1-1-1980

Measurement of the organization of memory.

John Anthony Bates
University of Massachusetts Amherst

MEASUREMENT OF THE ORGANIZATION OF MEMORY

A Dissertation Presented

By

JOHN ANTHONY BATES

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

DOCTOR OF PHILOSOPHY

September 1980

Psychology

MEASUREMENT OF THE ORGANIZATION OF MEMORY

A Dissertation Presented

By

JOHN ANTHONY BATES

Approved as to style and content by:

James M. Royer, Chairperson of Committee

Jerome L. Myers, Member

Ronald K. Hambleton, Member

Arnold Well, Member

Charles E. Clifton, Chairperson
Psychology

ACKNOWLEDGEMENTS

Special thanks are accorded to the following people for their assistance in the completion of this dissertation.

--Mike Royer, committee chairperson, advisor, and good friend, whose door was always open, whose support never wavered, whose criticism was always just, and who taught me by example what it means to be a good scientist

--my mother, who allowed me to dream great dreams, and who taught me to work hard to make the dreams come true

--my wife, Nancy, who sacrificed many of her own dreams to let me realize mine, who willingly shared both my successes and failures, and whose love and understanding were always freely given and gratefully received

--my son, Johnny, who tried very hard to understand why his Daddy spent so much time at school

--all the other members of my family, who encouraged and supported me throughout my graduate career.

ABSTRACT

Measurement of the Organization of Memory

(September, 1980)

John Anthony Bates, B.A., Youngstown State University

M.S., Ph.D., University of Massachusetts

Directed by: Professor James M. Royer

This dissertation described in detail three classes of memory organization models that incorporate varying degrees of declarative and procedural knowledge: concept, propositional, and schema networks. Next, the assumptions and characteristics of four popular techniques to assess this organization--word association, concept-similarity rating, tree construction, and card sorting--were discussed, and research pertinent to their application was reviewed. None of these techniques were found to adequately account for significant components of the theoretical memory models. Common methodological practices were also criticized for encouraging student performance based on something other than semantic memory structure, and for failing to demonstrate meaningful relationships between cognitive structure and student achievement. Alternative techniques were recommended for future research, such that practicable, more meaningful measurement devices

may be developed.

Two of these factors were incorporated in an original study to determine their importance to memory measurement: 1) effects of different orders of concept presentation on students' judgements about degrees of interconcept relatedness, and 2) utility of a prose passage to act as a context-setting schema for more consistent judgements about types of interconcept relations. Subjects were 30 undergraduate students enrolled in an introductory psychology statistics course, 26 undergraduate students with no prior exposure to statistics instruction, and 18 psychology graduate students who had completed two advanced courses on statistics. All subjects were administered a task comprising all possible pairs of 10 key statistics concepts, in both potential orders of concept presentation. Subjects were required to rate the degree of interconcept relatedness within each pair on a 7-point numerical scale. Undergraduate statistics students performed this task before beginning and at the completion of a 5-week sequence of statistics instruction. Other subjects were tested only once. During the postinstruction phase, half of the statistics students read a prose passage describing application of the statistics concepts in an experiment, prior to their engaging in the rating task, and the other half did not receive the passage. Half of both of the other

two groups similarly either did or did not see the passage. Subjects provided with the passage were instructed to base their judgements about the strength of interconcept relatedness on the rating task solely on the types of relations implied by concept usage in the passage. Performance of all subjects was evaluated against the performance of the statistics course instructor, who completed the rating task before beginning instruction.

Analyses of data indicated a reliable effect of instruction on memory organization. Statistics students, prior to instruction, did not differ from students who had never received training in statistics. After instruction, the statistics students were significantly more like their instructor in their ratings of concept relatedness, but did not reach the level of performance demonstrated by the graduate students. No effect on ratings was observed for order of concept presentation, and the possibility that this was due to the particular concepts selected for the rating task was discussed. No effect on mean performance was noted for the prose passage, and several expected relations between task performance and course achievement for the statistics students were not obtained. However, the prose passage appeared to affect the variability of student performance on the rating task, and one possible explication of this result was

presented. It was concluded that more research must be conducted before the importance of concept presentation order and the efficacy of a context-setting passage may be established. Further, the importance of demonstrating a reliable relation between performance on a structure assessment task and performance on standard achievement measures was emphasized, in order that such tasks may prove to be useful educational tools.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

To say that what we know is organized in some fashion within our brains appears so obvious as to be trivial. Indeed, the organized nature of human memory has been almost axiomatic in most discussions of human cognitive behavior, at least since the time of Plato and Aristotle (notwithstanding an interesting proposal by Landauer, 1975). Far less obvious, if the multitude of theories that have been advanced to delimit the components and procedures of this organization are any indicator, is how memory is organized (cf. Postman, 1972).

Despite the lack of consensus among psychologists regarding the exact nature of the structure of memory, various educational researchers have recently attempted to develop methods of graphically representing this structure. These efforts have been directed toward determining to what extent learners exposed to an unfamiliar content area are able to interrelate new concepts into a meaningful, functional structure. The basic notion has been that the more appropriate is a learner's structure, in terms of similarity to a theoretically optimum structure of the content area, the better that learner's performance will be on

tasks involving those concepts.

The importance of these efforts to educators is expressed by Anderson, Spiro, and Montague, who reflected on the common thesis of current cognitive psychology that "the knowledge a person already possesses is the principal determiner of what that individual can learn from an educational experience" (1977, preface). It follows that, if we possess a means of representing both the relevant concepts already known to a student and the manner in which these concepts are interrelated in the student's memory, then we can more effectively enhance that student's acquisition of new concepts in at least three ways. Instructional material may be internally structured so as to make maximum contact with and full use of the prior knowledge of the individual. Presentation of these materials may then be sequenced so as to better reflect the overall structure of the content area. Finally, regular assessment of structure during instruction may demonstrate the need for remediation, either to correct inappropriate or incomplete structural relationships among some subset of concepts, or to expand appropriate structures to encompass relevant concepts in other content areas.

The likelihood that such practical utility may be realized from any strategy to represent the structure of memory seems dependent on the validity of at least two assumptions. First, we must assume that there is a uni-

form system of representing and interrelating concepts in memory irrespective of the mode of concept presentation. Whether or not human memory comprises more than one representational system, and how multiple systems might be interrelated, are issues that have not yet been resolved (cf. Kosslyn, 1975; Kosslyn & Pomerantz, 1977; Pylyshyn, 1973). Indeed, J. Anderson (1978) has argued that an empirical resolution is highly unlikely. Instruction, however, frequently makes use of several modes of concept presentation (e.g., prose material, visual demonstrations, lectures, etc.). It may not be possible, and would surely be extremely difficult, to determine which of several modes used in an instructional sequence was most salient for a particular student when learning a new concept. Thus, if memory is multirepresentational in nature, then a method of representing structure that combines concepts learned via different modes may well generate a distorted picture.

A second assumption is that the structure of memory is sufficiently stable to permit meaningful measurement. That is, we must assume that structures, once established, maintain a relative degree of integrity, such that measurements taken at some point in time are reliable enough to allow reasonable predictions about future student achievement. This does not mean that cognitive structure must be conceptualized as being static. Indeed, one might point to

the apparent dramatic reorganization of knowledge that often accompanies a novel experience or exposure to a new analogy as evidence for a dynamic memory store. However, if structure reorganization is, in fact, very easily effected by seemingly inconsequential experience, then the resulting unreliability of structure measurement might seriously limit any potential educational advantages.

Strategies to measure the organization of memory clearly have the potential to be useful educational tools, provided the satisfaction of these two basic assumptions. It has been noted, however, that contemporary memory theorists are not in complete accord regarding how memory is in fact organized. Obviously, attempts by educational researchers have frequently not been at all explicit about the psychological underpinnings of their methods (cf. Glaser, 1979; Konold & Bates, Note 1; Perkins, Note 2). The purposes of this dissertation, then, are threefold. The first two sections of this chapter present a review of several recent models of the organization of human memory, and an examination of research approaches that have been most commonly employed in the measurement of cognitive structure, in order to make more explicit their relationships to or conflicts with these models. The third section delineates some important concerns that should be operationalized in future attempts to assess

cognitive structure if these methodologies are to both possess psychological validity and be of practical utility in the classroom. Included in the final section of this chapter is a description of a study conducted to investigate several of these concerns. The remaining three chapters provide the methodology and results of this inquiry, as well as a discussion of its outcomes.

Theoretical Models of Memory Organization

Winograd (1975) has provided a continuum onto which representational systems of memory structure may be placed, that describes the criticial features of such systems, and that allows an examination of their relative advantages and disadvantages. At one end of this continuum are systems that emphasize the declarative nature of memory. These approaches typically regard memory (specifically, semantic memory) as comprising an interrelated network of concepts, each concept in some way representing the salient features of an object or event in the real world. Thus, what is stored in memory is primarily a set of facts (as determined by individual experience), as well as how facts are related to other facts. The thought processes are applications of separately stored rules of logic, or some other general procedures, to some subset of facts for the purpose of

making deductions, but these processes do not directly affect the nature of the facts themselves. This system parallels the notion of knowing that, as described by Broudy (1977) and others.

A declarative representational system may offer several advantages. Such a system may provide some amount of economy in storage, because facts that can be used in a wide variety of contexts need only be stored once. If every new encounter with a previously learned fact in a different context were required to be stored separately, it seems likely that this would ultimately tax the capacity of any finite memory store. Of course, storing facts only once would require a considerable increase in the number of interconnections among facts necessary to adequately express fact usage in various potential contexts. Still, if each use of every fact had to be learned (rather than deduced), this inflexibility might severely limit any transfer of learning, especially the far transfer and figural transfer described by Royer (1979).

In contrast, the other end of Winograd's continuum emphasizes the procedural aspects of memory. The structure of knowledge, from this perspective, amounts to a set of specific operations (rather than facts) that allow the human organism to interact effectively with the world. Perhaps the most compelling argument in support of a

procedural organization of memory is the (apparent) existence of general heuristics that guide human behavior in many situations. Examples may include knowing how to act and what to expect when eating in an expensive restaurant, making judgements about the relative likelihood of potential events, and responding appropriately during an informal conversation with a close friend. Albeit that some of these heuristics have been demonstrated to be inappropriate (cf. Tversky & Kahneman, 1974), their very use suggests that the structure of memory includes more than a network of elaborated facts.

The following three subsections of this dissertation review several psychological models of memory organization that seem to inhabit different positions on the declarative/ procedural continuum. It should be noted, however, that none of these models could justifiably be labeled as purely declarative or procedural. All include or permit the inclusion of components to account for both aspects of memory phenomena. Their differences lie primarily in the relative priority ascribed to these contrasting features. In the first two subsections, models are discussed that emphasize the storage of interrelated facts, either as a hierarchical network of concepts, or as sets of semantic propositions. Finally, several recent perspectives with characteristics somewhat reminiscent of the precepts of Gestalt psychology

are reviewed. These approaches argue, to a greater or lesser extent, for the existence of distinct units of knowledge (generally termed _schemata_) that possess certain emergent properties not appreciable from a more molecular analysis of structure. Rather, emphasis is placed on those aspects of each model that seem especially relevant to the issues of whether and how research procedures to assess memory structure have been consonant with psychological theory.

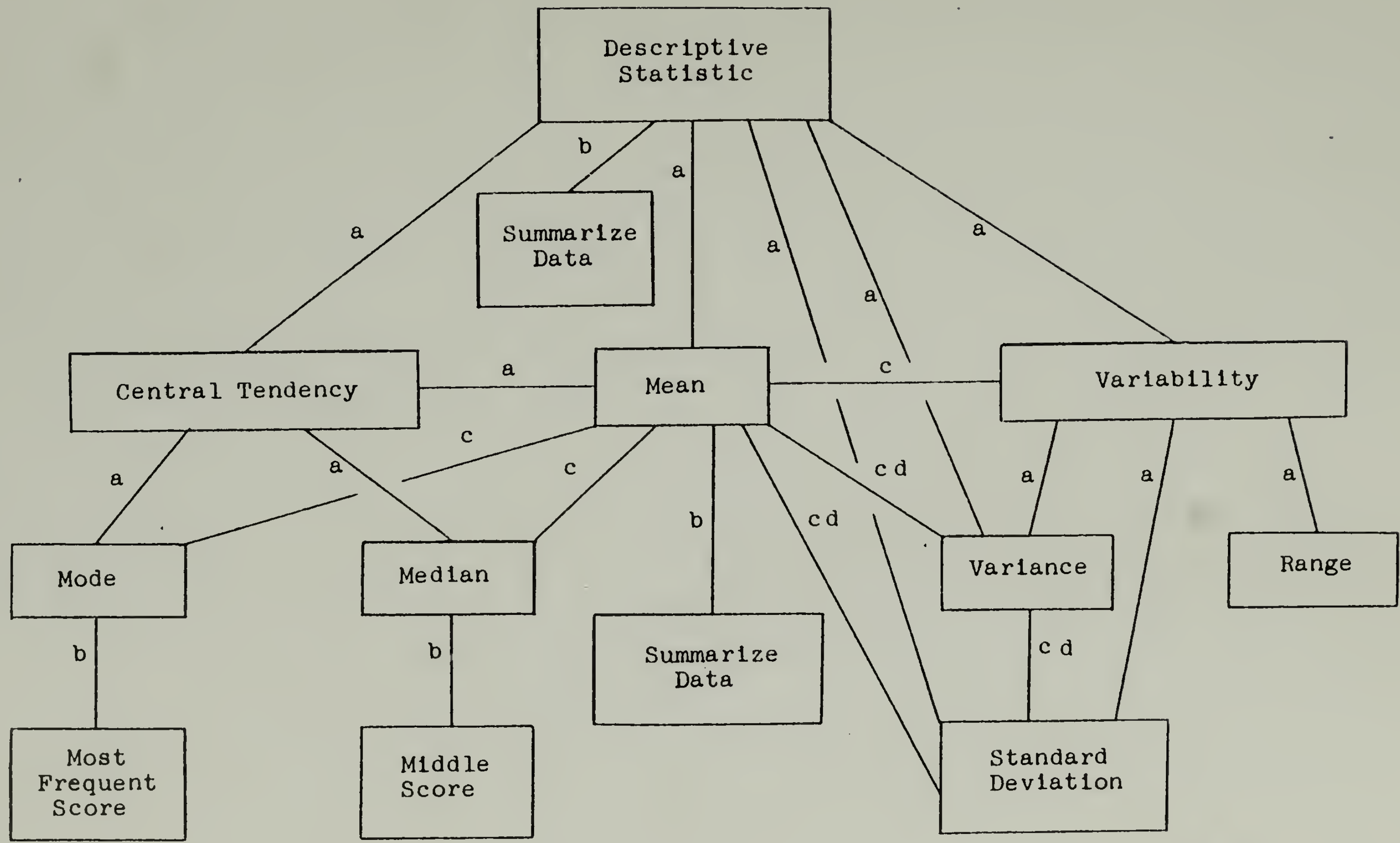
Concept network models of memory. The outstanding example of this class of declarative models is the semantic hierarchy theory first outlined by Quillian (1962, 1967, 1969), and later expanded by Collins and Quillian (1969, 1972). The original intent of Quillian's effort was to derive "a strategy for dealing with language in a computer" (Collins & Quillian, 1972, p. 348), and it was therefore somewhat restricted in its treatment of human memory phenomena. Collins and Loftus (1975) have recently clarified many of the ambiguities that were sources of criticism in the earlier form of this model (e.g, Anderson & Bower, 1973; Smith, 1976), and have attempted to align the model more closely with current psychological thought.

There are two fundamental units of information within memory, according to this model: the concept and the relational link. These units are illustrated in Figure 1, which

presents a much-simplified network of several concepts that would typically be included in an introductory psychological statistics course. Concepts (i.e., *descriptive statistic*, *mean*, *variance*, etc.) are represented as labeled nodes, and the lines interconnecting these nodes are bidirectional links. It should be noted that the labels for each node are not thought to be contained within semantic memory *per se*, but are rather stored separately in a lexical network of phonemic/orthographic information, and are themselves linked to their semantic counterparts. Indeed, concepts need not have any formal name in conventional language, as for example, the concept of "getting out of bed early on a cold Monday morning."

The relational links between concepts have two important properties. First, they indicate the types of relations learned between concepts. Collins and Loftus (1975) have postulated several general classes of links, although it is implicit in their discussion that many different kinds of links may exist in memory as there are observable relations among things in the world. This is because links themselves may be thought of as concepts (Collins & Loftus, 1975, p. 409). Four of the more obvious link types are illustrated in Figure 1. Most important of these is the superordinate/subordinate or *isa* link, which is the basis for the hierarchical structure of the model. The *isa* link

Fig. 1. Hypothetical Semantic Hierarchy for Basic Statistical Concepts.



[a]Superordinate/subordinate (isa) link

[b]Property link

[c]Conjunctive link

[d]Operational link

develops when it is learned that a concept is a member of a larger or superordinate class of concepts. Such links might normally develop, in the example, between central tendency and variability (the subordinate concepts) and descriptive statistic (the superordinate concept). Other potential isa links are also illustrated. The links between mode and most frequent score, and between median and middle score, indicate properties possessed by modes and medians. The link between mean and variability indicates that these concepts are frequently encountered in conjunction, in much the same manner as expressed by the notion of associative contiguity. Finally, variance and standard deviation are linked by an operation, in that the latter is the square root of the former.

The second property of links is that they indicate the criterialities of the relations among concepts. Criteriality is the importance of a given bidirectional link to the meanings of each of the linked concepts, and it may vary in value dependent on the direction traveled. For example, it may be more important for a statistics student to know that the range is a measure of variability than it might be to know that one of the measures of variability is the range. Within this model, a given concept is defined in terms of all the other concepts, superordinate and subordinate, with which it is linked; and the type and criteri-

ality of each of those links. The degree of relatedness between two concepts is not solely a function of the semantic distance between them (i.e., distance along the shortest path of intermediary links and nodes), as has occasionally been inferred by critics of this approach (e.g., Anderson & McGaw, 1973; Rips, Shoben, & Smith, 1973). Rather, concept relatedness, like concept definition, is an aggregate function of number, type, and criteriality of links.

Two additional related misinterpretations (and consequent sources of criticism) of this model are that the hierarchy imposed by isa links requires the "erasure" of subordinate concept properties once they are learned to be general properties of a subsuming superordinate concept (Anderson & Bower, 1973; Conrad, 1972), and that relations between a subordinate concept and a superordinate several levels removed can only be inferred via indirect intermediary links (Smith, 1976; Smith, Shoben, & Rips, 1974). This notion that their hierarchical model necessitates a strong theory of cognitive economy (i.e., each concept stored once and only once in memory) has been explicitly denied by Collins and Quillian (1969) and, again, by Collins and Loftus (1975). Instead, these researchers have argued that both superordinate properties and semantically distant superordinate concepts may be directly linked to subordinate concepts if those particular relations are

learned to be especially salient.

For example, Figure 1 indicates that the concept of mean is a subordinate of the concept of central tendency which is in turn subsumed under the superordinate concept of descriptive statistics, which is linked to a salient general property of summarizing data. However, mean is also directly related via an isa link to descriptive statistic, indicating that the possesser of this hypothetical knowledge structure has learned that means are themselves important to understanding descriptive statistics. In addition, the property of summarizing data has been learned to be directly related to the use of means, and need not be inferred via intermediary links (whereas such inference, because of a different instructional emphasis, may be necessary to come to the same conclusion for range). The critical variable determining whether superordinate concepts and properties are linked directly to subordinate concepts or must be inferred, then, is how these relations are learned, and not the positions they hold in an idealized taxonomy.

Several variations of the hierarchical network model of memory have been proposed. Glass and Holyoak's (1975) marker-search model also incorporates the notion of interconnected superordinate and subordinate concepts and properties. However, it differs from the Collins and Quillian (1969, 1972) and Collins and Loftus (1975) approach in at

least two ways. First, salient superordinate properties are not thought to be directly stored with subordinate concepts, but are rather accessed via "short-cut" links (called markers) from subordinate to superordinate. Thus, the marker-search model provides for stronger cognitive economy than does the previous model, since superordinate properties are stored only once, but it achieves this economy via a proliferation of markers between superordinate and subordinate concepts. The second difference between these models is that Glass and Holyoak's markers, although labeled to indicate type of relation, are not weighted to indicate relative property salience. This is in direct contrast to Collins and Loftus' (1975) explicit notion of link criteriality. However, the fact that short-cut markers are assumed to develop among nonadjacent concepts, through natural co-occurrence frequency, allows this model to explicate observed inverse relationships between production frequencies of concept pairs and reaction times to confirm these same concept relations (e.g., Glass, Holyoak, & O'Dell, 1974; Holyoak & Glass, 1975).

Another version of the concept hierarchy has been offered by Meyer (1970, 1973, 1975). His predicate-intersection model reflects the declarative notion of stored concepts embodied in the previous two models, but it also includes additional processes for confirmation or disconfirmation of concept relatedness. According to this

approach, concepts are stored in the form of superordinate categories and subordinate exemplars. Certain of these relations may be determined simply by affirming the existence of links between category and exemplar; however, some potential concept relations are not stored and must be computed. For example, the statement, "Some medians are measures of central tendency," could be confirmed by verifying the link between these concepts. In contrast, the statement "All medians are measures of central tendency," could be confirmed only by first verifying intersecting links, and then by comparing lists of defining attributes assumed to be attached to both concepts. This second step in decision-making was judged necessary by Meyer (1970) to account for observed longer verification times when subjects are presented with sentences of the form, "All X are Y," than if presented with sentences like, "Some X are Y."

Other memory models have also been proposed that have certain declarative features, but do not posit explicit hierarchical concept relations. The feature-comparison model developed by Smith (1976) and his associates (Smith et al., 1974) represents each concept as a set of attributes (features) that vary according to their salience as defining properties of that concept. Features with high salience are those that are learned to be essential for a concept's membership in a given category. Less salient features indicate

variable, general characteristics that provide a framework for creating concept prototypes, as that term has been used by Rosch (1977). Thus, a defining attribute of descriptive statistic might be the property of being a summative statement. A characteristic attribute of the same concept might be the property of being a real number, which would be a component of a concept prototype, the arithmetic mean. The feature-comparison model does not provide for concept interconnections in the form of links or markers. Instead, this approach requires that all concept relationships be computed on the basis of attribute comparisons.

A similarly nonhierarchical model has been proposed by Schaeffer and Wallace (1970), who conducted a series of experiments measuring the reaction times necessary to disconfirm the relatedness of semantically similar and dissimilar concept pairs. For example, subjects were asked to judge as same or different pairs like hemlock-daisy (semantically similar) and hemlock-parrot (semantically dissimilar). Because reaction times were found to be consistently longer to disconfirm semantically similar pairs, these researchers suggested that concept relatedness may be determined by an exhaustive comparison of concept attributes, rather than by any arrangement of concept nodes and interconnecting links or markers. That is, the many common attributes of daisy and hemlock slow the accumulation of

evidence against concept relatedness, whereas the far fewer common attributes of hemlock and parrot allow such negative evidence to accumulate more rapidly. Schaeffer and Wallace suggest that their model to account for these data "does not argue for a particular memory organization" (1970, p. 151). However, their approach does argue against the hierarchical network system of organization proposed by Collins and Quillian (1969, 1972) and Collins and Loftus (1975), as well as the marker-search model of Glass and Holyoak (1975).

Each of the above models of memory organization has as its most basic unit of information the concept. Each has interpreted concepts as being the semantic equivalents of words (usually nouns) or short phrases, incorporating attributes, features, or properties learned to have salience for concept meanings. The major difference among these models, for the purposes of this discussion, has been the nature of the organizational system in which concepts are embedded. Although the notion of a hierarchical arrangement appears to have the greatest appeal, possibly because of its logical structure and potential for cognitive economy, other, nonhierarchical systems have also been proposed. The next subsection reviews another general perspective on memory organization that incorporates a different, more molar unit of information, the proposition.

Propositional network models of memory. One major deficiency of the previously described memory models is that concepts, as they have thus far been defined, do not seem to be comprehensive enough units of information to capture the complexity of what must be a major portion of the human memory store: that which is learned via language, either verbal or prose. For example, a typical textbook definition of the standard deviation is, "The square root of the mean of the squares of the amount by which each case departs from the mean of all the cases" (Hilgard, Atkinson, & Atkinson, 1979, p. 606). Such a concatenation of concepts might decompose quite nicely into the concept hierarchy presented in Figure 1. However, it seems reasonable to suppose that the time spent and the methods employed by an instructor in explaining all the ramifications of these notions to a group of students will lead to the production of mental representations far more complex than a concept hierarchy.

Most models of memory organization advanced in the last seven or eight years have employed propositions as the basic memorial units to capture the richness of learning through language usage. A proposition may be thought of as an abstract conceptualization that represents the gist of relations among two or more concepts, where concepts are dictionary-like definitions of objects or actions. Whereas concepts

are always true (whether or not, in an absolute sense, they are correct), propositions may be evaluated, on the basis of previously stored propositions, as to their veracity. Also, propositions are conjoined through common concepts to form an elaborated network of conceptualizations in memory. Just as models of concept networks have been developed that are either hierarchical or nonhierarchical in structure, so are there differing models of propositional organization. Four of the more explicit of these propositional models will next be discussed in some detail.

In their recent book describing the current anthropological evidence of human evolution, Leakey and Lewin (1977) suggested that intelligence is a function of an organism's ability to mentally integrate its perceptions of its environment. Human language, they have speculated, has been a (or, perhaps, <u>the</u>) major factor contributing to the rapid advance of the species because of its unique capacity for representing perceptual information in a coherent abstract form. In much the same vein, J. Anderson and G. Bower have argued that language "permits men (sic) to exchange their experience verbally, to inform one another, to reinforce or punish or question one another, and generally to enjoy the many fruits of a technology for communicating with one another. ... (It) eventually facilitates the development of abstract conceptual structures that appear far removed

from a description of immediate perceptual experience. ... (It) plays a central role in our capacity for abstract thought" (1973, p. 155). Consequently, these researchers have developed a model of human associative memory (HAM) intended to have the capacity of expressing any idea that a human could formulate or understand--a system fully capable of the rich expression of Human Language, without being tied to the peculiarities of any particular language.

The most basic unit of information in HAM is the atomic proposition, which represents a single relation between two nodes. Atomic propositions are combined via relational links to form larger propositions that represent complete conceptualizations, or statements, and are stored as complete structures in memory. The links composing these complex propositions provide information regarding the nature of concept relations in the form of a predicate calculus specifying subject, predicate, object, context, time, etc. The nodes may be either of two kinds: concept or individual. Concept nodes refer to nonspecific, generic concepts stored in a definitional lexicon. Individual nodes refer to specific objects, events, or properties that are learned through personal experience.

An example is presented in Figure 2 of a simple propositional representation of an instructional statement regarding the origins of notions about normal distributions,

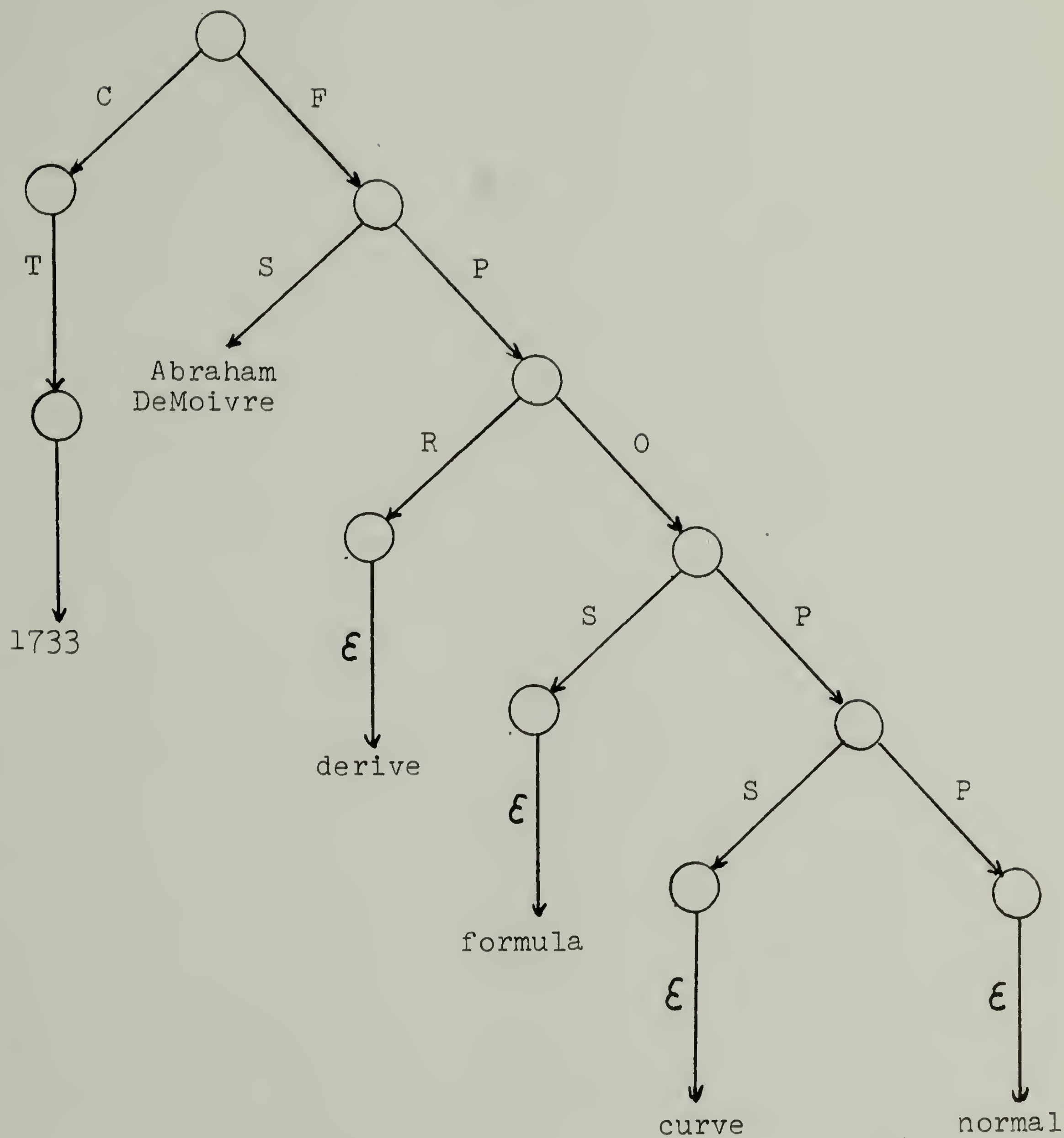


Fig. 2. Propositional Representation of the Statement, "In 1733, Abraham DeMoivre derived the formula for the normal curve."

as it might appear in HAM. Here, the year 1733 and the name Abraham DeMoivre are represented as individual nodes because they are the sole members of specific concept sets (unless, of course, a student is familiar with more than one 1733 or Abraham DeMoivre). Derive, formula, normal, and curve are represented as subsets (ε) of concept nodes because they are instances of more general categories of things.

Two characteristics of HAM are especially important to an evaluation of current memory structure measurement strategies. First, concepts by themselves are not thought to be the critical units of meaning in memory. Even atomic propositions are not considered to be complete, meaningful representations because they may specify relations that exist only within a context not apparent without reference to the complex propositions within which they are embedded. The second imporant feature of HAM is that, whereas a hierarchical arrangement of nodes is implied by the use of subset/superset relations, propositions are not thought to be hierarchically organized. Rather, propositions enter memory as production rules (via predicate calculus) that permit recreation of the gist of verbal or prose statements. Although these production rules are interlinked in an elaborated network, there is no provision in the model for denoting propositions of greater or lesser importance, other than the associative strength of their links as derived from fre-

quent use.

Because HAM did not fully satisfy the goal of representing any conceivable human thought or utterance, J. Anderson (1976) has developed another propositional model of memory, called ACT. Among the differences between HAM and ACT are several changes in representational formalisms, including the deletion of redundant relations and the addition of relations that would permit propositions unrepresentable by HAM, and the application, in ACT, of a diffuse search procedure rather than HAM's serial search. Also, ACT includes a complete network of productions that are thoroughly integrated with the propositional network. These productions are explicit representations of procedural knowledge, and serve as algorithms for the retrieval of information from memory for the purposes of problem solving and inference making. Productions may interact with the structure of the propositional network to the extent that individuals are biased to rely on particular problem-solving strategies, thereby strengthening (via use frequency) particular propositional relations.

The ACT model has been both praised as the most complete and explicit model of memory representation extant (Gagné, 1978), and criticized as being so comprehensive in its post hoc ability to accommodate data that it is empty of predictive validity (Wexler, 1978). Whatever the final evaluation of

ACT's merits and deficits, its propositional representation system is sufficiently similar to that of HAM for the purposes of this paper. Knowledge is stored in memory as an elaborated network of complete conceptualizations, rather than as a network of concepts. Further, these ideational units are not organized in a hierarchical fashion.

Another variant of the propositional network model has been provided by Rumelhart, Lindsay, and Norman (1972), and modified and extended by Norman (1973) and by Rumelhart and Norman (1975). This model, called the active structural network (ASN), is similar to HAM and ACT in some respects. Like both these previously discussed models, ASN represents declarative knowledge as a nonhierarchical network of abstract conceptualizations. Nodes representing concepts within a proposition may be either of two kinds: <u>type</u> nodes referring to generic concepts, or <u>token</u> nodes referring to specific instances of those concepts. Like ACT, ASN also incorporates procedural knowledge, but it does so in a manner different from Anderson's (1976) model. Whereas ACT has separate, although integrated, networks for propositions and for productions, ASN makes no representational distinction between these types of knowledge. That is, procedural and declarative knowledge are thought to be inseparable and thoroughly interactive. This integration of declaratives and procedures is achieved, in ASN, primarily through a very

explicit and comprehensive reliance on the decomposition of all statement predicates into semantic primitives. These primitives capture the essence of a relation among concepts, and specify the nature of concepts necessary to construct a complete conceptualization. For example, the sentence graphed in Figure 2 has as its predicate the past tense of the verb "derive." This verb may decompose into a primitive element within memory, DO. This primitive requires, to complete a proposition, the relationship of concepts specifying actor, object, and time. Thus, the original sentence, according to ASN, might be partially represented in memory as DO [Abraham DeMoivre, formula, 1973].

The notion that primitives direct the nature and number of concepts that may be interrelated within a proposition provides ASN with a much more dynamic memory store than has been suggested by previous models. As Rumelhart and Norman (1975) have put it, "...a sentence does not exist in memory after it has been interpreted; rather, the sentence is used to provide instructions as to how to modify the structures of memory to convey the deep, underlying components that comprise meaning" (p. 56). In the context of this paper, such a dynamic system may have serious implications for applications of measurement techniques, given the assumptions specified in the previous section.

The final propositional model to be discussed has been

formulated by Walter Kintsch (1972, 1974). Kintsch's model is intended to represent the meaning of entire prose passages, rather than individual sentences, and it differs from HAM, ACT, and ASN in many respects. Concepts are stored in a definitional lexicon, and are thought to be therein linked to a minimum of defining properties. Each proposition is represented as a predicate concept and its related arguments, which may be concepts or other propositions. Instead of organizing propositions into a semantic network, Kintsch has postulated that prose and discourse information, at least, may be represented by a text base that is a linear arrangement of propositions, ordered according to their importance to the meaning of an entire passage.

The relative importance of propositions in this hierarchy depends upon the amount of argument repetition found within a particular passage. Propositions containing the most frequently repeated argument are assigned the primary proposition in the text base when it is stored in memory. Less frequently repeated arguments relegate their propositions to lower positions in the text base hierarchy. Kintsch (1974) has provided some evidence in support of a proposition hierarchy by demonstrating that individuals are more likely to recall conceptualizations from prose material that have much argument repetition than they are to recall conceptualizations with little argument repetition. This result has

been supplemented by McKoon (1977), who demonstrated that text sentences containing frequently repeated arguments are recognized more quickly and accurately, after a retention interval, than are text sentences containing less frequently repeated arguments.

Two other features of Kintsch's model are especially worthy of note. First, Kintsch has rejected the use of semantic primitives, primarily on the grounds that the degree of decomposition of any word is purely arbitrary, and that human language is too complex to be based on a few primitive meaning elements. Kintsch (1974) has presented empirical support for these contentions by demonstrating that semantically complex words are no more difficult to recognize than are semantically simple words. Second, Kintsch has argued that the definitional lexicon is not absolutely precise. Concept meanings, according to Kintsch, "are something given to (a) word by its use in a particular context. Therefore, it may not be necessary to specify each word precisely in the lexicon, if one can show how a particular meaning could be elaborated on the basis of a given context" (1974, p. 10).

The propositional models of memory organization thus far reviewed are similar in their treatment of declarative knowledge as being more akin to interrelated, complete conceptualizations than to definitional concept networks. One may also note the increasing importance placed on memorial representa-

tions of procedural knowledge, especially in the systems proposed by J. Anderson (1976) and by Rumelhart and his associates (1972, 1975). The organizational models reviewed in the following subsection, although frequently incorporating propositions as their fundamental units of meaning, place an even greater emphasis on the procedural components of memory.

Schema models of memory. R. C. Anderson (1977) has reported a recent experiment in which subjects, who were either physical education or music students, read two prose passages and were tested for their interpretations of the passage themes. One passage had the dual potential themes of either a wrestler preparing to break a hold of an opponent, or a prisoner planning an escape. The other passage could have been interpreted as describing a group of people preparing to play cards, or as a rehearsal of a woodwind ensemble. Although the two groups did not differ in the amount of information later recalled about the passages, there were differences in ascribed meaning. Most of the physical education students judged the first passage to have a wrestling theme and the second to have a card-game theme. Most of the music students thought the first passage referred to a prisoner, and the second to a music rehearsal.

The results of Anderson's study, and many others of a similar nature, have been interpreted as evidence for the

existence of a more complex memorial system of representation than one might infer from concept or propositional models--a system that simultaneously provides the human organism with a more-or-less stable knowledge base, a method of inquiry to expand that base, and a general perspective for understanding environmental events. The label most commonly given to the units composing this memory system is schema (plural, schemata), as taken from Bartlett's (1932) studies of story recall. The notions embodied in that term are also generally characteristic of other frequently used labels, including frames (Kuipers, 1975; Minsky, 1975; Winograd, 1975), and scripts (Schank, 1975; Schank & Abelson, 1977).

A representative example of schema models has been provided by Rumelhart and Ortony (1977), and is a direct extention of the active structural network model discussed in the previous subsection. Like ASN, Rumelhart and Ortony's schema model makes use of a propositional representation of semantically primitive concepts as its atomic unit. Schemata are sets of atomic units that have become interrelated via links and predicate primitives, and which represent more complex generic concepts. For example, Figure 3 is a simplified representation of a schema indicating that "X" is a normal curve if it is both bell-shaped and symmetric. Of course, although these two properties of normal curves are repre-

sented in the example as atomic elements, they may well be decomposable into still more primitive subschemata.

Schemata, according to Rumelhart and Ortony (1977), have at least four important characteristics. First, they contain variables, or roles, that may be filled by different concepts at different times, much as a given predicate primitive in ASN requires that certain concepts be related to complete a proposition. These variables may take on any of a variety of values, dependent on environmental circumstances, provided only that the same relations exist among these values as are specified by the schemata. Second, schemata can embed within each other. In Figure 2, the dominant schema of NORMAL CURVE is a concatenation of the subschemata SYMMETRIC and BELL. Thus, one may comprehend an event by applying an appropriate schema as a whole, or, if need be, investigate the particulars of the event based on the components of that schema. That is, it is not necessary, although it is possible, to store or retrieve these structures separately. Third, as implied by the notion of embedding, schemata vary in a hierarchy of abstraction. Not only do schemata exist that represent simple concept relations, but there are also higher-order schemata that represent complex action sequences or prose organizations. Finally, schemata represent encyclopedic rather than dictionary knowledge. That is, they represent what is typically true, within limits imposed only by experience,

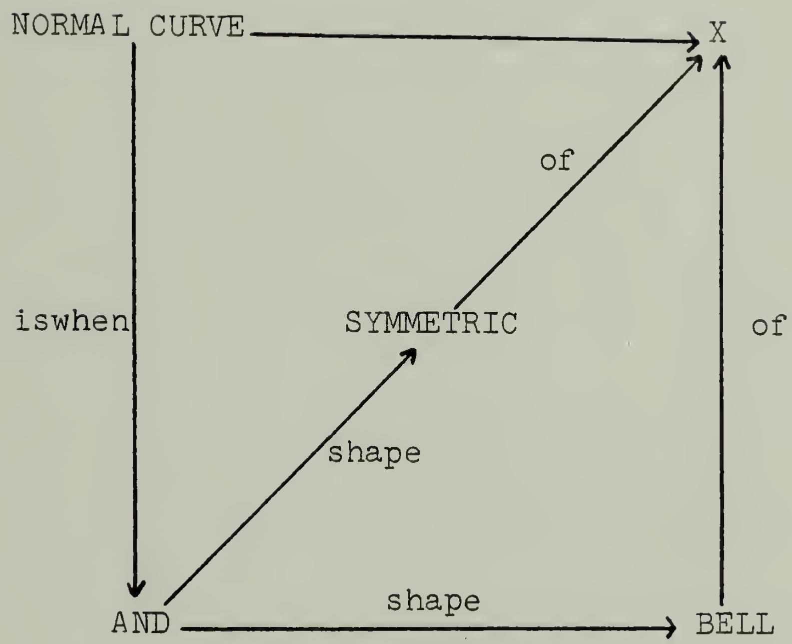


Fig. 3. Schema Representation of the Concept, "Normal Curve."

rather than what is necessarily true.

Rumelhart and Ortony's schema model is clearly the most dynamic conceptualization of memory organization thus far discussed. Schemata are viewed as able to be both constructed and generalized according to the dictates of personal experience. That is, schemata may be created out of a new concatenation of primitive elements if no available schemata are capable of subsuming some previously unencountered arrangement of variables. Also, if several separate schemata come to include, through experience, very similar primitive relations, then they may become interrelated to form a single, more abstract schema that describes the gist of each of its components. Rumelhart and Ortony (1977) provide the example of a generalized BREAK schema, created by links among and describing the common abstract properties of breaking a window, breaking a bubble, and breaking a promise. Schemata are thus thought of as providing a framework for understanding new events by abstracting relevant relations from the multitude of prior events. Moreover, they direct the perceptual search of the environment according to the variables and values that their structures comprise.

Additional interpretations and applications of the schema approach have been offered by several researchers. Schank (1975) and Schank and Abselson (1977) have theorized that information is represented in memory as <u>conceptualiza-</u>

tions, in the form of linked objects or property concepts and action (verbal) primitives. These conceptualizations are further linked in a linear arrangement, called an episode, that describes a cause-and-effect chain. In prose learning, episodes may be thought of as abstract representations of paragraphs, tracing the progress of plot and development of theme. Certain episodes may come to describe frequently occurring experiences, such as a visit to a doctor, attending a party, or driving a car. Such episodes are called scripts, and have properties as implied by that term: they specify appropriate roles, actions, and purposes. Like Rumelhart and Ortony's (1977) schemata, scripts permit comprehension of what otherwise might be ambiguous information by supplying a contextual framework. They also allow inference-making about missing information based on their component conceptualizations, and direct behavior when an individual encounters a familiar situation.

Kintsch and van Dijk (1978) have extended and somewhat modified Kintsch's earlier (1974) propositional model, including in their approach another perspective on schemata. They consider the arrangement of proposition in a memorial text base to be a microstructure containing a literal, albeit abstract, representation of discourse or prose material. A given microstructure may be collapsed into a macrostructure representing the gist of learned material, by the appli-

cation of certain rules for proposition inclusion or deletion. The agent that directs macrorule application is the schema. Not only do schemata determine which propositions in the microstructure are relevant to the gist of the text base, but they also direct retrieval of information according to contextual demands. That is, schemata respond to the nature of a task by directing retrieval of whatever subset of the original text base may be necessary. The composition and structure of schemata in this model are not clear, although a description of their use implies an arrangement of abstracted, generic knowledge about familiar events.

The schema construct, as it has been operationalized in these models, has several implications for educational research and practice. The notion that schemata provide anchors for the comprehension of new information may explicate a variety of reported prose and visual comprehension phenomena. For example, Brandsford and McCarrell (1974) have demonstrated the powerful effects of knowledge of context both on subjects' interpretation and on their degree of understanding of pictures and of prose passages. The activation of different schemata in memory, dependent on which context is ascribed to a situation, may account for such effects. Efforts have also been made to apply the schema notion to observed differences in the understanding and recall of various types of

prose structure (e.g., Bower, 1976; Rumelhart, 1975). Another important aspect of these models is that schemata, once activated by particular task demands, are thought to direct the search for and determine the appropriateness of information available in the internal and external environments. Schemata, in a sense, are viewed as providing the purpose for both the encoding and the retrieval processes (cf. Bobrow & Norman, 1975). That is, a given schema may be activated because of the presence of certain stimuli in the environment that satisfy the requirements of some of that schema's variable values. The schema may then direct further search of the environment in order to fill remaining variable slots, or it may be accessed for the retrieval of inferred information in the form of most probable variable values, where probability is determined by the frequency of prior encounters with those values in similar contexts. In this respect, schemata act as heuristic devices for comprehension and problem solving. This notion is consistent with Bruner's (1961) concern that formal education should strive to develop within students certain heuristics for the accumulation and application of knowledge.

As has been noted, these characteristics of schema models suggest a more dynamic memory store and a more thorough integration of declarative and procedural knowledge than any other of the reviewed approaches to the organization of memory. Whether current measurement strategies are capable

of accurately representing such an organizational system will be discussed in the following section.

Research Approaches to the Measurement of Memory Organization

Just as theoreticians have proposed a myriad of models for the structure of knowledge in memory, so have educational researchers generated a variety of methods to assess the nature of that structure. The following four subsections include a discussion of some of the more popular of these methods. Each subsection provides the assumptions on which the methods are based, descriptions of their general use, and brief reviews of attempts to apply them in investigations of cognitive structure. An extensive critique of these methods, *vis a vis* theoretical models of memory structure, is presented in the final subsection.

Word association. Word association has been by far the most frequently employed method of assessing cognitive structure in an academic environment. Typically, subjects in a word association study are presented with a list of stimulus words representing key concepts from a given discipline, and are required to respond to each word with whatever other word

that first comes to mind. Occasionally, subjects are limited to one response for each stimulus; usually, however, subjects may generate as many associates as possible, within some time limit. Subjects may also be free to respond with any word, regardless of its meaning or the nature of its association to the stimulus. Alternatively, responses may be constrained to associates from the same content area as represented by the stimulus words.

The basic notions of word association studies are 1) that the order in which associates for a particular stimulus are generated is indicative of their proximity to that stimulus concept in memory, and 2) that the extent of overlap of associates for two or more stimuli is indicative of the degree of stimulus concept interrelatedness (Deese, 1962). Based on these notions, the response orders and overlap observed in a given subject's word association data are typically combined into <u>relatedness coefficients</u> (cf. Garskoff & Houston, 1963) that reflect the overall associative strength between two concepts. Relatedness coefficients are calculated for each potential pairing of stimulus concepts, and entered into separate subject matrices. Individual subject relatedness matrices may be compared with an ideal relatedness matrix (generated from expert word association performance on some analysis of text organization), or the comparisons may be made between group performance (by creating a matrix of mean

or median subject coefficients) and the ideal structure.

Johnson (1964, 1965, 1967, 1969) was one of the earliest researchers to make extensive use of word association tasks in the study of cognitive structure. Subjects in his first experiment (Johnson, 1964) were high school students who either were currently enrolled in a physics course, had completed a physics course the previous year, or intended to take physics during the following year. Eighteen physics concepts were used as stimuli, and responses were not constrained, although only one response was permitted for each stimulus. As expected, current physics students generated the greatest number of associates that were also words from the stimulus list, and prior physics students gave more stimulus words as associates to other stimuli than did students planning to take the course. These results were judged as evidencd for the existence of an associative memorial network of physics concepts due to classroom instruction.

Relationships between word association performance and course achievement were also investigated by Johnson (1965, 1967). In the earlier study, he demonstrated that constrained, single-response word association performance may be enhanced when preceded by a content-relevant problem solving experience, whereas problem solving achievement may be enhanced when preceded by a word association task. However, no consistent relationship was found between problem solving

achievement and word association performance. Johnson (1965) attributed this result to a dependence of certain concepts on general language usage rather than on their more explicit meaning within the content area. In the later study (Johnson, 1967), high and low achievers in a physics course (as determined by course grades) were found to differ in performance on a constrained, unlimited response word association task. High achievers excelled in mean number of responses per stimulus word, and consequently generated higher median relatedness coeffiencients than did low achievers. These results were interpreted by Johnson (1967) as demonstrating that the constraints imposed on concept relations by explicit formulations in a content area like physics will engender similar relations in memory structures. Individuals differing in the extent to which these formulations have been internalized, as reflected by word association performance, will then differ in performance on content achievement measures.

In addition to relations between concept organization and problem solving, Johnson (1969) investigated the degree to which formal instruction may alter that organization. High school physics students performed a constrained, unlimited response word association task with terms taken from their course text, both at the beginning and at the end of a semester of instruction. Postinstruction performance was

found to exceed preinstruction performance in three respects: 1) more associates were generated for each stimulus concept; 2) associates tended more often to be concepts that appeared most frequently in the text; and 3) associates for any given stimulus concept tended more often to be terms that co-occurred with that concept in the text. These results were interpreted as reflecting students' acquisition, over instruction, of the rules of concept relatedness, as defined by concept usage in the formal language of the content area.

Although Johnson (1964, 1965, 1967, 1969) directed his research using word association measures toward assessing memory structure, he was not particularly explicit concerning the nature of that structure, beyond asserting that concept relatedness is a function of basic associative principles (e.g., contiguity, repetition, etc.). It is not clear whether the system of concept organization that may be inferred from Johnson's (1967, 1969) use of relatedness coefficients is a consistent, logical network resulting from active information processing, or is instead an arbitrary arrangement dictated strictly by the peculiarities of a particular text. One may wonder, in short, whether the word association data generated by subjects in these studies reflected their cognitive structures, or, instead, were the product of a simple response protocol based on concept co-occurrence frequency during instruction.

A more precise statement concerning the nature of memory organization has been offered by Richard Shavelson, who has been one of the most prolific of recent word association researchers. Shavelson (1974a) has defined cognitive structure as an elaborated network of concepts in memory. His conceptualization of this network is directly and explicitly derived from the memory model proposed by Collins and Quillian (1972), and includes that model's characteristic hierarchical interconnection of concept nodes, such that the meaning of any given concept is determined by it relations to other concepts. In addition to being more explicit about the nature of cognitive structure, Shavelson (1972, 1973, 1974a, 1974b) has also proposed a more informative method of representing <u>content</u> structure than Johnson's (1967, 1969) use of textbook word frequencies. This method, called <u>digraph analysis</u>, has been developed from procedures described by Harary, Norman, and Cartwright (1965) for the graphic representation of specified relations among concepts within any discipline.

Shavelson (1974b) has outlined three steps for the conversion of a text into a digraph for comparison with subject word association data: 1) identify the key concepts in a given subject matter (usually according to the judgement of a content expert); 2) identify all text sentences containing two or more of these key concepts, and diagram the sentences using some

standardized grammar (e.g., Warriner & Griffith, 1957); 3) transform diagrammed sentences into digraphs according to a set of conversion rules (cf. Shavelson & Geeslin, 1973). The resulting sentence digraphs indicate whether the relations among concepts are symmetric, and the number of relational links that must be traversed to arrive at one of the concepts from any other concept. For example, Figure 4 is a digraph of the sentence, "The median is the centermost score of a distribution." This digraph indicates that the relations between median and centermost score, and between centermost score and distribution are symmetric. It also indicates that two relational links must be traversed to move from median to distribution (or vice versa), but that only one link separates centermost score from both other concepts.

Once digraphs are constructed for all relevant sentences, they are collapsed into a superdigraph representing the relations among only the key concepts, with all nonkey concepts and relations deleted. The minimum distances in relational links between both key concepts of all possible pairs of concepts are then entered into a distance matrix. The distance matrix describing a particular domain of concepts may then be compared with a relatedness coefficient matrix produced from subject word association performance by directly computing Euclidean distances between the matrices; alternatively, hierarchical cluster (cf. S. Johnson, 1967) or multidimen-

Fig. 4. Digraph Representation of, "The median is the centermost score of a distribution.

sional scaling (cf. Kruskal, 1964) analyses may be performed on the word association data. The latter two procedures provide graphic, network-like representations of concept interrelations that may be compared with the structure of the superdigraph.

In his 1972 study, Shavelson applied digraph analysis to an investigation of the effects of five days of physics instruction on subjects' cognitive structures, as derived from their performances on pre- and postinstruction nonconstrained, unlimited response word association tasks. As expected, key concepts were interrelated more closely following instruction, and the group postinstruction relatedness matrix was nearer in Euclidean distance to the distance matrix derived from digraph analysis than was the preinstruction relatedness matrix. Subsequently, Shavelson (1973) reanalyzed these data to demonstrate that subject performance on a postinstruction achievement measure was related to word association performance. However, only a moderate correlation was observed (r = .35), and this only for the frequency of initial content-relevant associates to stimulus concepts. Shavelson (1973) provided no interpretation of this unexpected result other than to suggest that the word association acivity may have been sensitive only to early stages of structuring concepts and was not reflective of later, more complete concept processing in memory.

Geeslin and Shavelson (1975) have reported a further investigation of relations between word association structure and achievement. In their study, eighth-grade mathematics students were tested before and after a 10-day instruction period, and after an 8- to 12-day retention interval. Dependent measures included a constrained, unlimited response word association task and a 35-item achievement test. Multidimensional scaling of median relatedness coefficient matrices indicated that the students' cognitive structures were more like the content structure obtained from digraph analysis after instruction than they were before instruction. However, Geeslin and Shavelson (1976), much like Shavelson (1973), found no significant correlations between the correspondence of student structures to content structure and student performance on the achievement measure. They concluded from these results that acquiring an appropriate memory structure and being able to solve content-relevant problems may represent different aspects of learning.

Variations of the word association paradigm have been applied by other investigators to the study of cognitive structure. For example, Rothkopf and Thurner (1970) generated relatedness coefficient matrices not from traditional word association data, but from concept co-occurrence frequencies in pre- and postinstruction essay tests administered to high school physics students. Postinstruction data were found to

resemble more closely the naturally occurring frequency of concepts within the students' text, but this increase in structural similarity was not paralleled by an increase in student performance from a pre- to a postinstruction standardized, objective achievement test. More recently, Thro (1978) administered a typical constrained, unlimited reponse word association test to physics students, but compared resulting relatedness matrices to the word association performance of the course instructor, rather than to a textbook superdigraph. In addition to demonstrating an instructional effect on students' cognitive structures, this study is one of the few word association investigations to demonstrate a significant correlation between structure and course achievement. Unfortunately, Thro (1978) did not report the absolute magnitude of this relation; consequently, it is impossible to determine the educational significance of this result.

Word association was probably the earliest methodologically standardized approach to the measurement of memory organization, and it is still being applied in such research. The success of this technique in demonstrating changes in cognitive structure due to the effects of formal instruction is well-documented. Other alternative techniques have therefore frequently been evaluated according to their ability to generate similar representations of structure. It should be noted, however, that very few word association

studies (or, for that matter, studies using alternative measurement strategies) have demonstrated an unambiguous relation between what is purported to be an analog representation of cognitive structure and students' performance on content-relevant achievement measures. Whether this is a logical outcome, or if it instead suggests fundamental weakness in these approaches, are questions that will be addressed in a following section.

<u>Concept-Similarity Rating</u>. The methodology of structure measurement via content-similarity ratings differs somewhat from that of word association tasks. Typically, a small number of key concepts (i.e., 10-15) are selected from an instruction sequence, and each of these concepts is paired with all the others. Subjects are required to judge the degree of relationship between the two concepts in each pair according to a numerical, Likert-type scale usually anchored by "Very Similar" and "Very Different." Similarity ratings for all concept pairs may then be entered into individual subject matrices, or mean or median ratings across subjects may be combined to form a group similarity matrix. Finally, pre- and post-instruction matrices may be directly compared to determine instructional effects on congitive structure as indicated by systematic changes in similarity ratings; alternatively, they may be compared with some criterion similarity

matrix by means of multidimensional scaling or hierarchical cluster analysis to determine the effectiveness of instruction in assisting the creation of appropriate structures. The fundamental structural assumption of concept-similarity ratings is that the psychological distance between concepts in memory varies directly with the degree of their judged similarity.

Much of the early research involving the technique as noted in the previous subsection, was directed toward demonstrating results paralleling those obtained with word association procedures. For example, Johnson (1967) found significant correlations between concept relatedness coefficients and similarity ratings both for high achievers ($\underline{r}$ = .75) and for low achievers ($\underline{r}$ = .65). Similar results have been reported by Johnson (1969) and by Johnson, Curran, and Cox (1971) in comparisons not involving levels of course achievement. More recently, concept-similarity ratings have been directly applied in assessments of memory organization. Traub and Hambleton (1974) administered pre- and postinstruction rating activities, comprising all pairs of 13 psychometric and statistical concepts, to college students in an educational testing and measurement course. Analyses indicated that only about half of the mean concept-pair ratings changed significantly over one semester of instruction, and that most of those changes reflected greater judged dissimi-

larity between concepts. Also, multidimensional scaling indicated that students grouped the concepts into fewer categories after instruction. However, no control group was utilized in this study to clearly demonstrate that observed structural changes were indeed caused by instruction, rather than by practice on the rating task.

One of the more interesting and, perhaps, educationally relevant investigations involving concept-similarity ratings has been reported by Stasz, Shavelson, Cox, and Moore (1976). These researchers assessed the effects of teachers' differing cognitive styles on the concept structures of students also differing in cognitive style. The dimensions of cognitive style contrasted in this study were field independence (FI) and field dependence (FD), where the former is defined as being better able both to perceive objects as discrete from their backgrounds and to alter the organization of a perceptual field, and where the latter is defined as being less able to disembed objects from their backgrounds and tending not to change the organization of a perceptual field.

Stasz et al. (1976) determined the relative position of subjects on a FD-FI continuum according to their performances on several tests recommended by Witkin, Dyk, Faterson, Goodenough, and Karp (1962). Twelve FI teachers and twelve FD teachers each taught two FI and two FD students four 50-minute lessons on Mayan civilization. Both students and teachers

were tested on a rating task comprising 10 key concepts taken from the instructional material, before and after the instructional sequence. Multidimensional scaling analyses indicated that the postinstruction rating task performance of the FI teachers more closely approximated a criterion structure than did the performance of the FD teachers. A similar result was obtained in a comparison of postinstruction performance of FI and FD students. Also, students and teachers with the same cognitive style generated more similar structures via the rating activity than did students and teachers differing in cognitive style. Stasz et al. suggested from these results that cognitive style be given serious consideration as a parameter in Aptitute X Treatment interaction research, since the memorial organization of concepts inferred from concept-similarity ratings differed as a function of that variable.

Data obtained from the concept-similarity rating method of measuring cognitive structure have been shown both to parallel those from word association exercises and to be sensitive to instructional effects; however, few attempts have been made to unambiguously relate rating task performance with achievement measures. Johnson's (1967) investigation indicated only that grouping students according to the degree of relatedness between their word association and rating task performances did not differ from grouping according to their level of course achievement. Stasz et al. (1976)

administered pre- and postinstructional achievement tests to their subjects, but did not report correlations between these measures and rating task performance. Bates (Note 3) and Konold and Bates (Note 1) have observed significant correlations between these variables, but the methodology employed by these researchers (discussed in detail in a later section) differs in several fundamental respects from that of previous concept-similarity rating studies. Thus, this approach, as applied to the measurement of memory organization, has been no more successful than the use of word association activities in establishing relationships between appropriateness of cognitive structure and problem solving ability.

Tree Construction. Rapaport (1967) has developed another method or representing the semantic similarity among concepts that is especially amenable to multidimensional scaling procedures. This method, called tree construction, requires subjects to make judgements about the degree of concept relatedness relative to all others of a set of key concepts. Subjects are presented with a list of concepts, and are instructed to write on a separate page those two concepts from the list that are most closely related. A line is drawn between these concepts, and is labeled "1." Subjects then select a concept from the list that is judged to be next most

closely related to one of the first two concepts, write it next to its associate, and link the two with a line labeled "2." A fourth concept, of lesser judged relatedness, is next selected and linked to one of the first three with a line labeled "3." If at any time none of the remaining list concepts is judged to be closely related to any of the previously selected concepts, two list concepts that are themselves closely related may be written on the page and linked with an appropriately numbered line to begin a second tree. Concepts are added from the list to the tree(s) until all have been used. Subjects may then be required to connect each tree, if more than one have been created, to one of the others via an interconcept numbered link, such that only one large tree of concepts remains. Much like for word association tasks and concept-similarity ratings, it is assumed that the degree of concept relatedness indicated by the number and ranking of intermediate links is analogous to interconcept psychological distance in memory.

Shavelson (1974a, 1974b), Shavelson and Stanton (1975), and Preece (1976) have all demonstrated that tree construction exercises produce representations of structure similar to those obtained from word association data. However, Preece (1976) noted that tree construction did not approximate as closely an ideal concept structure generated from digraph analysis of a physics text as did either free or

constrained word association performance. None of these studies evaluated tree construction in terms of correlations with content-relevant problem solving measures.

Rudnitsky and Garlock (Note 4) have argued that digraph analysis of content concepts and word association activities may not be appropriate for the study of cognitive structure because these methods tend to mask differences in concept emphasis and order for different presentations of the same content. They have reported an alternative procedure, called graph building, that is a variation of tree construction. In this method, subjects are presented with a number of key concepts, each of which is printed on a gummed label. These labels are to be organized on a separate page into whatever configuration the subjects think best represents general concept categories and degrees of interconcept relatedness. Once subjects are satisfied with their concept arrangements, the labels are to be fixed to the page, and lines are to be drawn between all related concepts.

Subjects in Rudnistsky and Garlock's (Note 4) study were students enrolled in a college botany course. Seventeen concepts were selected from a 2-session instructional sequence dealing with plant growth and development. One group of subjects received instruction that structured the key concepts in terms of spatial and temporal plant characteristics (World-related unit). The instruction for another group of subjects

emphasized concept functions and compositions (Concept-related unit). Both groups were administered a graph-building exercise and a recall-type achievement test following the instructional sequence. Graph building performances of the two groups were compared with ideal concept arrangements, which were determined by content expert analyses of the two different sets of instructional material according to a method very similar to the subjects' graph building exercise. Rank-order correlations between interconcept distances for median-subject and ideal relatedness matrix cells indicated that high achievers in both instructional groups structured the key concepts more like experts than did the low achievers. Also, subjects tended to build graphs more like the ideal structure for the Concept-related unit than like the ideal structure for the World-related unit, irrespective of level of achievement or type of instruction. Rudnitsky and Garlock tentatively concluded that the Concept-related unit had a greater effect on cognitive structure than did the other instructional approach, and that identifiable groups of students (in terms of recall achievement) structure new information in qualitatively different fashions.

The primary appeal of the tree construction approach and the graph building variation appears to be the resulting graphic representation of structure that is at least morphologically consistent with many theoretical notions about memory

organization. However, too little research have been conducted to draw firm conclusions regarding the psychological validity of this technique. Of particular importance is how student tree constructions may relate to achievement measures involving higher level cognitive tasks than the rote recall of concept definitions.

Card Sorting. Administration of a card sorting exercise is very simple. Each of a set of key concepts is printed on a single index card. Subjects are instructed to familiarize themselves with the concepts on all the cards, and then to sort the cards into as many piles as might be appropriate, such that the concepts in any given pile are more closely related to each other than they are to concepts in any other pile. Individual subject relatedness matrices may then be constructed by entering a zero in a cell if the corresponding two concepts were placed into different piles, or a one if the pair was placed into the same pile. Group data may be compiled by summing subject matrices, or by recording the proportion of subjects placing each pair of concepts into the same pile. Just as in other approaches, it is assumed that observed groupings of concept cards are indicative of similar concept clusters in semantic memory. Consequently, these data are generally treated statistically in the same manner as are relatedness coefficient or similarity matrices in

word association and concept-similarity rating research.

Relatively few recent investigations of cognitive structure have employed card sorting as a means of evaluating concept organization. Shavelson (1974b) attempted to demonstrate that card sorting yielded a representation of structure paralleling those obtained from word association and tree construction activities, but had little success. Both of the content experts who were his subjects sorted 12 mathematics concepts into the same two piles, whereas the other two techniques both produced four superordinate concept categories. However, Shavelson and Stanton (1975) presented the same 12 concepts to a group of intern mathematics teachers, and were able to demonstrate at least visually similar structural representations for card sorting, word association, and tree construction.

In a more applied context, Hambleton and Sheehan (1977) administered pre- and postinstruction card sorting tasks and a postinstruction achievement test to ninth-grade science students, with key concepts taken from a 7-day instructional sequence dealing with atomic structure. Group matrices in this study were subjected to latent partition analysis (cf. Wiley, 1967), rather than to more typical multidimensional scaling or hierarchical cluster analysis. This procedure was utilized based on the assumption that the observed concept clusters for individual subjects in a card sorting task may not accurately represent the unobservable but underlying

categories that actually comprise those concepts. Latent partition analysis provides a means to derive such underlying structure from observed group performance. Results indicated that subjects who scored high on the achievement test generated card sorting data with fewer ambiguities (i.e., concepts frequently placed in more than one category) than did subjects with low achievement scores. However, a comparison of pre- and postinstruction data failed to establish an instructional effect for either achievement group. That is, subjects tended to sort the concepts into the same piles after instruction as they did before instruction. Nevertheless, Hambleton and Sheehan (1977) interpreted their results as supportive of card sorting activities and latent partition analysis of the resulting data in assessments of the effectiveness of instruction.

Although conceptually similar to the three previously discussed techniques, the card sorting method of measuring memory organization has not been as widely used in cognitive structure research. This is somewhat surprising given the ease with which data may be accrued with this procedure. In fact, it seems likely that far more concepts could be evaluated in less time via card sorting than via any of the other measurement strategies. Still, the research that has been conducted has thus far not found card sorting to be of particularly dramatic value.

A Critique of Measurement Strategies

In the first section of this paper, it was noted that very few structure researchers have been explicit about the theoretical foundations of their methods. In fact, the preceding review of those measurement strategies may be judged as indicating that, with one exception, researchers have provided only the most general set of assumptions regarding the nature of cognitive structure: knowledge is represented in memory in the form of concepts clustered and interlinked according to either contiguity or similarity. (Whether this lack of specificity is a significant omission, from the standpoint of educational applicability, will be addressed in the next section.) The single noteworthy exception to this perspective has been provided by Shavelson and his associates. Shavelson (1974a) has embraced the hierarchical network model of memory, as proposed by Collins and Quillian (1972), as the most appropriate theoretical treatment of memory organization. This view has been reaffirmed by Shavelson and Stanton (1975), and is at least implicit in the other reviewed investigations conducted by Shavelson, regardless of the particular measurement techniques applied. Although other researchers have not been so specific, it seems reasonable to suppose that they would generally share Shavelson's perspective, because they have so frequently cited

his research as supportive of their own investigations, and because they have so frequently employed his methodology. For these reasons, the four techniques discussed in the previous subsections will first be critiqued according to the characteristics of the hierarchical network model of memory. Of course, this is not to say that educational researchers other than Shavelson who have used these techniques have always done so based on this or any other of the reviewed theoretical models. However, if fundamental differences in perspective do exist, they have not often been clearly expressed in the structure research literature.

It may be recalled that the degree of concept relatedness in Collins and Quillian's (1972) and Collins and Loftus' (1975) model is dependent on three factors: the number, criteriality, and type of intermediate links. In contrast, all of the structure measurement techniques that have been reviewed generate structure representations based almost _solely_ on the first of these factors. The principle of link criteriality states that links between concepts, although bidirectional, are not necessarily symmetric. Collins and Loftus (1975) have discussed research in which reaction times to judge concepts as similar were found to vary dependent on the order of concept presentation. If reaction time in such a task may be considered as indicative of the strength of interconcept relatedness, then one could conclude that the

degree of judged similarity between two concepts may also vary dependent on presentation order. Although researchers making use of concept-similarity ratings routinely counterbalance concept order across subjects, no effort has been made to investigate criteriality effects for individual subjects by presenting them with both concept orders for all pairs and contrasting resultant similarity half-matrices. Moreover, the statistical techniques used to convert similarity matrices into graphic representations of structure uniformly require that the matrices be symmetric. If link criteriality is an important consideration, as suggested by the theoretical model, then this necessary assumption may have been consistently violated. Word association, tree construction, and card sorting are probably less subject to order effects, but studies involving these methods have also made frequent use of the same statistical techniques, and may have consequently also violated the assumption of matrix symmetry. Thus, a fundamental component of the hierarchical network model has been ignored, with no efforts to determine whether such an omission significantly affects structure representation.

Whether criteriality affects similarity judgements in a pencil-and-paper test in the same manner as may be inferred from reaction time studies is open to empirical verification. Far more critical is the failure of all the reviewed measure-

ment strategies to account for the _types_ of relations that may be perceived among concepts. Word association tasks, both free and constrained, appear to be especially deficient in this regard, because instructions to subjects allow associates to be generated based on virtually an infinite number of potential relations. For example, in reference to the hypothetical concept network presented in Figure 1, _variance_ may be offered as an initial associate for _standard deviation_ because the former is needed to calculate the latter, because both are measures of variability, because of a high concept cooccurrence frequency during instruction, or because of some other, possibly even incorrect, perceived relation. Word association tasks in previous cognitive structure studies simply have not been able to determine on what basis concepts have been related, nor have such studies demonstrated even limited control over the consistency with which particular classes of relation have been applied within or across key concepts.

Concept-similarity rating, tree construction, and card sorting fair little better in terms of controlling for types of perceived relations. The first of these methods expressly requires judgements to be made in terms of the degree of concept similarity, and descriptive anchors are provided for that purpose. However, two concepts from the same domain may easily be perceived as being more or less similar dependent on the context in which the judgement is made. The concepts

arithmetic mean and median may be thought of as very similar, because both are measures of central tendency, or they may be thought of as very different, because one is sensitive to extreme scores in a distribution and the other is not. Different similarity ratings in this case may not reflect different levels of sophistication as much as they do different purposes that are informing judgements.

Tree construction requires that concepts be selected from a pool in the order of strength of relation to other concepts. Once again, there is no guarantee in this method that relation strength is consistently expressed in terms of the same relation type. In fact, allowing subjects to construct multiple trees probably increases the likelihood that the nature of perceived relations will vary unpredictably as fewer terms remain in the unselected pool. Rudnitsky and Garlock's (Note 4) graph building variation and the card sorting technique both require that subjects simultaneously arrange all concepts in a manner that best describes the content area. One might assume that such an approach would maximize the likelihood that similar relation types will be applied throughout. Nevertheless, this still does not provide information regarding what relations are in fact being applied. Also, different resulting structure representations may again indicate little about relative levels of understanding, but instead may reflect differing purpose. This is illustrated, as reported in the previous sub-

section, by the conflicting structures obtained when the same card sorting task was performed by content experts (Shavelson, 1974b) and by intern teachers (Shavelson & Stanton, 1975). Based on those results, one might conclude that the new teachers were more sophisticated in their understanding of mathematics than were the supposed experts!

In summary, none of the reviewed structure measurement techniques, in their present form, are adequate for representing the organization of memory as conceptualized in the hierarchical network model, because none provide for two of the three link characteristics that, according to this model, define interconcept relatedness--link criteriality and link type. In addition, these techniques appear to be incompatible with the other concept network models as well. Whether in terms of markers, features, or defining and characteristic attributes, each of these models conceptualizes relatedness as a function of the context or purpose of the decision-making task. Current techniques have not demonstrated the ability either to precisely control the universe of potential perceived relations among concepts, or to distinguish among those relations that may have been perceived in any given structure measurement exercise.

If the reviewed measurement techniques are not presently equal to the task of representing declarative knowledge according to the notions of memory models designed for that purpose,

then they surely fall short of the mark as strategies to represent more procedural units of information in the form of propositional or schema networks. Each of the reviewed propositional models has in common with the other an emphasis on concept relations specified by the nature of their predicates. None of the four measurement strategies controls for or can distinguish among the particular predicates from which concept relatedness may be ascertained. Several of these memory models (e.g., ACT, ASN, etc.) explicitly postulate the existence of procedural knowledge units which actively specify the qualities and quantities of concepts that may be interrelated within propositions. Again, measurement strategies have not accounted for the potential effects that differing but equally correct and sophisticated procedural heuristics may have on resulting structure representations.

Schema models of memory have been shown to incorporate individual perspective into decision-making tasks even more intimately than have propositional models. Reference can be made once again to the story-interpretation experiment reported by R. C. Anderson (1977), involving music and physical education students, as an illustration of schema effects on inferred congnitive structure. It will be remembered that the two groups of subjects in this study did not differ in story recall performance, but differed widely in their perceptions of story theme. Imagine that these subjects had also

performed one of the four discussed structure assessment exercises incorporating a set of key story concepts. If resulting structure representations were found not to differ between the groups, then one might conclude, without access to observed theme judgements, that both groups had integrated the story information into memory in essentially the same fashion. More realistically, one might expect that their differing perceptions of story theme would cause the groups to differ in structure task performance, but this difference surely could not be ascribed to varying levels of understanding or sophistication because the groups also demonstrated equal recall performance. What this hypothetical extrapolation suggests is that, if units like schemata do exist in memory, then they may render uninterpretable any structure representations accrued from standard measurement techniques in their present form.

Thus, it appears that word association, concept-similarity rating, tree construction, and card sorting are presently not capable of generating graphic representations of memory organization that accurately reflect the considerations of modern memory models. However, these techniques have rather consistently been demonstrated to be sensitive to instructional effects. That is, subjects engaging in these tasks do seem to interrelate concepts differently after exposure to an instruction sequence. What, then, are the tasks measuring?

One possible answer to this question has been offered by Bates (Note 3) and by Konold and Bates (Note 1), who recently reviewed cognitive structure research from the perspective provided by Tulving's (1972) distinction between episodic and semantic memories. They have suggested that, although that body of research has had as an explicit goal, the evaluation of concept organization in an elaborated semantic network, traditional methodologies may have encouraged subject performance relying heavily on the recall of information with primarily episodic referents. Three procedural characteristics that may have contributed to this problem have been presented by these researchers. First, although a few studies have investigated structure changes over a full semester of formal instruction (e.g., Johnson, 1969; Thro, 1978; Traub & Hambleton, 1974), most have looked for differential structuring after only a few days of instruction (e.g., Geeslin & Shavelson, 1975; Hambleton & Sheehan, 1977; Johnson, 1975; Shavelson, 1972, 1973). There is little evidence to support the notion that semantic memory is indeed so amenable to rapid change. If it is not, then the basis for subjects' judgements about concept relatedness may have been temporally dependent remembrances of instruction. Stewart (1979) has made a similar observation. After discussing typical structure measurement procedures, Stewart concluded that, "neither Shavelson nor other associative mappers provide any evidence that temporal rela-

tionships are not what their techniques are measuring. They assume that temporal relations can only occur during paired-associate learning where the items to be learned are presented more or less simultaneously. Yet, it is possible to envision that concepts learned during a lesson or even during a course may just as well be associated only because they were presented at the same time" (1979, p. 399).

The second issue raised by Bates (Note 3) and by Konold and Bates (Note 1) is that subjects in prior structure research have typically been required to make judgements about relations between concepts that may have been directly taught during instrucion (e.g., Force = Mass X Acceleration). Although this procedure may be necessary to generate the complete matrices required by analysis techniques, it enhances the probability that certain relations may be judged on the basis of episodic recall of classroom events rather than on the basis of semantic inferences from cognitive structure. In fact, Perkins (Note 5) has provided evidence that an "ideal" arrangement of new concepts may be rote-memorized by students, with the result that their performance on a structure measurement task may be in no way indicative of their understanding of the course content. A third and related criticism offered by Konold and Bates is that few researchers have specified the taxonomic levels of the items included in achievement measures that have been correlated with cognitive structure. If

achievement test items have been written at relatively low cognitive levels (i.e., rote knowledge), or if such items have been near literal replications of classroom examples, then subject performance on these items may also, to some extent, be a product of episodic recall. Assuming that cognitive structure measures are tapping semantic structure, the frequently observed low correlations between structure (semantic memory) and achievement (episodic recall) would not be unexpected.

Konold and Bates (Note 1) have offered evidence in support of these arguments in the form of results from two studies conducted in normal college classrooms, using a methodology designed to increase the likelihood that subject performance would be primarily the product of inferences relying on semantic memory. Their results, replicated in a similar context by Bates (Note 3), have indicated that the length of instructional treatment, the characteristics of the cognitive structure measure, and the taxonomic characteristics of test items measuring achievement (cf. Bloom, 1956) are critical factors in the demonstration of meaningful instruction-structure-achievement relationships. That these considerations have not been operationalized in previous classroom-oriented investigations of memory organization is further evidence against interpreting the results of those studies as representative of semantic structure.

Current techniques for assessing the organization of memory have thus far been attacked on several grounds. In their present form, they do not appear to represent structure in terms of any contemporary memory model, they make use of statistical tools with basic assumptions that may have been repeatedly violated, and they routinely incorporate procedures that may encourage performance based primarily on temporal/spatial characteristics of instruction, rather than on semantic inferences. The next section of this chapter includes some tentative suggestions for what may be more appropriate methodologies to be applied in future structure assessment research.

Alternative Approaches to the Measurement of Memory Organization

One reasonable response to the foregoing criticism of structure measurement strategies is that it is not really relevant to the concerns of educators who may apply these strategies in the classroom. That is, whether or not word association, concept-similarity rating, tree construction, and card sorting are consonant with some theoretical construct is of far lesser practical importance than whether they provide useful information regarding students' understanding of course material. In short, do these methods provide data that

will enhance predictions about content mastery? After all, our current understanding of the underlying components of intelligence may be said to be not that far removed from Boring's assertion in 1923 that intelligence is whatever the tests test. Nevertheless, and despite recent popular notions to the contrary, knowledge about a student's performance on a standardized individual intelligence test can provide a teacher with useful indicators about that student's present academic strengths and weaknesses. May not the structure measurement techniques that have been discussed be of similar value, regardless of their lack of correspondence to theories about memory organization?

In order to respond to this sensible observation, one must reconsider the purpose of structure assessment, at least from the perspective of educational application. Three ultimate goals were suggested in an earlier section for the development of these strategies: 1) more facilitative structuring of text material, 2) more logical sequencing of instruction, and 3) remediation for students with inappropriate knowledge structures. But what is it that we hope to achieve by meeting these goals? If, as seems likely, our purpose is to make meaningful statements about how well individual students or groups of students will be able to apply what they have learned in a classroom to problems they will encounter within or outside that classroom, then current measurement strategies

have not been able to tell us what we want to know. In fact, only three of the twenty-one reviewed cognitive structure investigations (i.e., Thro, 1978; Bates, Note 3; Konold & Bates, Note 1) have demonstrated unambiguous correlations between students' performances on structure tasks (relative to a criterion structure) and on achievement measures. It would seem prudent for researchers who are actively attempting to validate methodologies for representing concept organization to seriously address this issue. If current techniques have not been able to consistently establish that what they are measuring is meaningfully related to students' understanding of and degree of sophistication in course material as reflected by their level of course achievement, then why should educators invest the effort to administer them? Indeed, Stewart (1979) has concluded that, for science curriculum researchers at least, these techniques are useless.

One need not, however, take such a pessimistic view of the future of cognitive structure research. For example, several alternative methodologies have been developed from traditional reaction time paradigms that appear to hold promise for the measurement of memory organization. Loftus and Loftus (1974) have reported a study that required beginning and advanced psychology graduate students to provide the names of psychologists who fitted certain descriptions. Reaction times to state the appropriate names were found to vary systematically with the

students' level of graduate school experience. This was interpreted as evidence for differing structural arrangements in the students' semantic memories. Meyer and Schvaneveldt (1976) have investigated semantic relatedness of concepts based on the reaction times necessary to judge sentences as true or false. Ratcliff and McKoon (1978) have been able to demonstrate the propositional structuring in memory of newly learned prose sentences by means of reaction times to recognize words from the study material when primed by other words also taken from those materials. Although these techniques generally require laboratory environments and involve equipment not often available to the classroom teacher, they may still provide useful methods for validating the results obtained from other, more practicable, structure assessment devices.

There may also be ways to alter the reviewed measurement techniques such that they are more consistent with contemporary memory models. Criteriality effects could easily be evaluated in concept-similarity rating studies by means of the suggestion offered in the previous subsection. Control over the types of relations perceived among concepts may be more difficult. Perhaps providing subjects in a tree construction exercise with a list of potential relations and requiring that numbered links be labeled to indicate which relation is being applied would generate more meaningful structure representations. Similar specified relations might also be incorporated

into word association and concept-similarity rating. Whether and how such additional information could be accommodated by traditional statistical techniques (i.e., hierarchical cluster analysis, latent partition analysis, and multidimensional scaling), however, is not immediately apparent.

A first approximation of experimental control over structure task purpose might be more easily attained by providing subjects with a short descriptive passage illustrating the interrelations among concepts included in the task. For example, an investigation of memory organization involving students in an experimental methods or statistics course might require that subjects first read a brief description of a hypothetical study in the form of an expanded abstract. This passage would imply, but not make explicit, some system of concept organization that is appropriate to the content area. Subjects would then respond to one of the standard structure tasks by associating the key concepts in the manner they believe may be inferred from this priming schema. Although the concepts to be structured would not be directly named in this passage, care would probably have to be taken to insure that appropriate degrees of relatedness could not be perceived based only on syntax or concept juxtaposition. Some empirical support for such an approach may be inferred from Johnson's (1965) finding that word association performance was enhanced, relative to a criterion structure, when preceded by a content-

relevant problem solving activity.

Whatever research methodologies may be developed to account for perceived purpose and relation types, future investigations of cognitive structure should include demonstrations that task performance is related to problem solving achievement. This is not an unreasonable expectation. Mayer and Greeno (1972) have reported that differences in the conceptual organization of two instructional sequences on binomial probability were observed effect qualitatively different problem solving abilities for college students. Eylon and Reif (Note 6) have noted similar results with hierarchical and linear arrangements of concepts from Newtonian physics. If altering the external structure of instructional material with otherwise identical content can result in differing levels or types of problem solving ability, then it seems illogical to contend, as have some structure researchers, that the <u>internal</u> (i.e., memorial) organization or knowledge may not be similarly related to achievement. Indeed, Greeno (1973) has argued that the structure of memory has a determinant role in problem solving. Given that some researchers have been able to demonstrate significant structure-achievement relations (i.e., Thro, 1978; Konold & Bates, Note 1; Bates, Note 3), and given that replications of such results are necessary to establish structure measurement techniques as valuable tools for educators, then attempts to verify these relations should be incorporated into

future structure research.

Careful consideration should also be given to the nature of the criterion structures against which subjects' structure representations may be evaluated. Three general types of criteria were mentioned in the previous review of structure research. The first was applied by Johnson (1965, 1969) in his word association studies, and involved generating ideal relatedness coefficient matrices based on the cooccurrence frequencies of key concepts in the instructional material. A major problem with this approach is that one is not sure how to interpret strong similarities between subject and ideal structures after instruction. On one hand, such parallel representations may indicate the acquisition of appropriate concept organization; on the other hand, this may reflect nothing more than the acquisition of simple associative response protocols based solely on how often pairs of concepts were encountered simultaneously during the course, irrespective of the contexts in which those associations occurred. That this second interpretation may sometimes be the more compelling is illustrated by the results of Rothkopf and Thurner's (1970) study, which involved just such a criterion structure. Although subjects were closer to the criterion after instruction than they were before, they demonstrated no parallel increase in knowledge about the course content on a standardized achievement test. Did these subjects

aquire meaningful semantic structure, or did they instead acquire a paired-associate mnemonic?

The second method of establishing a criterion structure is Shavelson's digraph analysis of text sentences. Stewart (1979) has criticized this approach on the grounds that reducing relations among concepts as expressed in text passages to digraphs tends to obscure the nature of those relations. This masking is compounded by creating superdigraphs that do not represent possibly important qualifying terms because they are not included on the researcher's list of key concepts. The nature of intermediate links is completely obliterated when the superdigraph is converted into a relatedness coefficient matrix. Stewart has also provided examples to illustrate that text sentences describing identical concept relations may be represented by qualitatively different digraphs, and that meaningful concept relations may be deleted from superdigraphs simply because another text sentence contains the same concepts with fewer, but possibly even trivial, intermediate links. Thus, using textbook superdigraphs as criteria for the evaluation of instructional effects ignores the factors of perceived purpose and relation type that have been argued as being critical to meaningful assessment of cognitive structure.

The final criterion type used in prior research has been the structure task performance of one or more content experts.

Thro (1978) has reported excellent results with this approach, using the word association data of the course instructor to generate an ideal structure representation. Konold and Bates (Note 1) have had some success using the concept-similarity rating data from a group of experts for this purpose, as has Bates (Note 3). At present, Thro's approach seems to be the most advisable of all the criterion alternatives. If the structure task employed controls for perceived purpose, such a criterion should provide a meaningful representation of semantic structure, because it is unlikely that an instructor's performance would be dependent on the peculiarities of any single text. Also, using a course instructor's performance as a criterion may have more psychological validity than would an ideal structure generated from a group of experts, because the latter representation may well be unlike the system of concept organization actually applied by any single expert.

Other methodological alterations, as suggested by Konold and Bates (Note 1) and by Bates (Note 3), may further enhance the validity of future cognitive structure research. In brief, these include lengthening the instructional interval and avoiding the inclusion within a structure task of directly taught relations, in order to increase the likelihood that judgements about interconcept relatedness are more dependent on semantic inferences than on episodic recall. In addition,

achievement tests with which structure exercises are to be related should be carefully constructed to insure that items are not replications of classroom or textbook examples and may not be correctly answered based only on rote-memorized information. Items should probably be written at least at the comprehension level of understanding, according to Bloom's (1956) cognitive taxonomy, and possibly at even higher levels, to maximize the likelihood that they will tap semantic structure.

This chapter has included a review of several approaches to the measurement of concept organization in memory, with the result that each of these methods has been judged to provide an inadequate representation of that structure according to the precepts of modern psychological theory. Moreover, the immediate educational value of these techniques, in their present form, has been severely questioned. All this has not been to denigrate the efforts of those researchers who have so diligently pursued what must have seemed at times to be an especially elusive quarry. After all, the hunt for the nature of human memory organization has been conducted for several thousand years without a unanimously accepted resolution. How, then, are we to best measure this construct? It is hoped that the tentative suggestions offered above will help to direct some answers to this question. The potential value of a measurement device that reliably and validly

assesses memory organization is obvious, and the search for such a device should continue. The remaining portions of this dissertation describe one such attempt to further lessen the distance to the quarry.

The Present Study

The previous section includes a number of suggested considerations for future cognitive structure research, in order that more meaningful and theoretically consistent representations of memory may be obtained. The study described in the following chapters was intended to provide information regarding several of those suggestions. Three different groups of subjects--undergraduate students enrolled in an introductory statistics course, undergraduate students with no prior exposure to statistics instruction, and graduate students who had successfully completed at least two advanced statistics courses--were asked to rate the degree of relatedness among a number of basic statistical concepts. Relatedness judgements were made on a modified concept-similarity rating task, and were compared with a criterion measure of memory structure obtained by administering the same task to the instructor of the undergraduate statistics course. Two factors of major interest were investigated, based on these relatedness ratings: the effects of interconcept criteriality on strength

of perceived relatedness, and the utility of a prose passage, acting as a priming schema, for directing the nature of perceived relations.

As noted earlier, most educational researchers who have investigated cognitive structure have done so apparently from a perspective on memory ogranization like that provided by Collins and Loftus (1975). One important factor in this model that these researchers have not accounted for is the effect that differing interconcept criterialities may have on relatedness judgements made by subjects who have been presented with only one order of concepts. There is some empirical evidence to support investigating such an effect: Collins and Loftus (1975) have reported that reaction times to judge superordinate and subordinate concepts as similar may vary as a function of which type of concept is presented first; also, Tversky (1977) has demonstrated that concept presentation order in a pencil-and-paper format can alter the perceived strength of concept relatedness. In the latter study, subjects rated the degree of similarity between what might be called superordinate and subordinate countries (e.g., USSR - Cuba, USA - Mexico, etc.), wherein superordinates were defined, a priori, as generally well-known, prototypical social systems, and subordinates as lesser known variants. The members of such pairs were consistently rated as more similar when the variant preceded the prototype (i.e., Cuba - USSR)

than when the order was reversed. Although Tversky (1977) interpreted this result from a theoretical perspective other than Collins and Loftus' (1975) semantic hierarchy model, it nevertheless supports the contention that concept presentation order may significantly affect representations of cognitive structure.

The existence of an order effect was investigated in the present study by requiring all subjects to rate the interconcept relatedness for all possible pairs of the selected statistics concepts, which included an equal number of general, superordinate terms and more specific, subordinate terms. It was expected that substantially different structure representations would be obtained, dependent on which pair-ratings were used to construct those representations. Specifically, presentation of concepts in the order, "subordinate-superorindate" was expected to elicit from the statistics and graduate students stronger relatedness ratings than was presentation of the same concepts in the reverse order. The logic of this prediction is based on the notion of interconcept link criteriality, but may also be derived from other memory models. Basically, what was expected was that, over the course of instruction, asymmetrical relationships between certain types of concepts would develop within the semantic structures of the statistics students, such that a given subordinate concept would be learned to be of more importance to the understand-

ing and application of another, superordinate, concept, than the latter would be to the former. To iterate an example used in a previous section, it seems reasonable to expect that students would be more likely to respond to the presentation of the specific concept of range with the associate general concept of measure of variability than they would be to respond to variability by conjuring up the notion of range. Because they frequently encounter and make use of the same concepts in similar contexts, the graduate students were also expected to demonstrate an order effect of the same nature. This effect, if observed for these groups of subjects, would support the notion that prior structure research has failed to account for a more than theoretically important factor in memory measurement. It would indicate that the analysis techniques so frequently applied in previous studies, because they are based on assumptions of symmetry in relatedness judgements, may not be appropriate, and might therefore generate distorted representations of cognitive structure.

Further evidence of a concept-order effect was expected to be obtained by comparing correlations between the statistics students' levels of performance on a postinstruction achievement test and their levels of performance on the concept-similarity rating task, relative to the criterion measure, for both possible orders of concept presentation. The

purpose of this investigation was to determine if one order of concept presentation would yield a better predictor of high-level achievement than would the other order, or if the best predictor might be obtained by providing subjects with both orders and averaging resultant ratings. Such information would be of value when using relatedness ratings in an applied setting to determine which students are sufficiently prepared to progress to the solving of complex problems.

Another consideration discussed in the previous section is that current structure measurement techniques do not adequately control for the types of relations that subjects may apply to their judgements of concept relatedness. This lack of control may lead to the construction of structure representations that do not accurately reflect any model of memory organization, and may account for at least some of the ambiguity observed in the results of the prior structure research. The present study incorporated a prose passage describing the research application of statistical concepts, which was to be read by half of each group of subjects before engaging in the concept-similarity rating task, and which was intended to provide a context for more consistent judgements about concept relatedness. It was expected that the rating task performance of those statistics students who were provided with this orienting passage would differ from the performance of the students who were not in at least two respects. First, the

concept-pair ratings of students provided with the passage were predicted to be more similar to the criterion ratings than would be the ratings of students not provided with the passage. This would be so because the criterion structure obtained from the course instructor, given his experience with the content area, would be based on similar and reasonably consistent perceptions of interconcept relation types. In contrast, the passage was not expected to affect ratings for either the statistically naive students or the graduate-student "content experts." In the former case, this would be so because the passage would not direct these subjects to an organization of concepts in their memories that did not already exist. Unless the passage were to teach a way to organize the concepts--a highly unlikely outcome--it should in no way yield ratings that were consistently closer to the criterion. No passage effect was expected in the latter case for a similar reason. That is, the passage should not act as a teacher; it should only serve as the basis for consistent judgements about concept relatedness. To the extent that the graduate students were already reasonably facile with statistical concepts and had actively applied them in a variety of research situations, the passage should have proved to be redundant.

The second way in which the passage was expected to affect rating task performance was to be demonstrated by

correlating structure scores (in terms of similarity to the criterion) for statistics students within each group with corresponding performance on two different item-types used in the postinstruction achievement test. The importance of specifying the taxonomic characteristics of achievement measures used in conjunction with cognitive structure research, experimental evidence illustrating this issue (i.e., Konold & Bates, Note 2; Bates, Note 3), and one potential theoretical interpretation of prior structure research (i.e., Tulving's distinction between episodic and semantic memories) have already been discussed in some detail. The present study included, as indicators of achievement, test items written at either the knowledge or the application levels of Bloom's (1956) cognitive taxonomy. If the passage acted as expected to consistently focus the basis of subject ratings on an inter-related set of semantic knowledge, then rating task performance was expected to be even better correlated with application-level achievement than it would be for subjects without the passage. However, to the extent that knowledge-level achievement, as defined in the present study, may be a product of episodic recall of classroom events, correlations between this measure and structure scores were predicted to be lower for both groups than would be the corresponding correlations with application-level achievement.

Demonstration of this differential relation between cog-

nitive structure and types of achievement was judged to be one of the more important of the expected outcomes for the present study. It would provide further evidence in support of the notions expressed by Konold and Bates (Note 2) and by Bates (Note 3) that a meaningful representation of structure can be significantly related to achievement that is similarly based on inferences from semantic memory. It would also illustrate the potential value of appropriate measures of cognitive structure as predictors of that type of achievement that is usually associated with meaningful learning.

Although not an issue of prime consideration, an indication of instructional effects on concept structuring was also expected to be obtained in the present study. This was to be achieved by administering the concept-similarity rating task to the undergraduate statistics students both prior to and on completion of their formal statistics instruction, and by observing the mean differences in similarity to the criterion measure between these two measures of structure. It seemed reasonable to expect that, regardless of whether these students were provided with the orienting prose passage, post-instruction performance would be more similar to the criterion than would be preinstruction performance. However, no appropriate control was used in the present study to establish the validity of an instructional-effect interpretation for such a result.

The following chapter provides a detailed account of the exact methodology employed in the present study. Chapter III includes a description of the analyses applied to evaluate the hypotheses and predictions presented above, and Chapter IV offers some interpretations of those results in terms of their implications for future structure research.

CHAPTER II

METHOD

Design

Two variables were directly manipulated in this study: degree of familiarity with basic statistics concepts, and exposure to a context-setting prose passage to assist judgements about the strength of interconcept relatedness. The first variable was investigated by selecting three different classes of subjects: undergraduate students currently enrolled in a psychology statistics course, undergraduate students who had never taken a course in statistics, and psychology graduate students with relatively extensive statistics experience. Each class of subjects was divided into two groups to assess the effects of exposure to the prose passage (described below). One group from each subject class received the passage prior to engaging in a concept-similarity rating task involving a set of key statistical concepts (also described below), and the other group did not receive this passage.

Statistically naive and graduate student subjects completed the rating task only once. However, this task was administered twice to the subjects enrolled in the statistics course, both before and after their being presented with an

instructional sequence involving the key concepts, to permit an investigation of instructional effects on cognitive structure. (The prose passage was presented only during the postinstruction assessment of structure for these subjects.) This class of subjects also completed a statistics achievement test (described below) after the postinstruction rating task, in order to compare task performance with achievement on test items at different levels of Bloom's (1956) cognitive taxonomy.

In addition to the two manipulated variables, one other factor was investigated in this study--the effects of order of concept presentation on relatedness judgements. Evidence for such an effect was sought by presenting all subjects with a rating task comprising both potential concept orders of all possible pairs of an equal number of superordinate and subordinate concepts. Superordinate concepts were defined as general, important considerations in inferential statistics, and subordinate concepts were defined as more specific factors that each might be subsumed under one of the superordinates. Two relatedness half-matrices were constructed from each subject's pair ratings. One half-matrix included those concept pairs in which the superordinate concepts were paired with and presented before the subordinate concepts, and those concept pairs with one order of superordinate-superordinate and subordinate-subordinate concept presentation. The other half-matrix included the remaining

concept pairs, identical to those in the first half-matrix, but with the reverse order of concept presentation. These half-matrices were expanded to create two symmetric relatedness matrices for each subject. Structure representations generated from these matrices were then compared to determine whether concept presentation order significantly affected these representations.

Rating task performance for all subjects was evaluated according to correspondence with criterion concept relatedness ratings obtained from the performance of the statistics course instructor on an identical rating task administered before beginning the instructional sequence. This instructor did not receive the prose passage prior to administration of the rating task, because his familiarity with the course content was expected to make his responses much less a product of text peculiarities or episodic remembrances than would be the responses of subjects with considerably less content-relevant expertise.

Subjects

The first class of subjects in this study were 30 undergraduate students enrolled in Psychology 240D (Statistics in Psychology). Twenty-six of these students took part in both the pre- and the postinstruction phases of the study, and four

others were available only for the postinstruction phase. Additionally, 26 statistically naive undergraduate students majoring in psychology, and 18 psychology graduate students volunteered to participate. The former group of subjects had no previous exposure to formal statistics instruction, whereas the latter group had successfully completed at least two graduate-level statistics courses.

Material

Context-setting passage. A prose passage, 330 words in length, was written to provide subjects with a context for completion of the concept-similarity rating task. This passage (see Appendix) described the design, implementation, and results of a study conducted by Pressley (1976) on the use of mental imagery, and was constructed in the form of an expanded abstract, implicitly involving all the concepts included in the rating task. None of the rated concepts were directly stated in the passage, and care was taken to decrease the likelihood that degrees of interconcept relatedness might be accessed via passage syntax or concept juxtaposition. The purpose of the passage was to provide a consistent contextual schema on which all judgements about concept relatedness might be based.

Rating task. The structure exercise in this study was a variation of a standard concept-similarity rating task, comprising all possible pairs of ten key concepts from inferential statistics. These concepts were selected according to the judgement of the experimenter, after consultation with the statistics course instructor, and included five superordinate and five subordinate concepts discussed within the five text chapters (Minium, 1978) that were covered by the statistics students during their instructional sequence on inferential statistics. Superordinate concepts were defined as those referring to some general but centrally important consideration in inferential statistics, and included the terms hypothesis testing, estimation, statistical significance, probability, and error. Subordinate concepts each referred to some more specific statistical consideration, and were judged to be reasonably subsumed by at least one of the superordinates. Subordinates included the terms region of rejection, confidence interval, Z_{crit}, alpha, and $S_{\bar{x}}$.

Concepts were paired in all possible combinations, resulting in 45 unique pairs. Twenty-five of these pairs included one superordinate and one subordinate concept, and twenty included either two superordinates or two subordinates. An additional 45 pairs were created by reversing the order of concept presentation observed for the first

45 pairs. Thus, the rating task was composed of 90 concept pairs, half with one order of concept presentation, and half with the reverse order. Pairs were then organized into eighteen sets of five, such that no concept was contained more than once in each set. Nine pair-sets included superordinate concepts preceding subordinate concepts and one order of superordinate-superordinate and subordinate-subordinate presentation (A sets), whereas the other nine sets included subordinates preceding superordinates and the reverse order of superordinate-superordinate and subordinate-subordinate presentation (B sets). Each pair set was printed on a single page, and pages were organized to compose two forms of a rating task booklet. Form 1 included, in order, five A sets, nine B sets, and four A sets. Form 2 included five B sets, nine A sets, and four B sets. The minimum distance between pairs containing the same two concepts was 22 pairs for Form 1 and 21 pairs for Form 2. The mean distance between like pairs was 41 pairs for Form 1 and 39 pairs for Form 2. Following each concept pair on all pages of both forms was a 7-point numerical scale of concept relatedness, anchored by the phrases "Strong Relationship" (7), "Moderate Relationship" (4), and "Negligible Relationship" (1).

Achievement task. Two knowledge-level and two application-level multiple-choice test items were written for each con-

cept included in the rating task. Knowledge-level items required either recognition of the correct definition of a concept from four alternatives, recognition of the correct concept from among four alternatives, given that concept's definition, or recognition of an important characteristic of a given concept, also from among four alternatives. Application-level items each required the solution, in prose or mathematical form, of problems directly involving one of the two concepts, and were also written in a four-alternative format. Care was taken in the construction of application-level items to minimize the likelihood that these items were iterations of examples or problems encountered by the statistics students during their instruction. For the purpose of calculating split-half test reliability, the resulting total of 40 achievement test items were organized in a counterbalanced fashion to compose the achievement measure. That is, both odd and even numbered questions contained an equal number of alternating knowledge- and application-level items. Further, two forms of the achievement measure were constructed, one with the reverse order of item presentation from the other. A complete set of the achievement test items is contained in the Appendix.

Procedure

Immediately prior to beginning the instructional sequence dealing with inferential statistics, students enrolled in the statistics course completed one of the two forms of the rating task. Subjects were told that the purpose of this activity was to provide information that might be useful in the construction and validation of a measurement device that would reflect how statistical concepts may be organized in memory. Subjects were then told that the nature of the task was to make judgements about the extent to which certain statistical concepts are interrelated. The format of the rating task (i.e., concepts grouped in pairs, judgements made on a 7-point scale, etc.) was explained, and several examples were provided. Subjects were required to rate all pairs on each page of the task by circling the number following each pair that best described the strength of that pair's interconcept relatedness. Unfamiliar concepts were to be circled, and ratings of pairs containing such concepts were to be made according to a "best guess." Subjects were not permitted to return to a completed page of the rating task once having turned to a subsequent page. All instructions (see Appendix) were printed on a separate page, and subjects were permitted to refer back to these instructions whenever necessary. Subjects were provided with as much time as

needed to complete this activity.

Five weeks later, following completion of the instructional sequence, these subjects completed the other form of the rating task, with identical instructions, but with two additional components. Subjects at this time were randomly organized into two groups. One group was provided with the context-setting prose passage, and was told to read it very carefully before beginning the rating task. This group was instructed to base all judgements about concept relatedness <u>solely</u> on the nature of relatedness that they thought was best illustrated by this passage. Even if they believed that the concepts might be related in other ways, they were told to restrict their judgements to the relation types implied in the passage. After reading the passage, they completed the rating task, referring back to the passage as needed. The other group of these subjects completed the rating task without exposure to the prose passage. Once subjects in either group completed the rating task, they were asked to answer all items on the statistics achievement test to the best of their abilities. It had been intended to provide these subjects with as much time as necessary for them to finish the rating task and the achievement test. However, an unexpected deluge of questions by the students regarding their upcoming course final examination severely limited the time available for their research participation.

Consequently, both components of the postinstruction activity were administered in only about 30 minutes.

At approximately the same time as the postinstruction measures for the statistics students, the statistically naive undergraduates and the psychology graduate students performed the rating task. As previously noted, half of the subjects in both these groups received the prose passage prior to beginning the rating task, and half did not receive the passage. Instructions to these subjects were identical to those given to the statistics students.

Scoring

The primary method of comparing subject and criterion representations of structure was to calculate the Euclidean distance, across all concept pairs, between an individual subject's relatedness ratings and the ratings of the course instructor. Euclidean distance was determined by summing the squared differences between subject and criterion ratings for each pair on the task, taking the square root of this sum, and dividing the result by the total number of pairs rated. Before calculating this value, subject and criterion ratings were normalized by converting them to z-scores, based on each individual subject's (and criterion's) mean overall rating across all 90 pairs. This was done in order

to control for varying subjective interpretations of the rating scale, such that all ratings would reflect the same relative degree of interconcept relatedness. Calculation of Euclidean distance has the disadvantage of generating a single numerical score for each subject, rather than a graphic representation of structure. Thus, it only indicates to what extent subject ratings differ, on the average, from the criterion, and does not provide direct information regarding the manner in which concepts are being inappropriately related. Nevertheless, this scoring method was used because of the relative ease with which it may be applied, and because it provides sufficient information for the investigations of prose-passage effects and of relations between rating task and achievement test performance. Multidimensional scaling routines were also applied to rating task data in an analysis of concept-order (criteriality) effects on structure representation.

CHAPTER III

RESULTS

The data in this study were analyzed to investigate three factors: the effects of formal instruction on concept relatedness ratings, the existence of a concept presentation-order effect on strength of perceived interconcept relatedness, and the usefulness of the prose passage as a means of providing a consistent context for rating task judgements. The results of analyses pertinent to each of these factors are described in the following sections.

Instructional Effect

At noted in the previous chapter, 26 undergraduate statistics students participated in both the pre- and the postinstruction phases of this study. It was expected that the concept-relatedness ratings of these students would be more similar to the ratings obtained from the course instructor after the students had been exposed to the instructional sequence than they would be before beginning the sequence, demonstrating that instruction had served to modify the manner in which those concepts were interrelated within each student's cognitive structure. One indicator of this effect was obtained by

by calculating the Euclidean distance, for both pre- and postinstruction data, between each student's normalized relatedness ratings and the normalized criterion ratings. (See section on scoring in the previous chapter.) Two preinstruction values were calculated for each student--one based on ratings for all 90 pairs of concepts, and one based on ratings of those concept pairs wherein both concepts were uncircled and thus familiar to the student. These same values were also calculated for postinstruction ratings, the latter value obtained by including only those concept pairs familiar to the students during the preinstruction phase. This procedure allowed the comparison of overall ratings, which is the traditional approach to investigating an instructional effect in cognitive structure research. It also allowed a comparison based only on concepts known to the students before beginning instruction, which would indicate the extent to which the instruction had modified the nature of pre-existing structure. This latter comparison seems to provide more meaningful evidence for instructional effects because it does not incorporate concepts that are not a part of a person's semantic memory structure.

The mean Euclidean distances across the 26 subjects who participated in both phases of the study are included in Table 1. It should be noted that 14 of these subjects were presented with the prose passage during the postinstruction

phase, and 12 were not. However, no differences were observed between the mean postinstruction distance values for these two groups; therefore, the postinstruction data were collapsed to provide a single group mean. As indicated in Table 1, the undergraduate statistics students (S) were closer in their normalized postinstruction relatedness ratings to the normalized criterion ratings, both for overall ratings (OR) and for ratings of concepts known prior to instruction (KR), than they were in their preinstruction ratings. Analyses of mean difference scores for both types of ratings indicated that these changes in perceived concept relatedness were significant, $t(25) = 6.67$, $p < .001$, and $t(25) = 3.30$, $p < .005$, respectively, thus providing some evidence that instruction had modified cognitive structure.

Other data pertinent to the existence of an instructional effect are also presented in Table 1--specifically, mean Euclidean distance scores (from the criterion) for the statistically naive (SN) undergraduate students and the statistically more experienced psychology graduate students (GS). Half of both of these groups of subjects also either were or were not provided with the prose passage before engaging in the rating task. As with the S students, however, mean distance scores did not differ between these two conditions within either group. Therefore, distance scores were collapsed across conditions to provide single group means.

TABLE 1

MEAN EUCLIDEAN DISTANCES BETWEEN SUBJECT AND CRITERION NORMALIZED RATINGS

	Subject Group			
	Pre-S	Post-S	SN	GS
Rating Type				
OR	.148	.124	.152	.112
	(.011)[a]	(.015)	(.008)	(.011)
KR	.355	.282	.361	---
	(.167)	(.129)	(.111)	---

[a]Values in parentheses are standard deviations.

Analyses of these data indicated no difference between pre-instruction-S subject ratings and the ratings of the SN subjects, whether comparisons were made between overall-rating distance scores or between distance scores based only on the ratings of pairs of familiar concepts. In contrast, post-instruction-S subject distance scores, as has already been noted, significantly differed from preinstruction scores; but, these scores did not reach the level of similarity to the criterion evidenced by the distance scores of the GS subjects. Postinstruction-S subject performance was significantly less expert-like than was GS subject performance, $\underline{t}(42) = 2.83$, $\underline{p} < .01$ (two-tailed). Such results are also consistent with the notion that instruction can serve to modify the nature of concept organization within semantic memory.

Supplementary evidence of an instructional effect was sought by calculating the correlations, for each subject in all groups, between the non-normalized ratings of concept pairs in one half-matrix and the ratings of the corresponding pairs (i.e., those containing the same two concepts) in the other half-matrix. It might be expected that such correlations would be relatively low for preinstruction-S subjects and for SN subjects, because these individuals would lack a well-organized memorial concept structure to direct consistent relatedness judgements. Postinstruction S subjects, however,

might be expected to demonstrate higher correlations because of their increased familiarity with the concepts, and GS subjects should provide even more consistent ratings, owing to their relatively greater experience with the concepts in a wide variety of contexts. Some support for these notions was obtained. The median correlation between preinstruction ratings of like-concept pairs was .43 for the 26 S subjects, and was .46 for the 26 SN subjects. This median correlation increased slightly to .52 for the same 26 S subjects, based on their postinstruction ratings, but did not reach the level of consistency in ratings demonstrated by the 18 GS subjects (Median r = .75). In comparison, the same correlation for the criterion ratings provided by the statistics course instructor was .76.

One final supplementary investigation of an instructional effect was based on a result reported by Traub and Hambleton (1974). In their study, these researchers noted that subjects tended to perceive increasing dissimilarities between concepts as a result of formal instruction. Such a notion of instruction serving to enhance a student's ability to discern subtle differences among a set of apparently similar concepts has some intuitive appeal. However, this notion was not supported in the present study. The mean relatedness rating, across all pairs, for the S subjects participating in both sessions of the study increased from

a preinstruction value of 3.68 (s.d. = .91), indicating a weak-moderate relation, to a postinstruction value of 4.71 (s.d. = .54), indicating a moderate-strong relation. This difference was significant, $\underline{t}(25) = 6.21$, $\underline{p} < .001$. When mean pre- and postinstruction ratings were calculated based only on pairs containing concepts familiar to the S subjects prior to instruction, no difference was observed. In this case, the respective mean ratings were 4.23 (s.d. = 1.10) and 4.34 (s.d. = .72). The difference observed between overall pre- and postinstruction ratings, then, seems to be best accounted for by a preinstruction response bias exhibited by most of the S subjects--specifically, a tendency to give relatively low relatedness ratings to pairs containing unfamiliar concepts. Whether a similar but opposite response bias would account for the result reported by Traub and Hambleton (1974) cannot, of course, be determined from the present study.

It was noted in the previous chapter that no appropriate control was used in this study to assure that changing responses on the rating task could be attributed solely to the effects of instruction. Nevertheless, the combination of reported differences in Euclidean distance from the criterion, strength of perceived relatedness, and consistency of relatedness judgements seems to severely detract from an interpretation based on other, uncontrolled factors.

Concept Order Effects

A major concern in this study was whether the presentation order of concepts within each concept pair would affect the strength of perceived concept interrelatedness. Based on the notion of link criteriality, one expectation was that a presentation order of "subordinate concept - superordinate concept" would result in stronger relatedness ratings than would the reverse order of concept presentation. Table 2 contains the mean difference values, for GS and postinstruction-S subjects, between non-normalized ratings of like-concept pairs for each of the three possible types of concept pairings included in the rating task: superordinate-subordinate, superordinate-superordinate, and subordinate-subordinate. These values were obtained by either 1) substracting superordinate-subordinate ratings from corresponding subordinate-superordinate ratings (Sb-Sp), 2) subtracting superordinate-superordinate ratings within one half-matrix of pairings from corresponding ratings in the other half-matrix (Sp-Sp), or 3) subtracting subordinate-subordinate ratings within one half-matrix from corresponding ratings in the other half-matrix (Sb-Sb). Once again, no differences were observed between these values for the passage and nonpassage conditions within either subject group, allowing the collapsing of data to generate single

group means. Data obtained from the SN and the preinstruction-S subjects were not included in analyses of concept order effects because both these groups demonstrated a high degree of unfamiliarity with the subordinate concepts. In fact, 23 of the 26 preinstruction-S subjects and 24 of the 26 SN subjects circled as unfamiliar three or more of the five subordinate concepts on the rating task. In contrast none of the 30 postinstruction-S or the 18 GS subjects circled any concepts as unfamiliar.

An inspection of the values reported in Table 2 would suggest that concept presentation order had little, if any, effect on relatedness ratings. In fact, none of these difference scores are significantly different from zero. An examination, across subjects, of the individual concept pairs of most interest to this investigation--that is, pairs containing one superordinate and one subordinate concept--also provided no evidence of a presentation order effect. For the S subjects, 13 of the 25 like-concept pair comparisons yielded stronger relatedness ratings when subordinate concepts preceded superordinates, and only one of these differences in strength of perceived relatedness reached significance for a two-tailed t-test. For the GS subjects, 14 comparisons yielded stronger ratings when subordinates preceded superordinates, but none of these differences was significant. Indeed, the one comparison of pair ratings for

TABLE 2

MEAN DIFFERENCES BETWEEN NON-NORMALIZED RATINGS OF LIKE-CONCEPT PAIRS

	Pair Comparison		
	Sb-Sp	Sp-Sb	Sb-Sb
Subject Group			
Post-S	.09	-.01	.09
	(.35)[a]	(.45)	(.45)
GS	-.03	-.07	.12
	(.33)	(.35)	(.38)

[a]Values in parentheses are standard deviations.

the GS subjects that did demonstrate a significant presentation order effect indicated a stronger relation when the superordinate concept preceded the subordinate.

These results are especially surprising in light of the criterion rating task performance of the statistics course instructor. Nine pairs were rated more strongly by the instructor when presented in the order "subordinate-superordinate," four were given stronger ratings when presented in the reverse order, and the remaining 12 pairs were given identical ratings for both orders. When superordinate-subordinate pair ratings were subtracted from subordinate-superordinate ratings, the overall mean difference was .60 (s.d. = 1.36). This difference was significant, $t(24) = 2.16$, $p < .05$ (two-tailed), supporting the predicted presentation order effect, and in direct contrast to the results obtained from analyses of the S and GS subject data. In order to determine whether this result was merely an anomaly, the rating task was readministered to the course instructor three weeks after completion of the postinstruction phase, and the same mean difference in ratings was calculated for these data. Once again, concepts were judged as significantly more strongly related when subordinates preceded superordinates, $t(24) = 2.20$, $p < .05$, the mean difference in this case being .40 (s.d. = .89). Perhaps the most reasonable interpretation of these data, given the results

observed for the S and GS subjects, is that they were the product of a response bias on the part of the course instructor. That is, because of his involvement in the design of this study and his consequent familiarity with the expected effects, he may have unintentionally differentially rated concept interrelatedness in the predicted direction. Even if this were not the case, however, insufficient evidence was obtained in the present study to demonstrate a reliable effect of concept presentation order on strength of perceived interconcept relatedness, at least in terms of mean relatedness ratings across subjects or across pairs.

A second planned investigation of presentation order effects was to involve correlating three measures of post-instruction-S subject performance on the rating task with performance on the application-level items included in the postinstruction achievement test. These three measures were the Euclidean distances between normalized subject and criterion ratings for 1) the half-matrix of relatedness ratings containing pairs in the order "superordinate-subordinate" (Sp-Sb), 2) the half-matrix of ratings containing pairs in the order "subordinate-superordinate" (Sb-Sp), and 3) the overall matrix of ratings containing pairs in both presentation orders. The purpose of calculating these correlations was to determine whether one order of concept presentation would yield a better pre-

dictor of high-level achievement than would the other order, or whether the best predictor of achievement could be obtained by presenting subjects with both orders of concept pairing.

Unfortunately, the unexpected time constraints placed on completion of the postinstruction phase of the study, as noted in the previous chapter, did not permit most subjects to answer all items on the achievement test. One-third of the postinstruction-S subjects completed 10 or fewer of the 20 application-level test items. In addition, a floor effect was observed for the proportion of application-level items completed that were answered correctly. The postinstruction-S subjects demonstrated a mean success rate on these items of slightly less than 27%, or about the level that might be expected to occur by chance for 4-alternative items. For these reasons, subject performance on the application-level items was judged to be unsatisfactory for use in analyses of concept order effects. Instead, the three different distance scores were correlated with each subject's combined performance on two classroom examinations that covered the same content included in the rating task. It was recognized that these tests may have been imperfect measures of the type of high-level achievement intended for this study, because performance on them was a product, to some extent, of individual pre-examination

study habits and of frequent practice with similar problems during instruction. Nevertheless, it was hoped that these examinations required sufficient application-level skills to serve the purpose of this investigation.

The correlations between achievement and Sp-Sb, Sb-Sp, and overall rating distance scores were -.122, -.177, and -.164, respectively. Surprisingly, although each of these correlations is in the expected direction--that is, each is negative, indicating that smaller Euclidean distance from the criterion was associated with higher achievement--none are significantly different from zero. Prior research incorporating considerations similar to those operationalized in the present study (e.g., Konold & Bates, Note 1; Bates, Note 3) has demonstrated a consistent significant relation between achievement and performance on a concept-similarity rating task. Why no such relation was observed in the present study is not clear, but this question will be addressed in the following chapter.

The final planned investigation of concept-order effects involved comparing graphic multidimensional scaling solutions obtained for each subject's Sp-Sb and Sb-Sp half-matrices of similarity ratings, and for the half-matrix obtained by averaging the ratings of like-concept pairs. The purpose for conducting this analysis was to determine whether scaling procedures applied in prior structure

research have been appropriate. In Chapter I, it was noted that researchers generating scaling solutions for concept-similarity rating data have consistently done so based on the assumption that such ratings will be symmetric irrespective of the order of concept presentation. One question raised in that discussion was whether the assumption of symmetry might be ill-founded, owing to the notion of link criteriality as expressed in Collins and Loftus' (1975) theory of concept organization within memory. Specifically, it was suggested that different orders of concept presentation might yield qualitatively different scaling representations of structure, and, consequently, different judgements about students' understanding of those concepts.

Separate multidimensional scaling solutions in one, two, and three dimensions were obtained for the three half-matrices generated from each postinstruction-S and GS subject's relatedness ratings, as well as from those of the statistics course instructor. Preinstruction-S and SN subject data were not included in this analysis because the high degree of concept unfamiliarity exhibited by these subjects would not have permitted meaningful interpretations to be made about the patterns of their scaling solutions. A three-dimensional configuration was consistently most appropriate for relatedness ratings across all half-matrices in both subject groups, in terms of Kruskal's

(1964) criterion of the "elbow" in a plot of dimensions versus stress. However, acceptable stress (i.e., <.10) was generally obtained when ratings were scaled in two dimensions, and the patterns of these two-dimensional configurations of the statistical concepts were more interpretable than were those of their three-dimensional counterparts.

Figures 5, 6, and 7 are two-dimensional representations of the scaling solutions for the Sp-Sb, Sb-Sp, and average criterion half-matrices, respectively. Two relatively distinct clusters of concepts seem to be included in both the Sb-Sp and average rating configurations. The first of these clusters comprises the superordinate concepts _hypothesis testing_ (HT), _statistical significance_ (SS), and _error_ (Er), as well as the subordinates _region of rejection_ (RR), _alpha_ (A), and Z_{crit} (Z). The second cluster comprises only the superordinate _estimation_ (Es) and the subordinate _confidence interval_ (CI). These clusters might reasonably be understood as describing a continuum of statistical concept application ranging from the testing of experimental hypotheses to the estimation of population parameters. In both configurations, the concepts _probability_ (P) and $S_{\bar{x}}$ (Sx) appear as outliers, not clearly associated with any other concepts. The configuration for the Sp-Sb ratings (Figure 5) also includes the larger of the two concept clusters, but differs from the other struc-

ture representations in that confidence interval is closely associated with probability, and estimation has joined $S_{\bar{x}}$ as an outlier. This visual comparison of the criterion structure representations seems to indicate that concept presentation order had little effect on the graphic scaling solutions for the relatedness ratings beyond a slight repositioning of one or two of the concepts.

The computer scaling program used to generate these representations of structure (KYST-2) did not permit a direct quantitative analysis of qualitative differences among the configurations; therefore, two indices of structure appropriateness were independently derived from the criterion data to allow such comparisons. The first of these was an index of clustering (I_c), obtained by 1) finding the centroids for the data points included in both the "hypothesis testing" and the "estimation" concept clusters described above, and 2) calculating the combined Euclidean distance from each data point to its respective centroid. Such an index would decrease, approaching zero, with an increase in the "tightness" of concept clustering around both centroids. The second index reflected perceived dissimilarity between the concepts included in both clusters (I_d), and was obtained by calculating the linear distance between the centroids of the two concept clusters. These two indices were then combined to provide a single numerical index of structure

(I_s) by dividing the I_d obtained for each scaling solution by its corresponding I_c. A larger I_s for a configuration representing one half-matrix, relative to the I_s of another half-matrix, would then reflect a combination of greater perceived strength of relatedness for the concepts within both clusters, and lesser perceived strength or relatedness between the clusters.

The respective I_s values for the criterion configurations presented in Figures 5, 6, and 7 are 6.74, 10.24, and 9.71. The relative magnitude of the first two of these indices is somewhat consistent with the results reported above for the difference in non-normalized criterion relatedness ratings between like-concept pairs containing one superordinate and one subordinate. That is, concepts were rated by the statistics course instructor as more strongly related when subordinates preceded superordinates; and, the configuration of concepts generated via scaling demonstrated tighter, better differentiated clustering for the Sb-Sp half-matrix than for the Sp-Sb half-matrix. For purposes of comparison with the criterion I_s values, Table 3 includes the mean I_s, across subjects, for each of the three two-dimensional scaling solutions obtained from postinstruction-S and GS relatedness ratings. Single group means are reported because no differences in mean performance were observed between passage and nonpassage conditions within these groups. As might be

TABLE 3

MEAN I_s VALUES BASED ON

TWO-DIMENSIONAL SCALING SOLUTIONS

	Matrix Scaled		
	Sp-Sb	Sb-Sp	Average
Subject Group			
Post-S	4.19	3.34	3.91
	(2.07)[a]	(1.93)	(2.11)
GS	5.15	5.09	5.39
	(2.40)	(1.54)	(2.18)

[a]Values in parentheses are standard deviations.

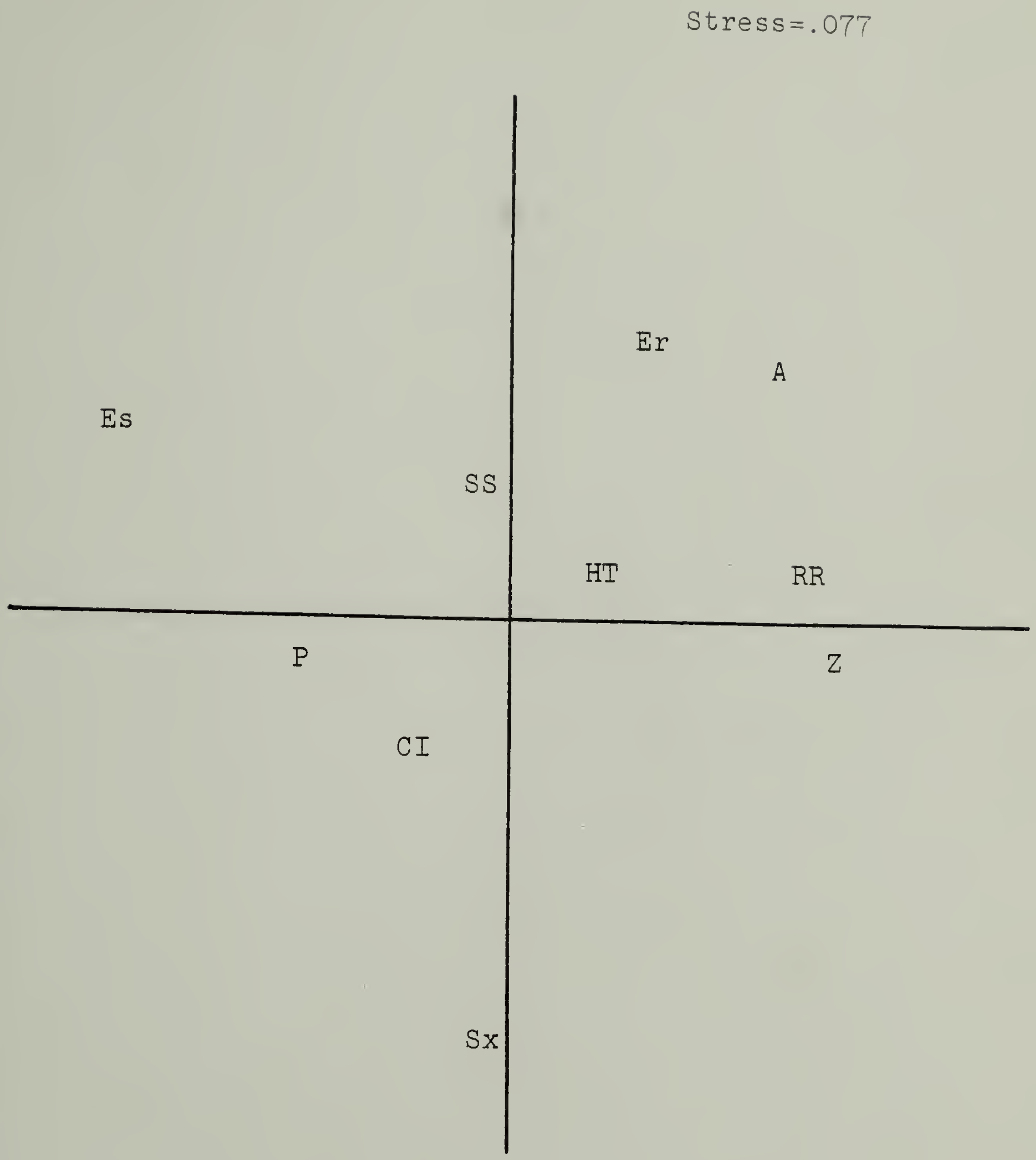


Fig. 5. Two-Dimensional Scaling Solution for the Criterion Sp-Sb Half-Matrix.

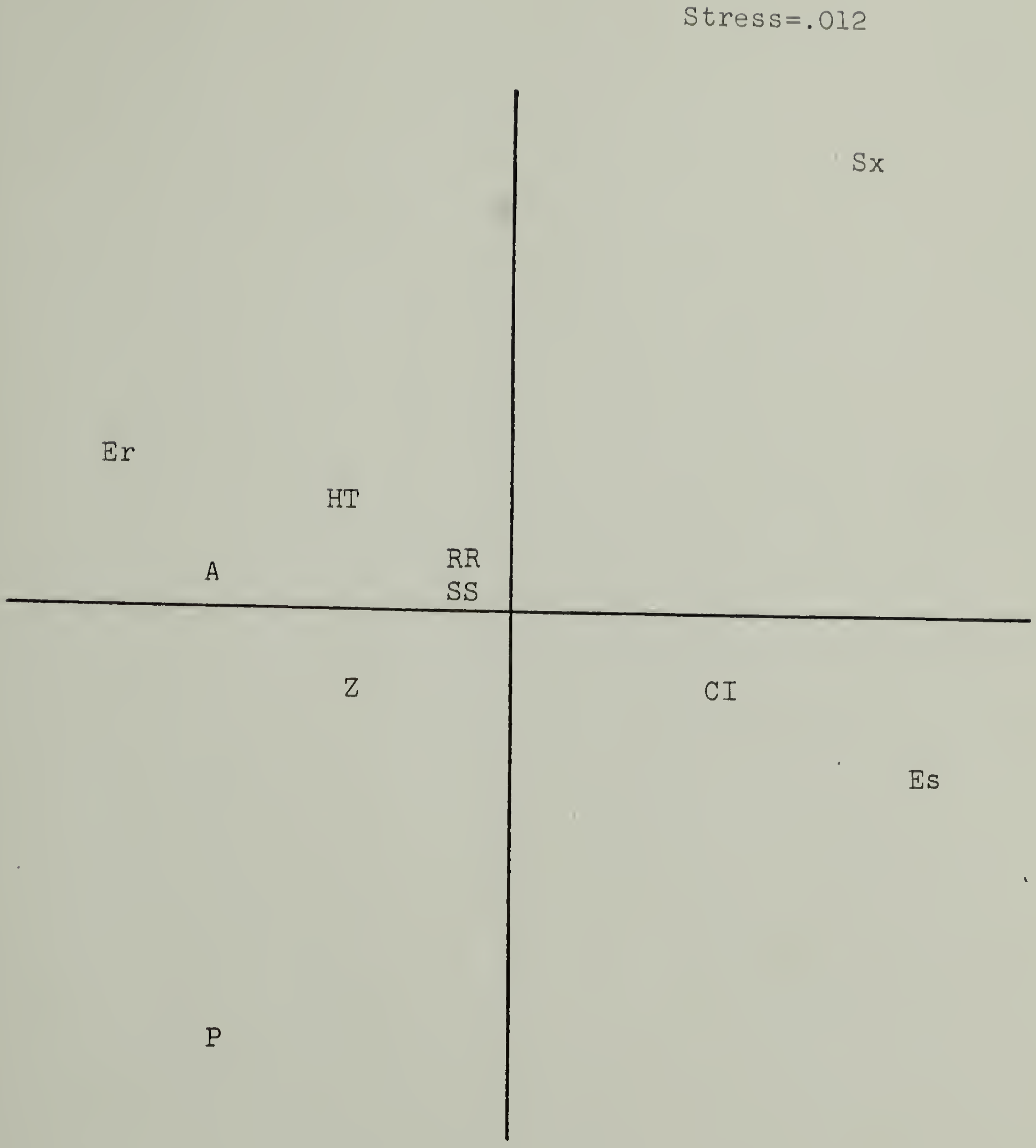


Fig. 6. Two-Dimensional Scaling Solution for the Criterion Sb-Sp Half-Matrix.

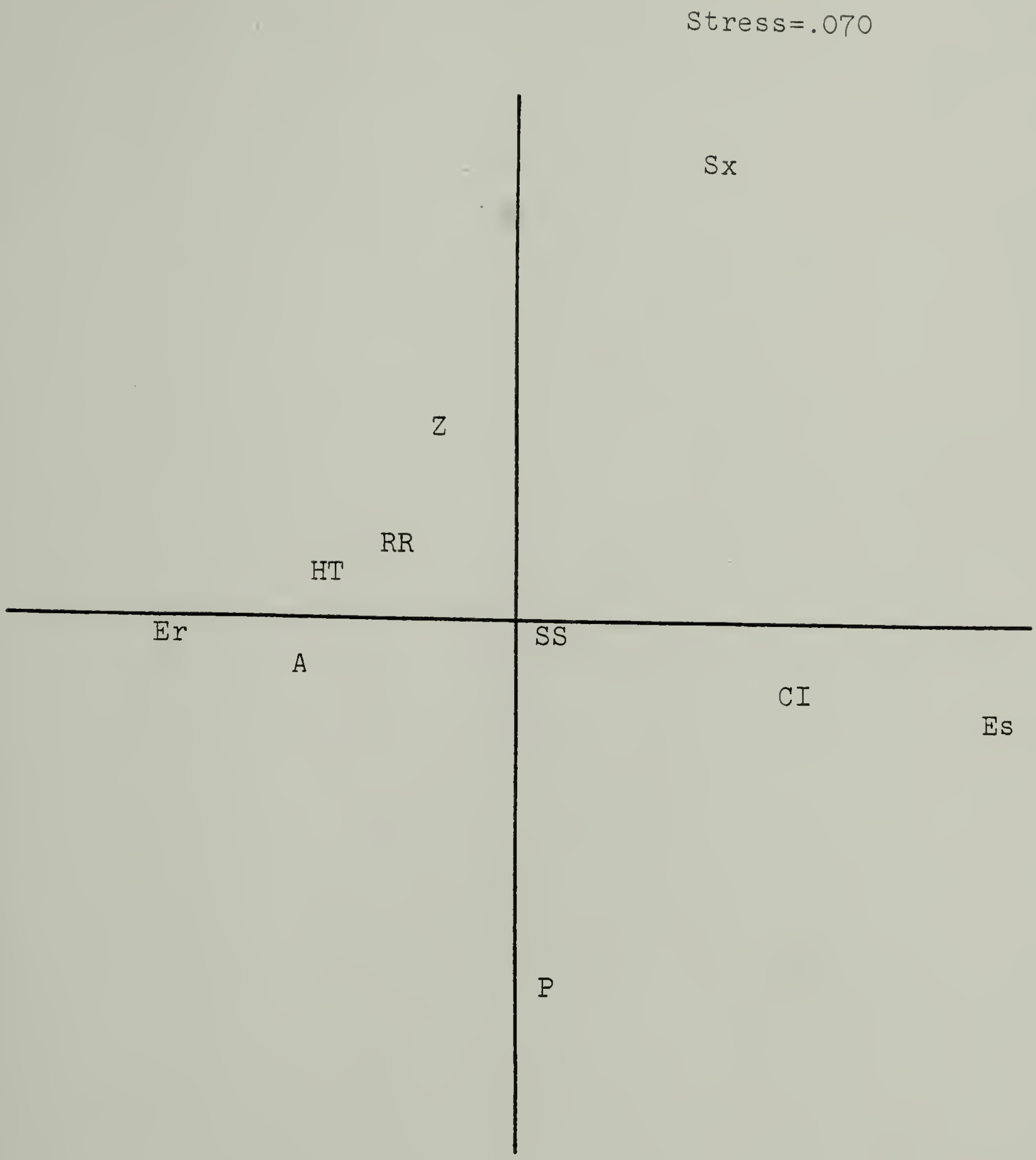


Fig. 7. Two-Dimensional Scaling Solution for the Criterion Average Half-Matrix.

expected, the I_s values for GS subject performance reflect significantly tighter concept clustering and greater cluster differentiation than do the corresponding values for the postinstruction-S subjects, both for the Sb-Sp half-matrix, $t(46) = 3.20$, $p < .01$, and for the average half-matrix, $t(46) = 2.27$, $p < .05$. The difference between Sp-Sb half-matrix I_s values, although not statistically significant, is also in the direction indicating GS subject superiority.

Figures 8-10 are graphic representations of the respective two-dimensional scaling solutions for the Sp-Sb, Sb-Sp, and average relatedness rating half-matrices generated by a representative postinstruction-S subject. The corresponding configurations for a representative GS subject are included in Figures 11-13. A visual comparison of these configurations seems to support the outcomes of the quantitative analyses of I_s data. That is, GS subjects tended to generate more tightly organized, better differentiated concept clusters than did postinstruction-S subjects. Despite these between-groups differences, however, the trend described by the criterion I_s values is clearly not present in the S and GS subject data reported in Table 3. Not only are the differences in mean I_s between the Sp-Sb and the Sb-Sp half-matrices nonsignificant <u>within</u> both subject groups, but the direction of these differences is also the

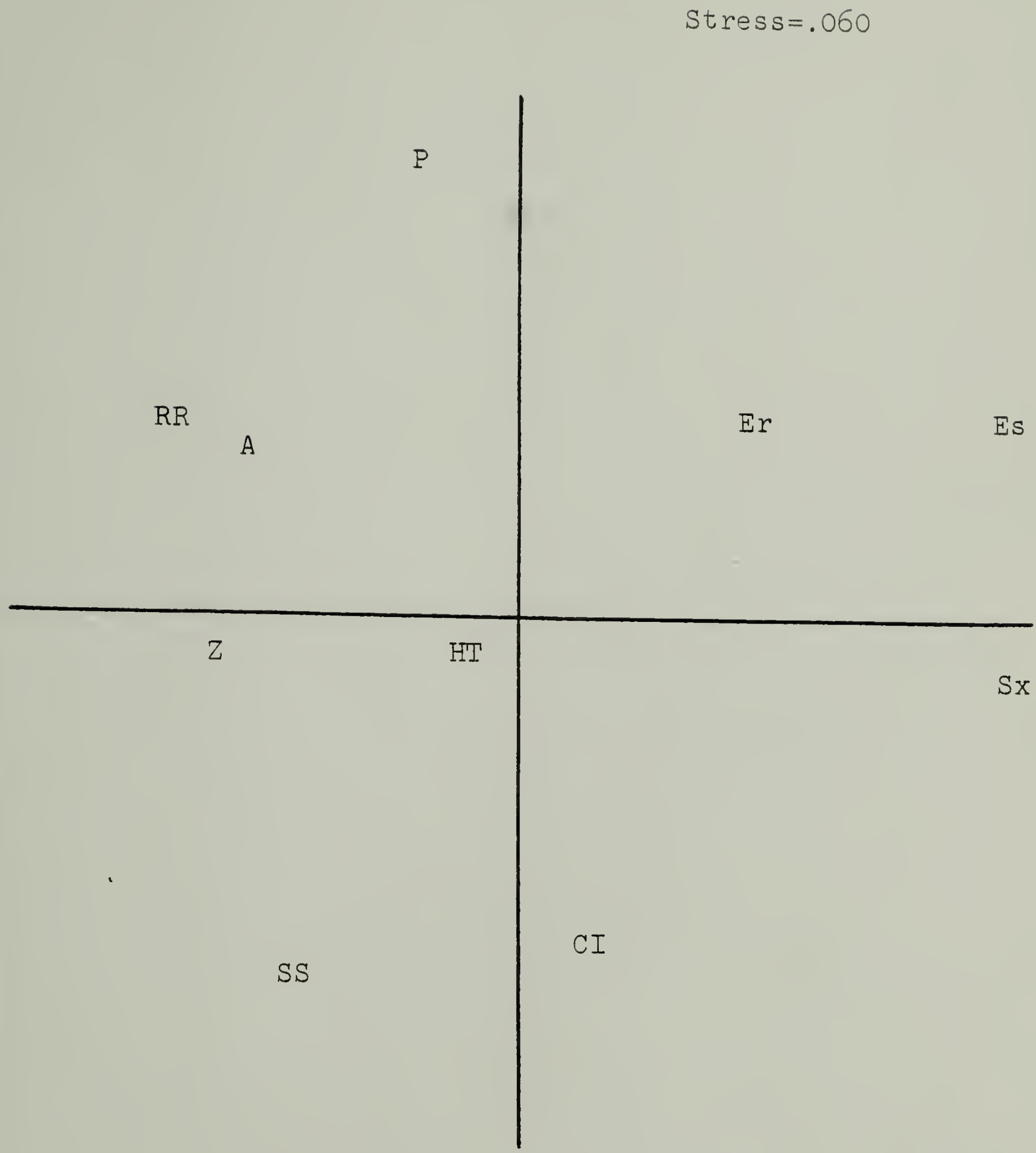


Fig. 8. Two-Dimensional Scaling Solution for a Representative Statistics Student's Sp-Sb Half-Matrix.

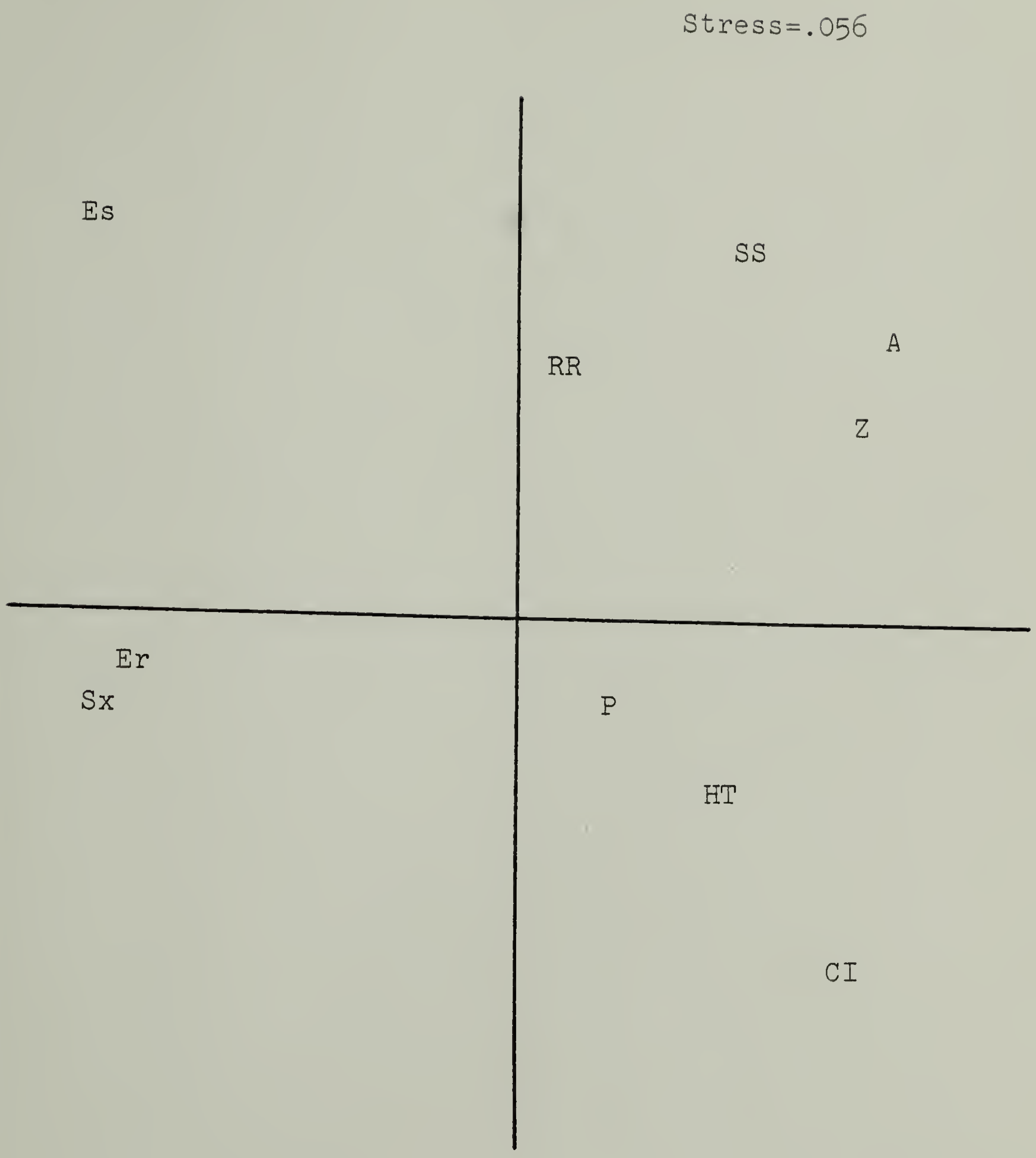


Fig. 9. Two-Dimensional Scaling Solution for a Representative Statistics Student's Sb-Sp Half-Matrix.

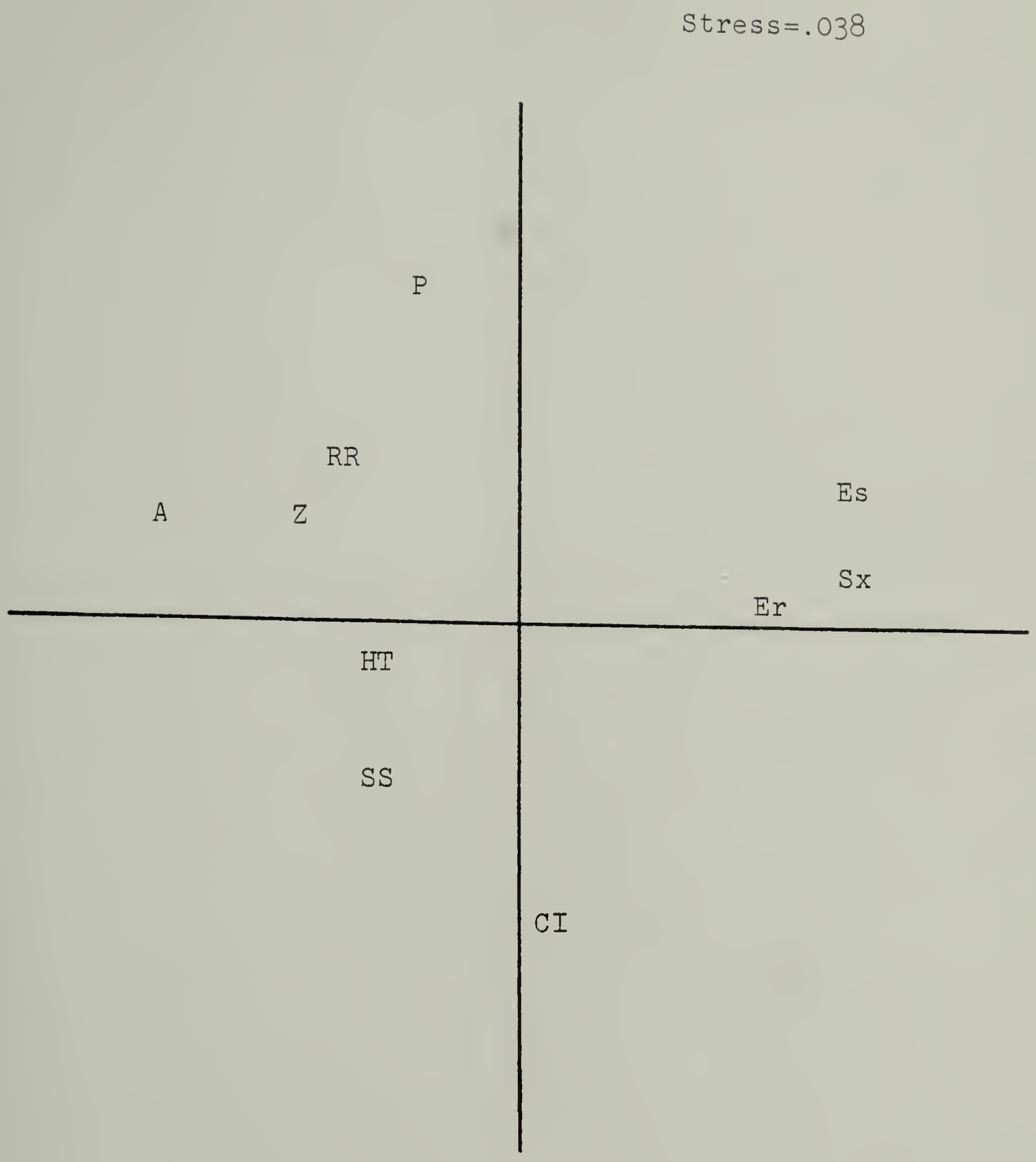


Fig. 10. Two-Dimensional Scaling Solution for a Representative Statistics Student's Average Half-Matrix.

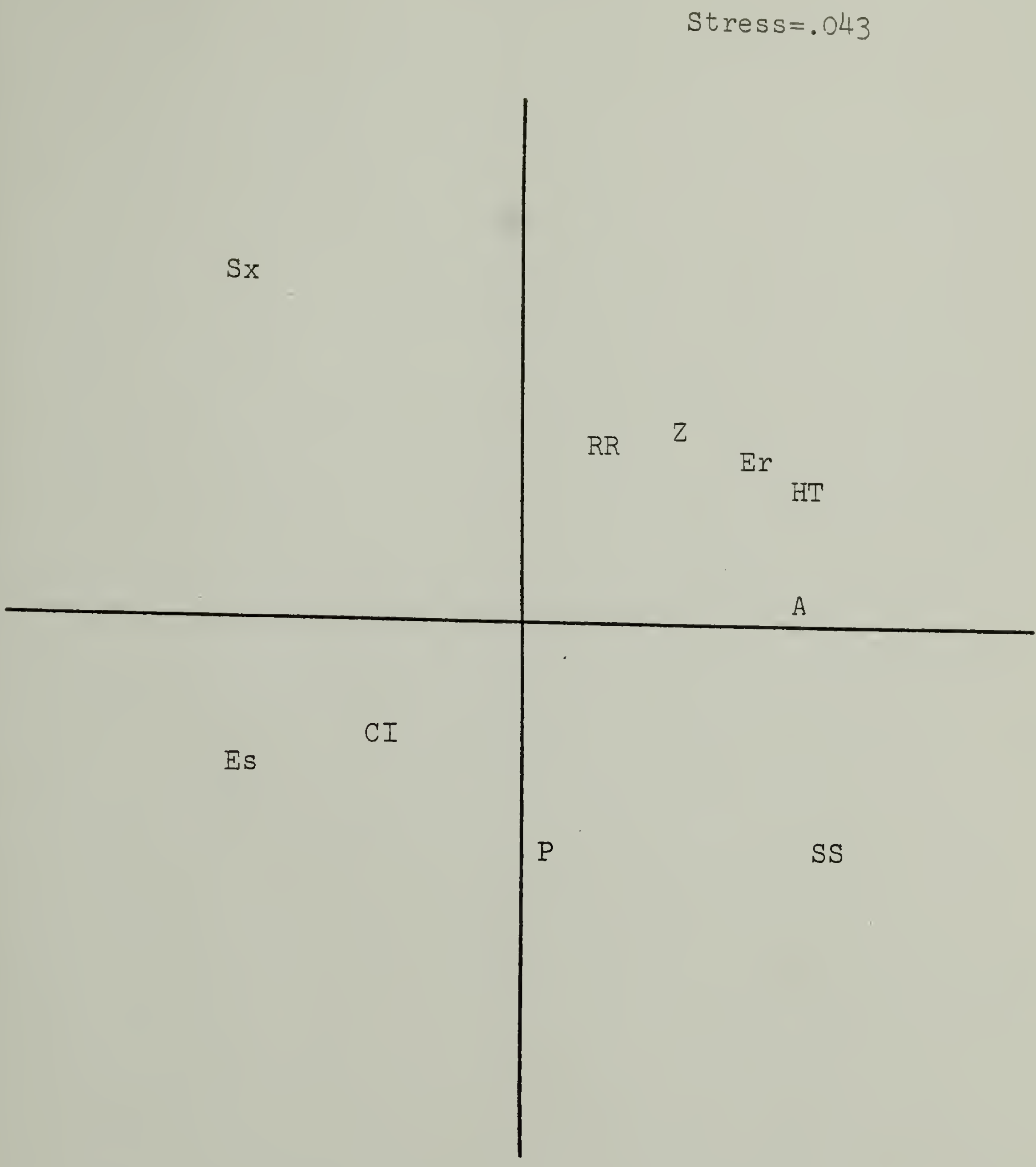


Fig. 11. Two-Dimensional Scaling Solution for a Representative Graduate Student's Sp-Sb Half-Matrix.

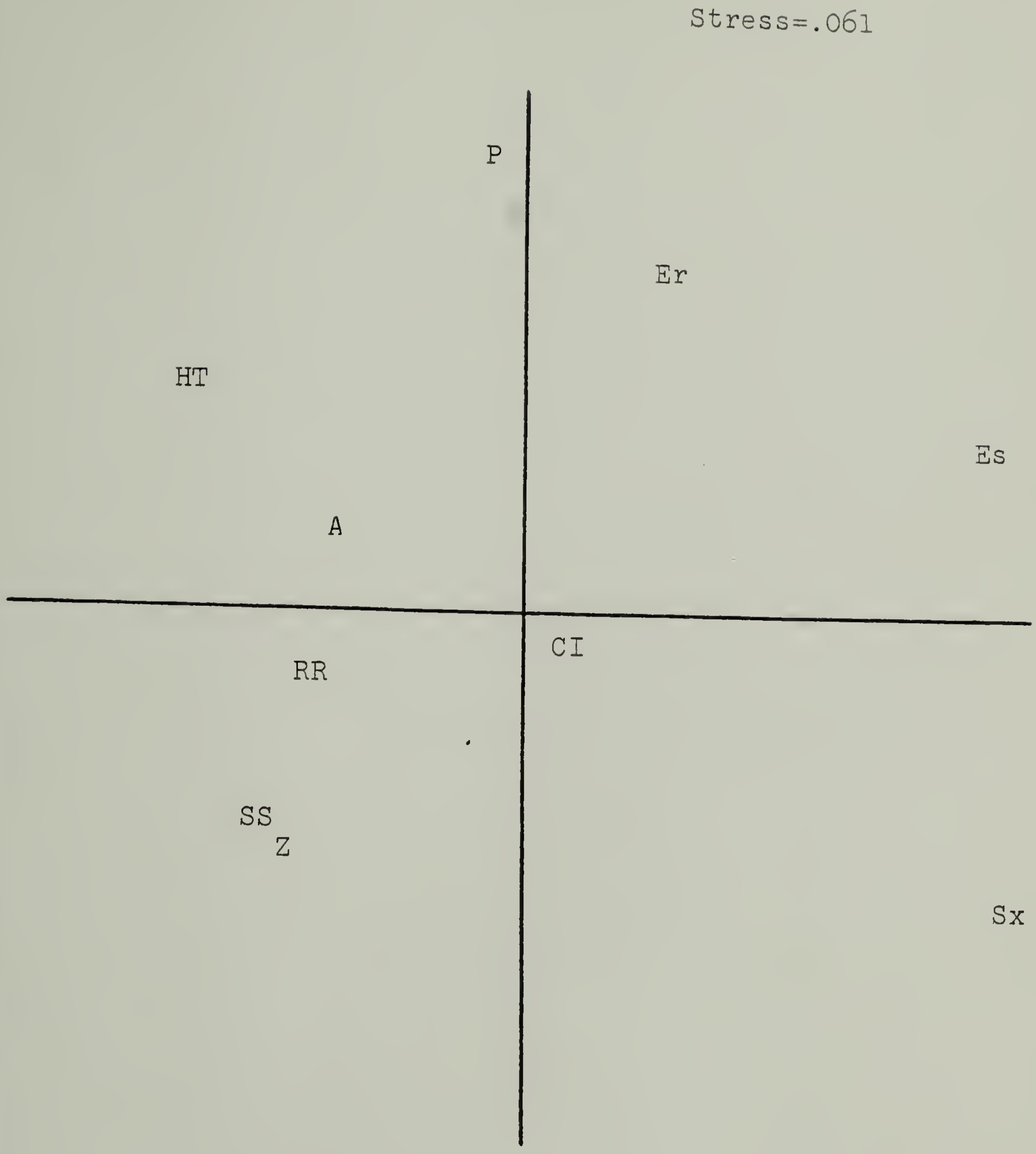


Fig. 12. Two-Dimensional Scaling Solution for a Representative Graduate Student's Sb-Sp Half-Matrix.

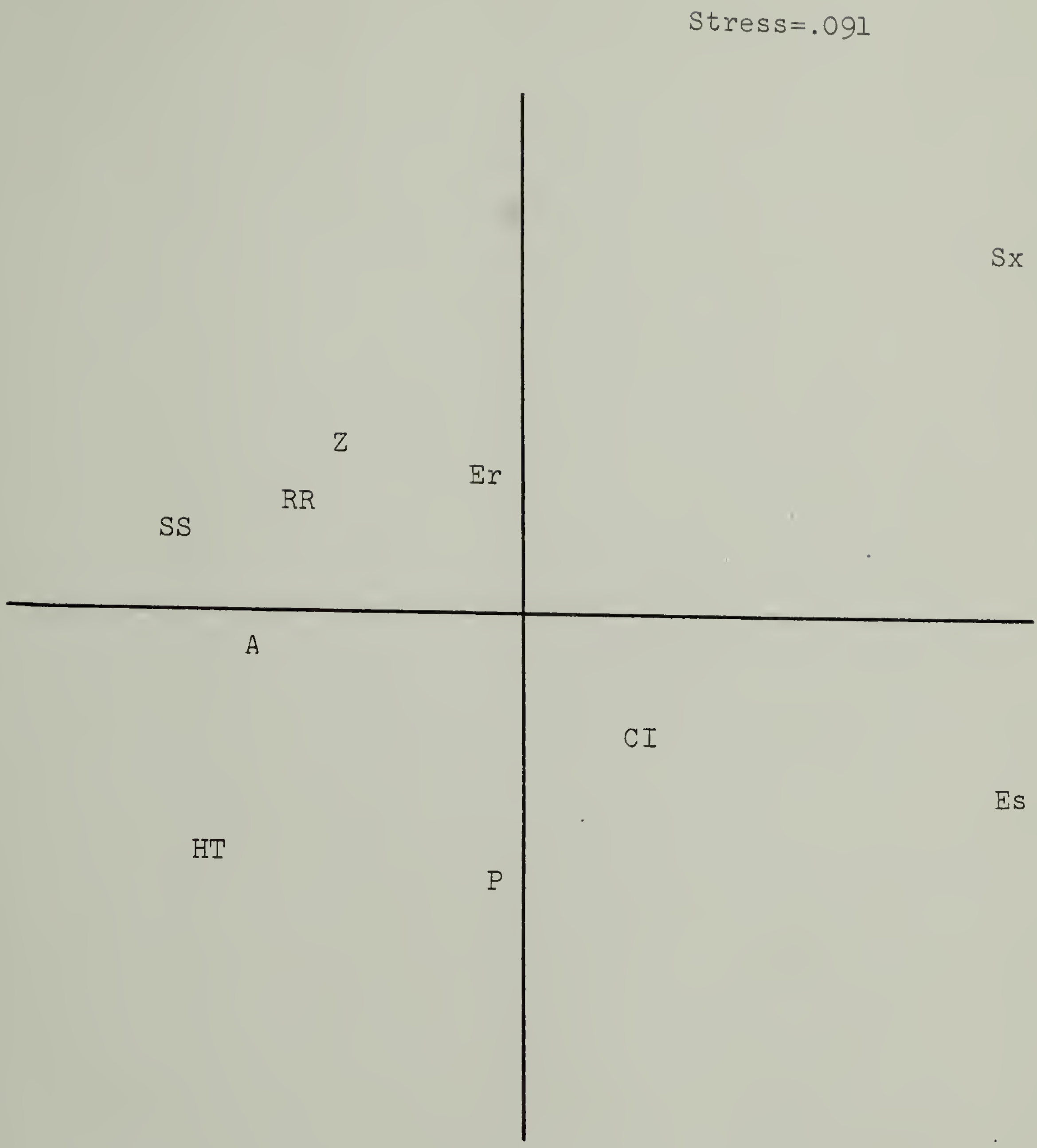


Fig. 13. Two-Dimensional Scaling Solution for a Representative Graduate Student's Average Half-Matrix.

opposite of that demonstrated by the criterion. If concept presentation order has any reliable effect on the scaling solutions obtained for concept-similarity rating data, the nature of such an effect is not at all obvious from the results of the present study.

Prose Passage Effects

The goal of providing subjects with the prose passage prior to performing the rating task was to encourage their use of a consistent frame of reference when making relatedness judgements. It was hoped that this specification of the context appropriate for task participation would limit the number of potential types of perceived interconcept relations, and thus control for a source of variability in relatedness ratings that may account for some of the ambiguity observed in results of prior structure research. Two indicators of a prose passage effect were investigated: differential rating task performance, relative to the criterion, for subjects using the passage in comparison with subjects not using the passage, and differential correlations between the rating task and achievement test performances for passage and nonpassage conditions within the postinstruction-S subject group.

It was expected that the 15 postinstruction-S subjects

provided with the passage would demonstrate rating task performance more like that of the statistics course instructor than would the 15 postinstruction-S subjects not provided with the passage. No passage effect was expected for the SN or the GS groups because, in the case of the former, the passage would not substitute for a nonexistent memorial concept structure, and, in the case of the latter, the passage would provide largely redundant information. Two measures of rating task performance have been discussed in the previous sections of this chapter--the Euclidean distance between subject and criterion relatedness matrices, and the index of concept clustering/differentiation derived from scaling solutions of relatedness ratings. Group means and standard deviations for these measures, based on overall relatedness ratings (in the case of Euclidean distance) or on average ratings of like-concept pairs (in the case of I_s), are reported in Table 4. A general trend across S and GS groups may be observed in these data that actually suggests slightly <u>poorer</u> performance for subjects who received that passage, relative to those who did not. However, analyses of variance indicated that this trend was not significant. Indeed, it has already been noted that, contrary to expectations, presence of the prose passage had <u>no</u> significant effects on mean group performance based on either measure of structure for any group of subjects.

TABLE 4

MEAN RATING TASK PERFORMANCE FOR PASSAGE AND NONPASSAGE CONDITIONS, BASED ON I_S VALUES AND EUCLIDEAN DISTANCES (E)

	Condition			
	Passage		Nonpassage	
Subject Group	E	I_s	E	I_s
Post-S	.124	3.43	.121	4.40
	(.016)[a]	(2.37)	(.011)	(.168)
SN	.150	----	.153	----
	(.009)	----	(.007)	----
GS	.114	4.72	.109	6.05
	(.008)	(2.26)	(.013)	(1.87)

[a]Values in parentheses are standard deviations.

However, there is some indication that the passage may have had an effect on the relationship between rating task performance and achievement for the postinstruction-S subjects. It was expected that if the passage acted to restrict the context of relatedness ratings to a consistent semantic network, then correlations between rating task performance and application-level achievement would be stronger than they would be without the passage. This would be so because both measures of students' understanding would be more likely to be the products of students' semantic memory structures. In contrast, rating task performance was not expected to be well-correlated with knowledge-level achievement for passage or nonpassage subjects, because this type of achievement may be largely the product of temporal/spatial associative relations within episodic memory.

For reasons already mentioned, application-level performance of the S subjects on the postinstruction achievement test was not usable for analysis. Instead, rating task performance, in terms both of Euclidean distances and of I_s values, was correlated with combined achievement on the two aforementioned course examinations. Although the same small number of knowledge-level postinstruction achievement test items were answered as has been noted for application-level items, the mean proportion of these items answered correctly

(.60) was judged to be sufficiently greater than chance to permit their inclusion in these analyses. Correlations pertinent to the above predictions are reported in Table 5, including values based on Euclidean distances between subject and criterion Sp-Sb (E_{sp}), Sb-Sp (E_{sb}), and overall (E_o) relatedness matrices, and on scaling solution indices of concept clustering/differentiation for the Sp-Sb (I_{sp}), Sb-Sp (I_{sb}), and average (I_{sa}) matrices. It should be noted that, according to the predictions, correlations involving E values should be _negative_, denoting a relation between decreasing distance from the criterion and increasing performance on the achievement measures. Correlations involving I_s values should be _positive_, denoting a relation between increasing concept clustering/differentiation and increasing achievement.

Of the twelve correlations between knowledge-level achievement and rating task performance, eleven are, as predicted, nonsignificant. The single significant correlation--between knowledge-level achievement and I_{sb} for the nonpassage condition--is very likely an anomaly. _Contrary_ to predictions, however, only three of the twelve course achievement-rating task performance correlations are significant. Nevertheless, three aspects of the particular pattern of correlations reported in Table 5 may provide some indication of a prose-passage effect. First, all three of these

TABLE 5

CORRELATIONS BETWEEN PERFORMANCE ON KNOWLEDGE-LEVEL TEST ITEMS (K) OR OVERALL COURSE ACHIEVEMENT (A) AND INDICES OF RATING TASK PERFORMANCE FOR POSTINSTRUCTION-S SUBJECTS

	Condition			
	Passage		Nonpassage	
	K	A	K	A
E_{sp}	−.213	−.337	−.234	.287
E_{sb}	−.178	−.292	−.341	.039
E_{o}	−.220	−.342	−.309	.183
I_{sp}	.215	.512[a]	−.290	−.119
I_{sb}	.085	.558[a]	.594[b]	−.015
I_{sa}	.230	.594[b]	−.130	.074

[a] $p < .05$

[b] $p < .01$

significant correlations were obtained from subjects in the passage condition. Second, all six of the course achievement-rating task performance correlations for passage subjects are, as predicted, larger in the appropriate direction than their counterparts for nonpassage subjects (although none of these differences are signficant according to Fisher-*z*' transformations). Third, *all twelve* of the passage-condition correlations, including both knowledge-level and course achievement, are in the predicted directions, whereas seven of the twelve nonpassage-condition correlations are in the direction *opposite* to that predicted.

It is not clear what factor(s) may account either for this differential pattern of correlations, or for the fact that rating task performance was, in general, so poorly correlated with course achievement in the present study, relative to prior research. One potential interpretation is, of course, that the very small number of scores composing each of these correlations ($n = 15$) yielded highly unstable r values, and that *none* of the correlations are therefore particularly reliable. Another possibility is that reliance on the passage to make relatedness judgements, although not affecting mean performance, still acted to increase the *variability* of performance, thereby increasing course achievement-rating task performance correlations in the predicted direction.

Some support for this interpretation is provided by an inspection of the relative standard deviations for rating task performance within the passage and nonpassage conditions. For I_s values, the passage-condition standard deviations were 2.25, 2.24, and 2.37 (for the Sp-Sb, Sb-Sp, and average half-matrices, respectively). The corresponding values for the nonpassage condition were 1.74, 1.56, and 1.68. A similar trend may be observed in the standard deviations of E values. These statistics for the passage condition were .021, .014, and .016 (for the Sp-Sb, Sb-Sp, and overall matrices, respectively), while the corresponding nonpassage values were .013, .014, and .011. Thus in five cases out of six, greater variability in rating task performance was achieved in the passage condition, relative to the nonpassage condition. Assuming that a significant relationship between measures of cognitive structure and achievement does exist--a not altogether unreasonable assumption, based either on theory or on prior empirical evidence--then the combination of small *n* and decreased variability in the present study may have masked this relationship for nonpassage subjects. Even so, however, this does not explain *why* greater variability in rating task performance was so consistently obtained within the passage condition. This issue will again be addressed in the following chapter.

Overall, the results of the present study have

revealed little conclusive evidence that presence of the prose passage acted as predicted: to provide a more consistent framework for making judgements about interconcept relatedness. No reliable indication that the passage enhanced mean student performance was observed. Correlational analyses, although suggesting the possibility of a passage effect on performance variability, did not unambiguously parallel the results of prior structure research. What these outcomes may imply regarding future application of this tactic to control for the types of potential perceived interconcept relations in similar research will also be addressed in the following chapter.

CHAPTER IV

DISCUSSION

Two goals were pursued in this dissertation. First, the assumptions and procedures of four popular techniques to assess the structure of human memory were contrasted against the defining characteristics of a variety of psychological models of memory organization. The purpose of this critical review was to illuminate areas where theoretical inconsistencies and methodological weaknesses might exist in these techniques, and to generate some strategies that might enhance their psychological validity and educational applicability. One outcome of this review was the observation that most structure researchers seem to embrace, either implicitly or explicitly, the semantic hierarchy memory model postulated by Collins and Quillian (1972) as the theoretical foundation for their approaches. However, it was argued that two important components of this model are not adequately accounted for by any current structure measurement strategy. One of these components is the notion of link criteriality, which hypothesizes that concepts may be interconnected in memory by links that are asymmetrically weighted, such that concept "A" may be perceived to be more important to an understanding of concept "B" than "B" is to

"A." It was argued that neither the procedures used to employ structure measurement tasks, nor the statistical techniques typically used to evaluate task performance make allowances for this hypothesized asymmetrical memory organization, with the potential result that these strategies may provide inaccurate representations of structure.

A second related component of the semantic hierarchy model not accounted for by structure measurement techniques is the notion that concepts are defined not only by the number and criteriality of relations that link them to other concepts, but also by the types of these relations. It was argued that students engaging in any of the reviewed structure measurement tasks may be able to perceive many different types of concept relations, and that these tasks provide no means by which the particular relations applied to make judgements about interconcept relatedness may be differentiated. This deficiency may result in the interpretation of differing student performance on structure tasks as indicative of differing levels of understanding of content material, whereas the actual source of performance variability may be students' perceptions of different but equally correct types of concept relations.

Thus, it was concluded in Chapter I that structure measurement techniques, in their present form, may not accurately reflect human memory organization, at least as that organiza-

tion has been defined within the memory model on which these techniques appear to have been based. Structure measurement techniques were found to be even more deficient in their ability to capture the essence of other, more complex conceptualizations of human memory. The propositional memory models reviewed in Chapter I each emphasized that interconcept relations are specified by the particular predicates within which concepts may be embedded. That is, it is the context within which a concept is being applied that determines the other concepts with which it may be related, and the nature of those relations. Structure measurement techniques, just as they cannot account for types of relational links, also do not distinguish among the particular predicates that students may apply to determine degrees of concept relatedness. In a similar vein, the notion of schema memory models that schemata may direct the purpose of cognitive activities was judged to be largely unrepresented by these measurement strategies. The fundamental problem here is, once again, that student performance on a structure task may be informed by a variety of schemata reflecting different aspects of concept application, but that the task is not capable of controlling for or distinguishing among these schemata. The ultimate result may be that what is judged to the inferior task performance may simply reflect alternative and equally sophisticated systems of concept organization.

Prior structure research was also criticized on the grounds that it has not been very successful in demonstrating significant relationships between structure task performance and course achievement. It was noted that, of the 21 reviewed studies on structure measurement, only three reported such relationships. Several possible methodological sources of this lack of reasonably expected outcome were discussed in some detail, and include generally brief instructional intervals, assessing concept relations that may have been directly taught during instruction, and using achievement measures that may have been directed at very low levels of cognitive processing (i.e., recognition of correct concept definitions). It was recommended that future structure research pay special heed to these factors, in order to maximize the likelihood that meaningful structure-achievement relations be obtained. Otherwise, it was argued, structure assessment techniques may prove to be of little practical value to educators.

The second major goal of this dissertation was founded on the extensive critique of structure measurement stategies presented in Chapter I, and involved experimentally operationalizing two of the issues raised therein: the effects of differing concept presentation orders on perceived interconcept relatedness, and the ability of a context-setting prose passage to constrain types of perceived relations, such that differing levels of structure task performance

might be more accurately attributed to differing levels of understanding of content material. This study also investigated the correlation between structure task performance and academic achievement, and incorporated features such as a relatively long instructional interval (i.e., five weeks) and an achievement measure directed at assessing application-level skills, in order that such a correlation might be obtained.

To test the former of these factors, subjects were presented with a concept-similarity rating task comprising all possible pairs of 10 key statistics concepts, irrespective of concept order within the pairs. Concepts were of two types: general, superordinate statistics principles, and more specific subordinate terms directly derived from the superordinates. It was expected that subjects would tend to rate concepts presented in the order "subordinate-superordinate" as more strongly related than they would when presented with the same concepts in the reverse order. This was not merely an idle speculation, but was based both on theoretical considerations (i.e., the criteriality component of the semantic hierarchy memory model) and on prior empirical evidence (e.g., Collins & Loftus, 1975; Tversky, 1977). However, this expectation was not confirmed. The lack of reliable support for a concept presentation order effect, nevertheless, should not necessarily be interpreted as a

disconfirmation of this variable's importance to cognitive structure research. One major problem with a nonsignificant result is that there is seldom a sure way to determine <u>why</u> no effect was obtained. It may well be that pencil-and-paper measures of structure are relatively insensitive to changes in concept order, but that other assessment techniques (e.g., reaction-time studies) are affected by this manipulation. Alternatively, the lack of a presentation order effect may be attributed to the particular concepts selected for inclusion in the present study. Although an effort was made to select concepts that, a priori, would demonstrate such an effect, a more rigorous analysis of content structure may have yielded an entirely different set of concepts that would have generated the expected results. Because the investigation of potential presentation order effects within a cognitive structure research paradigm appears to be unique to the present study, further experimentation within other content areas using other measurement techniques will need to be conducted before the importance of this effect can be established.

The second major concern investigated in the present study was whether some degree of experimental control over the types of potential perceived interconcept relations could be achieved by providing subjects with a context-setting prose passage prior to their completion of the rating task.

It was expected that, if such control were achieved, this outcome would be reflected by differential mean performance on the rating task as a function of subjects' differing levels of statistics expertise. No significant passage effect, however was observed within any of the subject groups. The absence of differential mean rating task performance due to passage use may indicate a generally impotent manipulation, or may be attributed to the particular passage developed for this study. The major interpretive problem, however, is not that no effects whatsoever were obtained via passage use. Rather, it is the nature of the effect that is so perplexing. Specifically, why should students' reliance on the passage for determining relation types have resulted in an apparent increase in performance variability, while not significantly affecting mean performance?

Of course, it may well be that the passage did not affect performance variability at all, and that observed differences were only anomalous by-products of the relatively small *n* available for analyses. One alternative, hypothetical explanation is as follows. Students provided with the passage were first required to integrate this information into their memorial concept structures before it could be used as a basis for relatedness judgements. Those students with well-organized structures could probably do so; and, the better organized their internal structures were, the more

likely that the passages would have provided redundant information. In contrast, those students with poorly organized structures, reflective of a general lack of understanding of the concepts, might have failed to recognize relevant passage information, and might also have inferred relations among concepts within the passage that did not exist. Consequently, it may be that these students--equipped, as it were, with incomplete statistics schemata whose variables were now filled with inappropriate values--performed even more poorly on the rating task than they would have without the passage. Such a hypothetical scenario would account for both the slight (albeit nonsignificant) decrease in mean performance of the passage group relative to the nonpassage group, and the increased variability in rating task performance for the passage subjects.

Unfortunately, this interpretation is purely speculative, and does not account for the general lack of meaningful structure-achievement correlations in the present study, given that prior research not involving the passage manipulation has demonstrated such relations. This is a particularly important point, because, as was established in Chapter I, the ultimate educational utility of cognitive structure exercises is dependent on their ability to act as predictors of academic achievement. It may therefore be of value to contrast the methodology and materials used in the present study

against those incorporated into more successful investigations, in order to determine which factors may be most salient to achieving this goal.

The studies conducted by Bates (Note 3) and by Konold and Bates (Note 1) were the most successful of reviewed structure research in terms of demonstrating structure-achievement relations, and served as models for the design of the present study. However, there are some clear differences between the former two investigations and the latter. Most obvious of these differences is the length of the concept-similarity rating task used to evaluate structure. In the Bates study, 36 concept pairs were rated; in the Konold and Bates study, the number was 40. The present study, owing to its investigation of presentation order effects, required subjects to rate 90 concept pairs. One standard axiom of testing is that an increase in the number of items will increase the reliability of a test. In contrast, it may be that the nature of the similarity rating task is such that too many pairs will <u>decrease</u> task reliability. This seems to be very likely when one considers that judging interconcept relatedness on a 7-point numerical scale is not a particularly interesting activity. It is quite possible that subjects in the present study tired of their task before it was completed, and that their responses became more and more the products of boredom and fatigue. This would certainly add a larger component of

error to each subject's indices of structure than would have been the case for a shorter task.

A second factor differing between the present study and its models was the measure used to assess academic achievement. Both Bates and Konold and Bates developed measures specifically intended to tap skills at or above the application level of Bloom's (1956) cognitive taxonomy. Although a similar measure was designed for the present study, the unexpected time constraints placed on the postinstruction phase yielded test performance that was ill-suited for analysis. Instead, the sole measure of achievement was student performance on classroom examinations. The problem with using this indicator of achievement is that it probably incorporated a variety of factors somewhat unrelated to a conceptual understanding of statistics--e.g., individual study habits, algebraic and arithmetic skills, prior exposure to similar problems as homework assignments or classroom examples, and ability to perform well in the "high-pressure" environment of graded examinations. To the extent that these factors were not dependent on the appropriateness of students' semantic structures of statistics concepts, their presence could also have acted to decrease structure-achievement correlations. In fact, some support for this notion was obtained from introspective observations made by several of the undergraduate statistics students regarding their own evaluation

of their abilities. These students reported that they had a good understanding of statistics--far better than might be guessed based on their achievement scores--but that they simply could not perform well on graded tests. Interestingly, they were among the best performers on the rating task, relative to the criterion. It is, of course, sheer speculation, but one might conclude from this evidence that the rating task provided a better measure of conceptual understanding than did course achievement!

The nature of the concept pairs included in the rating task is another area of difference between these studies. Because of prior evidence that increasing understanding in a content area may be associated with increasing ability to discern concept dissimilarities (i.e., Traub & Hambleton, 1974), both Bates and Konold and Bates constructed a large proportion of concept pairs wherein the two concepts might seem, to inexperienced students, to be strongly related, but which content experts would tend to rate as very weakly related. In contrast, few such pairs were found to exist in the rating task used in the present study. Indeed, the statistics course instructor rated only <u>one</u> of the 90 pairs as demonstrating a negligible relationship. Earlier, it was suggested that the tendency of the statistics students to ascribe higher mean ratings to pairs after instruction than they did before instruction might argue against Traub and

Hambleton's (1974) notion of the importance of discerning concept dissimilarities. An alternative interpretation would be that the particular concepts used in the present study should have been judged to be strongly interrelated, and did not provide students with the opportunity to perceive weak relations. If other concepts with fewer obvious strong interrelations had been included in the rating task, then the ability to discern dissimilarities may have become more salient both to task performance and to structure-achievement relations.

These three factors--an overly long rating task, an inappropriate achievement measure, and the lack of dissimilar concepts--may have contributed to some of the amibiguity in the results of the present study, and certainly suggest areas of concern for future cognitive structure research. Despite these procedural problems, however, relatively unambiguous results were obtained in a tangential but educationally relevant investigation conducted in this study--that is, the indication that formal instruction has a meaningful effect on students' interconcept relatedness ratings. Students clearly tended to rate the strength of relation among the statistics concepts in a fashion considerably more like that of their instructor after instruction than they did before instruction. Also, this change in perceived relatedness was not restricted to concepts about which the

18. A test of significance is conducted where $H_o: \mu = 0$, and $H_a: \mu \neq 0$, with alpha set at .05. After observing the difference between $\bar{X}$ and μ_{hyp}, the researcher decides to retest H_o with $H_a: \mu > 0$ and alpha set at .10. What is the total proportion of area in the regions of rejection for this experiment?

 A. .075
 B. .100
*C. .125
 D. None of the above

19. Given: $P(\bar{A}) = .25$; $P(\bar{A}|B) = .30$; $P(\bar{A} \cap \bar{B}) = .04$; $P(\bar{B}) = .30$. What is $P(A \cap B)$?

*A. .490
 B. .525
 C. .960
 D. None of the above

20. A researcher conducts a study to determine whether rats will run a maze more quickly if they are punished for mistakes or if they are rewarded with food for correct responses. Twenty rats are each run through two mazes, once with reward and once with punishment. The researcher wants to minimize beta error as much as possible, but does not have access to more rats or mazes. Which other procedure to reduce beta would be most appropriate and effective in this study?

 A. Increase the value of alpha from .05 to .10.
*B. Treat the data as if they were from dependent samples.
 C. Run each rat through both mazes for more trials to get a more stable estimate of completion time.
 D. Select a directional rather than a nondirectional alternative hypothesis.

students had no knowledge prior to the instructional sequence. Even the structure of previously learned concepts was significantly altered to more closely conform to the criterion as a result of instruction.

Moreover, the observed pattern of rating task performance, whether based on Euclidean distances or on indices of concept clustering/differentiation, may conflict with Stewart's (1979) notion that such techniques to measure memory organization necessarily reflect only simple associative relations learned via concept contiguity within some arbitrary instructional sequence. If simple associative relations were all that were measured in a rating task, it might be expected that the undergraduate statistics students, who had had direct access to the <u>specific</u> relations indicated by the criterion structure, should have performed more like the criterion than did the graduate students who had received their training from <u>different</u> sources. In contrast to this expectation, the undergraduates, although they were more like the criterion in their ratings as a result of their instruction, still did not reach the level of performance demonstrated by the graduate students who had not been provided with that particular instructional sequence. This result seems to suggest that structure measurement techniques have the potential to tap systems of concept organization more complex than simple temporal/spatial relations--specifically, networks wherein

interconcept relations are determined by an understanding of concept meaning and function.

In summary, the second major goal of this dissertation--to establish the importance of several factors in the assessment of memory structure--was not realized. Issues left unresolved include the potential effects of concept presentation order on representation of structure, and the availa ability of means to control for types of perceived interconcept relations. Several other specific areas of future inquiry are suggested by the results of this investigation. For example, stronger evidence against Stewart's (1979) contention that structure measurement techniques tap only simple associative relations could be obtained than what was inferred from the instructional-effect data reported above. It could be argued that, although there may be slight differences in emphasis, the content of statistics is not really subject to arbitrary instructional sequences, that statistics is a tightly organized, even circumscribed discipline, little affected by the approaches of different instructors. In that case, the graduate students in this study might be viewed as having received the same conceptual framework, irrespective of its source, but with far more concept repetition than that experienced by the undergraduate students. It could then be concluded that simple associative relations <u>were</u> all that was being tapped in the present study,

because the graduate students outperformed the undergraduates owing to their more frequent exposure to the key concepts in contiguity. This issue might be resolved by contrasting the structure task performances of two groups similarly disparate in experience, but in a content area more subject to variability of concept presentation. One of the social sciences might suffice for this purpose, provided that there is some agreement among its content experts regarding which concepts are fundamental to its mastery. If results parallel to the differential group performance noted in the present study were obtained, then the argument that structure tasks <u>can</u> measure more complex systems of concept organization than associative contiguity would receive strong support.

Future structure research must also and especially be directed toward establishing reliable relations between academic achievement and structure task performance. This point is worthy of iteration because without such demonstrations there is little likelihood that other, more specialized applications of structure assessment techniques--such as pinpointing concept relations in need of remediation, or matching student structures to appropriate instructional sequences--will ever be fully realized. In order to obtain these relations, far more careful consideration of the characteristics of appropriate achievement measures will be required than has been typical of most prior structure research. Meaningful

achievement is not best exemplified by recognition or regurgitation of rote-memorized concept definitions, and if this caveat is obvious to structure researchers, it has not generally been obvious in their research. A related area of investigation should be the determination of which structure statistics (e.g., Euclidean distance from criterion, multidimensional scaling solutions, indices of concept clustering/differentiation, etc.) best provide information necessary and sufficient to predict meaningful achievement. Several of these measures were incorporated in the present study, and although all seemed to demonstrate signficant instructional effects, there were observed differences in how well these measures correlated with achievement. Admittedly, these differences could have been anomalous; nevertheless, the need to establish the limitations and advantages of the various indicators of structure appropriateness remains.

This dissertation has attempted to determine areas in which techniques to assess the structure of memory need to be made more consistent with their theoretical foundations, to suggest some potential methods for achieving this consistency, and to empirically test these methods to determine if they enhance the utility of those techniques. The fact that many of the obtained results of this investigation failed to resolve these issues may be interpreted as indicating that the measurement of the organization of memory is not

a straightforward venture. One reasonable conclusion might be that all the evidence necessary is not yet in, that future studies should be conducted to address the issues this investigation has raised, and that such efforts are surely called for before structure assessment techniques may be applied with confidence in practical educational settings.

REFERENCE NOTES

1. Konold, C. E., & Bates, J. A. Applying the episodic/semantic memory distinction to the study of instructional effects on cognitive structure. Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978.

2. Perkins, M. R. A look at three approaches to "cognitive structure": Bridging the defintion gap. Paper presented at the annual meeting of the American Educational Research Association, Toronto, 1978.

3. Bates, J. A. The relationships between achievement test item-types and measures of cognitive structure. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1979.

4. Rudnitsky, A. N., & Garlock, V. P. The differential structuring of content in memory. Paper presented at the annual meeting of the American Educational Research Association, New York, 1977.

5. Perkins, M. R. Measures of cognitive structure: Do they really assess learning at the level of comprehension? Unpublished doctoral dissertation, University of Massachusetts, 1978.

6. Eylon, B., & Reif, F. Effects of internal knowledge organization on task performance. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1979.

REFERENCES

Anderson, J. R. Language, memory, and thought. Hillsdale, N. J.: Erlbaum, 1976.

Anderson, J. R. Arguments concerning representation for mental imagery. Psychological Review, 1978, 85, 249-277.

Anderson, J. R., & Bower, G. H. Human associative memory. Washington, D. C.: Winston, 1973.

Anderson, R. C. The notion of schemata and the educational enterprise: General discussion of the conference. In R. Anderson, R. Spiro, & W. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N. J.: Erlbaum, 1977.

Anderson, R. C., & McGaw, B. On the representation of meanings of general terms. Journal of Experimental Psychology, 1973, 101, 301-306.

Anderson, R. C., Spiro, R. J., & Montague, W. E. (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N. J.: Erlbaum, 1977.

Bartlett, F. C. Remembering. London: Cambridge University Press, 1932.

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

Bobrow, D. G., & Norman, D. A. Some principles of memory schemata. In D. Bobrow & A. Collins (Eds.), Representation and understanding. New York: Academic Press, 1975.

Boring, E. G. Intelligence as the tests test it. New Republic, 1923, 35, 35-36.

Bower, G. H. Experiments on story understanding and recall. Quarterly Journal of Experimental Psychology, 1976, 28, 511-534.

Brandsford, J. D., & McCarrell, N. S. A sketch of the cognitive approach to comprehension: Some thoughts about understanding and what it means to comprehend. In W. Weimer & D. Palermo (Eds.), _Cognition and the symbolic processes_. Hillsdale, N. J.: Erlbaum, 1974.

Broudy, H. S. Types of knowledge and purposes of education. In R. Anderson, R. Spiro, & W. Montague (Eds.), _Schooling and the acquisition of knowledge_. Hillsdale, N. J.: Erlbaum, 1977.

Bruner, J. S. The act of discovery. _Harvard Educational Review_, 1961, _31_, 124-135.

Bruner, J. S. _Toward a theory of instruction_. New York: Norton, 1966.

Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. _Psychological Review_, 1975, _82_, 407-428.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. _Journal of Verbal Learning and Verbal Behavior_, 1969, _8_, 240-248.

Collins, A. M., & Quillian, M. R. How to make a language user. In E. Tulving & W. Donaldson (Eds.), _Organization of memory_. New York: Academic Press, 1972.

Conrad, C. Cognitive economy in semantic memory. _Journal of Experimental Psychology_, 1972, _92_, 149-154.

Deese, J. On the structure of associative meaning. _Psychological Review_, 1962, _69_, 161-175.

Gagne, E. D. Long-term retention of information following learning from prose. _Review of Educational Research_, 1978, _48_, 629-665.

Garskoff, B. E., & Houston, J. P. Measurement of verbal relatedness: An idiographic approach. _Psychological Review_, 1963, _70_, 277-288.

Geeslin, W. E., & Shavelson, R. J. An exploratory analysis of the representation of a mathematical structure in students' cognitive structures. _American Educational Research Journal_, 1975, _12_, 21-39.

Glaser, R. Trends and research questions in psychological research on learning and schooling. Eduational Researcher, 1979, 8, 6-13.

Glass, A. L., & Holyoak, K. J. Alternative conceptions of semantic memory. Cognition, 1975, 3, 313-339.

Glass, A. L., Holyoak, K. J., & O'Dell, C. Production frequency and the verification of quantified statements. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 237-254.

Greeno, J. G. The structure of memory and the process of solving problems. In R. Solso (Ed.), Contemporary issues in cognitive psychology: The Loyola symposium. Washington, D. C.: Winston, 1973.

Hambleton, R. K., & Sheehan, D. S. On the evaluation of higher-order science instructional objectives. Science Education, 1977, 61, 307-315.

Harary, G., Norman, R. Z., & Cartwright, D. Structural models: An introduction to the theory of directed graphs. New York: Wiley, 1965.

Hilgard, E. R., Atkinson, R. L., & Atkinson, R. C. Introduction to psychology (7th ed.). New York: Harcourt Brace Jovanovich, 1979.

Holyoak, K. J., & Glass, A. L. The role of contradictions and counterexamples in the rejection of false sentences. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 215-239.

Johnson, P. E. Associative meaning of concepts in physics. Journal of Educational Psychology, 1964, 55, 84-88.

Johnson, P. E. Word relatedness and problem solving in high-school physics. Journal of Educational Psychology, 1965, 56, 217-224.

Johnson, P. E. Some psychological aspects of subject-matter structure. Journal of Educational Psychology, 1967, 58, 75-83.

Johnson, P. E. On the communication of concepts in science. Journal of Educational Psychology, 1969, 60, 32-40.

Johnson, P. E., Curran, T. E., & Cox, D. L. A model for knowledge of concepts in science. Journal of Research in Science Teaching, 1971, 8, 91-95.

Johnson, S. C. Hierarchical clustering schemes. Psychometrika, 1967, 32, 373-380.

Kintsch, W. Notes on the structure of semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Kintsch, W. The representation of meaning in memory. Hillsdale, N. J.: Erlbaum, 1974.

Kintsch, W., & van Dijk, T. Toward a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

Kosslyn, S. M. Information representation in visual images. Cognitive Psychology, 1975, 7, 340-370.

Kosslyn, S. M., & Pomerantz, J. R. Imagery, propositions, and the form of internal representations. Cognitive Psychology, 1977, 9, 52-76.

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 1964, 29, 1-27.

Kuipers, B. J. A frame for frames: Representing knowledge for recognition. In D. Bobrow & A. Collins (Eds.), Representation and understanding. New York: Academic Press, 1975.

Landauer, T. K. Memory without organization: Properties of a model with random storage and undirected retrieval. Cognitive Psychology, 1975, 7, 495-531.

Leakey, R. E., & Lewin, R. Origins. New York: Dutton, 1977.

Loftus, E. F., & Loftus, G. R. Changes in memory structure and retrieval over the course of instruction. Journal of Educational Psychology, 1974, 66, 315-318.

Mayer, R. E., & Greeno, J. G. Structural differences between learning outcomes produced by different instructional methods. Journal of Educational Psychology, 1972, 63, 165-173.

McKoon, G. Organization of information in text memory. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 247-260.

Meyer, D. E. On the representation and retrieval of stored semantic information. Cognitive Psychology, 1970, 1, 242-299.

Meyer, D. E. Verifying affirmative and negative propositions: Effects of negation on memory retrieval. In S. Kornblum (Ed.), Attention and performance IV. New York: Adademic Press, 1973.

Meyer, D. E. Long-term memory retrieval during the comprehension of affirmative and negative sentences. In R. Kennedy & A. Wilkes (Eds.), Studies in long-term memory. London: Wiley, 1975.

Meyer, D. E., & Schvaneveldt, R. Meaning, memory structure, and mental processes. Science, 1976, 192, 27-33.

Minium, E. W. Statistical reasoning in psychology and education (2nd ed.). New York: Wiley, 1978.

Minsky, M. A framework for representing knowledge. In P. Winston (Ed.), The psychology of computer vision. New York: McGraw-Hill, 1975.

Norman, D. A. Memory, knowledge, and the answering of questions. In R. Solso (Ed.), Contemporary issues in cognitive psychology: The Loyola symposium. Washington, D. C.: Winston, 1973.

Postman, L. A pragmatic view of organization theory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Preece, P. F. W. Mapping cognitive structure: A comparison of methods. Journal of Educational Psychology, 1976, 68, 1-8.

Pressley, G. M. Mental imagery helps eight-year-olds remember what they read. Journal of Educational Psychology, 1976, 68, 355-359.

Pylyshyn, Z. P. What the mind's eye tells the mind's brain: A critique of mental imagery. Psychological Bulletin, 1973, 80, 1-24.

Quillian, M. R. A revised design for an understanding machine. Mechanical Translation, 1962, 7, 17-29.

Quiilian, M. R. Word concepts: A theory and simulation of some basic semantic capabilities. Behavioral Science, 1976, 12, 410-430.

Quillian, M. R. The Teachable Language Comprehender: A simulation program and theory of language. Communications of the ACM, 1969, 12, 459-476.

Rapaport, A. A comparison of two tree-construction methods for obtaining proximity measures among words. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 884-890.

Ratcliff, R., & McKoon, G. Priming in item recognition: Evidence for the propositional structure of sentences. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 403-417.

Rips, L. J., Shoben, E. J., & Smith, E. E. Semantic distance and the verification of semantic relations. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 1-20.

Rosch, E. Human categorization. In N. Warren (Ed.), Advances in cross-cultural psychology (Vol. 1). London: Academic Press, 1977.

Rothkopf, E. Z., & Thurner, R. D. Effects of written instructional material on the statistical structure of test essays. Journal of Educational Psychology, 1970, 61, 83-89.

Royer, J. M. Theories of the transfer of learning. Educational Psychologist, 1979, 14, 53-69.

Rumelhart, D. E. Notes on a schema for stories. In D. Bobrow & A. Collins (Eds.), Representation and understanding. New York: Academic Press, 1975.

Rumelhart, D. E., Lindsay, P., & Norman, D. A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Rumelhart, D. E., & Norman, D. A. The active structural network. In D. Norman & D. Rumelhart (Eds.), Explorations in cognition. San Francisco: Freedman, 1975.

Rumelhart, D. E., & Ortony, A. The representation of knowledge in memory. In R. Anderson, R. Spiro, & W. Montague (Eds.), _Schooling and the acquisition of knowledge_. Hillsdale, N. J.: Erlbaum, 1977.

Schaeffer, B., & Wallace, R. The comparison of word meanings. _Journal of Experimental Psychology_, 1970, _86_, 144-152.

Schank, R. C. The structure of episodes in memory. In D. Bobrow & A. Collins (Eds.), _Representation and understanding_. New York: Academic Press, 1975.

Schank, R. C., & Abelson, R. P. _Scripts, plans, goals, and understanding_. Hillsdale, N. J.: Erlbaum, 1977.

Shavelson, R. J. Some aspects of the correspondence between content structure and cognitive structure in physics instruction. _Journal of Educational Psychology_, 1972, _63_, 225-234.

Shavelson, R. J. Learning from physics instruction. _Journal of Research in Science Teaching_, 1973, _10_, 101-111.

Shavelson, R. J. Methods for examining representations of a subject-matter structure in a student's memory. _Journal of Research in Science Teaching_, 1974, _11_, 231-249. (a)

Shavelson, R. J. Some methods for examining content structure and cognitive structure in instruction. _Educational Psychologist_, 1974, _11_, 110-122. (b)

Shavelson, R. J., & Geeslin, W. E. A method for examining subject matter structure in instructional materials. _Journal of Structural Learning_, 1975, _4_, 199-218.

Shavelson, R. J., & Stanton, G. C. Construct validation: Methodology and application to three measures of cognitive structure. _Journal of Educational Measurement_, 1975, _12_, 67-85.

Smith, E. E. Theories of semantic memory. In W. Estes (Ed.), _Handbook of learning and cognitive processes_ (Vol. 5). Potomac, Md.: Erlbaum, 1976.

Smith, E. E., Shoben, E. J., & Rips, L. J. Comparison processes in semantic memory. _Psychological Review_, 1974, _81_, 214-241.

Stasz, C., Shavelson, R. C., Cox, D. L., & Moore, C. A. Field independence and the structuring of knowledge in social studies minicourses. Journal of Educational Psychology, 1976, 88, 550-558.

Stewart, J. Content and cognitive structure: Critique and assessment and representation techniques used by science education researchers. Science Education, 1979, 63, 395-405.

Thro, M. P. Relationships between associative and content structure of physics concepts. Journal of Educational Psychology, 1978, 70, 971-978.

Traub, R. E., & Hambleton, R. K. The effect of instruction on the cognitive structure of statistical and psychometric concepts. Canadian Journal of Behavioral Sciences, 1974, 6, 30-44.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Tversky, A. Features of similarity. Psychological Review, 1977, 84, 327-352.

Tversky, A., & Kahneman, D. Judgement under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.

Warriner, J. E., & Griffith, F. English grammar and composition: Complete course. New York: Harcourt, 1957.

Wexler, K. A review of John R. Anderson's Language, memory, and thought. Cognition, 1978, 6, 327-351.

Wiley, D. E. Latent partition analysis. Psychometrika, 1967, 32, 183-193.

Winograd, T. Frame representation and the declarative/procedural controversy. In D. Bobrow & A. Collins (Eds.), Representation and understanding. New York: Academic Press, 1975.

Witkin, H. A., Dyk, R. B., Paterson, H. F., Goodenough, D. R., & Karp, S. A. Psychological differentiation. New York: Wiley, 1962.

APPENDIX

Instructions for Nonpassage Subjects

On each page of the booklet with which you have been presented are five pairs of concepts that have been selected from the undergraduate psychology instructional sequence on inferential statistics. Your task will be to rate how closely you think the two concepts in each pair are related to each other. The purpose of this exercise is to provide information on how students organize and relate such statistical concepts. Your performance will not be used in determining your course grade. It should, however, indicate which concepts are least understood by students in general, and thus allow for better use of class time. Follow the steps below very carefully, and refer back to them whenever necessary. You will have as much time as you need to finish this exercise.

Step 1 - When you are told to begin, open the booklet to the first page and carefully read the two concepts in the first pair. If you are completely unfamiliar with one of the concepts, circle that concept.

Step 2 - Determine how closely the concepts in the pair are related to each other, based on the numerical scale provided on the page. Of course, all the concepts are related, in that they are important to an understanding of inferential statistics. If you feel that the pair of concepts are related only in this very general sense, circle the 1 (negligible relationship) in the row of numbers following the pair. If, however, you think the concepts have a more specific relationship, circle the number that best reflects that relationship. If you think they are moderately related, circle the 4; if the relationship is strong, circle the 7. Circle the 2 or the 3 if the relationship is more than negligible, but less than moderate. Circle the 5 or the 6 if the relationship is more than moderate, but less than strong. For example, if these concepts had been selected from descriptive statistics, the first pair might have been "Median--Standard Deviation". Both these terms refer to descriptors of distributions of numbers, but you would probably decide that they are not otherwise very closely related. Consequently, you might circle the 1 or the 2 after this pair. If the pair were "Median-Mean," you would probably decide that the concepts have a specific and rather strong relationship, and circle the 6 or the 7. Complete this step of rating the concept relationships even if you are unfamiliar with one or both of the concepts. If you do not know the meaning of a concept, rate the relationship according to your best guess.

Step 3 - Follow the first two steps for each of the remaining concept pairs on the page. Make certain that you DO NOT OMIT RATING ANY PAIR. Once you have completed rating all the pairs, turn to the next page and follow these same steps for that set of concepts. It is very important that your later ratings are not influenced by your earlier ratings. Therefore, DO NOT RETURN TO A COMPLETED PAGE AFTER YOU HAVE TURNED TO A FOLLOWING PAGE.

When you have finished rating the concepts on that page, go on to the next. Continue until you have rated all the concept pairs on every page. When you complete this exercise, return these instructions and the booklet to the instructor.

Instructions for Passage Subjects

On each page of the booklet with which you have been presented are five pairs of concepts that have been selected from the undergraduate psychology instructional sequence on inferential statistics. Your task will be to rate how closely you think the two concepts in each pair are related to each other. The purpose of this exercise is to provide information on how students organize and relate such statistical concepts. Your performance on this exercise will not be used in determining your course grade. It should, however, indicate which concepts are least understood by students in general. Follow the steps below very carefully, and refer back to them whenever necessary. You will have as much time as you need to finish this exercise.

Step 1 - When you are told to begin, read the paragraph on page 3 of these instructions. This paragraph describes an actual psychological experiment, including the experimental question, the method of conducting the study, and the conclusions of the experimenters. The purpose of the passage is to provide a context for your decisions about concept relatedness in the remaining portion of this activity. The actual content of the passage is IN NO WAY related to the goals of this exercise, and you should NOT be concerned with the topic therein discussed. You should ONLY pay particular attention to the ways in which certain statistical concepts are exemplified by the data and conclusions expressed in the passage. Once you have read the passage, go on to step 2.

Step 2 - Open the booklet to the first page and carefully read the two concepts in the first pair. If you are unfamiliar with one or both of the concepts, circle the unfamiliar concept(s).

Step 3 - Determine how closely the concepts in the pair are related to each other, based on the numerical scale provided on the page. Your rating for the pair should depend on the type of relationship, if any, that you think is illustrated by the experiment described in the passage on page 3--that is, the type of relationship that the concepts would have within the context of psychological research. Whether or not any particular concept is illustrated in the passage is not important. What is important is to determine the way in which the concepts would be related in the general context of an experiment, and then to rate the strength of this type

of relationship. Of course, all the concepts are related in that they are important to an understanding of inferential statistics. If you feel that the pair of concepts are related only in this very general sense, circle the 1 (negligible relationship) in the row of numbers following the pair. If, however, you feel the concepts have a more specific relationship, circle the number that best reflects that relationship. If you think they are moderately related, circle the 4; if the relationship is strong, circle the 7. Circle the 2 or the 3 if the relationship is more than negligible, but less than moderate. Circle the 5 or the 6 if the relationship is more than moderate, but less than strong. For example, if these concepts had been selected from descriptive statistics, the first pair might have been "Median--Standard Deviation." Both these terms refer to descriptors of distributions of numbers, but you would probably decide that they are not otherwise very closely related. Consequently, you might circle the 1 or the 2 after this pair. If the pair were "Median--Mean," you would probably decide that the concepts have a specific and rather strong relationship, and circle the 6 or the 7. In any case, be sure to remember the context within which the pair should be rated--psychological research. If necessary, refer back at any time to the passage on page 3 to help you keep this context in mind.

Complete this step of rating the concept relationships even if you are unfamiliar with one or both of the concepts. If you do not know the meaning of a concept, rate the relationship according to your best guess.

Step 4 - Follow steps 2 and 3 for each of the remaining pairs on the page. Make certain that you DO NOT OMIT RATING ANY PAIR. Once you have completed rating all the pairs, turn to the next page and follow these same steps for that set of concept pairs. It is very important that your later ratings are not influenced by your earlier ratings. Therefore, DO NOT RETURN TO A COMPLETED PAGE OF THE BOOKLET AFTER YOU HAVE TURNED TO A FOLLOWING PAGE.

When you have finished rating the concepts on that page, go on to the next. Continue until you have rated all the concept pairs on every page. When you complete this exercise, return these instructions and the booklet to the instructor.

Prose Passage

Several educational researchers have recently been investigating how the use of mental imagery can affect what people remember about things they read. One of these studies involved selecting 86 8-year-old children from a suburban school system, and randomly assigning them to one of two groups. The first group was trained to "Make pictures in their heads" that illustrated a series of descriptive sentences. These children were told that such "mental pictures" would help them to remember what they were reading. The other group of children were given the same training sentences, but were told only to do whatever they needed to do to help them remember. After the training session, both groups read the same short story, and were told to do whatever they had done in the training session to help them remember this new information. Then, both groups were asked the same set of 24 short-answer questions to determine what they remembered about the story. The researcher expected that, if mental imagery did not affect memory, then both groups would remember about the same amount of information. On the other hand, if imaging <u>did</u> have an effect, the researcher expected that the children trained to image would remember more about the story than would the children not trained to image. In fact, the group that imaged averaged about 18.5 of the questions answered correctly (standard deviation = 4.5), and the other group averaged about 16 correct (standard deviation = 5.7). The average amount of time spent reading the story by the image group was well within the range of the reading time for the non-image group, so the researcher concluded that this time would be about 11 minutes for 8-year-old children in general. The researchers also concluded that there was less than 1 chance in 20 that imaging had had no effect on how much the children had remembered about what they had read.

Knowledge-Level Postinstruction Achievement Test Items

1. The correct formula for calculating the probability of event "A" is

   A. N(A)/n
   *B. n(A)/N
   C. N(An-1)
   D. A(n-1)/N

2. Based on sample data, an experimenter establishes a 95% confidence interval. This indicates that there is a

   A. 95% level of certainty that the sample mean is different from the population mean.
   B. 95% level of certainty that the sample mean is the same as the population mean.
   *C. 5% level of certainty that the population mean falls outside the interval.
   D. 5% level of certainty that the population mean falls within the interval.

3. In inferential statistics, estimation refers to the general problem of

   A. determining whether differences exist between samples and populations.
   B. precisely stating the size of a difference between two means.
   *C. determining population characteristics from sample data.
   D. precisely stating appropriate values for alternative hypotheses.

4. Which term is an estimate of the standard error of a sampling distribution?

   A. S_x
   B. $\hat{\sigma}_x$
   C. $\sigma_{\overline{x}}$
   *D. $S_{\overline{x}}$

5. If we know only the size of alpha in a given experiment, then we also know the

A. nature of the null hypothesis.
B. nature of the alternative hypothesis.
C. probability of committing a Type I error.
*D. probability of committing a Type II error.

6. The decision of whether to conduct a one-tailed or a two-tailed hypothesis test in a given experiment should be based on the

A. nature of the null hypothesis.
B. pattern of scores obtained from the experiment.
C. magnitude of acceptable alpha error.
*D. logic and purpose for conducting the experiment.

7. The difference, in standard deviations, between a sample mean and the mean of its sampling distribution exceeds the value specified by the experimental decision criterion. Such a difference in means

*A. is statistically significant.
B. specifies a confidence interval.
C. is probably due to random sampling error.
D. supports the null hypothesis.

8. Which term refers to that portion of a sampling distribution of means into which a given sample mean must fall before one may consider H_o to be false?

*A. Region of rejection
B. Confidence interval
C. z_{crit}
D. z_{obt}

9. Which is the term that specifies the minimum distance, in standard deviations, between a particular sample mean and the mean of its sampling distribution that is necessary before one may reject H_o?

*A. z_{crit}
B. z_{obt}
C. Confidence interval
D. region of rejection

10. If a researcher commits a Type I error, this means that the researcher has

*A. rejected a true null hypothesis.
B. accepted a false null hypothesis.
C. used a biased estimate of a population parameter.
D. selected a biased (nonrepresentative) sample.

11. If a researcher says that the result of his experiment is statistically significant, the researchers means that the result

A. supports the null hypothesis.
*B. probably did not happen by mere chance.
C. has important practical value.
D. cannot be generalized to a population.

12. Which term refers to the proportion of all possible outcomes in which a particular outcome can occur?

*A. Probability
B. Confidence interval
C. H
D. $S^{o}_{\bar{x}}$

13. In order to determine the appropriate value for Z_{crit} in a given experiment, one must first determine

A. the null hypothesis.
*B. alpha.
C. the confidence interval.
D. $S_{\bar{X}}$.

14. Which is an unbiased estimate of a population parameter?

A. S_x
*B. $\bar{X}$
C. S_x^2
D. None of the above

15. Which is the term that refers to the limits within which a population mean is likely to exist, as determined from sample statistics?

A. Alpha region
B. Criterion limits
C. Region of rejection
*D. Confidence interval

16. The correct formula for calculating $S_{\overline{X}}$ is

A. $\sqrt{S_X/n-1}$
B. $\sqrt{S_X/n}$
*C. $S_X/\sqrt{n-1}$
D. $S_X/\sqrt{n}$

17. Which term refers to the level of risk one is willing to take that one may reject a true null hypothesis?

A. Region of rejection
B. Confidence interval
*C. Alpha
D. Z_{crit}

18. In inferential statistics, hypothesis testing refers to the general problem of

A. determining population characteristics from sample data.
B. precisely stating appropriate values for null hypotheses.
*C. determining whether differences exist between samples and populations.
D. setting precise limits for population parameters.

19. The region of rejection for a given experiment

A. always falls in both tails of the sampling distribution.
*B. may fall in one or both tails of the sampling distribution.
C. always falls in a symmetric central portion of the sampling distribution.
D. may fall in a central portion or either tail of the sampling distribution.

20. If a researcher commits a Type II error, this means that the researcher has

A. rejected a true null hypothesis.
*B. accepted a false null hypothesis.
C. used a biased estimate of a population parameter.
D. selected a biased (nonrepresentative) sample.

Application-Level Postinstruction Achievement Test Items

1. Three coins are tested to determine if they are fair. Coin #1 is tossed 10 times, and 6 tosses come up heads. Coin #2 is tossed 100 times, and 60 tosses come up heads. Coin #3 is tossed 1000 times, and 600 tosses come up heads. For which coin are the obtained results <u>most</u> likely to be statistically significant?

   A. Coin #1
   B. Coin #2
   *C. Coin #3
   D. All results are equally likely to be significant.

2. A researcher conducts 20 separate experiments. Given that the null hypothesis is true in 10 of them, that the observed level of power in the experiments is .70, and that the experimenter rejects the null in 8 of them, what is the experimenter's observed level of alpha?

   A. .30
   B. .20
   *C. .10
   D. None of the above

3. All other factors being held constant, the absolute magnitude of Z in a test of the difference between two dependent means will be <u>most</u> influenced by

   A. increasing the number of subjects by a factor of ten.
   *B. changing from a nondirectional to a directional hypothesis.
   C. using sigma instead of an estimate based on sample data.
   D. calculating $\bar{D}$ instead of $(\bar{x}-\bar{y})$.

4. Given: $n = 10$; $S_{x} = 12.30$; $C(91.96 \leq \mu \leq Q) = .95$. What is the value of Q (rounded to two decimal places)?

   A. 96.78
   B. 99.58
   C. 107.21
   *D. None of the above

5. For which problem would the use of interval estimation be the <u>least</u> appropriate?

   A. Evaluating the potential benefit, in terms of student achievement, of a new method of instruction.
   *B. Determining the interrelationships among several emotional illnesses and three alternative methods of therapy.
   C. Predicting the percentage of Democratic voters who will vote for the Republican Presidential candidate.
   D. Developing a working hypothesis for research on the characteristics of a newly discovered chemical element.

6. Two-thirds of all null hypotheses are true. A fair coin is tossed; if the coin comes up heads, the hypothesis under consideration is accepted; if the coin comes up tails, the hypothesis is rejected. What is the probability of committing a Type I error for a given null hypothesis?

   A. 1/3
   *B. 1/2
   C. 2/3
   D. None of the above

7. Under which condition(s), if any, would the standard error of the sampling distribution of means be <u>exactly</u> equal to the standard deviation of the parent population?

   *A. When $n = 1$
   B. When $\underline{n} = N$
   C. When $\bar{x} = \mu_x$ <u>and</u> $n = N$
   D. None of the above

8. For which problem would the use of hypothesis testing procedures probably be the <u>most</u> appropriate?

   *A. Evaluating the effects of several different systems of reward on 2nd- and 6th-grade children's motivation to do homework.
   B. Establishing the maximum reaction time necessary for radar operators to respond to an unidentified aircraft.
   C. Determining whether using a new machine tool will increase worker productivity enough to counterbalance its cost.
   D. Investigating whether a new analgesic compound reduces pain more quickly than does ordinary aspirin.

9. Ten playing cards are randomly dealt face-down from a well-shuffled deck. The first nine cards are then turned face-up; all are spades and none is a face card (i.e., Jack, Queen, King, or Ace). What is the probability that the tenth card is either a spade or a face card?

*A. .372
B. .385
C. .465
D. None of the above

10. Which manipulation would have the greatest effect on the total area within the regions of rejection for a significance test of the difference between two dependent means, given that all other factors are held constant?

*A. Reduce alpha by a factor of two.
B. State a directional rather than a nondirectional H_a.
C. Increase the number of subjects by a factor of ten.
D. Calculate $\bar{D}$ rather than $(\bar{x}-\bar{y})$.

11. Which factor(s), if doubled, would have the greatest effect on the width of a confidence interval for the difference between two independent means? (Given all other factors remain constant.)

A. $\bar{x}-\bar{y}$
B. both n_x and n_y
*C. both S_x and S_y
D. $\mu_x - \mu_y$

12. An evaluation is conducted to determine what factors in 50 urban children, aged 6-12, contribute to the children's level of academic achievement. Forty potential factors are investigated, and, of these, three are found to have significant effects ($\alpha = .05$): level of parents' education; presence of father in the home; and, time spent watching television. What conclusion, if any, is most appropriate regarding these results?

    A. The three factors found to contribute to children's achievement are probably positively correlated with each other.
    B. More-educated parents are more likely to stay married, and also to encourage learning activities for their children.
    C. No conclusion is possible because the reported size of the effect is probably not of practical importance.
    *D. No conclusion is possible because the reported results could have occurred just by chance.

13. If alpha is set at .05 and H_a is directional, what must the value of the ratio, z_{crit}/z_{obt}, be in order to reject H_o?

    A. ≥ 1.65 or ≤ -1.65
    B ≥ 2.33 or ≤ -2.33
    C. ≥ 1.00
    *D. ≤ 1.00 and > 0

14. A researcher collects a sample of reaction times to name colors, in order to determine the range within which the true population mean is likely to exist. This research question is

    A. not answerable with inferential statistics.
    B. stated as an alternative hypothesis.
    C. a problem of hypothesis testing.
    D. a problem of estimation.

15. What is the minimum amount of information that would be both necessary and sufficient to correctly conduct an hypothesis test of the difference between two dependent sample means?

A. $\bar{X}$, $\bar{Y}$, S_x, S_y, n_x, n_y, r_{xy}, H_o, H_a, Z_{crit}

*B. $\bar{D}$, S_x, S_y, n_x, r_{xy}, H_a, Z_{crit}

C. $\bar{X}$, $\bar{Y}$, $S_{\bar{x}}$, $S_{\bar{y}}$, n_x, H_a, alpha

D. $\bar{D}$, S_D, n, H_o, alpha

16. Given the following information from a test of statistical significance:

H_a: $\mu \neq 0$; n = 5; α = .001; Decision: Reject H_o.

Which is the most accurate statement regarding this information?

A. The sample size is too small to apply the principles of the Central Limits Theorem.
B. The sample size is too small to demonstrate any meaningful degree of practical significance.
C. The population mean in question is probably very close to zero.
*D. The population mean in question is highly unlikely to be zero.

17. Given: $C(21.24 \leq \mu \leq 33.00) = .95$. What is $S_{\bar{X}}$?

A. 2.28
B. 3.59
C. 5.88
*D. None of the above