty to be used in repeated measures designs without having noticeable task confounds. Finally, a useful methodology has been worked out to determine other concepts that can be used in single-subject research. What is needed now is one more test for concept difficulty. The materials for constructional approach will be rewritten and tested in comparison to abulia and tau effect on all of the tasks of the typology. When this is completed, then the questions of transfer will be addressed. Hopefully, the final product will be a complete package that describes the tasks that are necessary and sufficient to teach college students and perhaps high school students the strategies needed to learn the behaviors in which professionals and

advanced students engage.

REFERENCES

Anderson, J.R. & Bower, G.H. Human Associative Memory. Washington D.C.: V.H. Winston & Sons, 1973.

Anderson, R.C. & Kulhavy, R.W. Learning concepts from definitions. American Educational Research Journal, 1972, 9, 335-390.

Andre, T. Does answering higher level questions while reading facilitate productive learning. Review of Educational Research, 1979, 49, 280-318.

Ausubel, D.P. The Psychology of Meaningful Verbal Learning. New York: Grune & Stratton, 1963.

Baer, D.M. Perhaps it would be better not to know everything. Journal of Applied Behavior Analysis, 1977, 10, 167-172.

Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H. & Kratwohl, D.R. A Taxonomy of Educational Objectives: Handbook I: The Cognitive Domain. New York: MacKay, Inc. 1956.

Bostow, D.E. The Experimental Analysis of College Instructional Practices: Some Mid-Course Corrections. Invited address presented at the Second Annual Mid-Western Association for Behavior Analysis Convention, Chicago, 1976.

Chase, P.N., Johnson, K.R. & Keenan, J.B. Survey of commerically available study materials. Unpublished manuscript, 1977.

Clark, D.C. Teaching concepts in the classroom: A set of teaching prescriptions derived from experimental research. Journal of Educational Psychology, 1971, 62, 253-278.

Deese, J. Psychology of Learning. New York: McGraw-Hill, 1958.

deLuna, S.V. Learning and Transfer in Hierarchically Structured Concept Attainment Tasks. University of Massachusetts, Amherst:

Unpublished Master's Thesis, 1972.

Ellis, H.C. The Transfer of Learning. New York: The Macmillan Co., 1965.

Engelmann, S. Conceptual Learning. San Raphael, Ca.: Dimensions Press, 1969.

Fisher, R. Statistical Methods and Scientific Inference. London: Oliver and Body, Ltd., 1956.

Furth, H.G. The influences of language on the development of concept formation in deaf children. Journal of Abnormal and Social Psychology, 1961, 63, 386-389.

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

Gagné, R.M. The Analysis of Instructional Objectives for the Design of Instruction. In, R. Glazer (Ed.), Teaching Machines and Programmed Instruction II: Data and Directions. Washington D.C.: National Education Association, 1965, 21-53.

Gagné, R.M. The Conditions of Learning. New York: Holt, Rinehart, and Winston, 1965. 1970.

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

Gagné, R.M., Mayor, J.R., Garstens, H.L. and Paradise, N.E. Factors in acquiring knowledge of a mathematical task. Psychological Monographs, 1962, 76, (7), 526.

Gagné, R.M. and Paradise, N.E. Abilities and learning sets in knowledge acquisition. Psychological Monographs, 1961, 75, (14), 518.

Gagné, R.M. and staff, University of Maryland Mathematics Project. Some factors in learning non-metric geometry. Monographs of Society

for Research in Child Development, 1965, 30, 42-49.

Galanter, E. (Ed.) Automatic Teaching: The State of the Art. New York: Wiley & Sons, 1959.

Gallup, H.F. & Warranch, H.R. Some Potential Problems With Commercialization of PSI Materials. In, J.G. Sherman, R.S. Ruskin, & R. Lazar (Eds.), Personalized Instruction in Higher Education Today: Proceedings of the Third and Fourth National Conferences. Washington, D.C.: Center for Personalized Instruction, 1978.

Harlow, H. The formation of learning sets. Psychological Review, 1949, 56, 51-65.

Hayes, W.L. Statistics for the Social Sciences. New York: Holt, Rinehart and Winston Inc., 1973.

Hersen, M. & Barlow, D. Single Case Experimental Designs: Strategies For Studying Behavior Change. New York: Perganon Press, 1976.

Hilgard, E.R., Irvine, R.P. & Whipple, J.E. Rote memorization, understanding and transfer: An extension of Katona's card-trick experiments. Journal of Experimental Psychology, 1953, 46, 288-292.

Hirshfield, G.A. Taxonomic Approach to the Evaluation of Secondary School English Classes. University of Illinois: Unpublished Doctoral Dissertation, 1969.

Hively, W. Programming stimuli in matching to sample. Journal of the Experimental Analysis of Behavior, 1962, 5, 279-298.

Hively, W. Domain-referenced achievement testing: Theory and practice. Unpublished manuscript, 1970.

Houghton, E. Precision Teaching. Workshop presented at Fernald School, Waltham, MA, August, 1979.

James, W. Principles of Psychology. New York: Holt, 1890.

Johnson, K.R. & Chase, P.N. Teaching Concepts, and Operations to High School and College Students. Workshop presented at the Fourth Annual Association for Behavior Analysis Convention. Chicago, Ill., 1978.

Johnson, K.R. & Chase, P.N. Instructional Applications of B.F. Skinner's Analysis of Verbal Behavior. Worktext for Invited Workshop presented at the Fifth Annual Association for Behavior Analysis Convention. Dearborn, Mi., 1979.

Johnson, D.M. & O'Reilly, C.A. Concept attainment in children: Classifying and defining. Journal of Educational Psychology, 1964, 55, 71-74.

Johnson, D.M. & Stratton, R.P. Evaluation of five methods of teaching concepts. Journal of Educational Psychology, 1966, 57, 48-53.

Johnston, J., O'Neill, G.W., Walters, W.M. & Rasheed, J.M. The measurement and analysis of college student study behavior: Tactics for research. In, J.M. Johnston (Ed.) Behavior Research and Technology in Higher Education. Springfield, Ill.: Charles C. Thomas Publishers, 1975.

Johnston, J.M. & Pennypacker, H. A behavioral approach to college teaching. American Psychologist, 1971, 26, 219-244.

Johnston, J.M. and Pennypacker, H.S. Strategies and Tactics of Human Behavioral Research. Unpublished manuscript, 1979.

Katona, G. Organizing and Memorizing. New York: Columbia University Press, 1940.

Keenan, J.B. & Grant, L. The Question of Relevance and Level of Learning: A Stimulus Control Question. Paper presented at the Fifth Annual

Conference on Personalized Instruction, Washington D.C., 1979.

Keller, F.S. Definition of Psychology. Englewood Cliffs, N.J.: Prentice-Hall Inc., 1973.

Keller, F.S. and Schoenfeld, W. Principles of Psychology. Englewood Cliffs, N.J.: Prentice-Hall, 1950.

Kintsch, W. The Representation of Meaning in Memory. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Kolb, J.R. Effects of relating mathematics to science instruction on the acquisition of quantitative science behaviors. Journal of Research in Science Teaching, 1967-68, 5, 174-182.

Kratochwill, T.R. Single Subject Research, Strategies for Evaluating Change. New York: Academic Press, 1978.

Lindsley, O.R. 100 Years of Learning Compared on Standard Acceleration Charts. Address presented at 87th Annual American Psychologists Association Convention, New York, 1971.

Markle, S.M. Empirical testing of programs. In, P. Lange (Ed.) Programmed Instruction. Chicago: University of Chicago Press, 1967.

Markle, S.M. & Tiemann, P.W. Behavioral analysis of cognitive content. Educational Technology Magazine, 1970, 10, 41-45.

Michael, J. Statistical inference for individual organism research: mixed blessing or curse? Journal of Applied Behavior Analysis, 1974, 7, 647-653.

Miller, L.K. & Weaver, F.H. A behavioral technology for producing concept formation in university students. Journal of Applied Behavior Analysis, 1976, 9, 289-300.

Morrisett, L.J. & Hovland, C.I. A comparison of three varieties of training in human problem solving. Journal of Experimental

Psychology, 1959, 58, 52-55.

Myers, J.L. Fundamentals of Experimental Design. Boston: Allyn and Bacon, Inc., 1979.

Perkins, M.R. Facilitative Transfer in Prose Learning of Elementary School Children. Unpublished Masters Thesis, 1976.

Pfeiffer, I.L. Teaching in ability grouped english classes: A study of verbal interactions and cognitive goals. Journal of Experimental Education, 1967, 36.

Polya, G. Mathematics and Plausible Reasoning, Vol. I: Induction and Analogy in Mathematics, Princeton, N.J.: Princeton University Press, 1954.

Postman, L. Transfer, Interference and Forgetting. In, J.W. Kling & L.A. Riggs (Ed.) Experimental Psychology. New York: Holt, Rinehart and Winston, Inc., 1971.

Reed, H.B. Factors influencing the learning and retention of concepts. I. The Influence of set. Journal of Experimental Psychology, 1946, 34, 71-87.

Rickards, J.P. Adjunct postquestions in test: a critical review of methods and processes. Review of Educational Research, 1979, 49, 181-196.

Rosch, E. Principles of Categorization. In E. Rosch and B. Lloyd (Eds.) Cognition and Categorization. Hillsdale, N.J.: Lawrence Erlbaum Assoc. Inc. Publishers, 1978.

Royer, J.M. Theories of Learning Transfer. Technical Report No. 79, University of Illinois at Urbana-Champaign: Center for the Study of Reading, 1978.

Semb, G. & Spencer, R. Beyond the level of recall: An analysis of complex educational tasks in college and university instruction. In, L.E. Fraley & E.A. Vargas (Eds.) Behavior Research and technology in Higher Education. Champaign, Ill.: Research Press, 1976.

Sidman, M. Tactics of Scientific Research. New York: Basic Books Inc., Publishers, 1960.

Sidman, M., Cresson, O., and Willson-Morris, M. Acquisition of matching to sample via mediated transfer. Journal of the Experimental Analysis of Behavior, 1974, 22, 261-274.

Sidman, M. and Stoddard, L.T. The effectiveness of fading in programming a simultaneous form discrimination for retarded children. Journal of the Experimental Analysis of Behavior, 1967, 10, 3-15.

Skinner, B.F. Are theories of learning necessary? Psychological Review, 1950, 57, 193-158.

Skinner, B.F. Verbal Behavior. Engelwood-Cliffs, New Jersey: Prentice-Hall, Inc., 1957.

Skinner, B.F. Teaching Machines. Science, 1958, 128, 969-977.

Skinner, B.F. Why we need teaching machines. Harvard Educational Review, 1961, 31, 377-398.

Skinner, B.F. Contingencies of Reinforcement. Englewood Cliffs, N.J.: Prentice-Hall Inc., 1969.

Stoddard, L.T. and Sidman, M. The effects of errors on children's performance on a circle-ellipse discrimination. Journal of the Experimental Analysis of Behavior, 1967, 10, 261-270.

Sullivan, H.J. Objectives, evaluation, and improved learner achievement. In, J. Popham, E.W. Eisner, H.M. Sullivan & L.L. Tyler (Eds.)

Instructional Objectives. AERA Monograph Series On Curriculum Evaluation, #3, Chicago: Rand McNally, 1969.

Watts, G., and Anderson, R.C. Effects of three types of inserted questions on learning from prose. Journal of Education and Psychology, 1971, 62, 387-394.

White, R.T. Research into learning hierarchies. Review of Educational Research, 1973, 43, 361-375.

White, O.R. & Haring, N.G. Exceptional Teaching. Columbus, Ohio: Charles E. Merrill Co., 1976.

Williams, R.G. A behavioral typology of educational objectives for the cognitive domain. Educational Technology Magazine, 1977, 17, (6).

Wolff, J.G. Effects of verbalization and pretraining on concept attainment by children in two mediation categories. Journal of Educational Psychology, 1967, 58, 19-26.

APPENDIX

A

TYPOLOGY CLASSIFICATION TEST

Sample Tasks

Please read each task carefully and completely before saying what category is exemplified. If you have any questions about a task, please ask the experimenter before you label it.

1. In describing the electrical waves that are produced by the brain on an EEG, the text states that no alpha is found in infancy and that it is normal to find no regular brain rhythm in newborns.
   The task is:
   Rick's sister recently had a baby boy. EEG recordings of the baby's brain indicate no regular brain rhythm. This is:
   a. normal for a newborn.
   b. an indication of damage to the Broca's area.
   c. an indication of damage to Wernicke's area.
   d. a sign of retardation.

2. The text tells a student that companies are now selling devices to control alpha waves. These devices turn on a tone or some other signal that the subject tries to keep on for longer periods of time. The companies claim that these devices help to relax people.
   The task is:
   Upon entering a psychologist's office you hear a soft tone coming on and staying on for longer and longer periods of time. The psychologist is sitting quietly in his chair with his eyes closed, two electrodes attached to his head. What is he probably trying to do?

3. After reading a passage that describes adaptive instructional sequences and exemplifies both adaptive and nonadaptive instructional sequences, the students is asked:
   Give an example of a nonadaptive instructional sequence and then change the example so that it exemplifies an adaptive instructional sequence.

4. After reading a passage on prompting, the student is asked the following question:
   Select a complex behavior of your own that consists of at least three sub-components. Describe a general strategy to develop a complete, logical set of instructions to use as prompts in bringing the behavior under stimulus control. List two instructions that would begin your teaching sequence.

5. The student reads a section of a chapter on how to determine reliability. She is given the following novel task: To qualify for flight training, Minerva took a test in rules of take-off and landing. She scored 22%. She asked for another form of the test right away, claiming that she knew the answers. She scored an 85%. Is this a reliable test?

6. After reading a prose passage that defines the concept "tau effect" the student is asked the following:
   Give an original example of "tau effect."

7. After reading two passages that discuss effective reaction potential and momentary effective reaction potential, but that do not compare or contrast them, the student is asked:
   Explain the difference between effective reaction potential and momentary effective reaction potential. Do not simply define the two, contrast them.

8. The student reads a chapter on test making. She is then given the following task among other tasks:
   Select any essay item in this study guide. Say whether or not it is an effective essay question. List the reasons for it being effective or not.

9. A student reads a chapter on the brain. Within this chapter a number of pages are devoted to the results of damage to different areas of the brain. Specific examples are given of older people suffering from strokes and particular areas of the brain are identified with

behaviors affected by the strokes.

The student is asked the following question:

John's grandfather suffered a stroke several years ago. When John last visited his grandfather, the older man seemed to recognize John and smiled. The old man said a few words, none of which were intelligible, yet he seemed to think that he had communicated with John. What part of the brain was probably damaged by the stroke?

10. After reading a section on Piaget in an Educational Psychology text, the student is asked the following question:

    Provide an original example from your interest area that illustrates one of Piaget's contributions to education.

11. The teacher lists on the board the procedures that need to be followed in order to teach a complex behavioral claim which includes links not currently available in the student's response repertoire. Then the students are asked the following question: What 3 procedures must be followed in order to teach a complex behavioral chain which includes links not currently in the student's response repertoire?

12. After hearing a description of the concept "tau effect" that does not include any examples or nonexamples, the student is given the following task:

    Say which of the following are examples of "tau effect":

    a. Todd saw a picture of Cheryl Tiegs modeling a bathing suit in Sports Illustrated, one day in 1973. When he saw her on Cover Girl makeup ads two years later, he told his girlfriend that he had predicted that she would make it big some day. Later that day he saw her on a talk show, and commented on how lucky he was to be able to see her twice in one day.

    b. Cindy likes bikini bathing suits on men. While at the beach one afternoon last summer, she saw a guy with a blue bikini. Two weeks later, she saw another guy wearing a blue bikini, and pointed out to her friend Liz that she had seen the same suit two weeks ago and thought that it would look good on her friend Joe. When

she went to the beach again the next day, Liz pointed out the same blue bikini on someone else, but Cindy said the suit she liked is light green.

c. Mary and Rod went to see "For Pete's Sake," starring Robert Redford. The next day, they went to see "Butch Cassidy and the Sundance Kid." Rod pointed out that Robert Redford was in both movies. Three months later, they went to see "All the President's Men" which also starred Robert Redford. Mary asked, "Isn't that Robert Redford? Rod replied, "Nah, that's Paul Newman."

13. The student is given definitions of various behavioral procedures including extinction. Then, the teacher presents the following task:

Define the behavioral procedure of extinction. Be technical.

14. After the student reads a passage that defines and exemplifies positive reinforcement, he is asked to complete the following task:

Define positive reinforcement in your own words. Then, offer an original example of the use of this procedure in the classroom.

15. Again, the student is provided with definitions of various behavioral procedures. The teacher gives her the following task: Which of the following example(s) illustrate extinction procedures?

a) Peter constantly stands up and shouts out the answer whenever the teacher asks the class a question. His teacher feels that the reason Peter doesn't raise his hand quietly and wait to be called on is because he is very insecure and needs attention. To reduce Peter's out-of-seat behavior the teacher decides to make an extra effort to pay more attention to him.

b) Wendy is getting sick and tired of listening to her co-worker's constant flow of chatter during working hours. She repeatedly finds herself trapped in conversation, and not wanting to be rude, allows Celia (her co-worker) to continue. Wendy's work is getting behind because of this problem, so she decides to politely ignore Celia's chatter during working hours. She does

listen and talk to Celia however, but only during lunch breaks and after work.

c) Laurie's dog has an annoying habit of begging for food while Laurie is eating. Laurie usually ends up feeding half of her meal to the dog because he makes such a fuss. She decides to slap the dog lightly on the nose every time he begs for food.

16. After reading about different kinds of reinforcing events, the student is asked the following question:
In your own words, what are the definitions of primary and conditioned reinforcers?

17. The student completes a reading assignment on schedules of reinforcement. He is then given the following task:
Say whether CRF or Intermittent schedules are operating in the following examples. Explain your answer.

a) Holly goes door-to-door selling Avon. On the average, about one out of ten people she visits places an order.

b) Our dog Amos begs for food every time we sit down to eat. He looks so pathetic, that we always give him something.

c) Mr. Chance purchases at least ten different lottery tickets each week. In the past 12 months he's won small amounts of money on 5 separate occasions.

d) Every student in Ms. Pro's classroom receives one half hour free play time on Friday when they pass in a completed Spelling packet.

18. A student has just finished reading Bloom's taxonomy and a chapter on writing objectives. She is given the following task: Select an area of interest to you (e.g. tennis, foreign languages, psychology) and write an objective(s) for someone who is just learning about the subject. The objective(s) should show that you, as the teacher, are considering the three domains and three characteristics of behavioral objectives. (You may need to write more than one objective to cover the three domains.) Be sure to label each part and identify each domain. Make sure you example is original.

19. An educational psychology student has just read a number of arguments against Conant's view that Ed. Psych. can only offer obvious answers to questions. The student is then asked to: Offer an argument against Conant's view that Ed. Psych. can only offer the obvious. Given an original example of your argument.

20. The student has just finished a passage that defines, but does not exemplify the procedure of reinforcing incompatible behaviors. She is given the following task:
Which of the following example(s) illustrate the procedure of reinforcing incompatible behaviors:
    a) Debi's mother would like Debi to stop biting her nails. She praises Debi's appearance whenever Debi holds her hands at her sides or has her hands folded in front of her.
    b) A child is giggling during a lecture, and other children are responding to her interruptions. The teacher sits her in a chair that is visually restricted for five minutes each time she giggles.
    c) Bob walks around school all day with his shoelaces untied. His teacher sees this as potentially dangerous. The teacher praises Bob when his shoes are tied, and if possible sends him on a special errand. She does nothing when his shoes are untied.

APPENDIX

B

DATA SHEET
FOR
TYPOLOGY CLASSIFICATION STUDY

Subject Name _____ Researcher Name _____

Date _____ Time _____

Next to the numbers below, write the name of the type of task you think each corresponding task represents. The five task types are: copy task, define task, exemplify task, example identification task and combination task.

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.
13.
14.
15.
16.
17.
18.
19.
20.

Add any comments or other categories that are not represented by the five presented here:

APPENDIX

C

SAMPLE COMMENT AND SCORING SHEET

Researcher _____ Subject _____

Date _____ Program Type _____

Concept _____

Tape #'s _____

Task #	Duration	Correct	Incorrect

COMMENTS:

APPENDIX

D

STUDY BEHAVIOR QUESTIONNAIRE

The quantity and quality of a variety of distinct study behaviors have important effects on how much and how well a person learns from written material. In order to get some indication of your typical study strategies we ask that you study the attached pages and complete the following questionnaire.

Some questions should be answered while you study, and some should be answered after you study. All questions to be answered during study ask for the actual time you spend engaging in various kinds of studying. Each time you engage in any kind of study, indicate its duration. Then, total the times you listed for each question. For example, your answer to question #2 may look like this:

2. Please indicate the amount of time you spend rereading any of the material, to the nearest minute:

   2+3+1+1+5+4=16 minutes

Please read the questions thoroughly before you begin studying, so that you know what information we are looking for. And please do not radically change any of your typical study strategies! Have fun!

STUDY BEHAVIOR QUESTIONNAIRE

S# _____

_____
(name)

_____
(date)

During study:

1. Please indicate the total time it takes you to read the material for the first time, to the nearest minute: _____

2. Please indicate the amount of time you spend rereading any of the material, to the nearest minute: _____

   Indicate the amount of time you spend making any of the following written study aids, each to the nearest minute:

3. outlining: _____

4. underlining, brackets, or marking any other parts of the material:
   _____

5. summarizing: _____

6. terms, names, definitions, important points, etc.: _____

7. examples, applications, uses, etc.: _____

8. fill-in questions: _____

9. multiple-choice questions: _____

10. short or long essay questions: _____

11. other: _____

12. other: _____

13. other: _____

14. other: _____

15. other: _____

16. Indicate the amount of time to the nearest minute you spend orally quizzing yourself with any written study aids you made: _____

17. Indicate the amount of time to the nearest minute you spend silently quizzing yourself with any written study aids you made: _____

18. Indicate the amount of time to the nearest minute you spend writing answers to any written study aids you made: _____

STUDY BEHAVIOR QUESTIONNAIRE

19. Indicate the amount of time to the nearest minute you spend orally quizzing yourself without your written study aids: _____

20. Indicate the amount of time to the nearest minute you spend silently quizzing yourself without your written study aids: _____

21. Indicate the amount of time to the nearest minute you spend writing answers to any nonwritten questions you ask yourself: _____

22. Indicate the amount of time you spend filling out this questionnaire: _____

After Study

1. Please indicate your level of disturbance during the study episode by rating any visual, auditory, or other distractions:

1	2	3	4	5
low		medium		high

2. Please indicate the time of the entire study episode to the nearest minute:

_____

3. Please indicate your accuracy in filling out this questionnaire:

1	2	3	4	5
low		medium		high

APPENDIX

E

PRETEST

Read each question carefully. Answer each question as completely as you can. If you don't know the answer to a question, write DK (for Don't Know).

1. Define "constructional approach" in your own words.

2. Give an original example of "abulia."

3. Say which of the following are examples of "tau effect":

   a. Todd saw a picture of Cheryl Tiegs modeling a bathing suit in Sports Illustrated, one day in 1973. When he saw her on Cover Girl makeup ads two years later, he told his girlfriend that he had predicted that she would make it big some day. Later that day he saw her on a talk show, and commented on how lucky he was to be able to see her twice in one day.

   b. Cindy likes bikini bathing suits on men. While at the beach one afternoon last summer, she saw a buy with a blue bikini. Two weeks later, she saw another guy wearing a blue bikini, and pointed out to her friend Liz that she had seen the same suit two weeks ago and thought that it would look good on her friend Joe. When she went to the beach again the next day, Liz pointed out the same blue bikini on someone else, but Cindy said the suit she liked is light green.

   c. Mary and Rod went to see "For Pete's Sake," starring Robert Redford. The next day, they went to see "Butch Cassidy and the Sundance Kid." Rod pointed out that Robert Redford was in both movies. Three months later, they went to see "All the President's Men," which also starred Robert Redford. Mary asked, "Isn't that Robert Redford?" Rod replied, "Nah, that's Paul Newman."

   d. Teddy likes to buy crackerjacks. He gets the prizes and then gives them away to his friends. Once he got two blue secret decoder rings in the same box. He went running to his mother and told her that he got the same prize twice. He gave one of the rings to his friend Joe. Three weeks later he was visiting Joe and asked him where he got the blue secret decoder ring. Joe told Ted that he had given it to him. Joe said, "I thought I had given you a red one."

   e. For one homework problem, Lisa had to factor the equation $4X^2 + 6X + 4$. In class the next day, the students had a quiz on which that problem was included. She solved the problem, and

thanked the teacher for putting the homework problem on the quiz. Two weeks later, on an exam the same problem appeared again. After the exam was over, the teacher laughingly asked Lisa whether any of the problems looked familiar. Lisa said that she didn't think so.

f. Jacques saw the painting "Crows Over a Wheatfield" by Van Gogh, in his art history class, and then later that night, while thumbing through Newsweek, he told his friend Sharon, an art major, that he really liked it. Later that year, Sharon and he were walking through the Lincoln Center For the Arts when Sharon pointed out "Crows Over a Wheatfield." She asked Jacques to name the painting. He said, "It's called Country Spring."

4. Roberta is an excellent tennis player. She feels great after returning a difficult serve or hitting a shot right down the line. She is usually able to make a number of such excellent shots in a game. Recently, she has started to spend more time hiking, and has found it difficult to arrange time for tennis. She has gone from playing four times a week to once a week.

   Say whether the above is an example of abulia, constructional approach, or tau effect. If it is one of these three concepts, justify your answer. If it is not one of these three concepts, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

5. Give an original example of the "tau effect."

6. Explain the difference between effective reaction potential and momentary effective reaction potential. Be complete.

7. Define abulia in your own words.

8. Say which of the following are examples of the "constructional approach":

   a. Bob was a top student in highschool, although he rarely studied. Things are a little different now that he has come to college. Every night after supper, the guys talk him into playing cards. Often the card games lead to philosophical discussions, which he admittedly loves. As a result, he spends every night playing cards for about three hours, and then expounding upon philosophical issues for another three hours. This leaves little time for homework. After getting a 1.3 cum last semester, Bob decides to see his advisor for suggestions. Together they decide that since he likes philosophizing so much, it might be best for him to arrange his schedule so that he studies right after supper, and discusses philosophy only after he has completed at least 90% of his homework. Since this arrangement, he has handed in nearly all of his homework on time, and has gotten above 90% on all of his exams and quizzes.

   b. Arthur disrupted his third grade class by swearing a great deal. His teacher decided to have a talk with him and the principal. They concluded that the shocked face of the teacher and the giggles of the class were encouraging Arthur to swear. The teacher talked to the class and asked them not to giggle anymore. She informed them that if they giggled when Arthur swore, they would not be able to have recess. She also tried not to look shocked anymore. This program effectively reduced Arthur's swearing; he has not sworn in class for over eight weeks.

   c. Greg was mentally retarded and participated in a workshop to learn to assemble transistor radio parts. He spent much of his time distracting the other members of the workshop. They often enjoyed the distraction, and kidded around with him. This caused the supervisor to believe that the kidding around supported Greg's distractive behavior. The supervisor decided that Greg should only be allowed to kid around with the other participants if he was on task for 15 straight minutes. This proved to be much better for all concerned, as Greg quickly learned to work diligently for 15 minutes and kid around for 5.

   d. Mrs. Johnson thought that Shawn talked too much in class. She decided to put an end to this verbosity once and for all. So Mrs. Johnson had the principal call Shawn down to the office to run an errand. Then, she told the class that they were to stop attending to Shawn when she talked for more than two minutes straight, and when she talked more than three times within each class period. She asked them to make sure that they continued to acknowledge Shawn's comments up to these points. Before the

end of 4 weeks, Shawn was contributing to the class with maximum effectiveness.

e. Gail and Darcy liked to go for a drink after work on Fridays. They find this to be a time to unwind and catch up on gossip. They frequently went to a local singles bar, but were constantly bothered by men looking for a pick up. It seemed that 9 out of 10 times that they went to a singles bar, they would get approached by obnoxious men. A friend of theirs suggested that they go to a gay bar. Gail and Darcy found that at a gay bar they could still enjoy each others company without being harassed by other customers.

f. Frank spent much of his time isolated in a corner. This bothered his parents, who would like to see him interact more with other people. They decided to talk to a psychologist. He determined that Frank was afraid of other people calling him names and picking on him. He instructed Frank's parents to give Frank money toward a bicycle every time he interacted with others. Because the psychologist did not specify the exact nature of the interactions, Frank got money for calling people names and verbally assaulting them. For the last three months Frank has not sat in a corner once.

9. Give an original example that illustrates how stimulus generalization affects effective habit strength.

10. Kim was reading the newspaper the other day and noticed a little blurb about Carter's energy policy. The next day, she saw another article on Carter's energy policy, and told her friend Carl that there had been two separate articles in two days about Carter's energy policy. A month later, in a STATEMENT-OF-THE-UNION address, Carter expounded further on his energy policy. Kim told Carl how she was glad to hear that Carter has expanded his energy policy.

    Say whether the above is an example of abulia, constructional approach, or the tau effect. If it is one of these three con-epts, justify your answer. If it is not one of these three concepts, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

11. Georgia, a three year old attending a local preschool, had a history of whining when things didn't go her way. For example, one day when she was building a house out of blocks, she accidently knocked it over. She immediately started whining, and her teacher came right over. This sort of behavior had been going on for a long while. The teacher wondered if her coming over to Georgia when she whined, helped support this behavior. To find out, the teacher told Georgia to raise her hand and ask for help if she was having trouble. Otherwise, Georgia would not get any help. This system worked fine: Georgia does not whine, and the whole classroom atmosphere is much better.

    Say whether the above is an example of abulia, the constructional approach, or the tau effect. If it is one of these three concepts, justify your answer. If it is not one of these, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

12. List the factors that affect habit strength.

13. Define the "tau effect" in your own words.

14. Kim grew up in a rural setting where she would say hi to everyone she met, whether she knew them or not. People would always smile and say hello back. When she went away to college, at Hugh State University, however, she found that when she greeted people that she didn't know, not only didn't they return her greeting, but they also avoided eye contact with her. On some occasions they acted as if she was trying to pick them up. She rarely says hi to strangers now.

    Say whether the above is an example of abulia, the constructional approach, or the tau effect. If it is one of these three concepts, justify your answer. If it is not one of these, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

15. Give an original example of the constructional approach.

16. Say which of the following are examples of abulia:

    a. Graelle was an elephant at the Metropolitan Zoo. Her trainer decided to teach her to lift her trunk, grab a hammer, and hit a lever. At first, the trainer would give her a peanut every sixth time that she hit the lever. Graelle spent a great deal of time every day trying to get the peanuts. One day, the trainer decided to give her food everytime that Graelle hit the lever. She ate the first 100 peanuts, hitting the lever very hard every time. Then she began to slow down, and finally gave up.

    b. Andrea is reading a novel, July's Mixed Blessing, by H.W. Chartier. It is an adventure story about a 14 year old youth, Brent, who survives a canoeing accident in the wilderness of

Northern Maine. His partner was not so lucky. The book describes many of his trials and tribulations as he tries to make his way back to civilization. At first, these new and exciting perils occur in nearly every chapter, and Andrea can't put the book down. Then, in chapter 6, Chartier starts expounding upon the wildlife that Brent encounters. These descriptions continue for 5 more chapters. Andrea never gets to chapter 11.

c. Sharon is a transfer student in electrical engineering. At her previous school she hated doing research projects, mostly because it was so difficult to obtain all the books that she needed. But she kept at the projects, completing 8 papers in her first year. When she transferred to the State University, she was amazed because everytime she looked for a book, it was there. However, during her first semester, she completed only two of the four papers that she was supposed to do.

d. John bagged groceries at the local supermarket. He always remembered to put cans on the bottom and eggs on the top. He even put soap items separate from meat items. Customers often remarked about what a good worker he was. He was never late and was always willing to work overtime. He even worked Saturday nights and Sunday when asked. Within a month, John was promoted to work the dairy case. Fewer customers got a chance to comment on what a good job he did. His new manager was also not so liberal with the compliments. John was still never late and did a great job stocking the dairy case. Eventually he became assistant manager of the store.

e. When Karen and Bob first started going out together they rarely went to the movies. They found that when they did, the shows were almost always a disappointment. Recently, however, they went to a couple of movies and they found them both entertaining. Now they go to the movies at least once a week.

17. Sony is learning the names of different animals. His older sister is looking through a book with him and having him name the animals on each page. She shows him a picture of a cow and he says, "cow." She shows him a picture of a cow a moment later and he says, "That's the picture you just showed me, and it's a cow." The next day they do the same thing. When she shows him a picture of a cow he says, "Gee, I think it's a walrus."

    Way whether the above is an example of abulia, the constructional approach, or the tau effect. If it is one of these three concepts, justify your answer. If it is not one of these three concepts, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

18. Given an original example that illustrates the two variables that interact with habit strength to make a response occur.

19. Ralph always beat up his sister, Annie. He constantly hit her and bruised her somehow. Every time he did this, his mother came rushing over and told him to stop. Otherwise she was always off somewhere else in the house, either cleaning or talking with a neighbor. She became very worried when one day Ralph broke Annie's arm. She decided to seek some help at the local Family Counseling Center. She told a counselor about the situation. The counselor suggested that shy may be reinforcing hitting his sister by running up and attending to him. The counselor suggested two ways to deal with the situation: First, she should make an effort to attend to the children when they are planing cooperatively. Second, if Ralph does hit Annie, his mother should put him in his room for 5 minutes, and should not say anything to him while she is bringing him to his room or while he is in his room. Since that time, Ralph has hit Annie only twice, and has not hit her for the last five weeks.

Say whether the above is an example of abulia, the constructional approach, or the tau effect. If it is one of these three concepts, justify your answer. If it is not one of these three concepts, rewrite the passage to make it illustrate the concept to which it is closest. Then justify the changes that you made.

APPENDIX

F

Abulia

(uh-boo-lee-uh)

Abulia is a term used to describe low rates of behavior that are caused by an abrupt change from frequent to infrequent reinforcement. Sometimes the number of times that we must perform a behavior before reinforcement occurs is too large to keep us behaving. The result is a decrease in the frequency that we engage in the behavior. Freud called the resulting low frequency of behavior abulia. Abulia occurs when a continuous or dense occurrence of reinforcement changes abruptly to a more intermittent or leaner occurrence of reinforcement. We use the term abulia to describe only low rates of behavior that are caused by a change from frequent to infrequent reinforcement.

Tau Effect

(taw effect)

The tau effect is a name for how time affects an observer's judgement of whether two identical objects or events are the same or different. People will say that two identical objects or events are identical (the same), if the time between the occurrence of the events is short. If the time between the occurrence of the identical objects or events is long, people will say that the identical events are different. As the time between the occurrence of an identical object or event increases, the number of times people will say that the identical object or event is different also increases. This difference in judgement, affected by time, is called the tau effect.

Constructional Approach

The constructional approach is a relatively new way by which we can change the problem behavior of an individual. Currently, most methods for dealing with problem behaviors focus on eliminating or alleviating a distressing behavior. However, since all behavior, including distressing behavior, is maintained by consequences that are desirable to the individual, it is possible to employ these desirable consequences to strengthen other behavior that is not distressing. The constructional approach focuses on changing problem behavior by teaching or suggesting other behaviors that are followed by desirable outcomes or consequences. First, the constructional therapist observes and/or interviews the client to determine the behavior or class of behaviors that lead to consequences that are disturbing to the client or others. In addition, the constructional therapist determines the desirable outcomes that maintain the problem behavior in spite of other disturbing

outcomes. Then, alternative behaviors that are maintained by the same critical desirable outcomes as the distressing behavior, but are accompanied by satisfaction rather than distress, are taught or suggested. In addition, the individual learns to perform the satisfying behavior in the same situations that the distressing behavior occurs. This approach can be used by professionals with behavioral, psychoanalytic, or other therapy orientations. Through self observation and analysis an individual can also use this approach to change his own behavior.