INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9334940

Development of students' cognitive structures in three disciplines

Streveler, Ruth A., Ph.D.

University of Hawaii, 1993

Copyright ©1993 by Streveler, Ruth A. All rights reserved.

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

DEVELOPMENT OF STUDENTS' COGNITIVE STRUCTURES

IN THREE DISCIPLINES

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

EDUCATIONAL PSYCHOLOGY

AUGUST 1993

By

Ruth A. Streveler

Dissertation Committee:

Frederick Bail, Chairperson
Martha Crosby
Peter Dunn-Rankin
Marie Iding
Shu Qiang Zhang

© Copyright 1993
By
Ruth A. Streveler
All Rights Reserved

ACKNOWLEDGMENTS

I complete this dissertation with deep gratitude and thanks to all those who supported me during this long journey.

To Fred Bail, Martha Crosby, Pete Dunn-Rankin, Marie Iding and Shu Qiang Zhang whose enthusiasm made writing this dissertation a positive learning experience.

To Mary Farmer, Mike Kirk-Kuwaye and Liz Peecook who were always there, always cheering for me, and always ready to help.

To Kathy Streveler and Lars Brennan who cheerfully accommodated my erratic hours and sometimes frazzled mood.

To my brother Dennis and my sister Mary who traveled thousands of miles to celebrate this accomplishment.

To my darling Glenn who always believed in me and supported me.

To the memory of my dear parents who, through their own curiosity, made learning seem natural.

ABSTRACT

The purpose of this study was to begin to integrate the findings of expert-novice research with studies of the students' cognitive structures. It is assumed that the findings of expert-novice studies can be used to predict the changes which occur in students' cognitive structures with instruction, as well as how the cognitive structures of low achieving and high achieving students might differ.

Instructors of three undergraduate college courses created a list of terms each deemed was central to their course. At the beginning and end of a sixteen-week semester, students in the respective courses 1) rated their familiarity with these terms, 2) clustered the terms and 3) described why each group of terms belonged together. The course instructors also clustered the terms at the beginning and end of the semester and wrote descriptions for each group of terms.

A percent overlap matrix between course terms was calculated and analyzed through multidimensional scaling. For each course, terms were assigned to groups based on the instructors' clustering at the end of the

semester. Students' clustering of terms was then interpreted with regards to the instructors' groups. The centroid of each of the groups was determined, and the mean geometric distance of points from the centroid was used as a measure of coherence of the groups.

The results of this study support the hypothesis that students' cognitive structures are more coherent and more similar to the instructors' cognitive structures after instruction. The question of whether or not the cognitive structures of high achieving students are more coherent than the cognitive structures of low achieving students is not settled. Though promising, the results of the study were not significant. The results do not support the hypotheses that more familiar terms will be clustered more coherently, or that a relationship exists between the depth of categorization and instruction or course achievement.

Implications of the findings of the study are discussed. Of particular note, is the supposition that the method used in this study could be used by classroom teachers to diagnose student misconceptions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

For over twenty years, researchers in cognitive and educational psychology have grappled with the problem of describing how cognitive structures change as an individual acquires knowledge in a domain. One approach to this question has compared the cognitive structures of experts and novices in various domains (McKeithen et al., 1981; Chi & Koeske, 1983; Gobbo & Chi, 1986; Chi, Hutchinson, & Robin, 1989). A second approach describes the differences in students' cognitive structures before and after instruction (Shavelson, 1972; Shavelson, 1974; Shavelson & Geeslin, 1975; Geeslin & Shavelson, 1975; Shavelson & Stanton, 1975; Champagne et al., 1981; Shavelson, 1985; Naveh-Benjamin, et al., 1986, 1989).

The results from these two approaches should be complementary and findings of expert-novice studies should predict how students' cognitive structures will change. However, no one has examined whether changes in students' cognitive structures can be predicted by expert-novice studies. An aim of the present study is to integrate expert-novice research with investigations of cognitive structure changes in students.

Many studies of both expert-novice differences and students' cognitive structures have been conducted in a few, well-structured domains such as chess (deGroot, 1965, 1966; Chase & Simon, 1973a, 1973b), and physics (Shavelson, 1972; Larkin, et al., 1980; Chi, Feltovich & Glaser, 1981). The present study investigates the differences in students' cognitive structures across three different domains: Physics, Educational Psychology and Women's Studies. This research design allows for some preliminary inter-domain comparisons to be made.

Statement of the Problem

In this study, changes in the cognitive structures of students will be examined by exploring the following five areas:

1- the relationship between students' cognitive structures and instruction;

2- the relationship between students' cognitive structures and course achievement;

3- the relationship between students' cognitive structures and student familiarity with course terms;

4- the role of instruction on student categorization of terms; and,

5- the relationship between course achievement and student categorization of terms.

Assumptions of the Present Study

The following assumptions are made in this study.

1- After instruction, students in a domain are more similar to experts in that domain than they are prior to instruction.

2- Students in a domain who receive high course grades are more similar to experts in that domain than students who receive low course grades.

3- This movement by students toward greater expertise can be detected in the course of one sixteen-week semester.

4- Measures of students' cognitive structures can be compared across domains.

Overview

The current study aims to integrate the findings of expert-novice studies with the research of students' cognitive structures. In order to provide an appropriate background for this investigation, Chapter II reviews the relevant literature in five areas: expert-novice studies, a discussion of the nature of expertise, studies of expertise in "ill-structured" domains, studies of students' cognitive structures, and issues related to the measurement of cognitive structure. The rationale for the present study is also outlined in Chapter II, along with the statement of five hypotheses.

In Chapter III, the methodology used in the present study is explained. The participants, procedure and data analysis used in the study are described. The results of the present study are described in Chapter IV.

Finally, Chapter V focuses on interpretation of the results of this study. The limitations of the study, recommendations for future research, and implications of the results on the field of education are also discussed.

CHAPTER II: REVIEW OF RELATED LITERATURE

In this chapter the relevant literature in five areas will be discussed: expert-novice studies, a discussion of the nature of expertise, studies of expertise in ill-structured domains, studies of students' cognitive structures, and issues related to the measurement of cognitive structure. The focus of the present study will be examined, and hypotheses will be presented.

Expert-Novice Studies

In the 1960s and 1970s interest developed in how the memory of experts might be different from the memory of novices in a field. DeGroot (1965; 1966) compared the memory of chess beginners and masters and found that master chess players had superior memories only when chess pieces were placed on a chess board in a *meaningful* pattern. Other studies of chess players (Chase & Simon, 1973a), experts and novices of other games (Reitman, 1976 [Go]; Engle & Bukstel, 1978 [Contract-bridge]), and some technical fields (Egan &

Schwartz, 1979 [symbolic drawing]; McKeithen et al., 1981 [computer programming]) have replicated these results.

With the superiority of experts' ability to recall meaningful domain-specific knowledge well established, research began to focus on the intriguing question of why this should be so. Among the first to tackle this problem were Chase and Simon (1973a, 1973b). Expert and novice chess players were shown a chess board and then were asked to duplicate the pattern of chess pieces on the board. Chase and Simon measured the time intervals between placement of successive pieces (which they call inter-response time (IRT)) and then examined the pattern of pauses. They assumed that unusually long pauses represented the boundaries between "chunks" of information stored in memory and found that chess experts can encode larger chunks of domain-pertinent information than can novices (Chase & Simon, 1973a). Inter-response time was also found useful in determining chunk boundaries in symbolic drawing (Egan & Schwartz, 1979) but IRTs could not be reliably used to determine chunks in the game of Go (Reitman, 1976).

McKeithen et al. (1981) studied differences between novice, intermediate and expert programmers in the ALGOL W computer language. Beginners, intermediate and experts programmers were asked to memorize 21, one-syllable "reserved words" for ALGOL W, such as "if," "then," or "string." Recall was cued on some occasions and uncued on others. The Reitman and Reuter algorithm (1980) was used to create an "ordered tree" organization, or hierarchical arrangement of terms, for each of the 22 subjects. The organization, depth, and similarity of trees were compared. Organization was measured using possible recall order (PRO), which refers to the number of different recall orders that could be generated by a specific ordered-tree. Depth was determined by the hierarchical organization of the tree. Similarity between trees was calculated using a formula to find proportion of similarity between two trees. A distance matrix between chunks was created to represent the distance between trees and multidimensional scaling was used depict the distances between all subjects.

Their study found no significant difference between skill levels in organization (as measured by PRO) or depth. However, multidimensional scaling analysis did show that experts' trees are more like one another than the trees of intermediates and beginners.

In the early 1980s the domain of choice in expert-novice studies switched from chess to physics, and the focus of study moved from recall to problem solving. Larkin, et al., (1980) found that experts in physics used a different representation of problems than did novices, (specifically, experts often used a pictorial representation) and they also spent more time constructing the representation than did novices.

In one of the seminal studies of expert-novice differences Chi, Feltovich and Glaser (1981) asked physics experts (advanced graduate students) and novices (undergraduates) to sort and then solve physics problems. Experts took more time to sort the problems during the first trial, but both groups were able to sort more quickly in the second trial. There were no differences in the number of categories produced by each group, and both experts and novices were able to reach a consistent sort within two trials. Probably their most important finding was that novices use "surface characteristics" to sort problems into categories (for example, "these problems all deal with inclined planes"), while experts use "deep characteristics" (such as "these problems can be solved using the Second Law of Thermodynamics.")

When solving problems Chi et al. found that experts often entertained a hypothesis early in the reading of a problem that was followed by the extraction from the problem of additional features that were used to confirm, reject or choose among hypothesized principles. However, novices seemed to be guided by surface-oriented schemata and in the course of problem-solving generated specific equations to solve specific problems.

Chi and her colleagues (Chi & Koeske, 1983; Gobbo & Chi, 1986; Chi, Hutchinson & Robin, 1989) also conducted a series of important studies of the structure of children's knowledge of dinosaurs. Chi and Koeske (1983), studied the dinosaur knowledge of a 4 1/2-year-old "expert." Over the course of six interviews the boy named 46 dinosaurs and responded to the question: "Tell me everything you know about them." This qualitative data was used to create semantic maps of the boy's dinosaur knowledge. In the semantic maps, links between dinosaurs, and links between dinosaurs and attributes of dinosaurs (such as whether or not they were meat-eaters) were represented by lines. The frequency of successive mention of two items was taken as a measure of the strength of the link between those items.

Dinosaurs were divided into two subgroups of twenty based on how well known the dinosaurs were to the boy under study. These subgroups were determined by the number of times the dinosaurs were mentioned in the boy's books as well as by the boy's mother's judgment about which dinosaurs he knew best. Links between dinosaurs and attributes of dinosaurs were determined by the order and frequency of mention of dinosaurs and attributes of dinosaurs.

Semantic maps of better-known dinosaurs and lesser-known dinosaurs were compared. The maps of better-known dinosaurs had a greater total number of links with other dinosaurs, greater strength of linkages, and stronger links within-groups than between-groups.

In an extension of this study, groups of children were studied who were considered experts or novices with respect the their knowledge of dinosaurs. Similar to the results found in sorting of problems by physics experts and novices, novice children also used surface characteristics to sort dinosaurs into categories, while expert children used deep characteristics to sort dinosaurs (Gobbo & Chi, 1986; Chi, Hutchinson, & Robin, 1989). It was also found that expert children used consistent rules to make groups (Gobbo & Chi, 1986) and that expert children were better able to place novel dinosaurs into correct categories (Chi, Hutchinson, & Robin, 1989). Gobbo and Chi (1986) proposed that experts' knowledge structures in the domain of their expertise are more cohesive and integrated.

The Nature of Expertise

By the late 1980's researchers were willing to tackle the subject of the nature of expertise (Lesgold, 1984; Glaser, 1985; Posner, 1988; Glaser & Chi, 1988). Glaser and Chi (1988, pp. xvii-xx) summarized seven characteristics they felt could be generalized about experts.

1- Experts excel mainly in their own domains.

2- Experts perceive large meaningful patterns in their domains.

3- Experts are fast; they are faster than novices at performing the skills of their domain, and they quickly solve problems with little error.

4- Experts have superior short-term and long-term memory.

5- Experts see and represent a problem in their domain at a deeper (more principled) level than novices; novices tend to represent a problem at a more superficial level.

6- Experts spend a great deal of time analyzing a problem qualitatively.

7- Experts have strong self-monitoring skills.

However, if expert skills are domain specific, how might the nature of the domain influence expertise? In 1980, Larkin et al., speculated that the results of studies in highly structure domains such as chess and physics were generalizable.

> "Expertness probably has much the same foundations wherever encountered. As in genetics, we learn much about all organisms by studying a few intensely. Chess, algebra and physics are serving as the *Drosophila*, *Neurospora*, and *Escherichia coli* of research on human cognitive skills." (Larkin, et al., 1980, p. 1336).

However, Glaser (1985, p. 12) later speculated that views of expertise were "probably biased by the highly structured domains in which is has been studied." Therefore studies of expertise and problem solving in less-structured domains must also be considered.

Studies of Expertise in Ill-Structured Domains

What defines an ill-structured problem or domain? Reitman (1965) defined a well-structured problem as being a problem whose solution was agreed upon by the community of experts in field. Ill-structured problems, on the other hand, have little agreement about their solutions. Both Reitman (1965) and Simon (1973), who also wrote about the nature of ill-structured problems, agreed that the well-structured versus ill-structured description of problems is not a dichotomy but rather a continuum.

If problems do not have agreed upon solutions then the characterization of experts as those who "quickly solve problems with little error" (Glaser & Chi, 1988, p. xviii) becomes very problematic. How can experts be characterized in domains where there are few correct solutions? This intriguing question has been tackled by researchers in such diverse fields as X-ray diagnosis (Lesgold et al., 1988), judicial decision-making (Lawrence, 1988), economics (Voss et al., 1989), instruction and testing in biomedicine (Feltovich et al., 1992), and military strategic thinking (Forsythe & Barber, 1992).

Voss, Tyler and Yengo (1983), studied the problem solving protocols of experts in the social sciences. They found that like experts in physics, social science experts spend more time than novices developing a representation of the problem. However, once the problem was represented, the social science experts began solving the problem in a method quite different from the method used by physics experts. While a physics expert would solve the problem through a series of equations, the social science experts offered an abstract solution and then began extensive argument development. This difference is attributed to the fact that problems in the social science do not often have agreed upon solutions (thus fitting Reitman's definition of an ill-structured problem) and a solution may be deemed correct by the strength of its supporting argument.

In a study of the development of problem-solving skill in the social sciences, Voss, Greene, Post, and Penner (1983) suggested that three types of structures were constructed as expertise was gained. First, conceptual networks were built and expanded. Second, causal relationships between factors are posited. And third, a hierarchical structure is developed which, they speculate, helps one to organize the large amount of information included in one's argument. Thus it would seem that hierarchical structure is present even in ill-structured problems.

Studies of Students' Cognitive Structures

The study of the changes in the way knowledge is structured is important to the study of expertise (Chi & Rees, 1983). If one assumes that the acquisition of expertise is on a continuum (Glaser, 1985) then students in a course should move closer to "expert status" after instruction. This movement in the direction of increased expertise might be reflected in changes in students' cognitive structures.

Shavelson and his colleagues studied change in the cognitive structures of students before and after instruction (Shavelson, 1972; Shavelson, 1974; Shavelson & Geeslin, 1975; Geeslin & Shavelson, 1975; Shavelson & Stanton, 1975; Shavelson, 1985). Shavelson predicted that, after instruction, the students' cognitive structures should more closely correspond to the content structure. Cognitive structure is defined as "a hypothetical construct referring to the organization (interrelationships) of concepts in long-term memory," and content structure is defined as "the web of concepts and their interrelationships in a body of instructional material" (Geeslin & Shavelson, 1975, p. 109).

It should be noted that there has been some controversy about whether or not the content and cognitive structures of students could be teased apart (see Phillips (1983) and responses by Greeno (1983) and Shavelson (1983)). However, the assumption that content and cognitive structure move closer together after instruction is prevalent in the literature on students' cognitive structure and is a predicted resulted for the studies described in this section.

Shavelson (1972) performed an experiment using 40 high school students who had not learned physics, but were interested in the subject. The students were divided into two subgroups of 20. One subgroup (the control group) received no instruction in physics. The other subgroup (the instructional group) read instructional material in physics for each of five days. Each day, both groups were given word association tasks using the same physics key words used to describe the content structure of their instructional material. Both groups were also given daily achievement tests in physics.

The content structure was measured by counting the frequency of key words in the instructional material (in this case, in physics) and then diagramming every sentence in the instructional material that contained two of the key words and converting those diagrams into digraphs according to a set of rules (Shavelson & Geeslin, 1975). Distances between items on the digraph was converted into a distance matrix.

Cognitive structure was measured by word association tasks. Students were provided with sheets which began with one of ten key concepts in physics. On each sheet they were asked in one minute to write as many other key terms that were associated with the first concept on the page. The data from these trials were converted into a relatedness coefficient (RC) matrix.

The correspondence between the distance matrix (representing the content structure) and the RC matrix (representing the cognitive structure) was determined in two ways. First, the matrices were analyzed using multidimensional scaling (Kruskal & Wish, 1978) and the plots of each matrix were visually compared and said to correspond (Shavelson, 1972).

A second measure of correspondence between content and cognitive structure was performed by comparing the

Euclidean distance between the distance matrix and the RC matrix. Euclidean distances between the matrices did decrease as predicted.

In another study with 34 high school students, this time using instructional materials in probability, (Geeslin & Shavelson, 1975), the same measures were used to measure correspondence between cognitive and content structure before and after instruction. The matrices were again analyzed with multidimensional scaling and by computing the Euclidean distance between the matrices.

The multidimensional scaling representations of the cognitive structure and the content structure were visually compared and felt to be consistent. A nonparametric analysis of variance test on the Euclidean distance between the matrices of control and treatment groups at the pretest and posttest revealed a significant difference ($p < .01$). These results were taken as evidence that cognitive structure had moved closer to content structure in the instructred group (Geeslin & Shavelson, 1975).

Champagne et al. (1981) studied the conceptual structure of 30 eighth-grade students studying geology.

Student pre- and post-instructional cognitive structures were measured using individual Concept Structure Analysis Technique (ConSAT) interviews. During the guided ConSAT interview a student is asked to arrange a set of cards containing important course terms "in a way that shows how you think about the words." The interviewer then questions the child about the relationships between terms and uses this information to draw arrows between concepts and describe the relationships. In other words, the child and interviewer create a kind of concept map of the child's understanding of the topic (in this case, geology).

The content structure was derived from the instructional material itself, through written and visual information (such as diagrams.) No attempt was made to use Geeslin and Shavelson's (1975) method of constructing the content structure of the material. Rather, the instructional materials were reviewed by three university professors who asserted that the materials "successfully maintained the scientific integrity of geology"(p. 98).

The ConSAT map created by each child were analyzed by classifying the map along six dimensions (Champagne et al., 1981, p. 105):

1- The size of the unit which is structured.
2- Relations between structural units are explicit.
3- Relations between structural units are scientific.
4- Degree of interconnectedness of relations between units.
5- Predictability between structural units.
6- Connections between concepts in structure.

Two raters than used these dimensions to analyze the ConSAT maps to measure change in the maps after instruction. Though no test of significance was performed the authors felt that the changes in the students' maps were consistent with what would be expected if the cognitive structures of the students moved closer to the structure of the content.

The cognitive structures of college students has also been measured. Naveh-Benjamin et al. (1986) used the ordered tree technique developed by Reitman and

Reuter (1980) to measure the change in knowledge structure at the beginning and end of an undergraduate course on the Psychology of Aging. Students were given a set of terms and asked to list these terms in an order that made sense to them. This task was repeated four times, in two cued and two uncued trials. Repetitions were used to increase the reliability of the measure.

The Reitman and Reuter algorithm (Reitman & Reuter, 1980) was used to determine organization, depth, and similarity between the instructor's structure and the students' structures. The amount of organization was measured by the possible recall order (PRO), "the natural logarithm of all different orders that can be obtained by traversal of a given structure" (Naveh-Benjamin, et al., 1986, p. 132). The smaller the PRO, the greater organization in the structure. Depth is a measure of hierarchical levels in an ordered tree. The greater the depth, the more hierarchical levels. Similarity provides a measure of resemblance between two trees, in this case a student's tree and the instructor's tree. Similarity is calculated by the formula:

$$\frac{\ln\text{ (number of chunks the two trees have in common + 1)}}{\ln\text{ (total number of chunks contained in both trees + 1)}}$$

where "ln" is the natural logarithm (McKeithen et al., 1981, p. 321).

In a pilot study, PRO, depth and similarity were all significantly correlated to final course grade and PRO and depth were significantly correlated. Their results also revealed an interesting interaction between organization and similarity to instructor's structure. Students with high organization but low similarity to the instructor did not exhibit high course performance. The researchers attempted to explain this by speculating that the high organization measure could sometimes be measuring stereotyping and repetition during the four trials.

In an extension of the pilot study, students were asked to perform the task on the first day of class, at the middle of the term and on the last day of the term. The PRO, depth and similarity at the beginning, middle and end of the semester and was compared for A students, B students, and a combined group of C and D students.

The A students were the most disorganized of all groups on the pretest. During the semester their organization, depth and similarity all increased. For B students organization increased, depth increased until the mid-term and then stayed level, and similarity increased. For the combined group of C and D students, organization, depth and similarity to instructor's tree all increased from the beginning to the middle of the semester, but then decreased at the end of the semester. It is interesting to speculate what might have happened around midterm time, when organization, depth and similarity all began to decline.

The authors concluded that the ordered tree was useful as an assessment measure. They proposed that low correlations between their measures of organization, depth and similarity, on the one hand, and course grades, on the other hand, meant that they were measuring something different than the course tests measure. Of the three measures, similarity to instructor was most sensitive to achievement.

In a second study by Naveh-Benjamin et al. (1989), students in another Psychology of Aging course were asked to complete the knowledge structure task using four repetitions of the Reitman and Reuter (1980) ordered tree method at the beginning, middle, and end of the semester. As in their previous study, ordered trees were analyzed using the Reitman and Reuter algorithm for organization, depth of structure and similarity of student's structures to that of the instructor.

During the semester there was an increase in all measures and all measures were positively correlated with course performance. In a reversal of the findings in their 1986 study, organization and hierarchical depth were significantly correlated with course performance. However, similarity to instructor's structure was not significantly correlated to course performance. The authors do not address the discrepancy between the results of their two studies.

Measurement of Cognitive Structure

A monumental number of psychological research studies on memory have concluded that an individual's memory is organized. What can be considered classic works by Miller (1956), Tulving (1972), Rumelhart (Rumelhart & Ortony, 1977), and countless others, have left little doubt that memory and knowledge are organized; that cognitive structures exist.

The spreading activation theory of semantic memory (Collins & Quillian, 1969; Collins & Loftus, 1975) has provided a useful and accepted construct for measuring the relatedness of information in memory, and thus became a tool for the measurement of cognitive structure. Words more closely associated in memory are retrieved more quickly than words that are more remotely associated, thus reaction time has been used as a measure of relatedness of concepts within memory. This method has been expanded to include the assumption that words recalled closely together in a word association test are closely related in memory.

Studies of the construct validity of measures of cognitive structure have also been conducted. Shavelson and Stanton, (1975) found that the cognitive structure representations of two math experts produced from word association, card sorting and digraphing were convergent. The cognitive structure representations were similar enough to each other to the expected structure (as represented by a text book) to be deemed valid measures of cognitive structure.

Champagne, Hoz, and Klopfer (1984), compared free sort, tree construction and word association as probes of students' cognitive structure of Physics concepts. They found that free sort and tree construction tasks had adequate test-retest reliability, but word association was not reliable.

They also compared the analysis of these three probes using three scaling methods: multidimensional scaling (using KYST (Kruskal, Young, & Speery, 1973)), hierarchical clustering (using the HCLUST program (Johnson, 1967)) and latent partition analysis (Wiley, 1967). This last method is a "multivariate scaling technique designed to reveal the latent categories that subjects presumably use to classify concepts" (Champagne, Hoz, & Klopfer, 1984, p. 12).

The three scaling measures were analyzed with regard to their sensitivity to differences in the students' distance matrices (which were determined by their task performance), the comparability between scaling methods, and their discriminant validity, or their ability to distinguish the instructed from the uninstructed students cognitive structures. In order to determine discriminant validity one must assume that changes do exist between instructed and uninstructed students' cognitive structures of a content area.

Champagne and her colleagues found that the three scaling measures did produce comparable structures. Latent partition analysis was most sensitive to changes in the students' proximity matrices. However, the discriminant validity of the measures was not settled. One cannot tell if this was due to the inability of the measure to discriminate between uninstructed and instructed students' cognitive structures, or whether no change was detected because no change in the students' structures occurred.

Reitman and Rueter (1980), argued that an ordered-tree method of analysis was a superior to the most common method of inferring cognitive structure, multidimensional scaling. The ordered-tree method was used both by McKeithen et al. (1981) and, Naveh-Benjamin et al. (1986, 1989) to measure cognitive structure. However, the ordered-tree method yielded inconsistent results for Naveh-Benjamin et al. (1986, 1989), and McKeithen et al. (1981) found that two of the measures the technique calculates (organization and depth) were not sensitive to skill level. These results cast doubts on the usefulness of the ordered-tree method for describing cognitive structure.

Focus of the Present Study

It is reasonable to assume that with instruction students' cognitive structures in a domain will become more like experts' cognitive structures. If this is the case, trends seen in expert-novice studies of cognitive structures should be reflected in changes in pre- and post-instructional structures of students. Assuming that course grades can be used as an indication of degree of expertise, one would expect that students with high achievement should be more like experts than students with low course achievement. Three findings found in the expert-novice literature are of particular importance.

- The cognitive structures of experts are more coherent than the cognitive structures of novices (Gobbo & Chi, 1986). In other words, links between items will be stronger within groups than between groups.

• Better-known, or more familiar knowledge is grouped more coherently than is lesser-known or less familiar knowledge (Chi & Koeske, 1983).

• Experts use deeper level characteristics to categorize items or problems in that domain, while novices use surface level characteristics (Chi, Feltovich & Glaser, 1981; Chi, Hutchinson, & Robin, 1989).

Prominent studies of students' cognitive structures have not addressed the three issues listed above. No major studies of students' cognitive structures have compared the tightness of between-group and within-group clusters of concepts before and after instruction. No studies could be found in the literature which compared students' clustering of more familiar concepts with those of less familiar concepts. Likewise, studies were not found which compared how the criteria students use to sort concepts changes with instruction, or varies with course achievement.

In addition, very few studies of changes in students' cognitive structures have been conducted in a naturalistic setting. Most studies have been done using a very short instructional time, and in a setting where study was self-directed using only written material (Shavelson, 1972; Geeslin & Shavelson, 1975; Champagne et al., 1981). Of the three major groups of researchers in this area, only Naveh-Benjamin et al. (1986, 1989) conducted studies in an actual academic course over a semester. And only Naveh-Benjamin et al. (1986, 1989) attempted to correlate students' cognitive structures to course achievement.

The present study attempts to investigate these questions. Specifically, five hypotheses are proposed which address the link between expert-novices studies and the studies of students' cognitive structures.

Hpotheses

Hypothesis 1

After instruction, students' cognitive structures are more coherent than before instruction.

Hypothesis 2

After instruction, the cognitive structures of students with high course grades are more coherent than students with low course grades.

Hypothesis 3

The degree of familiarity students have with course terms will effect the way these terms are clustered. More familiar terms will be clustered more coherently both before and after instruction.

Hypothesis 4

After instruction, students use deeper level characteristics to categorize concepts than they do before instruction.

Hypothesis 5

After instruction, students with high course grades use deeper level characteristics or criteria to categorize concepts than students with low course grades.

The methods used to investigate these five hypotheses are detailed in the next chapter.

CHAPTER III: METHODS

In this chapter, the methodology used in the present study is explained. First, the participants of the study are described, followed by a description of the procedure and data analyses used in the study.

Participants

Participants in the study were University of Hawaii at Manoa undergraduates who were enrolled in one of three courses: Educational Psychology 311 (Psychological Foundations), Physics 151 (College Physics), or Women's Studies 151 (Introduction to Women's Studies). These courses were selected because they represented a range of disciplines, yet were all introductory in nature.

The three courses differed not only in content, but also in size. Physics was the largest course by far, with an enrollment of 177 students. By contrast, 23 students were enrolled in Educational Psychology, and the Women's Studies course had an enrollment of 20 students. The difference in course size also had an effect on the amount of student-instructor interaction that was possible in the respective courses. In the Physics course, with an enrollment of almost 200 students, only a limited amount of interaction was possible between the instructor and individual students during class time. On the other hand, the much smaller Educational Psychology and Women's Studies courses were highly interactive.

In addition to a difference in course size, the method the instructors used to evaluate student achievement was also different among the three courses. Educational Psychology and Physics instructors used objective examinations as the primary measure of student achievement, while in Women's Studies, students were evaluated on the basis of their written essays. The distribution of course grades was also quite different for the courses, as is reflected in Table 3.1. Very few students received a grade of C or below in either the Educational Psychology or Women's Studies courses. However, almost 60% of the students in the Physics course received a grade of C or below. In fact 60 of 177 students (about one-third of the students) failed the Physics course.

Table 3.1
Grade distribution in Educational Psychology 311, Physics 151, and Women's Studies 151.

	Number of students receiving grades of				
	A	B	C	D	F
Educational Psychology (N=23)	10	12	0	0	1
Physics (N=177)	31	30	38	18	60
Women's Studies (N=20)	14	3	2	0	1

Not all students were able to complete all of the assigned parts of the pretest and posttest. In addition, not all students in the courses were present for both the pretest and posttest. Only data from the students in the courses who completed both the pretest and posttest were used in this study. Thus the actual number of participants was 14 Educational Psychology students, 74 Physics students, and 14 Women's Studies students, or a grand total of 102 student participants (see Table 3.2). Students were also asked to give permission to release their course grades to the researcher (see Appendix III and Table 3.3).

Table 3.2.
Number of students completing pretests and posttests.

Course	Number of students completing				
	Pre-test1	Pre-test2	Post-test1	Post-test2	Both Pre-test1 and Posttest1
Educational Psychology 311	21	21	15	6	14
Physics 151	178	124	79	32	74
Women's Studies 151	19	19	14	13	14

Procedure

Prior to start of the Fall 1992 semester, three experienced instructors from three different introductory courses at the University of Hawaii at Manoa were asked to select 30 terms they felt were central to their respective courses (see Appendix I, instructions to professors). Expert selection of key terms is the usual method of determining the content structure of a course (Shavelson, 1972; Shavelson & Geeslin, 1975; Geeslin & Shavelson, 1975; Champagne et al., 1981).

Each course was visited at the beginning and end of the semester. The pretest was administered in Physics 151 on the first day of instruction. Students in

Educational Psychology 311 and Women's Studies 151 both were given pretests during the fourth class period. As will be explained later, Educational Psychology 311 was also visited on the fifth class period. Instructors were able to allow about 30 to 40 minutes of class time for the students to complete the pretest. Students were asked to complete three tasks during the pretest:

1) Provide voluntary background information including a listing of related courses they had completed in high school or college. (See Appendix II.)

2) Rate their familiarity with each of the 30 key terms selected by their instructor. Students used a 5-point Likert scale ranging from "I don't understand this concept at all" to "I could define and briefly explain this concept to another person". (See Appendix II.)

3) Cluster terms. Students were given envelopes containing strips of paper listing each of the 30 important course terms as well as a plastic baggie containing strips of blank paper and paper clips. Students were asked to cluster the thirty terms in any manner they thought appropriate. They were free to create as many or as few clusters as they liked, and

could include any number of terms within one group. Participants were also asked to write a description of why each group of terms went together on a blank strip of paper and paperclip the description to each group of terms. Students were asked to repeat the clustering task a second time. In two of the three courses, the repeated clustering task was done immediately after the first clustering task. Because of time constraints, students Educational Psychology 311, repeated the clustering 5 days after the first clustering.

About a week after the students performed these tasks, each instructor was asked to cluster terms and state why the concept in each group belonged together.

Posttests were administered within the last two weeks of instruction. In two courses, Educational Psychology 311 and Women's Studies 151, posttest were given on the last day of instruction. The posttest in Physics 151 was administered on fourth to the last class period. Following the same procedure, and using the same terms from the pretest, students were asked to rate their familiarity of the terms (see Appendix III) and then cluster terms appropriately. Students were also asked for permission to use their course grades (see Appendix III).

Instructors were also be asked to cluster the terms at the end of the semester. All three instructors chose to do the final clustering at the same time the students were clustering terms.

Data Analysis

Multidimensional scaling analysis was used to measure the cognitive structures of students in Educational Psychology 311, Physics 151, and Women's Studies 151 before and after instruction. Students in each course clustered 30 important course terms (as determined by the instructor for the course) both at the beginning and end of a semester. The overlap of both student pretest and posttest clusters in each course was determined using the program PEROVER (Dunn-Rankin, 1983) which calculates the overlap of terms using data derived from the individual students' clusters. The resultant percent overlap matrices were then analyzed using the SAS program ALSCAL (Young & Lewyckyj, 1979).

Multidimensional scaling is a set of techniques that convert proximity data (such as similarity judgments or free clustering data) into distances that are usually displayed on a plot of points. As Kruskal and Wish (1978, p. 19) emphasize: "The central motivating concept of multidimensional scaling is that *distance* d_{ij} *between the points should respond to the proximities* δ_{ij}*.*" [Their italics.]

The metric stress is used to measure the goodness of fit between the proximities and distances. The lower the stress value, the better the fit of the model to the data. Though there is no test for adequate stress level, an accepted rule of thumb is that levels of stress around .15 or .20 are acceptable. It is expected that solutions with more dimensions will have a better fit to the model (and thus a lower measure of stress) than solutions with fewer dimensions. However, solutions with more dimensions are more difficult to interpret. Thus, for practical purposes, a balance is struck between stress and dimensionality of solutions. Levels of stress for 2-dimensional and 3-dimensional solutions were determined for each of the three classes. In order to be able to compare solutions between the three courses, a solution with the same number of dimensions was sought for all courses.

The ALSCAL program calculates the coordinates of each term input (in this case for all 30 terms input for each of the three courses) in each dimension of the solution. Therefore, in a 3-dimensional solution, coordinates for each term in dimension 1, dimension 2, and dimension 3 are determined. In a 3-dimensional solution the coordinates can be used to plot the solution in 3-dimensional space. The theory behind multidimensional scaling assumes that, given reasonable levels of stress, the physical distance between terms on a plot will be representative of the psychological distance between terms. Thus, it is assumed that the coordinates and plots of terms analyzed with multidimensional scaling are representative of the relationships between terms in students' cognitive structures.

The multidimensional scaling solutions generated by the students clusters were next compared to the way in which the instructor clustered the terms. This comparison was accomplished by assigning terms to groups according to the instructors' clusters and comparing how the students' pretest and posttest groups changed relative to the instructors' groups.

Terms for each course were assigned to groups according to the instructors' posttest clustering. For example, in Physics, the terms "diffraction," "interference," and "polarization" were assigned to one group, and "absolute temperature scale," "chaos," "entropy," "order at the molecular level," and "Second Law of Thermodynamics" were assigned to another group (refer to Table 4.4 in the next chapter).

The center or centroid of the set of terms in each of the instructor's groups was then calculated by determining the average of the x, y, and z coordinates of terms within that group. The distance of each term in the group from the group centroid was determined geometrically using the formula:

Distance of point from centroid =

$$\text{square root}((x_1-x_2)^2 + (y_1-y_2)^2 + (z_1-z_2)^2)$$

where x_2, y_2, z_2 are the coordinates of the centroid of the group, and x_1, y_1, z_1 are the coordinates of the term whose distance is being determined.

The average distance of each term in a group from the group centroid was then determined for all groups. The average distance will be used as a measure of the coherence of a cluster. The term coherence refers to the strength of with-in group links versus between-group links (Chi & Koeske, 1983). Thus a group of terms that are tightly clustered together and more distinct from other groups is said to be more coherent then a more loosely clustered group. Clusters which were tighter or more coherent should have a smaller average distance than less coherent groups.

A binomial test can be performed to compare the pretest and posttest distances to see if the posttest groups are significantly more coherent than the pretest groups. In this sense, the pairs of pretest and posttest average distances for each of the three courses are then viewed as being a series of trials (with each cluster being a separate, independent trial) and with the situation "pretest average distance larger than the posttest average distance" considered a success. The reverse situation, "pretest average distance smaller than the posttest average distance," is thus considered a failure. One assumes that, by chance, the probability of the pretest average distance being larger than the posttest average distance is .5. Using the binomial test, the probability of the pretest distances being larger than the posttest distance by chance alone can be calculated.

One can calculate the probability of a certain number of successes (B) given N number of trials using the following formula (Hays, 1988, p. 129):

$\underline{B} = Np^{r}q^{N-r}$, where N = the total number of trials

r = the number of successes

p = the probability of success

q = the probability of failure (also 1-p)

In order to test the hypothesis that students with high course grades would cluster the terms more coherently than students with low course grades, the grades students received in the three courses were needed. Instructors agreed to provide both the overall grade distribution, and the grades of individual students who had given their permission for their course grade to be shared. The three courses were very different in their grade distributions, as was seen in Table 3.1.

The numbers of students receiving each grade who participated in the study were lower than the figures listed in Table 3.1. Only students who took both the pretest and the posttest <u>and</u> who agreed to have their course grade shared could be used. With those two limitations accounted for, the actual number of students eligible to participate is listed in Table 3.3.

Table 3.3
<u>Number of students taking both the pretest and posttest who agreed to share their course grades</u>.

	Number of eligible students receiving grades of				
	A	B	C	D	F
Educational Psychology (N=8)	4	4	0	0	0
Physics (N=62)	14	17	15	7	9
Women's Studies (N=10)	9	1	0	0	0

Given the situation described in Table 3.3 the judgment was made only to use the data from Physics 151, which was deemed to be the only one of the three courses with a large enough grade distribution to warrant analysis.

Subsets of the Physics 151 posttest data were analyzed and compared. Data for students receiving grades of either A or B were combined, as were the data for students receiving grades of D or F. The rationale for this combination was to look at the ends of the grade distribution, while omitting the center (that is, the C students.) These subsets of data were analyzed using PEROVER (Dunn-Rankin, 1983) and ALSCAL (Young & Lewyckyj, 1979). The terms were assigned to groups according to the instructor's groupings.

Students in each course rated their familiarity of 30 important course terms at the beginning and end of the course, at the same time the clustering of terms took place. Students rated terms using the following 5-point Likert scale.

1 = I don't understand this concept at all.
2 = I only have a vague understanding of this concept.
3 = I can understand this concept if it is presented in context.
4 = I understand the concept quite well.
5 = I could define and briefly explain this concept to another person.

The mean ratings of familiarity at the pretest and posttest were assumed to measure the students' familiarity with terms.

When students completed both the pretest and posttest clustering of terms they were asked to write an explanation of why each group of terms went together. These handwritten student explanations were typed by a person unfamiliar with the study. In order to prevent bias, the lists of explanations were presented to two raters who were graduate students in Educational Psychology. The raters were given no indication of the explanations' pretest or posttest source. The raters next created criteria for distinguishing between deep and surface explanations.

In their study of expert-novice differences problem solving in physics Chi, Feltovich and Glaser (1981), defined surface explanations as those explanations which used explicit characteristics (for example, the physical configuration of a problem) to group problems. Deep explanations were those using implicit criteria characteristics (such as the laws of physics needed to solve a problem) to group problems.

In present study, however, abstract concepts, not pictoral representations of problems, were sorted. Therefore, the raters were unable to use the same criteria that Chi and her colleagues used to distinguish between deep and surface categorization of physics problems.

In order to establish specific criteria for deep versus surface explanations, the raters first independently rated a sample set of ten pretest and ten posttest explanations from each of the three courses. These samples were drawn from explanations not used in the final study (for example, from students who had taken only the pretest or posttest but not both). The raters also examined the instructors' explanations and agreed that acceptable criteria for deep explanations should also describe the instructors' explanations.

After the independent rating of the sample explanations, the raters met and compared results and agreed upon criteria for deep and surface explanations. Three criteria were agreed upon. Explanations were rated as surface if they referred to a structural similarity between terms (for example, "these all begin with 's'" or "these are all '-ism' words"), or if the student referred to their familiarity with the terms to explain the clusters (for example, "I don't have a clue" or "I know all these terms well"). In addition, all groups that were given no explanation at all were also rated as surface explanations, on the assumption that the student could not come up with an explanation of why the terms belonged together.

All explanations that were not rated as surface according to one of the three criteria of structural similarity, student familiarity, or lack of an explanation, were rated as deep explanations. For example, the explanations "statistical terms" and "studies and theories dealing with development concepts" were related as deep explanations in Educational Psychology. "These deal with movement and energy" and "temperature stuff" were examples of physics explanations which were rated as deep by the raters. In Women's Studies, "acceptable social constraints" and "reasons or causes of oppression of women" were rated as deep explanations.

Once these criteria were established, the raters independently rated the full set of pretest and posttest explanations for the three courses. The raters then met to compare results. Interrater reliability (99.5%) was determined by calculating the number of explanations which were rated deep by one rater and surface by the other. The average number of deep and surface explanations per student were determined. A t-test was used to compare the average number of deep and surface explanations in pretests and posttests.

CHAPTER IV: RESULTS

The results of the present study are described in this chapter. Results which pertain to each of the five hypotheses are listed in order.

Hypothesis 1

Hypothesis 1 posits that after instruction, students' cognitive structures will become more coherent than they are before instruction.

Levels of stress for 2-dimensional and 3-dimensional solutions were determined for each of the three classes. For Educational Psychology 311 and Women's Studies 151, the stress of a 4-dimensional solution was also calculated. A 4-dimensional solution was not determined for Physics 151 because the stress of the 3-dimensional pretest and posttest solutions were already quite low (stress = .134 for the pretest and stress = .120 for the posttest). Levels of stress for the pretest and posttest of each course are listed in Table 4.1.

Table 4.1
Stress of solutions generated by ALSCAL.

Course	
Educational Psychology 311	
Pretest	
2-dimensional solution	stress = .300
3-dimensional solution	stress = .203
4-dimensional solution	stress = .151
Posttest	
2-dimensional solution	stress = .295
3-dimensional solution	stress = .189
4-dimensional solution	stress = .143
Physics 151	
Pretest	
2-dimensional solution	stress = .235
3-dimensional solution	stress = .134
Posttest	
2-dimensional solution	stress = .209
3-dimensional solution	stress = .120
Women's Studies 151	
Pretest	
2-dimensional solution	stress = .250
3-dimensional solution	stress = .182
4-dimensional solution	stress = .105
Posttest	
2-dimensional solution	stress = .273
3-dimensional solution	stress = .151
4-dimensional solution	stress = .129

After analyzing stress levels it was determined that a 3-dimensional solution would be acceptable for the pretest and posttest interpretations of all three courses. Thus all data reported was derived from 3-dimensional solutions.

Since instructors clustered terms at the beginning and end of the semester, it was necessary to look at both the pretest and posttest clustering of terms by each instructor to determine which of the instructor grouping (pretest or posttest) should be used as a criterion for assigning terms to groups. Only the Educational Psychology instructor made no changes in clustering from pretest to posttest, thus both pretest and posttest clustering can be reported in one table (see Table 4.2). However, both the Physics and Women's Studies instructors did change their clustering of terms from the pretest to the posttest. Therefore, the Physics and Women's Studies instructors' pretest and posttest clusters are reported in two separate tables (see Tables 4.3, 4.4, 4.5, and 4.6). It should be noted that Tables 4.2 through 4.6 also report the instructors' description of each group, as well as the terms belonging to that group.

Table 4.2
Educational Psychology instructor's pretest and posttest groupings of important course terms.

Group 1: Cognitive concepts
- A-ha experience
- Accommodation
- Decentration
- Encoding
- Equilibrium
- Gestalt
- Metacognition
- Schema
- Working memory

Group 2: Psychometric concepts
- Correlation
- Normal distribution
- Reliability
- Standard deviation
- Validity

Group 3: Behaviorist concepts
- Aversive control
- Law of effect
- Operant conditioning
- Shaping

Group 4: Psychosocial development concepts
- Epigenesis
- Identity diffusion
- Moratorium

Group 5: Humanistic concepts
- Phenomenology
- Self-actualization
- Third force psychology

Group 6: Social learning concepts
- Reciprocal determinism
- Self-efficacy
- Vicarious reinforcement

Group 7: Motivation theories
- Drive theory
- Locus of control
- Need for achievement

Table 4.3
Physics instructor's pretest groupings of important course terms.

Group 1: Newton's 2nd Law
- Acceleration
- Action-at-a-distance forces
- Action-reaction forces
- Contact forces
- Force
- Inertia
- Mass
- Weight

Group 2: Properties that waves exhibit
- Diffraction
- Interference
- Matter waves*
- Polarization

Group 3: Momentum and its consequences
- Angular momentum*
- Impulse
- Linear momentum
- Momentum conservation

Group 4: Energy
- Conservation of energy
- Forms of energy
- Work

Group 5: Order
- Absolute temperature scale
- Order at the molecular level
- Second law of thermodynamics

Group 6: Disorder
- Chaos*
- Entropy*

Group 7: Fluids
- Bernoulli's Principle
- Buoyancy

Table 4.3 (Continued)
<u>Physics instructor's pretest groupings of important course terms.</u>

Group 8: Inertial reference frames are needed to describe electromagnetic waves
Electromagnetic waves*
Inertial reference frames*

Group 9: Density
Density*

Group 10: Ideal Gas Laws
Ideal Gas Laws

<u>Note:</u> * Indicates terms which the instructor grouped differently in the pretest and posttest.

Table 4.4
Physics instructor's posttest groupings of important course terms.

Group 1: All Newton's Law of Motion
- Acceleration
- Action-at-a-distance forces
- Action-reaction forces
- Contact forces
- Force
- Inertia
- Inertial reference frames*
- Mass
- Weight

Group 2: Light
- Diffraction
- Interference
- Polarization

Group 3: Relates to momentum
- Impulse
- Linear momentum
- Momentum conservation

Group 4: Energy
- Conservation of energy
- Forms of energy
- Work

Group 5: 2nd Law of Thermodynamics
- Absolute temperature scale
- Chaos*
- Entropy*
- Order at the molecular level
- Second law of thermodynamics

Group 6: Fluids
- Bernoulli's Principle
- Buoyancy
- Density*

Group 7: Electromagnetic waves
- Electromagnetic waves*

Group 8: Angular momentum
- Angular momentum*

Table 4.4 (Continued)
<u>Physics instructor's posttest groupings of important course terms.</u>

Group 9: Matter waves
Matter waves*

Group 10: Ideal Gas Laws
Ideal Gas Laws

<u>Note:</u> * Indicates terms which the instructor grouped differently in the pretest and posttest.

Table 4.5
<u>Women's Studies instructor's pretest groupings of important course terms.</u>

Group 1: These are structures, or social constructs, which produce systems of dominance/submission
- Capitalism
- Heterosexism*
- Patriarchy
- Phallocentrism
- Racism*
- Sexism*

Group 2: Both a method and metaphor for "looking at the world", which is produced through and validates Gr. 1
- Male gaze
- Surveillance

Group 3: Category of identity produced by Groups 1,2
- Difference *
- The other *

Group 4: These are strategies of Group 1
- Homophobia*
- Mystification*
- Oppression*
- Silence*

Group 5: Challenges to Group 1, to patriarchy.
- Feminism
- Gynocentrism
- Humanism

Group 6: Ideas that Group 1 has "owned" and that Group 5 rescripts, rewrites, takes!
- Knowledge
- Perspective*
- Power
- Subjectivity*

Group 7: The ways of seeing the world produced by and driving the movement in Group 5.
- Female gaze
- Female point-of-view*
- Gender*
- Multiplicity

Group 8: The strategies, the "activities", the possibilities produced through Group 5.

Consciousness raising
Empowerment
Narrative
Rescripting
Voice

<u>Note:</u> * Indicates terms which the instructor grouped differently in the pretest and posttest.

Table 4.6
Women's Studies instructor's posttest groupings of important course terms.

Group 1: Systems of power in a culture of dominance
- Capitalism
- Patriarchy
- Phallocentrism

Group 2: Challenges to Group 1
- Feminism
- Gender*
- Gynocentrism
- Humanism

Group 3: These are produced within the social conditions of Group 1
- Heterosexism*
- Homophobia*
- Racism*
- Sexism*
- Silence*
- The other*

Group 4: Strategies of those forms of dominance in Group 1
- Male gaze
- Mystification*
- Oppression*
- Surveillance

Group 5: Produced through Group 2
- Difference*
- Female gaze
- Multiplicity
- Perspective*

Group 6: Strategies of feminism
- Consciousness raising
- Empowerment
- Female point-of-view*
- Narrative
- Rescripting
- Subjectivity*
- Voice

Table 4.6 (Continued)
Women's Studies instructor's posttest groupings of important course terms.

Group 7: These are contended for by both Groups 1 and 2, but mean very different things by both.

Knowledge
Power

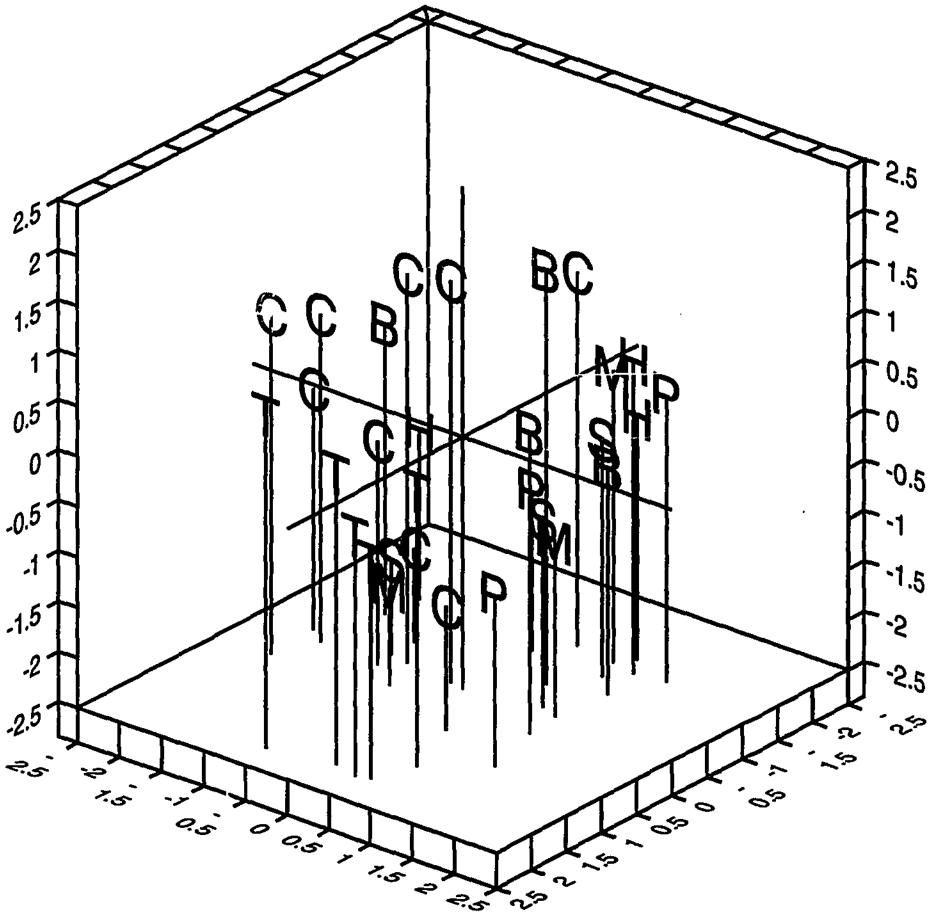
Note: * Indicates terms which the instructor grouped differently in the pretest and posttest.

The number of terms changed from pretest to posttest by the instructor in each of the three course was quite varied. As mentioned previously, the Educational Psychology instructor made no changes. The Physics instructor changed 7 of 30 terms (or 23.3% of terms) from pretest to posttest while the Women's Studies instructor changed 13 of 30 terms (43.3%) from pretest to posttest. It should be noted that, as was the case with the student clusters, the instructors did not have access to their pretest clusters while making their posttest groups.

Because of the variability in the Physics and Women's Studies instructors' responses, a judgment was made about which set of instructor clusters (pretest groups or posttest groups) to use as a criterion for comparison with student clusters. Because the instructors' posttest clusters were completed at the end of the semester, it was assumed that the posttest clusters would be a more valid measure of how the course terms related at the completion of the course. Pretest clusters might be viewed as the instructors' initial judgments about the most useful way to describe relationships between concepts, while the posttest clusters would reflect the actual relationships between course concepts that did develop during the semester.

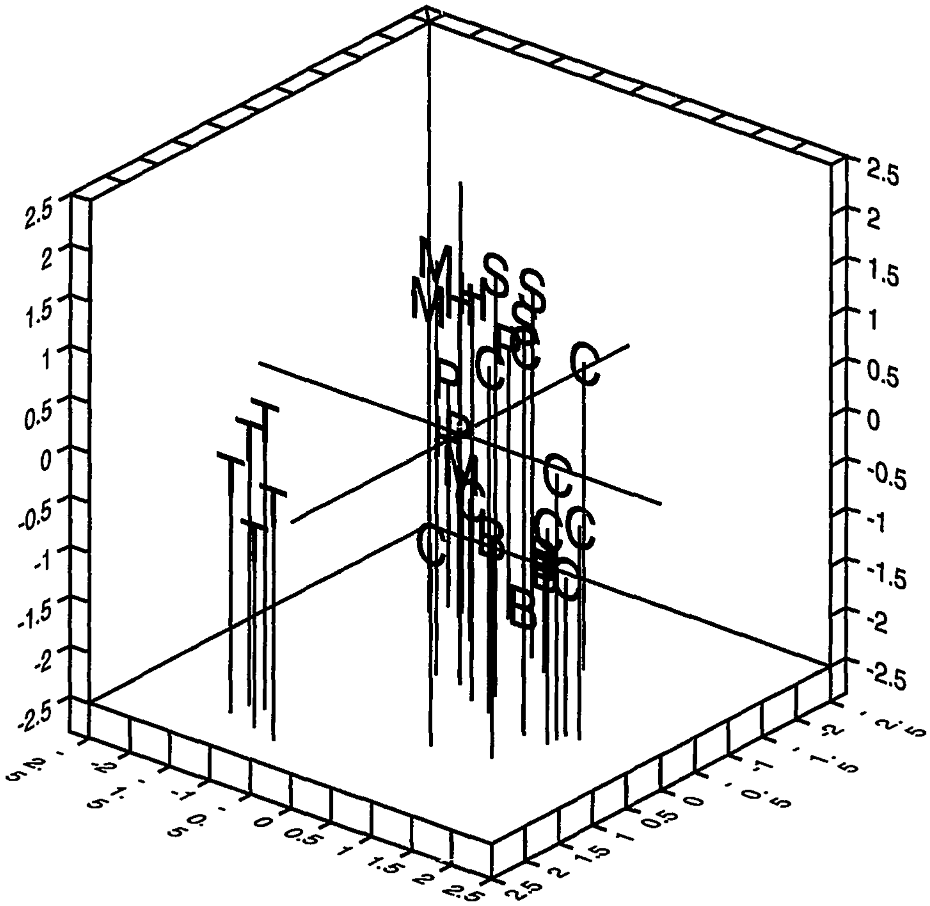
Plots of the pretest and posttest clustering of terms by students for each course are found in Figures 4.1, 4.2, 4.3, 4.4, 4.5 and 4.6. Because the symbols on the plots represent the categories the instructor used to group terms, Figures 4.1 through 4.6 can be used to visually assess the coherence of student clusters with regards to the instructors' groupings. This can be seen most dramatically in Figures 4.1 and 4.2 where the terms the instructor categorized as "psychometric terms" become much more closely clustered from the pretest to the posttest. (For plots of the individual 30 terms for each course, see Appendix IV).

Figure 4.1.
Pretest of students' clustering of important course terms in Educational Psychology.



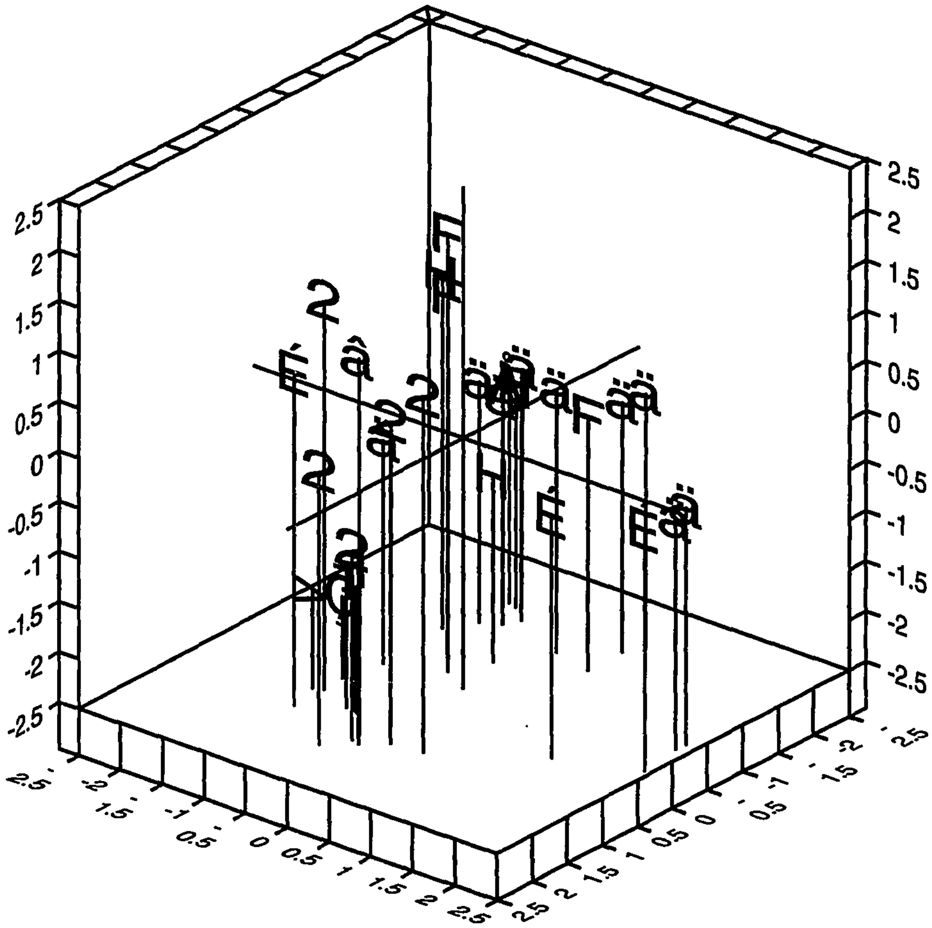
Legend
C = Cognitive terms (9 terms)
T = Psychometric terms (5 terms)
B = Behaviorist terms (4 terms)
P = Psychosocial terms (3 terms)
H = Humanistic terms (3 terms)
S = Social learning terms (3 terms)
M = Motivation terms (3 terms)

Figure 4.2.
Posttest of students clustering of important course terms in Educational Psychology.



Legend
C = Cognitive terms (9 terms)
T = Psychometric terms (5 terms)
B = Behaviorist terms (4 terms)
P = Psychosocial terms (3 terms)
H = Humanistic terms (3 terms)
S = Social learning terms (3 terms)
M = Motivation terms (3 terms)

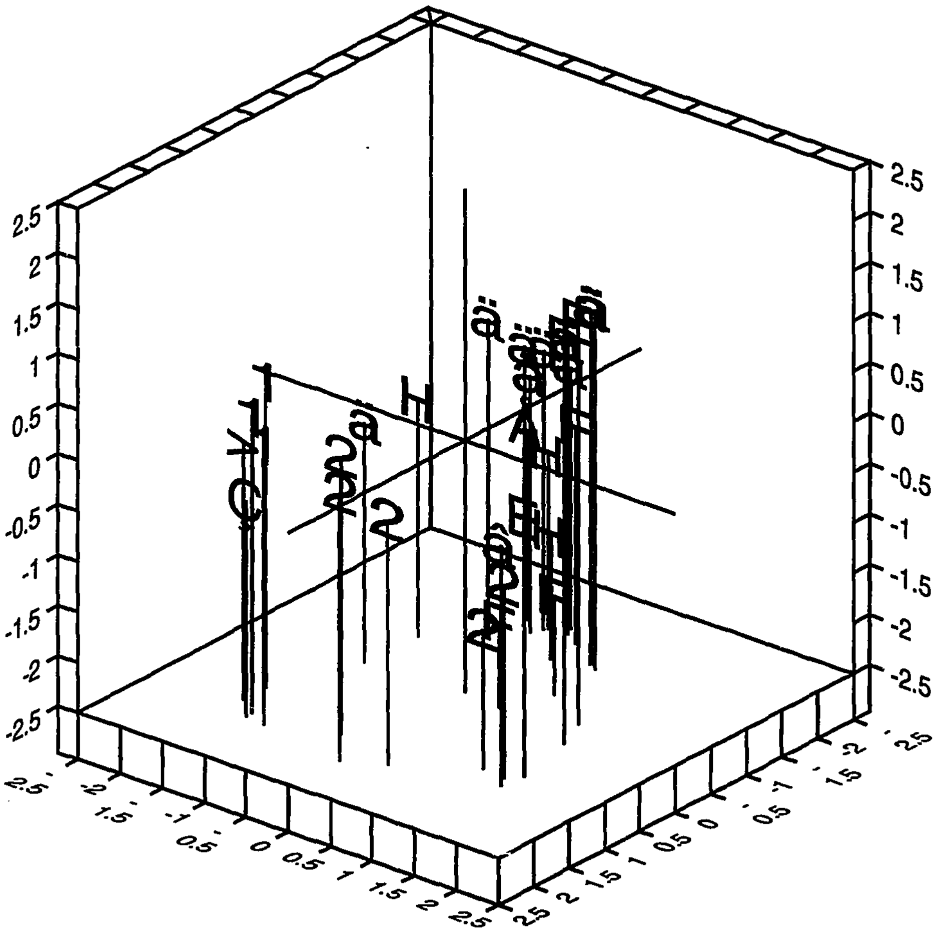
Figure 4.3.
Pretest of students' clustering of important course terms in Physics.



Legend

a = Newton's Second Law of Motion	(9 terms)
1 = Light	(3 terms)
H = Relates to Momentum	(3 terms)
F = Energy	(3 terms)
2 = Second Law of Thermodynamics	(5 terms)
E = Fluids	(3 terms)
C = Electromagnetic waves	(1 term)
Å = Angular Momentum	(1 term)
> = Matter Waves	(1 term)
å = Ideal Gas Laws	(1 term)

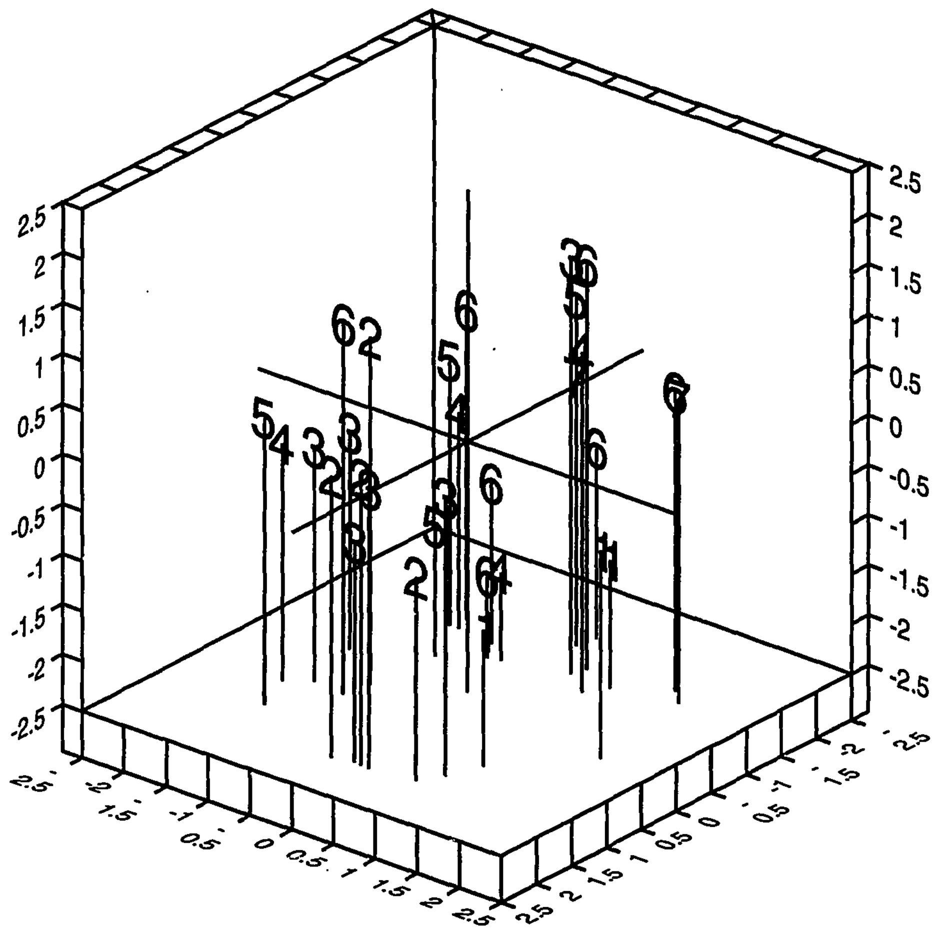
Figure 4.4.
Posttest of students' clustering of important course terms in Physics.



Legend

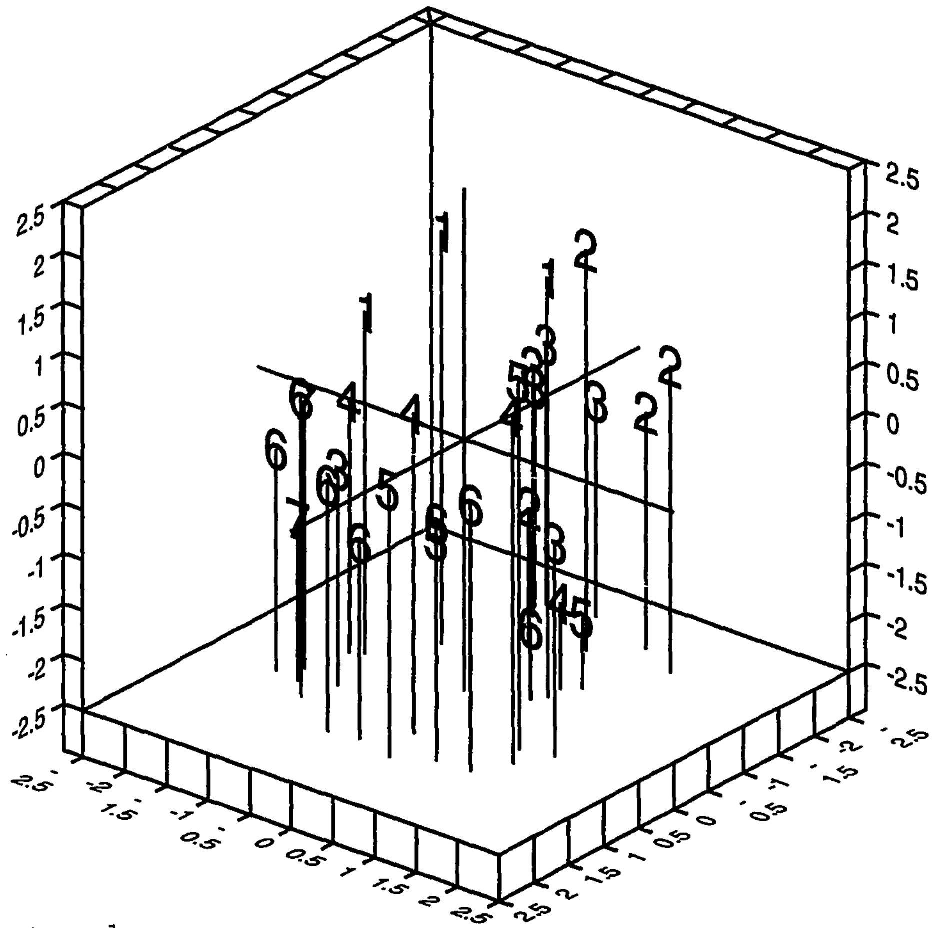
a = Newton's Second Law of Motion	(9 terms)
1 = Light	(3 terms)
H = Relates to Momentum	(3 terms)
F = Energy	(3 terms)
2 = Second Law of Thermodynamics	(5 terms)
E = Fluids	(3 terms)
C = Electromagnetic waves	(1 term)
Å = Angular Momentum	(1 term)
> = Matter Waves	(1 term)
å = Ideal Gas Laws	(1 term)

Figure 4.5.
Pretest of students' clustering of important course terms in Women's Studies.



Legend
1 = Gr. 1: Systems of power in a culture of dominance
(3 terms)
2 = Gr. 2: Challenges to Group 1
(4 terms)
3 = Gr. 3: Produced within the social conditions of Group1
(6 terms)
4 = Gr. 4: Strategies of those forms of dominance in Group1
(4 terms)
5 = Gr. 5: Produced through Group2
(4 terms)
6 = Gr. 6: Strategies of feminism
(7 terms)
7 = Gr. 7: Contended for by both Group 1 and Gr2, but mean
very different things by both
(2 terms)

Figure 4.6.
Posttest of students' clustering of important course terms in Women's Studies.



Legend
1 = Gr. 1: Systems of power in a culture of dominance
(3 terms)
2 = Gr. 2: Challenges to Group 1
(4 terms)
3 = Gr. 3: Produced within the social conditions of Group1
(6 terms)
4 = Gr. 4: Strategies of those forms of dominance in Group1
(4 terms)
5 = Gr. 5: Produced through Group2
(4 terms)
6 = Gr. 6: Strategies of feminism
(7 terms)
7 = Gr. 7: Contended for by both Group 1 and Gr2, but mean very different things by both
(2 terms)

Since the visual assessment of plots is much more difficult when a dramatic shift is not apparent, a quantitative measure of coherence was also calculated. As described in Chapter III, coherence of groups was determined by calculating the average distance of points in a group from the group centroid. Clusters which were tighter or more coherent should have a smaller average distance than will less coherent groups. The results are summarized in Table 4.7, 4.8 and 4.9.

The average distance of terms from the group centroid is smaller in the posttest than in the pretest for most of the groups. Because the normality of the measure of coherence could not be assured, nonparametric tests of significance were judged to be more appropriate in this case. Thus a binomial test was used to test the significance of the difference in the average distance of groups from pretest to posttest.

Table 4.7

Average distances of terms from the group centroid and variances of those distances for the pretest and posttest in Educational Psychology 311.

Group 1: Cognitive terms (number of terms in group = 9)	
Average distance from centroid of group	
Pretest	1.387
Posttest	1.126
Variance of distance	
Pretest	0.235
Posttest	0.419
Group 2: Psychometric terms (number of terms in group = 5)	
Average distance from centroid of group	
Pretest	0.586
Posttest	0.456
Variance of distance	
Pretest	0.080
Posttest	0.013
Group 3: Behaviorist terms (number of terms in group = 4)	
Average distance from centroid of group	
Pretest	1.337
Posttest	0.936
Variance of distance	
Pretest	0.115
Posttest	0.080
Group 4: Psychosocial terms (number of terms in group = 3)	
Average distance from centroid of group	
Pretest	1.349
Posttest	0.718
Variance of distance	
Pretest	0.028
Posttest	0.244
Group 5: Humanistic terms (number of terms in group = 3)	
Average distance from centroid of group	
Pretest	1.241
Posttest	1.304
Variance of distance	
Pretest	0.186
Posttest	0.240

Table 4.7 (Continued)
Average distances of terms from the group centroid and variances of those distances for the pretest and posttest in Educational Psychology 311.

Group 6: Social learning terms
(number of terms in group = 3)

Average distance from centroid of group	
Pretest	1.063
Posttest	0.676
Variance of distance	
Pretest	0.102
Posttest	0.132

Group 7: Motivation theory terms
(number of terms in group = 3)

Average distance from centroid of group	
Pretest	1.215
Posttest	1.199
Variance of distance	
Pretest	0.126
Posttest	0.226

Table 4.8
Average distances of terms from the group centroid and variances of those distances for the pretest and posttest in Physics 151.

Group 1: Newton's Law of Motion	
(number of terms in group = 9)	
Average distance from centroid of group	
Pretest	1.205
Posttest	0.660
Variance of distance	
Pretest	0.349
Posttest	0.193
Group 2: Light	
(number of terms in group = 3)	
Average distance from centroid of group	
Pretest	0.425
Posttest	0.319
Variance of distance	
Pretest	0.018
Posttest	0.020
Group 3: Relates to momentum	
(number of terms in group = 3)	
Average distance from centroid of group	
Pretest	0.897
Posttest	0.740
Variance of distance	
Pretest	0.026
Posttest	0.045
Group 4: Energy	
(number of terms in group = 3)	
Average distance from centroid of group	
Pretest	0.887
Posttest	0.821
Variance of distance	
Pretest	0.107
Posttest	0.091
Group 5: Second Law of Thermodynamics	
(number of terms in group = 5)	
Average distance from centroid of group	
Pretest	0.854
Posttest	0.829
Variance of distance	
Pretest	0.124
Posttest	0.082

Table 4.8 (Continued)
Average distances of terms from the group centroid and variances of those distances for the pretest and posttest in Physics 151.

Group 6: Fluids

Average distance from centroid of group	
(number of terms in group = 3)	
Pretest	1.434
Posttest	0.796
Variance of distance	
Pretest	0.372
Posttest	0.074

Group 7: Electromagnetic waves

(number of terms in group = 1)	
Average distance from centroid of group	
Pretest	not applicable
Posttest	not applicable
Variance of distance	
Pretest	not applicable
Posttest	not applicable

Group 8: Angular momentum

Average distance from centroid of group	
(number of terms in group = 1)	
Pretest	not applicable
Posttest	not applicable
Variance of distance	
Pretest	not applicable
Posttest	not applicable

Group 9: Matter waves

(number of terms in group = 1)	
Average distance from centroid of group	
Pretest	not applicable
Posttest	not applicable
Variance of distance	
Pretest	not applicable
Posttest	not applicable

Group 10: Ideal Gas Laws

(number of terms in group = 1)	
Average distance from centroid of group	
Pretest	not applicable
Posttest	not applicable
Variance of distance	
Pretest	not applicable
Posttest	not applicable

Table 4.9
Average distances of terms from the group centroid and variances of those distances for the pretest and posttest in Women's Studies 151.

Group 1: Systems of power in a culture of dominance
(number of terms in group = 3)

Average distance from centroid of group	
Pretest	1.055
Posttest	0.927
Variance of distance	
Pretest	0.082
Posttest	0.349

Group 2: These are challenges to Group 1
(number of terms in group = 4)

Average distance from centroid of group	
Pretest	0.908
Posttest	1.074
Variance of distance	
Pretest	0.156
Posttest	0.259

Group 3: These are produced within the social conditions of Group 1
(number of terms in group = 6)

Average distance from centroid of group	
Pretest	1.450
Posttest	1.297
Variance of distance	
Pretest	0.378
Posttest	0.505

Group 4: These are strategies of those forms of dominance in Group 1
(number of terms in group = 4)

Average distance from centroid of group	
Pretest	1.389
Posttest	1.519
Variance of distance	
Pretest	0.139
Posttest	0.129

Table 4.9 (Continued)
Average distances of terms from the group centroid and variances of those distances for the pretest and posttest in Women's Studies 151.

Group 5: These are produced through Group 2
(number of terms in group = 4)

Average distance from centroid of group	
Pretest	1.359
Posttest	1.279
Variance of distance	
Pretest	0.176
Posttest	0.316

Group 6: Strategies of feminism
(number of terms in group = 7)

Average distance from centroid of group	
Pretest	1.476
Posttest	1.182
Variance of distance	
Pretest	0.439
Posttest	0.455

Group 7: The are contended for by both Groups 1 and 2, but mean very different things by both.
(number of terms in group = 2)

Average distance from centroid of group	
Pretest	0.913
Posttest	0.547
Variance of distance	
Pretest	0
Posttest	0

Using the binomial test we can calculate the probability of the pretest distances being larger than the posttest distance by chance alone. These values are listed in Table 4.10.

Table 4.10
Binomial probabilities of the pretest and posttest distances from the centroid of assigned groups.

Course	Total number of clusters (N)	Number of clusters where average pretest distance is larger than average posttest distance	Binomial Probability
Educational Psychology	7	6	.055
Physics	6	6	.016
Women's Studies	7	5	.164
All courses combined	20	17	.001

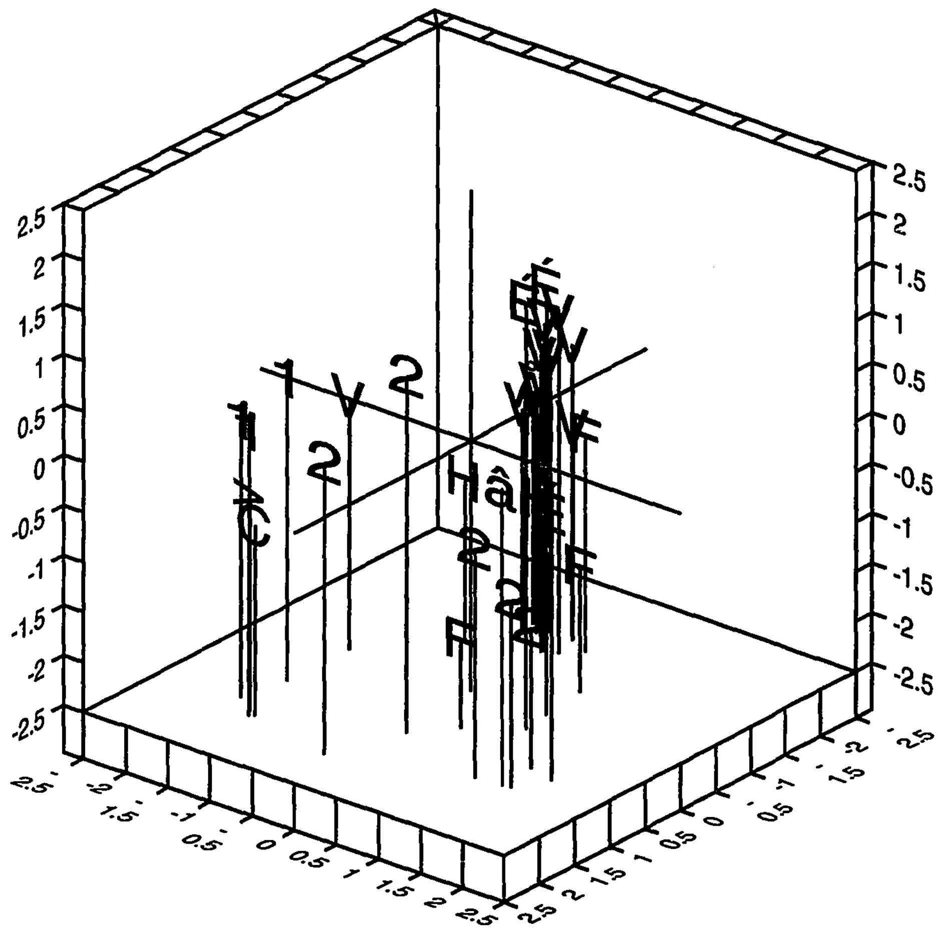
At the p = .05 level of significance, the results of Physics are considered significant, and the results of Educational Psychology closely approach significance. However, the binomial probability for the Women's Studies course is not significant at the p = .05 level. If the results of all three courses are combined the results are highly statistically significant.

Hypothesis 2

Hypothesis 2 posits that after instruction, the cognitive structure of students with high course grades is more coherent than the cognitive structure of students with low course grades.

As discussed in Chapter III, only the data from the Physics course was used to test this hypothesis because the students in Physics had the widest grade distribution. The clustering of a combined group of A and B students and a combined group of D and F students were analyzed using PEROVER and ALSCAL. The results of these analyses are plotted in Figures 4.7 and 4.8. These plots can visually analyzed to assess the similarity between the students clusters and the instructor's groupings. Plots which list individual course terms are in Appendix V.

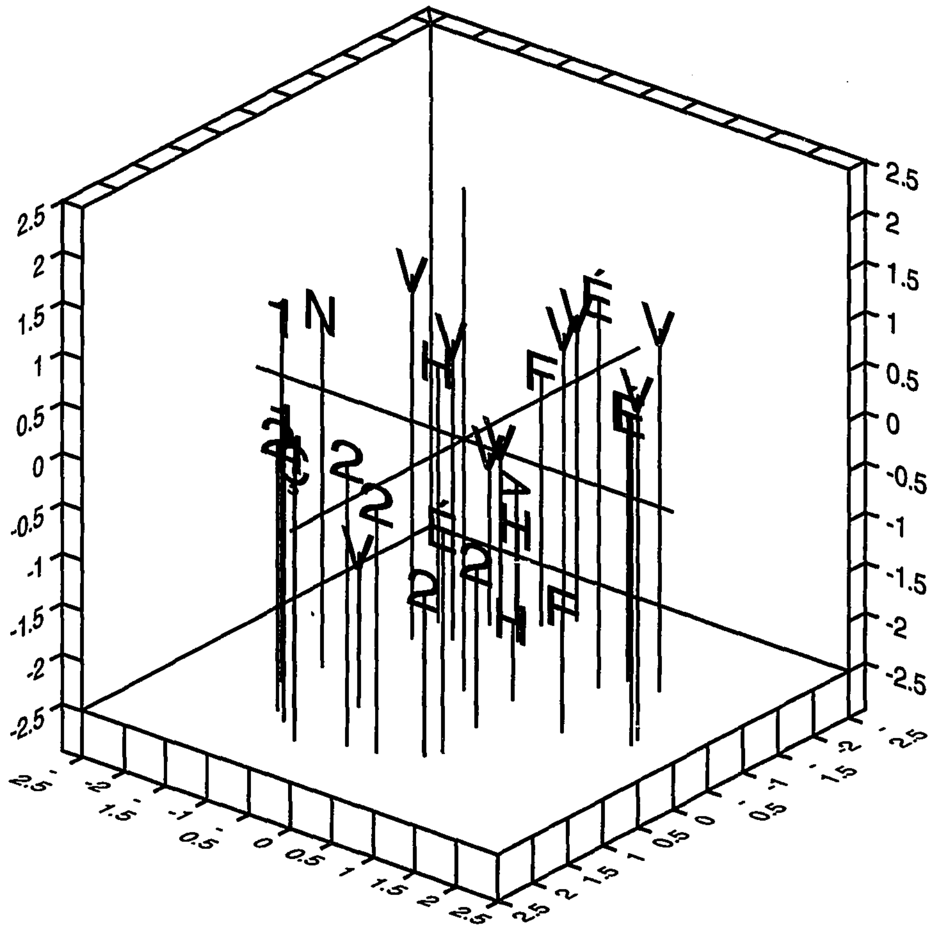
Figure 4.7.
Posttest of A and B students' clustering of important course terms in Physics.



Legend

a = Newton's Second Law of Motion (9 terms)
1 = Light (3 terms)
H = Relates to Momentum (3 terms)
F = Energy (3 terms)
2 = Second Law of Thermodynamics (5 terms)
E = Fluids (3 terms)
C = Electromagnetic waves (1 term)
Å = Angular Momentum (1 term)
> = Matter Waves (1 term)
å = Ideal Gas Laws (1 term)

Figure 4.8.
Posttest of D and F students' clustering of
important course terms in Physics.



Legend
a = Newton's Second Law of Motion (9 terms)
1 = Light (3 terms)
H = Relates to Momentum (3 terms)
F = Energy (3 terms)
2 = Second Law of Thermodynamics (5 terms)
E = Fluids (3 terms)
C = Electromagnetic waves (1 term)
Å = Angular Momentum (1 term)
> = Matter Waves (1 term)
å = Ideal Gas Laws (1 term)

The average distance of terms in each group from the centroid of that group was determined and the results for A and B and D and F students are listed in Table 4.11.

Hypothesis 2 predicts that the better performing students (here represented by students receiving grades of A and B) would have more coherent clusters than students who do not perform well in the course (represented by students receiving grades of D and F). Therefore in this case, a smaller average distance from the centroid for the A and B students would be considered a success. In five of the six groups, the average distance of the terms from the group centroid was smaller for the A and B students than for the D and F students.

The binomial test was again used in this comparison. The results of the binomial test were in the hypothesized direction but failed to reach statistical significance ($\underline{B}$ = .093). Therefore, the hypothesis was not supported.

Table: 4.11
<u>Average distances and variances of terms from the group centroid for the posttest in Physics 151; A and</u> B <u>students and D and F students</u>.

Group 1: Newton's Law of Motion
(number of terms in group = 9)

Average distance from centroid of group	
A and B students	0.621
D and F students	1.404
Variance of distance	
A and B students	0.169
D and F students	0.260

Group 2: Light
(number of terms in group = 3)

Average distance from centroid of group	
A and B students	0.351
D and F students	0.552
Variance of distance	
A and B students	0.012
D and F students	0.057

Group 3: Relates to momentum
(number of terms in group = 3)

Average distance from centroid of group	
A and B students	0.459
D and F students	1.037
Variance of distance	
A and B students	0.001
D and F students	0.139

Group 4: Energy
(number of terms in group = 3)

Average distance from centroid of group	
A and B students	0.922
D and F students	1.139
Variance of distance	
A and B students	0.140
D and F students	0.102

Group 5: 2nd Law of Thermodynamics
(number of terms in group = 5)

Average distance from centroid of group	
A and B students	1.050
D and F students	0.842
Variance of distance	
A and B students	0.132
D and F students	0.009

Table: 4.11 (Continued)

Average distances and variances of terms from the group centroid for the posttest in Physics 151; A and B students and D and F students.

Group 6: Fluids
(number of terms in group = 3)

Average distance from centroid of group	
A and B students	0.904
D and F students	1.152
Variance of distance	
A and B students	0.100
D and F students	0.078

Group 7: Electromagnetic waves
(number of terms in group = 1)

Average distance from centroid of group	
A and B students	not applicable
D and F students	not applicable
Variance of distance	
A and B students	not applicable
D and F students	not applicable

Group 8: Angular momentum
(number of terms in group = 1)

Average distance from centroid of group	
A and B students	not applicable
D and F students	not applicable
Variance of distance	
A and B students	not applicable
D and F students	not applicable

Group 9: Matter waves
(number of terms in group = 1)

Average distance from centroid of group	
A and B students	not applicable
D and F students	not applicable
Variance of distance	
A and B students	not applicable
D and F students	not applicable

Group 10: Ideal Gas Laws
(number of terms in group = 1)

Average distance from centroid of group	
A and B students	not applicable
D and F students	not applicable
Variance of distance	
A and B students	not applicable
D and F students	not applicable

The results of the binomial test led to the creation of an additional questions. Might students receiving higher grades have more distinct groups than students who received lower grades? To answer this second question, the average geometric distance of each group centroid from every other group centroid was calculated. Groups which are more distinct from each other would have a greater average distance from other groups than groups which are less distinct.

If students receiving higher grades have more distinct groups than students with lower grades, the average distance of each group centroid from every other group centroid should be larger for students receiving grades of A and B than for students receiving grades of D and F. As shown in Table 4.12, the average distance from other centroids is larger for A and B students in six of six cases. Using the binomial, this is a statistically significant difference (B = .016) supporting the idea that students with higher grades do have more distinct groups.

Table 4.12
Average distance of group centroid from every other group centroid for students receiving grades A and B, and D and F in Physics 151.

Group 1: All Newton's Laws of Motion

Grades	Average distance of group centroid from every other group centroid
A and B	2.099
D and F	1.714

Group 2: Light

Grades	Average distance of group centroid from every other group centroid
A and B	2.747
D and F	2.407

Group 3: Relates to Momentum

Grades	Average distance of group centroid from every other group centroid
A and B	2.125
D and F	1.954

Group 4: Energy

Grades	Average distance of group centroid from every other group centroid
A and B	2.064
D and F	1.741

Group 5: 2nd Law of Thermodynamics

Grades	Average distance of group centroid from every other group centroid
A and B	2.290
D and F	1.956

Group 6: Fluids

Grades	Average distance of group centroid from every other group centroid
A and B	2.481
D and F	1.976

Note: Groups 7 through 10 are not listed because each group contained only one term.

Hypothesis 3

Hypothesis 3 posits that in both the pretest and posttest, students' clustering of more familiar terms will be more similar to instructors' clusters than will student clustering of less familiar terms.

The mean ratings of familiarity at the pretest and posttest were assumed to measure the students' familiarity with terms. Despite the fact that these are self-report measures, results are consistent with what would logically be expected from a true measure of familiarity. In other words, average familiarity of course terms increased from the pre- to posttest. Pretest and posttest familiarity ratings were also highly correlated ($r(28) = .847$, $p < .001$ for Educational Psychology, $r(28) = .834$, $p < .001$ for Physics, and $r(28) = .820$, $p < .001$ for Women's Studies). Pretest and posttest familiarity ratings for all course terms are listed in Tables 4.13 through 4.15. (See Appendix VI for listings of the pretest and posttest familiarity ratings sorted in descending order by familiarity.)

Table 4.13
Pretest and posttest familiarity ratings for Educational Psychology terms.

Term	Pretest rating	Posttest rating
A-ha experience	2.929	4.214
Accommodation	3.286	4.214
Aversive control	1.714	3.143
Correlation	3.286	4.214
Decentration	1.643	2.714
Drive theory	1.429	2.643
Encoding	2.786	3.500
Epigenesis	1.286	2.000
Equilibrium	3.071	3.929
Gestalt	1.929	4.000
Identity diffusion	1.857	2.571
Law of Effect	1.929	3.429
Locus of control	1.214	2.643
Metacognition	1.929	3.143
Moratorium	2.214	2.357
Need for achievement	3.500	3.857
Normal distribution	2.786	4.000
Operant conditioning	2.571	4.214
Phenomenology	1.643	2.857
Reciprocal determinism	1.429	2.786
Reliability	3.214	4.286
Schema	2.786	4.537
Self-actualization	3.500	4.429
Self-efficacy	2.786	4.143
Shaping	2.500	4.000
Standard deviation	3.071	4.500
Third-force psychology	1.143	2.500
Validity	3.286	4.143
Vicarious reinforcement	2.000	3.643
Working memory	2.571	3.500
Mean familiarity rating	2.376	3.543
Variance of familiarity ratings	0.557	0.546

Table 4.14
Pretest and posttest familiarity ratings for Physics terms.

Term	Pretest rating	Posttest rating
Absolute temperature scale	2.568	4.027
Acceleration	3.230	4.432
Action-at-a-distance forces	1.622	2.792
Action-reaction forces	2.622	3.892
Angular momentum	2.149	3.541
Bernoulli's Principle	1.446	3.233
Buoyancy	2.446	3.878
Chaos	1.716	2.987
Conservation of energy	2.959	4.311
Contact forces	1.986	3.405
Density	3.351	4.392
Diffraction	2.284	2.770
Electromagnetic waves	2.176	2.548
Entropy	2.324	3.230
Force	3.122	4.324
Forms of energy	2.838	3.986
Ideal Gas Laws	2.384	3.784
Impulse	1.838	3.284
Inertia	2.581	3.703
Inertial reference frames	1.554	2.370
Interference	1.892	2.649
Linear momentum	2.189	3.973
Mass	3.446	4.581
Matter waves	1.671	2.554
Momentum conservation	1.851	3.838
Order at the molecular level	1.784	2.757
Polarization	2.297	2.662
Second Law of Thermodynamics	1.973	3.284
Weight	3.216	4.568
Work	3.068	4.311
Mean familiarity rating	2.352	3.352
Variance of familiarity ratings	0.341	0.474

Table 4.15
Pretest and posttest familiarity ratings for Women's Studies terms.

Terms	Pretest rating	Posttest rating
Capitalism	3.429	3.571
Consciousness raising	2.643	3.500
Difference	3.143	4.000
Empowerment	2.571	3.786
Female gaze	1.357	2.857
Female point-of-view	3.429	4.143
Feminism	3.143	4.286
Gender	4.214	4.571
Gynocentrism	1.429	3.643
Heterosexism	2.786	3.857
Homophobia	3.571	4.071
Humanism	2.714	3.429
Knowledge	3.429	4.214
Male gaze	1.286	2.926
Multiplicity	1.429	2.714
Mystification	1.429	3.357
Narrative	2.786	3.714
Oppression	3.214	4.143
Patriarchy	2.357	3.857
Perspective	3.929	4.071
Phallocentrism	1.500	3.000
Power	3.929	4.714
Racism	4.071	4.714
Rescripting	1.071	2.857
Sexism	3.714	4.286
Silence	3.571	4.429
Subjectivity	3.071	3.357
Surveillance	2.857	3.143
The other	2.429	3.214
Voice	2.857	4.000
Mean familiarity rating	2.779	3.746
Variance of familiarity ratings	0.861	0.328

Table 4.16
Pretest and posttest distances of terms from the group centroid in Educational Psychology.

Term	Pretest distance	Posttest distance
A-ha experience	1.895	1.363
Accommodation	1.323	0.702
Aversive control	1.197	0.936
Correlation	1.063	0.271
Decentration	2.189	2.076
Drive theory	1.434	0.526
Encoding	0.965	1.057
Epigenesis	1.278	0.284
Equilibrium	1.394	1.194
Gestalt	1.916	0.790
Identity diffusion	1.187	0.579
Law of Effect	1.121	0.860
Locus of control	0.715	1.535
Metacognition	0.590	1.305
Moratorium	1.581	0.453
Need for achievement	1.496	1.535
Normal distribution	0.626	0.515
Operant conditioning	1.098	0.580
Phenomenology	1.152	1.503
Reciprocal determinism	1.039	0.898
Reliability	0.366	0.511
Schema	1.097	0.596
Self-actualization	1.809	1.779
Self-efficacy	1.466	0.968
Shaping	1.918	1.367
Standard deviation	0.237	0.593
Third-force psychology	0.763	0.630
Validity	0.639	0.391
Vicarious reinforcement	0.684	0.163
Working memory	1.118	1.050
Average distance	1.179	0.900
Variances of distances	0.232	0.234

Table 4.17
Pretest and posttest distances of terms from the group centroid in Physics.

Term	Pretest distance	Posttest distance
Absolute temperature scale	0.858	0.971
Acceleration	0.590	0.356
Action-at-a-distance forces	1.042	0.490
Action-reaction forces	1.027	0.572
Angular momentum	none	none
Bernoulli's Principle	2.142	1.163
Buoyancy	0.653	0.716
Chaos	1.171	1.017
Conservation of energy	0.865	0.453
Contact forces	1.207	0.097
Density	1.508	0.509
Diffraction	0.345	0.122
Electromagnetic waves	none	none
Entropy	0.379	0.298
Force	0.816	0.387
Forms of energy	0.497	0.819
Ideal Gas Laws	none	none
Impulse	0.934	0.983
Inertia	0.323	0.446
Inertial reference frames	1.656	1.642
Interference	0.314	0.461
Linear momentum	0.685	0.464
Mass	2.072	1.026
Matter waves	none	none
Momentum conservation	1.072	0.773
Order at the molecular level	0.555	0.773
Polarization	0.617	0.375
Second Law of Thermodynamics	1.307	1.088
Weight	2.110	0.923
Work	1.299	1.190
Average distance	1.002	0.697
Variance of distances	0.298	0.138

Table 4.18
Pretest and posttest distances of terms from the group centroid in Women's Studies.

Term	Pretest distance	Posttest distance
Capitalism	0.854	0.509
Consciousness raising	1.053	1.266
Difference	0.716	1.017
Empowerment	1.640	1.560
Female gaze	1.825	2.097
Female point-of-view	2.078	2.046
Feminism	0.516	0.332
Gender	1.393	1.203
Gynocentrism	1.200	1.007
Heterosexism	1.148	0.999
Homophobia	1.202	0.975
Humanism	0.522	1.755
Knowledge	0.913	0.547
Male gaze	1.618	1.933
Multiplicity	1.278	1.432
Mystification	1.389	0.944
Narrative	1.113	1.371
Oppression	1.761	1.610
Patriarchy	0.854	0.509
Perspective	1.617	0.570
Phallocentrism	1.459	1.762
Power	0.913	0.547
Racism	1.297	0.896
Rescripting	1.804	0.463
Sexism	1.104	0.717
Silence	2.288	1.823
Subjectivity	1.067	0.947
Surveillance	0.786	1.590
The other	1.660	2.372
Voice	1.581	0.619
Average distance	1.288	1.181
Variances of distances	0.195	0.325

If Hypothesis 3 is true, and more familiar terms are clustered more coherently, then one would predict that the familiarity ratings of terms would be negatively correlated with their distance from the centroid of the cluster. In other words, one would predict that a term with a high familiarity rating would have a low distance from the centroid. One might also expect that, since the posttest terms are more familiar to students, the correlation between posttest familiarity and posttest distances would be greater than the correlation between pretest familiarity and pretest distances. Pretest and posttest distances of terms from the centroid are listed in Tables 4.16 though 4.18. Correlations between the pretest and posttest distances in each course were $\underline{r}(28) = .427$, $\underline{p} = .005$ for Educational Psychology, $\underline{r}(28) = .607$, $\underline{p} = 0.001$ for Physics, and $\underline{r}(28) = .478$, $\underline{p} = .007$ for Women's Studies.

For each of the three courses, the correlation between the pretest familiarity rating of a term and the pretest distance of that term from the group centroid were calculated. The correlation for the Educational Psychology terms was $\underline{r}(28) = -.017$, $\underline{p} = .928$. The correlation for the Physics terms was $\underline{r}(28) = .074$,

$p = .721$. And a correlation of $r(28) = -.170$, $p = .369$ was found between the pretest familiarity ratings and pretest average for Women's Studies terms.

The correlation between the posttest familiarity rating of a term and the posttest distance of that term from the group centroid was also calculated for all three courses. In Educational Psychology, the correlation between the posttest familiarity ratings and posttest average distance of terms from the group centroid was $r(28) = -.032$, $p = .866$. For Physics terms, the posttest correlation was $r(28) = -.043$, $p = .835$. The posttest correlation between familiarity ratings and average distance was $r(28) = -.432$, $p = .017$ for Women's Studies terms.

Only the correlation between the Women's Studies posttest familiarity rating and posttest average distance from the group centroid is significant at the $p = .05$ level.

Hypothesis 4

Hypothesis 4 posits that, after instruction, students use deeper level characteristics to categorize concepts than they do before instruction.

In this study inter-rater reliability was quite high (99.5%). The number of deep and surface explanations given by students of why terms belonged together are listed in Tables 4.19 through 4.21.

The average number of 'deep and surface explanations was compared from pretest and posttests in each of the three courses. Only one significant difference was found. There was a significant decrease from pretest to posttest in surface categorizations in the Women's Studies course ($\underline{t}(13) = 2.242$, $\underline{p} = .05$).

Table 4.19
Categorization of Educational Psychology students' explanations of why terms within a group belonged together.

	Average number of student cluster explanations Percent of total in parentheses		
	Deep	Surface	Total
Pretest	2.786 (53.4%)	2.428 (46.6%)	5.214
Posttest	2.615 (48.6%)	2.769 (51.4%)	5.385

Note: Number of students = 14

Table 4.20
Categorization of Physics students' explanations of why terms within a group belonged together.

	Average number of student cluster explanations Percent of total in parentheses Deep	Surface	Total
Pretest	5.459 (81.4%)	1.243 (18.6%)	6.703
Posttest	5.392 (80.8%)	1.270 (19.0%)	6.676

Note: Number of students = 74

Table 4.21
Categorization of Women's Studies students' explanations of why terms within a group belonged together.

	Average number of student cluster explanations Percent of total in parentheses Deep	Surface	Total
Pretest	4.214 (67.8%)	2.000 (32.2%)	6.214
Posttest	4.857 (80.0%)	1.214 (20.0%)	6.071

Note: Number of students = 14

Hypothesis 5

Hypothesis 5 posits that, after instruction, students with high course grades would use deeper level characteristics or criteria to categorize concepts than students with low course grades.

Using the same rationale as presented for Hypothesis 2, only students who received a grade of A or B and D or F in Physics 151 who had complete both the pretest and posttest and who had given their permission for their grade to be shared were eligible for this part of the study.

The rater categorization of the combined A or B student and the combined D or F student explanations of why terms were grouped together for both the pretest and the posttest are reported in Tables 4.22 and 4.23. Significant differences were not found between A or B and D or F students either in the average number of surface or deep categorizations during the posttest.

Table 4.22
Categorization of A or B and D or F Physics students' pretest explanations of why terms within a group belonged together.

	Average number of student cluster explanations Percent of total in parentheses		
	Deep	Surface	Total
A or B students	5.000 (82.9%)	1.033 (17.1%)	6.033
D or F students	5.625 (82.6%)	1.187 (17.4%)	6.813

Note: Number of students in study receiving a grade of A or B = 30. Students in study receiving a D or F = 16.

Table 4.23
Categorization of A or B and D or F Physics students' posttest explanations of why terms within a group belonged together.

	Average number of student cluster explanations Percent of total in parentheses		
	Deep	Surface	Total
A or B students	5.400 (83.5%)	1.067 (16.5%)	6.467
D or F students	5.500 (77.9%)	1.500 (21.2%)	7.063

Note: Number of students in study receiving a grade of A or B = 30. Students in study receiving a D or F = 16.

CHAPTER V: DISCUSSION AND RECOMMENDATIONS

The purpose of this chapter is to discuss and interpret the finding of this study in light of relevant research literature. The chapter is divided into the following sections: interpretation of findings organized by research hypotheses, a comparison of results across three disciplines, limitations of the study, recommendations for future research, implications for education and conclusions.

Interpretation of Findings

Students' cognitive structures and instruction

Michelene Chi and her colleagues have suggested that the cognitive structures of experts are more coherent than the cognitive structures of novices (Chi & Koeske, 1983; Gobbo & Chi, 1986; Chi, Hutchinson & Robin, 1989). In other words, information stored in the cognitive structures of experts would be organized into coherent groups whose members have stronger within-group links than between-group links.

If students move closer to being experts after instruction, then one could expect that the cognitive structures of students will be more like the cognitive structures of experts after instruction. If cognitive structures of experts are more coherent than the cognitive structures of novices, then one could predict that the inferred cognitive structures of students after instruction are more coherent than those inferred before instruction.

Students' cognitive structures were measured through student pretest and posttest clustering of instructor-determined terms analyzed using multidimensional scaling. Terms were assigned to groups according to the results of the instructors' clustering of the same terms at the end of the semester. The geometric center or centroid of each of these groups was then determined and the distance of each term in the group from the group centroid was calculated. The average distance of terms in a group from the group centroid was used as a measure of coherence. Using this scheme more coherent groups would have a smaller average distance from the centroid than less coherent groups. The binomial probability of these results were determined. For the

Educational Psychology course the binomial probability was $\underline{B} = .055$, a result very close to the generally used $\underline{p} = .05$ cut-off for significance. The binomial probability of the Physics course results were $\underline{B} = .016$, which is significant at the $\underline{p} = .05$ level. For the Women's Studies course results, the binomial probability was $\underline{B} = .164$, which would not be considered significant at the $\underline{p} = .05$ level.

What accounts for these differences between courses? One explanation might lay in the structure of the three different disciplines represented. There is some evidence that the content of some disciplines is more hierarchically structured than the content of other disciplines. For example, Donald (1982, 1983, 1986) found that some sciences, like physics, were much more hierarchically structured than some humanities, for example, history. It has also been suggested that expert-novice differences in ill-structured domains might be different from expert-novice differences in well-structured domains (Glaser & Chi, 1988) and that the whole concept of "classification" in ill-structured domains should be viewed in a different light (Feltovich et al., 1992).

Extrapolating on the definition of ill-structured problems as problems for which there is no consensus about a correct solution among the community of scholars, (Reitman, 1965; Simon, 1973), an ill-structured domain might be defined as a domain for which there is no consensus among scholars in the field about correct relationships between concepts in the field. Also extrapolating on Reitman's (1965) and Simon's (1973) supposition that classification of problems as ill-structured versus well-structured is a continuum, one could also expect such a continuum to exist when attempting to classify well-structured and ill-structured domains.

Looking at the three domains represented by the courses in this study, one could argue that Physics is the most "structured", in the above sense. Women's Studies would be the least structured, with Educational Psychology falling between the other two. One argument for this kind of classification could be based on the age of the disciplines. Physics, a domain which is several hundred years old, might be expected to have the most solidified structure, at least at the introductory college level. Educational Psychology is the next oldest domain and Women's

Studies might be considered to be the youngest of the three domains. With these younger fields, it would be reasonable to expect that their structures would be more varied and malleable. This prediction might be investigated by looking at the topics covered in a variety of introductory textbooks in each field. One would expect topics covered in a more structured field to be quite similar throughout texts. In a less-structured field, topics covered in textbooks might be more varied.

One might also expect that there would be more agreement in the way experts in a more structured field cluster important course concepts. Therefore, it might be useful to repeat the same clustering task completed by the student participants of this study with several instructors within the same field. Although this variation of the task was beyond the scope of this study, each of the three instructors did complete the clustering task at the beginning and then at the end of the semester and a comparison of instructors' pretest and posttest clusters can give us some insight into the solidity of the structure of the domain. Greater variability of instructors' clusters from pretest to posttest might reflect greater the variability, and hence a less solid structure, within a domain.

A comparison of the individual instructor's variability in clustering is somewhat compatible with the idea that Physics is the most structured of the three domains, and Women's Studies is the least structured of the three domains. One would expect least variability in the Physics instructor's pre- and posttest clusterings and the most variability in the Women's Studies instructor's clusterings. The Women's Studies instructor (with 43% of terms changed) was the most variable, however, the Educational Psychology instructor (with no terms changed) was the least variable. It is reasonable to expect that personality differences also may come into play here and may have accounted for this result. It should also be noted that the Educational Psychology instructor was familiar with the clustering method and therefore may have been more mindful of which terms were placed into clusters than the other two instructors, who had never used the technique.

In explaining the results obtained in this study it is also useful to look more closely at the groups where the posttest average distance was NOT smaller than the pretest average distance. In the Women's Studies course there were two groups where the pretest average distance was smaller than the results obtained during the posttest. One group with a smaller pretest average distance consisted of the terms: "male gaze," "mystification," "oppression" and "surveillance." In the pretest plot of this group, the four terms which make up the group are scattered rather widely and are spread apart from each other. In the posttest three of the terms ("mystification," "oppression," and "surveillance") are close to each other on the plot, while the fourth term ("male gaze") is widely separated from the other three terms. Thus it seems that three of the four terms in this group do cluster closer together in the posttest.

In the posttest plot, "male gaze" is very closely plotted to its semantic parallel "female gaze" (see Appendix IV). Posttest familiarity ratings of the 30 terms revealed that "male gaze" and "female gaze" were rated as the 3rd and 4th LEAST familiar term (see Appendix VI). This relative lack of familiarity with the terms, even at the end of the course, could have contributed to "male gaze" and "female gaze" being clustered by students according to semantic similarities (Romney, et al., 1993).

It is also noted that in this group of four terms, the instructor herself changed the clustering of two of the four terms ("mystification" and "oppression") from the pretest to the posttest. The students variability of clustering of terms in this group could reflect the instructor's variability of clustering of these terms.

The second group whose pretest average distance from the centroid was smaller than the posttest difference consisted of the terms "feminism," "gender," "gynocentrism," and "humanism." The instructor also changed the clustering of one of the terms in the group, namely the term "gender". The multidimensional scaling plots of pretest and posttest proximities (Appendix IV) reveal that the term "gender" does appear to move away from the other three terms in that group in the posttest. Again, the student placement of the term could be reflective of the instructor's own ambiguity about the relationship of the term to other terms in the field.

A portion of the small pretest average distance for this group could also be accounted for by the very close clustering of the terms "feminism" and "humanism" on the pretest. However, a third term "sexism" is also extremely closely linked to "feminism" and "humanism" on the pretest. Since the instructor would consider "sexism" to have a meaning very contrary to the terms "feminism" and "humanism," students may have clustered these three terms together in the pretest because of their structural similarity (all ending in "ism"). To support this argument it can also be noted the another structurally similar term, "heterosexism" (one that the instructor again would consider a term opposing "feminism" and "humanism") was also clustered close by "feminism," "humanism," and "sexism."

In the Educational Psychology course, only one group of terms did not have a smaller posttest average distance. That group, which the instructor defined as being a group of "humanistic concepts," was composed of three terms: "phenomenology," "self-actualization," and "third force psychology." Although there was not a very great difference between the average distances from the group centroid in the pretest and posttest (pretest average distance = 1.241, posttest average distance = 1.304) it is nonetheless interesting to speculate why this group did not follow the trend of a smaller posttest average distance that was exhibited by the other groups. In both the pretest and posttest, the three terms comprising the group were not equidistant. In the pretest, the terms "phenomenology" and "third force psychology" were clustered close together, with the term "self-actualization" at a greater distance from these two (see Appendix IV). In the posttest, the terms "self-actualization" and "third force psychology" were clustered very closely together, with the term "phenomenology" at a greater distance.

An examination of the plot of the pretest shows that the terms "phenomenology" and "third force psychology" seem to be part of a larger local group consisting of the additional terms: "aversive control," "drive theory," "epigenesis," and "reciprocal determinism." It is interesting to note that these six terms all fall within the bottom quartile of the pretest familiarity rating of terms. Using nearest neighbor analysis (Kruskal & Wish, 1973) one can speculate that "phenomenology" and "third-force psychology" are clustered closely in the pretest because they fall into a larger group of less familiar terms that students might have clustered together in an "I don't know" group. The contribution of the proximity of these two terms to the overall low pretest average distance might therefore be spurious.

In summary, although the hypothesis that students' cognitive structures become more coherent after instruction is supported, other influences such as structural relatedness of terms, student familiarity with terms, and the variability of the instructors' cognitive structure may also play a part in determining students' cognitive structures.

Students' cognitive structures and course achievement

Do students who receive high course grades have more coherent cognitive structures than students who receive low course grades? There is prior evidence that students with higher course grades have cognitive structures closer to that of the instructor than students with lower course grades (Naveh-Benjamin et al., 1986, 1989). Unfortunately, the grade distribution of the three courses was such that data from only one course, Physics, could be analyzed in the present study.

Five of six groups of terms did have a smaller posttest average distance by students who received grades of A or B, compared with students who received grades of D or F. The binomial test of these results (B = .094) does not support the hypothesis that students with higher final course grades have a more coherent cognitive structure than students who receive low final course grades. However, because of the low number of groups, only a result where all six of the groups of terms have a smaller average distance would be significant according to the binomial test.

It should also be noted that a subset of students were used in this study. Thirty-one of a total of sixty-one (or about 50.8%) of students receiving a final course grade of A or B in Physics were included in this portion of the study, while only sixteen of seventy-eight (or 20.5%) of student receiving an grade of D or F were included. Clearly the students with lower grades are underrepresented in the sample. This is due primarily to the fact that many more students who received low final grades were not present at the posttest compared with those students receiving high grades. This underrepresentation of students receiving poor course grades may have affected the results. Students who received a low course grade and who stopped coming to class (the group of students not represented in the sample) might be expected to be even more different from A or B students than the D and F students who diligently kept trying to do the course work.

Even with these caveats in mind, it is interesting to note that in only one group was the average distance of terms from the group centroid smaller for D and F students, than for A and B students. This group was labeled "Second Law of Thermodynamics" by the instructor, and consisted of five terms: "absolute temperature scale," "chaos," "entropy," "order at the molecular level," and "Second Law of Thermodynamics." Two of the five terms in this group were clustered differently by the Physics instructor in the pretest and posttest. In no other group of Physics terms was the instructor variability this high. If, as Nevah-Benjamin et al. (1986, 1989) suggest, student with higher grades have cognitive structures that are more similar to the instructor than lower achieving students, then it is possible that the ambiguity in the instructors' structure is reflected in the higher achieving students' cognitive structures. This may have contributed to a more diffuse clustering of this one group of terms.

When comparing the average distance of each group centroid from every other group centroid the binomial test did reveal a significant difference. The centroid of all six groups was farther away from other groups for A and B students than for D and F students, suggesting that groups of terms of better performing students are more distinct than those of poorly performing students.

In summary, although the within group coherence of groups was not statistically significantly different at the $p = .05$ level, the results are encouraging enough to invite further study of student cognitive structure and course achievement with a larger sample. A secondary analysis of the data revealed a statistically significant difference in the average distance of groups of terms from each other when comparing better performing and low performing students.

Students' cognitive structures and familiarity with course concepts

Are terms which students perceive as being familiar clustered in a way consistent with the instructor's clustering of these terms? There is evidence from the work of Chi and Koeske (1983), that this might be the case. In Chi and Koeske's study, a semantic map of a 4 1/2-year-old's dinosaur knowledge revealed that knowledge about well-known dinosaurs was clustered more coherently than knowledge about lesser-known dinosaurs.

In the present study, one would predict that if more familiar terms were grouped more similarly to the instructors' groupings, then the familiarity rating of terms would be negatively correlated to the average distance from the centroid of terms. Indeed a negative correlation was found between familiarity ratings and average distances in all cases, except the physics pretest. However, the correlations between the familiarity rating and average distance are generally low and insignificant.

A significant correlation was found between the posttest familiarity rating and the posttest average distance in the Women's Studies course ($\underline{r}(28) = -.433$, $\underline{p} = .017$). This result is difficult to interpret given the high variability of the instructor's clustering and the insignificant difference between pretest familiarity ratings and pretest average distances.

The results of correlations between the familiarity ratings and average distance from the group centroid do not support the hypothesis that more familiar terms are clustered more coherently. It is also possible that a successful result might be obtained were the question approached in a different way. Although the familiarity ratings and average distances seems to be internally consistent measures it is possible that correlating the two measures may not be an appropriate method for answering this question. In Chi and Koeske's (1983) study of dinosaur knowledge, the cognitive structure of less familiar and more familiar dinosaurs was contrasted within one individual. This suggests that rather than comparing students' groups with instructor groupings, it may be more appropriate to analyze the students' clusters without reference to instructor groupings. This can be accomplished by nearest neighbor analysis (Kruskal & Wish, 1973). In this kind of analysis local groups are created by looking at the plot and assigning terms that occur together into one group. Thus, groups are created according to student clustering, rather than by assignment to a group created by the instructor.

In summary, the results do not support the hypothesis that more familiar terms are clustered more coherently. It is suggested that another form of analysis, such as nearest neighbor analysis of multidimensional scaling plots, might be a more appropriate method for investigating this question.

Categorization of concepts and instruction

Chi, Feltovich and Glaser (1981) studied the categorization of physics problems by expert and novice physicists. They found that experts used deeper level or implicit criteria, (for example, the laws of physics needed to find the solution) to categorize problems. On the other hand, novices used surface level, or explicit characteristics (for example, the physical configuration of the problems) to categorize problems. A similar finding was found by Itano (1991) in the categorization of clinical cases by expert and novice nurses. Will students follow this pattern and use more deep level characteristics to group problems at the end of the semester than at the beginning of the semester?

In order to investigate this question, the kinds of statements students made to justify why groups of terms belonged together were analyzed and placed into one of two mutually exclusive groups. The average number of deep and surface explanations was compared from pretest and posttests in each of the three courses. Only one significant difference was found. The pretest and posttest surface categorizations in the Women's Studies course was significantly different at the $\underline{p}$ = .05 level, ($\underline{t}$(13) = 2.242), suggesting that at least in Women's Studies, there were less surface or explicit characteristics used to explain grouping of terms at the end of the semester. However, the average number of deep level explanations was not significantly different for Women's Studies. One should expect that if the hypothesis was true then the average number of surface characterizations would decrease and the average number of deep characterizations would increase from pretest to posttest. However, only half of this expected result was obtained.

These results do not support the hypothesis that the number of deeper level categorizations would increase from pretest to posttest. There are several possible reasons for this outcome. First, the classification of deep and surface by the raters may not have discriminated sufficiently between these categories. Second, it may be that the data available for analysis were not rich enough to be able to determine the true nature of the explanations. While Chi, Feltovich and Glaser (1981) used interviews with the participants in their study to determine the depth of their classification of problems, in this study the only data available were short written explanations of why groups of terms should be clustered together. Because the task was completed during a limited amount of time while students were in the classroom, the students were also under time pressure and often wrote very concise explanations of why the groups belonged together. It was then the raters' task to determine if these succinct, sometimes one-word explanations were deep or surface. This difficulty was compounded by the fact that some terms were both structurally related (which would generally be considered a surface relationship) and also related in a level of meaning (which would be considered a deeper level of categorization). For example, a student in physics might write the word "energy" to explain why a certain group of terms went together. It was then the raters' task to determine if this were a surface or deep explanation. To be consistent, any explanation that did not describe very obvious structural relationships like, "these terms all start with 's'" or "these are all '-ism' terms" were counted as deep categorizations.

Another possible explanation for the lack of a difference in deep and surface categorizations from pretest and posttest was that the task may have encouraged the students to use deep thinking even in the pretest. Students were asked to cluster terms in any way they saw fit, but the prior task was a familiarity rating, where students were asked to think about the terms as they related to that particular discipline. Perhaps this first task acted as a kind of "priming" and students continued to think about the terms as they related to the discipline (therefore in a deeper or more implicit way), even during the pretest. In support of this idea, recall that the pretest percentage of deep categorizations were high for all three courses. In the Educational Psychology pretest, 53.4% of the explanations were classified as deep. In the Physics and Women's Studies pretests, 81.4% and 67.8% of explanations, respectively, were classified as being deep.

Finally, it is also possible that, since the pretest and posttest were taken only a few months apart, sufficient time had not yet passed to see a difference in deep or surface thinking. In the study by Chi, Feltovich and Glaser (1981), a difference in categorization was seen in experts who had several more years of experience in the discipline than did the novices in the study.

In summary, the results do not support the hypothesis that students use deeper level categorization to group terms after instruction. It is suggested that there may have been insufficient data for the raters to accurately discriminate between deep and surface categorization by students.

Categorization of concepts and course achievement

Will students who receive higher final course grades use more deep level criteria to sort terms than students who receive lower course grades? Significant differences were not found between A or B and D or F students either in the average number of surface or deep categorizations during the posttest. These results do not support the hypothesis that students receiving higher final course grades will use deeper level criteria to categorize terms than students receiving lower final course grades. They are, however, consistent with the results obtained when a comparison was done between pretest and posttest clusters and similar considerations may apply for discussing the limitations of the results pertaining to both of these hypotheses.

In summary, although consistent with the results obtained in the previous section, the findings do not support the hypothesis that students with higher course grades use deeper level categorization to cluster terms.

Discussion of results across disciplines

This section will discuss a comparison of results across the three disciplines included in the study.

It is interesting to note that the results obtained from the Women's Studies students were consistently set apart from the results for Educational Psychology and Physics students. The Women's Studies students' cognitive structures were in least agreement with the cognitive structure of their instructor. The Women's Studies instructor exhibited the most variability of clustering of terms from pretest to posttest. The only significant correlations obtained between familiarity ratings and average distance was the correlation between the Women's Studies posttest familiarity ratings and posttest average distances. And finally, the only significant difference between pretest or posttest deep or surface descriptions of cluster was obtained in the Women's Studies course (between pretest and posttest surface explanations).

Why should the Women's Studies course be so consistently different from the other courses? An attempt to explain these differences was made earlier, based on the idea of ill-structured and well-structured domains. However, a discussion with the Women's Studies instructor about the aim of her course revealed a possible explanation which goes beyond the idea of a domain being ill-structured. The Women's Studies instructor's goal is to assist her students in becoming unsettled in their thinking and deconstructing certain ingrained concepts. In this kind of course, relationships are deliberately broken down by the instructor in the educational process. The instructor is deconstructing a students' cognitive structures rather than trying to impose a organized existing structure. A similar approach might be likely in disciplines like philosophy (Phenix, 1964).

Thus it is possible that not all disciplines strive to be solidly structured and hierarchically organized. Some disciplines may have as a goal to break down students' cognitive structures, so that they can be rebuilt without the restrictions of old prejudices and perceptions.

It is intriguing to speculate that the results of this study may have been documenting the deconstruction of students cognitive structures. If this is the case, then the assumption that students in a domain are more similar to experts in that domain after instruction than they are prior to instruction may not hold in all cases. It is possible that in some domains the intent of instruction is to disrupt current student thinking about some topics. Here, the instructor may wish students to vary their thinking from those of experts. The method of studying students' cognitive structures outlined in this study may be able to document cases when students' thinking (i. e., inferred cognitive structures) become more similar to the instructor's as well as when, hopefully by the intent of the instructor, they become less similar.

Limitations of the Study

This study has both theoretical and practical limitations. Theoretically, the study rests on the assumption that students' cognitive structures can and should be measured. As Phillips (1983) points out, students' cognitive structures can be confounded with other structures including the structure of the discipline the student is learning. Indeed, results of the current study strongly suggested that student cognitive structure was confounded by several factors such as familiarity or prior knowledge of students with course terms and content. Other factors such as the temporal sequence in which the terms were presented and the instructor's own clarity about course terms probably also affected the results. However, despite these confounding factors, the results of the current study make a case for the usefulness of exploration of students' cognitive structures.

In addition to the theoretical limitation just discussed, the current study had several practical limitations. One of the aims of this study was to conduct research on the cognitive structures of students in a naturalistic setting. Often studies of students' cognitive structures were conducted using a very short instructional time, (as short as 5 days) and using only textual material, rather than an instructor, to convey the information to be learned (Shavelson, 1972; Geeslin & Shavelson, 1975; Champagne et al., 1981). This study was conducted in actual college classrooms over the course of a semester.

Though this type of approach may increase the ecological validity of the results, it does introduce certain limitations.

One such limitation was the fact that the students were asked to complete the familiarity ratings and clustering tasks during the short periods of time the instructors could spare in their instructional schedule. This meant that the students had to understand the instructions quickly and may have felt pressured to finish the tasks. Indeed only 65.9% of the students were able to complete the repetitions of the clustering during the pretest and posttest and therefore data from the second clustering was not available for analysis.

Students also were not given course credit for completing the tasks. In each course, the instructor requested that the students complete the tasks thoughtfully. When introducing the tasks, the researcher presented some cognitive benefits of the tasks to the students. However, in the final analysis, it was the student participants' decision to complete the task thoughtfully and to the best of their ability. Some students may have been frivolous in their answers and this may have affected the results.

Perhaps more importantly, the timing of the pretest and posttest in the course schedule may have also affected the results. The posttests in two courses were completed on the last day of the semester. This is a time when students may be quite anxious about final exams, and may not have been giving the task full attention. This may have been especially true in the Educational Psychology course, where the posttest was administered directly after a lecture on a difficult subject was given. From the conversations of students during the posttest, it seemed clear that many students were still thinking about the subject material while taking the posttest. Students would sometimes stop their clustering task to ask other students or the instructor questions about some of the material that had just been presented.

Another effect of using real courses in different disciplines for the study was the fact that the course structure and outcomes of the three courses were quite different. The courses were selected for these differences, so that the results could be contrasted. Yet, the differences also make the results less homogeneous.

In the Physics course, with almost 200 students, there was less opportunity for interaction between individual students and the instructor than in the much smaller Educational Psychology or Women's Studies courses.

The method of student evaluation differed among the three courses. In Women's Studies, student essays and class participation were the primary modes of student evaluation. Educational Psychology and Physics primarily used examinations for evaluation.

The grade distribution in the courses was also varied. Physics had the widest grade distribution with a large percentage of students receiving low final course grades (over 44% received a grade of D or F) while there were few students receiving a grade of C or below in either Women's Studies or Educational Psychology. This affected the ability to analyze student data according to course grades in those two classes. In addition, students had the right to refuse to participate in the study, thus reducing the number of available participants in the portion of the study that compared students receiving high or low course grades.

Recommendations for Future Research

The use of the average distance from a group centroid is a promising measure of coherence of clusters. The results of this study support the hypothesis that the cognitive structures of students do become more coherent with regards to the instructors clusters over time. This agrees with the findings obtained by Streveler and Bail (1992) in a preliminary study of the coherence of students' cognitive structures.

However, a few modifications might be made in the procedure used in this study. First, a standardized measure of distance might be a more useful measure to compare distances within and across disciplines. The normality of this measure could also be tested to determine if would be appropriate to use parametric measures of significance rather than the non-parametric measures (namely, the binomial test) used in this study. Since the binomial test is very conservative, more significant results might be found using a more powerful test, such as the *t*-test. Using the *t*-test, the size of the differences between the pretest and posttest could be determined. It would also then be possible to compare individual groups within one course for significance and thus become more specific in describing and explaining how the clustering of groups changed over time or between achievement level.

Individual differences scaling analysis (Carol, 1972) of students receiving high and low course grades might also yield interesting results. Since individual scaling determines the dimensions different populations of subjects use as criteria for clustering, this type of analysis could provide insights into differences in the perceptions of high and low achieving students. The results possible with individual differences scaling do not directly address the questions asked in the present study, and thus it was not used here. However, individual differences scaling might be quite useful for further analysis of the data.

Implications for Education

As Champagne et al. (1984) pointed out, the method of clustering used in this study is one which can be readily understood and used by students in a course. The analysis of the results through multidimensional scaling is also straightforward.

Using the method outlined in this study, a plot of how students view the relationship between important course terms can be generated and presented to a course instructor. The instructor can view the plot to determine how closely the students' perceptions of relationships between terms matches the instructor's own perception. If students' groups are not consistent with the instructor's groupings, this may be a sign that the students are confused about the relationships involved. The instructor could then adjust the curriculum to reiterate these topics, perhaps stressing relationships that exist between concepts.

Close analysis of the plot of terms could also help the instructor pinpoint student misconceptions.

For example, Streveler and Bail (1992) found that students in a graduate Educational Psychology course persistently grouped the term "vicarious reinforcement" with other terms containing the word "reinforcement" such as "positive reinforcer," "negative reinforcer," and "intermittent reinforcement".

The student placement of "vicarious reinforcement" differed from the instructor's clustering of the term. The instructor grouped "vicarious reinforcement" with other terms related to observational learning theory, while "negative reinforcer," "positive reinforcer," and "intermittent reinforcement" were grouped by the instructor with terms relating to behaviorism. Thus the students' persistent placement, even at the end of the semester, of the term "vicarious reinforcement" with behaviorism-related terms could be seen as a misconception on the part of the students. Students might, for example, believe that because "vicarious reinforcement" is so structurally similar to other behaviorism terms that it, too, refers to behaviorism.

It should be noted that students were able to place other semantically related terms into groups on the basis of meaning. This supports the idea that "vicarious reinforcement" might be viewed as a student misconception.

The use of multidimensional scaling plots to help pinpoint student misconceptions parallels the use of concept maps to distinguish student misconceptions (Barenholz & Tamir, 1987; Feldstine, 1987; Hoz et al., 1987). While concept maps have the advantage of not needing to be analyzed by the somewhat esoteric method of multidimensional scaling, the method outlined in this paper can be readily used to analyze the data from a group of students. Multidimensional scaling analysis also lends itself to quantitative measurement, whereas the measurement of individual student concept maps is idiosyncratic and problematic (Lay-Dopyera & Beyerbach, 1983; Stuart, 1985).

Conclusions

In summary, the results of the present study suggest that:

1 - After instruction, students' cognitive structures are more coherent with regards to the instructor's cognitive structure than they are before instruction. It should be noted that other factors also influence students' cognitive structure and that in some disciplines, it may be the instructor's intent to "deconstruct" student thinking, resulting in cognitive structures which become more dissimilar to expert thinking.

2 - After instruction, high performing students may have cognitive structures which are more similar to the instructor's cognitive structure than low performing students. Results in this study approached statistical significance and were bolstered by the fact that higher performing students did appear to have clustered terms in more distinct groups than their low-performing classmates.

3 - Student familiarity with course terms was not significantly correlated to the way in with the instructor group the same terms. It is suggested that an alternate form of analysis, which does not take

instructor groupings into account, may be a more appropriate method for answering this question.

4 - The results of the study did not support the hypotheses that after instruction, students used deeper level categorization of terms than they do before instruction, or that higher performing students used deeper level categorization than lower performing students.

APPENDIX I: INSTRUCTIONS TO PROFESSORS

This study will investigate how students organize the knowledge they have learned in your class. Their organization will be measured by comparing how students group important course terms at the beginning and end of a semester.

Here's a summary of what you and the students will do. (After this summary each step is decribed in detail.)

• You will create a list of terms which you feel are important in your course. (More specifics on how to pick these terms will be given below).

• At the <u>beginning and end of the semester</u> **you** will sort these terms into as many groups as you think is appropriate.

• **Your students** will also group these terms at the <u>beginning and end of the semester</u>. This will take about 20-25 minutes.

I am planning to study how students at different performance levels go about organizing important terms.

Because of this I ask your permission to share individual and total grades in your course with me. I will of course also ask the students for their permission. This information will be strictly confidential.

I am repeating this study in other fields besides yours. Because of this I ask that you follow the procedures carefully, so that I can have as much consistency across courses as possible.

Here's a step by step run-down of what you need to do.

1) Prior to the beginning of the semester please create a list of terms which are important in your class. I will let you decide the exact number of terms, but a number of around 30 terms works well with the methodology I am using.

Here's some guidelines for creating this list.

1. Look through the glossary of the text(s) you are using to find terms you think are important.

2. Look at your syllabus to get ideas about what you will be covering in the class.

I'd like you to pick terms which you feel are **central** to your course. One way to look at this is to think of terms you think students in your course should understand well.

It is important that the degree of generality or specificity of terms be fairly consistent across courses. Obviously these terms can be put in very general or very specific ways. I'd like you to choose a medium level of specificity.

Let me use an example to illustrate what I mean.

Broad category ---- animal

Middle category --- dog

Specific category -- golden retriever

The desired level of specificity here would be 'dog'.

2.) **Please give the list of terms to me.** I will generate the materials to be given to your students. I will need this at <u>least two days before</u> the time it is administered to your class.

3.) Sometime during the first two weeks of class (at a time you deem is appropriate) I will <u>ask your students</u> to complete the following tasks:

- Rate their familiarity with the terms you have picked on a five point scale (from "I don't understand this term at all" to "I could define and explain this term to someone").

- Place these terms into groups (as many or as few as they desire) and write a one sentence explanation of why those terms go together.

- I will also ask students the following:
  - to provide **optional** demographic information (age, gender, class standing)
  - to list courses in the field or related fields they have already had in college or in high school (I will need to consult with you what these related fields might be for your course).

- to give me permission for you to share the course grades with me. (This of course is optional and may be rescinded at any time).

I am estimating it will take about 5 minutes to explain the procedures to the students and about 20 minutes for the students to perform these tasks. We can discuss how to best schedule this time. If at all possible, I ask that this be done in class. I will be glad to administer this.

4). I will also ask YOU to group these terms around this time.

5). Sometime in the last two weeks of class (at a time deemed appropriate by you) I will ask students to group the same terms, to explain why they go together.

6). I will also ask you to group the terms again at the end of the semester.

APPENDIX II: PRETEST DATA SHEET

Name ____________________ Social Security Number __________

Directions: Use the following **numbers** to **rate** your **understanding** of each concept listed **AS IT USED IN EDUCATIONAL PSYCHOLOGY.**

1 = I don't understand this concept at all.
2 = I only have a vague understanding of this concept.
3 = I can understand this concept if it is presented in context.
4 = I understand the concept quite well.
5 = I could define and briefly explain this concept to another person.

___	a-ha experience	___	need for achievement
___	accommodation	___	normal distribution
___	aversive control	___	operant conditioning
___	correlation	___	phenomenology
___	decentration	___	reciprocal determinism
___	drive theory	___	reliability
___	encoding	___	schema
___	epigenesis	___	self-actualization
___	equilibrium	___	self-efficacy
___	gestalt	___	shaping
___	identity diffusion	___	standard deviation
___	Law of Effect	___	third-force psychology
___	locus of control	___	validity
___	metacognition	___	vicarious reinforcement
___	moratorium	___	working memory

Providing the following information is **OPTIONAL**. Please note this information will be used for research purposes only.

Your age __________ Your gender (male or female) __________
Your ethnicity____
Your class standing (freshman, sophomore, etc.) __________
Your High School ______________________________________
List the courses in education or psychology in *high school*.
____
List the courses in education or psychology you have completed in *college*. ______________________________________

Please note: Your grades in this course will be shared with the researcher *for research purposes only*. Please notify the instructor if you do NOT wish your grades to be shared. Your participation in this study is voluntary. *Your participation or lack thereof will in no way affect your grade in this course.*

Name ______________________ Social Security Number __________

Directions: Please use the following **numbers** to **rate** your **understanding** of each concept listed **AS IT USED IN PHYSICS.**

1 = I don't understand this concept at all.
2 = I only have a vague understanding of this concept.
3 = I can understand this concept if it is presented in context.
4 = I understand the concept quite well.
5 = I could define and briefly explain this concept to another person.

___	absolute temperature scale	___	forms of energy
___	acceleration	___	Ideal Gas Laws
___	action-at-a-distance forces	___	impulse
___	action-reaction forces	___	inertia
___	angular momentum	___	inertial reference frames
___	Bernouilli's Principle	___	interference
___	buoyancy	___	linear momentum
___	chaos	___	mass
___	conservation of energy	___	matter waves
___	contact forces	___	momentum conservation
___	density	___	order at the molecular level
___	diffraction	___	polarization
___	electromagnetic waves	___	Second Law of Thermodynamics
___	entropy	___	weight
___	force	___	work

Providing the following information is **OPTIONAL**. Please note this information will be used for research purposes only.

Your age ___________ Your gender (male or female) _________
Your ethnicity ____
Your class standing (freshman, sophomore, etc.) ____________
Your High School ____
List the courses in physics, mathematics, or chemistry you took in *high school.* ____
List the courses in physics, mathematics, or chemistry you have completed in *college.* _____________________________________

Please note: Your grades in this course will be shared with the researcher *for research purposes only*. Please notify the instructor if you do NOT wish your grades to be shared. Your participation in this study is voluntary. *Your participation or lack thereof will in no way affect your grade in this course.*

Name ________________________ Social Security Number __________

Directions:Please use the following **numbers** to **rate** your **understanding** of each concept listed **AS IT USED IN WOMEN'S STUDIES.**

1 = I don't understand this concept at all.
2 = I only have a vague understanding of this concept.
3 = I can understand this concept if it is presented in context.
4 = I understand the concept quite well.
5 = I could define and briefly explain this concept to another person.

___ capitalism
___ consciousness raising
___ difference
___ empowerment
___ female gaze
___ female point-of-view
___ feminism
___ gender
___ gynocentrism
___ heterosexism
___ homophobia
___ humanism
___ knowledge
___ male gaze
___ multiplicity
___ mystification
___ narrative
___ oppression
___ patriarchy
___ perspective
___ phallocentrism
___ power
___ racism
___ rescripting
___ sexism
___ silence
___ subjectivity
___ surveillance
___ the other
___ voice

Providing the following information is **OPTIONAL**. Please note this information will be used for research purposes only.

Your age ____________ Your gender (male or female) __________
Your ethnicity ____
Your class standing (freshman, sophomore, etc.) _____________
Your High School ____
List the courses in ethnic studies, political science, or women's studies you took in *high school.* ______________________
List the courses in ethnic studies, political science, or women's studies you have completed in *college*. ______________________

Please note: Your grades in this course will be shared with the researcher *for research purposes only*. Please notify the instructor if you do NOT wish your grades to be shared. Your participation in this study is voluntary. *Your participation or lack thereof will in no way affect your grade in this course.*

APPENDIX III: POSTTEST DATA SHEETS

Name ____________________ Social Security Number ________________

Directions: Use the following **numbers** to **rate** your **understanding** of each concept listed **AS IT USED IN EDUCATIONAL PSYCHOLOGY.**

1 = I don't understand this concept at all.
2 = I only have a vague understanding of this concept.
3 = I can understand this concept if it is presented in context.
4 = I understand the concept quite well.
5 = I could define and briefly explain this concept to another person.

___ a-ha experience
___ accommodation
___ aversive control
___ correlation
___ decentration
___ drive theory
___ encoding
___ epigenesis
___ equilibrium
___ gestalt
___ identity diffusion
___ Law of Effect
___ locus of control
___ metacognition
___ moratorium
___ need for achievement
___ normal distribution
___ operant conditioning
___ phenomenology
___ reciprocal determinism
___ reliability
___ schema
___ self-actualization
___ self-efficacy
___ shaping
___ standard deviation
___ third-force psychology
___ validity
___ vicarious reinforcement
___ working memory

Please note: With your permission, your grades in this course will be used *for research purposes only*. Grades of specific individuals will never be published. Your participation in this study is voluntary. *Your participation or lack thereof will in no way affect your grade in this course.*

If you do NOT wish your grades to be shared please complete the bottom portion of this sheet.

I do NOT give permission for my grades to be used in this research study.

Name ____________________ Social Security Number ________________

Name ____________________ Social Security Number ____________

Directions: Please use the following **numbers** to **rate** your **understanding** of each concept listed **AS IT USED IN PHYSICS.**

1 = I don't understand this concept at all.
2 = I only have a vague understanding of this concept.
3 = I can understand this concept if it is presented in context.
4 = I understand the concept quite well.
5 = I could define and briefly explain this concept to another person.

___	absolute temperature scale	___	forms of energy
___	acceleration	___	Ideal Gas Laws
___	action-at-a-distance forces	___	impulse
___	action-reaction forces	___	inertia
___	angular momentum	___	inertial reference frames
___	Bernoulli's Principle	___	interference
___	buoyancy	___	linear momentum
___	chaos	___	mass
___	conservation of energy	___	matter waves
___	contact forces	___	momentum conservation
___	density	___	order at the molecular level
___	diffraction	___	polarization
___	electromagnetic waves	___	Second Law of Thermodynamics
___	entropy	___	weight
___	force	___	work

Please note: With your permission, your grades in this course will be used *for research purposes only*. Grades of specific individuals will never be published. Your participation in this study is voluntary. *Your participation or lack thereof will in no way affect your grade in this course.*

If you do NOT wish your grades to be shared please complete the bottom portion of this sheet.

I do NOT give permission for my grades to be used in this research study.

Name ____________________ Social Security Number ____________

Name ____________________ Social Security Number ____________

Directions: Please use the following **numbers** to **rate** your **understanding** of each concept listed **AS IT USED IN WOMEN'S STUDIES.**

1 = I don't understand this concept at all.
2 = I only have a vague understanding of this concept.
3 = I can understand this concept if it is presented in context.
4 = I understand the concept quite well.
5 = I could define and briefly explain this concept to another person.

___	capitalism	___	mystification
___	consciousness raising	___	narrative
___	difference	___	oppression
___	empowerment	___	patriarchy
___	female gaze	___	perspective
___	female point-of-view	___	phallocentrism
___	feminism	___	power
___	gender	___	racism
___	gynocentrism	___	rescripting
___	heterosexism	___	sexism
___	homophobia	___	silence
___	humanism	___	subjectivity
___	knowledge	___	surveillance
___	male gaze	___	the other
___	multiplicity	___	voice

Please note: With your permission, your grades in this course will be used *for research purposes only*. Grades of specific individuals will never be published. Your participation in this study is voluntary. *Your participation or lack thereof will in no way affect your grade in this course.*

If you do NOT wish your grades to be shared please complete the bottom portion of this sheet.

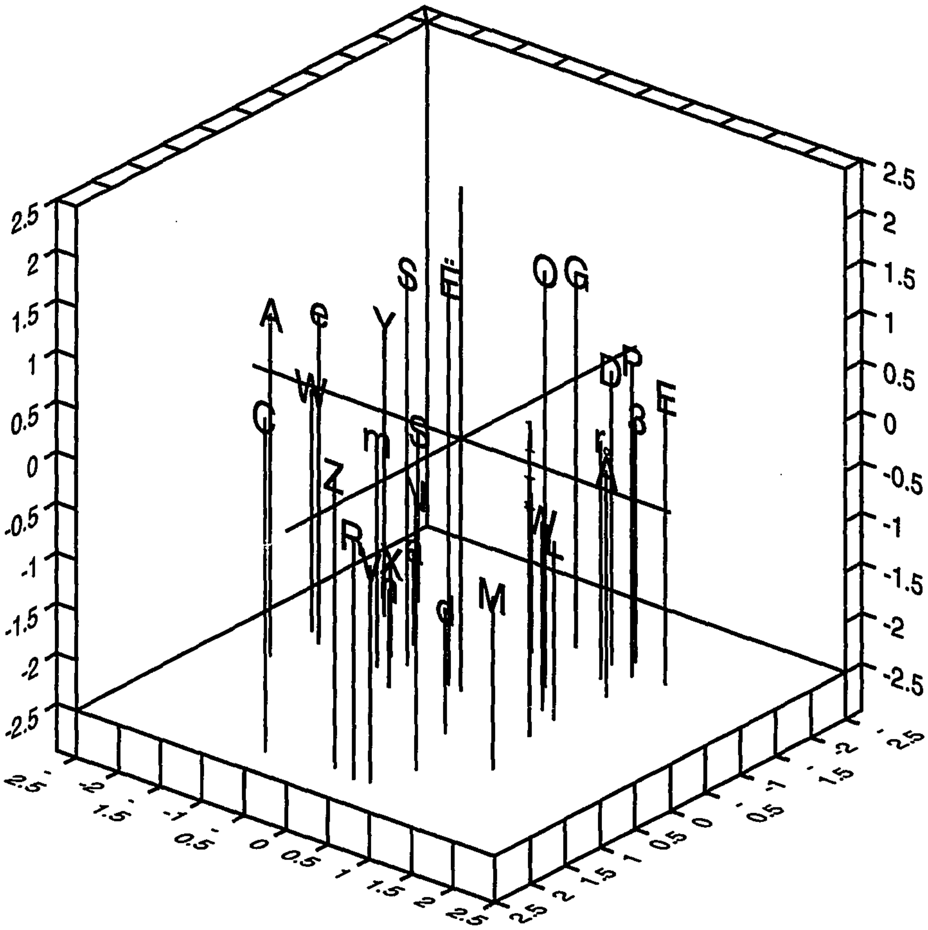
- -

I do NOT give permission for my grades to be used in this research study.

Name ____________________ Social Security Number ____________

APPENDIX IV: PLOTS OF IN OF INDIVIDUAL COURSE TERMS

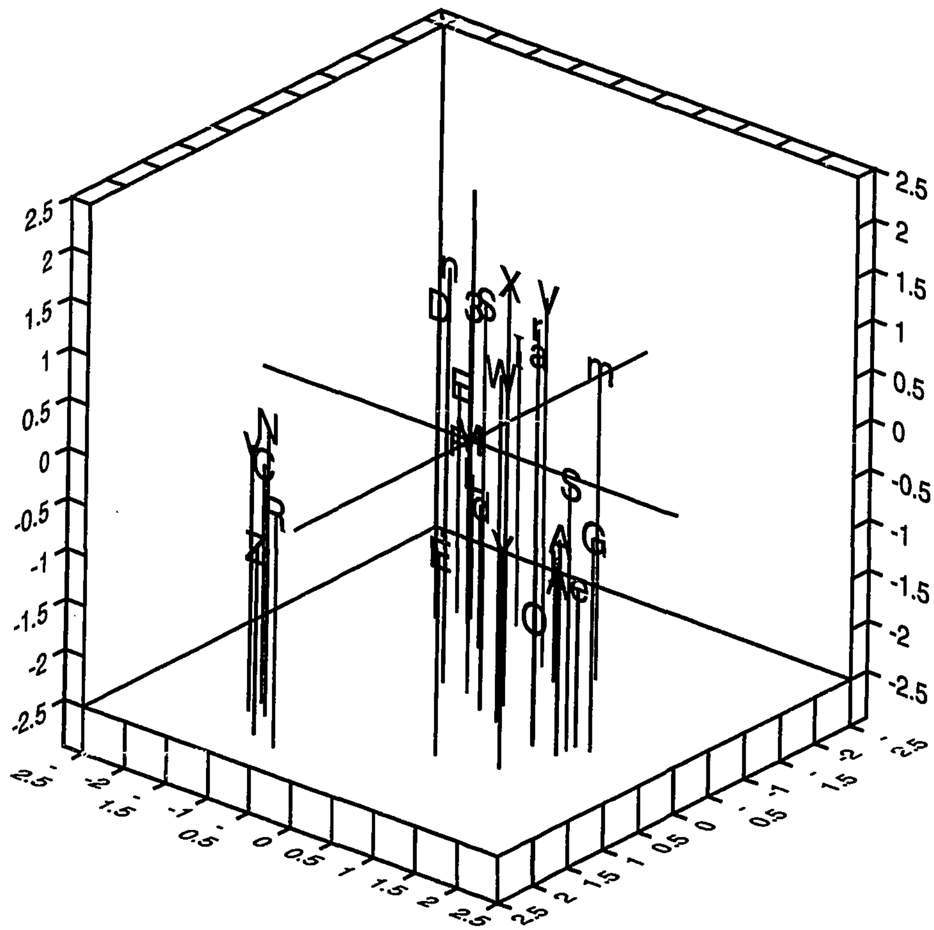
Educational Psychology Pretest



Legend

a = a-ha experience
A = accommodation
Å = aversive control
C = correlation
d = decentration
D = drive theory
e = encoding
E = epigenesis
Ë = equilibrium
G = gestalt
I = identity diffusion
l = law of effect
L = locus of control
m = metacognition
M = moratorium
n = need for achievement
N = normal distribution
O = operant conditioning
P = phenomenology
r = reciprocal determinism
R = reliability
s = schema
S = self-actualization
X = self-efficacy
Y = shaping
Z = standard deviation
3 = third force psychology
v = validity
V = vicarious reinforcement
W = working memory

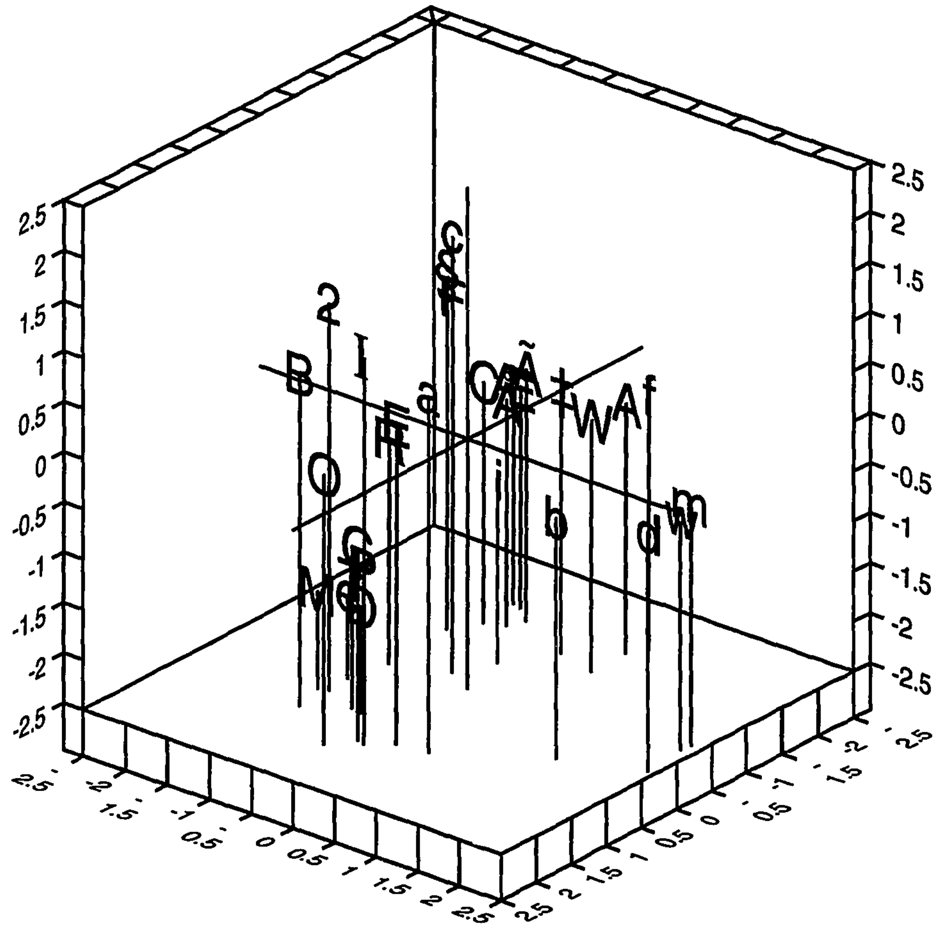
Educational Psychology Posttest



Legend

a = a-ha experience	n = need for achievement
A = accommodation	N = normal distribution
Å = aversive control	O = operant conditioning
C = correlation	P = phenomenology
d = decentration	r = reciprocal determinism
D = drive theory	R = reliability
e = encoding	s = schema
E = epigenesis	S = self-actualization
Ë = equilibrium	X = self-efficacy
G = gestalt	Y = shaping
I = identity diffusion	Z = standard deviation
l = law of effect	3 = third force psychology
L = locus of control	v = validity
m = metacognition	V = vicarious reinforcement
M = moratorium	W = working memory

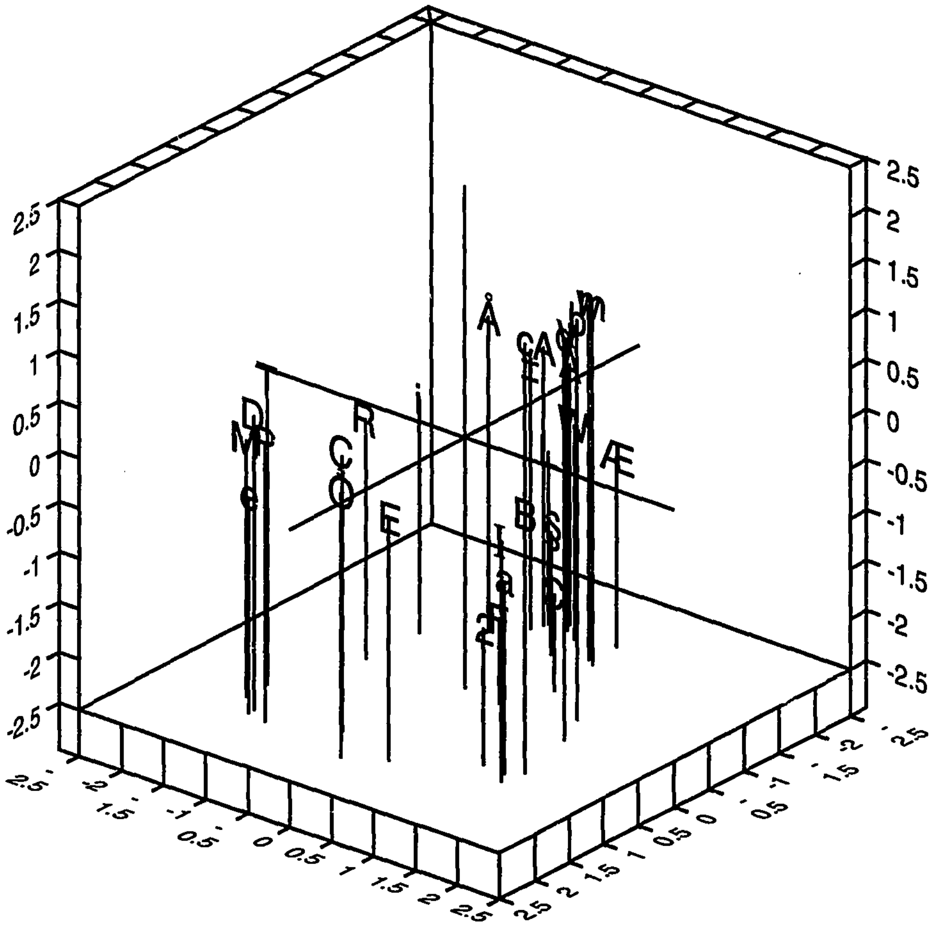
Physics Pretest



Legend

a = absolute temperature scale
A = acceleration
Å = action-at-a-distance forces
Â = action-reaction forces
Æ = angular momentum
B = Bernoulli's Principle
b = buoyancy
Ç = chaos
C = conservation of energy
c = contact forces
d = density
D = diffraction
e = electromagnetic waves
E = entropy
f = force
F = forms of energy
I = Ideal Gas Laws
i = impulse
‡ = inertia
R = inertial reference frames
T = interference
L = linear momentum
m = mass
M = matter waves
X = momentum conservation
O = order at the molecular level
P = polarization
2 = 2nd Law of Thermodynamics
w = weight
W = work

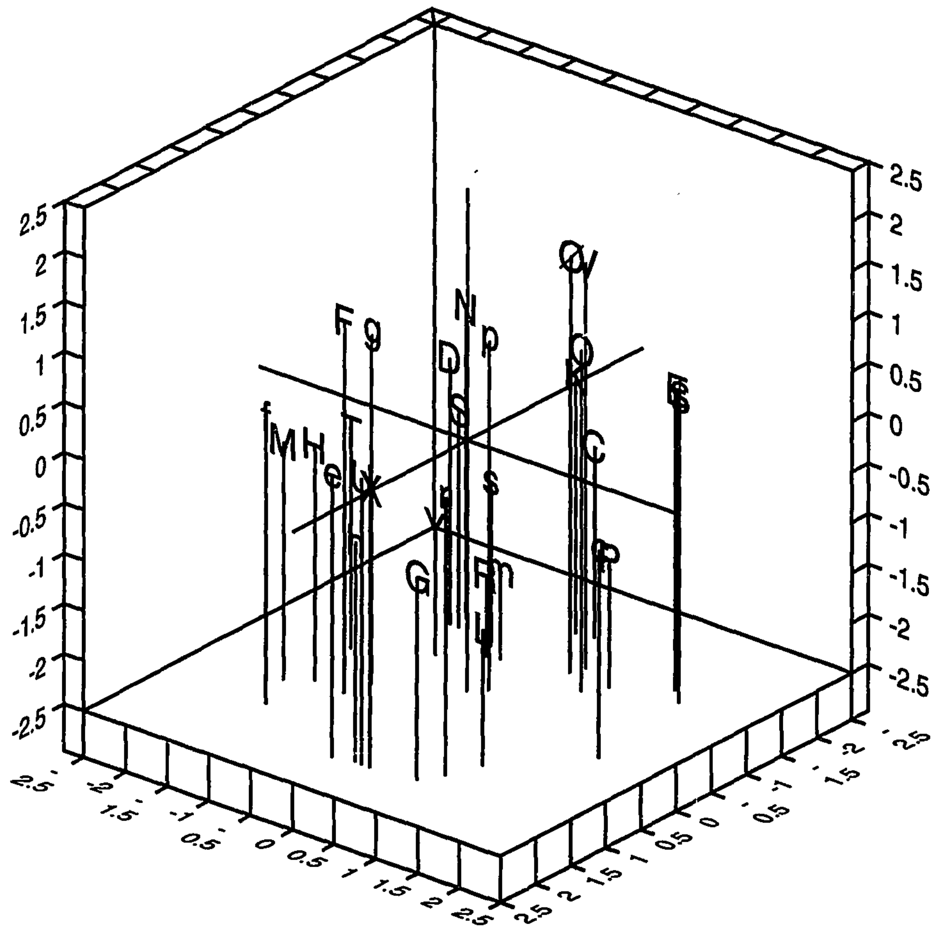
Physics Posttest



Legend

a = absolute temperature scale
A = acceleration
Å = action-at-a-distance forces
Â = action-reaction forces
Æ = angular momentum
B = Bernoulli's Principle
b = buoyancy
Ç = chaos
C = conservation of energy
c = contact forces
d = density
D = diffraction
E = entropy
f = force
F = forms of energy
I = Ideal Gas Laws
i = impulse
‡ = inertia
R = inertial reference frames
T = interference
L = linear momentum
m = mass
M = matter waves
X= momentum conservation
O = order at the molecular level
2 = 2nd Law of Thermodynamics
w = weight
W = work

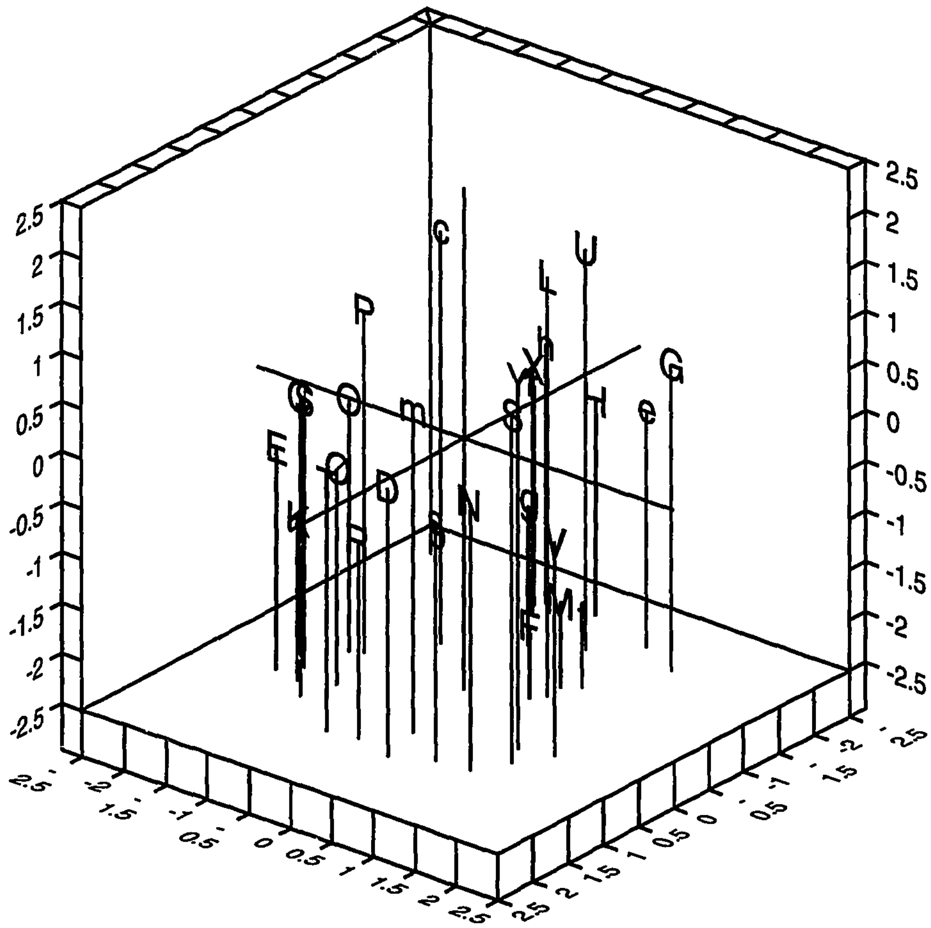
Women's Studies Pretest



Legend

c = capitalism
C = consciousness raising
D = difference
E = empowerment
f = female gaze
F = female point-of-view
e = feminism
g = gender
G = gynocentrism
h = heterosexism
H = homophobia
U = humanism
K = knowledge
M = male gaze
Y = multiplicity
m = mystification
N = narrative
O = oppression
P = patriarchy
p = perspective
L = phallocentrism
$ = power
r = racism
R = rescripting
X = sexism
Ø = silence
s = subjectivity
S = surveillance
T = the other
V = voice

Women's Studies Posttest

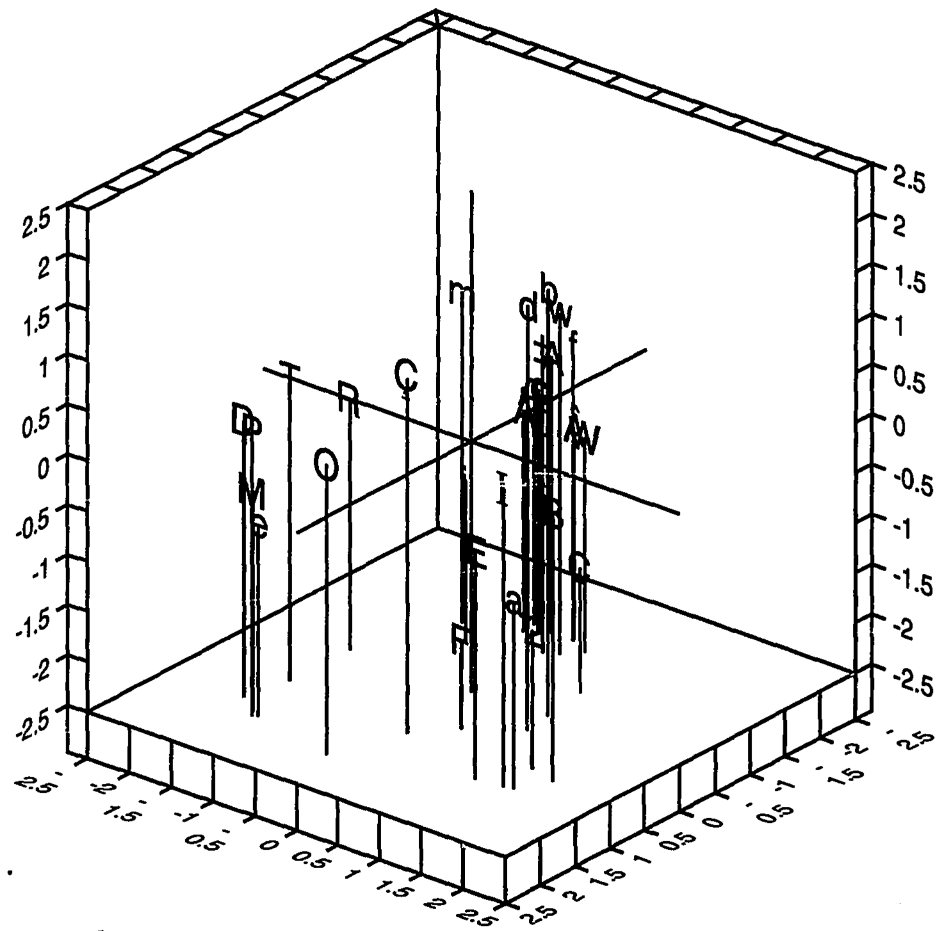


Legend

c = capitalism
C = consciousness raising
D = difference
E = empowerment
f = female gaze
F = female point-of-view
e = feminism
g = gender
G = gynocentrism
h = heterosexism
H = homophobia
U = humanism
K = knowledge
M = male gaze
Y = multiplicity
m = mystification
N = narrative
O = oppression
P = patriarchy
p = perspective
L = phallocentrism
$ = power
r = racism
R = rescripting
X = sexism
Ø = silence
s = subjectivity
S = surveillance
T = the other
V = voice

APPENDIX V: PLOTS OF INDIVIDUAL TERMS FOR A AND B AND D AND F STUDENTS, PHYSICS POSTTEST

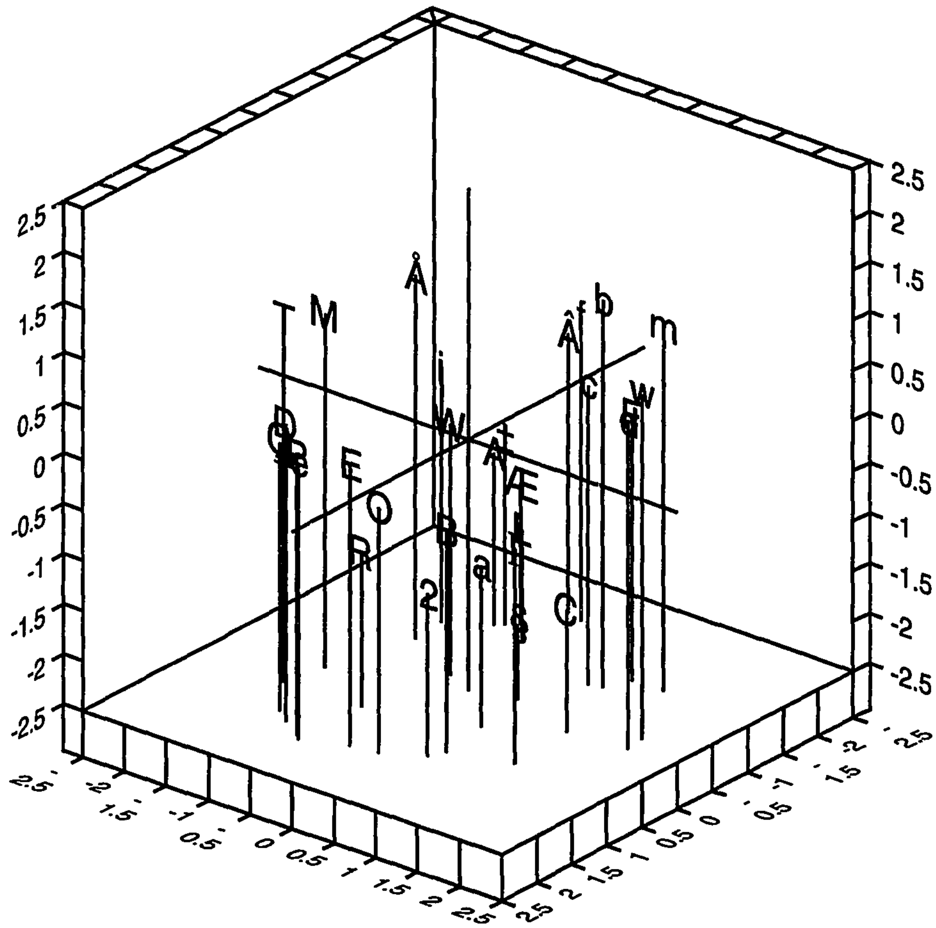
A & B Students



Legend

a = absolute temperature scale
A = acceleration
Á = action-at-a-distance forces
Â = action-reaction forces
Æ = angular momentum
B = Bernoulli's Principle
b = buoyancy
Ç = chaos
C = conservation of energy
c = contact forces
d = density
D = diffraction
e = electromagnetic waves
E = entropy
f = force
F = forms of energy
I = Ideal Gas Laws
i = impulse
‡ = inertia
R = inertial reference frames
T = interference
L = linear momentum
m = mass
M = matter waves
X = momentum conservation
O = order at the molecular level
P = polarization
2 = 2nd Law of Thermodynamics
w = weight
W = work

D & F Students



<u>Legend</u>

a = absolute temperature scale
A = acceleration
Å = action-at-a-distance forces
Â = action-reaction forces
Æ = angular momentum
B = Bernoulli's Principle
b = buoyance
Ç = chaos
C = conservation of energy
c = contact forces
d = density
D = diffraction
e = electromagnetic waves
E = entropy
f = force
F = forms of energy
I = Ideal Gas Laws
i = impulse
‡ = inertia
R = inertial reference frames
T = interference
L = linear momentum
m = mass
M = matter waves
X = momentum conservation
O = order at the molecular level
P = polarization
2 = 2nd Law of Thermodynamics
w = weight
W = work

APPENDIX VI: FAMILIARITY RATINGS OF TERMS, SORTED IN DESCENDING ORDER BY FAMILIARITY

Educational Psychology Pretest

Term	Pretest familiarity rating
Need for acheivement	3.500
Self-actualization	3.500
Accommodation	3.286
Correlation	3.286
Validity	3.286
Reliability	3.214
Equilibrium	3.071
Standard deviation	3.071
A-ha experience	2.929
Encoding	2.786
Normal distribution	2.786
Schema	2.786
Self-efficacy	2.786
Operant conditioning	2.571
Working memory	2.571
Shaping	2.500
Moratorium	2.214
Vicarious reinforcement	2.000
Gestalt	1.929
Law of Effect	1.929
Metacognition	1.929
Identity diffusion	1.857
Aversive control	1.714
Decentration	1.643
Phenomenology	1.643
Drive theory	1.429
Reciprocal determinism	1.429
Epigenesis	1.286
Locus of control	1.214
Third-force psychology	1.143

Educational Psychology Posttest Ratings of Familiarity, Sorted in Descending Order by Familiarity.

Term	Posttest familiarity rating
Schema	4.537
Standard deviation	4.500
Self-actualization	4.429
Reliability	4.286
A-ha experience	4.214
Accommodation	4.214
Correlation	4.214
Operant conditioning	4.214
Self-efficacy	4.143
Validity	4.143
Gestalt	4.000
Normal distribution	4.000
Shaping	4.000
Equilibrium	3.929
Need for acheivement	3.857
Vicarious reinforcement	3.643
Encoding	3.500
Working memory	3.500
Law of Effect	3.429
Aversive control	3.143
Metacognition	3.143
Phenomenology	2.857
Reciprocal determinism	2.786
Decentration	2.714
Drive theory	2.643
Locus of control	2.643
Identity diffusion	2.571
Third-force psychology	2.500
Moratorium	2.357
Epigenesis	2.000

Physics Pretest Ratings of Familiarity, Sorted in Descending Order by Familiarity.

Term	Pretest familiarity rating
Mass	3.446
Density	3.351
Acceleration	3.230
Weight	3.216
Force	3.122
Work	3.068
Conservation of energy	2.959
Forms of energy	2.838
Action-reaction forces	2.622
Inertia	2.581
Absolute termperature scale	2.568
Bouyancy	2.446
Ideal Gas Laws	2.384
Entropy	2.324
Polarization	2.297
Diffraction	2.284
Linear momentum	2.189
Electromagnetic waves	2.176
Angular momentum	2.149
Contact forces	1.986
Second Law of Thermodynamics	1.973
Interference	1.892
Momentum conservation	1.851
Impulse	1.838
Order at the molecular level	1.784
Chaos	1.716
Matter waves	1.671
Action-at-a-distance forces	1.622
Inertial reference frames	1.554
Bernoulli's Principle	1.446

Physics Posttest Ratings of Familiarity, Sorted in Descending Order by Familiarity.

Term	Posttest familiarity rating
Mass	4.581
Weight	4.568
Acceleration	4.432
Density	4.392
Force	4.324
Conservation of energy	4.311
Work	4.311
Absolute termperature scale	4.027
Forms of energy	3.986
Linear momentum	3.973
Action-reaction forces	3.892
Bouyancy	3.878
Momentum conservation	3.838
Ideal Gas Laws	3.784
Inertia	3.703
Angular momentum	3.541
Contact forces	3.405
Impulse	3.284
Second Law of Thermodynamics	3.284
Bernoulli's Principle	3.233
Entropy	3.230
Chaos	2.987
Action-at-a-distance forces	2.792
Diffraction	2.770
Order at the molecular level	2.757
Polarization	2.662
Interference	2.649
Matter waves	2.554
Electromagnetic waves	2.548
Inertial reference frames	2.370

Women's Studies Pretest Ratings of Familiarity, Sorted in Descending Order by Familiarity

Term	Pretest familiarity rating
Gender	4.214
Racism	4.071
Perspective	3.929
Power	3.929
Sexism	3.714
Homophobia	3.571
Silence	3.571
Capitalism	3.429
Female point-of-view	3.429
Knowledge	3.429
Oppression	3.214
Difference	3.143
Feminism	3.143
Subjectivity	3.071
Surveillance	2.857
Voice	2.857
Heterosexism	2.786
Narrative	2.786
Humanism	2.714
Consciousness raising	2.643
Empowerment	2.571
The other	2.429
Patriarchy	2.357
Phallocentrism	1.500
Gynocentrism	1.429
Multiplicity	1.429
Mystification	1.429
Female gaze	1.357
Male gaze	1.286
Rescripting	1.071

Women's Studies Posttest Ratings of Familiarity, Sorted in Descending Order by Familiarity

Term	Posttest familiarity rating
Power	4.714
Racism	4.714
Gender	4.571
Silence	4.429
Feminism	4.286
Sexism	4.286
Knowledge	4.214
Female point-of-view	4.143
Oppression	4.143
Homophobia	4.071
Perspective	4.071
Difference	4.000
Voice	4.000
Heterosexism	3.857
Patriarchy	3.857
Empowerment	3.786
Narrative	3.714
Gynocentrism	3.643
Capitalism	3.571
Consciousness raising	3.500
Humanism	3.429
Mystification	3.357
Subjectivity	3.357
The other	3.214
Surveillance	3.143
Phallocentrism	3.000
Male gaze	2.926
Female gaze	2.857
Rescripting	2.857
Multiplicity	2.714

REFERENCES

Barenholz, H. & Tamir, P. (1987). The design, implementation and evaluation of a microbiology course with specific reference to misconception and concept maps. *Proceedings of the second international seminar: Misconceptions and educational strategies in science and mathematics, 1,* 32-44. Itaca, NY: Cornell University.

Carol, D. J. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences, Vol. I,* (pp. 105-155). New York: Seminar Press.

Champagne, A. B., Hoz, R., & Klopfer, L.E. (1984, April). *Construct validation of the cognitive structure of physics concepts.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Champagne, A. B., Klopfer, L. E., DeSena, A. T., & Squires, D. A. (1981). Structural representations of students' knowledge before and after science instruction. *Journal of Research in Science Teaching, 18,* 97-111.

Chase, W. G., & Simon, H. A. (1973a). Perception in chess. *Cognitive Psychology, 4,* 55-81.

Chase, W. G., & Simon, H. A. (1973b). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121-152.

Chi, M. T. H., Hutchinson, J. E. & Robin, A. F. (1989). How inferences about novel domain-related concepts can be constrained by structured knowledge. *Merrill-Palmer Quarterly, 35,* 27-62.

Chi, M. T. H., & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology, 19,* 29-39.

Chi, M. T. H., & Rees, E. T. (1983). A learning framework for development. In M. T. H. Chi (Ed.), *Contributions to human development, Vol. 9, Trends in memory development research.* (pp. 71-107). Basel: Karger.

Collins, A. M. & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psycholgical Review, 82,* 407-428.

Collins, A. M. & Quillian, M. R. (1969). Retrieval times from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 240-247.

de Groot, A. D. (1965). *Thought and choice in chess.* The Hague: Mouton.

de Groot, A. D. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving* (pp. 19-50). New York: Wiley.

Donald, J. G. (1982, March). *The development of knowledge structures.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Donald, J. G. (1983). Knowledge structures: Methods for exploring content. *Journal of Higher Education, 54,* 31-41.

Donald, J. G. (1986). Knowledge and the university curriculum. *Higher Education, 15,* 267-282.

Dunn-Rankin, P. (1983). *Scaling methods.* Hillsdale, NJ: Erlbaum.

Egan, D. E., & Schwartz, B. J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition, 7,* 149-158.

Engle, R. W., & Bukstel, L. (1978). Memory processes among bridge players of differing expertise. *American Journal of Psychology, 91,* 673-689.

Feldsine, J. (1987). Distinguishing student misconceptions from alternate conceptual frameworks through the construction of concept maps. *Proceedings of the second international seminar: Misconceptions and educational strategies in science and mathematics, 1,* 176-180. Itaca, NY: Cornell University.

Feltovich, P. J., Coulson, R. L., Spiro, R. J., & Dawson-Sanders, B. K. (1992). Knowledge application and transfer of complex tasks in ill-structured domains: Implications for instruction and testing in biomedicine. To appear in D. Evans & V. Patel (Eds.), *Advanced models of cognition for medical training and practice.* New York: Springer-Verlag.

Forsythe, G. B., & Barber, H. F. (1992, April). *Military strategic thinking: Expert-novice differences in the structure and content of cognitive representations.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Geeslin, W. E., & Shavelson, R. J. (1975). Comparison of content structure and cognitive structure in high school students' learning of probability. *Journal for Research in Mathematics Education, March*, 109-120.

Glaser, R. (1985). *On the nature of expertise.* Pittsburgh, PA: Learning and Research Development Center, University of Pittsburgh.

Glaser, R., & Chi, M. T. H. (1988). The nature of expertise: Overview. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. xv-xxviii). Hillsdale, NJ: Erlbaum.

Gobbo, C., & Chi, M. T. H. (1986). How knowledge is structured and used by expert and novice children. *Cognitive Development, 1,* 221-237.

Greeno, J. G. (1983). Response to Phillips. *Educational Psychologist, 18,* 75-80.

Hays, W. L. (1988). *Statistics, 4th Edition.* New York: Holt, Rinehart and Winston.

Hoz, R., Tomer, Y., Bowman, D., & Chayoth, R. (1987). The use of concept maps to diagnose misconceptions in biology and earth sciences. *Proceedings of the second international seminar: Misconceptions and educational strategies in science and mathematics, 1,* 244-256. Itaca, NY: Cornell University.

Itano, J. K. (1991). *The organization of clinical knowledge in novice and expert nurses.* Unpublished doctoral dissertation, University of Hawaii at Manoa, HI.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32,* 241-254.

Kruskal, J. B. & Wish, M. (1978). *Multidimensional scaling*. New York: Sage.

Kruskal, J. B., Young, F. W., & Speery, J. B. (1973). *How to use KYST-2, a very flexible program to do multidimensional scaling and unfolding*. Murray Hill, NJ: Bell Laboratories.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science, 208,* 1335-1342.

Lawrence, J. A., (1988). Expertise on the bench: Modeling magistrates' judical decision-making. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 229-259). Hillsdale, NJ: Erlbaum.

Lay-Dopyera, M. & Beyerbach, B. (1983, April). *Concept mapping for individual assessment*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Lesgold, A. (1984). Acquiring expertise. In J. R. Anderson & S. M. Kosslyn (Eds.), *Tutorials in learning & memory: Essays in honor of Gordon Bower.* San Francisco: Freeman.

Lesgold, A., Glaser, R., Rubinson, H., Klopfer, D., Feltovich, P., & Wang. Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 311-342). Hillsdale, NJ: Erlbaum.

McKeithen, K.B., Reitman, J.S., Rueter, H.H., & Hirtle, S.C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology, 13,* 307-325.

Miller, G. A. (1956). The magic number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Nevah-Benjamin, M., McKeachie, W. J., Lin, Y. G., & Tucker, D. G. (1986). Inferring students' cognitive structures and their development using the "ordered tree technique". *Journal of Educational Psychology, 78,* 130-140.

Nevah-Benjamin, M., McKeachie, W. J., & Lin, Y. G. (1989). Use of the ordered-tree technique to assess students' initial knowledge and conceptual learning. *Teaching of Psychology, 16,* 182-187.

Phenix, P. H. (1964). *Realms of meaning.* New York: McGraw-Hill.

Phillips, D. C. (1983). On describing a student's cognitive structure. *Educational Psychologist, 18,* 59-74.

Posner, M. I. (1988). What is it to be an expert? In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. xxix-xxvi). Hillsdale, NJ: Erlbaum.

Reitman, J. S. (1976). Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology, 8,* 336-356.

Reitman, J. S., & Rueter, H. H. (1980). Organization revealed by recall orders and confirmed by pauses. *Cognitive Psychology, 12,* 544-581.

Reitman, W. R. (1965). *Cognition and thought.* New York: Wiley.

Romney, A. M., Brewer, D. D., Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science, 4,* 28-34.

Rumelhart, D. E. & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquistion of knowledge.* (pp. 99-136). Hillsdale, NJ: Erlbaum.

Shavelson, R. J. (1972). Some aspects of the correspondence between content structure and cognitive structure in physics instruction. *Journal of Educational Psychology, 63,* 225-234.

Shavelson, R. J. (1974). Methods for examining representations of a subject-matter structure in a student's memory. *Journal of Research in Science Teaching, 11,* 231-249.

Shavelson, R. J. (1983). On quagmires, philosophical and otherwise: A reply to Phillips. *Educational Psychologist, 18,* 81-87.

Shavelson, R. J. (1985, April). *The measurement of cognitive structure.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Shavelson, R. J. & Geeslin, W. E. (1975). A method for examining the subject-matter structure in instructional materials. *Journal of Structural Learning, 4,* 199-218.

Shavelson, R. J. & Stanton, G. C. (1975). Construct validation: Methodology and application to three measures of cognitive structure. *Journal of Educational Measurement, 12,* 67-85.

Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence, 4,* 181-201.

Streveler, R. A. & Bail, F. T. (1992). Free clustering of terms in an educational psychology course: A comparison of pre-test and post-test clusters. *Proceedings of the 14th Annual Conference of the Hawaii Educational Research Association, Winter 1992,* 47-52.

Stuart, H. (1985). Should concept maps be numerically scored? *European Journal of Science Education, 7,* 73-81.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* (pp. 382-403). New York: Academic Press.

Voss, J. F., Blais, J., Means, M. L., Greene, T. R., & Ahwesh, E. (1989). Informal reasoning and subject matter knowledge in the solving of economics problems by naive and novice individuals. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Esssays in honor of Robert Glaser,* (pp. 217-249). Hillsdale, NJ: Erbaum.

Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem-solving skill in the social sciences. In G. H. Bower, (Ed.), *The psychology of learning and motivation: Advances in research theory* (Vol. 17, 165-213). New York: Academic Press.

Voss, J. F., Tyler, S. W., & Yengo, L. A. (1983). Individual differences in the solving of social science problems. In R. F. Dillon & R. R. Schmeck (Eds.), *Individual differences in cognition.* (pp. 205-232). New York: Academic Press.

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

Young, F. W., & Lewyckyj, R. (1979). *The ALSCAL Procedure*. SAS Technical Report S-113. Raleigh, NC: SAS Institute.