National Library of Canada — Bibliothèque nationale du Canada

Canadian Theses Division — Division des thèses canadiennes

Ottawa, Canada
K1A 0N4

PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

- Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur

Richard Alan Yackulic

Date of Birth — Date de naissance

May 6, 1947

Country of Birth — Lieu de naissance

Canada

Permanent Address — Résidence fixe

327 Wathaman Place
Saskatoon, Sask.
S7K 4P7

Title of Thesis — Titre de la thèse

Multidimensional Scaling of Aptitude Treatment Interactions

University — Université

University of Alberta

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée

Doctor of Philosophy

Year this degree conferred — Année d'obtention de ce grade

1981

Name of Supervisor — Nom du directeur de thèse

T.O. Maguire

Date

April 27, 1981

Signature

[signature]

NOTICE

S

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

THE UNIVERSITY OF ALBERTA

MULTIDIMENSIONAL SCALING EVALUATION OF APTITUDE TREATMENT INTERACTIONS

by

© R. A. Yackulic

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

SPRING, 1981

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR R. A. Yackulic

TITLE OF THESIS MULTIDIMENSIONAL SCALING EVALUATION OF APTITUDE TREATMENT INTERACTIONS

DEGREE FOR WHICH THESIS WAS PRESENTED DOCTOR OF PHILOSOPHY

YEAR THIS DEGREE GRANTED SPRING, 1981

(SIGNED) ....

PERMANENT ADDRESS:
327 Wathaman Pl.
Saskatoon, Saskatchewan,
S7K 4P7

DATED December 16 1980

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled MULTIDIMENSIONAL SCALING EVALUATION OF APTITUDE TREATMENT INTERACTIONS submitted by R.A. Yackulic in partial fulfilment of the requirements for the degree of DOCTOR OF PHILOSOPHY.

Maguire ....

Supervisor

[illegible] ....

[illegible] ....

[illegible] ....

Dale Burnett ....

External Examiner

Date December 16, 1980 ....

ABSTRACT

The research reported evolved from the basic problem: can the acquisition of cognitive structure be measured with greater fidelity than currently possible using conventional achievement tests? The study focused on the cognitive structures acquired by a sample of university students studying a unit on descriptive statistics. A series of measures of cognitive structure (tests) were derived using multidimensional scaling techniques. The construct validity of the measures was then explored.

Two groups (and several subgroups) were used in the study. Since the tests were intended as measures of achievement (specifically of cognitive structure), the performance of a group of achievers (experts) was contrasted with the performance of a group who were relatively naive in the concept domain. Next, the tests' abilities to detect changes during learning were examined by testing the naive group after an instructional experience and comparing the pre, post and expert measures. The tests were also examined regarding their abilities to detect differences between groups of learners who received different instruction. The relationships between the derived achivement measures and a conventional classroom achievement test were also explored.

The construct validity of the derived measures was supported by each of the above investigations. Additional support was also provided by a brief convergent-discriminant validity analysis.

Previous research has established a link between cognitive style and achievement. Accordingly, the derived achievment measures were used as the dependent variable in an aptitude treatment interaction (ATI) study. Cognitive style (the aptitude) was assessed by the Group Embedded Figures Test (Witkin, Ottman, Raskin and Karp; 1971) while two levels of instruction (one emphased 'knowing how', the other emphasized 'knowing that') were employed as the treatment variable. Statistical analyses resulted in the rejection of the ATI hypothesis. However, instruction emphasizing 'knowing that' resulted in greater changes in learners' cognitive structures than did the 'knowing how' instruction.

Table of Contents

List of Tables

List of Figures

I. INTRODUCTION

A. Overview

Assessment of students' understandings of cognitive material has been a continuing problem for educators. By attending closely to instructional objectives of a behavioural focus, evaluators have been able to measure procedural knowledge (Ryle's (1949) 'knowing how') with success; however, valid Measurement of propositional knowledge (Ryle's 'knowing that') has proven to be far more elusive.

Yet as Mager (1973), Wight (1972), Bruner (1960), and others have suggested, the important intents of instruction are often propositional knowledges. Gagne and White (1978) have speculated that 'knowing thats' provide the interconnections among 'knowing hows'. They further hypothesized that these links provide the means for storing and accessing in memory intellectual skills.

Identification of propositional knowledge structures (also frequently referred to as cognitive structures (eg. Shavelson, 1972; Preece, 1976a)) has considerable pedagogical relevance. Essentially, a cognitive structure is the memorial storage of an individual's knowledges and intellectual skills. It includes not only stored concepts and their interrelationships, but also cognitive processes such as thinking, creating, and problem solving. Metaphorically, it might be thought of as a network of nodes and arrows where, in the simple case, concepts such as 'grass' and 'sidewalk' might be nodes connected by the propositional arrow-of-relation, 'beside'. Tolman (1932), Bruner, Goodnow, and Austin (1956), Gagne (1962,1965), Ausubel (1963), and Piaget (1970) are among the educators who have argued the importance of learning the propositional structure of a subject. Ausubel's (1960) advance organizer is anchored on the premise that cognitive structure entities can be identified and taught. Gagne's (1970) concepts of learning transfer and rule application are similarly founded in a cognitive structure model. Both lines of research operate from the rationale that knowledge of cognitive structures will facilitate instructional design.

Geeslin and Shavelson (1975) and Fenker (1975) are representative of researchers who have argued that student achievement might be assessed by comparing a learner's cognitive structure with a criterion structure derived from experts and/or instructional materials. Such a procedure may result in more valid measurement of 'knowing that' than is possible using other techniques.

Cognitive structure research also has potential in the realm of aptitude treatment interactions (ATI). Fenker (1975) and Nagy (1977) both speculated that estimates of cognitive structure might be appropriate dependent variables for ATI research. Successful inquiries in this vein may contribute to improvements in both instructional design and learning assessment.

During the past decade multidimensional scaling (MDS) has emerged as a viable method for estimating cognitive structures. Rappoport and Fillenbaum (1972), Shavelson and Stanton (1975), Fenker (1975), Preece (1975), Nagy (1977) and LaPorte and Voss (1979) are among researchers who have described cognitive structures using MDS. The resulting structures apparently have been useful for mapping content relationships during curriculum development. Nagy's attempt to explore aptitude treatment interactions was inconclusive, however. Research on the relationship between cognitive structures and achievement, similarly, has been less than fruitful (Fenker, 1975; Geeslin and Shavelson, 1975; Traub and Hambleton, 1974). The potential impact of successful research in these areas, however, is sufficient to warrant continued research.

B. Purpose of the Study

Succinctly, this study was concerned with the assessment of complex cognitive achievement. Currently popular procedures in which knowledge recall and problem solving are used as a basis for inferring the achievement of propositional relationships may introduce a validity gap which might be avoided by more direct measurement of cognitive structure. Johnson (1969) and Deese (1965) noted limitations of problem-solving tests to assess the acquisition of science concepts, while Preece speculated that

> Problem-solving tests are, perhaps, too drastic as tools for the initial exploration of cognitive structure, and a number of other techniques have been used which tamper less with what they are seeking to measure (1976, p. 1).

Further, estimates of learners' cognitive structures may prove to be more suitable dependent variables than conventional achievement test performance when investigating the effects of instructional variations: if the intent of instruction is the learner's acquisition of a cognitive structure, then that structure should be a principal concern. Aptitude treatment interaction studies involving cognitive style aptitudes might also benefit from such a dependent variable.

This study explored the above issues. Two measurement methods and several analytic procedures were used to estimate the cognitive maps attained by two samples of university students who were studying basic statistics. The impact of instructional variations (one instructional pattern was intended to favour 'knowing that' while the other instructional pattern emphasized 'knowing how') on the resulting cognitive maps was investigated in an ATI design employing field dependence as a cognitive style aptitude. A target structure was obtained from experts in the content area; this provided a criterion against which students' structures were compared. The congruence between students' structures and the criterion was related to student performances on a conventional achievement test. Finally, changes in students' cognitive maps during instruction were compared with achievement test performances.

The thrust of this thesis was to assess simultaneously the relationships among cognitive structures, instructional strategies, cognitive styles and achievement. Methods for estimating cognitive structures were also examined. Current research regarding these concepts and their interrelationships is reviewed in the next section. An empirical investigation of the nomothetic network is then reported.

II. REVIEW OF RELATED LITERATURE AND RATIONALE FOR THE STUDY

A. Introduction

In this chapter research pertaining to cognitive structures, cognitive styles, instructional strategies, and multidimensional scaling will be reviewed. Initially, cognitive structures, their role in concept attainment, and methods for estimating cognitive structures will be examined. The impact of cognitive styles on cognitive structures will then be explored. A brief consideration of an aptitude treatment interaction interpretation follows. Next, the relationship between cognitive structure and achievement will be considered. Finally, a nomothetic network amenable to empirical exploration will be presented.

B. The Nature of Cognitive Structure

Although cognitive structures as psychological constructs can be traced at least to Lewin's (1935) fields and Tolman's (1932) cognitive maps, the work of Bruner, Goodnow, and Austin (1956) is a milestone from an educational perspective. Unlike many of their connectionist contemporaries, Bruner et al were concerned with how man deals with a tremendously complex environment rather than with how man deals with a simplified stimulus-response laboratory situation. They observed, "were we to fully utilize our capacity for registering the differences in things and respond to each event encountered as unique, we would soon be overwhelmed by the complexity of our environment (Bruner et al, 1956; p1)." In order to successfully cope with this complexity, the human organism must attend to the similarities (rather than the differences) among events; humans must respond to events "...in terms of their class membership rather than their uniquenesses;" in essence, humans must categorize events.

In the same vein, Hanson (1958) stated "Knowledge of the world is not a montage of sticks, stones, color patches and noises, but a system of propositions (p.26)."

Pylyshyn (1973) defined cognitive structure in a manner which allowed the entity to be used as a primitive construct. A primitive construct is akin to MacCorquodale and Meehl's (1948) hypothetical construct. It is an abstract hypothetical entity possessing explanatory power and which does not require further reduction. The network composed would be

> best characterized as a descriptive symbol structure containing perceptual concepts and relations, but having the abstract quality of propositions rather than the particular quality of images (Pylyshyn, 1973; p. 7).

Concepts and relations in such a network need not correspond to words in an individual's vocabulary. Neither does the network require that propositions be expressable in language (natural or other). In fact, this very abstract representation of conceptual categories which Pylyshyn outlined would have the following implications:

> (a) that it does <u>not</u> correspond to a raw sensory pattern but rather is already highly abstracted and <u>interpreted</u>,
> (b) that it is not different in principle from the kind of knowledge asserted by a sentence, or potentially assertable by some sentence,
> (c) that it depends on the classification of sensory events into a finite set of concepts and relations, so that what we know about some event or object is formally equivalent to (i.e., can be reduced to) a finite (and, in fact, relatively small) logically independent descriptive propositions (p. 7).

Such a structure would be necessary for both 'knowing about' and 'knowing how'.

The highly abstract nature of Pylyshyn's cognitive structure is, at once, both powerful and hindering. Its being free of vocabulary and language requirements allows the model to explain the formation of concepts for which no verbal label exists; new concepts which trigger the creation of new variables are allowed; thoughts, emotions, perceptual events, and motor events are also within the model's explanatory realm.

Such a hypothetical entity taxes understanding. A concrete example certainly would be enlightening. Unfortunately the non-verbal non-visual nature of the structure precludes such concreteness. Indeed, many authors, seemingly, have been distracted from their theoretical investigations when they sought verbal or graphic metaphors. Gagne and White (197[illegible]) for example, in their efforts to anchor cognitive structures to observable events, ended up with a notion of propositional representation which was more akin to an intervening variable than to an hypothetical construct. In order to adequately 'explain' a range of behaviours, Gagne and White hypothesized four distinct types of memory representations. Descriptions of each type required both state and process variables. All four classes, however, can be explained more parsimoniously by Pylyshyn's paradigm.

At the risk of committing the same error, a reasonable metaphor for cognitive structure might be a nodes and arrows representation with features similar to those of models advanced by Kintsch (1972), Anderson and Bower (1973), Collins and Quillian (1972), Newell and Simon (1972) and others (some of these are discussed in greater detail below). Concepts are the nodes in the network. They are highly-interpreted abstractions of the environment. Among the entities they may represent are classes of things and events (e.g., cars, dogs), patterns of attributes (e.g., a strike in baseball), ideas (e.g., democracy), single events (e.g., D-Day), and single things (e.g., the earth). The nature of propositions which relate concepts may assume a variety of characteristics including logical or procedural associations, semantic or perceptual similarities, and episodic contiguities. Both concepts and propositions may have characteristics which are common among many people and other features which may be unique to a single individual.

This study employed Pylyshyn's notion of a cognitive structure. In addition to the characteristics noted above, cognitive structures are dynamic and evolving. New concepts and relationships may be added to and integrated with an existing structure through processes such as (but certainly not restricted to) Piaget's (Piaget and Inhelder, 1964) assimilation and accommodation, Ausubel's (1963) subsumption, Joyce and Weil's (1972) concept acquisition and concept attainment, Gagne's (1970) vertical and lateral transfer, and Wickelgren's (1977) chunking. Knowledge stating, thinking, and problem solving are among the processes of which the structure is capable.

Propositional relationships need not exist among all concepts in an individual's cognitive structure. Initially, a new concept might be related to only one other concept. As concept acquisition and attainment progress, propositional relationships between the new concept and previously existing concepts will be formed. According to this line of reasoning, an individual's cognitive structure might be organized into 'neighbourhoods' of interrelated concepts; neighbourhoods might also be interrelated. As a cognitive structure is increasingly elaborated, the number of relationships among neighbourhoods, presumably, will increase.[1]

----

[1] Interestingly, this paradigm of cognitive structure provides plausible explanations for two of psychologies perennial lemmas. Several investigators of multilingualism (e.g., Haugen, 1956) report among older people a tendency to

Although elusively abstract in nature, Pylyshyn's construct of cognitive structure has many of the characteristics of cognitive structures as intervening variables described by information processing researchers. Collins and Quillian (1972) presented a network model for long term-memory in which clusters of information are represented by nodes in an information network. The relationships between different information clusters are represented by arrows. In its simplest form, a node is a concept and the arrows indicate set relationships or attributes. Collins and Quillian and others have attempted to demonstrate their model in a variety of attribute and sentence learning studies, but have been met with only mixed success. They have also tested it in computer simulation with more favourable results.

Rummelhart, Lindsay, and Norman (1972), Newell and Simon (1972) and Frijda (1972) also have developed network models. In general, the models have been tested by learning recall studies. A conceptual structure for a set of learning

----

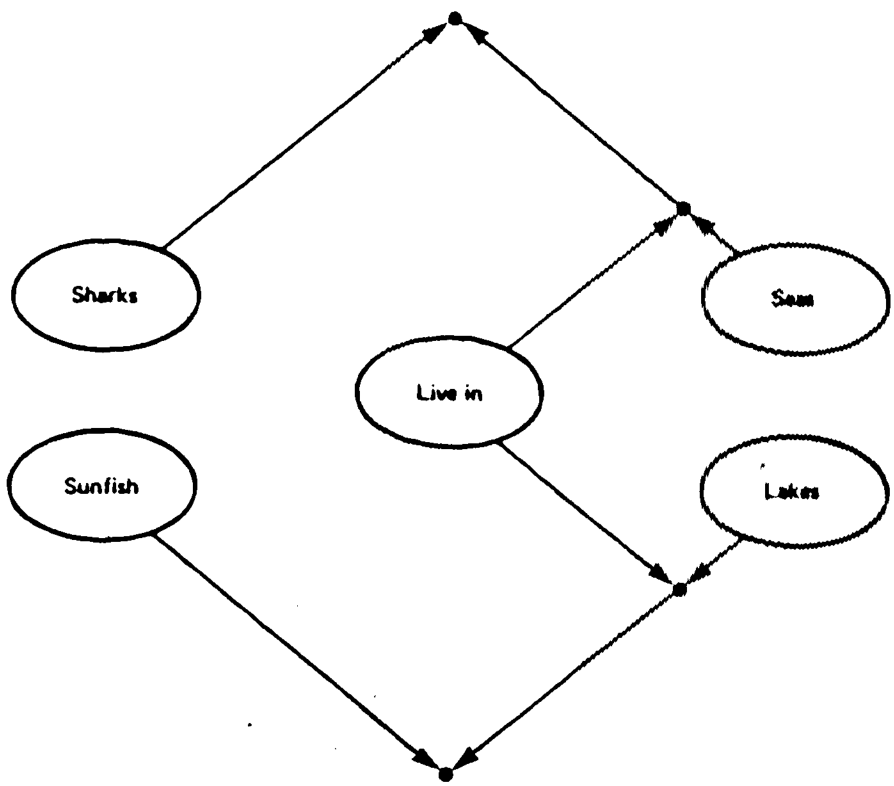
[1](cont'd)intermix vocabulary from one language to another. Perhaps as second (and more) languages are learned, concepts in the new language initially exist in a separate neighbourhood. As facility in the language develops, the conceptual neighbourhoods may become interrelated to the extent of overlap; a single concept might then have several labels which the individual employs interchangeably much as synonyms are used. Associative models of creativity (such as those spawned by Mednick, 1962) are also clarified by this view of cognitive structure. Mednick and his followers conceptualized creativity as the combining of disparate elements; establishment of propositional relations among distant neighbourhoods (or concepts) is, of course, the parallel cognitive structure process.

materials was hypothesized and the learner's recall of the material was studied in terms of the model. When subjects recall information that was unrelated to the information within the hypothesized structure, the model was revised.

Wickelgren's (1977) description of memory structures is also consistent with the cognitive structure model adopted here. He conceptualized associative memory as "records of events that are encoded and stored by networks of nodes (internal representatives of events) that are connected with each other by means of associations (p. 11)." Rather than a perceptual base, the nodes are characterized by conceptual and semantic qualities. New events may be incorporated into the structure through establishment of associations with existing elements. Processes such as chunking are suggested as ways of generating new principles, relationships and propositions.

Wickelgren regarded these abstractions (concepts principles, chunking, and propositions) as the important learning outcomes. Specific facts and examples functioned as the means for facilitating these ends. Performance was merely a fallible way of inferring learning. A diagram for depicting the memory for two propositions is shown in Figure II.1.

The next section of this study describes in greater detail the acquisition of cognitive structures. A more behavioural interpretation is also discussed.



A vertical associative memory for two propositions. Elementary concepts and higher-order concepts are both represented by specific elements. Vertical associations are used primarily to encode the components o· a higher-order concept rather than to encode temporal sequence.

(from Wickelgren, 1977)

Figure II.1 A Vertical Associative Memory for Two Propositions

C. Cognitive Structures and Learning

Bruner, Goodnow, and Austin (1956) presented a model for classifying events according to their features or attributes. The resultant categories were identified as concepts. The category (concept) could be defined either in terms of the attributes which defined category membership, or in terms of the set of elements which were members of the category. (Johnson (1972) has since labelled these definitions of a concept as the intensional meaning and the extensional meaning, respectively.) Three types of concepts were described:

1. conjunctive concept: the intensional meaning specifies the joint presence of several attributes;
2. disjunctive concept: the intensional meaning specifies either of two or more attributes;
3. relational concept: the intensional meaning specifies the relationship among attributes rather than the presence or absence of the attributes.

Within this model, categorizing was an essential element of concept attainment. The learner's task was to identify the attributes of the environment which were relevant for grouping events into categories; resulting organizations of exemplars according to their attributes were critical steps in concept attainment; the learner's goal in a concept attainment task was a structural representation of knowledge--a cognitive structure.

In subsequent works, Bruner (1960,1966) developed several principles for communicating a knowledge structure to a student. From the perspective of the present topic, the most important principles are those which deal with the structure and form of knowledge. The mode in which the knowledge is transmitted will greatly influence concept attainment, Bruner theorized. Citing developmental trends in children, he recommended presenting knowledge in a progression from concrete through graphic to abstract modes. The amount of information contained in the knowledge presentation will also influence concept acquisition: the more information an individual must deal with at one time, the more difficult will be the learning task. Bruner suggested that the knowledge structure should be presented in a form and sequence such that the number of processing steps required of the learner is minimized. This might be accomplished by employing a spiral approach--repeatedly exposing the learner to the same knowledge structure at increasing levels of complexity. Another approach might be to attend to the relationship among elements and to use a presentation mode which best depicts that relationship. The effective power of the knowledge structure was also important because it was this power which allowed the student to go beyond the facts to generate new propositions and solve problems. Indeed, Bruner suggested that the power in assisting the student to generate new propositions and to combine disparate elements was the most important feature of a knowledge structure.

Weil and Joyce (1978) developed a set of instructional procedures for concept attainment tasks. Their model suggests sequences and activities for facilitating the acquisition of concepts based on Bruner, Goodnow, and Austin's categories.

During the early 1960's, Ausubel and Fitzgerald (1961, 1962, 1963; Ausubel, 1960) performed a series of studies which focused on the influence of cognitive variables in learning and retention. They hypothesized that learning involves the positioning of new material within an existing cognitive framework and that this positioning is facilitated if the existing cognitive framework contains subsuming concepts which are relevant to the learning task. Employing brief introductory passages which focused the learner's attention on existing knowledge related to the to-be-learned material, Ausubel and Fitzgerald demonstrated the beneficial effects of organizing concepts. They concluded that the organizers acted as ideational anchors for the new material.

The effectiveness of advance organizers in facilitating learning is among the most powerful evidence for the existence of a cognitive structure.

Ausubel, of course, viewed learning through cognitivist spectacles. What-is-learned is an organized body of knowledge, a cognitive structure. Among the objectives of education is the learner's attainment of this structure and his use of the structure to generate propositions, to solve problems, and to enhance the learning of new information. "This knowledge, once acquired, is also in its own right the most significant independent variable influencing the learner's capacity for acquiring more knowledge in the same field (Ausubel, 1968; p.130)." With regard to the utility of a cognitive structure, at least, Ausubel and Bruner seem in close agreement.

Wickelgren (1977) cautioned against acquisition of huge quantities of knowledge and information. He argued not only that memory capacity was limited in theory, but that human beings might be operating close to quantitive learning and memory limits. Accordingly, he suggested emphasizing the acquisition of 'high-quality knowledge'.

> For example, it is of far less general value to know the superficial physical characteristics of various plants and animals than to know about nutrition, disease, first aid, the anatomy and physiology of the human body, and the general principles of living systems... The more a child in an intellectually rich environment learns about one thing, the less he or she learns about other things (p. 3).

Acquisition of general concepts and principles wh[illegible] provide insight into and understanding of facts and experiences is far more desirable. Chunking (the creation of a new concept to stand for a combination of previously defined concepts much like Bruner's going-beyond-the-information-given) is hypothesized as a powerful process for reducing large quantites of low-quality knowledge into smaller amounts of high-quality knowledge.

Although attesting the importance of goals such as reasoning, problem solving, creativity, and critical thinking, many educators and educational researchers have expressed little interest in efficient memory organization and processes, preferring instead the behaviourist model. Restricting learning outcomes to observable performances, they deny the acquisition of a cognitive structure as an end in the learning process. And, consistent with Dewey's (1916) caution that methods and ends are intricately related, focus on behavioural outcomes has spawned a complex of instructional methodologies which manifest behavioural principles. Weil and Joyce (1978) reviewed such strategies and provided guidelines for implementing them.

Although a critique of the behavioural model might, for a brief moment, be tantalizing, such a discussion would be tangential to the present problem. Of greater relevance is the effect on cognitive structure of instruction developed from a behavioural model. From a cognitive perspective, observable performances, clearly, are influenced directly by an individual's cognitive structure. Systematic performances will not be possible unless the propositions to which they are attributed have been adequately attained. Indeed, a person's performance is commonly the basis for inferences regarding concept attainment. Instruction which facilitates concept attainment ought, given appropriate motivation by an individual, result in improved performance. Further, if instruction has fostered the development of propositional relations, performance on transfer tasks (both vertical and lateral) should improve. Conversely, within the behavioural framework, performance is a function of the formal, temporal and reinforcement characteristics of the learning situation; concept attainment plays no role. Performance on transfer tasks will depend upon the similarities between the learning and test situations; transfer to unique situations will not be likely. In so far as lateral transfer involves considerable shift from the learning situation, prospects for this type of transfer are remote. The cognitive structures developed, similarly, will be influenced by the instructional strategy: direct attempts to elaborate propositional relationships in a structure might be expected to result in a richer structure than that resulting from behavioural-influenced instruction. Concepts presented contiguously might be expected to be related regardless of the instruction used; however, propositions among remote concepts (those from different neighbourhoods) ought be favoured by a cognitive-based approach.

Since Mager's (1962) influential book, Preparing Instructional Objectives, education has evidenced a decidedly behavioural focus. Learners' oft-noted failures to meet important educational goals might be interpreted within the cognition-behaviourism discussion. Champagne and Klopfer's (1980) observations regarding science education are pertinent.

> Science educators are nearly unanimous in professing the belief that problem solving and reflective thinking are important in children's learning of science in school. They advocate both the development of problem solving skills as an outcome of science instruction and the use of problem solving methods in instruction whenever appropriate.
> However, observations of science teaching in elementary and secondary school classrooms usually reveal that opportunities for students to engage in reflective thinking are all too rare (p. 4).

Unfortunately, although Champagne and Klopfer acknowledged the cognitive bases of problem solving, they proceded to analyze problem solving ability not as the mental states and processes which it no doubt is, but rather as behaviours, skills and observable competencies. They recommended that researchers focus on performance outcomes in problem solving situations. Such research would clarify the particular outcomes to be expected from learners.

Champagne and Klopfer seemed to have begged the issue. The learners' capabilities which they wish to address, quite simply, are mental states and processes and, as such, are not amenable to direct observation. Focusing on only observable outcomes will result, no doubt, in instructional strategies bereft of cognitive structure attention.[2] What they sought to avoid--"incongruities between what science educators say they want to teach, what they actually teach, and what they test for after instruction (Champagne & Klopfer, p. 8)"--seems a probable result.

----

[2] "Method means a way to a result, a means to an end, a path to a goal. Method therefore varies with the end to be reached. Without a clear notion of the end, we cannot proceed intelligently upon the journey toward it. (Dewey, 1916; p. 3)"

Gagne and White (1978), similarly, focused on performances as the outcomes of instruction. Although they argued from a three component model (instruction-memory structure-learning outcome), memory structures clearly constitute an intervening variable. Further, memory structures are not in themselves ends of the learning process, but are, rather, "merely antecedents that enable the human learner to display retention and transfer in terms of new performances (p. 187)." Having reviewed a variety of information processing models, they concluded that a single memory structure could not account for the diversity of learning outcomes. They identified four classes of models within which the domain of learning outcomes could be interpreted:

1. networks of propositions which facilitate knowledge stating.
2. intellectual skills which influence rule applications.
3. images which may be related to drawing
4. episodes which are presumed to be important for identifying action sequences.

The works of Rumelhart, Lindsay and Norman (1972), Anderson and Bower (1973), and Kintsch (1970) are cited as exemplars of the first model. From this perspective, a memory structure consists of nodes (concepts and information clusters) interconnected by propositional relationships. The structure facilitates knowledge recall as well as inferences regarding propositions which the learner has never directly received. Claiming that structures pertaining to intellectual skills have been less assiduously described, Gagne and White hypothesized a network model consisting of intellectual skills (rules) at the nodes and propositions relating skills as the links. In previous work Gagne (1968) has defined rules as statements of relationships among concepts. Unfortunately, rather than being content with conceptualizing intellectual skills, rules, and concepts as hypothetical constructs, Gagne and White defined these entities in terms of observable learner performances.

This, combined with a similar tendency to restrict learning outcomes to observable performances and a complete avoidance of literature which deals with memory structures as hypothetical constructs (e.g. Rappoport and Fillenbaum, 1972; Preece, 1976a), has resulted in Gagne and White overlooking the possibility that a richly-defined propositional network model is sufficient to account for intellectual skills. They seem to have fallen into a trap not unlike the learning-performance controversy of the 1950's: they appear to semantically confuse learning and performance. Mentalistic operations such as thinking, creating, relating, and inferring (which are possible within a propositional network model) may be adequate explanations for their intellectual skills. Apparently, the only distinction between the intellectual skills model and the network of propositions model is the emphasis on performance. But simply because observable performances (and, perhaps, even the cognitive operations which are their genesis) differ is not sufficient reason to postulate two or three or four parallel memory structures. Pylyshyn's (1972) arguments regarding parsimony would seem appropriate.

The tendency for instructional designers such as Gagne and White to be caught between cognitive and behavioural models complicates the development of instructional sequençes which manifest a single approach to learning. Unique among authors arguing the case for directly observable performances, Gagne (1976, 1977; Gagne and Briggs, 1974) has developed a detailed prescription for the teaching of complex (intellectual) skills. Unfortunately (at least from the current perspective), many of his guidelines seem to influence the acquisition of cognitive structures rather than directly influencing performances. The complications which result from conflating the two models will be more apparent in the discussion describing the development of instructional materials later in this document.

D. Assessing Cognitive Structures

As discussed previously, a cognitive structure is an hypothetical construct which refers to an individual's memorial representation of concepts and their interrelationships. Abstract entities invariably are difficult to estimate; their description customarily involves inferences from an empirical model. Such a strategy is employed when investigating cognitive structures.

Cliff and Young (1968) provide an example of how cognitive structure can be described empirically. They theorized that, for a particular stimulus set, an individual possesses a memorial configuration (structure). They suggested that the individual's responses to the stimuli could be described as mathematical operations on the configuration and that a non-metric scaling solution would be an approximation to the configuration. Their example demonstrated close agreement between the position of stimuli in the space determined by non-metric multidimensional scaling and the position of the stimuli resulting from independent, unidimensional judgements.

The adjectives were located on the circumference of a circle with the vertical axis apparently a favourability dimension; the nature of the horizontal axis is not so apparent. The vertical dispersion was similar to the scale positions of the adjectives derived from an independent favourability rating. Cliff and Young concluded that the MDS solution was a reasonable approximation to a memorial configuration.

Rappoport and Fillenbaum (1972) used subjects' judgements of proximity (similarity or dissimilarity in meaning) among terms to explore the semantic structures of two separate domains: colour names and the Have family of verbs. They "attempted to determine what sort of structure is required to accommodate adequately proximity judgements for each of two sets of related terms (p. 93)." Three methods of data collection were employed:

1. tree construction--subjects were presented a list of words and were required to construct trees in which the words were the nodes and branches connected interrelated nodes;
2. complete undirected (linear) graphs--subjects were presented a list of word pairs and were instructed to rank order the pairs in terms of the similarity of meaning between members of the pairs. Despite the seemingly high difficulty of ranking 105 pairs in this manner, test/retest reliabilities for individual subjects ranged from .432 to .901.
3. direct grouping or classification--subjects were presented a card deck with one word per card and were instructed to sort the words into piles on the basis of similarity in meaning.

Hierarchical cluster analysis (Johnson,1967) of the colour-names proximitry matrices detected differences between the judgements of male and female subjects. This was interpreted as demonstration of the model's ability to differentiate between populations. Young and Torgerson's (1967) MDS solution of the colour names matrices produced very orderly, two dimensional circular arrays. Rappopor t and Fillenbaum noted that the arrangement of the colours in the two-dimensional representation corresponds nearly perfectly with the arrangement of the colours on a hue circle.

Rappoport and Fillenbaum observed that the complete graph method was sensitive to the differences between groups detected by the tree construction task. However, although circular representations were possible using graphs, they were not possible using trees. Accordingly, the graph method is the preferred in instances when circularity is feasible.

Both tree construction and direct grouping yielded similar results for the Have verbs study, but the direct grouping task was much easier for the subjects and required less time. The representations were similar for both hierarchical clustering and MDS solutions.

Shavelson and Stanton (1975), in an attempt to ascertain the "correspondence between the subject matter structure and the representation of the subject matter in the cognitive structures of students (p.71)," estimated individual's cognitive structures for a mathematics domain using data collected from tree construction, word association, and sorting tasks. In the tree construction task, subjects were presented a list of concepts and were required to construct a graph connecting related concepts. The word association technique involved analyzing subjects' free associations to the concepts. Subjects sorted the concepts into piles on the basis of similarity in the sorting task. The data were analyzed according to an hierarchical clustering algorithm. Shavelson and Stanton concluded that a close correspondence existed between the structural representations based on data from experts and the structural representations of the concepts determined by a digraph analysis of instructional materials. They also found that representations from student data were similar across all three data collection methods, and concluded that "...the data support the interpretation that the methods are measuring a significant part of cognitive structure (p.80)."

Before progressing to further discussions of MDS-cognitive structure studies, consideration of the technique's construct validity is appropriate. Although an extensive review of MDS-cognitive structure literature might be both interesting and timely, such an endeavour is beyond the scope of this paper. A general overview is offered, with articles which are particularly pertinent to this study being reviewed subsequently.

The construct validity of MDS techniques to describe cognitive structures has been explored using four basic approaches:

1. varying methodology--a single group of subjects participates in a variety of data collection activities; various analytic techniques are used to provide structural representations; commoness in structures across data collection and analyses techniques provides validity support (Rappoport & Fillenbaum, 1972; Shavelson & Stanton, 1975; Fenker & Tees, 1976; Preece, 1976b; and Nagy, 1977).
2. varying subjects--groups of individuals differing in competencies in a subject area are identified;

variations in the cognitive structures derived provide validity evidence (Shavelson, 1973; Fenker, 1975; Preece, 1976a; and Nagy, 1977).

3. assessing intervention effects--structures for a group of subjects are assessed prior to and following an intervention intended to affect the structure; shifts between pre and post structures provide validity support (Shavelson, 1972; Traub & Hambleton, 1974; Fenker, 1975; LaPorte & Voss, 1979).

4. comparing with external criteria--cognitive structures derived from groups of individuals are compared with structures developed from content materials; similarity between structures supports validity (Geeslin & Shavelson, 1975; Shavelson, 1972).

Combined, the first three procedures provide convergent and discriminant validity evidence (Campbell & Fiske, 1959). The results have generally supported the construct validity of MDS techniques despite the following concerns:

1. several errors in analytic procedures--Preece (1976b), for example, criticized Shavelson's (1972, 1973, 1974) method for comparing matrices; Nagy (1977), in turn, accused Preece (1976a) of misinterpreting INDSCAL results.

2. the loss in precision resulting from analysis of group rather than individual data;

3. the crude means frequently employed to compare derived structures--of the studies cited, only Nagy (1977) used

a nonmetric goodness of fit measure. The remainder generally relied on visual interpretations.

4. uncertainty regarding the relative appropriateness of MDS versus hierarchical clustering analysis (Holman, 1972); and

5. controversy regarding techniques for analyzing individual differences data--Tzeng and Landis(1978) questioned the validity of Carrol and Chang's (1970) INDSCAL as a basis for inferring cognitive processes although they seemed less concerned about its use for the assessment of structures. They, as well as Rosler (1979) concluded that a recent modification of Tucker and Messick's (1963) Points-of-View is more suitable for describing interindividual judgement differences.

The fourth validation technique is related to criterion validity. Unfortunately, the criteria employed have tended to be of questionable relevance and validity; accordingly, firm conclusions regarding criterion validity are not possible.

Many of these studies investigated properties of cognitive structures relevant to the present study. Several of the more pertinent will be considered regarding the effects of instruction on cognitive structure, cognitive style and cognitive structure and the relationship between achievement and cognitive structure.

A series of studies reported by Shavelson (1972, 1973 1974; Geeslin & Shavelson, 1975) focused on instructional impact. In the 1972 research, twenty-eight high school students studied a physics text over a five day period and their daily word associations to fourteen physics terms were compared with similar responses collected from twelve control students studying a different content domain. He reported increases in relatedness indices during instruction for the experimental group and post-instruction differences between the two groups. Between group comparisons of derived group average distances between test sessions (the technique criticized by Preece (1976b)) also supported his conclusion that instruction affected structure. Unfortunately, Kruskal (1964) scaling of the relatedness matrices did not support this position. Shavelson speculated that the control group's existing concept meanings corresponded too closely to the concept meanings being taught.

Shavelson and Geeslin (1975) tried to eliminate the confounding effect of prior knowledge in an instructional intervention study involving grade eight students An experimental group (N=43) studied a programmed instruction text dealing with probability while a comparison group (N=44) studied a programmed instruction text on factors and prime numbers. Examination of the pre-test word association data indicated that both groups of students were unfamiliar with the ten probability concepts employed. Kruskal (1964b) scaling of post test data revealed that representations from the probability-instructed group agreed more closely with representation constructed by digraph analysis of the content domain than did representations derived from the control group. They concluded that this supported the position that cognitive structures are a function of instruction.

Geeslin and Shavelson also attempted to establish a relationship between achievement and cognitive structure. They used the digraph-derived content structure as a criterion for the to-be-learned structure. The cognitive structure data which they used to investigate the relationship were Euclidian distances based on the differences between an individual's cognitive structure and the criterion structure. They did not find as strong a relationship as they had expected. Geeslin and Shavelson's suggested explanation was that two different types of learning may have ocurred: the learning of a cognitive structure and the learning of problem solving behaviours. The achievement test, of course, measured the latter behaviour. From the perspective of the arguments presented above, such an occurrence is unlikely. If problem solving behaviour is not directly dependent upon a cognitive structure, at the least a cognitive structure will be developed in conjunction with problem solving behaviour.

Another plausible explanation might be that their criterion for a (learned) cognitive structure was inappropriate. The digraph analysis which they used to construct the criterion structure consisted of analyzing the instructional materials with reference to a number of critical concepts. They inferred the relationships among the concepts from the syntactic relationships in the written materials. The resultant content (subject matter) structure, then, was a function of the specific instructional curriculum and need not have been a good, let alone best, estimate of the conceptual network involved.

A better tack, especially when measuring tools tend to be crude and the inferential leaps large, might have been to follow the strategy one uses when determining the accuracy of a replica: measure the replica and the original with identically calibrated instruments. In this case, a better criterion might have been a representation derived from individuals who, according to some contemporary standard, have already acquired the cognitive structure being investigated. That is, the criterion (the original) would be a representation derived from experts in the subject matter following the same data collection procedures as used to obtain the learners' data. By measuring both the original and the replica with the same instruments, some calibration problems might be avoided.

Indeed, this rationale is implicit in Shavelson and Stanton (1975). Although chiefly concerned with validating various data collection techniques, they also attempted to establish the relationship between a structure derived from experts' data and a digraph structure of content materials. The minimal sample size (N=2) and use of group-average analytic procedures limited the value of the experts' structure as a target for learners. Fenker (1975), in a considerably more sophisticated study, tested a structure obtained from experts as the formal (target) structure for instruction. Eight faculty and graduate students in mathematical psychology judged the similarity among all possible pairs of twelve experimental design concepts (66 pairs) and nine measurement scale concepts (36 pairs). INDSCAL scaling (Caroll & Chang, 1970) yielded satisfactory solutions in two dimensions for both concept sets. Inter-rater agreement (assessed by correlating each experts similarity judgements with the interpoint distances in the group space) was judged to be high. The solutions were used as a formal structure in two experiments involving learner data.

In the first experiment, twenty students studying the concepts as part of an undergraduate statistics class completed the paired comparisons task before and after classroom instruction on course units containing the two sets of concepts. A combined analysis of expert and post-instruction data resulted in a poorer version of the formal structure. Fenker observed, "The addition of the students simply added considerable noise to the formal structure." Correlations between student judgements and the concept distances on the dimensions of the (expert only) formal structure were low. This, combined with low judgement reliabilities by the students, prompted Fenker to conclude that the degree to which the students understood the concepts was inadequate for meaningful assessment of their cognitive structures.

In the second experiment involving twenty-seven students enrolled in the same course, Fenker attempted to remedy the problem by encouraging students not only to learn the concepts, but also to consider how the concepts were related to each other. The procedure was similar to the first experiment. Students' judgements, in this instance, were considerably more reliable. INDSCAL analysis of measurement concept judgements (from seventeen students deemed sufficiently reliable) resulted in a three dimensional solution. Similar analyses of fourteen students' experimental design judgements provided a satisfactory two dimensional representation. Correlations between dimension weights of the formal structure and dimension weights of the student structure were high, prompting Fenker to conclude that the students utilized cognitive structures similar to the formal structure.

The relationships which Fenker reports between student structures and formal structures are remarkable. Implicit in the comparisons employed is the premise that the dimensions manifest in the student and formal structures are analagous. Since the group-space dimensions recovered by INDSCAL represent the concept properties and attributes attended to by the group when judging proximities, such a premise does not seem warranted. The student and expert spaces for the measurement concepts were not even of the same dimensionality. Further, in light of the failure to establish a relationship between learners' and experts' structures in the first experiment which was methodologically identical, the argument that the relatively weak manipulation--merely asking students to attend to the relationship among concepts--would result in such improved congruence between learner and expert structures is incredible.

Traub and Hambleton (1974) employed a different strategy in a study to determine the effects of instruction on cognitive structure. They reasoned that the MDS representation of students' cognitive structures should qualitively change during instruction. Avoiding the notion of a target structure, they examined changes in points of view (Tucker & Messick, 1963) in a pretest posttest study of fifty-three graduate students enrolled in an introductory tests and measurements class. They reasoned that the quality of students' cognitive structures should change during instruction. Students provided similarity judgements for all possible pairs of thirteen statistical concepts. Only one point of view was manifest in both the pre and post instruction data. The number of dimensions required to portray the structures changed, however Prior to instruction, four dimensions were necessary; after instruction, three sufficed. Traub and Hambleton argued that significant pre-post instruction changes in the judged similarity of some pairs resulted in the reduced number of dimensions required. Interpretation of the pre and post structures led them to conclude that although general similarity existed, the post test structures provided a sharper distinction between the two general classes of concepts being studied. Acknowledging the weakness of their one group design, they concluded that the differences between structures were suggestive of instructional effects.

The literature cited is less than overwhelming in its support of MDS as a technique measuring instructional impact on cognitive structures. Research implementing a cross-sectional strategy has been only slightly more positive in supporting the hypothesis that cognitive structures change and that MDS is sensitive in detecting that change. Preece (1976a) and Nagy (1977) both used quasi-experimental cross-sectional designs to investigate the cognitive structures of individuals having varying levels of education in the content/concept domain. Preece, studying fifteen mechanics concepts, collected word association data from five groups of subjects ranging from first-form students (mean age 12 yrs. 3 mo.) to university graduates studying to become physics teachers. Preece's inspection of the INDSCAL solutions revealed differences among the groups:

> There was clear evidence of semantic development going on from the least to the most knowledgeable groups. For the least knowledgeable groups (A and E), the clusters were less tightly organized...(p. 287).

Location of some critical concepts also shifted according to the knowledge level of the group.

Nagy's (1977) exploration of grade nine and grade twelve students' cognitive structures for scientific method concepts similarly resulted in representations which, on inspection, differed by grade. Goodness-of-fit measures (Lingoes & Schonemann, 1974), however, revealed too much similarity among the grades for Nagy to conclude grade differences.

Were it not for LaPorte and Voss (1979), the cited string of studies might militate against further research into changes in cognitive structure. They studied the effect reading a passage would have on undergraduate students' cognitive structures for twenty terms. Subjects were randomly assigned to three groups--two which read passages and a control group which did not--and similarity ratings for word pairs were collected before and after the passages were read. Half the subjects in each reading group received a narrative passage, the other half received a descriptive passage. The descriptive passage was expected to have a greater effect on subjects' cognitive structures than the narrative. Only results of the descriptive passage are discussed here. The control group performed the rating task twice, the two instances interrupted by a math task of duration similar to that required to read the passages.

As in many of the previously cited studies, inspection of the INDSCAL concept structures revealed group and time differences. Rather than analyzing the concept spaces for group differences, however, LaPorte and Voss examined the subject spaces. They performed separate discriminant analyses for each group space containing pre and post 'subjects'. Significant differences were found between pre and post subjects for both treatment conditions. There was no significant difference for control group data. Considering the relatively weak intervention (reading of a short passage) this finding is impressive. This approach to examining instructional effects warrants further examination.

This section has briefly analyzed several issu pertinent to the measurement of cognitive structure Although MDS analysis of word association, similarity ratings, tree graphing, and card sorting result in acceptable representations of cognitive structures, problems remain regarding the methodology's sensitivity to (perhaps minor) changes in structure. Use of goodness-of-fit measures and increased attention to INDSCAL subject spaces may prove fruitful.

E. Cognitive Styles and Cognitive Structures

For several decades educational researchers have been intrigued by the notion that cognitive style might be an important variable in the examination of student learning. Many studies (see Witkin, Moore, Goodenough & Cox, 1977, for a review) have investigated the relationship between cognitive style and performance measures of achievement; of present interest, however, is the relationship between cognitive style and cognitive structure acquisition.

An individual's cognitive style, although referred to by a plethora of labels by as many researchers, is the characteristic strategy which the person uses when perceiving, interpreting, and storing (in memory) environmental events. It is part of the dynamic aspect of cognitive structure and influences the way a person organizes and re-organizes his relational network. Although research in the area proceeds under a variety of labels and definitions (including: wholist/sequencer (MacDonald-Ross, 1972), conceptual level (Hunt, 1970), cognitive control (Gardner, 197[illegible]), [illegible]xibility of closure (Field & Cropley, 1969), scanning ([illegible]lzmann, 1966), levelling/sharpening (Lohrenz & Gardner, 1973), field dependence/independence (Witkin, 1950), and global/articulated (Witkin, Moore, Goodenough, & Cox; 1977)) there is sufficient conceptual and empirical overlap among many of the entitites to treat them as intervening variables manifesting the same construct. Hammond (1976) and Witkin et al have provided substantial reviews of the area.

Because of the active role cognitive style theoretically plays in an individual's cognitive structure, it has long and often been speculated that style is a determinant of the structure. This premise is manifest in the present study. Witkin's field dependent-independent concept is at once the oldest and most thoroughly researched--particularly with regard to education--of the cognitive style variables. It has been frequently related to observable learning outcomes as well as teacher styles. Accordingly, it is the variable of interest in this study.

According to Witkin et al:

> ...persons with an articulated cognitive style are likely to analyze a field when the field is organized, and to impose structure on a field when the field lacks organization of its own. Persons with a global style are more likely to go along with the field "as is," without using such mediational processes as analyzing and structuring (p. 21).[3]

This facility in the use of mediators should enable field independent people to organize learning materials which lack clear inherent structure, while field dependent people will experience considerable difficulty organizing such situations. Since cognitive structures are the resulting organization of learned material, this line of reasoning suggests that there will be differences in the structures acquired by individuals varying in cognitive style. Further, efforts by the instructional designer to organize the learning materials ought to facilitate cognitive structure acquisition for field dependent individuals but have little impact for field independent people.

With regard to the first premise, Witkin et al refer to a study by Stasz (1974) where learners' structures (presumably derived using MDS) for ten anthropological

----

[3] Witkin frequently uses the terms global and articulated as synonyms for field dependent and field independent, respectively.

concepts were examined.

> For field-dependent teachers and students, concepts clustered into a large, loosely organized group which included most of the concepts. For field-independent teachers and students, concepts clustered into small, tight groups with less overlap across groups (Witkin, et al, 1977, p. 9).

Support for the second premise was provided by Fleming, Knowlton, Blain, Levie, and Elerian (1968) in a study of list learning. Two lists of words were employed: one contained an advance organizer, the other did not. Field dependent subjects recalled significantly more words in the advance organizer condition. There was no significant difference for field independent subjects. Research by Schwen (1970) and Koran, Snow, and McDonald (1971) also supported the premise.

Cognitive style has been demonstrated to be influential in concept attainment. Nebelkopf and Dreyer's (1973) study of learning curves for a concept attainment task indicated different acquisition patterns for field dependent and field independent people. Learning curves for field independent people manifest discontinuous acquisition (no change in success rate for initial trials followed by a sudden improvement in performance) as predicted on the basis of the Bruner, Goodnow, and Austin (1956) hypothesis testing model. Field dependent subjects, on the other hand, seem to learn in a continuous fashion, gradually improving from trial to trial consistent with Woodworth's (1938) spectator approach. Style of learning did not influence ultimate performance, however: there was no significant difference between the groups on the number of trials-to-criterion.

MacDonald-Ross (1972) noted a similar variation in learner strategies although he did not relate it to field dependence-independence. Subjects were presented with a graphic representation of a knowledge structure which they were required to learn. The nodes in the graph depicted concepts and interrelationships were shown by connecting lines. During instruction, subjects were allowed to choose their own sequence for learning the concepts but they were required to indicate which concept in the network they were working towards. MacDonald-Ross identified two types of learners on the basis of their strategies: 1. those learners whose goal was the next concept in the network (sequencers); 2. those learners whose goal was several concepts away from their current position in the network (wholists). The sequencers progressed systematically from one concept to the next while the wholists seemed to approach several concepts simultaneously. Although the strategies of the learners differed, a post-test of achievement revealed no differences in level of achievement between the two strategies.

F. Aptitude Treatment Interaction

In light of positions established previously in this paper, the field dependence-independence findings are extremely pertinent. It will be recalled that direct attempts to interrelate concepts during the learning process (e.g., by employing advance organizers, spi ing, emphasizing propositional relationships) were expected to facilitate formation of an enriched structure. Instruction manifesting behavioural approaches was expected to result in an impoverished structure. The cognitive styles literature reviewed, however, suggests that the anticipated cognitive structure differences will occur only for field dependent people. Field independent people can be expected to organize the learning materials on their own and to acquire enriched structures regardless of the instructional approach. This, of course, is the essence of an aptitude treatment interaction (ATI) (Cronbach, 1957, 1975; Cronbach and Snow, 1975*). Generally, the dependent variable in an ATI study is a performance measure of achievement. Use of a cognitive structure as the dependent variable , seemingly, would provide a purer measure of learning. Accordingly, the study reported below used field dependence-independence as the aptitude in an ATI design. Two instructional strategies constituted the treatment variation.

G. Cognitive Structure and Achievement

To this point, acquisition of a cognitive structure has been presented as the outcome of learning. Although such may constitute the goals of the present educational system, it is clearly not the basis for evaluating students. Learning, of course, is commonly inferred from a learner's performance on structured tasks. If performances are determined by cognitive structures--as was argued above--then it would be reasonable to expect a relationship between measures of cognitive structure and conventional performance measures of achievement.

Research in this area is sparse. As an aside to his study of measurement and experimental design concepts, Fenker (1975) determined the relationship between learned structure and achievement. As the measure of structure-appropriateness, he calculated the proportion of variance in a student's structure which was common to the experts' (formal) structure. He correlated this with unit grades measured on a five-point scale (measurement concepts, r=.61; design concepts, r=.54). Unfortunately, his failure to describe the basis for the unit grades precludes drawing of even tentative conclusions.

The Geeslin and Shavelson (1975) study which investigated changes in students' cognitive structures arising from a study of probability, also examined the achievement-cognitive structure relationship. Although they demonstrated only a weak relationship between instruction and cognitive structure changes, a groups-by-time analysis of variance on achievement test scores revealed significant improvement in achievement for the group receiving test-related instruction. That performance improved but cognitive structures did not change, by itself, would be troublesome. Their technique for measuring structural change was rather crude, however, and may not have detected structural shifts that did occur. Further, little is reported about the achievement test other than that it consisted of 28 short answers and seven multiple choice questions and had an alpha coefficient of .83. There is little indication of the types of questions (knowledge, problem solving, etc.), the level of difficulty , or the relationship between the items and the concepts responded to during the MDS portion of the study. Review of the reported means and standard deviations suggests the test performance was somewhat lower and more diverse than might be expected on a unit test of achievement on a programmed instruction sequence which incorporated formative evaluation. Additionally, the data clearly lacked homogeneity of variance. Lacking further clarification of the test several alternative explanations for the results remain viable. Again, conclusions regarding the relationship between conventional measures of achievement and cognitive structure seem ill advised.

This remains an important issue. Performance on conventional achievement tests (whatever they are) ·s considerable credibility as the criterion for asse..ing learning. This alone would be sufficient reason to ·vestigate further the relationship. Even more important, ·owever, is the critical role that this relationship assumes in cognitive theory. Failure to establish the performance-cognitive structure link would raise serious concerns regarding:

1. the construct validity of measures used to assess cognitive structures;
2. the validity of performance tests used to assess achievement;
3. the validity of cognitive structure as a primitive construct.

Problems in assessing the relationship may be expected from at least two areas:

1. identification and crisp measurement of the appropriate cognitive structure attributes to use as indicators of learning (e.g., average distance shift occurring during learning, distance measures of congruence with experts' or content structures, goodness of fit measures between learners' and target structures, movement in a subject space, etc.);
2. construction of an adequate performance achievement test.

Both these issues are addressed below.

H. The Theoretical Network

The following theoretical network summarizes the preceding discussion. It is, of course,only a small subset of a much larger network. Both cognitive structures and cognitive styles are well established theoretical entities. Their roles have been well-elaborated and considerable empirical evidence verifies their existence. Further, convergence among several data collection techniques and a variety of MDS procedures indicate that at least crude assessments of cognitive structure are possible. The confusion of labels and theoretical speculations which abounds in the cognitive style literature suggests that consideration of field dependence-independence (of which there are several measures) as an intervening variable manifesting some aspects of cognitive style is a reasonable approach.

The relationship between cognitive structure and cognitive style is relatively unexplored. Some theoretical positions seem to suggest that cognitive structure is a superordinate construct consisting of two subcomponents: the stored network and process strategies. Cognitive style is consistent with the latter entity. Accordingly, cognitive style should interact with environmental events in the acquisition of the stored network. This relationship has not been verified. Speculation also exists that cognitive style can be influenced by environmental events. This relationship, similarly, has not been verified.

Two links in the network, between instruction and cognitive structure and between cognitive structure and achievement, are essential if cognitive structure is to be considered a primitive construct. Instruction ought to affect achievement by influencing cognitive structure (i.e., a-b-c). Accordingly, changes in cognitive structure ought to accompany changes in performance. Specifically, if what is to be learned is a cognitive structure, then an individual's cognitive structure for a particular subject domain will become more ordered during learning; further, if what is to be learned is a cognitive structure then individuals who have learned the same subject matter will share similar cognitive structures which vary according to individual differences such as cognitive style; and, finally, if what-is-to-be-learned is a cognitive structure, then a strong relationship should exist between present achievement measures and estimates of cognitive structure.

As might be expected, the observable portion of the model has been far better explored than has the hypothetical portion. Empirically verified relationships have been established between instruction and achievement, although the effect of different strategies of instruction is unclear The validity of measures of achievement more commonly is assumed than assessed; the measures are respected if for no reason other than their popularity and longevity. Overlap in the predictions of different instructional models makes it difficult to distinguish among the instructional strategies generated.

Several instruments have been demonstrated to measure field dependence-independence. This variable has been shown to interact with instruction to affect learning. Although degree of organization in the instructional situation apparently does not influence the achievement of field independent people, the achievement of field dependent people improves with increases in instructional organization.

The network is depicted in Figure II.2. A study exploring this network is reported below. In particular, empirical verification of the following hypotheses was attempted:

1. An individual's cognitive structure for a particular content domain will change during learning.
2. Field dependence-independence will interact with instruction during the acquisition of cognitive structures. Field dependent people will acquire more orderly structures when taught by a strategy organized according to propositional relations than when taught by a strategy which lacks such organization. The cognitive structures of field independent people will not be affected by these instructional variations.
3. For a particular domain, the cognitive structure of an individual who has mastered an instructional sequence will be similar to the cognitive structure of an individual who is an expert in the domain.
4. There will be a strong positive relationship between achievement and cognitive structure. Specifically, the more closely the cognitive structure of a student corresponds to the cognitive structure of an expert, the higher will be the student's level of achievement.

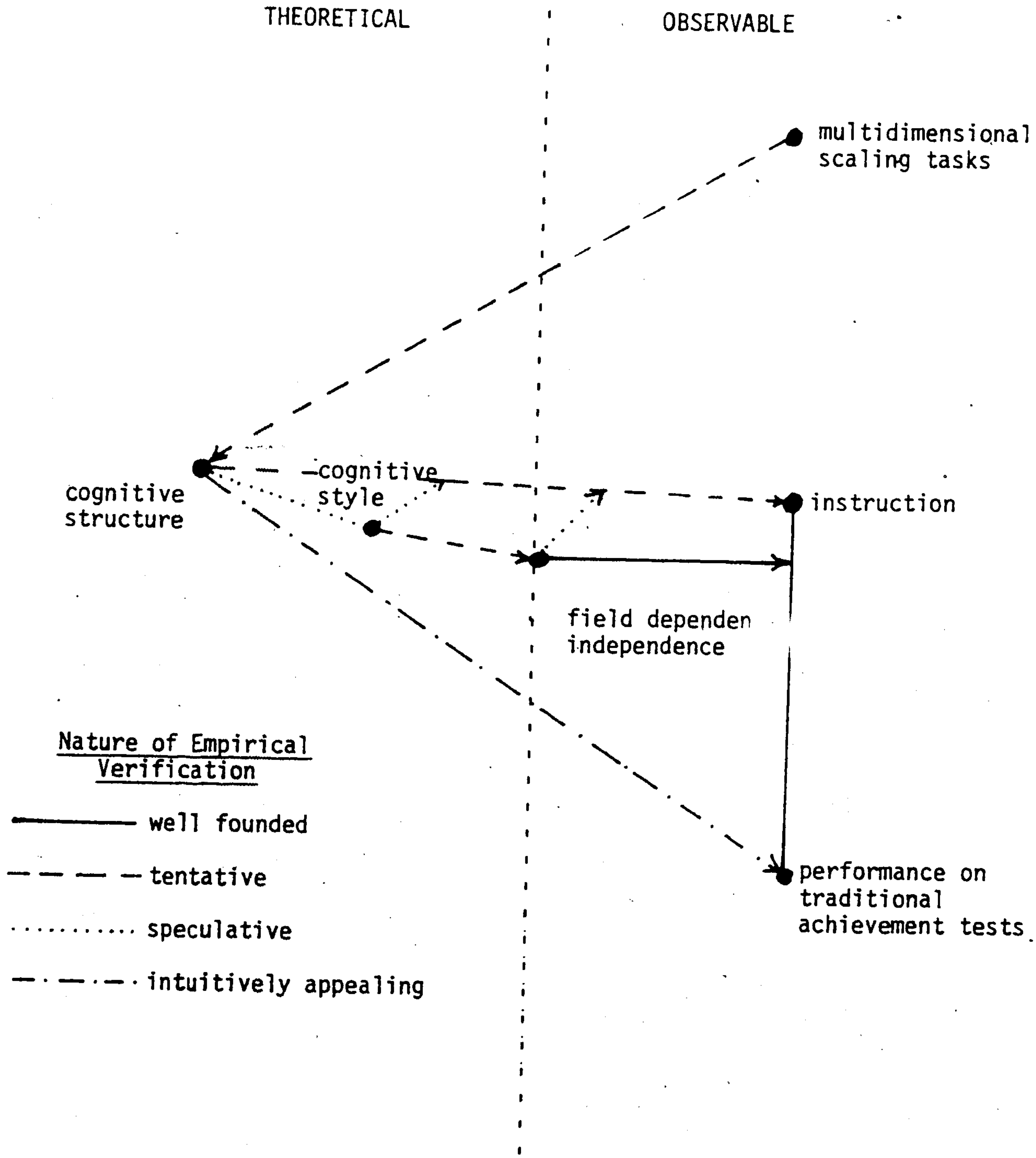


Figure II.2 A Subset of the Theoretical Network

III. ADVANCE ORGANIZER

In an attempt to minimize confusion and to provide the reader with a map of the study, the following simplified overview is provided.

The research reported evolved from the basic problem: can the acquisition of cognitive structure be measured with greater fidelity than currently possible using conventional achievement tests? A series of measures of cognitive structure (tests) were derived and their construct validity examined using the contrasting groups strategy: if a test can distinguish among groups which are known to differ on a construct, evidence pertaining to the test's validity is provided. Simultaneously, evidence pertaining to the construct's validity is also provided.

Two groups (and several subgroups) were used in the study. Since the tests were intended as measures of achievement (specifically of cognitive structure), the performance of a group of achievers (experts) was contrasted with the performance of a group who were relatively naive in the concept domain. Next, the tests' abilities to detect changes during learning were examined by testing the naive group after an instructional experience and comparing the pre, post and expert measures.

The theoretical network described in Chapter II suggested that the nature of instruction should influence cognitive achievement. Accordingly, the tests were also examined regarding their abilities to detect differences between groups of students receiving different instruction. The theoretical framework further implied that the learning instruction link might interact with a student's cognitive style. This interaction was also examined.

Additionally, since several tests of cognitive acquisition were developed, the convergence among the measures was also examined.

Construct validity , of course, is a "raising oneself up by one's bootstraps" concept. Examining the way a test functions provides evidence pertaining to the construct's validity as well as to the test's validity. This potential to explore the nature of the construct (cognitive structure) gave rise to sub-problems involving the description of the construct.

Succinctly, the derived measures of cognitive structure were examined from the following perspectives:

1. Can the measures differentiate people who have learned a concept domain from those people who have not?
2. Are the measures sensitive to changes during learning?
3. Are the measures sensitive to variations in instruction?

Additionally, the theoretical network was explored with regard to:

1. What is the nature of cognitive structure?

2. How is cognitive structure related to instruction?
3. How are cognitive structure, instruction and cognitive style related?

IV. METHODOLOGY

The design [illegible]ethodology of a study which empirically explored [illegible]tical network developed above is reported in this ch[illegible] Initially, the problems are reviewed, subject sa[illegible]g and assignment procedures are discussed, and the experimental design outlined. A detailed consideration of the treatment conditions is followed by a section on instrument development including reliability and validity concerns. Next, the experimental procedures are reported. Finally, the rationale and procedures for data transformation and dependent variable generation are reported.

A. Problems

The study reported below attempted to explore empirically the research problems identified in Chapter II. To summarize, these problems were:

1. Can cognitive structure be validly measured?
2. How do different teaching strategies influence the acquisition of cognitive structure?
3. Does field dependence-independence influence the acquisition of cognitive structure?
4. What is the nature of the cognitive structure which is

acquired?

5. What is the relationship between cognitive structure and achievement?

B. Subjects

Students enrolled in an undergraduate course in classroom evaluation at the College of Education, University of Saskatchewan, were used as subjects in the study. The experiment was integrated with the course's instructional procedures. The instructional treatments constituted a module on introductory statistics which was part of the course. Performance on the achievement test was considered during determination of final grades. To that extent, participation in the study was a mandatory part of the course. The evaluation course was a required component in the secondary education program at the University of Saskatchewan. Students generally enrolled in the class during their second year of undergraduate studies. All such students (N=97) registered during the 1977-1978 academic year participated in the experiment. Three sections of the course were offered. At least partial data were obtained from all students. Since data were collected over many class sessions, student absences resulted in complete data sets for only 64 students. Additionally, since the capacity of one of the analysis programs was limited to 79 subjects when in some cases more existed, a sample of 79 was obtained by randomly discarding students whose data sets were incomplete. The maximum number of students omitted by this procedure was five. The analyses reported below employed varying sample sizes. In general, these principles were followed:

1. Group-based measures were computed using all students for whom the particular data were available (subject to limitations of the analysis program).
2. Individual differences comparisons involving two or more measures were made using all students for whom the appropriate measures were available.
3. Several analyses were also performed using only those students whose judgements met a minimum consistency criterion and for whom complete data existed.

Sample sizes for each analysis are reported in Chapter V.

C. Experts Sample

Cognitive structure data were also collected from eleven University of Saskatchewan education professors each of whom had taught an introductory statistics module at least once during the twelve months preceding data collection. Checks of judgement consistency (described in Chapter V) reduced this sample to eight. This sample was heterogenous with regard to age, professional experience and academic rank.

D. Design

The research problems were investigated using a conventional aptitude treatment interaction design employing random assignment of subjects to two treatment levels. Following the recommendations of Kerlinger and Pedhazur (1973) aptitude was measured continuously. Measures of cognitive structure, the dependent variables, were obtained both pre and post treatment.

E. Treatment

Based on the researcher's previous experience with students from the same population, introductory statistics was selected as the content domain used during the treatment phase. Among the criteria considered during selection of the content domain were:

1. anticipated nature of the cognitive structure--Interpretation of results was expected to be facilitated if concepts to be learned were few in number, readily interrelated, defined a concept neighbourhood with integrity, and were relatively new to the students;
2. nature of intended learning outcomes--The variables investigated in the study required that intended learning outcomes include both performance skills and cognitive structures;
3. feasibility--Research procedures were expected to be facilitated if the content domain (a) could 'stand alone' but still be integrated with normal course

offerings, (b) could be taught near the middle of the thirteen week course (after student interaction, motivation and support had developed but before final exam anxiety had set in), and (c) could be taught in a relatively short period of time.

The chosen content domain rated favourably on each of the above factors.

Within the theoretical position developed previously, processes and characteristics of cognitive structure ought be relatively independent of the particular content being learned. That is, the formal features of cognitive structure ought be similar whether concepts learned deal with statistics or English literature. This issue has yet to be explored empirically, however. Accordingly, no strong claims of subject matter generalizability will be made.

Two self-instruction packages were developed using the format and some passages from Christensen (1977). Each package used different cognitive mathemagenics, the independent variable. One package (C) incorporated many of the features of cognitive instruction models, e.g., advance organizers, graphic and verbal presentations, spiralling, questions prompting new concept relationships, summaries of concepts and principles, etc. The other package (P) was aligned more closely with behavioural models of instruction and placed greater emphasis on skill development and practice. Although both packages contained conceptual and computational presentations, C emphasized the conceptual approach while P emphasized the computational approach.

Both packages consisted of separate xerox handouts for each of central tendency, variability, normal curve, correlation, and reliability. All handouts followed the same format: objectives, definitions, discussion, examples, and practice problems. The objectives and definitions components were identical in both packages. The major variations between the two packages were contained in the discussion and practice problem sections. Discussion passages in package P tended to be concise, highly structured and dealt with concepts as relatively independent entities. Conversely, package C discussions tended to be more lengthy, contained more diagrams, and focused on the interrelationships among concepts. P contained a greater number of practice problems than did C. Package C totalled 32 pages while package P totalled 30 pages.

With the exception of the practice problem sections, both packages entailed a relatively passive approach on the part of the learner.

The handouts were supplemented by tutorials conducted by the researcher in styles consistent with each package. Separate tutorials were held for each instructional treatment. Issues raised during C tutorials were dealt with through conceptual argument while P tutorials emphasized computational solutions to learner's difficulties. Eight one-hour tutorials were available for each instructional condition. Students were allowed to attend any three tutorials corresponding to their instructional package. An additional topic types of measuring scales, was presented in the tutorial setting. Student questions outside of the tutorials were [illegible]uraged.

A major difficulty encountered during the instructional development stage was the overlap between the cognitive and behavioural models inherent at the instructional level. Although the terminologies and theoretical underpinnings contrast adequately, distinctions at the performance stage are less clear. Behavioural objectives, for example, are clearly an important component in the behavioural approach. Nonetheless, they may act as advance organizers, a decidedly cognitive entity. Definitions of terms, although formally the same in both instructional approaches, may assume different roles in the two models: within a cognitive model definitions might be interpreted as propositional statements; a behavioural perspective might regard definitions as word chains appropriate for rote learning. Accordingly, the two instructional packages were undoubtedly less distinctive than had been originally intended. Samples of portions of both instructional packages are presented in Appendix A.

F. Instruments

Achievement

A pool of instructional objectives, consisting of those objectives used during instructional development and supplemented by additional objectives identified during an analysis of the learning materials, was generated. Two classroom evaluation instructors reviewed the pool with regard to its adequacy and completeness and their criticisms were used to revise the pool. A fifty-item multiple choice achievement test based on the objective pool was then constructed and the test reviewed for content validity and technical flaws by the same instructors. Slight revisions were made until the test was deemed acceptable by the judges. Following administration of the test to a group of students enrolled in the evaluation class during a previous academic term, an item analysis was performed and unsatisfactory items revised. Although items from nearly the entire difficulty range were retained, the majority of the items were discriminating ones of middle difficulty. Following post-instruction administration of the test to the experimental sample, a further item analysis was performed. The KR-20 for this administration was .74 which was considered acceptable given the nature of the objective pool and the relatively homogenous abilities of the students. No additional reliability or validity information was obtained. Appendix B contains a copy of the test.

<u>Cognitive Structure</u>

Since all the measures used in this study to assess cognitive structures involved judgements of concepts, the initial stage in instrument construction was the identification of a concept pool. A review of the content area and statistics textbooks resulted in a pool of approximately fifty introductory statistical concepts. Written free associations to each of the concepts were then collected from five graduate students who had completed several advanced measurement and statistics courses. The patterns and associations were then analyzed and a reduced set of thirty concepts which had overlapping free associations were identified.[4] The reduced set of concepts became the stimuli in a similarity sorting task which was piloted on two groups of subjects. One group (E) of subjects (n=10) consisted of graduate students who were familiar with the concepts while the other group (N) of subjects (n=10) consisted of students' wives, secretaries, and undergraduate students, none of whom was familiar with the concepts. The data were then analyzed using Wiley's (1967) Latent Partition Analysis (LPA, discussed below). Concepts which, in a six-partition solution of E data, evidenced confusion among partitions were purged from the study as were concepts

---

[4] Garskoff and Huston (1963), Shavelson and Stanton (1975), and Nagy (1977) are among the researchers who have collected word associations as the raw data for assessing cognitive structures. The technique is based on a concept network interpretation of cognitive structure (similar to that employed in this study) where word associations are considered 'near neighbours' to the stimulus concept.

which displayed similar partitioning in both E and N groups. Of the remaining 25 concepts, 15 which appeared to have the clearest clustering characteristics were selected as the basis for two of the cognitive structure instruments. These 15 concepts are referred to as the reduced concept set. The 25 concepts retained after the LPA review were supplemented by five additional concepts from the original set and this pool of thirty concepts (called the complete concept set) became the basis for the third cognitive structure instrument. The two concept pools are presented in Table IV.1.

Due to the abstract nature of cognitive structure and the desire to retain as much certainty and fidelity as possible in its assessment, construct validity checks were carried out on the cognitive structure measures. Three separate instruments were developed to assess the network among the concepts listed in Table IV.1. Convergence among the instruments, of course, would be interpreted as evidence of construct validity. All three instruments were based on the premise that judged similarity among concepts is inversely related to interconcept distances in cognitive structure; that is, when an individual judges the similarity among concepts, his judgements manifest the relationships among the concepts in his cognitive structure.

CARD SORTING TASK

1. correlation
2. error of measure
3. frequency
4. interval
5. mean
6. median
7. mode
8. nominal
9. normal curve
10. ordinal
11. range
12. rank
13. raw score
14. variance
15. z-score

reduced concept set

16. coefficient
17. criterion
18. deviation score
19. equivalence
20. internal consistency
21. norm
22. prediction
23. ratio
24. reliability
25. stability
26. standard deviation
27. statistic
28. true score
29. validity
30. variable

complete concept set

Figure IV.1 Concept Sets

This measurement activity requires that subjects sort a set of concepts into mutually exclusive and exhaustive subsets such that concepts in the same subset are more similar to each other than to concepts in other subsets. Inter-concept proximity is measured by the proportion of times two concepts are sorted into the same subset. Both the instructions and the task are relatively simple, allowing a large number of concepts to be judged in a relatively brief time.

Although card sorting techniques have been used frequently in research on abstract structures, reliability and validity evidence is not extensive. Rapoport and Fillenbaum (1972) reported agreement between results based on card sorting data and results based on a more-complex concept-tree construction task. Shavelson and Stanton (1975) also reported that similar results were obtained using card sorting, concept-tree construction, and word association data.

Test materials given to each student included a page of instructions, a set of concept cards, and a set of cards to be used to separate the subsets. The instructions asked the individual to consider the meaning of each of the concepts and then to sort the concepts into piles on the basis of similarity or closeness in meaning. Subjects were allowed to determine the number of subsets. Each concept deck was composed of IBM cards with one of the thirty concepts listed in Table IV.1 printed in 4 mm high letters on each card. All decks were arranged in the same random order. Each subject was provided with twelve separator cards and an additional supply was available during data collection. The number of the separator cards distributed with the concept cards proved to be sufficient. A sample of the instructions is contained in Appendix B.

No formal investigation of reliability was conducted. Wiley's LPA (1967) did provide estimates of confusion among partitions which might be construed as indicating inter-rater consistency. Such consistency, however, may be more a measure of conformity than instrument r 'iability.

SIMILARITY RATINGS

Similarity rating tests involve subjects rating the similarity (dissimilarity) between paired concepts. The instrument consists of an item for each possible pair of concepts in the study. The items are generally presented in the following format.

Desk:Chair

similar |---|---|---|---|---|---|---|---|---| dissimilar

Raw ratings are considered to be direct measures of interconcept similarities and are the bases for deriving proximity measures.

Although for university students the task is not a demanding one, the number of items in an instrument increases geometrically as concepts are added to the study. Fifteen concepts result in 105 items; thirty concepts would require 420 items. Accordingly, the number of concepts which can be investigated with this technique is limited.

Considerable research has been conducted using similarity rating data. As with the card sorting task, however, reliability and validity evidence is limited. Johnson (1967) reported a strong relation between word association and similarity ratings. Green and Rao's (1972) results based on direct similarities ratings compared closely with results based on derived similarities.[5] Nagy (1977) reported personal consistencies[6] in the .50 range although the meaning to be attached to this measure is not clear.

The instrument used in this study was based on the 105 concept pairs generated from the reduced set of 15 concepts listed in Table IV.1. Twenty-five of the pairs were repeated in the pool to provide a rating consistency check. The resulting 130 pairs were randomly ordered with the constraint that the repeated pairs appeared once in the first third of the instrument and once in the final third of the instrument. The items were presented ten to a page in a 14-page booklet. The first page of the booklet contained instructions for the task. A copy of the instructions and a sample page of items is presented in Appendix B.

----

[5] Derived similarities involve paired concept ratings on a set of bi-polar scales each of which taps a particular attribute. Direct similarities involve a single global rating for each concept pair.
[6] Personal consistency measures were obtained by having respondents rate a set of concept pairs twice and then correlating the two sets of ratings.

Personal consistency measures (Nagy, 1977) on the 25 repeated pairs were computed on post-test and expert data. Average consistency at post-test was 0.527 with 38 subjects having correlations greater than 0.50. Eight of the eleven experts had personal consistencies of greater than 0.60. Data for the remaining three were excluded from further analysis.

SIMILARITY RANKING

This task requires subjects to rank-order all concept pairs on the basis of similarity The test consists of a randomly ordered listing of all possible concept pairs and instructions ask that a 1 be placed by the pair in which the concepts are most similar, a 2 by the next most similar pair and so on until 105 by the least similar pair.

As with the rating task, the number of concept pairs to be ranked increases geometrically as concepts are added to the study. Unlike the rating task, however, the ranking activity places high demands on respondent's memory, attention, and motivation. Although Rappoport and Fillenbaum (1972) reported that by-subject test-retest reliabilities ranged from .43 to .90 after one month and that the instrument's results converged with findings based on tree-construction data, the technique has seldom been used in either semantic or cognitive structure research.

For this study, all possible pairs (105) generated from the reduced set of concepts were randomly ordered on facing pages on a test booklet. Instructions were similar to those described above. A copy of the instrument is contained in Appendix B.

Unfortunately, subjects responded very unfavourably to the pre-treatment administration. Fewer than half the subjects completed the task and several of those who did merely sequentially numbered the pairs as listed. Many subjects reported frustration and hostility. Accordingly, the ranking instrument was dropped from the study.

Field Dependence-Independence

The Group Embedded Figures Test (GEFT; Witkin, Ottman, Raskin, & Karp, 1971) was used to assess field dependence-independence in this study. The timed test consists of black and white line drawings of complex figures which have simple figures embedded within. Each of the complex figures contains one of 8 simple figures printed on the back page of the booklet. Subjects are required to trace the outline of the simple figure present in each complex drawing. The 23 complex drawings which comprise the test are subdivided into three sections: a set of five practice items and two parallel sets of nine items. Each set of items is timed separately. This and similar tests of field dependence-independence have been used extensively in research during the past thirty years.

Reliability of the test for this study was estimated using a split-halves procedure. Although internal consistency estimates generally would be inappropriate for an instrument which purports to measure an enduring, general attribute, research constraints precluded more appropriate procedures. Nonetheless, the split-halves estimate computed herein (r=0.78) was comparable to the split halves estimates reported in other studies which also computed satisfactory stability estimates (e.g., Witkin, et al, 1971).

G. Procedures

Treatment and data collection portions of the study were conducted over a three-week period during February and March, 1978. Subjects were distributed across three sections of the course. Two sections (I, II) had three one-hour classes per week while the remaining section (III) met for 1.5 hours twice a week. Pre-testing using the three similarities instruments was conducted during the first two class sessions with each group. The instruments were administered in a fixed order: card sorting, similarity rating and similarity ranking. All subjects completed the card sorting task in less than twenty minutes; the rating task required less than forty minutes; those individuals who complete the ranking task required approximately one hour. Prior to the end of the second class session, students were randomly assigned to the two instructional treatments and the instructional packages dealing with measures of central tendency were distributed. Regular class sessions were cancelled for the next week and one half (for sections I and II this involved five classes; two classes were cancelled for section III) and 16 one-hour-long tutorials were scheduled (eight per instructional treatment). Handouts for the remaining topics were distributed during tutor sessions.

Post-treatment data were collected during the third week of the study. During the second last class of the week, concept similarity measures were obtained (card sorting followed by similarity rating). The achievement test was administered during the final class session. The GEFT was administered during a regular class session three weeks later.

Subjects were aware, naturally, that the normal flow of the course had been interrupted and that a research study was being conducted. Before the study was begun, subjects were told that two different instructional packages were being evaluated and they were requested to confine their studies to the particular packages which they received. Though some subjects undoubtedly di ssed the packages amongst themselves, general compliance with the request seemed to prevail.

H. Generation of Variables

None of the cognitive structure instruments provided direct measures of cognitive structure or of cognitive structure changes. These measures were generated from the raw similarities data using multidimensional scaling techniques. Two general classes of variables were created: variables which were based on group-average data and variables which maintained individual differences.

Group average measures were used to address two of the major problems outlined at the beginning of this chapter: construct validation of the measures and the nature of cognitive structure.

The strategy used to explore construct validation included comparing pre instruction, post instruction and expert data. Construct validity evidence would be provided by discrimination of the groups. Card sorting data for each group was arranged into a joint occurrence matrix and this matrix was input into Wiley's (1967) [illegible] Partition Analysis. The resultant partition structure and the confusion (omega) matrix for each group became the basis for intergroup comparisons.

Kruskal's (1964b) multidimensional scaling was also performed on the joint occurrence matrix for each group and the resulting spaces were compared. Finally, the joint occurrence matrices were analyzed using Carroll and Chang's (1970) INDSCAL with groups assuming the role of individual subjects. Location of each group in the subject (group) space was then examined.

The remaining problems (i.e., what variables influence the acquisition of cognitive structure and what is the relationship between cognitive structure and achievement) were investigated using individual differences data. Raw data were provided by the similarity rating task. Initially, ratings for each subject were standardized to a mean of 10 and a standard deviation of 2. Separate INDSCALs were then performed on the pre instruction, post instruction and expert data sets.

INDSCAL produces two types of results: concept spaces and subject spaces. The latter were the principal interest in this study. The subject space, which shares the dimensions of the concept space, provides measures of the communality for each subject with the other subjects in the analysis. The distance a subject is from the origin is a measure of the fit between an individual's data and the group results. An individual's co-ordinates in the subject space indicate the salience or importance of each dimension in the person's judgements.

Subject spaces for the pre instruction and post instruction data using the expert group structure as a target were determined. Similarly, the subject space for preinstruction data on the post instruction group structure was derived. Measures of fit and subject weights for these spaces became the basis for the generated learning measures. Derivation of these variables is discussed in Chapter VI.

I. Summary

This chapter presented the rationale and methodology for a study designed to test the construct validity of several cognitive structure measures and to explore the theoretical network upon which the measures were based. Problem distillation, sampling, and the experimental design were discussed. Treatment conditions, instrumentation and experimental procedures were described in detail.

Adequate measurement of cognitive structure, a principal focus of the study, was addressed concurrently with two major issues:

1. the nature of cognitive structure which is learned;
2. the relationships between cognitive structure and achievement.

The chapter immediately following reports findings pertaining to the nature of the learned cognitive structure; Chapter VI reports results concerning the second issue. Findings regarding cognitive measurement are reported in both chapters.

V. RESULTS: NATURE OF THE COGNITIVE STRUCTURE

The nature of the cognitive structures for the introductory statistics concept domain is discusssed in this chapter. Initially, a framework for interpreting the results is established. A group by group report of the findings based on card sort task is followed by a comparison of the groups. Group average results of the similarity rating task are then reported. The chapter ends with a brief summary.

A. Interpretive Framework

The nature of the cognitive structure for the statistics concepts was explored using group-average data obtained in both the card sorting and similarity rating tasks. For the sake of interpretability (and admittedly at the expense of weakening the design) a reduced version of the experimental design outlined previously was employed. Subjects from both instructional conditions were combined into a single group and the groups' cognitive structures prior to and subsequent to instruction were estimated.[7] The cognitive structure of the expert group was also derived and

----

[7] This may have resulted in a post instruction structure which was a compromise for both instructional groups. At least weak evidence was found that suggested the two groups evolved slightly different cognitive structures. The evidence is reported below.

comparisons were made among the three sets of structures.

Initially, the card sort data for each group (pre instruction, n=84; post instruction, n=84 and expert, n=8) were re-organized into a joint occurrence matrix. The similarities matrices based on the similarity rating task were scaled--by individual--to a mean of 10 and a standard deviation of 2 and the scaled matrices then averaged for each group. The three joint occurrence matrices (30 concepts) and the three similarities matrices (15 concepts) were then analyzed using a variety of analysis procedures. In general, the analyses were intended to:

1. identify, for each group, concepts which were closely related; that is, concept neighbourhoods;
2. estimate the "crispness" (integrity, degree of resolution) and inter-relatedness of the neighbourhoods;
3. identify concept districts, that is, neighbourhoods which were closely related;
4. identify concepts which might have functioned as connectors between neighbourhoods;
5. compare the various groups on each of the foregoing.

Shepard (1974) noted that most methods of multidimensional scaling "have been based upon the assumption of an underlying space that is continuous and has well defined dimensionality (p. 411)." Although such methods are appropriate for mapping domains involving continuous physical variations, he observed that the procedures were not so well suited for conceptual or semantic domains which

"... appear to be inherently more discrete, categorical, or bipolar (p.411)." He also suggested that hierarchical clustering systems (eg, Johnson, 1967), although suited to domains which are hierarchically nested, are unable to represent psychological properties that correspond to overlapping subsets. As an example, Shepard noted that in an hierarchical cluster system "...once cat is grouped with the other household pet 'dog,' it can no longer be grouped with other felines 'lion' and 'tiger' (p.411)."

Accordingly, the combination of clustering and MDS techniques such as that employed by Shepard (1972) is not likely suited to representing cognitive categories in a relational structure.

An approach involving a combination of Latent Partition Analysis (LPA, Wiley, 1967) and Kruskal (1964b) multidimensional scaling appears to be free of the above difficulties. LPA provides a categorical representation of objects in a domain as well as measures of affiliation (1-delta**2) between an object and its latent category. Information regarding an object's tendency to cross to other latent categories is also available (the phi matrix). Estimates of a category's integrity as well as its overlap with other categories are determined (the omega matrix). Since the partitioning of objects is dependent upon the number of latent partitions requested rather than the category structure determined by analysis for a different number of partitions (that is, a different level), an object which is loosely affiliated with a particular partition at one level of analysis can shift its affiliation to another partition at a different level of analysis. Thus, in the cat example above, 'cat' and 'dog' might be partitioned together at one level of partitioning while 'cat', 'lion', and 'tiger' might be grouped together at another level.

By examining the partitions derived from at least two different levels of partitioning, one level involving many partitions, another level with few, identification of concept neighbourhoods, districts (related neighbourhoods) and concepts which seem to relate neighbourhoods should be possible. Intelligent speculations regarding the 'crispness' of the categories , that is the integrity or degree of resolution present in the LPA solution, also should be possible. In the present situation, for example, the crispness of the categories should increase as one moves from pre instruction to post instruction to expert data.

Additional support for the cognitive structure would be provided if Kruskal scaling of joint occurrence data produced spatial configurations in which the objects were located in groups similar to the LPA partitions. The stress measure would provide additional information regarding the structure's crispness: a structure which summarizes group average data based on highly idiosyncratic judgements can be expected to have greater stress than a structure derived from group average data evidencing high inter-subject agreement. Kruskal scaling of the omega matrices would provide additional insight regarding neighbourhood inter-relatedness.

With groups assuming the role of subjects, the concept space produced by Carroll and Chang's (1970) INDSCAL scaling of the joint occurrence matrices would be another, albeit compromised, estimate of concept groupings. Finally, the INDSCAL subject space should be a useful indicator of the relationship among the cognitive structures of the various groups. INDSCAL analysis of group average similarity rating data will provide additional independent estimates of group cognitive structures. Convergence among the many estimates, of course, would be evidence of the structures' validities.

Essentially, the latent partitions were interpreted as neighbourhoods (when the number of partitions was large) and districts (when the number of partitions was small). A particular concept's affiliation with a neighbourhood (that is, whether the concept tended to belong to only one neighbourhood or tended to belong several neighbourhoods and might act, therefore, as a connector) was determined by 1-delta**2. Diagonal elements in the omega matrix were interpreted as indicating neighbourhood uniqueness while off-diagonal elements were considered to represent relationships among neighbourhoods.

Neighbourhood characteristics were based on eight-partition analyses, while district characteristics were based on four-partition analyses. Among the factors considered when determining the number of partitions in the neighbourhood analysis were:

1. the scree test--a break occurred at either the eighth or ninth eigenvalue for all groups;
2. the number of eigenvalues greater than one (either eight or nine for all groups);
3. the sum of the first eight eigenvalues was similar for the groups;
4. the interpretability of the partitions;
5. the average number of manifest partitions (expert, 7.25; pre, 8.65; post, 8.35).

The number of partitions for the district analysis was determined arbitrarily.

Kruskal (1964b) scaling was performed for spaces ranging from one to four dimensions. Both Kruskal's stress 1 and stress 2 were computed for each solution and these values were considered when judging the adequacy of the structures. As Klahr (1969) suggested, however, comparison of stress values to those obtained from random configurations is more appropriate when there are no *a priori* notions about the spatial arrangement than when such notions exist. Since speculations regarding the structure being studied existed (based on the pilot study, textbook analysis, analysis of the learning materials, and the work of Traub & Hambleton, 1974), since comparison with independent estimates of the same structure were possible, and since the principal interest was interconcept distances rather than identification of spatial dimensions[8], stress values were of only minor importance when selecting solutions for interpretation. Comparability with other solutions, interpretability and ease of presentation were also considered and these factors weighed heavily toward the two dimensional solutions presented below. Stress values for the Kruskal analyses are reported in zable V.1.

B. Expert Results

Neighbourhood analysis (eight partitions) of the expert data resulted in a very crisp and readily interpretable latent partitioning. Seven conceptual categories--central tendency, variability, normative scores, scales, reliability, validity and true scores-were derived. The remaining category, which contained eight concepts, consisted of basic terminology (see Table V.2). With the exception of the terminology category, remarkably strong measures of affiliation were observed. Of the twenty-eight concepts distributed across the seven partitions, only four had 1-delta**2 of less than .75 and, as expected, these concepts seemed to bridge related categories. 'Norm', for example, tended toward both the central tendency and validity partitions in addition to the normative scores category in which it was placed. 'Criterion', in the

----

[8] It is worth noting that although semantic interpretations of dimensions can be influenced by the number of dimensions in the solution, inter-point distances are highly correlated among solutions of varying dimensionality.

Table V.1 Kruskal Stress 2 for Joint Occurrence Matrices

	Number of Dimensions			
Group	4	3	2	1
Expert	.227	.254	.309	.367
Pre instruction	.466	.510	.604	.676
Post instruction	.398	.432	.464	.545

validity partition, also tended toward the terminology group, while 'true score', in the category of the same name, evidenced some affiliation with the validity and normative scores category.

The concepts in the terminology category crossed several other categories. 'Correlation' was the most interesting of these concepts, tending toward the reliability, validity and true score partitions. The concept, apparently, bridges several neighbourhoods.

Review of the omega matrix revealed remarkably strong, consistent categories. The terminology category was the only one with a low diagonal value. Nonetheless, strong relationships among several categories were apparent.[9] These relationships are depicted in Figure V.1. The phi matrix is presented in Table V.2 while the omega matrix is presented in Table V.3.

District analyses (four partitions) resulted in the coalescence of the variability, normative scores, and true score partitions as well as the combining of the reliability and validity concepts. The basic terminology category disappeared, concepts from it being spread across the two previously mentioned districts and the measuring scales

----

[9] When a category has a large diagonal value as well as moderate off diagonal values, it indicates that the concepts in the category tend to be sorted as a unit, but that the unit tended to be combined with other subsets of concepts. A strong relationship was inferred between two categories if the off diagonal entry in the omega matrix was greater than .250. Values between .150 and .249 were interpreted as indicating weak relationships.

Table V.2 Expert Neighbourhood Phi Matrix

Concept	1-delta**2	1	2	3	4	5	6	7	8
range	996	894	010	034	007	007	015	-008	008
variance	996	894	010	034	007	007	015	-008	008
standard deviation	996	894	010	034	007	007	015	-008	008
nominal	[illegible]39	-036	993	-179	019	195	032	-012	-055
interval	993	021	1061	050	005	-081	-022	006	011
ordinal	993	021	1061	050	005	-081	-022	006	011
ratio	993	021	1061	050	005	-081	-022	006	011
norm	252	-239	031	433	234	060	238	-080	016
z-score	7[illegible]3	-018	-006	1000	-027	138	-004	027	-035
deviation score	815	130	011	1006	-018	-180	031	048	-024
mean	1001	006	007	-012	1003	011	-002	003	003
mean	1001	006	007	-012	1003	011	-002	003	003
mode	1001	006	007	-012	1003	011	-002	003	003
normal curve	448	-054	-080	525	-008	526	-112	-034	092
correlation	512	045	-060	-192	010	582	173	377	267
variable	247	-025	323	-245	-105	684	052	-066	-027
raw score	600	-049	221	363	-072	888	-110	-055	-029
coefficient	482	046	-027	-211	-030	892	310	097	117
rank	518	-093	015	162	-038	1067	082	051	-091
frequency	570	030	-141	046	075	1171	-139	-064	-082
statistic	543	068	-072	-129	064	1177	-111	-090	045
criterion	558	003	040	-143	-029	215	889	-122	049
validity	813	013	-023	059	010	-110	1077	105	-075
prediction	832	017	-024	036	002	-066	1150	-053	-068
equivalence	562	004	048	-119	-026	161	-016	724	050
internal consistency	1000	-008	-003	041	006	-048	-012	1022	-030
reliability	1000	-008	-003	041	006	-048	-012	1022	-030
stability	1000	-008	-003	041	006	-048	-012	1022	-030
true score	687	-106	124	298	-041	-268	228	-013	716
error of measure	1053	033	-033	-082	013	044	-111	-022	1155

* Decimal points have been omitted. Actual values are table entries x 10^{-3}.

Table V.3 Expert Neighbourhood Omega Matrix

Partition

	1	2	3	4	5	6	7	8
1	1202							
2	078	867						
3	406	214	775					
4	258	107	134	990				
5	181	095	226	159	408			
6	-012	023	016	007	097	681		
7	007	005	-024	-001	099	275	982	
8	212	117	372	092	167	196	165	854

* Decimal points have been omitted. Actual values are table values $x10^{-3}$.

Table V.3 Expert Neighbourhood Omega Matrix

Partition

	1	2	3	4	5	6	7	8
1	1202							
2	078	867						
3	406	214	775					
4	258	107	134	990				
5	181	095	226	159	408			
6	-012	023	016	007	097	681		
7	007	005	-024	-001	099	275	982	
8	212	117	372	092	167	196	165	854

* Decimal points have been omitted. Actual values are table values x10^{-3}.

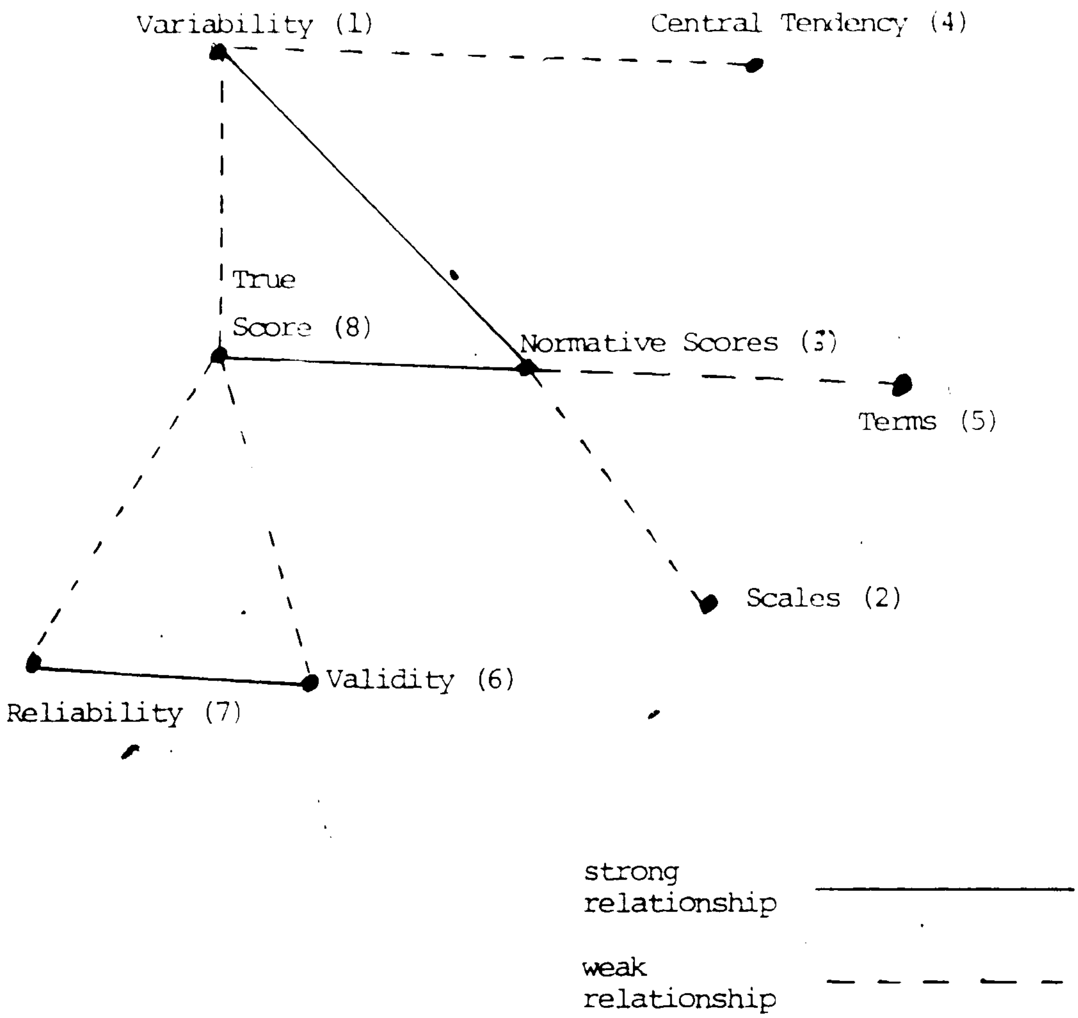


Figure V.1 Related Neighbourhoods: Expert Data district. In general, the terminology concepts displayed weak affiliations with the districts in which they were placed. They tended to cross several districts. Although such crossings might indicate that the concepts serve as bridges among districts, they could also indicate that the concepts are relatively discrete and are not closely related to any of the other concepts in the study. This latter position was supported by examination of the omega matrix. Despite the preponderence of crossing terms present in the reliability/validity and variability/norm districts, both diagonal and off-diagonal elements for these categories were low. The measuring scales and central tendency districts maintained their integrity and were both weakly related to the variability/norm district. Table V.4 contains the district-level phi matrix while the corresponding omega matrix is presented in Table V.5.

Although curriculum designers undoubtedly hope that instructional materials manifest the relationships inherent in a particular concept domain, the correspondence between the experts' partitions and the learning materials is still remarkable. The neighbourhood partitions match the subdivisions in the learning materials almost perfectly. The district partitions bear striking resemblence to the packaging and sequencing of the instructional units.

Kruskal scaling of the joint occurence matrix produced solutions with fair to good stress values (Kruskal, 1971 documentation) for configurations ranging from two to four

Table V.4 Expert District Phi Matrix

Concept	1-delta**2	1	2	3	4
criterion	149	531	086	042	025
coefficient	296	599	466	-033	056
prediction	211	646	064	-006	-036
validity	312	792	033	-007	-034
correlation	467	869	385	-073	055
equivalence	512	1025	-088	022	004
internal consistency	819	1297	-134	-021	003
reliability	819	1297	-134	-021	003
stability	810	1297	-134	-021	003
norm	096	084	250	146	151
true score	250	303	586	211	-171
rank	250	326	627	074	039
statistic	249	076	685	-065	178
frequency	265	037	727	-108	177
error of measure	280	267	780	029	-064
raw score	415	004	871	292	-031
normal curve	332	-024	948	017	-043
z-score	539	-046	1180	129	-131
deviation score	551	-088	1209	110	-143
range	584	-152	1247	-146	061
variance	584	-152	1247	-146	061
standard deviation	584	-152	1247	-146	061
variable	096	124	080	306	-007
nominal	793	057	-210	993	074
interval	976	-039	023	1084	001
ordinal	976	-039	023	1084	001
ratio	976	-039	023	1084	001
mean	985	004	-008	016	1015
median	985	004	-008	016	1015
mode	985	004	-008	016	1015

* Decimal points have been omitted. Actual values are table entries $x10^{-3}$.

Table V.5 Expert District Omega Matrix

Partition			
1	2	3	4
493			
049	381		
020	130	826	
015	174	095	955

* Decimal points have been omitted. Actual values are table values x10^{-3}.

dimensions. Although the scree test indicated that three dimensions was most adequate, little fidelity is lost when concept groupings in two space are examined. Interpretation of the dimensions is not recommended, however.

Consistent with LPA neighbourhood findings, central tendency concepts were closely clustered, as were concepts belonging to each of the variability, reliability, and measuring scales neighbourhoods.[10] The variability, measuring scales, and central tendency groups were relatively close together. Concepts from the remaining neighbourhoods were not clustered so neatly. Nonetheless, validity and correlation concepts were proximal to the reliability concepts, while 'z-score', 'deviation score', and 'normal curve' were close to the variability group. District groupings similar to those produced by LPA were present although 'variable', which had a very low district affiliation in the LPA, was closer to the variability/norm neighbourhood. Figure V.2 depicts the two-dimensional configuration with neighbourhood and district groupings shown.

Kruskal scaling of the neighbourhood omega matrix (Table V.3) produced neighbourhood proximities comparable to those manifest in the LPA results. The neighbourhood space and stress values are shown in Figure V.3.

Earlier the argument was advanced that people familiar

----

[10]Interestingly, these were also the neighbourhoods with the largest diagonal values in the omega matrix.

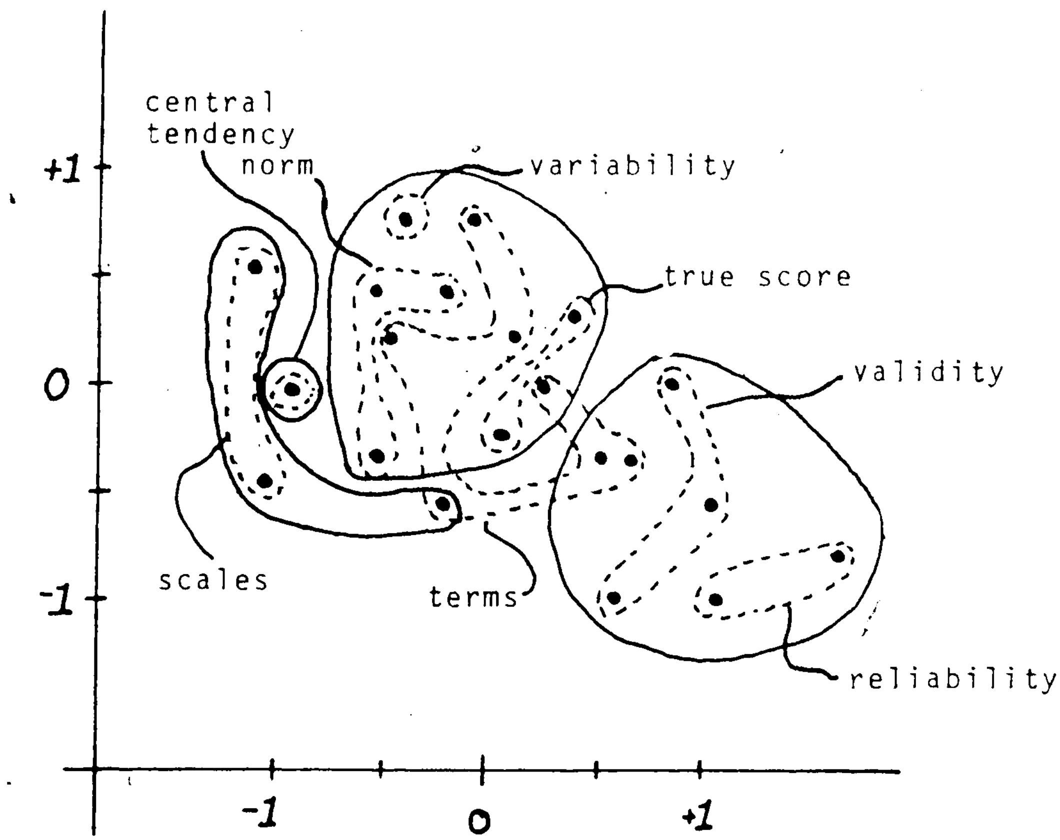


Figure V.2 MDS Solution: Expert S Matrix with a concept domain will share a well defined structure for that domain. The general agreement among the solutions above, which represent variations in data collection methods and analytic techniques, supports the validity of the structures presented as well as the validity of the methods used in their derivation. Further, the generally high affiliation and diagonal omega values from the LPA, low stress values in the Kruskal scaling, and high goodness of fit measures in the INDSCAL (reported below) all attest to the shared nature of the structures depicted.

One might still argue, however, that the structures described are unrelated to an individual's familiarity with the concept domain. Structural explorations for the pre instruction group, who presumably were much less familiar with the domain than were the experts, is reported in the next section.

C. Pre Instruction Results

Several of the neighbourhoods apparent in the expert eight-partition LPA were also present in neighbourhood analysis of the pre instruction data. Partitions corresponding to central tendency, variability, reliability and correlation were derived. The categories tended to include the same concepts that defined the categories in the experts' analysis. Additionally, a score neighbourhood emerged. The remaining ten concepts were distributed among three neighbourhoods which had no clear meaning. The

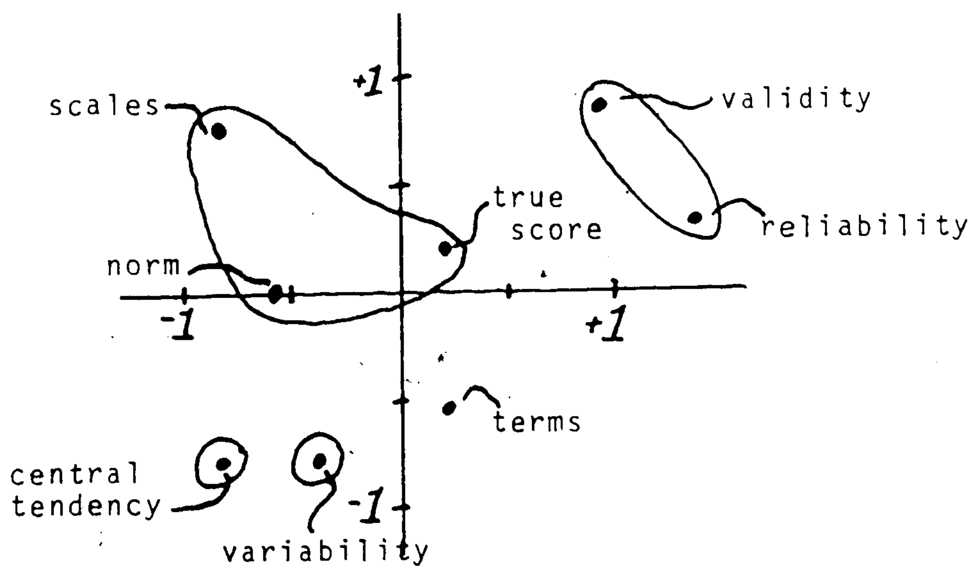


Figure V.3 MDS Solution: Expert Neighbourhood Omega Matrix crispness and measures of affiliation differed markedly from the expert analysis, however. Only the scores category--raw score, z-score, true score--was characterized by concepts with high 1-delta**2 values.

The values for concepts in the reliability category were also in the 0.6 to 0.7 range, however most of the remaining concepts tended to bridge several partitions. Diagonal values in the omega matrix were moderate to low for all but the reliability and scores partitions, suggesting a greater heterogeneity in the cognitive structures for the concepts than was manifest in the expert situation. A weak relationship existed between the variability and scores categories with 'deviation score' functioning as a bridge. 'Variable' and 'range' similarly seemed to link a category of general terms with the variability neighbourhood.

Several of the categories may be superficial. All of the words ending in 'ity' were grouped together as were all the 'score' words ('deviation score' sorted in the scores category with almost the same probability as in the variability category). Some subjects may have been sorting on the basis of similarity in the formal stimuli (appearance of the words) rather than conceptual similarity.

Despite some features common to the expert partition, the pre instruction partition evidenced considerable confusion among the group members regarding the concepts. Diagonal omega values tended to be low; off-diagonal values were near-zero. Apparently, the partition structure, although reflecting the dominant structure of the group, is not exten ively shared by group members. Further, the solution lacks the neighbourhood inter-relations present in the experts' partition. The phi and omega matrices for the neighbourhood solution are presented in Table V.6 and Table V.7, respectively.

District analysis resulted in a repartitioning of the concepts into a structure which was even less interpretable. Diagonal omega values, of course, were low. Two neighbourhoods--scores and reliability--changed only slightly (each added one concept). The remaining twenty-one concepts collapsed into two large but loose partitions. Table V.8 contains the district level phi matrix while the corresponding omega matrix is presented in Table V.9.

Kruskal scaling of the joint occurrence matrix produced the two-dimensional solution depicted in Figure V.4. The stress values were extremely high for all solutions attempted (number of dimensions ranged from one to four), indicating that pre instruction judgements cannot be adequately represented in a space with few dimensions (see Table V.1). This suggests that no crisp cognitive structure was shared among the students prior to instruction.

Kruskal scaling of the omega matrix resulted in the two dimensional solution presented in Figure V.5. The stress (0.1414, formula 1) was comparable to that expected for arrangements of eight random points in a two dimensional space (Klahr, 1969). Neighbourhood groupings in the space

Table V.6 Pre Instruction Neighbourhood Matrix

Concepts	1-delta**2	1	2	3	4	5	6	7	8
prediction	117	375	-029	090	019	024	-028	253	058
equivalence	241	803	040	154	-075	-032	-006	-030	001
coefficient	324	852	-024	-110	042	-092	063	047	244
correlation	580	1368	-045	-011	055	006	-067	-092	-093
raw score	664	-017	1016	-035	-140	-043	122	070	049
z-score	749	-073	1054	-015	060	036	-006	-030	-030
true score	739	006	1074	051	-096	-008	041	-035	-003
internal consistency	514	133	037	850	-067	023	095	-008	-013
validity	606	-015	-022	950	070	-017	-088	082	-006
reliability	728	-087	002	1055	016	006	029	022	027
stability	769	008	-008	1082	-026	020	022	-049	003
error of measure	257	073	132	168	739	-090	-187	041	-019
deviation score	647	-017	709	-030	747	-034	-314	-069	-003
variance	539	-043	-203	-004	890	-230	564	-052	076
standard deviation	483	100	-092	-028	1114	185	-161	024	-077
mode	242	-116	-015	005	-025	482	087	413	023
normal curve	290	234	034	-042	206	538	097	071	-055
norm	519	091	-061	-008	132	968	-130	043	002
mean	565	-034	020	016	-050	1070	029	-104	000
median	584	-093	000	019	-021	1081	052	-093	030
variable	261	031	-097	-011	346	-237	582	143	144
rank	233	125	105	-094	-067	233	590	140	022
range	510	-207	-093	-035	458	092	951	-046	-036
ratio	361	427	155	-097	-239	007	961	-085	-089
frequency	305	080	078	164	-294	017	1038	-048	-112
interval	437	-260	-003	013	019	-037	1193	-003	050
statistic	210	239	176	-098	-055	024	081	505	070
criterion	697	-069	-016	012	000	-055	-032	1301	-025
nominal	452	007	044	037	-008	108	052	-101	830
ordinal	827	-040	-011	-004	-019	-025	-043	009	1189

* Decimal points have been omitted. Actual values are table entries $x10^{-3}$.

Table V.7 Pre Instruction Neighbourhood Omega Matrix

Partition

	1	2	3	4	5	6	7	8
1	337							
2	079	667						
3	095	037	661					
4	100	170	043	403				
5	117	073	016	112	521			
6	114	031	042	195	120	332		
7	093	084	066	069	130	052	435	
8	104	053	007	062	125	111	076	605

* Decimal points have been omitted. Actual values are table values $x10^{-3}$.

Table V.8 Pre Instruction District Phi Matrix

Concept	1-delta**2	1	2	3	4
deviation score	548	971	291	-131	-053
raw score	627	1086	-087	014	-035
true score	714	1162	-155	-020	050
z-score	745	1188	-103	-003	-039
correlation	141	045	405	340	194
coefficient	121	036	491	293	027
error of measure	175	352	502	-139	194
rank	214	066	583	475	-074
frequency	187	-075	785	041	189
standard deviation	252	190	808	199	-021
ratio	231	047	887	139	-037
variable	256	-050	1083	-182	005
interval	327	-110	1196	-058	-033
range	446	-086	1359	032	-090
variance	486	-052	1514	-347	-028
prediction	068	022	134	307	192
ordinal	087	-003	288	396	-030
statistic	107	238	121	398	-007
criterion	076	111	-021	417	136
nominal	105	028	296	430	005
mode	[illegible]90	-003	-037	824	021
normal curve	2[illegible]0	088	275	831	-013
mean	48[illegible]	-032	-173	1364	-029
norm	49[illegible]	-038	-128	1369	-016
median	49[illegible]	-050	-131	1370	-036
equivalence	096	068	207	204	295
internal consistency	515	012	039	043	974
validity	600	-003	-074	-036	1069
reliability	708	-012	-021	-039	1157
stability	758	-037	-043	-032	1201

* Decimal points have been omitted. Actual values are table entries x 10^{-3}.

Table V.9 Pre Instruction District Omega Matrix

	Partition		
1	2	3	4
542			
075	248		
072	101	287	
045	049	034	533

* Decimal points have been omitted. Actual values are table values $x10^{-3}$.

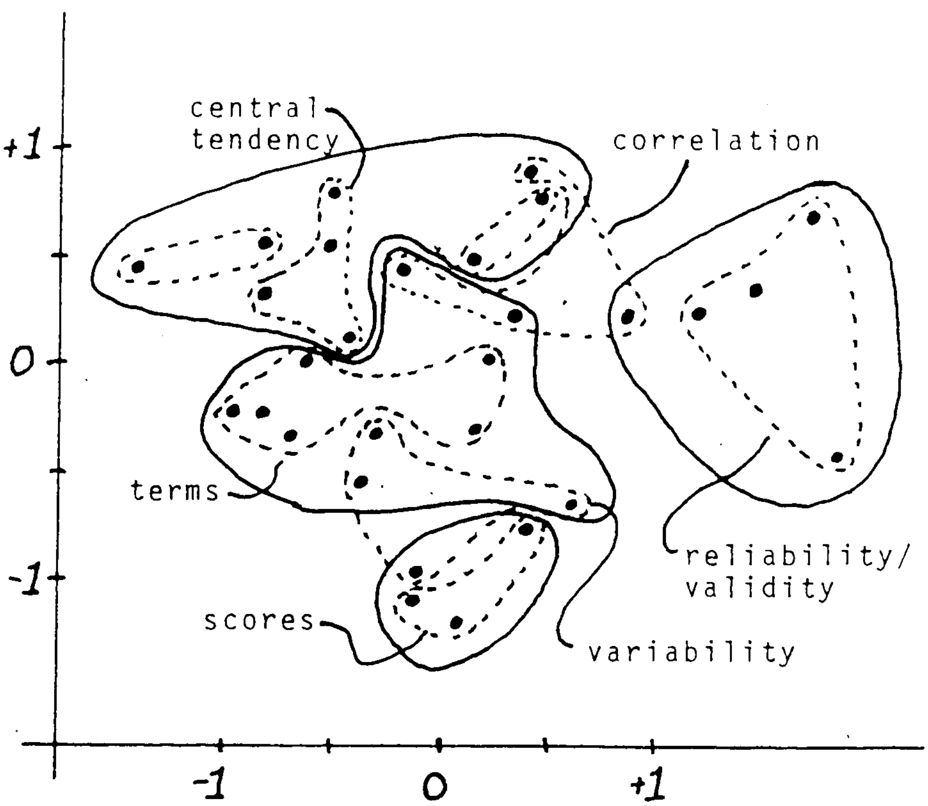


Figure V.4 MDS Solution: Pre S Matrix

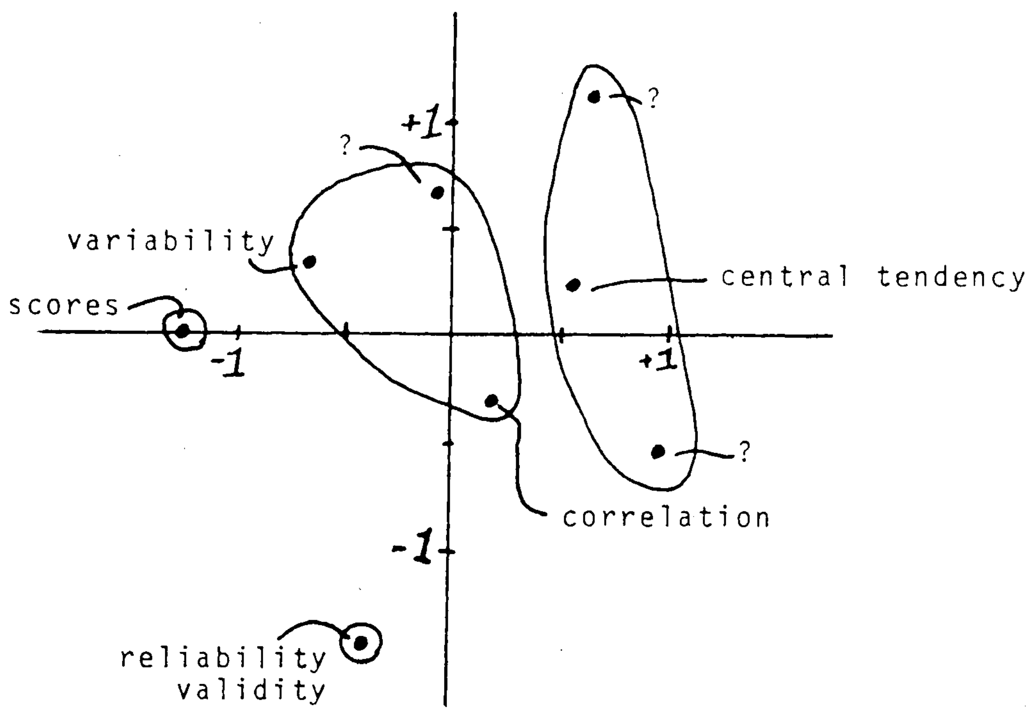


Figure V.5 MDS Solution: Pre Neighbourhood Omega Matrix are not readily interpretable.

Clearly, the cognitive structures for students prior to instruction lacked the crisp definition inherent in the experts' structures. Certainly, the learners were not completely naive in concept domain. However, low affiliation and diagonal omega values in the LPA, high stress values in the Kruskal scaling, and moderate goodness of fit values in the INDSCAL solutions (reported below) provide little support for a well defined cognitive structure which is common to learners before instruction.

Instructional impact on learners' cognitive structures and acquisition of structures which shared some of the features of the experts' structures would increase the validity evidence for both the structural representations and the methods used to derive them. Such evidence is reported in the next section.

D. Post Instruction Results

As might be expected, neighbourhood LPA of post instruction data produced a partition structure which shared features with both the expert and pre instruction solutions. Six interpretable categories evolved: central tendency, variability, reliability/validity, measuring scales, normative concepts, and true score. Although not as crisp as the expert solution, neighbourhoods tended to consist of the same concepts as the corresponding expert neighbourhoods. The variations that occurred tended to be in the bridging concepts. Partition affiliations also tended to fall between the corresponding expert and pre instruction values.

Diagonal values in the omega matrix indicated reasonable category integrities with only the normative concepts, correlation and prediction partitions evidencing substantial inconsistencies among individual structures.

Off-diagonal elements in the omega matrix suggested weak relationships among the true score, normative concepts, and variability neighbourhoods. 'Z-score' tended to bridge all three partitions while 'range' and 'deviation score' crossed the latter two. 'Error of measure' tended toward the correlation group as well as linking the true score and variability neighbourhoods.

District level LPA resulted in the carry over of two discrete categories (central tendency and measuring scales) from the neighbourhood analysis and the distribution of the remaining twenty concepts into two unstable casseroles. This was similar to the pre instruction result. Table V.10 and Table V.11 present the phi and omega matrices, respectively, for the neighbourhood solution while Table V.12 and Table V.13 present the corresponding information for the district LPA.

Kruskal scaling of the joint occurrence matrix resulted in a two dimensional solution with relatively high stress. Stress measures for the three and four dimensional solutions were only marginally better. At best, the structures were not widely shared by the post instruction group. Figure V.6

Table V.10 Post Instruction Neighbourhood Phi Matrix

Concept	1-delta**2	1	2	3	4	5	6	7	8
internal consistency	595	946	102	-031	049	-015	006	-004	-031
validity	587	979	-103	-016	009	063	-006	-007	008
stability	752	1096	-006	-025	-005	-030	-008	-003	062
reliability	802	1142	-132	079	-012	-055	010	013	-014
equivalence	252	372	414	034	-021	148	019	-012	077
correlation	484	063	1095	000	-068	-245	-002	-023	177
coefficient	603	-140	1232	010	-002	102	-007	-035	-074
error of measure	278	144	319	478	216	-016	031	001	-239
raw score	577	-107	-032	956	-077	133	007	013	061
true score	758	082	041	1141	-105	-023	024	005	-158
variable	247	002	298	-021	508	444	-040	036	-295
range	348	000	056	-118	674	-175	123	016	338
deviation score	517	-068	-158	317	740	030	-051	-055	274
standard deviation	681	-017	-045	-031	1192	026	010	-028	-023
variance	719	038	-012	-118	1285	-007	-033	-006	-149
prediction	199	106	309	-052	-058	525	-060	-046	126
criterion	293	085	-027	-109	-030	916	-037	011	111
statistic	428	-080	-076	106	020	1160	012	035	-083
frequency	235	002	147	-067	133	-066	403	124	243
median	812	-004	-007	021	-052	050	1037	009	-055
mode	814	010	-008	-006	017	-033	1048	-019	-058
mean	841	-002	-008	026	-002	-016	1054	-019	001
interval	407	-013	097	-080	239	-137	-006	777	083
ratio	499	009	294	049	-035	-070	-041	895	-025
ordinal	[illegible]84	-001	-125	003	-036	095	-006	1105	-055
nominal	[illegible]3	004	-130	011	-068	043	010	1106	017
z-score		-055	-149	610	181	-091	-030	-025	611
norm		-001	-060	-142	-121	453	216	-017	786
rank		=057	02[illegible]	110	084	-069	-040	085	817
normal curve		04[illegible]	07[illegible]	-102	-103	-064	-100	-031	1249

* Decimal points have bee[illegible] o[illegible] [illegible] Actual [illegible]alues are table entries x10^{-3}.

Table V.11 Post Instruction Neighbourhood Omega Matrix

Partition

	1	2	3	4	5	6	7	8
1	634							
2	155	415						
3	117	060	625					
4	028	110	192	503				
5	070	122	095	052	327			
6	-001	018	031	072	085	762		
7	015	062	052	068	084	129	569	
8	027	077	158	184	137	143	113	371

* Decimal points have been omitted. Actual values are table values x 10^{-3}.

Table V.12 Post Instruction District Phi Matrix

Concept	1-delta**2	1	2	3	4
frequency	221	438	015	182	308
median	802	1078	011	001	-083
mode	795	1078	002	-051	-047
mean	832	1096	-003	-041	011
criterion	064	066	222	195	170
prediction	076	-004	327	115	210
correlation	132	-036	426	115	321
equivalence	216	023	640	080	218
validity	549	009	1133	-035	-103
internal consistency	588	-001	1167	-023	-049
reliability	718	005	1295	-050	-100
stability	722	003	1297	-027	-097
interval	370	-025	-063	887	251
ratio	479	-084	100	1068	031
ordinal	644	-015	-061	1264	-144
nominal	653	007	-062	1265	-141
statistic	088	092	094	207	368
norm	193	367	030	186	375
coefficient	103	-047	299	140	375
variable	127	-068	146	092	629
normal curve	147	070	046	154	647
error of measure	197	-061	358	-040	694
rank	212	051	-071	192	818
true score	233	-071	299	-066	842
raw score	216	-036	068	-008	917
range	258	140	-098	007	954
z-score	378	-003	-050	-034	1244
variance	391	-073	-082	-089	1295
standard deviation	474	-023	-139	-103	1419
deviation score	529	-061	-161	-126	1509

* Decimal points have been omitted. Actual values are table entries x10^{-3}

Table V.13 Post Instruction District Omega Matrix

	Partition		
1	2	3	4
700			
005	437		
119	029	420	
075	055	072	252

* Decimal points have been omitted. Acutal values are table values x10^{-3}.

presents the two dimensional solution with LPA neighbourhoods and districts outlined.

Two-dimensional Kruskal scaling of the omega matrix resulted in a fair stress measure. The neighbourhoods generally were isolated although the proximity of the true score and variability neighbourhoods was evident. Figure V.7 shows the solution with LPA districts indicated.

Although the cognitive structures for students after instruction lack the crisp, readily interpretable categories and relationships evident in the expert structure, students clearly share a better defined structure following instruction than preceding it. Indeed, the trend toward acquisition of a structure similar to that of experts attests to the suitability of the expert structure as a target for both assessment and curriculum development purposes.

E. Relationships Among the Three Groups

The joint occurrence matrices were also analyzed using the INDSCAL procedure with groups assuming the role of subjects. Correlations between the final solution and the raw data (four dimensions, r=0.69; three dimensions, r=0.64; two dimensions, r=0.55) support adoption of the three dimensional solution presented in Figure V.8. Concept districts similar to those identified by the LPA were present. Interpretation of the concept space is of less interest currently than the locations of groups in the

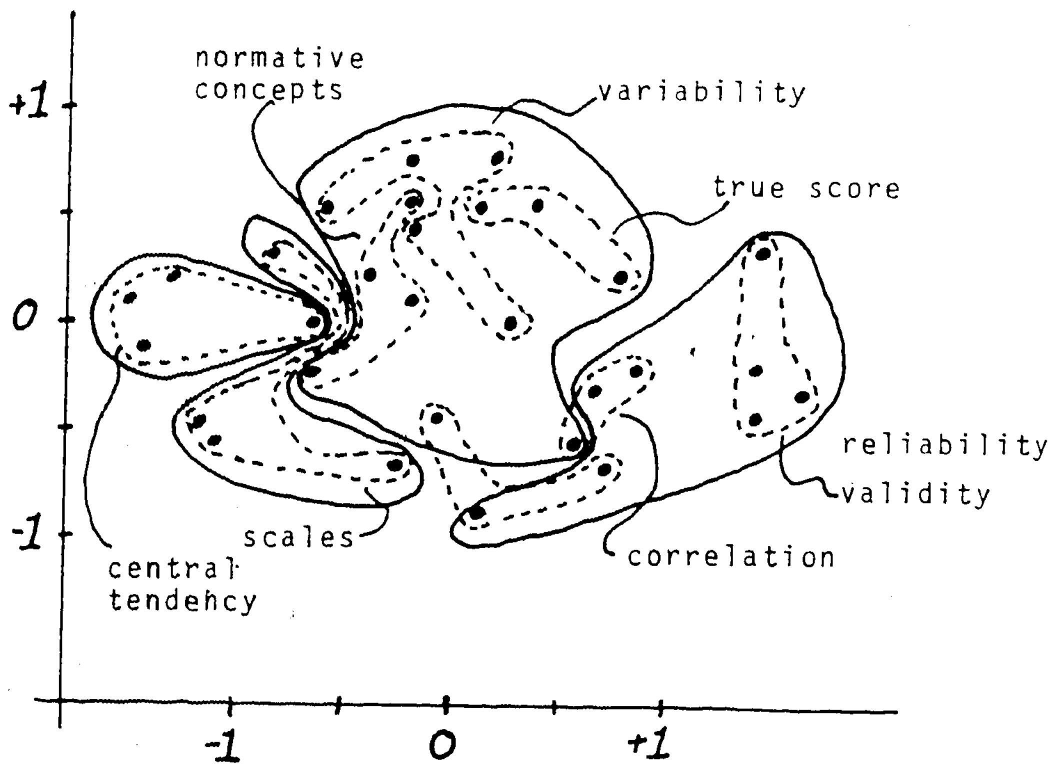


Figure V.6 MDS Solution: Post S Matrix

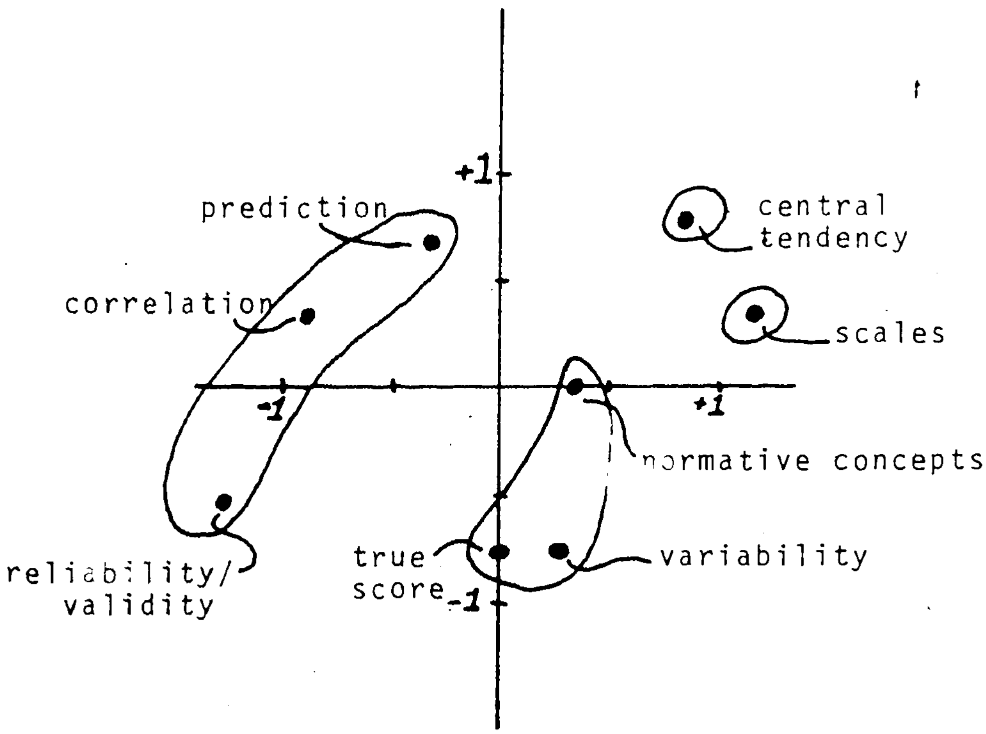


Figure V.7 MDS Solution: Post Neighbourhood Omega Matrix

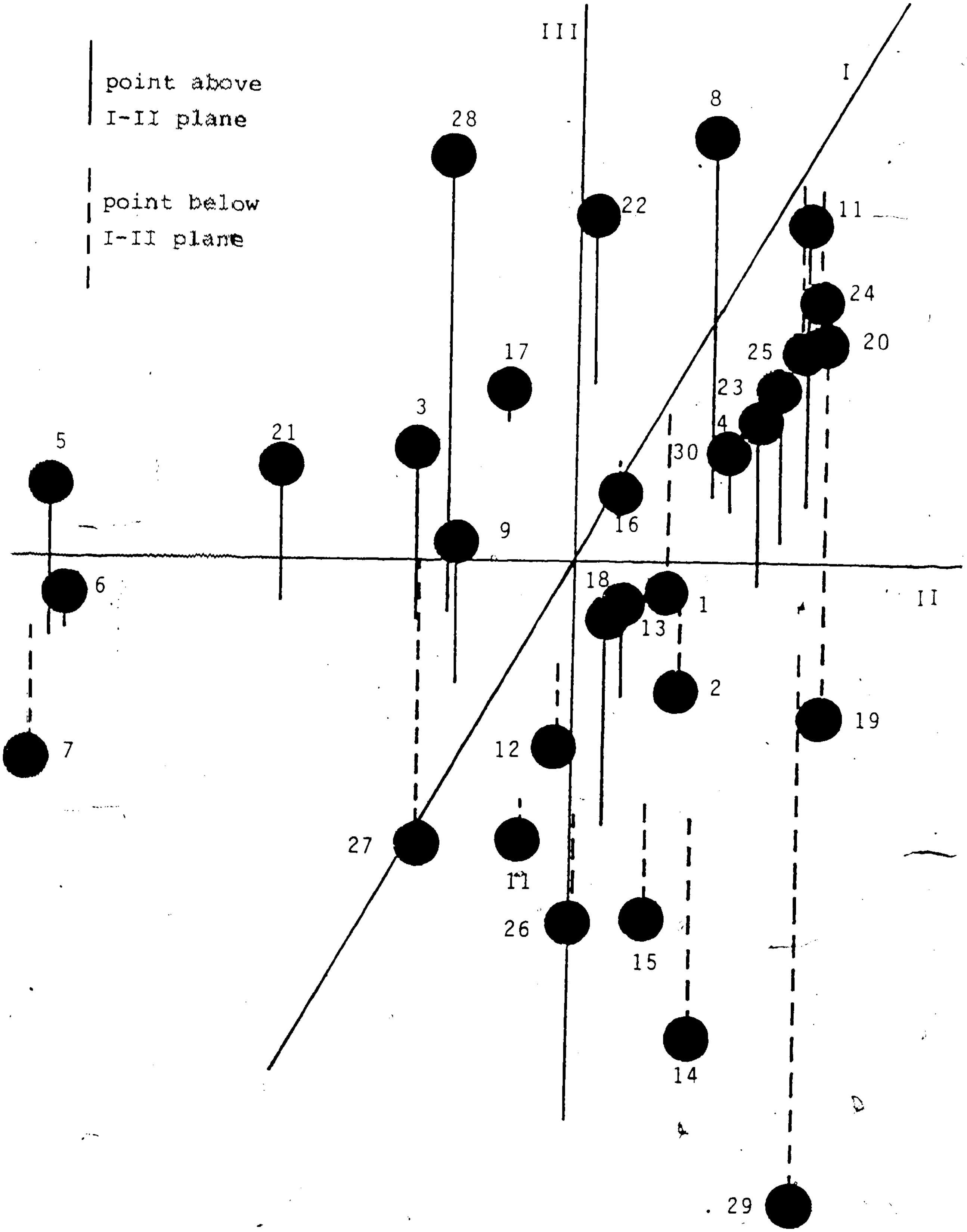


Figure V.8 INDSCAL Object Space: All Groups

1. correlation
2. error of measure
3. frequency
4. interval
5. mean
6. median
7. mode
8. nominal
9. normal curve
10. ordinal
11. range
12. rank
13. raw score
14. variance
15. z-score
16. coefficient
17. criterion
18. deviation score
19. equivalence
20. internal consistency
21. norm
22. prediction
23. ratio
24. reliability
25. stability
26. standard deviation
27. statistic
28. true score
29. validity
30. variable

Figure V.9 Concept Code for Figure V.8

subject space. Figure V.10 depicts the subject space while Table V.14 presents intergroup distances and goodness of fit measures.

Clearly, the expert group's judgements were most adequately represented by the INDSCAL solution, although the fit with pre instruction and post instruction data were also reasonable. As might be expected on the basis of the and Kruskal scaling results, the most disparate groups were the experts and students prior to instruction. By post instruction, the students moved closer to the experts. These findings are consistent with those reported above.

Despite the weaknesses of the research design, analyses of the card sort data indicates that the measurement procedures are sensitive to differences in cognitive structures and that cognitive structures change during earning. More evidence of a similar nature is reported in the next section which deals with analyses of similarity rating data.

F. Similarity Rating Results

The similarity rating data based on the reduced concept set was also analyzed at the group level. The judging task in the similarity rating situation differed markedly from that in the card sorting activity. During the latter task, subjects were encouraged to review all the concepts carefully, to sort concepts into categories, and to resort until the categories were satisfactory. The judgements (of

Table V.14 Inter Group Distances From INDSCAL Subject Space

	Group		Correlation
	Pre	Post	with solution
Expert	0.1445	0.0991	0.680
Pre		0.796	0.576
Post			0.653

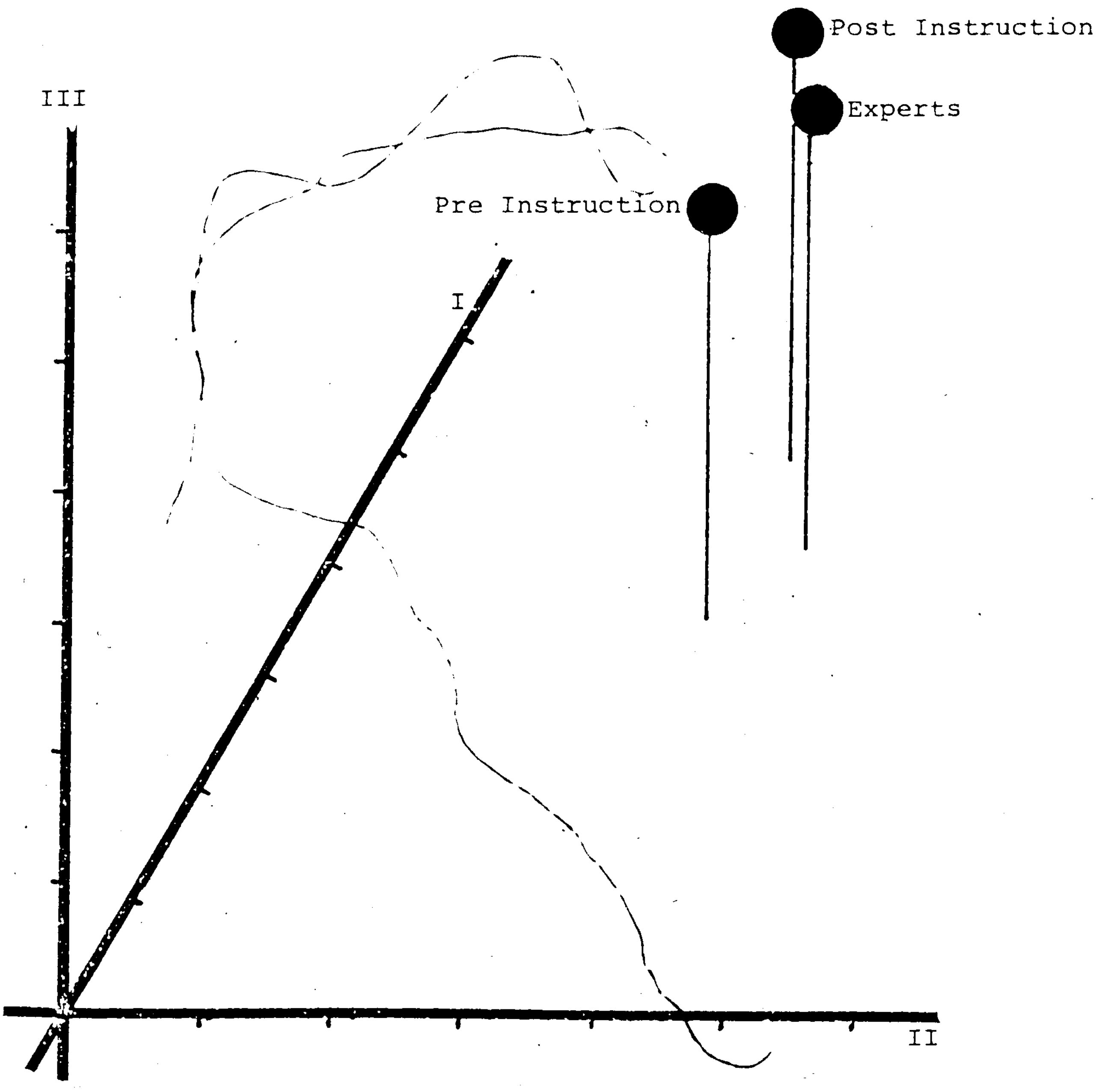


Figure V.10 INDSCAL Subject Space

each concept) were intended to be interdependent and the resorting possibility facilitated this. Further, the judgement, itself, is dichotomous in nature: concept _X_ is either similar to or not similar to concepts in a particular pile. Not surprisingly, data analysis produced concept relationships readily interpretable within a category model.

In contrast, during the similarity rating situation, subjects were encouraged to judge each concept pair as independently as possible and to indicate the degree of similarity between the concepts. Structures derived from the rating data, lacking the category crispness evident in the card sort structures, were not readily interpretable.

A major difficulty which confounds comparison of card sorting and similarity rating solutions is the variation in the concept sets used for each task. The large set of concepts used in the card sort activity contained several concepts from each of a few neighbourhoods. This clustering was evident in the LPA expert results. Subjects may have identified the neighbourhoods and used that information to faciliate the sort. The reduced concept set used in the rating task included few members of each neighbourhood, thus making it more difficult to identify the areas and use the information while making judgements. The change in the task itself (discussed above) may also have influenced the frame of reference used while completing the task. Knowledge of the categories, undoubtedly, would help subjects operationally define 'similarity'.

Fortunately, meaningful interpretation of the INDSCAL group structure was of lesser importance than were concerns regarding the adequacy of common group structures and general comparisons among the three groups. Table V.15 presents the goodness of fit measures (that is, the average correlation between an individual's raw data and the estimated data for that individual derived from the group solution) for each group. The goodness of fit between the pre and post instruction groups and the expert structure is also shown.

Although a three dimensional solution is reasonable in each case, the adequacy of the group structure differs across the groups. The high goodness of fit measures for the expert group further supports the position of a cognitive structure which is shared among people familiar with a concept domain. The decreasing goodness of fit values for the post and pre instruction groups indicates that single structures are less adequate in representing the cognitive structures for individuals in those groups. Nonetheless, the shared structure position seems more probable after instruction than before. These findings are consistent with those based on the card sorting data.

The correspondence between experts' structure and those of the pre and post instruction groups was also determined. Using the solve-for-weights option in the INDSCAL program, the fit between the experts' structure and the pre and post data sets was determined. The experts' structure fit the

Table V.15 INDSCAL Goodness of Fit Measures

Group	Number dimensions	r	r^2
Expert	4	0.718	0.523
	3	0.640	0.419
	2	0.548	0.307
	1	0.418	0.181
Pre	4	0.506	0.263
	3	0.449	0.211
	2	0.377	0.154
	1	0.280	0.090
Post	4	0.570	0.335
	3	0.515	0.277
	2	0.428	0.196
	1	0.307	0.104
Between Pre and E	3	0.343	0.125
Between Post and E	3	0.417	0.184
Between Pre and Post	3	0.415	0.178

post instruction data better than it fit the pre data. Goodness of fit measures for each group on the experts' target are included in Table V.15.

Although beyond the scope of the present study, correspondence among the structures may be of interest for instructional developers. Distances between corresponding concepts in the pre instruction-expert target diagram might indicate the degree of student understanding prior to instruction. Presumably, the smaller the distance between the concepts, the less the need for instruction. Post instruction and expert spaces might similarly be compared to identify concepts which require additional instruction. Further, interpretation of the spatial dimensions might be useful in planning instruction. Unfortunately, the complexity of the analytic methods undoubtedly will restrict these endeavours.

The pre group data also was analyzed using the post structure as a target. The goodness of fit (0.415) is considerably lower than that for the post data (0.515), suggesting a change in structure during the instructional process. These findings are also consistent with those reported earlier for the card sort data.

Statistical comparisons among the structures derived for the groups would have considerable appeal at this stage in the report. A simple test of statistical significance would provide (at least for the writer) welcome relief from (subjective) interpretations. Indeed, the Schoneman and

Carroll (1970) matrix matching procedure was considered. Unfortunately, the LPA, Kruskal scaling and INDSCAL solutions all suggest that inter-group differences seem to be more related to the clarity or definition of the cognitive structures than the spatial configurations themselves.

Succinctly, LPA solutions differed not so much with regard to partition membership, but rather with regard to the stability of the partitions. The major variations in the Kruskal scaling solutions were on the stress measures, or the adequacy of the group solutions. Variations among groups on the goodness of fit measure similarly existed in the INDSCAL solutions. The Schoneman and Carroll technique may not be sensitive to these variations.

Perhaps an example will clarify. Subject weights and correlations from the INDSCAL analysis provide information regarding the match between an individual's structure and the group structure. The weights indicate the expansion or contraction which must be applied to each of the dimensions in order to accommodate the individual's judgements. The correlation indicates the degree to which such accommodations will be successful. Small weights would result in contraction of the group structure. Since the correlation corresponds to the distance a person is from the origin in the subject space and the weights also define an individual's location in that space, contraction or expansion of the concept space is a meaningful adjustment when the correlations are not large. A space may have a particular configuration, but its adequacy vis a vis the raw data may be doubtful.

The Schoneman and Carroll procedure includes expansions and contractions. Thus a space which is contracted because it is a poor fit to the original data might match, after expansion, a space which adequately fit another group's data. Accordingly, the Schoneman and Carroll procedure seemed ill advised in the present situation.

G. Summary

The nature of the cognitive structure shared by each of the groups was described in this chapter. The expert structure was the most readily interpretable. The cognitive structure for the student group seemed to evolve from an idiosyncratic structure to one which was more commonly shared by the group and one which corresponded more closely to the expert structure.

VI. RESULTS: EFFECTS OF INSTRUCTION AND COGNITIVE STYLE

Analyses of group data support the positions that 1. group cognitive structures change during instruction, 2. the change is from highly idiosyncratic structures to ones which appear to be more common to the group, and 3. the change is also in the direction of the experts' cognitive structure. The utility of cognitive structure measures as the bases for assessing an individual's learning has yet to be explored.

This chapter reports results from the portion of the study which investigated the effects of instruction and cognitive styles on a variety of cognitive structure measures. The relationships between these measures and a more traditional test of achievement are also reported.

Previously the argument was advanced that acquisition of a cognitive structure should be facilitated by instruction which emphasizes the relationships among concepts. It was argued further that an individual's cognitive style would interact with the instructional approach in the formation of the cognitive structure. The results reported below address both these arguments. Only the similarities rating data were used in this part of the study.

Several cognitive structure-based estimates of learning were generated to supplement the more traditional achievement measure. The conceptual meaning, method of derivation, and potential weaknesses of each estimate are discussed in the next section. A description of the relationships among the variables follows. The impact of cognitive style and instruction on each measure of learning is reported prior to the chapter summary.

A. Derived Learning Measures

Agreement With Group Structures

Within the present perspective, learning--that is, acquisition of a cognitive structure--involves changing a learner's cognitive structure from a highly idiosyncratic or, in some cases, uninterpretable structure to one which is more similar to that of individuals who are competent in the domain. A measure of similarity between a learner's structure and a target structure (manifested by competent individuals), then, is a potentially worthwhile measure.

INDSCAL results include, for each subject, the correlation (or goodness of fit) between the individual's data and the computed structure. Essentially, the measure indicates the degree of communality between an individual's structure and the group's structure and, accordingly, is closely related to the distance between the origin and the individual in the subject space.

The extent of agreement between an individual's post instruction data and the post instruction group structure was used as one measure of learning. Average correlation for all students was 0.515. The correlations were converted to Fisher Z's. This variable is subsequently referred to as ZP.

ZP is subject to at least one major weakness. Essentially, it provides an indication of how close to the mean a particular individual is. If the mean represents the preferred structure, the measure would be a reasonable indicant of learning. Despite the shift toward the expert structure apparant in the pre and post instruction group structures, however, selection of the latter as the goal of instruction is difficult to defend.

Since both high and low achievers may vary considerably from the mean, and since ZP lacks directional information, ZP may be a rather crude measure of learning.

Acquisition of a structure which is similar to that shared by experts may be a more easily defended instructional goal. Accordingly, the correlation between each student's post instruction similarity ratings and interconcept distances in the experts' structure was determined. Average correlation for all students with the experts' structure was 0.4167; the corresponding correlation among the experts themselves was 0.6404. As before, the correlations were converted to Fisher Z's. This variable is referred to subsequently as ZE.

<u>Change in Agreement</u>

Implicit throughout this report is the position that learning involves acquiring a perspective of a set of concepts which is similar to that held by other people. In particular that might entail progressing from an idiosyncratic cognitive structure to one which has more in common with the structures of other people. Much of the group analysis (in particular the LPA omega matrices) supported this position. A reasonable measure of learning, then, might be the change during instruction of the congruence between a learner's cognitive structure and that common to a group.

Two measures of change in agreement were derived. Using the post group structure (that is, the three dimensional object configuration based on analysis of post group data) as target, correlations between each student's pre and post data and the target were calculated. Following conversion of the correlations to Fisher Z's, the pre agreement measure was subtracted from the post agreement measure. Hereafter this variable is referred to as DZP.[11] A similar procedure employing the experts' structure as target was used to derive the variable DZE.[12]

DZP and DZE may both be interpreted as representing an individual's change toward a common structure. DZP potentially has weaknesses related to those of ZP. Use of

----

[11] Of interest: Using a t-test, it was found that the mean of DZP was significantly greater than 0 ($p<0.01$) which was interpreted as indicating a significant shift toward an homogenous structure. The data were derived, of course, from a pre-experimental design, thus precluding attribution of differences to the instruction varaiable.
[12] Also of interest: the mean of DZE was significantly greater than 0 ($p<0.01$). A similar interpretation was made.

the post instruction structure (which is not likely the ultimate structure to be acquired) as target would result in confusing two types of individuals: those students who began instruction with a structure comparable to that which most learners eventually acquired and who, during instruction progressed toward a more superior structure; might be confused with students who, during instruction, developed misconceptions regarding the concepts. Based on pre instruction measures of agreement, there could have been few students of the first type. Further, since those individuals, presumably, were distributed randomly between instructional conditions, the problem should not confound conclusions regarding the instruction variable.

DZE was susceptible to a related problem. For some students the fit between their pre instruction data and the experts' structure was of the same magnitude as the fit between individual experts and the experts' structure. Accordingly, DZE would not have been sensitive to cognitive structure improvements for these students and may have been subject to a ceiling effect. Fortunately, based on the distribution of pre instruction agreement measures, few students could have been affected.

Change in Judgement Weights

The INDSCAL subject space also provides potential measures of learning. The distance a subject is from the origin corresponds to the fit between his data and the group solution. The further a person is from the origin, the better his struc `e corresponds to the group solution. A reasonable measure of learning, then, might be change in distance from the origin. Since distance from the origin is the same as the correlation measure previously discussed, such a variable would be the same as the DZ variables without the Fisher transformation. Such redundancy is not currently of interest.

When both pre and post points for an individual appear in the same space, however, Euclidian distances between the two points will reflect the changes in the individual's judging emphasis and, in turn, changes in his cognitive structure. DistP (based on the post group target) and DistE (based on the expert group target) were computed for each student and these variables functioned as additional measures of learning.

Figure VI.1 is a three dimensional graph depicting locations of C instruction students in the post target people space. Both pre instruction and post instruction points (represented by the tops of strings suspended above a plane) are shown. DistP is the distance between the tops of an individual's pre and post strings. Figure VI.2 graphs the corresponding information for the P instruction group. The pre and post strings seem to separate more in Figure VI.1 than in Figure VI.2 suggesting instruction differences. This observation is discussed further below.

DistE, of course, is the corresponding distance between an individual's pre and post locations in a subject space

Figure VI.1 Instruction Group C INDSCAL Subject Space

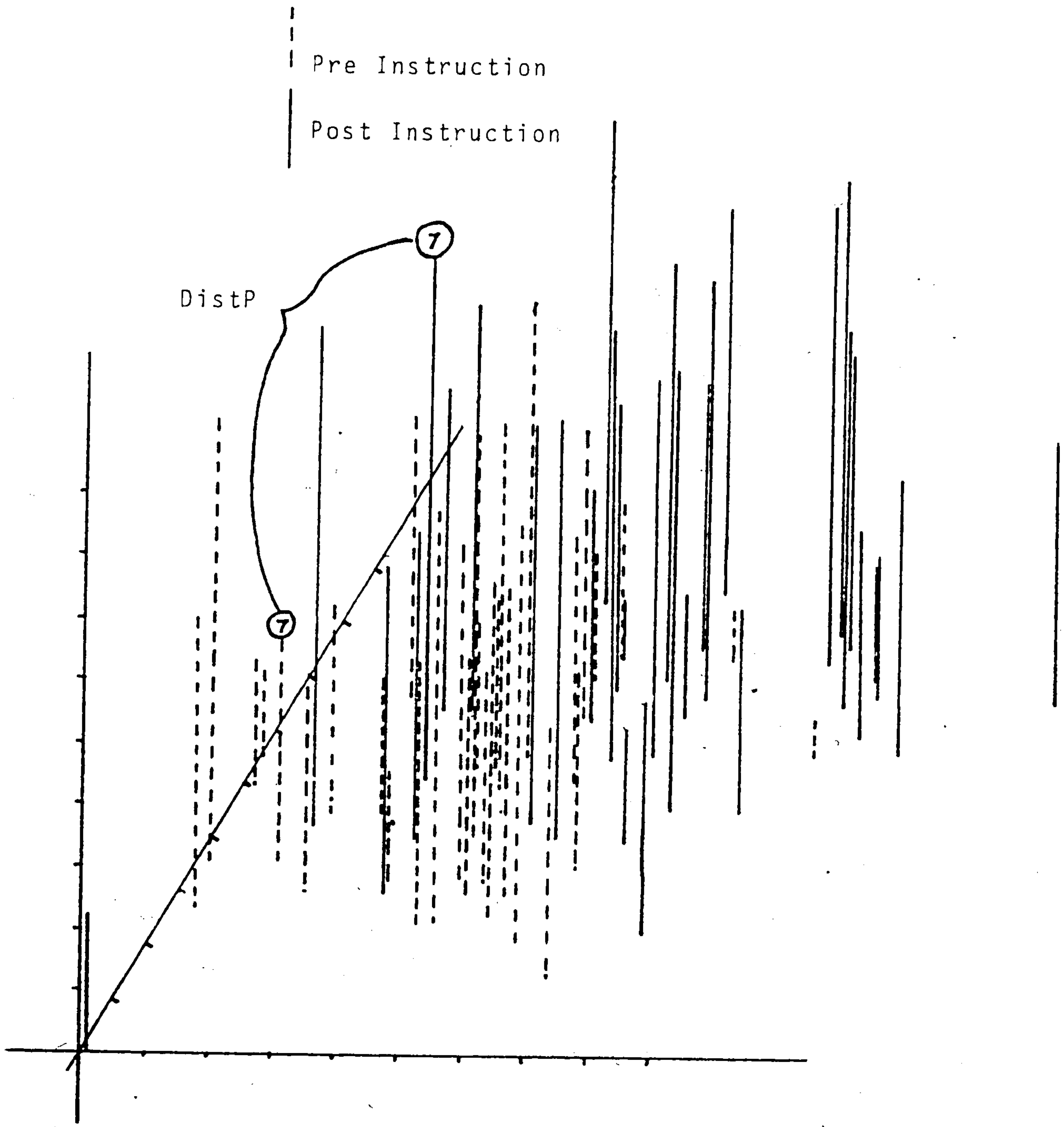


Figure VI.1 Instruction Group C INDSCAL Subject Space

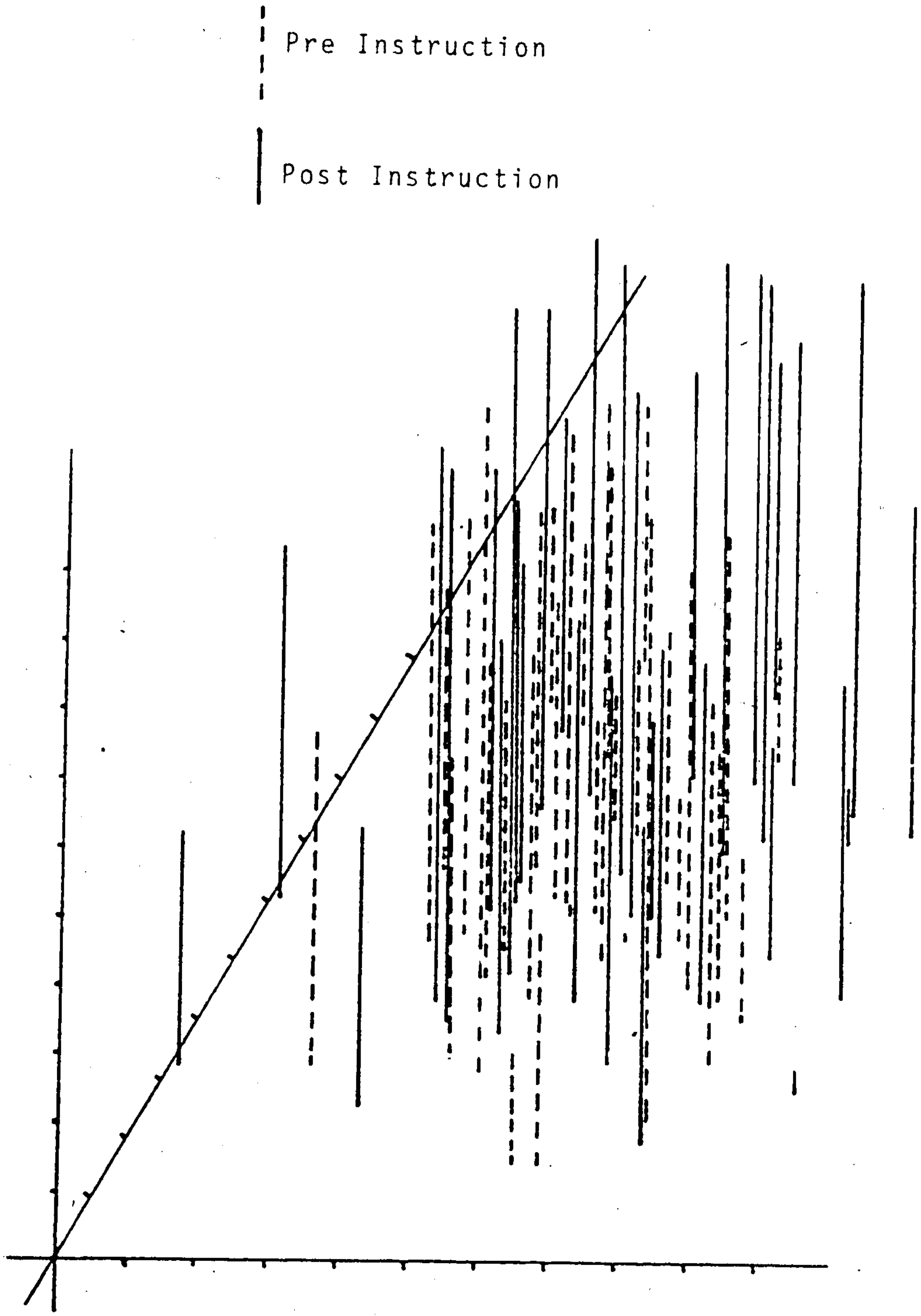


Figure VI.2 Instruction Group P INDSCAL Subject Space derived from the expert target.

Like the other cognitive structure measures, these variables are not without weaknesses. Whereas DZP can potentially underestimate (or even classify as 'unlearning') the learning of students who move from the group target toward a more superior structure, DistP and DistE err in the opposite fashion: all distances are interpreted as being directly related to learning, even those distances which manifest movement toward an uninterpretable structure. In light of the highly significant shifts toward the two target structures (the tests on DZP and DZE noted above), this is not likely a serious problem.

To review, six measures of cognitive structure learning were generated.

1. ZP--a measure of agreement between an individual's ratings after instruction and interconcept distances in the post group structure;
2. ZE--a measure of agreement between an individual's ratings after instruction and interconcept distances in the expert group structure;
3. DZP--a measure of change in agreement with the post group structure; specifically, the difference between ZP and the agreement of an individual's pre-instruction ratings with the post group structure;
4. DZE--a measure of change in agreement with the expert group structure; specifically, the difference between ZE and the agreement of an individual's pre-instruction

ratings with the expert group structure;

5. DistP--the distance, in the people space generated from the post group structure, between an individual prior to instruction and his location following instruction;
6. DistE--the distance, in the people space generated from the expert group structure, between an individual prior to instruction and his location following instruction.

Means and standard deviations for each of the derived measures are presented in Table VI.1.

Conceptual review of the six derived variables suggested their a priori ordering on the bases of sensitivity and suitability as measures of learning. DZE and DZP seemed to be the best measures, followed, in order, by ZE, DistE, DistP and ZP. Accordingly, the best tests of the aptitude-treatment-interaction and effects of instruction hypotheses might be those which employ either DZE or DZP as the measure of learning. Similarly, these two measures will be favoured when interpreting construct validity information.

Since considerable conceptual independence exists among the various measures (particularly between those based on the expert target and those based on the post instruction target) the relationship among the measures provides method-convergence validity data. Table VI.2 is the correlation matrix for the learning variables, including the achievement test. The table is discussed later in this chapter.

Table VI.1 Means and Standard Deviations

Variable	Mean	Standard Deviation
Ach	28.97	5.16
GEFT	12.75	3.69
ZP	.583	.150
DZP	.139	.185
DistP	.255	.099
ZE	.459	.126
DZE	.100	.156
DistE	.191	.079

Table VI.2 Correlations Among Variables

$N_{GEFT} = 61$

$N_{other\ vars} = 73$

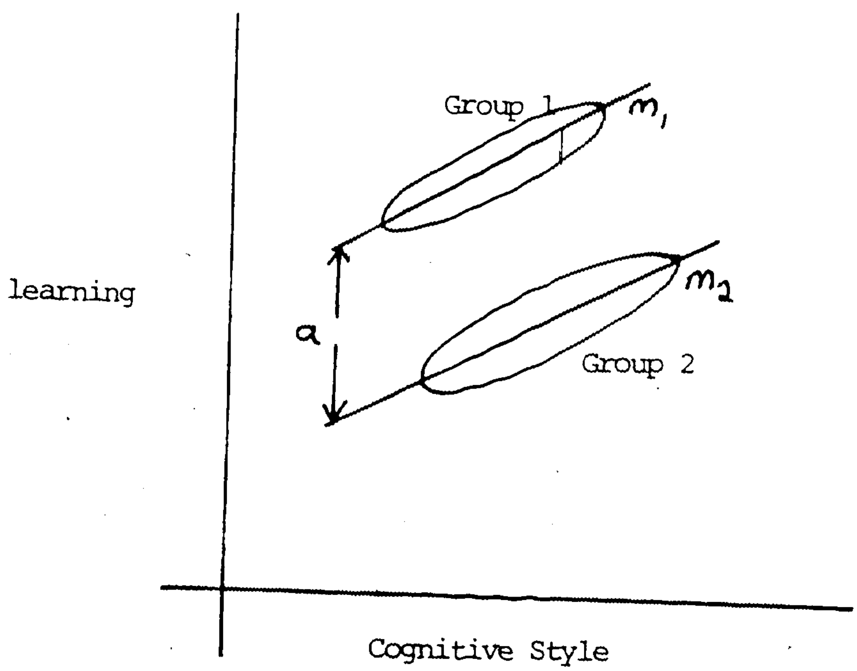
	ACH	ZP	DZP	DistP	ZE	DZE	DistE
GEFT	047	270	215	071	174	104	-093
ACH		347	278	220	290	266	201
ZP			842	302	653	478	302
DZP				438	531	512	293
DistP					066	191	-031
ZE						773	378
DZE							508

* Decimal points have been omitted. Actual values are table entries $x10^{-3}$.

B. Aptitude Treatment Interaction Analysis

On the basis of the theoretical discussion presented earlier, an hypothesis regarding the interaction between cognitive style and instructional approach was advanced. Cognitive style (field dependence/independence) was expected to influence the acquisition of cognitive structures for individuals in the P instruction group whereas it was expected to have less influence on the cognitive structures of the C instruction group. Specifically, a positive relationship between cognitive style and measures of learning was predicted for the P group while a neutral relationship was expected for the C group. Further, while the cognitive structures of field independent students were not expected to be influenced differentially by the instructional strategies, the structures acquired by field dependent students who received C instruction were expected to be superior to the structures of comparable students taught by P instruction (see Chapter II, section H).

This hypothesis was explored separately for each measure of learning by analyzing the regression of learning on cognitive style for the two instructional groups. The regression lines are shown in Figure VI.3. Differences between the slopes of the lines (m), of course, is evidence of an aptitude treatment interaction. Differences in group means (a) on the learning measure is indicative of the effect of instructional approach.



ATI Test: $m_1 \stackrel{?}{=} m_2$

$m_1 > m_2$

Figure VI.3 Test of Aptitude Treatment Interaction

Of the six derived learning measures only DistP was characterized by an aptitude treatment interaction ($p<0.05$). The ATI on the traditional achievement also approached significance. Results of these tests are reported in Table VI.3. The pattern of the regression lines was the same on both variables: cognitive style and achievement were inversely related in the C instruction group while a weak, positive relationship was present in the P group.

Although supporting the hypothesis of an ATI, this finding is not completely compatible with theoretical expectations. Rather than the C instruction bringing the achievement of field dependent people up to the level of the field independent, the C instruction may have affected detrimentally the achievement of the field independent group. Of course, the possibility also exists that P instruction was considerably better than expected. Figure VI.4 depicts theoretically expected slopes while Figure VI.5 depicts the slopes for the DistP analysis.

DistP was not expected to be the best measure of learning. Further, considering the many tests for homogeneity of regression which were conducted, the findings are less than convincing that an ATI was present.

The effect of instruction on each measure of learning was also investigated. Tests of differences between means for the two groups were conducted. The findings are reported in Table VI.4. Students receiving C instruction were expected to display greater shifts in their cognitive

Table VI.3 Aptitude Treatment Interaction Results

Variable	a priori ranking	df	$F_{homogeneity}$ of regression	p
ZP				
ZP	6	1,57	0.107	0.745
ZE	3	1,57	0.044	0.834
DZP	2	1,57	0.025	0.875
DZE	1	1,57	0.227	0.636
DistP	5	1,57	6.150	0.016
DistE	4	1,57	0.038	0.846

Equations for variable with significant interaction effect:

C instruction group

DistP = 0.01045 GEFT + 0.39077

P instruction group

DistP = 0.80501 GEFT + 0.15905

Table VI.4 Effects of Instruction

$N_C=35$

$N_P=38$

Variable	$\bar{X}_C$	$\bar{X}_P$	df	t	p
ACH	28.971	28.974	71	−0.00	0.999 (2-tailed)
ZP	.579	.586	71	−0.20	0.420 (1-tailed)
ZE	.470	.449	71	0.73	0.235 (1-tailed)
DZP	.176	.105	71	1.65	0.052 (1-tailed)
DZE	.129	.073	71	1.54	0.064 (1-tailed)
DistP	.256	.254	71	0.11	0.454 (1-tailed)
DistE	.214	.170	71	2.45	0.008 (1-tailed)

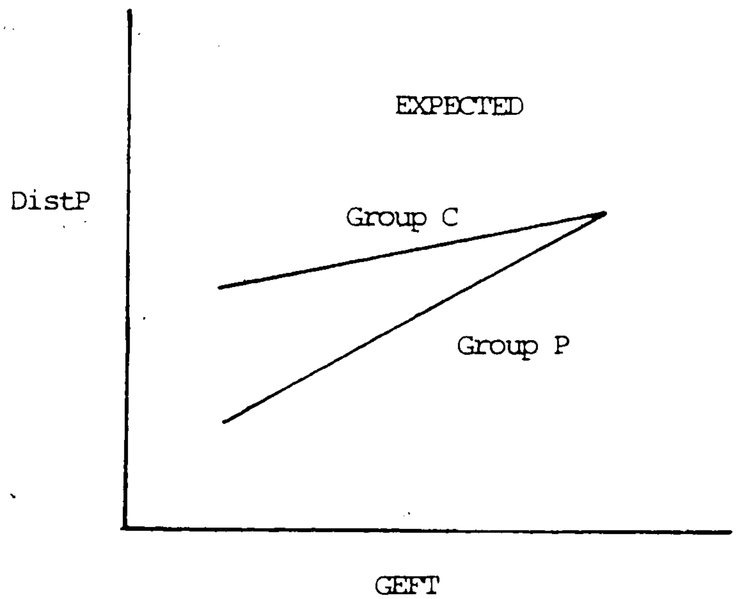


Figure VI.4 Expected ATI Regression Lines

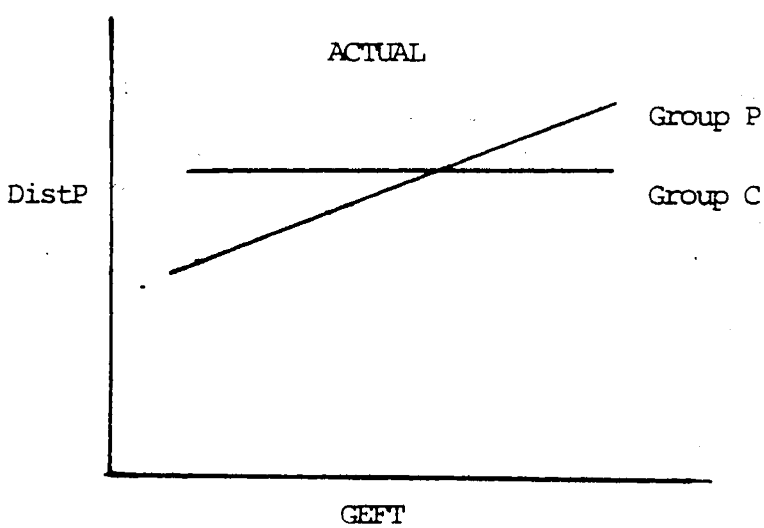


Figure VI.5 Actual ATI Regression Lines

structures than were students receiving P instruction. Further, the congruence with the expert cognitive structure was expected to be superior for the C instruction group. Considering the crudeness of the learning measures, a somewhat liberal approach to Type I error might be appropriate (0.10). Accordingly, group differences were indicated on DistE, DZE, and DZP. All of these variables were expected to be reasonable measures of cognitive structure acquisition. In each instance the observed mean of group C was higher than the mean of group P. Since DistE, DZE and DZP all measure change in structure, these results were interpreted as supporting the theoretical hypotheses. Concept relation-based instruction resulted in greater changes in learners' cognitive structures than did computation-based instruction.

Another important hypothesis was not supported, however. Cognitive structures acquired during concept relation-based instruction were expected to correspond more closely to the experts' structure than were cognitive structures acquired during computation-based instructi Group differences were not observed on ZE, the measure of fit between a learner's structure and the experts' structure.

The findings provide construct validity evidence for the derived learning measures. A construct valid measure should be able to distinguish among people who differ on the construct. This feature of the derived measures (of changes in structure) was demonstrated.

Failure to detect group differences on the traditional achievement test was also of considerable interest. Since P instruction involved greater attention to practicing skills (similar to those tapped by the achievement test) than did C instruction, the P instruction group might have been expected to excel on this measure. However, C instruction not only was superior in facilitating cognitive structure formation, but also was as successful as P instruction in developing the more performance oriented skills tapped by the traditional achievement test. This finding may be of considerable relevance for research into instructional strategies. Perhaps much of the confusion which surrounds assessing the relative effectiveness of various instructional strategies could be resolved by reconsidering the learning measures.

C. Achievement and Measures of Cognitive Structure

Correlations among the various measures of learning are reported in Table VI.2. Although all the derived measures were positively related with the achievement test, none of the correlations was particularly high ($0.20 < r < 0.35$). The greatest, achievement with ZP, was only 0.35. The crudeness of the measuring devices again should be considered, however. All the correlations were significantly greater than zero ($p<.05$). Similarly, the derived measures were inter-related, these correlations being somewhat larger. With the exception of DistP, all the correlations among the learning measures were larger than any of the correlations between a learning measure and the Group Embedded Figures Test. This pattern, combined with the patterns of group differences reported previously, supports the construct validity of the derived measures. Further, the derived measures were based on group cognitive structure estimates which converged across data collection and analysis methods. In sum, the validity of the derived variables as measures of cognitive structure acquistion seems promising.

DistP was the only measure which did not correlate with the other derived variables in the expected fashion, evidencing no correlation with either ZE or DistE. On conceptual grounds DistP was anticipated to be a poor measure of learning. Results of the ATI analysis on DistP, further, was not consistent with analyses of other variables. Accordingly, doubts regarding the variable arise from three perspectives. Its use as a measure of learning is ill advised.

DZE and DZP were both expected to be reasonable measures of learning. The patterns of their relationships with other variables supported that position.

D. Summary

The findings reported in this chapter generally supported both the theoretical positions advanced previously and the construct validity of the derived learning variables. The derived variables were all inter-related and, considering the coarseness of the measures, the relationships between the post target and expert target measures were particularly interesting. Failure to detect the anticipated ATI may be attributed to a variety of explanations besides poor learning measures. The presence of consistent instruction group differences on the derived measures combined with no differences on the traditional achievement test was of particular interest.

VII. Conclusions and Recommendations

Several major problems were addressed by the study reported above. Implications of the research for each of the problems are discussed in this chapter. Recommendations for further research also are presented.

A. Nature of Cognitive Structure

Several models of cognitive structure were reviewed in Chapter II. At that stage, a concept network model (Pylyshyn, 1973) was identified as a sound, parsimonious model for considering cognitive structure. Among the important attributes of the conceptual network model is the potential to represent both propostional relationships and the dynamic character of cognitive structure.

The results reported above are compatible with this perspective of cognitive structure. Interconcept distances derived using the various MDS procedures are interpretable, generally, within the network paradigm. In light of the many instances which have reported similar findings (eg: Cliff & Young, 1968; Preece, 1976a; Shavelson, 1973), the present results are not particularly startling.

Spatial configurations may not be the most apropriate models for cognitive structures, however. Confusion abounds regarding the apropriateness of spatial (and hierarchical) representations of semantic structures. Shepard's (1974) arguments in this regard were discussed previously. Latent partition analysis was used in this study in an attempt to accommodate Shepard's concerns. The procedure involved neither dimensional interpretations nor hierarchical assumptions. LPA of the experts' data resulted in cognitive structure representations which were consistent with Bruner's (1966) neighbourhood concept and Wickelgren's (1977) chunking model. Tight, well-defined concept neighbourhoods were observed. Relationships among neighbourhoods were also apparent. The clarity of the LPA results is evidence of a cognitive structure which is shared by people of similar competence.

The addition of concepts and propositions to an individual's cognitive structure has been addressed by many theorists, including Bruner, Goodnow, and Austin (1956); Joyce and Weil (1972); and Wickelgren (1977). A common perspective seems to have evolved. During learning, a new concept is initially added to the cognitive structure by attaching to a concept which the person already knows. During latter learning stages, the new concept is integrated more fully within the existing structure and super concepts (chunks or neighbourhoods) may result. Also as learning progresses, a cognitive structure will evolve which is similar to the structures of other people. This model was supported by the student data. The attaching of concepts, elaborating of relationships, chunking and evolution of common structures were evident in the results.

Pylyshyn's (1973) model of cognitive structure suggested involvement of the structure in the thinking process. This aspect of cognitive structure was not investigated in the present study. Accordingly, a reasonable problem for new research might be investigation of cognitive structure changes which occur during problem solving. The data collection and analyses techniques employed above may be adaptable for such investigation.

B. Cognitive Structure and Instruction

Inherent throughout the present study was the position that cognitive structures are learned and, hence, should be susceptible to instruction. The impact of instruction on cognitive structures, however, was not clearly demonstrated by the studies cited in Chapter II. Although LaPorte and Voss (1979) reported cognitive structure changes resulting from instruction and Traub and Hambleton (1974) noted shifts in cognitve structure which may have been instruction-related, other research (eg: Shavelson, 1972; Fenker, 1975) was contradictory.

Clearer findings were observed in the present study. Instructional effects were noted on both the card sorting and similarities data obtained from students. Both data sets also evidenced the acquisition of cognitive structures similar to those shared by experts. The findings were consistent using both group and individual data.

Citing the instructional suggestions of Ausubel (1968), Bruner (1966) and others, a thesis was developed in Chapter II that the nature of instruction should also influence the cognitive structure which is learned. In particular, instruction which attended to cognitive mathemagenics was predicted to result in cognitive structures which differed from those learned during performance-oriented instruction. The results reported in Chapter VI support this thesis.

These findings (which were based on the derived learning measures) are particularly interesting when contrasted with the failure of the traditional achievement test to detect instructional differences. Research into instructional effectiveness perhaps has been constrained by an inability to assesss cognitive structure changes. The derived measures may prove valuable in this area.

Closely related to the issue of instructional strategies and cognitive structure acquisition is the problem of learning transfer. A potential advantage of explicitly teaching for cognitive structure acquisition is (the theoretically expected) improved transfer (Champagne & Klopfer, 1980). The influence of cognitive structures on the transfer of learning was not investigated during the current research. However, in light of the present findings, such investigation may be a logical next step.

C. Cognitive Structures, Instruction and Cognitive Style

Witkin, Moore, Goodenough and Cox (1977) reported that cognitive style influenced the nature of the cognitive structure which an individual acquired. Further, several researchers (eg: Schwen, 1970) have reported that individuals of differing cognitive styles are differentially affected by variations in instructional strategies. Accordingly, the hypothesis was developed earlier that cognitive styles (in particular, field dependence-independence) would interact with instructional strategies in the acquisition of cognitive structures.

The results reported above failed to support this hypothesis. A variety of explanations are plausible. The derived individual measures of cognitive structure tended to focus on changes in structure and, accordingly, may not have been sensitive to acquisition of different cognitive structures. Since the card sorting data were the primary bases for considering the nature of the acquired cognitve structure and these data were not analyzed from a cognitive style perspective, additional analyses employing the cognitive style variable may be interesting.

Another possibility concerns the students involved in the study. The possibility exists that use of university students as subjects has resulted in a sample in which cognitive structure, cognitive style and instruction are not related in the same fashion as in the general population. Perhaps university students (being, in general, particularly successful learners) have developed learning strategies which have a more powerful effect on learning than does cognitive style. Research involving less adept learners may be enlightening.

Of course, the possibility also exists that cognitive style (or, particularly, field dependence-independence) is unrelated to cognitive structure learning. Certainly, the cognitive style literature is sufficiently ambiguous to warrent mention of this as a plausible explanation.

D. Cognitive Structure and Achievement

Previous research (eg: Shavelson & Geeslin, 1975; Fenker, 1975) failed to establish a relationship between measures of cognitive structures and traditional measures of achievement. Demonstration of such a relationship is essential for cognitive structure theory. Although the presently reported relationships between the derived measures of cognitive structure and the traditional achievement test were not large, considering the crudeness of the measures and uncertainty regarding the validity of the achievement test, the pattern of the relationship amongst the variables tends to support both the construct validity of the derived measures and the theoretical network which relates cognitive structure and achievement.

E. **Measurement of Cognitive Structure**

One of the major research questions concerned the validation of measures of cognitive structure. Evidence was provided which supported the validity of the measures employed. The derived learning measures distinguished among the groups as predicted; they also converged in the expected fashion.

A question remains, however. Are the measures of any practical significance? Certainly, they seemed to tap an aspect of learning (concept relations) not assessed by many traditional achievement tests. This feature may be important when measuring learning in some concept domains. Nevertheless, several factors militate against use of the derived measures in instructional settings:

1. The basic data collection instruments are crude, at best. Internal consistency on the similatity rating task was low. More appropriate reliability information is simply not available for either the similarity rating or card sorting procedures.
2. Derivation of the learning measures from the raw data is a complex tast.
3. The measures lack clear meaning.
4. Traditional achievement tests are not so seriously flawed as to warrant such a complex, potentially ambiguous approach.

The study's contribution to the field of measurement is somewhat greater when viewed from a research and development perspective. The appropriateness of the expert structure as a target for learning was supported. Maps and concept relationships identified by the LPA-MDS methods may prove valuable in both curriculum development and curriculum evaluation. Additional attention to the interpretation of INDSCAL concept spaces would also be fruitful. Use of the expert target procedure may be a reasonable approach for inclusion in large scale curriculum and instruction evaluation projects.

F. Summary

A study which investigated the construct validity of cognitive structure and its measurement as well as the relationship among cognitive structure, instructional strategies, cognitive style, and achievement has been reported. The findings generally support MDS methods for assessing cognitive structure. With the exception of the results involving the cognitive style variable, the study provided empirical support for a theoretical network developed from a cognitive structure model of learning.

REFERENCES

Anderson, J. R. and Bower, G. H. Human Associative Memory. Washington: V. H. Winston, 1973.

Ausubel, D. P. The use of advance organizers in the learning and retention of meaningful verbal material. Journal of Educational Psychology, 1960, 51, 267-272.

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

Ausubel, D. P. Educational Psychology: A Cognitive View. New York: Holt, Rinehart, and Winston, 1968.

Ausubel D. P. and Fitzgerald, D. Organizer, general background, and antecedent learning variables in sequential verbal learning. Journal of Educational Psychology, 1962, 53, 243-249.

Bruner, J. S. The Process of Education. New York: Random House, 1960.

Bruner, J. S. Toward a Theory of Instruction. Cambridge, Mass: Harvard University Press, 1966.

Bruner, J. S., Goodnow, J. J., and Austin, G. A. A Study of Thinking. New York: Wiley, 1956.

Campbell, D.T. and Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.

Carroll, J. D. and Chang, J.-J. An analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika, 1970, 35, 283-319.

Champagne, A. B. and Klopfer, L. E. Verification of a causal model for students' achievement in mechanics in a coollege physics course. Paper presented at NARST Annual Meeting, Boston, 1980.

Chomsky, N. Current Issues in Linguistic Theory. The Hague: Mouton, 1964.

Christiansen, H. B. Statistics Step by Step. Boston: Houghton Mifflin Company, 1977.

Cliff, N. and Young, F. W. On the relation between unidimensional judgements and multidimensional scaling. Organizational Behaviour and Human Performance, 1968, 269-285.

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

Cronbach, L. J. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684.

Cronbach, L. J. Beyond the two disciplines of scientific psychology. American Psychologist, 1975, 30, 116-127.

Cronbach, L. J. and Snow, R. E. Aptitudes and Instructional Method. New York: Irvington, 1975.

Deese, J. The Structures of Associations in Language and Thought. Baltimore: Johns Hopkins Press, 1975.

Dewey, J. Method in science teaching. Science Education, 1916, 1, 3-9.

Fenker, R. M. The organization of conceptual materials: a methodology for measuring ideal and actual cognitive structures. Instructional Science, 1975, 4, 33-57.

Fenker, R. M. and Tees, S. Measuring the cognitive structures of pre-school children: a multidimensional scaling analysis of classification performance and similarity estimation. Multivariate Behavioural Research, 1976, 11, 339-352.

Field, T. W. and Cropley, A. J. Cognitive style and science achievement. Journal of Research in Science Teaching, 1969, 6, 2-10.

Fleming, M. L., Knowlton, J. Q., Blain, B. B., Levie, W. H., and Elerian, A. Message and design: the temporal dimension of message structure. Final Report. Bloomington, Indiana: Indiana University, 1968.

Frijda, N. H. Simulation of human long-term memory. Psychological Bulletin, 1972, 77, 1-31.

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

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

Gagne, R. M. Learning hierachies. Educational Psychologist, 1968, 6, 1-9.

Gagne, R. M. The Conditions of Learning, 2nd ed. New York:

Holt , Rinehart and Winston, 1970.

Gagne, R. M. The learning basis of teaching methods . In N. L. Gage (Ed.), The Psychology of teaching methods. Seventyfifth Yearbook, National Society or the Study of Education, Part I. Chicago: Chicago University Press, 1976.

Gagne, R. M. The Conditions of Learning, 3rd ed. New York: Holt, Rinehart and Winston, 1977.

Gagne, R. M. and Briggs, L. J. Principles of Instructional Design. New York: Holt, Rinehart and Winston, 1974.

Gagne, R. M. and White, R. T. Memory structures and learning outcomes. Review of Educational Research , 1978, 48, 187-222.

Gardner, R. W. Reliability of group test scores for cognitive controls and intellectual abilities over a one-year period. Perceptual and Motor Skills, 1973, 36, 753-754.

Garskoff, B. E. and Huston, J. P. Measurement of verbal relatedness: an idiographic approach. Psychological Review, 1963, 70, 277-288.

Geeslin, W. E. and Shavelson, R. J. An explatory analysis of the representation of a mathematical structure in students' cognitive structures. American Educational Research Journal, 1975, 12, 21-39.

Green, P. E. and Rao, V. R. Applied Multidimensional Scaling: A Comparison of Approaches and Algorithms. Hinsdale, Illinois: Dryden, 1972.

Hammond, A. Assessment Contexts in Relation to Physics Achievement and Cognitive Style. Unpublished Ph. D. dissertation, University of Alberta, 1976.

Hanson, N. R. Patterns of Discovery. Cambridge: University of Cambridge Press, 1958.

Haugen, E. Bilingualism in the Americas: A Bibliography and Research Guide. University of Alabama Press, 1956.

Holman, E. W. The relation between hierachical and Euclidian models for psychological distances. Psychometrika, 1972, 37, 417-423.

Holzman, P. S. Scanning: a principle of reality contact. Perceptual and Motor Skills, 1966, 23, 835-844.

Hunt, D. E. A Conceptual Level matching model for co-ordinating learner characteristics with educational approaches. *Interchange*, 1970, *1*, 68-82.

Johnson, D. M. *A Systematic Introduction to the Psychology of Thinking*. New York: Harper and Row, 1972.

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

Johnson, S. C. Hierarchichal clustering schemes. *Psychometrika*, 1967, *32*, 241-254.

Joyce, B. and Weil, M. *Models of Teaching*. Englewood Cliffs, N. J.: Prentice-Hall, 1972.

Kerlinger, F. N. and Pedhazur, E. J. *Multiple Regression in Behavioural Research*. New York: Holt, Rinehart and Winston, 1973.

Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of Memory*. New York: Academic Press, 1970.

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

Klahr, D. A. A Monte Carlo investigation of the statistical significance of Kruskal's non-metric scaling procedure. *Psychometrika*, 1969, *34*, 319-330.

Koran, M. L., Snow, R. E., and McDonald, F. J. Teacher aptitude and observational learning of a teaching skill. *Journal of Educational Psychology*, 1971, *62*, 219-228.

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

Kruskal, J. B. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 1964b, *29*, 115-129.

Kruskal, J. B. and Carmone, F. *How to use M-D-Scal (Version 5-M) and other useful information*. Bell Telephone Laboratories, Murray Hill, N. J., 1971.

LaPorte, R. E. and Voss, J. F. Prose representation: a multidimensional scaling approach. *Multivariate Behavioural Research*. 1979, *14*, 39-56.

Lewin, K. *Dynamic Theory of Personality*. Trans. by D. K. Adams and K. E. Zener. New York: McGraw-Hill, 1935.

Lingoes, J. C. and Schonemann, P. H. Alternate measures of fit for the Schonemann-Carroll matrix fitting algorithm. Psychometrika, 1974, 39, 423-427.

Lohrenz, L. J. and Gardner, R. W. Cognitive control in recall of similar visual designs vs nonsimilar thematic materials. Perceptual and Motor Skills, 1973, 36, 627-631.

MacCorquodale, K. and Meehl, P. E. On a distinction between hypothetical constructs and intervening variables. Psychological Review, 1948, 55, 95-107.

MacDonald-Ross, M. The problem of representing knowledge. Paper presented at the Structural Learning Conference, Philadelphia, 1972.

Mager, R. F. Preparing Instructional Objectives. Belmont, Calif.: Fearon, 1962.

Mager, R. F. Measuring Instructional Intent: or Got a Match? Belmont: Calif., 1973.

Mednick, S. A. The associative basis of the creative process. Psychological Review, 1962, 69, 220-232.

Nagy, G. P. Construct Validity of Cognitive Structures: A Multidimensional Analysis. Unpublished Ph. D. dissertation, University of Alberta, 1977.

Nebelkopf, E. B. and Dreyer, A. S. Continuous-discontinuous concept attainment as a function of individual differences in cognitive style. Perceptual and Motor Skills, 1973, 36, 665-662.

Newell, A. and Simon, H. A. Human Problem Solving. Englewood Cliffs, N. J.: Prentice-Hall, 1972.

Piaget, J. Piaget's theory. In P. H. Mussen (Ed.), Carmichael's Manual of Child Psychology. New York: Wiley, 1970.

Piaget, J. and Inhelder, B. The Early Growth of Logic in the Child. New York: Harper and Row, 1964.

Preece, P. F. W. Science concepts in semantic space - a multidimensional scaling study. Alberta Journal of Educational Research, 1976a, 22, 281-286.

Preece, P. F. W. A note on the comparison of cognitive structure and subject matter structure. Journal of Research in Science Teaching, 1976b, 13, 351-352

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

Rappoport, A. and Fillenbaum, S. An experimental study of semantic structures. In R. Shepard, A. K. Romney, and S. B. Nerlove (Eds.), Multidimensional Scaling: Volume II Applications. New York: Seminar Press, 1972.

Rosler, F. Identifying individual judgement differences: INDSCAL or three-mode factor analysis? Multivariate Behavioural Research, 1979, 14, 145-167.

Rothkopf, Z. E. The concept of mathemagenic activities. Review of Educational Research, 1970, 40, 325-336.

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

Ryle, G. The Concept of Mind. London: Hutchinson, 1949.

Schonemann, P. H. and Carroll, R. M. Fitting one matrix to another under choice of a central dilation and a rigid motion. Psychometrika, 1970, 35, 245-255.

Schwen, T.M. The Effect of Cognitive Styles and Instructional Sequences on Learning a Hierarchical Task. Unpublished Ph. D. dissertation, Indiana University, 1970.

Shavelson, R. J. Some aspects between 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 representation of a subject matter structure in a student's memory. Journal of Research in Science Teaching, 1974, 11, 231-250.

Shavelson, R. J. and Stanton, G. C. Construct validation: methodology and application to three measures of cognitive structure. Journal of Educational Measurement, 1975, 12, 67-86.

Shepard, R. N. Psychological representation of speech sounds. In E. E. David and P. B. Denes (Eds.), Human Communication: A Unified View. New York: McGraw-Hill, 1972.

Shepard, R. N. Representation of structure in similarity data: problems and prospects. Psychometrika, 1974, 39, 373-421.

Stasz, C. Field Independence and the Structuring of Knowledge in a Social Studies Minicourse. Unpublished master's thesis, Rutgers University, 1974.

Subkoviak, M.J. The use of multidimensional scaling in educational research. Review oif Educational Research, 1975, 45, 387-423.

Tolman, E. C. Purposive Behaviours in Animals and Men. New York: D. Appleton-Century, 1932.

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

Tucker, L. R. and Messick, S. An individual differences model for multidimensional scaling. Psychometrika, 1963, 28, 333-367.

Tzeng, O. C. S. and Landis, D. Three mode multidimensional scaling with points of view solutions. Multivariate Behavioural Research, 1978, 13, 181-213.

Weil, M., and Joyce, B. Information Processing Models of Teaching: Expanding Your Teaching Repertoire. Englewood Cliffs, N. J.: Prentice-Hall, 1978.

Wickelgren, W. A. Learning and Memory. Englewood Cliffs, N. J.: Prentice-Hall, 1977.

Wight, A. R. Beyond behavioural objectives. Educational Technology, July 1972, 9-14.

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

Witkin, H. A. Individual differences in ease of perception of embedded figures. Journal of Personality, 1950, 19, 1-15.

Witkin, H. A., Moore, C. A., Goodenough, D. R., and Cox, P. W. Field-dependent and field-independent cognitive styles and their educational implications. Review of Educational Research, 1977, 47, 1-64.

Witkin, H. A., Ottman, P. K., Raskin, E., and Karp, S. A. A Manual for the Embedded Figures Tests. Palo Alto: Consulting Psychologists Press, 1971.

Woodworth, R. S. Experimental Psychology. New York: Holt, 1938.

Young, F. W. and Torgerson, W. S. TORSCA, a FORTRAN IV program for Shepard-Kruskal multidimensional scaling analysis. Behavioural Science, 1967, 12, 498.

APPENDIX A

SAMPLE OF INSTRUCTIONAL PACKAGE:

COGNITIVE APPROACH

MEASURES OF CENTRAL TENDENCY:

MEAN, MEDIAN, MODE

Objective: Given a set of measurements, compute the mean, median and mode.

Introduction:

When you receive a score on a class test you may find it difficult to interpret how well you have done unless you know something about the scores of the other students in the class. If you are given some idea of the average or typical score on the test, you are in a better position to interpret your own performance. You have probably used the notion of average or typical score in other situations, too, perhaps to describe your general level of performance at university, the income of teachers, or the daily temperature during March. In such situations you may have calculated the arithmetic average--like you were taught in elementary school--by adding up all the scores and dividing the total by the number of scores you had.

There are additional methods for identifying the typical value for a set of scores. The methods are similar in that they all summarize a set of scores by determining a value which is representative of all the scores. The values which result are called measures of central tendency; they are used to describe one of the basic features of a set of scores.

In this lesson we will define three measures of central tendency: the mean, the median, and the mode. Methods for computing each of the measures will be presented, as will guidlines for selecting and interpreting each of the measures.

Definintions:

measure of central tendency: A measure of central tendency is a number that represents the central or most representative measurement in a set.

mean: The mean is the arithmetic average of a set of measurements. We obtain the mean by dividing the sum of the measurements by the number of measurements in the set. The formula for computing the mean is

$$\bar{X} = \frac{\sum X_i}{N} \longleftarrow \text{sum of all the scores}$$

$\sum$ means 'sum'

where

X_i is the score of a particular individual i

N is the total number of individuals in the set

$\bar{X}$ is the mean of the set of measurements

median: The median (denoted by Md) is the middle number in an ordered set of scores. If there is an odd number of scores in a set, there is one and only one middle number, the median. If there is an even number of scores in the set, then there are two middle numbers. By convetion the median falls halfway between them. To compute the median, order all the scores and find the middle one. If there are an even number of scores, calculate the average of the two middle scores.

mode: The mode is the score which occurs most frequently in a set of scores. It is possible for a set of scores to have more than one mode. To determine the mode find the frequency of each score; the mode(s) will be the score(s) with the greatest frequency.

Discussion:

Measures of central tendency (i.e., the mean, the median and the mode) are different kinds of averages that can serve as numerical summaries of a set of measurements. All three measures define the centre of the set.

MEAN. The mean of a set of measurements represents the physical centre of the set (like the centre of gravity). If the scores represent units of weight placed on a real line, the mean represents the fulcrum or point of balance. Suppose we use the formula given above to find the mean of the set of numbers 1, 1, 2, 3, 4, 4:

$$\bar{X} = \frac{1+1+2+3+4+4}{6} = \frac{15}{6} = 2.5$$

The number 2.5 is the point of balance, or mean, for these values. The scores are pictured in Figure 1.

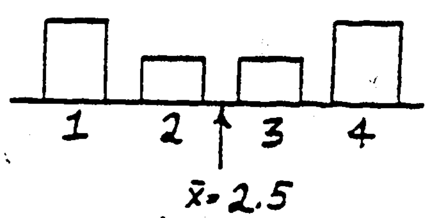


Figure 1

Let's calculate the mean for the set of measurements 1, 3, 4, 7, 8, 9, 9.

$$\bar{X} = \frac{1+3+4+7+8+9+9}{7} = \frac{41}{7} = 5.857$$

The point of balance has been located on the real line shown in Figure 2.

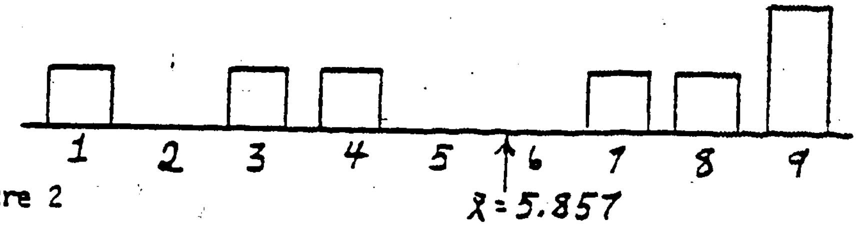


Figure 2

Now, if we refer to Figures 1 and 2, we see that if any of the weights were shifted, the mean, or point of balance, would also shift. Thus we can see that the mean is sensitive to the magnitude of the numbers in a given set. This sensitivity can occasionally be a drawback to the use of the mean as a measure of central tendency. For example, we calculate the mean for the set of numbers 1, 2, 3, 4, 20 as follows:

$$\bar{X} = \frac{1 + 2 + 3 + 4 + 20}{5} = \frac{30}{5} = 6$$

Figure 3 shows the point of balance for this set of scores.

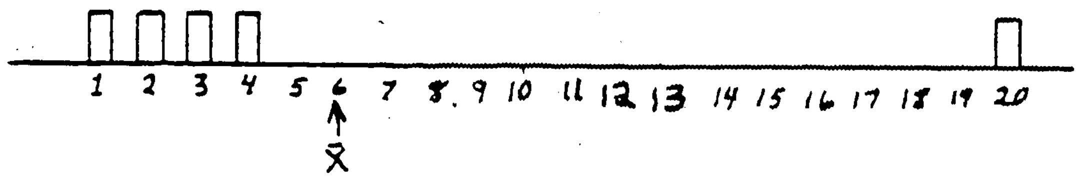


Figure 3

In Figure 3, note that the mean is nowhere near the centre of most of the numbers, but has shifted to the right in order to balance the extreme value 20.

When there are extreme numbers at one end of a set, the mean is quite sensitive to them, and shifts so that it may not be very representative of the majority of the measurements. In such cases it is useful to consider the other measures of central tendency.

Recall Stephens scales of measure. Notice that when we calculate the mean that we are actually adding and dividing distances. We deal with the scores as if their magnitudes had at least interval qualities. Strictly speaking, then, the mean can only be calculated for scores based on interval or ratio scales; although you have no doubt seen means calculated from ordinal data.

MEDIAN Let us use examples to illustrate how to find the median. Consider the numbers 4, 1, 6, 20, 3. By ordering them we get 1, 3, 4, 6, 20. The middle number, or median, for this set is 4 (N = 5 is odd).

1, 3, 4, 6, 20
*
Md

Note that the unusual size of the extreme value of 20 in this set of numbers hasn't affected the location of the median value.

Now suppose we are given the measurements 6,2,17,3,4,10,11,11. Ordered they are 2, 3, 4, 6, 10, 11, 11, 17. The median is 8 which is halfway between the two middle numbers, 6 and 10 (N = 8 is even).

2, 3, 4, 6, 10, 11, 11, 17
8
*
Md

Again consider the measurements 11 20 32 1 10 11 7 12. Ordered they are 1 7 10 11 11 12 20 32. The median is 11 since the two middle numbers are 11.

1 7 10 11 11 12 20 32
*
Md

To compute the median we do the following:

1. Order the scores in size from smallest to largest.
2. Determine whether the total number of scores is odd or even.
3. If the number of scores, N, is odd, find the middle number; this is the median. If N is even, find the two middle numbers, then compute the mean of these two numbers to find the median.

When we compute the median, notice that we make use of the relative position (order) of scores but do not use the magnitude of the score. A score which is above the median has the same influence if it is 100 units above the median as if it was 1 unit above the median. Accordingly, we require only that the measuring scale permits ordering of scores and can therefore calculate the median for ordinal, interval or ratio scales.

Mode If, in a set of scores, one particular value occurs more frequently than any other, this value is called the mode. If two values have the samefrequency, or approximately the same frequency, the set is said to be bimodal. If three values have the same frequency, trimodal is the term (and so forth).

For instance, suppose we have the set of scores 1 3 4 4 7 3 3 4 2 7 8 3 1. By organizing these data into a table we get Figure 4.

	measurement	frequency
	1	2
	2	1
mode --->	3	4
	4	3
	7	2
	8	1

Figure 4

In Figure 4 the mode is 3 since the number 3 occurs more frequently than any other value in the set.

Consider this example: by putting the set 22 26 27 27 23 23 27 22 28 22 28 30 29 25 ito a table we get Figure 5.

		measurement	frequency
bi modal	mode -->	22	3
		23	2
		25	1
		26	1
	mode -->	27	3
		28	2
		29	1
		30	1

Figure 5

It is easy to envision examples in which the mode is not at the centre of the data, but is located at one end or the other. For example, in Figure 4, if the 27 occurred only once, then the mode would be 22. When this occurs, it doesn't seem appropriate to call the mode thecentre of the scores. However, it can still serve meaningfully as a representative number.

When computing the mode, the only information we use is the catego (or value) of a particular score. We require neither the qualities of order nor distance. Accordingly, the mode can be determined for data arising from any of the four types of measuring scales.

Examples:

In each of the following exercises compute the mean, the median, and the mode.

1. Calculate the average weight (in pounds) of the 1977 Intramural Flag Football champions. The weights of the players are as follows:

   125 303 163 175 181 190 215 178 163 178 186 171 178

   $X_i = 2406$

   SOLUTION: The ordered weights are: 125 163 163 171 175 178 178 178 181 186 190 215 303. $N = 13$ (odd). From this ordering we see that the median is 178 pounds since 178 is the middle number. The mode is 178 pounds since 178 is the number that occurs most frequently. The mean is $\frac{2406}{13} = 185.1$ pounds.

   Mo = 178 pounds
   Md = 178 pounds
   $\overline{X}$ = 185.1 pounds

2. The following set of numbers indicates the length of time (in minutes) that 12 teachers spent on coffee breaks during one afternoon.

   15 47 53 23 17 32 14 27 34 19 27 26 $\Sigma X_i = 334$

   What is the average length of a coffee break?

   SOLUTION: Ordered the coffee breaks are: 14 15 17 19 23 26 27 27 32 34 47 53. ($N = 12$, even). The two middle numbers are 26 and 27. The median, which falls halfway between the two middle numbers in an even-numbered set, is 26.5. The mode is 27. The mean is $334/12 = 27.8$.

   Mo = 27 minutes
   Md = 26.5 minutes
   $\overline{X}$ = 27.8 minutes

3. Suppose that a teacher records the number of verbal outbursts that a particular student makes on a series of consecutive days. The number of

outbursts per day is: 4 5 3 5 7 3 4 3 2 3 3 5 1 1 3 5 6 2 5.

SOLUTION:

outbursts per day	frequency
1	2
2	2
3	6
4	2
5	5
6	1
7	1

The median for the data is 3 (N=19, odd). The mode is also 3, although 5 is a close runner-up. We might call this set of numbers bimodal since two distinct values stand out above the others in frequency. The mea s 70/19 = 3.68.

Mo = 3 outbursts per day
Md = 3 outbursts per day
$\overline{X}$ = 3.68 outbursts per day

Practice problems:

Compute the mean, the median or the mode for each of the following problems.

1. The following set of numbers gives the time (in minutes) that it takes each of nine students to complete an arithmetic problem. What is the mode? The median? The mean? Does the mean have a practical interpretation?

   1 3 2 2 3 4 2 1 3

2. Suppose that the following numbers represent the yearly incomes of seven people in a neighborhood. Compare the median and the mean income. Which is the more representative number?

   7,500 12,500 4,500 15,000 9,000 30,000 8,500

3. The following data represent the number of children in each of ten families that live within a school district. Which is most representative, the mean, the median, or the mode?

   2 8 1 1 4 5 2 3 0 1

4. The following data represent the results of a final exam administered to eight students in a testing course. Calculate the mean and median of these scores.

   87 79 93 69 75 91 84 77

5. Over a period of five days in February the number of students in attendance at a large junior high school was:

   630 638 635 640 645

   What was the mean and median attendance?

6. When checking the cummulative record file, a teacher discovered that the students in her class had the following I.Q.'s. What was the mean I.Q.? The median? The mode? Compare the three measures. Why are they different? Which single measure best represents the I.Q. scores?

student	IQ	student	IQ
Fred	83	Jane	74
Susan	185	Beverly	80
John	78	Mike	83
Linda	74	Dan	77
Dave	80	Judy	185
Frank	82	Christine	74
Janice	77	Ralph	76
George	74	Harold	78
Nancy	81	Calvin	80
Sharon	78	Herb	81

7. If 7 was subtracted from each of the scores in Question 4, what would be the new values for the mean and the median?

8. If each of the scores in Question 4 was multiplied by 3, what would be the new values for the mean and the median? What is the relationship between the new and original values for the mean? The median?

Comparing the mean, the median and the mode.

The mean may be regarded as an appropriate measure of central tendency for interval and ratio variables. The magnitude of each of the scores is incorporated in its calculation. The median is an ordinal statistic. Its calculation is based on the ordinal properties of the measurements. If the observations are arranged in order, the median is the middle value. Its calculation does not include the particular magnitudes of the scores, but merely the fact of their occurrence above or below the middle value. Thus the sets of numbers 5 7 20 30 90 and 18 19 20 21 22 have the same median although their means may be quite different. The mode, the score value with the greatest frequency, is a nominal statistic. Its calculation does not depend on either the magnitude or the order of the scores, but merely on their frequency of occurrence.

A comparison of the mean, median and mode may be made when all three have been calculated for the same set of scores. If the set of scores is presented graphically, the mean is the point on the score axis which corresponds to the centre of gravity of the set. The median is the point on the score axis above which (and below which) half the scores fall. The mode will be the point on the score axis which corresponds to the highest point on the curve.

If the graph of the scores is symmetrical about the mode, the mean, the median and the mode will coincide. IF the graph is shifted ('skewed') the measures will not coinc- (See Figures 7, 8, 9 and 10).

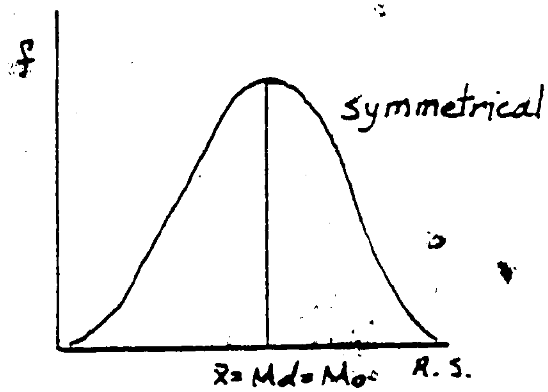


Figure 7

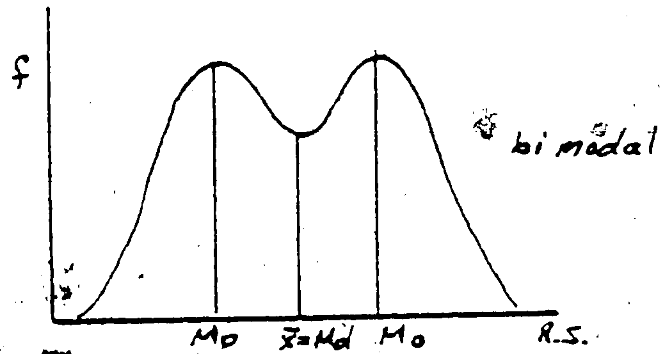


Figure 8

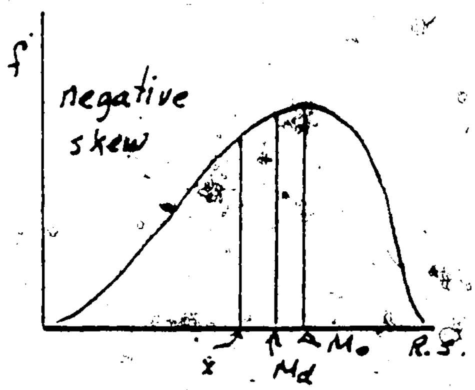


Figure 9

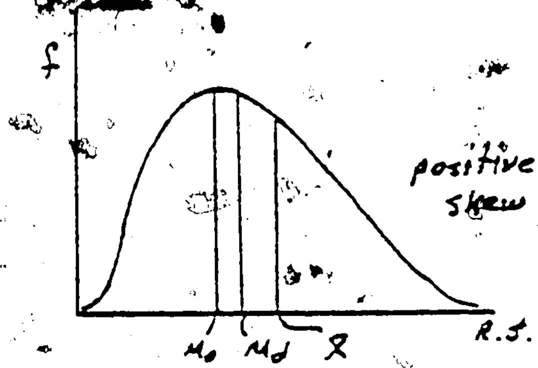


Figure 10

A question may be raised regarding the appropriate choice of a measure of central tendency. In practical situations this question is rarely in doubt. The mean is usually preferred to either the median or the mode. It is rigorously defined and easily calculated.

The median, however, may be preferred in some situations. Scores may occur which are atypical of the other scores in the set. Such scores may greatly affect the mean. Consider the scores 2 3 4 4 7 8 10 11 86. The score 86 is quite atypical of the remaining scores and its presence greatly affects the value of the mean. The mean is 15, a value greater than all but one of the scores. The median is 7. Under such circumstances the median may be preferred to the mean.

For strictly nominal measurements the mode is the only measure of central tendency that can be used. It is rarely used with interval, ratio, or ordinal variables where means and medians can be calculated.

Sometimes, the very nature of the measurements and the use which they serve will influence our choice of the mean, median or mode. For instance, if you were the manager of a shoe store and were interested in the sizes most often purchased, the mean and the median could yield useless fractions; the most useful measure of central tendency would likely be the mode.

We may also have measurements which cannot be readily summarized by a measure of central tendency. Calculating the mean (or the median or the mode) of a beauty contestant's measurements would be of little help in describing the young lady.

MEASURES OF RELATIONSHIP:

CORRELATION

Objective:

Expla the way in which correlation can be used to estimate the relationship between two sets of scores.
Interpret correlation coefficients of varying magnitudes.

Definition:

correlation coefficient: A correlation coefficient is a measure of the relationship (literally, the co-relation) between two sets of scores. In testing, the correlation will generally be between two sets of scores collected on the same persons--for example, between high school grades and university grades, or between scores obtained on two administrations of the same test. A correlation coefficient of +1 indicates a perfect relationship between two sets of scores; a coefficient of 0 indicates no systematic relationship between the two sets of scores; while a coefficient of -1 indicates a perfect inverse relationship between two sets of scores.

$$r = \frac{\sum Z_X \cdot Z_Y}{N} = \frac{\Sigma(X - \overline{X}) \cdot (Y - \overline{Y})}{N\ S_X \cdot S_Y}$$

where:

r is the correlation coefficient between X and Y

Z_X are the standard scores for all the people on test X

Z_Y are the standard scores for all the people on test Y

N is the number of people

Discussion:

If we have two sets of scores from the same group of people, it is often desireable to know the degree to which scores are related. For example, we might be interested in the relationship between science test scores and students' average marks. (Do people who do well in science also tend to do well in other subject areas?) We might also be interested in the relationship between IQ and grades, or IQ and financial success. We could also be interested in the reliability of a test: that is, whether our test is consistent in its measures of student achievement. We could also be interested in determining if high school grades were valid for predicting university success.
Although such questions might never be answered with absolute certainty, we can shed light on the problems by examining the association between the sets of scores.

In particular, we would like to know if there is a systematic relationship between scores on one of the measures with scores on the other measure. That is, are high scores on one measure accompanied by high scores on the other measure? Or, more precisely, is the ordering of the individuals by the first measure related to the ordering of the individuals by the second measure?

Examine the following table. Notice the association between study time and final grade. Also notice the relationship between grades and height.

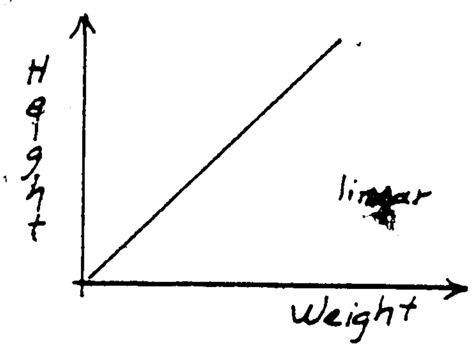
Student	Hours of Study / Wk	Final Grade Econ 100	Height (in.)
Paul	12	76	72
Fred	8	62	70
Alice	15	78	64
June	18	85	66
Andy	22	96	74
Peter	19	90	68

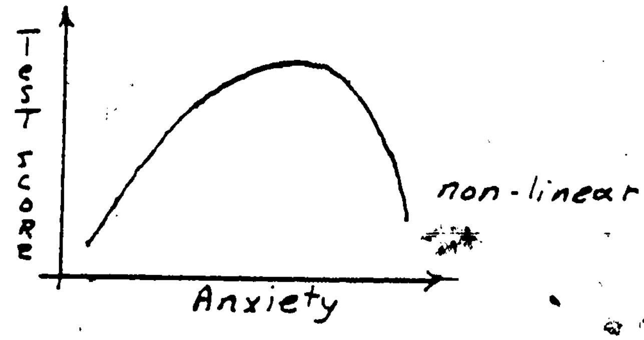
Glancing at the table, we see that students who are above the mean on hours of study also tend to be above the mean on final grade. And students whose hours of study are low, also tend to have low final grades. When considering height and final grades, however, there does not seem to be any tendency for high values on grades to be associated with any particular height.

Suppose we have another student, Barry, but for him we only have his final grade in Econ 100. Could we make any predictions concerning his study habits? If his final grade was 92, is he more likely to study 20 hrs. per week or 10 hrs. per week? Our table suggests that high grades and high hours of study tend to go together, so we can predict that Barry is likely to study quite a lot, perhaps 20 hours per week. On the other hand, knowing Barry's final grade does little to help us estimate his height; there doesn't seem to be any systematic pattern between grades and height according to the table.

A correlation coefficient is a means of summarizing the relationship between two sets of data. Its magnitude indicates the certainty with which we can predict a person's score on one variable if we know his score on another. If people who are far from the mean on one measure (X) are also far from the mean on the other measure (Y), and people who are near the mean on X are also near the mean on Y, then the magnitude of the correlation coefficient will be near 1. If there is no tendency to be far from the mean (or near the mean) on both, the correlation coefficient will be near 0. The sign of the coefficient indicates whether people tend to be the same direction from the mean on both variables—resulting in a positive r — or in opposite directions from the mean on the two measures—resulting in a negative r.

Of particular interest to a teacher building a test might be the relationship between how well students perform on the first half of the test compared to how they do on the second half. Or how they do on the odd numbered questions compared to how they do on the even numbered questions. Or how they perform on the test today, compared with how they performed on the test last month. If the first half of the test ordered the students in the same sequence as the second half of the test, we would have evidence for the test's reliability (consistency). Similarly, if the students were ordered the same way tested today as they were by the same test last month, we would have evidence for the tests reliability.





The logic of the correlation coefficient can best be seen if scores are expressed as standard scores (Z-scores). If the two variables to be correlated are designated X and Y, and their corresponding standard scores as Z_x and Z_y, then the correlation between X and Y will be:

$$r = \frac{\sum Z_x \cdot Z_y}{N}$$

where N is the number of pairs of scores (each pair of scores will be obtained from a different individual). You can see that if corresponding (from the same person) standard scores are both large and have the same sign, the value of r will be large and positive; if both are large but have opposite signs, the value of r will be large but negative; if both are small or some pairs have negative signs and others positive signs, the value of r will be near zero.

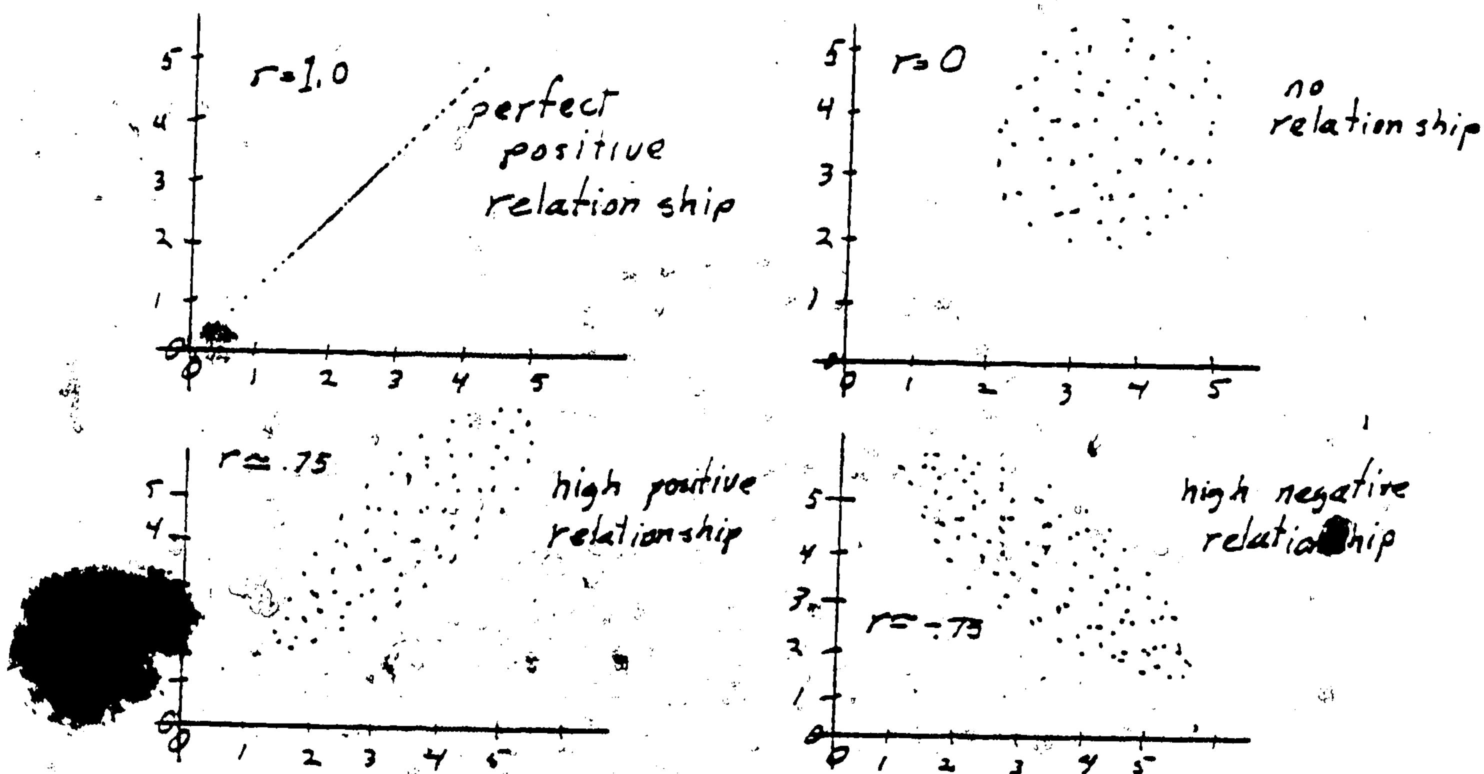
In the second formula given under the definition, the method for converting raw scores to Z-scores has simply been substituted for the Z-scores.

$$Z_x = \frac{X - \overline{X}}{S_x} \qquad Z_y = \frac{Y - \overline{Y}}{S_y}$$

The value of r can range from +1.00 to -1.00. The absolute value of the coefficient tells the strength of the relationship; the greater the absolute value, the greater the correspondence between the two sets of scores. Thus when $r = 1.00$, the scores on Y are completely predictable knowing the scores on X. If we knew what a person scored on X, we could predict perfectly what his Y score would be. If $r = 0.00$, the relationship between the pairs of scores is random; knowledge of one score does not help us to predict the other. The sign of the correlation coefficient tells us the direction of the relationship; thus, the coefficients of +.70 and -.70 represent equally strong relationships, albeit relationships in the opposite direction.

The relationship summarized by the correlation coefficient can be shown by the scatterplot of the test scores. If there is high correlation, the scores will cluster around a straight line. The orientation of the line with respect to the axes will indicate if the coefficient is positive or negative. (See Figure 2).

It cannot be overemphasized that r indicates only the degree of relationship between two variables; it does not indicate causation. If variables X and Y are highly correlated, there are at least three possible explanations: (1) X causes Y and thus changes in X result in changes in Y; (2) Y causes X and thus changes in Y result in changes in X; or (3) X and Y are both influenced by something else. In other words, a correlation coefficient is not a sufficient basis for inferring causation.



Each dot indicates 1 person.

Figure 2

Example:

Suppose that two arithmetic tests were administered to a class of students. The class mean on Test A was 20 and its standard deviation was 2. The class mean on Test B was 25 and its standard deviation was 4. What is the correlation between Test A and Test B based on the following 5 students? (note: these are only five of an entire class of 30)

Student	Test A	Test B	Z_A	Z_B	$Z_A Z_B$
Alice	22	29	1	1	1
Beth	23	27	1.5	.5	.75
Carol	21	23	.5	-.5	-.25
Debby	20	27	0	.5	0
Ellen	19	25	-.5	0	0
					$\Sigma Z_A Z_B$ = 1.5

$$\sum Z_A Z_B = 1.5$$

$$r = \frac{\sum Z_A Z_B}{N} = \frac{1.5}{5} = .3$$

Example:

Let's calculate the same coefficient using the other fourmula.

Student	Test A	Test B	$(A - \bar{A})$	$(B - \bar{B})$	$(A - \bar{A})\ (B - \bar{B})$
Alice	22	29	2	4	8
Beth	23	27	3	2	6
Carol	21	23	1	-2	-2
Debby	20	27	0	2	0
Ellen	19	25	-1	0	0
					12

$$r = \frac{\sum (A - \bar{A})(B - \bar{B})}{N\ S_A\ S_B} = \frac{12}{5 \times 2 \times 4}$$

$$= \frac{12}{40} = .30$$

Practice Problems:

1. What is the correlation coefficient between the following sets of scores?

   Social Studies: 11 [illegible] [illegible] 14 15
   Language: 58 57 [illegible] 62 61

2. When comparing scores on a mid term test with scores on the final test, a teacher calculated the following:

   $\sum Z_M\ Z_F = 22 \qquad N = 30$

   What is the correlation between mid-term and final test scores?

3. When checking the cummulative record file, a teacher sensed that her students' performances seemed closely related to their previous years' marks. She calculated the following. Were her suspicions supported?

   $\sum (X - \bar{X})\ (Y - \bar{Y}) = 12 \qquad N = 25$

4. The correlation between Spelling and Grammer was calculated to be .80. The sum of the product of the Z-scores ($\sum Z_S Z_G$) was 24. How many students were used in the computations?

SAMPLE OF INSTRUCTIONAL PACKAGE:

PERFORMANCE APPROACH

MEASURES OF CENTRAL TENDENCY

MEAN, MEDIAN & MODE

Objective: Given a set of measurements, compute the mean, median and mode.

Introduction:

We often need more concise ways of summarizing information than a simple listing of a set of scores. Certain summary statistics can be calculated in order to provide a general description of the entire set of scores. One of the characteristics which helps to describe the set is the general location of the measurements, that is, a score value which seems to represent all the scores. Statistics which fulfill this function--describe the general location of the scores--are called measures of central tendency. There are three commonly used measures of central tendency: the mean, the median, and the mode. All three of these measures convey information about the 'average scores' in a distribution, although each of the measures interprets 'average' in a slightly different way.

Definitions:

central tendency: A measure of central tendency is a number that represents the central or most representative measurement in a set.

mean: The mean is the arithmetic average of a set of measurements. It is the sum of all the measurements in a set divided by the number of measurements in the set. For example, the class mean on a literature test would be the sum of every student's score divided by the number of students. Symbol: $\bar{X}$

median: The median is the middle number in an ordered set of measurements. It is the point above (and below) which half the measurements occur. For example, on the same literature test, the mode would be the score which divided the class into the top and bottom halves. Symbol: Md

mode: The mode is the number that occurs most frequently in a set of scores. It is the most common score. A set of measurements may have more than one mode.

Computational procedures:

mean: $\bar{X} = \frac{\Sigma X_i}{N}$

where: $\bar{X}$ is the mean

X_i is the score of individual 1

N is the number of individuals

means - sum of

ΣX_i is the sum of all the scores example:

Student		Height
Mary	(X_1)	64 inches
John	(X_2)	70 inches
Paul	(X_3)	73 inches
Fred	(X_4)	72 inches
Judy	(X_5)	67 inches

$$\bar{X} = \frac{\Sigma X_i}{N}$$

= (Mary's ht. + John's ht. + Paul's ht. + Fred's ht. + Judy's ht.) ÷ 5

$$= (X_1 + X_2 + X_3 + X_4 + X_5) \div 5$$

$$= (64 + 70 + 73 + 72 + 67) \div 5$$

$$= 346 \div 5$$

= 69.2 inches

example:

The scores for 10 students on a spelling test were:

15 20 21 24 25 18 22 30 25 20

What was the mean test score?

$$\bar{X} = \frac{\Sigma X_i}{N} = \frac{15 + 20 + 21 + 24 + 25 + 18 + 22 + 30 + 25 + 20}{10}$$

$$= \frac{220}{10}$$

$$= 22$$

median: To compute the median

1. order the measurements from smallest to largest
2. if there is an odd number of measurements (if N is odd) select the middle score. It is the median.
3. if there is an even number of measurements, then there are two middle numbers. By convention, the median falls halfway between them.

example:

What is the median of the following scores?

6 2 17 3 4 10 11 11

1. Ordering the scores we have 2 3 4 6 10 11 11 17
2. N = 8 is even
3. The two middle numbers are 6 and 10. Halfway between them is

$$\frac{6 + 10}{2} = \frac{16}{2} = 8$$

The median is 8

example:

What is the median weight of the following football players?

weights: 125 303 163 175 181 190 215 178 163 178 186 171 178

1. Ordered weights: 125 163 163 171 175 178 178 178 181 186 190 215 303
2. N = 13 is odd. Middle score is 178.
3. The median is 178 pounds.

mode: To compute the mode

1. count the number of times each score-calue occurs
2. the mode is the value which occurred most frequently

example:

The following numbers represent the number of minutes that each teacher in a school spent on coffee breaks during an afternoon. What was the modal coffee break?

15 47 53 23 17 32 14 27 34 19 27 26

Coffee break (minutes)	Frequency
14	1
15	1
17	1
19	1
23	1
26	1
27	2
32	1
34	1
47	1
53	1

The mode is 27 minutes

example:

Suppose that a teacher recorded the number of times that a student swore over a series of days. The observations are recorded below.

Number of Swearings per day	Frequency
1	2
2	2
3	6
4	2
5	4
6	1
7	1

The mode is 3 swears per day although 5 swears per day is a close runner-up. Since two distinct values seem to stand out above the others, we might identify two modes and refer to the scores as being bimodal.

Exercises:

Calculate the mean, median and mode as required in the following exercises.

1. Compute the mean, median and mode for the following sets of scores:

   a. 22 26 27 27 23 23 27 22 28 22 28 30 29 25

   b. 10 3 3 9 7 4 6 9 6 7 6

   c. 2 12 8 2 5 8 6 7 6 5 6

   d. 12 19 24 24 42 29 36 18

   e. 2 19 24 24 56 39 36

2. Add 5 to each of the scores in 1b. What are the new means median and mode?

3. Each of the scores in 1c is multiplied by 3. What are the new mean, median and mode?

4. A teacher counted the number of times per day that students in her class went ot the drinking fountain. Calculate the mean, median and mode. What do you think is the most representative number? Why?

   5 1 4 8 3 5 2

   6 4 2 8 1 5

   7 3 3 7 6 8

5. The following are scores on a final exam administered to eight students in a test construction course. Calculate the mean and the median.

   87 79 93 69 75 91 84 77

6. The following data represent the number of children in each of ten families. Calculate the mean, median and the mode. Which is the most representative number?

   2 8 1 1 4 5 2 3 0 1

7. A typist recorded the following number of errors per page. Calculate the mean and median number of errors. Which is most representative?

   1 3 20 2 4

8. Suppose that the following numbers represent the yearly incomes of seven people living in your neighbourhood. Compare the median and the mean incomes. What is the more representative number?

   7,500 12,500 4,500 15,000 9,000 30,000 8,500

MEASURES OF RELATIONSHIP

CORRELATION

Objective: Explain the way in which correlation can be used to estimate the relationship between two sets of scores.
Interpret correlations coefficients of varying magnitudes.

Definitions:

correlation coefficient: A correlation coefficient is a measure of the relationship (literally, the co-relation) between two sets of scores. In testing, the correlation will generally be between two sets of scores collected on the same persons--for example, between high school grades and university grades, or between scores obtained on two administrations of the same test. A correlation coefficient of +1 indicates a perfect relationship between the two sets of scores; a coefficient of 0 indicates no systematic relationship between the two sets of scores; while a coefficient of -1 indicates a perfect inverse relationship between the two sets of scores.

$$r = \frac{\Sigma Z_x \cdot Z_y}{N} = \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{N S_x \cdot S_y}$$

where:

r is the correlation coefficient between X and Y
Z_x are the standard scores for all people on test X
Z_y are the standard scores for all people on test Y
N is the number of people

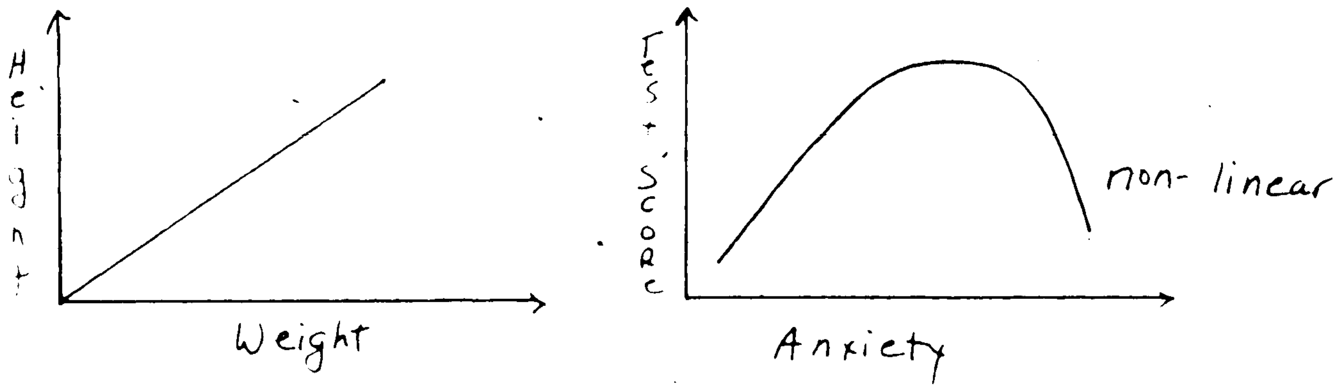
Discussion:

We are often interested in the relationship between two sets of scores. Questions often arise such as:

Is there a tendency for people who are short to also be light?
Do students who do well in math also tend to do well in science?
Do bright students tend to require less study time than dull students?
Do students who do well on the odd-numbered items on a test also do well on the even-numbered items?

Although such questions might never be answered with absolute certainty, we can shed light on the problems by examining the association between the sets of scores. In particular, we would like to determine if there is a systematic relationship between the two measures, that is, if knowing a person's score on one of the measures, we can predict his score on the other measure.

There are several different measures of correlation. The one most frequently used in educational testing is the Pearson Product-Moment correlation degree to which the association between two sets of scores can be represented by a strait line. (See Figures 1 and 2).



(Figure 1)

The logic of the correlation coefficient can best be seen if scores are expressed as standard scores (Z-scores). If the two variables to be correlated are designated X and Y, and their corresponding standard scores as Z_x and Z_y, then the correlation between X and Y will be:

$$r = \frac{\Sigma Z_x \cdot Z_y}{N}$$

where N is the number of pairs of scores (each pair of scores will be obtained from a different individual). You can see that if corresponding (from the same person) standard scores are both large and have the same sign, the value of r will be large and positive; if both are large but have opposite signs, the value of r will be large but negative; if both are small or some pairs have negative signs and others positive signs, the value of r will be near zero.

In the second formula given under the definition, the method for converting raw scores to Z-scores has simply been substituted for the Z-scores.

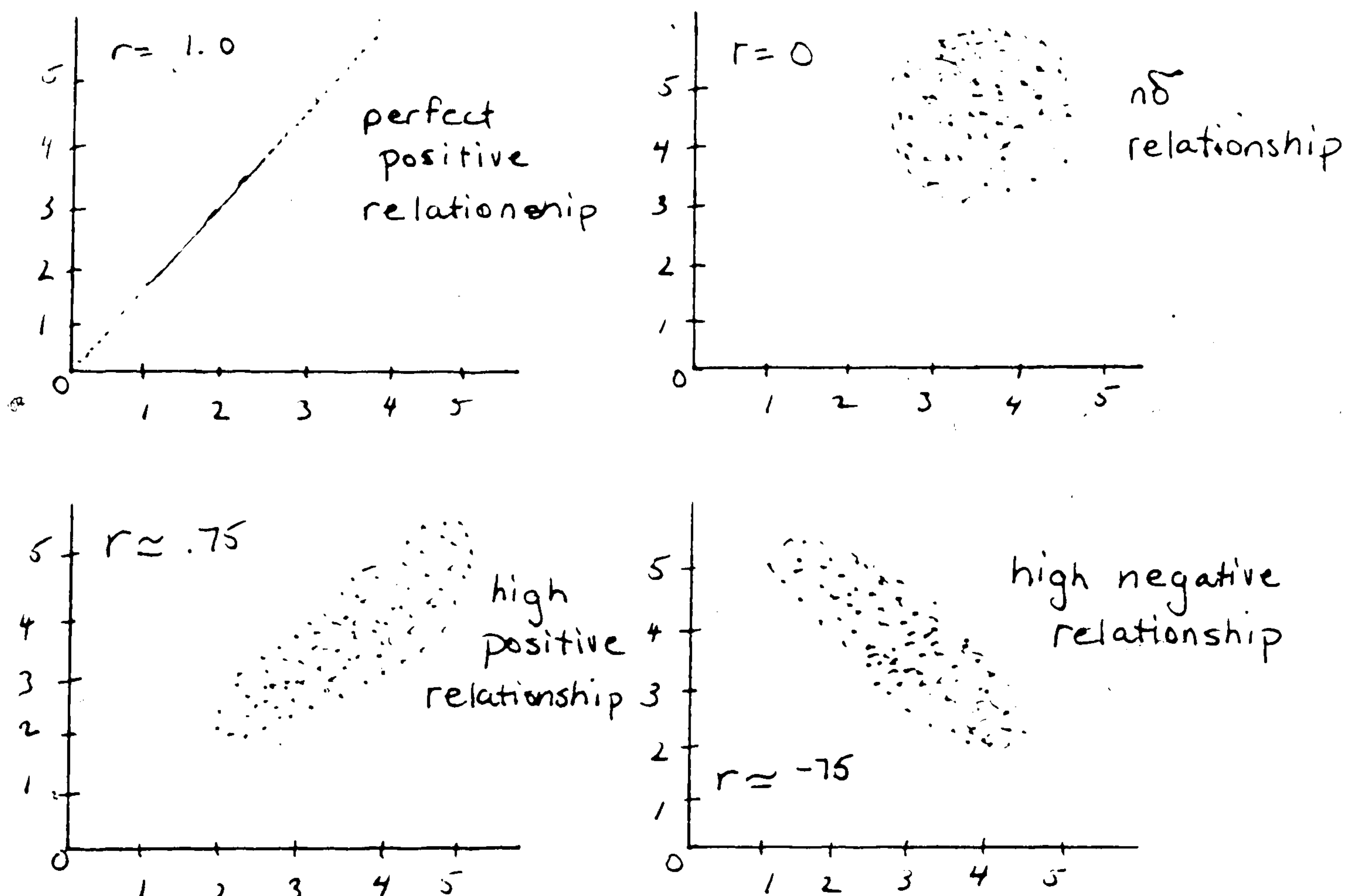
$$Z_x = \frac{X - \bar{X}}{S_x} \qquad Z_y = \frac{Y - \bar{Y}}{S_y}$$

The value of r can range from +1.00 to -1.00. The absolute value of the coefficient tells the strength of the relationship: the greater the absolute value, the greater the correspondence between the two sets of scores. Thus when $r = 1.00$, the scores on Y are completely predictable knowing the scores on X. If we know what a person scored on X, we could predict perfectly what his Y score would be. If $r = 0.00$, the relationship between the pairs of scores are random; knowledge of one score does not help us to predict the other. The sign of the correlation coefficient tells us the direction of the relationship; thus the coefficients of +.70 and -.70 represent equally strong relationships, although relationships in the opposite direction.

The relationship summarized by the correlation coefficient can be shown by the scatterplot of the test scores. If there is high correlation, the scores will cluster around a straight line. The orientation of the line with respect to the axes will indicate if the coefficient is positive or negative. (See Figure 2).

It cannot be overemphasized that r indicates only the degree of relationship between two variables: it does not indicate causation. If variables X and Y are highly correlated, there are at least three possible explanations: (1) X causes Y and thus changes in X result in changes in Y; (2) Y causes X and thus changes in Y result in changes in X; or (3) X and Y are both influenced by something else. In other words, a corelation co[illegible] is not a sufficient basis for inferring causation.

EACH DOT INDICATES [illegible] ON



(Figure 2)

Example:

Suppose that two arithmetic tests were administered to a class of students. The class mean on Test A was 20 and its standard deviation was 2. The class mean on Test B was 25 and its standard deviation was 4. What is the correlation between Test A and Test B based on the following 5 students?

Student	Test A	Test B	z_A	z_B	$z_A z_B$
Alice	22	29	1	1	1
Beth	23	27	1.5	.5	.75
Carol	21	23	.5	-.5	-.25
Debby	20	27	0	.5	0
Ellen	19	25	-.5	0	0
				$\Sigma z_A z_B$	1.5

$z_A z_B = 1.5$

$$r = \frac{\Sigma z_A z_B}{N} = \frac{1.5}{5} = .3$$

example:

Let's calculate the same coefficient using the other formula.

Student	Test A	Test B	$(A - \bar{A})$	$(B - \bar{B})$	$(A - \bar{A})(B - \bar{B})$
Alice	22	29	2	4	8
Beth	23	27	3	2	6
Carol	21	23	1	-2	-2
Debby	20	27	0	2	0
Ellen	19	25	-1	0	0
					12

$$r = \frac{\Sigma (A - \bar{A})(B - \bar{B})}{N S_A S_B} = \frac{12}{5 \times 2 \times 4}$$

$$= \frac{12}{40} = .30$$

Practice Problems:

1. What is the correlation coefficient between the following sets of scores?

   Social Studies: 11 12 13 14 15
   Language: 58 57 59 62 61

2. When comparing scores on a mid-term test with scores on the final test, a teacher calculated the following:

   $\Sigma z_M z_F = 22$ $N = 30$

   What is the correlation between mid-term and final test scores?

3. When checking the cummulative record file, a teacher sense that her students' performances seemed closely related to their previous years' marks. She calculated the following. Were her suspicions supported?

   $\Sigma (X - \bar{X})(Y - \bar{Y}) = 12$ $N = 25$

4. The correlation between Spelling and Grammar was calculated to be .80. The sum of the product of the Z-scores ($\Sigma z_S z_G$) was 24. How many students were used in the computations?

APPENDIX B

ACHIEVEMENT TEST

Name:

EDPSY 218B

I.D. #:

March, 1978

MID-TERM

A. Yackulic

Time allowed: 55 minutes

Total marks 50

Each of the questions or incomplete statements below is followed by several answers. From these you are to choose the one alternative that answers the question or completes the statement correctly. Use a LEAD PENCIL to record your choice on the answer sheet provided. Attempt all questions. No penalty will be applied to incorrect answers. Each question is worth one mark.

1. Fred scores 80 on a teacher-made math test. Jill scores 40. The BEST statement regarding their performace is:

    a. Fred knows twice as much math as Jill
    b. Jill has accomplished fewed than half the objectives
    c. Fred passed the test while Jill failed
    d. Fred scored higher than Jill on the test

Questions 2 - 7: Classify the following measuring situations according to the type of scale being employed. Record your answers using the following code:

a. nominal
b. ordinal
c. interval
d. ratio

2. Students participating in intramural activities are grouped into Red House, Blue House and Black House.

3. For each score on a test, the teacher reports the number of standard deviations the score is from the mean.

4. The five most beautiful women are selected in a beauty contest (that is, first, second, ... to fifth place).

5. A principal counts the number of students in attendance each day of the school year.

6. A centigrade thermometer is used to measure air temperature.

7. A teacher constructs a test and uses it to measure achievement in English literature.

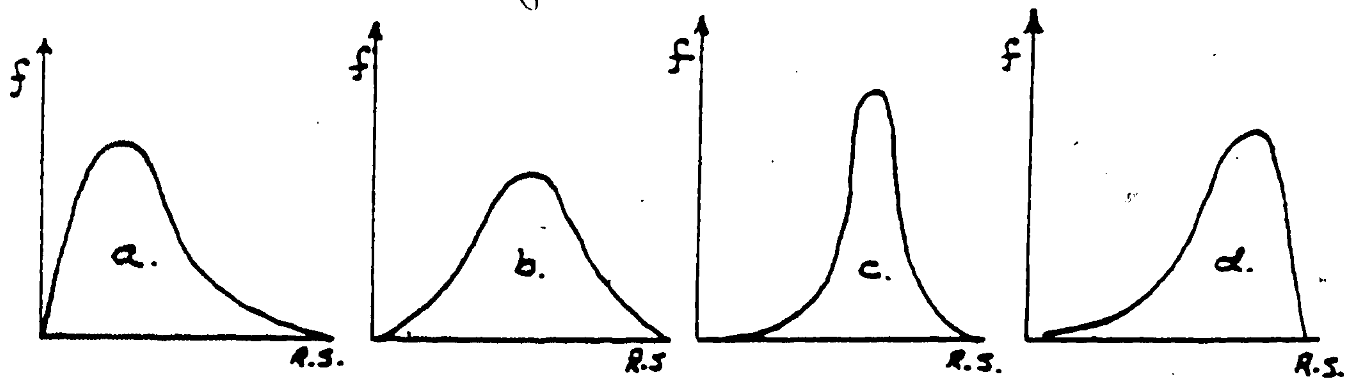
8. Mode:median:nominal:

    a. ordinal
    b. interval
    c. ratio
    d. z-scale

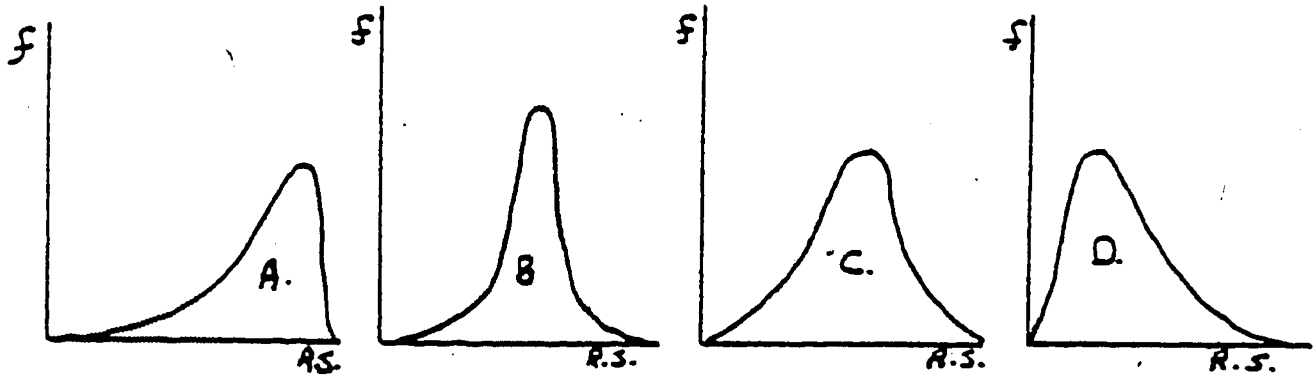
9. The following frequency table represents student grades on a scale from 0 to 4.

Grade	Frequency
0- .9	20
1.0-1.9	40
2.0-2.9	40
3.0-4.0	20

Which of the following shapes depicts these scores?



10. The following distributions depict the performance of a single class on four different Social Studies tests. The range in performance was the same on all tests.



On which test did the mode exceed the mean?

a. A b. B c. C d. D

11. From question 10, which test has the smallest standard deviation?

a. A b. B c. C d. D

12. From Question 10, which test will likely have the lowest reliability?

a. A b. B c. C d. D

13 - 15 Given the following test scores:

15, 16, 17, 18, 19, 20, 21, 24, 24, 24, 25

13. The mean is:

a. 15
b. 17
c. 20
d. 22
e. 25

14. The median is:

a. 15
b. 17
c. 20
d. 22
e. 25

15. The mode is:

a. 15
b. 17
c. 20
d. 22
3. 25

16. When copying a set of scores—1, 3, 5, 7, 9—a teacher erroneously copied 70 instead of 7. Her error should have increased the:

a. mean, median, and mode
b. the mean and mode only
c. the mean and median only
d. the mean only
e. the median only

17 - 20 A student in a senior typing class typed 10 pages, recording the following number of errors per page:

3, 2, 0, 1, 0, 1, 2, 0, 1, 0

17. The mean number of errors per page was:

a. 0
b. .5
c. 1
d. 1.5
e. 2

17 - 20 (continued) A student in a senior typing class typed 10 pages recording the following number of errors per page:

3, 2, 0, 1, 0, 1, 2, 0, 1, 0

18. The modal number of errors per page was:

a. 0
b. .5
c. 1
d. 1.5
e. 2

19. The range was:

a. 0
b. 1
c. 2
d. 3

20. The standard deviation was:

a. 0
b. 1
c. 2
d. 5

21. When extreme scores occur in the test results, the measure which best reflects the average score is the:

a. mean
b. median
c. mode

22. Mean:standard deviation:: median:

a. variance
b. semi-interquartile range
c. covariance
d. z-score

23. Sum of all scores:mean::sum of deviations squared:

a. mode
b. median
c. standard deviation
d. variance

24. Which of the following sets of scores has the largest range?

    a. 3,6,3,1,4
    b. 5,9,9,5,5
    c. 3,4,7,5,5
    d. 11,7,9,9,10

25. Which set of scores in question 24 has the largest variance?

26. What is the approximate variance of the following test scores:

    3,2,4,1,2,5,4?

    a. .6
    b. 1.7
    c. 3
    d. 12

27. In what units (e.g.: pounds, cm, ft) is the standard deviation of a test expressed?

    a. items
    b. standard units
    c. deviation units
    d. percentages

28. For a mastery test consisting of 50 items, you would expect a test variance of:

    a. close to 0
    b. about 30
    c. close to 50
    d. insufficient information to estimate

29. The mean of a distribution of z-scores will always be equal to:

    a. the standard deviation of the raw scores
    b. the mean of the raw scores
    c. one
    d. zero

30. On a spelling test Mary received a z-score of -.5. The test mean was 50 and the test standard deviation was 8. What was Mary's raw score on the test?

    a. 42
    b. 46
    c. 50
    d. 54
    e. 58

31. A test is given to 100 students in a large psychology class. The sum of the squared deviations about the mean is 2,500. The test mean is 52 and Janice scores 57. Janice's raw score is equivalent to a z-score of:

a. −2
b. −1
c. 0
d. +1
e. +2

32. You are asked by the principal to name one of five students for a science award. You were given the following records. Note that each student was given a different science test.

Student	Test Score	Test Mean	Test S.D.
Mary	100	85	15
Lou	100	125	10
Ted	61	40	7
Murray	140	100	20
Rhoda	62	50	12

Other things being equal, to whom would you present the award.

a. Mary
b. Lou
c. Ted
d. Murray
e. Rhoda

33. According to the table in question 32, Mary's performance was about the same as:

a. Ted's
b. Lou's
c. Murray's
d. Rhoda's

34. Which coefficient of correlation indicates the greatest degree of relationship between two variables?

a. .10
b. −.87
c. −.07
d. .76

35. Two tests—A and B—were given to a class of 30 students. $\Sigma(A-\bar{A})(B-\bar{B})=900$, $S_a=5$, $S_b=10$. What is the correlation between A and B?

a. .30
b. .45
c. .60
d. .75

36. The correlation between height and weight was found to be .31. It was later discovered that the weight scale was systematically recording 5 pounds heavy. The weights were adjusted (scale reading minus 5 pounds) and a new correlation coefficient was calculated. The new coefficient would be:

a. less than .31
b. .31
c. greater than .31
d. insufficient information to predict the new value

37.

Student	Hours of Study	Mid-term Test Score
Pierre	1	80
Joe	2	90
Jimmy	3	85
Anwar	4	85
Richard	5	80

According to the above table, the relationship between hours of study and test scores is:

a. positive linear
b. negative linear
c. curvilinear
d. random

38. A correlation of .70 was found between height and weight in a large sample of girls aged 10 to 16. If the correlation had been computed separately for each age group, it would have been:

a. lower than .70 among the 10 year olds
b. higher than .70 among the 10 year olds
c. .70 for each group if each group has the same number of girls

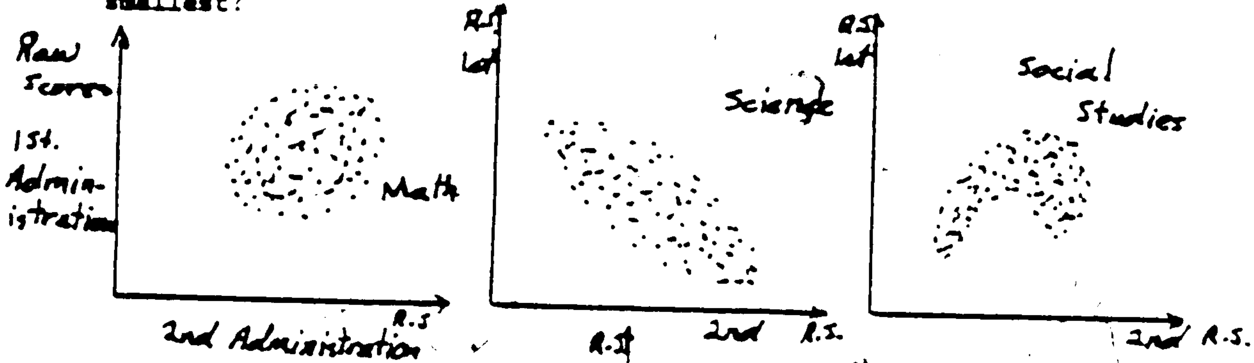
39. A small standard error of measurement indicates that:

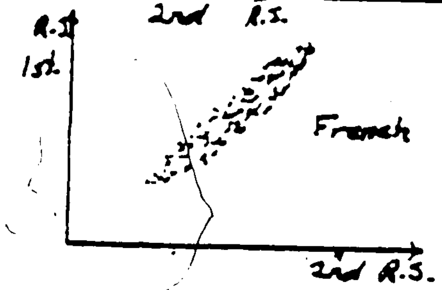
a. the test is highly unreliable
b. the obtained score probably approximates the true score closely
c. the standard error of measurement units are smaller than the test score units
d. little confidence should be placed in the obtained scores

40. On a test of Social Studies achievement John receives a score of 80 and Ted receives a score of 78. The standard error of measurement is 3.5. It may be correctly stated that:

a. there is probably not a genuine difference in the achievement of the two boys
b. John has achieved more in Social Studies than Ted has
c. the test did not reflect individual errors
d. the reliability of the test is questionable

41. For which of the following exams would the standard error of measurement be the smallest?





a. math
b. science
c. social studies
d. french

42. Pierre receives a score of 70 on a test of science aptitude. (test mean=45, test reliability=.6) The best estimate of his true score is:

a. 45
b. 50
c. 60
d. 70
e. 80

43. For a test of fixed length, as the reliability of the test decreases, the standard error of measurement:

a. increases
b. remains constant
c. decreases
d. does not fluctuate systematically with reliability

44. An educator is interested in establishing the equivalence of two forms of a literature test. How could she best measure this kind of reliability?

a. By administering both forms to a group and correlating their scores
b. By administering each form to a different group of students and computing a KR20 for each form
c. By administering each form to a different group of students and computing the split half reliability of each form
d. By having a group of experts in literature decide whether the content of both forms was equivalent

45. If a test has an $S_t^2=12$ and $S_e^2=8$, the test's reliability is:

a. .33
b. .40
c. .60
d. .67

46. Judy scores 2 S.D. above the mean on a math test of low reliability. What would happen to her score if she was retested with the same instrument two weeks later (assume her math skills did not change during the interval)?

a. her second score would be higher than her first
b. her second socre would be about the same as her first
c. her second score would be lower than her first
d. her second score cannot be predicted from the information given

47. Students' scores on the odd-numbered questions on a test may be compared with their scores on the even-numbered questions on the test in order to determine the test's:

a. internal consistency
b. equivalence reliability
c. stability reliability
d. predictive validity
e. concurrent validity

48. If a test very accurately predicts success in a given sequence of instruction, for the purposes of that instruction the test is said to have a high:

a. criterion validity
b. content validity
c. test/re-test reliability
d. objectivity

49. Scores on a new French test were correlated with scores on a standardized French test administered immediately after the new test. What kind of validity are we attempting to establish?

a. content
b. construct
c. predictive
d. concurrent

50. Which of the following statements is true regarding the relationship between reliability and validity?

a. a reliable test is likely to be nearly as valid
b. a reliable test is not always a valid test
c. a reliable test may be completely invalid and a valid test completely unreliable
d. a test will be at least as valid as it is reliable

CARD SORTING INSTRUCTIONS

Read the words on the accompanying cards and carefully consider the meaning of each of the words. Arrange the words into piles on the basis of similarity or closeness in meaning. All the words within a pile should be more similar in meaning to each other than to the words in other piles. You can use as many piles as you wish and can have as many or as few words as you like in any pile—all the way from a large number down to just one. When you have sorted the words, look over your piles and make any adjustments or changes you feel are appropriate.

When you have finished, place a pink card on top of each of the piles, then stack the piles one on top of another in a single deck. It does not matter in what order you arrange the deck so long as a pink card separates each of the smaller piles. Place a rubber band around the entire deck.

Remember, be sure to look carefully over all the words before starting and then, in terms of similarity or closeness in meaning, sort the words into as many different piles as you feel appropriate.

SIMILARITY RATING TEST

In the following task you will be presented with pairs of concepts. You are to consider each concept in a pair and then determine how similar one concept of the pair is to the other. You are to record your judgement on a scale from "very similar" on the left, to "very different" on the right. Each item will look like this:

VERY SIMILAR — SOFA : COUCH — VERY DIFFERENT

If you feel that these concepts are similar, you would circle a point on the left side of the line. The more similar you think the concepts are, the further to the left you would place your circle. Here's another example:

VERY SIMILAR — CBC : OREGON — VERY DIFFERENT

If you think that these concepts are different, you would circle a point on the right side of the line. The more different you think the concepts are, the further to the right you should place your circle.

If you think the concepts are neither very similar nor very different, you would circle a point near the centre of the line, depending on how similar you think the concepts are. For example:

VERY SIMILAR — DESK : CHAIR — VERY DIFFERENT

Try to be analytical in your judgements. Remember that similarity between concepts has to do with the number of features which the concepts share.

Below are the concepts which appear on the next few pages. Read the list carefully and recall what you know about each of the concepts. You may not know precise meanings of all the words; rely instead on what you currently understand about them.

correlation	median	range
error of measure	mode	rank
frequency	nominal	raw score
interval	normal curve	variance
mean	ordinal	z-score

Work quickly, first impressions are important. Use the entire scale; do not be afraid to circle the end points. Occasionally some concept pairs will be repeated. Please rate these without trying to recall your first answers.

Remember, you are to base your judgements on similarity--use the LEFT side of the scale to indicate SIMILARITY and the RIGHT side of the scale to indicate DIFFERENCE.

THERE ARE QUESTIONS ON BOTH SIDES OF THE PAGES.

VERY SIMILAR — VERY DIFFERENT

INTERVAL : NORMAL CURVE

CORRELATION : NOMINAL

Z-SCORE : NORMAL CURVE

FREQUENCY : MEDIAN

RANK : NOMINAL

ERROR OF MEASURE : MODE

MEDIAN : RAW SCORE

CORRELATION : VARIANCE

MEAN : VARIANCE

ORDINAL : MEDIAN

SIMILARITY RANKING TEST

Carefully consider the meaning of each word listed below:

correlation

error of measure

frequency

interval

mean

median

mode

nominal

normal curve

ordinal

range

rank

raw score

variance

z-score

On the following pages you will be given a randomly arranged list of 105 pairs of words, where each of the above words is paired with every other word. GO CAREFULLY THROUGH THE LIST AND THOROUGHLY STUDY ALL THE PAIRS. Then write 1 by the pair which is most similar, 2 by the pair which is the next most similar pair, 3 by the next most similar pair . . . and so on until 105 for the least similar pair.

WORK SLOWLY AND CAREFULLY; THIS IS A DIFFICULT TASK; TAKE YOUR TIME.

___ interval : raw score

___ normal curve : correlation

___ error of measure : range

___ nominal : interval

___ frequency : variance

___ median : z-score

___ variance : ordinal

___ mode: correlation

___ mode : ordinal

___ interval : mean

___ variance : nominal

___ mean : nominal

___ interval : z-score

___ interval : mode

___ median : error of measure

___ ordinal : nominal

___ z-score : range

___ rank : ordinal

___ interval : error of measure

___ variance : error of measure

___ range : rank

___ z-score: raw score

___ correlation : ordinal

___ rank : error of measure

___ range : frequency

___ variance : rank

___ correlation : nominal

___ frequency : median

___ error of measure : mode

___ correlation : variance

___ ordinal : median

___ rank : raw score

___ range : normal curve

___ rank : median

___ nominal : raw score

___ error of measure : raw score

___ rank : z-score

___ normal curve : variance

___ mode : nominal

___ mean : median

___ ordinal : z-score

___ normal curve : ordinal

___ correlation : error of measure

___ mean : ordinal

___ ordinal : error of measure

___ mode : rank

___ ordinal : raw score

___ frequency : nominal

___ mode : mean

___ correlation : interval

___ normal curve : frequency

___ raw score : correlation

___ range : correlation

___ frequency : interval

___ frequency : ordinal

___ median : mode

___ ordinal : range

___ frequency : rank

___ raw score : normal curve

___ mean : correlation

___ median : normal curve

___ nominal : normal curve

___ raw score : variance

___ normal curve : rank

___ rank : interval

___ range : mean

___ frequency : correlation

___ correlation : rank

___ mode : frequency

___ error of measure : z-score

___ ordinal : interval

___ z-score : frequency

___ range : mode

___ raw score : mode

___ raw score : frequency

___ variance : median

___ z-score : correlation

___ variance : interval

___ raw score : mean

___ interval : normal curve

___ z-score : normal curve

___ rank : nominal

___ median : raw score

___ mean : variance

___ correlation : median

___ z-score : mode

___ nominal : z-score

___ variance : z-score

___ median : interval

___ error of measure : normal curve

___ nominal : range

___ error of measure : frequency

___ mode : variance

___ raw score : range

___ interval : range

___ range : variance

___ nominal : median

___ z-score : mean

___ mean : rank

___ normal curve : mean

___ mean : frequency

___ median : range

___ normal curve : mode

___ nominal : error of measure

___ error of measure : mean